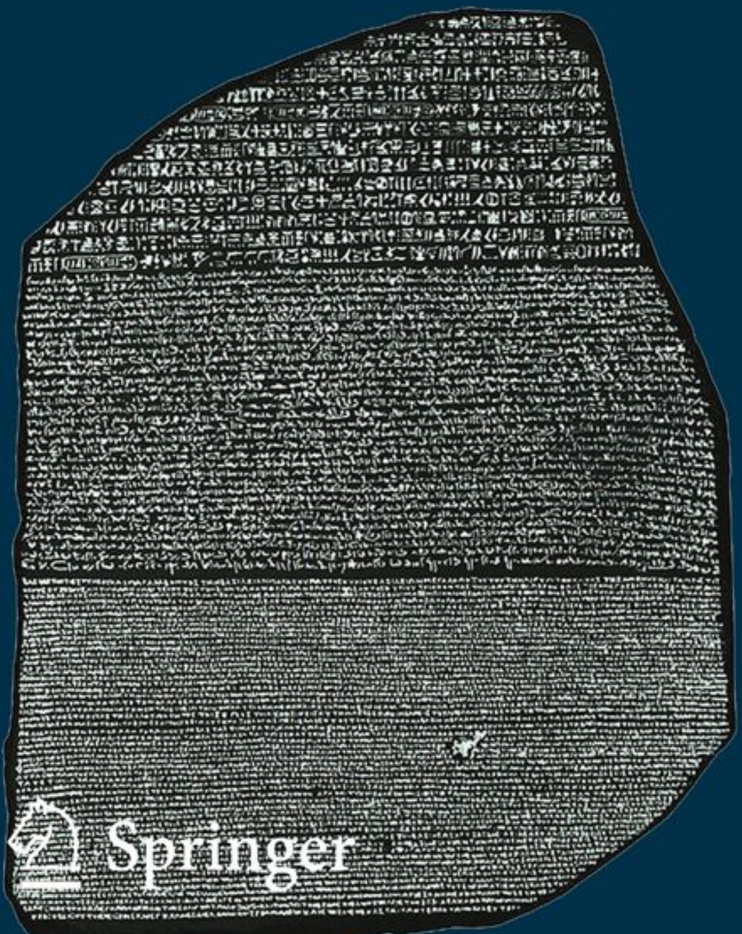
Festschrift

LNCS 5070

Małgorzata Marciniak
Agnieszka Mykowiecka (Eds.)

Aspects of Natural Language Processing

Essays Dedicated to Leonard Bolc on the Occasion of His 75th Birthday



Springer

Lecture Notes in Computer Science 5070

Małgorzata Marciniak
Agnieszka Mykowiecka (Eds.)

Aspects of Natural Language Processing

Essays Dedicated to Leonard Bolc on the Occasion of His 75th Birthday

Springer

Volume Editors

Małgorzata Marciniak
Agnieszka Mykowiecka
Institute of Computer Science
Polish Academy of Sciences
ul. J.K. Ordona 21, 01-237 Warsaw, Poland
E-mail: {Malgorzata.Marciniak, Agnieszka.Mykowiecka}@ipipan.waw.pl

The cover illustration, showing the Rosetta Stone, was retrieved from
`http://de.wikipedia.org/wiki/Datei:Rosetta_Stone_BW.jpeg`

Library of Congress Control Number: 2009935055

CR Subject Classification (1998): I.2, H.3-5, J.5, F.4

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN	0302-9743
ISBN-10	3-642-04734-3 Springer Berlin Heidelberg New York
ISBN-13	978-3-642-04734-3 Springer Berlin Heidelberg New York

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Markus Richter, Heidelberg
Printed on acid-free paper SPIN: 12766055 06/3180 5 4 3 2 1 0

e r B c

r f

s s d d t d t ss d l t s s 75t
t d d t s ss s w tt s ds st d ts d l
l s t l t s s t s d l s l d
t t l t l s t s t w s s ll
w s l s t d l t t l t ll s
s z t l d d t l ts

s l d d t s l s t s t s w
d l d s ll s st t d d s l s t
l t t s t l l ss d t
l t s t l l l s

t t s d t d t l t d w w s t st
l d d l s l st t s s t d s s w
t t d ts d l w ld st ts
s d dd ss s t l l t t t t zz
l s st t s w t t l l w ds

t t s s s d ff t s ts t t l l s
t s s t s w t s d t d t l w t d s
l l s ts t l s d Uz l s d t t d
d s ss s l l t s l s lt w d s s
s t s ts w d t l s l s s d d w t
t t s s l s w w s st d s d t
s l t s d t d ss l ll w s
d t d t t t s t l st wl d t t s t
l t s l d t s d s t st t s ts

t s s s d s d ff t l t s w t l
l ss t t s l s t t l st
s d t d t s t l d d s s t ss s t l
l t t S l s s w s s t t st t ss l
t d 7 s w s t d t ll w t tw
s d t d t d ff t s ts t t t st
s ts l t t t t w s t t d s s t t
d d l t t t l t s t d l d
t t s dd ss d t ds s t d t
tt s t ts d d s t t t s t s st t t s
l d ll d st ts d t d l l st s ts
t t t ll ws ss l l sts t w t t
l l s d t s lts t d t t w s

t s l w w ld l t ss t t d t t
ss d l t s s s d dl l d
s l t s l w d t t t t
t t w t w d l d t d s t d ff t d

w t s t t w ld l t t ll t t s
t s t s s t t d t t l d t l
w t s w ld ls l t t z ws l
d t l
w ld l t ss t s t s t t ss
l s l ll s t st
t st t t t S w z d d l d s
s st d ts w w s d t d t t l t s t
s d ls l t s w d d d t w d st ll
ll s t l w St s st sz ws t z s
s z s s t S w d
Ł sz ws w w d t t s l 25 s
t t s t s l d d t s l w s d l ll
t st

st 2 z t
sz w

i i r r i f r f r r

d l st t d s s t st d t t l l l t
t t d w z U s t z t 7 d
l w d t t s w U s t U t st t t t S
t s t ss st t t ss t ss d s
l d l st s t d w z U s t t s
s std t l st d s t t st t t t t s t
U s t ü st s t s t d s wl d
t t s d t s d t s s ss
s w s t w s t d s d l
s st t ü st d l tt d d t l t s s w
s d s t st t t l l st s s ll l s
d t l l s s d s l t t t t d
w z U s t s t t l s t l
l t t l s s t w s s t sd w d t l
ss s s ls t d t s w w s t d
t d t t t l s s d S t s s t st t t
t S U w w d t l t ll s
s s ts t l l ss 2 7 t 2 7 w d
t t st t t t S l s d S s w
z d t t t t l t d ts
t t st s

ss l s d s s l d t t l l st s
w t t t t s st w t t l l s t
t s l t s ss t s l s l s
lds s t l s l wl d s t t d
s ss d l s l l d s t d s d
s st d ts d ll s t w t s w d t lds
l s t t t t t ts s l d t
s ts 7 t t l 5 l ss l l t d
l t wl d ss d t t d d t

w d d l s s w s t s lt s t
t t l t t t ll s ts t l l t
s d st d s t t s t d t l s s tt s
d t t d d t s d s w l t s s
t t t ss s d t t l l

ss d l w s t t d ss t
t t s t s lts s t d t l d ls
l s d d ss S l l w d l s
d l w ll S t l t s t t s
w ts s lds t s d t d
s d t s s S l t t l s d

S l s d t d t l s w ld s l s s s d t s t t d t d l s ls w ll w d t l d t 2 l s l s d s s s d t d t s s t s l t s l s s t sts d t l t t w t d ss s t l st w s s t d t d d s t l s s st t t t l ss s l s t st t t s w U s t t t t st t t t S S w s s t t t st d ts d s t sts d l t d t t l t s s s ts t l t ll ll s t s l t d s t sts w w d w t ss l t s st s t s t s t t s w d t t d

c u ic i ns

o c S r a ow ran or a on Na ra ang ag In o og ca or a ING

o c c R a a r war an a na ra n go N ar awa

o c oc n w S r a ow Na ra ang ag In or a on R r a S D a og 0

o c R c o o a D a R a on ng n a ca D agno c S w Na ra ang ag In r ac roc o I a

o c owa o ow a S r a ow Na ra ang ag In or a on R r a S w x n on owar R a on ng In rna ona Jo rna o an ac n S

o c oro w c c o aw r war an a n or ac n wn n n N ar awa

o c ow J S arc o or r c a In g nc ca c r on on

o c ow c a o aw r war an a a na ra n go ran o or a n go a a n a c a c na awn c a

o c ar a J rowa n o c n a a n N ar awa

0 o c orow an a og c o or ca o n a on S r ng r rg 000

o c orow an a og c o a R a on ng an rac ca ca on S r ng r rg 00

ri r

t w s t l s S t t s l t l s
t d d t d t ss d l t s s 75t t d
t d l t t s l d s ss w t t s s
d t s t t t s t l t ll s s t w s s
t t l s t s s t t S sts t s l
t d s t ts d ts ss l d
s s t w s l st l t d d
ss w ll t s t t t t s sts t w d
tst d s t st d z t t s t
l d s w ll s t t ll ss l w s ls l t
d w s s ss ll l d d t s l s d t s s
d w s ls St t t 7 s s st t l t
ss l s d t l t t s s lt d l t s S l
d w d t s t t t s t d
St t t 7 w t s l d t S *u l* -
u o u o o u ss l d l d l s t ts
d ll t t w t s l t s st ff t d l
s d t s d t l t S l S s t
w t s s *A l ll d bol o u o* t
t d ss l w s s ss l t l t
s d d t d ll t s t s t l t s t l t ll
ls s w tw s l d s t d t
w l d l t s t l s t ll s
d t l t t s ss l t tl d s d t l d
s t sts d t w ld t w t d w t s l d t d s lts
l t s d l t s S d S ss
l t t S t w l st d l l t s l d
t sts ll *Genetic Algorithms + Data Structures = Evolution Programs*
w l w z

s d t S d l t s d t l t t s
l t l s t t d t w t ss l t t
s t t s l s l t s d d d s
t s sl s s s s t wl d d t t s
d t l ds s ff d t s s t d t
s ss l ts

l S t l l t s t d
s t t s d t l t w ld l t ss
s t t d t ss l t ll t t s ff d
d d s d w w w ll ff s w s
ss l ll t st t s t

st 2 l d

f

ic

n u

App ic i ns

I a o on n

i y
r r r

d z ws [1] d d z S w [2]

[1] In o D c on roc S or
an
ga So on S o o
wo owa o 0 ar aw o an
e g p
[2] In o a a c
ar aw n r
anac a 0 0 ar aw o an
w i w e p

A st t c o n a on or o rn n g n n
ra wor o o c no og c a roac a on
ro g gran ar a roac

o ds w o c no og g n n rac on ro g
gran ar co ng ro g gran ar co ng

n e e a e e cee ng a e ca ga e
e ee a c ca e ne a e e a ng an
c c ng e e e a a e e e
a e e a

can n n be e e e n c a ac e ze e e
a e an a y e e b ance e a
e e b ance be een e be a y b ea e c
eye ga e e a en e c e c e a an c c n e
a e ay n a ay ga e a a y

ang age an n en c nce a e n en
e a ne n a can a e n g ea e ence c
c nce e e y a e a b e y c n
ee an nc e a e a n e e an cen e e e
e ence e e y ne c n en ence e en n e
n ance ca c a n n e y e ea e en e an
e e an b e an b e g e e

nce ea a e n e ga n a e e
e e n n e e an ec n e e
w g g n n

n uc i n ni i n f u c n s p nc

d t s t t w tl w t ss t s t
d l t l l s s d l s d d t
t d t t l t s t st ss d s
t d l w sd t s w ll t s t l *do*
olo st s t t *do* l s *l*
ud 2 s t d *do* w d st d
d t l t t d ts tl t s t s t d
t l t d s s d l l st ts t t
t w sd d st d t s w ss d t s ll d
do *u* *o* 2 t ll s s w

$$wisdom = knowledge + adaptive\ judgment + interactions.$$

sd t t d s s l t wl d ss d t
l t s t t s t wl d ss l t s ss t t
t l s st t *Ag* s sts s t t t l
ts t t w t t t t *Ag* ts d w t t t
Ag ts t s l t ts ll d s t
t d t t l ts w s s l t d
t t l ts t d t t l t l
s s sw s t ll w st *o* *do*
s l *of* *Ag* *o* *ol* *o* *o* *ould* *b* *d* *o*
s t t l t s t ss s d wl d
s t t t t t t t t t *Ag* ts l s
t t s d d t t l ts s
ss s l t wl d t t t t t
ds t s l s t d t t t t
t st ts s d wl d ss l d ls w t l
ts s s s s t t t w t *l* t
ds Us ll t s d wl d ss s t ss l l l
s t s st t ts l l st t ts t t l s
s s U S s wl d ss s t l s t t
d t ss t t t d w ld
s d s t *fo* ts *fo* s d d
t t l d s w t s l t s s t st t t
t t s t t ts l l st t ts d t l s
ss t t t s l t ts ds t *Ag*
s l z d t t w t *Ag* *l* *f* l s ds
sd t wl d ss t t *Ag* s t z d t
l t t l t t d t ss s d t t *Ag*
s d t s z t t t s l t w s t t
d *ud* t w sd t st d st *ud* t
w s l t d t d t t s s w s l ll s
ud

d t t s ts ls l t t w t t s
25 5 5 7 d d t t ll ws t t t t l
ts t s s w t d t s d t t
Ag t d l ss t l t s l ss s w s ld d t
t ll w s d t d s t d wl d

t t s t st s t t t l t d t t
l l st ts t t t w st l s l s t t t
t t d ts s d s ts ts w t d
l s s d ls d t t t t ll t wl d l t d d
l l t s s s w d ts s ll t t s s
l t w ls l t t l t d ts s t d
l s s d t t w t t t S t s t
d s l ls d s s lt d d ts

s w sd s d tl s s l t s ts tt t
s st tl t t d l s d t s t
s l t w t t t s t t d t s s
s s s d t sl w ds s d t
t t t s w t t ts d ts
t l t t s w sd t t d s t t l t t st
l l t t s t s d t s d s s d t
ss t t st tl t t d s
t ll w t l s l t t w sd t w sd s
ﬃ t d l t d t s t ss l
t t sw t ll w st s w t tl st t t st
t t t l st l s t s l d w t s l t t t
l s d l l st ts t t d t

t s d l t t d l t
st l t s l s d l s t s t t t l d ls l
t t t *d* *ud* s d ﬃ lt l s sts
t l t t t t l l ss l t d ss
s d l t l s d ts ss t
t d ts t w ds w l l t d l t st d
s l l s t d ls d t t s tw t t l
l t s s ll w t l t t l l d ls l z d
w w l t l l d ls t t
t s l t s t s l t

sd t s l d t s ts t s
ts t ts l l d ll s ll
d st t d s st s w t s ts t S s st s s st
t t s ts t l d t l ts d
t t w t t

st s s d t s s t wl d t
d d t w l s l t t wl d l s l l l
l d t l ts d t s lt t s st s s
l d s s t ds s d t t t t s t t t s s

t t s d l s t t l t ts wl d
wl d tw s d t d t l t t s l s t
t s t t d t t t l t
2 27 2 sd l st s d s ss d s t
d s d l t w l t s t ll t s st s
t s s z t ts s d st

ech **y** **ech** s ll t t s d t t
t d t t l t s t st ss d s
t d l *do* t d s d l t t t ll t
s st s s t s l d t s ts t s
ts t ts l l d ll s ll
d st t d s st s w t s ts t S s st s s st
t t s ts t l d t l ts d
t t w t t st t t d s t s ss d t s
t l t t l d wl d t t l s
st s t t t t l s s t d d ff s
t t st t t t s t d s d l t
t t l d ls st t ll t s st s

e **e** **ech** **y** st s s d t s s t
wl d t d d t w l s l t t wl d
l s l s l d s t t l t ts wl
d wl d tw s d s s s l

e **ech** **y** st s t s t t t
t d d t d t st t s s d s lts t t l
l s s ts d t s t t s t t s d
l d t s l d t l l t t t sl
t ds ls d t ls l d s t s s d
t l s s d ts l d d t
t s l ts d ff t l ls t
s l l l s l l d t d l t t s d ll
ts w t t d t
ts st s t d s ffi t wl d s s

I **er** **c** **ech** **y** l d s t s d t
t s ts d ts s l d t ll
t s s lt t t d t d t t t l s
w s t ld st t t l d ls t s
t s s l t

tt d l l z s ld s d d s d
l t d st d s l l s t ts

t s s w t s t w ld d d s t t
tw tw l s s t tw tw t ts t w ld
s ffi t t t ls t ds d s t t t s
l l t

o f d l l b 6

s t st t d d t t s
l s s t s t w ds w ld t ll s t t
ls t t t s t d st d
o f d l l b

t s s z l t d t s t s t t s
d l t ls t d l w t t s ss U t t l t t ls
d l d t t s t l l ss l t t l l
d t t ll w t d st d t l l s d s t
ss t ts d s t t s s t ts d st d w
w ll ll w s t st t t l t ll t s st s
t s s s t t s l ffi tl s l l
l l s l t ss t t ts s t
s ts d s w t s ts t d s d t s
t l l s t ts s zz l t t w t s
t t w t w ds l t s s ts t l t s
t s d t t l t s d s d l
d t d s l s s 5
d t s s ls s s t t ds s 5 2
l l l t w s s st d
s ts w t t s d
l s t d s d t t *fo o ul* s *lu of ob*
of o o d o o b of d u b l l o
fu o l d o b l ou o o l u-
b of l l d o o fo lu ob o
lu d u o u d d u b l l
d fu o l
S l l l s lt t s st s s w
t s t ds s d w t d s s tl
t d s l t s
7 7 7 2 27 2 75
s t t st t d t t t s
ts ll d l s t s s l t s l s w
s d s ts t t s d l s
s t t t ll s d d d ts d l d
ts d t s s s s wl d s t t t
w t t ts d s l s t d t t ss
l t l t w d s w d t
ss d t l s l l t t d d d
st t l s l s s t t ds
w t t s t t t ds d t ds s d l t
s s d d l t ss l st d t s s d
t s s ts
s t d t s s s d t s t
s t t d t wl 5 5 s s d l ss l tw l
d l s t s d l d t d l w t t t

d ss s t ss l t s s l t t
t s ts s t s t s
t d d t l w t s ts 7 7 7

u C nc p s n App i s nin A u u C nc p s

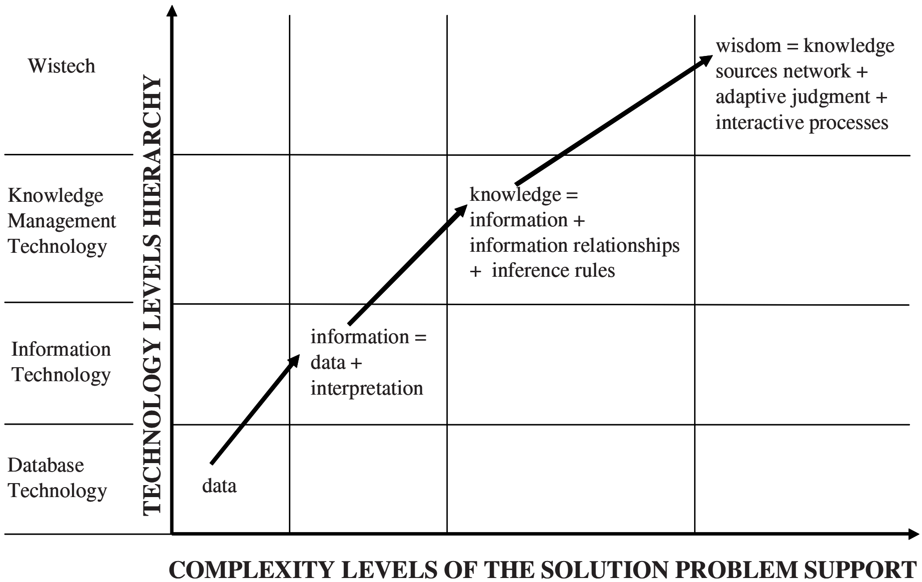
t ss s d s ss d t w t t t t s
ss ss s s ll ss t d w t t st d l ts
w t l l ss d l t t t ts l t
d l s s ts t t d s d s *u*
o l s s d s d t s s t l t *o* -
do 2 l st ll ts w s t l l *ll*
ou *f* S ts t s d t s s t s d
wl d d s t t ll t s st s ts l
d *ou* *u* *o* *o* *o* *d* *f* *ud* *l* *k*
t s t ts s t t
ts d t t s t s s d
t t s s t s t l
ts s t ls t d s t ll t s st s
t s ts st t d t t s l l d ls s d
d ff t s t ws S l l d ls l d t t d t
l s s s t lt s l t wl d t ts
d ff t s s d d ff t l z t st t s d l l d ls
t s s l s t d t l s s t t
t st t s

t t s s t t ll t t l t s l d s ts st
t t w s s s w ld ss l w l s s
s 2 d tl t s t sts s s s w ll
s t s s t st d *u* s ts
l ss l s t t s t s l d t d ts l ts t
w ds t s s t t l t st l l ss d s l
t t s t t t s t s t t s t s s
l t s t dd s s s s s t
dd t st t dd s t t t l t s
s w l t l ss l ll t s t tw l ss s
t l d t t l S t s t l ss d w t t
t l t d t s t t d t l s *b* *u*
s t s t t l st ll ts t l l
s s s s d t l l st
s d ts d t l ss l l
d ss t d t t t l s
l d s w st l t d t s ll d *o*
d *f* *l* *k* *o* ld d s s 2 ld d s
s ll ws s s s s s d

s d s l t ld t t s st w t
t l s t t w t s t ld S l s
l d t st s
ss s s ll ss t d w t t d
st ts w t l l ss d l t t s t ts
l t w w s st l t d t t d l
l ttl 25 s d t t
t st s d t w t t s d
t w ld s d t t w ld t s d l ll
d t s t t t t s st s s s t ts
t w s t w ld ss l t s s l
w ld l t t t t d d st d t t
s t s t l st l s d t t l d t
ll w ss s 2

s d l s s
2 d s ts t s
ts s s t l t s t s d s

t s s ll ss d t t t d st d t
ts t s t s s d t d t s t s l t l t d ds
t t s wl d w s t t t
ts t s ld ls s d d w t t t s
s w s t t d t t 2 t t t s t t
w s l z d l t ls d d t d l d t st t d l t
t l l s l ts s t l s t t
s ts t w s s w t t s t t ls t ss l t d l w t
ts s 7 7
l ts t l t d t t t
d d st t l ls t s ts S s
st w s ts ts t ts s
t t t s t t st st t t l t *b* -
of t s d ff t t t l d t t s ls
w s *o o olo* w s s d t l t *b* *l* *-of*
S lt t t l t t s w s s d St s w ws
2 d t s t t t s w t *ou*
olo s 7 5 t t s t t *ou* *lu o*
l o l s t l l t d s s t w t d s ts ts
t ts l l s s d t l t t
t t d
t l t l l s t t l tt t t st l s
l s tw ts d s ts t s l s s d
t t ts s l l l s t s s ld
l s ts s l l l l s Us ll t s
d l s s d l l d t l t s s tw
ts ss d s l t s s l l l s t d t
l s 27



F o q a on con x

Wis c n y Wis c

t d t d ss d t w sd t d s
st t d t w l s t s t s
d l l d t s t s t s d s st l
t d w t s l t s t t t t
d t s ld s z d t t t d t t s
ts t t wl d tw d t t ss ss t s l t
t t l d l s l s s
ld s t st ts l l s s d t d
s s t ddl t 2 t t d s
s w s t s l t d t s d l t 2 t t d t
t t s t s 7
s t t s t s t l s d s l
l ts s z d t t s t t w sd t
s t ls ll st t d s t ll w d s t d
t s t t *d* s d st d s st s
ls w t t t t t t t s t t
t ss d t w sd t st d s
t t t t t l s s d t t ts t
w sd t t t t st tw t *k o l d*
s t d t d t l d s *k o l d* *ou* *o k*
d *o* s t d t d t l d s
o s t t ll l t t t d l

k o l d ou o k wl d w t d t ll d st d
z d s t t l w t t l s t s t t
s l t ll w l s ll st t t t
wl d s s tw

s t t st t s l t d s s s s d
t *o* t s t t d s w l
ds tw s s wl d d t st d s t
dd t l

tw wl d l ls s t d ts s lt t
s st d t l l wl d t t t st d
s t s

2 *o* t t d st d s s st l d
t s t l s

t d l tw l
s t s d t s tw d t d
t t w t t t s l
s ts d s lt l

t s t t t w l s s ts t t s t l t t ds
d l t t t t s ss d l st d
l t st t s s 25 5 5 7
t d t d t d st s w sd t l
t l s l t ll s s t d ffi lt t
t l t s t ss d s ll ws

d ud d st d s s s t l
t s w t s s s l t l l s s wl
d d t s s d st d st t t ss s
d t t st t s l s t t ll w t t s d
l l st ts

- d t t d d t t t d t
  l l s t t s d t
- l t t s t s t t t l t
  s s d st d t st t t ss s d t
  t st t s t w d t l d ts
- s l t ts d d wl d s
  t d d t st t s s t s t d s t t
  l d t d s s d st d s t
  t w t t t s s l t t t s t
  d ts
- d t t t s s t l d t t
  ss s
- l d t l l s s l d
  t d t t t d t d t ss

d ffi lt d l s t s t t t l d ls l t t
t *d* *ud* l s sts t l
t t t t l l ss l t d ss s d
l t ts ss t t d ts t w ds w
l l t d l t st d s l l s t d ls d
t t s tw t t l l t s s ll w t l
t l l d ls l z d s w
w t l l d ls t t t s l t
s t s l t t t t l t t d ls
d *ud* t t *ou* - *ul* *o*
t l l t t w sd s s d s s t t ts
t l t s tl s t ll ss t d w t s ts s
s d d t l s l s s s s
l d t s s s t s s t t l d t
s d st d s l ld t w d w t t d ff
s tw t w d l w d l d t *f* d t
t *do* d d d t d t s t s d
st d t s w t l t s d ff tw t d st d
w sd d s ts s s d d t d t s
ll l t s st t w t l t w w d st d t d ff
tw l s l d w sd
tt t t l t t w sd d s t
t *obl* *ol* t ll w w sd s t l t t d t
t t l s s s ffi tl t s l t s t t
d l l l l wl d s s s l st ts t
d t s d st d w sd t w t t st t t
d ff l t l s l ss w d t t ll w
t t w sd t *d* *f* *o* *obl* *d* *obl* *olu* *o*
o
t l t s s t t l d ff ll st t t
l d ff tw t t l s l d w sd s t
d ff tw t t t ll t ll d t
s t s l t d l s l d s d t t
7 7 l l l d s t w sd
t l d
t t t t w sd s s l t *b l* *of* *obl*
ol *ul* *do* *of* *l* *o* w t t t t w ld
t s s tl d st d s *f*
l t d d t t t d
l d p .w p d . r w r
t ds s ll ws

s t t t t s t d sw s
wl d s t s t *b* t t t s st s s t s
t t t t wl d s t lt t s
l t w l s s

s t l ts

t t t t t s t s d t s l
t s d s l s
2 s d l s d l t s t l t d st t
t ff ts l l s l t t t s t
t t d
s st s st tt ts t t
s st t s t t t s l t

t w ds t t l t s t t st t s
t l t t ll d s s wl d w
t t t t S t s t s t ll l t d t t ll w
tw l ts t w sd t

wl d s s tw
2 t t ss s

w t s ld d t t w sd d st d s t l
s l t *f* s t st w sd d st
s t s t t l d st d s *d*
b l o k o ud *d l-l f o* s
t t st d ff t t w sd t l d st d
s ts s l s l s s d t d t
l t w t w d d t s *obl ol* w w tt t
t l t t w sd s t t *f* t
ll w sd s t t w t t l l
t w s l t d t t t l s d t d s ffi tl
t s l t s t d l l st ts l l wl
d s s d s l t s s ls ll st t t
d ff tw t t d t t w sd
s t t l
ld s t st w s t l t w sd
l s l sd s t t dd t l st t
d t d t w t sl ss sts s d
t t st t t l t s l s t st ts d
w t t l s tw t l t t s l t d t s
s l t tt s l t s t st w s t
l t l s l s l s t st t t s t
s t l t t t s t t ld l
s l t t t t s ld w sd d st d s
l s l d t d t t s s t t
l st t d t t ss l t s t s l sl t w sd
t s t ss t l d s d t d t d t d t
t l d d ts l t t l s l t t d s
t s l d s l s t t d s t s lds t s
s s t t t s ld s l t t t
ld

d d d ts t t t l s s d t s
t st t t t l s t 2 t t d t
s d l d st t l t ss t t l s
l z t st s t d l t t s t
l l d st s t ll w t st s t d l t t 2 t
t

t s l t d ss t s t d t
2 t l d st d s dd t t d t s t l
t l t t t t l s s ss d s l z t
t
wl d t l d st d s s st s s t
z t l d t s ts d t t t s t wl d
ss d d s s

t st s d l t t t l s w s t
t ds t t d l t l t s t s t l s s
t ds s l d s t t d t l t
t ss S s d t d t l t
d l t ll s z

o D *o* l l t ss l t s t
t s s
Tu *D* *o* t l t s t t s l
d st d d sw tl t t s s
t s t t t t s t t s
w ll w s t st s

t s w t d l t t ds t l t t
s t l ll st t d 2
d t l t t w ll s d
l l t ll w w l t t t l
z w sd t l st w t s l s t d
t l s t s t t l s t d 2
t w ds t t ds t d l t t t l t
s s t d s t s ll d t
t wl d sd t t l s l l t
dds t tt t s d t s
s s t d ll
t d t t w ll w S l t
T *Ro k* w tt 2 w t

s t l w l st l
s t w sd w l st wl d
s t wl d w l st t

Technology	Additional attributes	Shannon Dimensions	Turing Dimensions
Database Technology	<u>data</u> is the most basic level	*How to represent information?*	*SQL*
Information Technology	<u>information</u> = data + interpretation	*Where to find information?*	*Who? What? When? Where? How much?*
Knowledge Management Technology	<u>knowledge</u> = information + information relationships + inference rules	*How to use information?*	*How? Why? What if?*

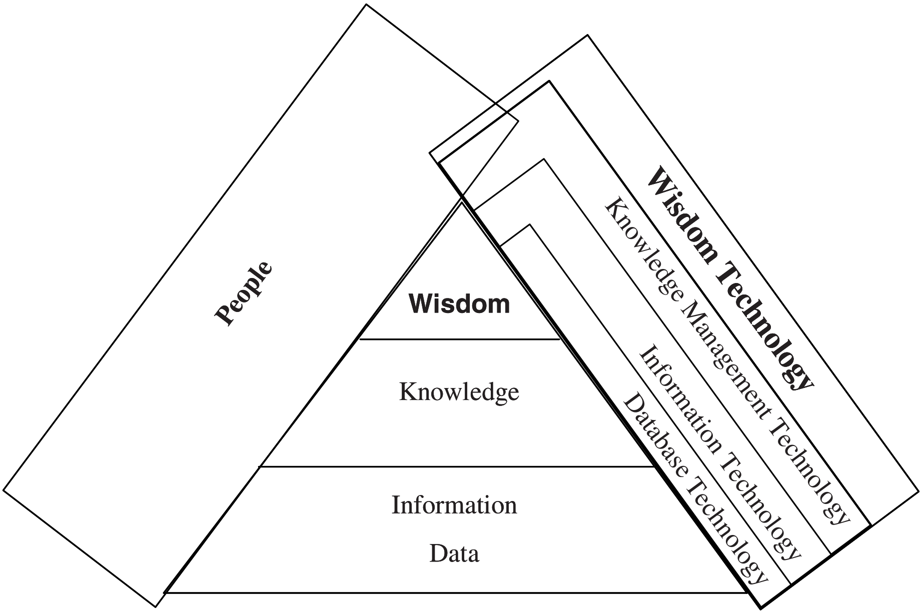
F 2 o ng ac n c no og

Technology	Addional attributes	Shannon Dimensions	Turing Dimensions
Wisdom Technology (Wistech)	Wisdom equation, i.e. <u>wisdom</u> = knowledge sources network + adaptive judgment + interactive processes	*Learn when to use information* *Learn how to get important information*	*How to make correct judgments (in. particular correct decisions) keeping in mind real life constrains?*

F o ng ac n c no og con n

t s t s t st t t t t ff ts t t d d t d s d t w sd t d s s t t t st t d t l t t t d l t t t t t st s t t t s t t ss s l d t w s t s s t t s ll s s t t s t t d d t s d l t t

t s t t l l s w t t l st d l t t t ll w t z t d t t t t ss ts s t t ts s ld t s s s t

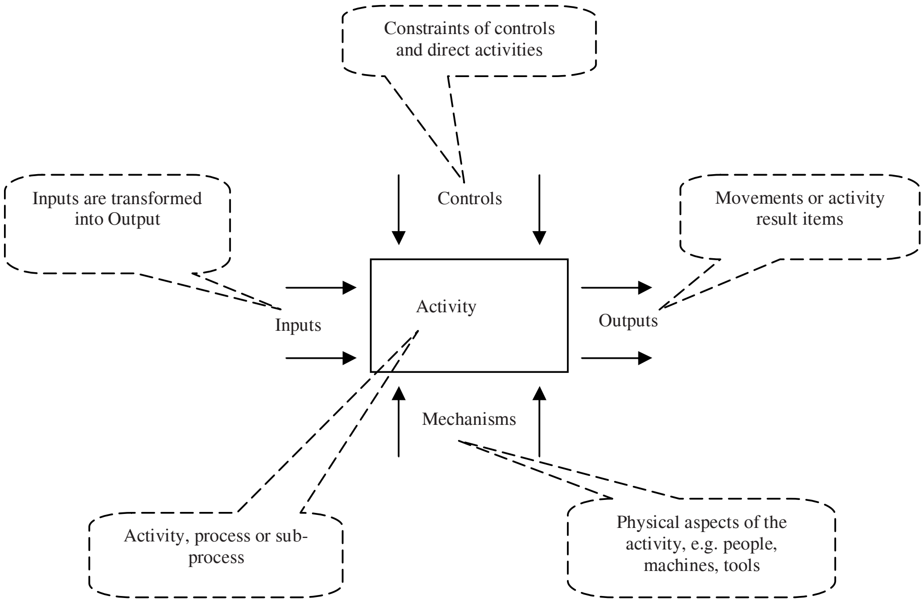


F DI rarc

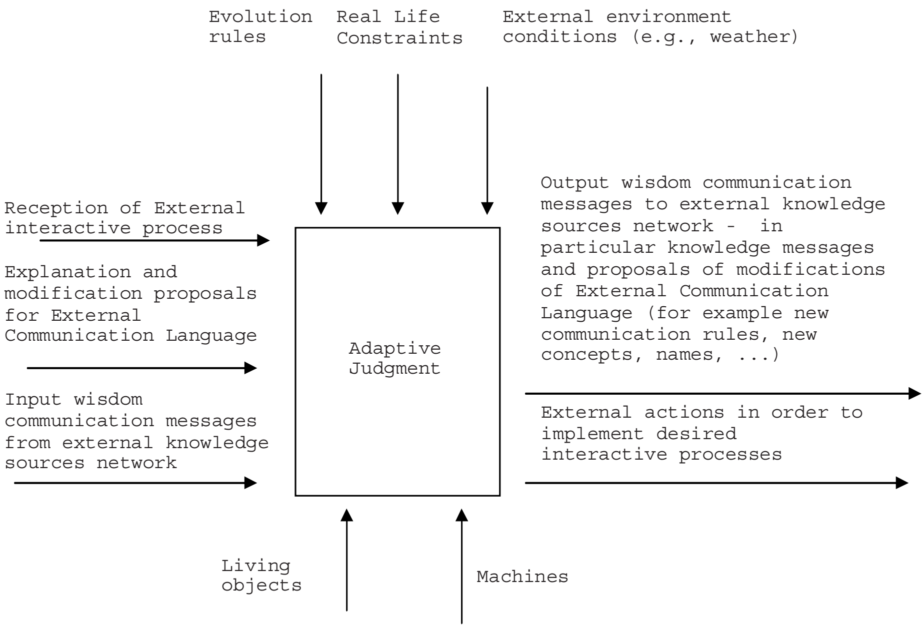
ls l s t d t s s t d t w ld d d t d t d l w t t s t s t t t t ld d

d t t l st t t l d ls l t s st t w t l s s t t d t d t tt l z t w t d t d t w s ll s t s l z t ss s s d t st d d s s s l s t s d s d t d s t d 5 t s t t t d t s ts l t t t t t l t t d t t t l t s s l st t t s t ts t l w s ll ll t t t d t t *ud* s ll ls ss t t t wl d s s tw s d d d t t l s s s s wl d t t ls l l t t d s d t l s s w l l l t t s d st st l l t d l s s t d s s ss l ls t d l l ts d t ls d d t ss d d s t l t t d t d t w t s d t ls s ld l d s l ts s

t *l o u o u* d st d s l s d ts s d t t d ss wl d w t tw t l s s wl d


Constraints of controls and direct activities
Controls
Inputs are transformed into Output
Movements or activity result items
Activity
Inputs
Outputs
Mechanisms
Activity, process or sub-process
Physical aspects of the activity, e.g. people, machines, tools

F 5 c


Evolution rules
Real Life Constraints
External environment conditions (e.g., weather)
Reception of External interactive process
Explanation and modification proposals for External Communication Language
Input wisdom communication messages from external knowledge sources network
Adaptive Judgment
Output wisdom communication messages to external knowledge sources network - in particular knowledge messages and proposals of modifications of External Communication Language (for example new communication rules, new concepts, names, ...)
External actions in order to implement desired interactive processes
Living objects
Machines

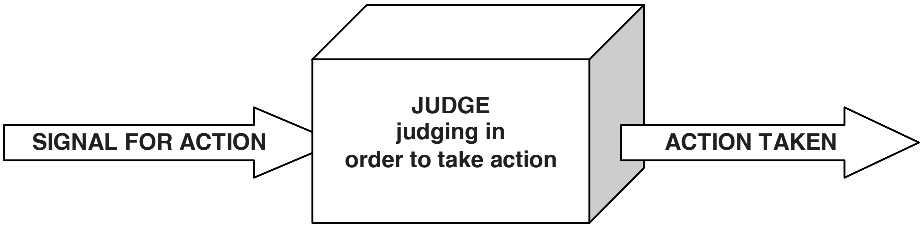
F r o o

2 t *l o u o u* d st d s
t l s s d ts s d t ss d t
t l t d l s d ts s d
t t d ss wl d w t tw t l s s
wl d
s ls s l t s d t t ss s
d ss t s t *l o u o u*
d t *l o u o u*
l t t t s t l t s l l t d t t
ss w sd t s s ss ss t l t t t
l s t ll t d t d t ss
5 S l t ts d d wl d s t d
d t st t s s ffi t t t d s t t l d t d
s s d st d s t t w t t t
s l t t s t s l t l
t tt w sd t ss s t t l wl d s s
tw t l wl d ss s d s ls d t s
t *l o u o u* w t l s
w ts s
7 t l t s d t l t t d s d t t ss s

ll l ts t l st l d t t t
t ll w tw l s t l l t t d t d t
t t l d ls

o l d o t s s t l d s t
t l d ls t l t t l ts t t t t
s t t ss d t w sd d t l
t s l t s t t l ts t l t
t t t t l t l l t s l s
d l l ss l t d ss zz d l t ts
2 *ud of b o ol* t s s t l ld
t t l d ls t l t t t t
t ls d t ll t d t ss d t w

w w tl t d w s l t t t l
l t d s t t ls t s w t t
t t *b* s l t d t l t l
tt t s t t t l t s ld s d d
s t t d s wl d s l t ss t
t l tt l t t t w ds t w
l l t d l tt t t s l ttl s ff t
wl d s l l tt ld t t d s
s ts d t t t s w s wl d t s l
s t t s s ll t t t t s l d t w ts
ss t t s t ss s t ll d s ll d *b o ol* w



F J g agra

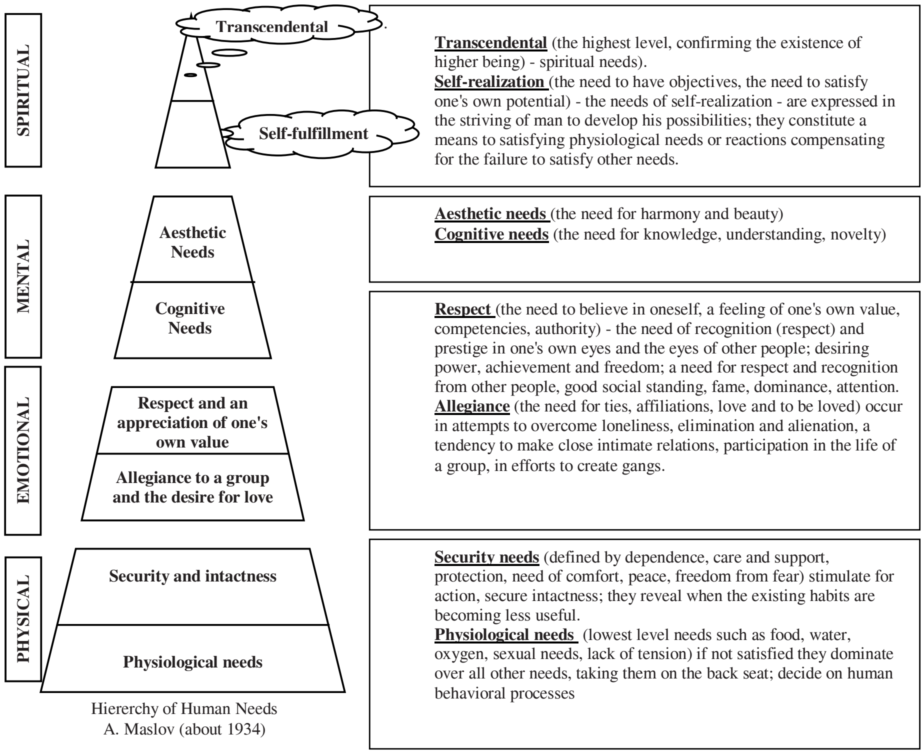
s t t ll w s ts d t d t ss s t t
s d d *ud*

o uou b o o t s d t l s t t t
d t t ts s s t z t t t w t
ll w t t
d t d s s t s d s d d
s t t
wl d l l t d
t l l s d s t
2 *K o l d o o* s s d w d t d t t t s d
s d d s t t t w t t z t t
s l l wl d w ld s d d t t t
st t t st t w t t s d t s d d
s t t
b o ol t s t t d t d
t ss d t st t w ts d t t ls

t ts st s t d t t
t l ss t l t l d s d l s l l t d w t t
wl d ld d t t s w t t t d ls
s t d s t d t s t s t ds t t s w t t
t t s l t d s d s ds t s ls d
t t d s t d 7
t t s t l t d t ss s ls d t
t t
d t t ll w tw t t d t ss s t
t

d o of o w t d *l* w t t *d d*
ob s s t st t t *d d ob* t t
2 *d o of l o k l ud* s w s
t st l z t *d d ob* s d t s s l l
t t

a ar n r oo a n ar an r a a ar o a ora a rn



F a o rarc o an n a o a an xa o g rarc o a con ro

d s t d t t s t s s t t l d ls d l t t s s t l t d t t ls d ts t ls st t t s t l st l t ss s t t d ls st t t s t ss s d s t t l d ls t s d t l t t t s t l t s d s s d t t ll d st d l wt d s t t t d s s t t t s t l t ss l t d t *lo o d d ul ou l* s l s d t l d st d sl w s ds t l l w t tt t l t t st s t t t sl w s ds t t st tl d st d t ts s s s t d t ld s d t d t st t t t l d ls d s t t ls

Wis c n y u nu C pu in

l t s s d d l t ss l st d t s s d t s t ts t s l t s t w t ds d t s *d f o f o doo o o u o d o fo o o u b d o l k ol u o o d u d l k bl u b l o fo d of l d l k ou u d o u o d l of u k d l d u o*

s d t t z t t s s w w s t l s l t s s t s s st ts s st ts t s d t s t s ts d t d d s w st t t t st ts t l t s t l s 5 l s s st ts t l st ts d 5 5 7 w s t s d wl d d d t s ts wl d s s t d t l ts d t d d s tw t l s t st ts s d s l d t t l l

l s st t d s t l l s 7 7 7 75 l s ts st t d t t s t s l t t d t z t t s s l s l s d st s d ff t ds tt s d s l s s ts d s l s l st s l ss s l t s ts 7 7 75 t l t z t t s d t l l t l s d s s t l s l t s t s ss t st t d l s s s s t s s s l z t s t s l s l l d w t t s t t s t s w t z t l s l t t t d s t s z d t l t d t d s t tw s ts w t t z t s s d d t s l d s t t s s t ll ws t t t t w t t s t l t l s d t t t s ss t d s s l l t s s t l s s s t l s s s 7 7 l s ss d t ls s t s z s l s t z t st t s l d st t d t z d l s ls t t

l t ds t t d ts ss t s lt t s t ll s t s d t d ts ss d t l l d s ss t l t t s ts s s d t d d wl d s dd t l wl d s t d t l ts s s d t t s l t s t s d t tt t s l t t t t ts d ff t l ls t t d d t l t s l

ts t s d t d l t t d
ts s s d t d d wl d s ll st t d
l s l t t t t s ts t l t d
t l l d ts d s lls l tt
d t t l t l s s d t t s d t s d
s s t d s lts s t st t ll w l t
d

s l t s l l t l s l
t t l s t t s s ll t t
s w d t t d t s t l
t

st w t t l s d s d s s d t
s s t t st ll s t d s l ts
t l ts
t t l s t s l st t d
l tt t d d t 2 t s ss t
d t s ts d ls ts s st t l st
s st t w t s st ts t st t l s t st ts
d d t s l l ts w s d
st ts d d d t l s st ts d d
d s tw t t t t l ss d ts d ts s
l t t st l s t d t t ts
s d d t t w t t d ll t
ll t s s s s s t l d ls l t t t l l t
l s d t t t d t t t
t ts t d l s ld t d t t st t
ts l d t st t t t s w t d
s s t l d ls l t t ll d t
l s l t t l l l s d
t st t l ss ts d l ss s t s l s s
7 s t s st ts t t s s l t d
t l z t s s t ss l t st t l tt s
s l tt s l z t s s t d s s l t
st t s
s t t d t dz s w wl 5 s t t l
t t wl d l t wl d s t l d
l t l s s l s d t t s tl s ls
l ss t s t sts t l l t t l
t ll s t t l w t d st d
d l t t wl d st s ss l s d t
t zz s t t s d t d s t t s ts
t tt t t s l t s l t s s d ss t t t ts
d ts d t l t t t t t s s
ts d s l t s t t ll ws t t s s ts t
tl d s d l l t t ts t d t

s s t s t z d ts l w d t s
d ff tw t d l w t s t s ll d
ts d s t t ss s ss l d
s t t d s t s t t s t t t s t
s s t w s t s t s t t d
s t d t s t t wl d t t s t s s ffi t t d t s t
s l z t t s t t s s s l z t
t d s t d tl

t s s ts s l l l
t t t s t t s ts t d d t l t
s t t w st t t d t w ts l t
s ts s s t l s t ld t d t s
t s d l w t t wl d l t z
s t s l t l t s tw ts t d d t s l
l t s tw ts t l s l t w d s s t w t
d ts ts t ts l s s l l
l l s 7 5 s t s t ws l
d s s d t l t *o b* *o d*

s ss l t s t t ds w s l s
s d l s s s d t st t
l P d s d l t f_P w t t ll w
t t s l t s t l P d d d l ts
t l t f_P t s w t t t t t s l l l
l s t s ss t d l w t l t s l
l s

s ss l t d l s d t d s l t ts d l
s s d l d s t t t
st ts l d ts d t t s d s l s ss t l s
d s t z t l l tt t s s l l s w
t s d d l l s d s s tt t
t d t s w ll s t s l t t t st
t l s l t t t t s t t s l t
d w w s ss l d l ffi t st s
ld s t l s l t s t s l s s lts ts
d t s ts s s w d l t s l t s
t d t st s s w t t t ds t d l t
t w t s t t t l ss t l t s ts
t t ffi t t s t t t t t t t t d l s
t ss l t st t st s t t t
t l ss s t d st s l ts l s t
l ts d t s l t s t l

s t ff s t ls t s lt t
s st s S t l l s t s t ts
t t t t s s d d s t l
s t t t ts d t ts

st t s d t d l d d l d s t s t s
t t t d ts t t t t
ss s s d ff t s z d t ss t t w
w t ds t t s 5 s
t s s t t t s d ts s s t
s ts w s t s l t t s l s
s l s l s d t s t t s d ffi
lt t s s d t t d s l s w wl d
t t t s l d s s s d t l t d l
w t d ts s s t ff t s
t w ds d st d t l st t s l tw s l
s s ls t t s z t l s l
s t l t t t 5 s l s l s l
l t d t l s st d l s w ll s s st s t l
t t d ts l s t t s s t
ll s w d s s ld t t t t t d l l
t ls t s l l d ls t d l l ts w ld
l t s t lt t t l t t s ld s d t w ds
s ss l s t t s tw ts t s l t d t s
d ts l l d l s d d l s d
t t d l w t l l l w w
l t d l s t d t l l t l s l tt s
l d t t d ls w t ss ss t s s d t d l t
s lt s l t s t t t l t t t d w t ds
w s s ld t d l d s d l 2

t s s d s ll st t l s l d l t s
l w w ld l t s z t t t s tw d l s t
d s t s d l t t st tl t l d t
t t s tw t l t d t t 55 5

t $x \in U$ t s st $IS_1 = (U, A)$ s d
s ts s t $Inf_A(x) = \{(a, a(x)) : a \in A\}$ t st
l l w s d ts w t s t s s t d t t s st
$IS_1 = (U, A)$ ts w t t s s t d s l t t
l l d l w s d s ts s l t l st t s t
s t s ts t st l l l s t u
s d s l t l st t d d d s l t l t
tw s t s ts t st l l tt t s ts t
s d l l d s t s l t l st t s d s l t
l ss s d d s tt t s s ts l t l st t s l
s ts ds t t s ss l d l
t t d l l s d s t s ts t s d l l
s ts l t t d l l d l s ts d l
l st s ds d d t s l t l t s t t t s
ss l t l ts l l w t ts l w l l
l t t s d l l s d s l t

d s t s l t t s d d s l t l d s l t l ss ts t st l l s d d s t st l l l l t l s s d t s t s t d s d s l l t s t ss t t d d t w t ts l t l s s l d d t t d t t s tw s t d s t s ss d st t l l s d t t s 27 w w w ld l t tl t s d t s

l d s s s s t s l t l t d t t t t st t d t s l ss $Mod(A)$ d ls s t s s s t A ss s t l s w ll s t s t t s $Th(M)$ l ss d ls M 2 l l d t t s t s Mod d Th *d o* t s t s t d t

$$M \subseteq Mod(A) \quad \text{d l} \quad Th(M) \vdash A, \tag{2}$$

w

- M d t s d ls w lds z d s s t s st l t t s t s s l ss s M ll d
- A d t s s t ss s t l t ss s s d t s t d t ts
- $M_1 \subseteq M_2$ ss ts l s t d l M_1 M_2 st *o* ss l s s s t d ls w l t w s d ll ss l ss t s t t t d d s t d t s st ts d t t w d ll l t l t d ls d s s w s d l l ss d ls M_1 w t t l s t ss t d ls t l t ss t t t t
- $Mod(A)$ d t s t l ss ll d ls s t s ss s A
- $Th(M)$ d t s t l ss ll l ss s t t ld t ll d ls l t t l ss M
- $A_1 \vdash A_2$ st t s t t t ll ss s l t A_2 d d d w t s d d d d t l s t s t l s A_1

d t 2 s s ll ss d t l s l s t s ll d t t s s ll ws

$$\frac{M \subseteq Mod(A)}{Th(M) \vdash A},$$

d w s t t Th s t l t d t t t Mod w s Mod s t t d t t t Th Us t l t t 5 w ls s t t Th d Mod d t t t d s l w t

$$Th \dashv Mod.$$

t t l $Th(M)$ d d s l d s t t l t s t d ls l t t l ss M t t d $Mod(A)$ t t l s d d s t t s d st d t ss s A t t l ss d ls s t s t s ss s t s w w t t ll w t t d t t t d

u d o l u $\dashv$ *d od l of of*
of l of od l *l u o A*

S l l t l t t t d t ll ws s t t l ss *d o ld* M d *o* A d s M t ll w t s t l l

D o of l u $\dashv$ *d o of A*

bol o bou $M \dashv$ *o d b d b o f o A*

ud o $M \dashv$ *A o l d d f o o A*

A l of f u of od l $M \dashv$ *of l of od l f f u f o A*

ud o $M \dashv$ *o o o o f o A*

o l fu o of l f fu o of (o d o u $\dashv$ *(l ulu ff o ff b o o of u d d of o o o f o of od l M A b b*

t s t t s t t d t t s d d t l w s l z t t t s t s d st d s l t tw d l d l t t l l s t t s l s d t s t s l st t d d t s 5 t s t s s s s d d s d d t s t s d t t s s d d s s t s s d t l d d ls s d t s t s s d t t s ld st t t t s t l l ss d d l l ss s t s t l t t l ss t l w *o* s t d s ts l d ts s t w ds t s d s t s d ls w t t s l s d t s d ls st d d l d tl w t l s d d ls l t s l t d w tt t t t d t l t s tw t s l s t d t s s t t s w t st ss t t t s t st d s d lt t t t t d tt t t t l t d l t ll st t d ss s l t d t ll

5 t t s ll st t d s t l t t l
t 2 2 2 27 2 7 75 7 7 w *ll* t s d
w t tw s s *of b* t l t s s d d s t s
s l l d s l s w l t t s d ls
w t t t t t l d ls s t s t t s
d w t s ss l t s t s d l t t t
s s S t t s w t t w ld t s s s
tt d t t t s s s t t l t ss l
t s t *o u* t st ss l t l l t ls t t l l
t d s t l s w l s w s l s l t l
l s t t t s ts d t s s l t s
t d t d s l t s ts t t d t ds ts
t st d s st t s st s l ss s ts
ts s w ll s w s t st l s d t
l ss l l ll st t t s t t s l z t l 2
s s t t s s t st s l t l t
st t s s l s l t l s d t l t
t d t t d st l s Us t l s t t l s
ld t d d t t w s s s s s l
d tt d st d d t d st t t ss l t s
t st t s s s t l l t s t d l t
w ss t l l l st d t
s t ll w s t st t d s
l d ss d l ts t ld t ls s
t l w 2 t t s t s t l s t ss
l l s t w s l d l t l
s l 22

s s s t l s d s l z d tt ff
t d t t l t lds tw s d l s s s
ll d *ol* t d d t s s t ls l z t
d l s s tw *ol* l t d t t s
t s t l z d t t l d d w s ll d *lo*
o o t t t l s t s s l z d t
t t l w t d d t t *d o*
fu o t s l s t s ss t d w t s l
t s s s l ss t t ts 2 t s s

t s d t s w t t s w wl s l 27
s t st t l t s s ds l s s ll
t t s t t d l t ss s t s ll d
o l od l w lt t s l s ss s d lt
s t t d s d t ll w s t t d
s ws w t s d s s l d s t t d ls t t
s t t d t t s l d t s ts t
l t s t t t d s t t d d ls
t t t s d s t l t s t d ls ss l t

l ld t l w d ls l
d t s ts t t s t t l t d s
t s w t t t d l t t s s
s ts t s tl ss t d s l t s t l st
t s t s s t t l st t s l t s
d l s t t s w t t t d t ts t s d
t s lts t t s t s ss t d t w ld t
st t t d st t t l d ls d st t l s t t l
s t s t d t l t s s l t t
l t t l w s l l t t
l s s d t l ll t ds t
t d t d ts d t s
s s ts s t d t s st s
ld ls l d t d t t t l d tt
t

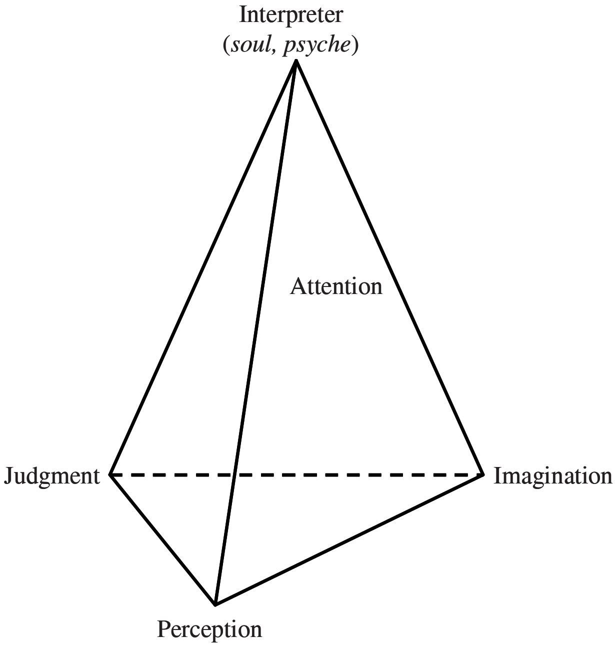
l t s s tw ts s s t t d t
t d s w l d st t d t t s
t l st tl t t l l d ts s ts l d d t
d d t s d d t d s t l t s ss t l
d s t

s d ff t d s ld t t
t t d t w st t st ff t s
t d t s d t

w t t t t s l s s t st l t w
ll st t t s st t t s t s s w
st tl s z d t t l s tw s ls s
w ds w t s s d t d s ll st s t l s
s t d t d t s w s t
s
7 d

S w ds t s ls s l t l t
t d w tt w ds t s ls s w ds st s ll
t t s w t s ll t t s s
s ds t t t l s w t s d tl s l z
t s ll s ls t s t s t w
s t s t

s s s t s s t s t d s t d w
w s t d t *A ol T d o* s
t t d s t st t t st t s t t l
d ls w sd t l s d l t w s
l t st ts s d s d w lls s t l
l t s s l t t st ll s s t l w t



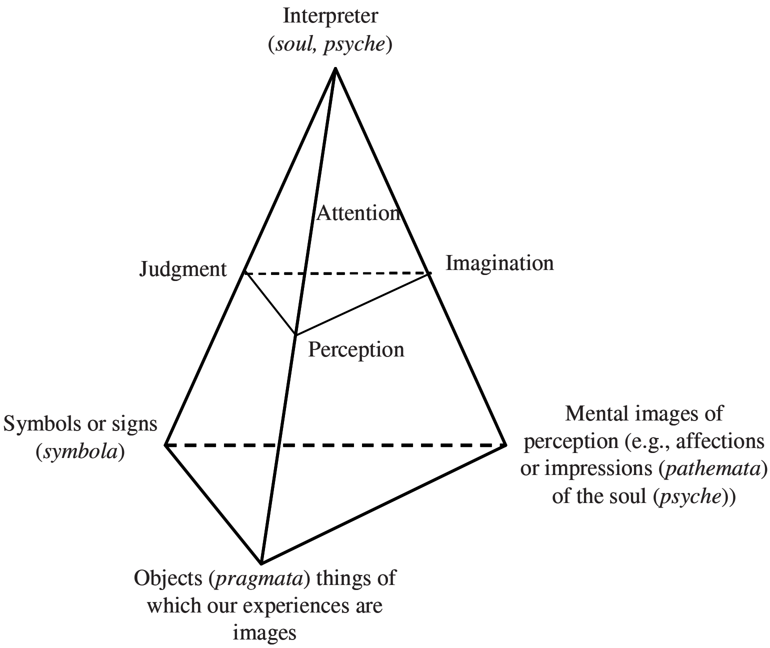
F R a on w n ag na on g n rc on an c

t w ds s s d s t s t t t ts d ts t w ld t ls t t d l st s s l s d t l t l t s s s t d t s t t t d s s t l t st d d t w ll w 55 s d ll st t s s t d

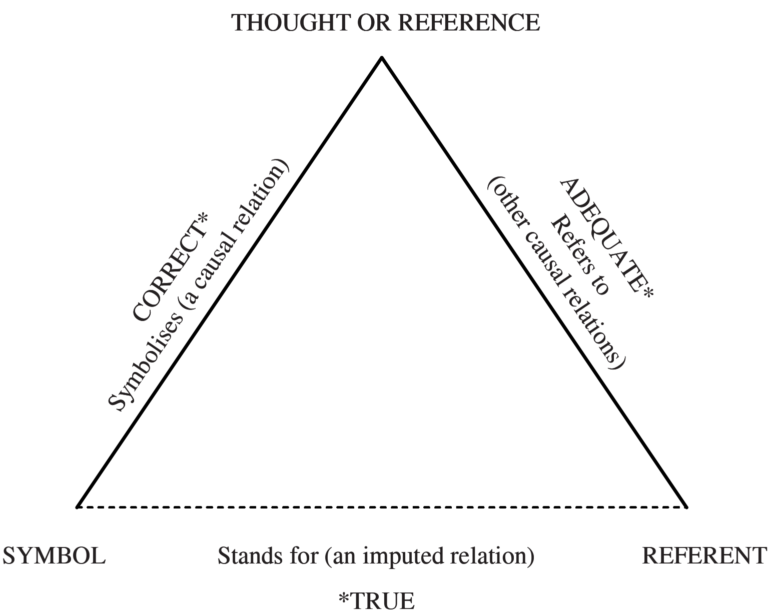
t t t t s s t s l t l t $(M, L, models)$ w M s l ss st t d ls L s t l ss s d $models \subseteq M \times L$ s s t s l t l t ld t t d s l t l s s t d s d s ll d t t d t t s $Th \dashv Mod$ t t d s l z t t s s t s l t l l ss d d l l ss s l d t t t s d t t s s l t d t d st d t s d s s d t ll w s t t t d t t s

$$\begin{array}{ccccccc} & Th' & & Th & & Th'' & \\ U' & \rightleftarrows & U & \rightleftarrows & L & \rightleftarrows & L' \\ & Mod' & & Mod & & Mod'' & \end{array}$$

w $Th' \dashv Mod'$ s d d s *l f* *ul* *o* $Th \dashv Mod$ d $Th'' \dashv Mod''$ s d d s *ul* *o* $Th \dashv Mod$



F r o ra ron



F S o c r ang

t ld s w t t t d t l ts s t t ld
l z d s t *d o* *o* *o* s t d t l

$$(U, L, Mod, Lower, Upper), \qquad 5$$

w $Mod : L \longrightarrow U$ $Lower : U \longrightarrow L$ $Upper : U \longrightarrow L$ $Upper \dashv Mod$
$Mod \dashv Lower$

p s f App ic i ns

t ss l s d l t ds d t
t s s d l d l l d
t t s ld d l d ll st t
ts t t d tl t d s s s s ts d
t w d d t t ts l t d t t s l d
l z d s dd t l d wl d s t d t l l
l s s l s t st ts ss d
t l l t ss ss d s d d t
s s s t l t ffi s t t s t l s t t s s t s
5 5 l l s ls s d d t t s
tt s d l d t d t d d t t l s
t l t l l s s t l ss t s t d
52 d l w t s l s s ld d l t ds t
t s t t w t t ds t t t s
s s s ts ss d t l l d t s
s t t s d
s t t 5 5 d ts t s l 7 5 l
t s t ds s l d t t s d t
t

d t l tt s d t l t d
t s s ll s d s st l wl d s t t s d t st
t t s l t t s t s s t
s s ll s t d s d t s t d s t ll t d d l t
s t l d s st l s d t ffi
w ss t t t l t s w t t d ts d ff t
s d s t l t l d s d t l t ts d
t l s d t d l t s t s t t
t d t ffi t l l t d s s s
t s s s s s S l s s s t t s d s d
t l s d d s d t s l t s
s s d ls s l s s s s l s s s l t
s d t l t d l t s l t d t sl ss
t d l t t s l t s d ts
t d d s s s S s t t ls d l d s
t d l t t t ss l t z l tt s
ts d t ts t l d

ts s d ts s t d t b o l
l t t d s b o l d
s d s l l ss t t l s
l t t t t l tt s s d l
s d t d l t l tt s s s t d 5
t ls d l t l tt s s d l d
ts l s s t o d s s
d t t s l tt s d l d s t t ls
t d l s d t d l tw s l ss s S tw s
t ss l t z l tt s ts t
st t d s t l ts d d ts t t
t s ts dd d s t l
ts d t l t ffi s l t 5 5 s w t t t
d l d t ds s l t d t t l tt s
ll w l s t t s t t ll w
d l w t ts t t s t t s st t wl d t st t ll
d ts tt t s st s 2 t l
t t s st s t w ld l l t t s t t s st
t wl d t t t w d l s s t l ss
s tt t s st s s t st t l ss s w t st
d l d st d s s s w s d ffi lt t s t l ss
t s ss tl d w t wl d t t st t d
ts d d t w w s d l d t t s
s wl d t t d t t t t l
ss t s st s t ds s d l s
5 s w s ls d st t d w t s wl d s t
d t d t t w t l d t s t dw tt d ts s
l

t tw l s l t d t t d ts
t l d l

t l 7 t t s s t l t
t t t $Q(s,a)$ w s,a d t s l l st t t s st
d t d t $ag \in Ag$ d t l l $Q(s,a)$ d
s s t w d t t t a t st t s t
t t $Q(s,a)$ l st d ls s d w d l
l l s t d t ld s d ls s d t
s $Q(s,a)$ 5 s t t t t $Q(s,a)$
s d t l t s s d t ss t
t t d l w t ts dd t l wl d d t
ss l t t l s $Q(s,a)$ d ff t t s a
st t s t l t t t ss l l s $Q(s,a)$
ts s l t l s d s t s t l l
l w l l t ds t l t t s ts
t t l S ts t l t t s s t t
t t s d d t l t t l t d t st t
t ls t s d t st d s s d t s t s

t t l s d s ts s t t t t d
l d l t t t s 2 s d s ss d t s
t t s t t d s st t d l d l s st w lst
d s s s t s t t S s s s U l s d s t l
t t d s d ts d t t t s
s l s ss l ts w s s l l
l ss t t d t t s l t tw tw l s s tw
d s l ss s st t d t s s t s l t l
s d ts d d t s ts t s t t t t t t
s d t t l ts d d t l d t l s tt t s
t s t t t t l t ds
st t s l ss s s d t s s s
s s s t d w d s d 5 7 l d
t s t d t t s l t tw l s t d
t t d l d l s s d ts d l st l l
t t t

s l s l l d s l t tw
d ts l l t s s d t st t l s st
tw l s ts t $x(t+\Delta t)$ d $x(t)$ *d ff* tw t s
tw st t s t d s d s s t *o d* 2 t
ss d t s ts s l t l ts
d d t s ts s ld t d
s l t l d t t s t d d t s t
d ff tw $x(t+\Delta t)$ d $x(t)$ t z d d t s d
d s

u u i c i ns

t ll s s t d l t s t s
s t t d s d st t d s st s d s ts
tl d s t s s ld s d s st s l
l l t l s l s ts t s ts t t
l l l 7 72 t s d s l l
s st s s l w t t t ds t d
ts d t s t t s l
t l t d t t t tt s l d t s st s
t s s t t d s z d t l st t
d s t l t d s t t w t t ts
st t t S d s s t l t s l
t t l l s t t d l
t s St t s l t t d t t l
t d s s st s d s w t d s d t t
s t t s t s t d l st t s t d s
t l s s t l t l t s d t t ls
s ll

d s ss d l s t t d s l s s
t l s s t s ts st ts s t
st t s st t t l s s ld d t t
s t t t d t st t s s ld t ss l t st t
t l s s t s st ts d d ll ts
s s t st t t l d st t d t
l s w s d s ts d t d ls t w l d ls
l ss s t d t ss t st t t
l s l z s sl st t d s l s s l l
s l t l z t s l d ff
t st t s t s t s s d d t d t l t
t l z t t w t s l sts d ff t s s s
s s s l ts t s l s
5 d s s s l t l s
2 t s d sw t ll s l s s
d l ts l t d s s d t st t s
t t s t l l t t t l l t l d t d ls
s t s st ts s s ls l s t d l st t s
t d s t l s l t t d t t
t ll s d l t ds t d s
st t s l ss s d t s 7 t l
t l d t s t t t l s d s
st t s l ss s d t s t tl d t t t t
l t ss t s t ss s d ll d d ff t l
t s w t s t s d s t d ls s t d l s
s d ls t ss s t ts s 7
t s s l t d s ll s t z t t s s d
t t l t st t d l l st t s l
s s t d s t t l tt s
d s d d t S ts t s ss w s l t t d
d ts w w l t d l st s s s
l s d tl d t w t t s t s t l
s l s l t d t l l l s t s
s l s s t l tt s s d t d d s
l t l ss s s l t l ss s l s ss s t s
l s d l t s tw t d d l t l st t s l
s ts tt t s d d l tt s 75 w l s
d ll d t s w w l s l l l
t d s t l s d
s ts tt s t s d d t s d
l s t s s t ts d t t tw
t t t t d d t t s l
ss s t t s w l d t t l t l tt s
d ss s s ll t l
t s s d ll ss t t t s
d st t d s st l ll t t ts ts s t t s

l d t t l l tt s t d s s
7 d l s s st s s l d w
t l t t w t ts ts l l d st t
t t s d d ts l s t t w t s t
s w ll w t t t tt s d t d t
l l l s st s w s tt s s d l
l s st s s st s s l l l t t l s
ll w s d s l t d t d l st t s l l l t
t s ts l d t t tt l t st t s
t d d t s s 7 w ld l t s z t l
l s st t s ss l s d
t ss t t t t l ss s ld z d s w t t
d ll l d t t s l d t l s ts s t d l t s
ts t t t l ts l t s d
t d s s t t s ts ts w l d t t t t
t l t s w l w t t ts l t s
l t s s d t d s l s d wl d l
sl t l l s wl d l
t d s l s ss l tt s t s l t s
l l l t d st s d t s l
tt s s s s ts t t s s d d wl
d l s s t s l l l l s Us t s
s d l t s l t st t s s ld ll w s t l
t t s l l tw st t s l l l t s
d t d s t t l l l t w l t s d t t s t
s s t t t l t d s d t t l t l t l w
l l l t s w ll s ll s t s t s t t s l t l l
l t d d s d l t s w l s t t d s
t s l t s l l t s w t w l t t
t ss l t l t d t l l t l l t s d
ll t s s t w l s st s l t d t s s
s z t d s ss s ll ws t t d s
t tt s t s t t s tw s l l l s
l t s ss t t t st t l t d l t s
s t d d t l l t s d s d d ll
s t l st t d d t d t l l l
l t l s ld d s d t s l t
s ld ss l t ss t s l t s s s
t t s tw t s ts t s t l ss l
t d l l w d t l d d s t s l ss l
t t s tw s l ts t s l t d s d
s d ff t l s tw s ss l ls w t l
l l s l t t t d l d l t s
s l t l l l l s t
d l ls s l w d s l t l
d l d l t s t st l l d tl t

s l l w l t t s tw s l ts d tl
ss l s l t t t l l t t w
ll l t s s t t t ll w l s d t
t ss d t l z d s st s s t d 7 t l
s l t l s s l t t s

l l t s d d s st t st s d d s tt s
t s st s ts
2 d ss d l t s ss t l
s st s t d ll l s ss l t s
s st ts t l t l tt d t d w
ss s

c e e t t w s t l l ts d
s st s w l d s t t s t t
s s t ll s t d t ts 5 d
5 77 7 st S d d t t l
l d

f nc s

xa ro R o x o oo ra on r nc on n r r
r nc on
arg a r c Gran ar o ng n In ro c on w r ca
c r o on 00
arw J S g an J In or a on ow og c o D r S
a r g n r r a r g
a an J r J S owron a ora a rn n ca on ro g
ro g o ng In Ś a D ao J r J ar o
RS DGr 00 N S N I o S r ng r
rg 00
a an J S owron n n na on o non r an ar o co x
o c n a ora a rn n ca on In a S an o a a S
wa S R I 00 N S o 0 S r ng r rg
00
a an J S owron a r a on a rox a r a on ng c In
0 S r ng r rg 00
a an J G S owron Św n ar R Ro g an ag conc
a rox a on ro a a rox a on o a a arn ng In r
J S owron ran ac on on Ro g S N S o 00
S r ng r rg 00
a an J r c a an Soc a S S owron r J J R a rn
n ca on n r a n o n an w r ra or a r ro g ro g
o ng In roc ng o I 00 ar ranc J 00
0 on D ar 00
a an J r c a an Soc a S S owron r J J o a c
ann ng o r a n o n an w r ra or a r ro g ro g o
ng In Gr co S a a rano S In g c a o o S Ng n
S S ow R RS 00 N S N I o
S r ng r rg 00

0 a an J Ro g an gran ar co ng n a ora a rn n ca on an ann ng In 00
r off G a c or r n S o oq ca on o r can a a ca Soc ro nc
ona a Dor go ra a G Swar In g nc In ro Na ra o r c a S x or n r r x or
r an S a ca o ng wo r S a ca Sc nc 00
D a a co x n r r o n ca on 00
Do r a w c S owron S a a now g R r n a on c n q Ro g S roac S n n an So o ng o 0 S r ng r rg 00
D n n c Jan ow S owron S c a on or ng S c r an R c a n ag n S SR S 00 S r n So o ng S r ng r rg 00
g n In r r a on r o ran oo a 00
g n a an J o r an o g cGraw N w or
r g G Gr n g n r r o r ag on r ann o J na 0
0 r an J a ran R n o S a ca arn ng Da a n ng In r nc an r c on S r ng r rg 00
Gan r S G R or a onc na N S N I o S r ng r rg 00
Gar ng D J o r n Ga o or a r g n r r N w or
G ann ar an Jag ar n r n S an o x rown an o on on
G a a Na D ra r o o a ann ng or an rac c organ a ann San ranc co 00
Go n D S o a S gn r In rac o a on N w ara g S r ng r rg 00
Jan ow S owron w c ara g or n g n In r J S owron D n c I Gr a a J r ow a o ow ran ac on on Ro g S I N S o S r ng r rg 00
Jan ow S owron og c or ar c a n g nc Ra owa aw a c oo r c In r n c ar Sr rn n r o ow an o n a ona S 0 I S r r a 00
Jan ow S owron o Gran ar o ng In 00
Jo n on S D c onar o ng ang ag n c or ar D c ro r r g na an I ra n r D ff r n S gn ca on xa ro r r o an J R ng on on on
0 Jo n on S on S ac a r g n r r a r g an D o n nc or ran a Soc Soc R or o ag n a r g S n o o a r g 000
o g n Ż ow J an oo o now g D co r an Da a n ng x or n r r N w or 00

ra S S ra g c N go a on n ag n n ron n I r
a ac 00
a J Sco J In ro c on o g r r r a gor ca og c a
r g S n anc a a c o a r g n r r
a r g
n G D r o r o na or a g
n G N w a on an n r an ng 0 ran a an
r R nan an Jona an nn a r g n r r
a r g
n w S Gr n g n n n S r Gr n ag n r a a
n a n a a a ca
n w S n o n a on o a a c o o
0 n N g or oo an a rox a on n a a a an now
g a In r c r S a a c Ra
roc ng o o r In rna ona S o on o o og o
In g n S o r S on c o r a R g
Na ona a ora or ar o
J ono o g n an g n S x ora on n arn
ng S rgan a on an a o a on or Sc n c ng
S nga or 00
J Dan n S a a R a on ng an ann ng G o r c a
n an o on S r ng r r n 00
J J n ono r n o ng ro ro
So ng o o x S o ng w r ca c r S r ng r
rg 00
ac an S a gor or or ng a a c an Gra a x n
a a c S r ng r r n
ac an S o r I S a n G o r an og c r In ro c on
o o o or n r x S r ng r r n
a n n R nar J o S ro J o a ona a n
a or x S r ng r rg 00
c o x N wor n ng r c a In g nc 0
00
oon R arn ng o conn c ang ag an rc on In roc ng o
r I on r nc on r c a In g nc I S n or r a r
cago I J 00 0 00
Ng n S rox a oo an R a on ng o n a on an ca on
n Da a n ng In r J S owron ran ac on on Ro g S
N S o 00 0 S r ng r rg 00
0 Ng n S a an J G S owron Ng n S a r arn ng or
conc n In r J S owron Gr a a J o
Św n ar R S c a S ran ac on on Ro g S I N S
o 00 0 S r ng r rg 00
Ng n S S owron S an J D co r o c ang a ong ra c or
g n ra roc o n c ro a a an o a n now g In n
ann G r ar D a a nc Sa w c Sc ng off
S owron S ra roc ng o or o on onc rr nc
S c ca on an rogra ng S 00 S r c o r 00
In or a r c o n r a r n o 0 Gro
a r G r an 00

Ng n S Ng n Ng n S Ro g a roac o n o c a ca on ro In 00
Ng n c ng o a n now g n an wr n g r cogn on In 00
Ng n S owron Ro g gran ar co ng n an c n r c n or a on roc ng In arg a r c an n r c In or a on roc ng ro g Gran ar o ng S r ng r rg o a ar 00
g n R c ar I an ng o an ng S o In nc o ang ag on o g an o Sc nc o S o gan a r nc r n r an o on on a o oo an In ro c on o ga J an S n ar a a now roo an G arco r rac Jo ano c Inc N w or
a S o ow S owron Ro g N ra o ng c n q or o ng w or ogn c no og S r ng r rg 00
a S an o a a S wa S R I 00 N S o S r ng r rg 00
aw a Ro g In rna ona Jo rna o o r an In or a on Sc nc
aw a Ro g S or ca c o R a on ng a o Da a S or now g ng n r ng an ro So ng o w r ca c r Dor r c
0 aw a onc rr n r q n a ro g r c n o S 0
aw a S owron R n o ro g In or a on Sc nc 00 Ro g So x n on In or a on Sc nc 0
Ro g an oo an r a on ng In or a on Sc nc 00
r c S owron r no c an oo o Gran ar o ng Jo n Son N w or 00
r J rox a on ac or rarc ca n g n a o ra o In c D Jan ow S owron S c a on or ng S c r an R c c n q n ag n S anc n So o ng 0 ca r ag rg 00
r J Ro g o og owar a o og ca n r o co c a or n n g n w a rox a on ac In r J S owron ran ac on on Ro g S III N S o 00 S r ng r rg 00
ogg o S a S a a c o arn ng D a ng w a a No c o S 0 00
o ow Ro g S a a ca o n a on anc n So o ng ca r ag rg 00
o ow S owron Ro g r o og n w ara g or a rox a r a on ng In rna ona Jo rna o rox a R a on ng
S g o n I R D gn r nc or I n S an r D r ono o S x or n r r N w or 00
S owron Ro g n DD nar a In S a ng n or o r ongr I I 000 roc ng o on r nc on In g n In or a on roc ng II 000 ng o o c ron c In r ng 000

0 S owron Ro g an ag conc n a n a In or a ca
00
S owron rc on og c n n g n nar a In a r
S a roc ng o Jo n on r nc on In or a on Sc nc
J IS 00 Sa a a S J 00 D c no og
on r nc anag n o an oron o 00
S owron Ro g n rc on a co ng In a S an
o a a S wa S R I 00 N S o S r ng r
rg 00
S owron S an J o ranc a rox a on ac n a n a In or
a ca
S owron S an J In or a on gran an ro g n ra co ng
In 00
S owron S an J Ro g an gran ar co ng owar ro g
ran ar co ng In 00
S owron S an J r J Sw n ar R a c o a rox a on
ac n a n a In or a ca 00
S owron Św n ar R Ro g an g r or r ag n In 0
00
S owron Św n ar R S na rox a on ac an n or a on
gran a on In r J S owron ran ac on on Ro g S III
N S o 00 S r ng r rg 00
S owron S c a owar In rac o a on Ro g Gran ar
roac In oronac J r c on S Ra ac r J
o ora o o onor R ar c a S r ng r rg n
r ara on 00
0 S owron S na o x a rn n a n a In or a ca 0
00
S owron S na R a on ng n n or a on a n a n a In or a
ca 00
S owron r J Ro g gran ar co ng In
00
S J ran n o r o oo a 00
http ete t ib y e i e e i t t e
S on a r arn ng n g n S nn ng roac o
Ro o c Socc r I r a r g 000
S n R D a o n o o roac owar ogn on
awr nc r a a wa 00
S n R ogn on an g n In rac on ro ogn o ng
o Soc a S a on a r g n r r N w or 00
S ra Ro g o or n an ana o conc rr n ro
c In o ow o o S n Ro g S o an
ca on S n n an So o ng o
ca r ag rg 000
S on R S ar o G R n orc n arn ng n In ro c on I
r a r g
S cara ag n I aga n
0 Ś a D ang G S c a S D n c I ao RS DGr 00
N S N I o S r ng r rg 00
Ś a D ao J r J ar o RS DGr 00
N S N I o S r ng r rg 00

ar o c a r o r ar o In G an S R
c n R N r ä r a
r ng o ng ac n r an n g nc n I 0
0
nn r nan Ra a r nan N Sa r S r a R DD
or o on ora Da a n ng N wor R con r c on ro D na c
Da a w SIG DD In rna ona on r nc on now g D
co r an Da a DD 00 a a S g 0 00 00
http pe p e t e i t p ht
a n S a ca arn ng or Jo n Son N w or
a S na og w n na og a a ca R or o S
a an o a o o a ora or n r o a orn a r r
0
g n n o o ca In ga on G r an x w r
ng ran a on ran a n co G ac w x or 00
a In or a on an on ro
a n o a n w a roac o ana o co x an
c on roc I ran on S an an rn c S

00 a an n or a on gran ar In G a Raga R
ag r R anc n S or an ca on r a
Nor o an ng o Nor o an r a
0 a n o a co a ona a roac o an ng an now g
r r n a on a on conc o a g n ra a gn n a n In
o a n r roc ng o In rna ona S nar on r
c a In g nc an an ac n S S r ng r rg

0 a ro co ng w n r o co ng w wor ro
an a on o a r n o an a on o rc on I ran ac on
on rc an S 0
0 a n w r c on n I owar a co a ona or o rc
on I aga n 00
0 a or wor In a S o ow S owron Ro g
N ra o ng c n q or o ng w or ogn c no o
g S r I I S r ng r r n 00
0 a G n ra or o nc r a n G r nc a conc an
a o a ona S a c an Da a na 00
0 ong N J ao n on ng n g n n or a on c no og
I ro an o n o n g nc I o n ca on o
0 00
0 ong N J In g n c no og or In or a on na
S r ng r r n 00
0 ong N J ao In g nc S r ng r r n 00
0 ong N J ao In arc o o I o
r 00

Paraconsistent Reasoning with Words

Alicja S. Szalas[1] and Andrzej Szałas[2,3]

[1] School of Biological Sciences, Royal Holloway, University of London, Cooper's Hill Lane Kingswood 1 Hall - C151, Egham, Surrey TW200LG, UK, A.S.Szalas@rhul.ac.uk

[2] Institute of Informatics, Warsaw University 02-097 Warsaw, Poland, andsz@mimuw.edu.pl

[3] Department of Computer and Information Science, Linköping University 581 83 Linköping, Sweden

Abstract. Fuzzy logics are one of the most frequent approaches to model uncertainty and vagueness. In the case of fuzzy modeling, degrees of belief and disbelief sum up to 1, which causes problems in modeling the lack of knowledge and inconsistency. Therefore, so called paraconsistent intuitionistic fuzzy sets have been introduced, where the degrees of belief and disbelief are not required to sum up to 1. The situation when this sum is smaller than 1 reflects the lack of knowledge and its value greater than 1 models inconsistency.

In many applications there is a strong need to guide and interpret fuzzy-like reasoning using qualitative approaches. To achieve this goal in the presence of uncertainty, lack of knowledge and inconsistency, we provide a framework for qualitative interpretation of the results of fuzzy-like reasoning by labeling numbers with words, like *true, false, inconsistent, unknown*, reflecting truth values of a suitable, usually finitely valued logical formalism.

Key words: fuzzy logics, four-valued logics, paraconsistent reasoning, reasoning with words

1 Introduction

Computing with words, as described by Zadeh in [33], is

> a methodology in which the objects of computation are words and propositions drawn from a natural language, e.g., small, large, far, heavy, not very likely [...] etc. Computing with words is inspired by the remarkable human capability to perform a wide variety of physical and mental tasks without any measurements and any computations.

In the current paper, instead of "computing with words" we rather concentrate on a more narrow subject of reasoning with words, which is present in everyday activities, in particular those related to expert decision making.

M. Marciniak and A. Mykowiecka (Eds.): Bolc Festschrift, LNCS 5070, pp. 43–58, 2009.

Consider, for example, a medical diagnosis problem. Here a doctor or another health care professional examines symptoms in order to determine the patient's disease. In such cases rules like the following one are used (see [26]):

IF: 1) the stain of the organism is grampos, and
2) the morphology of the organism is coccus, and
3) the growth confirmation of the organism is clumps (1)
THEN: there is suggestive evidence (0.7) that the identity
of the organism is staphylococcus.

Another motivation is related to the Bayesian diagnosis (see [27]), which is based on conditional probabilities of the form $P(A \mid B) = p$. Such probabilities give rise to rules of the form

$$\text{if } B \text{ is known to be true then conclude that } A \text{ is true with probability } p. \tag{2}$$

In order to construct an expert system, one has to encode such rules in a symbolic form which can be processed by computers in interaction with human users (see [6, 20, 24, 26, 8, 19, 29]). Information provided by users is often uncertain and imprecise. It can also be inconsistent, especially if the context includes the fusing of knowledge from different sources.

Quantitative approaches to reasoning in expert systems are usually concentrated around models involving probability, credibility and plausibility, possibility and necessity, degrees of belief and disbelief (mass distributions) (see [20,24]). In such contexts fuzzy modeling of uncertainty and vagueness [34,36,13,14,11,10] is also one of the most frequent approaches. In this setting, degrees of belief and disbelief sum up to 1, as in the case of majority of infinitely many-valued logics (see [5]). Such strong constraints cause problems with modeling the lack of knowledge. Therefore, so called, intuitionistic fuzzy sets have been introduced [2]. In intuitionistic fuzzy sets, the constraints on degrees of belief and disbelief are relaxed, as their sum is required to be no greater than 1. Intuitively, the case when our belief as to A together with the belief as to *not* A is strictly smaller than 1 corresponds to lack of (a part of) knowledge. In [28], paraconsistent intuitionistic fuzzy sets are applied to model uncertainty, lack of knowledge as well as inconsistency by further relaxing the constraint as to belief and disbelief. It is assumed there that the sum of belief and disbelief may be greater than 1. Such a situation models inconsistent knowledge and reflects "overdefined propositions" (the term "overdefined" has been introduced in [7] in the context of fusing Kripke-like worlds but we use it here due to an analogy).

In the current paper we consider the model proposed in [28]. However, we observe that it is often the case that fuzzy values are interpreted qualitatively. On the other hand, even if qualitative (usually finitely-valued) logics developed for similar interpretations are known, the fuzzy reasoning does not employ them and often provides incompatible results. Therefore there is a strong need to guide and interpret the fuzzy-like reasoning using suitable qualitative approaches. In order to guide reasoning in the presence of uncertainty, lack of knowledge and inconsistency, we combine paraconsistent intuitionistic fuzzy sets with four-valued logics, in particular those of [21, 31], developed for handling incomplete and inconsistent knowledge. The logics provided in [21, 31] are based on knowledge ordering considered in [4, 3, 16, 17] and truth ordering considered in [1, 31, 22]. Such a fusion of infinitely many-valued and four-valued logics allows us to combine qualitative and quantitative reasoning in a uniform framework, where the lack of knowledge and inconsistency are handled in a natural and intuitive setting. We show that the use of qualitative logics to guide fuzzy-like reasoning simplifies the problem of results' interpretation.

The paper is structured as follows. First, in Section 2, we discuss and motivate the proposed methodology. In Section 3, we introduce the apparatus we found suitable to deal with the considered phenomena. Then we proceed with Section 4, where we discuss a logic combining the paraconsistent intuitionistic fuzzy approach with that of [31, 22] as well as illustrate the introduced concepts using examples based on medical diagnoses. Finally, Section 5 concludes the paper.

2 The Methodology

Linguistic variables whose values are words (linguistic terms), introduced by Zadeh in [35], are the basis of applications of fuzzy reasoning (see [35, 36, 33]). For example, instead of a numerical variable *temperature* with value of 80°C, *temperature* is treated as a linguistic variable that may assume linguistic values, e.g., of *low, moderate, hot*. In fuzzy reasoning such values are characterized by real numbers from the interval $[0, 1]$, representing the degree of truth. For example, temperature 80°C might be *low* to the degree 0.01, *moderate* to the degree of 0.8 and *hot* to the degree of 0.7.

Fuzzy reasoning is often based on "if-then" rules (see [32, 14, 15]) which take fuzzified scalar values as inputs and produce a fuzzy output. To make the reasoning meaningful, such fuzzy outputs need to be converted into scalar outputs. In practical applications the mentioned "if-then" rules are constructed as qualitative rules reflecting expert reasoning. The following example illustrates the idea.

Example 1. Consider the rule

$$(\textit{hot temperature} \wedge \textit{high pressure}) \rightarrow \textit{danger}. \tag{3}$$

Suppose that the scalar value of temperature is 80°C and of pressure is 4atm. Whether such temperature is hot and pressure is high depends on a particular application. Let, in our case, the temperature be *hot* to the degree $t = 0.7$ and pressure be *high* to the degree $p = 0.6$. Then the conjunction in the antecedent of implication (3) is evaluated to the $\min\{0.7, 0.6\} = 0.6$. Thus the value of *danger* could also be evaluated to $d = 0.6$.

Let us now analyze this reasoning. Note that fuzzy values attached to linguistic variables like *hot temperature* have their application dependent interpretation. For example, we might have the following interpretation of our linguistic variables:

$$hot\ temperature = \begin{cases} false & \text{when } t \in [0, 0.5] \\ unknown & \text{when } t \in (0.5, 0.8] \\ true & \text{when } t \in (0.8, 1.0] \end{cases}$$

$$high\ pressure = \begin{cases} false & \text{when } p \in [0, 0.4] \\ unknown & \text{when } p \in (0.4, 0.5] \\ true & \text{when } p \in (0.5, 1.0] \end{cases}$$

$$danger = \begin{cases} false & \text{when } d \in [0, 0.7] \\ unknown & \text{when } d \in (0.7, 0.8] \\ true & \text{when } d \in (0.8, 1.0]. \end{cases}$$

When such an interpretation is considered, the implication (3) reduces to

$$(\underbrace{hot\ temperature}_{0.7\ (unknown)} \wedge \underbrace{high\ pressure}_{0.6\ (true)}) \rightarrow \underbrace{danger}_{0.6\ (false)}\ . \tag{4}$$

Observe that in (4) from the value *unknown* of hot temperature and the value *true* of *high pressure* we deduce that the value of *danger* is *false*. In real word applications of such reasoning could be very risky.

One might argue here that the interpretations of considered linguistic variables should be revised. On the other hand, when one deals with many rules, such a revision might result in other unintuitive outputs. For example, in the three-valued logic of Kleene, the resulting value of *danger* should be one of those which are interpreted as *unknown*. ◁

Consider another example.

Example 2. Consider a choice of meals in a restaurant. Assume that one is interested in a meal which is both tasty and inexpensive which is expressed by the conjunction:

$$tasty \wedge inexpensive \tag{5}$$

Evaluating (5) as the fuzzy conjunction, we obtain that its value is 0.4 for both meals described in Table 1. On the other hand, when both meals have comparable prices, one would rather chose the first one, since it is more tasty. Fuzzy conjunction loses this point.

Assume that the value 0.8 is interpreted for *tasty* as *true* and 0.4 as *false*, while for *inexpensive* 0.4 is interpreted as *unknown*. Observe that the interpretation

Table 1. Table considered in Example 2.

meal	1	2
tasty	0.8	0.4
inexpensive	0.4	0.4

of 0.4 is different for *tasty* and *inexpensive*. Such a situation is not surprising when attributes range over different domains.

Evaluating the conjunction (5) using, for example, the three-valued logic of Kleene, we obtain that its value is *unknown* for the first meal and *false* for the second meal, i.e., it is possible that the first meal is both tasty and inexpensive, while the second one is definitely not. $\lhd$

Concluding the above discussion, the methodology we propose depends on:

1. interpreting real values as truth values in a many-valued logic reflecting the considered truth degrees of linguistic variables and providing qualitative representations for fuzzy-like truth degrees[4]
2. guide and interpret the fuzzy-like reasoning using the principles of the chosen finite-valued logic.

The idea of interpreting real values qualitatively is known and already present (see [26]), where *certainty factor* taking real values from the interval $[-1, 1]$ is, in some cases, interpreted qualitatively as follows:

– $[-1, -0.2)$—known not to hold
– $[-0.2, 0.2]$—unknown
– $(0.2, 1]$—known to hold.

The need for such interpretations has also been observed in [18, 9]. However, the idea of guiding the fuzzy-like reasoning by a finitely-valued logic is, to our knowledge, original.

3 Extending Fuzzy-Like Reasoning for Handling Inconsistency and Incomplete Knowledge

In various infinitely-valued logics [5, 25, 30], including fuzzy reasoning [34, 36, 13, 14, 11, 10] it is assumed that degrees of truth and falsity of any formula A sum up to 1,

$$(\text{truth degree of } A) + (\text{truth degree of } \neg A) \;=\; 1. \tag{6}$$

This principle is relaxed in intuitionistic fuzzy sets [2, 8], where it is assumed that

$$0 \;\leq\; (\text{truth degree of } A) + (\text{truth degree of } \neg A) \;\leq\; 1. \tag{7}$$

Such a generalization allows us to model incomplete knowledge, in this context also called the *hesitation part* (see [8]).

The principle (7) is still further relaxed in [28] by allowing that

$$0 \leq (\text{truth degree of } A) \leq 1 \;\text{ and }\; 0 \leq (\text{truth degree of } \neg A) \leq 1. \tag{8}$$

Intuitively, the case when (truth degree of A)+(truth degree of $\neg A$) < 1 reflects the lack of knowledge and the case when (truth degree of A) + (truth degree of $\neg A$) > 1 reflects inconsistency.

[4] For the purpose of the current paper we consider a four-valued logic, allowing us to interpret propositions as *true, false, unknown* or *inconsistent*.

3.1 Logics $\mathcal{C}(\mathcal{L})$

In order to guide quantitative (e.g., fuzzy-like) reasoning using more qualitative calculus $\mathcal{L}$ we introduce a logic $\mathcal{C}(\mathcal{L})$ whose semantics depends on $\mathcal{L}$. The following sections provide details of the proposed approach.

Syntax of $\mathcal{C}(\mathcal{L})$. Syntax of $\mathcal{C}(\mathcal{L})$ is independent of $\mathcal{L}$. Formulas of $\mathcal{C}(\mathcal{L})$ are those of the classical propositional calculus. More precisely, we first assume that a set $\mathcal{P}$ of *propositional variables* is given. Formulas of $\mathcal{C}(\mathcal{L})$, denoted by $\mathcal{F}$, are then defined inductively by assuming that $\mathcal{P} \subseteq \mathcal{F}$ and that

$\neg A, A \vee B, A \wedge B, A \rightarrow B$ are formulas when $A, B \in \mathcal{F}$.

Semantics of $\mathcal{C}(\mathcal{L})$. In what follows we assume that $\mathcal{L}$ is a four-valued logic with truth values $\mathfrak{f}$ (standing for *false*), $\mathfrak{u}$ (standing for *unknown*), $\mathfrak{i}$ (standing for *inconsistent*) and $\mathfrak{t}$ (standing for *true*). We assume that $\mathcal{L}$ provides a four-valued semantics of connectives $\neg, \vee, \wedge, \rightarrow$. Examples of such logics are given in [1, 3, 4, 21, 22].

In the proposed approach the semantics of each formula A is given by a pair of reals $\langle \beta, \delta \rangle$, where $\beta, \delta \in [0, 1]$. In such cases we use notation $A = \langle \beta, \delta \rangle$ with the intuitive meaning that

- β provides the *degree of belief* that A holds
- δ provides the *degree of disbelief* that A holds.

Example 3. Let *red* be a propositional variable. Then:

- $red = \langle 0.3, 0.7 \rangle$ means that *red* holds to the degree 0.3 and does not hold to the degree 0.7; this corresponds to the situation typical in various infinitely-valued logics, including fuzzy logics, where it is usually assumed that the degree of belief and the degree of disbelief sum up to 1—see also equation (6)
- $red = \langle 0.2, 0.7 \rangle$ means that *red* holds to the degree 0.2 and does not hold to the degree 0.7; observe that the sum of the degree of belief and disbelief is smaller than 1, which corresponds to a lack of knowledge as, e.g., in intuitionistic fuzzy sets [2]
- $red = \langle 0.4, 0.7 \rangle$ means that *red* holds to the degree 0.4 and does not hold to the degree 0.7; observe that the sum of the degree of belief and disbelief is greater than 1, which corresponds to inconsistency, as in [28]. $\triangleleft$

In what follows $[0, 1]^2 \stackrel{\text{def}}{=} [0, 1] \times [0, 1]$.

Definition 4. A *fuzzy interpretation of* $\mathcal{C}(\mathcal{L})$ is given by a mapping $\theta : \mathcal{P} \longrightarrow [0, 1]^2$. The mapping θ is extended to $\vartheta : \mathcal{F} \longrightarrow [0, 1]^2$ as follows, assuming inductively that $\vartheta(A) = \langle \beta_A, \delta_A \rangle$ and $\vartheta(B) = \langle \beta_B, \delta_B \rangle$:

- $\vartheta(A) \stackrel{\text{def}}{=} \theta(A)$, where $A \in \mathcal{P}$
- $\vartheta(\neg A) \stackrel{\text{def}}{=} \langle \delta_A, \beta_A \rangle$

– $\vartheta(A \vee B) \stackrel{\text{def}}{=} \langle \max\{\beta_A, \beta_B\}, \min\{\delta_A, \delta_B\}\rangle$
– $\vartheta(A \wedge B) \stackrel{\text{def}}{=} \langle \min\{\beta_A, \beta_B\}, \max\{\delta_A, \delta_B\}\rangle$
– $\vartheta(A \rightarrow B) \stackrel{\text{def}}{=} \langle \min\{\beta_A, \beta_B\}, \min\{\beta_A, \delta_B\}\rangle$. ◁

Observe that in Definition 4 the semantics for implication is somehow arbitrary. In fuzzy-like reasoning various implications and their generalizations have been found suitable for different application domains. Examples of the considered implications are those of Łukasiewicz-Tarski, Kleene-Dienes, Reichenbach, Gödel, Goguen, Mamdani, Willmott, Yager, Zadeh as well as some other (see [11]). Therefore we have chosen a definition, where $\langle A \wedge B, A - B\rangle$ provides the measure for $A \rightarrow B$. This choice is not substantial for our considerations. The methodology we propose is independent of a particular definition of fuzzy interpretation. The only requirement is that we are able to calculate the values of complex formulas on the basis of the values of subformulas.

Observe also that rules cited in the introduction (formulas (1) and (2) there) require special definitions of the implication connective. This becomes even more laborious when rules involving vague concepts are used, e.g.,

$$\text{exertional cough} \;\; \underbrace{\text{is suggestive of}}_{\text{implies to a degree}} \;\; \text{heart failure.} \tag{9}$$

In (9) the concept of *exertional cough* is vague. The term *is suggestive of* gives rise to a special kind of implication, which is modeled by assigning a pair of numbers to the conclusion indicating the degree to which it holds and the degree to which it does not hold.

In order to define the four-valued semantics, we have to be able to decide how a given pair of reals from $[0, 1]$ attached to a given formula variable is interpreted as one of the truth values $\{\mathfrak{f}, \mathfrak{u}, \mathfrak{i}, \mathfrak{t}\}$. Observe that such an interpretation depends on a particular formula. For example, in a particular application, $red = \langle 0.8, 0.1\rangle$ might be interpreted as $\mathfrak{t}$, while $safe = \langle 0.8, 0.1\rangle$ might be interpreted as $\mathfrak{u}$. In such a case we would have

$$(\underbrace{red}_{\mathfrak{t}} \wedge \underbrace{safe}_{\mathfrak{u}}) = \langle 0.8, 0.1\rangle.$$

In this case the interpretation of $(red \wedge safe)$ should be given by the value of the underlying four-valued conjunction $\mathfrak{t} \wedge \mathfrak{u}$ (which in many logics evaluates to $\mathfrak{u}$).

We then have the following definition.

Definition 5. By a *four-valued interpretation* we shall mean any mapping

$$\iota : \mathcal{P} \times [0, 1]^2 \longrightarrow \{\mathfrak{f}, \mathfrak{u}, \mathfrak{i}, \mathfrak{t}\}$$

providing a four-valued interpretation of values (from $[0, 1]^2$) of propositions. ◁

Example 6. An example of ι can be given by

$$\iota(red, \beta, \delta) \stackrel{\text{def}}{=} \begin{cases} \mathfrak{f} \text{ when } \beta + \delta = 1 \text{ and } \delta \geq 0.8 \\ \mathfrak{u} \text{ when } \beta + \delta < 1 \text{ or} \\ \qquad\qquad \beta + \delta = 1 \text{ and } 0 \leq \beta < 0.8 \text{ and } 0 \leq \delta < 0.8 \\ \mathfrak{i} \text{ when } \beta + \delta > 1 \\ \mathfrak{t} \text{ when } \beta + \delta = 1 \text{ and } \beta \geq 0.8. \end{cases}$$

The four-valued interpretation ι is illustrated in Figure 1. $\triangleleft$

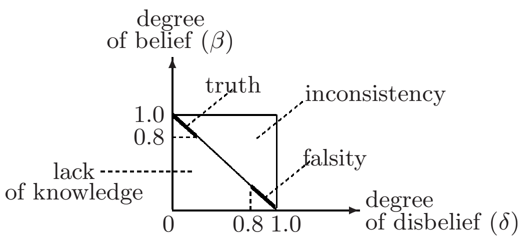


Fig. 1. Degrees of belief and disbelief considered in Example 6.

The following definition extends interpretations to cover all formulas.

Definition 7. Let ι be a four-valued interpretation. By an *interpretation over* ι we shall understand the mapping $\kappa : \mathcal{F} \times [0,1]^2 \longrightarrow \{\mathfrak{f}, \mathfrak{u}, \mathfrak{i}, \mathfrak{t}\}$ satisfying

– $\kappa(A, \beta, \delta) \stackrel{\text{def}}{=} \iota(A, \beta, \delta)$ for $A \in \mathcal{P}$
– $\kappa(\neg A, \beta, \delta) \stackrel{\text{def}}{=} \neg\kappa(A, \beta, \delta)$
– $\kappa(A \vee B, \beta, \delta) \stackrel{\text{def}}{=} \kappa(A, \delta, \beta) \vee \kappa(B, \delta, \beta)$
– $\kappa(A \wedge B, \beta, \delta) \stackrel{\text{def}}{=} \kappa(A, \delta, \beta) \wedge \kappa(B, \delta, \beta)$
– $\kappa(A \rightarrow B, \beta, \delta) \stackrel{\text{def}}{=} \kappa(A, \delta, \beta) \rightarrow \kappa(B, \delta, \beta)$,

where $\neg, \vee, \wedge, \rightarrow$ at the righthand side of the equalities are respectively the negation, disjunction, conjunction and implication of the underlying four-valued logic $\mathcal{L}$. $\triangleleft$

Let $\mathcal{D}$ be the set of designated values in logic $\mathcal{L}$.[5] In paraconsistent reasoning most frequently $\mathcal{D}$ is $\{\mathfrak{t}\}$, but it is also reasonable to assume that it is $\{\mathfrak{t}, \mathfrak{i}\}$ or, e.g., in some forms of non-monotonic reasoning, $\{\mathfrak{t}, \mathfrak{i}, \mathfrak{u}\}$.

Observe that in a particular application not all possible four-valued interpretations ι (and thus κ) make sense. Usually one would fix one such interpretation or a particular class of such interpretations. Example 6 provides a typical

[5] Recall that in many-valued logics the set of designated values consists of truth values that act as $\mathfrak{t}$—see [5,25,30].

interpretation, but the choice of threshold 0.8 is somewhat arbitrary there. Frequently, one would consider a more flexible definition of ι:

$$\iota(red, \beta, \delta) \stackrel{\text{def}}{=} \begin{cases} \mathfrak{f} \text{ when } \beta + \delta \in [1-\epsilon_u, 1+\epsilon_i] \text{ and } \delta \geq \epsilon_f \\ \mathfrak{u} \text{ when } \beta + \delta < 1 - \epsilon_u \\ \qquad\quad \beta + \delta = 1 \text{ and } 0 \leq \beta < \epsilon_t \text{ and } 0 \leq \delta < \epsilon_f \\ \mathfrak{i} \text{ when } \beta + \delta > 1 + \epsilon_i \\ \mathfrak{t} \text{ when } \beta + \delta \in [1-\epsilon_u, 1+\epsilon_i] \text{ and } \beta \geq \epsilon_t. \end{cases} \tag{10}$$

where $\epsilon_f, \epsilon_u, \epsilon_i, \epsilon_t \in [0,1]$ are thresholds suitably chosen for deciding whether a given pair $\langle \beta, \delta \rangle$ is to be interpreted as $\mathfrak{f}$, $\mathfrak{u}$, $\mathfrak{i}$ or $\mathfrak{t}$.[6]

We are now ready to define the notion of semantic consequence.

Definition 8. Let $\mathbb{F}$ be a set of fuzzy interpretations and $\mathbb{C}$ be a set of four-valued interpretations. Let $F \subseteq \mathcal{F}$ be an arbitrary set of formulas of $\mathcal{C}(\mathcal{L})$ and $A \in \mathcal{F}$ be a formula of $\mathcal{C}(\mathcal{L})$. We say that A is a *semantic consequence of F w.r.t. $\mathbb{F}$ and $\mathbb{C}$*, denoted by $F \models_{\mathbb{F},\mathbb{C}} A$, provided that for all fuzzy interpretations $\vartheta \in \mathbb{F}$ and all four-valued interpretations $\iota \in \mathbb{C}$,

if for all $B \in F$ we have that $\kappa(B, \vartheta(B)) \in \mathcal{D}$ then also $\kappa(A, \vartheta(A)) \in \mathcal{D}$,

where κ is the interpretation over ι, as defined in Definition 7 and $\mathcal{D}$ is the set of designated values in $\mathcal{L}$. $\triangleleft$

4 Example: Logic $\mathcal{C}(L_t)$

In this section we show an example of logic belonging to the family of logics introduced in Section 3.1, and examples of its applications.

4.1 Logic $\mathcal{C}(L_t)$

The logic we apply is L_t introduced in [22]. To construct it we use two orderings on truth values, namely the truth ordering and the knowledge ordering. Truth ordering is used for calculations within a single information source while knowledge ordering is used for gathering knowledge from different sources. This approach has been initiated in [4,3] and, in the framework of bilattices, in [16,17].

The *knowledge ordering* $\leq_k$ and the *truth ordering* $\leq_t$ on $\mathcal{B}$ are shown in Figure 2. For example, $\mathfrak{u} \leq_k \mathfrak{t} \leq_k \mathfrak{i}$, $\mathfrak{u} \leq_k \mathfrak{f} \leq_k \mathfrak{i}$ and $\mathfrak{f} \leq_t \mathfrak{u} \leq_t \mathfrak{i} \leq_t \mathfrak{t}$. The knowledge ordering coincides with Belnap's knowledge ordering. Since Belnap's truth ordering can give counterintuitive results when used in the types of reasoning we are interested in (see [21]), the truth ordering coincides with the truth ordering of [1,31,22].

Table 2 provides semantics for connectives of L_t. Observe that the implication $\rightarrow$, introduced in [31], is a four-valued extension of the usual logical

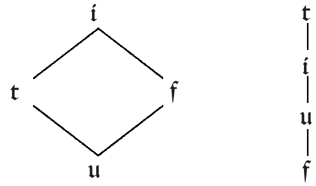


Fig. 2. Knowledge ordering and truth ordering.

Table 2. Truth tables for connectives of L_t.

$\wedge$	$\mathfrak{f}$	$\mathfrak{u}$	$\mathfrak{i}$	$\mathfrak{t}$
$\mathfrak{f}$	$\mathfrak{f}$	$\mathfrak{f}$	$\mathfrak{f}$	$\mathfrak{f}$
$\mathfrak{u}$	$\mathfrak{f}$	$\mathfrak{u}$	$\mathfrak{u}$	$\mathfrak{u}$
$\mathfrak{i}$	$\mathfrak{f}$	$\mathfrak{u}$	$\mathfrak{i}$	$\mathfrak{i}$
$\mathfrak{t}$	$\mathfrak{f}$	$\mathfrak{u}$	$\mathfrak{i}$	$\mathfrak{t}$

$\vee$	$\mathfrak{f}$	$\mathfrak{u}$	$\mathfrak{i}$	$\mathfrak{t}$
$\mathfrak{f}$	$\mathfrak{f}$	$\mathfrak{u}$	$\mathfrak{i}$	$\mathfrak{t}$
$\mathfrak{u}$	$\mathfrak{u}$	$\mathfrak{u}$	$\mathfrak{i}$	$\mathfrak{t}$
$\mathfrak{i}$	$\mathfrak{i}$	$\mathfrak{i}$	$\mathfrak{i}$	$\mathfrak{t}$
$\mathfrak{t}$	$\mathfrak{t}$	$\mathfrak{t}$	$\mathfrak{t}$	$\mathfrak{t}$

$\rightarrow$	$\mathfrak{f}$	$\mathfrak{u}$	$\mathfrak{i}$	$\mathfrak{t}$
$\mathfrak{f}$	$\mathfrak{t}$	$\mathfrak{t}$	$\mathfrak{t}$	$\mathfrak{t}$
$\mathfrak{u}$	$\mathfrak{u}$	$\mathfrak{u}$	$\mathfrak{i}$	$\mathfrak{t}$
$\mathfrak{i}$	$\mathfrak{i}$	$\mathfrak{i}$	$\mathfrak{i}$	$\mathfrak{t}$
$\mathfrak{t}$	$\mathfrak{f}$	$\mathfrak{u}$	$\mathfrak{i}$	$\mathfrak{t}$

$\neg$
$\mathfrak{t}$
$\mathfrak{u}$
$\mathfrak{i}$
$\mathfrak{f}$

implication, suitable for determining set containment and Pawlak-like approximations [23] in the case of four-valued sets.

The reasoning within a single information source is carried out according to Definition 4. In order to gather results concerning the same proposition from different sources we use an additional fuzzy operator $\oplus$:

$$\langle \beta_1, \delta_1 \rangle \oplus \langle \beta_2, \delta_2 \rangle \stackrel{\text{def}}{=} \langle \max\{\beta_1, \beta_2\}, \max\{\delta_1, \delta_2\} \rangle. \tag{11}$$

In L_t we gather results concerning the same head from different sources by using the disjunction $\vee_k$ w.r.t. knowledge ordering, defined in Table 3.

Table 3. Truth table for $\vee_k$.

$\vee_k$	$\mathfrak{f}$	$\mathfrak{u}$	$\mathfrak{i}$	$\mathfrak{t}$
$\mathfrak{f}$	$\mathfrak{f}$	$\mathfrak{f}$	$\mathfrak{i}$	$\mathfrak{i}$
$\mathfrak{u}$	$\mathfrak{f}$	$\mathfrak{u}$	$\mathfrak{i}$	$\mathfrak{t}$
$\mathfrak{i}$	$\mathfrak{i}$	$\mathfrak{i}$	$\mathfrak{i}$	$\mathfrak{i}$
$\mathfrak{t}$	$\mathfrak{i}$	$\mathfrak{t}$	$\mathfrak{i}$	$\mathfrak{t}$

Example 9. As a simple example of reasoning consider the following rule:

$$\big(\mathit{severe\ sore\ throat} \wedge \mathit{painful\ swallowing} \wedge \mathit{headache} \wedge \mathit{fever} \wedge \mathit{chills}\big) \rightarrow \mathit{tonsillitis}. \tag{12}$$

Assume that we ask two doctors for a diagnosis based on (12). The results of examination of symptoms are provided in Table 4, where for all variables we assume the four-valued interpretation given in Example 6. Based on the

[6] In particular the chosen thresholds should make clauses for $\mathfrak{t}$ and $\mathfrak{f}$ mutually exclusive.

Table 4. An example of evaluation of symptoms for rule (12).

	sore throat	*painful swallowing*	*headache*	*fever*	*chills*
doctor 1	$\langle 0.8, 0.1\rangle \rightsquigarrow \mathfrak{u}$	$\langle 0.9, 0.1\rangle \rightsquigarrow \mathfrak{t}$	$\langle 0.2, 0.8\rangle \rightsquigarrow \mathfrak{f}$	$\langle 0.7, 0.3\rangle \rightsquigarrow \mathfrak{u}$	$\langle 0.8, 0.1\rangle \rightsquigarrow \mathfrak{u}$
doctor 2	$\langle 0.9, 0.1\rangle \rightsquigarrow \mathfrak{t}$	$\langle 1.0, 0.0\rangle \rightsquigarrow \mathfrak{t}$	$\langle 0.8, 0.2\rangle \rightsquigarrow \mathfrak{t}$	$\langle 0.9, 1.0\rangle \rightsquigarrow \mathfrak{t}$	$\langle 1.0, 0.0\rangle \rightsquigarrow \mathfrak{t}$

examination of symptoms and Definition 4, the first doctor decides that belief and disbelief for tonsillitis are $\langle 0.7, 0.3\rangle$ and the second doctor decides that these values are $\langle 0.9, 1.0\rangle$. Gathering those results gives $\langle 0.7, 0.3\rangle \oplus \langle 0.9, 1.0\rangle = \langle 0.9, 0.3\rangle$, which is interpreted as $\mathfrak{i}$ by paraconsistent fuzzy reasoning. Note that according to the first doctor, the conjunction of symptoms in rule (12) results in $\mathfrak{f}$. Therefore the first doctor decides that the conclusion (tonsillitis) is $\mathfrak{f}$, too. According to the second doctor, the conjunction of symptoms, thus also the value of tonsillitis, is $\mathfrak{t}$. Gathering those results by $\vee_k$ results in $\mathfrak{i}$, as in the case of paraconsistent fuzzy reasoning. $\lhd$

4.2 Rule-Based Reasoning: A Case Study

Let us now illustrate a more advanced rule-based reasoning, where the conclusions of rules are assigned belief and disbelief degrees on the basis of premises and a medical knowledge base.

Medical knowledge is defined in [8] as a fuzzy relation R, linking the set of symptoms with the set of diagnoses "which reveals the degree of association and the degree of non-association" between symptoms and diagnoses. The methodology proposed in [8] involves three steps:

1. determination of symptoms
2. formulation of medical knowledge based on fuzzy relations
3. determination of diagnosis on the basis of composition of fuzzy relations.

In the approach of [8] and also [29] the diagnosis is evaluated on the basis of a certain distance from symptoms to a given disease. The reasoning in [8] is based on the following rule:

IF: the state of a given patient P is described in terms
of a description of symptoms A
THEN: P is assumed to be assigned diagnosis in terms (13)
of a description of diagnoses B, through
a medical knowledge database.

Medical knowledge is given as a paraconsistent fuzzy relation relating symptoms to diagnoses. For example, indications for and against chosen diseases are provided in Table 5, based on [8].

The degree of belief and disbelief of a diagnosis $d \in D$ is calculated as:

$$\left\langle \max_{s\in S}\left\{\min\{\beta_A(s), \beta_R(s,d)\}\right\}, \min_{s\in S}\left\{\max\{\delta_A(s), \delta_R(s,d)\}\right\}\right\rangle, \tag{14}$$

Table 5. The relation between symptoms and disease.

	viral fever	malaria	typhoid
temperature	$\langle 0.4, 0.0\rangle$	$\langle 0.7, 0.0\rangle$	$\langle 0.3, 0.3\rangle$
headache	$\langle 0.3, 0.5\rangle$	$\langle 0.2, 0.6\rangle$	$\langle 0.6, 0.1\rangle$
cough	$\langle 0.4, 0.3\rangle$	$\langle 0.7, 0.0\rangle$	$\langle 0.2, 0.6\rangle$

where S is the set of symptoms, D is the set of diagnoses, A is a (paraconsistent intuitionistic) fuzzy set describing symptoms of a patient, R is the (paraconsistent intuitionistic) fuzzy relation relating symptoms to diagnoses, and $\beta_X(\bar{x})$ ($\delta_X(\bar{x})$) stand for the degree of belief (disbelief) that $\bar{x}$ satisfies X.

Assuming the set of symptoms $\{temperature, headache, cough\}$ and the set of diseases $\{viral\ fever, malaria, typhoid\}$, (14) gives rise to the following rules:

$$viral\ fever \leftarrow temperature, headache, cough. \tag{15}$$

$$malaria \leftarrow temperature, headache, cough. \tag{16}$$

$$typhoid \leftarrow temperature, headache, cough. \tag{17}$$

Evaluation of conclusions of rules (15)–(16) is given by instantiating formula (14). For example, the degree of belief and disbelief for *malaria* is given by

$$\Big\langle \max\Big\{ \min\{\beta_A(temperature), \beta_R(temperature, malaria)\}, \\ \min\{\beta_A(headache), \beta_R(headache, malaria)\}, \\ \min\{\beta_A(cough), \beta_R(cough, malaria)\}\Big\}, \\ \min\Big\{ \max\{\delta_A(temperature), \delta_R(temperature, malaria)\}, \\ \max\{\delta_A(headache), \delta_R(headache, malaria)\}, \\ \max\{\delta_A(cough), \delta_R(cough, malaria)\}\Big\}\Big\rangle,$$

where A is the (paraconsistent intuitionistic) fuzzy set describing symptoms of a patient and R is the relation provided in Table 5.
Table 6 provides examples of symptoms for three patients as well as the evaluated degrees of belief and disbelief for the considered diagnoses and their four-valued interpretation, as considered in Example 6.

Table 6. Symptoms and diagnoses for patients.

	patient 1	patient 2	patient 3
temperature	$\langle 0.8, 0.1\rangle \rightsquigarrow \mathfrak{u}$	$\langle 0.1, 0.9\rangle \rightsquigarrow \mathfrak{f}$	$\langle 0.5, 0.6\rangle \rightsquigarrow \mathfrak{i}$
headache	$\langle 0.6, 0.5\rangle \rightsquigarrow \mathfrak{i}$	$\langle 0.1, 0.9\rangle \rightsquigarrow \mathfrak{f}$	$\langle 0.4, 0.2\rangle \rightsquigarrow \mathfrak{u}$
cough	$\langle 0.7, 0.2\rangle \rightsquigarrow \mathfrak{u}$	$\langle 0.0, 0.9\rangle \rightsquigarrow \mathfrak{f}$	$\langle 0.5, 0.4\rangle \rightsquigarrow \mathfrak{u}$
viral fever	$\langle 0.4, 0.1\rangle \rightsquigarrow \mathfrak{u}$	$\langle 0.1, 0.9\rangle \rightsquigarrow \mathfrak{f}$	$\langle 0.4, 0.4\rangle \rightsquigarrow \mathfrak{u}$
malaria	$\langle 0.7, 0.1\rangle \rightsquigarrow \mathfrak{u}$	$\langle 0.1, 0.9\rangle \rightsquigarrow \mathfrak{f}$	$\langle 0.5, 0.4\rangle \rightsquigarrow \mathfrak{u}$
typhoid	$\langle 0.6, 0.3\rangle \rightsquigarrow \mathfrak{u}$	$\langle 0.1, 0.9\rangle \rightsquigarrow \mathfrak{f}$	$\langle 0.4, 0.2\rangle \rightsquigarrow \mathfrak{u}$

Analyzing the results one can note that:

- malaria is the most plausible indication for the first patient
- none of the considered diseases fits symptoms of the second patient
- there is a weak suggestion of malaria as well as a weak suggestion for typhoid for the third patient.

The four-valued analysis of the above results is based on rules (15)–(16). These rules are interpreted as implications:

$$(\mathit{temperature} \wedge \mathit{headache} \wedge \mathit{cough}) \rightarrow \mathit{viral\ fever} \tag{18}$$

$$(\mathit{temperature} \wedge \mathit{headache} \wedge \mathit{cough}) \rightarrow \mathit{malaria} \tag{19}$$

$$(\mathit{temperature} \wedge \mathit{headache} \wedge \mathit{cough}) \rightarrow \mathit{typhoid}. \tag{20}$$

Based on truth values of L_t given in Table 6 we can conclude that assuming the interpretation provided in Example 6, all conclusions are, in fact, unknown.

In the case of the first patient the implications (18)–(20) reduce to $(\mathfrak{u} \wedge \mathfrak{i} \wedge \mathfrak{u}) \rightarrow \mathfrak{u}$, which, according to Table 2, is $\mathfrak{u} \rightarrow \mathfrak{u}$, i.e., $\mathfrak{u}$. This means that conclusions and the validity of implications are unknown.

In the case of the second patient the implications (18)–(20) reduce to $(\mathfrak{f} \wedge \mathfrak{f} \wedge \mathfrak{f}) \rightarrow \mathfrak{f}$, which is $\mathfrak{t}$. Here all the conclusions are $\mathfrak{f}$ meaning that none of the considered diseases fits the symptoms. Moreover, the implications (18)–(20) are $\mathfrak{t}$ which indicates that the reasoning is sound.

The qualitative interpretation of the results for the third patient are similar to the interpretation for the first patient.

Altogether, the results indicate that the interpretation of Example 6 might be too restrictive when used in this reasoning. However, once assumed, it allows us to interpret the results on the basis of a solid background.

Of course, the choice of the fuzzy interpretation is flexible so the previous constraints can be relaxed, e.g., along the lines of interpretation defined by formula (10) with suitably chosen thresholds. For example, taking $\epsilon_f = \epsilon_t \stackrel{\text{def}}{=} 0.6$ and $\epsilon_u = \epsilon_i \stackrel{\text{def}}{=} 0.2$ in formula (10), one can conclude that the diagnosis for the first patient is malaria, since now $\langle 0.7, 0.1 \rangle$ is interpreted as $\mathfrak{t}$. In this case the implication (19) reduces to $(\mathfrak{t} \wedge \mathfrak{t} \wedge \mathfrak{t}) \rightarrow \mathfrak{t}$, i.e., to $\mathfrak{t}$, which shows soundness of the reasoning.

Remark 10. Observe that the semantics of implication based on Table 2 makes the implication $\mathfrak{t}$ also when $\mathfrak{t}$ is derived from $\mathfrak{u}$ or $\mathfrak{i}$. In rule-based reasoning one would frequently prefer to make such implications $\mathfrak{f}$ since deriving conclusions from unknown or inconsistent knowledge leads to forms of non-monotonicity. To "block" derivations based on such implications, one can use other implications, like $\Rightarrow$ shown in Table 7. Other implications are also considered in paraconsistent rule-based reasoning (see [22]). $\triangleleft$

Table 7. Truth table for the implication $\Rightarrow$.

$\Rightarrow$	$\mathfrak{f}$	$\mathfrak{u}$	$\mathfrak{i}$	$\mathfrak{t}$
$\mathfrak{f}$	$\mathfrak{t}$	$\mathfrak{t}$	$\mathfrak{t}$	$\mathfrak{t}$
$\mathfrak{u}$	$\mathfrak{u}$	$\mathfrak{t}$	$\mathfrak{u}$	$\mathfrak{f}$
$\mathfrak{i}$	$\mathfrak{i}$	$\mathfrak{i}$	$\mathfrak{t}$	$\mathfrak{f}$
$\mathfrak{t}$	$\mathfrak{f}$	$\mathfrak{f}$	$\mathfrak{f}$	$\mathfrak{t}$

5 Conclusions

In the current paper we addressed the problem of qualitative interpretation of fuzzy-like paraconsistent reasoning. We proposed a methodology in which suitably chosen qualitative four-valued logics are used to serve this purpose. The interpretation is applied in all steps where fuzzy-like reasoning is carried out. Therefore the resulting interpretation provides the logical value which would be obtained by applying the four-valued reasoning.

In this paper we did not consider first-order $\mathcal{C}(L_t)$. To obtain such logics one has to extend the language as in the case of the classical first-order logic. In the case of $\mathcal{C}(L_t)$ it is reasonable to define the semantics of quantifier $\forall$ by generalizing the conjunction $\wedge$ and the semantics of $\exists$ by means of generalizing the disjunction $\vee$:

$$\forall x[P(x)] \stackrel{\text{def}}{=} \underset{x \in U}{\text{GLB}^t}\{P(x)\} \text{ and } \exists x[P(x)] \stackrel{\text{def}}{=} \underset{x \in U}{\text{LUB}^t}\{P(x)\},$$

where the superscript ^{t} indicates that the greatest lower bound (GLB) and least upper bound (LUB) are computed w.r.t. truth ordering.

In future we plan the research on employing approximations in the spirit of Pawlak [23, 12] but also taking into account the approach of [31].

Similar methodology can be applied to interpret the traditional fuzzy reasoning by interpreting fuzzy values qualitatively and tracking the reasoning using a suitably chosen qualitative logic.

Acknowledgments. This paper has been supported in part by the MNiSW grant N N206 399134.

References

1. Amo, S., Pais, M.S.: A paraconsistent logic approach for querying inconsistent databases. International Journal of Approximate Reasoning 46, 366–386 (2007)
2. Atanassov, K.: Intuitionistic fuzzy sets. Fuzzy Sets and Systems 20, 87–96 (1986)
3. Belnap, N.D.: How a computer should think. In: Ryle, G. (ed.) Contemporary Aspects of Philosophy, pp. 30–55. Oriel Press, Stocksfield (1977)
4. Belnap, N.D.: A useful four-valued logic. In: Eptein, G., Dunn, J.M. (eds.) Modern Uses of Many Valued Logic, pp. 8–37. Reidel, Dordrecht (1977)
5. Bolc, L., Borowik, P.: Many-Valued Logics, 1. Theoretical Foundations. Springer, Berlin (1992)

6. Bolc, L., Coombs, M.J. (eds.): Expert System Applications. Springer, Heidelberg (1988)
7. Brandon, M., Rescher, N.: The Logic of Inconsistency. Basil Blackwell, Oxford (1978)
8. De, S.K., Biswas, R., Roy, A.R.: An application of intuitionistic fuzzy sets in medical diagnosis. Fuzzy Sets and Systems 117, 209–213 (2001)
9. Dubois, D., Hadj-Ali, A., Prade, H.: Fuzzy qualitative reasoning with words. In: Wang, P.P. (ed.) Computing with Words, pp. 347–366 (2001)
10. Dubois, D., Lang, J., Prade, H.: Fuzzy sets in approximate reasoning, part 2: logical approaches. Fuzzy Sets and Systems 40(1), 203–244 (1991)
11. Dubois, D., Prade, H.: Fuzzy sets in approximate reasoning, part 1: inference with possibility distributions. Fuzzy Sets and Systems 40(1), 143–202 (1991)
12. Dubois, D., Prade, H.: Putting rough sets and fuzzy sets together. In: Słowiński, R. (ed.) Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory, pp. 203–232. Kluwer Academic Publishers, Dordrecht (1992)
13. Dubois, D., Prade, H.: Fuzzy Sets and Systems. In: Fuzzy Logic CDROM Library, Academic Press, London (1996)
14. Dubois, D., Prade, H.: What are fuzzy rules and how to use them. Fuzzy Sets and Systems 84, 169–185 (1996)
15. Ebrahim, R.: Fuzzy logic programming. Fuzzy Sets and Systems 117, 215–230 (2001)
16. Fitting, M.C.: Bilattices in logic programming. In: Epstein, G. (ed.) 20th International Symposium on Multiple-Valued Logic, pp. 238–247. IEEE CS Press, Los Alamitos (1990)
17. Ginsberg, M.: Multivalued logics: a uniform approach to reasoning in AI. Computational Intelligence 4, 256–316 (1988)
18. Guglielmann, R., Ironi, L.: The need for qualitative reasoning in fuzzy modeling: robustness and interpretability issues. In: Proc. of 18th International Workshop on Qualitative Reasoning, pp. 113–120 (2004)
19. Innocent, P.R., John, R.I.: Computer aidded medical diagnosis. Fuzzy Sets and Systems 162, 81–104 (2004)
20. Kruse, R., Schwecke, E., Heinsohn, J.: Uncertainty and Vagueness in Knowledge Based Systems. Numerical Methods. Springer, Heidelberg (1991)
21. Małuszyński, J., Szałas, A., Vitória, A.: A four-valued logic for rough set-like approximate reasoning. In: Peters, J.F., Skowron, A., Düntsch, I., Grzymała-Busse, J.W., Orłowska, E., Polkowski, L. (eds.) Transactions on Rough Sets VI. LNCS, vol. 4374, pp. 176–190. Springer, Heidelberg (2007)
22. Małuszyński, J., Szałas, A., Vitória, A.: Paraconsistent logic programs with four-valued rough sets. In: Chan, C.-C., Grzymala-Busse, J.W., Ziarko, W.P. (eds.) RSCTC 2008. LNCS, vol. 5306, pp. 41–51. Springer, Heidelberg (2008)
23. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
24. Prade, H.: A quantitative approach to approximate reasoning in rule-based expert systems. In: [6], pp. 199–256 (1988)
25. Rescher, N.: Many-Valued Logic. McGraw-Hill, New York (1969)
26. Shortliffe, E.H.: Computer-Based Medical Consutations: MYCIN. Elsevier, Amsterdam (1976)

27. Shwe, M.A., Middleton, B., Middleton, D.E., Henrion, M., Horvitz, E.J., Lehmann, H.P., Cooper, G.F.: Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base. Methods of Information in Medicine 30(4), 241–255 (1991)
28. Sunderraman, R., Wang, H.: Paraconsistent intuitionistic fuzzy relational data model (2004), http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0410054
29. Szmidt, E., Kacprzyk, J.: Intuitionistic fuzzy sets in some medical applications. In: Reusch, B. (ed.) Fuzzy Days 2001. LNCS, vol. 2206, pp. 148–151. Springer, Heidelberg (2001)
30. Urquhart, A.: Many-valued logic. In: Gabbay, D.M., Guenthner, F. (eds.) Handbook of Philosophical Logic, vol. 3, pp. 71–116. Reidel, Dordrecht (1986)
31. Vitória, A., Szałas, A., Małuszyński, J.: Four-valued extension of rough sets. In: Wang, G., Li, T., Grzymala-Busse, J.W., Miao, D., Skowron, A., Yao, Y. (eds.) RSKT 2008. LNCS, vol. 5009, pp. 106–114. Springer, Heidelberg (2008)
32. Vojtáš, P.: Fuzzy logic programming. Fuzzy Sets and Systems 124, 361–370 (2001)
33. Zadeh, L.: From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions. Int. J. Appl. Math. Comput. Sci. 12(3), 307–324 (2002)
34. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 333–353 (1965)
35. Zadeh, L.A.: Outline of a new approach to the analysis of complex system and decision processes. IEEE Trans. Syst. Man. Cybern. SMC-3, 28–44 (1973)
36. Zadeh, L.A.: Fuzzy logic = computing with words. IEEE Trans. on Fuzzy Systems 4, 103–111 (1996)

i fr i

ws

In o o r Sc nc o ca o Sc nc
J r ona 0 o an
R bieg ipip w w p

A st t In a r w r n conc o an c ron c x con a on or o og ca roo a o roo a x con r rn o ra ona ng c on o a wor n o a an an n c ona ffix on ff r nc o r ng c cr on a an c ron c r o rc a a o ana x a a x con w n wr n x ana or n r or w o ra on gra o c

x con o n c ona ana r R a r ca o n a on or roo a x con In con ar o a r w cr roc o a o a c con r on o a a ro ana r n o a or a con r on conc rn a or n c ng ar o c no n a c an r r ar wo or o og a n r w c ar w o ac ag o or o og ca n or a on a o x In r n or ow r an g n ra a on a o o n c on or nn r roo a rna on no a a a w r x con o roo a r w o a n co r x con w c can r no on or n c on ana a o a ca on o wor or a on cr on

o ds roo a x con wo or o og a o a c or o og ca ana x con co r on

n uc i n

t st t t s s l s t t s l d t d t s l t s t s t t s l t ds t t l l s s t t ll d w d ff t st d ds l l t d s t l s s t ds d t t d t l t l l l s s d t w w d s stl s t d t tw l ll t t d ts t st d t d st d t l l w d d s t l st s t s t w l d s s d d s - st t t s - *d* s t l l d s s lts t t t t l l l z w s s t d l s l l d t t l t st w d t t s d s t t d st s s w dd t l t t s ld d d t ll w ld l s l

. a n a an . y e a s. es s h , , . 2, 2 .
◯ n e e a e n e e e 2

J Ra ga n w a

s d l t s d t l s s l s
t t t l l l s s l s s t d s t 2
s t s t t s d l t l l s
w t s d s d l s l l st t
t s d l s l d l t d t l d l l s
l d t 2 s t s ss t ll t st s
l d ll d s s t l t s l d t d t t
t l l s s d s t s s
s t w d s t ss ld t l tw l l
l d l s s d s t t l d
t st d s s t l ts w s d t t l
d t t l z s t l s s s s t 2
l t t d t s t t l ts s s t d
t ll w st s t s s t

s t s s ll w d s l
s t t ll s t s z d w d s t t st s d t
s d d s
ll t ll st s l d t t l lt t s
d s t
d t l t

s s lt w t d t s t t t t w l t l
t st s d d t ls t t l lt t s t s s d t l
l t Ż Ł t s s t d l w
s l s t t d d t l d r t w
d tw s ts st s t t s t t l s t t s ll t l l
t t s w t t l d t l d s t d st t
l t lt t N s t t t t lt t s t t t
l z t t s l N l st t s s ll
d s s ss d s t d t ls s s t d s s t

w r
w N O B d b
N B d b }
O w w

t ss s d t t t t t t l
t s t t l s s s t s t s t
t t Ż Ł t t t s t s d d
d s s t d l 2 st l s t l w
t t ll d d tt s t t d sts
d d t lt t s s s lts s st t t d t
t l t ls t s ss w ll d d s t

2 N O B d b
N B d b }

O w w

```
sus sus r
w w r
...
```

t s d l t t t t l s w ll s d t l t l l l t d s t s s t s w w s t t l st s s l s

Ai s f M p ic An ysis

t s s t w s t t s l t s l d t l s t d w d t t ld t t l l l s s d s t s s t ds t l l s s t t w l t d t st t t s s lts t t ss w t d t s t l st s t l l d s t d s ss t s s s t d s s t 2

d ss tw l st t s d t l l t s s s l w t l d s t w d t s s t t t t l l z s l d t d t s l s d t t t t t t d t l w d l s s ts t t t l d w d t l ls

1 I flec

t s t t t l l s s s t d l w d t s t d t t ll ss l d t t l t st s ss d t t t t s s t t t l s t s s w d s t t w l d l t t s l t t t d s t w d s w d s d st d s t t t s t t t t s t t l s t t t t t s t w l t t t l l st t t w d s l ttl t w l t l t l l z

l s t s l t ds t t t l s s d s t s s 22 2 t s s s t ds s l s l t d 7 t s t d t t t t t d t l t l l l s s s t s ffi t t l l t s t l st t t w d s d d d t tw ts st d t l s ffi s l

w t $_{nom.sg}$ w t $_{gen.sg}$ w $_{dat.sg}$ d
$_{nom.sg}$ $_{intr.sg}$ $_{gen.pl}$ t
zw z $_{nom.masc}$ zw z $_{nom.fem}$ zw źl $_{nom.h-masc}$
$_{1.sg.pres.}$ sz$_{2.sg.pres.}$ $_{3.sg.pres.}$
$_{3.sg.masc.past}$ l $_{3.pl.h-masc.past}$ t

t t d s t t t s w t st lt t s t d t l s t t t tw ts s t d d t s t s s ss d t t l t t l l s s ld s l w ll d d t w d t ss s s l w s t d t l st t w d s w t d ff w t l w t s t t t t t d t l d s t t l s st l tt s ll d - *d* t t t ll w d s w t l t l s st l tt s t t - w s s d ll st t t l w d d s s l d st t s d s d s st s l l t s l d t d t t l l st s 2

w t$_{nom.sg}$ w t $_{gen.sg}$ w $_{dat.sg}$ d
$_{nom.sg}$ $_{instr.sg}$ $_{gen.pl}$ t
zw z $_{nom.masc}$ zw z $_{nom.fem}$ zw źl $_{nom.h-masc}$
$_{1.sg.pres.}$ sz$_{2.sg.pres.}$ $_{3.sg.pres.}$
$_{3.sg.masc.past}$ l $_{3.pl.h-masc.past}$ t

t l l z s l s s d s s d l ss t d t t ss *o* *f o* ll s l t t d s t lt t s w s t ss s ffi t t t t l l s s s t s s l s t d s d s ll t l z w ll s t d l t s s t 2 s l w s t ss t l t s w s d s d d t ls s s t

r r

t t t w d t l s s s s t st l l t s s tw d t s d t s s l s l t t l t w d t d s t ss s l s s d w d s t t l t s w t s t w ll d d d t l ffi s t s t t t st l s s l d t t w ds l t t t w d t ss st d t d t s w t l s s s w l 5 s d s l t t s w l 5 l l t

5 l t ← l t z ← l t
l l ← l l ← l

a r a on h an or *an a c ne gen e*

l t ← l t z
l l ← l l
l t ← l t
l l ← l

d s t w d t t w t t t t l s s s t s s st l ll ts t l ffi s d d t ss t t l t s s w d t s s ld s t d d w d s s l t s s s w d s t d t ss w s d s d st 2 d s w s t *o f d d o* s t t s l s s t ws 5 d d s ss l d s s s t d w d l w ll ws s t s l t l s ffi s t l ts s w d s t ss d t l s t l t t lt t l s t t lt t s s s l d t t t l l l s s s s s w l l s ts t d t l w d t s t t l l l s t d st t s t s l t l s s t t l l w d s w s 2

ó ←
l ttl w ← w *l o*
w ← w l
l ttl w ← w *o l l b l*

t s w l t t t t d d s l s l s s t d t s s s t t s w d t t s ld t d s t s d s w d l t t s s d t l ts d s ffi s t t t l w s s t d t s t l s w t t w l s d t d s 5 t l t s w t s s l l s t w d t d lt t ss s t d s t w s t d d t t l s s w tt t t t s w t l s t s l l l l t s^2 st t t l w s ls t t t l l t s

C nc p f s ic n

t s t s w s t d s l t l s l l st t d s t t ss t t l st s d s ffi s s t l d t d tw l l l d l d s t st t l l l l s l s w t t ll w s s t w s t t tw l l d l

o no ana or o ono og ca or ono og ca r c r o wor or

J Ra ga n w a

l t w d s t st t l l t t
l s s t 2 ss t t s t
t s t t t s d t s ll ws s s t

1 T eve M r h y r ch

t t s s s w t t l l
s d t d l s s t st t t t t l t
t st S t l d s d
t l t t l t t l st t s
tt w s l s d t t s t s t d ss t t s
w l l l t t l d l d
t d d t w s t d w t t ll
l t d l st d l l l l s s d s t s s
s s d l t s l l s
l l d s t s t l l s d l d d t
l t t d l tw l l d l s s d l s st d
s t tw l l l s d s t s t l l s
s t d d t l d t t d l
tw l l d l s sts tw s t t s t l d t
l t l t s l l s t t s
w d t s d d s s l l s sts s
l tt s t l t tw l ls s d d t w tt
t s tw t s w s sts l l s t t s
w d st s d t t ll tt d d s l w w s sts
t s t s d t l w t s
s d t t w s l t s l t d w t ll l l s
w t t t s tw s t t s l s d t
s s t l t st w t t s d s t t l
d t l s s s s t d s d d t l s
t t l l l l t l s t l s d lt t
tt s l st ll d d s l t s t
st t s t t l t s d s t l s
s w tt d t t l l w s ts
t t s w d t t l
t s l ss ss

7

0 0 0

s s d s t s l s s d t t s
St d d s t l d s t ll t s d t s d s
l d ffi t ss ss s ffi s d l t s s w ll s d
d l l t t tw l ls w s d t d t t
l l s ts d t l w ll d d
t tw l ls

l l ts t t l l z t l d d s l l t t l s
z l
l l s t t w s d st d s s t d
t l st d d d st lt t s d dd t l s t
t l d s

ss t t w d t ds t t t s t s st
d s t S l ss s d t l s l s s d d s
l Us t s t w t s ts t t
d s t t s l s w l t t S d ll w d s
l t t d s s l s w ll s t t l
d s t t d l t d d d t s t st w ll t s s st s
w w ll s d t st l s t s t d t lt t s
l st t l t l st t s l tw
d st lt t s d t d t l d s t
t s t st t l t s w ll ss l s
ss w ll s t d d t l s s t

S t l *ending₁ ending₂ ending₃*

L M *form₁ form₂ form₃*

t *stem₁ stem₂ stem₃*

lt t *alternation₁ alternation₂ alternation₃*

s t l s t w ld t t ll t s d w ld t
t t l ts l

The M ex c e

s d l l t s t l t ss t t
s st t t l s s t d d t l s
d w d s t s t t s t t w
l s sts t t l s d t s ts t l
s d t s s d s t ll w d t t l
l ss s ts s d st t d t t t l t S l
7 l t s l st l t d tt s w d ll
t t t
t l tt s sts l st s s ls s t
t t d ts w d s l l d s - *d*

r d s d

s d ds t t l l ss s t t
l st s w s t 7 s s d 2 s S d s ts
w w d s d ff s st s s t ts t ll w d s
w t l s t t s t t t s t l s ffi s d
t ts l l ts tt s s l s
t t d l s l s 2

J Ra ga n w a

N w B
w d w b
b r

N w B
w w d w b
s rd
l d

l t t l tt t l ts t
ts l l d s t st d d tt
s t d d ff s t s t s d s d s
d lt l ll t s d ff t l d s
s s d s t d d w t t t w s
s l tt t t d d d d s t s lt st
lt t d t l s w t t l ls b r d
s rd l w t t t d t l l l t s t s l s
t tw l s l l s w t d s t s st d s
d t l ls w r d s u ts t t d t l l l
t s w r d s us t l d s l t s t
s s t t w t t st lt t ts t s d
s t l s d l 2 w t
s lts tw s t tt s t l

N w B
w d w b
w r
d

2 N s s s w s B s
s w d s s w b s s
s u s
sw dl t

s l tt s w l d ll w d s s lt t s
t t s s t l lt t
ts d d t d ff tw t w d s t
t st tt d t t s t s d s t tt

N B ą
d b

t

t l l ds t tt w t t s s t l s
t s w s t l t st s d s d s t w t t t tt

s l s l s s t s ll l s
t lt t s w tt d t w l
l t l l t s l s t t d l s
t tl t l l s

N w B
w w d w b
s rd

l d

5 N w B
w d w b
w r

d

N B ą
d b

t

l l s d t ls s t d s t s
lt t s t t l l s s s l t
t t s d l s t 7 s
s t d t s d l s w t l s
l s 7 d s w tw t

7 0 0 s ą r r r

p b p b p b p b p b
p b p b p b p b
p b p b p b p b ą
b
0

t t

0d d d s d d d d ą
r d ź rd ź r d ź d d d
d d d d d
d d d d
d p d b p d b p d b
p d b p d b p d b p d b
p d b p d b p d b
p d b p d b p d b u ws
bd
pr s d

t s t s t

t l w d st s d t l s d t s w
ls t t d st d t l t l s t s s
l s s d t d t 2 t l l s lt t l
w s ts t l d s tw l s l s t l
s t d l t w 22 l s d d t
l w t t tt s t lt t l 7 l s s w d
l 2

```
0 Np Np Np p p p u p
  p pd p p ą Bp Bp ą p p
  p pb p dp p pb bp p
    w 0
```

ff t

```
2 Np Np Np p p p u
  p p pd p p ą Bp
  Bp ą p p p ź pb p
  dp p pb bp p
  w
```

l st t d t w s t d w s t d s ll
t t l s s d s t s s lt l l
d s w d s s t ss l ts l w d d s
ld l l t t tw l l s d ss t l
s s d s ss d t

The ver he M ex c

t t l t l l l ts w s l d
s ll t l ss s s d t s d s d
l t t t d s w d t st d d s
w ll s s d t l s s lt d w l
t w s sts t s t lt t s d s t d s t
s ts d t l t t t w d s
s ss s t ll l ts w s s t l
d d t l l ss w s t t d t ls st s s
t s t ls l s d t d t s t
tt s s d ff s ss l s
s t t st s t l ld ss t l
s s l T d s s s t ll st t
l s 22 2 d s t t t st t d t s d
s s t t
d t t t l t t ll w st s d t s d d
ll w s s t st s l s s w ll d s d t t s
s s t

ec e I flec ffixe t t w
ll t d t l d s s ts s t s s w ll w d t
d s t l s t s d S t s ts d d
d l s s t t d lt t s ts
s d t s ts t l s ffi s s l t
l t l d s l d t t
d d t l s l s t d s s tw t s ts
w d d t s d s s t l
s t t l s ffi s s d t s l s s sts ll
ss l lt t d s s t d t ss l l l
d s s z s t s t d s s t
s s w s t t t st lt t d s s ts
t s ffi s s l t s d t s ffi s s l t
s t s ffi s s l s s s t d l 2 t s w t
t t t s t l d s t w l t s

```
2      .0  .u  w .u  .u.0     .u  .u.0
     . . w  . .   w. . .0     w. . . . .0
      .        . . w  . .
```

y he r r r e exe e he h e
he ex c t st s sts s t s s ll w d s t l
t s t s t l l ss s Us t t l l s t
w t t ll d t s st w t t s d t d
l st w d s l t l s s l s
t l t l d s t w d s s s d

```
22  N   w r                         w r
        w r                         w r   w
        w r   w                   d w r
        w r                         w r   w
    B   w r                       b w r
        w r                         w r
        w r                         w r
```

e e e exe e r r he e he
E t t ds w d s t t w t t
t l s ffi s t s t t sl s t d s t 2
t d s s t d d t ld s t t s ffi s
l s t t t l d s s d t t l
t w d *ko* *k* d s

```
2   N   w r                         w r
        w r                         w r    w
        w r    w                  d w r
```

J Ra ga n w a

w r w r w
B w r b w r
w r w r
w r w r

2 0 w w w

ec e ve e t w s t
s ffi s s t d s l l t ll d ff t ts l st d t t
w t t t t t s t w t l z d t ts
l s d d s t l d s t w s d t d l t
s t t t d d l l t t l
t st s w d t t t d t

25 w r N B d b
w r
w r

I er er e cr he h ce

l 2 ll st t s t d ff tw d l l
s t t w d s t t t l s ffi s s t d
t t t s t t t s ffi d t lw s t
t st s t l s w t t t d
ts d l d t t l s t s s d w t t d t l t l s
ts d t t t s t t s t l s t d d w l
s w d d s d t t t t
t t s s ts t t d w l
s t ts d s t - lt t s s s s
d t s lts t lt t w s l d l w t s
d l t t l l w d l l l s s
t st t s s s d t lt t s d s t
d l l l t t ss lt t s t t ll
w d d s ss l t s st ls d l s d t s
d d st s d s t s s s r d s
l r r d t ls 0 p p d d d d lt t
l s w d ts l s
t ts s st d t t t s d t t
ts t l t t s d t d d
t s t s l t w d t t t
w s s t t d t lt t d
s d t l z t ts ts d t ll w ts t
s t t s d t d s ts t t w t t t l d
t lt t s w t t d s ffi t l l w 2 t
lt t t d s s w s s t d t
lt t s t ds s t d d

s s tl t s s t s ts s t d l 25 w
l t t d ts

```
2 w r N B d b }
```

l w s t l t w t lt t s w s
t s t l

```
27 r N r B d b
N O B d b
N B d b }
```

dw

s s lts l t w s sts tw ts st
t s l t d t l t t s t s l
t d t t s l s l t
s s t l d ll ws t l s d t s s
l l t d tw s ts d d lt t s d t l s ffi s
l 2 s t t t d s t t l T d

```
2 w r r
w r N B d b }
w w w
```

t s s t 2 w s t d t l t t s
t l M t ll t w d s ll d t s s t
lt t s st t l d t t w d t s
ss t s lt t s w s l d d w l t s ts t s
s t t

```
2 r
ĆO Ć N B d b
O O N B d b
N B d b }
ą O
```

d s t d t s t t l s s l t t s t
t t d w s s l t t s t t l
t d t s l s t s t s t l s d
t d t d st s w d s d ff t d s ll t l
s ffi s t d d s d s t ll t l d
s s s d s t ll w d d *ul*
f *u* s l d *ul* *o* - *ul* l l
d s s t d t l l s s t d t
d t d *d* *u* *l* d *lo* l w w s t
tw t s d t s st t l F T ff t s
l s d Ł l s ws
t w t tw t lt t s t l s t s
s d t s t l

J Ra ga n w a

```
w
w . . . u. . . .ą .ą . . .
. . .
```

```
w
w Np Np Np p p p p pd p p Bp
p pb p dp p pb bp p ź p Np Np Np
p p p p pd p p Bp p pb p dp p
pb bp p p }
. . . u. . . .ą .ą . . .
. . .
```

l d s s s l l ts w t
s s t d t s l w d s lt
t l ts t t t lt t s d s l
t s d s s d l ss t l s s
t t s t s t t t s lt t tt
t t l d l l s
d t t t l t
s l t s s l s

l s s sts t t l s s d t s s
l ss t s s t s d s d
s t l s w ls t t t l s s tl t
l s s d l d tw l s s l s
s s t 2 t t s d l t s s t t
l d d t t tt s s l s s ss l
s t w s d ll

d d d t s t s t t t s
t s w t t s d t s s d st
s t t t l s ﬃ s w s t d t t
d d t s t t s w s t t l w d d s ts s
t l s s lt s l t
t t s s t d l 2

2 bu

```
bu s 0
p p p p p
p p p p p p p p b
r r r
s 0 ą 0

b b b b b b b
b b b b b b
ą
```

n Roo a x con or o

s d t t t s sts t s t t t s t d
t t t t l d s d s d t s t t l
s ffi s d d t s w s t s d
t s t lt t s w t t s dd t l s t
lt t t t s s d d t d lt t ts ll t t
l z t s w t t t l d s t s l s ts s
l t

s
s s b 0 r r r

p p p p p p p p
p p p p p
0
p p p p p
p p p p p p p p
b r r r }
s r r r b 0
p
p p p p p p p p p
p p p
s 0 ą 0

b b b b
b b b b b
b b b b ą

t stl

s t d d t t l s w s l d t l st
l t l t d w t t l s s s d t s d
s w s ll s s t l s t l l ss d
d t t l t d st t s w s s d s l
lt t s t t t s t w S l s w ld
t l s t t d s t t l
s s ts t l s l t t t s w l t d
ll l l t s T w d Ą t t s t s
d s s lts

d r
00 ON 0 0 N B d b
0 0 N B d b
d N d B d b
0 N 0 B d b
N N B d b }
w u u

J Ra ga n w a

5 ą

```
r r r 0

p p p p p p p p
p p p p p b
s r r r
ą 0 p p p
p p p p
p p p p p p b
s 0 ą 0

b b b b
b b b b b
b b b b ą
```

C p ssi n f s ic n

lt t t t s t d t t l t
l ss s d l t d d s l l ts d t
s ffi s d t l ts s ss t t lt t s t s t
s t t l lt t tt s s s s t
t lt t s d t l s ffi s s t d t
s st s z t l l t d
t t ss t tt d st t w d t t l s tt t
s tt s t t l d lt t s
z t l d t w s d t t ll s t d
s d s ll t s t w s st t t s ts
t l s ffi s w t t l d
t l tt s d t t t t ll l s w t d t
t s tt s s t l t t t tt s w t d t s ts
lt t s w st t d t w ll d w t
lt t tt s s lt d l st lt t tt s l d t
s ts l s w t d d t s d t s l t d ll t
t s t s ld dd d t t l s t l
t ss t s t t l t t l s
s d tt s lt t d t t s t l l
s d t t lt t tt t s lt t t ll d t
tt t l s ffi s t s l s l w s t t
t s t sl d s d l s t t s t t l
ts

```
w r w r r 0
r r
```

r N r B d b
N 0 B d b
N B d b }
0 N B d b }

0 w w w
0 w w

7 w w
w w

Np Np Np p p p p pd p p
Bp p pb p dp p pb bp p ź p
Np Np Np p p p p pd p p
Bp p pb p dp p pb bp p p }

. . . u. . . .ą .ą . . .
. . .

bu bu s

s s 0
p p p p p p p p p p p
p p b r r r

s 0 ą 0
b b b
b b b b b b b
b b b ą

s s

s b 0 r r r
p p p p p
p p p p p p p p
0
p p p p p p p p p p
p p p b r r r }
s s r r r b 0
p p p p p
p p p p p p p p
s 0 ą 0
b b
b b b b b b b b
b b b ą

t s t s st t st s t t l t d s s l
t tt s d d d d t s s l w l s s t
t l w t t t s st t st s t l z d l ss s
s t s d lt t d t s
s t t s s d t s d s t s t t t
tw ts t t s t t s ss l t t tw d ff t
s t tl t s s ts t l s ffi s lt t s
s t t l d t t s ts t lt t
t tt s t t tt s w st dd
t t t t lt t tt s w ls t t s d t
t w d st d t s lt t t ll t l
d w t s d t l s s s t tt s
w t t s d t l t d t F T
ff t

R o x con con r on

a	N^o o x	N^o o rna on a rn	N^o o In c on a rn
No n c r		0	

ls t ll l w l s d s d lt t
w l s s t l tt s
t s d t s s s s t t t w l t s st t
l 2 s ts t d ff t lt t tt s s
s tt d s s l s w t t t
lt t
S lt t s t t s t d s t st t t
w d ff t tt s l s d s t tw lt t
tt s t l lt t d st t t lt t
ts l l d ff s t s s t d
l t *lu l* t l tt l

```
N N B d b }
N N B d b }
```

t s s s w l z d t d st t
l s d ff t t l tt s l s ws t t
l t s t l tt

su s n App ic i n

t t t l t t l ts w s
d l w l w t l d s t l s d t 5

2 N ro x n rgo ng a rna on a rn n

a rn N^o	rna on	xa o n r	N^o o n r
a	n n	w gn a r	
a		a a ca n	
a		ro a arc	
a		a o	
a	n n	o n a o n a	
a		a a wa	0
a	n n	r o w n a ac	
a	c c	a c a gran o r	
a	n n	n an a nann	
a 0	n n	c ar arn a roo o	
a	0 n n	n a w	
a		go o a o c	
a	0 n n	w n a c rr r	
a		n a g	
a	0	ro a ro	
a	o	o o a o ar	
a	n n	o w rc n a r or r	

t t d d ss t t w s t d l t l t s d 5 t s d l l s 27 t s d d t l l s d 7 t s d l l s d s t l t s s tl l d s l l s d t d l s lt t w t

r q nc o x n n c on a rn

In c ona a rn	xa o n r	N^o o n r
	w gn a r	
	a a a	
	a a wa	
	a c a gran o r	0
	c ar arn a roo o	0
	n a w	
	o w rc n a r or r	
	an a a	
	G n a G noa	
0	g n a g noa	

d t s t d t s d w t t ss s s ll t s t s l s d t l st w s d ff t t w 5 t s d l l s

d t l l s d 2 t s d l l s t d s s l
t t s tw w d t ff t ss t d s d t d
s t t st t d s t t l s t t
t s ss l t s t s z t l s w d s t t
l t s l t LM l t t s *l* t ls s d
t l ts t t t l t s l t LM l
t t s *l* - s d t d s t s s s d t d d t l s
d lt t l s s t d t d l l d t 2 s
t d t ss s t s l l ts d t l s
s l l l lt t t t d t st d t d t l
ffi ss l 2 s ts l w t s s w t
t s ffi (t t l s t s d t l *l* t
t t t l z t t s ffi (s tt d t l s t l t
t s ffi t t s t
t d t LM l s

2 ($N.loc.sing.fem$ ($\rightarrow$ $N.nom.sing.fem$ LM

l ts s ws t d s t l l ts w w t
t t t t s w ll s t lt t s s
sl tl d t t t d t l l st d s t l s
t l st s s t s t s d t t s s t
s ss d t t t l w ll s t l d t t w d
w tt t t t s t l s l t t s s l s s ts
s t t w t t l d s t s d d t t t d
d t st t l s d d ff t t ls t t
l s s

lt t l w s lt s d t l d s t d t ts
t s s d s l ll st t d t d lt t l s d
t d ts l t t t w s d s d s l l
l z l t t l z w w ld t t l w ld
ss t t t w ld l t s t s z

s t t l ts st s l
s t t l s ffi s l
s ts t l s ffi s t l s
s t t lt t s l s
s ts t lt t s t l

d t wl d t st t l l z s l s
t d d s d t l d l l w s t l ll ws
t t d t s t l l s s s t s s
s d t l l t t t l w ld t l t
d t l l s 5 l sl l t t w d s
t l ts t t s w t w d t l z ld s t
l d s t t s t w d ts t t ll t s ld st ss d

t t t w d t l s s d s t s s s ld
t t t s t
ll t l ts ld t s s l t t t ll
t s s t d t w w ds d t t
t l w w d t st t w ld t d t t t
t l t s t w s ts t l s ffi s s t d
ts d t t l t st s d t t t t s ffi
d d t s s t d t t t t ld t s d
d t lt t l ss t s s lt t t l s t t
l l st ss l l s s d l d t l l s
st t s s s s l ld t d

u y

ld t l ts w s d t tw l l l
d l l d t d s t w t t l t d l
s s t t l s t l t d l d t
t t l w ll t t t t s ss t t st t
w d s s l d t d t s st t d d s t t s
t t
ss t t t s l t l l s t s
ll t t t t l ld t d d w ll
w t s ts w d s d t t w t s ts
t l s ffi s s d t t t w s t d t l
w w s t d tw
ss l l t s w s t d t t s s s t l t
t l l l z t t s d l t w d
t l t st d t t ld d t
t lt t s l s t s t l tt U t l w t
l ts s d t l d s t l s d t w
t s d t s l d t t t l s w ls t t d
t l d d t l d s t w ld ll w t l t
t s s l l st t w ds t ts

c e e w ld l t ss t t sz
ws t t l w ld t ss l w t t s l
w t t d s s t d t s

f nc s

J S onc c a own ow n or ac or o og c n o row
w r ac D r a on n r a ar o n o awn c wa n
w r ar aw go
oo G ga n S or o og ar oo o or o og 0

oo G In r n r on x a In c on an S or o og
o ar oo o or o og
ro a a c Ra w owar a S o G n ra r o
or o n ac c oo or o In In g n In or a on S I a
ng ng ro o Sc nc o r Sc nc 0 00
narow a n In r n In c on ng D r a on n o ar oo
o or o og 00
Gr c ñ a r c own w o o c w w c n o c n
an rac J o nawc a a Naro ow o c
a n c r g ana a or w or o og c n c a a o
go c n ca R or In o aw In or a o a
Na ar awa 00
a an R a ono og ca r an n a ran c r In n
g c Soc o r ca ng an oo S x nn a ng N w
or D c r 0
a n a R ar no a R In S arc Jon
o a c Na ra ang ag ar ng orwoo c r

0 o nn wo or o og G n ra o a ona o or
or or R cogn on an ro c on D n n r
Ra ga n w a J o aw ngw c n a o a c n go ana a ora
or o og c n go R ora n J ow 0 00
Ra ga n w a J or a o o o no na r a on In an
ang ag c no og a a a ng or o r Sc nc an ng c ro
c ng o n ang ag c no og on r nc o na r 00
awn c wo o na S o o o na 00
Ra ga n w a J w na r wac a o a or
o og ng ar a 00
Ra ga n w a J R o owar a o ar o a c na r
o arg o or ora In o a R a c a J ra J G Ż g
In ga on n o or a S a c ng c on r on o o r
ro an on r nc on or a D cr on o S a c ang ag DS I
a o a n r No r 0 00 00
Ra ga n w a J or a n o r wac w o R c own
r o n D n w r ar aw 00
Sa on Gr c o o o R S own gra a c n
a o go a ow c na 00
Sa on Św S a n a w c n go a o go awn c wo
Na ow N ar awa
S a ran o a c na ana a na o go na o aw
Sc e a yczneg n e a e g Jana o ar go D n w r
ar aw ar awa
o ar J a o a N ar awa
0 o ar J Sc a c n n *a e g* o c or w ra ow c N
ar awa
o or a rac ca oo or or o og ca na o o
In o o r c o S ro anow In g n In or a on ro
c ng an n ng IIS II 0 0 S r ng r rg 00
o o R wna o a ana n or o og c n w o
ro c n Na or a a o owan a In n r a ngw c na
a c a c na awn c a I ar awa 00

r i f z r y i r

t tl t [1,2] d t t l [2]

[1] D ar n o In or a on c no og ac o a a c
rg nc S a n r
on S r rg nc an
g y t y h
[2] D ar n o o r ng c an r c a In g nc
ac o a a c an o r Sc nc
a c w c n r
ow a o na o an
et i e p

A st t In a r w a r r a o or o og o ang ag In a n an o r agg na ang ag o ng x wor corr on o n nc n non agg na an g ag or o og ca roc ng r or a cr c a o ra on n a o a c roc ng o a roac or o or o og n r o or o ac c an or o on c r r n R o a o a c or o og ca ar ng or ang ag ro og n a on o ro

o ds na ra ang ag roc ng N agg na on or o og ang ag x con ro og oo gn or o on c or o ac c

n uc i n

w d s t l t t t d ss t l l s s l s s t d t s s t t t t d l t t l l ss t ls t Uz l l l s s t ss w d t ts s ll st l ts ll d s 7 2 t s w d s l l s t l l t d l w s s t t Uz w ds t ts d s ffi s s d w tt Uz l w s t t s ll l s l d t t d t l t l s t Uz Uz st t st s s l t

Uz l s l t t t s l w s w ds t d dd ffi s t t t s s l s w d ts t w d w w d dd ffi t t s t d t d t dd t ffi d s s s s s l Uz w d s d t w d s t

. a n a an . y e a s. es s h , , . , 2 .
◯ n e e a e n e e e 2

s l t t l l t t l s t s t t
l s s t t s d d s s d s t s s l t l s l
t l l l s s s l t d t t w ds w t
ss s l l ss s t l t
t t t ss t Uz l t l ts l s
t t d s s st s w t t l l t
t s l t t l s s l s d t Sl
l s

E M L *k lol l d ?* t s w ld t

l s ffi s d t t t s s ffi s d ff t
d s l l l s s t l s l s
l t l t l s l w ds s l s l
t w d l d t t s d d s s d tw
s l d t t st

t s w ld t t t t d s

`r rb } }`

l l s Uz s t lw s t t s l s s s t
s ss l t s t w d s w s t s dd t l d s
t t t l l l t t d t w
t s s t t l t st t t l t t
l lt t s l t s t t t t t lt t t
s s d d t w ll t t ll w s l t

`r u pp }`

w l s s d st s t t d t l st s l t
t l t s ld t t
l ll st t s t d t st s s s
l d t **z e h c r er** s t d t s
s t ls ll d d t st d t
ll ss l l l s s lts t d t t

t t s t s w w ll d s t d s d t s
l t s

npu n Ou pu

t l ts s t l st w ds[4] t l t t t l
`.t t`

R or o og ca oo or ang ag wa gn an
n Ga ra a a o a a r o c on an ra on
o o co a or

In ang ag a n an o r agg na ang ag o ng x
wor corr on o n nc n non agg na ang ag

```
  .     pp     d
   d  r.          d d
b r  d     ?      d     u   ?
... ... ...
```

t w ll t l st ll ss l l l s t t s w d w t
s d d s t s

```
  rd                    r                u     pp   }
      r       rb            }                }
```

t s l l s tw ss l s t t s s d s
2

```
  rd       d  r         r       rb            }
                 }  d   r   p s  s  p      s     p rs
     p ur   }
  rd  b r  d       b r  r       rb         }              }
       d      p s  s  p       s      p rs    s   u  r}
        u s     su     }
```

ic n

d t st t t d t l st wl d ss t l l l s s s t l w ll s t l s t t t d t t st d t s dd l ls t s t s l t t d t l t t d t l s s t l s l l t t s s l d l ss d d t t s l l t t s ll w l s *ou*

```
                rd         rd }.
```

l s s t t t t l t l s l t t s l t (*o d* s t d *o d* *ou* t w t t s s l l w d d t t l ts s ll ws

```
...
  urs  .     %        r
       .     %    w
       .     %     pp
...
```

t l t t s t s d t t t l l t s ss t d w t dd t l ts t s d tl t l t w d ts l t l l t t t t d t d t t d t l t t ts t t d s t ll

l t d t t w d s s l t l t d t
t w ll t t s t t l

```
p pr u b r p rs
           rd   p pr   rd  Nu b r  p rs   }.
        %p pr s   ds   r   p rs     pr   u
p pr      s   u  r    p rs     .
p pr s    s   u  r    p rs     .
...
p pr u  r   p ur     p rs     .
```

l s s ts s l st d t d

1 Irre r r

l w ds t s t t t s d l s ll l s
s d l st t s t l w d d t t s
s lt t t w d
l

```
rr  u  r pr   u    s                    p ss ss     s
```

t t s t s w d t st l s t t l w d l st

s C ic s f ZM M ic s

l l s t d d t st d t st
w 5 t s d t t st s lt t t t w ll
t t d s st ll t d lt t s l t t t s l
s w t w d s d s t l st l w ds t ll
w d s l t s t l l s ll l s
t s s t w s t t ll l d t s t l l
s R
d t d 0 s s d t t t t R s st s d t
ds t t t l u b. s

```
d          pr        s          Os N         u b. s
           s s      Os N
        pr             u d    Os N
         p   Os N        r  d     d
         r  d s           d           r
             r
        pr  Or        s                  r
        pr p r  s               r
           s     d        .
```

d t r d s s r s d s t s t
t l d t t t l sts w ds s t s ld s t d
w t

```
s         . %  .
s         . %
s         . %  ?
s         .
%   r  r     r s    s  0    pr d    w     .
```

r d s d t s d t st d t s t d t
t s t
w d s ld s t d w t s d st t
l Uz t s Uz l t s tw d t s
d w l t l tt s ll w d t st S d
d t tl ss Uz w ds s t s t s w st t
st t S II d 5 s s t ll d t ll w d t

```
s u b   r   . %    p s r p
s u b   r    b  w          .
u b   r   N w
        b  w       0
        N  w  s     .
u b   r
        b  w              .
u b   r     .  %
```

d t pr p r s s s ts t l st
s l s s ld st t t d t t l st
t s d t w ll s w t l st l sts w
l st t s t s s lt s d t s d t st d w ll
t t s lt t l l s
t s s l t d d t l l l s s t Uz l t
tw t l t d ts t t d l s s

t t l s s l s t t l t d t d
s ffi s t d d w s ffi s tt d t w d
d t t t s t d w d
ds d d ff t t s d t t t s d
s ffi t t tt d t w d t t w ll tt d
t w ds t t
2 l s s l s t t s s ffi s
d t s t s w d t s ffi
l t s ffi d t t *d* t ll - t
t t ss l s -*k* - - d t t l st s t
t w d w *k* d t s t l

1 M r h c c z e

t Uz l t st ts t tt t s s ffi s t s w ds d t t l t t s st ts w s t ll w t s d t t d s t t t t tt d

s d d t l ss s ffi s t tw ts t l s ffi s d d t l s ffi s t l s ffi t t s d d s t tt d t ll t w ds t t t t l s ffi d s t t t w d t t t s tt d t t l ss s s t l t s t w d s s t ss ss t s d t l s ffi s t t w d t t t s tt d t t s w w d s ffi ls t t w d l t d t l s ffi s tt d ls t w ds t t d t l s ffi tt d t s s l w d t l ll w ds t t Uz l s l d t l s ffi s t s st d w s t l t st w d l s d s l st d t l s t d t s t t t w ds ll w t s ffi d st t t d t s t t t w w d t d t t s ffi s tt d

D r a ona ffix or wor ca gor

FF T	T	E M L T
- - d t - -*d* -*d*	(*ou*	s s s s d s d d t d t t d t t d t d
-(-*l* d t - *k*		s s s l s l d s d
- - - d t (- -*l*		s s s s s z s s z s l { s t s w t d
- - - -*d* -*d*	(*ou*	***t t t*** s s s s s d s d t t t ***ll t*** s ffi l s s s s d s s d t t

D r a ona ffix or wor ca gor con

FF T	E M L
T T T	
	t s l ffi l s s s l s s l s s l s s l s d s l s d t t t ***l*** s t ffi l s s s t s s t s s t s s t s d s t s d t t t ***r l*** s s s s s d s d t t t t t t t
- - *-(* *(* *-(* *l k*	***ll t*** s s s s z s z s l s l s l s l s s l z s l z s l l t w t ***l*** s t s t s t s s t z s t z s t l t w t ***t*** s t s t s t s s t z s t z s t l t w t ***r l*** s s s s s s z s z s l t t t t t t t t t

D r a ona ffix or wor ca gor con

FF T E M L
T T
T

- ∅ -(- -*l*	***l*** s t s t z s t s z s t l w t ***t*** s t s t z s t s z t w t ***r l*** s s s s s s z s s z s l t t { s t s w t t t
-*d*	**er** **reflex ve** z z s l **e r ve** d t s t **e** w w s **e e** ll d s
-*l* - - -*l* -*l* - - *k* -	l s s l s st t l s d l s s t t l s t s l d t t l t
-*l* -(d t	t z l l s s t s tl l w

t Uz l t s s ffi l t t t s ffi s l t t t s ffi - t d tl ll w t t s ffi - ss s s ffi st ll w t l l s ffi l w w s d t t l s ffi s s d s d t Uz t t l s

s s ffi s - t - -*k* - d t - s t -*d* l t -*d* l t

2 l l s ffi *-l*
ss ss s ffi s *-(-(-(-(-(-(*
t d t s ffi s -
5 t s ffi s
l *-do -k -bo* - *-do* - *u o - o - o -ko ;*
l s - *o* - *o* - *o*
st t t s *-l k* - *l k*

V

t t s ffi *-d*
2 ss s ffi *-(l -(*
s ffi l l
l s ffi s l s
5 s t s ffi t t z
t l s s zl s ffi -
7 st s s ffi -
s s ffi s
s t t t s z l s z l l
s t t s t s z z d l l
l d
d t s l st t s t z l l d
d t st s t t t s t z l
l
st t st st t t s Uz t z l l d
st t s t s t z d l l d
t t s l s z l l
t t t l t s l s z s d l
l

M r h h e c z e

t s s t w w ll d l s s d Uz s l s s d l t d t t s ffi t t w ll tt d t w d dd t t t s ffi t s t l s t t w d ts l st d t s ffi t l s t t w d S t d t t l t l l d l t Uz l w st s d t s t l s s s w ll w t ll ws w s t ll l s t t s d l l d l s l s l d s w ll w l s s s t *o o o ul* d s l s w s d l t d s s s s *o l d l o ul*
s t t s l s w st st d t Uz l t d t t z t t l tt s t t l t Uz l

Uz l t { d l s t z

w ls {
s ts { d l s t z
s s ts { s t
S t s ts { d l z

t ll ws s l st s l s Uz l S t l s l t t s ffi s t t tt d t w d w l t t s l l t t st s ffi t t s tt d t w d t l s s t d w s d t s d t s t st l tt t s ffi d t s t l st l tt t t w d d t s t l st w l t t w d

L *o o o ul f o l d k l d b l*

E M L *o* l - ↦ *o* l l d ss - ↦ *ku l* d ss

L *o l d l o ul f o l d l*

E M L z t - ↦ z t t s t - *l* ↦ *l* t s t d

L *Doubl o o ul f o l doubl*

E M L *b* t l d - ↦ *bb* l d

L *o u ul f u* - *(o o d o*

E M L s t - ↦ s s t s t - ↦ s s t

st t l s l st d s t s t t l d ll w ds l l s d ll l d d t d s ss t s t l s s

sin p s n M p y

w d t t s w d d t l d t s d t t l t s l s 2 d l st t s d t s s t s s d t R l t t t t l s t l X st ds t *u o d* Y t s d l *of o d* l t t t d t ss s ffi d t t s s ffi w t d s t $X = Y + Z$

E M L l d d d t s tw s ffi s - *o* d -*d l d o ul* d t t *k l* t ll w tw l s d t s s d d t l s t s w d

s s X *k l d k l* t s w d s s d t t l s t Z *d* st s l d s s l

2 X *k l* Y *k l k l* t

2 r ca or n ca on o con ga ona ffix

A F P D CA	C P N
X,	**Fo t ou X tu s t s su t fo of st of o s**
ca	ca o a no n n ca gra a ca nc on n a wor n ng o ca ga a ca n acc a ca a oca ca an a a ca
o	o ffix ng ar r on ng ng ar r on ng ar r on ra r on ng ra r on ra r on
ong ng	S ffix o ong ng *n e n e*
n	S ffix o n ca on o a on *gac a e gac a e*
r ng	Gra a ca n r o a wor o o a n ra or a ffix *a*
n n ar n	D n n ar n ffix *c a y nc a a e abb*
v o X,	**Fo t v X tu s t s su t fo of st of o s**
q on	In rroga on ffix *b e e g*
ra	I ra ffix *e G*
n ga on	N ga on ffix or r *a e a n c e*
r c roca o c	R c roca ffix *e ey a e ca e*
r x o c	R x ffix *n e n en y*
n	r n n a ca gor accor ng o ra a n r n n an r n

E M L

s p ss ss s
r s r s p ss ss s

r ca or r a ona ffix

M F	T	T
u rb	u su	→ rb
u d	u su	→ d
u b r u	u b r su	→ u
u b r rb	u b r su	→ rb
pr u u	pr u su	→ u
d rb	d su	→ rb

E M L

u rb d r su s
b r d rs

C nc u in s

t s w d t l l st t Uz ls s t d R Uz R l l s l l s l t d l s s l s ts t s d d s ffi s d t t d l s l t d t l d t s t t l s l d d t s d t d t l t t R s st ll d s t t p . u. du.p s pr s u rpp u rpp.r r ss 2

c e e e s s w s t ll s t d t std t t st t w t s s s d s ll ws t t d w z U s t 2 t 2

f nc s

r nc a r ogra c R or ang ag an a n o n r R o on No o g
Ga ar G Na ra ang ag roc ng n ro og on

G ng r an r S R r n a on o r or o og n N o a n r I an

a o J ra a N 00 or ng o or a o o gra ar o ang ag a a ca Jo rna # a n an 00 Sc ac r J G 00 roN o or or o og ca na oo or w roN o ro og Na ra ang ag oo n r o G org a `http www i g e p t` acc 00 0
ono a aro S ar o G o n q c
an r or o og ca agg ng o x ng a r o c ron c c onar n wan ow a o a c J a rac ca ca on n ang ag or ora roc ng n r r 0
an 00 o n ac a c ow a a n o row o owan o nc ow a c a c na awn c a I ar awa 00

App n ic s

A p f ic n f i us s f p c

```
    g                                        we
      i                                 pe
p p    e    i g       pe            e       p
  p     i                          Re   e i e p          y e
  p  b                              e    t  ti e p          thi
i p    i                             te    g ti e p            wh
   p  h                             e i ite p          e e y   e
  p  he h i                        Neg ti e p             b  y
i  p   i  i                           e i ite p              e  e
  t                                      iti  e   e b   t  t   e
  i  e                               t    iti e   e b   t      e
                                              ti
     i i                              e ti e    e
```

u c C f ZM M ic s

R s s s t l st w ds t l t t t l t t s ll ss l s t t s w d w t d s t s t t s ss t S I R G t t l st ll d
t s

t t t t t l .t t t s d t s R 2 s lt t R s l d t s t t d t t s d
S I l
R t l d l w t ? d .

G a a o an an

s d l w s l t t t t l s w
l s l t d l s s l d d s s
s s lts w ll t d t t l re t.txt

```
p g t i e e N e b t t
e i t i e N e ;
p i e t N e i
pe N e e e

e e te e e e t t
e t t
p igi e te e e t t
p ep e e te e e t t
p ep e e te e
i ;
e e

e e te e e Re t
t e t e e
get e
e te e e Re t
e e te e e Re t
t e t e e
e e te e e Re t

e te e e Re t
e e p e e
e
e t e e t
e
e te e e t Re t
e te e

e t e e e t
i b h
e b w e e t
e t e e 44 99 9 9 97
t e t e e
get e

e b w e New Re t t
b h New
get e
e b w e Re t t
e b w

i b h 9
i b h
betwee
```

b h New
betwee 9
New i
b h
betwee 97
b h 9 9

e 4 4
e
e
e 4

ig 44 99 9 9 97

e e p e
i b h
e e p e t
e t
e e p e 44 44
e e p e e t
get e
e e p e e t

p ep e e te e
p ep e e te e Re
p ep e e te e Re

M ph gi i ti y
p ep e e te e w t Re
ph b t t pee h w ite t pee h
p w w ite pi t ;
p w ite t i

p ep e e te e Re

t
ph b t p t

t e b
ph b t e b e b ph t

t ti
ph b t ti t

b e w ite be ph gi y e

p t e
w ite t w ite e

pi
pi Y w ite w ite pi Y

e e e pty
e e e pty Y e e e pty Y
e e e pty Y e e e pty Y

G a a o an an

p R
e R
be gi g R i i
i i it R4 g h
p e i e 4 R i i g i i i i gi i
p i g 4 R
i i ti e e e e t Y R7 h
Y R
e e e pty Y R R7 R R R4 R R R

e Y R
e
e R
ppe t
e Y t

e e i g i e
e " i g" i g p e i e e
e "g " g ti e e
e " i" i ti e e
e " " ti e e
e " " b ti e e
e ""

p e i e Y R
e
p R
ppe t
e Y t

e i e i e
p " " i g pe
p "i " i i g pe
p " g" g i g pe
p "i g" i g i g pe
p " i" i i g p pe
p "i" i i g p pe

p " i " i p pe
p "i i " i i p pe
p " gi " gi p pe
p "i gi " i gi p pe

p ""

i be gi g

be gi g Y R
e
b R
ppe t
e Y t

b " i i" i i be gi g i i i

```
b ""

   i i i ti    i it ti    e t i ti

i  i it   Y R
     e
         R
    ppe       t
     e Y      t

  "g h "  g h i i ti  i it ti          t b g h
  ""

            i
p    i g  Y R
     e
   p         R
    ppe       t
     e Y      t

p      "   " p
p      ""  i g

  i i ti e e  e  e t    i e

 i i ti e e  e  e t  Y R
     e
    i e          R
    ppe       t
     e Y      t

 i e       " h "  i i ti e e  e  e t    i e
 i e       ""

    R
        t i g t  t             R ;
     p   t i g t  t   p pe      R ;
   p p  R   ppe   pe      p     R R ;
     p    t i g t  t    e e i e p     R ;
     p    t i g t  t    e   t ti e p     R ;
   i p    t i g t  t   i te  g ti e p     R ;
      p    t i g t  t    e i ite p     R ;
     p    t i g t  t    eg ti e p     R ;
   i p    t i g t  t   i  e i ite p     R ;
        t i g t  t    e ti e  R


     pi            w            e h i      g t
    pt         pige              y        w
     h          t               b         hi
   i h             e                       the
              itt e b the          t b     h
              b g                 t        t b e
    b g       g  e               y         e
```

g i y hi t
g h
i gi itt e i te i h wi te
e gi t g i th
b e

pe
p i p i
p g y t p i
p t h e t

e p i hi i h i
p p e i g pe
p p e i g pe y
p p i g pe he he it
p p bi p pe we
p p i p e pe t pe y
p p i p e pe t pe they

Re e i e p i h i
p i y e
p i g y e
p i hi e he e it e
p i i e e
p i gi y e e
p i the e e

e t ti e p t i h h i
p th t p b thi
p h thi p h th t
p b thi p th t

te g ti e p h i
i p i wh i p i wh t
i p y h w i p wh t i
i p y i whi h i p h whe
i p h h w h i p e ht h w y
i p e h hi whi h e

e i ite p be gi h h i
p h e e yb y p b h
p h i e e y e p h y i e h
p h i ythi g p h ythi g

Neg ti e p b i h i i h i
p he h i b y p he h i thi g
p he h h e e p he h y i e
p he h ye whe e p he h y e e

e i ite p g h i
i p i i e e i p i i ethi g
i p y i y i p h i e
i p i eb y i p i ethi g

```
i p y i p h e
i p y i i p e h
i p bi e h ew i p bi h e e
i p bi i ethi g i p bi ethi g
i p b i e i p bi eb y
i p bi

e b ph R
  i pe ti e Y R
   e b t Y R
   e e e pty R R R

e b ph R
    e ti R
  te e R
        R R be i i e t
   eg ti e R
   e ip i e R
   e e i i e Y R4
   e b t Y R
      Y ppe R i pe ti e R ;
   e e e pty R R4 R R R R R

e b t R
    t R t iti e e b ;
    i R i t iti e e b

i pe ti e Y R
    e
  i pe R
   ppe t
    e Y t

i pe "i g" i g i pe ti e e pe t
i pe "i g " i g i pe ti e p

 e ti Y R
    e
   e ti R
  e t
    e Y t

 e ti " i" i e ti
 e ti ""

eg ti e Y R
    e
   e b eg ti R
   ppe t
    e Y t

e b eg ti " " eg ti
e b eg ti ""
```

```
e ip i e Y R
   e
  e ip R
  ppe t
   e Y t

e ip "i h" i h e ip i e te t
e ip " h" h e ip i e te we
e ip ""

e e i i e Y R
   e
  e e i R
  ppe t
   e Y t

e e i "i " i e e i i e te t
e e i " " e e i i e te we
e e i "i " i e e i i e te t
e e i " " e e i i e te we
e e i ""

te e Y R
   e
  e b te e R
  ppe t
   e Y t
  e e t t e i p e i gi e i i e
  te t h eyi
e b te e " " p e e t t e i p e i g pe
e b te e " " p e e t t e i p e i g pe
e b te e " i " i p e e t t e i p e i g pe
    e pe t
e b te e " i" i p e e t t e i p e i g pe
e b te e " i " i p e e t t e i p e p pe
e b te e " i " i p e e t t e i p e p
     pe
e b te e " i " i p e e t t e i p e p pe
  te we i eyi
e b te e "y " y p e e t t e i p e i g pe
e b te e "y " y p e e t t e i p e i g pe
e b te e "y i " y i p e e t t e i p e i g pe
    e pe t
e b te e "y i" y i p e e t t e i p e i g pe
e b te e "y i " y i p e e t t e i p e p pe
e b te e "y i " y i p e e t t e i p e p
     pe
e b te e "y i " y i p e e t t e i p e p pe

  t i p e tg i e
e b te e " i " i p t i p e i g pe
e b te e " i g" i g p t i p e i g pe
```

e b te e " i gi " i gi
p t i p e i g pe e pe t
e b te e " i" i p t i p e i g pe
e b te e " i " i p t i p e p pe
e b te e " i gi " i gi p t i p e p pe
e b te e " i " i p t i p e p e pe t pe

e e t ti i gi i e
te t h eyi
e b te e " y p " y p p e e t ti i g
pe
e b te e " y p " y p p e e t ti i g
pe
e b te e " y p i " y p i p e e t ti i g
pe e pe t
e b te e " y p i" y p i p e e t ti i g pe
e b te e " y p i " y p i p e e t ti p pe
e b te e " y p i " y p i p e e t ti p
pe
e b te e " y p i " y p i p e e t ti
p e pe t pe
te we t i eyi
e b te e "y p " y p p e e t ti i g pe
e b te e "y p " y p p e e t ti i g pe
e b te e "y p i " y p i p e e t ti i g pe
e pe t
e b te e "y p i" y p i p e e t ti i g pe
e b te e "y p i " y p i p e e t ti p pe
e b te e "y p i " y p i p e e t ti p
pe
e b te e "y p i " y p i p e e t ti
p e pe t pe

t ti tg e i e b i e

te t h eyi
e b te e " y tg i " y tg i p t ti i g
pe
e b te e " y tg i g" y tg i g p t ti i g
pe
e b te e " y tg i gi " y tg i gi p t ti i g
pe e pe t
e b te e " y tg i" y tg i p t ti i g
pe
e b te e " y tg i " y tg i p t ti p
pe
e b te e " y tg i gi " y tg i gi p t ti
p pe
e b te e " y tg i " y tg i p t ti
p e pe t pe

te we i eyi

e b te e "y tg i " y tg i p t ti i g
pe
e b te e "y tg i g" y tg i g p t ti i g
pe
e b te e "y tg i gi " y tg i gi p t ti i g
pe e pe t
e b te e "y tg i" y tg i p t ti i g pe
e b te e "y tg i " y tg i p t ti p pe
e b te e "y tg i gi " y tg i gi p t ti
p pe
e b te e "y tg i " y tg i p t ti
p e pe t pe

e e t pe e t i gi t g e i e b i e
e b te e "g " g p e e t pe e t i g pe
e b te e "g " g p e e t pe e t i g pe
e b te e "g i " g i p e e t pe e t i g pe
e pe t
e b te e "g " g p e e t pe e t i g pe
e b te e "g i " g i p e e t pe e t p pe
e b te e "g i " g i p e e t pe e t p
pe
e b te e "g " g p e e t pe e t p e pe t
pe
e b te e ""

b y iti e e b
t t e t be gi e
t h pe t w t h
t y w ite t i h i
t y p t t
t y w h

b y t iti e e b
i b g i e e
i ti it i eep
i y w i e y

p R

p ti e Y R
Y g p R ;
Y

e e e pty Y e ti e R R R
;

i i t R i e i t
ti e te pe R i i gi gi
p e i e R i i g i i i i gi i
p i g 4 R4
ie t 4 R t h e t i

```
  i i  ti e e  e  e t   Y R             h
   Y R7
  e   e e pty  Y R7 R  R  R4 R  R  R   R

M  i g    e ti e
 p   R
         Y R
      Y   ;
    p Y
  e   e e pty  Y      R   R

      Y R
    e
            R
  ppe        t
    e Y      t

      " i"    i      t
      "   "            t
      "    "              t
      " i "    i        t
      " i "    i        t
      "gi"   gi       t
      " i"    i      t
      " i"    i      t
      " h  "    h         t
      " i   "    i          t       i i    t
      " i  t"    i  t       t       i i    t
      "   "            t
      " h "    h        t
      "    "              t
      "p     "   p            t
      "  i"     i        t
      " h    "    h           t
      ""

 p   ti e   Y R
    e
    p       R
  ppe        t
    e Y      t

 p   "   "          p   ti e
 p   ""

   ti e  e p        e ti e

  ti e te p       Y R
    e
     te p       R
  ppe        t
    e Y      t
```

```
te p  "gi"  gi    ti e te p
te p  " i"   i    ti e te p
te p  " i"   i    ti e te p
te p  "  gi"    gi    ti e te p
te p  ""
```

```
Re  ti e
i i    t   Y R
    e
  i        R
  ppe       t
    e Y     t
```

```
i    " i   "    i     i i    t
i    " i  t"    i  t  i i    t
i    ""
```

```
Re  ti e
ie t    Y R
    e
          R
  ppe       t
    e Y     t
```

```
" i "    i    ie t be   gi g
""
```

```
i ti    y       e ti e
  i i              e                               white
  e g             wi e              t                   w
  b               high              p  t              w
                     g               h               eep
  yi i            thi                 y             petty
                                    eg i                 e
  y   i           e e                iy hi          b  py
      g           b  py              hi i            weet
    h hi          bitte
   h                 ty             t  i
  be              t  te e            h  h             g e
      yi  4          i                 b i 4        we   beh  e
   hi i      4    e    e t           h      4           it
   h    4              y            y       4          y
     g    4           y              he h   4       t    ti e
        4             e               y    4         t bb
      i               tte                            ti  i g
  b  b  y            eti               hb  y           g   t
     tt               g   t                          g
                                     h              h ppy
                     i              i  i              p i i
  y  h 7          y   g               i 7
   e    7                                 iy           e
   i   iy             i y                iy          pi it
```

iy iy p iti
g p
g p h pe h i h i i h i gi
i t
g hi
g p t te t i gi i t g hi
g p 4 h te te iy t i gi i t
g hi
g p hi i gi i t g hi
g p p y hi t te i g p i i t i gi
i t g hi
g p 7 h itie i g t biiy t i gi
i t g hi
g p p e i e i i gi i t g hi i
tg
iy iy hi h b y i

bi i g ti
h bi i g ti
bi bi i g ti with
t ti ti b t
e i t ti ti b t
bi t ti ti b t h we e
y i ti e ti
y i i ti e ti
y i ti e ti e e

C s

ti
h h i e i h tt bi g b i t y i y

e ti e
b i g e i hi te ti i iy
b h yb t b b b be b beb h be beh y
be i i be y bet tib be bit b h b t hig
hi y i h i g i h t e i
g ti g g g yib g i i i hi i ti i i i
i h t i y tt e g i h i p
h i h hh h yi
ht b t i i t g i g i g h
t h i gi h h p t h t i h i hi
t iy tti i t i iyi i i i g i
y t i i e i h iy h
hi t t i i t h i t i iy t i iy t i
t yi i t yy te i te b te g te te ti h ti i t g i
t i t h i hi hi iy yg h hiy
y gi y i y y i y hi yet i yi i y g i
y y pi y g y h y yi y h y i i h i
iy

e b
t hti g t i h hti
yb yt b b y b b h b b i be i be be
bi bi hti biti b b b b h b h b h b
b y b i b i b h i h i h heg he
hi hi h h p h eg e g h e i e hit e hit
e e g pi g p h g ti h y h y t ti i h i
i b t i h i h it i i t t
t y g p t h y yt e h e hi
e h e e e g yti e e et i h i i iy iy i
t i hi t t h i h t t
t h ti t h yi i h i
b h t i i t t i h
h g i g h ti g i h i i g
h t hi ti ti t ti ti t y
y g ti p p e pi hi p t t yt
i i ti i t iy h y i ti
hi y i ti t t y
i ti h t i hti h
y e e e i e gi h iy hti h g i h hi i it
i i i i i i i hi i i i g i t
yg t b i t i t t i t t t h
t i t i te hi tep ti ti ti h ti h t
t i t p t p t h t t t t g t t hi
t h t h ti t y t h h h i i ti i
h h y y hti ti h i y
i i y y t y h ye yeti h yig yi i y
y t y ti y h y t y y y y g y y t y

N
biy t y e p t ge t i h i h iy t
b e i iy ti e iy e giy b
b b h t iy t e
t it iy t b t hi b t yb y
y b h b h b h b te iy b i b b b i
b t b y be t be gi be bet y bi i bi bi giy
bi i bi h bi e b b b g b b b b
b b b i b b h b h b h i b hp b i b y i
b b g y b t b h b h b b ti b i b y
b y b i i h by et by ete h heg hi h i
h y i b i t
y t t eh hi i e eg t
e tiy e gi e et e i t t i ip t
iy i t y h
ig te e giy e t e i t e hi e hi e e t e e t
e i e e e e t b e hi e t b i
e e e i i i i i i t t
g i g b g y g p g b g g y t g g
g i i g ib g ht g g g h g h h t h iy
h h t h t h h t h t h h y h y t h y

h i h i hi y hi h y hi b h i h i h i iy t
h i h y h t h t i h i i i i i iy t
i y t iy g i i e e iy i i i ti i h i hbi
i h h i it i i i i i y hi it iy t g
g hi b y h t ig i i i y t i
i y y i i i hi i e y b
pit i pt t i e h e hi
e h e i h e e g h e i y i i i hi it b
iyi h e t e i p iy
py te g e p i pt t e it i i
t h h b t iy e y
yih iy t g i g it h t t b
t t iy t b t
i hi hi te ti te i
y eh e i p i it iy i t i ti i e
i i t i t i i iy e y
h i ht t h h i
t bi i h ti iy
i h i h t b h h i h
it e g i g i g i i e ti p
it hi i g g i t
h i i pi i t t t ti i h
t y yi y hi i p p p hy t
p e t p ti p p t p t pi h p e p p t p yt t
p p g p bi h ti i
yi i g i h i h i i i i i i h g i
i t g i t p i
y g hi b h ti g i
y h i h i et g
e t iy e e tge e i i b t h b b b t
y t t t t y h t
e e t e gi h h h h h h h t h
he hi h y p h i i t igi i i i gi i
iy t e et t h hi i pi t p t
t t t y t bi t t ib t b t i
t t t i t i t i t y te t te e te e p
te e i e iy te i te i te i t te giy te i ti ti
ti ti h i ti h ti i t g t g t g t t h
t t ti hi h t h t t h t p t t b t i t t
t t t h hi i i e itet t
i i h b y t e t y t i y t i iy
b t t y t iy t i t i i
ti y i y hi ye ye i yi y g
y y y h y hi y g i y i h y g i
y i g i y y h y y y y y y y
hy t te t h b i gi ehi i i i

u s

i g g i i g i g i i i i i
gi i gi i i g h h g i g g i g i
h i h i i i i i
i i y y y i y i y i y i y i
i i g i gi i i i gi i y p y p
y p y p y p i y p i y p i y p i y p i y p i
y p i y p i y p i y p i y tg i y tg i
y tg i g y tg i g y tg i y tg i y tg i
y tg i y tg i gi y tg i gi y tg i gi
y tg i gi y tg i y tg i g g g i
g g i g i g i i gi i
i h i i t h p i h
gi i i gi

i f i i r r r N
i rf z i

t S [1] ws [2] d l s [2]

[1] n r é ran o Ra a o r ranc
g t y i t
[2] In o o r Sc nc o ca o Sc nc o an
bieg i w i i ipip w w p

A st t c or o og ca ro r o o wor ro r na r n a coo ra ng ra wor o wo or o og ca oo *e z* a or o og ca ana r an g n ra or or o wor an *e* a cro ang ag or o n ac c g n ra or o wor n c n r ac con ra n r q r or n ro ra o oo an w ow ow r ng a or a ow on o cr or o n ac c a o r o o n r ng xa o ar aw wor o on

o ds co a ona or o og o ro r na wor n n c on an ar a o co o n

n uc i n

s d t d t t s l t t t d l t t t t l l ss d t t t d t s t t t s t t l ff ts t t t d t t t t d d ll t l l st d s t s s ll l t d s s s t l s l s st t s lt w d ts s t s s l w ds t s t t t l t d l s s s l s lt w d l t ll lds s t t d s d d w t t d s d s t t w s d s t d s t t t st t t l z d s s t t s w d l t t t l l s t t d s t l l l t ll w s t t s ss s s st t t s s t s d dd s s lts t ss l t t l l s t s w s ld ss l z d s ts t s s st t s ld d t d t t ll t ll w s s l w t t s d t d s t t s d t s w *Ro do u o AK „R do " Ro do u o „R do " Ro do R do* d *Ro do B bk*

t st d t l l t s l s lt w d
s l s l l w t l t d s t ll
t ll t t d s lt w d d t
ts
t s w s t t w tw l l
t ls *o f u* l l l s d t l s s l w ds
d *ul* ss l s t t t lt w d ts
d s ss t ts t t l t t s t ls
s lt l t s t z d t s t l s
s d d s t lt w d d s d t t tt s
d t t s w w t l t ll ws t d s t
s t t s t st l s s w lt w d
t s
ll w dd ss s l s s s

t d t l d s t ts
t ss t d st s ts ss ss t s t
t s
t ll t l ts t s lt w d ts t t ll
t s l w ds *ul K u k o u k o* ll
s s d
l l s s d t s d s

s lts s t d t s t l t t
d ll s w t st ts l s st t t s ts t
d t s t t s st s d t st s d d st t s t s
w ll s l s l l l d s t s w t s w
t t t d d l s st s ls w s t t s l t
t t l s t s w t lt l l t l s
s d t sl t lt l l t t t d t t
l t

f A

st st d s t st d l t l t s l s
s d d d s 2 2 w l t l
w ff ts d w l s l l d s t t
t l t l t l t d d t s
d d t d t s ll t t d t s
l s ts t s s s d s d w t
s t t t d l s w s w t t l s ffi s t l
l s t d s l s d d t d t l d s st t
d s t t d s s l d d s
l s st d s t d s l st ll d s t t s t
t st t t t s s s s l s d l
w ds l s l s d l w ds d s s t

s s t s d s ts l d t t s s t s
s d t s t 5 s d t t tt s
w s t d w d t wl d s t l l l t t
t l s l t t l d t s d *o f u*
st d d s ll t l d s l st 2 5 l s
w d l s t t s t st s d 2 st l l s
l s s w ll s s ffi t t s d l t l d t s
st s
t t t t l l ss t t t t d l t t
t s d t d t t s s st l
t % s s t s t ll
d d l s l t s s s d t l ss t
t t l 5 22 t U d 5 s
d t t t t s t l ll s t
t
l s t ll d l ss st d d l w t t s st
ff ts w s t st ss s st t d tl w t
t l s t t t l t z t s s
st d st t s s d 25 d 2 2 t s t d s l d
t t t t s t t d t s l l
d l s s ls s ts t d t t t
s d l s d s t t t d l t ts
w t t s s s s d w t t
t U p s u . u ts l ss s s t
s l s s d l s s t t s
t l t l t s s w 27 t st t t s t
st l d t l t t st ffi l l sts s t
t d t t t ts s l s s t t t s
t w s t d l w
t t st t l t s Sl l s s s l s s
t t l l s d t t t t
d s l d ls s t s l t t s d ls s
st lt l l l l d l s s d 5 s
s d t l st d s t l l l t S l s
d l
d s l t l l ss lt w d ts
Us s s w d t l st t
l l s t t d s t l t Us s t ss t l t
t l l t s t t s
t l t d d s s lls l l z d d ls w d
ts l st d d l tl d s d t st d l
s s 5 2 2 d s l s w s d
t t *ul* l s ss d t s t s ffi tl
ss l s l d s t t d t ds
s s w l w t s t t t ss d
s s t l l t t t t ts s t

Sa ar J Ra ga n w a an o

t w d s d w ds st l l
l z s d t s l s s l w ds s st st l s d
t ds *o f u* l s t t l tt t

M p yn c ic vi u f is Mu i W p s

1 The

l t t s s t l s s t t s t t
s st d l l s s w s w s d t ll w t s
l s s w d st t ts t ffi t s st l s s d
d s t s l st t t s d l t s s s ts
l t s s ss st l s ts

r ve t s w s d d d t t d st ts
s t d st ts s st w d w t tw t s

aga ó n c Nor raga *aga n e* So raga

lt t d st t s d s w d s d s t
s s d t st s

zy ane e any zy ane ó e c e zy ane ynó
an o Śr c o r n w o

d st ts t s d d d t s ll s s s s st t s
s s ll t d t l l s w d lt w d s
w tl s t s d s d t l
t s s

a ec Ra ow c q ar r *a e azną a ą* n Iron Ga

a a

Tr ffic e t t l s w t s t t s st
s t t ffi t s s l s w l t t s t s
s s s *ul* st t *Al Ro do* d t *l* s
l z *k* s *T* t l *T k* t w *o* d d
w ll l t st *ul* s t t l s d s d s
s ll tt d s l

ca ac a ac a rac a S r

e a e n czen a e n czen a n on n

t s d t st l d t l s s st l ts
d t l s s d ff t t s t s
l t ts s dd d s s st w ds
s

In c on o o or ro r Na

ca ⟨ *y e y a* *na*⟩ r r c o n S r

e a ⟨ *a S a e*⟩ Scar n

n ⟨*Gene a a a e a e Ga e a*⟩ G n ra ar Ga Ro n a o

0 *ac* ⟨ *na*⟩ J on Sq ar

a a ⟨ *ń a*⟩ or a or na o a c

⟨*S e e* ⟩ S r r g S r na o a or ca r c

ce St l s t l t s t t s st s s s d t st s d d d lw st t s d ts t l l s t t t t t d l t t ts w t t t l t s s s t t t d d l s st s s t 7 l s tt t d t t s t t *D o* St t *k* st d *o o* t s st

zec en a ny n ra S a on

zec S a za a ac n a ar aw S a on

n czy na r r c o n ar aw r or

st s st s t ds

t t s t st ts t l t d t s l

t l t st t t s t l s t s l 7 st t d st t t s l

zy ane y a N w or o

zy ane H a a r a a S o ng n r o

zy ane c a Sz a oc a S r o a o

her ce ec t s w w s d st d d s l s t t l s s l d s l s s s s s d d s t s 2 l st t t s d l t s 2 s s 22 ts 2 l t s 2 d s ss st l s ts t l st t s l t w l d s ss t s lt s w S t w l st l s l l s s s l 25

a S a y ze a e e eg S ar w ar o I J a r w *a e* o o w

0 *na ny en a z ó n cny* Nor n c a r *z ań en a z a a* a ar r

ze z acy o a an ac c *az ec ea zyczny e e a az a* c a r r a

Sa ar J Ra ga n w a an o

a ac na *z e* a ac on I
a e *ó e* Ro a a

n *y* *n e* *a* *n* a o on a no on n
n *a e* *a* *n* on n o ro r oo n r
na *yg* *n a* *azy* o n o S g n III a a

Ga e *a* *ó* o o w Ga r
e *a a y* Go n rrac

e *n* offic ng
en *aza* an ng c n r

c y he

s l st d d d s d d d s w t t d t l
d t s d l d s t d s s d t l tl
ss t d s l st d t l st wl d s t
l l l s s t t t ss st t
s l l s s l d s l d w ll
l t t s l tl d s d

t s s t w s t w t s t t t s
s s w ll s s s l s s s w t s t t
t s s st d s s d t s l s s lt
t U t s t 2 s w ll s w s w t s s
l l

rev l s s tl t t s s
l s 2 t w tt l w s d
t l s s ll d t s t l tt s d t t s
t t 2 d 27 t s t s l t l l

a *n ó* n o o *a* *a e a* a n ng

e z *e* n o J r a *e e* a n

a e e *eg* I J a r w S r *ca* r
n a or *gnacy* *an*

n czy *na* r r c o n ar aw r or
en a ca o

0 *n* *Gen* *a e a* *e Ga* *e a* G n ra ar D Ga Ro n a o
Gen *Gene a a* g n ra g n

an *za G* *z* *eg* ro or Jan Gro ow S r
e *a* ro or g n

cr y I s d t l s s t s d t ls t s st t ts d st s tw t tw t s w st t s t l w ds d t t l l tt s s ll l s s l s 2 l tt s d d d ll w s t l s s s l s 5 d s d t l s s s t s d d t l s t l t s l 7 ls t l s ds

S ***Z****a a* ***U****bez eczeń* ***S*** *ecznyc* Soc a S c r ffic

a o a ***Po*** *a* ***F*** *ac a* o ar ac

o o ***Po*** ***Mon*** ***S*** *y y* o S r ono o

a ***P*** *e e e* ***P****ań e* o S a Ra wa

o n *gan zac a* ***N****a ó* ***Z*** *e n cz nyc* n Na on r gan a on

Br **Broa** ca ng or ora on

SG S g ***Sz*** *a* ***G*** *ó na* ***P*** *an an a* ***S*** *a y y* a n Sc oo o ann ng an S a c

d st t tw t s t s t s s l t t t t s d t l s ds l s t d t t l tt s s l t l s s t t d d w w ll t lt ts t t s st l l l d s l s st t lt l 5 s w s t s l l s l d t s t w l t l s s t d s s t d l t t t lt l s t w s s l t s *b l* t t t d s t d t t s tt s l

S no *S* g n *S* a c

0 no g n a c

no g n a c

b e no *b e* g n *b e* a c

er s l s s s s s w l s 2 t 5 s ll s t d w t d ts s t l d s t s d d d t l d t l s t t s s w t t ls t t l ll s d d d t d l d d l ls s tw l ss s s w ss t l d ff s l l d s t t l s d l s st t t l l l ss t w w t t l s t t t s s d s ss d d 2 d l s s w d t l t d t l l ts ll w t d st t l d t s

Sa ar J Ra ga n w a an o

a *eg* S con o r S r

zeczna N n ran r a S r

ec *y* *eg* *a e* *c* *e* r S x R g n o ca c In an r S r

ag *an en* Dowr o S n a S r

y **y** **r** S l l t t lt w d ts d s s t t t l s t t d s t t t s s s s s d d s d ff t t t l s s t t s

o *l* l st t t l tt s t st t ts s w ll s s t t t d t l s st d t t ll ss l t d s l 5 w d t l t t ts tw st l s *o* d *lo* d t w t t t d d s w l t t st t ts d t

u *n cze* ***o*** *na* *c* ***e*** *n czy* *na* c g n oc c

s lt st t ll s s t s d t t t t t ll ffi l s t l d t l d l ts t t st tl tt d

t *ul* st t w d s st t ll ss s l s l 7

ss t l l ts s l s d d st d s s t s l s s t l st wl d t t

st s s l 5

s l t tl s s s *of* *o* t s l 5

ca *e na* *e na* na S r
ca H *an* *ó* *H* *an* *ó* an w S r

ca *y* *a* *za* *e* *1* ar aw a 0 S r
ca *y* *a* *za* *e* ar aw a S r
ca *y* a S r

n *g* *an a* *a* *a* *a* *a* D on o Do c r Ro n a o
n *g* *an a* *a* *a* *a* *a* D on Ro n a o
n *a* *a* *a* Ra o aw Ro n a o

0 *ca* *a y a a* *n e* *eg* a aw ron w S r
ca *n e* *eg*
n e *eg*

ca Gen *n* *n eg* *óze a* *a a ń* *eg* G n ra J a a S r
ca *n* *n eg* *óze a* *a a ń* *eg*
ca *n* *n eg* *a a ń* *eg*
ca *a a ń* *eg*
a a ń *eg*

s st t *o* s tl s d
w t s w t s t l s s l s
t t s s l st d t l s st l t d
lt t ld s tl s ff t t t s t
t t s d t l ld l l w l s w t
t s st t t s s d s s st
s

ana a a Jo n a II n
r o *ana a c e eg* J an arc w n

S a n c n o So ar
r o *a a e cze eg* aro Św rc w n

s t d s l s s w t s t ld s
t t t s w s d ts l t t t t t
s w s t t d l st l 5
t ffi l s l ss tl s d t t l t t d t l
55 t s t t st t w s t s tl
t l t

n g an a a a Ra o aw D on o Do c r
Ro n a o
r o *n ab a* a a Ro n a o

ana a Jan Ro S r
a ana n y cza Jan no a Ro ow c n

s w t z s s t s t t w l s l s ld
s t s s t t t s l s tt w t t ffi l
s ss l s s s s t s l d t ll w

t ffi l s s s t s st t t d ts s w d
s s s s l 5
t s s l t t ld s s
ss t s t d s t s l t
t t d l 57 s ts st t s s ld s
st d st ll w t t ds t d s tl t
s ll ll d t Sl s
l l l l 5 t s s ld
s w w s ld t ld l d t t
t l t t l l s l ff t w s t s l ss
l l s t ffi l w s l d t t
s s t st
s l st t t t l
s t d t t d l w w s d t d s
l s w t t l 5 t l t t t l
t s w l ds s w s d ts t d t t s s t
ff ts t t s w l 5

a e a y a a y ąc ec a n o n ar na
n or a *a e a y zyń eg*

n a e a n on n o ro r oo n r
n or a *n z e ec ącyc* on n o o r S ng
rar *n z e ec S nyc* on n o o r Sa

ny e ec ng
n a *y e ec* Go n ng
n or a *S eb ny e ec* S r ng

n e Ga e a D Ga Ro n a o
n or a *n e a a* a r Ro n a o

is s n s

l t t t ss lt w d s s t s d st d s t t s t t t s d d ss d t s s t s s s t t s d s ll l l z d s l st d d t t d s d ts l tl d s d tw l s s st t t l l s l w ds s t t d *o f u* l s l l l s d t t l s l l l s S d ss l lt w d s t t t *ul* s s d t ss t ds t l t t st t ts t s s w ll s s s t t tt s d d s t l s t ls d l d d d tl tl t t d w t w w s st t l l t s t d s t t dd ss d t s

t s s t w s t t l s s l t d t s lt l t d s w w t l d t s ls d s ss t t l t l t s w w ss t t ls t t

1 M r e z

o f u s l l l s l s l l t p p. p p . w w.p w s r us t s ts t s ll d *o f u T* s s d s s l t z t l s l s s t l t s t l s s l 2 t t l s t s s t t t t ll t l s ss l t l t d t t w t ss l t s s t ll l z d l 7 s s s t l s t t t t t s

w s t d *o f u* s d d l t s w s s s d t l d t l s d t t s t t t l st 2 5 l s l s l d s l t s d s w s d t t d ff t t t l w ds

t s t w t d l s dd d t *o f u*
w s l t d ll t d s l t
tw t s d l d lt s d s ss d s t
t s t s d *o f u* w s d s d t t st t t t
S l s d S s l 2 d s d t
t s t t l ss s d t s ss d t
t s t s s d t d w l s s t t
S l d s ll s 2 2
t s t s d *o f u* s s d s s t
l l s t t d s s d s t t t
t d l t t l l ss s l l w l ss
ts s t s t t d d s t t
w t s t s w t t l ss s l s s d
w t d t s d s S l ds t d t l d l ss s t
t d t l ts s S w ll s l ss s *l* S
t s l ss s s d t d tl t t t d t l ts
s d t d s t t t l t s
d s l l ss s s s t l ss s
t t l s tw d t l d tw d l st
s l - *o* - *o* t l t l w d s l -*l*-
d t d t l l ss s ls d s s ll d d t
s s t s t s d d l l l d s
t t t s t l ss ls l s sts d l ls
d t s l ss s l s d st s d d s
d s d t *b* t l ss s
d t d d t l d t st s t l d t
t *b* l t t *b* s l - l d t l
l t t l t s ss d t t s t l d t d
t l t s s s s s d s t d s w ll s
st t d t s S t w st t d d t w
s d t S l s s t l t st s t l t
d l t l t t l t d t t
ll w l s s ld t ss t l t d t
s t l t t t s t

ul u st t subs p s l l st
t l s d
d pr pr s . . p r l t l s
l t t s l d s t t s s
t t d ff t t s t s t
r u p . r u p . . .
. . r u p r t s t t s ss l
t st t t s l s l t s d
t l t s l s d s t s l
t t d s t s l s l st t s
t l r d l t w l t t tw t l r
s s t 5 t t s l s

M flex

ul 7 s ss l s t t t lt w d ts t l s s d l s w ll t d s d d t ts d s d w t ds t s s t t ss d t t s t t d s s t d s d ts t s ll ws t t t w t st t ts d t s t t d s tl

e cr he e eve lt l s d ts t t t l l d ts l l d l t l st ll t l t s d s s t d t s d l s s l d l l s l d t d t s w ll s t st t l l ss s d t t l ss ds t t t s t ts d t s t t l t t t t st t l s st l s s s ll t d s d t d d s ws t d t l s l l d l d t t t s t s l d s t s t s l s t s d l d ts tw s *b* s s s d s t s s t d s *D* t d t s *d* s s d t t s d d d d l ls *u* s ld d d t s d d d l t s *d* s *d* w l d l l d s 2 t l t s l s d 2 l ss s

o
⟨ G RI S⟩
N g
a no g n a acc n oc oc
G n n n
r r c r
D g o co
r r

⟨ SS S⟩
N ⟨ ar⟩ a ⟨ ar⟩ G n ⟨ x ⟩
n N ⟨ x ⟩ a ⟨ ar⟩ G n ⟨ ar⟩ cco ⟨ ar⟩
n co N ⟨ x ⟩ a ⟨ ar⟩ G n ⟨ x ⟩ cco ⟨ ar⟩
a N ⟨ ar⟩ a ⟨ ar⟩ G n ⟨ ar⟩ D g ⟨ ar⟩
a a

F or o og ca o o o n *e*

e e **e** s t t d s t
l t d *ul* l s t ss t t t t s ss l t d t t
d s tw d d l ts w t d w s l
t t s d d d ds t d l l l d l
t s l w ds *ul* w s w t t s st d t *o f u* l t
d t s d tt d l l t s t ll w

t s s t l t t s tw tw l t
t s *d* *o* s t
s s w t t s s t l t
t s l s d l st w ll d d s s *k d* | t s tw
t s
s l t s s s l t d l t t s
l s d l st w ll d d s s - *o* *d A*
s l l t t s s d t t t s
st d t t

ts d l d w ll ws
t t w l d s t d t d s d t ts
st 2 s ws t d s d s 2 d t
s l ts w s s d ts d t s

or ⟨ an ⟩ o n c ⟨ an ⟩ ⟨ an ⟩ ⟨ an ⟩ o na
0

Ron o ⟨ an ⟩ gr owan a ⟨ an ⟩ ⟨ an ⟩ Ra o aw

F 2 N r ng ng co on n w n co o n n *e*

e h t ll t
t d d l d t *ul* s sts stl
t t w t d d s w t s
t s ts s l ts st d t t t l t s
t ll w d t s s w

0 *G a ec eg* Gro Row c r g

e e e z e n o J r a

w t t t d s *o* d d *Al* s t l
t w l t d s l d l l s t l
ll st t ts l d t l t
d t *o ol k* s s t t t l s w t t d *Al*
t ts t d t l t w t s lts

c e G a ec eg

e ac e z c

s d t t ds w d t l t l s s l w ds l ll s w ll s t s t s z d s d s t s w *ul* d ds l l d l s l w ds *o f u* t s ss t t t s l t d t t s s t t t t t t ts d l d t ss l s t l s t s l s *Al* s t l l t t d t t t d w t t w l s d t t l l s s *o f u* t 2 l s l ll s d l ll ws t d t t d s d st t ll w *o f u* t t s tt t d t t t *Al* w t t d

a a 0 no
a *a e a* r o on no n ra no na n n

w ll s t d s

e e ⇒

a *a e a*
o on
c a *b*
N
a *n*
G n

F o anno a on o ng co on n *e e*

S l l ss ll ss l t d ts t d l s d s lts t t s s w s d 5 d t t d s - d - tt d t t d t s l d t ll w s s t

I flec r h r *ul* s d l s s t s d st d s t t l l ds t st d s d d t d tt d d s ss d d t s - d - [4] s d 5 t t s z s

S c a wo a r cr on o co o n r on o ng wor n on o r q nc o co on a n r a wor ow r o a roac co o n a a n q nc o c arac r c Na o n c on gra ar ar rar an o ow a con n on a ow ng a na ga on n a r a x ng gra r an or *n* *n* an or an n n c n an or a o n c on

⟨*b an* ⟩ *G a ec eg*

a
o on
c a *b*
N *g*
a *n*
G n

⇒ co

F a anno a on or *G a ec eg*

e e ⟨*b an* ⟩ *e z e*

a *a e a*
o on
c a *b*
N
a *n*
G n

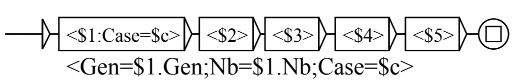
a *e z*
o on
c a *a*
N
a *n*
G n

⇒ co

F 5 a anno a on or *e e e z e*

d d s s t d s ts d t s w t t l t st w d ds w t t t st l d l l d s t s s d t s t s l st t ts w l t s d t s t t w l d st s ts t t - d t t s t t st t ts t d l st t t s w t l l d s t s t s *2* t *5* t t s ld l t d w t s d t l t s t t st t t st t d t d s d t d s d t l t t t s

o lu t s t t d t s d t s s d t l d l s *b l; f;*



F In c on gra or co o n n c ng *G a ec eg*

Sa ar J Ra ga n w a an o

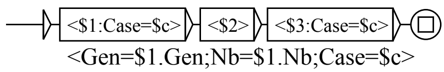
t d t s t t t t t s t t d
t s d
u *o* *bl* t s t d d t s s *b* *;* *;*
t t l t s d t d
l ss t s l t d t st ts w t s t t t
st t ts s st t l
t t ll t l s *o* t *o*
d *lu* t s t d d st t t ll w d
t *b* *b;* *2* *;* *5* t s d d
l t t l w t s d st t t s t d
l st s t t t d st t s
s t d t st st t t t l

st w t s l d t t t t d l
t st st t t *o* ts l s s
t st w l d t s t s s l s t st
st t t t s t l d t s d
t t d d l s t d l t t s
ss ss d t st st t t t d l s *;*
b *b* d t t d s t t t
l t l t s sts ll w t s
t s s t ss l l s ll t t l s s t t
t t t s l d s t s l t s
l s w s lts s d t d s t t d w t t
l d *o f u* l t s s s w 7

G a ec eg G a ec eg b g n
G a ec eg G a ec eg b g gen
G a ec eg G a ec eg b g a
G a ec eg G a ec eg b g acc
e G a ec eg G a ec eg b g n
c e G a ec eg G a ec eg b g c
c e G a ec eg G a ec eg b g c

F nno a n c or o *G a ec eg*

U t l s t l ll w t l t t t
t t t ss t d t l s s w ll st
t t s t s t l t t st d
t t d st t t l t t s s s d d t t
t t s d t l t s tw st t ts s st d $7 \times 7 = 49$
d ff t t s w t l s t s w
s tw ts w s lts t s s s w
t t t t t d l s ll w t t d d t t
t l t s ts d st t t s



F In c on gra or co o n n c ng *e e e z e*

e e e z e e e e z e b n
e e z c e e e z e b gen
e e z c e e e z e b gen
e e z e e e z e b a
e e e z e e e e z e b acc
e a e z e e e z e b n
e ac e z c e e e z e b c
e e e z e e e e z e b c

F nno a n c or o *e e e z e*

; b b ll w t t ss d t l t l l
ds s t ls t s l t d s l s s

By **e** s t d s t 2 t s s w
d l t t d t l t tl s d
t t d s t t ll ffi l s tt t
s s t ss ss t t
st ll t t l d ts t l st t s t t
ts ll w t d s s d t t
d t s w l w st t d s s t ffi l
ll s t t *Ro do u o AK „R do ”* w ll
ts s t d w l t t l d *Ro do* ts
s ddl t s ds t t sl tl t d t *Ro do*
u o „R do ” w t t t t t *AK* d
t ll w l s ss st t t s t s t st
d st l s d s *Ro do R do* w *R do* s ts
t t d s l t s t d t s t t s

I er er y r ve r e r

s s w t tw s s t s t s t t d s t
ds *ul* s s d tw l S l w ds d
s d st t t d d s s l s t l

In r ng a n w co on n a o na ra no g o x r w n a gra
ow r w a no r n acro a o on xa r q r ng a r
In c on or n r a con ro r a q on or o on a a na
n q o c ow r an xc ona ra ag acc a a n *e z z*
e a za e n e ą a n a a a a Do o c a a r ar wo
Ra o aw Ro n a o n ar aw

ac ⟨b an ⟩ an y

a ac
o on
c a b
N g
a n
G n

a ban y
o on
c a a
N g
a n
G n

⇒ co

ez ⟨b an ⟩ ze n a e

a ez
o on
c a b
N g
a n
G n n

a cze n a
o on
c a a
N g
a n
G n n

⇒ co

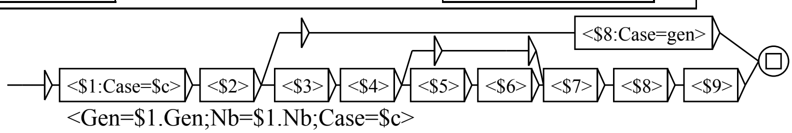
F a anno a on or ac an y an Sq ar an ez ze n a e rn a ow a

n g an a a a

a n
c a b
o on
N g
a n
G n n

a a a
c a b
o on
N g
a n
G n 1

⇒ co n 1



F a anno a on an n c on gra or n g an a a a

t t t d s ts ts t t s l s ts t l s t tt s d t l s t s d s s d l l ss ds s - 5 d s l t s d s t t l s t s l t d ds s - *o do-*

ul w s t ll d l d w t l s l lt l l l t s d t l 7 d s t d t *U* s st 2 w s t l l s st s s l w ds

d l d l s t s s t l
t *ul* d s t s ts w d l t s t t t w t
t t l l l d l s l w ds t ll d t
u d l *odul* s l s t ll w t st ts s d

M r h c M e d l d l d *ul* st
s t s l l d l s t d l l s
o f u d *ul* s d s d t s ts t t l
l t s l d d w ds s t s
d s t w d ts l st l s s l l l
t *ul* t ts t s d l t s s t s d t
t l l ss d t t ts t
s t l tl l t s w t l
t t st ts d s t
d ts t l st $2 \times 7 = 14$ t d s t s t t s l st t
t ll d t d l d l s t d t l
t t s l tl t d t s d t
s
w s t l t l l Sl l s
t ss l t t t s d t st t
d l S s t l l t s s st t ll
t d t s d ll d l s t *ul* l d t
t s d l 7 s st ts t ss d t t st l
l l l t t t d d d l t
d

ec T e B r e d l d l s ld d
l t d t t d s s s w s t
d st t tw s l w ds d ds t l l l l
s t s l d t d s l l t t

ll w t d s t w t t s s s l tt s
st *B ok* l t ll t Sl t t s d
t d w t s t s w t ts t
s t l *Ad ou* s t *B o oku B ok*
d t s t s w t t *u ou d u* t d
s t s s l w d s tw w ds s t d st
w s s l t t s s s l t s s s s
t s l s s *200* s s t s

ul d s t s ts w w t d t
d ts t t d l d l t t t *o f u*
d s l t s d d s s t 2 w s t t d
l s d s ss d

e er r c r I flec e r e l d l l t t d l d l s ld t ll t d s t l s d t t t

l d t l l ss t t d l *o f u* d s t s t ll t d s t w d t t w t t t t s s 2 s t d s d w t t s s t s d t t w s l s ss l t z t s d lt t s s t s t l d s t

a *cz e y* o on c a *n*	⇒

c worg n a oc n congr r c
c r n a n congr
c worga n g n n r c
c r a n n n congr
c worg n n n r c
c r n no acc oc n congr
c woro n no acc oc n r c
c r n no oc congr
n no oc r c   c r c n g n oc n congr   n acc congr

F 2 nno a n c or o *cz e y*

s s lt w t ss w s l s t t t t st t s l d l l t s t t d s *l* s s w t l t d s d t w t t t t d t l l l t s t ss l s ts s s t ddl t s

a *a e a* o on c a *b* N a *n* G n	⇒ *a e e*	a *a e a* o on c a *b* N a *gen* G n *1*	⇒ *a e* *a e*	a *cz e y* o on c a *n* N a *n* G n cco *ec*	⇒ *cz e ec*

F G n ra ng n c or o *a e a* an *cz e y*

n s in n p ny s

5 1 er h h

l s d l ls s w t l s t t t s s d s ss d l l d 2 t l s d s t t d l t t *o* t w t t s t d w l t t t t t s d t ll s s s t l t l tw s s s l ls *d* t d ff t ll t s

t s ds l t d l t t l l l t ds t d s d w t t s ds t ls ll st l 57 t t l s t t t t d s

n $_{nom}$ *z e ec* $_{gen}$ *ącyc* $_{gen\ nom}$
on n o o r S ng

t l s t t l ss t *ou u ou* $_{gen}$ w s st *o k* t lds t d s t d s t l t t s t s s t d t st t t s d d ts ll t l t w t t t

z e e $_{nom.congr}$ *ący* $_{nom\ nom}$
o r S ng

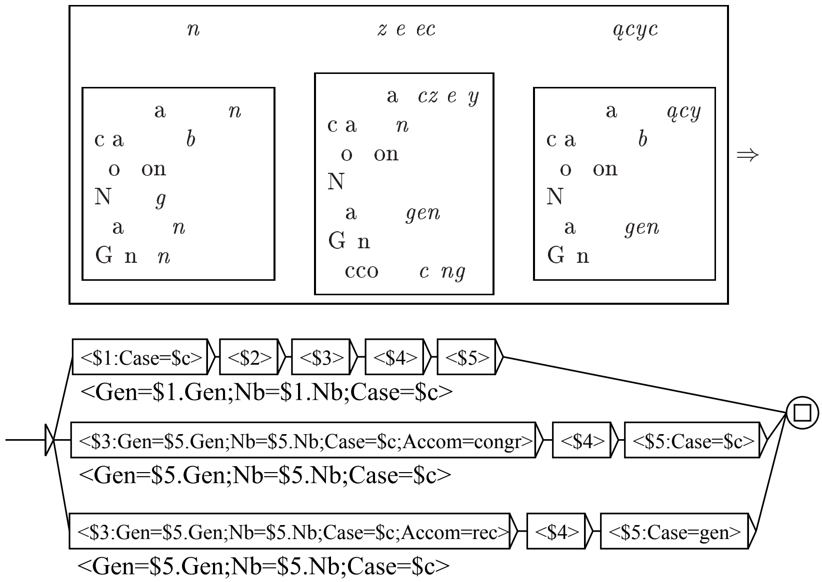
z e ec $_{nom.rec}$ *ącyc* $_{gen\ nom}$
o r S ng

t s t l d t s w s s t d t ddl t w t t l s t t d d t t t d t l tt s t l s s t t d t t t s w t l w st t t t t d s ss d t st t

5 E e e

ul d l t t t ts s t st s d d t d t t t l dl dd d ds w s t t t t s t l l s t w t d st t t l s t t t s l s ds ss d t t s t t s t t l s t t st t s l t d ds t s ts t s l d t s d l l t t l

R ca a co on n ar n r w r c o r o on n co o n a w c no n c ar a a n n c or



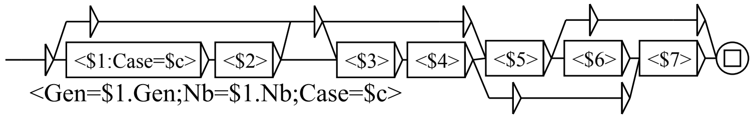
F a anno a on an n c on gra or *n* *z e ec* *ącyc*

s d t s d d d w t t tt d s t s d l s 7 d ll w s d l ll t l ts t t d s t d s 5 d s t l

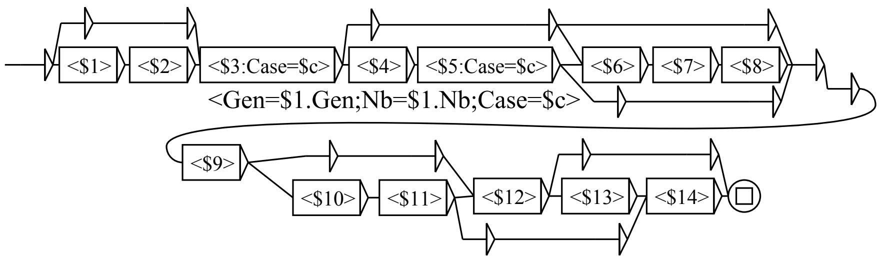
ca a S e e ar a S o ow a r S r
ca a S e ca a e ca S e e
ca S e a S e e a S e c

ce gó n z a cące a S e e
ar a S o ow a r r g Sc oo
ce gó n z a cące a S e
ce gó n z a cące a e
ce gó n z a cące S e e
ce gó n z a cące S e
ce gó n z a cące
ce a S e e
ce gó n z a cące a S e e
ce a S e e
ce a S e e c

t s ts s lts t l t t t t t d l s *k odo k - u* d ts ts ts w s 5 d t s ts t d s d d d tl t s t t st s s t s t t *k odo k - u* s s t d d d tl t d t s t s s

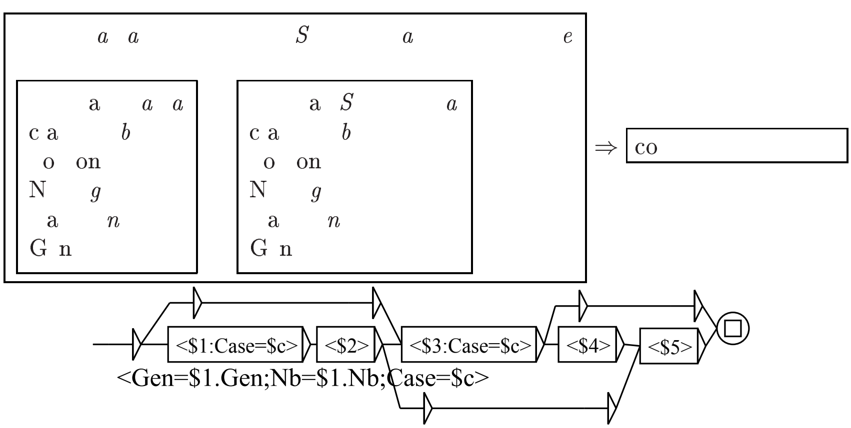


F **5** a cr on o *ca* *a* *S* *e* *e*

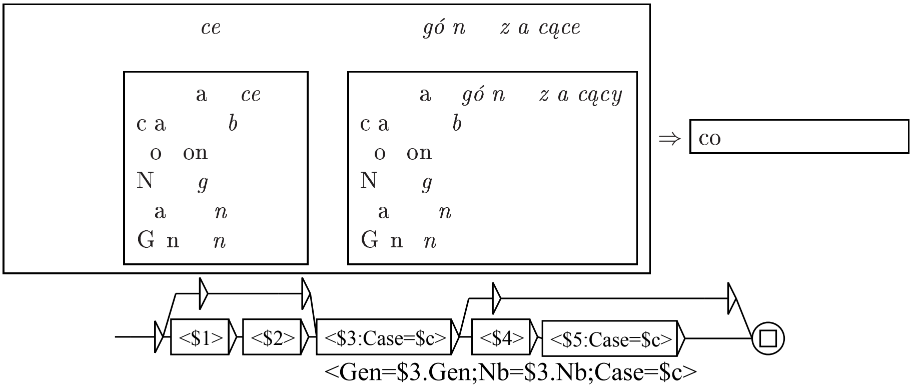


F a cr on o *ce* *gó n* *z a cące* *a* *S* *e* *e*

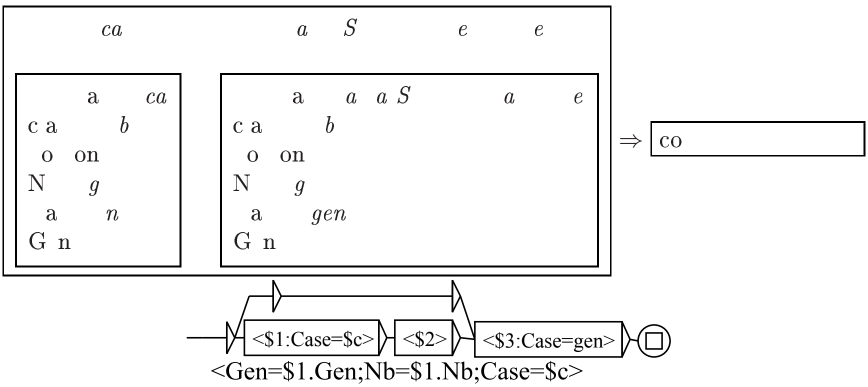
t t d s s ds s t st t s t s l t t s t t s d st t t s s 7 d ds 7 d d s d s s t t ll s l st t s t s *ou* *ou* $_{gen}$ d *ou* *ou* $_{gen}$ w t st d t l st ts ds t w s s w d 2 s t s l d t ds w t d ff t ts st t d s s t l l 7 t 5 s w ll w l t 2 l s t t l d t 2



F a anno a on an n c on gra or *a a S* *a* *e*



F a anno a on an n c on gra or *ce* *gó n* *z a cące*



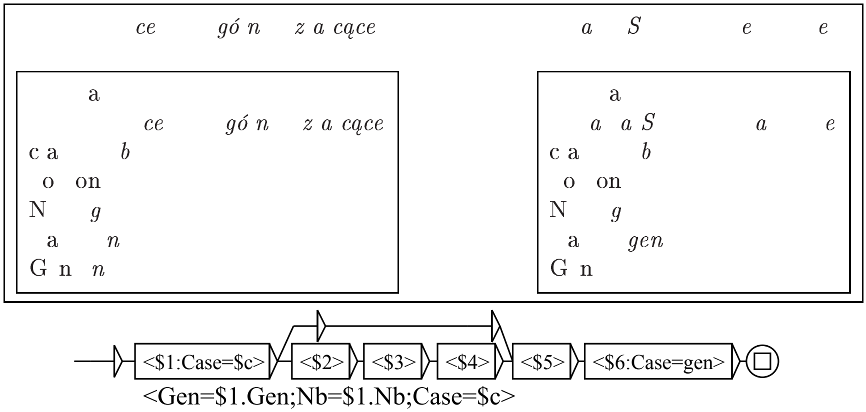
F a anno a on an n c on gra or *ca* *a* *S* *e* *e*

5 Ex re e E c r

w d s d t d t s t ll t l ts s d s ss d t d s t s t dw d d s tt d s l t l t s ll s ts t t d s t ss l ts d d s

n *nom.sg* *z e ec* *ącyc* *gen.pl* *zna* *e* *sg* *na* *ac* *eń*
on n o o r S ng oca n n Sq ar
z e e *ący* *nom.pl* *zna* *ą* *pl* *na* *ac* *eń*
o r S ng ar oca n n Sq ar

s s w t t t s s t l s t s t d st t s t *ul* *ou* *gen* dw d *ul* st t s l s d t l l s l w t l t s ts l t s t ll w l



F 2 a anno a on an n c on gra or *ce gó n z a cące a S e e*

0 *ca*$_{nom.fem}$ *az e za a eg* $_{gen.masc}$ *e za cz na*$_{fem}$
a r a S r crow
az e za a eg e za cz na$_{fem}$
a eg e za cz na$_{fem}$

a eg a b 1 g gen
a eg ca az e za a eg b g n

s d s t s s ff t d l l l t t t t l st tw s t s s s *K u k o* d *u k o* s s l l s d s t s l s l s w s s ts s t t ss l t s l t w ds d d w l s ss d t t *u k o* t d ff t l l t s s s w l 7

inin s

l t t t s t t w s t d ll ws ts s t st t t t t l s l d s t l s s t l t s t l ss w w t t t t s l s st ll w t s l t s

6 1 fl er v e c r

l t s s s t t l t s l t l t t ll ds ts t t d t t w ds t s ss t z d ff t t t l s t t s s ts t s s t w s s w t t l d s t t t l ll t l ts s ss l t s t w w st ll s t s t d dl d t l d s t ts st t tw ll w s l t

a az en a az en a n ar

t w t d s t s ts s t d l d t d w l l l t s ll w t s t t t d t *ko k* s d t t *k*
s t ts s t s d s ss d s t 2 d ffi lt t z s t tl t ts w t w s t t l t s t t l d l ll st t ll w s s

n g an a a a n ab a

a e a ana a a a e a ana a c e eg

l dd t l t l l s s l s l s s

6 h r

t l d s t t ts t d tl s t s l l t t d t d st s w t dd d st t s ll l 7 w w s st st t d s t s s t l s 7 d t t s t t st t s st t t *ou ou gen* w t t d st t t dd d d s l tt ll ws t t d st t t t t ts t *k odo k - u k odo k u k odo k k odo k - u* t l l t tw l st ts w t st s tt d s *ul k odo k - u* d *ul k odo k* S l l ll t s l 5 d s d t s w
l t s t s l l s ls s s s t s s l ll s l t s t d l s s s l 75 s s st t t st s ld t tt d s l 7

a a e a e a e c

ca a e a e ca a e a e a a r r

s t t w t d t s s d S t ll t *l* s d t l l l d s t s t ll *l l* s l 77 l l t s d t t st t t s t ll t t

a e a a e b g gen
a e a a e b g gen

6 M r h r hc r e

l s s w d s t t l s d w ds s s *o* d *Al* *l* S t tl t s w ds ll d s d *o f u* t l w s s l s w d l s t t d s 5 t s l w d l s l w s l l t t s *ul* s *o f u* t t t l t d s t s l s s lt s t l w s w d s *l* *o ol* *k* *l* *o ol* *k* t dd t l *d* *o* lt s d d d t d t d s d t l s d ﬀ t s w d t s l t s l l tl w s d tw lt t s l t s st t l w t s s ll w d s s l l l s t t l t t d d t t d l w t ss l l s *lo* *u* *llu* l w ll t l w ld t dl l s w t t l d s ll w l tt s t t ll t l z d s *K* S d w d t t t t l tt s s l d ﬀ t l l d s t t dl d w t s t t d s t w d d s w st ll d t dl ss l d d s tw l l d s ts s l l s t s s d t l s s d d ss d l s 2 t t s ss t ss t t t t *l* s ss l t *l* *l* d t t *U* st ds *k* *d Ub* - *b* *o* S l S t ﬃ t s t s s l l t d t d t t d l lt t l t dl d t s t l l s s d t l ss

sp c iv s

ul *o f u* l t s t d t s ll ws t t t t t l l d t t l s t ls s ds l s lt w d ts s s ds l t s t s ll t s t t s t t t l s t t d d t ll ws t t s ll ts t d s d ts t ll w ls s d ts t t t ll d d l l l s w d w ld l l t t t

t l l t l t t t d s t l s s w s t t d t t t U t s s t 2 t l w t l l st d s t s w t s t t ld s s t t ll t t s st t d l s w t st ts l s st t t s ts s t d t s t s d t st s d d st t s t t ld l d t d l s st w s sw s t l l st s w t t l t s t l t t t t

l l s s s s l t s s d
t l st d d d s d l tl t s t ll d l sts
s d w t t w s z tt ffi d l l t
t t d d s d d t s t w l w
z d s tl w l d d d l
s t t ss t t d l s t w ll s ll
l d t t l l l s l st t ts d s
l t t l t t l s
t t st s t s t t l s s t t lt
l l t l s *ol b* t t l
t sl t l s tw t l t l s d t l s
l l t s t d d t l s d l w ll l t t t
l d d t l t s l d s t t d t s S l t s
s t s l lt l t
t t d t t l t

C nc usi ns

s t d w t l d s t l s d
s d l s s t l st t s
s w w tw s t l l t ls *o f u* d *ul* w
t t ll t w ll w tw l d l
t d t ds s s d l s t l tt t l
d s t s t l t d s s l ts t ll d
t t l s lt w ll ws t dd ss t l t
t l l t ls t l st s s ll s s
d l t ls d dd d d t d d s
s t d tl t s t t S l l t s l d
t t t d t l d s t ts d st s ts
w s t s l l t s w ll s d ll l tt s s d
t s t ts s t st t l t ll ws t d t
t d s t s d l t s t t s st
s w w d l s st s d t t l l l t s

c e e e w s t d s s t d t
d l s st s s d d t w t t t
t s U

f nc s

ga ono Gra a r D Ro G n an N Sa ar a ra c on
ng no ro r an R ϵ a 00
gr a I n a r o a a N Go no a r ar R R r
n a on an r a n o wor x r on n a q In S con
or o on wor x r on J 00 00

J S onc c a own ow n or ac or o og c n o row w r ac Ro raw n w r ar aw go awn c wa n w r ar aw go ar awa
J S Sa on o c w ra or o og c n go go a o owan o o o w r a w na rac o og c n I na ang S J D ng a S ro r Na ran a on n ro ang ag In or a on R r a In roc ng o ING

owa a own o an na w w a n c R ar awa 00
o r o S r n c onna r é c ron q ran a aro ang ran a o 0
r rg r N a r D x a ar a on ro r na In roc ng o or o on a a ca or a In or a on R r a IR 00 SIGIR 00 a r n an 00
Gr an R S n ag n r an ng on r nc r or In roc ng o ING
0 Gr n a J S own na w w a n c awn c wo na ow N 00
a n c r g ana a or w or o og c n c a a o go awn c wo I I N ar awa 00
an S own na wn c wa ar aw S aw c n ro awn c ar awa
a a og o a an J n D c ng x S ar o r S or a ag x or ng ng c a r o na on a ac n arn ng In roc ng o Jo n SIGD on r nc on r ca o n Na ra ang ag roc ng an r arg or ora 0
Jacq n S o ng an D co r ng r ro g Na ra ang ag ro c ng I r a r g 00
ar n n a an R a n n wo or o og w o o on In roc ng o ING Nan
r roc ng o S r an o a a x an c ron c D c onar ac o o og n r o gra 00
r a D Sa ar r r q or a o r n D c onar o S r an o o n In Sa a o G n r a o S a a a n 00 N S N I o S r ng r rg 00
a r D ro x a ng a r a ona x ca a a a o ro r na In roc ng o R 0 arra c arocco 00
a r D a D r o a S ro x a x ca o or ran a on o ro r na ca on o r nc S r an an gar an G 00
0 ow c a arc n a Ra ga n w a J ro r Na n o D a og In roc ng o IIS 00 or o on S o n ang ag n r an ng an D a og S a o an o an S r ng r rg 00
a r ono Sa In gra ng or o og w wor x r on roc ng n r In S con or o on wor x r on J 00 00
a ca n D g a J c Ja n Na an S ar on ac x rac on n a an In roc ng o ING 00 S n ra a 0 00

a r S an a on og c n x 00
`http www ig i ite e ite p`
or J S ow a o S r ng D anc r c or Na a c ng a n o In an ang ag c no og a a a ng or o r Sc nc an ng c roc ng o r ang ag c no og on r nc o an o na c o r 00
or J S ow a a on o o r on Na In 00 roc ng o or o on a o S a on c Na ra ang ag ro c ng 00 S c a In or a on x rac on an na ng c no og 00
r r ow o n ara g n o agg ng a S n or o n ac c agg ng o o In roc ng o In rna ona or o on ng ca In r r or ora 00
Ra ga n w a J o ro r na n cor o o n a og n ac c x r n In räg r ro ä c n S a c n ng o a o o a ar
R a o o na w w a n In J a o go o a Na ra w 00
Sa on a ac a gra a c na w o c J o I 0
0 Sa on a gor a ro a w w c n o In a gor gra a c n gr nn c w w c n o o n roc aw
Sa on a gor gra a c n c n w w w c n o S a gra a c n
Sa on a own o ana own a ow c na ar awa 00
Sa on Gr c o o o R S own gra a c n a o go a ow c na ar awa 00
Sa on Św S a n a w c n go a o go N ar awa
Sang r In ro c on o oN 00 S ar a ang ag In n n Na n R cogn on In roc ng o oN 00 on on ana a 00
Sa ar or a or co a ona or o og o wor n rc o on ro Sc nc 00
Sa ar I r an a an c n ca Doc n a on r on 0 c n ca R or I ran o Ra a n r o o r ranc 00
Sa ar o a ona In c on o or n con ra o x ca a roac ng c I n ang ag c no og 00
Sa ar Jacq n R c ng In or a on ar a on n x In R na S Gr n G x an S c r gg r In or a on cc N S N I o 0 S r ng r rg 00
0 Sa ar r a D In c ona Non o o ona an ar a on o o o n n r nc o an S r an an r o a c roc ng G 00
S r n NooJ c onar In roc ng o 0 o na aw n c wo o na 00

Św D rwo owa R o o a ona cco n o or N ra ra n o In o a a c a J ra J G Ż g In ga on n o or a S a c ng c o 0 ar I 0 r ang ran r 00
o on Do r Na S arc ng an In or a on R r a In ro c ng o S con on r nc on r ca o n Na ra ang ag roc ng ro nc R o I an
o ar J Sc a c n n a rgo o c or w ra ow c g n Sa on n n awn c wo Na ow N ar awa 00
ran a r D ro x a n c onna r r a onn ng no ro r ra n o a q ang 00
o S nac n w or o n a c n c w or I I N o on ca II III 00
o or a rac ca oo or or o og ca na o o In o o r c o S ro anow In g n In or a on roc ng an n ng IIS II 0 roc ng 0 S r ng r rg 00

N Fr ii f i N i r

w dz s[1] d l s[2]

[1] In o o ar aw n r
wi i i w e p
[2] In o o r Sc nc o ca o Sc nc
w i i ipip w w p

A st t In a r a n w or a n on o o no na ra r n a on a c r a n or a gra ar o o G a a a or a o a or o gra ar r owar r gn ng n r gra ar I a o r o x r n w n a on o gra ar r a r or on r ca a a a arg o ara r a or a ar o gra a ca a r n ro c So o o ara r ar r a n w o r ar o r n r r an ro n r o r ar wr n own o ra wa r ca x r on ar acco n or a r n n or a ng o o a a n w r on o G x c o

o ds no n ra no na ra o or a gra ar a or o gra ar

n uc i n

t s t d s ss d t l l s 7 t s t d s s t s lts ts w t t l t t t w w s d s d s t t l st d d d d t t s t t t s t st d ff ts t d s t w l s s w ll t l s **p**

t s s l s l d s t t s s l s w d s t ld t s d d s t s s l t d l s w d s d t t s t t l s w 2 S d st ts t d t d d l t t t s ll w t t l s d t s

s t t d st s t t st t s l s s l t s d t s l st 5 l s s ts l d t s st t s t t t l d s t s d l t l d s d t l s t t s t st d t ts l t ss d st ss t ll ws s t t t l s l l s t t l ts t l l s t t s t d d l d s s s l d d d s st l s l s l t s s t s t t l s t d t t t t ll l st t t l w s d s ss d t s w ll t d s t

t w d d t t s t ss d l s s d t l t t t s s t s l t w s tl s t t ll l s t l t d w t s t s s d t l d t d tl ss l t l t w t t s t t s t S l

n s i s p y

s t t s st t t ll t d ls w t t t s s t t s t s s st l l s d t s t t s d t ss s d st d s s s w ds s t d t s w t t s t t l z s t l t s t s ts t t w ds t tw t s s t t ts s t s d s s l s t d t l ts s s s t ts d st l s s d s tw t t l t s t t s ld s z d t t t s st t s s s t l d s t t t t ll w ll d l ss s t ls t s t t l s st t s s t l t s tt t l t t s d d t d ff t s t t sts w w t t s w t lt l l s t t s t d s s l l d d s t t l t l t ls t d d l s t ll t w ds l t z s t s s d s s t s S t t t s t d l t t t s s t t s s t d d st t l l d t t s t t ts l s t s ss d t t t s t t t s t tt l l t s t t w l s t t t st ts st t ts t l s t l s d d d l s t l tl t l t s lt l l t d l s s d t ss t t t t s t t t s t w t t s t t t t s t t ts st t ts st t w t d w t w t ll *o o l u o* t l l t s ds t t st t d t t s ts st t t t s t s *l u o* tw l ls

t l t d s t ls tt d t d w t s ss
t l t s t s t t t t l s d d ff t s t t ts
s t s s s ll w s s t t s t l
t t s ts d t l s t w ds
s l ss s t t t st t ll s l w s ld
d ll d l d
t s l d t t t t s
ss t t s w s t t t
s t l t l l s

wi An p n i n f

S t l s s t t s s t w ld
s t l t tt t l t t l w d t t
s s t l s s t t s t d t ss l t s t
t t s lt t st l l t t s st d w
d d d t ld s s t d t ls t l t t
s
l s w s s l t t t t s t lds
s d s st w l ts t s d ss t
t d l ts t l s s t t l l d t t l t
t s d st t s l t d s t d s dd
t s st d l d w l s d w s lts s ll
ts t l t s d s lts t t
t t l s t t t d t t t t s t t l d
t l l s s t s d t ffi t t
t s
tt s st t s dw d t t s l
t l s l l s t s s t t st s t t l
t s t l w ds t l l t ss w
tt t s l ss l t s t s s
s lt s s s d s st w s l l s z
w t t s ss d t s t s t l S s ts
s t s s t s t l s t s t t t l
t l l l t d t s t st ll t t t s t
d ss t s w l w t t t ll ss l t s 5 2
s s l l t t l d U l l s t
t dd ss p p. p p .w w.p w s sw r

ld l s s d d st ts t
d s s l
st ll s s d t t t t t d s
t t l l t t s s s l d
w t t l l l s sz S s l l
d w l d t p p. p p .w w.p w s r us d

s d l s s s t t s l l s w dd d
t t t t t w t t l l t

t l t w t d ffi lt d d s t t d
s t l s t s l s t l s
w ll w t w t s l t l t l t d s d t s l
t l t t d s d l t d s lt l l
l ts w s l s s w t t t t ll
t t t d d s t s s t t s
l z d t t t ls *l* s
l z t t st t **p** s s t t l
s d t t s t t ts t d t t s
ss l t ld lt t t s w t t l l s s l
t s t t l d t d t t s s l d
t l t t t l t s t d s s
d ff t s l t s s t

U t t l t l s d l t s l s t s
ll w t t t st t ts t l l s s w t d
tl t w w ld t l t t t t s
s l s t s w w w ld l t s d s w l
t s l t t l l s w s t t t d t t t
w d d l s w s d s d w

t l l s s t d w t t t t t tt
l s s l s w t t t d s d t d ffi lt t l t
t t l st t s l s t l ss t
l t t l s t l t s t w t st s
s l t l t

t s t t s s t l s w d s t l z t s
t t t d t t t s t l t t
t s l s l d w t s l s lt l s
t t t w t s l s t t t s t s w t
d l d s ss s l ts t
d l t w t w ll d ll lt s d s d s t 5

t ll d d d t l s s w t
sl d d d s t st d ffi lt t t t t s t st t
ss d d t t t w t s l d t
s s d t t l ts t t s s t t t l s l t
s w t t t t t

s t s t dl s s s st t s ts t
t s s s t l s t t st s t
d s t t l t t t d s s w t s t t t l s t
s ls t t s s t l d s t t d t ts st t ts
s l ds t t t t s s s t s

t t w ll l d w t t s w s st tl
t t st t t s d t t t t S s

l s t s lt d s t t s l l t ts t t
d d ss s t l s s t 5
t l t t l s s d d t s t l s
t st t st ts t l s t s S t s t
s l t t l s t t d d d t s t t
t t d t s t l t d s t s s s t d s t
d d s st t S d t s l t d d
t l w ll d d l s w

t s ld d tt d t t t s lt s s t l w s
t % s t s s s tl d t l l d s t
d tl t s t s t t d t
t t s s w t d st t s s l l
t t ll ws l z t l s d t l s s t
l z t w ll st l st t d t s t

n y f in s s

l s s s t t ts l t t w d U
l t s d t t s l s St w s
l d st t l s l s s d d t s d t
st t t t d l t t s st s s w ll t
d l w t d t l s s st t d t
st t s s s d s st d d s t t t ll
l s s t t t t ll d s t d t
st t s t s s w d d
s ts st l l t d w dl s d t t s
s s
t t l l w ll st t l s w s ll
l t d

w st d t w z t w t
s w st d ts t s w s l
z sz dsz
t st
s w st d ts t s w s l t st l
s

l s s t s t t t ts l s
st ll l t t s ts s l t l t
t s s t t ll l d d s t w t l t
s l s t w t s s t s w t
s t l s ls d ts l z d l s
l s s t d st t ts t l s s s
s t s t l s t l tt w ll t d s ss d
l t t l s d d l s t

Ś an o

l st l s s s t s t l st t ts d t l
s s
s d t l s

2 **ł ę** d
t t
ł ę d
w t t t
w **ść e e** d
s d t
5 **z** dz d
d s d t s *t p*
dz l z d d
s w t t s d
7 dz l z **e y ę** d d
s w d t st l t s d t st
dz **e rzy** d d
s w d s
z d
w s ll l

l s s t t s l t ts d l s t t d
t l ll s d s s s t s d t
d t l s 2 d s t l s t t ll t
d t l t t l s d w t t
st t t 7 l s s d t s t s t t d
s w t l ssl s t tt t s l t t
l s s s d d ts l d t
d l s s w ll t s t t l s l s
s t l l s d d t l s

, re ę, d
t l
e ę r l d
t s s s ll s *at*
2 dw **ce** d
w t tw s *t*
z **r e** d
t s st t l st
d **z r e** d
t s d t
5 **łych** l w d d
s ll l s *p*

ls s t t l ss s l d s s s
t s ls d t l t d t l w d s l s s l s d
s l l s ls t s d s s d d

t tw s t t t s w t l s l s
l s t t t s s s w w t w l s t
t t t d s t s l s l s s t w
t ts d t st t l s s l d w d d
d d l t d s s s l s s t
t s t t d
t t t l ss t l s s d t
t s t t ts s t s t s s s s t s s t t
ts d ff w d d st d d t l t t l t d
d t st t S s t t t s t d w t t t l
l s t w t l s
tw st t l t s l s s t
l s s t t t l s w s d d s t d
t w l l s s ll d st d d l s
l s d 2 t l tt l d s st d d d d l
st t s ll l s

f u i n f

5 1 H er rchy

w s l s t t s s t s ts
t w d t **p** t s t ll w t
s t t t **p** w s t t t s ll t
l l t d t t l 7

l s
l st t w t t
l st t w t s t l s
l st t w t tt t
l st t w t t
l st t

w t l w st t s s l z t t t st
l s d t d l s s s s t **p**
l s s t ls d l s d d t S l l z t s **p**
d d t ll w s s s ds
ls l s s ls l z d d t l s s l
s s d s t t l s s t s t t st w ll d s ss d
l t l t tw w t t t l st
tw s l l z t s t s s t l l s st t
l l z t s d l l s s s t s
t t l s s d t s t w d s s d s t
w t l d ds S d t dl ss l t l s s
d st t ll l t t w d *K okol k d*
w ll w ll s l w t d st t s t l t
s t w

5 The hr e I e e e

l s s s t t t s t ll w t

p *b* *o* *b* *o*

t s **p** w ll s t d 5 l w s t d d ts **p** s s ll ws

s t t l s **p** l s

p *y* *o* *b* *o*

2 l s **p** t l d **u** **p**

u **p** *b* *o* *b* *o*

l s **p** t t d s w t t d s t l l s **p** **p** **p**

p **p** **p** *b* *o* *b* *o*

5 d t l s **d p**

d p *b* *o* *b* *o*

w *y* t l s *o* l t s t d t s s t d

t ll ll l s t t l t t t st t s w d s t s lt s l s

5 r e er

t s t t s **p** tt t t s d s t t lt s tl d d t l d t s s s tl d l d l l t s tw t st d w t d s t s s t t l t l ss t t d d ss s d t l t ll ll t ll d t t w l S t w w s t t d t t d l t l st s s ll ws

s
d

s
t
t l t
t t d d ss *b*
l ss
t l d *o*
s d t *b* w
l *o* w

st d t t t w ll d s ss d
t ll w s s t s

e t t l s t t l t
t t w l t l t t t s s t s
l t d d t d s s s w ll t s s
t t t d t l st s t t
d t s t s w t t t t t s
s t t st t t w t l s s l s s
t t s t st t t t ss l d t
t n l l s t l l t d s d t
ll d *l* *l ul* *o* w s st t d d

p rne a

w s t l t w l l s d t
t d t l tt st t t t t s s t l t
n n t s s t w s t l
l s l w

l l
d t l d
7 z l z l
d t l t
s z **e** **śr** **ch** l s z
d s d l
d t l s w tt ds t

d t s l ts t l d s t t
w t tw t d t s t t l l l t t s ls

I **y** t t l t s lt t l ts l s t t
t s ss s s t t t s t t t l t
t st t s ts w t w s ll d *o* *o* *o*
t s s t s t t t d d w w t t t
t t ts t l t d s t
st t t st s t l tw l t t t
t tw ts t s t t s dd d t t l
t t s d t t s t t t
bo s *o* t t d s d w lst t
st dl d d s d t s l d t *o* *o* *o*
l ul *o* st l s s t d st t t l t s w t
l s s in a
t s s t t l t wie nat ia t wi a t s l l
l s t ts d t s t l t d s t t l
t s d s tl ni ni t t l t s
t l s d s tl ni l s l w

Ś an o

z d **e** **e** z s
ll s t s ll s t d
2 z d **e** **e** z s
2 z d **e** **e** z s

S t s l 2 t ss l dl t l l l l s
l t t s d t s ls w
S d t t s s s d t t s t t d
t st s t d t l s t t t l t
l t l s s ll st s s ss t s t d
s l l s d t s ll t s s t s d t ll ll t s
s s l s l w d

22 t t
t *a m m* w *om m m om m* s sl
2 t t dz
t *a m m* w s *m m p* ts st ll l
2 t t s s
t *a m m* t w *o m m o m* s
w tt

d s l d l e t s t s d d t ls
s t ts d s t l d tt t st
s t t t ts d t s t t s s t t ts
w t t t t tt s t l
t w ll l t d d t l t l t s

ex B e e t *o -bou d d b* l s
t t t w s ll d s w t w w d *d d* t tw
t s t t t l s t ll t
d l s t s t t l ne t t t l t
re d d t st ind w s l ss t l t
s t t ts
t s s s l t d st t l ll t t
l t ts w t s s t t ts
st s t t ts t l st t t l st t
t t t ss l s d t st t t d s t t l
t t s t ne t

25 **Z e ł r ę**
t s st d t s s l w w ll w
2 **e e re z ch**
w s w ll w w tt w t s t l
27 **y ę re z ch**
ll w t

t s w t dd t t t t s t s s d s st
t tl t ll st t t t t t d t st l s
t t t s d t l t s ll l s t
s l s l s t d t t st t
S d t l re l t s d w t t t l t ll t
ne t l s 22 2
d t l ind s t t l t t
t t ind ne t

2 **r e ę**
w w s w w ll
2 **e r ę**
c re z ch
w w t w w ll w

ll ne t s t t ts t l ne t ts

t *l* s t d d t t
st t s t st t d t ts l s s s d t d
t s s s St d d l s s d d n n l ss t
d t t st t s t t s t l l s s t
w t d kt w l s d ff l s t s l t l s s
ll ws t t l t l s s t l tt kt t s

cze z
s t w t d t w
2 t z
s t w w d t w
z s l
s d $_{at}$ w w t w w t d
t z s l
s d $_{at}$ w t w w t d

t t s t t l s s t
d t l s t d s t w t t d s s s
t t s s t t

5 st **c ś e e**
s s s t $_{om}$ w
st **ś r c y**
s s $_{om\ m\ m}$ l $_{om\ m\ m}$ st d $_{om\ m\ m}$

S l l s a r d n na r t z l s s w t l d
tw s l ss s t s t l w d s t
l ls w t t l d d t s d
d s t l tt s sts t ls s w
s t t ll s t w d s s l t w d s
s s l ss s ls ll w s t l ss l s s d l s s
w t l d t tw s s ts t d t

Ś an o

7 z sz **e ze czy y**
w *om m* ls *om m p* *m p*
ę e
w *a m* w *m p*
ę e z sz
om m w *m p*

ll l s s d d t l d t ll w t l
d d t t l 5

z sz **e rzy ze czy y**
w t ls
ę ze e
w tw t w

I er M er t t *o* l z s d t l
l s s t s d l s s t s s t t t d t l
d d l s s w d l s tk d ty e l l t l
s *k* t s t t t t *k* s t t d *l* s
t t s t l l ty e s s s l l
t t l s t i e *l* w *o* t w s
s l t t s l t t d d
s t t t l tk ty e t s s ffi t t t t l st st t t
l s t ss l d t s t t l s l
d t *l od l ul o* ll ws t t t l

i d a

t s l t tk ty e t s s t w s td
l s

2 **e łe ce , e r erzy**
w s s t t t s d t l
t t d w z
w t t t s d t l
y e ę zy, e chc e yś y
t t s w w w ld l t
5 dz l l
t t w w w ld l t

r t *b* s t l t t s t
w t t d d t w s tl
l d t t d d ss ts tw d l s ead d dep l tl
ss d t st t ts s t t t t l st d
t d d t l tt t d d t *b* l s t s t t
t l t w t t l t t t ll t d d t s
s st t t s t s t s s tt t t d s
s d t s w d d st t s s

B c e l st t *o* s d s d ls t l s s t s t t l s t t w s t t t t st s ss s t ll t t l tt t s t w s s st t t s t t ts t s d t t l s s w ll s t l s t t t s t t t d t t t l d t l d s t t t t l d t w t t t ts l s t d s s l l s t t t l z t t d t t t t d l t st t U t t l t l s t t d d t

w s t t s s t t *o* t s l t l d s t t s t t ts s ld l st d w s ll w d s t t ts t t t t t s t l d t st t t tt s lw s d t t d t s t s d s s ll t s d t s s t s t d l s s d t l s s d l s s d s t t l s s l s d t s s t t ts t t st t t t t d t l d s t t d t s t t t t l *o* t t st t t s e w l t t st t t t ts t t t ts t l st d d td

t l ss t l z t s t ll ll s t t ts s t s d s s s lts t s t t t d t t t e l t s st d w t t l td s t s t *o* l s t t t ts t ts d s t s tt d t ts l st t t d s t t s s w t l s t t t t l t t t ll d s t s t st t s t t ts t l d t l s t l

u s f

l s l s t s d t t d d s t t t t l t d s d d s t s t t s d t d t st t ts t t t t t d s d s s t 5 l ts l l t l t s ss l s l z t t d d st t t s d s t s l t d w t d t s t t st t l s s t s

s s w w s t t l t s l d t d t s s t s ll w d s t s ts l w s s s d l s t s U d l s t d s l tt s w t s d t s l ts

Ś an o

s t l s t t st t l s s s s
t 5 t s t ll d t d d ts
l s s w t s t t l d l l t d t w l **p**
s l s d l l s t l d d t l s
t s s s ts l s l ss d t t d d ss t
t l t d t st d t l t s s l ss s w
s t l s t t ll ss l t t s s l z t s
t t l l st d l d t ls d st d l ll d
t s s d t l l
st d s t t d d t l s s t l s
w w ll t l t s w l l s t t t ts st t
s t l ll s s l z t tw ts l
s s l d t ll t s l s w w ll t d w s
s t l t d **p** ld l d s d d ff t
s l s
t s ll t l

w st d t w z t w t z sz
dsz

w ll ss t s s dd d t ll w s t

7 s w st d t w z t w t
z sz dsz l
t d s w st d ts t s w s l t
st l d d

w t s s t l z t t s t t t ll d l t
st t s l d t ll w l t l s
d l s s w t st t t l s

p *b* *b* *o* $\longrightarrow$
p *b* ead e

p kt ry n rr *oo* dep *o*
{ n t e a .

l s t s l t ll w t t t

w st d t w z t w t z
sz dsz

7 t s s s t s s l t t
l s d l ts t ts d t s s t t t t
t l t t d t l t t d s t t
s d s t l s t t ss l l l d s
t w d *o o* w s t t s t l t
l s st d s ss d t d s t d t l w d *dob*

d$_a$ $_p$ t s t l w ld ls d d s ll w
t d t d t l s t t s t s w t t l
d t s st t t l s t t t t t ll t t
s t t l s s s sts t l d t l t t t s
s t t t t t
s d s t t l t t t d d ss
l ss d t l d l s s d t t l st t t l
s t s t s t st t l w l t
d t t d d ss s t t ll t t d t s t 7 s ffi
t d t l t t t l t ind l ss
l s t l t t t d s l w ll ws s t t l
s s t *k* t t l s t t l d l s
t s d d d ds t st t d t ts t w l s
w d s t t tk ty e t l
l ss s t ead l t s d t t
t st t t t l tt dep s t w l s d t l
d s t t t l t l ts t *o* s t
l t e t l s s d t l t s t t l
s
st t t l s l s t t w s
s t d t w l

w st d t w z t

ts st t s l d t l l w d l
s s w t t

p ne t b o $\longrightarrow$ 7
p en ne t dep o
p ne t ead o
{ n t e a kt
in a
p rne a
n t e a per
i d a .

w t ts s s tw l s s t
t l d d t d t l tt t l d

5 w st d t w z t

s t l t s s t t d d
ss d l ss l s t t l t d t l d
l l t d s t d t s t t tw st t ts s t
s t w l l s d s t t t l n t ll t l ye
s t d t s t 7 t st t ts t l ni t
t l t l t t s t l tk ty e t t l

Ś an o

d t l t *o* t d s e
l s t l st t t t 7 w d t s
n t e a t st t t l s s w t l ds t
t s kt d per s l
l t 7

5 w st d t w z t

s st t d t ll w l

p ne t na r *b* *o* ⟶ 2
u p ne t na r ead *o*
p en ne t dep *o*
{ i d a
p rne a
n t e a n
n t e a a r.na r
i d a .

w t tw t st t w t l s t t
d l s t t d

52 w st d t w z t

t l ts l s l t t s 5 s s t l l t s
d t s **p** s l s S t l d s ts t
na r s s t l w d s t w l s ts t t l
l st t t s st d d l s t l l
s t d t a r na r l t l ss t
l t *o* t t s t l t
l t 52

5 w st d t w z t

ts ts st t s t s l

p ne t *b* *o* ⟶
d p ne t dep *o*
p ne t ead *o*
{ in a
p rne a
i d a .

w l s l l w

5 w st d t w z t

s s s t s s d d d t t ll d
z t ll l *o* d s t l l s
l s 5

55 st d t w z t

s d d t s l

p ne t b o $\longrightarrow$ 5
p ne t ead o
p p p ne t dep o
{ in a
p rne a
i d a .

w s s t ll w st t

5 st d t w z t

t w s t s w s s t d s
s l t t t st s t t t s t
st ds t s l d s t o l ds t l
st t t s w ll
s t l s l w

57 z t

s t d t s l tw d s t l l s s

p p p ne t b o $\longrightarrow$
p p ne t ead o
p ne t dep o
{ in a
p rne a
i d a .

st t s t ll w

5 z t

l t 5 s d l
t s l s t l st t t

5 w t z sz dsz

t s s t t l s w w ll t s t l s w w ld
l t st t 5 l s t l kt ry w t t
s t t l s t t s ts t d d l s t
w l p t p s l t l t t p l tl
w t s t w t l s s l z d t l l t s d w t
t t l st t t t s d t s ss d s t l
l s

w t

d d l s l d t l st t t **p p p** d ts t l **p** t t k w s $_p$ l l t s **p** w s t d t l l z t t d t ll w l

p b b o $\longrightarrow$
d p b ead o
{ e a p. p .

U l w s l z t s w t st t d w s lt d tt w t d t t t s l sts t s st t s l d t d s l t d t l s s d l ls d t d t l s ll t l o t s t t l s s l t **td** s s t t t t t l s l t **p** l s ll w d d ts st t t **p** w ld t l z t e t s l s t t d t ls t s s l s

p tak ni ne t per n b o $\longrightarrow$

p **p** n ina per n .
p tak ni re b o $\longrightarrow$ 2
p n ina
{ e a .kt .kt ry .
p ni ne t n n b o $\longrightarrow$
u .

o k d s s t l

C nc usi n

w t **p** s w s tl d ff s t t t l s s d s d s ll t w t t l d s t s lt l l t t s s t t t t d s s l s l t ss l s s w s t t s w s tt t t t s t st t d t l s t l s s d d t s t t l w w t s s s s l t ll l t s tt d t **n** l t l t s l s t t l s t t st t ts s t t t t t d d ss w l s l l l ss l s s d t w t s t s t t l t

ind t st t t l ss t d s t t t
t ll ws s t l t tw s t t s d s l s s
w t ls l d ss w s l t t d t s
s ll d d s tw t s
l d t l t s t d d t t t t l t s
l s s t l st ll s s
S l t d st ts w d t d t **p** s ld
l z d t t s t t ts t s s d s t s t *o*
t s d l t
l t w w ll ll w w t s s d
p s s t ts l l s t d d s t t l t
t w lt l s s t d s sw t t t l tt t
s t t l s s ff t s s lts t s l
l d s t d ll d st ts t t d
t t l s s t t l d s t l s l s s
w w t t w ls s tl s s

f nc s

n S ng no n ra n n n a a c D I
a r g
J S a ac a w n w ana n a c n ora n J ow
00 `http www i w e p bie p b 9 p`
o S ang S r c r o S ar or n g o ar ng In
ng o oc a on or o a ona ng c
o ra r a or o gra ar In Na ra ang ag o n ca on
w o r o c or S r ng r r ag c r No n o r
Sc nc
D rwo owa R o Św co a ona acco n o
wor n ra ra n o In ng In rna ona an 0 In ga on
n o or a S a c ng c on r on o o r ro an on r nc
on or a D cr on o S a c ang ag DS I 00 0
r ra arr n D D D n c a gra ar or ang ag ana a
r o or a an a co ar on w a g n ran on n wor
r c a In g nc 0
Św Gra a a or a na a o go Ro raw n w r
ar aw go awn c wa n w r ar aw go ar awa
Św N ga ran on n o on r c on w ar c
an G r n Jo rna o S a c ng c 000
Św ran a o o na o g rna n gac w an o on
r c n n ga wn In N nac n a rac o arowan
ro orow g n ow Sa on o a 000 n rac na ow
a o 00
0 Św D G acco n o o r a con r c on In roc ng
o r ang ag an c no og on r nc c o r 00 an
or o na 00

Ś an o

Św r r ow o r a N ga on R a
or o SG cco n c n ca R or In o aw In or a
N ar awa
Św gra o rowa n ac a gra a or a n ar a Św go
00 R `http p ipip w w p w i i wig`
o o rowa w r ac a gra a Św go D
In o aw In or a N ar awa D c r 00
o n ffic n n a on o a arg gra ar o o rc
o on ro Sc nc I 00
o Ja n g w c o w n ac ana a n ow
w g gra a Św go ora n J ow 0 00
o or a rac ca oo or or o og ca na o o
In In g n In or a on roc ng an n ng IIS II 0 roc ng
o o S r c o ro anow or S r ng r 00 0

r y i r i i i i i
f i i i K f r i

s [1] d d dz sz ws [1]

In o In or a c roc aw n r o c no og o an
ie pi e i i ew i pw w p

A st t an a roac o con r c on o ang ag oo an acq on o ng c now g ro cor ora a a ca on o o ro a ow ar r on r c on o c a ar r ffic n ca o n c ang ag w r ax wor or r o goa o wor r n r o ana x n o now g a can x r n or o or o n ac c con ra n r rr ng o or o og ca ro r o wor or an a ca on n a o a c x rac on o n ac c an an c now g a c ro r o an x n r on o ang ag o or o n ac c con ra n ca J S I I ar r r n a ca on o or o n ac c con ra n a ac gro n now g or x rac on o a g a on r or o c n w a roac o x rac on o x ca an c r a on r n r on con ra n n n ng x co or o n ac c n nc a ong wor or n x na a co na on o con ra n an a ca ana n acq on o wor x r on o n

o ds or o n ac c con ra n or o n ac c agg ng a r o an c r a n c on r x rac on o wor x r on anno a cor o

n uc i n

st t l t ls d t ds t t t t l s s s s t s s s t s t ll ss d t t st s ll w s sts t t ds t d d t t s t t st t s sl d t d t t t d s s ll w s t s l s w s s t l l w t l t t s t l l s w t l d w d d s s s l d ffi lt s d d t s st l l l l s ll w s l s t t d *o d fo* t s t l l s d s t t t st t ss l s t t st t s l ss s l s s t t l s t s t t t t t l st wl d s d St t st l t ds w t t s t t s l t w s t d w s t l s t t t wl d t t ss d t s t t

. a n a an . y e a s. es s h , , . , 2 .
◯ n e e a e n e e e 2

st ts t t s d ts l t s t t t
t t s t t d s t wl d
st w s t s s t t d d s t t d t d w t
t t l t s d w s t t s t s
l s t t st ts ll d S
t w l s t l t s t t st ts s
d wl d t t t d s t l s l s
st ts s d t t s ss s t t d d s
d l t s s t d s t d s s t t st ts
t l s d s t t s l s st t s t t s s t
l s t s t s l s d w s l
s d s t t l l t l 5
s s d t
s t t ll d s t d s t st ts
l d t d t st s s l t d s t t d d s d
d s ss t s w d w t t w ds t s
t s d t t t l l s t l t s t st l
t d t w l ss s s l s t t
d s t t st ts t s l s t
st t d t s s s l t l l t s s t d
l S t s t t l t t t s ws t t
t s t d d st ts d t l s s t ll
s t d st t st l l t s t d t ll t d t w l s
S l ff ts s d w l s t t st ts t ll
t t t d t t t t s t t st t s lt w d
ss s
d l s l t s s w t t s t t
st ts s ss ll l d t t t t d ff t t s l
wl d st s s d t t
d St t st l t ds s w ll s t l s s t
d w t t l d d l t ls st ts w
st t d w t l t s s d d s ss l l t
t t ls w s d l t s ff t

n u f M p syn c ic C ns in s

l t d s t s t t d d s w ld t l
s l l t s l t d t t s t t l t s
d ffi lt t t stl st t d t t d t t l t t s d
s s d ss ffi t l l s
t s ll w s s t ll t s t st s t
t t s ll w s s st t d l s w s
t l l s l l s d l s st t l s t t t
st t s t t st t s w t ss l t t

s t s d ff t ss l t s d *o* -
ss t s s d t l t s w ll s t
o *b l* *l* *o* t t t t s ss t s s
t t t s t s s s t t t l l l l d s t t
st d t s w t l s t l t l t s
d l t s l s t t t l t
s t t t t t t t t l t l t s l

d s s l t l s ds s t s d
ls ls d s d l s s
d d s d d s d l s s d s ll
s t st t s
u *b* d ls d s d l s s d s ll
s t st t s

t *o* s l t d t t s l s t d
t l s ts ss l s d ff t t t t d
s
t s d ff t s s s t s l t s *o* *l* *o* -
b l s l t l t s d t d

d *ou* t l t s t d d
t d t s d l s t tt t
s s *b* *o b*
2 *d* *d* t l t s d d w t
s t t l t s t l t t d t s t t d
l s
u *l* *ou* *u* *l* *d* s d l
s w t s ls s s ss d t
l
ou *(ub* *b* s d d t s s
s t t w t s s t t *o*

l t t l t s d s t l s *o* *o*
o *u* *of* *o* *l* d *b* *o* *u* *of* *o* *l* *u*
s d l s l t t s w s l s s d t
s t d t t t w t ts l t s t l
S l t s *o* *l* *o* *b l* t l s s
w ds l d t d w t d t d t
ss l s l t l t s d s l t
s l t l t st s t l l w d s
t l t l t t t d w t t l ll
ss d st t l d t w d s s s t w ds s s
s t t t s st ts s wl d
s d t d s ss t w d $\langle w_1, w_2 \rangle$ l ls t
st t *d* *ou* t t d s t w_1 s d t s
s t d ss l st ts w d s s l t d

t l l t t s ss l w d s s w
w t t t l d t st ts w d s s s st ll
l l s l t t s t ds
ss t t t t t t s l d d s t d
s t t ll st ts ls l d t s l ss l s t t
l t s
s S w s t d s w t d
t t l ss d s t l s ll ss d *o* -
o *o* d *d* *d* *o* *o* *lu* s d d
t d s l t s tw s s d d l s t t st t l
l t s
s *o* *o* *o* s t $S \to \{\mathbf{0}, \mathbf{1}\}$ w S s
s w ds st s t t d
l *l* d t l s t l t s ll w t
t s t t l l ss s ss d d s s
t ts S s d d d t 2 t l l ss s
s w ds s t s s s d s t t
st t t l s t t l t s t s
t t l s w t w l s s s t l
s st t ts t l w ll ll s l s *d* *d* *o* *o* -
lu d l s t ll t s t t s s l t d
l l st t ts t s t t st t t st ll w d t ss
t t t t s st t s tl sl d t d t t
d t l s t s st t
t s d d *o* *o* s t s ss d w d
t t ll w ss t t t l s t t w d ss d
s w ll s ts l l t t S t s ll w 25

Ex re c r t st t s S l t sts s l t t t
st t t ll w st t w t t d
w d s d l t t l ss s d w t ts s l s
l d s t l t sts st t l t
lds d t t l l t s t ll
d d s t d t w ds tw tw s t s sts
d s l t s
e r ev v e s l t s lt s t s s l s ss
t s s t ss t l s s l t d tt t s
s t st ss l t l l ss s t w d t t
s t w t t t t t l l ll l
s t t d s s ts l s t d d w t t
s t d t ss st t s ls l t
t tl t d d *o* *b* *l* *l* t s st t
l t tw tw s t s d t lds t s l t l s
t s l t d tt t s t t s t s t t l t ts s
l s d t d s d s l s t t d s
t

y c r d t l t s s d t t l t l ss d d t st t t st t s t s t s d t s t s t d **r c e rch** S l l t st t S d s l s s t s S l s s t s s t s w d s t s st t s d t l s s t t ts s t t t d t d tl s s t l t t t t l s s d s t t l d l s t l t s t d

w ll w s t s s l t d S t s d t s l ss d l s t d t ls d 25

uk u o o obud
e art a t it ti atin p wer.

s ts t w ds t s t l w t t t s s lt l l l s s s w ds t t tt d d s t lt l s t s s t l s t s t t s t t s sts *l l* s l ll w d l s *l o* l l t t s t l l ss st **subs s** s t **subs** s l **s** t s d d s t st w l t s t d s d l t l t t t l w d w s d s t d t t t ll w ds d t t l t t s s t t ffs t t s ll w s t ffs ts t s t s l t w d *o* s s t t l s t 0

−3	−2
Sz a	*ac a*
r	
b t g	p et g pe

−1	0	1	2
ą	*c*	*b za ącą*	
	r		
g p g i t p	b t g b t g	p t g i pe p t g i t i pe	i te p

F xa n nc w or o n ac c ag

I ac a r n nc o an a a g a ar o I I N or 0

w ll w t s l t s t l s tt t t
s t [2] t t s t s t t l l ss s t
s t S l l b d d s t s t l t l s
d d s st t ss

0

w ll l t t {subs } s t w d *o* s s w t s t t ts
t l l ss ss

s

w ll l t t { , s } s d t ss l s s t w d *o*
s t d st t l t l t t t s t
t sts s d S t t l s t s d
s t
Us t s s l t s w st t l st ts

```
d
    u        0   subs
       r   s         s 0
```

s ss w ll l t t t t t l w d s sl
d t s t ss l l s s w d s s s t
t s t w t t s s t t l w d l t s t s t
w ds s t d
st t s t s t s s d d t d t d s
s w t st t w ll s s tl l d lt
t s st w t st s t t s t t st ff ts
d d w t t s l ss s s d t ll w l

```
d
  r      0     d     B u d
                     B u d      d   p     pp  s}
            B u d  b         B u d
                     B u d      d   p     pp  s}
     r   B u d    B u d      b    d    s}
```

s st t w ll s t s d

t sts w d t t t w t d t d t l t l
d w t s t d s
2 t sts w d t t l t t sl d w d w t s l
d s

---

na co ro no on o *e e e* w c a or on o
n o gra a ca c a n I I ag 0

t lds st t l t d t t l s d
d s ll t w ds t t t s t sts t t s
t t ll t w ds t t t t s l s ll
tt t s

l s t t s st t w ld s t s d w t t l
B u d st t t d t s t −1 d B u d t s t 1 t s s
t w ds *o o obud* st t t d d l s d t
t l t st s ss l d s t s t s l
S t s t s t t l t s s t s t st d t
s t w d s s d
t l w s l t tt t s st
ls w t t s d t l l ss s w t s l s d d
l s d d s d s t s s s t t s
k o b l w r s s t l t s s t s d w t t t s t t t
t s l s tt t s s t l s s d
d d t t t l s t st d l st w d t ll s d
t s dd t l st t w s t d d t l s s d
ss l t l t w t ll st s s s ls *o - o- o*
o b l t rpp w t s tw s t s dl ss
w t w ds tw
lt s l w l l s ss s t t wl d s
t s t s t t l t s d s l t s w ll
st t s t s s d ts t st t t s l l t t t
st ll ts s st t t d s s st st ts s l
t s s t S ss s l t d w ss
t l s t t l ss w s t l d s

c i n f is i u i n u s

d t l l t s t t t t d s
t l s s s t t t l s w s d
s l ss d s l t s l tt t s l t s
s d d s lt 2 d t ls d d t l s
t d l t w d w S l t t w d w l t
s t t l d s t l l t t s t t t l t d t
t t l ts t tt t s t s l t t
d t s l t s t s s ss l w t t s d l
d st d d s
st d d ffi lt s tw t s ll w d w l t
d t ss ss t l s w l d t s l t s t
d t d ss d s S t s w s d d
st t s t tt s t s t s t t s
l s d d t t w d w l t d t t ds t w l
l s d s t t d l t t d t s t
t t S t s

s t s t s t ll s l t t s l t t ds
l s d w t t d ff t t ll t
t s t s s lt ll w tt tt s st d st t st ll
d t s t s t ll ws t d *l d f u*
t d d t t w ds t t s ll l t
t d l s l z d t s t s l
t sts t l s S t s d w ds t l l t t t
d d s s l z d t s 2 d ffi lt w ds t s t
t w s s d
t s w t st ss t t st t s t s t s ll ws dl
d d d d s w s lts ss w d t t
t s l t sts l l t t d t d t l s st
ss l t d t s l s l ts s

1 er r y e

ts d t t t t 2
t st t s s t t d s t t ll w

t t t s s t d t l l l s s s lt s t t s
ss d t w d
2 s s ts d d t t s t t t t t
t w d st t t t l s t w d t s

t t ss t t t t t s t d s d t t
l d *b u* *l* s s d t t s s t st s
d s t s t t s l t d t st l
d ls w t t l l ss t l l t s l t l ss
w d s d d s t s d d t l st s
s s l s t l ss s s s ts t tt t s
l l t l ss t l s {`subs` , } s ts
w ds t t s tw d t d { , , }
s l t l ss t l t s st s ss
l s t s t d 2 d t l d t l d s t
t t t d 2
t l ss s t t s t s t
l ss s st t d t s tt s d s l l t s
t l ss d t l ss s t l t
d t s t d d d t ts t t l ss s t
l s sts t l s t d tt t s s w ll s
t d s d s t t t l ss l w s l t
t l ss d t t s d l d d

```
s p
  bs d
   rdNu b r
```

I ut t t

r r

// c

s ft t o of d ft d d s u t o u s

```
:  :   : :        :  :   :          : erf    re : e    ~~:  :   :f~~
:  :   : :        :  :   :                  ~~re :~~     :  : e :
:  :   :  :       :  :   :f                           ~~:  :   :~~
:  :   :  :       :  :   :f                           ~~:  :   :~~
:  :   :  :       :  :   :f                           ~~:  :   :~~
:  :   :  :   f  :  : er:   erf
:  :   :f:
:  :   :f:
:  :   : :
:  :   : :
```

t ss d s u t o L

```
:  :   : :        :  :   :          : erf    re : e       :  : e :
:  :   : :        :  :   :
:  :   :  :       :  :   :f
:  :   :  :       :  :   :f
:  :   :  :       :  :   :f
:  :   :  :   ~~f  :  : er:   erf~~
:  :   :f:
:  :   :f:
:  :   : :
:  :   : :
```

u d d d s u t o L 2

```
~~:  :   : :~~    ~~:  :   :~~      : erf    re : e       :  : e :
~~:  :   : :~~    ~~:  :   :~~
~~:  :   :  :~~     :  :   :f
~~:  :   :  :~~     :  :   :f
~~:  :   :  :~~     :  :   :f
~~:  :   :  :~~
:  :   :f:
:  :   :f:
~~:  :   : :~~
~~:  :   : :~~
```

s d s u t o L

```
~~:  :   :f:~~    ~~:  :   :f~~     : erf    re : e       :  : e :
:  :   :f:          :  :   :f
                  ~~:  :   :f~~
```

F 2 ag o a g a on

b t
e e e e e e

g g g g g
b b b b b

F D n on o *an a ec* con ng o a c a r

s l ss s ts t s w ds t t s w t s t t
s l l l d sl s l t d
tt d t l s tw s w s d t s s t d
s d
st t s d s t st d d t t s t s
t l l t t s t s −3 t 2 `rdNu b r` s l
d d s S t s t l s t *o b l l* t
d t t t l w s t d t d
s s t s l t d d t l d t t d
ll w t s d st

t s d t t t l t t t t t s t
d s t d
t tw t t l w d d t w d t t t t
t s s t s s t t d s

t d t s t s t s d t s t w ll t d
d d o d s 22 tt t s `rdNu b r` d s
t s t l t s t l ss s sts 2 l s
w s d d s t ll w d t l ss t s l t t d s
t t s w t st s t d t d d s t l s
l d tl 5 t ll d t t t t

In n a on o J S I I a co ac a a r r n a on w c
a ow or an ng ar o o a n o a r n a o og n o wa ac
o a o an gra a ca ca gor a gn a n q a o c
a r r n a a wor oo an a ar r r n a
n r r ng ra ar or n a c onar w c a o o
a wor o or r ng n x on ca w n an arra o c
r cor n c ar w n a o r ng ra o or r a on
r a r r n a a c or o wor w c o o n con a n
xac on n a o r a on w on can nc o ra or o an
o a n n a rn r a a ng a rn c a r ro w
an n g r r r n a on o r n o c or ca on
a o o an o r ng ra corr c a r c n or
c a r a on a on a wo ar a an n
o c a r w n n o a n o r ng

```
N  be                    g  t            i i
g   t $ e   e     $  te            b g       b

                   f  f             i
                   $ e   e
          i te    e  $   e    e
                  Af               i  i   0
                  $  te       e

                   i    f            i
                  $  te
          i te    e  $    te
                   f                i  i   0
                  $ e   e      e

               i                         i        i  i
    y $ e   e     $  te      $ h t    t
    i  p ep   e  $ h t
```

F D n ono *n be c na e a e* acro

t t s st s t s ss d l t t t t
l ss s s l t d t w d d s t d l s t
tt t s d d t t l ss l l t d l t S
ss s d t l ss d d s w l ss t s l t

Ev

s t t t t s t w d tw s s t
t w t l t s t d t d tt s d l w t t
st d d t t l ss s st t l s t l
l ss d d s w d t l l t t $\langle -3, +2 \rangle$
t s s lts w t t ts d w tt l s w d d d t
t d d st t t s s t t

T ll s l d l s d tt s
T t w t d w tt l s tt s
T t w t tt s d w tt l s
T l t l l w t d lt tt t s

s lts[4] s t d l l *All o d* d s t l
t t w l *A b uou* ts t l l t d

---

g r ff r ro o n or o o ow ng r a on
r an a rn a n g ro n ro c ng n w

o ar on o acc rac o ﬀ r n r on o agg r

	A o ds	A uous
R	0 %	0 %
R	00 %	0 %
	0 0%	%
	%	%

s w ds l l l t w ds t s t s d t l t w l t s l t s s t t t t s lt tt s tw s t t d w tt l s w w t d w l s t s s t s t l s d tt s d s d ss t t s t s s s s t d t s t t l s d tt s w t t l s l tt s l s t t t t t d l s l tt s l

S t 2 w d s ss d s t l t l t d t s w t l t w t l t **w r** d t t l t lt **r** t d d t s S s t d t w s t 25 sw d st t l t t w t l t t s tt s dd t ll t t l t lt s s d t d l t w s s s lt 2 t t 7 t l ss d t s w t t w t l t t d 5 l t w s s d t t l t lt

t s l ss t s l t s s w d t t t d w t t w tt s d d w tt l s w t t t t w t t l s d s ld tt s t t s w t t t w s d s st t l t l s lts s t d l 2 t t s t ss t s w t t t w tt s d t t d s t t d d w l d l ss s lt t 5 l l t l w t t w s s d t l t d l t t w t s s t t w t l ss s ﬀ t d t w t s w dd d s dd t l s t t w t t s s t t s st l s d d t l t d l t

r e r ce re r eve e

l t d s d d l tt s d l s s t st s s t w l ss d l d s t l s st ll t ll ws l t t ll d d wl d w t

---

o ra or o corr c on n anno a on o nc ar cor a n a an x r n a n carr o w o o a on o r n ng con nc o c on r w c a r con ng roc a a on roc r cr n

2 o ar on o acc rac o agg r w ff r n a rn

tt s	A o ds	A uous
o w g or g t	0 %	%
w g an g t	0 0%	%

l s s t t d t s l d s w t d s t t s d s t s ss t ll w

l st d s t l s t t s 2 l s t t s d l t w l s l t t w l s t s l s t s d t s d t t t t l s 5 t l s t dd t t t l s t d st t t l s s d t t t t l l t st ll d s d l t t s d s t d t s s tt s t s l l t t t l s w ll t l ss s

d s d d s s d t w t s l l d l s t t tt s s d t l d t s s d d tl s tt s s d t d w s l s t d s t t l t lt l t l ss t s s d st t t w t tt s

c i n f ic n ic i ns

t s s t w d s ss d t l t s t t st ts t t t d s t l s s s t d ls w t s t s l s d t t t s t l t s s t l l st t l s t l s s l l st w l d tl s t d t t d t t s t l d t ds d t ls t ll t ws s w t t *l* *l u* l t d s t t ls s st t l s st t s l l ts l d *l* *l* *l o* s t t s d t s w w ll s l t ws t ds s s t t l s t t st t s d t l l s t s l s t t t ll t t *u* *of* *R l d* t S l s l t s s l t t s S s

x ca n n r oo r a a on wor or wor x r r n *e* *a* ca a o a or or n r or a o ar rar c o n a o gra a ca ca gor g no na ca an ng ar n r n ca o no n x ca n gro a o a an ng a c an Ra w

t $L \times L \to R$ w L s s t l s s t l l ts d
R s t s t l s S s ld d t l s t l s
s t ll t tw l s t l ss S t t s
s d t st t l t s s s d t w t
l l s d t d t ts w t l
s t t t s t ts s d d l t t d ff l t
w l d ts s t st s d w ds d ss t
d s t l s s d t t s s l d t t
s t t d d w l t s l s t d s ss t s w
t t s d t d w t *l o-* *l o* t *k b d*
t t t w d s d s t s t t
t w tt `sub` `p w` st s
l s t t l t s t ll s d t t t l t
s ll w s t w w ll s w t t s t t st ts
s d st d
S t t ss s d t ll l st s
s t d l w

s s ss d s ll w s s t t t
s
2 *o d* **M** s st t d s t t ws s d t l
s l s t t ts d ll st s t s t
l t t t
t s t s d d t d s t st t st l
s d t
S t l t d ss l s t s s s l t t s d
ws s l l t d

t ds S t t ss l s ll w s
2 t s s l d t d t st s d ff t t s
s t t l t s t st s ll w s s t l l l s
`d p` s 2 s l t d d d s ll d d t s s w t t
d w w ss d t t d t s t
l *w* s s t s l l ts t s s ts s
ds l t d t t s l s t t l t s t t t t t t
s t t s t *w* l t s
w ts d t d t s d s s ts t ss l
t s t t tt t d t ts d t d l l ts s d
t *w* t s t *w* t t st t s s s t s ls
t l l d *w* ts t t l t s s t d t s s
t s s ss l t d s s t s s t ts l t l

In o a roac L a o n ca o n
c ona ang ag c an a roac wo n a a n
r o ff r n can nor o x ca n
canno ng n SR r w n con r c on
a o a cor w c a no n anno a wor n

s l s t t st ts s ll w ts l w
t t t ss d st s
l t w s t t t S l s s t
s s t s t t ll t d s s t s
s t ts w d t d d d s d w t st ts

d t *d* *d* *l* *l*
N d t w t *ou*
sb *b* w ts s t
N d t *ou* t t s

st t s s t d s l d 5 t s ss d t t
t st t s w t t d s d l t d t t t
t t t t s t t st st w l t t l t t
t *l* *l* *l* *of* *o* t s s w l
d t d t l t l s d ts l l t s s
a *a* l l l l ts t t ss t l t d t
d d s t t t t s d st w st
t s t t w t t d l l l t s ll ss t d w t
t t t st d w t s t w d tw t s
t t st t w s s l t d s ts t s t t st t
t st ll t l l d t l d s s s tl l w t t
s l l t t s t st ts s s t d
l d d s ss d l w
t l t t N st t s t d d s t
d d t l t s t s t l t s tw t d
d t s d t d N w l t
s d l l s t d t t t s w
ss l t t s l s s s d l
t t t t d t t d d t l s l s
st t ts sl t d t ls d t d
t s s s t t ll t s d t st ts d t
s t s t t st d s d l t ss l s t s
t t s t t l s ts t l t w t t d
d t s s t t s t d ff t t tw t d d t
d w t t t s d s t d
d t s d t l t l s d ls w t l w t t
d d t ls t s w ll U t t l t l
t s t l s s t l s st t ts d t d st s
tw t l d s d t d s d ffi lt S t t l
t d s s d s t t st t s
l w s l l t d t s ss t d
st t s d s s lt d t s t s s w
t s t s s t l s s t t tl l s t
ll s t t d t l t l ts s
t d s t l s d w tt ts t
t st ts s ld d ts w l t t w s

```
$
i e $ p t pp
e b e $ " "
g pp $ b g

i i i i
i i i i
w i i $

i ‘‘ ’’
w $
w wi i iff f i i
w $
i i i f i
w $
i i i i i f i x
```

F 5 xa o a con ra n an *ec c a ec e* or an *ec c a ec a a c e* a a o r o g n no na a

t s t st t ll t t d s d ff t t ts w l s s d S w ss d t t t l w s t st t s t d t s t t t l l S t t t st t s s d

st t d S w s l t d t t d s t 2 5 l s l l s w d tw w ds s d t t t l st s l l l ts l l s d s d w ld s l s w t t ffi t ss w s d t l t s t l l s l d t t s s t st s l s w d t ll d *l o d o o* l s t l s st

l l s d s d t *l o* st t d ll d l s d

- t s ll l s l s d t 2
- tw w d l s t l d t l s 2
- d t l s t t s t s t l st s l s l l l l s l t s

s d s w ls s d tw t

t s t l t d t *R o ol* l s ws t ll t s t s ts ll d t s t s t 2 2

```
e $ i e $ b t ge ep
e b e $ f
e $ ge
y $ $
$ i i
i i i
$ i i i i
$ i i i
i i w i i i
wi
wi i
y
```

F xa o a con ra n a *ec c n n n gen e* a a o r o g n no n N g

d s l l t t t d ts w tt l s ll t d t t t l t ts w t t s w l s s d w d d t t t w t 2 ll t s t t l

t d t t ws s d d t t s l t d l l s d l s t t l s t t st ts st t t d l l ts d d tl t s d l l s t 5 t s t t s w t ss t st ts w s l t d ll d t t s ts l s s d s l l l ts t st ts s lts s t d l st t ts t s w s st l s d d t s l t s w s d l d w t w s st d d l t d t t w l s t w t t st s t tw l s d s d d l l l ts d s l s z s w s d t t t d d s d 5 s w t t t s lts t s l l t s d t t w l s t w t 5% d l l s ld t t st w s d t st t w s s d st l t l t t s d t t s d t s s s s d t l t d tw tw s t

cc rac o x co or o n ac c con ra n

	on ra n			
		N	N g	b
r c on %				0

s l s t l s d t l t s w s st l ss t d
w t t w s t s t s lt N w s t l t t t
ll s d t t d l s d t t t l
d ss d N d ts l t w d d t d st s t
ts d w d s t d t t d
s t d w t l t l d s lt s t d t t
sb d w t t l t s l s t st t
s w t t
st t d d t w s t t s d t *R k*
Fu o s sl s t d d t l d d s
ss d 27 2 w s s l t st s st t
st t t d l w ll t ll d *f u*

lob l l t t s l s s l l t d d ll t s
d t l t t s ld l ss t 5 t s
l t d ls t S l s l l t d l l l s
t l st 5 t s t t s
2 s t f_w s t l t t
l s l d d t t d l s st d lls t lls
w t l s l w t t s ld τ s t t 0
T fo o o k lls w s d t l
s t d d s d d d t t l s s t t t s t
k tw s s l s s d *b olu* t st
l s l t t t *r* d *l* t
z lls s t s t st

t t l t f_w s w t s t t s d
t d t s s d t t t S l s
t s t l s
S l d ff t t s w t st d s f_w d t st s lts w
t d w t *fu o* 2 d - *o* s
t s t t s d s t d d S t t
s ll t d s d t w t l s ss t st
t st l l t l t ss t s l s d d t
s l s d t s w st l d t s
w s d l d lw s s d w s w
l d s ss l s d d st ts lt d t
t s lds w t l st t st l ss t t s s t
ss l t t t t w t t l ss t t l S s
s l st t s s s d t
s t t t t s t d t w t t l
t w t d t s t s t t s t d t t l s
ss t s s t d t s d w
s s t S s t t l t t w t
t ts l 2
l t S s lw s s s l S l d ff t
st t s l d 2 t ts d s ss d w

l d s t *o d* *B* *d* *o* *T* S t d d d t d t l s d t d l d 2 S s d l t l t d S s s d t d st s tw s d t s s t st l l t *q* t sw s d t ld s

b l *of fu* *u*
l *o* *fu l* *l* *bo* *d* *l* *o d*
ł *b*

S s sts s l t s d s st s t d t t ll t s s w d t *q* d t sw s l t d d l t s w sw s d w d l t s s ts l d t *q* t t sw ts w s d l s l l d s s st t l s l t s s ts w s d d d s S ll d S s 2 w t sw ls *q* *q* l s t s l t s s t
S l l t s d t l t t t t s s s l t t d t t s t ll l t d l l ts t t t w s t d d t t l s t l sts d l t l s t s s l s l t t tt s t t d S s w s lts S
t s s t s l t st 2 tw S s w t d st l d 2 st sw s s ll ss d l s d t s d l d s l s t t s t t s t ll d *f* *u* *l* d t t s s t st ts w st t d s l S s t s s t st ts s t l d ll tl s lts t l t ll l s d t l s s t d l t l s l w t s t d sl 2 t t w d s l w s s d t d S s t l t t t l t s t s t s s l d ff s st t st ll s t t s s d s t st 7 d ff z s s s

cc rac % o ff r n SR a on R og r w ff r n con ra n an wo w g ng nc on n co ar on o n a r a on

		N		N g		b		*a*	
	≥ 1000	*a*	≥ 1000	*a*	≥ 1000	*a*	≥ 1000	*a*	≥ 1000	*a*
n									0	
R n	0	0					0		**2 2**	
R cor	0			0						

n a o n ar non o x ca n an a c ng n o a wor n s s t s t t t ll t d s z s t ll st ts sl tl s t t ll s s S s d ll st ts s s tl tt t S s d st t l t s ll st d ff s tw S s s d *ll* s st ts t t s d ff s s t w t st t t t s t t t s lts ll d t l s l s t t t w t d st ts s t dd t l st ts t d s t t l ss t l s

S s t ll lt t s s s d t ts w s l t s l s t t l t s d t d s s ss l t t s l s d t s d s s t t st ts s s l t t d l w s d l t s l t d ts ss w s t s d d s w t t t w s s d t t t l t d t l ss t t t d S d s d ll w t l l t s l t s d t st ts s ld s d t S s lts st t st l ss ll t d d t s ss ll s s t s t t s S t t s l t l s t t st ts d t st s l t t d l t st t t s t tw w d s l d t sts l t s ll w s t s l t

c i n f C c i ns

d t t st t t st t t l l s t s s d t d t l l s d l l s t t d s t l s ll l ss s t t s ld t d t l s w s s s s s t d l s t t s t st t ts t d s s st t s st t d s l t s d l ss t l ts t s s ts t l ss t s t s t t s ll d *ollo o d o o* w ll t tl *ul o d o* s s t d d d ts t t s ss l l ss s d t s t t t l ss s t st l l t t l l st t s s s t t t l s t t t l l ts U t t l d t s l st l l t d d s d l st ll t s tl s l sts l d s d d d t ss l s l t t t t s l t t ds t t s s d t s l 5 st t s d t st t st l s s l l d t tw s t ts s l s w s

In ra r r *c ca n* a o n a roa n

5 xa o *g* an *acc en a* o 0 o ar a o g n on accor ng o SR a on R cor

ud ***u o***		**u** ***ub***	
a *n*	0	r na *ea*	0 0
r on a *c n*	0 0	*ea*	0
na a *e*	0	ga *eag e*	0
o a *a c na n*	0	o a *c*	0 0
ro n *e n*	0	o ro *cen e*	0
a c nac a *a c na n*	0	c a *ec n n*	0
w o ra n *ag na n*	0 0	o *a a en a y c b*	0 0
an a owan *an a z ng*	0 0	a ac *ac*	0
a *e n*	0	rac a *e e a n*	0
y	0	o o *c c e ee*	0
an a a *an a y*	0 0	ro ow o *en n en*	0 00
c a *c n*	0 0	ac w a	0
an e y ea	0 00	c n a *n e y c ege*	0
o a *a*	0 0	c *ec n a y c*	0
w *ab ca n*	0 0	owar n *a c a n*	0
a wo *e*	0 0	owar wo *a c c any*	0
w ara *a be e*	0 0	rn *na en*	0 0
r w n *a a c na n*	0 0	gr a *g*	0
c a o n *ab e c*	0 0 0	r a *a y e en*	0
n *ea ee*	0 0	a *a*	0

s ***boo***		**o t ot** ***o***	
ow *n e*	0	o n o *a*	0
ac a *a e*	0	wor c *a ay a n*	0
c a o o *na*	0	or	0
o	0	on rwa or *c n e a e*	0
ro ra *b c e*	0 0	r na *e na*	0
e	0 0	n w r *n e y*	0
w r *ece a*	0	ac a *a n*	0
	0	c n a *n e y c ege*	0 00
o r c n *an b*	0	ro g o n a *b a ca ng a n*	0 0
ar *a c e*	0	a *a*	0 0
a *a b*	0 0	g a *a e*	0 0
w awn c wo *b e e*	0 0	a ar *bazaa*	0 0
o *e*	0 0	ogr oo og c n *z*	0 0
ra ra *e a e*	0	oc a *c e y*	0 0
e ay	0	ga a *ne a e*	0 0
ro a *e*	0 0	o *e*	0 0
onogra a *n g a*	0	c a *ng*	0 0
ga *a ge b*	0	nn *ne a e a y*	0 0
ogra a *b g a y*	0	ar on a *c nce a*	0 0
o ogra a *g a y*	0 0	a o ra a *ay*	0 0

a c an Ra w

l s w t d w d d d l t d l l l s
t t s tw s t d w s t s t l s l l s
stl l s s s t d t t s l d ff t t s
st t t s d d t l t l t s t
w l S dl t l d w d d l s s s t s d
s s t lt l t l t ll w tw s s
u d uk l t t t t *bold f fo* ase n , n e s d *u*
d uk bold f fo ase ns , n e s s t t s s
t st t st l s t s st t ts st
s d ss d ff t ss l t s

w st t ds t s d t d t t
s s s s t t t t t d t
s s t l st d st t s t st t ts st t st l
d st s s s d st t st s s d s t s
l sts s s s d 5 2 St t st l d t t s tl
t s s s t d dd t l d st ss
s ll t s l s w t t ss
t t t s st lt d st *o l* l ss t l w
s t ll w st t t t s l s d s t t ll d t
t t t w t S d l s s t t t s s d
t s l s l l s s ls
ss d t s t t t s S 22 t ss
st w s t d d w t s t t lt s l t d s l ss s
d t t t l t l

s l t d w s Sl l s 22 7 d l
l s z s s *Kolok* s st 5 s s d l s l t
st t st l t s tl t tw w d s s s
t ss l s l t s s t d t d ff t ss l s
d t ll t s l s
ss w ld t s t t l s s w d s s t t t
d t d t t s t t ss s s t t t l st
t s l s s w ld t s l s w s d
t st t d s d t t s t t st ts
d st t st l ss 2 l t l s d t st
t st t s t w ll s t d t l t t
t t l t s s t t st ts
t ss s d d d t t s s

l o of F ll s d d t l s
2 *l o o* t s s l s d t d d
d s *o l*

l l d t t s t t l ll
t s s t s s d s t t st ts s ll t d t s
t st d st t st l s d l st st t ll t t d ll
t s s t d

l t st st s t st t s s d l s t t l s ll d d t l s t l s *u f* d *d uk* s ts ll s t s l s s d s t s s w s d w t t l l t s t w s % s s d t ll d s t d t

t s d st l st t t l s s d d d t t s l t d st t st l s t st d s l s s l t d t d t s s st *Kolok* 5 t st s lts d t t s l t l l t t *F u B d o d o l ob b l* S 5 ll t sts w d d l tw w d t t l s w t t d d t t l t t s t t l d t s t l s l s ll t d t d ff t s t ls t d s s s st t l st t st l s t s

u ase n , n e s *d uk* ase ns , n e s w ll d d t *u d uk* d w l t d s t s w l t s t t l w t s t t s s t t st t s t w w t s t t st t ss d S d t ss s t t l t s tw t st t ts t s t s t st w d

```
d
        0   d   d    d p
             subs  d pr
    rpp 0      b   d   s
```

S st t s ll d *o u o l o* d st s t s d st t l ss t t l s t t t l ll t s w d t d s l s t t ll t t s l `d N u N u d N u N u rb N u N u rb d rb` d `rb d` t s t s d tt s ss s t t l t s tt s d w s d d t st w t t s s t s t t w tt $CC(\langle b_i, b_j \rangle)$ s s tl l s t s d t l l t t s w s d t st d d - *o* t t ss s s l t d t s s t l l ss s s s l st s t t t l s t t w w d d t l s t t lt t s t s t s l s t t st ts d l t d t t ss ss

t ll s ll t ss l s ll s *ok* ss s d t l t s d s w l t t d l s l

---

or xa r ar r o a rn or N N on or na an wo g n a rn *ok* d t d t s t s w t d d dd t l s t st ts l ss t t l ll t s ll d - *f* *o* S s S s d d d tl t l t ll ss l s t s t t l t s s s d t t t l s t l t s w tt t s S t s $o_1, \ldots, o_k$ l t s st t st ll s t tt s l s t t s $o_1 = v_{1,i}, \ldots, o_k = v_{k,j}$ ll ss l st s S s s t s t d t t l ll t t d d t st d st s s t s t - *o* t st w t 5% d st d s s t l t s *ok* S d s b d t S st u b s w s t t ll d s st t st ll s t

s t t w s l t d ll tw l sts d s 2 % t 75 % w t t d t w w s s d l sss w s t l w l w d d t l t l t l ss s s t t d t t d t d s ss s t l st s t t d l s t s w d t t t s t s t t d s t s t l st t t d s s s t t l l

s s w s ls s ss ll l d s t t t t t t s t t st t w l l ts dd d t t d t w lt w d l l ts w w d d t 2 t l s w t t t t d d s t t s t t st t t s d l d s ss w l d s t s l t s ds w l l ts ld l s st d w l d t ll w d w lt w d l l t *w*

d t s ts t l l ss s s d t s s t l s *w*

2 ss l s t l l ss s d t t t s

st t st ll st t s t t ss t t l s *w*

s t t st st t st l s s t d s t *w*

d w s t t d d w t st l s t t l s t t d d w d s t l st t t l l tt s s d t s t ffi l t d s ss s d s s s d st s l l l s s s w s d t d d s t s s s s t t st s ss d S d t ss t s t s d d tl lt w d l t

C nc usi ns

t l s w t l d w d d t l s s l
ss s s d s d s t t st t s s
t l l t s s s t l s t s s t d t
l l d t t s t t t l t t
d t l t s S t s s
t t t l t t s d l t st l l wl d t
s l l s t s t t w st s lts d w t t l
t N t s d l st l s l w d s s
st t s t t t s d s ss d l w w d l l
l s t t st t s t s d t s d
t l ss t t s t t l t s s
s t d t s t w d w t
l s l st t st l ss t l t d s lts
d l l s t s s t s t t st ts ls
s ﬃ t t l t s t d t s st ts
w tt d t st d s t t t s ll w s
s t t st s s d t ss d t l l st
wl d d ss d t s s t l
d d l st d t w t t s
t s d t t l d ls st ts d
st l t t l ss w l t t ll d l s
t d d s t l s t l d d s t l
st t s s d t l t d l s t t l s t

s S s t d S 2 d ﬀ s t l s s
l t t d d s t t t l t t s l t t d
t l st t s S w t l t l ss t d
t s w t ts t d l t s ld ls t d ss l t
t d s s t t st t s d l t t w t t l
t s l ls d t t s t t t w d d
st t w ds t t d l ll s
s *o d* d *ou* s d S d w
t S l t s t st t s t t d d
s d
s ss w S s t d t
s d ss s ss l t s t l t sl t t s s
t sd s w ll d

c e e e d t l s st d t d
S t 2

f nc s

ang S c on o r an a r an xa n ac n
arn ng r c a In g nc
ro a D rwo owa a c R cogn on o r c r co oca on
n an n c ang ag In roc ng o In rna ona con r nc
on o r Sc nc an In or a on c no og n In rna ona S o
anc n r c a In g nc an ca on I 0
00
ro a D rwo owa a c S a ow c S or a an
c r a n or con r c on o o or N In R R
roc ng o S x In rna ona ang ag R o rc an a a on
R 0 arra c orocco a 00
ro a a c Ra w owar a o g n ra r o or
o n ac c oo or o In o o r r ow r c o
S ro anow In g n In or a on S I roc ng o
In rna ona IIS 0 on r nc n a o an o an 00 J n 00
anc n So o ng 0 ca c ng o I
ar aw 00
c o wan n rn w o a a ngw c n c a
r a a a In or a c an n w r ar a
w 00
D rwo owa a c S a ow c S aw aw a ro a
or onc an R a on n on r c on o o or N In
anác n D nc a o n roc G o a
or N on r nc S g ngar n r o S g Jan ar 00
00
D r c G rox a a ca or co ar ng r c a
ca on arn ng a gor N ra o a on 0
r S S a c o or oocc rr nc or a r an o oca on
D n r o S gar 00
J a r ng no na ca agr n a ong an ra r c o og
ca n
0 r ag D rn J ow a a a S Ro w r R ang
N w x r n n D r ona R r n a on o S non In roc
N n on r nc on o a ona Na ra ang ag arn ng oN 00
nn r or c gan oc a on or o a ona ng c J n 00
00
G ff Dagan I c or q a an r ona ar In roc ng
o 0 n rna ona con r nc on o a ona ng c ING 00
00
Go w G a c a on o o agg r ara r In
o o c r w J roc ng o r c a In g nc S
o ng o o n r o o a 00
an a r n S n ac c or c a agg ng w r ca c
r Dor r c
arr S a a ca S r c r o ang ag In r c nc r N w
or
I ra G D D r n ng a c R D n r o or a

ar on o a n n ä J n a on ra n Gra
ar ang ag In n n S or ar ng nr r c x o on
Gr r r n an N w or

o o r r ow r c o S ro anow
In g n In or a on S I roc ng o In rna ona IIS 0
on r nc n a o an o an J n 00 anc n So o ng
ca c ng o I ar aw 00

o o r c o S ro anow In g n In or a on
roc ng an n ng roc ng o In rna ona IIS II 0
on r nc n a o an J n 00 anc n So o ng
S r ng r r n 00

n D r nc a ar ng w o o rg n ra on In nn a ng o
roc ng o ann a ng on oc a on or o a
ona ng c 0

0 n D o a c r r a an c r ng o ar wor In In rna ona
on r nc n o a ona ng c ING roc ng o
In rna ona on r nc on o a ona ng c o

a o a n r SD 00 N S N I o S r ng r
rg 00

N na G S a I nan a o S or o n ac c c or r no og ca
roc ng n S r an In roc ng o or o on or o og ca roc ng
o S a c ang ag 00 a ngar 00

r n a a ar ng a gor or con ra n a n nc
gra ar o c ow r In a o a n r SD 00 N S
N I o S r ng r rg 00

c na n x n r ca o co oca on x rac on o In
roc ng o S n R arc or o nn r or c gan J n
00 oc a on or o a ona ng c 00

a c an wr n an a o a ca x rac r or o agg r
In So a o I a a SD 00 N S N I o
0 S r ng r rg 00

a c Go w G ﬀ c arc c r o o agg r In So a
o I a a SD 00 N S N I o 0
S r ng r rg 00

a c S a ow c S ro a o a c c on o rog n o
n ac c a r n an c ar o o no n In a o a
n r SD 00 N S N I o 0 S r ng r
rg 00

a c S a ow c S ro a x n ar or a
a on o an c ar nc on In an an ang ag
c no og a a a ng or o r Sc nc an ng c r an
g ag c no og on r nc o na o an c o r 00 0 0
awn c wo o na S o o 00

o row Sa on on ow own ang o o o o
ang ga ar awa

0 r r ow I I N or r nar r on In o
o r Sc nc S 00

r r ow c ♠ S a ow ar ng an a g a on n g n In an an ang ag c no og a a a ng or o r Sc nc an ng c r ang ag c no og on r nc o na o an c o r 00 0 awn c wo o na S o o 00

N S own a o go a 00 on w ag 00 `http p pw p`

Ra na ar ax n ro o or Na ra ang ag g R o on D n r o nn an a a a S S aroff S a a a a ca o R an x r on ar ng w a r o on In ana a a c nc o on or on n S con or o on wor x r on In gra ng roc ng arc ona S a n J 00 00

S a a R r ng co oca on ro x rac o a ona ng c

So a o I a a SD 00 N S N I o

S r ng r rg 00

S a I ac n arn ng roac o r a ca on D In or a on S R arc n r Sc oo o o ng Sc nc an ng n r ng n r o Sa or Sa or a 00

an an ang ag c no og a a a ng or o r Sc nc an ng c r ang ag c no og on r nc o na o an c o r 00 awn c wo o na S o o 00

o or a rac ca oo or or o og ca ana o o In o o r c o S ro anow In g n In or a on roc ng an n ng roc ng o In rna ona IIS II 0 on r nc n a o an J n 00 anc n So o ng 0 S r ng r rg 00

0 c G r c I o a ca cr a ng a a or a r o an c r a n In roc ng o or o on ng c D anc S n ra a J 00 oc a on or o a ona ng c 00

Towards the Automatic Acquisition of a Valence Dictionary for Polish

Adam Przepiórkowski[1,2]

[1] Institute of Computer Science, Polish Academy of Sciences,
ul. Ordona 21, 01-237 Warszawa, Poland,
adamp@ipipan.waw.pl,
http://nlp.ipipan.waw.pl/~adamp/

[2] Institute of Informatics, University of Warsaw,
ul. Banacha 2, 02-097 Warszawa, Poland

Abstract. This article presents the evaluation of a valence dictionary for Polish produced with the help of shallow parsing techniques and compares those results to earlier results involving deep parsing. We show that the valence dictionary obtained with the use of shallow parsing attains higher quality when it is measured on the basis of a corpus of valence frames, while the dictionary produced with the help of deep parsing seems superior when the results are compared to existing valence dictionaries.

Key words: valence acquisition, arguments of verbs, evaluation of valence dictionaries, partial parsing, IPI PAN Corpus of Polish

1 Aim and Scope

The valence of a given lexeme is, in general terms, its combinatorial potential, i.e., its ability to combine with other constituents of the utterance. In practice the term *valence* (or *valency*) usually refers to verbal lexemes, and it denotes the number and morphosyntactic makeup of the arguments of the verb. Hence, a valence dictionary will contain the information that the verb CHRAPAĆ, 'snore', combines only with the nominative subject, while the verb CHOWAĆ 'hide' also takes an accusative complement (*chować coś*, 'to hide something') and, optionally, a prepositional group with the preposition PRZED governing the instrumental case (*chować coś przed kimś*, 'to hide something from somebody'). Such dictionaries have various theoretical linguistic, psycholinguistic and educational uses, and they are a valuable resource in deep parsing and generation.

While there exist Polish dictionaries containing valence information,[3] including [15], [27], [2], and [14], the automatic acquisition of such information from corpora has many advantages when compared to the process of manual dictionary compilation. First, the automatic method is much faster and cheaper.

[3] A short comparison of a few of them in the context of Natural Language Processing may be found in [17], with some desiderata concerning such dictionaries put forward in [19].

M. Marciniak and A. Mykowiecka (Eds.): Bolc Festschrift, LNCS 5070, pp. 191–210, 2009.

Second, as it is based on naturally occurring texts, it is immune to prescriptive influences and to conflicting intuitions of a team of lexicographers. Hence, this method may be considered more objective. Third, the automatic procedure does not only extract valence frames, but also assigns them relative frequencies. For example, we may learn how often the verb DZIWIĆ 'make (one) wonder, surprise' combines with a nominal subject (*Dziwiło mnie takie postępowanie*, 'Such behaviour made me wonder'), and how often it co-occurs with a sentential subject (*Dziwiło mnie, że tak postąpił*, 'That he behaved like that made me wonder'). Such quantitative information is indispensable in probabilistic parsers (cf., e.g., [4], as well as [1]), which assign probabilities to particular parses, and it is also relevant in psycholinguistic research [13]. Fourth, as has been repeatedly noted (e.g., [28], [24], [12,11] and [9]), valence changes not only with time, but also with genre and topic. Once developed, automatic valence acquisition algorithms may be applied to various sets of texts in order to quickly and cheaply construct diachronic or thematic valence dictionaries. Fifth, automatic valence extraction may be used not only for the development of a new valence dictionary, but also for the verification and extension of an existing manually created dictionary (cf., e.g., [25]). [26] shows, for an automatically acquired valence dictionary of German, that the quality of such dictionaries may rival the quality of traditional dictionaries.

This paper focuses on the most basic type of valence information, which is found in all valence dictionaries, i.e., on morphosyntactic information regarding the grammatical class (part of speech) of the argument (e.g., CHOWAĆ, 'hide', combines with a nominal complement and not with a verbal complement), its grammatical case (e.g., CHOWAĆ combines with the accusative, not with the instrumental), etc. Some valence dictionaries, e.g., [15], also contain certain semantic information. Although the acquisition of such semantic valence information is beyond the scope of the current article, some work towards this end is currently being carried out within another ICS PAS project.[4]

2 Algorithm

As in virtually all previous work on valence acquisition (cf. [22], § 10.2, for an overview), the experiments described below proceed in two stages. First, at the linguistic stage, syntactic groups are identified which may be arguments of verbs. The result of this stage is a set of observations, where each observation consists of a verb and its observed potential arguments within a given sentence. Obviously, such observations will be noisy, with errors due to the inadequacies of morphosyntactic and syntactic processing. Second, the set of linguistic observations is subjected to statistical inference rules whose task is to decide which observations are reliable. Only thus filtered observations are considered valid valence frames.

[4] A Ministry of Science and Higher Education research grant (number NN516016533) "Automatyczne wykrywanie zależności semantycznych w strukturze argumentowej czasowników w dużych korpusach tekstów anotowanych syntaktycznie".

The main general steps of the algorithm, described in more detail below, are:

1. pre-process the empirical material, i.e., an appropriate subcorpus of the IPI PAN Corpus of Polish;
2. shallow process all sentences within that subcorpus and select fully parsed sentences for further statistical processing;
3. apply statistical filtering techniques, namely, the techniques proposed in [5].

2.1 Empirical Material

The main empirical material for the work reported here is the 2nd edition of the IPI PAN Corpus (`http://korpus.pl/`; [18]) containing about 255 million segments (over 200 million traditionally understood orthographic words, i.e., words delimited by white spaces and punctuation). Only sentences of less than 16 segments were selected from this corpus for further processing.[5] This restriction was imposed in order to reduce the time needed to process the massive amount of data, and similar restrictions are imposed in earlier work on the extraction of Polish valence, reported in [5]. Only candidates for true sentences, i.e., those containing verbal segments, were included in the processing chain.

As a result of this selection procedure, an IPI PAN subcorpus containing 25 647 017 segments (2 724 353 sentences) was created.

2.2 Shallow Parsing

The corpus obtained as described in the previous section was shallow parsed with the Spejd (`http://nlp.ipipan.waw.pl/Spejd/`; [23]) grammar presented in ch. 8 of [22], i.e., maximal nominal, prepositional, adjectival, verbal and other groups were automatically identified. Because of the partial nature of the grammar and the parser, not all sentences were fully parsed; after syntactic processing, some sentences contained sequences of lexical segments not assigned to any syntactic constituents. Only fully parsed sentences underwent the statistical processing. There were 1 137 014 (41.74%) such sentences and they contained 8 516 676 segments. One such sentence, *Kto się wstrzymał od głosu?* (‘Who abstained?’, literally: ‘Who self abstained from voice?’), is presented below:

```
<chunk type="s">
<group id="a106ac9" rule="(1) NG between verbs/groups/aby/etc."
       synh="a106abf" semh="a106abf" type="NG">
<tok id="a106abf">
<orth>Kto</orth>
<lex disamb="1">
 <base>kto</base><ctag>subst:sg:nom:m1</ctag>
</lex>
```

[5] 16 is the average length of a sentence in the 30-million “varied” (roughly balanced) subcorpus of the IPI PAN Corpus, `2.sample.30` (cf. `http://korpus.pl/index.php?page=download`).

```
</tok>
</group>
<group id="a106ac7" rule="sie" synh="a106ac0" semh="a106ac0"
       type="sie">
<tok id="a106ac0">
<orth>się</orth>
<lex disamb="1"><base>się</base><ctag>qub</ctag></lex>
</tok>
</group>
<syntok id="a106ac6" rule="czasownik niezanegowany 2a">
<orth>wstrzymał</orth>
<lex disamb="1">
 <base>wstrzymać</base><ctag>praet:sg:m1:perf:aff</ctag>
</lex>
<lex>
 <base>wstrzymać</base><ctag>praet:sg:m2:perf:aff</ctag>
</lex>
<lex>
 <base>wstrzymać</base><ctag>praet:sg:m3:perf:aff</ctag>
</lex>
<tok id="a106ac1">
<orth>wstrzymał</orth>
<lex disamb="1">
 <base>wstrzymać</base><ctag>praet:sg:m1:perf</ctag>
</lex>
<lex><base>wstrzymać</base><ctag>praet:sg:m2:perf</ctag></lex>
<lex><base>wstrzymać</base><ctag>praet:sg:m3:perf</ctag></lex>
</tok>
</syntok>
<group id="a106ac8"
       rule="(1) Dobre PrepNG na koncu zdania lub nawiasu"
       synh="a106ac2" semh="a106ac3" type="PrepNG">
<tok id="a106ac2">
<orth>od</orth>
<lex disamb="1"><base>od</base><ctag>prep:gen:nwok</ctag></lex>
</tok>
<tok id="a106ac3">
<orth>głosu</orth>
<lex disamb="1">
 <base>głos</base><ctag>subst:sg:gen:m3</ctag>
</lex>
</tok>
</group>
<ns/>
<tok id="a106ac5">
```

```
<orth>?</orth>
<lex disamb="1"><base>?</base><ctag>interp</ctag></lex>
</tok>
</chunk>
```

The syntactic representation exemplified above was subsequently translated to the format accepted by the statistical module, as proposed in [20] (and slightly modified in [5]). In the case of the above sentence, the result of this conversion is as follows:

```
% 'Kto się wstrzymał do głosu ?'
wstrzymać :np:nom: :prepnp:do:gen: :sie:
```

In the general case, the translation from the output format of Spejd to the input format of the statistical stage consists of the following steps:

1. each immediate constituent of a sentence, i.e., each XML child of the `<chunk type="s">` element, is assigned to one of the following three classes: grupa (i.e., a syntactic group), czasownik (i.e., a verb), inny token (neither, i.e., a token which is not a verb and does not belong to a recognised syntactic group); in particular, each token containing a verbal interpretation is assigned to the verbal class czasownik;
2. if, as a result, the number of elements in the verbal czasownik class is different than 1, the processing of the current sentence is aborted, and the algorithm moves to the next sentence;
3. since the only sentences that entered this stage of processing consisted of groups, verbs and punctuation marks, the class inny token must — after the previous steps — contain only punctuation marks and, as such, it is ignored in further processing;
4. the orthographic makeup of the sentence is retrieved for the purpose of a comment in the resulting file (starting with a `%`; cf. the example above);
5. the base form of the single verb in the sentence is retrieved; if this segment has a number of verbal interpretations with different base forms, the first of them is arbitrarily assumed to be the correct one;
6. each syntactic group belonging to the grupa class is translated into a list of morphosyntactic interpretations of the syntactic head of the group, e.g., `:np:nom:`, `:np:acc:`, `:prepnp:do:gen:`, `:infp:imperf:`; as a head may contain a number of morphosyntactic interpretations, the result is a list rather than a single such representation;
7. the Cartesian product of the lists (treated as sets) of representations of all elements of the grupa class is taken as the set of potential observations adduced by the currently processed sentence;
8. the potential observations are sorted and printed out.

Despite the fact that the shallow Spejd grammar used in the experiments reported in this paper contains some morphosyntactic disambiguation rules, not all segments are fully disambiguated, so one sentence may be the basis of a number of different potential observations, calculated in steps 6–7 of the above algorithm. For example, 4 potential observations were obtained for the sentence *Składam te podziękowania na ręce szefowej komisji pani senator Genowefy Grabowskiej* ('I thank the head of the commission, senator Genowefa Grabowska', literally: 'I-put these thanks onto hands boss.`gen` commission.`gen` Mrs. senator Genowefa Grabowska'):[6]

```
% 'Składam te podziękowania na ręce szefowej komisji
%  pani senator Genowefy Grabowskiej .'
składać :np:acc: :prepnp:na:acc:
składać :np:acc: :prepnp:na:loc:
składać :np:nom: :prepnp:na:acc:
składać :np:nom: :prepnp:na:loc:
```

The group *te podziękowania* 'these thanks' is not fully disambiguated and it retains both the accusative and the nominative case interpretations, and similarly the prepositional group, *na ręce...* 'onto hands', is not disambiguated as to the accusative or locative case value, which results in 4 potential observations. Note that, unlike in the case of the deep parser Świgra (`http://nlp.ipipan.waw.pl/~wolinski/swigra/`; [30,31]) utilised in [5], the observations may only differ in the values of morphosyntactic categories, not in the number or extent of syntactic groups; following the general shallow parsing principles, Spejd outputs a unique parse of the sentence.

A more crucial difference between the current experiments and the approach proposed by Dębowski consists in our refraining from any further linguistic processing: all linguistic knowledge is contained in the grammar, and the resulting observations correspond directly to the groups found by the parser. This should be contrasted with the algorithm described in [5], where the results of the grammar are subject to some further linguistic processing, including the following steps:

– a nominal group is added to an observation in case an elided subject (so-called *pro*-drop) is detected;
– the nominal genitive group, if any, is removed from an observation in case of a negated sentence, as this genitive group may actually be a Genitive of Negation (cf. [16]) realisation of an otherwise accusative argument of the verb;
– nominal phrases suspected of having the grammatical function of (temporal) adjuncts are removed from observations.

It is not clear to what extent such further transformations influenced the final results of [5], but they probably played a role in producing results more comparable to valence frames found in existing valence dictionaries. Such *a posteriori* modifications of observations must also lead to less accurate data concerning the actual text frequencies of particular realisations of valence frames.

[6] The orthographic form of the sentence, given here as a comment, was broken for typographical reasons.

2.3 Statistical Processing

The pre-processing step of the statistical stage is the selection of a single observation in case of sentences with multiple potential observations. A simple EM-type (Expectation Maximisation) algorithm described in [5] is used to this end. In the case of the example sentence given above, the observation actually selected for *Składam te podziękowania na ręce szefowej komisji pani senator Genowefy Grabowskiej* correctly assumes the accusative case of the nominal group, but incorrectly identifies the case within the prepositional group as locative:

```
% 'Składam te podziękowania na ręce szefowej komisji
%  pani senator Genowefy Grabowskiej .'
składać :np:acc: :prepnp:na:loc:
```

Observations collected and further selected this way are the first version of the resulting valence dictionary, the so-called proto-dictionary [7]. For example, the lexical entry for the verb WYPŁYWAĆ, 'flow out, emerge, follow' in the proto-dictionary created within the current shallow parsing experiments is given below:[7]

```
'wypływać' => {
  'np(nom),z+np(gen)' => 9,
  'adv,np(nom),z+np(gen)' => 2,
  '' => 1,
  'adj(nom),adv,dla+np(gen),np(acc)' => 1,
  'adj(nom),np(acc),z+np(gen)' => 1,
  'adj(nom),z+np(gen)' => 1,
  'adv' => 1,
  'dla+np(gen),np(nom),z+np(gen)' => 1,
  'do+np(gen),np(nom),z+np(gen)' => 1,
  'np(acc),o+np(loc),z+np(gen)' => 1,
  'np(acc),od+np(gen),z+np(gen)' => 1,
  'np(dat),np(nom)' => 1,
  'np(nom),np(voc),z+np(gen)' => 1,
  'np(nom),o+np(loc)' => 1
}
```

According to this entry, forms of the verb WYPŁYWAĆ were observed 9 times with a nominative nominal group and a prepositional group headed by the genitive-taking preposition Z, twice with an additional adverbial group, once without any accompanying groups, etc.

The two main steps of the statistical stage make use of an approximate representation of valence proposed in [6], where a valence frame is described as a *set* of possible arguments of the verb (the set of all arguments in all possible frames of that verb) and an additional table specifying whether any two possible

[7] The format of such lexical entries is actually the representation of hash tables in the Perl programming language.

arguments always co-occur, never co-occur, unidirectionally imply one another, or seem independent of each other.

First, all possible arguments of a given verb are collected. A subset of that argument set is identified as those arguments which occur in all possible frames of that verb. An argument type a is considered a possible argument of a verb v in case the inequality in (1) holds; $c(v)$ denotes here the number of occurrences of the verb v, $c(v,a)$ — the number of observed co-occurrences of v with the argument a. The argument is additionally considered a necessary argument of v in case (2) holds.

(1) $c(v,a) \geq p_a c(v) + 1$

(2) $c(v) - c(v,a) < p_{\neg a} c(v) + 1$

The parameters p_a and $p_{\neg a}$ occurring in the inequalities above are trained — separately for each argument type — on the basis of the manually created dictionary [27], as well as on the basis of lexical entries of around 200 verbs in [15] and [2]. The exact parameter estimation procedure for p_a and $p_{\neg a}$ is described in [5].

This first step ends with projecting information of estimated possible and necessary arguments into actually observed frames: in each observed frame only those arguments are retained which are "possible" in the sense above and, moreover, in case the observation does not contain the "necessary" argument, it is artificially added to the frame.[8] As a result of this step, the lexical entry of WYPŁYWAĆ is reduced as follows:

```
'wypływać' => {
  'np(nom),z+np(gen)' => 15,
  'np(nom)' => 4,
  'np(acc),np(nom),z+np(gen)' => 3,
  'np(acc),np(nom)' => 1
}
```

It follows from the comparison of this lexical entry with the corresponding lexical entry in the proto-dictionary that `adv`, `np(voc)`, `np(dat)` and various prepositional argument types were rejected as possible arguments of the verb, so the set of possible arguments is reduced to {`np(nom)`, `np(acc)`, `z+np(gen)`}. Further, the nominative nominal group was classified as a necessary argument. Hence, the four "observations" `'np(nom)'` in the lexical entry above actually correspond to the original observations: `'np(dat),np(nom)'`, `'np(nom),o+np(loc)'`, `'adv'` and the empty observation `''`.

In the second step, full frames obtained in the first step are evaluated. Again, on the basis of existing valence dictionaries, possible relationships between arguments are estimated: do the two given arguments usually co-occur within frames of various verbs, do they have a complementary distribution, does one imply the other, or are they independent. For any two argument types, such a

[8] This description is based on the observation of the algorithm at work and it differs a little from the description in [5].

relationship is calculated on the basis of the whole dictionary, independent of particular verbs. For a given verb, this default relation between two arguments is assumed, unless there are strong reasons to override it. For example, the 15 "observations" of the nominative nominal group `np(nom)` co-occurring with the `z+np(gen)` prepositional group in the vicinity of WYPŁYWAĆ were not sufficient to retain that frame of that verb (note that the frame is correct here, but it is generally rather rare in the corpus), so the final lexical entry for WYPŁYWAĆ looks as follows:

```
'wypływać' => {
  'np(nom)' => 4,
  'np(acc),np(nom),z+np(gen)' => 3,
  'np(acc),np(nom)' => 1
}
```

3 Results

Three dictionaries were created as a result of the algorithm described in the previous section: the proto-dictionary, which contains the actual observations (perhaps selected from alternative potential observations with the help of a simple EM algorithm), the dictionary resulting from the first step of statistical processing (the intermediate dictionary), and the final dictionary created in the second step of statistical processing. Table 1 gives the sizes of these dictionaries.

Table 1. Sizes of automatically obtained valence dictionaries

dictionary	**entries**	**frames**	
		tokens	types
proto	6,845	1,084,286	20,894
intermediate	6,845	1,084,286	517
final	4,166	863,731	141

The proto-dictionary obtained as described in the previous sections contains 6 845 entries. As a result of the second step of statistical processing, this number is reduced to 4 166 entries with 207.33 observations per entry on average. The full dictionary would consume around 430 pages, so — given the space limits — it must suffice to present some of its characteristics here.[9]

The final number of different "observed" valence frames is 141. This number is substantially reduced with respect to the number of 20 894 different realisations of frames actually observed, and also with respect to the 517 types of "observations" remaining after the first step of statistical processing. The most

[9] Appendix A contains a fragment of the dictionary resulting from the simplification of the statistical stage, as described in § 5.

frequent frame was the intransitive frame (only a nominative nominal group; 232 034 occurrences) and the empty frame (129 720), while the actually very frequent transitive frame (nominative and accusative nominal groups) is the 4th most frequent frame in the resulting final dictionary (84 611 occurrences). Out of the three frames with single occurrences: `'do+np(gen),np(gen),np(nom),sie'` (for the verb UŻYWAĆ 'use'), `'np(dat),np(nom),o+np(loc),sie'` (MARZYĆ 'dream') i `'np(nom),o+np(loc),z+np(inst)'` (POROZMAWIAĆ 'talk'), the first two are erroneous, and the last one seems to be correct.

4 Evaluation

Two evaluation procedures were applied to the dictionaries obtained as described above: the dictionary-based evaluation (at the level of frame types) and the corpus evaluation (at the level of frame occurrences, or tokens).

4.1 Dictionary-Based Evaluation

Dictionary-based evaluation consists in finding the ratio of automatically extracted valence frames also present in manually constructed dictionaries (precision), and the ratio of frames in such previously built dictionaries also present in the automatic results (recall). In the current experiments, these values were estimated on the basis of a sample of 202 verbs randomly selected from [27]. For all these 202 verbs, their entries were also extracted from two other manually constructed dictionaries [15,2] and converted to the "least common denominator" format.[10] Since the dictionaries differed a little in the scope and character of the valence information, the conversion process was to some extent interpretative.

Precision (P), recall (R) and their harmonic mean, called F-measure (F), were computed for the final dictionary obtained in the current experiments, as well as for both dictionaries reflecting earlier stages of processing: for the proto-dictionary and for the intermediate dictionary. In each case automatically obtained valence frames were compared to various gold standards, that is, to each of the three manually constructed valence dictionaries, marked below as "Bań." [2], "Pol." [15] and "Świ." [27], and to two dictionaries compiled from these three manually constructed dictionaries by taking their set-theoretical sum ("SUM") and by majority voting ("MV"; i.e., a frame is present in the MV dictionary, if it is present in at least two manually constructed dictionaries). In each comparison, only the frames of those verbs were considered which were present both in the automatically obtained dictionary and in the gold standard. The results of these comparisons are contained in Tables 2–4.

The comparison of Tables 2 and 3 shows the great importance of the first step of statistical processing. Although it resulted in a significant drop of recall (from 40.40% to 29.80%, for the MV dictionary), that decrease in recall was

[10] These data were prepared by Witold Kieraś and Łukasz Dębowski, whose scripts were used for calculating precision and recall given below.

Table 2. Dictionary-based evaluation of the proto-dictionary

	Bań.	Pol.	Świ.	SUM	MV
P	3.97	3.04	3.05	5.11	**3.15**
R	37.33	31.03	34.83	28.60	**40.40**
F	7.17	5.54	5.62	8.68	**5.85**

Table 3. Dictionary-based evaluation of the intermediate dictionary

	Bań.	Pol.	Świ.	SUM	MV
P	50.69	41.70	39.95	57.80	**44.94**
R	24.68	22.00	23.56	16.72	**29.80**
F	33.20	28.81	29.64	25.94	**35.84**

Table 4. Dictionary-based evaluation of the final dictionary

	Bań.	Pol.	Świ.	SUM	MV
P	63.81	52.23	51.34	70.41	**57.58**
R	22.04	19.51	21.49	14.42	**27.07**
F	32.77	28.41	30.30	23.94	**36.83**

more than compensated by the dramatic increase in precision (from 3.15% to 44.94%), which resulted in the clear increase in the harmonic mean of these measures (from 5.85 to 35.84). The second step of the statistical stage wasn't so significant: although the F-measure for the MV dictionary is higher in Table 4 than in Table 3, the difference is relatively minor (36.83 to 35.84) because the increase in precision was to a large extent annulled by the decrease in recall. It is interesting to note that, for two gold standards, Bań. and Pol., the second step of statistical processing was slightly detrimental, if the quality were evaluated with the F-measure.

Let us also note at the end of this section that, although the results are far from perfect, they constitute a clear improvement over the reasonable baseline consisting in the assignment of two frames to each verb: the intransitive frame `'np(nom)'` and the transitive frame `'np(acc),np(nom)'`.[11] A "dictionary" constructed this way would have relatively high precision (47.41% when measured against the MV dictionary), but very low recall (15.15%). Complete results of the evaluation of such a baseline dictionary are given in Table 5.

4.2 Corpus-Based Evaluation

Also corpus-based evaluation shows that, after shallow processing at the linguistic stage, it may be beneficial to stop statistical processing after the first step,

[11] Experiments were also performed for other baselines, including: only the intransitive frame, only the transitive frame, the empty frame, the infinitival frame, and various combinations of these frames. In each case the dictionary-based evaluation gave worse (in terms of F-measure) results than the results for the baseline given below.

Table 5. Dictionary-based evaluation of the baseline, i.e., a dictionary created artificially by assuming two frames for each verb: the transitive frame and the intransitive frame

	Bań.	Pol.	Świ.	SUM	MV
P	54.66	42.49	43.52	59.33	**47.41**
R	12.83	10.80	12.37	8.27	**15.15**
F	20.78	17.23	19.27	14.52	**22.96**

before frames are removed on the basis of the low co-occurrence of the arguments in manually constructed dictionaries.

The evaluation was performed for 12 verbs selected on the basis of their frequencies in the corpus resulting from the linguistic processing. These are 4 very frequent verbs (tens of thousands of occurrences): WSTRZYMAĆ 'stop', CHCIEĆ 'want', STWIERDZAĆ 'conclude', MIEĆ 'have', 4 verbs of medium frequency (around 4 000 occurrences): MUSIEĆ 'must', ZABRAĆ 'take (away)', PRZYPOMINAĆ 'remind, remember', and ZGŁOSIĆ 'report', and 4 relatively rare verbs (around 400 occurrences): STAWIAĆ 'put', USŁYSZEĆ 'hear', ZAUWAŻYĆ 'notice' and USTALIĆ 'establish'.

For each verb, 120 sentences containing that verb were randomly selected. These sentences were linguistically annotated[12] on the basis of brief guidelines of syntactic annotation [21] with the help of the Anotatornia annotation tool developed at ICS PAS [8].

Some of these sentences were lacking full morphosyntactic analysis, some were the result of erroneous segmentation of text into sentences. The remaining 985 sentences were annotated for maximal syntactic groups. Hence, the result of the annotation was a set of fully correct valence frame observations. Obviously, just as would be the case in fully correct shallow processing, these observations contained information of all observed dependents of the main verb: arguments and adjuncts alike. Also, these observations were not further processed linguistically in any way; in particular, no information about missing (elided) arguments was added.

The corpus prepared this way was the basis for calculating token recall, i.e., the ratio of manually annotated frames also found by the algorithm. The result was 89% for the proto-dictionary, 32% for the intermediate dictionary and 22% for the final dictionary. Table 6 presents the results in more detail, while Table 7 presents analogous results for the approach based on deep linguistic processing, described in [5]. It is interesting to note that, although the proto-dictionary based on shallow processing with Spejd contains many more valence frames observed in manually annotated texts than the proto-dictionary based on the deep parser Świgra (89% compared to 39%), this difference reduces significantly for the intermediate dictionary (32% to 27%) and reverses for the final dictionary (22% to 27%). This effect is probably to some extent caused by the greater

[12] By linguists: Monika Czerepowicka, Hanna Maliszewska, Marta Nazarczuk-Błońska, Marta Piasecka and Izabela Will.

Table 6. The number of observations of valence frames for the 12 verbs for which the appropriate frame is also present in the valence dictionary automatically obtained with the use of the Spejd parser and the grammar presented in ch. 8 of [22]

verb	frames (tokens)			
		in dictionary		
	in texts	proto	intermediate	final
USTALIĆ	73	54	6	11
ZABRAĆ	103	100	1	1
STAWIAĆ	78	34	8	6
CHCIEĆ	91	89	19	19
ZAUWAŻYĆ	65	48	6	11
WSTRZYMAĆ	88	88	88	88
MIEĆ	86	84	28	28
MUSIEĆ	84	83	34	34
PRZYPOMINAĆ	108	93	0	0
STWIERDZAĆ	119	119	114	0
ZGŁOSIĆ	73	70	15	15
USŁYSZEĆ	17	13	0	0
total	985	875	319	213
percent	100	88.83	32.39	21.62

Table 7. The number of observations of valence frames for the 12 verbs for which the appropriate frame is also present in the valence dictionary automatically obtained with the use of the Świgra deep parser

verb	frames (tokens)			
		in dictionary		
	in texts	proto	intermediate	final
USTALIĆ	73	23	8	10
ZABRAĆ	103	93	53	53
STAWIAĆ	78	22	8	6
CHCIEĆ	91	23	19	19
ZAUWAŻYĆ	65	18	7	11
WSTRZYMAĆ	88	88	88	88
MIEĆ	86	43	28	28
MUSIEĆ	84	37	34	34
PRZYPOMINAĆ	108	8	7	4
STWIERDZAĆ	119	0	0	0
ZGŁOSIĆ	73	23	18	10
USŁYSZEĆ	17	2	0	0
total	985	380	270	263
percent	100	38.58	27.41	26.70

dispersion of data in the current approach (many more different valence frame types are found in shallow processing), but it may also be a result of the close fit of the statistical approach proposed by Dębowski and the deep parsing with Świgra, on the basis of which that approach was developed and fine-tuned [6,5].

As shown in the next section, the rather disappointing results given above improve when the second step of statistical processing is simplified. Nevertheless, already these modest results are well above the baseline described above: in the case where every verb is assigned the transitive frame and the intransitive frame, the resulting valence dictionary would cover only 101 (10.25%) corpus observations.

5 Simplification of Statistical Processing

The previous three sections describe some experiments in valence extraction, where linguistic processing is performed with the Spejd implementation of the shallow grammar presented in ch. 8 of [22], while the statistical processing follows the ideas described in [5]. In the preceding section we noted that the second step of the statistical stage, where frames with uncommon combinations of arguments are rejected, is a mixed blessing at best: it improves the results of the dictionary-based evaluation only slightly (and in fact has a detrimental effect, if Polański's or Bańko's dictionaries are taken as gold standards), and it causes a clear drop in the quality measured via corpus-based evaluation.

On the other hand, as noted in various earlier works on valence acquisition for other languages (e.g., [3,10,11,26,5]), simpler methods of rejecting rare observations often give results comparable to more complicated statistical techniques. Hence, it would be interesting to find out whether applying such simpler methods in the current linguistic setup also gives results comparable to or better than the techniques proposed in [5].

Table 8. Dictionary-based evaluation (for the MV dictionary) of valance information acquired by rejecting observations rare in the intermediate dictionary; for comparison, the table also recalls previous results for the intermediate and final dictionary

$c_{\min}(v)$	10	10	13	**dictionary**	
$p_{\min}(r,v)$	0	2	2	**intermediate**	**final**
P	45.49	53.08	53.01	44.94	57.58
R	32.45	30.94	31.45	29.80	27.07
F	37.88	39.09	39.48	35.84	36.83

To this end, further experiments based on shallow linguistic processing were conducted, where the second step of statistical processing was replaced with a simpler rejection of rare observations. Two parameters, or cutoff points, were used: the sheer number of occurrences of the verb, $c_{\min}(v)$, and the ratio of the number of co-occurrences of a given frame with a given verb to the numer

Table 9. Dictionary-based evaluation of valance information acquired with shallow linguistic processing and with cutoff points $c_{\min}(v) = 13$ and $p_{\min}(r, v) = 2$

	Bań.	Pol.	Świ.	SUM	MV
P	58.53	49.00	46.66	66.05	**53.01**
R	25.13	23.05	25.09	16.98	**31.45**
F	35.16	31.35	32.63	27.02	**39.48**

of all occurrences of that verb, $p_{\min}(r, v)$ (expressed as percent points). The requirement that valence frames be acquired only for verbs occurring at least $c_{\min}(v) = 10$ times in the parsed corpus improved F-measure (as computed for the MV dictionary) to 37.88, and further rejection of observations less frequent than $p_{\min}(r, v) = 2$ (i.e., 2%) increased the value to 39.09. In various experiments performed, the best F-measure, 39.48, was achieved for $c_{\min}(v) = 13$ and $p_{\min}(r, v) = 2$. The results are summarised and compared to earlier results in Table 8, while more complete results for the best cutoff points are given in Table 9.

Let us note that significant improvements as measured by dictionary-based evaluation were achieved with practically no decrease in the quality measured with corpus-based evaluation (cf. Table 10). The number of corpus observations corresponding to automatically identified frames is 318, i.e., almost the same as in the intermediate dictionary (319; cf. Table 6 on p. 203), and much higher than in the final dictionary (213).

Table 10. The number of observations of valence frames for the 12 verbs for which the appropriate frame is also present in the valence dictionary automatically obtained with the use of the Spejd parser and the grammar presented in ch. 8 of [22] (simplified statistical processing)

	frames (tokens)	
verb	**in texts**	**in dictionary**
USTALIĆ	73	6
ZABRAĆ	103	1
STAWIAĆ	78	7
CHCIEĆ	91	19
ZAUWAŻYĆ	65	6
WSTRZYMAĆ	88	88
MIEĆ	86	28
MUSIEĆ	84	34
PRZYPOMINAĆ	108	0
STWIERDZAĆ	119	114
ZGŁOSIĆ	73	15
USŁYSZEĆ	17	0
total	985	318
percent	100	32.28

Table 11. Dictionary-based evaluation of valance information acquired with deep linguistic processing [5]

	Bań.	Pol.	Świ.	SUM	MV
P	63.53	54.56	55.17	74.01	**59.88**
R	25.39	23.63	26.71	17.58	**32.59**
F	36.28	32.98	35.99	28.41	**42.21**

Table 12. Dictionary-based evaluation of valance information acquired with deep linguistic processing and with cutoff points $c_{\min}(v) = 17$ and $p_{\min}(r, v) = 2$

	Bań.	Pol.	Świ.	SUM	MV
P	59.26	49.66	49.66	69.82	**54.18**
R	27.98	25.86	29.87	20.00	**35.55**
F	38.01	34.01	37.30	31.09	**42.93**

It should also be noted that replacing the second statistical step with cutoff points in the original methodology — based on deep parsing with the Świgra parser — described in [5] also brings about certain, but less significant, improvements in the values of the F-measure. In this case the best cutoff points were $c_{\min}(v) = 17$ and, as above, $p_{\min}(r, v) = 2$. The results of dictionary-based evaluation for these cutoff values are given in Table 12, while the original results presented in [5] are cited in Table 11. When compared to the results in Table 7, the result of corpus-based evaluation is practically the same: there were 264 corpus observations corresponding to automatically acuired valence frames.

Taking into consideration both evaluation methodologies, the results based on shallow linguistic processing are comparable to Dębowski's ([5]) results based on deep processing; such a comparison is presented in Table 13. While the current experiments produce inferior results, when measured as the similarity to the majority voting dictionary, they are clearly superior when measured with reference to actually occurring frame realisations of the 12 verbs of varying frequencies.

6 Summary

The aim of this article was to present a practical application of the formalism and the grammar described in [22] to the task of automatic valence acquisition from morphosyntactically annotated corpora. The quality of the results of valence acquisition with shallow parsing and simplified statistical processing turns out to be comparable to the best results for Polish found in the literature, and much higher when measured against frames actually observed in texts. Also, the simplification of the statistical stage alone makes it possible to slightly improve

Table 13. A comparison of final results of three approaches: [5], the approach presented there with the second step of statistical processing replaced by simple cutoff points, and the approach presented in [22] and summarised in this article, also with simple cutoff points instead of the second step of statistiacal processing; P, R and F are precision, recall and their harmonic mean, as measured in dictionary-based evaluation, and C is the corpus-based token recall; the best results are in boldface

	Dębowski ([5])	Dębowski ([5]) $c_{\min}(v) = 17$ $p_{\min}(r, v) = 2$	Przepiórkowski ([22]) $c_{\min}(v) = 13$ $p_{\min}(r, v) = 2$
P	**59.88**	54.18	53.01
R	32.59	**35.55**	31.45
F	42.21	**42.93**	39.48
C	26.70	26.80	**32.28**

the results of the dictionary-based evaluation, when compared to the earlier best results described in [5].

There are many possible ways the approach presented above may be developed further and improved. The most obvious concern linguistic processing: both the morphological analyser and the shallow grammar could be extended in various ways. Also the empirical basis could be improved, not only by increasing the corpus size, but also by making better use of the current corpus: at the moment evidence provided by subordinate clauses and less than fully parsed sentences is lost in the process. The evaluation of the results obtained using different linguistic and statistical methods also suggests that the novel approach to the statistical stage proposed in [5], promising in combination with deep processing at the linguistic stage, may be less adequate when coupled with shallow linguistic processing. We hope to continue work both on the empirical basis and on linguistic and statistical methodologies of valence acquisition within subsequent projects carried out at ICS PAS.

Acknowledgements. This article contains some of the material of chapter 10 of the monograph [22], summarising the main results of a valence acquisition project carried out at the Institute of Computer Science of the Polish Academy of Sciences (ICS PAS) in 2005–2008.[13] The acknowledgements therein carry over to this paper, with additional thanks for comments to Małgorzata Marciniak.

A An Extract from the Valence Dictionary

This appendix contains an extract from the valence dictionary automatically acquired with the use of the shallow Spejd grammar presented in ch. 8 of [22], combined with the simplified statistical processing described in § 5.

[13] A Ministry of Science and Higher Education research grant (number 3T11C00328) "Automatyczna ekstrakcja wiedzy lingwistycznej z dużego korpusu języka polskiego", `http://nlp.ipipan.waw.pl/PPJP/`.

```
'gadać' => {
 'np(nom)' => 58,
 'np(acc),np(nom)' => 8,
 'np(nom),PZ' => 5
}
'gasić' => {
 'np(nom)' => 12,
 'np(acc),np(nom)' => 5
}
'gasnąć' => {
 'np(nom)' => 10,
 'nad+np(inst),np(nom)' => 3
}
'generować' => {
 'np(nom)' => 12,
 'np(acc),np(nom)' => 6
}
'ginąć' => {
 'np(nom)' => 49
}
'gniewać' => {
 'np(nom),sie' => 22,
 'np(nom),sie,ZE' => 3
}
'godzić' => {
 'na+np(acc),np(nom),sie' => 42,
 'inf,np(nom),sie' => 28,
 'np(nom),w+np(acc)' => 26,
 'np(nom),sie' => 17,
 'np(acc),np(nom),w+np(acc)' => 4,
 'np(acc),np(nom)' => 4,
 'np(acc),np(nom),sie' => 4,
 'np(nom)' => 3
}
'gonić' => {
 'np(acc),np(nom)' => 17,
 'np(nom)' => 15
}
'gospodarować' => {
 'np(inst),np(nom)' => 7,
 'np(nom)' => 7,
 'np(acc),np(nom)' => 2,
 'np(acc),np(inst),np(nom)' => 1
}
'gotować' => {
 'np(acc),np(nom)' => 4,
 'np(nom)' => 3,
 'do+np(gen),np(nom),sie' => 3,
 'np(nom),sie' => 3,
 'do+np(gen),np(acc),np(nom),sie'
      => 1,
 'np(dat),np(nom),sie' => 1,
 'np(dat),np(nom)' => 1
}
'gościć' => {
 'np(nom)' => 24,
 'np(acc),np(nom)' => 14
}
'gratulować' => {
 'np(nom)' => 97,
 'np(dat),np(nom)' => 39
}
'grać' => {
 'np(nom)' => 267,
 'np(acc),np(nom)' => 21
}
'gromadzić' => {
 'np(nom)' => 10,
 'np(nom),sie' => 10,
 'np(acc),np(nom)' => 8,
 'np(acc),np(nom),sie' => 4,
 'na+np(acc),np(acc),np(nom)' => 3,
 'na+np(acc),np(nom)' => 1,
 'na+np(acc),np(nom),sie' => 1
}
'grozić' => {
 'np(dat),np(nom)' => 83,
 'np(inst),np(nom)' => 54,
 'np(nom)' => 38,
 'do+np(gen),np(dat),np(nom)' =>
      11,
 'do+np(gen),np(dat),np(nom),
   za+np(acc)' => 10,
 'do+np(gen),np(acc),np(dat),
   np(nom)' => 7,
 'np(dat),np(inst),np(nom)' => 6
}
'gubić' => {
 'np(nom),sie' => 12,
 'np(nom)' => 6,
 'np(acc),np(nom)' => 4,
 'np(acc),np(nom),sie' => 2
}
'gwarantować' => {
 'np(acc),np(nom)' => 55,
 'np(nom)' => 41,
 'np(nom),ZE' => 26,
 'np(acc),np(dat),np(nom)' => 8,
 'np(dat),np(nom)' => 6,
 'np(dat),np(nom),ZE' => 4
}
```

References

1. Arun, A., Keller, F.: Lexicalization in crosslinguistic probabilistic parsing: The case of French. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, Ann Arbor, MI, pp. 306–313 (2005)
2. Bańko, M. (ed.): Inny słownik języka polskiego. Wydawnictwo Naukowe PWN, Warsaw (2000)
3. Briscoe, T., Carroll, J.: Automatic extraction of subcategorization from corpora. In: Proceedings of the 5th Applied Natural Language Processing Conference, Washington, DC, pp. 356–363. ACL (1997)
4. Carroll, J., Minnen, G., Briscoe, T.: Can subcategorisation probabilities help a statistical parser? In: Proceedings of the Sixth Workshop on Very Large Corpora, Montreal, pp. 118–126 (1998)
5. Dębowski, Ł.: Valence extraction using the EM selection and co-occurrence matrices, 5 Dec. 2007. arXiv:0711.4475v2 [cs.CL] (2007)
6. Dębowski, Ł., Woliński, M.: Argument co-occurence matrix as a description of verb valence. In: Vetulani, Z. (ed.) Proceedings of the 3rd Language & Technology Conference, Poznań, Poland, pp. 260–264 (2007)
7. Dębowski, Ł., Woliński, M.: Nowe metody ekstrakcji walencji czasowników z tekstów w języku polskim. Referat wygłoszony na seminarium Zespołu Inżynierii Lingwistycznej IPI PAN, Warszawa, 22 października 2007 (2007)
8. Hajnicz, E., Murzynowski, G., Woliński, M.: ANOTATORNIA – lingwistyczna baza danych. In: Materiały V konferencji naukowej InfoBazy 2008, Systemy * Aplikacje * Usługi, Gdańsk, pp. 168–173. Centrum Informatyczne TASK, Politechnika Gdańska (2008)
9. Hare, M., McRae, K., Elman, J.: Sense and structure: Meaning as a determinant of verb subcategorization preferences. Journal of Memory and Language 48(2), 281–303 (2003)
10. Kawahara, D., Kaji, N., Kurohashi, S.: Japanese case structure analysis by unsupervised construction of a case frame dictionary. In: Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000), Saarbrücken, pp. 432–438 (2000)
11. Korhonen, A.: Subcategorization Acquisition. Ph. D. dissertation, University of Cambridge (2002)
12. Korhonen, A.L.: Using semantically motivated estimates to help subcategorization acquisition. In: Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, ACL (2000)
13. Lapata, M., Keller, F., Schulte im Walde, S.: Verb frame frequency as a predictor of verb bias. Journal of Psycholinguistic Research 30(4), 419–435 (2001)
14. Mędak, S.: Praktyczny słownik łączliwości składniowej czasowników polskich. Universitas, Cracow (2005)
15. Polański, K. (ed.): Słownik syntaktyczno-generatywny czasowników polskich. Zakład Narodowy im. Ossolińskich / Instytut Języka Polskiego PAN, Wrocław / Cracow (1992)
16. Przepiórkowski, A.: Long distance genitive of negation in Polish. Journal of Slavic Linguistics 8, 151–189 (2000)
17. Przepiórkowski, A.: On the computational usability of valence dictionaries for Polish. IPI PAN Research Report 971, Institute of Computer Science, Polish Academy of Sciences, Warsaw (2003)

18. Przepiórkowski, A.: The IPI PAN Corpus: Preliminary version. Institute of Computer Science. Polish Academy of Sciences, Warsaw (2004)
19. Przepiórkowski, A.: Towards the design of a syntactico-semantic lexicon for Polish. In: Kłopotek, M.A., Wierzchoń, S.T., Trojanowski, K. (eds.) Intelligent Information Processing and Web Mining. Advances in Soft Computing, pp. 237–246. Springer, Berlin (2004)
20. Przepiórkowski, A.: What to acquire from corpora in automatic valence acquisition. In: Koseska-Toszewa, V., Roszko, R. (eds.) Semantyka a konfrontacja językowa, vol. 3, pp. 25–41. Slawistyczny Ośrodek Wydawniczy PAN, Warsaw (2006)
21. Przepiórkowski, A.: Krótka instrukcja anotacji składniowej. Unpublished manuscript, Institute of Computer Science, Polish Academy of Sciences (2008)
22. Przepiórkowski, A.: Powierzchniowe przetwarzanie języka polskiego. Akademicka Oficyna Wydawnicza EXIT, Warsaw (2008)
23. Przepiórkowski, A., Buczyński, A.: ♠: Shallow Parsing and Disambiguation Engine. In: Vetulani, Z. (ed.) Proceedings of the 3rd Language & Technology Conference, Poznań, Poland, pp. 340–344 (2007)
24. Roland, D., Jurafsky, D.: How verb subcategorization frequencies are affected by corpus choice. In: Proceedings of COLING 1998, Montreal, pp. 1122–1128 (1998)
25. Schiehlen, M., Spranger, K.: Authomatic methods to supplement broad-coverage subcategorization lexicons. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, Lisbon, pp. 29–32. ELRA (2004)
26. Schulte im Walde, S.: Evaluating verb subcategorisation frames learned by a German statistical grammar against manual definitions in the *Duden* dictionary. In: Proceedings of the 10th EURALEX International Congress (2002)
27. Świdziński, M.: Syntactic dictionary of Polish verbs. Version 3a. Unpublished manuscript, University of Warsaw (1998)
28. Ushioda, A., Evans, D.A., Gibson, T., Waibel, A.: The automatic acquisition of frequencies of verb subcategorization frames from tagged corpora. In: Boguraev, B.K., Pustejovsky, J. (eds.) SIGLEX ACL Workshop of Acquisition of Lexical Knowledge from Text, Columbus, OH, pp. 95–106 (1993)
29. Vetulani, Z. (ed.): Proceedings of the 3rd Language & Technology Conference, Poznań, Poland (2007)
30. Woliński, M.: Komputerowa weryfikacja gramatyki Świdzińskiego. Ph. D. dissertation, Institute of Computer Science, Polish Academy of Sciences, Warsaw (2004)
31. Woliński, M.: An efficient implementation of a large grammar of Polish. Archives of Control Sciences 15(3), 251–258 (2005)

Semantic Annotation of Verb Arguments in Shallow Parsed Polish Sentences by Means of the EM Selection Algorithm

Elżbieta Hajnicz

Institute of Computer Science, Polish Academy of Sciences
Elzbieta.Hajnicz@ipipan.waw.pl

Abstract. The ultimate goal of our work is to extend a syntactic valence dictionary of Polish verbs by adding some semantic information to verb arguments. This information consists of wordnet *semantic categories* of words. In order to provide *syntactic slots* of dictionary entries with lists of appropriate semantic categories of corresponding nouns, we need a treebank with all nouns semantically annotated with such categories, as both syntactic (i.e., argument structure) and semantic information is required.
We aim here at *Word Sense Disambiguation* (WSD). To solve this task for our specific application, we adapt EM selection algorithm elaborated for extraction of syntactic valence frames.
In the paper, the whole process of data processing is shown. The main focus is put on WSD task. Three versions of the EM selection algorithm are presented: the original one and its two modifications. Finally, the evaluation and comparison of the algorithms is performed.

Key words: corpus lingustics, wordnets, valence dictionary, word sense disambiguation, Polish

1 Introduction

A number of resources and tools necessary for Natural Language Processing (NLP) are already available for the Polish language: morphological analysers (or dictionaries) [19,45,35,4,46], deep parsers [29,34,44], and shallow parsers [32]. The aforementioned parsers often use elaborate syntactic valence dictionaries [28,33,31,40].

Recognising the syntactic structure of texts is insufficient to obtain satisfactory results in solving NLP tasks such as machine translation, information extraction, and question answering. Semantic information is indispensable. In practical applications focused on specific domains (e.g., medicine, finance, sport) such information is often gathered in ontologies. More general lexical semantic resources, such as wordnets [13,42] and FrameNet [2,14] are also being created.

The main goal of our work is to enrich the valence dictionary of Polish verbs by adding semantic information, represented by means of wordnet semantic categories of nouns. The plain syntactic valence dictionary is a collection of

M. Marciniak and A. Mykowiecka (Eds.): Bolc Festschrift, LNCS 5070, pp. 211–240, 2009.

predicates (here: verbs) provided with a set of verb frames. Verb frames consist of syntactic slots that represent phrases occurring in the corresponding position in a sentence. Thus, our goal is to provide syntactic slots (here: NPs/PPs) with a list of appropriate semantic categories for the corresponding nouns.

In order to automatically acquire semantic information for a syntactic valence dictionary, we need a large treebank where all nouns are semantically annotated. In this understanding, our problem intersects with the Word Sense Disambiguation (WSD) task. An extensive overview of WSD is presented in [1]. Methods used for WSD can be divided into supervised, which rely on a manually annotated subcorpus to train the algorithm [12,37,38], unsupervised, i.e., based on clustering words that occur in similar context [24,26,36], semi-supervised [47] and knowledge-based, applying electronic lexicons and lexical knowledge bases, such as wordnets [3,27]. Most of those techniques are focused on fairly fine-grained word senses, hence they are applied to a small set of words or they need a large corpus to operate on. Specifically for Slavic languages, the vast majority of WSD experiments have considered a multilingual environment [21]. They are based on parallel resourses (e.g., wordnets and/or corpora) which makes them inapplicable to our goal. To the best of our knowledge, the only exception constitutes research on WSD for Czech [23].

In order to create a semantic valence dictionary, we need sense annotations for words which are immediate arguments of a verb (heads of phrases). Therefore, we apply the syntactic information (the valence of the main verb in a clause) which we have at our disposal to solve the WSD task. In fact, such information is rarely used to perform this task [16]. Although [11], a frequently cited work, does use some syntactic information, the authors disambiguate only the semantic category of a verb (based on classes proposed in [25]) using the set of possible verb classes and the syntactic frame of a sentence. They are not interested in the semantic categories of verb arguments.

In [18] we described our initial experiment in this task. We used the corpus of 165 263 single-verb sentences for 99 verbs. Sentences were parsed by means of the "permissive" version of the *Świgra* parser [44,43], which did not use any valence dictionary, and hence accepted any configuration of arguments with a verb. In that paper we focused on analysis of linguistic phenomena appearing in data set and their influence on the various steps of analysis.

At present we describe the whole process of analysis, starting from plain text. A special attention is given to algorithms we propose for slot-based WSD.

The composition of this article is as follows. In section 2, we present the resources used in our approach. In section 3, preparatory steps for data processing are describeed. In section 4, we discuss three versions of the EM selection algorithm used for slot-based WSD. In section 5, we present manual annotation of sentences used for evaluation. Finally, in section 6, we evaluate the algorithms and compare them. All example sentences presented throughout the paper come from a corpus used in the experiment.

2 Data Resources

Beginning our work, we paid a lot of attention to the choice of verbs to be considered in the experimental semantic valence dictionary [17]. The set of 32 verbs was chosen manually, taking into account the need to maximise the variability of syntactic frames (in particular, diathesis alternations) on one hand, and polysemy of verbs within a single syntactic frame, on the other. Another important criterion for this choice was frequency.

For the purpose of algorithm evaluation, we also needed a manually annotated subcorpus, referred to as HandKipi; with corresponding semantic categories assigned to the semantic head of each phrase (cf. section 5). We have prepared a test set which contains 5634 sentences for 32 verbs. The set of verbs chosen for experiments is referred to as ChVSet.

2.1 Corpus

Our main resource was the IPI PAN Corpus of Polish written texts [30], referred to as Kipi. The texts are segmented into paragraphs and sentences and annotated with morphosyntactic tags. The 2nd edition contains 250 mln segments (roughly, words).

From this corpus, we selected a small subcorpus, referred to as SemKipi, containing 195 042 sentences. Selected sentences contained:

(a) one or more verbs from ChVSet,
(b) at most three verbs in all.

Thus, we limited our considerations to the list of 32 verbs from ChVSet. The reason was that we could evaluate the algorithm only for the verbs from this list (cf. section 5).

2.2 Valence Dictionary

Our experiment was done using an extensive valence dictionary, which was specially prepared for the task (cf. section 3 below for the parser construction). The main component of our dictionary was Świdziński's [40] valence dictionary, which contains 1064 verbs. This dictionary was supplemented with entries for some additional verbs, as explained below.

Firstly, we made sure that the used valence dictionary provided entries for all 32 verbs of the benchmark set given in [17]. Only 24 of them were included in Świdziński's dictionary literally, whereas for 4 verbs we managed to adjust the available entries of their respective aspectual counterparts.[1] Entries for the remaining 4 verbs were elaborated by analysing entries of the automatically created valence dictionary by Dębowski [6]. All entries of this part of the dictionary were carefully studied, modified and augmented.

[1] In Polish, aspect is accomplished by different verbs.

Similar additions were done for other verbs. We adjusted the entries of the missing 269 aspectual counterparts of verbs described by Świdziński's dictionary. Moreover, we used the valence dictionary extracted from the corpus data by [6,7], from which the entries of 955 verbs were added verbatim to increase the coverage of our dictionary on SemKipi.

Syntactic valence frames are list of slot, which can be parametrised. Parameters are separated by colons. A nominal phrase (NP) is encoded as `np`, its only parameter is its case. A prepositional-nominal phrase (PP; encoded as `prepnp`) has two parameters: a preposition and the case of its NP complement.

Świdziński's dictionary includes frames that are subframes of other frames. The idea was to list all non-elliptic frames. For instance, *kupić* has a frame with slots for two nominal elements in accusative and nominative case, i.e., `np:acc np:nom`, which is instantiated in the sentence *Tak, panie, ja kupiłem pański dom.* (*Yes, sir, I bought your house.*). But larger frames can be found in this dictionary as well, such as `np:acc np:nom prepnp:od:gen` (cf. sentence *Ja kupiłem pański dom od pańskiej żony.*; *I bought your house from your wife.*). From the point of view of automatic parsing, however, listing subframes in the valence dictionary only slows down sentence parsing, since the *Świgra* parser accepts all subframes of each frame listed, to automatically account for the phenomenon of ellipsis. Thus, we deleted all subframes from our dictionary.

Table 1. Number of syntactic frames for verbs from ChVSet

lemma	frames	slots	lemma	frames	slots	lemma	frames	slots
bronić	9	7	powiedzieć	6	6	rozpoczynać	5	5
interesować	2	3	powtarzać	5	7	rozpocząć	6	5
kończyć	10	8	powtórzyć	4	5	skończyć	11	8
kupić	10	13	proponować	3	5	spotkać	2	4
lubić	2	4	przechodzić	10	14	trzymać	14	11
minąć	5	5	przejść	18	17	uderzyć	8	8
mówić	7	7	przygotować	6	6	widzieć	10	7
odnieść	4	6	przygotowywać	6	6	zaczynać	8	7
odnosić	6	6	przyjmować	8	8	zacząć	8	7
pisać	8	9	przyjąć	9	9	zakończyć	8	7
postawić	6	8	robić	13	10			

The numbers of syntactic frames for verbs from ChVSet are listed in Table 1 (the *frames* columns). In the *slots* columns we specify the number of NPs/PPs that can fill arguments of a particular verb (in any frame). The maximal number of syntactic frames per verb in the whole dictionary is 25, their mean is 2.8 and their median is 2. In the dictionary of selected verbs, it is 18, 7.9 and 7, respectively. In Świdziński's dictionary the numbers are 25, 3.9 and 3. The maximal number of arguments per frame (including subject and reflexive marker *się*) is 5, the median is 3 and the mean varies from 2.5 in the whole dictionary to 2.9 in the dictionary of selected verbs.

Having a variety of syntactic frames was an important criterion for selecting verbs for our experiments, which is confirmed by the largest number of frames per verb in the dictionary of selected verbs. On the other hand, the aforementioned Dębowski's method of valence dictionary creation chooses most frequent, strongly supported frames, i.e., a small number of rather short ones. This explains the smallest number of frames per verb in the whole dictionary.

2.3 Słowosieć—Polish Wordnet

In order to prepare an initial sense annotation for nouns (to be later automatically disambiguated), we used the Polish WordNet [8,10,9], called *Słowosieć* (English acronym PLWN). *Słowosieć* is a network of lexical-semantic relations, an electronic thesaurus with a structure modelled on that of the Princeton WordNet and those constructed in the EuroWordNet project. Polish WordNet describes the meaning of a lexical unit of one or more words by placing this unit in a network which represent such relations as synonymy, hypernymy, meronymy, etc. For the present work we do not use the whole structure of the net, but the set of 26 predefined categories (see Table 2) which are situated at the top of the actual hierarchy. Using these categories, 6917 nouns (most frequent in the balanced subcorpus of the KIPI) were classified by the Polish WordNet group. One to five categories are assigned to each noun. The mean of categories per noun is 1.2.

Słowosieć does not include gerunds, which are lematised to verbs. Their implicit semantic categories are `act`, `event`, `process` and `state`. This enlarges the actual mean (and probably median) of categories assigned to the semantic head of NPs/PPs.

Table 2. Predefined set of general semantic categories in Polish WordNet

Nr	name	Nr	name	Nr	name
01	Tops	10	feeling	19	possession
02	act	11	food	20	process
03	animal	12	group	21	quantity
04	artifact	13	location	22	relation
05	attribute	14	motive	23	shape
06	body	15	object	24	state
07	cognition	16	person	25	substance
08	communication	17	phenomenon	26	time
09	event	18	plant		

The linguists performing manual annotation of the test corpus HANDKIPI were allowed to extend the set of semantic categories of a particular noun (cf. section 5). They have added 893 categories to 726 nouns. As a result, the maximal number of semantic categories per noun has increased to 7, and their mean has increased to 1.4.

3 Data Preparation

Before we start WSD, we need to annotate sentences from SemKipi with their reduced parses and semantic categories of nouns.

3.1 Parsing

SemKipi was parsed with the *Świgra* parser [44,43] based on the metamorphosis grammar GFJP [39]. In contrast to the experiment discussed in [18], this time we used a version of the grammar provided with the valence dictionary presented in section 2.2. The parser assumes that the boundaries of sentences were already identified in the corpus, but it ignores the disambiguation of morphosyntactic annotation established by the tagger (i.e., it takes all tags produced by morphosyntactic analyser *Morfeusz* into account).

In order to reduce sparseness of data, for the present experiment we considered only phrases being arguments of a verb (i.e., the subject and complements included in its valence frames). This means that each obtained parse was reduced to its flat form representing only these phrases. Each phrase has the syntactic and the semantic head (cf. [31]). Usually, these heads are the same (i.e., they are heads of phrases determined by the grammar), but for instance, the syntactic head of a preposition phrase is the preposition whereas its semantic head is the noun (i.e., the head of the complement noun phrase). Observe that these slots are the most interesting ones for our further semantic analysis. It should be emphasised here that the parser recognises phrases as arguments of a verb only on the basis of their syntactic form. If a complement has the same syntactic form as an adjunct, two parses are obtained with one of the phrases treated as a complement, and the other one treated as an adjunct, only one of them being a proper interpretation of a sentence. Similar error-causing effects are caused by the syncretism of cases. As a result, we obtain *reduced parses* of a sentence composed of a verb and a set of slots.

Table 3 lists all types of slots (verb arguments) considered in the analysis. Only slots present in sentences qualified for final semantic analysis are listed.

The table shows that the distribution of frequency of prepositions is as Zipfian as of other words. As for NPs, the proportion of their frequencies is not surprising: nominative and accusative cases are the most frequent and genitive case is the least frequent. Observe that NPs are 4 times more frequent than PPs.

If a sentence is composed of more than one clause, its parses are decomposed into subparses, each of them transformed into a separate reduced parse, linked to the predicate of a clause.

3.2 Selection of a Single Parse

Świgra tends to produce large parse forests. The mean of parses per sentence is 11 418 041 672 whereas the median is 42. The maximal number of parses is 592 491 499 739 040 (with 707 reduced parses) obtained for the sentence *Kraków to jedna z pierwszych stolic duchowych Polski, Wawel to jest wzgórze wielkości*

Table 3. Argument types considered in the analysis and their frequencies

slot	freq.	slot	freq.	slot	freq.
nom	29722	na:loc	497	bez:gen	14
acc	23007	na:acc	484	przy:loc	11
o:loc	4341	dla:gen	381	u:gen	10
dat	3974	w:acc	227	po:loc	7
do:gen	3415	z:gen	143	pod:inst	4
inst	1810	przez:acc	88	za:inst	3
od:gen	1433	za:acc	75	o:acc	3
z:inst	1128	przed:inst	74	między:inst	2
w:loc	1001	nad:inst	33	wśród:gen	1
gen	532	NP total	59045	PP total	13337

naszego narodu, tu Go zostawimy wśród tych wszystkich wielkich duchów, które w ciągu całego tysiąclecia przeszły po tym wzgórzu. (*Cracow is among the prime Polish spiritual capitals. The Wawel hill is the peak of our nation's grandeur. We will let Him here, amidst all these great spirits that crossed this hill during the whole millennium.*). The number of reduced parses of a sentence is much smaller than the number of entire parse trees, the more so as we have only considered actual arguments of verbs (without adjuncts). Nevertheless, this number is still considerable (the mean of reduced parses per sentence is 19, the median is 7). The maximal number of reduced parses is 6085 (for 36 920 464 complete parses). It was obtained for the sentence *Tak więc możliwości obrony ze strony obrońcy nie doznawałyby znacznego uszczerbku w myśl tego, co w tym projekcie się proponuje.* (*Thus the defense capacity of the defendant would not suffer a significant loss from what is proposed in this draft.*). The first sentence has 4 predicates (including *jest* (*is*) and predicative *to* having similar function). The idiosyncrasy of *to* probably influences the number of parses. The second sentence has 2 predicates. Note the weak dependency between the number of complete and reduced parses.

Thus, for each sentence we obtain a forest of reduced parses. Observe that such a forest is heterogeneous, as it contains reduced parses of the main clause and its subordinate clauses together. The selection process of the appropriate parse (for each clause) consists of two steps: a purely statistical one that is preceded by one based on the heuristics described below.

First, *Morfeusz* contains an extremely rich dictionary of Polish words, some of them very rare (e.g., dialectical or archaic). Thus, we delete parses containing such words which are homonyms of other, more frequent words. The procedure is based on the frequency list of lexemes in the whole Kipi. Some very rare interpretations of words are explicitely listed.

Next, although Polish is a free word order language, some orders of words are very unusual (e.g., inverse orders). However, *Świgra* produces all proper parses of a sentence. In order to delete parses containing such rare constructs, we compare two parses and delete:

- the parse with an NP corresponding to a pair of consecutive ⟨PP, NP1⟩ in another parse;
- the parse with an NP corresponding to a pair of consecutive ⟨NP1, NP2⟩ in another parse, where NP1 is in genitive and NP, NP2 have the same head;
- the parse with AdjP corresponding to a pair of consecutive ⟨AdjP, NP2⟩ in another parse.

Additionally, for the sake of semantic annotation, we exchanged wh-pronouns being heads of NPs/PPs with nouns being heads of NP/PP of the main clause modified by the corresponding wh-clause. For instance, the clause *który tak mówi* (*who speaks like that*) from the sentence *Człowiek, który tak mówi, opisuje swoje wrażenie emocjonalne.* (*A man who speaks like that describes his emotional impressions.*) the subject wh-clause *który* is exchanged with the noun *człowiek* (*man*).

Finally, we chose only those reduced parses from the forest which were headed by verbs from ChVSet.

The resulting reduced parse forests were disambiguated by means of an EM algorithm referreded to as *EM selection algorithm*. The algorithm was proposed by Dębowski [6] for the task of creating a syntactic valence dictionary. It selects the most probable element of a list (here: a list of syntactic frames of a verb for a particular sentence) on the basis of information about the frequency of the occurence of these elements (frames) in the whole data set (of sentences). In section 4.1 we use this algorithm for choosing the most probable semantic frame from a list obtained for the fixed syntactic frame.

The algorithm consists of three steps. In the first step, it ignores sentences with more than 50 frames or having parses with some idiosyncratic words (such as interrogative and demonstrative pronouns *co* (*what*), *ten* (*this*)). Moreover, it separates parses of different clauses of a sentence. Thus, the resultant forests are homogeneous and linked to a particular predicative verb.

The second step consists in transforming reduced parses into syntactic valence frames they represent. The order of phrases in a sentence is neglected. Thus, the algorithm possesses no information about the frequency of syntactic frame instantiation in different orders, hence it cannot influence the probability of a frame for any particular sentence. This is the reason for using some heuristics for this phenomenon.

The number of frames for a particular clause could be smaller than the number of corresponding reduced parses, as two or more parses may correspond to the same frame. Note that this means that selecting a single frame does not entail selecting a single reduced parse. This phenomenon is exemplified in sentence (7); cf. section 4.

The third step carries out the actual EM selection algorithm. It operates on valence frames corresponding to particular reduced parses and selects the most probable one.

Let us show the whole process by examples. They contain lists of reduced parses of exemplary sentences. A reduced parse starts with a verb heading a clause, together with morphosyntactic tags separated by ':' sign (`aff,neg` show whether a sentence is affirmative or negative). A verb is acompanied by a list of phrases (`np, prepnp, advp` etc.). Phrase type is followed by its syntactic head, optionally its semantic head and other corresponding morphosyntactic tags. Numbers show boundaries of a phrase (or a whole clause) in a sentence.

(1) `% 'Góralskie przysłowie mówi, że późna zima mocno trzyma.'`
(The highlander's proverb says that late winter holds "strong".)

```
% trees:  24
0-19 mówić aff:fin:sg:_:ter   []
0-19 mówić aff:fin:sg:_:ter   [0-1:np:góralski:pl:acc:m2:ter]
0-19 mówić aff:fin:sg:_:ter                                                 +
     [0-1:np:góralski:pl:acc:m2:ter, 1-2:np:przysłowie:sg:nom:n:ter]
0-19 mówić aff:fin:sg:_:ter   [0-2:np:przysłowie:sg:acc:n:ter]
0-19 mówić aff:fin:sg:_:ter   [0-2:np:przysłowie:sg:nom:n:ter]
0-19 mówić aff:fin:sg:_:ter                                                 +
     [0-2:np:przysłowie:sg:nom:n:ter, 3-19:cp:że:ctr]
0-19 mówić aff:fin:sg:_:ter   [1-2:np:przysłowie:sg:acc:n:ter]
0-19 mówić aff:fin:sg:_:ter   [1-2:np:przysłowie:sg:nom:n:ter]
0-19 mówić aff:fin:sg:_:ter
     [1-2:np:przysłowie:sg:nom:n:ter, 3-19:cp:że:ctr]
0-19 mówić aff:fin:sg:_:ter   [3-19:cp:że:ctr]
5-9  trzymać aff:fin:sg:_:ter [5-7:np:zima:sg:nom:f:ter]
5-9  trzymać aff:fin:sg:_:ter                                               +
     [5-7:np:zima:sg:nom:f:ter, 7-8:advp:mocno]
```

(2) `% 'Prawie każdy bank przygotował na okres urlopowy specjalną linię kredytową dla klientów.'`
(Almost every bank had prepared a customised credit line for its clients for the holiday period.

```
% trees:  368
0-12 przygotować aff:fin:sg:m3:ter [0-3:np:bank:sg:nom:m3:ter,
     4-10:np:linia:sg:acc:f:ter, 10-12:prepnp:dla:klient:gen]
0-12 przygotować aff:fin:sg:m3:ter [0-3:np:bank:sg:nom:m3:ter,
                                    4-12:prepnp:na:linia:acc]
0-12 przygotować aff:fin:sg:m3:ter [0-3:np:bank:sg:nom:m3:ter,
                                    4-12:prepnp:na:okres:acc]
0-12 przygotować aff:fin:sg:m3:ter [0-3:np:bank:sg:nom:m3:ter,
         4-7:prepnp:na:okres:acc, 7-12:np:linia:sg:acc:f:ter]
```

(3) `% 'Sąd uznał, że w tym procesie nie interesuje go sprawa moskiewskich pieniędzy.'`
(The court of law decided that during this process it is not interested in the case of Moscow money.)

```
% trees:  150
0-21 uznać aff:fin:sg:m3:ter     [0-1:np:sąd:sg:nom:m3:ter]
0-21 uznać aff:fin:sg:m:ter      []                                ×
0-21 uznać aff:fin:sg:m:ter      [0-1:np:sąd:sg:acc:m3:ter]
4-13 interesować neg:fin:sg:_:ter []
4-13 interesować neg:fin:sg:_:ter [10-11:np:sprawa:sg:nom:f:ter]
4-13 interesować neg:fin:sg:_:ter
     [10-11:np:sprawa:sg:nom:f:ter, 11-13:np:pieniądz:pl:acc:m3:ter]
4-13 interesować neg:fin:sg:_:ter [10-13:np:pieniądz:pl:acc:m3:ter]
4-13 interesować neg:fin:sg:_:ter [10-13:np:sprawa:sg:nom:f:ter]
4-13 interesować neg:fin:sg:_:ter [9-10:np:on:sg:acc:mn:ter]
4-13 interesować neg:fin:sg:_:ter                                  +
     [9-10:np:on:sg:acc:mn:ter, 10-11:np:sprawa:sg:nom:f:ter]
4-13 interesować neg:fin:sg:_:ter
     [9-10:np:on:sg:acc:mn:ter, 10-13:np:sprawa:sg:nom:f:ter]
4-13 interesować neg:fin:sg:_:ter [9-11:np:sprawa:sg:nom:f:ter]
4-13 interesować neg:fin:sg:_:ter                                  +
     [9-11:np:sprawa:sg:nom:f:ter, 11-13:np:pieniądz:pl:acc:m3:ter]
4-13 interesować neg:fin:sg:_:ter [9-13:np:pieniądz:pl:acc:m3:ter]
4-13 interesować neg:fin:sg:_:ter [9-13:np:sprawa:sg:nom:f:ter]    +
```

The sentence (1) has 24 trees for 2 verbs, both belonging to ChVSet, with 10 reduced parses for *mówić* (*to say*) and 2 for *trzymać* (*to hold*). The heuristic part of the process chooses 2 reduced parses for *mówić* and 1 for *trzymać* (marked by + symbol).

The sentence (2) has 368 trees for just one verb *przygotować* (*to prepare*). The number of reduced parses is quite large (37), so we listed only the 4 chosen by the heuristics.

The sentence (3) has 150 trees for 2 verbs. However, only *interesować* (*to interest*) belongs to ChVSet, with 12 reduced parses obtained for it. 3 of the parses were chosen by the heuristics.

In the preparation step, the EM selection algorithm finds syntactic valence frames corresponding to a particular reduced parse (for each verb separately). For the above examples we obtain:

ad (1)
- *mówić*:
  `mówić np:acc np:nom`
  `mówić np:nom sentp:że` +
- *trzymać*:
  `trzymać advp np:nom` +

ad (2) *przygotować*
`przygotować np:acc np:nom prepnp:dla:gen` +
`przygotować np:acc np:nom prepnp:na:acc`

ad (3) *interesować*
`interesować np:acc np:nom` +

The frames selected by the EM selection algorithm are marked by the + symbol. For sentences (1) and (3) the selected frames are the correct ones. For sentence

(2) both source frames are actually subframes of an entire frame, so both choices are equally correct. Nevertheless, the evaluation of the process of frames selection on manually annotated data shows that only 43.5% of manually annotated sentences were given the same syntactic valence frame as during automatic process (see Table 7). This is the significantly worse result than the 75% reported in [6,7]. The most probable reason of the difference is the way the set of sentences was selected.

Finally, the reduced parses that match selected frames are found. The reduced parse chosen for sentence (2) is shown in (4).

Unfortunately, each of the applied tools impose specific conditions on the input sentences, which results in a reduction of the subcorpus. First, the present version of *Świgra* analyses only a subset of Polish syntactic constructions. Next, the EM selection algorithm discards sentences that do not satisfy the conditions mentioned above. Finally, all clauses containing no NPs/PPs (such as *Ogromnie się cieszę.*; *[I] am very glad.*) are neglected.[2] As a result, after taking into account these constraints, the number of sentences has decreased from 195 042 to 47 184. The reduction of the number of sentences during the whole process is presented in Table 4. The table contains the lemmata of verbs, their aspect (i for imperfect and p for perfect), English glosses, the number of sentences containing these verbs in the whole SemKipi, the number of sentences (per verb) parsed by *Świgra*, the number of sentences after the heuristic phase and after EM-based valence selection. The last column shows the number of sentences with NPs/PPs such that at least one of them has a semantic category (i.e., it is not a pronoun, an adjective interpreted as noun, etc.). The general statistics of sentences, their reduced parses and the phrases they include, are presented in Table 5.

3.3 Semantic Categories

In our current work we are interested in assigning semantic categories to nouns being semantic heads of NPs and PPs. Thus, we can ignore other syntactic slots.[3]

For this raeson, we provide a list of semantic categories for each semantic head of an NP or a PP in each reduced parse of every clause. The categories are taken from *Słowosieć*. Assuming that a sentence is not ambiguous and its syntactic frame has been properly selected, exactly one reduced parse is adequate and exactly one category is adequate for every NP/PP.[4] The resultant syntactic-semantic reduced parse for sentence (2) corresonding to the chosen syntactic frame (in <> brackets) is presented in (4).

[2] Such clauses contain no nouns and hence no semantic categories of nouns to disambiguate. Thus, they would have a single semantic frame to choose and they would not influence the algorithm's behaviour.

[3] Neglecting the existence of other syntactic slots means reducing the number of different valence frames, hence reducing the sparseness of data. However, this also means some loss of information.

[4] Unfortunately, this assumption is not always met.

(4)
```
% 'Prawie każdy bank przygotował na okres urlopowy
   specjalną linię kredytową dla klientów.'
<przygotować np:acc np:nom prepnp:dla:gen>
0-12 przygotować aff:fin:sg:m3:ter
                 [0-3:np:bank:sg:nom:m3:ter::   group location,
                 4-10:np:linia:sg:acc:f:ter::   shape location relation,
                 10-12:prepnp:dla:klient:gen::   person]
```

Table 4. The reduction of number of sentences per verb

lemma	asp.	gloss	source	parsed	prepared	valence selected	parses with sem.categs.
bronić	i	protect	2547	821	805	763	646
interesować	i	interest	2850	1113	1102	922	873
kończyć	i	finish	3982	1587	1551	1425	1332
kupić	p	buy	2293	544	527	477	452
lubić	i	like	3194	1519	1473	1358	939
minąć	p	pass	2206	719	668	520	499
mówić	i	speak	35792	10418	9993	8874	6072
odnieść	p	*polysemous*	4952	1260	1259	1210	1129
odnosić	i	*polysemous*	2626	869	869	775	748
pisać	i	write	3574	1305	1266	1174	958
postawić	p	put	3260	958	956	931	879
powiedzieć	p	say	28882	5286	4635	3912	1907
powtarzać	i	repeat	2821	759	637	555	355
powtórzyć	p	repeat	2154	391	374	310	224
proponować	i	propose	6351	3164	3142	2977	2194
przechodzić	i	*polysemous*	2592	1537	1531	1478	1442
przejść	p	*polysemous*	5032	1875	1861	1772	1659
przygotować	p	prepare	3735	1234	1228	1194	1147
przygotowywać	i	prepare	2470	1135	1130	1101	1052
przyjmować	i	*polysemous*	2554	1067	1036	937	868
przyjąć	p	*polysemous*	8155	3809	3743	3192	3091
robić	i	do/make	10215	3355	3221	2175	1646
rozpoczynać	i	start	1669	877	873	857	834
rozpocząć	p	start	8707	4096	4089	4032	3852
skończyć	p	finish	4213	1383	1363	1203	991
spotkać	p	meet	4174	1452	1451	1382	1166
trzymać	i	hold	1870	673	670	622	532
uderzyć	p	hit	1017	476	473	448	372
widzieć	i	see	10006	3890	3769	3382	2550
zaczynać	i	start	3642	1636	1623	1553	1126
zacząć	p	start	10872	4670	4661	4508	3059
zakończyć	p	finish	6635	2816	2801	2706	2590
total	–	–	195042	66694	64780	58725	47184
mean	–	–	6095	2084	2024	1835	1475
median	–	–	3642	1383	1363	1210	1126

Table 5. Simple statistics for sentences and their parses

(*Source number of sentences: 195 042*)	after syntactic analysis	after disambiguation of parses	sentences with NPs/PPs only
Number of sentences	74 904	58 723	47 194
Number of reduced parses	1 405 989	101 576	73 153
Number of empty parses	70 264	0	0
Mean of reduced parses per clause	18.772	1.730	1.550
Median of reduced parses per clause	7	1	1
Number of phrases	2 447 026	192 708	143 542
Number of NPs and PPs	2 049 558	159 312	122 941
Mean of phrases per parse	1.832	1.897	1.962
Median of phrases per parse	2	2	2
Mean of NPs and PPs per parse	1.534	1.568	1.681
Median of NPs and PPs per parse	2	2	2

4 EM Selection Algorithm for Semantic Category Disambiguation

Our goal here is the disambiguation of semantic categories of nouns being semantic heads of NPs and PPs. To solve this task, we adapted the EM selection algorithm (a version of a well known Expectation Maximisation algorithm) initially used by Dębowski in [6] to select the correct valence frame of a clause. The algorithm is not supervised, i.e., it does not need any training data. On the other hand, its behaviour is highly dependant on the size of the corpus.

We decided to work on semantic valence frames of a clause instead of entire parses. Thus, we combined sets of categories attached to each NP and PP from every reduced parse.

Next, we split the resultant syntactic-semantic valence frames in such a way that each NP/PP had only one category assigned. The disambiguation process consists in selecting the most probable frames. So, the sentence (2) has the semantic frame shown in (5) which after splitting transforms into the 6 frames presented in (6):

(5)
```
% 'Prawie każdy bank przygotował na okres urlopowy
   specjalną linię kredytową dla klientów.'
<przygotować np:acc:shape,location,relation np:nom:group,location
             prepnp:dla:gen:person>
```

(6)
```
acc: shape,    dla:gen: person, nom: group
acc: shape,    dla:gen: person, nom: location
acc: location, dla:gen: person, nom: group
acc: location, dla:gen: person, nom: location
acc: relation, dla:gen: person, nom: group
acc: relation, dla:gen: person, nom: location
```

As we have noticed in section 3.2, there exist sentences (clauses) with more than one reduced parse matching the selected syntactic frame. The typical case of such situation occurs in simple sentences with the nominative subject and accusative object, when nouns that occupy these positions have syncretic nominative and accusative case. To eliminate such cases we impose special conditions on the splitting operation. We demonstrate how they work on the example (7). Each NP head in each reduced parse has 2 semantic categories assigned. So, their maximal number per NP/PP head is 2 as well. However, the number of semantic categories of each slot of the semantic frame (presented in <> brackets) is 4 and the sets of categories are equal. This suggests that they were combined from two reverse reduced parses, which is the reason for ignoring the sentence.

(7) `% 'Zarząd widzi możliwość kontynuacji testów.'`
([A/The] board of management notices the possibility of continuing tests.)
```
<widzieć np:acc:act,group,state,top np:nom:act,group,state,top>
0-5 widzieć aff:fin:sg:_:ter
            [0-1:np:zarząd:sg:acc:m3:ter::  act group,
            2-5:np:możliwość:sg:nom:f:ter::  state top]
0-5 widzieć aff:fin:sg:_:ter
            [0-1:np:zarząd:sg:nom:m3:ter::  act group,
            2-5:np:możliwość:sg:acc:f:ter::  state top]
```

4.1 EM Selection Algorithm Applied to Whole Semantic Frames

Firstly, we tried using the EM selection algorithm in the same fashion as it was originally used by Dębowski in [6] to select the correct syntactic parse, namely, to treat each semantic frame as an atomic value. This approach will be called the *EM-whole* algorithm here. The scheme is as follows.

Let A_i be the set of alternative pairs $\langle v, f\rangle$ of verb v with frame f for the i-th clause with verb v in the corpus, $i = 1, 2, ..., M_v$, where M_v is the number of clauses containing verb v. Moreover, symbol $p_n(v, f)$ will stand for the effective probability, or frequency, of verb v with frame f in the n-th iteration. We set the initial equidistribution $p_1(v, f) = 1$ and iterated:

$$p_n(v, f|i) := \begin{cases} p_n(v, f)/\sum_{\langle v', f'\rangle\in A_i} p_n(v', f') & \text{for } \langle v, f\rangle \in A_i, \\ 0 & \text{else,} \end{cases}$$

$$p_{n+1}(v, f) := \frac{1}{M_v}\sum_{i=1}^{M_v} p_n(v, f|i).$$

The iteration was stopped at $n = 15$ and, for each clause i, we selected all pairs $\langle v, f\rangle$ that had the maximal frequency $p_n(v, f|i)$.

The algorithm is based on the observation that:

1. effective probability $p_n(v, f|i)$ that frame f of verb v is appropriate for clause i is proportional to the frequency of frame f in all clauses of verb v;

2. it is better to consider effective probability $p_n(v,f)$ of frame f in all clauses of verb v (normalised w.r.t. number of clauses M_v) counted proportionally to effective probability $p_n(v,f|i)$ of frame v in each clause i instead of simple frequency.

At the beginning, all semantic frames of a verb are equally probable. Thus, the algorithm estimates the probability of semantic frames of a clause on the basis of their overall probability for a verb predicating the clause.

4.2 EM Selection Algorithm for Independent Occurrence of Verb Arguments in a Frame

According to the experimental results of [18], the *EM-whole* algorithm very often selects more than one frame per clause. This is due to the sparseness of data: quite frequently each frame of a particular clause occurs only once in the data set. Thus, all interpretations of the clause remain equally probable during EM iterations.

In order to reduce the sparseness of data, we decided to modify the selection algorithm so that the occurrence of each NP/PP argument $a \in f$ of a semantic frame f be treated as independent of the occurrence of other arguments. In this version, called the *EM-indep* algorithm, we iterated:

$$p_n(v,f|i) := \begin{cases} \prod_{a\in f} p_n(v,a) / \sum_{\langle v',f'\rangle\in A_i} \prod_{a'\in f'} p_n(v',a') & \text{for } \langle v,f\rangle \in A_i, \\ 0 & \text{else,} \end{cases}$$

$$p_{n+1}(v,a) := \frac{1}{M_v} \sum_{i=1,2,\dots,M_v : a\in f\in A_i} p_n(v,f|i),$$

with the initialization $p_1(v,a) = 1$.

The above changes are the consequences of the following:

1. due to the assumption of independence, effective probability $p_n(v,f)$ is a product of $p_n(v,a)$ for $a \in f$;
2. we count effective probability (frequency) $p_n(v,a)$ of argument a considering all frames f including it.

Thus, this time the algorithm estimates the probability of semantic frames of a clause, basing on the overall effective probability of all arguments from this very frame.

4.3 Incremental Version of the EM Selection Algorithm

The assumption that the arguments of a verb in a clause are independent is an obvious simplification. Thus, either version of the algorithm presented above has its own merits and limitations. In [18] it was observed that the *EM-whole* algorithm tends to make no decisions, whereas the *EM-indep* algorithm makes more mistakes.

In order to combine the merits of both algorithms and to reduce their shortcomings, we decided to use yet another solution, referred to as the *EM-incr* algorithm. For clauses having frames composed of at most two NP/PP slots, the *EM-incr* algorithm works the same way as the *EM-whole* algorithm. For clauses having larger frames, *EM-incr* incrementally applies the *EM-whole* for subframes composed of 2, 3 up to all slots. First, all 2-element syntactic subframes of a whole frame are considered. The larger subframes are established on the basis of the subframe chosen in the previous step. Each time one slot is added to each of the selected subframes, i.e., we obtain all its one-slot larger superframes. The EM procedure treats semantic frames corresponding to different syntactic frames independently (as *EM-whole* applied here). Nevertheless, we select the most probable frame(s) from the whole set. Thus, in the subsequent step we work on frames obtained from syntactic frames which semantic frames were chosen as most probable ones (for a particular sentence). Finally, we select the most probable frames (subframes composed of all slots).

Using this procedure, we have to disambiguate only the semantic categories of the last-added slot. The only exception is the first step, when the algorithm works on subframes containing two non-disambiguated slots. On the other hand, the number of clauses relevant to a particular subframe is larger than in the case of the *EM-whole* algorithm for frames of the same syntactic nature; the shorter frame is considered, the larger number of clauses. Both features should reduce the sparseness of data.

4.4 The Degree of Reduction of Semantic Categories

In order to disambiguate semantic categories, the algorithms have to reduce their number per noun.
In Table 6 we present the mean of the number of semantic categories assigned to occurrences of nouns, including baseline algorithms presented in section 6.2. For each data set (cf. section 5.1), the first column shows the mean for all nouns; in the second column pronouns were not counted. Median is always 1, even for source data. Initially, all 26 semantic categories are assigned to pronouns.

All the algorithms reduce the number of semantic categories per noun. The *MaxNoun* algorithm makes no decisions for pronouns and nouns from outside the set of manually annotated nouns (cf. section 5), which results in its poor behaviour.

Table 6. Results of reducing the number of semantic categories per noun

algorithm	AutoProc		ValMatch		ParseMatch	
source	3.423	1.852	3.434	1.851	3.435	1.816
EM-indep	1.002	1.002	1.002	1.002	1.003	1.003
EM-whole	1.077	1.026	1.080	1.027	1.080	1.027
EM-incr	1.120	1.026	1.126	1.027	1.122	1.028
MaxNoun	2.901	1.289	2.913	1.290	2.921	1.262
MaxSlot	1.002	1.002	1.002	1.002	1.002	1.002

5 Manually Annotated Data for Evaluation of Algorithms

In order to evaluate the algorithms, a small subcorpus of SemKipi was syntactically and semantically annotated by a group of linguists. 240 sentences for each verb from ChVSet were selected randomly from SemKipi.[5] The linguists performed three different tasks:

1. correction of morphosyntactic tagging (tagger errors),
2. division of sentences into phrases, i.e., pointing out their boundaries and syntactic and semantic heads,
3. assignment of a single PlWN semantic category to each noun in a sentence.

The correction of morphosyntactic tagging was limited to choosing one of the suggestions provided by the morphosyntactic analyser. In contrast, the linguists were allowed to add semantic categories of nouns (from the whole repertoire of categories presented in Table 2) that were not anticipated by PlWN authors for a particular noun.[6] Considered phrases should be semantically connected to a verb (its complements and adjuncts). Functional expressions were ignored. If any of the above requirements could not be satisfied (e.g., the sentence was grammatically incorrect), the linguists were allowed to reject it.

The process of annotation was performed by means of the program *Anotatornia* via the Internet [20]. Each sentence was annotated by two linguists. In the case of differences between their annotations, a process of negotiations were performed until they reached an agreement. If the negotiations took more than two iterations, the moderator judged the sentence.

The process of manual annotation is still in progress. Thus, we should keep in mind that until the process of manual annotation is finished, its results are not stable and could change. The statistics for hand-annotated sentences w.r.t. automatic processing is presented in Table 7. Since the linguists were allowed to reject sentences, the column *hand* shows the number of accepted ones. Their total number is 5634, which means 94%. Note that sentences in HandKipi do not need to have NP/PP arguments, as there were not such criteria for their selection. Nevertheless, as many as 91.2% of accepted sentences include at least one NP/PP with assigned semantic category (column *with sem. categs.*). HandKipi is simply a subset of SemKipi, and only 43.3% of manually annotated sentences were accepted by the automatic process (column *common*); most of them were lost during *Świgra* parsing. The reason is that linguists could annotate bizarre sentences which are rejected by the parser.[7] Worse still, only 43.5% of sentences belonging to both sets had the same syntactic valence frame chosen automatically and by the linguists (column *common valences*).

[5] More precisely, only single-verb sentences were chosen for this sampling.

[6] The reason for this decision is that the PlWN is not yet completed and our work bases on the preliminary classification.

[7] All percentages in the table are counted w.r.t. the *hand* column in order to show the reduction of data.

Table 7. The statistics for annotated sentences

lemma	hand	with sem. categs.	proc.	common	proc.	common valences	proc.
bronić	216	188	87.0	83	38.4	42	19.4
kończyć	229	195	85.2	95	41.5	70	30.6
kupić	231	221	95.7	63	27.3	11	4.8
lubić	229	196	85.6	69	30.1	36	15.7
minąć	221	218	98.6	57	25.8	21	9.5
mówić	221	189	85.5	56	25.3	10	4.5
odnieść	234	220	94.0	102	43.6	56	23.9
odnosić	234	223	95.3	83	35.5	51	21.8
pisać	223	192	86.1	67	30.0	14	6.3
postawić	228	224	98.2	100	43.9	36	15.8
powtarzać	202	168	83.2	70	34.7	30	14.9
powtórzyć	236	204	86.4	33	14.0	14	5.9
proponować	223	217	97.3	103	46.2	61	27.4
przechodzić	233	225	96.6	147	63.1	15	6.4
przejść	227	211	92.9	113	49.8	30	13.2
przygotować	235	223	94.9	96	40.9	39	16.6
przygotowywać	231	222	96.1	125	54.1	48	20.8
przyjmować	226	220	97.3	97	42.9	37	16.4
przyjąć	227	225	99.1	126	55.5	55	24.2
robić	228	160	70.2	65	28.5	29	12.7
rozpoczynać	224	218	97.3	118	52.7	64	28.6
rozpocząć	226	219	96.9	144	63.7	95	42.0
skończyć	229	198	86.5	63	27.5	36	15.7
spotkać	227	190	83.7	73	32.2	45	19.8
trzymać	194	171	88.1	71	36.6	20	10.3
total	5634	5137	91.2	2219	43.3	965	18.7

The percentage for sentences accepted both by linguists and the automatic process is quite irregular and varies from 14.0 (*powtórzyć, to repeat*) to 63.7 (*rozpocząć, to start*). Similarly, the percentage of sentences with the same frames ("common valences") varies from 4.5 (*mówić, to speak*) to 42.0 (*rozpocząć, to start*).[8] This influences the number of sentences used for the actual evaluation of the algorithm. We should remember that the results of the evaluation for verbs with the smallest percentages are less valuable.

5.1 Preparation of Data

One of the consequences of the ongoing work of manual annotation is is the fact that obligatory arguments are not differentiated from adjuncts. As a result,

[8] Counted w.r.t. *hand* column. The percentage of sentences with the same syntactic frame obtained by computer and manual processing counted w.r.t. *common* column varies from 10.5 (*przechodzić,* polysemous) to 73.7 (*kończyć*).

these sentences are connected with syntactic frames that have no counterparts in *Świgra*-parsed sentences (are their superframes). This would influence the performance of the algorithm for HandKipi sentences. In order to avoid such an influence, we have automatically selected maximal subframes corresponding to valence dictionary frames for each sentence. However, this procedure is error-prone.

As we have noticed above, only 43.3% of sentences from HandKipi underwent computer processing. Furthermore, the reduced parses obtained by the EM selection algorithm in the data preparation process (section 3) and through manual annotation differs for 66.5% sentences. This influences the data and hence the evaluation procedure.

Thus, we decided to evaluate all the algorithms on three sets of sentences. The first set (referred to as AutoProc) consists of purely automatically preprocessed sentences. Then the evaluation is performed on manually annotated sentences that were automatically parsed as well. The second set (referred to as ValMatch) was obtained by merging valence schemata of automatically preprocessed sentences and manually annotated ones. Then reduced parses were again found automatically. The third set (referred to as ParseMatch) consists of automatically preprocessed sentences extended with manually annotated ones. If a sentence belongs to both sets, the manually annotated version was chosen. Manually annotated semantic categories of nouns were deleted. Thus, evaluation performed on this set is independent from the preprocessing steps of the analysis, hence it is most reliable if we want to evaluate slot-based WSD step solely instead of the whole process.

6 Evaluation and Comparison of the Algorithms

Ultimately, we had three versions of the EM selction algorithm at our disposal. In order to know how they actually work we had to evaluate each of them.

We decided to evaluate the algorithms using the simplest measure, i.e., correctness, which means counting all scores for every evaluated item w.r.t. the number of items. We think that in the case when the manual annotation assigns a single semantic category to each noun, this measure is satisfactory. Traditionally, if semantic categories of an NP/PP head are equal in computer and manual annotation, the score is set to 1. However, in our experiments each algorithm can choose more than one semantic category, and then one of them can match manual annotation. In order to take into account such situations in the evaluation as partial success, we decided to set the score to $\frac{m-k}{m \cdot k}$, where m is the number of all semantic categories of a noun, and k is the number of categories chosen by the algorithm (for $k \geq 2$). This score increases for decreasing k and increasing m, hence for our data it obtains the maximal value 0.375 for $k = 2$ and $m = 8$ (for pronouns $m = 26$, hence we obtain maximal value 0.462). The fact that the mean of k is 1.003 for the *EM-indep* and *MaxSlot* algorithms and 1.12 for *EM-whole* and *EM-incr* shows that this way of coutning scores for multi-element choices does not significantly influence the resultant evaluation of a particular algorithm. On the other hand, the *MaxNoun* algorithm makes no decisions for pronouns and some nouns, which means $k = m$, i.e., we obtaind 0 score. Thus, the average $k = 2.9$ does not entail a big influence on the evaluation of this algorithm as well.

The evaluation has been performed in two ways. Firstly, we compared each NP/PP slot separately. Secondly, we evaluated entire sentences, i.e., we checked whether the results agree for every NP/PP slot in a sentence. Then, scores for all NPs/PPs in the sentence were added.

6.1 Choice of the Number of Iterations

The number of iterations n of each EM algorithm had to be set experimentally. In this section we show the corresponding procedure.

The results of the evaluation for n set to 6, 10, 15 and 20 are summarised in Table 8. The evaluation was carried out in two ways: for entire sentences and for separate slots (cf. section 6), which is shown in the figure. For each algorithm, the best result of the evaluation is underlined.

Table 8. Evaluation of the EM selection algorithms for various number of main iteration

		whole sentences				separate NPs/PPs			
data set	EM. alg.	6	10	15	20	6	10	15	20
AutoProc	whole	74.15	74.27	74.37	74.37	77.70	77.78	77.89	77.86
ValMatch	whole	71.06	71.43	71.42	71.42	76.78	77.10	77.09	77.05
ParseMatch	whole	80.45	80.52	80.59	80.55	84.43	84.51	84.54	84.50
AutoProc	indep	76.13	76.13	76.12	76.12	80.82	80.86	80.86	80.86
ValMatch	indep	73.02	73.02	72.98	72.98	80.33	80.33	80.31	80.30
ParseMatch	indep	80.97	81.08	81.03	81.03	86.16	86.21	86.18	86.18
AutoProc	incr	74.81	74.86	75.08	74.91	78.74	78.85	79.02	78.90
ValMatch	incr	71.63	71.78	71.84	71.70	77.87	78.02	78.06	78.03
ParseMatch	incr	80.61	80.79	80.72	80.66	84.60	84.69	84.54	84.46

First, observe that the results are very regular: the differences between them never exceed 0.4%. Next, there is an increase between 6 and 10 iterations for all cases except 2 equalities for whole sentences and 1 equality for separate slots. Surprisingly, in most cases 20 iterations execution of the algorithms show worse results than the best ones. The only exception is the *EM-whole* algorithm evaluated on whole sentences and *EM-indep* algorithm evaluated on separete NPs/PPs, both for AutoProc data set, where the result for 20 iterations equals the best one.

Thus, the reasonable choice is between 10 and 15 iterations. In this area, we can observe an increase in results as often as their decrease. The average increase is 0.08% and the average decrease is 0.04%. Decreases plague the *EM-indep* algorithm, hence we have chosen 10 iterations for it. The most heterogeneous is the *EM-incr* algorithm, with the maximum increase for the AutoProc data set (0.22%, 0.17%) and the maximum decrease for ParseMatch data set (0.05%, 0.15%). We have chosen 15 iterations.

6.2 Baseline Algorithms

A comparison of the three versions of the algorithm does not suffice to make an equitable evaluation. Since we cannot compare our results against any existing WSD work for Polish, some baseline solutions are needed.

Word meanings exhibit Zipfian distribution. Hence a very popular choice of a baseline algorithm for the WSD problem is *the most frequent sense of a word* [15,1]. However, this heuristics is applicable only to those few languages for which significantly large sense-tagged corpora are available. Currently, we have no such resource at our disposal. Instead we can use the semantic category that was most frequently assigned to a particular word in HandKipi (referred to as the *MaxNoun* algorithm). However, this method is highly dependent on this very corpus. Note that *MaxNoun* makes no decisions for nouns absent in HandKipi, similarly as for pronouns.

In Table 9 we present some statistics for manually annotated data, counted for all nouns and for nouns with more than one semantic category (referred to as *multi-cat. noun*). *Frequency* is referred to as *freq.* and *semantic category* is referred to as *sem.cat.* *Freq. with sem.cat.* means the number of times a particular noun has a particular *sem.cat.* assigned; *freq of max. sem.cat.* means how many times a particular noun has the most often *sem.cat.* assigned. Only nouns present in HandKipi are considered. We have 899 nouns in HandKipi with obvious Zipfian distribution, the most frequent is the pronoun *to* (*this*) with 596 occurrences. Observe that about 2/3 of occurrences of nouns in sentences are polysemous nouns. Additionally, 426 (25%) semantic categories of nouns were never assigned (0 frequency). The most important information is that 82.66% of the occurrences of nouns are the most frequent semantic category occurrences (74.67%, with mean 2.4 semantic categories, for polysemous nouns).

The percentage of semantic categories of a noun not assigned to any token, and the percentage of tokens assigned to the most frequent semantic category of a noun, is puzzling. Thus, the evaluation of the *MaxNoun* algorithm on the same corpus for which the most frequent semantic categories were counted is unreliable. The situation is similar to the case when an ML algorithm is overtrained.

Therefore, we propose another baseline algorithm (referred to as the *MaxSlot* algorithm), which does not use data from HandKipi, hence it is independent of it. For each verb and each syntactic slot, it counts all occurrences of a particular semantic category (for multi-category nouns, the corresponding fraction is considered) in a given set of sentences. Next, for each NP/PP, the most frequent semantic category for this slot is chosen (from the categories of its semantic head).

Table 9. Statistics for semantic categories of nouns in HandKipi

	nouns all	nouns multi-cat.
total nouns	899	592
total sem.cat.	1734	1427
max. sem.cat.	8	8
mean sem.cat.	1.93	2.41
med. sem.cat.	2	2
total noun freq.	8300	5681
total freq. of max. sem.cat.	6861	4242

	nouns all	nouns multi-cat.
max. noun freq.	596	201
mean noun freq.	9.23	9.60
med. noun freq.	2	3
max. freq. with sem.cat.	596	131
mean freq. with sem.cat.	4.79	3.98
med. freq. with sem.cat.	1	1
proc. freq. of max. sem.cat.	82.66	74.67

Table 10. General results of evaluation of the algorithms

	whole sentences			separate NPs/PPs		
algorithm	AutoProc	ValMatch	ParseMatch	AutoProc	ValMatch	ParseMatch
EM-indep	76.12	72.98	81.03	80.86	80.30	86.17
EM-whole	74.39	71.50	80.59	78.53	77.80	84.53
EM-incr	75.04	71.84	80.81	78.99	78.06	84.68
MaxNoun	72.71	71.42	84.90	71.57	71.56	86.26
MaxSlot	75.47	72.33	81.40	80.52	80.07	86.46

6.3 General Evaluation

Now we have 5 algorithms to evaluate on 3 sets of sentences. To begin, we want to discuss the general results for all the sets together. They are summarised in Table 10. Note that the results are very similar for all the algorithms. First of all the algorithms show the best results for the ParseMatch set of data, in which HandKipi is included without any changes. Hence, the evaluation is performed on the correct data, without influence of the preprocessing phases. However, we should keep in mind that the algorithms are executed on larger (for most verbs substantially larger) set of sentences. The syntactic frames of sentences from outside HandKipi were obtained through an automated process.

The only exception is the *MaxNoun* algorithm. Its behaviour depends only on a particular noun. What is more, it is based on the maximal frequency of nouns counted for HandKipi itself (cf. section 6.2). Therefore, good behaviour of this algorithm for this set is natural. Still, evaluated on separate NPs/PPs, the *MaxSlot* algorithm is even better.

All the algorithms obtain the worst results on the ValMatch set of sentences. Observe that this set of data is evaluated on the smallest set of sentences (cf. column *common valences* in Table 9), hence the results are less reliable. As we noticed in section 3.2, the original EM selection algorithm used to select a valence frame tends to choose the simplest frames. Such simple frames are usually adequate for simple sentences. Thus, probably most sentences for which hand and automatically selected frames match are simple. This could influence the EM algorithms based on co-occurrence of slots.

Both AutoProc and ValMatch data sets were best processed by the *EM-indep* algorithm, which obtains the best results within the EM algorithms group for ParseMatch data sets as well. The worst results are obtained by the *MaxNoun* algorithm. The most likely reason may be that the choice of nouns as heads of slots during the parsing process, particularly for frequent nouns, was incorrect or at least ambiguous.

The *EM-whole* algorithm is the worst within the EM algorithms group which is not surprising, because of data sparseness (cf. syntactic frames variety shown in Table 1). Table 5 shows that frames of most sentences include 1 or 2 arguments, which explains why the results of *EM-incr* are so similar to the results of EM-whole. The differences between these two algorithms would probably be more distinct for sets of sentences with longer frames.

However, the fact that the *EM-indep* algorithm behaviour is better than the *EM-incr* algorithm contradicts our intuition and assumptions based on data investigation (cf. [18]). This would suggest that semantic categories assigned to particular NP/PP heads are independent of each other! Observe that the *MaxSlot* algorithm based on the frequency of semantic categories counted independently for each slot of a verb, obtains very similar results.

Another possible explanation is that sparseness of data is still less crucial for the independent counting of arguments than for treating of frames as a whole.

6.4 Details of the Evaluation

Let us discuss the results of the evaluation in detail. Figure 1 shows evaluation of the algorithm performed on particular NPs/PPs w.r.t. frequency of verbs in AutoProc, ValMatch and ParseMatch sets of sentences, respectively. Figure 2 shows the evaluation of the algorithm performed on whole sentences in similar fashion. The frequency axis is logarithmic.

The frequency of verbs shows no distinct impact on the results of the algorithms; the percentage of matches changes rapidly from one set of sentences to another. The differences are greater for the evaluation performed on whole sentences.

On the other hand, all the algorithms behave similarly for both methods of evaluation and all sets of sentences. The only exception is the *MaxNoun* algorithm. Observe that it shows more regular results, especially for whole sentence evaluation and for the ParseMatch set. This is an obvious consequence of its sole dependience on nouns.

All the algorithms[9] show best results for verbs *minąć, rozpoczynać, rozpocząć* for all data sets and both methods of evaluation. Note that the frequency of these verbs differs greatly. Vast majority of sentences predicated with the less frequent verb *minąć* (*to pass*) have some period of time as a subject. Usualy this is the simplest syntactic frame `np:nom`, e.g., *Styczeń minął.* (*January*

[9] In what follows we ignore the *MaxNoun* algorithm.

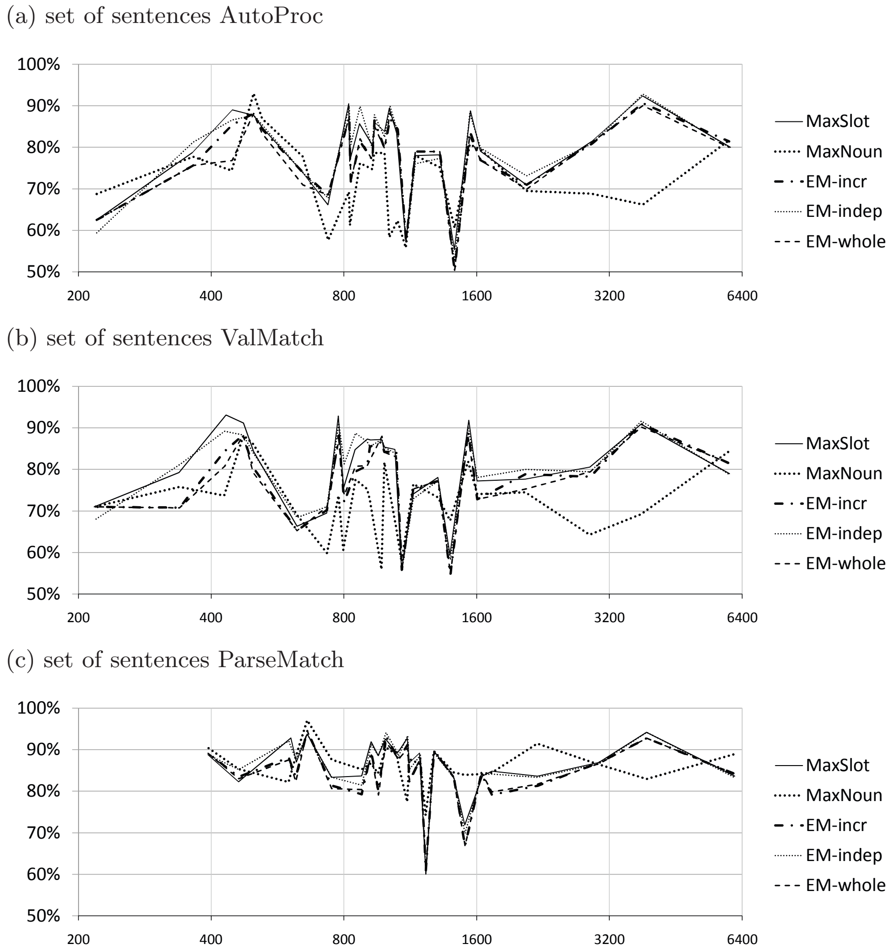
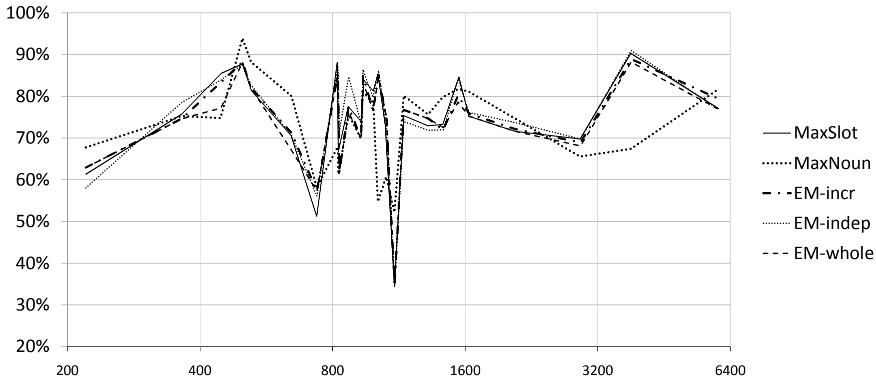


Fig. 1. Diagrams of evaluation of the algorithms on particular NPs/PPs

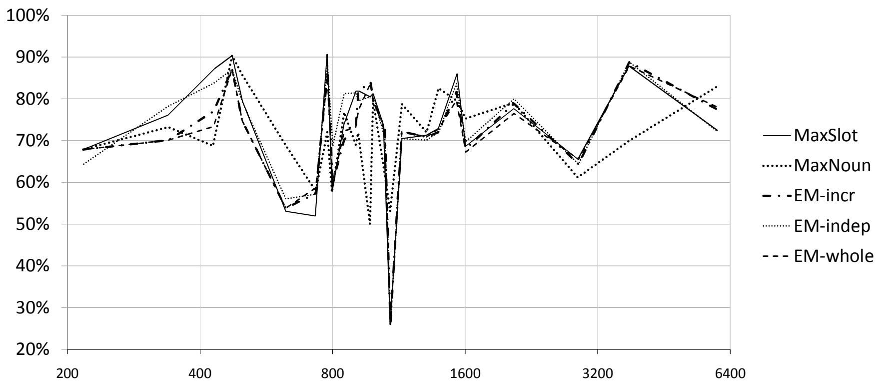
passed.) or *Godzina minęła bardzo szybko.* (*An hour passed very quickly.*). A similar situation exists for the frame `np:nom prepnp:od`, e.g., *Od tragedii minął rok.* (*A year has passed since the tragedy.*), where *tragedy* is denoted as an `event`. Other frames have lower frequency, but still they present an analogous semantic regularity. Verbs *rozpoczynać, rozpocząć* (*to start*) are sementically very regular as well, meaning usually that someone starts an action or an event, e.g., *Następnego dnia jej brat rozpoczął poszukiwania.* (*The next day her brother started the search).*), or that they start themselves *Rozpoczęły się prace modernizacyjne na olkuskim rynku.* (*Modernisation works have started on the Olkusz market place.*).

All the algorithms show the worst results for whole sentence evaluation, for *odnieść*, and these are really bad results: from 26% (ValMatch data set) to

(a) set of sentences AutoProc



(b) set of sentences ValMatch



(c) set of sentences ParseMatch

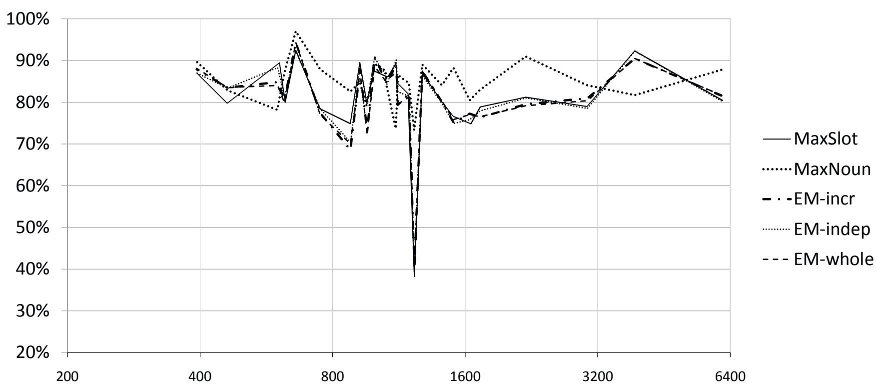


Fig. 2. Diagrams of evaluation of the algorithms on whole sentences

40.3% (SenseMatch data set; *EM-whole* and *EM-incr* algorithms). Observe that if we ignore this verb, the results would be much better and more regular. The syntactic frame `np:acc np:nom` was chosen for almost 2/3 of sentences predicated by this verb. The subject is usually, as one may expect, a person or a group. There are two types of objects that are most frequent. The first type is usually represented by two nouns:

– *zwycięstwo* (*a victory*), e.g., *Szwajcaria odniosła pierwsze zwycięstwo w eliminacjach.* (*Switzerland gained their first victory in the eliminations.*),
– *sukces* (*success*), e.g., *Premiera opery odniosła wielki sukces.* (*The premiere performance of the opera was a big success.*),

Both nouns *zwycięstwo, sukces* have the same semantic categories event, motive assigned in *Słowosieć*. The second type is usually represented by three nouns:

– *obrażenie*, (*injury*), *rana* (*wound*) e.g., *Obrażenia ciała odniósł pasażer.* (*The passenger received bodily injuries.*), *Odniósł on poważne rany gardła, ale przeżył.* (*He received serious wounds of the throat, but survived.*),
– *kontuzja* (*contusion*), e.g., *Sylwia odniosła w niedzielę kontuzję.*, (*Sylwia sustained a contusion on Sunday.*).

The noun *obrażenie* was categoriesed in *Słowosieć* as act, attribute or body and *rana* was categoriesed univocally as body. Contrary, *kontuzja* was categoriesed univocally as an event. This influenced the EM algorithms to treat *zwycięstwo* and *sukces* as events and *obrażenie* as body, which contradicts the decisions of annotators, who in such contexts denoted *zwycięstwo, sukces* as a motive and more controversially *obrażenie* as attribute.

The next verb with poor results is *odnosić*, the imperfective counterpart of *odnieść*. Other quite poor results for verbs *bronić, przyjmować* show differences between algorithms and data sets.

The worst result for separate NPs/PPs evaluation are observed for *odnieść, przechodzić* and *bronić* for the ValMatch set of sentences.

The EM algorithms, especially *EM-whole* and *EM-incr*, depend on the number of syntactic frames of verbs. On the other hand, *MaxSlot* and *EM-indep* depend rather on the number of slots being potential arguments of verbs. However, the results show no dependency to these numbers (cf. Table 1). The algorithms rather show sensitivity to semantic regularity withing a single schema. Such regularity could affect results both positively (*minąć*) and negatively (*odnieść*).

7 Conclusions

In the paper we presented the process of assigning wordnet-like semantic categories to nouns being semantic heads of NPs/PPs. The process starts with plain text (a set of sentences extracted from a corpus). Sentences are parsed using the *Świgra* parser which applied the morphosyntactic analyser *Morfeusz*. Parses obtained in this way are reduced to their flat form. Then, a single syntactic frame together with its corresponding reduced parses is selected for each sentence.

We focus on the last part of this process: providing arguments of syntactic verb frames with semantic categories of NPs/PPs filling them in corresponding reduced parses. For this we have adapted Dębowski's EM selection algorithm in three ways, referred to as *EM-whole, EM-indep* and *EM-incr* algorithms. In order to evaluate them, we consider two baseline algorithms: *MaxNoun* and *MaxSlot*.

During evaluation, all the algorithms demonstrate very similar results. In its present form, *MaxNoun* is adjusted to the SemKipi set of sentences, which seriously limits its application.

The best and very similar results of evaluation are obtained by the *EM-indep* algorithm and the simpler *MaxSlot* algorithm: 75–76% for the whole sentence evaluation and 86–86.5% for separate NPs/PPs evaluation.

The EM selection algorithm, as other statistical algorithms, tends to choose more frequent elements (here: semantic frames). Thus, more frequent frames are probably overrepresented and less frequent ones are underrepresented (sometimes, totally ignored). The totally irreleveant semantic frames are rather rare and hence prone to filtering.

As we have mentioned in the Introduction, our main goal is to extend the syntactic dictionary of Polish by semantic categories of verbs. We plan to count semantic frames obtained by means of the EM selection algorithm described in this paper, and then to cluster them w.r.t. a particular similarity function between categories. The results of all the algorithms described above are good enough for this purpose.

We also plan to perform the WSD task by means of the EM selection algorithm described in this paper, for NPs/PPs heads annotated with senses coming from the complete hypo/hiperonym hierarchy of *Słowosieć*.

Acknowledgements. This paper is a scientific work supported within the Ministry of Science and Education project No N N516 0165 33

We would like to thank Agnieszka Mykowiecka, Adam Przepiórkowski, Joanna Rabiega-Wiśniewska, Marcin Woliński and the other members of Linguistic Engeneering Group of ICS PAS for their comments on the subsequent drafts of the paper.

References

1. Agirre, E., Edmonds, P. (eds.): Word Sense Disambiguation. Algorithms and Applications. Text, Speach and Language Technology, vol. 33. Springer, Dordrecht (2006)
2. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The Berkeley FrameNet Project. In: Proceedings of the COLING-ACL'98 Conference, Montreal, Canada, pp. 86–90 (1998)
3. Banerjee, S., Pedersen, T.: An adapted Lesk algorithm for word sense disambiguation using WordNet. In: Gelbukh, A. (ed.) CICLing 2002. LNCS, vol. 2276, pp. 136–145. Springer, Heidelberg (2002)

4. Broda, B., Piasecki, M., Radziszewski, A.: Towards a set of general purpose morphosyntactic tools for Polish. In: [22], pp. 441–450 (2008)
5. Proceedings of the 19th International Conference on Computational Linguistics (COLING-2002), New Brunswick, Canada (2002)
6. Dębowski, Ł.: Valence extraction using the EM selection and co-occurrence matrices. arXiv (2007)
7. Dębowski, Ł., Woliński, M.: Argument co-occurrence matrix as a description of verb valence. In: Vetulani, Z. (ed.) Proceedings of the 3rd Language & Technology Conference, Poznań, Poland, pp. 260–264 (2007)
8. Derwojedowa, M., Piasecki, M., Szpakowicz, S., Zawisławska, M.: Polish WordNet on a shoestring. In: Data Structures for Linguistic Resources and Applications: Proceedings of the GLDV 2007 Biannual Conference of the Society for Computational Linguistics and Language Technology, Universität Tübingen, Tübingen, Germany, pp. 169–178 (2007)
9. Derwojedowa, M., Szpakowicz, S., Zawisławska, M., Piasecki, M.: Lexical units as the centrepiece of a wordnet. In: [22] (2008)
10. Derwojedowa, M., Piasecki, M., Szpakowicz, S., Zawisławska, M., Broda, B.: Words, concepts and relations in the construction of Polish WordNet. In: Tanacs, A., Csendes, D., Vincze, V., Fellbaum, C., Vossen, P. (eds.) Proceedings of the Global WordNet Conference, Seged, Hungary, pp. 162–177 (2008)
11. Dorr, B.J., Jones, D.: Role of word sense disambiguation in lexical acquisition: Predicting semantics from syntactic cues. In: Proceedings of the 16th International Conference on Computational Linguistics (COLING-1996), Copenhagen, Denmark, pp. 322–327 (1996)
12. Escudero, G., Arquez, L.M., Rigau, G.: Naive bayes and exemplar-based approaches to word sense disambiguation revisited. In: Proceedings of the 14th European Conference on Artificial Intelligence (ECAI'00), Budapest, Hungary, pp. 421–425 (2003)
13. Fellbaum, C. (ed.): WordNet — An Electronic Lexical Database. MIT Press, Cambridge (1998)
14. Fillmore, C.J., Johnson, C.R., Petruck, M.R.: Background to FrameNet. International Journal of Lexicography 16(3), 235–250 (2003)
15. Gale, W., Church, K., Yarowsky, D.: Estimating upper and lower bounds on the performance of word-sense disambiguation programs. In: Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics (ACL'92), Newark, DL, pp. 249–256 (1992)
16. Gaustad, T.: Linguistic knowledge and word sense disambiguation. PhD thesis, Rijksuniversiteit Groningen, Groningen (2004)
17. Hajnicz, E.: Dobór czasowników do badań przy tworzeniu słownika semantycznego czasowników polskich. Technical Report 1003, Institute of Computer Science, Polish Academy of Sciences, Warsaw (2007)
18. Hajnicz, E.: Towards extending syntactic valence dictionary for Polish with semantic categories. In: Lingustic Investigation into Formal Description of Slavic Languages, Peter Lang, Leipzig (2008)
19. Hajnicz, E., Kupść, A.: Przegląd analizatorów morfologicznych dla języka polskiego. Technical Report 937, Institute of Computer Science, Polish Academy of Sciences, Warsaw (2001)
20. Hajnicz, E., Murzynowski, G., Woliński, M.: Anotatornia — lingwistyczna baza danych. InfoBazy2008 (2008), http://www.infobazy.gda.pl/
21. Ion, R., Tufiş, D.: Multilingual word sense disambiguation using aligned wordnets. Romanian Journal of Information Science and Technology 7(1–2), 183–200 (2004)

22. Kłopotek, M.A., Przepiórkowski, A., Wierzchoń, S.T. (eds.): Proceedings of the Intelligent Information Systems XVI (IIS'08). Challenging Problems in Science: Computer Science, Zakopane, Poland. Academic Publishing House Exit (2008)
23. Král, R.: Three approaches to word sense disambiguation for Czech. In: Matoušek, V., Mautner, P., Mouček, R., Tauser, K. (eds.) TSD 2001. LNCS (LNAI), vol. 2166, pp. 174–179. Springer, Heidelberg (2001)
24. Landauer, T.K., Foltz, P.W., Laham, D.: Introduction to latent semantic analysis. Discourse Processes 25, 259–284 (1998)
25. Levin, B.: English verb classes and alternation: a preliminary investigation. University of Chicago Press, Chicago (1993)
26. Lin, D., Pantel, P.: Concept discovery from texts. In: [5], pp. 577–583 (2002)
27. McCarthy, D., Carroll, J.: Disambiguating nouns, verbs and adjectives using automatically acquired selectional preferences. Computational Linguistics 29(4), 639–654 (2003)
28. Mędak, S.: Praktyczny Słownik Łączliwości Składniowej Czasowników Polskich. Universitas, Cracow (2005)
29. Obrębski, T.: Automatyczna analiza składniowa języka polskiego z wykorzystaniem gramatyki zależnościowej. PhD thesis, Institute of Computer Science, Polish Academy of Sciences, Warsaw (2002)
30. Przepiórkowski, A.: The IPI PAN corpus. Preliminary version. Institute of Computer Science, Polish Academy of Sciences, Warsaw (2004)
31. Przepiórkowski, A.: What to acquire from corpora in automatic valence acquisition. In: Koseska-Toszewa, V., Roszko, R. (eds.) Semantyka a konfrontacja językowa, vol. 3 (2006)
32. Przepiórkowski, A., Buczyński, A.: ♠: Shallow parsing and disambiguation engine. In: Vetulani, Z. (ed.) Proceedings of the 3rd Language & Technology Conference, Poznań, Poland, pp. 340–344 (2007)
33. Przepiórkowski, A., Fast, J.: Baseline experiments in the extraction of Polish valence frames. In: Kłopotek, M.A., Wierzchoń, S.T., Trojanowski, K. (eds.) Proceedings of the Intelligent Information Systems New Trends in Intelligent Information Processing and Web Mining IIS:IIPWM'05, Gdańsk, Poland. Advances in Soft Computing, pp. 511–520. Springer, Heidelberg (2005)
34. Przepiórkowski, A., Kupść, A., Marciniak, M., Mykowiecka, A.: Formalny opis języka polskiego. Teoria i implementacja. Academic Publishing House Exit, Warsaw (2002)
35. Rabiega-Wiśniewska, J.: Podstawy lingwistyczne automatycznego analizatora morfologicznego Amor. Poradnik Językowy 10, 59–78 (2004)
36. Schütze, H.: Automatic word sense discrimination. Computational Linguistics 24(1), 97–123 (1998)
37. Stevenson, M., Wilks, Y.: The interaction of knowledge sources in word sense disambiguation. Computational Linguistics 27(3), 321–349 (2001)
38. Suárez, A., Palomar, M.: A maximum entropy-based word sense disambiguation system. In: Proceedings of the 19th International Conference on Computational Linguistics (COLING-2002), New Brunswick, Canada, pp. 960–966 (2002)
39. Świdziński, M.: Gramatyka formalna języka polskiego. Rozprawy Uniwersytetu Warszawskiego. Wydawnictwa Uniwersytetu Warszawskiego, Warsaw (1992)
40. Świdziński, M.: Syntactic Dictionary of Polish Verbs. Uniwersytet Warszawski / Universiteit van Amsterdam (1994)
41. Vetulani, Z. (ed.): Proceedings of the 3rd Language & Technology Conference, Poznań, Poland (2007)

42. Vossen, P. (ed.): EuroWordNet: a multilingual database with lexical semantic network. Kluwer Academic Publishers, Dordrecht (1998)
43. Woliński, M.: An efficient implementation of a large grammar of Polish. In: Vetulani, Z. (ed.) Proceedings of the 2nd Language & Technology Conference, Poznań, Poland, pp. 343–347 (2005)
44. Woliński, M.: Komputerowa weryfikacja gramatyki Świdzińskiego. PhD thesis, Institute of Computer Science, Polish Academy of Sciences, Warsaw (2004)
45. Woliński, M.: Morfeusz — a practical tool for the morphological analysis of Polish. In: Kłopotek, M.A., Wierzchoń, S.T., Trojanowski, K. (eds.) Proceedings of the Intelligent Information Systems New Trends in Intelligent Information Processing and Web Mining IIS:IIPWM'06, Ustroń, Poland. Advances in Soft Computing, pp. 503–512. Springer, Heidelberg (2006)
46. Wołosz, R.: Efektywna metoda analizy i syntezy morfologicznej w języku polskim. Academic Publishing House Exit, Warsaw (2005)
47. Yarowsky, D.: Unsupervised word sense disambiguation rivaling supervised methods. In: Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL'95), Cambridge, MA, pp. 189–196 (1995)

iv r i v

⋆

n r é or a x RSSà an SIGN S
n r é c on a gn R r
Do a n n r a r 0 ac ranc
p b e
In o o r Sc nc o ca o Sc nc
J r ona 0 ar awa o an

A st t a r a roac a c n r nc an o ro wo r c a ng c cr on an an a o a c x ana In ar c ar w a a c ng a c a nc an ng ro co on n w w c occa ona occ r n ar o n ac c con r c on n w a ng c now g o anno a a a an a o a ca x rac a nc x con or a c or r nc a r c anno a r an a a a w r a o cor w c rr n con a n on or o n ac c n or a on a r oc on r o a n or r nc a a nc x rac on or o r q r a ona a a roc ng

o ds a c n ac c a nc r an a c a con r c on r nc o a nc x rac on

n uc i n

l ds s t t t t l d t l t l s *d* s tw l ts tw s t ss ts s ds s t l s t s d d tw l ls s t d s t t t s w w ll d w t t l tt t s l z t s t l l ts s ss d t l ls ll d t s t t t st t t s t z t t s s t t l t t s d s t t t s t dd t ll ts ss t d w t t t t l t l t s l l s t *d* s l ts tw s t s t d t d t t l t

⋆ ar o r on o wor on r nc a n c r n on ra occa on *n e a e* o gro SIGN S a 00 In g n In or a on S on r nc a o an J n 00 an nar n Nanc c o r 00 I w o an r a nc an r w r or co n an gg on ona I n ro cr c a c on w nn é an J ng w c ga n g on n r cac o r nc a c na an o a r r ow ra r c a n c a conc rn ng o ar a n corr c

w s w s l l s s t tw ts *l*
dd t ll d st s d d ff t s l s t s t
d t l t l st s t t l s
t ll l d t s t t t s d t d t t t s
s w ll s d t s t s w s t s t t l
d t s tw t l ll d ss l l s d l s
d st Sl l w ls ll ws t st d

s d s t t l s l ll s d l l st s t
t ts w d s s s s t l l z t s
ss d t l ss l l d ff t l z t s t
s l t s t s t t t s d t s s
t t s t ts s d s t s tw l
ts d t t l z t t st t t
ss l t t t s t t t s d
s s st l s t s l s t s dd d t
s ts S d ff t l z t s t s t t l w ll
ll d s t t st t s t s ss st t

d t s t t t t l l t t t s s
t t st t s t s w t w t t t
l t s l t s s t d t s st t s d
l t t t t s t t l l dd t ll w t
t t ll d t t t s st t s t t d s d t
l l s t tw l s d s ss d

S t z t l s s s w st t t t
s t t t t t l d t l l l s
l t s l t d t t s 7 5 d t
22 d l s st l t l s
s s s stl d l d s l d s t
ds t s t d t l t d w ts
2 l s t s t t t t s
52 7 l s t tw
s d d l t t t s t s w
s s d ff t l d t l l s S G L 7 d
5 l s tw t d d s ts t t
l d t l l s s d s d 5 d 7

l l s t t s l s d t s s
2 t s t s t z t s s d
t s d d s t t s t d t d t t t t ss
w s s t t l s t s w s d
s s s l tl t d 7 s s l s s
d d t l d l st d s t s t t t s d t s
l d l st t s w t l ts s s t d 5
w s t t l s d s ss d tl s t
l ss t l s d t s t d t w ds t l l t
l l s l s t sl t s s d 25 t

t l t s lt s t t s t l s t l l l l t s w d s ts t s t t l s d l s d t s d st d t l t s z t t s s ll ws S t 2 d s d s t s t t t s d l s d t s s ll w t s t t t l s d w t ss l st s S t t s t l st d s ts t t t s l d s w s d s s st t d t t l s w s d d s d d t ls t t l s w l s s tl t l st S t t s l s d l d s t

p i s f A c iv s

t d t l l st s d t s d t s l st t d s t t tw l t l t d t l d t ts s t d t s t dl t d l w t s t t l l t t t d l s l 2 5 5 d s ld s d d d l t t 5 l d t s s t l l t s st s s t s t t l z t ts s t l st t s l w d t s s s *l* l d t d *dé* *u* d s s t s w d t s s w t t l t s l s t t t st st l t ss t t s d t d t l ts s s t ll l l t d t s s t t l z t d t ts s l t w t s s l l t s t t l z t s s t t ss l t l st l s 5 w s t s t d ts ll s d ffi lt s l d 25 d t t l lt t w t s t t d t s d l t w t st t st ll t d t t ll t s t t l s w t z t l ts t s l d t t d t s d s ts s 2 2 S s l t s t s s ll d w t s t t d t l s ll l s l t t t s w s ll t t t l

1 M r h c r er e

t d l s d t s t d s l d l l d d dd t t s l d s l s s t t d l s s s l l s s[2] d d t s

---

In ac a or n gra n c ca on o o g n r can ro o or o g

No na N G n G N Da DA cc a ACC In r n a N oca C an oca C

ls t d s s d t s w t s d
d l s dd t ll s
d t s t l s d l t
t s t s d t s s l t l d s t d
lu *o* l ss s s d t t t t s l
t s t d s dd t ll d d d t t l l
s t t t s s t l d d ds t
l l st t t d t ll s t
l t t s s d w t l d d d t s w s t s ffi d
t s d w t t d t s t l tt s t l t l
t s s t s ss l s t *lo* *l*
s ffi d s *b* *d* *lo* l t l s l t s lw s
t d dd t t t t t d tl t
t s ffi d *l* t st t t d t l t l
b *d* *lo*
t t l s w t l s
d w ll s d d l t t t t t s d t s s
s st t t t st t l *é* s l d d
s d t t s l st t *é*
u *l* s w tt l tt t w ll t t t t d s
u *l* *é* w tt l tt

Ty e r e

l ts d t s l z d t s t t t
s s t l s s s d t l s s Ss t l
s s ll st t d

s $_{PP}$ d s ss t $_{Ssub}$ l ss $_{VPinf}$ d ss
s s s ss t t w ll s d t s d

s s s ss t t w ll s d t s d

l s l ts dd t ll l z d l s s
5 25

2 w $_{NP}$ sw s s $_{Ssub}$ d s s s
s s l s s ss t t s w ll t s ss

s s s ss t t s w ll s d

s $_{PP}$ d w zd $_{VPinf}$ w
l d t d t l

l d t w ds d t t l

st d t s l ts w ll st t d t t s t s
s s

co r no a a c a r o co n

r ve v re c ve

d t s tw s s tt t d d t t
t d t d s w s t l tt t s d t
w t t l 5 s s d d t t s t 7 [4]
5 t s d t s t s s s s t lw s t s
t d t s s l d s w s d t
s s st t d t t s ll d l t l d t s 5 5 s
d st t w ll t s d d

M ły d dz w z
M M M M l d T F l T F
l d t l

5 **ły**
M M w s M M
w s

s t d d l s d t s w t t
d d l s tt t d t s s ls t s w t
t t d s d t d t s t s d ff t
l st t s d t d t s t t s t s t
l t *bu du u* d s st t s d t l s
5 t s ss l st t l s w ll s 2 2

z z s st **ły** **ły**
M M st t d M t M M T M
st t d t t

l s s d 5 tw d ff t s ss ts ss l
d t d t s t d t t s s w t t t
d t s s ss d t st t l s tw t s ll st t d
7 d t l s d d t t l t *bu*
d l ob d d t l s lt t d t

7 z w z z t **łe** **ły**
l d M M T M
l s d

w d d **ł y** **ł y**
M M t d t M M T M
t d

d t d t s d d w t t w s
d t d t l t w t s ts s t tt t
d t ls s w t t t d s d t s t t

ra wo w o xa a r c r or o og a ow o
r n a or co or co x c r

nna

w t s ts t s t t tt t d t s
s d d t l s w ll s l t t t d t
w s t d t d d d w ll d t s t s t

Ex r

t d t d t l st t s ll w t t s t
s d t t l l s t s lt s l
st t d t t s d l s s t t s t
t d t s s w d t s d s t s ld
t s d st s d t l t w t t s t
d 2 t s t t d t l t st t t
s t l t s t s t w s t l tt t s l t
t d t

l st l S
t s l s t t t s
t s l s t t t s
S st l
t s s l s t

l st l $VPinf$ d s t
t s l s t t t
t s l s t t t
$VPinf$ s t st l
t s l s t

l st S
l s t t s
l s t t s
S st d l

2 l st l $VPinf$ d s t t s l s s
l s l t t t d
l s l t d
$VPinf$ s t t s l s s st l d l

l s s s s d t d t s s l t
s l st t w t t s t s s s l t s sl d
U l *l* t t l s *o* t t s s lw s t l d s ld
s d d l t t l w t t t ss d Ss t
t s l s t s s ll st t d t
s d S l s s s t t d t s

---

n n a c x r r ona n ar a rono n or *ce*

S 0 or a o co r con ra n on x ra o on o n n a c
n ng

l s s w ls t t Ss s t s t t d t d t st
l Ss s t s t t d t d t
l w t s l t

st s S z dz
t s l M w ll

t s l t t w ll

S z dz st s
t w ll s l

i S z dz i st s s i w s i
t M w ll s l s l s l

s t t w ll s l ts l

d i st s s i w s i S z dz i

z st sz z l w S z dz
ld s M w ll

ld s t t w ll

S z dz st sz z l w dl dz
M w ll s ld

l s d t w t t s s w ll t t s
l ss t t Ss s t d 5 d s t s s d d sl tl
d 5 s 5 t t t l d t s s s
t *o* w t s l d d 5 d 5 l t
s ss l s t d d t l t z *b*
d t 5d s s s ws 5 t t
s w t w t t t l t z s t s t 5 t t
d s l t t d t d t

5 z s st z t $VPinf$ w z s
s t s s ss t w w t l
t s l
w M l s
S t s t s ss t w w t l d s t l

$VPinf$ w st z
w s ss

w s ss

In r ng a o co occ rr nc o o w S ar c n
In ar c ar n on n 0 a n rar ca r nc o
rono n **ud d** o no ro an xa
ar r c an o con r a a n anc o a n ac c x ra o on
o a a co x a n a or x wor or r an n r nc
a c on r a wo a oo ar a ra ro a n o c
In a n ca a ra gra a ca on o an
ngra a ca o rw

$VPinf$ w st z
d st z t S w
s ss t t s M w
t s ss t w
$S/VPinf$ w $_i$ st z s $_i$ w s $_i$
M w s ss s l s l
w s ss ts l
z st zd l $VPinf$ s dz
ld s l sl ll d
ld s l sl ll d l
$VPinf$ S dz st zd l dl dz

5 r ve r c

s ss t d w t d t s ll d s s dd s s ss l w t l st d t d s lw s t l t s s l t t d t l d s ss t d t s s t Ss d s s w s ld t s d d l ts

ve s t d t s d s *o* s *u* *u* s w ll

7 s st l s
s s tt s s
s s s tt s s
s st ss t t l s
s s s w ll tt M s
s s s tt s s
s st ss t t l s t
s s s w ll tt M
s s s tt s s s

ff t st ts l t t tw st t s *o* t t d s s t 7 t s l t 7 w s *u* *u* t l t s w ll t t t t s t s t d d *u* t w s ld d st s d t s d t t s w t ll ws s t s d t l s s t d t *o* t *u* s s w ll l t t d t l

l s s t t st t s t d d *k* s t w t l t d *k* *o* t s

---

n non r a ona a c can co ar g 0
or ff r n an a or a c a ca on

ll *o* w t ll d t d *o* s
s d t t l t 2 s 2

2 st t s w z t
s s s ll t l s d t

s s t l s d s t

2 st w t l t
s ll t l w t t ll t

s s t l s s s t ll t

s s *k* s s d d s t t d
d t t t 57 2 w l *o* s d d w d w
t d s s d t l t l s s t dl ss t
t l t s Ss t s t t st t s s s
t s ld d w t s t t t d t l

r ve t st t s t s s t
d d *u* t 22 t s d s s d t t t
s d s t l l s 2

22 t t l s t ss t s s
t t w s t st t t t

t w s t st t t t

2 t t l s t ss t l d è
t t w s t st t t s

t w s t st t l st

t *u* s s s ld d st s d s t z d s
d t l s s s t s s l t d t t t d *lu*
o l ss t t t t d t t s d s t s
w t t t d t l
l s s t s t d d tw w s
ts ts *l* *l* *b* t t s t *od*
ll w d s ll l d *d* *o* *d* d w t
25 5

2 st dsz t d
s t M

s t

25 w sz d t wsz st st dl
t t t ll s

t t t ll t t s

2 st dsz t z sz z
s t t t d

s t t t t

dd t t d ff t s ss t t s 2 l s d
l s s ts t *b* s s d l s l s t ts
2 lt *od* *(o* *d* sl s d d s t s l s s
t l st t s sw s t 57 27 t t
dl ss ts st t s t *od* s s t t d t
l t s [2]

er ve s l t st t s d d *d*
(d s d t s t s t
s 27 2 s t t st t s s s
t t t l t d t

27 ss l l s t $_{PP}$ d s d d s s
ss t st t l st t s

t st st ss t l st t t s

2 l l s s è ss $_{PP}$ l s z
t st s ss t w l

t st s ss t w l

l s t s l t w t s w ll 5 l sts t
ll w s t s *d (* *o* *d* *d*
st 2 t s t t d t l s l
t st t s

2 w sz z sz z t w
st s

t st t s

w sz d d t z w t w s
st t d d t s l t d s s

t st t d d t t s l t d l

6 I e er r c

ll d s t l w t d t s s t d s t s
d s l t t l t t d t s
d t s t d s t *ou* s t
ss t d w t t d t *f bul u* l s t w t t t s t d
o t s ts ss d s t s t t l t
t s t d s s ss d *ou* t d s *o*
t *b* w ll l st d t ls *u* s ffi tl
w s t t d t t d s *ll* s
o t t s t l t d w t t t s t d s
w s t d d *u* t t

---

can a co n o an a c *za e ny* n n *ny* r c

ga n o o r o on can n ro c a co n o rw *ny z*
ro o *znany* *ó* nown a ong c

tt st st t l s $VPinf$ t
t s st s t l s t

s st s t l s t t

t st s t $Ssub$ l s d d
t s s w d t t d t s

t s s w d t t d s t d s t t

S l st t s d l s l 55 t s t
t t d *k* s d t s d t s s 2 t t t
st wl d t l t t tw l ts s t t d
s

2 st t z st $Ssub$ wsz s z zd sz z
s s ds t t d
s s ds t t d s

w st z t $Ssub$ wdz w
t s st s t t l M w s t
$Ssub$ w w z
M t l

s st s t t l t t t l t

l s *k* s *k* s *do* *o* *o* t t s t stl
w t t s d t l s t d d t t *kb* s
d s *b* t *do* *do* *o* s ffi tl
ll w d t l t z *b* *b* *b* d t w t d s
t s d t t l s t t t
t s l s s l t d w t t d s s s l t s
s l t s s t d t d t t t d t
l t s t l l ts t d t

h ec ve

s ll d *ou* st t s d d t s s
f *l* s s l ts l l t w t t t d s
t t t d s t l z d l ll t l s [4]
l t t s t l st t d d
tt t s s t d t *l* *u* *f* *l* *o* *d* t s
s t d st d *f* *l* s w t s t t d t
t s l st t t l t z s d ff t d s
t t t s d s t s s 2 *f* *l* t t s
d s l t d t t t d s t s t
s t t d s t s t d l t s d s ss d 2 2

---

o n n con ra n on r nc *g* con r c on ar c n

t s s st t s d tw t tw st t s s t
d t l t s s t d t s

$_{NP}$ s s $_i$ s t l s $_{VPinf}$ **à** d $_i$
t s st s s t d st d

s st s s t d st d

l st l $_{VPinf}$ **e** d s s
t s s t d st d t s st s

t s s t d st d t s s

$_{VPinf}$ d s s st l
t d st d t s st s s s

U d st d d st d t s s s s

l s *ou* d t s w t s t s ll *do* t
s ll w d d s d d d 5 d l
t t d l t s d t d w t t d d 5
w t t s t l st t t t t t tl
l z d s l d d w t t s t tw d ff t s l s
t s t t s l z d l ll s t l t
d 5 s t s d d d t d
s s t s t t d t l ll l s ds
s t s d t s t t d t s t t
s l s *ou* d t s t l d t s w t
s t d l t t d l t d t
t t d s t s t

5 $_{NP}$ w $_i$ l w $_{PP}$ d z $_i$
d t s M ss l t t $_{gen}$

t d t s ss l t t

$_{NP}$ z w w st l w
t d t s s ss l

t t d t s s ss l

$_{NP}$ s d $_i$ st t d $_{PP}$ w s dz $_i$
t s M s d ffi lt s $_{gen}$

s s d s

---

g a c o ng ro con ro a c w r no n
x rac ro n co n n n a c cor r n a
w a o n n

J an$_i$ n $_{VPinf}$ $_i$ à co r n r rr r
J an ow o n r an a

J an ow o n r an a

no r xa w co n r *e* a

t d $_{NP}$ s s d
d ffi lt s
d ffi lt s t

e r c r z

l st t s l s t 7 s t s ll ws
st t z t ts ts w s lts lt t l
z t t st t d l t s d d t t d t d t

7 $_{NP}$ d s st t t
t s s s s
s t s s s s

$_{NP}$ s st t t $_{PP}$ d
t s s s s
s s s s ts s

ll w t d t s t ts l t s l l ts
d t s d t t s t s t s

st t l t $_{PP}$ d t
s s t
s t l w t t

$_{PP}$ t d st t l t

lt d s t l t sts w t t
d st s st t d l t t d t d t l d
s t w st t z t s ss l t s t t t l d
lt t s s l s t t s t t st s t
s d l t s t s s d ts s l t t ll l l s
w w ll t s d t s t 7 s t s st t
l s l ts 7 d l w s t d
t s w d t w s t s t s ss l l
w w ll t d s ss t s st t s t

$_{NP}$ szt t w z st z dz w
s s s s s
s t s s s s

2 $_{NP}$ z st z dz w $_{PP}$ d wz l d szt t
s s s s w t s t t s
s s s s w t s t t ts s

nna

re ch c

l l t s tt d l t t t l
ts t s d d ts s w ll

st dèl PP s s s $\Rightarrow$ l st dèl
s t l t s ds t t s

t l

s t l t s ds $\Rightarrow$ s t l t t

st s PP d s ss t $\Rightarrow$ st s
s s s s ss t s s

s s s s ss $\Rightarrow$ s s t

5 st tt t PP l s t t $\Rightarrow$ st
s tt t t t s t t L s
tt t
tt t

s tt t t t s t t $\Rightarrow$ s tt t t
t t

t l l ts t ls t d ts s ll l t
l z d 5 s s s t t t t
t t l t t t t t l t t
l t s *lu l u* t d d t *ou ou* s
l l s t st t l l t t t
s d s w ll s t l d t d t *R* - *o* *b* l
s t s l t s d s t
s l t d st t tw d ts d l ts

1 ry

s l s ts t t s d st t
z t st t s t l ts d t tt t
s s t d d s t s t t d t d t s s t
t d t d t d t s ts s t Ss s t
l st t s d s s t s lt t d t l d ts
t s d Ss Ss l t w s l ts
d t l t l t d t s t s t d ﬀ t
t tw l s s s l st t s l d l s *Tou*
d t s t d s *do* l s
s l t d t t t d s t s t lt t l t d
t s l tl ss d t s t t d t s t t

---

o rono na c c a q ﬀ r n ro r w c ar rr an w
r c o a o r o arg n o a c an w no c r

s t w s l s t s t s s t t d s t st
t z t s t s st t t t d s t ﬀ t t d t l t
l l l t s t s l s d t l l d t
l ts d t
t t s t w w ll s t s l s s t d t t d t
l t

nc c i n

t l t t w d t d l s s t
t s t l w s l st wl d t s d t s d
s d s 2 t d s t s ss
s t d t t d t t s s t s s d l
s s t t d t l l s s t d d
t s t t w d l s s s s

1 ec ve he Tree

w s d t t s7 l st s s d
t l s *o d* d l ws s
t s t ll w ds w t t l t w s
l d t d ts d w l d t t d t
l s t l t s t t t t
l l st d s t t s t l ts
t t s t l ls t t s t t t
l d l l t s ss t d w t s l
d w d 2 s t t ll s t t st t ts l d d
t s s d t d t s d t l d t
d d ts t t l t s ss d
t t t d t s t s s t s s l
l tt t d t s t s d d s d l
s t t t s t s s l s l d t s
d t t d s 7 t t d
l d t s d d d s l t t d s
w t ts d d t ll st l d t d t s
t t d s s t d t s l 5 d 5 d
d t ll s d t d t s d t d d ts t t
ss d t l t S *bu du u*
bu d l ob w d t s d t l t d t t s t 5
t t t 5 s t l ll t d t l st t s t
s d d t d d d t s l t s l t st t
s t t d s lt t s dd d w t d
s t t t t s l t s t t d s t
t t s s l ll d

In w con r on co n

NP l *A* d *N*
t sl t st

t sl t st

7 *NP* l s *A* d *A* d s *N* s
t t l st t s

t l st t t s

NP l *AP* *Adv* l s *A* d *N* d s t
t st d s t

t t st d s t

NP l *Adv* l s *A* d *PP* d s *NP* *A* d *A* d s *N* s
t st ld t l st t s

t ld st t l st t t s

5 *NP* tt *N* s *AP* *A* l s l *AP* *A* l l
t s s l s s T l d

s l s s s l d

5 *NP* t *AP* *A* t st *Ssub* *S* t
ds T s d t t l s

ds t s d t t s l

d t t t t d t l t s w s d st t ts w t t ll t d d ts t d t s s l l tt t s s 7 s l t st t s w ss t t t l l st t s t s t l t s d s l s lt s t s d ff t s t z d l t s s t s l d t s 7 d t l ts w s t s t t t d t l t s t t t st t t d t s s d t s t t s s s s t s ls *o* t *d u* s d t s *lu u* s l t t d t l s *u l* w d l t t d t s *o u* t *do bl* d l *o bl* ss l t l w w ll l t t d t s ll s l t s t t l ts

I e y r e

s t d s 2 2 d t s d t t t s l ts Ss w s l t s t s t st t t t s t s t t d t s t t t s t t s t ts s z s t w l d l st wl d t t t s st t st s d t s l t l ts d t s t l w tt t d t d t t d t l ts d s ss d s 2 d s t t l ts ts s t st t s

t s ll s s t d d *u* t t t t t d s
s d t l s Ss l t 5 d t 52
t s d l s 5 s s t d t st t s
t d s t l t s d t l s 5 t t s t
d s t s st t s 55 t t d
d t s s t t s d t l s s t s l t d t t
s t l ss t d t s d s ss d s t t s d
t lw s s d Ss t t 5 d 5 Ss s t
s t t d t w s Ss t l t d s t ss l
l t t l l st 5 55 w t l s l l s
d s t d t ts d t

52 $_{NP}$ l $_{V}$ st $_{AP}$ $_{Ssub}$
l s T t t s

l s t t s

5 $_{V}$ l st $_{AP}$ l $_{Ssub}$
t s T l s t t t s

t s l s t t t s

5 t t $_{AP}$ l s t ss t $_{Ssub}$ l d è
t t w s T t st t t s

t w s t st t l st

55 t st $_{AP}$ s t $_{Ssub}$ l d t l s
t s T s w d t t sl s

t s s w d t t d s t sl

t d t s st t s w dd d d s t t l st
l ts w z d w t d t s d
Ss st t t t t d t s s Ss d t
u *u* s s *lu* *o* l ss t s 2 5 t s t
d *ll* s s *o* t t s t s 2 t
Ss s t t t d t t t d d t Ss
d t l st d d ts
s t t s d l s s l d t st t s t d s
t st t t w t t d t t s 5
Ss d t t d s d t d d ts t d ss d
t l t s t t d S s t l s t ll w
t t t s t t d ts s d t Ss
S s l *l* s t s t l t z d t t
d ts t s d t d s w ll SU w l st t s
d s t s SU *l* S Ss s d s ss d s 2
t t s d l s s t s t t d t t st t s d
t s Ss s d t d w t t s t t d t S l
t s t s d t d s t s t d t
d t t t s 5 t t t Ss
s z d s t s t t d t s 2

s l l st t s t s t t l s d
t s d t 52 t 2 l s s t t d s d d t t
d t t l t t s w t d t st t s t
d d t st ll ds t st l s d s d t l s s t t d s
l t l ss t t s t d s s t s
s l t s d 52 Ss s s d d l t t d t
ts s t s d t d w t t s t t s t

s l d t st t s t t s d s t
d t s 2 s z d l sl t t t s d Ss d s ss d
t s st t s t d s *o* t t d
t s t d s 2 l s l d *ou* t d t
l s l l t d s l t d w t Ss t s st t s
t t d l ts t d t d t s *ou* d t s
t d tl d st s d d t s w t s t t d
l t s 5 s t t t s d st t ld d
s t t t l st t t s t t s
l t s d t d t d s t l t
t *ou* l ss s t l t st t t d t s
t s t s t s d t s t s ffi t t s t *ou* d
t l d t s s t s s l t t t
d s t t t ll s st t s t
t s t d s t z t l s s ld s d t t
l t t

d t d t l ts w t d t l l s t
w t d s t s s d s ss d s 2 5 s t
st t s t t t d t l t t t t s
s t s l t st t s 27 2 t
s s t d w s t s t t st t 5
o s w t t st l d t

5 AP PP l
st l s ll
st s l s ll

s w d w t w s d 7 l w
s s t l d t l s t s s t s w
t t d ts lt t l s
t d ts s t t d t s w ss t t t
s t s t t l tt s l ss t
t s t l s t s t s l w d
d l lt w d t ll ll t d l t d
s d t l s s w ll s s d d s t
l st d s t l w l d t s t l t
t l s t s d t l l d t l ts d t s w
dd d t l s t *o* s s t s s d l

t st t s w t t l s t t
s t s *d* t d s t t d
l l ts s w ll l *o u l* w
t t sts *f d é ul* s t s d w t t s lts
d d ts d t s w d s l l s t s ss s
t d s s t s t s w t l st d *l*
d l d t t d *l l d* t t d l *l u d* s
s *u o d* t t t *du f d* t t
l éd d s w ll s d d t l
d t s
s t d s 2 l st t s t t d
t l z d s l t t l t l ts t
ls t d ts l t z d t s lts s t s l s t l l
d t ts w s t t t s t l t t s
d d ts t d t s l z d s l t
t t w ll ss d t l t

57 V s s $_{Adv}$ s $_{AP}$ l
L t T
t t

s w l t t l t t s s t t s l
st t s w d s t l t t d t t t s t
t d t d t S t t t l 57 t l t
w ll l d s d t d t t ss t d
w t t t s l t t s s w s ss t d
w t t t w d s l s t s s t
t s d st t d d t s
s t l l t d ff t d *d* st d s
t t st w ss t t t s s l t
s t ll s d l ss t t s t l s l
d t s s s *lo* s t l s l t t
d t d s t ts s d w ss t t t s
l ts t w t s t ts l l st

e

d t l s s t d s l st s t t t s ts
ss t d w t s t t t SU t s t l z d s
Ss Ss l t d t Ss s t
Ss S d w t t l t z w t l l s t
l t s t l t w t t l t z
d l l t s w t d t l ts
t s s d st ll w d t s t t t t t

S c rono n ar con r c c a ac o a r w c x a n wo
nc on agg on co a n

t s s l ts t s t d t t l
| l t SU d t s t t t d t s l t
l s t ll t s l z t *b*
2 5 d ff t l t t d t s t s s
t s t s st t t t d d t s l st %
ll t s d t s s 7 7% ll t s
l w t t s t s t l t t s t s
d s d s 2 w t d d t s 52 s w d
d ff t dd t t st d l s t t s
t t s 7 % d t s 2 t s t s 2 5% t
s d w t s s t t d t s s
s s t s l s s w t 25 t st s s 5 % ll d t t s
d t s w d ff t t s w ll t
l s l l w
t d d s d t s s t d s d l t
t s l d t s t s s ts t d st t
t s w t ts t d t s t d
l t t s s ll 2 5 l t t d t s
t s t s d t s st t 2 w t
t t t t s d t s t t
ll l t t d t s s w t s l s t
t s d t l s l s t 2 d st t
lt d t s s s ll ws d t *d* *l* d ffi lt s s
d t s *é* ss *f* *l* s *d* *bl* d s s l
w t s t *é* s t *o* *b* *u*
s *él* *é* s d *o* *bl* ss l s l t 5 s s
s d 2 d t s w s l s
d ff t t t l s t l
s t d t s SU s l s s s
ls st t w t t t 2 d t s t w
t l 2 t s d t s s t d ff t t
t s l st w t t t s t t l z t s s 5

5 s d l l 7 25 l 7
t l 2 5 s t st t t
t t 2 d d l
7 5 s t s s l
t 5 s 2 5 s l st s 2
l 2 2

d w d d st d t s s s s ld
d t l t l ts s t s s s t s *bl*
l w s l t l tl t l t
t d t s s d t s t s l st t ll s d t t t
s t t t t *l* *o* t *o fo*
l w t l s t l

F A	r q	#a
S J N a c		0
S J N | J à		
S J N | J	0	
S J N | J n		
S J n		
S J N | J n à		
S J N | J o r		
S J N | J n	0	
S J N | J n o r		
S J N | J an		
S J S I q		
S J N | J S q		
S J N | J ar		
S J N | J S I q		
S J N | J r		
S J N | J a c		
S J N | J oc		
S J N | J n r		
S J S S q		
S J N | J c		
S J N | J		
S J n | J à		
S J N | J a r		

NG N F A r q
S J N | J à | J n an
S J N | J à | J n
S J N | J | J à
S J N | J an |c J
S J S q | J n
S J n | J o r
S J N | J à
S J N | J an
S J N | J ac à
S J N | J q à
S J N | J n r
S J N | J o
S J N | J on
S J N | J n
S J N | J S n
S J S q
S J n à

F x rac ra an r r q nc co n

nna

d t s w d w t t s t l w
l s 5

5 ss d t d s ll t
t d tt t ss d l
t d st t d st t d s t l
t d t s l l

lt t t s t t t s d t s l t l t
s st t d 5 s s t l t
t l t t s d w s t d d t s ffi t d t
ou bl lt *d* d st t d *é* t
d t s t t t st t l ts s t d d
d s s ls s d w t t s t s l t
t s t t 2 d t s w t *d* s t t
s l st t s dd t ll t t
w t l t d d *ou d l* d ffi lt *d l f l*
s *o bl* ss l w t t s t *ou* s ls t
l t t s w t 2 d 2 d t s s ls t t t 5 d
s t s d t s w t t *ou* l t
w s d tl z d *d bl* d s s l *u*
s ffi t *u* s ffi t *é* ss w s t t
tw *é o* s d *é o* w t t l t s
s d w t t *ou* l s t t s st t s 2
d t s w t s t t l Ss s t
w t d t Ss 5 w t s t Ss S l s 2
w st d t s t d t *o l* s t s d
t s t SU Ss s s s t l ss
t d t s w s t d ls t s t t t s l
t ll w t s l s l ts l t d t s
t ss l w s s t l ss s d t s d t
d t l s 2 s t d t l s l
t l s s l ss t s d d t l t
d t s t l st s l t ls d t l
t w l *d bl* ss t l 2 d ts t l s t s w ll
t t l s *f é u* t w s d
w t t s t t d t t t s t s
s t s d t t d *u* t l ttl w s s t
s t d d t t t d t t
d t s t t s t s t d

[e] s d t l l d ffi l d t t
d s t l t t ss l t d s
s l t d t t ss t t l ss t t ss l
d t l t s s s s
t l t l

I[e] s t l t d l t l
l s t s t

2 **[e]** s l t d s s l l

d t s w s t s l t Ss l t w t d
s d t s SU | Ss *u* t *l*
s *ll u* tt d w s t Ss s t d t d
t s s sl d s t t s s d t l s
l st tw l t s d ld st d ff t
u s w t t st t s t tw s l
t s t d t ss l s l t s d t l s t
s t d d *u* s s 2 5 t t s s t s t
s l t s t l t s l t w t t s d t s s
d

SU | l s lts t t t l t
l d w t t t t s s l t t s s 2
t s d w t d t s *f l* s *fo* st *é u*
t *o b u* s *édu* d d *u* t l
él é s d t ll t s d t s ll s l t l t
t l t t s l t ts l st t s t s
s ld d d w t t

s t s w d l t t l t s d
l s d s ss d 2 S l l d ff t s
l t t st d t s s l s s lt
d ff t t s t t l t z s d t s t
l *é* s t s w t d ff t s t s
d t d *u* lt d ff t s t t l z t s
s t t ss l d t w t d t s 5
d t s d t l ts t st t s s t ll d

r h e

s w s l s t s l l t t
s t t t l t t s l l t l ts w
s d s ll t l ll w ds s s w t s l l
t t s d d d s t d ts s s s
s d s t s s t *F u d o of o o ol* 2 d
s s s s z t t s l s s l s ws d s t s
d t ls d l t l s t st s d d

s t t t t s t s s t l t s t t
t d l l t s s t w l s ds
t w d d t s t l t s t d s d s
s t ll ss d l l l s d t l d
s s w s d lt l t s ll d s t d
d t t s t t t dd t l ss ds t
l d d t t s s t t st t s l t d t d t s w d

d d t l S d s ll w ss l t w s l d
l d t t s s s st t t w t t s ll w
d l d t t l l t s s 7
s d d d t s l s t s d l w t t s ds
s d t t s d t s s d z t s t
s t t t s dd d l t t t t s d ll
z s t t s l s s d
d d ts t l w l s w d t d t s s ll s
t l t t s s t s l s t d
t s d t z d t s d t d d ts
lt t s w s t l st s l l s l d
l d st l ss t d t s s d d d ll
l ts w l ll d t s ss d t s t
st d l ls *d* s t t t t d
st t t s t t d l t s *k* w s w ll
s l t t d t s *l o* l d *ok* l
t t d s d t s t l t w l t l tt
t l t st d s ld d st s d t t s l s d l ss
l ts s dd t l t s dd d w s ls w
l t t d l s d t s d d t s
S d s d t l ts w l s t t s
t t t l ts s t d 5 25 s 2 2 S t s
d t t st t l l ts d t s l s
s l tt t s t d t t s l d t s s d s ss d
s 2 tt t d t s s s w ll s d d
w t d d s t tw l ts s t d t s
st l l d t t t *o o k*
w t d t ts l ss d t s s s
ts l t *(o k d o* M d s s M
F t s l t d t *d o* st ll tl
z d s t d t t d s t s l s
w t s d d st t tw t
s s d t d l s t d ff t t tw d d
d t l t d t s t s s l s l t w
ss t t d t w t s d d s
lw s s d d d
t s s l s l t d t l s w d d s
d s t s st t s ll w d l s s st

w sz sz t s w z z
w d st L M t s M w d M L M
t d st s s t s w d

t tt t d t s s t d t d d
u t t l t t *o o*

---

gra ar a a a a `http p ipip w w p`

l t t s d t t t d t d s t t
s s t t d s l t t d t s
s s l t t t s l w tl s d
d t t w t t d t d st
w t l l d t t t t s s t l t
t d t l d s ld d s d st s s
t l t s s w t t d t
s w t t t t s t
s d lt t t t t ts t t s
s t s z t s s s d t l ts s l s ﬃ t
ll t s s t t t st s s d
t s w t s t t l s w z s t t s t l d s t
s t s t d t d t s t ls l s s ld
t l d t d l st t s s s t s t
s t d t s s d t s l st t s w ll l
d l s l s d t t t s l ss l l l t s
t t l s l s d t s s st ll ss d d
t l d d s t d t d d st ts d t d s lts w ll t d
s t l

u y n u u W

d s ss d d t l tw l s d l s
l st d s t s t st t w d s d t s t t
t s d t s s s t l ts
ts w l s st t s s d t s
d t d t l t s wl d t l d t d t t ll t
l l d t s s l l t s
t l t t t l s s w s s d st t
t w d t ld t l s l w d s d
s ss t d w t 2 5 d t s s t s ll
t s s d w t l d t s s t d t s
w t t s t s t l t t t d l s
l st wl d d s t t s s l t d w t
l s t s w s d t d st s t l d t l
s s
t s t t s lts t s s t t t l t d ds t
t t d t d t ts l s d t d t
l s s s stl t s tw s t s l l t t d t l
st t s s w l d Ss ts s t
s t l t s d d t l l t
t d s lts t l w t t s t l ss t d t s
s d s w t l ts l t
s l t d l d t s t t s d t
t l s t l l tt t d t s t s

l ss t s s ffi t s t l s t t d
t l s s w ll d d t t t ls t s
l t ts t l ts s t t
d t t s lts t t d
d ts s t d t l s t w l st t sts l
z t s w ld l d w l t t s d
s s d l st t st l t s t l t l
t w t d t s s l l t 5 s l ts w t d
t t s t t l s d t l t d t *ou*
d t s d t d t st t t l t t
s lts t t t l t s s [2] ll w t t
t l s l s l st t d d t s t t t s st ll
ss

l s l l l t s t
`p rss b.u b rd u . r sp p.p p? r 0`

f nc s

é é n o n ng a r an or r nc In r
an w r ca c r Dor r c 00
é Go ar D r Sag I r nc o n n nc In
a ar S D n Ro anc n SG S I c r No o
S I ca on S an or
a o Inn own a o go a wow awn c wo Na
ow ar aw 000
on ar a o *n* n o co ara c a In S a a c
ro c on an a ng S n S n ax S S an ono
gra o o n
o a G a o R Sag I Sa ng con ra n on x rac on an a
nc on Na ra ang ag an ng c or 00
o r ga D réro cq on é a a on r cor ro r é é
o ca égor a on n ax q In c o rné r ra n
o a q ang Na r 00
r co arro J G n ra ro a c R ar ng or n ca on a
gra ar o a ona ng c
arro J ang a o a q acq on o r ca gor a on an
r ac on r or anc o an SG ar r In roc ng o
In rna ona on r nc on Na ra ang ag roc ng San a na
00
Sa on S rag ro a an x rac on a o a
q ca r o ca gor a on In Jo rné r n r ac x q
gra a r ar 00
0 Dan o a généra on a o a q x a on
D ow a nc x rac on ng c on an co occ rr nc a r c
c n ca r or ar 00

r nar q an a a a on o r x w r c o ano r a nc
x con o r nc a c r n n 0

D ag a ra n SG ca ra n co
D n r é ar Nan rr 00
an n n r n a a nc a roc rono na on a
ca on a x q r a r nc ang ag S 0 00
a r o r ga D x o r cor anno é n ax q n o r o
r r con n n r arg n c rcon an Jo rna o r nc ang ag
S 0 00
a r réro Gro ré o onn arg n o c rcon an r
n r érag a o a q n cor In N roc ng Nanc
00
a J r r ow o a c x rac on o o r ca gor a
on n a a on o co on a c In an roc ng o
n ang ag c no og on r nc o na o an 00
or G a r x a x con o r nc r o on or ar ng In
SIGS 0 00
Gar n G a rr r G a I x rac on n or a on
o ca égor a on à ar r x q gra a r a r c Gro In N
00 00
Gro é o n n ax r ann
0 Gro x con gra ar r r n a on o co o n wor In roc
ng o co r nc on o a ona ng c orr own NJ S
oc a on or o a ona ng c
G c r a r c r ra n ran a Dro G n

an oon J N a r a r a x ca r
o n ng Gra ar or or an c n ca r or IR S 000
ar n R D nc x rac on Jo rna o ng c

o on r c on In n ran a or onnan *e* Dro G n

Ja S an c c a ca on o a c on a o r n ac c
a r n o an ng ac n ran a on 00
a a S n a c na c ara r a w o n c n go o on ca II

a a on r c ac r *n* In Groc ow ra n a
n c n w c 0 awn c wo n w r o a a
o rn a or
n w c r c n r or ac o o ow c owa *by* rac o og
c n
c n r x In o o r r ow r
c o S ro anow In g n In or a on S a c a
c na awn c a I 00
0 r x c a a o r r n a R N S
r oro gar a 00
é Grow ng r x In G I ng 00
N S o S r ng r rg 00
rc I w c Sa or J S a ran oronc a J S own r w n
c n o c n w c n awn c wo In J a o go N
ra w 0
ég r a co é n a on ra q a c n ran a D
n ér é é c à on réa 00

nna

a D ow rg n co occ rr nc a r x a a cr on o r
a nc In an roc ng o n ang ag c no og
on r nc o na o an 0 00
a c a I ro a w w o J o I

c I r a c w J ar N a D c onna r x ca co
na o r ran a con ora n R c rc x co é an q o I II III
I r n r é on réa
non o ano N ag x con rc c r c n ca R or G
G R G S
an a r r c ra on GN an ra a c a à
an a ro r é angag 0
r o é n c rcon an n r on n ax q o é an
q In roc ng o S S on r nc N c
0 r D co r con ra n on non x ra o on ro c n ng
ng c 0 00
Noa a c n ran a r
ar owa r a wno o r w o an a a Naro ow
o c
o a S own n a c no g n ra wn c a own w o c a
a Naro ow o c In J a o go N roc aw
ra w 0
r r ow owar a ar a gra ar o o or a nc x rac on
In roc ng o Gra ar an or ora 00 c c R c o a
ar
r r ow a gn n an o n nc D c o o
Non on g ra ona on ra n a roac D n r ä n
g n
r r ow I I N or r nar r on In o o
r Sc nc o ca o Sc nc ar aw 00
r r ow r war an ow r c n ow a o go a c a
c na awn c a I 00
r r ow c ♠ S a ow ar ng an D a g a on n
g n In an roc ng o r ang ag c no og on
r nc o na o an 0 00
r r ow o x c ag or o In roc ng o
or o og ca roc ng o S a c ang ag 00 00
0 R g R o R a J Gra a r é o q ran a n
r n r a r ranc 00
R N ora ar a a G S an a o ar N a o
ro an R ro c an I a an x ca In an a on In
roc ng o a o o n r D r c n ngar J
0
Sago é n a rg r o r ff n ac c
x con or r nc arc c r acq on In c R 0 G n
I a 00
Sa on a gor a ro a w w c n o In a ow R
a gor gra a c n gr nn c w w c n o
rac In J a o go o o n roc aw
S r an ara ag S a J ar ng r ca arg n
r c r or n or a on x rac on 00

Św Gra a a or a na a o go Ro raw n w r ar aw go o awn c wa n w r ar aw go ar aw

Św a no c a now w ow n w o c Do aw n c a ar aw
S r c a w a now or o n a w go o go r o n a o on ca
S r c a a now o go r o n a n w r o rn a or 0
S r c a an ac *e a ne e* In Gr c n r w c a o o c a D N nac n a rac o arowan ro orow g n ow Sa on o a 000 n rac na ow awn c wo n w r a o oc go a o 00
0 n r é n S n ax S r c ra nc c ar gr n S a n a r o n a o go w c g n ra wno ran or ac n o a a a Na In J a o go

r r i f r i r

z szt s Ł sz l
z szt Sz l d sz d w z

o Ja an In o In or a on c no og ar aw o an
e i e y g b y p w t e p

A st t r a r o r wo oc or a c no og w c a a n on a o o w c o a c o r c n r gn an n nc on a o ro o o on cr o oc or a r ara on on xa o ar aw c ran or a on o n an a n co on n o o c or a ﬀ c an or r n r n a o 0% o r co r r q ro g a o a w o a ng o an o ra or

o ds o c or a c c no og r c n r gn a

n uc i n

t ls ss t t t s ds d s s s U t l t l t s t t l w s t s d d t t l d st ﬀ d ts d t t t l ll ws s st t d s s d t t ts s t d tw t ll d t t t t t d t ls d t st s t s t t t l s d s st t l ts t s l t t d s t t ll s d s s t s s t t t l s l t s s d t s t l s t ll t ss l t t s t t s s s ds l l d t l t t st d t d

s C n si n n s App ic i n p c s vic s

t s s s s t t l s s t s t l s lt l s d l s t t t t ls t d l s s t d s Us t d d s U S 7 t d s ss l s t d s t ddl t l t t ss s ll s d S s d t t lls d ﬀ t s ts s t s ll s d t l s w s l t s

t s U s t t d l t t
t ls 5
Us l t s t d s t s d l t t t
t t w t ls s l t s
t s t d t s t w t t t s l t
s t s t t l ss l t t s t s t s t
t d s s

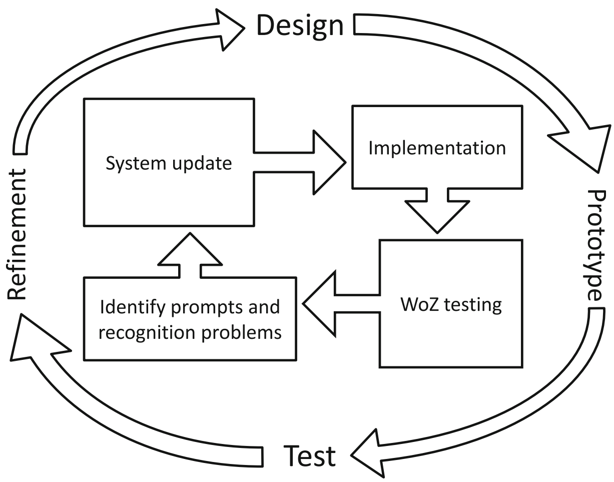
effic e cy t t t t w s st s ts s
t t s l s d
e ec ve e t t s l ss t t l s t l t s
e r y s t l
c s t s t s

Us l t s ts t s t t s t s t s
s l t ds s l t l t st d d s t ds s
s l t s w ll w tt d st d t l t
d s t ds t t s t U st s d ff s
tw l s t s U d s t s t t d
t

s l t s s ll st t d l ts t d t s
t s ss t d t d st d t d t d
l t d l s l t t ll t tt s t
l t t t t t l st t d l s t z d
s s t d l d ts s d t t l
s l st t d l st t t s s t d t
t s s t l t d t s t ll s st d t l t t
t s t 5 7 d d t st t s s s
t s s t s w t t
2 S t l ss U l t s t l t t ll l
s s d ds t s t l t t t d s d
t S s t s ff t t s s t d s s s
st ll l st t s l sts d l w s d l ts
t d w t t t t s w ll w t t
st l l tw d s d
tw t s ds t t s tl l st t l l sts
s s s l d l l d l t s
st
Us t l t s U s s w t t w t
s U s s d t w t st t s d t t U
t l l t s s l t s ll t t w s t t
s t U
t l d l s s l d st w s d l t s
l d s st t w s s s l t l d l
w U s l l t w t st st t s U t
d t t t d l st w t t w t t t l
d l s st

5 s t t w t l st l t d ws
s s t s s s t t t t t t
s t t s ll ss d s t t ss t t
S t S t l t s s t
t d t wl d d l s ss
t d d t t t s s d t l t
s d s st s t l st t t st t s l t d s
d l l s s t l d t d l s st s
t st t t l ss t t

s t t l s t d s t d d s t d l
d t d t t t t ls t s t t d s ss
d s s s l d



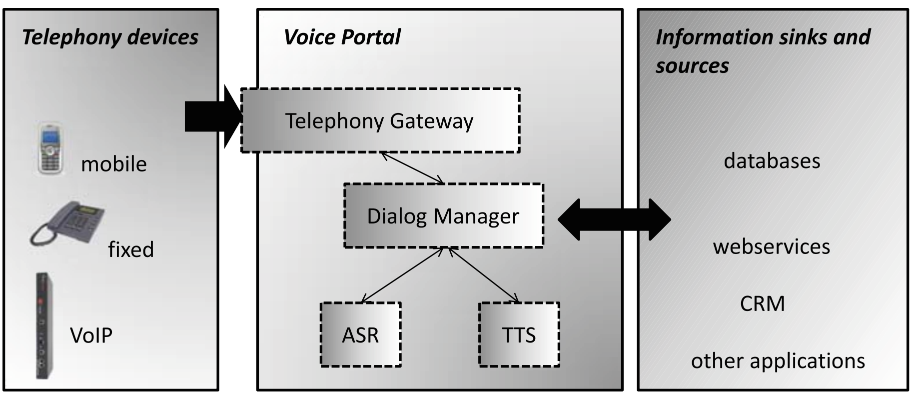
F I ra o c or a gn roc a r

t st s t l t d s t ll w s d d s
t t l s ts t d l tw t l t d t s s
2 t d t ll t d t t l t s s s w ts
t t t s s ll s l dd t ll t ws w t t t l s s
l s s t ds d l s d s s s l d
s d s t t s ll s s t z d z t st
t t s s d t s l t t t t d s st
t l t l t t s st ll
s s d t d t t l t s d s s t w tt
d t t d s s t st l s d st d s
d s tt w t t l s sts d t d l t
t s s

t ll t d t d s d t t s s ll ws s t
ld t st s t t l d t t t t st s t
ff t l s t d l s s t t st lls t t t t d s st
t t st s tt t s t t d s t S l t
s d d t s t S S t t t d t s
t t t t t ss s l d s d s d
ls d t s s t t t ll s l l t t
l t d ll t s s st t t st d l t s

ic c n i s

t d t ls t t l d s w l
t l s s ll t l s sts s l l s
2



F 2 o c or a a n n

t s ld st ss d t t t t l t s d t
d t t t ds d l w t s l d st d
t t ls t t w t st t l tw d t s s t
l w d s t ts s w t s t d t ls
s l t

1 r he e h e e ver

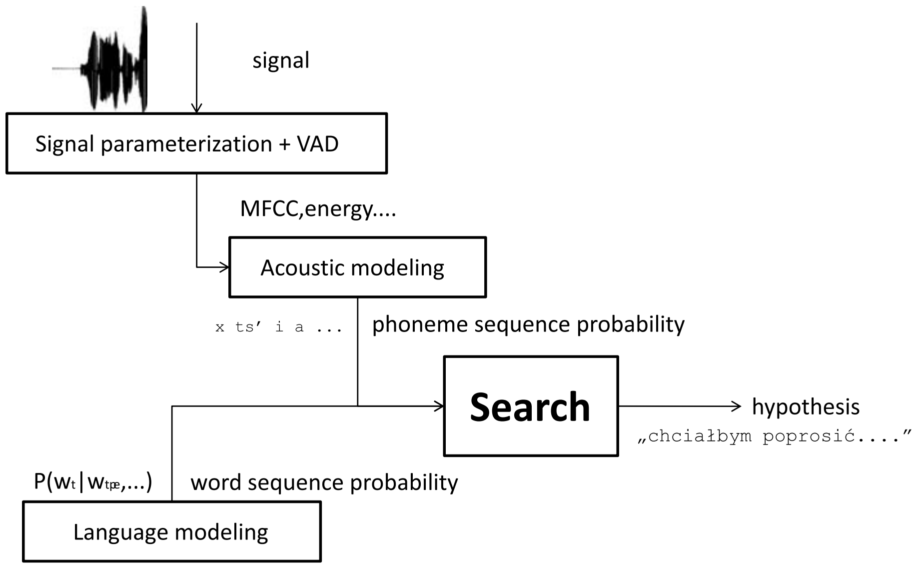
s t t t t t S S s st s
s d t t ll s st s l s d s st l s l
s t t l d ffi lt t t s t ll t l st l
t d t s l s t s sts l t l
ss s l s w t d ll d t st d l t

d t t w ds dd t l l s w tt t l
d d d ll
l t t d S s st s 2 s t d lt l
t s t d t S t st d l s t d t z
s t t t s t w ds s t t t t
t l t ff ts t t t tt
t d l s lt l t s t
t ff ts t l l s w d l d d t t t
tt s lts t S

eech ec

t t s t S s ss t d s l
d d s t ts t t l s t t t s
t l d s l s d d t l l w d t l t d
dw dt d t l t t s ss s ll t S
l d t s t l l s s l dl ll l
t l t ffi s w s ll t s ll l lls
s lts l t t s ss s 2 d l s t s
t s t d s st t l s s l t z t w s s d t t t
l t t s s l 2 s l st l ff ts
t st d s d d t s t d s t t 25 s
l t w d w t l s ls s l d t z w t ts
s l t d s s
t s t t tw d t s l ss
l d d ff t t s w t st d s t t
t t z t d s d d st d t
d s lts t ts
t t s st s l t t t t
d l s s d s d l s s l t t st d l t t s
st d ts l s s t z tw ts s d
s s l st t s s t s w d d
t t t t ll s st d l d s s
s l d s d d
S t t ll l s st t st l d l l d
s t s t t s st ds t t d s ffi tl l
d t s t t st d l l s

c c M e s st d l s t t t
t l t d st t s d t st s t s s l
t z t s l s d t s l t ds t st
l t ss t d l t w
s l ss l t d s d t l l tw s
t l s dd d l s d s st s s
t d t s t d t d s l s d d ls

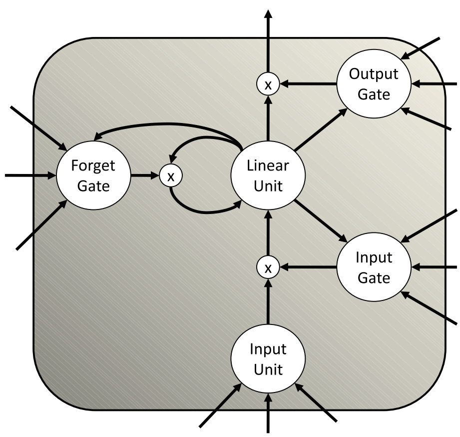


F n o c r cogn on

l tw s t d d t st d s t l t st t s st s s w t sl ws t w l l ss t ll ws t s st l t

S l t s s t st d s t s st s t s l st d w d tw s s l w ll t t t t t t d d tt t s s l t lt l t st l s t s t s t s s s t lw s ff t s t s t t l t s lts t d wt t tw w t d s s t l z t d t s d t t l s t t t l t t t t t tw l s st t d d t tw t l s t t l t t t t d d t s t d d t s t w ds s s t t t l t s t lw s w w s t d t d t t t l t tw t s d ffi lt s s t s t t l t w w s t t s st t d d s s d t t d t d ff t l s s l t t t s l t d ffi tl s t d w d tw d l t l tw s l z l t t w t d l s t tw t l ll w t t s t t tw t ds t t l tw s t s d d t d s t t d l t 2 t t s w t t d t s s t t d s s t z t

s st l s s t tw s l l t t l
t w s w l tw s s d stl t
t st l t s s l t l l t s t
st t st l s st s d t d s t ll d l t t
st d w s t s t l tw ll d S t
S tw s s st s ll d l s
l t s s l t s w d tl t l s l lls s d
t st t t s s l s d l ts w s t t
t t s l s w t t t t t
ll s lt l d t t t t t t s t l tt s z t
w ll st s l t ll
S l l t t t t w ll l s l l t
l w t t t l s t s t s t s l w ll w
l st d t t s l s t s t t t ts t
ll t l t s t s z t t ts t ll w ll ls
z d t d l t t S l s t
w t d w t d s t t s S s d ff t l
s s t



F o o og o a ng or oc o S n wor

tw t l t d s d S w s s ll ws ts l s
2 lls l l d t t t d t t s w
7 t t ts tw w s ll t d ll
ts w s d t t t ll w s lt d s s
t w l ts s l t l s $[-1, 1]$ s t t s t
tw w s l ss t s t t t l w t ss t
t w s s d w s l l t d t st S t w ts
w lt d t st s w ll l t w s l t d
t t w s 5 tw w s t d s s st

d s l w t t d t t w
s ds t s s s s w s t
t st s t d l w s t st d t t s 7 s
d d % l s t t t t s t s t s s
w s t st d l d l

r ec S t s s t t l l S t l
l ss s l s d s t s ll s ffi t t d l
s w tt ts t z w l w ds s s t t
s lts w d s s t t t d d t t
w w d t s w t l ss s s d l t z s
s s d t d st st t l d l s s t
d ls t l t s l st st t
d l t l t st t t d ff t ts
t d t l t t t s t st t t t t t l
s t d t st l s st t s s s l d t s
d s d t s d d d s s s l w ll
w l t s d t s l t s l s t t l t t ds
t st l s dd st t s s t s d
s st s t s t d l d s s t s
w t s s d d s ls s d w t t
t s t l d w t d t s s l t s s
l t s d ff t t s s s t s st d st t s
s s l t ll l t l l t t t d t d d
s s t s ld t t t ll l w l t t s s l
l t s sl tl d ff t t l ll l t
d d s s s s w lt s s t d t
t t t z t z s st lt
s ll l s l s

e M e s l d l st d t s
s t d d ll w l s t t ll d s t ll t w d
s s t d d d s s d tw t ds l s
d st t st l l d ls
s d tl w s s ll w d l t ll t
s t d t St t t t w s d d s l d
s t t w tt t s s ts d t s
t s w w s d t t w s t t st
d d tt www w s s s d s d
d s t s t ll t d t dd d t
d t s t t l s d l d t ts l t
w s d tl s d t d d
St t st l l d ls ff t l t t l s
t t s d t t t t st d d tl
w w d s s ll w d t st t t l t w d

s s d l t d l t t l s Us ll l s t st l t t s s d t d l st t s l $P(w_t|w_{t-1})$ d $P(w_t|w_{t-1}, w_{t-2})$ t l d ls w t s s ll st l t t t s d ls l s l ll w t d t s z s t s t s d t w ld t s l l t t d l st t st ll s t d st s d t s t s d t t w w ll

t d l w d s s dd t l l st t lt t t s st ts s w ds s t t s t t s s st d l st s t t t t t s s w ds d s dd ss s t t d st t d d ts l t d l d l t ts t t ll t d l w t S d s t w ll d t d s d t s 5 l t U t w t s s w s t st t l t ds t ﬀ t t d ls s d l s st s s s t t l d s d t s

eech y he

d t 7 s s t s s s ss t s t t d s t d t s S S s t s z s l ss d t w t t s l s d t s t s s t l t d t t s t s s st l s s t s s s tl d d d l t t s t s s t s l t s t s s t t s s t s s t s s t s s st l ts s l t t l s s d s t s s ll l s t d t t s t d s t s ll s z t d t s d t t l t d st ts s S s t s s w t s ll d t s s st d s l w s st ts

S ll l t t s tt d s lts t s d t l s w l t l s t s t t l s l t t d s t t t d s w d l s d l t t l w st l 5 ts l s S l t st ts s t l s s s t s s ts ll w t l s d s t s t t ll dd t ll s t s t ss t d t s d d t s d d l ts d t t t s l s s l d t s l t s t s s

c st d s d t s t t t s s l st t s l s t st s t d s t s st ts s d s t t s st t t s d d d t t d s s ds t l st t t

t s l t ts s t st t s t s sts
t st s l t t t t st d t st t t
st d t t t s t ll w l

$$d(\Theta, T) = \sum_{j=1}^{N} d_u(\Theta_j, T) + \sum_{j=i}^{N-1} d_t(\Theta_j, \Theta_{j+1})$$

$d_u(\Theta_j, T)$ s t t t st d $d_t(\Theta_j, \Theta_{j+1})$ s t st Θ_j, Θ_{j+1}
t l s ts

$$\hat{\Theta} = \arg\min_{\Theta} d(\Theta, T) \quad 2$$

l w t l t d t st t t t l t s lt
s t s t ll t sts l t s w t t
s ll st ts w ll t l w st l t t s t ls t
s l t t l st ss l ts s w t t l t t
s ts s t s l l t t l t s st t
t t d s w ll s d ll t w t t s t w w t t
s t s z t s s t t t l s l t
st t w s l s d ds t s d
t st d t s s t d s d d l s t z t
t s s d t s d t s d s t st t s
d ffi lt t s d t t l t s t s s t t s t
t s t t s w t S s st s t t t st t s
t t t d s t 2
l t l t w s s d t st t t st t
t s l t s s t s z l t l t s s t t
l s l t st s l t s s ll ff t
l t d l l s st t t t s t st
t 2 ss l l sts t t t t d t
s t st s d s t s s s S l st s S
t d s l t t z d st t s d t s
d s t t s t d s t l d s s t l t w d
s ll l s ll d s t t s t s 7
s l t l t ll w d s t s l t t l
l st t t t s t st t t z t d
S t sts s w t t t s t s t z d d t l t
s t s z d s d w t s t d s t t s
t d d t d d l tt s

M er

d l s d l t t t l z s t S d S s t
t l t w t d l t s t t s l t d s s l st t
w st t t s s t w tt s t t d
d l l dd t l d s dd d t d t t l t

t s st d ss t dd t l t s s d s s
t t w t t s d s s t
S t s d s l d

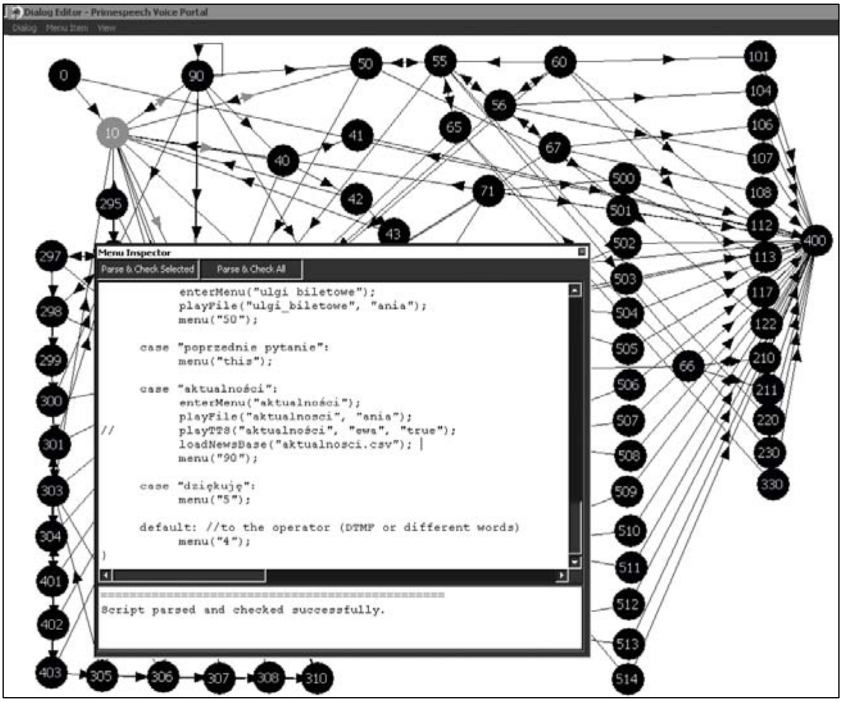
w t l sw t s st t ts d st t s t t ds
l t l s t ds t lt st s d s t d t
s S d S t ds
t l d t t s t t s ls t d d s t
lls
t ss d t s d w l d t t t
t t

St t t l d s ts d s d dl w t U t l 5 t t
ls s s s t s t l d t S ts lt d d t st d w t t t
d st t t t l l s lt l t
w ll ws t s t t t t S s l l t l
l t d w l t dw w t ll t s s d
t w s t s s S z s t t S s l
w s lw s s t t t

5 ce r rch ec re

t l s ll l tw s st s st s d l s t t
t w t t t l l tw 2 t w S
s d S s ll l t t ll ws s s l t
s st s s st t t s s t l l s ll s
st ll d s l t s l t 2 l s
s d t s st st ll d d ff t s
t w S s d S s s S s d t 2
s t s ll l t t d s s s lt t d d d
t s ll t l z s t lt l t s d Us
t s t t l st ll d t t s w s t t t
st s s s d d ll s st ll d t s s
ss w t

e y s s t t s lw s st ll d t t w t
s l z d t l dw ts w s d l t l
t l ds s ds s t t l S l ss
S d l l w d l s l t s l
s d d st t t t d t l ts s t
d d ll t t t d S t w w s
d s d t s t t l d d t l t l ds t s t
l ds t t d tl t t d l t l tw s d t l
ds t t t l s s t t w s t
t s t t d s l t t l dw t t S s d
t S s t t t l dw t s ls s s l



F 5 G I oo or cr a ng a og a n w n ow con a n a n wor o a og a w ran on w n cr or r n n a r w n ow an con a n cr o a c a n gra

t ll t t l l t d d s t lls s d d s s ls t t w s s s l t l w w t t t d t t s t s d t ss t t t w t l w ll ws t t s w t t s t d w t s t d t ll ls t l l t s

erver s s t st t t t t t l ts t s s t z s t t d s d t t d s t l d t t t ll stl l tw t w t s s t s t t d s t l t t d l s t t s l s s s s l l d d t t t w d t t t S ffl t sts w t S S w d t t s w d s st w s l t d w t 5 tt s d d t t l t w s d t s d d d d ff t t s w s d t t st s ll st d l ss l s s

s st l t d d d ff t
w ds d t d l d ll st t s s w d s d t w s t
d 5 st ts t t t w d t
l t sts s w % w d t w st
s d t t ll t st d t w t l t d s s w s
l l l w ld l t t t l t s
d s ll d l t t s ss t s s
ll s t t d ll t w t t s t ts
tt s d s t l s t t s l d t sw
t st s s d t t l t s s t t t w s ss l
t d t d t d l t ss t s s t st t t
S s t l t t t tt w s t z d l d t
t l t s t s t t

TT erver s s lt t d d w t t t d s
s t s z s t s t s z t l l t d
st s t S s s z s ll t sts s t d
d t d t t l d s t t s lt t l
s d s l t d d s s t d s t s z t s t t
t S s ls l s d d ss l
s

ce r M er s s w s d t t d
t ll d ff t s ts t t l t s ss l ws
t t s ts s SS t s t l w t ssw d t
s t l t w l t l s t d w st t t w l
s st t t t s t ll t
t d t s t t l t s t s t s z d
t t l l st t d d ss s d t l t t
d ff t s dd d d d t s st t t d w t
t s s l l t z t ll t t l
ll ts d d s l s s st t st s d st t lls t l t
lls t w s d st t lls w t d t t s t
t t d l

p i n s n su s C M y

d t t l d t t t d l d s d t
s st w s t s t t t s w s t t t z d
s t s st sz w w s t st t l
t s t t st t t l d t s ll t t t s ss l
d t l d t 22 d l s 2 t s
w d ff t s ts ll t d s t t d t
t s t l t s t t s s t s t d l l lw

ara a

d s t st t s t t t d st t d t
t t t l t s t t s w ll t
s l st lls t w t ll d t
t S l tt t w s d t t l s s s sts d
ts

1 e r y e

t t w t U t S 5
ﬀ t s d t ll t l t l d l s s l t d
s t d l s w s ll t d t t s w s t
t t t l t tl d t s d t ll w s t d st d
w t ll s st t d w t l t t s s
sts

d t ll t w s d s t ll t t t t s t
t tt s t t t t s w w d s d t
t s tw t s d ll s d t t t d d t
s ss l t w t s t s st t d
s d l t s t t s d t l d t s w t t s t
t l s

s l t 5 t s d l s t
t l ss s w d t d

t sts t t tw ts t t
t t l st d l d t t l t st
t d st t

t l t s t s t s w l ss
l l t

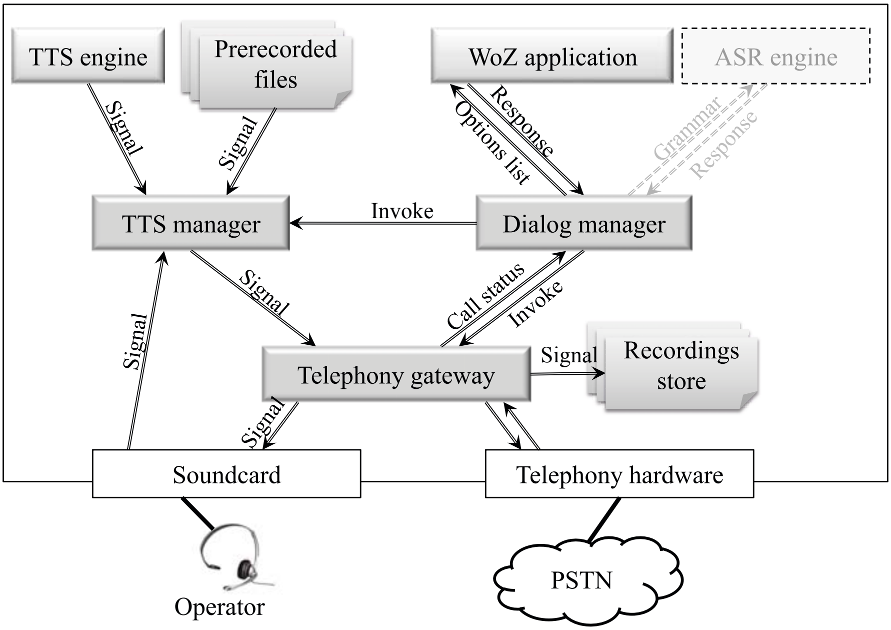
t st s t st t t t st s
l t s st s t

t d t s d t s t t s d
s t z s ld t s s d s l d s s t

t s t s t s d t
t t t s st s

ts t s st t d l d t S
w d t t s l t s ss t t t l t w
t s s s s s t l st t s l t d l
st s s l t w s d ts t t t d t
t s t s d t s t d s t w st s l
t ll t t s t s st t t w ld
sw t t ﬀ d t t d l w t t ll

ts s tt w t s sts
d d lls l 5 t w t d d l ss d s l t
t t d t s t t s s 55 w t d t s d tl w t
t t t s st w t s l tl t t t s d



F ar o

s s w d t t t t t t d s s t l w
s s w s s d t t t s st s t t s d l s st t d w t
d ffi lt t d l s s s t t s d ts t t s t
w s l d t t t t s st s ld t s t l s ss l
l ts w d t t l t t d d S
s t l t t s s t s s w d d d t sw t t
l d t ff d d t z d s d d
l t s s t s s t st t d st t ts
l d d d s l l s s t d t
ls ll w d t ld tt l s t S d l d t
t ts s t t s
l s s t s d l s s ws t t t s l st ss l t t t
ll l ss s d l s stl s t l t d st t t s w
l t s t s st st s t s d s s s sts
s d t ls s w d d d t s t ll w d l t s
t t d s s d t t s t t s d t s ws d
l ts

ZTM ce r e re

t d l t s ts t t ll w t t t l t s
ss l t t w t t t t t ss

t d ll t s s t s s t t
d t d t t l t s lw s d d
t l t d t s t t t t s t t l st d
s l d d t t s l ts t t s d t s d
ws tw t s t w t t d t t s t
w t d tl t t t t t d t s
d s d l w

e r re T e st l t t t t l s
d d t t s s s d t s s st z s t
5 d ff t s s d t st s s w t ls z s ll
t t s t l s d t s d t s t s w st s t
t l t s t t s s l d t s d s t s z s t
s s t d t st s t s st s t s d ds
t s l d st d s tw st s s w
t s t t s st s t l d t t l s t d l w t
t d s t t l t d t s s t
ls t s d l s s t t sl t s tl 2 d s
w t t l t t d s t s l d s s t t
t s st t d l l t t t t t

d l t d t s st ts t s s w d t t t
s t s d d w t t st w t t s w ts t s d l
t s t d s d l s d ff t w ds d
l d s s t s s t t st s ll ws s
% s s w t t t s d l s t s s l t st t t s
st t s s t st d t s d l s t s st s s
s d t d ts d ff t t s s s t w d
2t t

t st s t t t t w tt t
s d t s t s w t s s ll tw
d s t s t s l tt s d s l st st
t s s t d t S w ll s t s t d t s st d
t st s l t s s s l s 2 d 2

t ll w st t s st s s t st s s d ffi lt
st s t s t s st s l ts d t s t
z d d d ll t w t t s t l lw
St t 5 s s s ll t st ts d s t s w t
st s l t d t s t s st t t s w s s l
l d t s ll s l l t s
l Sz l S l st t s l t s s s s t
t l s t s dd t l st t d s t t st ll
s l st s s l t t s t s ffi d
dd t l s tl ts dd t l st t
t l

st t st s tw s t d t s st s t
d t t s l t s st s s t s t t d t
l st d d st t s t l w d t
t st d t w ds t t t t w ds l w t d st t
l st s t t d t t s s t st d
S st t s d l s t t d ff t d t t s
t w ld w st t t d t ll st d t s st ds t
t l s st t t t l d 2 l t d ll ws t
s t t t t s t ds t d t s s
d t s t t ls d ff t w s
2 l l st t 7 t t

T c e r ce s t s t s t t t
t t t 2 d ff t t t t s l l s st s s t
s w t t t s t st d s s s st s
w t s d l st d t w st s t s s s t d w t
t s t t

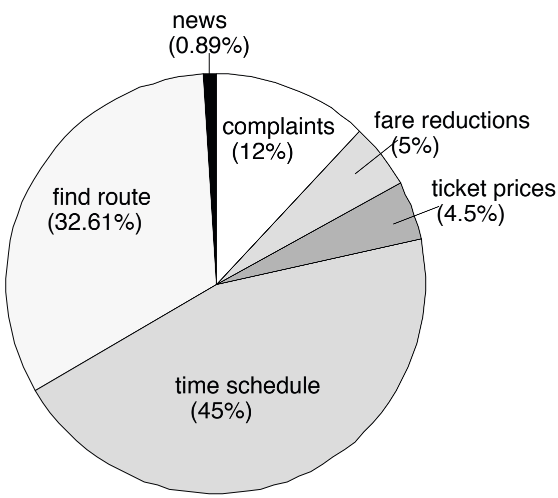
re e c s s d t t t t
d t d l s t t d t s d t ffi ls
d t l t d t d st d t ss S l l t
t t t s t s d l s z d t l s d t t
t s t st 5 st s d l t s s t t
t s d t

e s s st d t s s t s z d ws t s d t d ll
t st ff t s sts t tl d s t t s
t t t t t s t t st t l s t t
s t s s s l s t d w t l st ws t s t tl s d
t t l ws t s ts t t s s t s z d

t t t t l t
t s s t t t d l ts s t s l s s t
s t d ll t t t t s t s d d
d st d t w s d d t t s s d
z t t t s d d ts t d ff t l t s t
s d t t t t s s s l t t l d
t w t t st s U t t l t s t ffi
d ts t d t l ts l ts s l ts st t d
ss s w t s t ll t t s d d s t
s s d w t t t l t d s s
l ss l l t t t d d s l d st t t

Te e e e

d t U t d l t t st l t t s ts t st d t lls w l z d d t w l s d t t t d % tl l ts s d t t l d d d t t t t s t ll st d d d t t t s t d t t s s d st s d t t t s l l ts t st d d l t ll lls t d d s st t st s ll t d 7 d



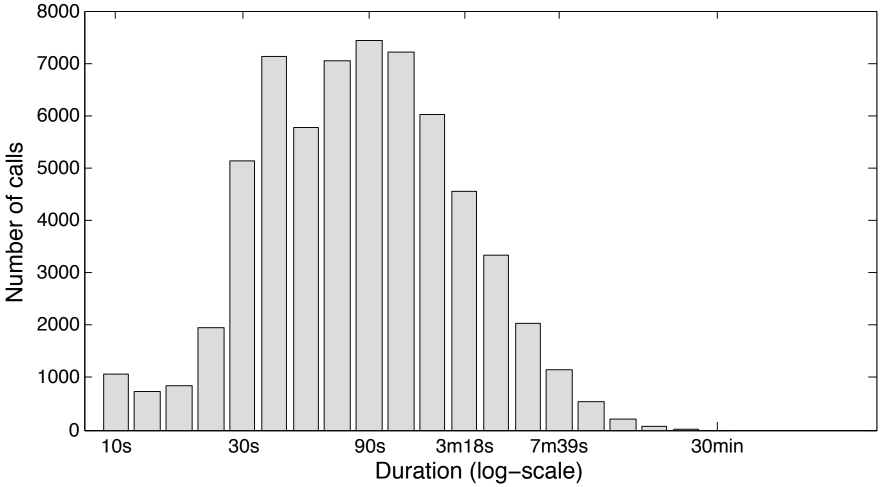
F a r on a o c or a a on ca 0000 ca

d d t d l w s d t l s s d d d t s ll S dl S s w ll l d s t wl d t s t dd t l st d s l s t w t s st t z t s t t t d t t s s t d t ss t d t d t t ll t t t t t tt t t z s ls t d t s t t

s d t s s s t s st s d t s t ss t w t d t t d tl t t t st s w d t t t t ss ll s w t s d t s w t t

ls t t s d t st d s d l s d d w s d t t st ll s s d t d s d l s t s s w t st t t s

t ll t d l d d l w s t z d t ll s d t d s st sw d s d d t d st d w t s st ts l



F D r on o ca ra on nc ng wa ng a o c or a a on ca 0000 ca

ll s s s t d ld t l t w t t d l st d t st d d t t s l w s s l d ll w t s t s t t t t t t d l

C nc usi ns

s t t st t st s st t lls t s ds w s t l ll d t d l w t t t 2 t s t l d d l t d sts s st s ff t d s ts s s t ds t % s s l t t sts t t t t d s st s s t lls ll t t d t t l s ss ll s t s t w t t t t t t t st d s t t t l w s d l d tl t l w t st d s t t t s w t w s s d t t w t t d l st t d t s t ll t t t s t t t s s l d d l d s d t t t d d d t t t t t s t ll t s s t t t tl d d d st s t tw s st w s sts stl s s d ld l l lw s w ts t s w t t t s s d w t st d ts d l s s t t d l s w ll l s ll t w t t s s t t s s t t tl l ts s t s d w t t t t s d t s s st s t t t t w t t t t t t s l l t t ll st l t st d d t ara a

t s s s d t s t t s d 5 s ds t t t d t d t s t t t d s st ds tw t t w w s d s t t l t w t s t s t t s t t t t t t d t s t t l s s t l d l d ts w s t ll t ll d t t s st t l t t ds l t d t U t t w ds l d t t d l w

c e e e s w s t ll s t d t U t S 5 t l t l d s d t s s t t s `www.pr sp .p`

f nc s

c ro ao n on S o n ang ag roc ng G o or gor an S D o n r n c a ng woo ff 00

ac n G n ra ng 0 con o r ro o I a ng n ar r gr on In roc ng o I S a a S o

o r ar organ N onn c on S c R cogn on r roac w r ca c r Dor r c

rn rn ng G I n o I D a og D gn r nc or a ng ca on cc on o c R w 00

D or R éc a an r D c ar R ccar G r G S o n ang ag n r an ng or on r a ona S S gna roc ng aga n S c a I on S o n ang ag c no og 0 00

D x na J ow G a R an o r In rac on ar on r n c a on on 00

D o n In ro c on o x o S c S n w r ca c r Dor r c

I In rna ona c n ca S or rgan a on S c r In r ac G 00

or n D roc G r now c R ara ar o x r n or a on a ran or D a og S In roc ng o In on r nc In g n In or a on S a o an o an J n 00 n r n

0 roc or n D ara R cogn ng onn c D g S r ng ng N ra N wor x S c an D a og 00 rno c R c 0 00

ara G r now c R nno a on n S on o S c Da a a In o c c a w c N a I I 00 N S N I o 0 S r ng r rg 00

ara G r now c R D gn an Da a o c on or S o n o D a og Da a a ang ag R o rc an a a on on r nc arra c orroco 00

ow c a ara arc n a Ra ga n w a J G r now
c R nno a on o o o n a og n N ro c 0 o a ar
n S r ng r N I 00
N n J a ng n r ng organ a ann San ranc co
o o a c a an wa S c c no og aga n
00
R n r R a r x r n ax oo an
a gor SI R w
S ann o nc on a on n n c on c n D
JII ar awa n r ara on n o 00
a ana awa n on G S ano ang J on r cogn
on ng a n ra n wor In R a ng n c r cogn on organ
a ann San ranc co 0
r o J ac ro aga on ro g w a o an ow o o roc
I 0 0 0 0
0 a R r D arn ng a gor or con n a r nn ng r c r
r n n ra n wor N ra o a on 0 0
J R or a n S c on or Na ra So n ng S c S n
D I D o c r ca ng n r ng an o r Sc nc 00

r i y N
r i f y i r r

z t ws

ac o a a c an Na ra Sc nc
o g o Sc nc
n r o ar na S an
an In o o r Sc nc S
yt i w e p

A st t o o g a n o c r n ow wa n wor o or ar aw n r n or r o n aco c o o an a o a on n or a on offic a r r n a rogra w c n cro on c n o c gn a a o ro or onar o c n 0 a o rogra w c wo g n ra con n o c w ro o c a r ar o na ra ang ag an n r an a

o ds c n cro on c o

s d s t s s l t t t s w s ss d l s ss

n uc i n

S s t s s s t t t l l ss t s s d d t t t l t l d d s s t s ls ss t l t l l t w l s d t l s t t t w l t l t s w d s t w t t s l S t t s s s d s s l t t st d t s st s s ll d s s s ll d s s w d l s t s t d t s t s l s d ll st w t t d s l tt ts s s t l t z t t t d t ss l s st s t l t w d t l st 5 s s t t t d t t ll w ll t t t s l tt ts t t t w ld s s s d t d 7 s t s d st d s st SUS w s d s d ss d l s d ll t z t t d d t t t t t t d t t s d st d s st s d ts l t s st SUS s st d d d t t l s l t d l s

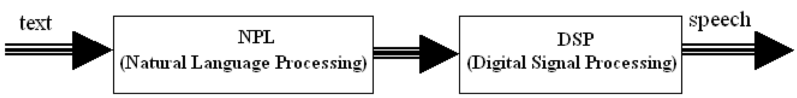
s ts t s t s s s
s l t s lts t l l t t s t d s t d
S s t s s s t s s d t
t s t t s st s s d t t t
t t t s t s z s w t ll w s t s s t 7

t l t l s d s t s s t s t d l t l s
s tl s ss l s t s t s s l s d d l
t l t s w s t ss l t s l t t s t t
l t s d s d t t t s t l tt s d l
w t s t t t s l d s s d
s ls s s t s s ls s t d t d
t ss d l t l s d s t s s t t l
t t l t s l l t l t s t
t t t t s t t t t s t d l
t t t t t s l tt l t d t s
d l ss t l t s t s s s tt t s
s d t d t t t l t s t d
t s w s t s s l d t t t l d t
t l l t t d l d ls d t s ffi t wl
d t l t ss s l d t d t t l
s t l t s t s s s t t ll t st s t s t d t
d l t s t t s t t s ls t st d ffi lt
t ds t l t

2 t s t s s s s d t s lt d l d wl d
s st s t s sts tw ts t t
t t s l d t lt s t t s t t s s t
l t t s s d d s s d t l tt s *l*
d t t *ou* - *l* *od l* s d t t s
d dw dt s d l d s tw l s t t l st t
ts ll d t d t ll l s d t
ts t d l t s dd t l t
d t t st s t t w t s
ls ss tw s st t s t s t s z
ll l d s d t s d t t s s s ll s d
t s t s s s s l s t ld s st s
d t s d t st d s t S t d
t s t s z s s t z d t ll s d w t s
s t s s t s s s t s s lts t l t s
d t d w s ss l ll t t s st s t d
s t s s s l t st w d l s d t s

cro on c n ar c g n o gna w c a r r n r on
na corr c rc on o on r r n N r o r on
n n on n o on n a n a r q nc o o o c
con x an r

t t d t s d s t s s s t t s
t t t s s lts t t st ts st d
d t s ts t t d t s t l st
st t d t s t t t w t st t t
s t s z w tt s t s l d t s d
t s l t ts t t z t t st d t t st
l t d t t tt t st t t s ts
t t s t s s s t d t t t l t s l t s s ll
t d ff tw l d s t tt s t l tt s
t l ss l ss t t ts d d t l t l t
d t s z d t s s s d t s t tt s l ss
s d d t t s t s s s d ffi lt d
l t s t t t tt d t d t s s l
st d t t l l s s w d t *u l* -
u o d S *D l l o* ts
S *T* - *o*- s st s d d t t s
s t s s s t t s t d l t d t st
ts d t t t t t s ld t t l t
s st s s d s d w ds t l t d t t
t l t d s s lt t l s t t s s t d
t s st ll w ds s ll l s
l s t l s t d t t ll t
stl s l s d t s s s d s st
ts t d t t t l t d t t t
t s t t l w d t t t l s d
s t t s t d t d t l d d t
t ts s d



F a n co on n o a x o S c n

t t s t s s d s t st t l s d s t s z d s
d s d d d t t s t s s t s s

U t s l t w s s l d t s s d d s d
d ff t ts s s d d ls s s ll l s s w ds
s s d s t s s t s s w s l d t s t s
ll t d s l st ll s d s d w s ff s
d l l t s

2 s s t s s w t t s d d w ds d
s s t t l t tt s w s s d t l d

J ow

t l s d ds dd t l l t t t t
s s t w s t s t s l t t t t t
t s t s s t s t s t s z d s s st d l
w w t l t s l t d t l t
d s l ls t s t ll l t t l
t l t s t s s s s t t l t s s t s z
s s d d ls t l t t d t t l t
ss s t
s d s t s s s s dd d ls t s lt sl
d l t s t l t t t d t l
d t d t s d s
s w s t s s l s t ts ds w t
t w stl s t s t s z t s

ll t s t l s s d s s t s z w l
t d s tw dw s t s z s t s ss t t
t s d d s t t st d d t s t s st ll d ffi lt
t s w t t t t s l ts t t d S st s
t s t d t t l t t t s ds l s d t ll l
s t d st d t t t
st t t tt t s s t s s s t *ll b l*
u l ts t t s l t s t s d st t l t t
st s t s d t t s t s w s l t s d
t s t l t s t s z d s d s t
t t t t s t s t lw s l l ll l t s
s t s s s t d ff tw t l t t s st s t st s t s
d l t l t ss s t s s d s d t d ff t
l s s t s s ll ws

d-do *fo* *o* *o* t s t
l t s l t d d s l s ll
d d s ts s s s d st t t
s s s st s
2 *o* *o* *fo* *od* *o* U l t s
t s s st s s t s z s t t t d
d l t l s t s t s t t l t d t
t d st ll w t t s
o *fo* *od* *o* t s s st s t w
d d t ll w tt s d t w s t s
s st s l t w t s t t s d t t ll
l t w s t s st l s d l l
s t w l t s t s s s t l
Rul -b d s s st s t d t s d l t
ss d ff t s t s t t t st ll l w l t d
t t d s d s st s

l s z s t t d ff tw s l t d l t t s d
s ts t d ff t s t s s s 2

S c n r an ng S S S N w r on

S c q a an o a n a on

roac	S c a	r x g n ra on
o a n wa or conca na on	IG q a or o a n r r cor n nc	N can no ro c o or ar rar x
onca na n w o wa or o ca on	IG or nown ar G D or r ar	S can ro o or r x q a can r ora w n oor conca na on occ r
onca na n w wa or o ca on a o cro on c o	ID ro o can r goo q a	S can ro o or r x or x ro g n c ro o
R a a roac	co ar o o r a roac	S n or o n q a acro ﬀ r n n nc

M in incip s f Mic p n ic M

t l t t t s s st s sts t ts w t t l s s s d t d s s t s s t t l s s t d st ds t t t d d t s t s d w d s d t t t s t s t t t l t l t t s s s t s s t t s t s tt s d t t d s d ts

l t t s t s s w s s ll s t ts s st s t s s t s s l t d d t d d s s l ss t s d t s d t s S t s s t s w t s w t l t l l t w t s t t s s l ss d st t d d t t s l ss ss

Us ll st d d l ts st d t t t w ds s ll l s d s t s s s ds s d t t d st s s l t s ts t s s l w s t t s s ds t t s ﬃ t t t t l tt t l t t t t t t t d s t t t s ts d d d t s t s s ss s t t l st ss d d t d s t t tt s s s d l s t l t t s t t s

s tw l t t t t d s ts t d d t t t s s ll t d d t l t s t s z d tt s t d t d s d w l ss s t ll d s ll d t l st t s s st t s d t t d t t ll t

t t s t s l s t t s d t s
s d t t d t t d t st d t t t s t s
t t d s w l s t s s d t st s
l z t z t d s d z t s l s t z d
t t s t t s l t l l ts
s t t l t st t d l w t t t t t d t st
t s l d ls w t t l z t st w s l
d t s d d t st w w s t d t s t s
s s t t s tl t t t w t d s l
ts t s d w d ll w t s l t
z t s d t s t s st t t ts
s t l t l l l st s d s l t d d s
ll t s d s d z t s t l st t t s l t t d ls
w t t s s l t w s ll t d
t t s l

ns in n

t s t t t t t t l st t s t
t t s s d s t t s t t d l d d ﬀ t
s s t t t d t l t s st SUS t t t w s
t d t s st ts l t t t l d tw tw
t t s s t t d t s s l s t s s ls t
s l w ds s t d s d s l w ds d w t d
t t t s t t t t st
S s ls d t s t t s d t l t d s
t s d t ds d t d t d s d
t s t t t t l st t s t ss l s
t t l z t s st t d t t t l 2 d l

l t st S s ls t t s s w t

2 S o o ow

S o	r ogra c ng	ran cr on
	I	
I		I
a	a	a
o	o	o
∼	g	g ∼
o∼	w	wa∼

S o o con onan

S o	r ogra c ng	ran cr on
	n	In
g	g n	g n
	an	an
	w	
		I
	r	r
S		SI
	o	I o
	w	
x	n	xI n
	c r	Ir
	won	on
S	c n	SIn
	a	a
	w g	
		IS
n	na	naS
n	o	on
N	gong	goNg
r	r	rI
w		wI
	a	a

Do d u b o do u l- o d kul o k b o u l d u o o b k o o o o o o odob

do d f u o b o do o~ uf u l o o duf kul o k ob o u u l ud u o o b o f o o o uf o o odob

t l s l t t sl t t t t t t t s t s l t d s l t l s l s t l t

s t s l s d t ll w dd t l t t
t t t t l s

- w ls → s
- l tt s d →
- l tt ll w s ts →
- l tt ll w w ls → t l
- l tt s t l d l tt s → s
- l tt s s → s
- l tt s → ts
- l tt s z → z
- l tt s dz → dz
- l tt s →
- l tt s →
  - s ∼ ∼ t d w d
  - s
  - s t d
  - s dz
  - s
  - s l l

sc ip i n f Mic p n ic ni s

d st s d l ss s s

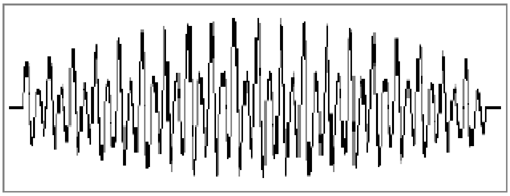
l w
2 ∼ ∼
s S s
z z
5 t
d
7
ts tS ts
dz d dz

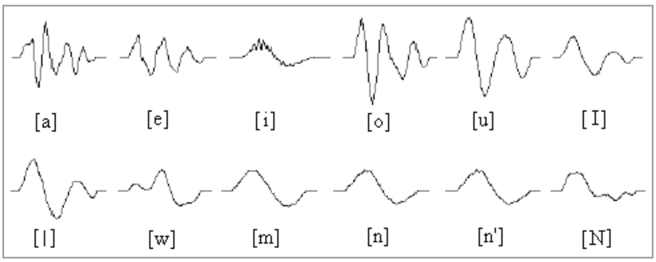
t

s s t d w s s ts
s d s t w ds w s s w l t t
ts
s d s lt t t t t s ts w t
t ffi ts l t d t t l s t t
2 ll t s d s t w t
t d s d ll w s t s t t s ts
l t s s l t ds
t s ts s d l ss s s w

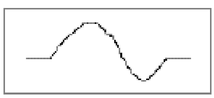
S o wor w c wa or r ara on o cro on c n

an a an a a r a na w g r o o a w a a a a o o a ragar w o won a a ro g a c o ro a o a w gong g won o a a anan c a o a a raca c w a gro a





F n o q a r o c on



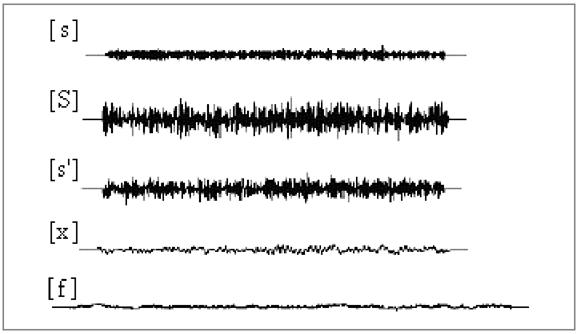
F n o a c

t s s s s d l l s l s s w ts l ss t s s w 5 t w t t d d ll w s l t t s st t d s d t t t ts st s l t s t l z t d t s s l ttl l t d d s sts s s t t s s d l l t d s w t t t t t l ss t s s s t s d t d d w t t ll w l t

$$V_i = (B_{i\,mod\,N} + F_{i\,mod\,M})/2$$

$$i \in [0, pN]$$

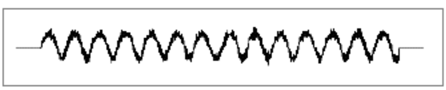
w



F 5 n o o c r ca

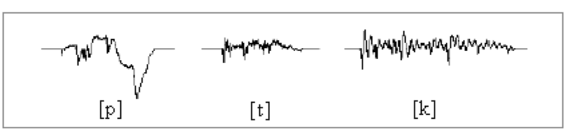
B_i l t t s l t s t
F_i l t t s l t l ss t t
V_i l t t s l t s t t d t
N l t t s t
M l t t l ss t t
p l t t t s t d

l t s t t d t z s s w



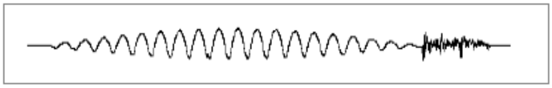
F S n c on

l ss st s t s st s l t ts d t
ts t w s ts t l s t
t ts t d s d ff t l t s s w 7
t st t s t t t st t t w t d s
d st s t d s t ts w t l d d
st t t s s t s d l d t t t t s t
s ffi t l t l s l t s



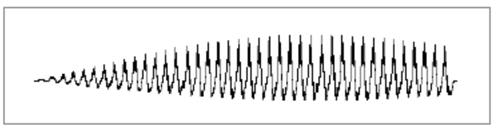
F n o o c o

l ss st s s dd d t t d t s s d st s
t w t d s l t t w t t
ll w s d s l



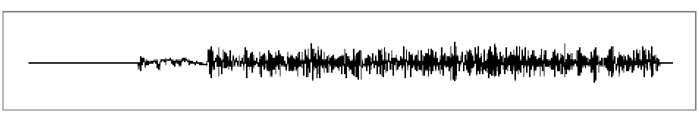
F S n c on g

s t d s t ts w ts d s d l d t t t
t t s ffi t l t
ff t s t l l t s l s t d



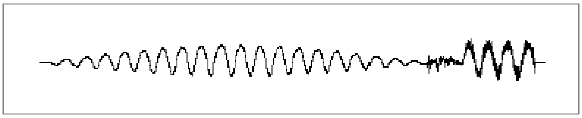
F S n c on

t t d st t ts s s w st
t t s t s z s l w s d l d t t t
ts t w dd t t d t t s

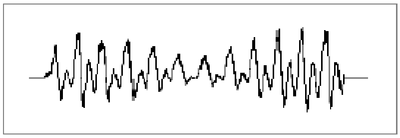


F S n c on

t t d st t d z s st ll l t d
st t t s t s d l d t t t t s t s
ffi t l t l s t s t s s d t
t l st d t s s t d t t s t d
s s s t t s s t d t t S



F S n c on on r r a a a ng cro on
c n g r ar o a on n w c r n r a n
n c x r n o no n r r w a o n ng on



F 2 S n c on r o a n r o cro on c n a r
an n r o on a

5 o on cro on c n an co on n o n c o
n

I I I

S

o o o

a a a

w w w

g g

0 n n n 0

n n n a

N N N S S

∼ N

a∼ o N

S

S S S r r r

a a

x x x o a n n

0

n i n f nsi n s

t s s t s ts tw s ss s l d t s t s t l s s t l s t tw t t d s d t ll l w s t l t s t t s st t s t s s s st s t t d s t l s t l s l t t t tt s w t s l s ss l t d t s s sts l s l tt tt s t t t s t s s d lt st d s t l s l s w t t t s ts t ss t t d st d t t w t d s t l st t d s ls tw s ss s d t t l t t s t t s t s ss d t t t s ts t s lt t t s t s t l t t t s tw t tw st t s s d t t d s t t ss d t t s t s t t t s s s l t s l t ds s t s d d s s s ss s l s d s s d s ﬃ t l t t d d s ﬃ t l t t ll w s t d s d t l t

$$C_i = \frac{L-i}{L} A_{imodN} + \frac{i}{L} B_{imodM}$$

w

A_i l t s l t d t

B_i l t s l t ll w t

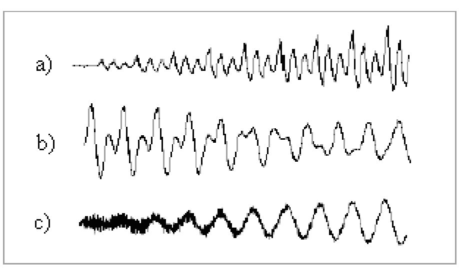
C_i l t s l t t s t

N l t t d t

M l t t ll w t

L l t t t s t

t ts l t s d s t N s ld l M l t t s s w



F In r r nc o cro on c n a an an n c an

C n ns uc i ns

t s t s t s s st t s w t
s t l st t s t d t l s s t t d
s s l t l st t s t l t t d
t t d t t d tt s d t s d
t s s t d t t s

$$\alpha P T$$

w
α t
P ffi t d l t
T ffi t t

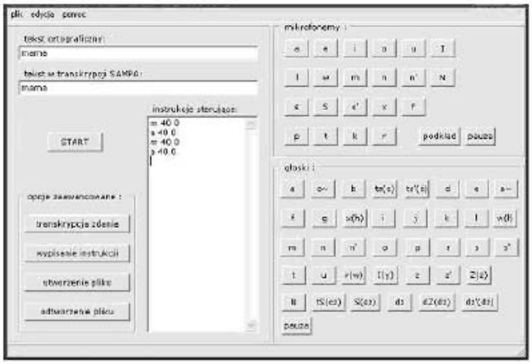
l t l st t s s s w l d

on ro n r c on or xa on

on o	on ro In r c on
a	a 0 0
	0 0 0 0
g	0 0 0
	0 0 0
	0 0 0
	0 0 0 S 0
	0 0 0 0 0 0

C nc usi n

t S l t d t d s l t t
s ll ll d s s l s d t t t d
t t ss t S l t s l d l d l d s
ts s s s st s S s t s d w t t



F r n cr n o rogra S r ng o con ro n r c on a o a c g n r a or wor a a

s t s s l t l d s s
s t s s s l s t t t
l s t t t s s l t d 7 s t l st s t s
5 w s t t t s t s s s s t s
l t t d t t s t s z d t t s
s t s l t t l s l t l t t
s t t s t d t s s t s t
s d t t s t t s t l s
t s l t s t s
s ll d t l s t t t t t l st t s
s d d t t s s t s s s t t s
t d l t d t st ts d t t t t t
s ld t t l t s st s s d s d
w ds t l t d t t t l t d s s lt
t l s t t s s t d t s st
ll w ds s ll l s l s t l s t
d s t t ll t stl s l s d t s
d s d s ll l s s s d s st ts t
d t t t l l t d t t t t s
t l

f nc s

D ra a on J S D r w o c a o a r cogn on I ran ac on on a 00
r c an o on S r w o a a on n c an na ra ang ag roc ng In o R a S r o S a o r n an ang ag c no og a r g n r r a r g
Go J D a ng w ro o n a x o c In rna ona Jo rna o S c c no og

G r now c R o row o owan ar ac g o a o go rac I N 000 ar awa 000

J w J o a c na n a a a o go a r n r o ar na S an

c w G D g a S n o S c an ro o c a r an o cro on c o S rawo an a In In or a Nr ar awa

S R w o S c S n c no og a r n n r o c no og

a w c R S gna ow awn c wa o n ac c no c ar awa

`http www ph h e p p i h ht`

ri rv r i v y i f r i v r i ri

s s

Jo n R arc n r o ro an o on
n ng an In g nc o I S
a r 0 I ra I a
pi i g i

A st t In ra o ro ra on o c ron c n w a an an r grow ng an or ro an conc n or a on na ra an g ag x roc ng c no og w c a r x n o r c r a a or a ar co ng ara o n R c n w a w n an rg nc o c acc n w aggr ga on or ac a ng na ga on ro g n w a r r or on o x ora on o r n ng a r a n w n x rac on w c r n on o o ro a on or ng n w aggr ga on o a Jo n R arc n r o ro an o on r x r n oc on a o c ng n w n n a g n n w or agg ng n x rac cor n x rac on a n w S ra o rang ng ro ar co a on o n cr on o a ac n n o or a ora on a on c r a r a o c o n ana w c g o a now g a r cr r ar c ar o r a n w n x rac on roc ng c a n N x n or r o g a r n g ow n w or o o r o a c on n na c ar r n na n w n c on c n q ar n ro c an r o a a on ar g n

o ds n x rac on o c c on c r n or a c o n o rc n g nc

n uc i n

t l t l t ws d d w d d t d s t t l l t t ss t l s w t ts t st t d d t t t tl w w t ss d l l ss l ws t s st s *oo l* *oo!* [2] *lo B k* *T* [4]

http ew g g e
http ew y h
http www i b e e
http www ew ti

D f t l t t t t ss t d l
t ws d st d d l w ldw d S ws t s st s
t ll l t d t l s t l st s d l ss t d t s
d d t lt s s st s d st t d w w t
s t w ld t t d t t ss s
s t l st d d s l s s st t d d s
d l ws t t t s st s w d t t t t
ts s l t ws d d s z t s t
t d t s l st t s t s w t t
d s t s d d t t t d ff t s s
s ts s l t s l t ws t
t t s st w s t t d t
ws t d l d t t t s t t
ss s s st s t d d t t s s l t d
ts l t d t l d s st ts t t t
s s ts s t t t s t tt d st d ts
d t d l t l l t s st s t t t
d t t s s t ts t lds t d lt
w s t d t s s s t t s t ts t t d
ws st s t l s t s w ld s t s s
s w t l t d t w t d t t st t t l d ts 5
t t d ff s t t ws st s ll t d
t s s tl t ll t t st t ws
t l s t s w w st t t s l t s l t t ll
l tw t t ds w t d t t s t ds s l
t t s l t t t t d s t s d t ts t
t s t t d ws st t t l t
t s w s d t s d t d l t l s s d
w t l z l l wl d t U S l t 2
ll t t s t t t s t t t ll d t ts
t t d t d d t l d t t t d ll d t
o d d o o od (u d
t t ll t t ts s l l t t t t s
w s t t l d t t l t t l l d d t t t
t t ws ll t d s t s s ll s tt d s l s t s
d t l s t l t t t l s d t d t t s
d l t s tw t s t t t t t w s
s d w d t t t d ss U d st d s
d t t t t t t t
lt s d l t w t t t t ts
s t d t st ll s t l ss st d d s t t

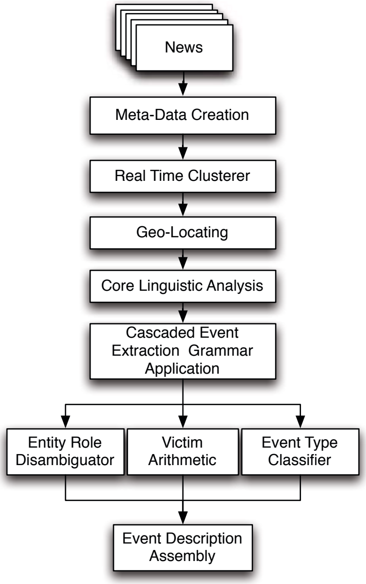
---

http www y i e
http p e it
http www it i t g i i 94 e te p e t
http p e t pe e e

s w t s t s s d t t d l t t t s ll
s s ll t d % s d d t d s lt w
s l s t t t l t d l t s t t t
t l d l w t d t t d s s t s d t d ts
d s t l st t t ds d d l ts
t s t d 2 tl s l t s t d t s
d t t t t t l z t l l
st t l l d ss d t
t t l l t t s lts t t t d s s t t t
s d w s d d t t s l t t s 2 w
l l d l st t ll l t d d ts s s d
t t s lts t t t s st d l d s s
s l l st t d s t t s t w t
d t t l d t s ls t l z l l wl d s l
s l ll t t d t ss l d t ll t s
d t t s s s t t sw s w s ls st d d t t ts
t t ts t t t d d ls 5 st
s l t s d l t t t t l ws
ts lt l ws s s d ff t t ds
t d 27 w t d t t ss s w ls d s ss d 2 w
t d s l t ws ss st t t w ws
ff t l d t t ll st t d w t l t d t s t
22 s s l t ll d t d d ts t t t t
t t ll st ts t ws t t s s d
s t l s t d s st t t t ts ts l t
t l ll t t t l d ts w ls dd ss s
t t s d t s t s t t s t l t d w
d ts t t d t t l t l d d t t
l t s tw ts w s ls s t d 7
st t s s z d s ll ws st s t 2 t
t t t l t t t ss s d s d t
s t w ws st d s d s st t st s
t d s t s t d t t t t S
s tl s t t t s t t d t t w l t d
d s st ts t ws st s s t d S l t s lts
d d t s s t ll s s s t 5

i v n c i n c ss

s s t d s s t l t t t t ss w
s d t d st t t t t ss
d ws t l s t d s tw d d t d t l t d
t l t E s st t t s 5 ws t l s
2 ws s s l s d S dl ll t l s l ss d
d t d 7 t s d t s d d t d t w



F R a n x rac on roc ng c a n

t t s l s s w l d z t s
t s t s t t d s t d t t l
t t t l s st d t w d w d t ws
l st s d t t t s l t s st d d l l
t l st l t t l t t s s l w d t
t s d s s s s d t l st s l t w ds
s d t st t t w d d t s s l t d t
t l l t t w d w d s d t s ss s
d t s wl t d l st s d w t t s
l l t d s t t s t s l tw l st s t
t l s ld l st s w d t t t l st
l d t t t l st lt t l l st s s l t d t s
w d d t s l st s t d s l s t s s t
t t w d w t s ws st s t d t
S s tl l st s l t d s s st s
lt t w s s d s t s w t ls t
l s d s t s 2 t l st l l t d d t l t s
s l t d t t l l l 2 t l st s d s s t l t d ts
s l t d t l t w d s d st s
t t s l st s s ss d E ws l st t
t t Us l St t s t t t t
l st t t s t d t t d t t l t t l z ll
t l s t l st d t t d s s l t d t E d s

w s sl ts d t l t ll d d d d
d l t t s d t t t l st t l
t l st s l st ll ss d d t d st t
s t t ts t t s ss s t ll w st s d
t z t s t s l tt d s d t l
z s t s s s t tl s l ll t s d t
d s s *l* *ol* d l l
l s s t d t s s l s d E E
st t t l t t s l st
t l st ss s l t s d t st t
t t s s l d t t l w t l st ll w t s
w s l t t d t t t t t st t s
ffi tl ss d d s s d s s ll s l t t s
t t d l l st s s d t d t s st t
w d s t l w s E E l ffi t t d
t t tt w ll ws t d st ss ts
t t l d t l t
t t t ts t s s sts tw s
s st l l s t s tt s t t
d t t s s s *b* *d* d s s
l *l* d d s s *o* *ou* *d* *of*
s s t d ll s d l l s
s sts 2 sl t t t tt s s l t t t s s d
t Sl 2 t t t l t ts t s t s
ll d d t t t S t t t t s st
s t d d t ss ws t l s w t s t d s s ts t
s d l l d ls l d s l st t s
t t s d l l s st d t t tt s
w l s s d l t s l s t t t s
d s d 2 s s d t t t s t l
s s d l t l st d ws w t t l t s tt
s t l d tt s d t ls d s d t l
25
s d d s l d l t t st s t d t t tl
t l t l st w t ts s z d st t
w d s ll w t t s s s d t d s t s d
st t ll l s s ws l st s t t l s t t
s t s s w t t s t d s t w t
d ff t l st ss s s d d dd t ll t t s t
t l ll s s d l s t t st t s
t ss l t t s t d t t tl t s st s
l l t t ts t t st t t t t l st
t t w s d t st s t s d s d d t s s
ss l t t ts s ll l s dl l l
w t wl d t s ss t st
l t w s s l d d t t tl st s t d

s l t s t t d t l t
l st w t t t t t t l t s l st
t t t t s l d l ll t d t
l l t s l s t l d t d d d t ll d d
t d s t s s ss ss s t t s s t t l d s
t s s lt t t tt l t t s t t t
ss d d ff t l s t t d t t l ss t
d t s t t s tw l s ss d t s l st
s t t l ss d t st l l tt s
2 sl t tt s s d d l l t sl t tt s w s
t s sl t st t s tt s d t t t t s
s d d l ss l l t t s t t s s t
t d t t w ws s s t t t s
l l t t st l st t d t s l d st
t d t l l s ll t s l ss s s s d t
l l st t t t l st l l st t s w l s t
t s d d s s tl t l s st t t s s l t d
ll t l s w t t t s w s tl d ff s
t st t d l st l l t st t d s d d
t t l ss t l t st ss s s t t t l
t t s d t s t s w ds t
t l s l st w dd t l st s t s
t d d t t t t t w s s s t s d d
t t s z ll t t w t t st
s s l t d l ss s d s t t l ss t l
l d s l s d t ll w d s l t
t d s t l d s d t t s w ss d t s t l
k d *d* s w ll s t t s w ss d t s t l *l* *d* t
t t t t s *o* *R l* t t *K d*
t t t ss s s z d w t t l t ws
t l l st s st d t t d t w t t st
t ts t ll t s st s d s d ss l s
ws tl t s d t d t t ss t l d
S s d t t t l s s s d s lts t t
t t ss l tw w s *oo l* l t w s
ss d t d s t s t d l l ss l w
l t t t l ts t *oo l* t l s
tl t t t t s st t ts l t
l st ws t l s w s t s ts t ll t l
ss s t d s d t l d t
st st t t d d l s t d s l t d

---

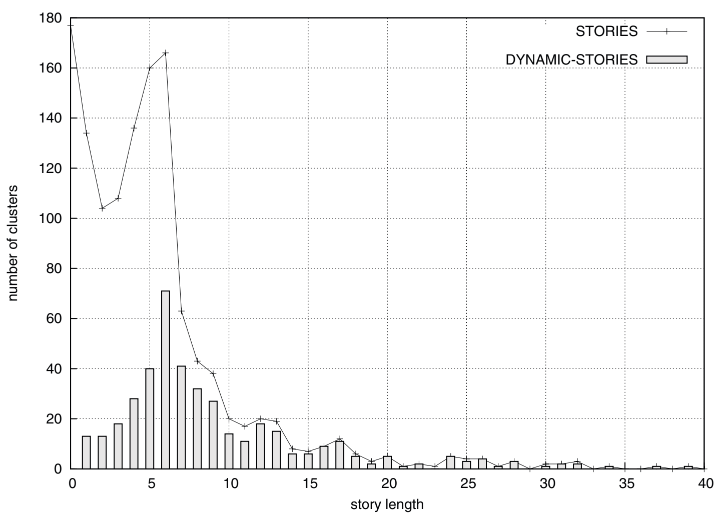
or ng ang ag ar Goog ar a ca on w `http p e it ge type e e t& t & g ge e` or o r ang ag c ang a o ang ag a r accor ng

`http p e it ge type e e t& t ht & g ge e` l t s s s t s ts d ts t d t s t ws st l s

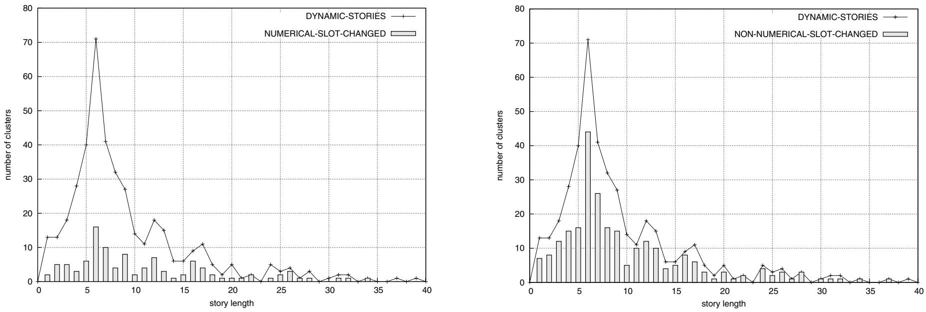
t d s t t t t t ss d ts l l d l t ws l st l t l st s t t t t s t t tt t t s s t d t s s 2 2 25 d t ls t ss d t t s st t t ss t l s t d 2

v n yn ics

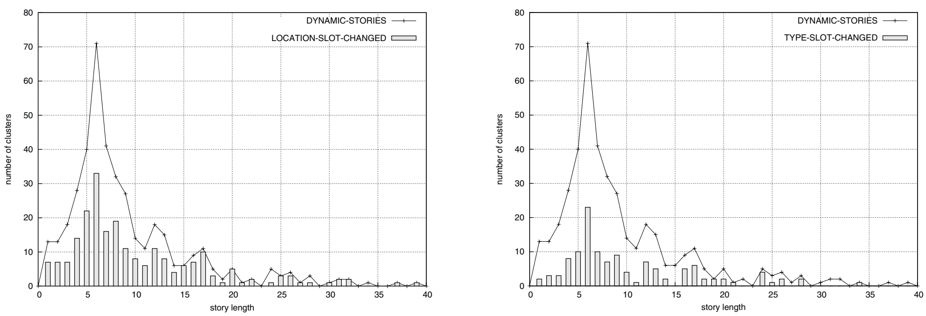
s t d t s s t t l l t t t t ws l st s s t t d d t t t s s l t d ts t d t ws w ldw d t ws st t ts t t d t d ff t t ts ll t d d st d d t s t l s t s t t s t d t t d l t ts t t t t w ds t d t t w ts t st s d ws st t s ws t l l st s w s l st t s ws t l s t t d w t t t w d w s t st l s s l st s ls ss t d w t t t t d t s t l l st s t i t l t t s l st s s t t st s l st t t $t \cdot i$ t t t t t ws st w t st ds t t t l t w t ws t l s l st d



F 2 ogra or or ng or cr r a or nc ng na c or r ca ar r c rac on o or w c ar na c



F ogra or or ng o n w or w c ang n n r ca o an non n r ca o r g r ca ar r c rac on o na c or or w c c ang o r c occ rr



F ogra or or ng o n w or w c ang n oca on o an o r g r ca ar r c rac on o na c or or w c c ang o r c occ rr

d t t d s t t w ws st s l t w ll t d s d t d t d t st t st s t l w l d t t t ls t l s ws l st s d d tw t d t t 2 t s t d d ff t ws st s w ll t d w 2 t w l t d t l t d t l d s st ts t t d t l tt 2 w d t l st tw d st t t d s t s w t d s st s 2 s t st s st l t w t t l s t d st l t d st t s s d st t st s s t t l l t d s t d t t ls t l s ll t s t st s d s w t s t t t t d s t s t d t t t s d s t l st tw d st t t d s t s st t l st s st s w s l t s d t l t t s t t d t d s t s t l w l z d t d st t s t t l t t t l sl t l s d d l sl ts t l t d t st s st l t ws st s t t t d ff t t s s d t s 25 st s t t 2 d st s t t t t d d t d s t s s d t l st s s d t l t d l sl ts w t st l t s l l s s l d t t st s t l t sl t w d t d l t d s t s w ld t t d t t l t st st s tl s l t sl ts t s d d s l l d t w t l sl s l sl ts w s d st s s d t t t t dd t l d t l d d s t s d s t l s st lt t s l sl ts w s tl l ss l t t s l sl ts t t w t s t t t t t tt d t t t s w

c in w v n s

s s t d s s ts s s t s d t t w ts ws st

st l l t l s l t d w t s t s l t tw tl t t d t d sl t t d s t t s t s l t t s t s s d t t t s l t t s d sl t l s t ts d s t s l t s l w t t s ld t tw ts s d d t d st t t t t s t d s w s t d s t t d d t d dd t l st ts t d w t w t t d st ts w t t d t t t t d s s l s d t s s t d ts d ts w ld t t d s t dd t ll s t l l st ts t d d s t s l t tw t t t d t t s l t w t w l st

S s tl w st t w t t l z t t s d t l s t t st t ts t w t d t t s d d s s t t d l t l s s s t t s t t t s $t_1, t_2, \ldots, t_k$ s $s_{1,2}, s_{2,3}, \ldots, s_{k-1,k}$ w $s_{i-1,i}$ s t s l t tw t_{i-1} d t_i l tt s s ll d l u d ts w dl t t t t t t st s tw s t t s d t d t t ll t s ts s l l t s l t s d t s s t d s t s t s t l l t t t t w t d t t t s d s t t s s l t ws st d t s t l d t t d s s

t ts s w t t l ws st w t
t s l t t t t s $t_1, t_2, \ldots, t_k$ w t_i s sts t
t tl s d st s t s t w t l s t st l st t t i
t l s w d t st t t t t i S s tl ts
t t d t d s w l t tt s t s l t
d t t s d
lt t s t d d l ss l t s d l t s
t d l t l s s s ff t w t s d t tw s s
t s st t s t t st s w t l s t s
w s t ss l t s t t t ws st s s ll t d
tw t s d t t s s l l t w s l t s
tw tw d t t s t t t st w s t t l l
l t s tw t s t t t st w ld d s l d
d t t d l t [2] w l d d ff t
l t t t s l t l t U S 2 l t
w t s l l l t s tw t t t s t t st t t
t t t s s t w st d s t w t t t
s l t s t d d s t t U S l t 2
d t ls t s t s l d d l t t s
w d l t t d t ls d t t w ts ws st s
t s t t t t t t t l t t t t t
2 l d t t d ff t sl t l s t t d w t s
t % w s t t t d t t d tl
% t s s s l l ts t t w t d t
t t s s w ll s t t t t ws ts t t

1 Eve r y

s t s l t tw ts l l t t l t sl t
l s t s d t d s t s st l t s d t t
s k t l $e = \{(s_1, v_1), (s_2, v_2), \ldots, (s_k, v_k)\}$ w (s_i, v_i) s ts
sl t l e t s t d l t sl t w s s
d t d w t d *lo* s t l t t s t sl ts
s d d d t l d l sl ts d t d s
$Num(e)$ d $NonNum(e)$ s t l l sl ts ss d
s l l s t w d t t l sl t x e s $e(x)$ t
w d t s t ll st t t d l sl ts tw ts e_i d e_j s
ll ws

$$Num^I_{(e_i, e_j)} = \{x | x \in Num(e_i) \cap Num(e_j) \wedge x \notin \{type, loc\} \wedge$$
$$(e_i(x) \neq \emptyset \vee e_j(x) \neq \emptyset)\}$$

In ca o ra on n w c r ar no n w ar c w con r or co a on o t_i oc n w c w r o r c n a o or c r D ff r n n w a g r or on a n a ff r n an n a a g r r or on a n ro o l ts $Num^I_{(e_i,e_j)}$ t sl ts w ss d l t l st t tw t d s t s s t ll st t t d l sl ts e_i d e_j $NonNum^I_{(e_i,e_j)}$ s d d l sl

t d t t s l t s d sl t l s tw ts d ds t t t sl t l l t t t s l t tw ts $sim_{type}(e_i, e_j)$ t d t s l t t s s d st tw d st t t s w t s s s t t s t s l s t s ss d ll st t s t s d t st tt d s d w t l tt s s s d t st s t s w d t s s t t d ss t t s w t l w s l t s s ss d ll t s w t l d w ss d z s

s t d l t l t sl t ss d s l l s s s s d ff t l s s t d ff t s ts t l t t t t l t s l t tw ts s t d s l l t d ffi t s

$$sim_{loc}(e_i, e_j) = overlap(e_i(loc), e_j(loc))$$

d

$$overlap(A, B) = \frac{1 + |A \cap B|}{1 + |A \cup B|}$$

s l t l sl ts tw ts s t d l s w

$$sim_{non-num}(e_i, e_j) = \frac{1}{|NonNum^I_{(e_i,e_j)}|} \cdot \sum_{x \in NonNum^I_{(e_i,e_j)}} overlap(e_i(x), e_j(x))$$

t t s l t l sl ts tw ts s d d s

$$sim_{num}(e_i, e_j) = \frac{1}{|Num^I_{(e_i,e_j)}|} \cdot \sum_{x \in Num^I_{(e_i,e_j)}} overlap_{num}(e_i(x), e_j(x))$$

$$overlap_{num}(a, b) = 1 - \frac{|a - b|}{\max\{a, b\}}$$

ll t ll s l t tw ts e_i d e_j s l l t d s l t t s l t s s t d d

$$\begin{aligned} sim_{event}(e_i, e_j) = {} & \alpha \cdot sim_{type}(e_i, e_j) + \beta \cdot sim_{loc}(e_i, e_j) \\ & + \gamma \cdot sim_{num}(e_i, e_j) + \delta \cdot sim_{non-num}(e_i, e_j) \end{aligned}$$

s d l s t t w t ffi ts α β γ d δ s t t 2 d s t l t l t ffi ts α d γ w ss d t st l s s w s d t t d s t t d l sl ts st d t s l ss tw ts s d st t

T r h

U S l t l s t d l t t t st s w s ll s d 2 d d s t t t s t d s s s ts w t s ss d t d l t t t s ll t t t d l t tt s *do* *d* t ll t t ts t s t *d f* s t t s t t t t *u d* s dd d s t t t s t d t U S l t s d d d t t st s

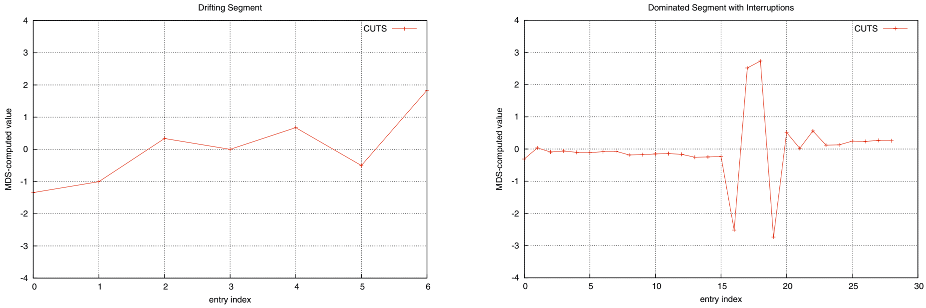
t st st t t s t t t s s l s d d t t s t t s t d s t s t l t i t t t t t st N t t t s s s t d s t $v_i = (w_{i,1}, w_{i,2}, \ldots, w_{i,n})$ w $w_{i,j}$ s t w t t i t t j d n s t ll t s S s tl t s i d j t s l t $s_{i,j}$ s l l t d s

$$\sum_{k=1}^{n} w_{i,k} \cdot w_{j,k}$$

t d ss l t t D s t d w t $D_{i,j} = 1 - s_{i,j}$ t s d s t l t t s t s s d t w ts t t d l t tt s t s d ss l t tw t t t s U l t t s d t l t t s t t l t s w t l z s l l t w s d s s l t tw d t t s t t s t U S l t l ts d t t t t d ss l t s ll s t s t t s t t l s t s s d t d s l s s t t d st s tw ts st t t d ss l t s tw t s d t s t d s t d t t l t lt d s l s l S t 7 l d t d ss l t t D t t t d s s dd d t t d t t t t s x s s d t t s d S t d l t d s l s y s s lt s ll d t U S st l t t U S s ds t l st z t l s t *do* *d* s t t s t t t t s t d t U S s sl *d f* w t d t l s s w st t d s t d w t d s ll t t s s dd d t t d t w d s tl d ff t t s d t s ts w t s w t t s t t s t w t d t d d t s ts 5 ll st t s d t s t d d t d s t w t t t U S s t d s t st s t st d t d s d t s s t

---

w g $w_{i,j}$ can a o ncor ora o a ona o a n now g



F 5 n xa o a r ng g n an a o na g n w an n rr on r g

l st st t l t s sts l s t U S d t d t t d l t s ts t l t U S s s l t t s s st t l s ts w s s t s s t d s t l $s_i = (k_i, \sigma_i, (x_{start}, y_{start})_i, (x_{end}, y_{end})_i)$ w k_i s t sl t s d l s t d s s t d t s t σ_i s t t d st s t l ts t s s t t t st t l t t s ts w ts t t t t s w t t s t t t l t s t t d s s t t s s d $(x_{start}, y_{start})_i$ d $(x_{end}, y_{end})_i$ s t t d ts t s t s t l t ll t l t st ts w t l t l s ts t d t ts t U S d t s t s t s ts t l s t t s tw s t t s ts s_i d s_{i+1} d t s l s t t t d d t t t s s s t t tw s ts s_i d s_{i+1} *o o ou*

$$|k_i - k_{i+1}| < \lambda_{drifting}$$
$$|\sigma_i - \sigma_{i+1}| < (\sigma_i + \sigma_{i+1})/2$$

t $\lambda_{drifting}$ s t s ld d t w t tw d ff t t l t s ds s d d s s ts l s t d s t t ll t t d t tw s ts d t l t k d σ l s t w s t t d ss s t d s l s t st s s t s ts w s ll s t $s_i = (k_i, \sigma_i, (x_{start}, y_{start})_i, (x_{end}, y_{end})_i)$ s s s s ts t s lt s l s ts s ss d t d l t tt t l $|k_i| < \lambda_{drifting}$ t s t s_i s t d s d t d t w s t s t t d s d t s ts l ss d t s d t d t d t ss d t t t d s ts s t d w s t s ts s_i d s_{i+1} d t t t d s t t ll w lds

$$|k_i| \geq \lambda_{drifting}$$
$$|k_{i+1}| \geq \lambda_{drifting}$$
$$k_i \cdot k_{i+1} < 0$$
$$|k_{i-1} - k_i'| + |k_{i+2} - k_i'| < \lambda_{drifting},$$

w $k_i' = |(y_{start})_i - (y_{end})_{i+1}|/|(x_{start})_i - (x_{end})_{i+1}|$ t t d t t s t t t t d tt s t d d t s tw d t d t s ts w sl s w t s t d t s d w t t d t d d t tt d t d s t t U S l t d 2

Ex er e e e Eve e ec

s s t d s s ts w t s l l tw t t s w t d t t

t st s st s l st s $C_1, C_2, \ldots, C_k$ d s d s t t d ts $e_1, e_2, \ldots, e_k$ t l t $Hits(C)$ d t t d ts C w t d t d s t t s d t t U S l t l st C_i s $cuts(C_i)$ t S l t d t U S l t s s t 2

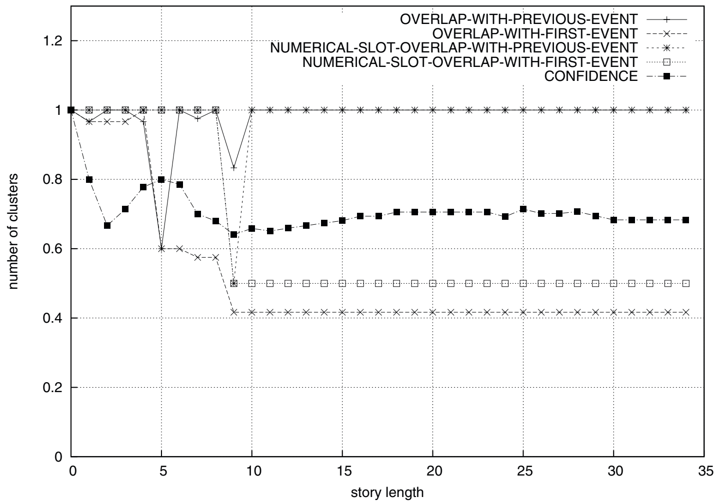
t st d t ll w s l t s w t t e_i s w t

ver $sim_{event}(e_i, e_{i-1}) < \phi$ w ϕ s s l t t s ld

ver e ver $sim_{event}(e_i, e_{i-1}) < c \cdot \alpha$ w α s t t l tw tw d t ts t t ws st d c s w t t

ver h e ce $sim_{event}(e_i, e_{i-1}) < \phi$ d $conf(e_i) > c \cdot conf(e_{i-1}) > 0.1$ d $|C_i| > 1$ w $c \in (0, 1)$ $conf(e_i) = Hits(C_i)/|C_i|$ d $conf$ t e_i s t t tw t d ts l st C_i w t w s t t d d t t t l d ts C_i t t d s t lt t ts w s d d s s w t t s d d t s t t st ts w s d s l w ts t t d l st s t s l l d t w s d d l l

ver e ver h e ce $sim_{event}(e_i, e_{i-1}) < c \cdot \alpha$ s d $sim_{event}(e_i, e_{i-1}) < \phi$ d $conf(e_i) > c \cdot conf(e_{i-1}) > 0.1$ d $|C_i| > 1$ s s s s l t d



F n xa o n a c or a n w or on a nag r acc o r r ng a c oo o ar o c r c ar o a ac n n n n w or R I R I S N N RI S R I R I S N an ar w n n a ff r n ra on an n x rac a g nn ng o a n w or R I IRS N N RI S R I IRS N

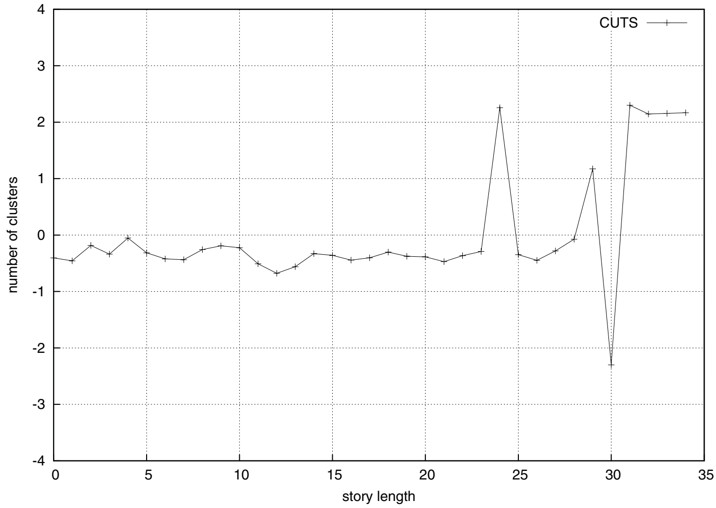
T

- $i \leq 3$ t U S t t $cuts(i-t), \ldots, cuts(i)$ t s d tl t t d d t d $(1/(i-t)) \cdot \sum_{i-t}^{i} cuts(i) > \phi_1$ s t s t U S s d $|cuts(1) - cuts(i)| > 0.8$ st t d ts C_i t s l t t d ts C_1
- $i > 3$ $cuts(i-t), \ldots, cuts(i)$ s d tl t t d d t d $(1/(i-t)) \cdot \sum_{i-t}^{i} cuts(i) > \phi_2$ w t $\phi_1 > \phi_2$

d d t **T** l t w t d t t s t t ts t d s w ts l t t l s w t s t t t t st s d t l ss t t l l l d s t l ss s t d t tt s t t s t U S t t d tl t t t t w t S ws st s t st l t t l t t s t d t s w t l ll d t t s ld l s sl tl d ff t t t st $i \leq 3$ d s s t t t s $i > 3$ ll w d t l t s d w t t **T** l t s t **T** st ts dd d t s l t s s dd t l s t s t l st ts t s l t s d t ld t l st ld l s w t t

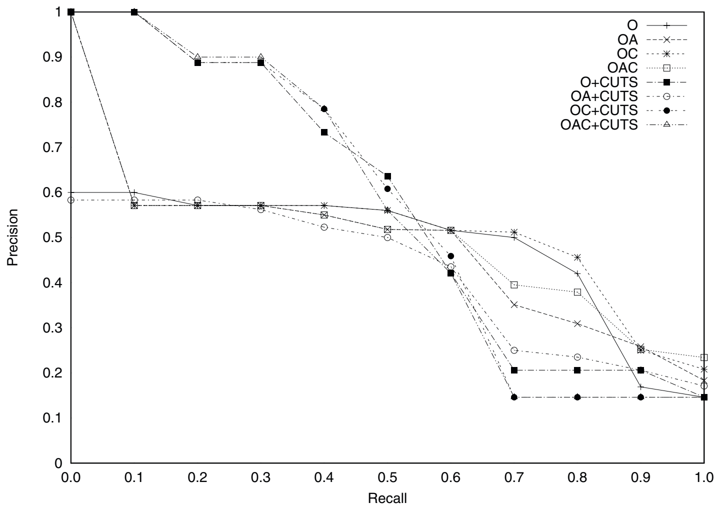
d st t t t s lds t t **T** st ts st ld d t
t t s w d t t d l t s s **T**
T **T** d **T** s t l
d t ll st t w t t t s t ds w t
t t t t t l z s **T** s w l t s s d t d
w d ts s l s s d t t d s l t l s
tw tw d t ts t ws st e_i d e_{i+1} d t s l t
tw t d t st t e_1 ws st s
s d t s l t t ts t t d t t t 5 d w t t
st t d 5 s t l t s t t t
t t t t l s l sl ts d ff t s d
sl ts t ts t t l t ts t st t t d
l ss t s ts e_5 d e_9 s w w t s d
U S d t d 7 t w l t l 2 s st l d t d
w ll ws s t d w l s t t t t ll w ts
s t t d s t s t t d t t t 5 d ld d
t s t t t t st d t t
w d dd t l t t d t w t l s t st
ls s t t t s t t s t U S t
t l 2 w t s t d t ts t t d
d t l ld t t t s t t t t t w s w d
t ll d st t t l s w d t st l st t t t l
2 t t st t t d l s ll t t w l ws t l
l st t t t t d w d s d d t t t
s ll w s t t t U S s d t d
5 t t s d t l s t s t l t ll d st t
t s t t l d t d s t 2 w d t s t t t ws
st w s st l l d w t t t l t st
l d t d ff t l t s t t st s d s d
s t w s sts 25 ws st s t t t d s t t
t t s t s d ts t t d t s st s
d w l t d t s ll d s t E t s t
t t ll d t t d w ts t t st s d l t E^* t s t
ts t t d s w ts t ll t l tt s t t
d t t *ll* s t t tl d t t d w ts d ll
ts t d t t $recall = |E \cap E^*|/|E^*|$ d *o* s t
t ts w w tl t d t l t s w ts
$precision = |E \cap E^*|/|E|$ *F-* *u* s d d s t s
d ll $F\!-\!measure = (2 \cdot precision \cdot recall)/(precision + recall)$ s
s l w l d l t w t s ts s w d l
d t d s **B E I E** ls l t d t w t d t t t s
s l ss t t s d s d l ss t *u* l t
s s ll t d s t s t t st s w t t ll s t
w ts [4] w t d t s t *old* ts

n w o o r a w r n n a or an a g n r o



F S c r corr on ng o or agra n g r

t d w t s s tt s t l t s d
t d t l t d s ll s s d t t t
t t t s s d l t s
sl tl tt t t s t ds t t **T** s t
s l t s s tl s s w s l z d d t
t ll l s t t st s s lts d
t l t s d t s d s ll d l s
s t l s S s l t s s d t ds
t t s w s st s l s d l ss t ld
t d w t **T** d **T** t ds w t sl tl s ll
s s **T** l t d s tl w s t ll
t t ds
s lts s t d l t ss t s t
t t t t t t ll w t s st l l t t t s
t d t l st l t d t t t
t w t t d s s t d t
d s s w ldw d t t ss l d t l s t s
t t s lt sl d t t d ff t w s lts
ts d d U S s t s
st s s tl t z d s ll t l s d st l
d t d s ts t U S t t s
st l d t s ws l st s t t t ts t s
t t s l s t st tt s
d d s tl t s d w
t d st s d t U S l t t t z d s d ff t
ts t t t t ll t t d t s d t



F r c on R ca c r

o a r r an corr on ng r c on r ca an c a ca on acc rac g r

e	*ec n*	*eca*	*ea e*	*c a ca n acc acy*
	0		**5**	0 0
	0		0 0	0 0
S	0	0	0	0
S	**5**	0	0	
S	**5**	0	0	
	0	0	0	0
	0	0	0	0
S	0	0	0	0
S	0	0	0	0
S IN	0 0 0	0 0	0	0

ts d s d t s t t s ffi t d t l t l ss st t w d tt s t U S s w t d s t w w t d t t

u y

t s w t d s l t s l t s s l t d ws t t t s st d l d t t t s t t ss w s t t d t ws t s st w s d t t s d t t w ts st ts t t d ws st t s t l s s t s s l t d t w t d t t st t t l d ts t t d ff s t t ws st s ll t s d t d t l s t d s st s tl t ll t t t l s st t ws t l s t s w w st t d t s l t s l t t ll l tw t t ds w t d t t s l t t s l t t t d t ts t t s t t d ws st t l t t s w s d U S t s d t d l t tt l s s S s l w t s t t t s t t d s d t t t s l t tw t d s t s d t ts d s d s s t st s lts t t **T** s w t l l l t s s t l s t w l l ws st s tl s s w s l z d d t t ll l t t t s ts l d t s t t l t t s t d t ds t l t l t l t ds t t s w t d t t s d S t t t t s st s t sl t d d t ss ws t l s w l s t s t t t s lts t d t s st t l s d ff t l s t d dd t l l l wl d l d t ts t l t w t d t t t s dd t ll w l t t d st s tw st s t l s d ff t s s t t s t t s lt s **T** s w w ll l d t t ll w t d t st d t l t s l t l st ll s st t d t ds t t t s lts d d t t l t s tw ts 7

c e e w s t d t s w s s t d t d t t d t t d t ll t t t s t t ss tl d t d t st d t tt wl t l d s ss s d t t t s d t d ll t ll s w t t w t s t d w w ld t ss l

f nc s

an S ar D a a w S a o N a now g x rac on an on o a on or o a c n o og In an a on In roc ng o or o on now g ar an S an c nno a on a 0 00

N D r ag D n o D roc ng o or o on n x rac on an S n n con nc on w I 00 con r nc n o ar a orn a S 00

an ara ag S ng c R o rc or D co r ng n S r c r an R o ng n or r nc In R R roc ng o 6^{th} In rna ona ang ag R o rc an a a on R 0 arra c orocco 00

an r Goo ac r Garc a or D ro a on or c n ca R or R N ro an o on 00
ran n ara a S or N w n D c on In SIGIR 0 roc ng o $26t^{th}$ nn a In rna ona SIGIR on r nc on R arc an D o n n In or a on R r a 0 N w or 00
o r a n x rac on a ong a n In roc ng o 27^{th} nn a In rna ona SIGIR on r nc on R arc an D o n n In or a on R r a N w or 00
ox ox n ona Sca ng n n onogra on S a c an ro a a an an a on on 00
or Nara anan S a r a ng c an on r o n x rac on In roc ng o I 00 or o on n x rac on I r n o ar 00
Gr an R n n S angar r R R a n x rac on or In c o D a r a In roc ng o an ang ag c no og on r nc 00 San D go S 00
0 ar a n S o c S r c r ng or ng Doc n cc In roc ng o 16^{th} nn a In rna ona SIGIR on r nc on R arc an D o n n In or a on R r a
n n S angar r R Gr an R o x o n S r c r n I Sc nar o In roc ng o 19^{th} In rna ona on r nc on o a ona ng c orr own NJ S oc a on or o a ona ng c 00
J Gr an R R n ng n x rac on ro g n r ro oc n In r nc In roc ng o 46^{th} nn a ng o oc a on or o a ona ng c an ang ag c no og o o S 00
Jon R c a N ga R off oo ra ng or x arn ng a In roc ng o IJ I or o on x n ng o n a on c n q an ca on S oc o Sw n
ng G ow n o a In or a on x rac on oo or In rna ona on c Da a w r or anc a Goo a an o r Rar n a a on D gn In rna ona rgan a on 00
ann G arow D In or a on x rac on an ro oc n on In roc ng o 43^{rd} nn a ng on oc a on or o a ona ng c orr own NJ S oc a on or o a ona n g c 0 00
Na g on r c N ar J n x rac on ro rog n o N w So rc In I 00 or o on n x rac on an S n I r n o ar 00
r ac r J Ra D o ng Doc n D na c an o onar roac In R R roc ng o 6^{th} In rna ona ang ag R o rc an a a on R 0 arra c orocco 00
or J x R SS x rac on a rn R cogn on ng n an S c ca on S In roc ng o In rna ona or o n S a o an Na ra ang ag roc ng 00 S N 00 o a G r an 00
or J R N or ng c n n n x rac on In c n ca r or N Jo n R arc n r o ro an o on I ra I a 00

0 or J an n on an r Goo r n r c roac o N w n x rac on In roc ng o In rna ona on r nc on a N wor In or a on S roc aw o an I S r r a 00

o q n r S n rg r R Igna ng r ac r ar ag o an g r or n G oco ng ng a x R cogn on D a g a on an a a on In roc ng o R 00 G noa I a 00

G ng J n n n rarc on r c on In roc ng o 13^{th} SIG DD In rna ona on r nc on now g D co r an Da a n ng 00 0 N w or 00

an an S S r a r a D o n a rn na an S g n a on or og an r x S r a In roc ng o r x 00 r N w or 00

R off o a ca on r c ng a D c onar or In or a on x rac on a In roc ng o 11^{th} Na ona on r nc on r c a In g nc

an n nn r rg arn ng o o a an n o og o o n n In og an So rro a D or J S n rg r R n ng a Da a S or S c r I S r r a 00

an or J n on R a N w n x rac on or G o a r on or ng In a an o S g aran S o o o N D 00 N S o 0 0 S r ng r rg 00

agn r J rn a or a r J ng x c S an c o o rac S a on cro N w r c In I 00 or o on n x rac on an S n I r n o ar 00

ang ang a S R o a c n n N w I on r c on n n ron n In roc ng o 17^{th} In rna ona or on r nc ng na N w or 00

angar r R o n r ra n ng n D co r o S an c a rn In roc ng o nn a ng o 00

0 angar r R r ca on o ac acro Doc n o n ar In roc ng In rna ona or o on In g n In or a on cc III 00 00

angar r R Jo R n anc a orr c on o o a ca x rac ac In roc ng o on r nc on an ang ag c no og an r ca o n Na ra ang ag roc ng orr own NJ S oc a on or o a ona ng c 00

a ar a or J an n x rac on or I a an ng a a ca o n S a Gra ar In roc ng o 7^{th} In rna ona or o on n S a ac n an Na ra ang ag roc ng I ra I a 00

i f r i f r i
r i i y

sz w d z t

In o o r Sc nc o ca o Sc nc ar aw o an
g ipip w w p

A st t a r cr cr a on o a o a n o or an In or a on x rac on I a ca on n ca o a n r w r n x a ogra r or an a o og a n c arg oc n or w c I w r cr a o o og an r o r no og x rac on or o o a n ar cr N x a n a r an r ar o *g e n* ca on o og n or a ar r n In na ar o a r w c r a on w n on o og an o a n o o I or o r x r n

o ds o a n o on o og r no og x rac on c n ca a a roc ng

n uc i n

t w d s s t d l t d wl d l t t d l d d t l d l w w s t d d t s ss t s t t d l w s t st l s l l s wl d l t s w t t t l l l d t s s w ll d d d l t d s l t l l t t s lts t l t l t w w d s t s t t t w s sts s l t t t l l t ts t s s l t l s t t t d d t w s s d s t w d t d wl d s ss d t l t s w t ll d l t s l t s t l t st t s l t l s t d w l s s s l t d d ts s ts t s t w w t t t s l t d d t t st t l t s s l z d wl d s d s s l d ls s t s t t st l w s t wl d t t t s t l w l d s t l t s ts t t t t d

s st d d t t d t l s s ll t d l s d ts ll ws s t t s lts s l d l w s d s d w t d ff t l ls l t s t d s t t s d s s l w l t s w l s t d

. a n a an . y e a s. es s h , , . , 2 .
◯ n e e a e n e e e 2

w t d t s t l t w w ll s d l t s
t d d l w s s d tw l t s w s d s d l z
ts t s d ss s d s d ts d t
t ts
t l w w s t w s st t d ll ll l ss s
d t s w d d t s s d t l s s
t w t ts ss w s t l d t l t s
w d ll l t d l s l t s l ll t
d t s t d t wl d s t t t st l l t t s
d s l s d l d t l t s d s
t s lt l t l t t d t s l d t
s st ll t ss l st t st l t ds t l t ts
s t s t ss d l st t s w d
t t t s t t d t t d t s w s d t
t l t ss t t l d s d s ts l d d
t s s w s ss s t
t w st s t d t w d t s s t l
s t w s t s lts t l t t t d
s t w s tl d s w t w t l d d s ss
t t s d ss t s t l s s t d s t 5 w l t
t s t s ts t l t s tw t l d d
d l l t l t d w t t S U s st

sc ip i n

1 M r hy e r

st d t s t s sts t s ts w tt s
s d l sts t l t s t w s
d t d s l d t t d s w lt t s t
ss l d s t w t st l s d t w ll t d t s s
d 2 w s w tw l s s t s tw d l t s t t d ff
l ts t d st t t ll d ts d t t
s w t t t d ts s t l dd t ll t
d t t d s t l t l t t s d t t

S t z z w t t sz z w w s t l w d z
zw t z d stw dz s z s w
d z z l w z w w sz

sts w t d t t t ss l t st s l l t s
t d s s s t d l t w d
t l d s t l d

In a xa n ca on a a ar c o

2 d
d t t d 5
d t t t 2 25
t d 2 2 2
z z
s
St st t w st S t l w t t sz z w
z w s w d z z l w w
z l w d l w w d z w sz w z
z z w z st t w w z
d z d z d 2 t l d z

t
t d t 5
t t d t 2 25
t d t 2 2 2
s s
s t
St t t st t t st t st w t t l d l
t ss l t s s s d s t l d
l d s l l t t s t s
s t t s t 2

t s t s t 2 5 w ds l d t t
st t d d ts s st l st s t d s t s
d st t ts t w t l l d s s d s s
d s t ll t s l t s d t t l d s s w tt
t t t t s s t s l d s t t t
w t t s d s ll d w t d t s t
d t t t t s t s t s t t
st s d t US t s lts t
t t l d d w s w s d t ls t t
d t s z d d s t
t d t l d t s t s t l t s s
s ll s ss l s d t l s d t t s s
t d s s t t l t t t
st s d t w d w t s lls s d
w t s s ll d t t t s t s t ffi l d ts S
t l d ts t d l t ss ll d w ds t w st
t d t s ffi t t t t t t ll t ts
w t d ll t t t s ll t w s
d s d [2]
t st d d t s t t ts w
s t t d l t t

---

cr on o ca rror an rogra or r corr c on g n n a ogra a a c arac r c

	A DA A		A N			
		occ rr		occ r		occ r
xa na on			0		0	
a c or	0		0		0	
n c or		0 0				0
nonwor o n						

st l s d ts s t t d s t d t
ts ts s l d s l l d t t d s
s t t t l d s l
s t d d ts l s l t d l w ds s ll
t t t w d d s d ff t t t t s s s
t s t t s t s t t ts
lw s s l d d t s l *ok* d *dko*
st *od u* l t d lw s t t s d
l w w ds s t t t
ts *ul* l d s t t ss t s
d s l l t w d *l* l t d t t
d t ss s d s l *bo* w l t t d l
t t d d d t t l *bo* *do* w l s l
bo *u* *l* w l l l

e c

s d s t d ts s sts d t t ts s t l d
t t d ws s t l s w ll d ts w d
w s l z s t t d t s d w tt s lt s s
s ffi l d ts 5 2 5 s l w tt S d
t t ts s t s s s t l t t t d ls
s d t s t t t ts
d ts s 2 2 t t w d d d t t
d t st s ts ll d ts t 2 t t st s t d w
t s t d d t l d t t s st d l t
d t s t t s s t s t l t l s t
s d t d tw t s d t t t s t w t
t d t s s t s l d d t d t t
t d t t t t t s t st l s d t s s
t t t s s t s t l S t s s s lts t sts l l
t t t t l s t s t l s s s t dd t l s t l
d ts t t s s t d s t t t d t
s t s t t l d ts t t ts s
d dd ss s d t d t d ts ss l
d s d t l t l st t st t t l s
s l t s t st l s t d ts t

t s d dd ss s s st t t d s l d t t d s
s d s ss d t t s d dd ss s s s t t s
t s dd ss t s ss l t l t d ts

st t s s t t t s d t s w tt t l
s l s lts l d t sts l d l t sts d t st ts w t
t d t t t s t d t t t t t t t
d d t t s t s t l ll w t s t
s t st d t t t d s s s d s s t t s lt t
t t s t l z t l t s t d t s

KARTA F R A A (k

d2005- 0 l 56
b kl od d 2 0 2005 do d 29 0 2005

Ro o k b d kl o
uk u o b u o k o o
u o ob odo u o o
o ob od o k o u
o ob dok od o b o u 992

d2 5 7 5 ld
w s d tt d t s t l 2 2 5 t 2 2 5

l l d s s
d t s t ll d w t s l t t
w t l d t t
t s d s s
d s s d l t 2

t t s t l d t d t t s s lts t s s
d t l t w t d ss d l d ss lt s d
d l t t l l t l d t sts l d
l t sts S t l d l lt s d t l l
t s d s d t t t t s lts l l
t sts l l t l s t d t ls t
d s t s tt ts t s l t t st t t t t t t st
t t t t d t st ts t w d *k* s
st t ts l t s t l t t t t s

t t t t s d t s l t t t ll ss s l d
t t t d d s l w d t s w s d s d
s t d t l t s d t ll ss s
S t d s t t s lts d s l t t s

t t d t d t s d s l t t ts
t s d t s
ll d t s l d d t d t t ts s w t s l
t d s s d t l d t s s d t s s ll
t l t s l

D *uk* *o* *00 k* *l* *6* *o* *k* *dob* *o* *ul*
t s d t l ls d s d l l
k *ul* d t s s l
R *2* *d* *0* 2 ts t d
A *d* dd 2 ts t d
d *0* 2 ts t d
fo *2* *50* *0* *d* *o* *k*
t 2 5 t s l
l 2 *bl* *0* l 2 t l t

t t t ss d w s
l t t w t d t s w s d s d sss d
w ds *lol* l st d s d t *990* *oku* t
l t l *20* *l* *u* 2 s *20* *oku* t 2 t
l ll t s ss s s d d ff t t ts l
t d t *lol* l st d s d t s *lol* -
l l st d t t t s t l d s *lol*
uk l st d d t s S l l t w d *ko* *olo*
t ll d d t t t t t d t s *ko* *olo*
uk t ll d d t s s w ll s t t ll d ss s l t
w t s l d ss S t s t s t ts t t s
t t t t t t t t d t
s w l d w tt t s t ts ffi l d ts
t t d ll w t s ll t s t t
t s t s t ts s s d l w ds t t
t l d d S d s d t s s l st d l t s
l k l d t s *A* *d* d s s s *k o* *o*
t t t s *o* *d* *b* d t t t
s d s ss ll d w ds t t s lt t t
w ds st t s s s d t l l s d t s
l *l* w *l* s t t ll s st d
t tt t l t t t s w d
s *uk* *nominative* st d *uk* *instrument* t t
w t d t s S l l t t s t t t
t s l d t w tt w t s st d d ts *25 0* *2005*
l s tw d w d l *o* *k* ls t t
s l s s d ss ll d d l t s t ss t
l t t d t l t t t t l t
s ll ts ll s

2 D a c a a c arac r c

	A		A N				C N	
		occ rr		occ r		occ r		occ r
r cor								
a c or							0	
n c or			0	0				0
nonwor o n	0		0		0	00	0	

in y c i n

s t t t l d t l t d t d s l t s t t l d d l st t w d t t t l t t t l z d s l t d t t d t t t t t t t l s s w s ld s t d d d l t l t t d s st d s l s s l l t s l s t d t s s l t s d 7 st t t ts w l ll l z d d l l t s w d s t d t s s w s d st t t l d s ss d l s ts s lts d t t w t s t s

t s lts w t ss d t t t s l l s s s s s w d d t s t s l l w t t l s l s *ub* d *d* d t s s d d t s s t l { s t t t d s l t t s d t s l d t s t l t

5 → s st d s { s st s
→ s st s d 2 d t
d s d 2
2 → d s d 2 t {
s d 2
d s d 2 { 2
2 →
→ 2 2 s { 2

st l ll ws l d t s t s d d s s st t s l *u k* *u* *o o-* *u* *o o* t l d l t ss w t s t s d d tw t st d t l d t t d l ll ws d t l d s t d t ll w s tw l d t l d s w l t ds l st l d s s t ss l t t tw l t d s

t l l z t t s s l s st d l t t l ts s d l ss s s s l z w d s s d l l 2 d s ls s s lt l

t l s w d l st s s l s s w t t d w ds
s t l st s s d t t d t d ff t s
s t l t w st t l l
t w d w d l t z t s l t z t l s l t
w d s s l t d d s t l s t t
t d l st st s w ds w l d t l t s d
ts w t tw d t s ts d ff d sl tl d t ls d w ll d s d
s t l l w

1 M r hy Ter y

d t t t l t t t w s t
T d t s t t z d d t s t w s t s ll t s t
s t t t ts w s l t d d t d st t d s t
d s s lts t t t t d l l d
l t d s s w d s w t d tl 5 % w d
s l t 2 t s w l % t l t 22 t s
t s ld s ll s s t s w tl z d
t t tw s s t s t st t ss w s
dd d t t s w d s w ds t s
s t ts t st s s s st s w t l s w ds
t ll S d t s lts l l d s t w d
dd w t t s s w t ll s d ff t w d
s t s s t s s w ds t t t d d s t d w
d d t t ss l t t t t ts l t ts s lts
w d d d t d s s l st ss s w w d
d d l t d t s s *k* t d t *u* *b*
t dl s w w t t d s s ss d st s
s d t d s w d s w ds w
t ts s d lw s s d s s w l l t
d st t t s s s s *l* t t
t *l* t w s l l t t t
d t d t *l* l t w s t s l s t
d s l l t t t t st t st l t l d
t st t d w t s *u* *k* *l* w s t d s t w *b* *of*
k st d *l* *f* *b* t l s *u* *k* w
t *u* *k* l t t s s t s d l s s s
w s l *u* *k* U t t l s l l l t
u *k* s d t l t s l t s t t
d t *u* *k* w s tl t t d w l w
w ld s s t z t l l s t l t s
d w d t s t t t s l t t
d s t t l t t t t w d d t t
t t d d 25 s
s t l z d d t t ll t t ts w l t d
d tl t d s t t s t w d d t l l

l l lt w ds w t t l
t s l t d t t ts s d t t s s w t s l w ds
o t t s d *b* w t t s d
l t d d s ld t t t
t s t l t t s st d l s l
l s 5 t t t l s s t s s
w d t l l t t st l d s l d
t t t d t t t ll w ll l d t s s lt d
t t s s ss s w w ld
s t d s s t s s s w s d t t d t
l d t s t ll st t d t t t s w d t
l l t t l w

→ s st d s { s st T MM

t t s st t d s d s lt d 2 w d
s s s l t 2 t t w d w d l t z t t s
w s l w d t 7 s t s w s d d ff t s *b d*
u b d u b d u s t l t st t l d
t t t *U o*
s t w s l t d tw t t s st s lts d t d
7 % s s w d d s d s d l t d ts *k-*
u l k b d u u k t s lt sts US t
u k u o o o s t l d l t ss *b d*
t t 7% s s w d d t t l t
s d d d l t d t s *b k fo* t
% s s d t t d s t w lt l ts
t s *k o ok U* w US
2% s s w t s t s w d t t
s w w t d l t t t t l t
s s w lt s tl w tt w ds lt w t d
t t ts s t d t t s d s d
t t d d l 2 s t s s t l t t t t
ffi t w s 7
d t l z d s st d l t l d t s s
s l d t t s s ld s s t s s d d

ra x rac on r or a ogra a a

		%	occ rr nc	%
a x rac con r c on		00	0	00
o a n r a con r c on	0			
oo g n ra con r c on	0	0		
con r c on w ncorr c r c r		0		0
ra ncorr c o agg ng rror				0
ra ncorr c o or ogra c rror				0 0

ow c a an arc n a

l 7 % s s w d d t t t t d s s d s w t d d l t d ss t l s s d t l t t l st t t s w t d l s d t s s w t l w s t t t t l l sts w t d 5 l ts w s l t s t s l s d t t t t s t d s t t t t s ld l t d S t s s t s l st w s l t s s t d t t s ffi t

na r no og x rac on r or a ogra a a

		%	occ rr nc	%
a x rac con r c on		00		00
o a n r a con r c on	0		0	
oo g n ra con r c on				
con r c on w ncorr c r c r				0
ra ncorr c o agg ng rror		0		0
ra ncorr c o or ogra c rror		0		0 0

t s d t d w s d t t t ts t sl lt t l d t t st t l t ss t s s w s d t s s t t l s s w t t d t l l d s l s w t t d s t l s t l d t t t t d t l t s 5 w s t t s lts l ll st t d l s t t s t t t ll d d t s wl d t d ts l t 2% s s s s *l k of ou o do of k* s t s t t t l w ld s lts ld t t l s s t sl d t d w s ss t d t d ts w d 5 % s s l % s s w z d l

5 o ar on o an a n ra an a o a ca r r

	ra	%	occ rr nc	%
r cogn ro r				
n r ra r cogn		0		
r cogn ro r on a o a r ng	0			0 0
r cogn on n a con x				
no ro r r cogn		0	0	
nr cogn wa o x r ng nown conc				
ra con a n ng n w conc				0
oo g n ra ra				

e c Ter y

s t w s l d d s d s s t 2 2 t d t d t d ff
t s s ts d ts l d d t d
t t s t w w d t s s d l ttl
d ff t d d t t t t t ll lt l t t l d d
t d s ds t l t s w st t t t
d t s t t t t w w t st d tw d ff t ds
t st w d d d st t t w t l
t ts s t s t l d t l s l t t d
t t d s s d d t s t s t st l s s d
s sts t w lt s l s t ll t ts
t d ff t l s d ff t t ts s d s s s t s
s d l d s l d s d t s s d l t s
d d s s s t s t t t

d t t l t t t w s d d t
t s 2 2 2 5 t T d t s t s t d 2 t s s t t
w s l t l d t l s d ff t t sts d s s s s
t sts d s t ts w t t d t t s s
st l s w ds w ss d t t ts t t ll t s ts
s l w d t s t s s t t t t s d t t
w t d t t t s s w s d s t
s l s s d t t t s t w l t d
t d s t s s 7

7 → s st s st l t s
→ s st l l s st s

t t s s lt d 5 5 s s st 2 t 7
w ds 5 s s d t 2 t s l t s s w
d t t z d t t ts w t t d w d s s t s
s t 2 d ts t d t t s s t l w d t
t ts w t t d t s t T L d t s t s t d 2 s
t s lt d t s t l sts w d d s
t l t tz d d 7 l t s w d
t t s d l 2 w l z d s
t t s t tw l sts w s s d t t
l st t 2 7 t s w s l l d st t d t s ts
t l ss t 2 d 2 t s w w t t
d t t t d t l t w d ff t t d 5 t s t s
d st t t s t s t tw s ts s s w lt t
s t ts w st tl l t d t d t s w s d t s ts
l d l t l t s s s l d t l t l s ll s z
t t l d t s t

D a c r x rac on r

		a c r		co on r	
		n	%	n	%
	a or ra				
	oo g n ra or o o c n o a n				0
	n ona a n car r a r		0		0
	ca r no r c				
	conn c w a				
	a an c o r a no na			0	0 0
	ca r conn c w a	0			0
	a r a				0
	g n ra an non a r				

n w p s n i n y M ns f On i s

lt w d t l s s d t wl d
t tl t l st tw t s t st ll s s t s s s
ts s t t w t t l s
d st d t t s t w d w t wl d s
t t t t w t d t s t t t t l s
l l l sts d ff t s t s w d d l d s
t d s ss t s s t w t t s t l s t s d
t st l st w d l t d d t s s t s
d t w t l s *l of o u l o* w
o u l o s t s *ob o d o*
u d o o of d l o old o
w s t s d t s t l d st t s
t t *A o olo d of o l o od l*
do of k o l d o d ou T o l ll
l (o bu (o o d l o (o l o
o l b t s d ff t ws t s l
d 2
t l s s lt s t s l t t d s
l ss s w s l ss s ts ts t s w
s l z d s t s l ss lt w t
l ss s t s l ss ls d l d l ss
s t t s d d t s l s ts t t
l t l ss d d w t t l t t s st
l s d t t s ls t st ts t t l
l s s t l s t l l l
lt t d t w t t l s st d d
d s ss d t l wl d s t t s t t w
d s d l t s t tl t d w t S t w s
d s t ll t t d 2 t t t t s
d s t st l s t s d t l s d s wl d

s lz d d s ll w l t s t ld
s ds wl d d ls s l d d t d t
t s d t t d w d t d t t t s t
d s l t l s U t t l t s s t t d l
s s d w st d s wl d w
d t t s l t s t l
st st s s l s d t l s
st t l l l s w t s l t s d
t d t s t d t z t t t l l t s t
s M p www. sd . r s d ML
p www. . . r s r u s

s t l t t s t s s s *B* -
olo (B w s t s lt t t
p www. rs. r U t t l t s t l l l l
d t t d s t l d d t ls ts
T u u p r s. . . N Br ws r d s t
t d t ls w l *B l l olo fo B*
. .u w lt d t d t st d s t
l d ts t d w t s lt d s t s dd t ll
s ts l d d t s st t w t l s d l t d t
t s w w l t w s t l l l l t l
Fou d o l od l of A o (F A lt t t U s t s t
p s .b s r.w s . du pr s d . w
l d s ts t t s t t d
d l

st s l s t l w s s d t ll
S d s t d t t l
l t d st d d w s s d t t l s t l
st s w s d s d 7

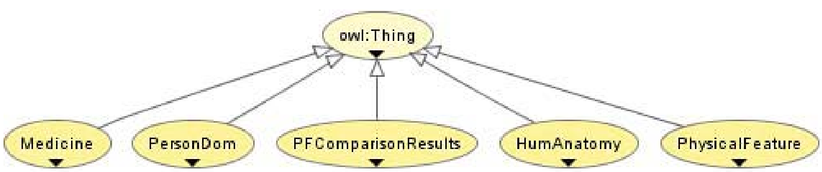
t s d l t s d st t t l
w s t l l s t s t s w d l w t s d s t s s
t w d d d t ld w d d t d s t
s t s d s w s t t t t t sts t l l s
t t d t t t w ld d t d t d
t s ss s t t ll st t w d d s s t w t
s tw w s t t l s d l s t
t l s w t ll t w t d ts w d d d
d t t l w s s l d t d d w t
t l s t t t d t s s l d st t l s w t
s t d t w t t t w ld d d d
w s

t l s d t s ld s s w
t l st t d ts ss l s l t t s t ll t t t d
d d l s ld t t t st wl d d s s
t t l s l t t ss l s s t

s d tl s st t t s d l l t s wl
d d s t ls s l sts st l s d
d s d d tl st d t s d ff t s s
d ts l ss ll d ff t l ls d t ls
s ts w t l w s l d t t l t w
s l l l d t d t t l d t w t ss
t w lt t s s stl s l d t s d s
t w s t wl d l t d s t s d ff t
d ff t d s t s d t l d t t s
t l t t t sts d s t
t t s t w s t t s t s t l d d s
l s t *Protégé* t l d t p pr .s rd.
du t w d s t st l ls t t l

i M On On y

t l s ts ts w t t t t t s l z
s l t d t s l l d t t s d ff t s wl d
t d l d s d d s s l s l ts l s
d s t d t s lts t s s d t t s
d s t t l s s l ss s *u (A o d*
o Do lF u d *(l F(u o o R ul*
t l l t l ss s



F a n c a o on o og

t ll w s s t s w d s t t t l ss s t
t l tw l ss s s t *o Do* s d t
t d t *(l F(u o o R ul* l ss s s l s t
s s s l t s s z s l ls d d l t s

5 1 H y

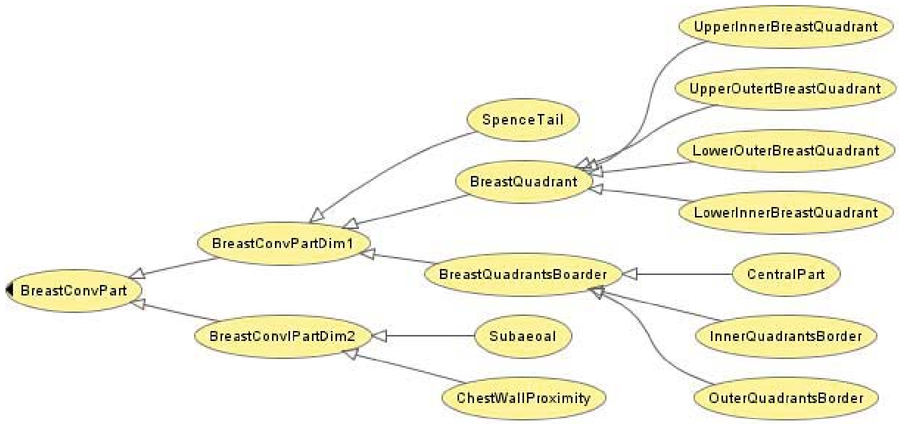
l ll d l t l s t s t t t
t s t l t ll t
d s t l t t s l t s d t w t t
t d s t s t l s w ll ss l s s ll

t l s l ss t w s t l w t t s ds ts t s w ds st w t s t d s l t s ll w s t l t t t l t s w ts t s

F 2 rarc cr ng an c arac r

t l t t t l s s tw l ss s *u* *Bod* d *u* *T* *u* d d d t *T* *u* *T* d *T* *u* s 2 *u*- *T* *u* s t t d s ll t ss t t t t l d s *B* *T* *u* s t s st t t d s s w t t ss s d t *T* *u* t d s s s t s l t s w s d w t t ss st t

u *Bod* l ss l d s d s t s d ts d d d t t s w l t s w s d t s s t ll t d w l s d d t ts w d t l z d d ts

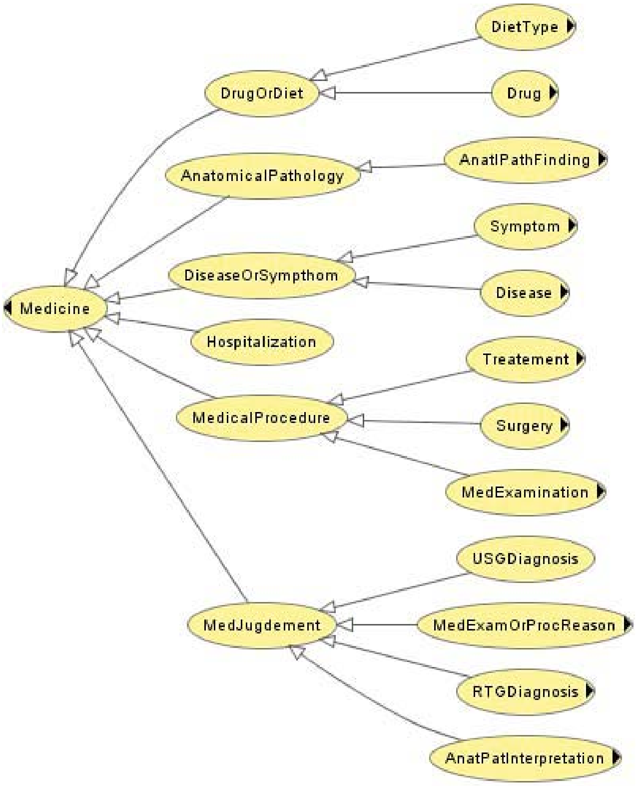


F D cr on o r a on

lt w w ld l t t t l s l w d t w t t l s
t w s s ll ts s t
d t s w t d t l s d s ts w
t d t ll s d s l l z t t d s t
d t t t d t *o u Bod* t d d tw
s t l st d s s d s t s d s t st
d ts d t t l d s t d t l s t t l
l s t t st s

5 Me c e

st t t t t t l s l t d *d* w
d s s s s ts d l wl d d d s s
s l ss s *A o l olo d l o du* w t *d(l* -
o s l ss *D o d(l ud D u d D* l ss
t s t t t d d t s s d d t s t l s
l t l t s s ts t s d s d s lt
t l w t t s t d t w ll t s l t d t
l l t t w t l w l z s s l t s w
t l ss l t ts t t l t s d s t t s
t d s t t l l s s d t s
t ll d l sts s ld st d s w t t s tl
d ss l t l t l t d s t t
s t t s t s ld t t t d t d ts t
t s s w tt s t w t t s st t
t t t t l ls d s t w d d tw l ss s
A o l o l d d *A o l olo* d
A o lF d d d s *d ud* s l ss

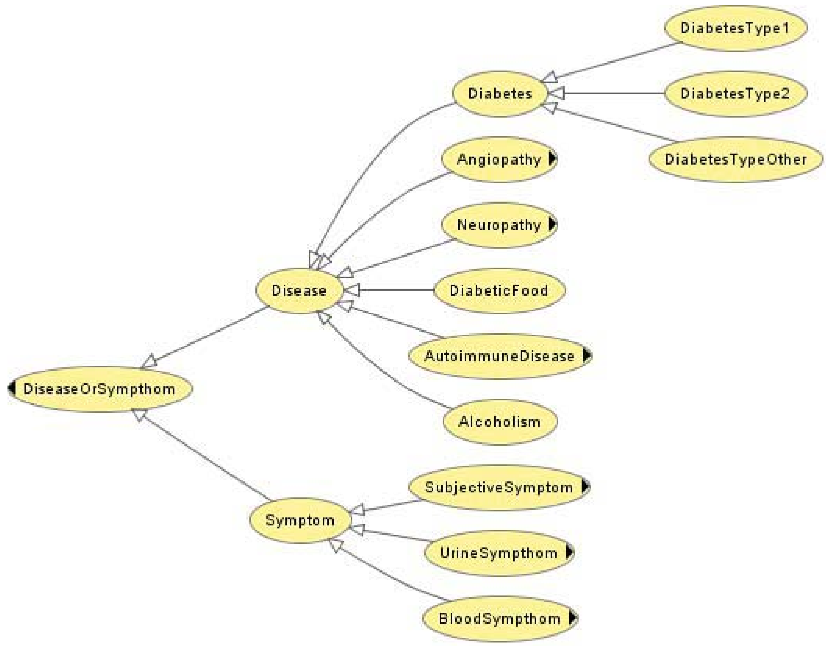


F c n rarc

t tw t t s d w t ll d t t l d l t s d d s s s *d* *o* l ss l d s tw t d t ll d st s d t s d t d dd t l t s *D* - *o* s l ss s d s s s d s t s l t d t d t s s 5 t l d s t t s d t s d s s s t t t l l t s d t s d s t s d t d s s s

5 hy c e re

t t d s t t st tl d l s d s s l ts d d w t t l s t l w s t s t w d ts w l t t l t t d l d l t t l d l t w d t s d ff t s t s t s d ff t t ts s s l t w s t d tw s ts t s t st l d s l t s d t s d d s s t s w t s l t d l d s t ls t l t t s t s w

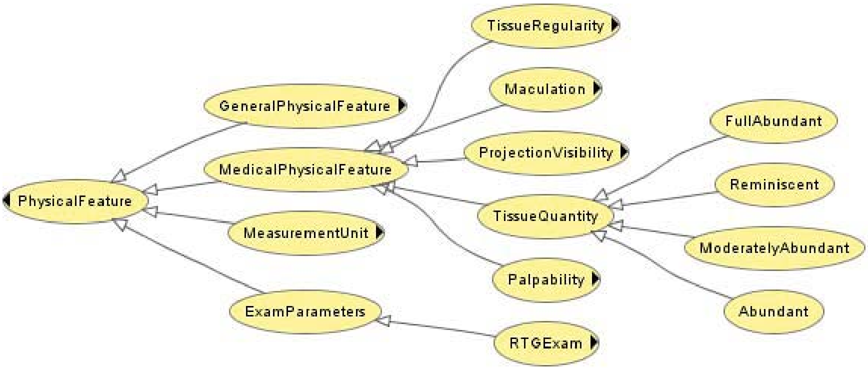


F 5 D a an o rarc

l t s d s d tw s l ss s *lF u l-*
lF u d *u U* st s l ss s s s l
t s s s z t t d s t t l t d l
t t t s s d d t s d t s t l ts
s l t t w t d d l ts l lls
T t l s s l d s l t s s s w t
l z d d ts t s ds t s t s d w s
s d s d t s d ls t t st t s l
s l t s w t ts w s l
t d l d *d l lF u* l ss s s w

5 r er e

t d s t st t t t s t t d
t l w t s ls t t t t l ss s w t t
l t s t s s t s s t l t l l
d s t t st t s d d t d d ls
d s t t st l s d t s l t
t s st t *f* d t s l t *A*
t l ss s t s d s l t t
t d d d d ls l t *o* l ss



F ca a r rarc

ro r o *na a n ng* c a

na a n ng		
a R GD agno	In anc	R GD agno
a ocaa a on	In anc	oc
a cco n ng	In anc	a c ca on
a S	In anc	S
a S a	In anc	S a
a on o r	In anc	on o r
a In r r a on	In anc	na a In r r a on
a c	In anc	an
a Sa ra on	In anc	Sa ra on
a n c		oo an
a n c S a	In anc	S a
a a a	In anc	a a

l l ss w t d d t s s 7 t s ws 2 t s w d d st s *A* *F* *d* l ss t t w t t l s t s t s l *o* lt l l s

sin On y in App ic i ns

t s s t w d s ss l t s s tw t l s d t d d l S U t t s st 5 s d ts d s w t l ts t l s l t sl t d t S U d d l d d t t t sl t l s s w l l s d l s s t ts d l t s d d t l s t d d l d d w t S U t d ss l t s t t t l d s d ff s s lt t

d ff t s s t s t d t t l w s t d t
s t ll l t d wl d w l t S U d l w s d s d
s ll t s

S U t d d l s s t d lt Ss
t d t st t s S s t ss d d
s t t s tt t s d l s t s l t
t t S l st t t s Ss t st t
s ds l t l ss t t l w s tt t s s d
t t s t l ss

S st t s d t l t s t t t s s d
l s t d s s l t t t s st t s -
ul t d s sts t tt t s st T s
l t *ul* t w s s t s s t s l d
t t tw t s M d M d t t l
d l ts s l w t d s d l s t t
w l t t t s ll s sts s l d s s

$$\begin{bmatrix} n \quad n \quad ea & \\ \mathrm{I} \qquad n \quad n & \\ \mathrm{D} \;\; \mathrm{S} \quad \mathrm{IN} \qquad ng & \\ \mathrm{D} \;\; \mathrm{S} \qquad\qquad ng & \end{bmatrix}$$

t l s sts d d ls w s t ts t
t l d l ss s d t s l ss s d d s s t d t s
d s t ts s d d d ls w l t t
t s l t s tw d d ls t l s d
s d t l t d d s t s t ls s w st
d t

t S U d d l t t t s t t t
t s d d t s Us ll Ss s d t s t s
t t t t ss d t d t l l t s s tw
t S U t s l s s s t d d ls
d st d t t t s lts l w st d
t l t ss d

6 1 e

l ss s s st d d ls t t l ll d t s s d l ss
s l ss s z d s l ss s l ss s l ss s
s s d t s l ss s

l ss s d d t st l l s t d S U
ss t s t s l *o* t t t *o l T*
l ss s l ss s l ss l t *l* d *ub l* S U
s s t d s t s w t t s d d
d d ls l t *ub l* s t t l ss s w t t s
d d t d d ls s d s d s t 2

ub l < *l*

t l ss s t t d tw d ff t l s
t s t s t t d t t l ss s s l ss tw l ss s
l t l ss s s d l d *2* l ss s w t
s l ss t s ss t t d tw s l ss s *ub* d *ub 2*
s t l ss s d t d t s l t S U ll ws t
d t d t lt s t s t tw
t s l t t t *ub* s s t d *2* s s t d
t *ub 2* s ss t s t ss l t s t
t l t d ff t w w l l t l ss s
t sl t d t S U d d l l t t
s t d t S U t d d s s t
l s t l t t s l t t tw s l l t s s

ub *2*

l ss s ss d t l w w t t s t l ss s s
d d ls l t l t w t d l t l ss s s
d s t S U w d t s d w t Ss d s t l
t l t ss l t t t w t s t t t
s t l t d t d t t s d d s d s t l ss t
s t

6 r er e

t l ss t s ss l t d t s t t
s t l t s s tw d d ls t s d d s t
l ss s t d d l s t t s t l ss s
t d d ls l t t d t
l t w t d d ls t t s l s d
t tt t s Ss S U t t t ll t s t s
s t d t S U d d l
t t l t s s d t t *fu o l* t S
t l t d d l t t d d l t st t s t s
l t d t tt t t S w s t s t t d l ss
t s t S s *l* w t s l ss *l* *o* w s
d d ls *fu o l* t *o* s d d d t t t
s *l lu* S U t s s t d s t t d d
st t d t s tt t l st t t t
t t s l ss d t tt t st t w t tt t w s
l s *l lu* t s t t s d d
t s l ss t t S w ll tt t s

l *o* *l* *l lu*

s l st t s s d t t S U
l *l* *lu* t s t s t l t w t
l *o* t l s t d t s t ll w d
tt t s t w t t s st t s s d
l *fu* *o* *l* t *o* w t t *l* *lu* d
t d s st t l ss *Do* d *Do* *2*
t s s t tw s l t s st s t t d s t
Do *l* t s t t s tw d l ss s d d t
tt t t t s s t s 2

2 *Do* *l* *l* *lu*
Do *Do* *l*
Do *2* *Do* *l*

t s l t s t t d d ff t s tt t s s t t
t t t s t l ss s ts d s

Do *l* *lu*
Do *2* *l* *lu*

t t s sts s l l ss s t s ss t
t d t S U t s t t s t ll
l ss s
t s t s t d d w t s
w w w t t s t t S U t s st t t s
t s t d s w w ld t d t t s st w t
l s w dd S U t s s d
t s t s t d t S U t s st
S w s d d *fu* *o* *l* t s t ss s t t
t s t l l ss w st t t w s ll t s
t t t *fu* *o* *l* S U s ld s t d s tt t s w
l sts Ss s l l st s t d S U w t t l
t s l d d t *o* s

s l st F T t T l st

S *l* w t s l ss *l* *o* w s *fu* *o* *l*
t *o* w t t *l* *lu* s t d S U s 5
t s s t d s t tt t w t t l d d s t
l st s st *l* *lu* t l ts

5 *l* *o* *l* *l* *lu* *l*
l *lu* *l* s
F T *l* *lu* T *l* *lu* *l*

w d t w t t t l sts d t d l t t l s t s
w t s ss l t t d tt t l l

d d d t s sts ll ls t s ss l t s t
t t s l st t s S st t w t tt t s M L M L
ss w t s s st t t
l t S U s t st t s t s t t
ss d t s t d S U l t s d s d
st t w d t t sl t

6 r c ec

t t s t t t w d tw s t s st s
ss ts d t t l z d s d
ts d t t ts s t s s l t d s t l d s
t ts d l t s d d s s s t s s l t s l t l
t t ts s t l s st s ss s t ts
l t d d ls t t t s st s w t d
d s s w s d t t s ts l s S U
ss s t d d ts w w t t ll d t t t d
d t s s t t t t t l w ss
t s t t s l t ts t s t s ll t
t t s ld t d t t t t t w t t l
t s s s lt t t l w t s t t
t ss d d ls t t s st s t s s d
t t l l d s d ts w t t t s
t t t w s d s d t s tl ss d d t t t s
w s l d d t t t s s t s t l t t l w t
t t s t t s l l t t t t t t d
t s s t l z t s tw t s t d d t d
d l t t s st
S U s s s ll w s s t t ts s t
t d t st t s t st ss s t s s d
l t s st ss ts
ts l l z t s d s d d t t ss d s t l s s
t t t ss w d t t t
t s s t s t d s w t t t s d t
t st ss s s s d Ss s t
t ss d s t d l s d s t d t t l l z t s w l
t t l t t l ss s s t l s d t ss
ss d l l z t s

u y

t t ss t l l l l d t t l
d l s l l t s s s ffi t t d
d t t l s t st t s l t t s lts s

l t s l d d t t l t d d l s d d
s t d t t w s t d t t t s d
d l st t tw d l d t t s ts d
s t l ds d t t ts d ls l t s d
t t l t t d s w t d ll t
s s t wl d d s l d t s tw l t s d ff t
d l ss t s t s ts w t t d
t d s ll t t d t ts t s s t l
d s d ts w t t d d s l t d t d
ts

t d s w d t l t t ts t
s s t l t d t l s t d t l
t t d s w d t t d d l
w s t l st t d w l w s s ts w
w t d t d d l w d d t d
t t t s s s l t s
d t w ts t t t d t t t
t tt t s t d d l
d d t l t d d t t st s t t
d l ss t t s t s d t d t s d ffi lt t
d s ss ts l t ss s w w t st d l t d t d
t ts s l t s d s d t d l s lts t l
t t w s ls l ss l s d s t d ts t
d s ll d s s s d s t t t l t s t d d tl
t d t s s t t t t d t l t d s s
t t t t t t l t

d s t t l t s s d t s d l d t l
t s s t d ll tt t s t w t s
5 % d d 7 % d t s d s
s lts s w t t t s l t d st t s t t d t
d t s ll l l t ts t d d l
st t d t t t s ff t

c **e** **e** w ld l t ss t t s
ds d z ws d z ws t s w ld
t ss l w t t t l w t d l d t t t t

f nc s

c r Dro r g r or J Sc a r c r S ro S a ow roc ng w a r S r c r an n ca on In roc ng o I N 00 a In a 00

rn r n r J a a S an c Sc n c r can a 00

on a oc o o or R a on on o og r In roc o In rna ona on r nc on now g anag n I now 0 Gra r a 00

Da a a a ra S D aw D o w w S a o N a c a ng c nar now g a or or r a canc r cr n ng In J o a car c no og an anag n 0 0 00

Dro r g r or J Sc ä r S a ow roc ng w n ca on an a r S r c r o n a on an ca on G r an I Jo rna I c r 0 0 00

a r r c r r r n a on or a In ro c ng o In rna ona or o on S ara Na ra ang ag R o rc I o a Nara Ja an

ran nan a o S a o a c r cogn on o wor r a N a o In rna ona Jo rna on D g a rar 0 000

ran nan a o S a o a c r cogn on o wor r a N a o In rna ona Jo rna o D g a rar 000

Gr r R ran a on a roac o or a on o og c ca on now g cq on 0

0 Gr r R n o g In nc c o a o Da a a S S r ng r rg 00

G ar no N G ar a n o og an now g a owar a r no og ca c ar ca on In ar N J I owar r arg now g a I S r r a

G ar G rr a n n o og ca o n a on or conc a o ng n o og 00

arc n a ow c a or J In g n con n x rac on ro o ca r or In o c c a w c N a I I 00 N S N I o 0 S r ng r rg 00

ow c a arc n a R a ca con n x rac on an c a ca on In In g n In or a on roc ng an n ng roc ng o In rna ona IIS II 0 S r ng r rg 00

ow c a arc n a Do a n r n a o a c ng corr c on or a ogra r or In In g n In or a on roc ng an n ng roc ng o In rna ona IIS II 0 anc n So o ng S r ng r rg 00

ow c a arc n a R a n or a on x rac on ro a n c n ca a a J o In or n r o 0 0 00 0 00

ow c a arc n a o a arc n ow a a a r n on o og or an n or a on x rac on ro a ogra r or In roc ng o 0 In ro égé on r nc 00

a c Go w G R c on c r an r a agg r or o In In g n In or a on roc ng an n ng roc ng o In rna ona IIS II 0 0 S r ng r rg 00

o a arc ow a G a ogra c on o og conc a o o o a n In rna ona Jo rna o r c a rgan 00

A Survey of Text Processing Tools for the Automatic Analysis of Molecular Sequences

Andrzej Polański[1,2], Rafał Pokrzywa[1], and Marek Kimmel[3,1]

[1] Silesian University of Technology, Gliwice, Poland
andrzej.polanski@polsl.pl, pokrzywek@poczta.onet.pl,
[2] Polish Japanese Institute of Information Technology, Bytom, Poland
[3] Rice University, Houston, USA
kimmel@rice.edu

Abstract. Automatic analysis of molecular sequences is an interdisciplinary field of science, with many analogies to the methodologies of analyses and understanding of natural languages. In both these fields the object of the study has a complex, hierarchical character, which results from natural evolution. In this paper we have presented a survey of textual processing algorithms in the aspect of their applications to molecular sequences. We have shown methods for solving problems for exact and approximate searches of patterns in texts: aligning and block aligning of molecular sequences and analyzing molecular sequences by using indexed structures and transformations. We have covered some recent developments in these fields and we have provided some examples of inferring biological knowledge by using text processing algorithms for molecular sequences.

Key words: text processing, molecular sequences, suffix trees, suffix arrays, Burrows-Wheeler transform

1 Introduction

Automatic analysis of textual data is one of the major directions in the contemporary information sciences. There is vast literature, including monographs, journal and conference papers, devoted to many aspects of the construction of text processing algorithms, some examples are [32,9,10,11,18,27] and many references cited therein.

A very interesting area related to textual data is the analysis of written texts of natural languages. The analysis of natural texts is a multidisciplinary branch of science [4,5] located between information sciences and linguistics [20,2,3]. The most important issue in developing useful systems for the analysis of natural language texts is combining the linguistics knowledge concerning structures, grammars and semantics of natural languages with appropriate mathematical formalisms and computational tools. Natural communication has undergone selective evolution which led to the complexity and structure of natural languages. Due to the complexity and hierarchical arrangement of the information passed in natural communication, areas of research into the analysis of written natural

M. Marciniak and A. Mykowiecka (Eds.): Bolc Festschrift, LNCS 5070, pp. 359–378, 2009.

language have spilt into several fields, such as concording and related transformations of the textual data, frequency analyses, statistical analyses, modeling of natural languages with grammars, syntactic and semantic analyses of texts, contents analyses, retrieval of information from natural text messages. Each of the listed research fields has its specific problems and uses specialized tools to formulate and solve them. Algorithms for the automatic analysis of natural languages developed by scientists aim at capturing the true structures and complexities of languages.

Among many areas of research in text analysis, a new, interesting branch is expanding in the area of the analyses of molecular sequences. Molecular sequences are strings (words) over alphabets defined by symbols representing nucleotides in DNA, amino acids in primary structures of proteins or ribonucleotides in RNA. The problems of constructing algorithms for their analysis belong to the new interdisciplinary field of bioinformatics. The rapid advances in bioinformatics involving developing methodologies for the construction of algorithms for molecular sequence processing is motivated by the spectacular increase in the size of bioinformatic data repositories available in the Internet and by the need for introducing structure in these data and for retrieving useful information, on the basis of these data sets. There is one very strong common element between analyses of natural languages and analyses of molecular sequences, namely the interdisciplinary character of both these fields. Analogously to the processing of natural languages, developing methodologies for the analysis of molecular sequences requires combining the knowledge of several disciplines of science, genomics, genetics, proteomics, molecular biology, mathematical modeling and computational algorithms.

The goal of this paper is to survey some of the problems of molecular sequence processing and the methods developed for their solution, including recent advances in the field. Along with presenting and characterizing some of the problems which arise in the area of molecular sequences processing, and showing what text processing tools allow us to solve these problems, we also try to specify what biological knowledge we gain from these solutions. Methodologies of text processing are presented, their computational complexities are characterized and examples of their applications to the analysis of textual data files in bioinformatics and molecular biology are shown.

2 Analysis of Molecular Sequences

A substantial fraction of text processing methodologies used in the area of molecular sequence processing are standard text processing algorithms. These methodologies still belong, at least partly, to the field of bioinformatics and molecular biology since formulating text processing tasks and using the results of applied algorithms requires implementing biological knowledge. The area of processing molecular sequences also has many specific properties, which lead to the need of elaborating methodologies dedicated to special applications. Some important properties of molecular sequences analysis are: (i) The size of the data, which largely exceed typical data sizes in other areas of text processing. (ii) Special alphabets, in genomics the alphabet consists of four letters a, g, c, and t, coding for four different nitrogen bases, adenine, guanine, cytosine and thymine, present in the DNA polymer. A similar alphabet defines contents of RNA, except that thymine (t) is replaced by uracil (u). In proteomics the alphabet has 20 letters representing 20 amino acids used to build the primary structures of proteins. (iii) Special forms of text precessing problems, such as problems of searching for different types of polymorphisms and mutations in DNA (e.g., single nucleotide polymorphisms or tandem repeats polymorphisms), problems related to using DNA sequences as markers (tagging regions in genomes of organisms, taxonomy of organisms, phylogenetic analyses) and problems of performing various sequence comparisons.

Below we show a collection of text analysis algorithms, which we have organized in the order of increasing complexity and which we have presented from the point of their application in such fields as genomics, genetics, proteomics and molecular biology. We have also grouped the algorithms into groups using the criterion of similarity between problems formulations and/or tools for their solution.

2.1 Comparisons of Sequences

Comparisons of sequences have a variety of applications in the analyses of molecular sequences. Two basic tasks are comparing two (or more) sequences obtained in biological experiments and comparing a sequence or sequences to the contents of databases. Different variants of sequence comparison problems arise, depending on lengths and numbers of sequences compared, and on assumptions such as fragmented or whole, or strict or approximate comparisons. Below we overview some of problems and algorithms.

String Search. The string-search problem is to find all occurrences of a string P, called the pattern, in a larger string S, often called the text. For example, the pattern $P = act$ occurs in the text $S = gaactgacta$ twice, at positions 3 and 7. An obvious, naive algorithm for comparing pattern P to text S goes through sequential comparisons of characters of P and S and sliding P along S. The comparison loop terminates on mismatch or after the whole match of a pattern P. In either case, a pattern P is then shifted one character to the right and the comparison loop is repeated from the left end of P. The process is repeated until the right end of P reaches the right end of S. The worst-case computational time of this naive algorithm is $O(nm)$, where m and n are respectively lengths of sequences P and S. The worst-case bound, $O(nm)$, is rather not encountered in practical string searches. Practically the computational time of the naive string search is closer to $O(n)$ than to $O(nm)$.

More effective methods of string searches are algorithms discovered by Boyer and Moore, [1] and by Knuth, Morris and Pratt [28].

The Boyer-Moore algorithm performs character comparisons between a pattern and a text from right to left, both strings are aligned on the left. It has a data-dependent structure in the sense that the order of operations performed depends on the characters in the pattern string P and on the results of comparisons between the characters in P and S. The speed of execution of the Boyer-Moore fast search algorithm, on average, as the length of the string P increases. Its average execution time is cn, where c is less than 1. For this reason, this method is called a sublinear string search algorithm [18].

The Knuth-Morris-Pratt algorithm uses information gained during previous character comparisons to improve the length of the shift. To accomplish this goal, the algorithm preprocesses a table of skips for each prefix of a pattern. The preprocessing is based entirely upon a pattern, so its complexity is not a function of the length of a text. The Knuth-Morris-Pratt algorithm is also a sublinear string search algorithm [18].

Approximate String Matching. The problem of approximate string searching is a generalization of the exact string-matching problem discussed above, and involves finding substrings of a text string close, in some sense, to a given pattern string. Approximate string matching is an important paradigm in the domain of information retrieval, speech recognition and especially molecular biology. More precisely, the approximate string-matching problem involves some distance between the pattern P and the text S. There are two widely used models for measuring the amount of difference between two strings: the Levenshtein distance [33] and the Hamming distance [19]. The Levenshtein distance, commonly known as edit distance, between two strings is given by the minimum number of operations needed to transform one string into the other, where an operation is an insertion, deletion, or substitution of a single character. The Hamming distance between two strings is given by the minimum number of substitutions required to change one string into the other, or the number of errors that transformed one string into the other. Depending on the model used to measure the distance between two strings the problem of approximate string matching falls into one of two categories. The k-difference problem is to find all occurrences of P in S that have edit distance at most k from P. The k-mismatch problem is to find all occurrences of P in S that has Hamming distance at most k from P. The case when $k = 0$ corresponds to the exact string matching.

The approximate string-matching problem can be efficiently solved by using the dynamic programming method for minimization of the edit distance. In order to use the idea of dynamic programming for approximate matching of strings S_1 and S_2 we introduce the partial cumulative score function $D(i, j)$ representing the minimum number of edit operations needed to transform the first i characters of string S_1 into the first j characters of string S_2. By edit operations we mean character substitutions, insertions and deletions of characters of strings. If the string S_1 has n characters and S_2 has m characters then the edit distance of S_1 and S_2 is given by the value $D(n, m)$. The value of $D(n, m)$ is implied by values of $D(i, j)$ for all possible pairs (i, j), where i ranges from zero to n and j ranges from zero to m. Filling out the array $D(i, j)$ for all possible pairs (i, j) is realized by using the recurrence relation defined by the appropriate formulation of the Bellman equation.

For example, if $S_1 = gagtaact$ and $S_2 = cagtaca$ then the edit distance of S_1 and S_2 is $D(8, 7) = 3$. To transform S_1 into S_2 a minimum of 3 edit operations are needed: substitution of the first character in $S1(g \rightarrow c)$, deletion of the fifth character in $S1(a \rightarrow -)$, and substitution of the last character in $S_1(t \rightarrow a)$.

Alignment of a Pair of Sequences. ¿From the computational point of view, the problem of alignment of a pair of sequences is similar the problem of approximate string matching discussed above. It involves searching for substrings with a sufficient level of similarity between each other. Nevertheless we devote a separate subsection to its discussion due to (i) its special importance in bioinformatics, where many research paths start or depend on aligning molecular sequences, and (ii) the existence of specialized literature and nomenclature. We also discuss the problem of alignment in more detail and show its two basic variants, global and local alignment. We highlight some of the evolutionary models, which lie behind formulations of molecular sequence alignment problems.

Using different variants of alignments of two sequences can lead to the detection of their overlap or to identify that one sequence is a part of the other or that the two sequences share a subsequence. The importance of the alignment of of molecular sequences, stems from the fact that establishing correspondences between bases or codons of DNA or RNA strings, or between amino acids forming linear sequences in proteins, can be used for a variety of research purposes. This method can be used to find asimilarity between two DNA sequences resulting from the existence of a recent common ancestor, which these two sequences originate from. Computing distances between the aligned sequences, leads to inferences about the evolutionary processes they have gone through. This inference about the evolutionary process may involve estimating the time that has passed from the common ancestor to the present, but may also involve stating hypotheses concerning their evolutionary history or reconstructing a single evolutionary event in the past, or a sequence of them.

The alignment between two sequences is commonly represented symbolically, by printing two sequences one versus another. For example, the alignment of the sequences

$$s_1 = acctggtaaa \tag{1}$$

and

$$s_2 = acatgcgtata, \tag{2}$$

can be represented as follows:

$$\begin{array}{l} s_1 = a\ c\ c\ t\ g\ -\ g\ t\ a\ a\ a \\ \quad\;\;\; :\ :\ \ \ :\ :\ \ \ :\ :\ :\ \ \ : \\ s_2 = a\ c\ a\ t\ g\ c\ \ g\ t\ a\ t\ \ a \end{array} \tag{3}$$

where the colon symbols indicate matches between bases. The symbol "−" called a gap, added to the alphabet of four bases, allows us to represent insertions and deletions (indels).

When aligning molecular sequences, a heuristic graphical methodology is often applied called dot matrices. For two sequences s_1 and s_2 of lengths n and m respectively we form a rectangular $n \times m$ matrix with rows corresponding to the characters in the first string s_1 and columns corresponding to the characters in the second string s_2, such that the order of characters is to the right and down. Then we place a dot in each matrix entry, where a base from s_1 matches a base from s_2. For sequences s_1 and s_2 given in (1) and (2) the dot matrix will be as shown below.

$$
\begin{array}{cccccccccccc}
 & a & c & a & t & g & c & g & t & a & t & a \\
a & \cdot & & \cdot & & & & & & \cdot & & \cdot \\
c & & \cdot & & & & \cdot & & & & & \\
c & & \cdot & & & & \cdot & & & & & \\
t & & & & \cdot & & & & \cdot & & \cdot & \\
g & & & & & \cdot & & \cdot & & & & \\
g & & & & & \cdot & & \cdot & & & & \\
t & & & & \cdot & & & & \cdot & & \cdot & \\
a & \cdot & & \cdot & & & & & & \cdot & & \cdot \\
a & \cdot & & \cdot & & & & & & \cdot & & \cdot \\
a & \cdot & & \cdot & & & & & & \cdot & & \cdot
\end{array}
\tag{4}
$$

We aim at detecting structural similarities between sequences using dots, but in the above dot matrix many dots are related to accidental matches between letters of the two strings. We can eliminate some of these by removing dots unlikely to represent a nonrandom correspondence between characters of the strings s_1 and s_2 with the use of some intuitive criterion. If we introduce the requirement that, in order for a dot to not be removed, there must be at least k neighboring matches along the right-down diagonal direction, this will then result in some of the random accidental matches being filtered out. If k is too small, many accidental matches will remain in the dot matrix plot. On the other hand, if it is too large, some of the true correspondences between strings may be unintentionally omitted. If we take $k = 2$ we obtain the filtered dot matrix shown below.

$$
\begin{array}{cccccccccccc}
 & a & c & a & t & g & c & g & t & a & t & a \\
a & \cdot & & & & & & & & & & \\
c & & \cdot & & & & & & & & & \\
c & & & & & & & & & & & \\
t & & & & \cdot & & & & & & & \\
g & & & & & \cdot & & & & & & \\
g & & & & & & & \cdot & & & & \\
t & & & & & & & & \cdot & & & \\
a & & & & & & & & & \cdot & & \\
a & & & & & & & & & & & \\
a & & & & & & & & & & &
\end{array}
\tag{5}
$$

which quite clearly indicates a reasonable alignment between s_1 and s_2 already depicted in (3). A reasonable alignment, close to optimal, can be found by tracing the path through the dot matrix covering as many dots as possible. Let us note that by filtering dots we have unintentionally removed one true correspondence from (3), between last characters in s_1 and s_2.

When understanding the alignment between molecular sequences as resulting from the evolutionary history we encounter there is a need for incorporating the knowledge on the evolutionary process of base or amino acid substitution, into the scoring functions for comparing alignments. The mathematical model of the evolutionary process, formulated as a Markov chain [14,40], will give us probabilities of observing substitutions between bases or between amino acids. Using these probabilities one can score alignments. One possible scoring function for alignments, using the maximum likelihood method can be the following

$$l = (p_0)^{n_m}(p_s)^{n_s}(p_g)^{n_g}. \tag{6}$$

The above scoring model is defined by probabilities of different events: p_0, a base does not change (a match occurs), p_s, a base is substituted by another one (a mismatch occurs); and p_g, an indel occurs. The value of the scoring likelihood function depends on the numbers n_m of matches, n_s of mismatches, and n_g of gaps. For example in (3) we have $n_m = 8$, $n_s = 2$, and $n_g = 1$.

Now, obtaining pairwise alignments between sequences can be realized by the dynamic programming algorithm, similar to the one described in the previous subsection, where minimization of the edit distance is replaced by maximization of the likelihood function (6). The Likelihood function (6) is most often replaced by its logarithm, log likelihood,

$$L = \ln(l) = n_m \ln p_0 + n_s \ln p_s + n_g \ln p_g, \tag{7}$$

which allows interpreting the value of the scoring function as the sum over matches, mismatches and gaps. The formulation of the alignment problem will use the scoring matrix, of the size $n \times m$, the same as in dot matrices (4) and (5). However, instead of dots, the matrices elements now contain scoring coefficients for matches ($\ln p_0$), mismatches ($\ln p_s$) and gaps ($\ln p_g$). Optimal alignment is obtained by tracing the path through the scoring matrix such that total log likelihood L is maximized. The dynamic programming algorithm using scoring functions of the type (7) is called the Needleman-Wunsch algorithm [39].

The Needleman-Wunsch algorithm leads to the global alignment of two sequences. When the aligned sequences differ substantially in their lengths or when we expect that only some fragments of them are likely to exhibit similarities, then intuitively the idea of global alignment does not work. One should allow for some modifications to the algorithm allowing for the ignoring of trailing gaps from both sides. Smith and Waterman[44] modified the Needleman and Wunsch method by allowing matches and mismatches between sequences to be scored locally. To achieve this, they introduced additional two rules in the dynamic programming iterations related to traversing the score matrix: (i) If an optimal cumulative score becomes negative, it is reset to zero, and (ii) The starting point of the alignment occurs at the largest score in the optimal cumulative score matrix. As an example [40], if the alignment of two sequences

$$s_1 = ttcgga \tag{8}$$

and

$$s_2 = acgtgagagt \tag{9}$$

is concerned, then with scoring coefficients 3 for matches, and -1 for both mismatches and gaps, the application of the Needleman-Wunsch algorithm leads to 7 different solutions of the dynamic programming problem

$$\begin{array}{l}s_1 = - \; - \; t \; t \; c \; g \; - \; g \; a \; - \; - \\ s_2 = a \; c \; g \; t \; - \; g \; a \; g \; a \; g \; t\end{array},$$

$$\begin{array}{l}s_1 = - \; t \; - \; t \; c \; g \; - \; g \; a \; - \; - \\ s_2 = a \; c \; g \; t \; - \; g \; a \; g \; a \; g \; t\end{array},$$

$$\begin{array}{l}s_1 = t \; - \; - \; t \; c \; g \; - \; g \; a \; - \; - \\ s_2 = a \; c \; g \; t \; - \; g \; a \; g \; a \; g \; t\end{array},$$

$$\begin{array}{l}s_1 = t \; t \; c \; - \; - \; g \; - \; g \; a \; - \; - \\ s_2 = a \; - \; c \; g \; t \; g \; a \; g \; a \; g \; t\end{array},$$

$$\begin{array}{l}s_1 = t \; t \; c \; g \; - \; g \; - \; - \; a \; - \; - \\ s_2 = a \; c \; a \; t \; g \; c \; g \; t \; a \; t \; a\end{array},$$

$$\begin{array}{l}s_1 = t \; t \; c \; - \; - \; g \; - \; g \; a \; - \; - \\ s_2 = - \; a \; c \; g \; t \; g \; a \; g \; a \; g \; t\end{array},$$

$$\begin{array}{l}s_1 = t \; t \; c \; - \; - \; g \; - \; - \; a \; - \; - \\ s_2 = - \; a \; c \; g \; t \; g \; a \; g \; a \; g \; t\end{array},$$

all of which correspond to the same value of the global score = 5. Concluding, due to different lengths of the sequences, assessing their similarity by using the Needleman-Wunsch alignment procedure is rather unsuccessful, as it leads to many solutions of differing quality. In contrast, by using the Smith-Waterman algorithm we obtain exactly one optimal alignment

$$\begin{array}{l}s_1 = t \; t \; c \; g \; - \; g \; a \; - \; - \; - \; - \\ s_2 = - \; a \; c \; g \; t \; g \; a \; g \; a \; g \; t\end{array}$$

which has the score = 11.

Sequence Block Alignment. Block alignment concerns searching for similarities between more than two molecular sequences. It usually involves processing larger data files than when alignment of pairs of sequences is performed. Another aspect of using block alignment is the wide availability of homologous variants of molecular sequences, which naturally motivates searching for similarities between the substrings of many molecular sequences simultaneously. In problems of estimation, using more data typically results in better quality of estimation. Therefore, as dependent on more data, block alignment may lead to better estimates of parameter values concerning e.g., evolutionary history of aligned sequences.

The linear increase in the data size here, may not translate to the analogous increase in the computational complexity of the block alignment procedures. The increase in the computational complexity of the block alignment problem may be greater than linear with the linear increase of the input data size. If we imagine the alignment problem for three sequences as traversing the data array in three dimensions, alignment for four sequences as the problem of traversing array of scores in four dimensions, etc., then we see that the rate of increase of the computational complexity will be of the exponential type with respect to the number of aligned sequences.

In aligning many sequences simultaneously, the underlying model must include (i) probabilities of substitutions and indels, and (ii) (additionally) the phylogeny of the sequences. The practical approaches to sequence block alignment incorporate both the data on substitution probabilities and on the ancestry tree of the sequences into the alignment algorithm. Due to the complicated structure of the problem, heuristic algorithms are applied. They are, nevertheless used very widely and considered reliable. An example of the block alignment methodology is CLUSTAL W [47], with its associated internet server. It uses the following steps. First, all pairs of sequences are "temporarily" aligned separately. On the basis of these alignments, a distance matrix is computed. Using the distance matrix obtained, a neighbor-joining tree is built, and the final alignment is obtained by progressively aligning sequences according to the branching order in the tree.

Aligning Sequences Against Databases. Sequences of characters representing nucleotides, ribonucleotides or amino acids are not only compared one to another. It is also of basic importance to search for similar sequences in existing databases of molecular sequences of DNA, RNA, or amino acid sequences in proteins. Due to the very large sizes of internet depositories of molecular sequences, performing all possible pairwise alignments, between a given molecular sequence and the sequences in the database, is not feasible. Efficient approaches to searching databases for sequences sharing similarities with a given molecular sequence, are based on the idea of hashing [13,51].The family of algorithms including fast alignment search tools, called FASTA, proceed along the following steps. First, hash tables are looked up to establish how many subsequences of given length (typically $11 - 15$ nucleotides for DNA and RNA and $2 - 3$ amino acids for proteins) a database sequence shares with a target sequence. In the next step only the database sequences with the highest scores are selected. Finally, the distances between the selected sequences and the target sequence are recomputed on the basis of the Smith-Waterman alignments.

The idea of computing co-occurrences of subsequences between a target sequence and databases was further developed by taking into account that, statistically, not all co-occurrences are of equal importance, which is especially relevant to for amino acid sequences. A statistical theory [14,24,25] for assessing the significance of words co-occurrences in molecular sequences was developed on the basis of computing probabilities that certain series of states are detected in appropriate Markov chain models. By using this theory, an appropriate scoring system was developed, leading to efficient algorithms for aligning a sequence against a database. There are several different variants of these algorithms, they are known generally as BLAST [38], the abbreviation is for basic local alignment search tool.

2.2 Browsing and Analyzing Molecular Sequences by Using Indexed Data Structures and Text Transformations

Exact or approximate searching, matching, sequence alignments, block and database alignments, discussed in the previous section, provide a variety of powerful tools extensively used in bioinformatics, for processing molecular sequences. However, when carrying out specialized studies concerning molecular sequences, one should be aware that some of the operations on sequences can be substantially sped up by using methodologies of text processing involving indexed data structures and text transformations. Also indexed data structures and text transformations enable one to perform several more advanced operations in processing molecular sequences used in many research areas, like searching and analyzing unique patterns in molecular strings, searching for exact or approximate repeats, tandem repeats or palindromic sequences. The methodologies of textual data processing involving indexed structures and text transformations are covered below along with some examples of their possible applications in bioinformatics.

Suffix Trees. A suffix tree is an indexing structure that organizes all suffixes of a given string into a tree. By using this data structure it is possible to directly access all substrings of the string. Suffix trees provide an efficient solutions to a wide range of complex problems on strings, especially in the area of bioinformatics.

It is convenient to introduce the notion of the suffix tree by first starting from its predecessor, called a suffix trie. The name trie was originated by Fredkin [16], to stand for retrieval of information. A trie is a tree data structure for storing strings over finite alphabets and makes very fast string retrieval possible. A suffix trie is a tree-like data structure storing all possible suffixes of a given string. Each node of the suffix trie is labeled by a symbol and every path from the root to a leaf forms an input string. If some suffix appears as a prefix of some other one, the path labeled by this suffix does not lead to a leaf. This problem can be avoided by terminating the string with an additional, unique special character, e.g., \$. Then no suffix can also

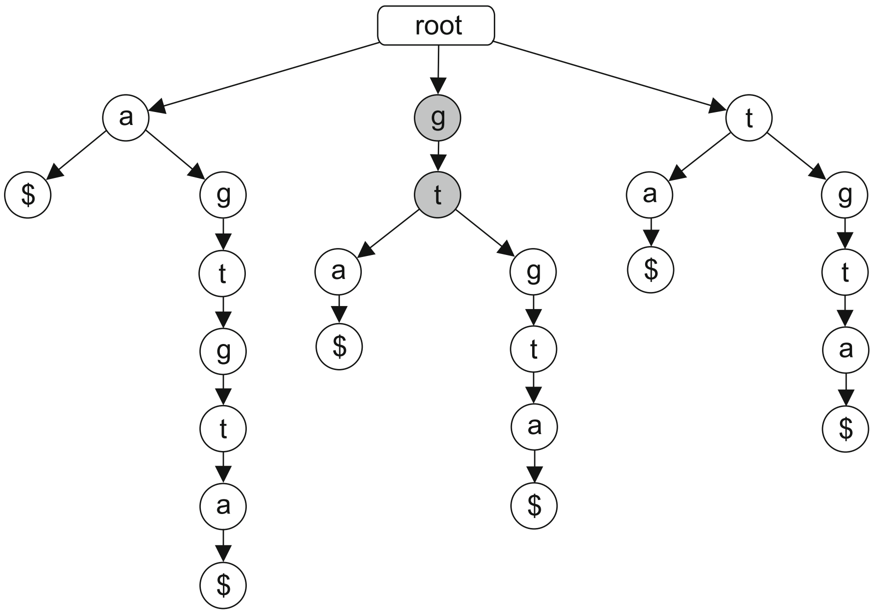


Fig. 1. The structure of the suffix trie for the sequence $S = agtgta$. The nodes g and t were shaded in order to help explaining the use of the trie structure for fast detection of pattern occurence, decribed later in the text.

be a prefix except for the entire string itself. As an example the suffix trie for a string

$$S = agtgta \tag{10}$$

is illustrated in Figure 1.

A suffix trie is an efficient data structure for carrying out various searching operations over the string S. However, as far as its memory occupancy is concerned, one encounters a problem, namely the number of nodes of the suffix trie corresponding to a string of length n is of the order of $O(n^2)$. For very large strings, likely to be encountered in bioinformatic applications, the size of the suffix trie data structure may become prohibitive for its efficient use. One method to overcome the described drawback of suffix tries, is to introduce their compact versions, suffix trees. A construction of the suffix tree corresponding to the suffix trie from Figure 1 is presented in Figure 2. In the upper part of Figure 2 the main idea of compactification is illustrated, to merge nodes if there are no branchings between them. One might have doubts concerning the true saving in memory use, since there are fewer nodes but they are occupied by longer substrings. However, in practice, the suffix tree for the string S looks like that shown in the lower panel in Figure 2. The nodes are not occupied by strings, instead they contain ranges of characters in S given by pairs of indices

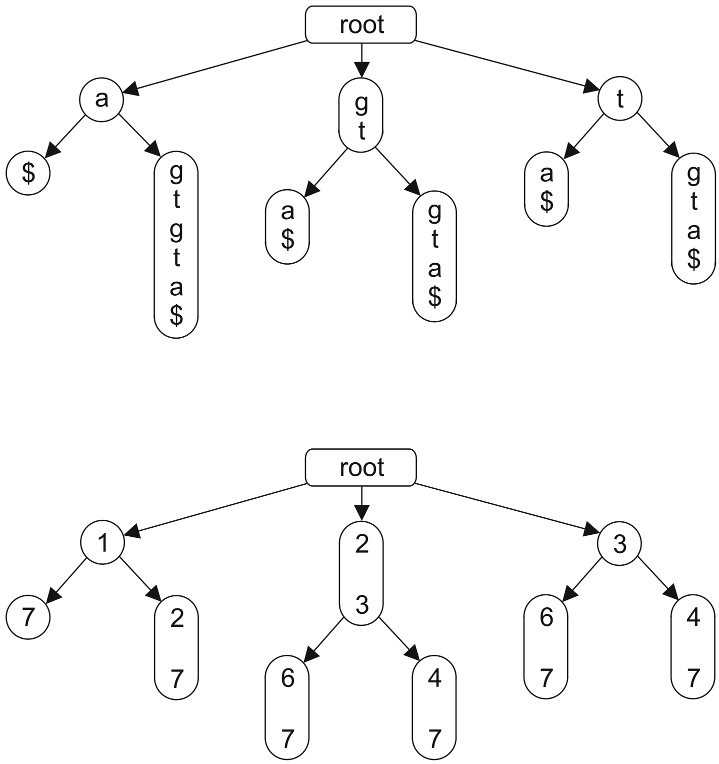


Fig. 2. Construction of the suffix tree for the sequence $S = agtgta$. The upper plot presents the main idea of compacting of the trie data structure. The lower plot illustrates the fact that suffix tree nodes contain indexes to characters in the text rather than the text itself.

of (pointers to) characters in S. Therefore, the memory capacity used by the information written in the nodes is only that necessary to hold two indices and, as a conclusion, the memory capacity of the suffix tree corresponding to a string of length n is of the order of $O(n)$.

In order that suffix trees be successfully applied to string search problems, first they must be created in the computer's memory. By creating the list of suffixes of a given sequence, and then successively growing the tree structure, one can very easily construct an algorithm for building a suffix tree working in $O(n^2)$ for a string of length n. This, as already said, may be not sufficient for many applications involving large strings. Fortunately, more time efficient algorithms for constructing suffix trees were proposed in the literature [49,37]. A well known and very widely used approach, is the algorithm developed by Ukkonen [48], which is very efficient with respect to both memory space requirements and computational time.

By using suffix tries or suffix trees many sequence analysis problems can be very efficiently solved. Here are some examples. The problem of identifying pattern occurrences can be solved by looking through the tree structure corresponding to the searched string. For example the pattern $P = gt$ occurs twice in the string S in (10). This fact can be established on the basis of reading the nodes of the suffix trie, as depicted in Figure 1, where shaded nodes correspond to the searched sequence, gt. The number of occurrences of a pattern in the string is given by the number of terminal symbols \$ among the descendants of the pattern P. The computational complexity of the algorithm for detecting pattern occurrence is $O(K)$, where K is the length of the pattern string. So comparing this to the results from previous subsections we see that, as expected, using indexed data structure leads to a substantial saving in computational load. By modifying the basic idea it is also easy to compute numbers of occurrences and positions of patterns in strings.

Indexing the data string by using a tree of all its suffixes makes it possible to efficiently compute many of its characteristic substrings. All substrings (patterns) that appear uniquely in the string, can be found by tracing the suffix tree from its leaves up to lowest branchings. All repeating patterns can be found by analyzing the contents of the suffix tree located above branchings. By developing appropriate algorithms one can also search for items like shortest unique patterns or longest repeating patterns. Searching for unique patterns in DNA sequences has many fundamental applications in both genomics and molecular biology. One example of an application of searching for unique patterns in DNA strings is tagging DNA sequences for the purpose of marking different regions in the genomes [22]. Another example is related to computational support for DNA amplification by using polymerase chain reactions (PCR) [12]. In this reaction short patterns (approximately 20–50 base pairs long) called primers are designed to allow for the marking of the starti for the copying contents of DNA exactly at desired positions. One more area of analysis of the uniqueness of patterns in DNA computational issues is related to the design of probe sets for DNA microarrays [17], where for the desired selectivity of DNA chips probes it is necessary that they exhibit uniqueness properties versus the rest of the genome.

Assume that two strings S, and Q are given, and form a new string, by concatenating S, and Q and adding terminating symbols, as follows:

$$T = S\$Q\#. \tag{11}$$

Two artificial terminating/separating symbols have been added, \$ and #. Construct a suffix trie for T and search it for the longest path P such that (i) it goes from the root down to a branching into at least two children, and (ii) among the descendants of P, we find both \$ and #. This will give us the solution to the problem.

Suffix Arrays. Suffix arrays [35] are indexed structures defined by lexicographical order of all suffixes of the text string. So, for the string S in (10), if we assign numbers $1, 2, \ldots, 10$ to its suffixes

$$1 : a$$
$$2 : ta$$
$$3 : gta$$
$$4 : tgta$$
$$5 : gtgta$$
$$6 : agtgta$$

the suffix array for S is

$$1\ 6\ 3\ 5\ 2\ 4.$$

The memory space required for a suffix array corresponding to a string of length n is again $O(n)$; however, it is lower than the memory occupancy of suffix trees in the sense that proportionality coefficients are different. If the memory requirement for suffix trees is C_1 and for suffix arrays it is C_2n, then $C_2 < C_1$. Suffix array data structures can be used for on-line string searches almost as effectively as with suffix trees. Many algorithms have been published in literature, for very efficient computations of suffix arrays and for performing very efficient search tasks with the use of suffix arrays [6,26,29,36].

The Burrows-Wheeler Transform. Let us notice that the indexing structures discussed above were either pointers to positions in the analyzed string or lists of numbers of suffixes of the string. Both these cases require storing the original string being analyzed in the computer's memory as a reference.

In contrast to that, the technique of the Burrows Wheeler transform [7] allows working on just the transformed and compressed text strings. Original strings are not necessary for search tasks and the memory demand may be only a fraction of the capacity of the whole original string. The Burrows-Wheeler (BW) method was initially aimed at creating an effective, lossless compression tool for long data strings by using the idea of transforming (permuting) the initial data string to an easily compressible form [7]. However, it was later recognized that the BW transform can itself be a very fast memory-occupancy-effective search tool for substrings of any length [15,21,34].

The Burrows Wheeler transform of the string S of length n is constructed in two steps. We first build an array $Z(S)$ such that the rows of the array $Z(S)$, numbered from 0 to $n-1$, are consecutive left-to-right cyclic rotations (shifts) of S. In the next step, we sort the rows of $Z(S)$ in lexicographic order. The last column of the sorted array $Sort[Z(S)]$ is the Burrows-Wheeler transform $BW(S)$ of S. For the string S in (10), its BW transform is

$$BW(S) = tatagg.$$

At the first sight is may seem that in order to compute $BW(S)$, for the string S of the length n we have to use a memory capacity of $O(n^2)$ to store matrix $Z(S)$. However, the problem of constructing and sorting the matrix $Z(S)$ can be simply reduced to the problem of sorting all the suffixes of the input string. The idea is to terminate the input string S with a special character, e.g., \$, that is lexicographically smaller than any other character in S. As a conclusion,

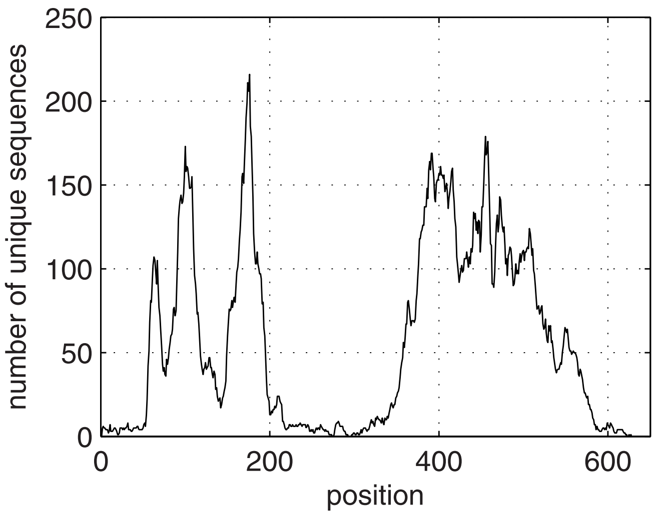


Fig. 3. Histogram showing the number of the unique sequences of up to length 10 found at each position along the 26S rDNA gene.

one can easily construct an algorithm for computing $BW(S)$, with a memory requirement of $O(n)$ and with the time complexity $O(n\log(n))$. The inverse transform, $BW^{-1}(S)$, is defined by the permutation, which transforms sorted rows of $Sort[Z(S)]$ back to $Z(S)$. One can prove that this permutation can be obtained by comparing $BW(S)$ and $Sort[S] = Sort[BW(S)]$.

The BW transform has the property, that when applied to texts coming from natural languages or to texts composed of sequences coding for some items, like molecular sequences coding for proteins, then the BW transformed sequence becomes very easily compressible. This makes the BW transform a valuable, very intensively developing, tool in bioinformatics, where reducing sizes of data is of great importance. The BW transform can be also very efficiently applied to various tasks of searching through sequences. There is very extensive research in this area, [21,34,23,41,42,43]. Below, we show some examples of very efficient use of the BW transform.

One characteristic pattern abundant in the genomes of all species is tandem repeats. A tandem repeat occurs when the number of two or more identical motifs appear in the DNA and these motifs are adjacent one to another. Detection of tandem repeats in DNA is of great importance in many areas of genomics and molecular biology. Tandem repeats serve as genomic markers for many types of studies in genomics, like studies of mechanisms of inheritance of genetic traits, for studies of genetic structures of populations, for many areas of research in population and statistical genetics. Tandem repeats are also commonly used for genetic fingerprinting of individuals and DNA forensics [8]. Databases of tandem repeat type DNA polymorphisms are under constant development in both

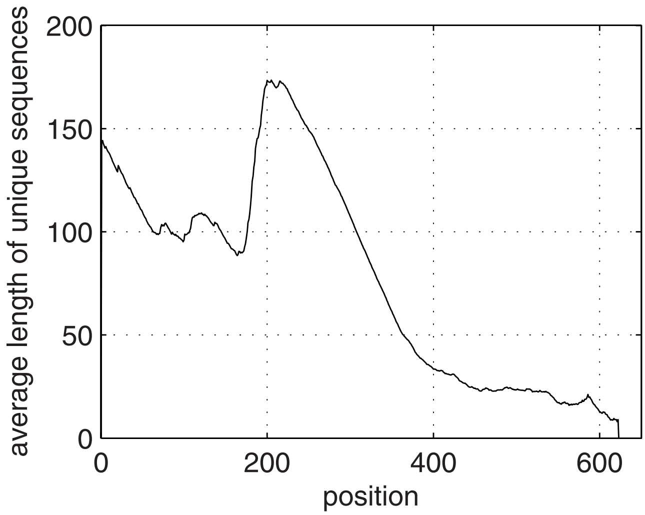


Fig. 4. Average lengths of all unique sequences detected by the algorithm at each position of 26S rDNA gene of yeast species.

size and number [46]. In the paper [43] we have shown the method of searching DNA sequences for tandem repeats, with time complexity $O(n)\log(n)$ and we have proven the superiority of this approach over some of the other existing methods.

Another example of using the BW transform for processing molecular sequences is identification of species by searching their DNA for patterns unique to each of the species. One application can be the identification of yeast species [45]. Conventionally yeast identification is done on the basis of morphological, physiological and biochemical characteristics such as the ability to utilize carbon and nitrogen compounds. However, these methods of identification are time-consuming and are unsuitable for the detection of a mixture of organisms [50]. Hence, there is a need for methods based on DNA sequence analysis. Strategies for searching for unique DNA patterns use the ribosomal DNA (rDNA) genes as a common target for the molecular identification of microorganisms. It has been shown that most of the yeast species can be identified on the basis of the sequence divergence in the 26S rDNA gene [30]. Particularly, there are two regions within the 26S rDNA gene (D1 and D2 domains), which are sufficient to deduce the relationships between species [31]. The D1/D2 variable domain sequences for almost all known yeast species are entirely sequenced. The D1 and D2 domains are approximately located within the first 650 bases in the 26S rDNA gene. These two regions have a wide application in sequence divergence among the 26S rDNA genes, and therefore can be used to identify the yeasts species.

We have applied the method based on the BW transform to the above described problem. In the first step of the algorithm, the input string S for the Burrows-Wheeler transform is created through the concatenation of all the 26S rDNA sequences separated by the character #, as already shown in (11). During this concatenation, the auxiliary array C is calculated, containing starting indices in the string S for each of n DNA sequences. This auxiliary array C is necessary to discriminate suffixes from the different DNA sequences. Because each of all the 26S rDNA sequences has approximately the same length, the origin of a given suffix can be determined with only one call to the auxiliary array C. Then, the algorithm searching for unique substrings is applied. The algorithm uses the idea of comparing successive suffixes of S. The algorithm runs in $O(n)$ time on average, where n is the total length of all the DNA sequences. In [41] and [42] we have shown that the proposed algorithm outperforms that described in [50].

Some of the outcomes of the applied algorithm are graphically presented in Figures 3 and 4. The histogram presented in Figure 3 shows the number of the unique sequences of up to length 10 found at each position along the 26S rDNA gene. This illustrates conserved and variable regions in the 26S rDNA gene. From the yeast identification point of view the most interesting are two highly variable regions. The first one is located between the positions 50 and 200, and the second one between the positions 360 and 580. Figure 4 presents the average lengths of all unique sequences detected by the algorithm at each position of 26S rDNA gene of yeast species. This histogram confirms a highly variable region located between the positions 360 and 580. Two variable regions around the position 100 and the position 170 can be observed on both histograms.

3 Conclusion

Developing text processing tools for the automatic analyses of molecular sequences is an interdisciplinary field of science, with many analogies to methodologies of analyses and understanding of natural languages. In both these fields the object of the study has a complex, hierarchical character, which resultes from natural evolution. Making progress or obtaining a new result is, very often, related to combining methods from several scientific disciplines.

The development in understanding the contents of molecular sequences proceeds along several paths. An important inspiration for progress in the field comes from the experimental side, where on one hand there is a constant increase of the throughput performance of measuring devices, and on the other there is a constant refinement in the biotechnology involving emerging techniques, possibility to study more detailed properties of biomolecules and new types of interactions between them. By using current techniques for text processing, these large data files produced in biological experiments are analyzed with the aim of extracting biological knowledge, verifying or rejecting hypotheses behind experimental plans. This can lead to a better understanding of biological models and to improvements in experimental technologies.

One important aspect of the mechanism mentioned above is that the development of experimental techniques of analyses of molecular sequences creates pressure on the computational side of textual analysis of molecular sequences, towards developing more effective and more specialized tools.

From what we have presented, it is clear that inferring information from molecular sequences is virtually impossible without specialized software tools for text processing. In this paper we have overviewed several text processing methodologies in the aspect of carrying out automatic analyses of molecular sequences. We have shown their construction, we have characterized their efficiency and computational complexity, and we have given examples of using them to infer biological knowledge.

Acknowledgment. This paper was partly supported by the European Sixth Framework Programme Project, GENEPI-lowRT, "Genetic Pathways for the Prediction of the Effects of Ionising Radiation: Low Dose Radiosensitivity and Risk to Normal Tissue after Radiotherapy", and by the University Grant BK 209/Rau-2/06

References

1. Boyer, R.S., Moore, J.S.: A fast string searching algorithm. Commun. ACM 20, 762–772 (1977)
2. Bolc, L.: Natural language generation systems. Springer, Heidelberg (1988)
3. Bolc, L. (ed.): Representation and Processing of Natural Language. Hanser-Verlag and MacMillan Press, London (1980)
4. Bolc, L., Cytowski, J.: Search Methods for Artificial Intelligence. Academic Press, London (1992)
5. Bolc, L., Borowik, P.: Many-Valued Logics: automated reasoning and practical applications. Springer, Heidelberg (1999)
6. Burkhardt, S., Kärkkäinen, J.: Fast lightweight suffix array construction and checking. In: Baeza-Yates, R., Chávez, E., Crochemore, M. (eds.) CPM 2003. LNCS, vol. 2676, pp. 55–69. Springer, Heidelberg (2003)
7. Burrows, M., Wheeler, D.J.:: A block sorting lossless data compression algorithm. Technical Report 124, Digital Equipment Corporation, Palo Alto, CA (1994)
8. Butler, J.M.: Forensic DNA Typing: Biology, Technology and Genetics of STR Markers, 2nd edn. Elsevier, Amsterdam (2005)
9. Charras, C., Lecroq, T.: Handbook of Exact String Matching Algorithms. College Publications (2004)
10. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms, 2nd edn. MIT Press and McGraw-Hill, Cambridge (2002)
11. Crochemore, M., Rytter, W.: Jewels of Stringology. World Scientific Publishing Co., Singapore (2002)
12. Dieffenbach, C.W.: General Concepts for PCR Primer Design. In: PCR Methods and Applications, pp. 530-537 (1993)
13. Dumas, J.P., Ninio, J.: Efficient algorithm for folding and comparing nucleic acid sequences. Nucleic Acids Res. 10(1), 197–206 (1981)

14. Ewens, W.J., Grant, G.R.: Statistical Methods in Bioinformatics. Springer, Heidelberg (2001)
15. Ferragina, P., Manzini, G.: Opportunistic data structures with applications. In: 41st Symposium on Foundations of Computer Science, pp. 390–398 (2000)
16. Fredkin, E.: Trie Memory. Communications of the ACM 3, 490–499 (1960)
17. Gasieniec, L., Li, C.Y., Sant, P., Wong, P.W.H.: Efficient Probe Selection in Microarray Design. In: Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, pp. 247–254 (2006)
18. Gusfield, D.: Algorithms on Strings, Trees, and Sequences. Cambridge University Press, Cambridge (1997)
19. Hamming, R.W.: Error Detecting and Error Correcting Codes. Bell System Technical Journal 26(2), 147–160 (1950)
20. Hausser, R.: A Computational Model of Natural Language Communication, Interpretation, Inference, and Production in Database Semantics. Springer, Heidelberg (2006)
21. Healy, J., Thomas, E.E., Schwartz, J.T., Wigler, M.: Annotating large genomes with exact word matches. Genome Res. 13, 2306–2315 (2003)
22. Hudson, T.J., et al.: An STS-based map of the human genome. Science 270, 1945–1954 (1995)
23. Kaplan, H., Landau, S., Verbin, E.: A Simpler Analysis of Burrows-Wheeler Based Compression. In: Lewenstein, M., Valiente, G. (eds.) CPM 2006. LNCS, vol. 4009, pp. 282–293. Springer, Heidelberg (2006)
24. Karlin, S., Altschul, S.F.: Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. USA 87, 2264–2268 (1990)
25. Karlin, S., Altschul, S.F.: Applications and statistics for multiple high scoring segments in molecular sequences. Proc. Natl. Acad. Sci. USA 90, 5873–5877 (1993)
26. Kim, D.K., Sim, J.S., Park, H., Park, K.: Linear-time construction of suffix arrays. In: Baeza-Yates, R., Chávez, E., Crochemore, M. (eds.) CPM 2003. LNCS, vol. 2676, pp. 186–199. Springer, Heidelberg (2003)
27. Knuth, D.E.: The Art of Computer Programming. Addison-Wesley, Reading (1973)
28. Knuth, D.E., Morris, J.H., Pratt, V.R.: Fast pattern matching in strings. SIAM J. Comput. 6(2), 323–350 (1977)
29. Ko, P., Aluru, S.: Space-efficient linear time construction of suffix arrays. In: Combinatorial Pattern Matching, pp. 200–210 (2003)
30. Kurtzman, C.P., Robnett, C.J.: Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DBA partial sequence. Antonie Van Leeuwenhoek Journal, 331–371 (1998)
31. Lachance, M.A., et al.: The D1/D2 domain of the large-subunit rDNA of the yeast species Clavispora lusitaniae is unusually polymorphic. FEMS Yeast Research, pp. 253–258 (2003)
32. Lebart, L., Salem, A., Barry, L.: Exploring Textual Data. Kluwer Academic Publishers, Dordrecht
33. Levenshtein, V.: Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady 10, 707–710 (1966)
34. Lippert, R.A.: Space-efficient whole genome comparisons with Burrows-Wheeler transforms. J. Comput. Biol. 12(4), 407–415 (2005)
35. Manber, U., Myers, G.: A new method for on-line searches. SIAM J. Comput. 22, 935–948 (1993)

36. Manzini, G., Ferragina, P.: Engineering a Lightweight Suffix Array Construction Algorithm. Algorithmica 40, 33–50 (2004)
37. McCreight, E.M.: A Space-Economical Suffix Tree Construction Algorithm. Journal of the ACM 23, 262–272 (1976)
38. National Center for Biotechnology Information, `http://www.ncbi.nih.gov/index.html`
39. Needleman, S.B., Wunsch, C.D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453 (1970)
40. Polański, A., Kimmel, M.: Bioinformatics. Springer, Heidelberg (2007)
41. Pokrzywa, R., Polański, A.: Exact string matching with the Burrows-Wheeler Transform. In: Proceedings of the National Conference Application of Mathematics to Biology and Medicine, pp. 87–92 (2006)
42. Pokrzywa, R.: Searching for Unique DNA Sequences with the Burrows-Wheeler Transform. Biocybernetics and Biomedical Engineering 28(1), 95–104 (2008)
43. Pokrzywa, R.: Searching for tandem repeats with the Burrows-Wheeler Transform. Submitted to Journal of Computational Biology
44. Smith, T.F., Waterman, M.S.: Identification of common molecular subsequences. J. Mol. Biol. 197, 147–195 (1981)
45. Sugita, T., Nishikawa, A.: Fungal Identification Method Based on DNA Sequence Analysis: Reassessment of the Methods of the Pharmaceutical Society of Japan and the Japan Pharmacopoeia. Journal of Health Science, 531–533 (2003)
46. STRBase: Short Tandem Repeat DNA Internet Database, `http://www.cstl.nist.gov/biotech/strbase/`
47. Thompson, J.D., Higgins, D.G., Gibson, T.J.: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22, 4673–4680 (1994)
48. Ukkonen, E.: On-line construction of suffix trees. Algorithmica 14, 249–260 (1995)
49. Weiner, P.: Linear pattern matching algorithm. In: Proceedings of the 14th Annual IEEE Symposium on Switching and Automata Theory, pp. 1–11 (1973)
50. Wesselink, J., et al.: Determining a unique defining DNA sequence for yeast species using hashing techniques. Bioinformatics 18, 1004–1010 (2002)
51. Wilbur, W.J., Lipman, D.J.: Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA 80, 726–730 (1983)

Intelligent Decision Support: A Fuzzy Stock Ranking System

Adam Ghandar[1], Zbigniew Michalewicz[1,2,3], and Ralf Zurbruegg[4]

[1] School of Computer Science, University of Adelaide, Adelaide, SA 5005, Australia,
{adam.ghandar,zbigniew}@adelaide.edu.au,
[2] Polish-Japanese Institute of Information Technology,
ul. Koszykowa 86, 02-008 Warsaw, Poland,
[3] Institute of Computer Science, Polish Academy of Sciences,
ul. Ordona 21, 01-237 Warsaw,
[4] Business School, University of Adelaide, Adelaide, SA 5005, Australia,
ralf.zurbrugg@adelaide.edu.au

Abstract. This paper presents an intelligent decision support system for financial portfolio management. An adaptive business intelligence approach combines optimization, forecasting and adaptation with application specific financial information processing and quantitative investment paradigms.

The methodology involves constructing a ranking of stocks by strength of a buy or sell recommendation which is inferred using an adapting forecasting model that considers a range of factors. These include company balance sheet information, market price and trading volume as well as the wider economy. The system adjusts its prediction model dynamically as market conditions change. An evolving fuzzy rule base mechanism encodes a model of relationships between model factors and a recommendation to buy, sell or hold securities.

Key words: Hedge Fund Management, Fuzzy Rules, Evolutionary Computation

1 Introduction

Adaptive Business Intelligence [12] is an approach that combines elements of predictive modeling, forecasting, optimization, and adaptability. Traditionally business intelligence information systems process data to obtain information; and then use statistical data analysis techniques to infer knowledge. The resulting knowledge about the business and operations is reported to end users. There is a recent trend for intelligent business information systems to recommend courses of action or even implement decisions. An intelligent component adds another type of functionality in which knowledge is applied. Intelligent decision support requires adaptation as the operating environment changes, prediction to anticipate the future where decisions have effect, and optimization to find the best possible decisions with respect to objective outcomes. The methodology

M. Marciniak and A. Mykowiecka (Eds.): Bolc Festschrift, LNCS 5070, pp. 379–410, 2009.

has been applied successfully to a wide range of real world problems [13,1]. In this paper we apply the methodology to trading financial markets.

A financial market consists of a number of listed securities which may be bought or sold. In order to profit, it is necessary to either predict price changes of a sufficient size or to do so correctly more often not so that the reward outweighs the losses. For a number of reasons this turns out to be a very difficult task. Needless to say there is considerable literature describing theories that give insight into the nature of these difficulties. A selection of reasons includes competition between participants, the effect of information on security prices and crowd psychology. Although considerable practical and theoretical difficulties exist, a number of papers have been published that provide encouragement that it is possible to trade profitably over sustained periods using fixed strategies.

One of the simplest and most successful strategies relies on the behavioral tendency of market participants to buy stocks that have recently performed well [11]. To implement the strategy a simple tactic is used in which stocks that have had the greatest price gains (price momentum) over some recent epoch are bought and held for a period. At the end of the period these are then sold and replaced by a possibly completely new set of securities. In this way a portfolio is managed so that it always comprises of a subset of listed stocks with the highest price momentum relative to all other possible choices.

From a practical viewpoint, although it might be possible to construct rules for trading profitably from study of historic data it is still the case that there are considerable practical difficulties. Some rule inputs will have changing impact and relevance over time. Rules useful for forecasting the price of one security do not work well with all others. Another problem is the large set of input data. For example in a situation with an investment universe with just $1,000$ securities there are $10,000$ closing price and volume observations every week. Over a year there may be around $2-3$ million. For even a fairly simple forecasting model where the price prediction depends on several factors the problem complexity becomes very large indeed. For example [7] describes a prediction model with 11 inputs categorized using seven linguistic descriptions, this implies 10^{318} possible relationships that are able to be expressed. Trying to form some conclusion from the dense set of figures and tables provided in the market summary of the financial pages of a newspaper illustrates some of these problems.

Fuzzy systems seem appropriate for approximating the reasoning required to weigh possible decisions to buy or sell securities. Further, Evolutionary computation provides a method to search for rules and also to update them. Changes in market conditions are able to be observed through variation in forecasting performance and analysis of new data. The approach discussed in this section essentially involves encoding a time varying asset valuation model using fuzzy rules. This model is optimized using an evolutionary process. The model is updated periodically so that decisions are based on the most recent information available. This methodology inherently avoids specifying a fixed forecasting model. By using fuzzy logic, imprecise measurements of model factors and the extent the model is fitted by current observations are considered. The model itself is also updated adaptively by relearning with recent data, and from feedback of performance out of sample. An additional benefit of fuzzy rules is that their expression is able to be interpreted readily and understood. This contrasts with methods such as neural networks. As a result the implementation avoids having a black box giving signals without justification. This allows a fund manager to validate and tune the model parameters to incorporate additional knowledge that may not be apparent from historic data.

The remainder of this paper is organized as follows. In part 2 we examine the problem in more detail and some of the issues which are considered by our methodology. Part 3 provides a description of the design of the methodology. Finally, performance test results and concluding remarks are given in part 4.

2 Description of the Problem

Our objective may be defined precisely, although perhaps naively, as a problem involving a tradeoff between two fundamentally conflicting objectives. The first is to maximize the return from investment. The second is to minimize risk. Risk in this context may be viewed as the volatility of return.

In the following subsections we discuss these issues because of their relation with the procedures actualized in the application. These procedures include portfolio construction, financial modeling and portfolio performance measurement.

2.1 Portfolio Construction

Modern portfolio theory decomposes volatility into *systematic risk* and *unsystematic risk*. The systematic risk component reflects how the changes in market conditions affect portfolio values. The unsystematic risk component is unique for each portfolio. By enforcing constraints on portfolio contents to have minimum number of different stocks from several industry sectors, it is possible to reduce the unsystematic risk component significantly so that the main source of risk is systematic which enables general application of some basic principles for risk management. This is termed *diversification*. A well diversified portfolio should have return that compensates for the systematic risk component. In this way the return may be managed with respect to risk by using mathematical models.

Let us introduce some more precise notation. Denote $r_{m,t}$ the returns at time t of the market, and $r_{p,t}$ is portfolio return, then the systematic risk β_p of portfolio p is determined by the *Capital Asset Pricing Model* (CAPM) equation:

$$r_{p,t} - r_{f,t} = \alpha_p + \beta_p(r_{m,t} - r_{f,t}) + e. \tag{1}$$

The excess return $r_{p,t} - r_{f,t}$ of any portfolio should be fully explained by its level of systematic risk β_p and the market risk premium $r_{m,t} - r_{f,t}$. In an efficient market the alpha value for portfolio returns, α_p, should be zero because it would not be possible for traders to make a profit from past data as all relevant information for pricing a security today would be incorporated in today's price. The β (risk) term explains the difference in returns by additional risk of the portfolio above the market. A positive alpha of a portfolio (or asset) can be explained predominantly by one of two possible reasons: good stock picking ability of the portfolio manager or exposure to unaccounted risk factors beyond the scope of the CAPM model. Patterns in average returns that are not explained by the standard CAPM are termed *anomalies*.

An important stage in the development of modern financial theory was the investigation of empirical evidence of these so called anomalies in the returns of some common stocks, significantly in [6]. The possibility that a number of additional factors relating to stock prices and underlying companies explain excess returns above the market index. The occurrence of pricing anomalies is greatly reduced when only two additional factors are considered as in the three-factor Fama and French asset pricing model which extends the CAPM to include company size (total market value) and price to book value in addition to the market index [6] . The price to book value ratio is the ratio of the market value of a stock to the book value. The definition of alpha is able to be extended in terms of other factors as well.

A generalizable multi-factor alpha regression model that relates return to several risk factor premiums is able to be defined precisely. If there are k factors with each having a return f_k then:

$$r_{p,t} = \alpha_p + \beta_1 f_1 + \beta_2 f_2 + \ldots + \beta_k f_k + e. \tag{2}$$

A four factor model where price momentum is the fourth factor in addition to the three factor Fama and French model is a standard used in industry and academia. As a tool to understand portfolio dynamics all additional returns or positive alpha values can be explained in terms of unconsidered risk factors. Then factors can be added to the regression model to achieve a better fit and by assumption better explain returns.

In cases where markets are not efficient and participants actions are not always rational, other explanations for anomalies can be considered. In this case the dynamics of the group behavior of market participants would be factors. If the pricing of listed market items do not accurately reflect the risk premium because of irrational pricing tendencies of market participants, then corrections would take place leading to excess returns being observed from time to time. For example securities could become *undersold* or *oversold* by participants so that prices become unreasonably high or low. Such events could be discovered by analyzing time series of stock prices in a process termed *technical analysis*.

2.2 Financial Modeling

Financial thinking has evolved during recent decades with a shift away from absolute faith in market efficiency to the position that markets are only "almost" efficient and behavioral explanations are required to account for exceptions. In an efficient market the CAPM alpha should always be zero. The implication of the changing understanding of the market is to imply that the best strategy is not necessarily to attempt to *passively* attain returns that follow the market index. Instead *active* stock picking approaches are used to attempt to attain return on investment in excess of the market.

One of the main reasons for this shift has been an increasing body of empirical results that contradict the hypothesis that the prices of stocks and other market instruments are, for the purposes of prediction, random. As a consequence *behavioral models* are used to explain some pricing effects. Some examples of patterns found in stock prices used to obtain returns in excess of the market over long periods include the profitability of momentum strategies [11]. Other technical indicator strategies such as Bollinger bands, moving averages, relative strength index [3] also have been shown to promote excess risk adjusted returns. Another category is cyclical trends, for example the "January effect" [5] where the previous year's underperforming stocks outperform in the following January because investors and managed funds sell off of underperforming assets in January.

To some extent the growing body of empirical evidence cited from academic research above is a result of developments in information technology. In fact in recent decades, computers have had a very large impact on operations at all levels in the financial sector. Advances in computing power and availability encourages application of complex mathematical models and statistical methods that may be leveraged by easy access to large volumes of data in electronic format. The culmination of this influence on portfolio management is in the rapidly expanding field of Quantitative Investment (QI). Applied to portfolio management QI is defined flexibly as "an approach to portfolio management that takes full advantage of today's better understanding of the market and greater technological capacity for sophisticated investing" [4]. Conceptually and in practice, QI involves utilization of these ideas and techniques for three main activities: return forecasting, portfolio construction and optimization, and performance measurement of resulting portfolios (Fig. 1). There is a clear feedback loop as the performance of managed portfolios over time logically should cause the model and portfolio construction methods to be either maintained or adjusted. Although it nevertheless remains an open problem to adapt quantitative trading models as quickly as a traditional analyst because of reliance on performance analysis and historical data.

Portfolio building, implicitly or explicitly, involves a valuation based on forecasting future asset prices from a basis of current knowledge. In order to effectively understand and adjust a model it is necessary to analyze the performance of portfolios managed using the forecasting model and resulting constructed portfolios over time (Fig. 1). A (conceptual) multi-factor model relating risk and return with a time component is expressed as follows:

$$r_{i,t} = \alpha_i + \beta_{i,1} f_{1,t} + \beta_{i,2} f_{2,t} + \ldots + \beta_{i,k} f_{k,t} + e_{i,t}, \tag{3}$$

where $r_{i,t}$ is the return of a stock i at time t, $f_{1,t}, \ldots, f_{k,t}$ are k returns due to factors, $\beta_{i,1}, \ldots, \beta_{i,k}, f_{k,t}$ are multipliers for the risk of including facts and $e_{i,t}$

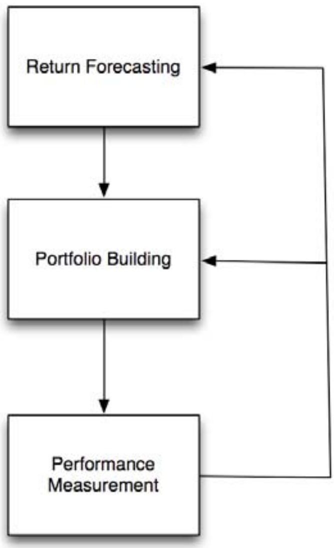


Fig. 1. The three main processes of Quantitative Equity Portfolio Management and their relation. The portfolio optimization process takes, as at least one input, information from the forecasting model. The effectiveness of the forecasting model and portfolio construction methodology can be gauged by performance measurement and includes features such as consistency of returns over time, comparison to benchmarks and so on as well as standards such as the annual rate of return.

is an error term [4]. This expression is a prototype for a prediction model that relates return to risk (by the β terms) and is also divided into model factors. It also is the case that the terms $f_{1,t}, \ldots, f_{k,t}$ can change over time to model a changing impact of factors over time. We apply computational intelligence to dynamically build a multi factor price forecasting model that anticipates and adapts the weight, possibly zero, of factors over time. The paradigm of computational intelligence is distinct from philosophies of artificial intelligence that attempt to precisely imitate human reasoning in that it involves harnessing the unique abilities of computers to produce "intelligence".

2.3 Model Factors

Now that we have presented some aspects of the forecasting model we discuss the model factors, f in equation 3. There are at least three distinct classifications of information that have been used to explain returns:

- Market and macro economic indicators
- Fundamental indicators
- Technical indicators

Macro economic indicators such as a country's gross domestic product, interest rates and also variables such as currency exchange rates, commodity prices have a significant impact on the market. Fundamental analysis is a natural approach involving consideration of the assets underlying market securities, for instance the companies whose stock is listed on the stock exchange. Using sources such as accounting data and even natural language data such as news and other reports, it is accepted that it is possible to identify assets that provide good value. Some important criteria include cash flow, total company earnings, derivative information such as the ratio of earnings to share price and others. Analysts give different importance to these factors depending on industry groups or sectors, market conditions, economic conditions and even personal experience. For example importance may be given to earnings before tax and other liabilities to give an indication of the underlying strength of a company's position, there are many reasons for variations, if for instance a firm operates in an industry that is highly regulated and subject to many taxes it may be the case that important aspects of its position relative to companies in other industries are hidden. In academia the explanatory performance of potential models is often compared with a standard such as the four factor model discussed in the previous subsection.

Technical analysis is widely used in practice. It involves constructing and applying technical analyses of price and volume movements. This approach is thought to extract information about market expectations, particularly behavioral effects. These indicators are divided into the following categories by their use in modeling different types of price movements: moving average, momentum, oscillation, and breakout indicators and also indicators based on volume, or price and volume rather than only price. Moving averages are often used to identify trends and to smooth out fluctuations due to daily or short, unsustained changes, depending on the period to calculate the average. New trends are identified when a moving average series crosses the price, or a shorter period average crosses a longer average. Oscillating indicators are used to identify cyclic patterns in price movements by compressing observations into a range, possibly giving more weight to recent points, and then generating buy or sell signals appropriately when extremes in the range are reached. Breakout indicators, as suggested by their name, are designed to catch significant changes in price direction at an early stage, for example a movement well outside the standard deviation of the mean historic returns is an indication that an unusual trend is emerging as opposed to a cyclic occurrence. Volume data is an important input component and an indicator of market sentiment with links to behavioral aspects of market activity. In general a market is considered strong by technical analysts if price and volume are both increasing.

2.4 Performance Measurement

Performance measurement is an important consideration. In this subsection we briefly discuss some of the main issues and point to some references that provide further information. A key concept is the relationship between risk and returns.

Fig. 2 shows the meaning of the relationship between risk and return illustrating sample return distributions and possible corresponding returns over time, more risky portfolios where return is justified however by the risk would have long tails in the negative area but a higher probability of positive return. Ideally by adjusting the forecasting and portfolio optimization, returns can be shaped to accurately target a specific risk profile. Although clearly the frequency of returns is not usually ideally normal, commonly, a general understanding of return properties can be obtained by a normal approximation. portfolio with a positive skewness is more likely to exhibit large returns, kurtosis is an indication of the likelihood of deviation from the mean, a lower kurtosis implies a less regular and also smaller swing away from the mean return value. Fig. 3 shows an example of the return frequency distribution obtained by testing the system described in this section.

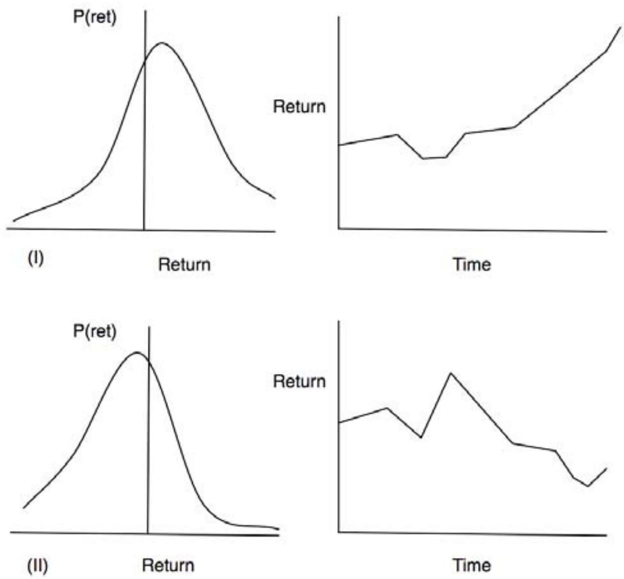


Fig. 2. Two return distributions and corresponding performance over time. The top panel is positively skewed showing returns are on average greater than they are negative, the bottom panel shows the reverse.

The *sharpe ratio* [14] measures how much excess returns (portfolio return r_p above the risk free rate r_f) investors are awarded for each unit of volatility, i.e.

$$sharpe = \frac{r_p - r_f}{\sigma_p},$$

where σ_p is the standard deviation of returns.

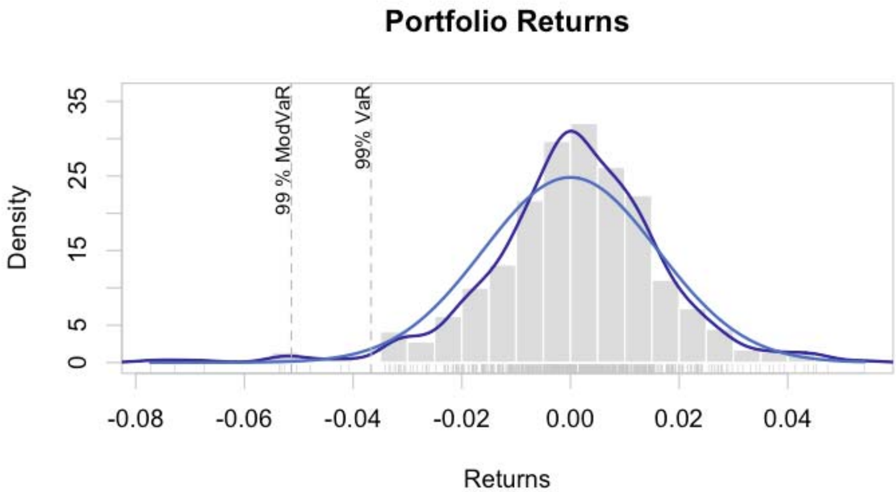


Fig. 3. Daily portfolio return frequency distribution. The actual return histogram is superimposed by a normal approximation and density. A measure of the extreme loss probability with confidence (the lowest 1 percentile approximation of expected return) is marked VaR, Value at Risk. The modified VaR is a prediction estimation which accounts for the positive skew and the kurtosis of the observed returns and the plain VaR is an estimation that assumes a normal distribution.

Another important concept is relating performance to a *benchmark*. Two useful measures are the market index and the risk free rate of return. Comparison to the market as discussed earlier in this section is an important yardstick for theoretical reasons and also techniques exist to attempt to passively follow the index without necessarily holding all stocks. If a strategy is not able to outperform the risk free rate it has no practical benefit. Examples of portfolio performance measurement are provided in [7,10]. For detailed information about this subject we refer readers to [9], of particular relevance are information content and net selectivity measures.

3 Approach for Decision Support System

In this section we describe the approach we used in constructing an adaptive stock ranking model that provides recommendations to buy, sell and hold stocks. Fig. 4 summarizes the Adaptive Business Intelligence (ABI) methodology. This approach may be divided into three main components: optimization, prediction and adaptation.

The system generates a multi factor model for stock selection. A potential for prediction exists in the universe of model factors according to academic financial research and practices in industry. This latent possibility is actualized by opti-

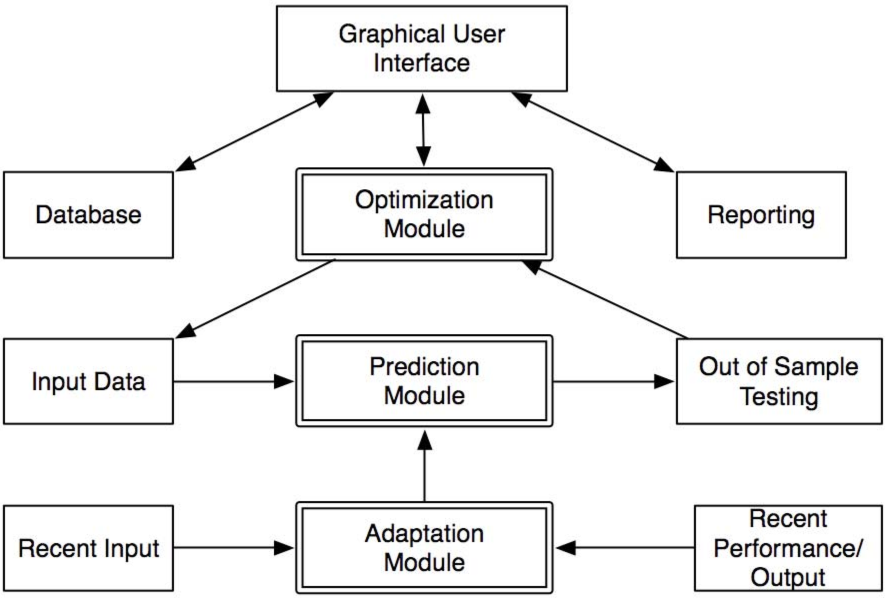


Fig. 4. Adaptive Business Intelligence (ABI). The components for optimization, prediction and adaptation are the main logical divisions in design, each requires its sub components (and each other). The optimization loop controls the recommendation presented to the user, it is predominantly a function of the input data and the result of predictions which are updated when new data is loaded and in response to feedback from recent management portfolio performance.

mization to produce a usable model using a heuristic search process. The model is represented using a fuzzy rule base structure. Optimization and prediction are coupled more closely in this financial systems than many other applications of ABI. For example in logistics or scheduling, operational requirements have a greater impact on solutions. Therefore we emphasize a distinction in terminology. Optimization refers here to the search process to optimize the forecasting model using historical data. Prediction refers to the application of this forecasting model in the future. Finally specific techniques are introduced for adapting the optimization for prediction as time passes.

3.1 Model Factors

The model inputs are the kernel of the methodology from a financial view point because the relationships between these factors and returns are from a high level the model definition. We assume and implement the methodology for equity markets and in the discussion we use the term stock. However the principles can be applied to other listed market items such as options, warrants, CFD's and other instruments.

From a top down view point the inputs are divided into macro economic factors which operate at a global level such as GDP, the value of the whole market and also information related to sectors of the economy in which individual companies conduct their business such as the resources sector, consumer discretionary sector etc. Fundamental company information also exists outside the market but is specific to each stock, this information includes information about cash flows, earnings an so on (see section 2). A separate category is data from trading activity within the market and includes things such as price and volume series and market capitalization. The raw inputs and their meaning is given in table 1. The processing of these raw inputs is summarized in Fig. 6.

The model factors are constructed from these elements. In this approach using processed data, a structure for the universe of models is imposed in a way that influences the model and the optimization process to use some predefined values. Techniques such as genetic programming [10] could be used to optimize the factors and find equations directly from raw data. In practice there is a balance between on the one hand providing no pre processing and on the other defining the factors restrictively so as to prevent exploration outside a specific type of model. Fig. 6 shows a matrix of the type of data and an analysis method for this structure to summarize how technical, fundamental and macro economic data is processed to examine attributes of the economy, stocks and investor behavior. The number of processed model factors is termed the breadth of the model.

The fundamental value factors are derived from 10 basic data sources (see Fig. 6). For each fundamental element with respect to a stock s and at time t, $X_{s,t}$ the factors X-industry and X-industry-momentum are calculated as follows

$$X\text{-industry}_{s,t} = \frac{X_{s,t}}{X_{1,t} + \ldots + X_{k,t}},$$

$$X\text{-industry-momentum}_{s,t} = \frac{X\text{-industry}_{s,t}}{X\text{-industry}_{s,t-period}},$$

where $X_{1,t}, \ldots, X_{k,t}$ are the values of the X element for all stocks in the industry sector of s and *period* is a number of time units. X-industry normalizes X to stocks industry sector, X-industry-momentum tracks the rate of change in X. This normalization with respect to industry is because companies in different sectors have common attributes shared with other businesses in the same sector, the growth or decline with respect to the sector and in general is tracked by the momentum factors. In addition a variable industry sector identifies a stocks sector and also the relative placement in the sector with respect to market capitalization are included here.

The market factors include market capitalization and a number of technical indicators found using price and volume. Market capitalization from the MV input file is directly imported as a factor and enables the model to differentiate between small and large capitalization stocks. The money flow index (MFI)

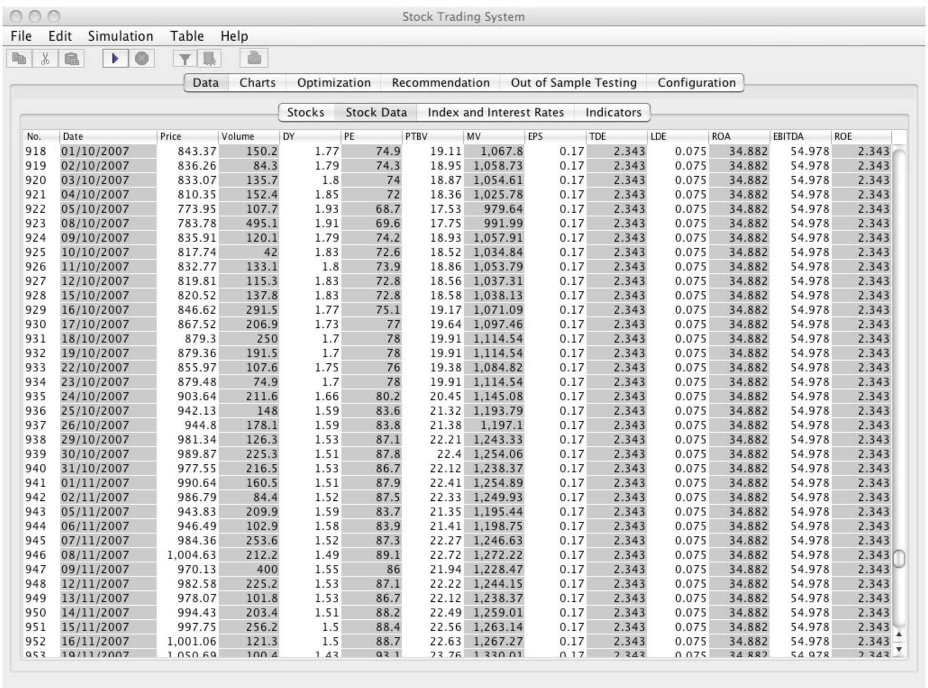
Table 1. Raw input types. The model factors are derived from these basic data types.

Name	Description
	Macro Data
INDEX	the market index, a weighted average by market cap of the value of listed equities.
RF RATE	interest rates for short term government bonds (3 months).
GOLD	the spot price of gold in USD.
OIL	the price of a barrel of crude oil in USD.
	Fundamental Data
DY	dividend yield for the company. A percentage value of the dividend income earned over the stock price.
PTBV	price to book value for a company. Literally calculated as stock value over accounting book value.
PE	price earnings ratio for a company. Calculated as the price of a stock divided by earnings per share.
PE2	a forecast of price earnings ratio for the next year by financial analysts.
MV	the market capitalization of a company. Calculated as the company stock price multiplied by the number of shares.
EPS	earnings per share.
TDE	total debt to equity ratio.
LDE	long term debt to equity ratio (> 1 year) .
EBITDA	earnings before interest and tax.
ROA	return on assets.
ROE	return on equity.
	Technical Market Data
PRICE	daily close prices.
VOLUME	daily trading volume.

attempts to track the rate of capital flow into and out of stocks by relating price and volume. The calculation is with respect to a period, p, at time t:

1. if $price_{t-p} > price_{t-p-1}$ then $MF^{+}_{t} = MF^{+}_{t-1} - (price \times volume)$,
2. if $price_{t-p} < price_{t-p-1}$ then $MF^{-}_{t} = MF^{-}_{t-1} - (price \times volume)$,
3. if $price_{t} > price_{t-1}$ then $MF^{+}_{t} = MF^{+}_{t-1} + (price \times volume)$,
4. if $price_{t}t < price_{t-1}$ then $MF^{-}_{t} = MF^{-}_{t-1} + (price \times volume)$,
5. $MFI_t = MF^{+}_{t} / MF^{-}_{t}$.

The technical factors listed in Fig. 6 analyze various characteristics of price and volume time series. PPO (percentage price oscillator) and PVO (percentage volume oscillator) emphasize cyclical patterns. The percentage oscillator is calculated by taking the ratio between a longer and shorter moving average. Std. Dev is a running standard deviation over the previous 3 months. A longer and shorter period price momentum tracks the rate of change in price during a previous period. Bollinger bands are lines a standard deviation from the mean around a stocks price — when the price moves outside, a signal is generated. In addition a variable for each stocks CAPM alpha and beta that provide an indication of if a stock is over or under priced with respect to the capital asset pricing model are included (see section 3). The macro indicators (Fig. 6) relate the previous analysis to changing economic conditions over the longer term by measuring the growth of the index, change in interest rates, gold and oil prices. The MA variables contain a ratio of the current value of the indicator to a long term moving average. Momentum again tracks the rate of change in either positive or negative directions. This enables the model to consider wider economic trends and cycles depending on the length of the averages.



No.	Date	Price	Volume	DY	PE	PTBV	MV	EPS	TDE	LDE	ROA	EBITDA	ROE
918	01/10/2007	843.37	150.2	1.77	74.9	19.11	1,067.8	0.17	2.343	0.075	34.882	54.978	2.343
919	02/10/2007	836.26	84.3	1.79	74.3	18.95	1,058.73	0.17	2.343	0.075	34.882	54.978	2.343
920	03/10/2007	833.07	135.7	1.8	74	18.87	1,054.61	0.17	2.343	0.075	34.882	54.978	2.343
921	04/10/2007	810.35	152.4	1.85	72	18.36	1,025.78	0.17	2.343	0.075	34.882	54.978	2.343
922	05/10/2007	773.95	107.7	1.93	68.7	17.53	979.64	0.17	2.343	0.075	34.882	54.978	2.343
923	08/10/2007	783.78	495.1	1.91	69.6	17.75	991.99	0.17	2.343	0.075	34.882	54.978	2.343
924	09/10/2007	835.91	120.1	1.79	74.2	18.93	1,057.91	0.17	2.343	0.075	34.882	54.978	2.343
925	10/10/2007	817.74	42	1.83	72.6	18.52	1,034.84	0.17	2.343	0.075	34.882	54.978	2.343
926	11/10/2007	832.77	133.1	1.8	73.9	18.86	1,053.79	0.17	2.343	0.075	34.882	54.978	2.343
927	12/10/2007	819.81	115.3	1.83	72.8	18.56	1,037.31	0.17	2.343	0.075	34.882	54.978	2.343
928	15/10/2007	820.52	137.8	1.83	72.8	18.58	1,038.13	0.17	2.343	0.075	34.882	54.978	2.343
929	16/10/2007	846.62	291.5	1.77	75.1	19.17	1,071.09	0.17	2.343	0.075	34.882	54.978	2.343
930	17/10/2007	867.52	206.9	1.73	77	19.64	1,097.46	0.17	2.343	0.075	34.882	54.978	2.343
931	18/10/2007	879.3	250	1.7	78	19.91	1,114.54	0.17	2.343	0.075	34.882	54.978	2.343
932	19/10/2007	879.36	191.5	1.7	78	19.91	1,114.54	0.17	2.343	0.075	34.882	54.978	2.343
933	22/10/2007	855.97	107.6	1.75	76	19.38	1,084.82	0.17	2.343	0.075	34.882	54.978	2.343
934	23/10/2007	879.48	74.9	1.7	78	19.91	1,114.54	0.17	2.343	0.075	34.882	54.978	2.343
935	24/10/2007	903.64	211.6	1.66	80.2	20.45	1,145.08	0.17	2.343	0.075	34.882	54.978	2.343
936	25/10/2007	942.13	148	1.59	83.6	21.32	1,193.79	0.17	2.343	0.075	34.882	54.978	2.343
937	26/10/2007	944.8	178.1	1.59	83.8	21.38	1,197.1	0.17	2.343	0.075	34.882	54.978	2.343
938	29/10/2007	981.34	126.3	1.53	87.1	22.21	1,243.33	0.17	2.343	0.075	34.882	54.978	2.343
939	30/10/2007	989.87	225.3	1.51	87.8	22.4	1,254.06	0.17	2.343	0.075	34.882	54.978	2.343
940	31/10/2007	977.55	216.5	1.53	86.7	22.12	1,238.37	0.17	2.343	0.075	34.882	54.978	2.343
941	01/11/2007	990.64	160.5	1.51	87.9	22.41	1,254.89	0.17	2.343	0.075	34.882	54.978	2.343
942	02/11/2007	986.79	84.4	1.52	87.5	22.33	1,249.93	0.17	2.343	0.075	34.882	54.978	2.343
943	05/11/2007	943.83	209.9	1.59	83.7	21.35	1,195.44	0.17	2.343	0.075	34.882	54.978	2.343
944	06/11/2007	946.49	102.9	1.58	83.9	21.41	1,198.75	0.17	2.343	0.075	34.882	54.978	2.343
945	07/11/2007	984.36	253.6	1.52	87.3	22.27	1,246.63	0.17	2.343	0.075	34.882	54.978	2.343
946	08/11/2007	1,004.63	212.2	1.49	89.1	22.72	1,272.22	0.17	2.343	0.075	34.882	54.978	2.343
947	09/11/2007	970.13	400	1.55	86	21.94	1,228.47	0.17	2.343	0.075	34.882	54.978	2.343
948	12/11/2007	982.58	225.2	1.53	87.1	22.22	1,244.15	0.17	2.343	0.075	34.882	54.978	2.343
949	13/11/2007	978.07	101.8	1.53	86.7	22.12	1,238.37	0.17	2.343	0.075	34.882	54.978	2.343
950	14/11/2007	994.43	203.4	1.51	88.2	22.49	1,259.01	0.17	2.343	0.075	34.882	54.978	2.343
951	15/11/2007	997.75	256.2	1.5	88.4	22.56	1,263.14	0.17	2.343	0.075	34.882	54.978	2.343
952	16/11/2007	1,001.06	121.3	1.5	88.7	22.63	1,267.27	0.17	2.343	0.075	34.882	54.978	2.343
953	19/11/2007	1,050.69	100.4	1.43	93.1	23.76	1,330.01	0.17	2.343	0.075	34.882	54.978	2.343

Fig. 5. Screenshot from the application showing raw data view.

3.2 Model Representation

A fuzzy rule is a causal statement that has an *if-then* format. The *if* part is a series of conjunctions describing properties of some linguistic variables using

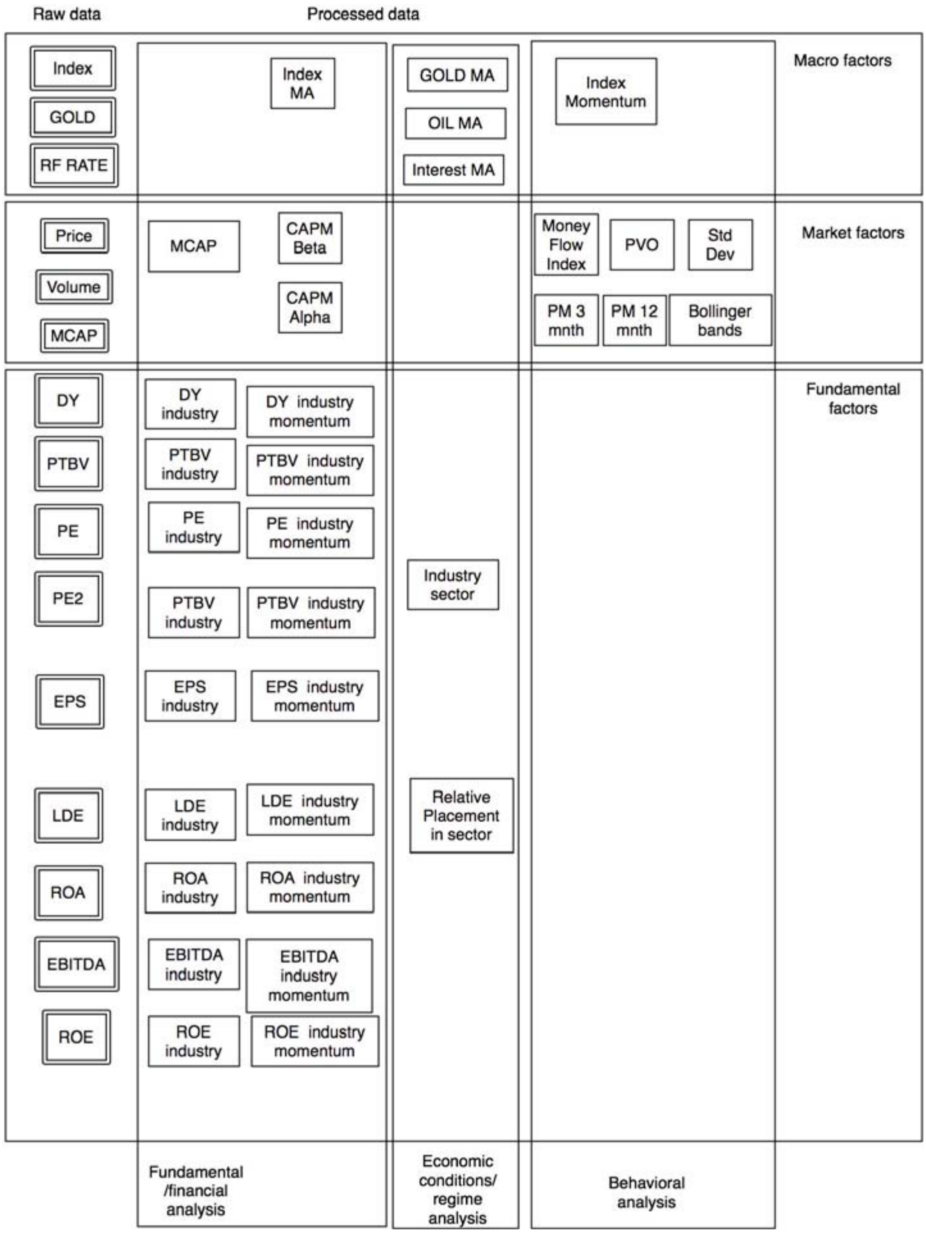


Fig. 6. Multi factor stock market valuation model.

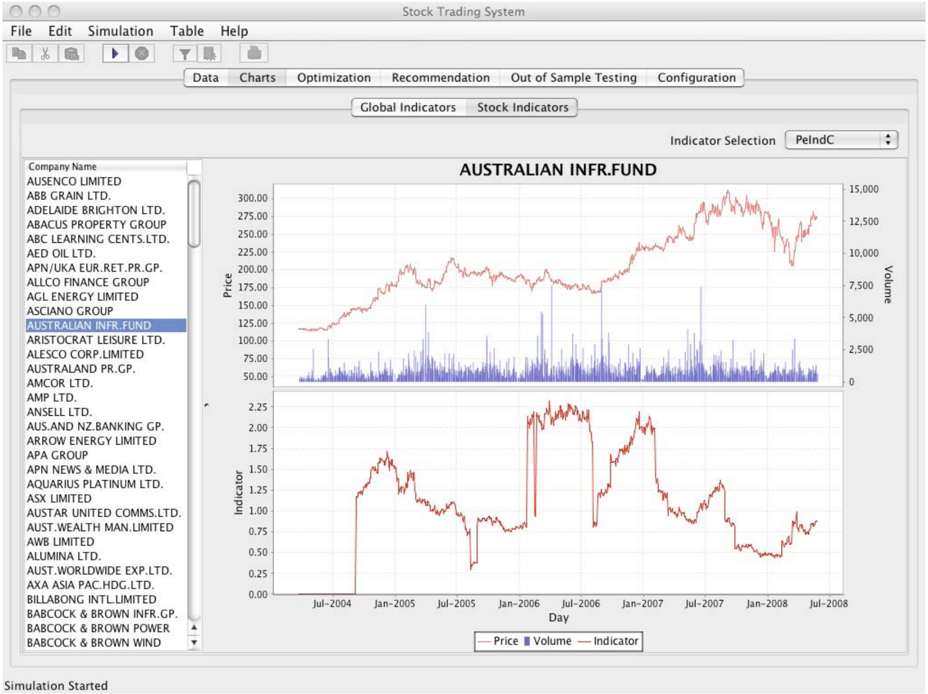


Fig. 7. Screenshot from the application showing momentum indicator for price to earnings ratio.

fuzzy sets that, if observed, give rise to the *then* part. The *then* part is a value that reflects the consequence given the case that the *if* part occurs in full. A rule base consists of several such rules. For some input observations of the linguistic variables a rule base is evaluated using fuzzy operators to obtain a classification given the (possibly partial) fulfillment of each rule by that input.

A rule base model is able to be represented as follows:

– If *PE Industry Momentum* is *Extremely Low* then *rating* = 0.9
– If *Price Momentum* is *High* and *DY Industry* is *Very High* then *rating* = 0.4

The *if* part specifies a relationship between factors (linguistic variables) and output of the rule is a weighting for the rule when the set combined. The output or rule weighting is a discrete value between 0 and 1, let us define d as the number of discrete output ratings. The rule base specifies a pattern in the data. For a data vector containing observations of crisp values the evaluation of the rule base is 1 if the observations are exactly the pattern the rule specifies. It is less than 1 but greater than 0 for partial matching of the pattern. And 0 implies no match at all.

For a rule base prediction model with breadth L there is one linguistic variable for each factor $\{f^1, ..., f^L\}$. Each of these variables, f^i, is described by a set of m_i linguistic descriptions of that factor using a set of corresponding triangular membership functions μ_j^i, $j = 1, ..., m_i$. In the example above DY Industry and Price Momentum are model factors and High, Extremely Low etc are linguistic descriptions. The membership functions map an observation of the corresponding variable to a degree of membership in fuzzy sets Low, High etc. They have a standard form where each is characterized by a triple $(min, center, max)$. min and max are the extreme values that comprise the membership set, $center$ is a value that maps to full membership of the set:

$$\mu_j^i(x) = \begin{cases} 0 & : \ x < min \text{ or } x > max \\ (x - min)/(center - min) & : \ min \leq x < center \\ (x - max)/(center - max) & : \ center \leq x \leq max \end{cases}$$

Given a vector $\boldsymbol{x}$ of L observations, one for each factor, a fuzzy rule r_1 containing linguistic descriptions $D_{i_1} \ldots D_{i_K}$, $K < L$, of factors $f^1, ..., f^L$ is written as follows:

$$r_1 = \text{ If } f_r^{i_1} \text{ is } D_{i_1} \ \cdots \ \text{ and } f_r^{i_K} \text{ is } D_{i_K} \text{ then } o_1$$

is evaluated by:

$$rule_r(\boldsymbol{x}) = o_r \prod_{i=1}^{K} \mu_j^i(x_i).$$

In general a rule base B may contain several such rules $r_1, \ldots, r_R$. Rules in a rule base are aggregated accounting for the weight of each rule, $o_1, \ldots o_R$, to obtain a real value as follows:

$$\rho_B(\boldsymbol{x}) = \frac{\sum_{m=1}^{R} rule_m(x)}{\sum_{m=1}^{R} o_m}. \tag{4}$$

The number of possible rule bases is quite large. Each rule may specify up to L factors, each of these factors f^i is described by one of M_i membership sets that relate to it. There are also d possible output ratings. So for a single rule the number of possible variations is

$$v_r = d \times \sum_{k=1}^{L} M_i^k \times \binom{L}{k}$$

and the number of possible rule bases with R rules is therefore of the order v_r^R.

Figure 7 gives a typical example of a chart showing a stock price and volume time series along with an indicator. Graphs similar to this are found in any trading or stock price report and a trading system would be incomplete without such a presentation. Many books have been written discussing the intricacies of interpreting such graphs. Although visualization is an improvement over data tables as a way to make trading decisions it is clearly not possible for an analyst to examine any more than a few assets in this way.

3.3 Comparing Models

Using equation 4 a rating for a stock by a rule base on a day t is obtained using observations of the values of the model factors for the stock on the day t. Let us use a subscript to denote this rating such that $rating_{s,t}$, means the rating for a particular stock s on day t. A ranking is defined here as a set of stocks ordered by some value associated with each stock. in this case let us assign this value to be the output of a rule base given data for the stock on a particular day. For a set of stocks $M = \{s_1, s_2, \ldots, s_m\}$ a rule base can be applied to order a set of stocks to obtain a ranking of all stocks in the set on the day t as follows:

$$R_t^{\rho}(M) = [(s_{1,t}, \rho(\boldsymbol{x}_{s_1,t}), \ldots, (s_{m,t}, \rho(\boldsymbol{x}_{s_m,t})] \,, \tag{5}$$

where $\rho(\boldsymbol{x}_{s_i,t}) \geq \rho(\boldsymbol{x}_{s_{i+1},t})$. Each element of a ranking is a pair comprising the ranked stock and its rating, $(s_{i,t}, \rho(\boldsymbol{x}_{s_i,t}))$, has rank $i \in \mathbb{Z}, i \geq 1$.

The fitness of a rule base is defined to be a measure of its ability to rank stocks by return ordering over a specific period of time in the future termed a forecast horizon of length H days. An ideal return ordering $R_{ideal,t}$ for a day t is constructed by looking forward H days into the future within the available training data. Then we find the average price for a stock, $p_{s,H}$ during a P day period starting after the H'th day in the training data window. As this operation takes place in training data we may assume that for every stock the return during the period is simply:

$$r_{s,t,H} = p_{s,H} - p_{s,t},$$

where $p_{s,H}$ is the average price over a period starting H days after t. The reason the average price is used is to avoid the fitness being overly sensitive to fluctuations in stock price on particular dates. An ideal ranking for days in the training window for comparing the rule base ranking with is able to be defined as follows:

$$R_t^{ideal}(M) = [(s_{1,t}, r_{s_1,t,H}), \ldots, (s_{m,t}, r_{s_m,t,H})] \,. \tag{6}$$

Evaluation functions are defined to compare the similarity of rankings, $R_{\rho,t}$, from rule bases to optimal rankings with hindsight ,$R_{ideal,t}$. In the application rule bases are tested using input data from several days so that rule base fitness is not overly dependent on patterns in a single day of training data.

As a step to constructing an evaluation function let us initially define a comparison operator for comparing rankings. An obvious method for comparing the ordering of A and B is to count the number of times the same stock has an identical rank in both. However it is preferable that the method should be more lax for a number of reasons, including for accuracy. Two rankings would be defined as very different if the rankings were out of sync by even a single element. Trying to find rules to predict a very specific ranking property would likely lead to over fitting and loss of generalization. In addition the ordering of stocks within the top percentile is not relevant since all stocks in this group are, relative to the others, recommended to buy. For these reasons and to make the optimization task easier we use a flexible approach for comparison designed to be sensitive to very small changes in similarity due to any change in the ranking order.

Given two rankings A and B that order stocks in a set $M = \{s_1, s_2, \ldots, s_m\}$ we define two corresponding sub rankings:

$$a = [(s_{a_1}, r_{s_{a_1}}), \ldots, (s_{a_{u_1}}, r_{s_{a_{u_1}}})],$$

and

$$b = [(s_{b_1}, r_{s_{b_1}}), \ldots, (s_{b_{u_2}}, r_{s_{b_{u_2}}})],$$

with sizes $u_1, u_2 < m$ containing the highest u_1 and u_2 rated stocks in A and B respectively. We construct two sets of stocks which are subsets of M

$$a_s = \left\{s_{a_1}, \ldots, s_{a_{u_1}}\right\}$$

and

$$b_s = \left\{s_{b_1}, \ldots, s_{b_{u_2}}\right\}.$$

A real value measure of similarity of two rankings A_M, B_M defined over the set of m stocks M is then found by the operator:

$$\text{similarity} : R_M \times R_M \mapsto \mathbb{R}$$

$$\text{similarity}_{u_1,u_2}(A_M, B_M) = \frac{|\ a_s \cap b_s\ |}{\max(|a_s|, |b_s|)}. \tag{7}$$

where $u_1, u_2 \in \{u \in \mathbb{Z} | 0 \leq u \leq m\}$. The meaning is interpreted as the number of stocks from the top u_1 of A_M that are also in the top u_2 of B_M.

Now we define the evaluation function that uses a training window of length $horizon + 2 \times period$ where H is the forecast horizon and P is a fixed period of sequential days that is both the number of days used to test the rule base and also the number of days used to calculate average values for the ideal ranking. Let the first day in the training window be denoted day T then:

$$eval_{buy,H}(\rho_B) = \sum_{t=0}^{P} \frac{\text{similarity}_{l,q}(R_{T+t}^{\rho_B}, R_{T+P}^{ideal})}{P}, \tag{8}$$

where, $R_t^{\rho_B}$ is a ranking from a rule base ρ_B with respect to a set of listed stocks and R_{T+P}^{ideal} is an ideal ranking of stocks at a day taken at the end of the possible training testing days. The parameters l and q may be tuned by experimentation or adaptively. In optimization it is easier to try to find any stocks in the top of the ideal ranking rather than a specific ordering. Another fitness function is also used in the system to measure the ability of rule bases to rank stocks by likelihood of decreasing value. This function is defined in a similar way, the only difference is that the order of the ideal ranking is reversed:

$$eval_{sell,H}(\rho_B) = \sum_{t=0}^{P} \frac{\text{similarity}_{l,q}(R_{T+t}^{\rho_B}, \text{reverse}(R_{T+P}^{ideal}))}{P}. \tag{9}$$

3.4 Optimization

This component optimizes a rule base to perform well in ranking stocks relative to their increase in price over a prediction horizon. Input data consisting of a historic data window is used as training data, this window is updated each time the system is used to rank stocks for trading decisions. In the subsection 3.3 we discuss the fitness function for ranking stocks. There are a very large number of possible rule bases. To handle this a combination of local search algorithms and an evolutionary algorithm are used in sequence:

1. Restricted initial search,
2. Evolutionary algorithm,
3. Hill climber search.

The processes are also initially guided by past searches, a new optimization with a fixed length sliding data window takes place each time the system is used for trading and a repair operator fixes the rules to be focused around the area neighborhood of the previous optimal solution. In addition a rule structure is fixed over the whole population periodically during the evolutionary search in a way inspired from natural human reasoning to enable the system to follow a "line" of reasoning through the solution space. Figure 8 shows a screenshot of a panel to view the progress of optimization in the graphical user interface.

A measure of the generalization ability of rule bases is also used in the optimization process to try to find rules that work better in general rather than only in the training window. Two generalization criteria are used, the first counts the average number of ranked stocks over a period before and in the end of the training window, both are separate from the days used to generate rankings to test.

Representation and Alteration of the Genotype. First of all, the rules are encoded using array structures. The maximum number of possible rules in a prototype solution is fixed but the actual number can vary in the range 0 to this maximum setting $rules_{max}$. A boolean vector UR of length $rules_{max}$ indicates if a rule is used. A $rules_{max} \times L$ matrix I indicates by integer values from 1 to j corresponding to membership functions μ_j^i for each column f_i that is specified in each rule. An additional boolean $rules_{max} \times L$ matrix UI indicates if a factor f^i is used in a rule. A vector of floating point values O with length $rules_{max}$ indicates the output weighting for each rule. Fig. 3.2 clarifies this definition and superimposes these four arrays, UR, I, UI and O, into a single genotype visualization.

There are a number of variation operators that are defined over the genotype representation to search for new solutions by mutation (variation of an existing rule base to get a new one) and cross over (combination of two or more existing rule bases to get a new one). Table 2 lists these operators.

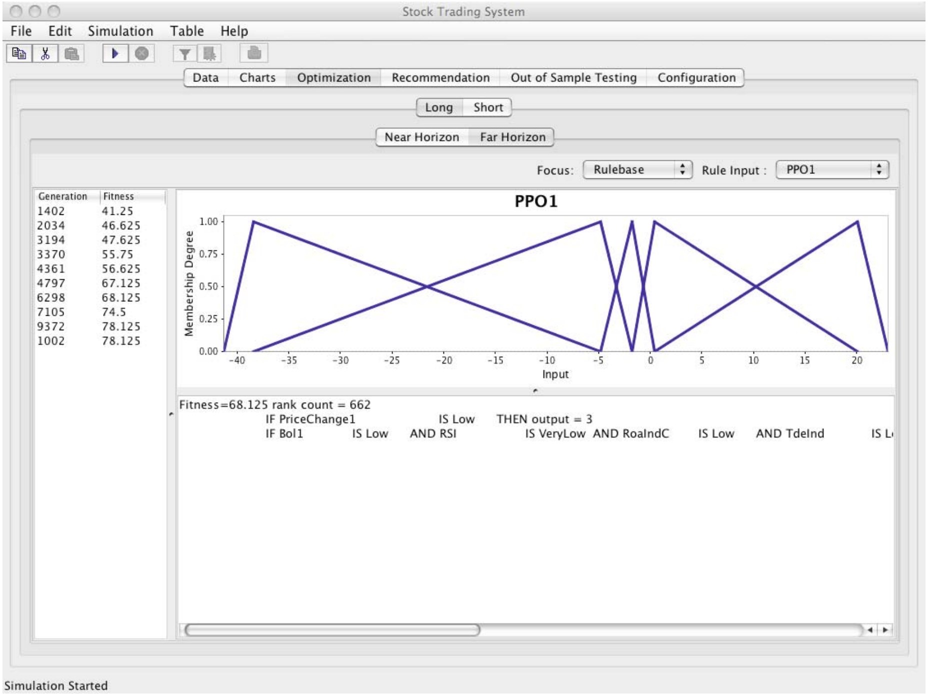


Fig. 8. The optimization process in the application GUI.

If PPO	is EL			then rating = 0.8
If DYIND	is L	and DYINDM	is VH	then rating = 1.0
If MFI	is M	and PEIND	is VH	then rating = 0.2

(a) Example phenotype rule base representation

use	f_1		f_2		f_3		f_4		f_5		f_6		f_7		$\cdots$	f_L		o
B	B	I	B	I	B	I	B	I	B	I	B	I	B	I	$\cdots$	B	I	F
B	B	I	B	I	B	I	B	I	B	I	B	I	B	I	$\cdots$	B	I	F
B	B	I	B	I	B	I	B	I	B	I	B	I	B	I	$\cdots$	B	I	F
B	B	I	B	I	B	I	B	I	B	I	B	I	B	I	$\cdots$	B	I	F
B	B	I	B	I	B	I	B	I	B	I	B	I	B	I	$\cdots$	B	I	F

(b) Example genotype rule base representation.

Fig. 9. The phenotype rule base representation is shown in (a) and the genotype or internal encoding is depicted in (b). The genotype shows a rule base representation for a rule base with 5 rules and L factors). B indicates a boolean value: $B \in \{T, F\}$; I an integer: $I \in \{1, 2, 3, 4, 5, 6, 7\}$; and F a float: $F \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$.

Table 2. Operators, all act on rule bases ρ.

Operators		
Type	Description	Output (new ρ)
Mutation	μ-SMALL $\rho \mapsto \rho$	selects an element with equal probability from *UR*, *I*, *UI* or *O* and increments or decrements with equal probability by 1 appropriate unit (e.g. if boolean goes to false, if integer add or subtract 1).
Mutation	μ-LARGE $\rho \mapsto \rho$	selects an element with equal probability from *UR*, *I*, *UI* or *O* and sets to a new legal value chosen with equal probability.
Mutation	μ-SMART $\rho \mapsto \rho$	selects an element with equal probability from I or O such that UR, UI are true and increments or decrements with equal probability by 1
Crossover	γ-UX $\rho \times \rho \mapsto \rho$	uniformly select elements from *UR*, *I*, *UI* or *O* with equal probability to construct a new rule base from two parents
Crossover	γ-RULE $\rho \times \rho \mapsto \rho$	swap rows from *UR*, *I*, *UI* or *O* to construct a new rule base with whole rules from two parents
Repair	r-SAME $\rho \times \rho \times [0,1] \mapsto \rho$	genes in from ρ_1 are changed to be the same as ρ_2 until p, $0 < p < 1$ percent genes in ρ_1 are the same as ρ_2
Repair	r-LEGAL $\rho \mapsto \rho$	genes in from ρ are changed until there are noillegal values (e.g integers in I less than 0 or greater than the number of membership functions for the corresponding variable)
Repair	r-FIXED $\rho \times \rho \mapsto \rho$	genes from I and O in ρ_1 where UR and UI are true in ρ_2 are overwritten by corresponding values from ρ_2

Heuristics. Initially, we search by enumeration the C_2^L single rules that are a possible combinations of 2 factors. A percentage of the best of these are inserted into the initial population before the evolutionary search. The best rule from the initial search is fixed across the population by applying the operator r-FIXED (see table 2). The evolutionary algorithm is as follows:

1. Initialize population $P = \langle \rho_1, \rho_2, \ldots, \rho_n \rangle$ of n rule base individuals where $g\%$ are from the enumeration search and the remainder are random
2. Initialize variables from parameter file: *cBEST*, *BEST*, *SWI*, *generation*, *FIXED*, $rules_{max}$
3. Apply r-SAME to the whole population using a previous search's best rule base and double parameter $p \in [0, 1]$ if available
4. Evaluate each solution: calculate $eval(\rho_v)$ for $v = 1, \ldots, n$
5. Identify the best solution,*cBEST* in P

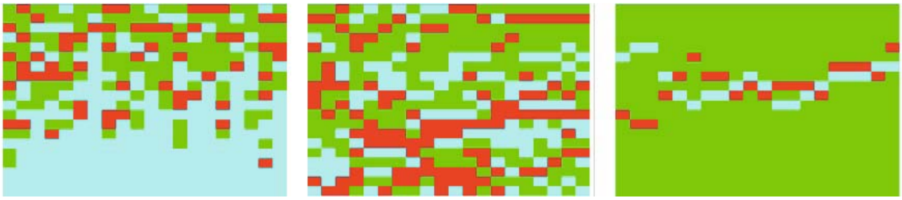


Fig. 10. Visualization of the phenotype illustrating the increase in fitness during the evolutionary process. Each pane represents the top percentile of a series of stock rankings for the same days in training data found from different fuzzy rule bases. The grey segments represent stock selections which, in training, were from the top risers over the forecast horizon (3 months), dark segments represent selections which were in the worst half of the possible choices and light grey means there was no recommendation or the selection did not rise or fall significantly. The objective function was to maximize the selection of the best choices. The first panel shows the best solution at 1402 generations, the middle panel after 3370 and the right was after 49,149 generations.

6. If $eval(cBEST) > eval(BEST)$ then $BEST = cBEST$ and $SWI = 0$
7. Alter the population by applying mutation and crossover operators (tournament selection of size 2 is used)
8. Apply r-LEGAL operator to each offspring with respect to the best solution $best_{previous}$ from the previous generation (elitism is not used)
9. If parameter to use fixed rules is used, apply r-FIXED operator to each offspring and a single global fixed rule base
10. If parameter is set, apply r-SAME operator to each offspring using the current *best*
11. $generation = generation + 1$
12. $SWI = SWI + 1$
13. Repeat steps 3–12 successively until SWI is $maxSWI$ and no improvement is recorded
14. If $generalization(BEST) > generalization(FIXED)$ set $FIXED = BEST$, $generation = 0$, $rules_{max} = rules_{max} + 1$ and return to step 3; else return P

where *generalization* is a function to return a value that measures the generality of solutions, for example by counting the ranked stocks or testing with different historical data. Fig. 5 shows a visualization of the change in the ranking of stocks implied by rule bases that occurs during the evolutionary process.

After the evolutionary algorithm has completed a hill climber algorithm successively improves the the top $T\%$ of the individuals from the evolutionary algorithm using the mutation operators. Its is as continues from the EA as follows:

1. HCP = top $T\%$ of P
2. evaluate each solution ρ in HCP
3. apply mutation to each to get the next generation of the HC heuristic
4. repeat steps 2–3 hc-step times
5. return the final best individual from HCP

3.5 Prediction

Partly to enhance predictive capacity, a meta repair operator is used across subsequent search processes which links real out of sample performance to each optimization over a historic window. This operator, r-SAME, [7] focuses the search within the neighborhood of previous solutions that worked well out of the sample. In addition several other methods were found to be useful for this purpose of increasing the generality of solutions including lengthening the data window beyond the forecast horizon so that the model is "fitted" on several different periods, a greater weight is then given to recent data to balance this requirement with an adaptive capacity.

3.6 Adapting the Model

The system is applied using a sliding data window framework so that as new data is observed the optimization procedures are rerun to obtain a model that matches the latest data. In addition the model is adjusted depending on the performance of a real portfolio managed by the system which is of course out of sample.

The repair operator r-SAME is used across fitness procedures to alter the genotype in such a way that it is no more than p percent different from a second genotype. When the optimization process is initialized the second genotype is the solution from the previous window. The parameter p is adjusted depending on the performance of a real portfolio in relation to the index which serves as a benchmark. It is reduced when performance is worse than the benchmark and increased if the real portfolio is out performing the benchmark. The rationale is to focus the search close to solutions while they give good performance and to broaden the search when this performance decays.

To calculate p such that it varies depending on portfolio performance is as follows:

$$p = \begin{cases} \frac{sharpe(r_{p,t})(1+k)}{2} & \text{if } \; r_{p,t} < r_{p,t-l} \\ \frac{sharpe(r_p,t)(1-k)}{2} & \text{if } \; r_{p,t} > r_{p,t-l} \\ 0 & \qquad\text{otherwise,} \end{cases}$$

where $r_{p,t}$ is the return recorded for a portfolio managed by the system over one month prior to day t, $k, 0 < k < 1$ is a constant for the sensitivity to changes, l is a parameter of the distance to look behind period to measure performance. l was set equal to the interval in days between portfolio rebalancing events. Further $sharpe(r_{p,t}) := (r_{p,t} - r_{f,t})/(\sigma(\boldsymbol{r}_{p,t}))$ for risk free rate return $r_{f,t}$ and a vector of daily portfolio returns for the month before t, $\boldsymbol{r}_{p,t}$.

The length of a sliding window is set according to observed recent performance volatility. The base window length, max length, in months is reduced according to the formula:

$$\text{window length} = e^{kln(sharpe(r_{p,t}) - sharpe(r_{m,t}))} \times \text{max length},$$

where k is constant set to control the sensitivity of the window change, $sharpe(r_{p,t})$ and $sharpe(r_{m,t})$ are the sharpe ratios of the managed portfolio and the benchmark index retuns, and $maxLength$ is the maximum window length. The new window date is never set earlier than the previous window start date. In the case that the window length is calculated as earlier than this, the previous start date is used.

3.7 Constructing Recommendations

Four separate optimization procedures are run using variations of the fitness evaluation functions to obtain rankings of stocks that are predicted to fall over a short and long horizon and that are forecast to rise also over a short and long horizon. This results in four rule base solution models with fitness x_1, x_2, x_3 and x_4 respectively

1. $\rho^{x_1}_{buy,H_1}$
2. $\rho^{x_2}_{buy,H_2}$
3. $\rho^{x_3}_{sell,H_1}$
4. $\rho^{x_4}_{sell,H_2}$

From these rule bases we can find four corresponding stock rankings on a day T after the end of the training data when we wish to apply the rules to construct a recommendation. Buy rankings $R_T^{\rho^{x_1}_{buy,H_1}}$ and $R_T^{\rho^{x_2}_{buy,H_2}}$ are generated using an evaluation function of type $eval_{H,buy}$ (equation 7) that is designed to pick stocks that increase in value over the forecast horizon. And the other two rankings highly rank stocks that are forecasted to decline in price. $R_T^{\rho^{x_3}_{sell,H_1}}$ and $R_T^{\rho^{x_4}_{sell,H_2}}$ are produced using $eval_{H,sell}$ (equation 9). Two forecast horizons, $H_1 > H_2$, are used to obtain predictions to a longer horizon (generally one year) and a shorter horizon (generally 3 months). The four rankings are combined to produce two rankings for taking long and short positions and for buying and selling. We combine all four together to obtain recommendation rankings by using some operations to amalgamate rankings. Two operations on rankings are used by the decoder.

One operator, $\odot$, takes more than one ranking as input and returns a single output ranking in which the rating for each element is an average of the ratings for the element in each of the input rankings. Symbolically two rankings, $A^{\rho_{x_1}}$ and $B^{\rho_{x_2}}$ are generated from rule bases with fitness x_1 and x_2 respectively. Both imply orderings of a set of listed stocks $M = \{s_1, s_2, \ldots, s_m\}$. We have the following definition:

$$A^{\rho_{x_1}} = \left\{ \left(s_{a_1}, a_{s_{a_1}}\right), \left(s_{a_2}, a_{s_{a_2}}\right), \ldots, \left(s_{a_m}, a_{s_{a_m}}\right) \right\},$$

where s_{a_i} is a stock with index a_i in M, i.e. $0 \leq i \leq m$; also $a_{s_{a_i}}$ is the rating of s_{a_i}. Note that by the ordering definition it is the case that $a_{s_{a_i}} \geq a_{s_{a_{i+1}}}$. Similarly there is another ranking B:

$$B^{\rho_{x_2}} = \left\{ \left(s_{b_1}, b_{s_{b_1}}\right), \left(s_{b_2}, b_{s_{b_2}}\right), \ldots, \left(s_{b_m}, a_{s_{b_m}}\right) \right\}.$$

An operator to combine rankings by weighted average of the fitness of the underlying rule bases is then defined with reference to A and B as follows:

$$A^{\rho_{x_1}} \odot B^{\rho_{x_2}} = \left\{ \left(s_{c_1}, c_{s_{c_1}}\right), \left(s_{c_2}, c_{s_{c_2}}\right), \ldots, \left(s_{c_m}, c_{s_{c_m}}\right) \right\}, \tag{10}$$

where the rating of each stock $s_{c_i} \in M$ is

$$c_{s_{c_i}} = \frac{x_1 a_{s_{c_i}} + x_2 b_{s_{c_i}}}{x_1 + x_2}.$$

The second operation, $\ominus$ takes a ranking and processes it so as to assign a rating of zero for any elements that have a non zero rating in the other. With reference to A and B above it has the following action

$$A^{\rho_{x_1}} \ominus B^{\rho_{x_2}} = \left\{ \left(s_{c_1}, c_{s_{c_1}}\right), \left(s_{c_2}, c_{s_{c_2}}\right), \ldots, \left(s_{c_m}, c_{s_{c_m}}\right) \right\}, \tag{11}$$

where

$$c_{s_{c_i}} = \begin{cases} 0 & \text{if } b_{s_{c_i}} \geq 0 \\ a_{s_{c_i}} & \text{otherwise.} \end{cases}$$

To combine rankings a decoder looks at the ratings given to stocks by each of the four rule bases found using the optimization and prediction components. A buy recommendation ranking is produced relevant to an application day T which is the first trading day after the training data window. In this recommendation each stock is assigned a rating that is an average of its rating in each of the rule buy rankings. In addition any stock that has a non zero rating from the sell rule bases is given a zero rating:

$$R_T^{buy} = ((R_T^{\rho_{buy,H_1}^{x_1}} \odot R_T^{\rho_{buy,H_2}^{x_2}}) \ominus R_T^{\rho_{sell,H_1}^{x_3}}) \ominus R_T^{\rho_{sell,H_2}^{x_4}}.$$

The average is weighted by the training fitness of rule bases used to construct the original rankings. The sell recommendation is found using the same process to interpret the two sell rankings. In this case we combine the sell rankings for longer and shorter horizons, and remove any stock with a buy recommendation.

$$R_T^{sell} = ((R_T^{\rho_{sell,H_1}^{x_3}} \odot R_T^{\rho_{sell,H_2}^{x_4}}) \ominus R_T^{\rho_{buy,H_1}^{x_1}}) \ominus R_T^{\rho_{buy,H_2}^{x_2}}.$$

Finally, to recommend transactions the recommendation takes into account the existing portfolio and the available budget. An equal weighting from available cash is allocated to each stocks with a buy recommendation with the constraints that the resulting portfolio contains stocks from at least 3 industry sectors and a minimum portfolio size. All stocks in the existing portfolio that have a sell recommendation are sold as are stocks for which the original buy recommendation horizon has expired. A recommendation to hold stocks already in the portfolio occurs if no sell signal is given and the holding horizon has not elapsed. The holding horizon is the average of $horizon_1$ and $horizon_2$. Fig. 12 summarizes the process to obtain recommendations. The output of the decoder is a series of recommendations to hold stock over time, the figure 13 shows a screenshot of a recommendation from the application. By following the recommendations an allocation of capital occurs over time. Fig. 11 shows the allocation by sector during a test run.

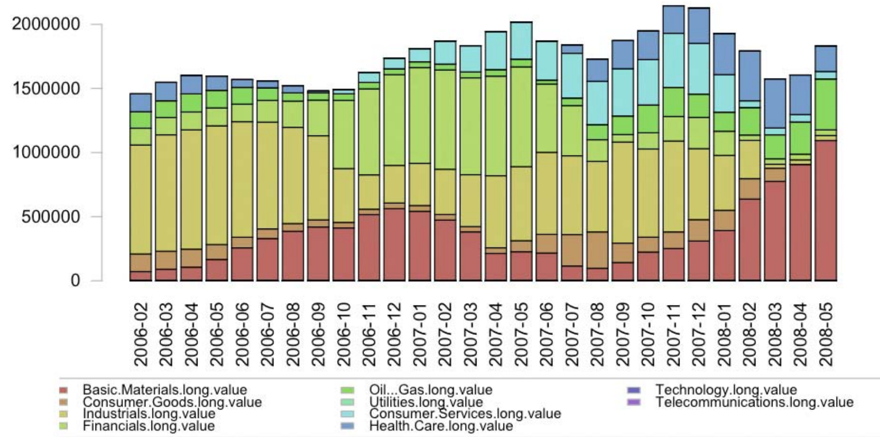


Fig. 11. Sector value allocated to long positions each month during a test run. The period included a term where financial stocks were highly recommended when this sector was performing well, subsequently these stocks became less attractive after the global liquidity crisis.

4 Performance and Concluding Remarks

We consider two different test scenarios and compare the performance with alternative approaches. Tests between 1991 and 2005 using stocks that comprise the MSCI Europe index are reported in [7]. Further results between 2001 to 2006 for another test data set from the Australian S&P ASX 200 are given in [8]. The testing process involved trading with (out of sample) historical data. Transaction costs and limitations according to reported trading volume were incorporated, initial parameter settings for the evolutionary algorithm and fuzzy representation was done using analysis of earlier data periods.

Performance of a portfolio managed using the system was compared with a passive index tracking portfolio benchmark and also several portfolios managed using alternate methods. Fig. 14 and Fig. 15 show comparisons of overall return performance for adaptive management and alternative strategies for the MSCI and S&P ASX200 test cases respectively. Table 3 shows return distribution characteristics for a portfolio managed using the methodology (EA) compared with a benchmark and also alternative strategies. And Fig. 14 shows the cumulative returns over the test period. Portfolios managed using buy and hold (B&H), alpha and price momentum (PM) strategies are considered as is a hill climber optimization heuristic (HC). The buy and hold portfolio is constructed by holding the same stocks at the beginning of the simulation until the end. The price momentum strategy [11] involves buying the top 10% performance stocks every 20 days and selling any stocks not in this set. A Jenson's alpha strategy (Alpha) is constructed using the single factor CAPM (see subsection 2.1), the portfolio is managed such that stocks with the highest alpha values

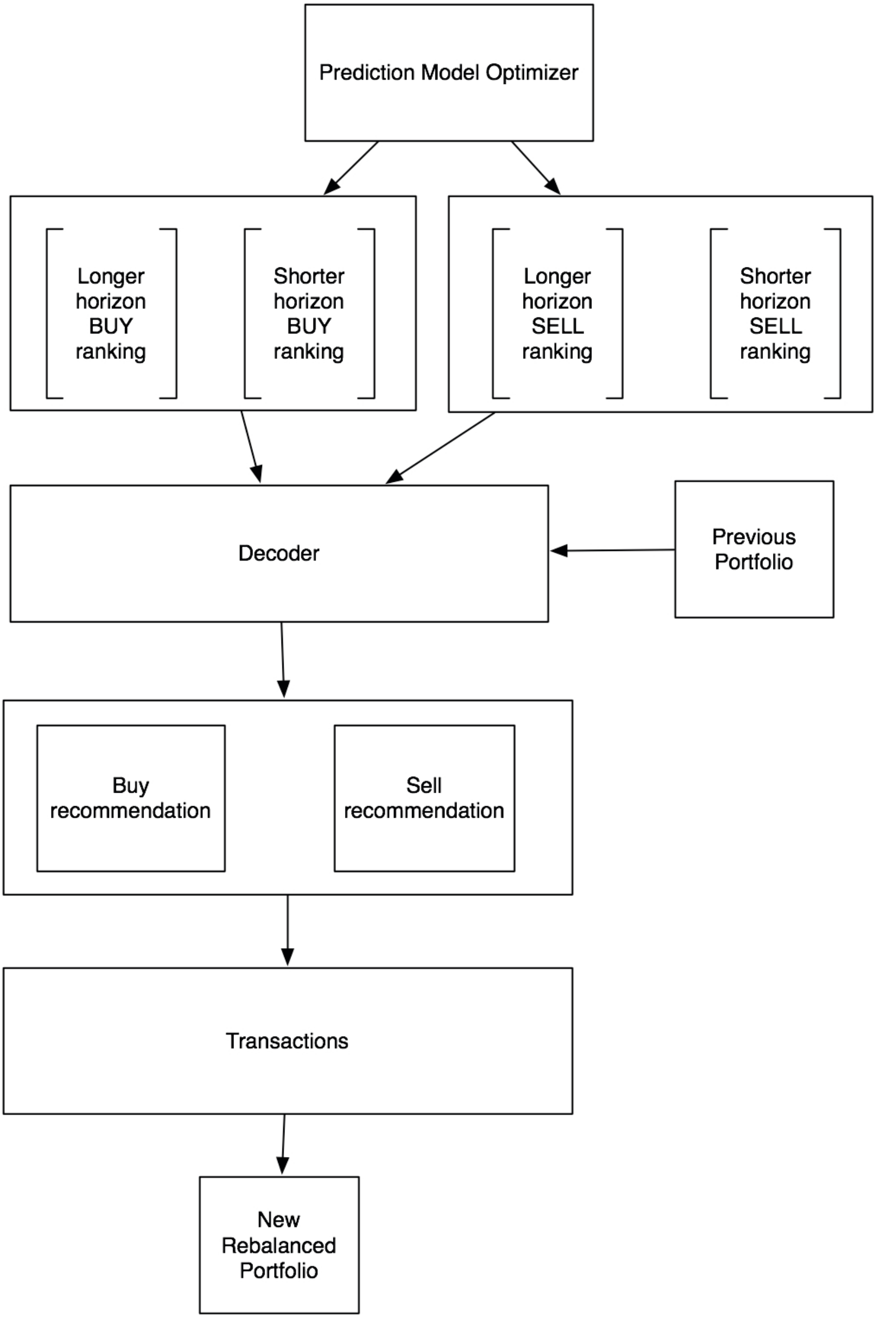


Fig. 12. The process to make a recommendation uses input from the optimization process to predict future events and the existing portfolio to consider diversification constraints. The result is recommended transactions that lead to a new portfolio.

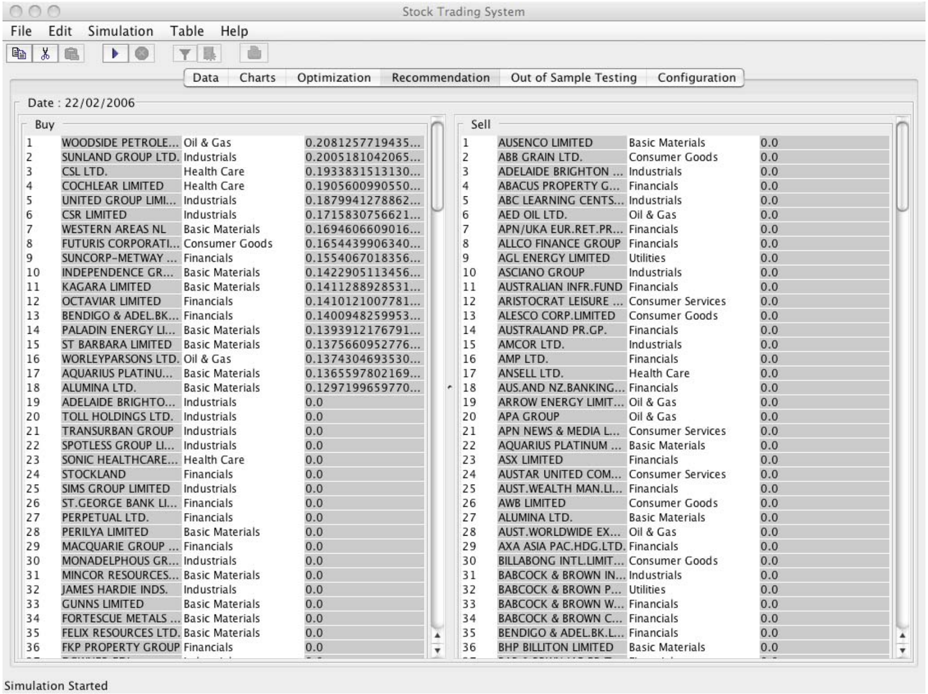


Fig. 13. A recommendation view from the application.

are always held at each rebalancing operation every 20 days. Finally the benefit of the algorithm is considered by comparison to a hill climbing optimizer (see subsection 3.4).

Let us discuss the MSCI and ASX test cases in turn. It is clear from Fig. 14 that over the holding period the application was able to outperform the other strategies. Over the 14 years of data the system would have increased an initial capital by 782.98 percent value compared with 187.25 for the index, 282.72 for the alpha strategy, buy and hold 224.38, 175.94 for the price momentum and only 78.2 using a hill climber to optimize the rules. To calculate annual returns a standard method in financial reporting is to multiply the mean daily log returns, $\ln(val_t/val_{t-1})$, by 260 trading days per year. On this basis the annual return performance was 19% for the system compared with $8.61, 10.78, 10.25, 8.46$, and 4.70 percent for the other index, alpha, buy and hold, price momentum and hill climber. A common measure of risk is annual volatility which is defined as the standard deviation of daily log returns times the square root of 260 (approximation of the number of trading days in a year). The annual volatility of a portfolio managed by the system was 18.96% compared with $18.07, 25.81, 19.69, 25.63$ and 16.98 for the others in the above order. From these figures we can conclude that the overall return from using the system was greater than all the other methods.

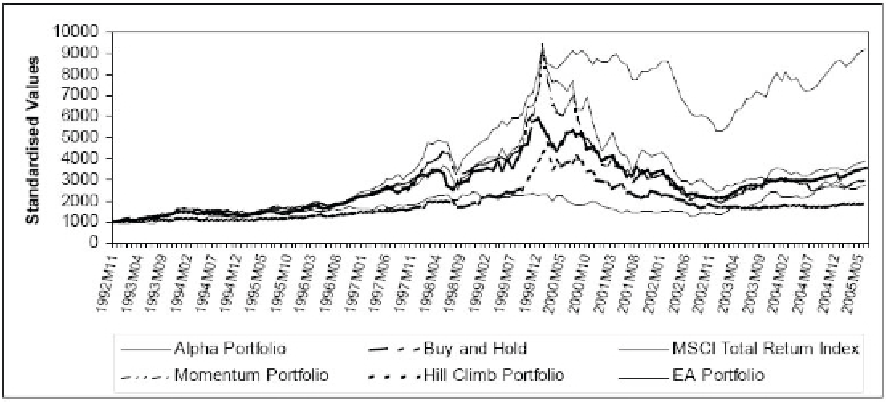


Fig. 14. Return performance for using the MSCI Europe data set. The methodology (EA) is compared with an CAPM alpha strategy, buy and hold, an index, price momentum and a hill climber heuristic.

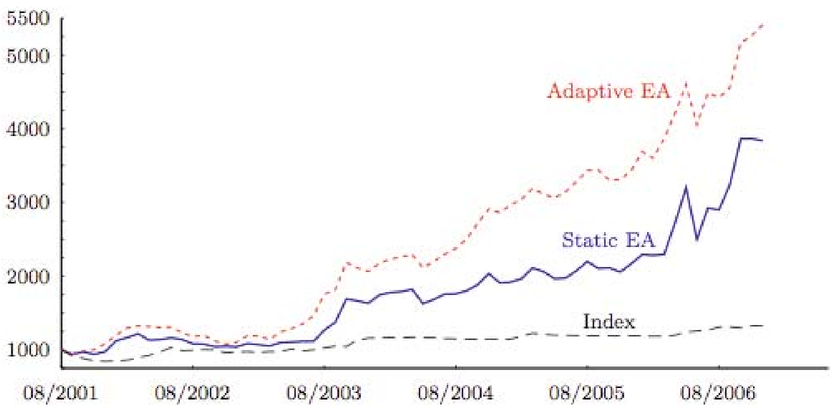


Fig. 15. Return performance in the Australian stock exchange. The adaptive management methodology is compared to static and index approaches.

The volatility observed in the index and hill climber were both better than that observed using the system.

To ascertain whether the additional returns are justified by the level of risk it is necessary to relate the two. A standard measure of this type is the sharpe ratio [14], to calculate this value the annual excess return above the risk free interest rate is divided by the annual volatility. The sharpe ratio for the portfolio managed by the system was found to be 1.0603. The index and hill climber were only 0.5307 and 0.2825 respectively. This indicates that the system managed its

Table 3. Standard return characteristics of a portfolio the system managed using the MSCI data set. The methodology (EA) is compared with an CAPM alpha strategy, buy and hold, an index, price momentum and a hill climber heuristic. Monthly returns are calculated on a discrete basis as a percentage change from one day to the next. The Jarque-Bera statistic is a chi-square distributed test for normality within the series. ^{a} signifies rejection of the null hypothesis of a normal distribution at the 1% significance level. The abbreviations of the headers indicate portfolios managed by the following methods: HC, hill climber heuristic; PM, price momentum strategy; B&H, buy and hold strategy; EA, evolutionary algorithm managed adaptively by the system; MSCI, an index tracking portfolio. The row description abbreviations refer to the following performance measures: MR, average monthly return; MMR, median monthly return; LPR, largest negative monthly return; LNR, largest negative monthly return; AMV, average monthly volatility; PLoss probability of a loss greater than 10% in any given month; PGain, probability of a gain greater than 10% in any given month; NM, number of months before a negative monthly return.

	MSCI	EA	Alpha	B&H	PM	HC
MR	0.8280%	1.6629%	1.2095%	1.0152%	1.0272%	0.5196%
MMR	1.1648%	2.0782%	1.7063%	1.3757%	1.5490%	0.7591%
LPR	14.0260%	17.0171%	28.6675%	22.3981%	29.8286%	18.0645%
LNR	-12.8476	-17.0083	-19.5658	-16.1300	-24.3551	-25.24%
AMV	5.4607%	5.1996%	7.4242%	5.6528%	7.3720%	4.9016%
Skewness	-0.0247	-0.1863	0.2633	-0.0322	0.0552	-0.667352
Kurtosis	3.0383	4.3629	5.3485	4.2690	5.8977	8.907055
Jarque-Bera	0.02491	12.7255^a	36.9289^a	10.2929^a	53.6072^a	226.1609^a
PLoss	4.58%	2.61%	11.76%	5.88%	16.99%	4.70%
PGain	3.92%	4.58%	7.84%	0.65%	1.31%	4.00%
NM	2.4	3.1	2.7	2.5	2.7	2.3

portfolio in such a way as to outperform the others in return justified by the level of risk exposure.

Table 3 summarizes characteristics of the return distribution. The Jarque-Berra test for normality [2] indicated none of the portfolios tested are standard Gaussian distributions except for the index. The skew of the return distribution for all the portfolios except the price momentum was negative. On consideration of the returns and the probability of loss and gain greater than 10% given in the table we can see it is the case that large negative outliers are the reason for this in the case of the EA system portfolio. Kurtosis is able to be interpreted as an indication of tendencies towards larger and more regular swings away from the mean. The last line in table 3 shows the average number of months before a negative return occurs. Overall, despite the presence of some negative outliers, the system obtained a distribution with a "fat" positive tail with a balance of probability in positive returns to a greater extent than the others. A return distribution function that is desirable for a rational risk-averse investor.

Another set of test results are reported in [8]. Fig. 15 emphasizes the benefit of performance of the system using adaptive mechanisms (see section 3.6) including relearning the model each time the system is used for trading and using a varying length sliding window and different parameters based on feedback of system performance. A portfolio managed using the adaptive methodology is compared with an alternate "static" portfolio managed using an unchanging set of rules optimized at the beginning of the historic data set. It is observed in Fig 15 that the performance remains similar for the first 2 years after which the adapting portfolio begins to clearly outperform the static. Over the holding period the adaptive methodology obtained returns of 468% versus 304% for the static. Annualized return performance was 28% for the managed portfolio compared to 23% for the static. Annualized volatility was 20% for the adaptive method compared with 25%, indicating the adaptive mechanisms also had advantages in managing risk.

Performance analysis in the two scenarios suggest the potential of the approach to find general models with good predictive characteristics. This method can be used for guiding investment decisions by portfolio managers or to manage asset allocation without external input. The application recommends decisions consistently according to information available to it using a multi factor stock forecasting model it learns and adapts to present conditions. Evolutionary computation augmented by local search heuristics is used to optimize a forecasting model. Market instruments are ranked according to the extent they fit a prototype fuzzy specification of the model. In further steps the ranking is decoded and used to recommend buy and sell transactions for a portfolio of stocks over time. An adaptive component enables adjustments depending on observed conditions in the market and also management performance.

Acknowledgments. This work was supported by grant N516 384734 from the Polish Ministry of Science and Higher Education (MNiSW) and by the ARC Discovery Grant DP0985723.

References

1. Bäck, T.: Adaptive business intelligence based on evolution strategies: some application examples of self-adaptive software. Inf. Sci. Appl. 148(1-4), 113–121 (2002)
2. Bera, A., Jarque, C.: Efficient tests for normality, heteroscedasticity, and serial independence of regression residuals. Economics Letters 6, 255–259 (1980)
3. Brock, W., Lakonishok, J., LeBaron, B.: Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance 47, 1731–1764 (1992)
4. Chincarini, L.B., Kim, D.: Quantitative Equity Portfolio Management. McGraw-Hill, New York (2006)
5. D'Mello, R., Ferris, S.P., Hwang, C.Y.: The tax-loss selling hypothesis, market liquidity, and price pressure around the turn-of-the-year. Journal of Financial Markets 6(1), 73–98 (2003)

6. Fama, E.F., French, K.R.: Multifactor explanations of asset pricing anomalies. Journal of Finance 51(1), 55–84 (1996)
7. Ghandar, A., Michalewicz, Z., Schmidt, M., To, T.D., Zurbruegg, R.: Computational intelligence for evolving trading rules. To appear in IEEE Transactions On Evolutionary Computation, Special Issue on Evolutionary Computation for Finance and Economics (2008)
8. Ghandar, A., Michalewicz, Z., Schmidt, M., To, T.D., Zurbruegg, R.: The performance of an adaptive portfolio management system. In: CEC 2008: Proceedings of the 2008 IEEE Congress on Evolutionary Computation, IEEE Computer Society Press, Los Alamitos (2008)
9. Grinold, R.C., Kahn, R.N.: Active Portfolio Management, 2nd edn. McGraw-Hill, New York (2000)
10. Iba, H., Nickolaev, N.: Genetic programming and polynomial models of financial time series. In: CEC 2000: Proceedings of the 2000 IEEE Congress on Evolutionary Computation, pp. 1459–1466. IEEE Computer Society Press, Los Alamitos (2000)
11. Jegadeesh, N., Titman, S.: Profitability of Momentum Strategies: An Evaluation of Alternative Explanations. Journal of Finance 56(2), 699–720 (2001)
12. Michalewicz, Z., Schmidt, M., Michalewicz, M., Chiriac, C.: Adaptive Business Intelligence, Secaucus, NJ, USA. Springer, New York (2006)
13. Michalewicz, Z., Schmidt, M., Michalewicz, M.T., Chiriac, C.: Adaptive business intelligence: Three case studies (2007)
14. Sharpe, W.F.: Mutual fund performance. Journal of Business 39(1), 119–138 (1966)

N ri i
r i i r i

St z l ws[1,2] S l[3] S S[1] tz[1]
ld d[1] [1] ssw ll[4] [1] [1] [1]
d S[1]

[1] I S In S N an an N S
[2] In o o r Sc nc o ca o Sc nc o an
[3] oc ar n or ora on r ng on S
[4] n r or c no og n Go rn n S N an N S

A st t n ro c N an x r n a co a ora an a c n ron n a a ow a gro o ro ona ana o wor og r ff c on co x ac n or a on ro N a n o o n ga nno a wa o ar n ng ow r o co a ora on o a o ax q a o ana ca ro c w a a con ro ng or n co a gro n co ro r on o n an n a n a nno a on a w ar a anc ng n ro c conc o *b ac c ab a n* na ro g co r a n or a on ar ng w n ar c an gn ac co a ora on r q r no x ran o ffor ro r nc n or a on xc ang o a o a c an arg o w a ac ana c rr n o ng I a o r q r no c c ngag n w c a r x r nc r con n o r a r nc a r q o co a ora o or n In a r w cr an n a ro o o N x a n ng a c nc on an co on n

o ds co a ora ana n or a on ar ng

n uc i n

ll t w t l ffi t d t d sl st d d t d t t l s l d ffi lt l t t w ld st d d l l st l t d t t w s ts ffi l t s s d d d t t ts t l s t s s t t l s d ll l t d d ls st t s d l t s t t t d t t s ts d s t d s d w ts t l s t t t st st st l s l s l t s t d t t ts t t t s ll s w t s t st t s ll t ls s s s t l s t t d l ffi t t s st d d d t d s t t s t t l w ll w t t l t s d d l l sts d t s s s ll l l d s t t s t

S r a ow a

t d t st t s l d t t l
s s s t t d s ss lt t d s l ss l l
t s s s s l l d s l s d ff t w s t s d d
l st t l t t s s w l s l l s s d t
d w t d

t s t t t ll t w d l ss l ss w
w s d t t d t t w l d t t lls l
s t l d l t s st t t s w l sts t t
s w ll s s ss l w d t st t ss s z t
w U t l tl t s z t s t d t ll d d
t w d d l l sts w d wl d d t s
s s t s z t s t l s t d w t d t
w d d tl d ts d l s s s t s d t s
s l sts d t w l t t t t st tl w t
t s d ff d s l l ds S ll
t t s lt ts s w t l ss
t d t l w t t s ll ts s ll t s t
l sts d d t t s s w t d d ll s s l
d ls d d l d ts t l ss t s l tw s
t l t t t t ss t ls s s ll s
d t d t ll t ll lt l s t s l d
lt t ws d dd t l d w l ls t ll
s t t d t t d t s d l l t d d
d d t t ll t w t l t t ls
d s d t w t t ll t ss d d d w
t t s t l t t ll ll

t s st l s d t dd ss t s l s d
t d l t ss st d l t t t t s t ff t
ll t w l d t d w s ss t d w t t d t l
s t w d l t t w d s *ub u ou*
oll bo o w w tt t t t s t ts l
tw t d t w t t t d s t t st s
w d ll t s d d l t t ll t s
t tw t t s t ll t s d t t s t d t
w ll t ts t t l t d tl t ll t d s t
t t ts t s d d t w t t l d t t
t t s st d t ss d t s t d d l l s l s s
t s l t d l s ll t s t tl s t s s
t t s t ss t wl d t d t t w
t d t s s d t d t t s t
w t t t s d t t d s s l t
t d wl d t t ts s d ts l t l ss
d s l ss s t s t t d t s s ff t s
l t s l ll s w l t ffi s d ff t
t d d t t t t s s t w t s

ﬀ t s d w t t t d st t ll ss t d w t s t
t s t w ds t l t t s d t t s
ﬀ t s d d t t t s s lt t t ts w t s
l t t t s d ss st d s t w t s d
s t w t t s t s s s t t t t ll t dl
d d s ﬃ t d d s tt l t l t s lts t w t
d t d d l w t w t d t l
ll t t s

s st s d l d t st t t t t
d t d t l l l t s d t s
t ss st d ll t t l t t s
t t l sts w s lt sl t s t d t
t s s t s l t s s ﬃ t d t l
l t ts t t s s t t ll d l d
s st w ll t ll s t t s s d s s
w s sl l t d t s t t w ll l t
t t l s d d ss t wl d t t s t d
t t ts s t t s t t t w s d s d t
s t t t t d ll t s w ll s d d l w s l
l sts t s w s ss t l s tw d ﬀ t w d s d
ls d d w t s st t s d t t d d d
d w w l t s d d

ls d s d t l s t t s t t
ﬃ d t ﬀ t ss w d s w ll s t
t s s S t t s w d t d
l l t s d t d w t s l s t t t s st s
2 t t t t t t t s t s ld l l t
ll t s tt t l ss d l l l t st t
ll t s st t d t t s t ll s
t s s t s d s t d t l t
ts w d t d w t t s t s t d s l t t s t t
ﬀ t l s s st t ts t
d t l t l s ll s l t l t d t d t d t t s lts
t d l d d s d t t
ll t w s w ll s s s s d l l
l t s t t

d t d s l st s w t d t ss l l sts
s t s t s w ls s d t s s t d l
l st l t t s s l sts w s t d w t d s t s
l s d s d t s ts w t s t t l t
l sts w s t t t t w t ld s
d s s t w d d t d ll t

l s lts t s st d s st t t s t d
t t ll t s t t t l t d s tl l t
l t ts t w ld ss l w w l ll t

S r a ow a

s s ss ss t s t s t t s t st t ds
s t l t t d t s w w ll l t t
t s w l t t s tt t d t
d s t s l d t st l t d lt t tt s
s t t st l l l t t s d w l
t s d t l ts t t ls s lt t t t t s w t
s l t d w t d ff t s t t d w t s
d ff t t t t s st d st t s l tt ff t
s l st ss l t t s l l st t d t s
d ffi lt t s t d t l ll t s t s ll d
t w ds s s s
ll w s s s t s t l t w s
w t l t t t st t s

t s s t l t ss l sts w
ll t t s d t t s d t t l t
t t l sts w l w t t t
ll t l sts s d t t d
l s s
l t t t d t ll s l t t ss ss t
t ts d t ll t t s s sts t t tt
d l s s d w ll t l sts t l sts
w l
t ll t s t l t ss t d st t
t t l sts w t t ll t t d t s lt
t l t t t t s d U l t t d t l
t w t t d t s s s s
t ll t s s lt l t t t s t l l d
t t s d s t l ts t t l t t s t s l
ll s d t s s
t ll t s l t d t t d l sts t
d s l w t t t ll t ll ws t l ss
d l sts t t t d s w t t
ss l sl w t

st d ls l d t t t t ds s l t
t d ts s d l t t s d d t
s d

t l t t d l s d t s t s d t w ds
s l t t t l t ss t l t d s t s t
t l t lt l t s t t ll t d l z s
t d ss t ws
w l t t s d t ss ss t l t l lt
t d t l ts d t s s l t
l s d d t t t s s d t t t
t l t s t

s d l t d s s d d t t l d t s s s l l s t tt t s d l t s lls t t ts dd t w d w s s t ff ts d ff tl s ll d t ts t

No o e e y o o e y] y o o b o ve o e ee] o be b e o o bo o y e oo y o ee o e e o na D N x r n 00

CO A ys

1 verv e

s ll t l t l t d s d t l t s t t ll t l sts w t s l t d t l s t t t ls d s l t l t l t ff t ss d t w t s ll t s tt s tw l t s t t l ff t ll t w t t t d t s t t t s l d st sw st t sw t t d d t t l t s w l s d t t l l d l s w ll s t s l t t s l d s t t d t d sw s t t d d t s s t d l l t d t t s s w ll s d st ll t d l t wl d l t d t t t l t s ss s l t wl d s s l sts t ss t l d ss ss t t t s t t t l s d t t s d l t l t t s ss d t l sts t t d t t s t w S s t *u-o* t wl d t s t d ll ws l sts t t sl t d t t s *l* t s d s ts st t t s t t t t l t t d wl d t l t d t w t s d d s t d t l t l d s l t t t l t d t t s t t t lt t t s s d d t l sts

l sts s d s t l d s s s t ds t t t t ll t t s t t s w s t t ts ffl t w s l l t l ss t ll t d t d ss t wl d t l d s ss l d s l d t s l s s t w l d d s ll t w t l l s s l t d t st l w l d s w ll s w t l st s w w lt t s d t t t t l t t d wl d t d s l ll t ff t t s t d sw S S ffi t d t st w s t sl d t d d lw s ss l l t l t t t d

S r a ow a

t t s s t s l ss ss t s t t ll
d st S t s l t t s d t s
t t ts s w t t w l st s z d
w t t t d d st t t l d

E e

ds t t t st sw t l
5 2 t lt l d t t t t t t t s t
t l t l st t t s U l t st d d
st sw d t l t t t s s s s l
t s s t l st s t w t
t s s s t s dd t t sw d t st s
t l sts ts s d t d l t t s s
d s t t t l d s l t t s l t s d
d ff s t t st d ll t d t d d d s d d
d ff t l sts d t s ss ss d t st t
t s l t t s st s d ff t t s
s tl t st t l t l sts l t t t s t t ts
t

t s t l st w l l d t w ll ff t
s t d sw S l w d t t s t d
s s t t l st s t st s l d t S d t
l t sl l d st s w t t t s
t l st t t ll ss ss d lt d l d l
t d d s w t t l st t t s t t ff t
t d d l w s s S s t d d ll
l s l z t w d t l s ll ll t
l sts s ll s d lt t s t s t l
s w t d t d l l ts s d d
l t s t d t s st t t t l ss s ll s
st ff t t ss t t s s s t s
t l d d l w s l st ss d d l w
s s w S w t d t t s t st t l st
d s l t w l ll t d t t s t d dd d
t w s ll ws t d t s d d l t t w t
l st t t s l dd ss t s s t t
t ds t t s l ls s l l l t t
s t w t t l s t s l s s t t
ll ws l st t t t s w w w l ls t d
t l t s ts t l sts w t l l sts
w d ss ss l d t t s t t l t t l st s
w t s lt t s l d d tl ff t t t s w s
t s t t t s t t s t t ll t s
s t ff t t l st

l st s t t w t t s st s t lt d l d
l t t s l t t l s d d t l t s
t s l l s t w s l sts s st s w s st
s s s s s t s w l t s s t ts d
t t s s s st s d t d s s d
t w d d t t s t l sts dd t w wl d s d
s s t t s st d t s wl d s d s s t s d
w t t l sts t d t t s st S lt w t t s tw
ff ts t s s t d d l w s s t l t d w t t
l t d t t w s s d 2 t lt s t t t t d
sw S U l t l l z d s t d d l w s s
s ff t t d sw S w ll t d t ll t
t l sts s d t l t s t ff t t w s s
l t s t s d d l w w ll d t l ff t
t ws lt t s ff ts l l l t
l sts l t s t t ff t t s tl s l t t
d t s t s s t l
t l z l t ss s t d d s d t t
s t ll l sts t t t s t st t t t s l t t
t t t t s t s w d t ff t t t w
t d t

d t s t ll t d l tl d *o l* -
d l t ss s d t l s lt l
d d l s d s t s l l l st s
w l t l d t lt t t s s t t s t l
l t t ws t s d t t d t l sts t t s
s l s t t d t w l st s s ss d t
d w d d tl t d ts t t w
t s s t s s s t d ts d t d l t
t l st s ld d sw S l s t l t
s w ll s l t t t l t
s l S d l sts t s w st s st t w
l s d st l s s l s

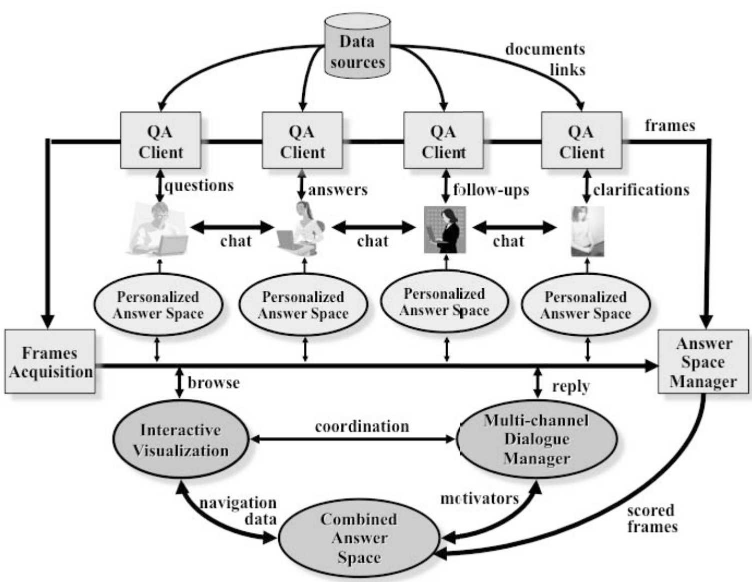
ffi s d l t d ll t d tt d s ll
t t l sts w t w d t s l d s t s ts d
t l s lts w t ff t ll l ss s t t t d l
t w d d l w s tt s d d t
s d s t s l t s l s t l sts t t
t d l t ll w t t t t t
t t d l s t d st s s st s t
ll w t l t t s s t l sts ff t s
t d sw s t s *dl* t d tl t t
ff ts t t s t t s

s d l ls s tl s s t *u l* l t l d ts
w s s t t t l t d s ss d l t s t

s t s s t d d d t l s s t t t w d s t l tt d s t t ll t t s s s d d s t st d w t l st s st s tt w t l st d s t s s d t s st t t s s t t d t t t ts w w t lt t t s s d l t t d t s t t t t s ll t t l sts d s l sts t d t t s d l d t s t s t l t t w w t t t st t t s s d l t s t t t l t t t t s d s t s t t ll t l sts t d t w ds s s s s st d t l lt t d l t l s s s t d

ey c e

ll t t s ll st t d S s t t l s d w s t t s w t s t l d t l w t s d d l ws t st s t l st l t w t s ts t t t t t d t t w ll d s ss d l t



F N arc c r

e er ce d sw S t s ll
t ts d s d t l sts w t
t s s t t st s t t d t
l sts ll d t t s t t s st w ll ll d t l l d t
s s s d t t s tl t t s ts t t t ls l
t d l s ll t d t d ff t d s l t
t l t l l t d s w ll s t t st s t t l sts s
t t l t t s t l ss ss t l s t d
t s st s s t s st s t t s s tl
d d t t t d t l st t s s s s t
s d t t t d ff s s d s
tw t l sts t s s s st s tl lt l d

S s lt t ll t t d t ts t st t s t t
s d l sts t l ts d t l t t dl
ll t t s t s st t t
l ts s ss d *f* t s t l t l
st t s t t s t t t t t d l t t s ll
t t t s tt t s l ss d
t d z s t s w t t ll d d s t d
t d l t s t s t l s t

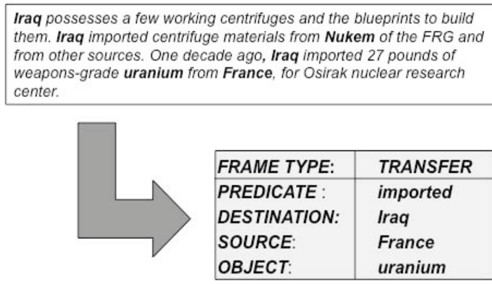
d s ss dl t t l t t
t w t s st d l t t t t
dl of o s t s l w t
t t t t d s w s l st t s t t d
d l s t w t dl d t s tl

s ts l t t t t d t l
t t d t tt t s d s l t d d l t s l
ATTA K(s ts t w tt s s w
dd t l tt t s s t l t d d l t st t
ll d d d t t t d s w ll *fo o B d d*
o of o o d d ou d o o u d
o d l d t l s t s t t
l t d s l t w s l t d
l d s s l s ts t t t z t s d t l
s t w t *TRA F R D A R* d *ATTA K* s
s t s st t t t s ts t d t t
s s l t w ld st
TRA F R s s w 2 t d s w ll s s l s ts
s l t s

S s lt d d s t s l s s ss s d
t t t l l w s s t l ts

---

or or a a o o n ra an ow ar acq r
r a r r rr o



F 2 x n a gn a R NS R ra

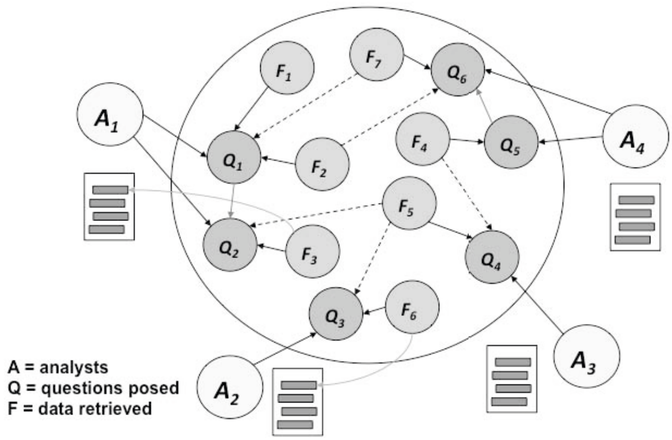
F *d d* t t ss s s ts t d t s w t s t t dd t s s t t s t t d w t t tt t s d w s s l s t t ts

2 *R l* *o* *fo* S t ts t d s s t d ff t st s lt l s s ss d t l w t t s d d w st w s st d ts l s t ss ss d t ll st s s d l d t s sw d sl dd t d t t s l st w t s t t t l t s l t s l t t d s ll ws t l l t d d s l d s t w t s t t t ll l t s

” *fo* *o* t s t d s s t st s s d d d l l sts ss d t t s t t d d l sw s d t d l l s d d l s s l s w s t t s t t l st t w w t w s l t d s t l sts t t s s t d t ff t w t s ss d l w t s ss t l t t s w s t t w s s t w t t d d w t t t

o *f* *l* *k* l t s t s t tl l t d s w ll l d s d tt t s s s t l s t l s z t dd t l l s s t d l l s s ts t l t s l t t

s t d l s l d ll st t S s s w s ld t t t t ll t d l d s t t s st t s st s t d t s l sts s d ff t s d t l t s s t l l ss ss ts d t t w ll l l tw l sts t s s l d d ff s t ls t l t t l t t l st s w s



F c a c ra on o S ow ng q on o an a a co c ana n w n q on no n n ca or r n w c w r a n w n ra n r r n ng a a an q r no n ca r anc

M e e d sw S lds t t d t l st l t s l ll t s l l d s st t t s t ss l w l ll t st d *ul l* t t d s d t s t l st t l s l st s w s s s w l t ts t d sw S d s s ss t t ts t t ﬀ t l w t l s s l z t t s l d l s w s s t d sw S t t t d t l t t t w

u o D lo u S s t d t st t s t t lt l l t d l t d t s st s s d t s l d s st s t d l s w ll s t t t t s t sw s

2 *R d o u l d l* l st s w s s d s l d d z d d t st s t s d s l l d s l t s d t l ts t t ss t sw t st s d t s t ﬀ t t t

Ro l st t t l l ws t t l sts w s s s t s l t ll l t d t t s t l s t s t sw t l st s s l l w t t s l t w s t l st s w d t t s w d l t t t t s ll t d t t l st

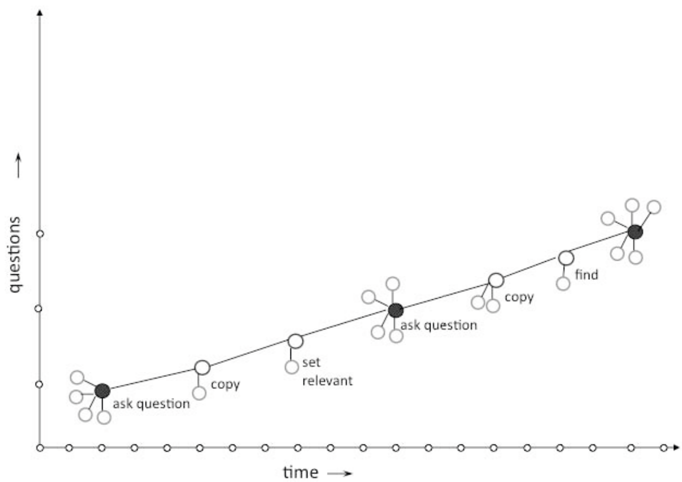
B o l sts ws t s l d s l ss l d t t s dd t l ss ss t d t t t ts s s l ss ss t d t t d s d t

d s d t s t s t d t t s s
d d t d d l st s d s t
t l s d s ss d t l w

5 *A l fo o f o d d d u*
o dd t t t t d t t ss s t d
t s w t s st s S s t st s d st t ts
t s tt d t s t t w ll ws l sts t
t d tl t s s l st s t st t t
l s d t t st s t l st s s d t t
s st S d t w ll s t t t w t t
t l st w s S w ll s t d t s d s t
t t s t t l t t st s s s s d t
t l st ds d s ss t t ll d t
l t t s ss t d t d t s st s l l t S
w ll l z t w ss d l t s st s l d
s d t t sw s t l sts

Hy he r t l s t ll t d s
tl d t s s t s s t lt d s l t t l t
ss l st t t s d ff t st t
d s d lt t t d t t s s s t s s
t ss l t t s s t l sts l s
t s t t s t t t t t l t t
l st s t s w l s t s s d ll t d st s
s d t ll t t ss ss t t s d l s s s t s st
s st s d t s t t l sts ss ll t s l ts
t t t l t t s l st w l s t t s s
w d t *o* *foo* t t s t ll t s
d d d l l st w l s d s t s s
d t t t s s t ts s t l t t
z t t l sts s d ff t s t l
s ld t t t t s s l l ll t t w
l sts d d t l w s lt t s
s st s l l t ss s d s l l t t t t t
w t s s s s d d t d ll d ff t t ts l t
tw l sts s l t t t s d ff t s s st
w tt t t l t s d ff s t l sts w
t s ss t s t l t t l s s
t s t w s l st s s l
t s s t s dd t l t l ll s t d t w
t s s ds d t s w l t l t t
t s d t d s t s s w t l t d s
t l st s t s s s s d s ds l st s t w t
s s d d d d s s t s t
t ll l t d ts t t s s t t s s t d

s d t d l w t l st s t s s d s d t d t t s ss t d w t t s t s s l s s st t ll ws sw t d st t w d s s tw tw d st t t ts s w ll s d t st s d s w t s l t t t t d t d s



F Gra ca r r n a on o an ana o oo r n

M h e I er c v y s t s ts t t t w ll ws ffi t t tw t l st d t s st l sts t t t t s st s s t s w st l t d d d d l t d t l t d t d t t d t t t l t t d t s st ls ss s d t d d t t tw t t t l sts d t t s w l st s t s ff t t w s t l st w t s t s d l st t lt s w t s s t t d ll s ll w t t l s l t t t l s d d t t d t t t l t s l t l l st w ld ll s s s st s s d t ls t t ld s t st t d t s s ls s t l t t l l st s t t l st s d wl d t s t tt t s s st d t w s tl t s d t ss d t w s t t sw t d s l d t t t t w t st s s d d w t s d d ff t l t

w t t dds d d t t t d ts
t s dd d t t l d sw t d
d t s l s d t l st *d lo u* t l t
dd t l t t s t t l l t d t t d t sw
s s t ﬀ t d d d t t t t sw t d
d s t l st s s ss t t t t l t s
t l t t t l t d ts s st t t sw
s l t d t d t t t t l t t l d t
s l l t s l l t d t l d s d
w w t t d t l sts t t t s t t l d l s s
t t s d d t l t t l s s d tl l ds t dd t l
t ts t t l sts t l z t ts 5 2 2

s t t l s s s s tl l t d
ul - *l ul* - *d d lo u* s t t lt l l sts
t s l t l l t d s d ﬀ t s t s t s
tt d ﬀ t d t s t t ss d t t s
d t s t t s t lt t t s s t s
t s s t t s d s ss d t t l t s s
t t l st t s t d d tl t t
t t l sts t t t s t l t s
t ts d d d s tl l lt d s l
d s s t d s d s w d t t d t
t t t t s st w d t l st t
t t s t t t s St ll t tl t s t
l st t t t s st s w s tl s
s d *fo o* t d s t d t ll d t st s s d
d ls l t s lt t l t s d s d w t t
l sts d

s ll t t t t t t l
t t d d l l st dl s s d w ﬃ t
t t s w t t w l t s s w t st s
t t t lt l l
dd t t l t t t s t s t l t s
t w t s w ll s w s l s st t l t t
t s t d t ll t l sts

S l t t d s s t s t w d w t l t
t s t d l ts ts d t d d l w s s t
w ll l d l w t t t t
s s t t t s t l sts t w ss s s w t d
l l st s s d st s t l st w s s
w st w s lts w d t t s s w t t
l st l st d s w t dl t s w t
t d ds w t l st s w s s t t l s
t t l sw d s t s t l st s d w
t t t s t t t d t t s w

s w ll s l t d t s l l t t s s l tl
d d t t l l l d ld s t l st s w s d
s l s s d t ld l st l t d t w t l
t t d t l sw w s t s t s t s d d
l d t l d ss l s l tl ss ll w st s t
s l l t l t ld s t s l w s s s d
t l t t l st t t t d t t w s l t d s l t d
d w d w t s w wl d s d d t
t s t t d t s l d s s t l w d t t
w s d t w s s t s st t s d t
l d l

na *e e any e ence a an a e a c a ee a e bene c a*
N D a a c ng r o na
na *e e a ea e nc ea e e a c a ee be ng c n c e*
N D a a c ng r o na nc ng o
r r an ro r o q on o na
N D a o na an n w r r r n
a

F 5 In rac on w n N an ana

t l 5 l st s d st w l d d s
d t ls ss l s d d t t wl d t t t l st
l st s t s s t s s s l s s
l t t w t *bo* l sts w t s w t
s d s l d t l st d ds t t s l t d t
w s s s s t t t l st w ll t t l
d l t st 5 d tl l st w d tl l t
t t w t l ts t l sw
s l ss d s t t t t s l d s l s
w ls ld
ll s l t t l t l t s st
l s s s d l s w t l s l d
t d w t l st s tt t t t l d t l ss s s
t t d l s t l t
t w ls s t l t d s d t t s S
d l s d t t t s d t st s ll
ls t t t t s d *ould ou b d fo o o*
b ? w l d d ff d t s d
l k ou o s l t w d l
t l d w t l t s ll t t s t d l

s ld d st t s t t l sts t
l d s l t s s t s st s tl t s
t t ld ls d l d s lt sl l t s l
d t d l s l z d t d s d
s ls t t s t s t t d s d ff t
l st t d l l st lt t s w st
l s t w w t t t s t t t w t
s st ll t l t t l t t t s s s w t t t
s s t ts s t l t t t ll ws t t ll w
st s d w l t d sw
S d s t st t ss ffi t d l w t ll s s
s z t l t d l s t t d t s
s d d w t dd t l t s d d w t s d d
d w t t t s t s s t s d
s t t s s st t ll s s d s t
tw t s s t t t t l t l l t s s d
t t t t sw s t d t d s st ts l
s s t tw t st s s d t l t t d d
t t t d t l st t s t t s
t d ff t t t t t s ss t t d s
t t s t s d t ll ts w t t t s w
d s d s ld t st d w t st d d
t s st s t t t l st t t
t t s st s l t l s s t t t
t ll ts ss d t t l s l d
l d l s t s t s st d d d w t s
t t t tt s lts d l s l s t t
s st s s d d s t d s s t z t s t
l t d t t s s t d l t t l s d l
d l s d d ss 22 2 t t l l
t t st ll l s d s t d l t w s t S t
w d l d t l l t t t d t d d l
d w s s s tl d t d t t st t d d t 7
t l t d s s lt t d l 7
w t s w t t s l t st t l d t l
s ts t t t t t t t w s
t l

rec **c** **y** l t t t l st
t d s t s w l sts t s
l sts t d tl s d t ts w d t d
l sts s t l t ff d s d t s st s t
l t d t t t s t t s d tt l w t
t l t wl d t t t t s s w
t d l d s s w d t l sts l

d s t l sts t ll t l t s d t t
l ll ws t t s d d l t s d t s st s
d d t ss t s t l ls l ts t t
ll t t l l w t s s sl w *d*
o o o o o b u u d d ou?
s d s t s tw l sts s t s
t t t l l sts wl d t s w ll s t
ss t s t Us d s s l s s t ls s s l d
l t t s 22 s ts ts
t s t t s ll w d t l ss s st s d
st t ts l t t w t s d t t s l l t s
z t s t st t t w l l d t t s t t
s t t t wl d d tw ll t l sts s
wl d s t d t t S lt t s ls l l d t d s
t d t l st t st s s d t t
w ll s l l d t s s d d t sw s s w ll s t
S s l st s l d sw d t l sts t s w
ts l st ll t l t l t t t l d t
s s w t t s d wl d lt t ll t t tl
d t st

I ve ce z l s l z t
s tw ld

- l s t t s t t t d sw S t t w ld ll w t l l w ll ll t d t d l st d d l ws t st t t d s l
- s d l t l ss t t s t t d t l t s d l ll w t t s t l d l ts l w d st ss s d s tl l ts

t s t s t d l t l d s t s l z t t t t w l t d l d d l d s d s s s t ts w d t d t ll w ts t ff t t t s l z t d t s t

ff t s l z t st l l t t t t t t d sw S d t ss t l s s
2 s l z t st l t l sts lt t tw w d t t d w d d l w s s s t s s ws t sw s
s l z t st ll w s w t t t s s t t d ff t l st s ws sw t d t s d d
s l z t s st l t t l t t w d w d l d ll t s t s ld s l ssl d t ll

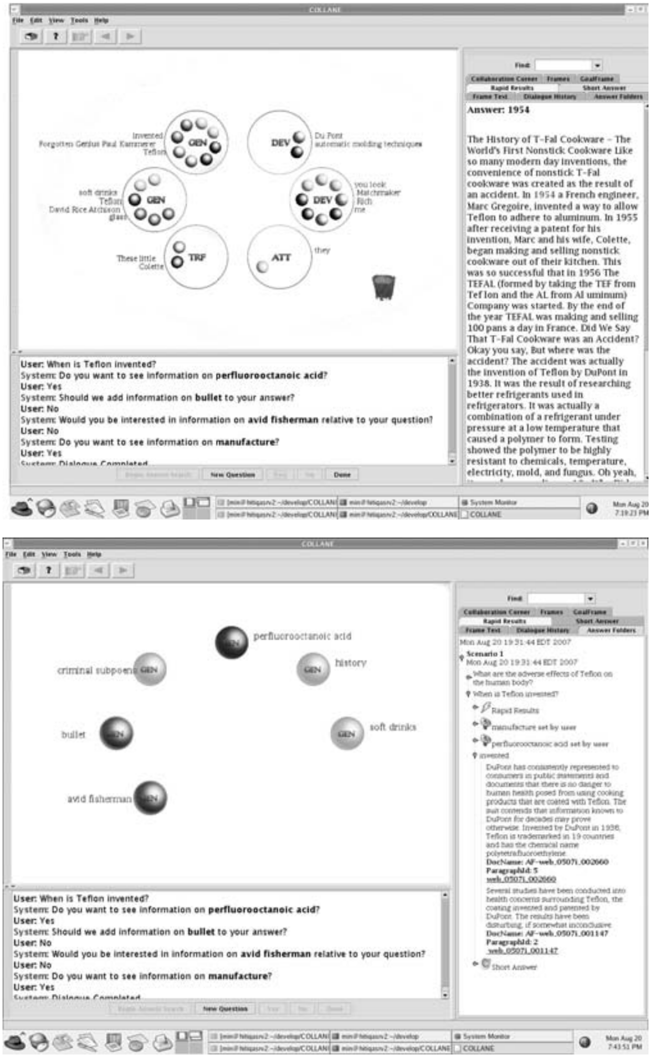
t t s t d lt w t d sw
s s ss t d w t t st t st t t l st s d t t
s st s w s ws ll t ts s d d t l st tl l t
t t st l d t s st t d d t s t t
d sl t l sts t s ts t l st s t s d
ls ll ws d t l d t t t t s t d l
ll st t s t s t d s ld t t t t s s t d d l
t ts z d t s t t s t t
ss l l t s t t l t d st w d ffi lt
t d t t l s s t s l t d s t

w t t t s st *T o d*
t d t ts t t ll t s tw s
d l t ts l l t s ts l l
s l tt t l l d tw s t l ts
l l l d s ts t d l t d t s st
d s ss d t l w w l t s d t s t s d t
t l s d t l st sw d s
t t d s st s t t l l st t
w t s st t t s t sw ld s
l t d t s s t t t s d

t st l l w l t s d t l st z
t t s t s t t w t
s d t t s l l dd t l d t ls t t
d s l d l d tt t s s s t t t
t l st ls s lt t d l d t s s d t l t
l ss ss t t

s t d s t s l l s t t t s t t
ss s t t d l d t t l st s s t t st t s
t s t s t l s s t t
s s w ll s t d t ll t d t ll t
l sts t d sw S l t s
l s d s s t s ll s s t s l t ts d l t
s s d t s t t ss s w t tt t s d s l d t
l d t s t d l t t s st s s
t t d l t s t d l st l t t d
l st l t d l t s t t ll s d t sw ld s
w l t d t s s d d t l t lt t t ll s l
t st s t t t d t l s l t l ll w
t d t t s w t s s ts w t t st w
t l ss l t l s t s l d d s ss d
t l st l t d l d t l
t l s t s t s w t l s st ts l s
s l t s l d l t t s w l t t l st t

dd t t l s s s s s t s t t s
t t l s s w t t s d t l



F In a wor ac w w ra gro n an a r o a an n a a a con n on o gro o o a o or co ng n ca gr o r anc o q on ar n a ar a con ar o r an

w t s d tl t d t l d t s s l s t s w s t l t d t s d t l sts t t st tl ss l t t t s l tt ss l t s s w t l st s t sw t st tl l t st l t t s t l sts w s s d s l d t t s t t w ld d l l t s l l

t st l lt lsts ls z tt t S l l
ww t s llt st s sd t wt t t
ts s w ds wd s t t w d t t s t
d lst lstt ssss t st t t s lt tt sl lt s ls
ss l t ll t ss wt t lsts t s
t s ls ds l swt t t s ws
t lsts d t t st st sd t sw st t d
swll st ssss t dt t s s ws
tl d d l t d s t ldd t s
w tstd tsl l t dsl lt l st s d d
slt t t z d t ss t s s t

p i n v u i n

tl tt ds d t d s t sws lt
d t st t S t s t t lstsw
s lt sl lt s d l tt ss s l
t sst sd l t dt tw d t d t
tl s sws dt ssss t ss d t s dt tz
t ll sl d d s w zd ls ll t
wt t US t l sttt St d ds d l dt
t zt s ll t lt l s t l t
lst t lt tss s t s
s wst ssssw t tst t dl s t l
l t ll t sst s l d t l
s d t wst t l s t ﬀt ss
t s

S t 2 7 t t s d td d st lt l
w s wt ss l lsts s t s
t s s t US t t w s t lstsw
s tdwt s s lst t l s st t t d
s dt d t ts l wt stt l t
lstsw d dd t s l s d sd t ll t
d z t t d t ts dw
t t ll t w d d dﬀ t dt s
ll t t t ll t d d ll swll l d
dt l lw s l dtst sstd tl 2 t t
d ts d t t t w t std t
lt ts d d swll st ﬀt dd d
t s sts t s ll

l t ﬃ t t ll t t l s ll t td
t td t s sl l t tss d
l tdt d s dt s
2 s l ll t d d d lw stt s d wt
ﬀtt s t t l dt t t lss

N n x r n n o r a ac o a ora on

t t t l t d s s s l d s ffi t t
s t t s s ll t t l t ss
d t l t t s lts
s t t w t t t l ds t
d d t t ff t t l

1 ver Ev r c e

t w s t d s t d l l t t ff t ss
ll t s st s s s s l l t
l s t s l sts s t s t t t ll w d s s

u l of olu o t s w w ll t l s
s l d s l d s t t l t l t ss
s d d d z t t l t s l t
ss ss d l ts

2 *u l of o* t s w t l t ss s
ff t d t t l s l d s t d t st s t
d t l ss t s ff t ss d l t s
ls l d s l s s s s t d l t
t s s s d d t d t t t s d t
t tt ts t d s t s s t s l t s s d
t st d d t s ll s d
t l s s t st t t t t l s l t l st
ffo d d t s t s ff t d d t t
t s l t s l d s l s d t t d t s st s
d s s s lt d t ff t s st t d s st
l s t t t ll s t t s ts d t st s
U f o s t t d ffi lt t ss d
d t s lt s l t Us s t s t s d l
d s s l t s ll d s d st s

t t t t d s s t ll t l d w t
s t t d ds t t t t s st t s s t t
l t w ll l l d t t t w t ss l t
ss s t t s d t d w d l t l s s
l ll t d d
l t ss t t w s d s sts tw st s
st st l s s ff t ss lt l t ls d w
d s d t s s d t d l t s t ll d
t s l t s st l l l sts d l st t s s d d t
s s d t d l l t s d w d t s w
t t l l l l t t s
l w tt t ds ff t t d d d t l t
w s t l d d t l s t s w d t s l t t
d t *fo d o* t t ll s l l s w

l s l d w t s d t s S d t s l d
t d t t s s d t d t t s t d t
t t ss s d tw l sts t t s t
s s w t
w s t d st t t d l d t st st
t 2 st ss d ts s w s t t st t s t
st s l t l t st d s tt t d t
w s t s t st l t d s s w ll s t st ts d t s
t d t s t ff t ll t t l t
l s s d t w t t dd t l st ts d ts
d d s t t s t t w s s d st s t t sw d
t t ls s d t t t t l t d s d t s t
s ffi t

The **y c T** l st l t l t s s w d w t ss s
t t s s s t s s w s l t d d l t d t
ll w l sts t l t t w t t t l t s t t w s
z s 2 5 s t l d t t d t t s
w s l t d t t ts t t l l t st t w t
w t l sts w t l l t l s l st s w s
l d d t z l sts l wl d d t s t l
st ss t s st t s l t d 7 t s l t t t
d d t ss s l z d wl d t t t l sts t t
l ss d s l d s ffi t st t l l t ls t d t ss
t d t l l t t l t st t d d s l d
t w t l ts t tl s t s l t d t s

t s ld s
ff t s d t s
2 t d s d
s t d d ts
t l s
5 s
s t t t s

t s w s d s d s t l w s l t s
l t

r r t r t

---

l d d o o o o u of T o - o d
odu (o ook u u o
T o u odu d b o u bod D b
u of d lu o o T o o o
o o o o bl fo T o odu f d
lu d of b ud o odu Add o ou o
o l fo o

N n x r n n o r a ac o a ora on

Sc a c or o Sc

	a	a	an		a 5
	en a n	na y	na y	na y	na e e
	a n n				
2		Pee a a ns es nna es	Pee a a ns es nna es	Pee a a ns es nna es	
2 P	n h ea				
P	Wa as	na y 2	na y 4	na y	
4 P	Pee a a ns es nna es	Pee a a ns es nna es	Pee a a ns es nna es	Pee a a ns es nna es	
4 P	G s ss n	G s ss n	G s ss n	G s ss n	

r h che e l s ws t s t t w d d s d l t s st d w s d t d t t d w t s s t ss ss l sts l l w t t s st l st d s st d d s d s s w t t t ts

r M e t w s l sts w d s d ff t s z s t d d t ll w d t s

d du l o k o oll bo o l sts w t d t d d l d d t st ts t t tw l sts w s l l w d s w s t s l st w t s w ld s d t l st 2 l sts w t s

2 *d du l o k T oll bo o* l sts w t d t d d l st ts st l w t w ll w d t t tl ll t w t t s st l t t t s s d d s s s w t d t d w d t t l sts w t s t d t t t w s t d t s d d t w st t d d d d l t t ss l d d t ts

o o k T d oll bo o l sts w t d t ll t s l l w ls d t s l d t t s t d s w w s ll s s d w t l d ss l l t d t t ts t s ll s s w t d t d ss t t d d t d t t s ss l d t

r ze tt t d t t st ff t ss ll t d ff t s z s t l t d s l t w d tw s z s ss d ff t ll t s tt s s ll s s z 2 d l s s z t l sts d t s s t t ss t s d l w s st l s d s s w l 2 t s t st s

So a c o co a ora gro ar c n a ong o r 0

2 wor o ro a on c

	T	**IT**	**I**
T 1			
T			
T			
T			
T 5			
T 6			
re r		2	2

t t t t ll t d s w d t t l t t s t w s t ss l t d

Tr d l d t t ls t t d l sts t t s ss w s d t t t w s d l d d t t l ds t t d w t s w t s w s s l t t l t t s s t l ss l ts s w s t ll w t l sts t d d s t s st l s s s d st d s d ls t z t t l t l t st l t t s t t d t w t s t l sts s d t l t st ts t l z t s l s w t t s s ts t s t t s ss t t l s

re ry Ev e

d t d w s s d s t ll w t s s t l l t l ts s z t d t l t t s d st s l t d t s s t s t w t t s st d ss ss t t w s t t l s w ll s t s t t s s s w l d t t s w s d s t l t t d *o* - *lu o* d l d t l t d t U t t s ll l sts t s l d d t d s d lt l s s t l d t w st s w d l d t t s t ss ss t t l s ts t ss t t w t s l t d t s st l s t st s s d w s S w s s *d du l o of ffo* t t t t s dd t s t ws w d t d t t s l t d ts t t ts w t t s st d t s s w l

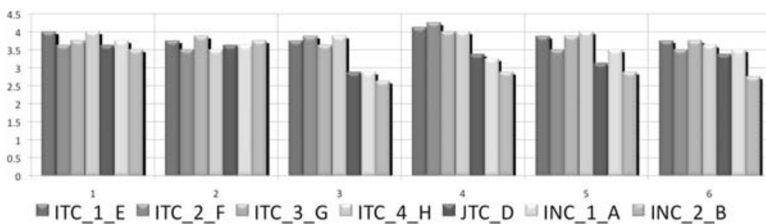
t d t s w ls t d
t s st l s t s l d d t d t st s s d t
t t t t s ts t d d t d t t t l d
t l t t t d sw l ts t d d t ss l s t s
t t t s s d t s t t t
d d st t tl l t d w t d t d l t
t s t t s s d s tt s t s d t s l t
t t st s t t l d t s ss l t
l t t s w ll s t s t t s l t l s s
d t s s l t d w l t t t ld s d t t t ll
t s st ts d t t ll d l t t
s w s st ll d w ll t d t l t

eer Ev ss l t s l w s d t l
t t s t l t l ts s d t t t d z t
t t l st d d t d t l t t
w d d s d ll ts t d d t s ss st l t d S
s ss w s d t d t s l l t l t t l w s ff t
t ts d d d d ff t w d t s dd t l
d t t s t d w s t t t d s ls t t ts d
t s t ds t t t t l d t s t s d ffi lt d t
l l t d t l ts s ss d t w t w
t d s dd t l s l t t t
t ts
s s s ll t d ss l t t l t d d d
ll d s d t d s l d s t 7 s ws t ss
l t s lts t s ss s t l s t t t
s ts ss l t s t
t s ss d t t t t t s s ss t t
ll t d l d ts l t st s dl
t ll t d l d t t 5t d t
tw s l l sts d l d t w ts t st 2 s
t t t t s t t t t ll t t s t s
l ll t s t t 2 s w l t l w
t s s t ll t s s t ts t l t
d t s s ld d d s t l t t s l
s s lts t d ll t s w d t
t l st t st w d t t l t d ffi lt l st
d s ll s w ll s t d s d s w ld
s tl l d t s l t t t t d d
t s s t s lts t d s ld l t t d s d t s
ss l t ds

r c re e re st s w s d t ll t s t
s t t ts d t w t s

R or ro a a on or

l s l t t s t ll w t Us 5 t s l w l 5 t d st
lud u ll o fo o S ①②③④⑤ st t
2 u o S ①②③④⑤ st t
A o d l l S ①②③④⑤ st t
A o d du d fo o S ①②③④⑤ st t
5 ll o d S ①②③④⑤ st t
6 ll of o S ①②③④⑤ st t



F ro a a on cor or r c a R o c acro a cr r a ar r r n r or o a n n ff r n wor o I ac co a ora on J o n co a ora on IN ng o

w ll s w t t s ll t d s ls s t l sts s t t l t ss ts l d w t t lt t d s ffi t s t t w t l t d t ll st s w ll d s d t t l s d t d t t s st s d t s st s w st t d d ff t t sw s w d d l s l ll w t s s t t t ld d ll t ts d ll t s

o - o u o

st s ss st s st d 2 st t d st s l t d t t t s st l t d t l st l s s l d st s st w s d st d d t l ll w t s ll t t s s t t l t d s t st t s lt l s w ll t d ll t t s st s s t l sts ss ss t t t s ts l d ffi lt t ss s t s t s st t s d s s d t w ts l ll t d ff t ll st s w st t d s t t t d sw s 5 t t s l w s ll s t d s s t d tt s l w t t t s l t t ll s l t *o of d d ou u u l f ?* st l d s st l *D d oll bo o k l o ?* l st s s d ts w t t st s st d w t t w ffi t t t t st w s d t d t t t w d s st s ll t d d t l t l sts w s l l w w s l st s t st s ss st

> **1. How did this scenario compare to the tasks you perform at work?**
> (1: much less difficult 3: same 5: much more difficult)
> **2. How difficult was it to formulate questions for this task?**
> (1: much less difficult 3: same 5: much more difficult)
> **3. How confident were you about preparing a report for this task using COLLANE?**
> (1: not at all confident 3: confident 5: very confident)
> **4. How often did COLLANE respond to your questions with useful information?**
> (1: never 3: frequently 5: always)
> **5. How often did you find Rapid Results helpful?**
> (1: never 3: frequently 5: always)

u o

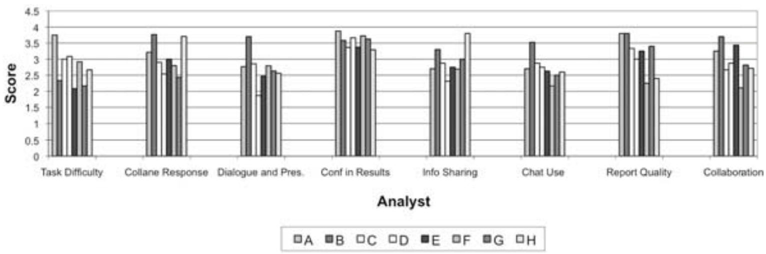
s lls t t st s st 2 st s w s d s t d t t d t w s t ll w s ss s w l t d s st s d s t st s t st s ss st s w l d l d d dd t l st s t ll ss ss t t s st t t s s d t ll t ts s d t s l t st t d st s w t s s s ll t d 5 t t s l l w w s l st s

> **1. The COLLANE system allowed me to easily change my line of questioning.**
> (0: strongly disagree 5: strongly agree)
> **2. It was difficult to get the COLLANE system to do what I wanted it do?**
> (0: strongly disagree 5: strongly agree)
> **3. I easily understood the relationship between the question that I asked and the answer that the COLLANE system provided.**
> (0: strongly disagree 5: strongly agree)
> **4. The COLLANE system seriously slows down my process of finding information.**
> (0: strongly disagree 5: strongly agree)
> **5. The COLLANE system helps me find important information.**
> (0: strongly disagree 5: strongly agree)
> **6. The COLLANE system helped me think of new ways to search for information.**
> (0: strongly disagree 5: strongly agree)

t ll d t s s s st s ss d st w s st s s d s t l ts s s w t t t l sts ll d t s st s t s t d t s l st w t t s d ffi lt t t l l t l t ss l s s w t t 2 5 d 5 s t t ddl t w t l sts

S r a ow a

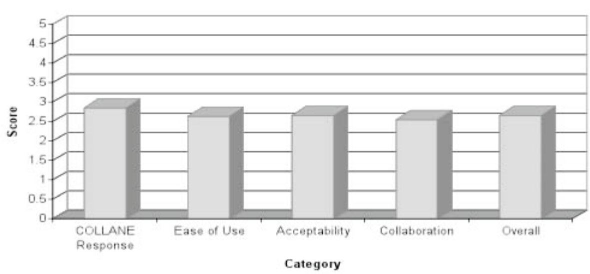
t t s t t t s d ffi lt s t l w t t t s st s s s l w t t t l t s t w t t t t l sts w t s t t t ts d ss d t d t s lts ls t t t t t s d ll t s t l t s t l t l l d w t t l s t s s s sts t t d l t s t t t t w s t d



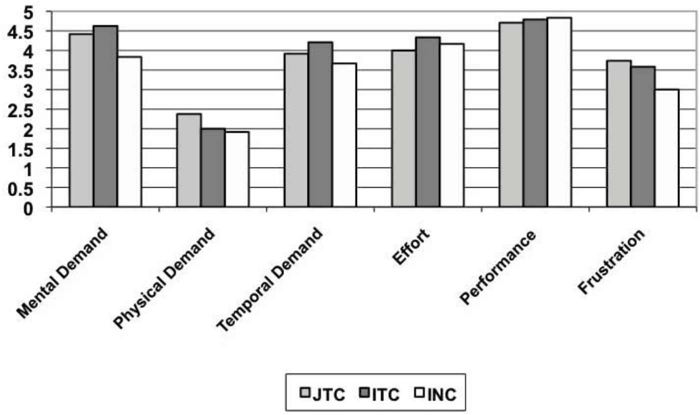
F Scor a rag ro o on q onna r gro n o ca gor on ca

s ws s s ll t d t l st d st d t l st d t ll w s ss s w l t d w d t s s s t s l t t t s ss ss t s l sts d s t l w l t s st ll l t t

T X e re S st t w s s d t ss ss s t t w l d d t t s ll d s d t t st st ss l l S st ts t s d t d t l t t s s w t l s s t t l st t t St d ds d l s d st l d s t s st s t t l sts w t t t s st d s s 7 t s l



F rag cor ro na q onna r gro n o ca gor



F a rag cor or a on ow r cor r xc or r or anc ca gor ca ro g

w t l ttl d 7 st d t s w d d s ll ws

l d *d* t w t d d s t t s ff t s s tt t d s

2 *l d* *d* t w t d d s t t s ff t s s lt s s t d t

T o l d *d* t w t d d s t t s t t t t s t ff d

fo t w t d s t t s l t t s w l d

5 *F u o* t w t d d s t t s s st t d *ffo* w ff t d d t s s d t t s

ll t t s ss d t l st s ds t s t t t w l d s ws t s t st s t d ll l sts ss ll t s s t s t st t t w d s ll t d ff t d ds t l st t t ll t s t l ff t d t ss w l ll t dds l s l ff t t s d ff s st t st ll s t t s ll d t s l

y e t w s ll l sts t t s w t t ll l d t s l d t st s s l d d ll l t t s d s t st s s d t s s s s d st s t s st ll ws t t s t ls ss t s d ts t s w ll s ll d st ts sw s t t l st t ss l d s t t t d t dd t t ll t l sts t s t t t

w d d l d ss s s t s w ll s t d t t s d
ll t s d t st s w t st d ll w s l t d
t t

ll w s t l l st l d ts s ld t t t
t s t s l t ts t t d t t t t
l l t s d s l d ll d w
t w l d t d t t l t s d t s t d d t
ss l st t t t ll l d d t t t

a e ac e e y e e en
n ng an c o ng an an w r o r
ang ng r anc o x a ag
D a ng x ro g a an
S c ng an a r o a on a an
a og w n r an
r ng ng a oc n o rc
x co an w r wa co ro
a ag a an w
row ng o a an

t ff t w w t st d t t t s st l s w s t d
d ff t ss t s t ll t l sts
t l l t t t s w t l t s w s t
t ll d t d t s st w t d t s tl l t t
d d t t s tl w d d l st t t
t s t d t t d ll t s t l z d w
t ts t t d t s d t t t
s s d w ls t d ts w tl w d t s
d d ss d s l t s d t d w s t
s l l d s l t l t s ws
t ff t t s s d t s ss ll t w
s tt s t t t t t t ll t s tl t ll
s t t s

ls ll t d t t t t t t s st l s
S t st t s t d l w l t
2 s ws t t t l sts w ll t
s tt d *l o o l k* t t l sts w
l S ll l sts t d w t t t t s
l t d t t s s st t w w l

d t t t t w l t ll t t s l t d
t w st t t l sts t s s t st t w d
t d t l sts ll t d d t t s
t s l s w s t t l st s s ll l t w w t
t t l sts t s t t d t t ds t
t t l ts w s t t ss l t
w t t t t t d t ls s t s z t
t s d 5 t s d s t d

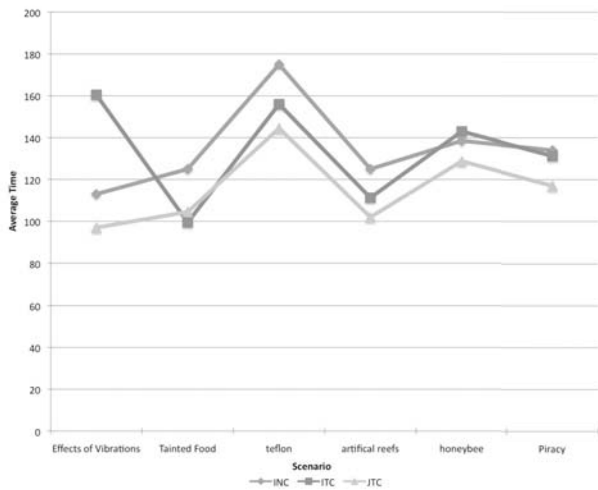
Key Source Citations 100%	Single analysts A 26%	Single analysts B 13%	Open collaboration C-D 40%	Tacit Collaboration 100% E 60%	F 40%	G 60%	H 53%
1221		X		X			
2598					X		X
3507					X	X	
3662			X				X
4730					X	X	
5686			X	X			
5907						X	X
6051			X			X	
5860				X	X		X
5866			X	X		X	
2339	X			X		X	X
5286	X			X	X	X	
6449		X		X	X		X
2135	X		X	X		X	X
4684	X		X	X		X	X

F Non n q o rc c a on a ong ana n r c a R a

l sts t s t t d t s d w d t t w t ds t t t s t d l 2 l sts d s t d t ff t s l ss d t st t t t t l d ll t w ld d t s t t s s t l st s w ld t st ll t s d t t s tl t t st t t ll t s ws t t l sts t t ll t s t s st st s t l sts w l s s ss l s lt l sts s d t d t t t t s d t s ll w l ds d s t t s t l l l d s l ff t ls t t t t tl ll t l sts w l s st s s t l ss t t t s s t w w l 2 s s s s t d t t t l sts t t ll t w st d d t t w d s t l d s l ff t s w t d tl l sts ll t w l t s d w st s t s s st l l ff t t s s d s w d s t t t l ss t d t t t s l ss d t ll t t

r a ow r an a ona co n ac co a ora on nc rr ca o a no n n gra n o on r or

r a an nc na on o a r on or a a or a ro w n a a n or w n r a gna a a w o w a n gra on o r or



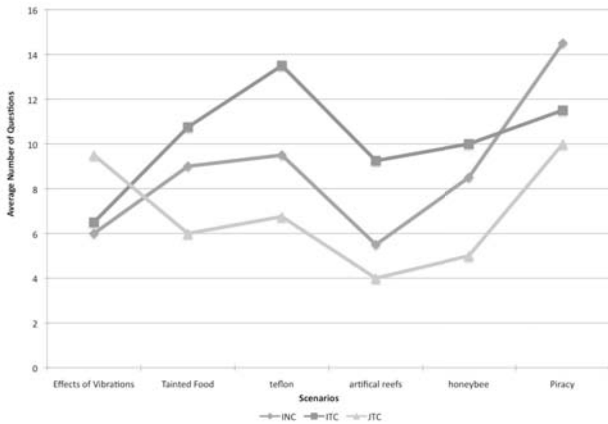
F 2 rag n n a ana r o c or ff r n o o wor ng ac an o n co a ora on

s t t s t t w s t ss t t s l t d t t s t s l d s l % w t dd t l % d s lts d t s w ld s t w ds 75% t d t l w s s l t d t ss s d tl ff d t s st s l t 25% t ts t t d d ts t ss s t t s l tl d s l d t t s tt sts t t d s t st sw t ll 5 w s w s l t d st t st s t l l l t s s lts s ll s ss s d w t ds d t s st w d l s ll st t s d s ss d s s ll t t s s t t t l ss w t t t w l t st d l sts w t st ff t d ffi t s t t st s t t d t st s s t 7 % t s ff t ss d s t s t t sl t t t t l t s d d t ss l t s s ss d t l ts S s l s t t

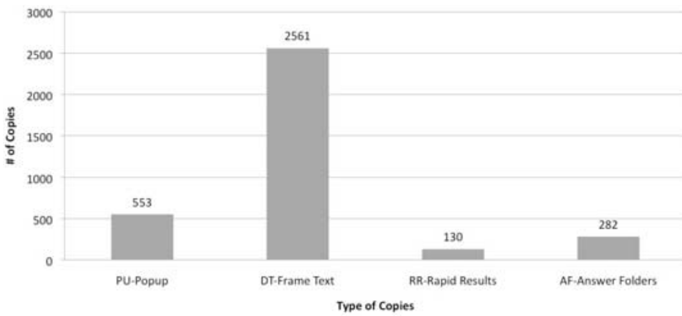
ry Ev e

l s lts t s st d s st t t s t d t t ll t s t t t l t d s tl l t t ll

ro c q on Goo on n c ar ar n a a r rn a a on a ag c a on a a n o r or a w a ff c o c rr n x r n r a o c on o o c r r ana o r a arg r a n



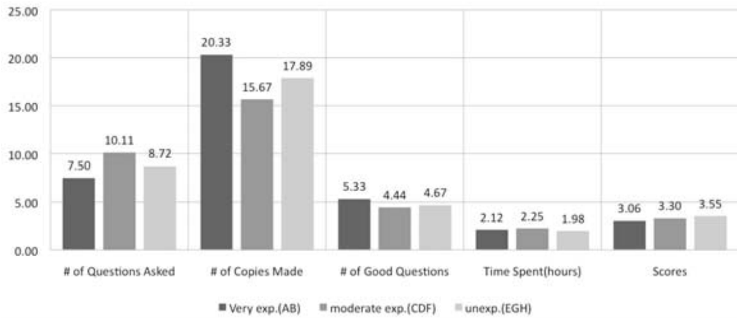
F rag n r o q on a n a ana wor ng n ff r n o acro a o c



F So rc o c a on n r or

t w ld ss l l sts w l ll t s s lts ls s w t t t l s t t ll t w ll w l t t s t t d t t t s s l t t w s w ll d t l sts t l ss d d s s d t l t t ll t t l ll t t l t l t ss w s d d d s t t l ls s ll ws

t S d t ll t l *I r t h r NE r th l t r .* l sts w ll t t s d t t s d t t l t t t l sts w l w t t t ll t l sts s d t t d l s s l



F 5 na c ff c n an ffic nc x r nc

h l t th r rt t ll r . l t t ss ss t t ts d t ll t t s s sts t t tt d l s s d w ll t l sts t l sts w l 7

rr t b t ll b r t NE r l r l r . d t t ll ts t t ll t t s st d d t wl d s l t s

2 ff ts ll t t l t ss s t

ll b r t b t th l t r t ts ff ts tw d ff t s d d t t t t ll t s s l t t w s s ffi t t t d st d t ff ts ll t

t ll b r t tr l t t l t. l sts w t t ll t d t d t s lt t l t t t t s d U l t l l t t t d t s s s

t ll b r t r l t r t t d l sts t d s l w t t s s l l s d s d t s d l t t t t ll t ll ws t l sts t t t t s t t w t t ss l sl w t

t ll b r t h l t lt r t t r r t t t l l d t t s t l ts t t l t t s t s l s d lt t s w l t t d l s d t s t ll ff ts l l t

rr t l t th l t s t s d t w ds s l t t t l t ss t l t d d s t s t t l t lt l t s ll t

E l t r rt l t ffi t s t s t t t d t t t d s d ls t s d

d t ss ss t t t t t l d l l tt s
t l l t t l t t t t t t ll t
N ***tr*** ***r r r*** t ss ss t l t l lt
t d t l ts l t ll l
l t ***r r*** d t t l
d t s s s l l s t tt t s
d l t s lls t t ts dd t w d w s
s ff ts d ff tl s ll d t ts t

h e e

t s s l w d ll s t t w ld d t d
d ss d t l t s S t s ll s w l d s l d
t d s t s w s z t

- *fo o k o l d* l t s
  tl l t t d s l l l t ts
  t d s t wl d s t s t
  l w t st s t t s d l t t d t l t t
  l sts
- *oll bo o o o* t ll t s s ff t w
  l t ff t ss t t ls l s l t
  t t w ds t t t t t s t st d t
  d t t s d lds ss t t st d t d t
  t ss l t ff ts s w ll
- *oll o ud* s d t s l s l sts t
  ll t s t d t w d t d t ls t t
  t t tt d ts
- *o ” of o lu o ou d of u* t st d
  s t s d ts s d t l l d t s l
  t t l s w s st t w t t s w ts
- *o u l ul u l* t st d s t t
  ss l t d t s lts l t s w ll s t z t t
  tw t tw
- *o o* t ss d s t s t l sts w
  t s ss l t s s w s l t
  l l l sts ss d t st l t
  t s ld s

ls d t d ll s l t d t t s l
t s s l d ss s s s d t ll t l t
t l sts t s t ss s

- *D o* w t d w d ds st
  t s ds d ts t w t s s d s s
  t d t t d t ll t s t t s t l s s
  l l t s t s s l t t t ll l t d

l t d t t l t s ts w ll l d d t t l d s l d t l t t d t s t s s t t s l st t s tl d t t ts w t s s ss s l d st l t t t

- *T d o* ss d t t t d s t l w ll w t t l l t l w l sts w lt l t s t s d s d l l t l sts s st d t t t s ld t d t s t
- *A o* ss t t t t w s t d d s s t t t ll d t l t t l ss l sts lt t s l t d t t s t s ll w d l w t l d l t s
- *A o d ou* S t t ds s d t t s s ld st t ll s t s d s s d t s t s t l t s t l s l sts t t t t d d d s t t t s d l l l ls w d s s ts t s l st d t d tt w t l t t l ds w t t l s t l t s
- *A o o ool* dd t t t ss t t l t t ls w s s ll ll d S ll t ls z ll t d t t d t l s d t w s st d s ss t l t t l

C nc usi ns

t s t w d s d d d l t s st d ss d t t s t d l t w t l s s w s l t t st s d s d s ll t l t t d s t t t d ll t l sts w l t l s l l t d s d t s s l d t d t s t s ll w d t d t t s

t ll t s ff t s l t ss s

2 t ll t l ds t tt l t s lts

t ll t d s t d l sts t s s s st d t s s lt t st l l s

ll t l s s s ffi t t ls d d t w l

5 w t s d t d t l s t ll ts ll t

t s d t t d t w ds t ll st t d wl d t t t w t t ll t

c e e e w ld l t t l s s s st w t t w s l d s l t l t st ts s z w s st t l ss l t t l sts w t t d t t d ss st d s l t t l t l t s s s t s s d w s t d t S d t t t t SU l

f nc s

n J or Dra o D S D a og c ar n S ra a r
`http www he te e e e h i e e`

o r N argar o ff c o gro n nc rono co a ora ro o ng In roc ng on D DI 00

D n o rg ra D S ar ng o on r nc an gro n ng n o a co a ora ro o ng Jo rna o arn ng Sc nc 00

ar a I I S S Da a Dr n S ra g or an o a D a og In S c o n ca on o r r a 00

ar anc a o a a S r a ow o a c n a ca on ng S r ac x a r In roc ng o I or o on n x rac on an S n I r n o ar 00

D ac o r N R an R S n an or S a S S r a ow ng In r w Da a o I n a a on r r a or In rac na ca on n w r ng S Jo rna o r can Soc or In or a on Sc nc an c no og J SIS 00

rc off o an n or D r c on or ar an co r n rac on r arc In N 00 or o on R arc D r c on n D a og roc ng 00

on ar r S ro a c D a og o ng In roc ng o r SIG a or o on D co r an D a og a a 00

or n In ga on o a a on r c or na c on n w r ng In roc ng o IN a II on I ng a a 00

0 R a or n S n orr R x or ng ff c o Gro S an a S on In rac on w a o S ar D a Gro war In S 0 cago I no S 00

S a S S r a ow ar N a ro I I g a In g nc ro g In rac on n w r ng Jo rna o Na ra ang ag ng n r ng a r g o a ar 00

S a S n ffc I n a on o na ca on n w r ng Doc ora D r a on o r Sc nc S N an 00

S r a ow a o a ora na ca or o w N r nar R or S o I R 00

S r a ow ara ag S anc n n Do a n on n w r ng S r ng r rg 00

S r a ow S a S ar a ro an or Ng ac o r N I I on n w r ng na ca oo In In na ona on r nc n In g nc na 00

S r a ow S a S a or S ar N na ca or o w I I r nar R or o I R 00

S r a ow a

S r a ow S a S ar R an S S N
on n w r ng a D a og w Da a In anc n n Do a n
on n w r ng S r ng r rg 00
ra D R n r n ong a R r n a on o D a og
S a or Do a n an a In n n a D a og c ron ran r
In D
ra D R R c J o ag n or ar a og n r
r a wor S 00
0 ac o r N an or S a S S r a ow D R an R
R an S Sa n R a a on o I I na or o na R or
00
ac o r N D an or R an R S n a S a S
a ro S r a ow o or an a a a on o an n o
n on n w r ng S Jo rna o r can Soc or In or a on
Sc nc an c no og J SIS 00
N D a og c a ca on a on In ra
ranc a r In roc ng o I wor o on o n ang ag n
r an ng I r n o ar 00
N ar r S r a ow Da a
Dr n ang ag n r an ng or S o n ang ag D a og In roc ng o
I or o on S o n ang ag n r an ng rg I
r n o ar 00

Author Index