自然语言处理
week-9

凌震华

2024年5月9日

□Information Extraction (信息抽取)

- □Named Entity Recognition
- □Relation Detection and Classification
- □Template Filling
- □Biomedical Information Extraction

Information Extraction (IE) Overview

- This topic is just a series of reuses of existing techniques to solve specific problems
  - Partial parsing/chunking
  - ML sequence labeling
  - Classification/ambiguity resolution
- Application of NLP techniques

Named Entity Recognition (NER)

- Find and classify all the named entities 命名实体 in a text.
- What's a named entity?
  - A mention 提及 of an entity using its name.
    - *Kansas Jayhawks*
  - This is a subset of the possible mentions...
    - *Kansas, Jayhawks, the team, it, they*
- Find means identify the exact span of the mention
- Classify means determine the category of the entity being referred to

NE Types

Type	Tag	Sample Categories
People	PER	Individuals, fictional characters, small groups
Organization	ORG	Companies, agencies, political parties, religious groups, sports teams
Location	LOC	Physical extents, mountains, lakes, seas
Geo-Political Entity	GPE	Countries, states, provinces, counties
Facility	FAC	Bridges, buildings, airports
Vehicles	VEH	Planes, trains, and automobiles

NE Types

Type	Example
People	*Turing* is often considered to be the father of modern computer science.
Organization	The *IPCC* said it is likely that future tropical cyclones will become more intense.
Location	The *Mt. Sanitas* loop hike begins at the base of *Sunshine Canyon*.
Geo-Political Entity	*Palo Alto* is looking at raising the fees for parking in the University Avenue district.
Facility	Drivers were advised to consider either the *Tappan Zee Bridge* or the *Lincoln Tunnel*.
Vehicles	The updated *Mini Cooper* retains its charm and agility.

Ambiguity

Name	Possible Categories
Washington	Person, Location, Political Entity, Organization, Facility
Downing St.	Location, Organization
IRA	Person, Organization, Monetary Instrument
Louis Vuitton	Person, Organization, Commercial Product

[$_{PERS}$ Washington] was born into slavery on the farm of James Burroughs.
[$_{ORG}$ Washington] went up 2 games to 1 in the four-game series.
Blair arrived in [$_{LOC}$ Washington] for what may well be his last state visit.
In June, [$_{GPE}$ Washington] passed a primary seatbelt law.
The [$_{FAC}$ Washington] had proved to be a leaky ship, every passage I made...

NER Approaches

- As with partial parsing and chunking there are two basic approaches (and hybrids)
  - Rule-based (regular expressions)
    - Lists of names
    - Patterns to match things that look like names
    - Patterns to match the environments that classes of names tend to occur in.
  - ML-based approaches
    - Get annotated training data
    - Extract features
    - Train systems to replicate the annotation

Hand-written Patterns for Information Extraction

- If extracting from automatically generated web pages, simple regex patterns usually work.
  - Amazon page
  - <div class="buying"><h1 class="parseasinTitle">(.*?)</h1>
- For certain restricted, common types of entities in unstructured text, simple regex patterns also usually work.
  - Finding (US) phone numbers
  - (?:\(?[0-9]{3}\)?[-.])?[0-9]{3}[-.]?[0-9]{4}

Rule-based Extraction Examples

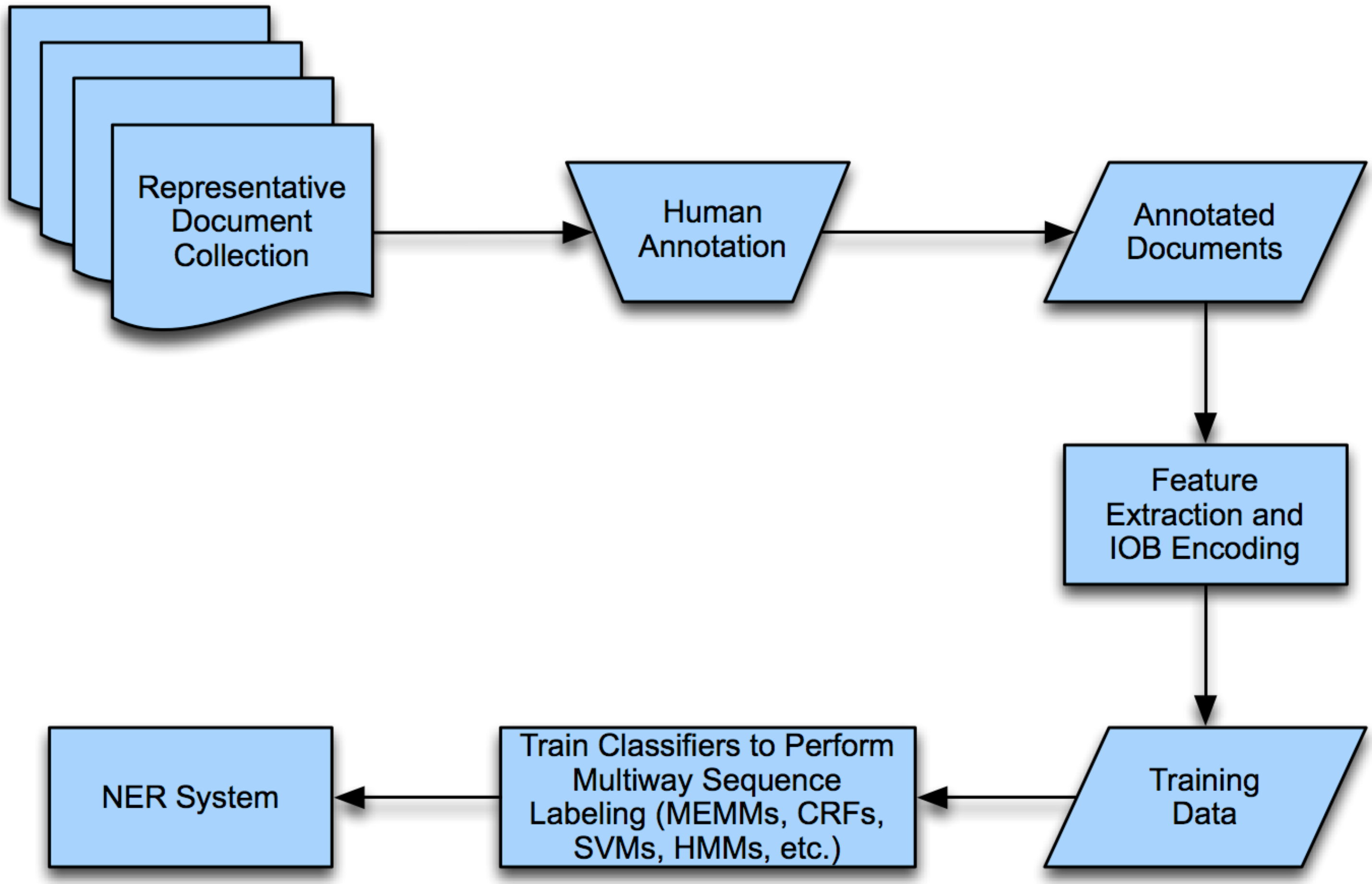
Determining which person holds what office in what organization

- [person] , [office] *of* [org]
  - Vuk Draskovic, leader of the Serbian Renewal Movement
- [org] (*named*, *appointed*, etc.) [person] Prep [office]
  - NATO appointed Wesley Clark as Commander in Chief

Determining where an organization is located

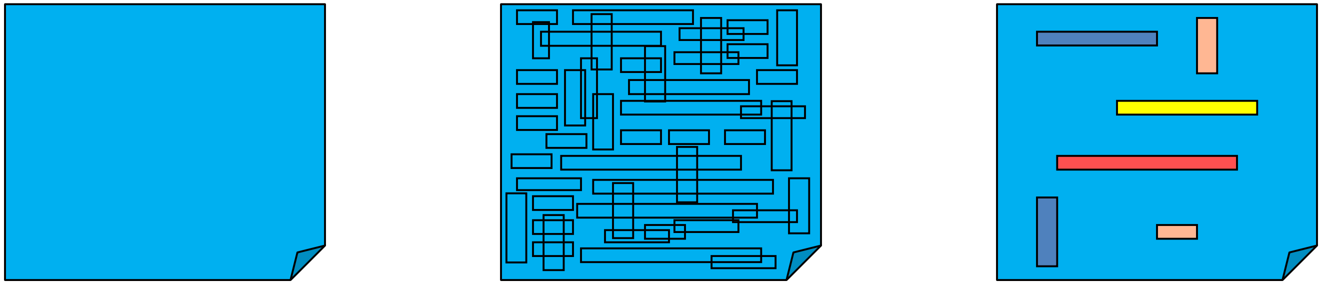
- [org] *in* [loc]
  - NATO headquarters in Brussels
- [org] [loc] (*division*, *branch*, *headquarters*, etc.)
  - KFOR Kosovo headquarters

ML Approach


Representative Document Collection
Human Annotation
Annotated Documents
Feature Extraction and IOB Encoding
Training Data
Train Classifiers to Perform Multiway Sequence Labeling (MEMMs, CRFs, SVMs, HMMs, etc.)
NER System

Naïve use of text classifier

- Use conventional classification algorithms to classify substrings of document as "*to be extracted*" or not.



- In some simple but compelling domains, this naive technique is remarkably effective.

‘Change of Address’ email

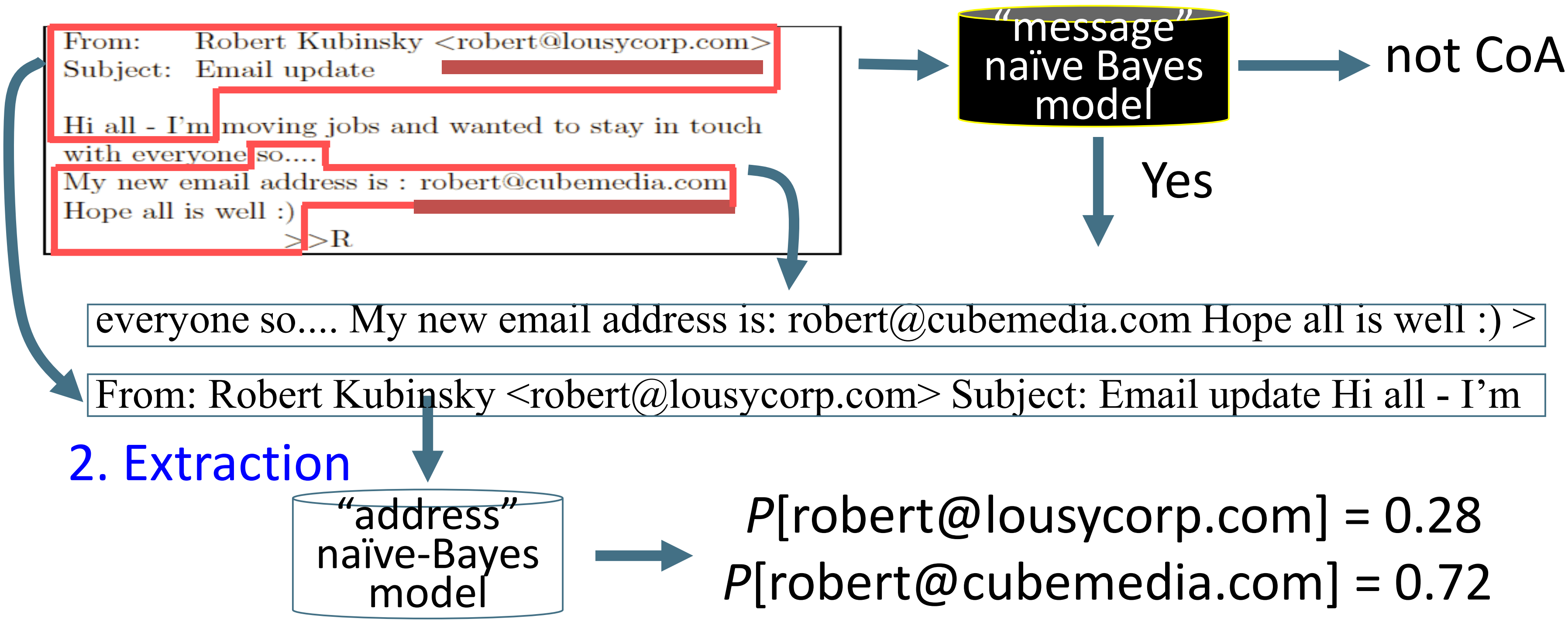
From: Robert Kubinsky <robert@lousycorp.com>
Subject: Email update

Hi all - I’m moving jobs and wanted to stay in touch with everyone so....
My new email address is : robert@cubemedia.com
Hope all is well :)
>>R

Change-of-Address detection

[Kushmerick et al., ATEM 2001]

1. Classification



The ML sequence model for NER

Training

1. Collect a set of representative training documents
2. Label each token for its entity class or other (O)
3. Design feature extractors appropriate to the text and classes
4. Train a sequence classifier to predict the labels from the data

Testing

1. Receive a set of testing documents
2. Run sequence model inference to label each token
3. Appropriately output the recognized entities

Encoding for Sequence Labeling

- We can use the same IOB encoding here that we used for chunking:
  - For N classes we have 2*N+1 tags
    - An I and B for each class and a O for outside any class.
  - Each token in a text gets a tag.

Encoding classes for sequence labeling

	IO encoding	IOB encoding
Fred	PER	B-PER
showed	O	O
Sue	PER	B-PER
Mengqiu	PER	B-PER
Huang	PER	I-PER
's	O	O
new	O	O
painting	O	O

Features for sequence labeling

- Words
  - Current word (essentially like a learned dictionary)
  - Previous/next word (context)
- Other kinds of inferred linguistic classification
  - Part-of-speech tags
- Label context
  - Previous (and perhaps next) label

Features: Word shapes

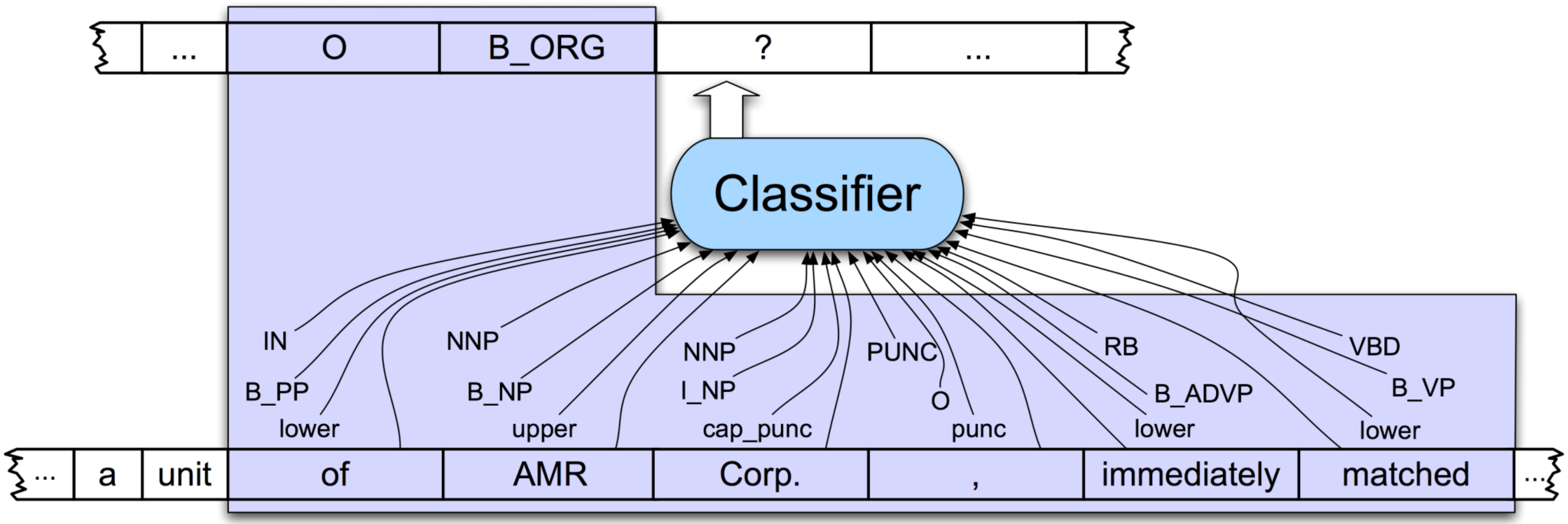
- Word Shapes
  - Map words to simplified representation that encodes attributes such as length, capitalization, numerals, Greek letters, internal punctuation, etc.

Varicella-zoster	*Xx-xxx*
mRNA	*xXXX*
CPA1	*XXXd*

NER Features

Features				Label
American	NNP	B_{NP}	cap	B_{ORG}
Airlines	NNPS	I_{NP}	cap	I_{ORG}
,	PUNC	O	punc	O
a	DT	B_{NP}	lower	O
unit	NN	I_{NP}	lower	O
of	IN	B_{PP}	lower	O
AMR	NNP	B_{NP}	upper	B_{ORG}
Corp.	NNP	I_{NP}	cap_punc	I_{ORG}
,	PUNC	O	punc	O
immediately	RB	B_{ADVP}	lower	O
matched	VBD	B_{VP}	lower	O
the	DT	B_{NP}	lower	O
move	NN	I_{NP}	lower	O
,	PUNC	O	punc	O
spokesman	NN	B_{NP}	lower	O
Tim	NNP	I_{NP}	cap	B_{PER}
Wagner	NNP	I_{NP}	cap	I_{PER}
said	VBD	B_{VP}	lower	O
.	PUNC	O	punc	O

NER as Sequence Labeling



MEMM inference in systems

- For a Conditional Markov Model (CMM) a.k.a. a Maximum Entropy Markov Model (MEMM), the classifier makes a single decision at a time, conditioned on evidence from observations and previous decisions
- A larger space of sequences is usually explored via search

Decision Point

Local Context

-3	-2	-1	0	+1
DT	NNP	VBD	???	???
The	Dow	fell	22.6	%

Features

W_0	22.6
W_{+1}	%
W_{-1}	fell
T_{-1}	VBD
T_{-1}-T_{-2}	NNP-VBD
hasDigit?	true
…	…

CRFs [Lafferty, Pereira, and McCallum 2001]

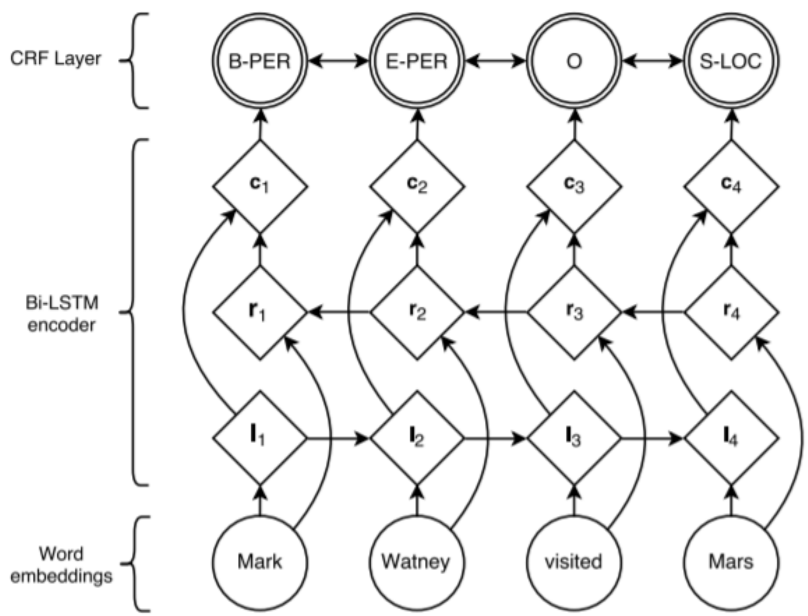
- Another sequence model: Conditional Random Fields (CRFs)
- A whole-sequence conditional model rather than a chaining of local models.

$$P(c|d,\lambda)=\frac{\exp\sum_i \lambda_i f_i(c,d)}{\sum_{c'}\exp\sum_i \lambda_i f_i(c',d)}$$

- The space of c’ s is now the space of sequences
  – But if the features f_i remain local, the conditional sequence likelihood can be calculated exactly using dynamic programming
- Training is slower, but CRFs avoid causal-competition biases

Neural approaches

- Character-based word embeddings
  - Helps with suffix, prefix features
- Use bi-directional LSTMs with a CRF output layer
- Active research area
  - "Neural architectures for Named Entity Recognition", Lample et. al., NAACL-HLT 2016. pages 260-270



NER Evaluation

Task: Predict entities in a text

Foreign	ORG
Ministry	ORG
spokesman	O
Shen	PER
Guofang	PER
told	O
Reuters	ORG
:	:

} Standard evaluation is per entity, *not* per token

NER Evaluation

- As with chunking it is a bad idea to evaluation sequence labelers at the tag level.
  - Most labels are O; so just guessing O gives a learning algorithm a lot of credit.
- So we need to evaluation P/R/F at the entity level.
  - But we may not care equally about all kinds of entities
    - So we might weight them differently in the evaluation routine.

Precision/Recall/F1 for IE/NER

- P=TP/(TP+FP)
- R=TP/(TP+FN)
- F1 = 2PR/(P+R)

	Positive	Negative
True	TP	TN
False	FP	FN

- The measure behaves a bit funnily for IE/NER when there are *boundary errors* (which are *common*):
  - First Bank of Chicago announced earnings …
- This counts as both a FP and a FN
- Selecting *nothing* would have been better
- Some other metrics (e.g., MUC scorer) give partial credit (according to complex rules)

The Full Task of Information Extraction

As a family of techniques:

Information Extraction =
segmentation + classification + association + clustering

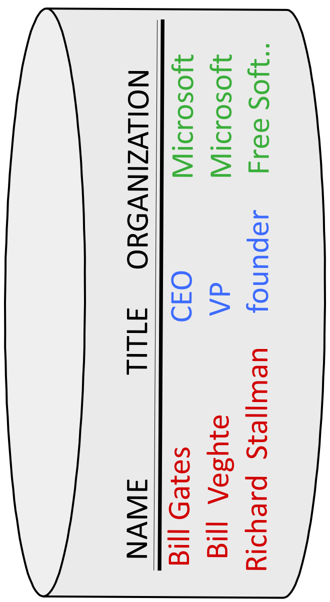
For years, Microsoft Corporation CEO Bill Gates railed against the economic philosophy of open-source software with Orwellian fervor, denouncing its communal licensing as a "cancer" that stifled technological innovation.

Now Gates himself says Microsoft will gladly disclose its crown jewels--the coveted code behind the Windows operating system--to select customers.

"We can be open source. We love the concept of shared source," said Bill Veghte, a Microsoft VP. "That's a super-important shift for us in terms of code access.“

Richard Stallman, founder of the Free Software Foundation, countered saying…

Microsoft Corporation CEO Bill Gates
Gates Microsoft
Bill Veghte Microsoft VP
Richard Stallman founder Free Software Foundation



Slide by Andrew McCallum

Landscape of IE Tasks:
Document Formatting

Text paragraphs without formatting

Astro Teller is the CEO and co-founder of BodyMedia. Astro holds a Ph.D. in Artificial Intelligence from Carnegie Mellon University, where he was inducted as a national Hertz fellow. His M.S. in symbolic and heuristic computation and B.S. in computer science are from Stanford University.

Non-grammatical snippets, rich formatting & links

Barto, Andrew G.	(413) 545-2109	barto@cs.umass.edu	CS276
Professor. Computational neuroscience, reinforcement learning, adaptive motor control, artificial neural networks, adaptive and learning control, motor development.			
Berger, Emery D.	(413) 577-4211	emery@cs.umass.edu	CS344
Assistant Professor.			
Brock, Oliver	(413) 577-0334	oli@cs.umass.edu	CS246
Assistant Professor.			
Clarke, Lori A.	(413) 545-1328	clarke@cs.umass.edu	CS304
Professor. Software verification, testing, and analysis; software architecture and design.			

Grammatical sentences and some formatting & links

Dr. Steven Minton - Founder/CTO
Dr. Minton is a fellow of the American Association of Artificial Intelligence and was the founder of the Journal of Artificial Intelligence Research. Prior to founding Fetch, Minton was a faculty member at USC and a project leader at USC's Information Sciences Institute. A graduate of Yale University and Carnegie Mellon University, Minton has been a Principal Investigator at NASA Ames and taught at Stanford, UC Berkeley and USC.

- Press

Contact

- General information
- Directions maps

Tables

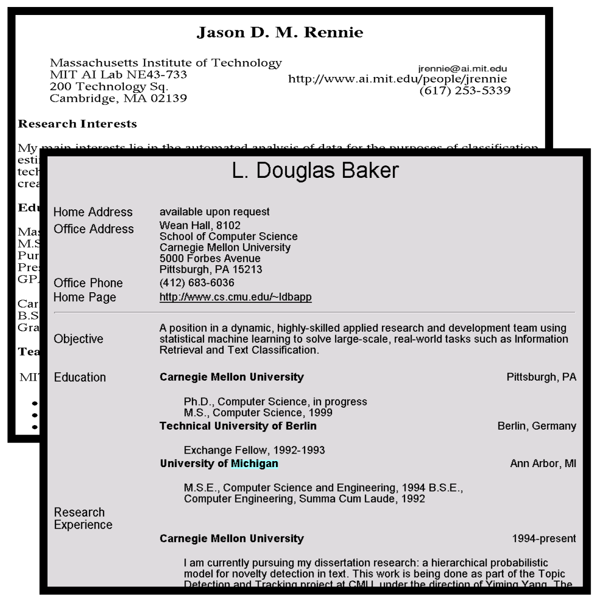
8:30 - 9:30 AM	**Invited Talk: Plausibility Measures: A General Approach for Representing Uncertainty** *Joseph Y. Halpern, Cornell University*				
9:30 - 10:00 AM	Coffee Break				
10:00 - 11:30 AM	Technical Paper Sessions:				
Cognitive Robotics	**Logic Programming**	**Natural Language Generation**	**Complexity Analysis**	**Neural Networks**	**Games**
739: A Logical Account of Causal and Topological Maps *Emilio Remolina and Benjamin Kuipers*	116: A-System: Problem Solving through Abduction *Marc Denecker, Antonis Kakas, and Bert Van Nuffelen*	758: Title Generation for Machine-Translated Documents *Rong Jin and Alexander G. Hauptmann*	417: Let's go Nats: Complexity of Nested Circumscription and Abnormality Theories *Marco Cadoli, Thomas Eiter, and Georg Gottlob*	179: Knowledge Extraction and Comparison from Local Function Networks *Kenneth McGarry, Stefan Wermter, and John MacIntyre*	71: Iterative Widening *Tristan Cazenave*
549: Online-Execution of ccGolog Plans *Henrik Grosskreutz and Gerhard Lakemeyer*	131: A Comparative Study of Logic Programs with Preference *Torsten Schaub and Kewen*	246: Dealing with Dependencies between Content Planning and Surface Realisation in a Pipeline Generation	470: A Perspective on Knowledge Compilation *Adnan Darwiche and Pierre Marquis*	258: Violation-Guided Learning for Constrained Formulations in Neural-Network Time-Series	353: Temporal Difference Learning Applied to a High Performance Game-Playing

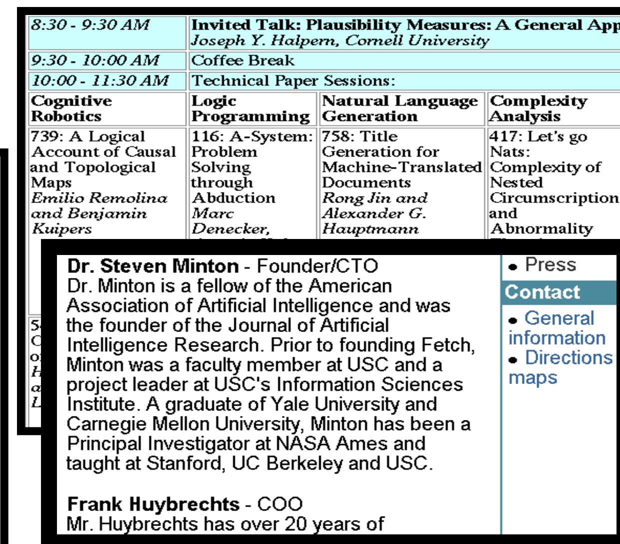
Slide by Andrew McCallum

Landscape of IE Tasks

Intended Breadth of Coverage

Web site specific	**Genre specific**	**Wide, non-specific**
Formatting	Layout	Language
Amazon.com Book Pages	Resumes	University Names





Slide by Andrew McCallum

Landscape of IE Tasks :

Complexity of entities/relations

Closed set

U.S. states

He was born in Alabama…

The big Wyoming sky…

Regular set

U.S. phone numbers

Phone: (413) 545-1323

The CALD main office is 412-268-1299

Complex pattern

U.S. postal addresses

University of Arkansas
P.O. Box 140
Hope, Al

Headquarters:
1128 Main Street, 4th Floor
Cincinnati, Ohio 45210

Ambiguous patterns, needing context and many sources of evidence

Person names

…was among the six houses sold by Hope Feldman that year.

Pawel Opalinski, Software Engineer at WhizBang Labs.

Slide by Andrew McCallum

Landscape of IE Tasks:

Arity 数量 of relation

Jack Welch will retire as CEO of General Electric tomorrow. The top role at the Connecticut company will be filled by Jeffrey Immelt.

Single entity

Person: **Jack Welch**

Person: **Jeffrey Immelt**

Location: **Connecticut**

Binary relationship

Relation: **Person-Title**
Person: **Jack Welch**
Title: **CEO**

Relation: **Company-Location**
Company: **General Electric**
Location: **Connecticut**

N-ary record

Relation: **Succession**
Company: **General Electric**
Title: **CEO**
Out: **Jack Welsh**
In: **Jeffrey Immelt**

"Named entity" extraction

Slide by Andrew McCallum

- Information Extraction (信息抽取)
  - Named Entity Recognition
  - Relation Detection and Classification
  - Template Filling
  - Biomedical Information Extraction

Relations

- Once you have captured the entities in a text you might want to ascertain how they relate to one another.
  - Here we're just talking about explicitly stated relations

Information Extraction

CHICAGO (AP) — Citing high fuel prices, United Airlines said Friday it has increased fares by $6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit AMR, immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL, said the increase took effect Thursday night and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York

Relation Types

- As with named entities, the list of relations is application specific. For generic news texts...

Relations		Examples	Types
Affiliations			
	Personal	*married to, mother of*	PER → PER
	Organizational	*spokesman for, president of*	PER → ORG
	Artifactual	*owns, invented, produces*	(PER | ORG) → ART
Geospatial			
	Proximity	*near, on outskirts*	LOC → LOC
	Directional	*southeast of*	LOC → LOC
Part-Of			
	Organizational	*a unit of, parent of*	ORG → ORG
	Political	*annexed, acquired*	GPE → GPE

Relations

- By relation we really mean sets of tuples 元组
  - Think about populating a database.

Relations	
United is a unit of UAL	$PartOf = \{\langle a,b\rangle, \langle c,d\rangle\}$
American is a unit of AMR	
Tim Wagner works for American Airlines	$OrgAff = \{\langle c,e\rangle\}$
United serves Chicago, Dallas, Denver, and San Francisco	$Serves = \{\langle a,f\rangle, \langle a,g\rangle, \langle a,h\rangle, \langle a,i\rangle\}$

Relation Analysis

- We can divide this task into two parts
  - Determining if 2 entities are related
  - And if they are, classifying the relation
- The reason for doing this is two-fold
  - Cutting down on training time for classification by eliminating most pairs
  - Producing separate feature-sets that are appropriate for each task.

Relation Analysis

- Let's just worry about named entities within the same sentence

```
function FINDRELATIONS(words) returns relations

    relations ← nil
    entities ← FINDENTITIES(words)
    forall entity pairs ⟨e1, e2⟩ in entities do
        if RELATED?(e1, e2)
            relations ← relations+CLASSIFYRELATION(e1, e2)
```

Features

- We can group the features (for both tasks) into three categories
  - Features of the named entities involved
  - Features derived from the words between and around the named entities
  - Features derived from the syntactic environment that governs the two entities

Features

- Features of the entities
  - Their types
    - Concatenation of the types
  - Headwords of the entities
    - *George Washington Bridge*
  - Words in the entities
- Features between and around
  - Particular positions to the left and right of the entities
    - *+/- 1, 2, 3*
  - Bag of words between

Features

- Syntactic environment
  - Constituent path through the tree from one to the other
  - Base syntactic chunk sequence from one to the other
  - Dependency path

Example

- For the following example, we're interested in the possible relation between American Airlines and Tim Wagner.
  - *American Airlines, a unit of AMR Inc., immediately matched the move, spokesman Tim Wagner said.*

Entity-based features		
	Entity_1 type	ORG
	Entity_1 head	*airlines*
	Entity_2 type	PERS
	Entity_2 head	*Wagner*
	Concatenated types	ORGPERS
Word-based features		
	Between-entity bag of words	{ *a, unit, of, AMR, Inc., immediately, matched, the, move, spokesman* }
	Word(s) before Entity_1	NONE
	Word(s) after Entity_2	*said*
Syntactic features		
	Constituent path	$NP \uparrow NP \uparrow S \uparrow S \downarrow NP$
	Base syntactic chunk path	$NP \rightarrow NP \rightarrow PP \rightarrow NP \rightarrow VP \rightarrow NP \rightarrow NP$
	Typed-dependency path	$Airlines \leftarrow_{subj} matched \leftarrow_{comp} said \rightarrow_{subj} Wagner$

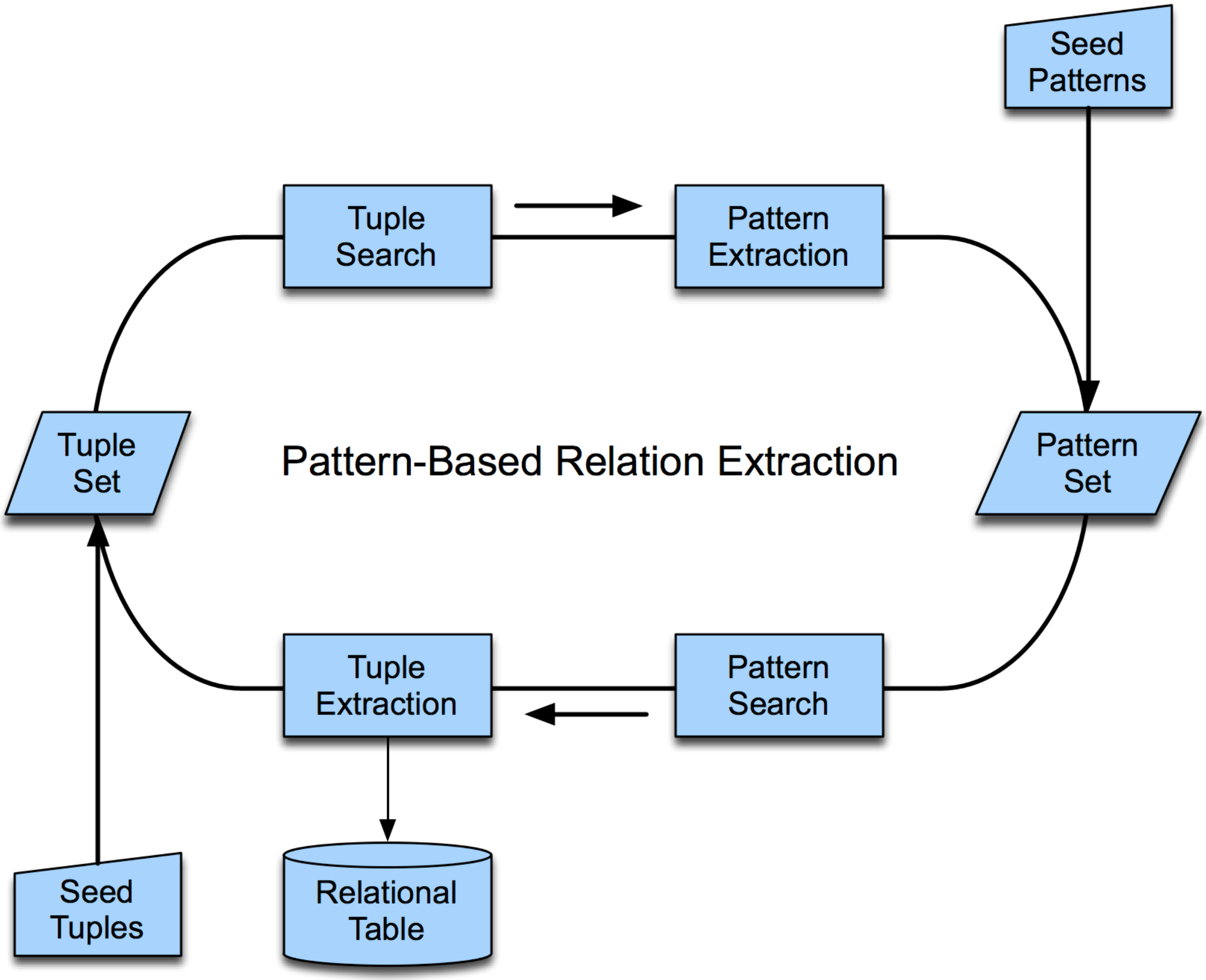
Bootstrapping Approaches

- What if you don't have enough annotated text to train on.
  - But you might have some seed tuples
  - Or you might have some patterns that work pretty well
- Can you use those seeds to do something useful?
  - Co-training and active learning use the seeds to train classifiers to tag more data to train better classifiers...
  - Bootstrapping tries to learn directly (populate a relation) through direct use of the seeds

Bootstrapping Example: Seed Tuple

- <Mark Twain, Elmira> Seed tuple
  - Grep (google)
  - "Mark Twain is buried in Elmira, NY."
    - X is buried in Y
  - "The grave of Mark Twain is in Elmira"
    - The grave of X is in Y
  - "Elmira is Mark Twain' s final resting place"
    - Y is X's final resting place.
- Use those patterns to grep for new tuples that you don't already know

Bootstrapping Relations



- Information Extraction (信息抽取)
  - Named Entity Recognition
  - Relation Detection and Classification
  - Template Filling
  - Biomedical Information Extraction

Template Filling 模板填充

- For stories/texts with stereotypical 模式化 sequences of events, participants, props etc.
- Represent these facts as slots 槽 and slot-fillers 槽填充: templates 模板 (frames, scripts, schemas)
  - Evoke the right template
  - Identify the story elements that fill each slot

Airline Example

FARE-RAISE ATTEMPT:	LEAD AIRLINE:	UNITED AIRLINES
	AMOUNT:	$6
	EFFECTIVE DATE:	2006-10-26
	FOLLOWER:	AMERICAN AIRLINES

Template-Filling

- Two approaches
  - Cascades of transducers
    - FASTUS https://www.isi.edu/~hobbs/fastus-schabes-jul95.pdf
  - Supervised ML as Sequence Labeling
    - Two approaches
      - One seq classifier per slot
      - One big sequence classifier

Resolving coreference

John Fitzgerald Kennedy was born at 83 Beals Street in Brookline, Massachusetts on Tue 29, 1917, at 3:00 pm,[7] the second son of Joseph P. Kennedy, Sr., and Rose Fitzgerald; I turn, was the eldest child of John "Honey Fitz" Fitzgerald, a prominent Boston political figu was the city's mayor and a three-term member of Congress. Kennedy lived in Brookline fo and attended Edward Devotion School, Noble and Greenough Lower School, and the Dexter School through 4th grade. In 1927, the family moved to 5040 Independence Avenue in Riverdale, New York City; two years later, they moved to 294 Pondfield Road in Bronxville, New York Kennedy was a member of Scout Troop 2 (and was the first Boy Scout to become Preside Kennedy spent summers with his family at their home in Hyannisport, Massachusetts, and Christmas and Easter holidays with his family at their winter home in Palm Beach, Florida. 5th through 7th grade, Kennedy attended Riverdale Country School, a private school for b 8th grade in September 1930, the 13-year old Kennedy attended Canterbury School in New Milford, Connecticut.

Practical Issues

- Language-specific features
  - – Character classes
  - – Capitalization Patterns
- Instead of IOB, use IOBES
  - – E=end of entity; S=singleton word entity
- Specialized knowledge resources
    - Gazetteers 地名词典
      - – Manually developed per domain

Rough Accuracy of IE

Information type	Accuracy
Entities	90-98%
Attributes	80%
Relations	60-70%
Events	50-60%

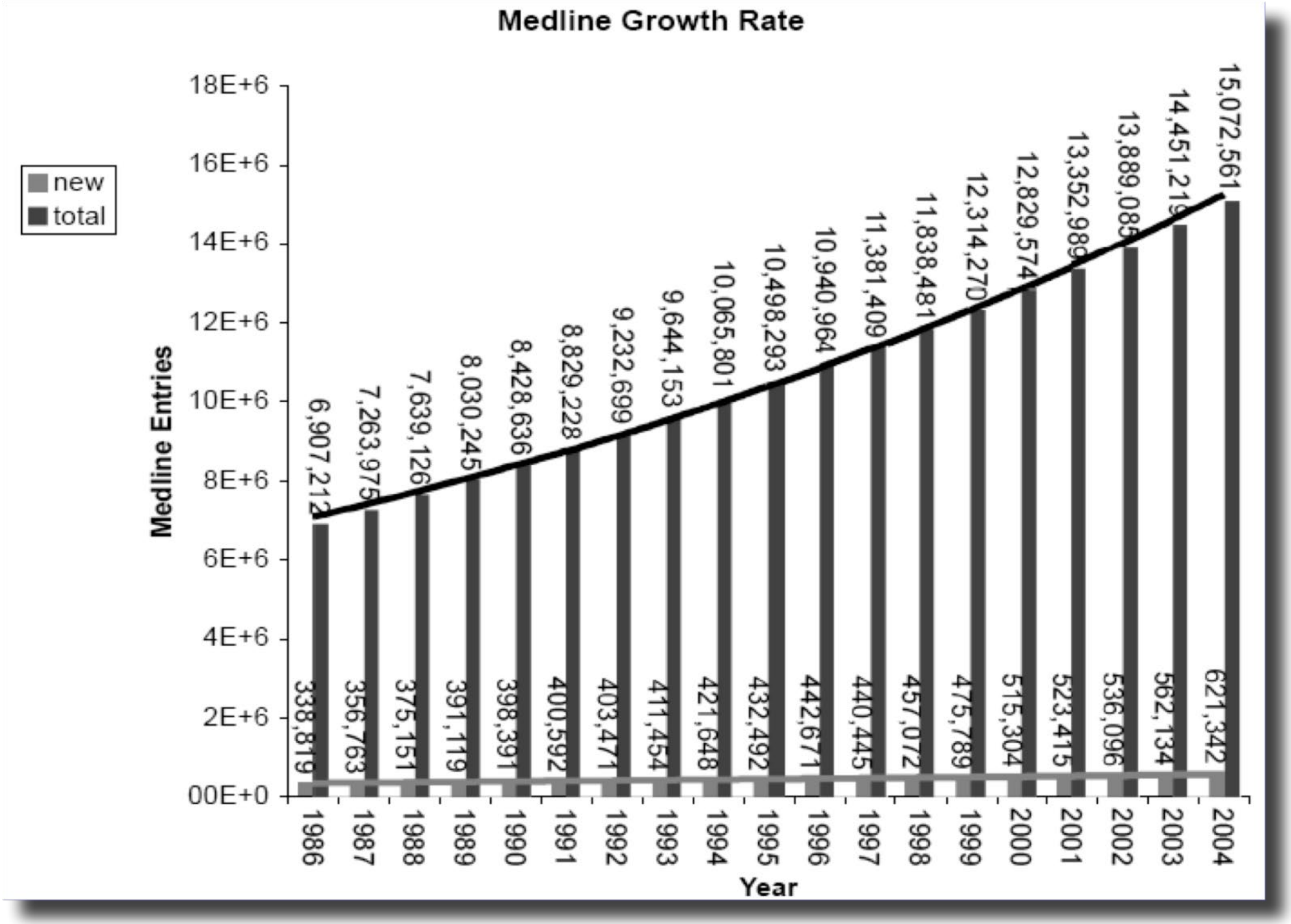
- Errors cascade (error in entity tag → error in relation extraction)
- These are very rough, actually optimistic, numbers
  - Hold for well-established tasks, but lower for many specific/novel IE tasks

- Information Extraction (信息抽取)
  - Named Entity Recognition
  - Relation Detection and Classification
  - Template Filling
  - Biomedical Information Extraction

Bioinformatics NLP

- An example domain
  - Very important
  - Practitioners care about the technology
    - They have problems they're trying to solve
  - Lots and lots of text available
  - Lots of interesting problems

Lots of Text


Medline Growth Rate
new
total
Medline Entries
18E+6
16E+6
14E+6
12E+6
10E+6
8E+6
6E+6
4E+6
2E+6
00E+0
Year
1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004
6,907,212
7,263,975
7,639,126
8,030,245
8,428,636
8,829,228
9,232,699
9,644,153
10,065,801
10,498,293
10,940,964
11,381,409
11,838,481
12,314,270
12,829,574
13,352,989
13,889,085
14,451,219
15,072,561
338,819
356,763
375,151
391,119
398,391
400,592
403,471
411,454
421,648
432,492
442,671
440,445
457,072
475,789
515,304
523,415
536,096
562,134
621,342

Problem Areas

- Mainly variants of NER and relation analysis
  - NER
    - Detecting and classifying named entities
    - And also *normalization*
      - Mapping that named entity to a particular entity in some external database or ontology
  - Relation analysis
    - How various biological entities interact

Bio NER

- Large number of fairly specific types
- Wide (really wide) variation in the naming of entities
  - Gene names
    - *White, insulin, BRCA1, ether a go-go, breast cancer associated 1, etc.*

Bio NER Types

Semantic class	Examples
Cell lines	*T98G, HeLa cell, Chinese hamster ovary cells, CHO cells*
Cell types	*primary T lymphocytes, natural killer cells, NK cells*
Chemicals	*citric acid, 1,2-diiodopentane, C*
Drugs	*cyclosporin A, CDDP*
Genes/proteins	*white, HSP60, protein kinase C, L23A*
Malignancies	*carcinoma, breast neoplasms*
Medical/clinical concepts	*amyotrophic lateral sclerosis*
Mouse strains	*LAFT, AKR*
Mutations	*C10T, Ala64 → Gly*
Populations	*judo group*

Bio Relations

- Combination of IE and SRL-style relation analysis

(22.27) [$_{THEME}$ Full-length cPLA2] was [$_{TARGET}$ phosphorylated] stoichiometrically by [$_{AGENT}$ p42 mitogen-activated protein (MAP) kinase] in vitro... and the major site of phosphorylation was identified by amino acid sequencing as [$_{SITE}$ Ser505]

Relation Extraction: Disease Outbreaks

May 19 1995, Atlanta -- The Centers for Disease Control and Prevention, which is in the front line of the world's response to the deadly Ebola epidemic in Zaire , is finding itself hard pressed to cope with the crisis…

Information Extraction System

Date	***Disease Name***	***Location***
Jan. 1995	Malaria	Ethiopia
July 1995	Mad Cow Disease	U.K.
Feb. 1995	Pneumonia	U.S.

Relation Extraction: Protein Interactions

"We show that CBF-A and CBF-C interact with each other to form a CBF-A-CBF-C complex and that CBF-B does not interact with CBF-A or CBF-C individually but that it associates with the CBF-A-CBF-C complex."

$$\text{CBF-A} \xleftrightarrow[\text{complex}]{\text{interact}} \text{CBF-C}$$

$$\text{CBF-B} \xrightarrow[\text{associates}]{} \text{CBF-A-CBF-C complex}$$