

# Curs 5-6: Cum este făcut un sistem de întrebare-răspuns?

Tehnici de bază în  
Prelucrarea Limbajului Natural

Cu un număr de slide-uri oferite de Adrian Iftene

# Domeniul

## Prelucrării Limbajului Natural

- Își propune să studieze limbajul uman ca mijloc de comunicare om-mașină, pentru a intermedia comunicarea om-om, ba chiar și mașină-mașină
- Scopul ultim: transformarea imensei acumulări de cunoștințe umane cuprinse în biblioteca planetară într-o zonă de explorare automată
- Instrumentele de lucru: teoretice, empirice și aplicative
- Un domeniu pentru lingviști și informaticieni, dar mai ales pentru lingviști-informaticieni

# Tehnologia limbajului natural

- Limbajul vorbit
- Limbajul scris
- Limbajul în corelație cu alte modalități de expresie (multimodalitate)

# Tehnologiile limbajului vorbit

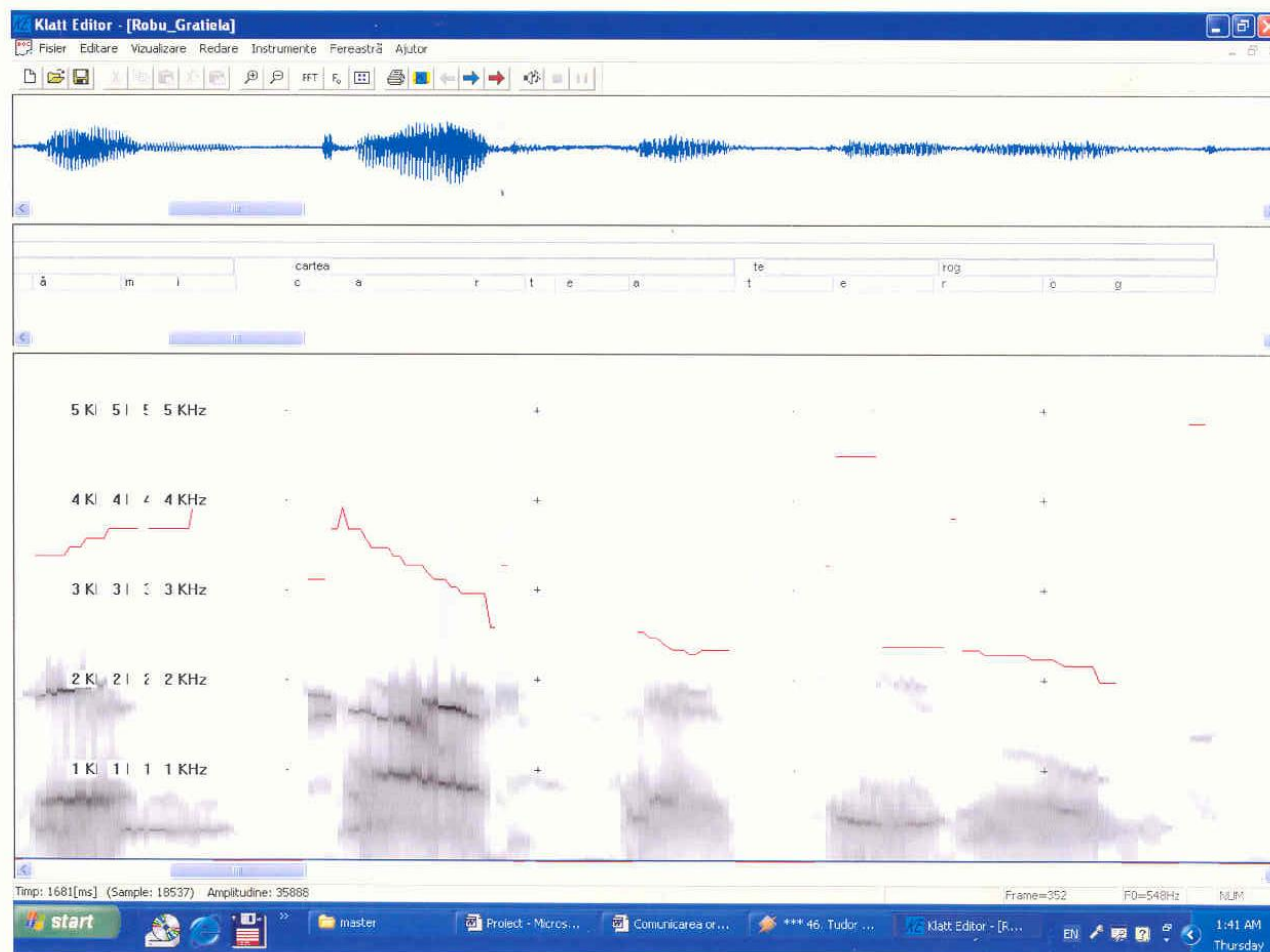
- Interpretarea vocii
  - reprezentarea semnalelor vocale
  - recunoașterea vorbirii
  - analiza prozodiei
  - recunoașterea vorbitorului
- Sinteza vocală

# Prelucrarea semnalului sonor

Robu Grațiela-Andreea

Masterat Lingvistică Computațională – anul al II-lea

Facultatea de Informatică, Univ. „Al. I. Cuza”



Prin bunăvoința Acad. H.N.Teodorescu

# Tehnologiile limbajului scris

- Prelucrări sub-sintactice (nivelul cuvintelor și al grupurilor)
- Prelucrări sintactice (nivelul propozițiilor și al frazelor)
- Prelucrări la nivel de discurs (ce se află dincolo de frază?)

# Tehnologiile limbajului scris

- Analiza și înțelegerea limbajului
  - prelucrări sub-sintactice
    - unitățile lexicale
    - granițele de frază
    - granițele de propoziții
    - partea de vorbire și marca morfologică
    - lema
    - numele de entități
    - grupurile (nominale, verbale, prepoziționale etc.) și atracțiile lexicale (colocații)

# Fraze

Solicitat să comenteze un editorial recent al lui Dinu Patriciu, în care acesta preciza că nu crede în social-liberalism și să aprecieze dacă, astfel, a dat o lovitură de imagine USL, Antonescu a spus că nu știe dacă Patriciu s-a referit la USL. | El a precizat că USL este o uniune social-liberală pentru că reunește un partid social-democrat cu un partid național liberal. |



# Propoziții

Solicitat să comenteze un editorial recent al lui Dinu Patriciu, | în care acesta preciza | că nu crede în social-liberalism | și să aprecieze | dacă, astfel, a dat o lovitură de imagine USL, | Antonescu a spus | că nu știe | dacă Patriciu s-a referit la USL. |

# Unități lexicale

Solicitat să comenteze un editorial recent al lui  
Dinu Patriciu , în care acesta preciza că nu  
crede în social-liberalism și să aprecieze dacă ,  
astfel , a dat o lovitură de imagine USL ,  
Antonescu a spus că nu știe dacă Patriciu s - a  
referit la USL .

# Lema și partea de vorbire

Solicitat	solicita	vb.
să	să	conj.
comenteze	comenta	vb.
un	un	art.nehot.
editorial	editorial	sb.
recent	recent	adj.
...		

# Adnotarea morfologică

- English

```
0
1  He  he  subj:>2    @SUBJ PRON
2  did  do  v-ch:>4    @+FAUXV V
3  not  not  neg:>2    @ADVL NEG-PART
4  know know main:>0  @-FMAINV V
5  her  she  subj:>6    @OBJ PRON
6  name name obj:>4@-FMAINV V
```

- Romanian

```
<TOK ID="TOK478" lemma="Nu" pv="Particle" Type="negation">Nu</TOK>
<TOK ID="TOK479" lemma="ști" pv="Verb" Type="main" Mood="indic."
  Tense="imperfect" Person="third" Number="singular">știa</TOK>
<TOK ID="TOK480" lemma="cum" pv="Adverb" type="int_rel">cum</TOK>
<TOK ID="TOK481" lemma="ei" pv="Pronoun" Type="pers" Person="third"
  Gender="feminine" Number="singular" Case="accusative">o</TOK>
<TOK ID="TOK482" lemma="chea" pv="Verb" Type="main" Mood="indic."
  Tense="present" Person="third">cheamă</TOK>
```

# Nivelul morfologic: cuvintele flexionează

mănânc

mănânci

mănâncă

mâncăm

mâncați

mănâncă

Cuvintele sunt formate din  
părți fixe și părți variabile

# Nivelul morfologic: cuvintele flexionează

mănânc

mănânci

mănâncă

m\_\_âncăm

m\_\_âncăți

mănâncă

Cuvintele sunt formate din  
părți fixe și părți variabile

# Nivelul morfologic: cuvintele flexionează

mănânc+\_

mănânc+i

mănânc+ă

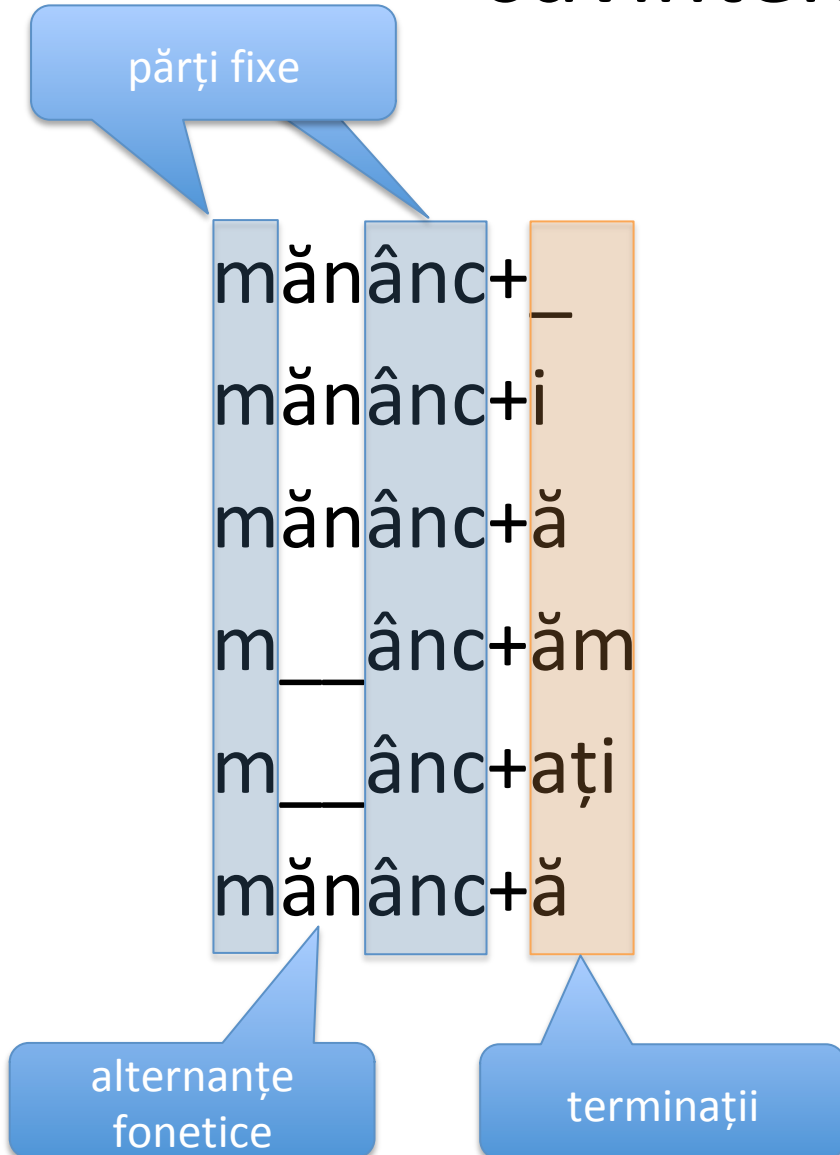
m\_\_ânc+ăm

m\_\_ânc+ați

mănânc+ă

Cuvintele sunt formate din  
părți fixe și părți variabile

# Nivelul morfologic: cuvintele flexionează

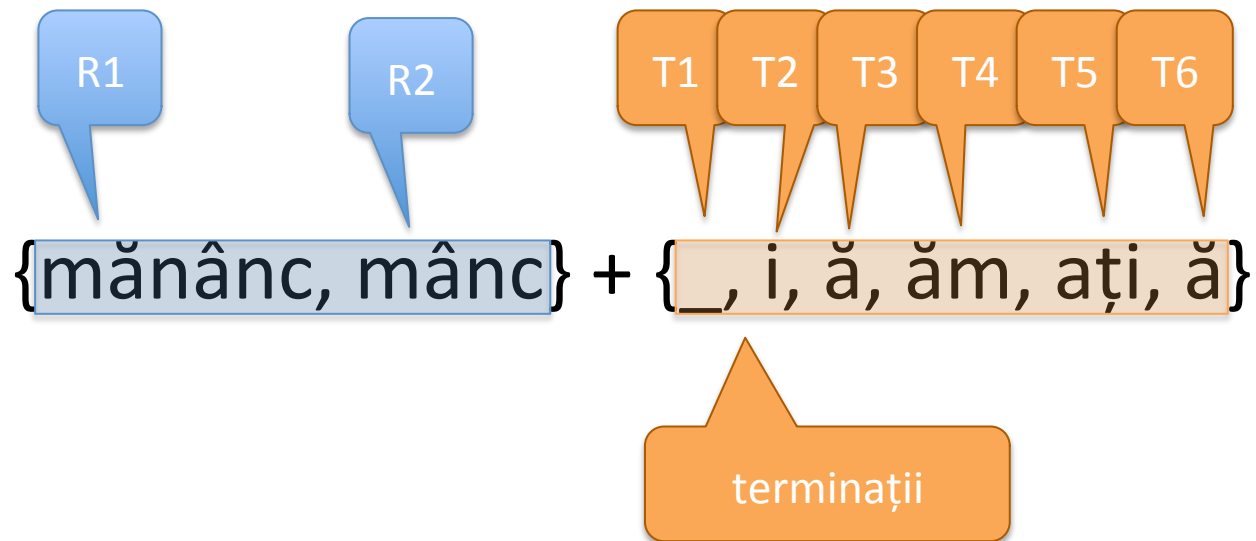


Cuvintele sunt formate din  
părți fixe și părți variabile

# Morfologie fonetică



# Nivelul morfologic: cuvintele flexionează



R1=>{T1, T2, T3, T6}

R2=>{T4, T5}

Morfologie  
paradigmatică

# Nivelul morfologic: cuvintele flexionează

mănânc  $\Leftrightarrow p=1, n=s$

mănânci  $\Leftrightarrow p=2, n=s$

mănâncă  $\Leftrightarrow p=3, n=s$

mâncăm  $\Leftrightarrow p=1, n=p$

mâncați  $\Leftrightarrow p=2, n=p$

mănâncă  $\Leftrightarrow p=3, n=p$

Abordarea brutală:  
memorează toate formele  
flexionate ca atare

# Grupuri nominale

Solicitat să comenteze [un editorial recent al lui  
[Dinu Patriciu]], în [care] [acesta] preciza că nu  
crede în [social-liberalism] și să aprecieze  
dacă, astfel, a dat [o lovitură de [image]]  
[USL], [Antonescu] a spus că nu știe dacă  
[Patriciu] s-a referit la [USL].

# Adnotare la leamnă, grup nominal și sintaxă

```
<NP ID="NP903" HEADID="W3190" VERBPOS="W3191">
  <W ID="W3190" POS="PRON" NUM="SG" GENDER="M" ROLE="SUBJ"      LEMMA="he"
  LINK="W3191" LINKTYPE="subj">He</W> </NP>
<W ID="W3191" POS="V" ROLE="+FAUXV" LEMMA="do" LINK="W3193" LINKTYPE="v-ch">did</
W>
<W ID="W3192" POS="NEG-PART" ROLE="ADVL" LEMMA="not" LINK="W3191"
LINKTYPE="neg">not</W>
<W ID="W3193" POS="V" ROLE="-FMAINV" LEMMA="know" LINK="W3189"
LINKTYPE="main">know</W>
<NP ID="NP1188" HEADID="W3195">
  <NP ID="NP904" HEADID="W3194" VERBPOS="W3189">
    <W ID="W3194" POS="PRON" NUM="SG" GENDER="F"      ROLE="OBJ"
    LEMMA="she" LINK="W3195" LINKTYPE="subj">her</W> </NP>
    <W ID="W3195" POS="V" ROLE="-FMAINV" LEMMA="name" LINK="W3193"
    LINKTYPE="obj">name</W> </NP>
```

# Entități de tip nume

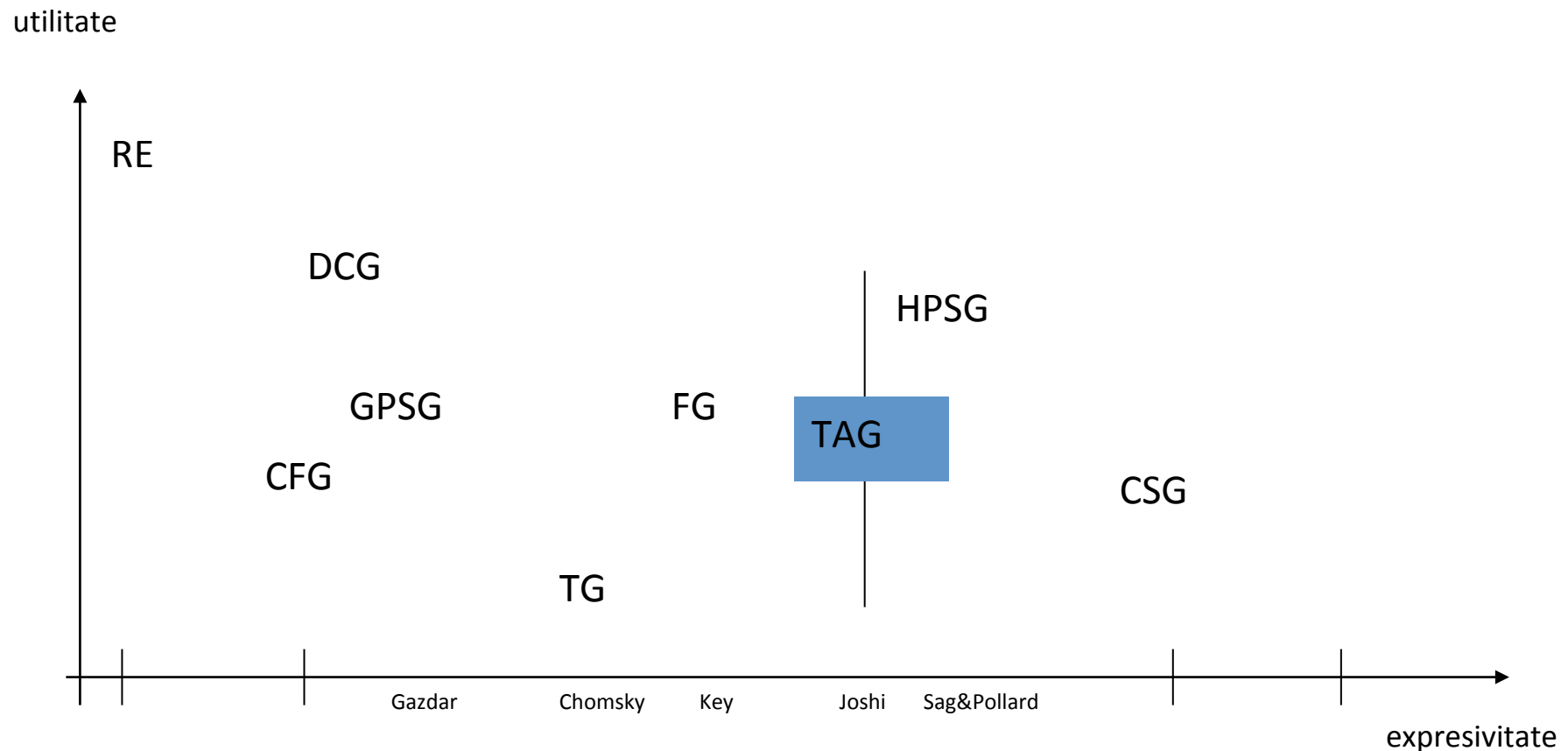
Solicitat să comenteze un editorial recent al lui **Dinu Patriciu**, în care acesta preciza că nu crede în social-liberalism și să aprecieze dacă, astfel, a dat o lovitură de imagine **USL**, **Antonescu** a spus că nu știe dacă **Patriciu** s-a referit la **USL**.



# Tehnologiile limbajului scris

- Analiza și înțelegerea limbajului
  - prelucrări sintactice
    - formalisme gramaticale
    - parsarea → structura sintactică a frazei

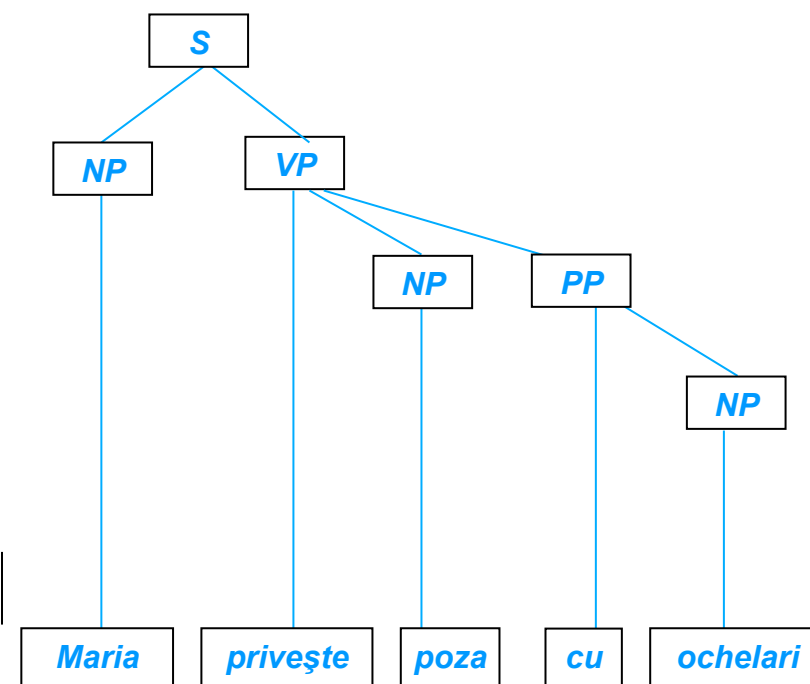
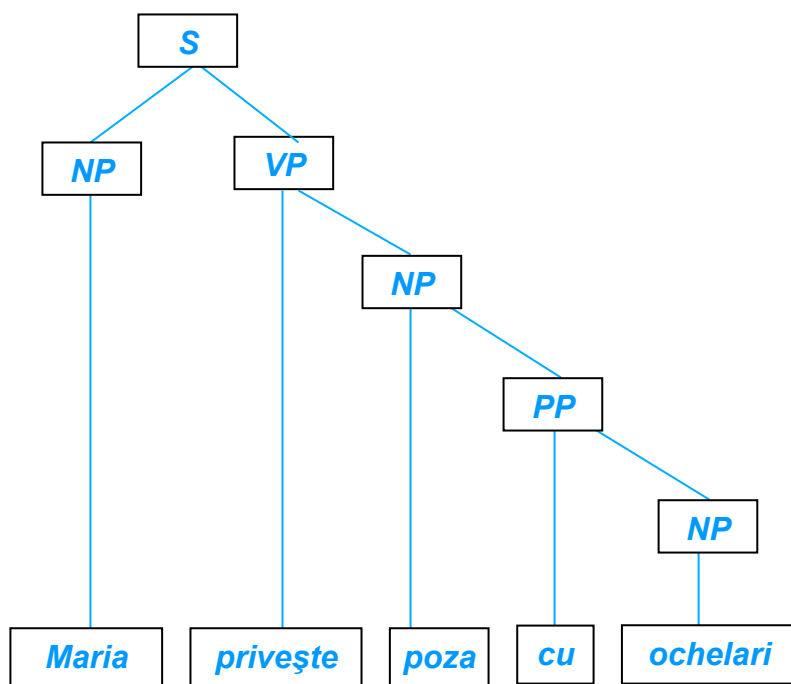
# De la expresii regulate la gramatici de constituenți conduse de *head*



# Gramatici de constituienți.

## Ambiguități sintactice

*Maria privește poza cu ochelari.*





# Gramatici funcționale de dependență (FDG)

- Structura de arbore:
  - Nodurile sunt cuvinte
  - Arcele exprimă relații sintactice între cuvinte

# Relații funcționale în FDG

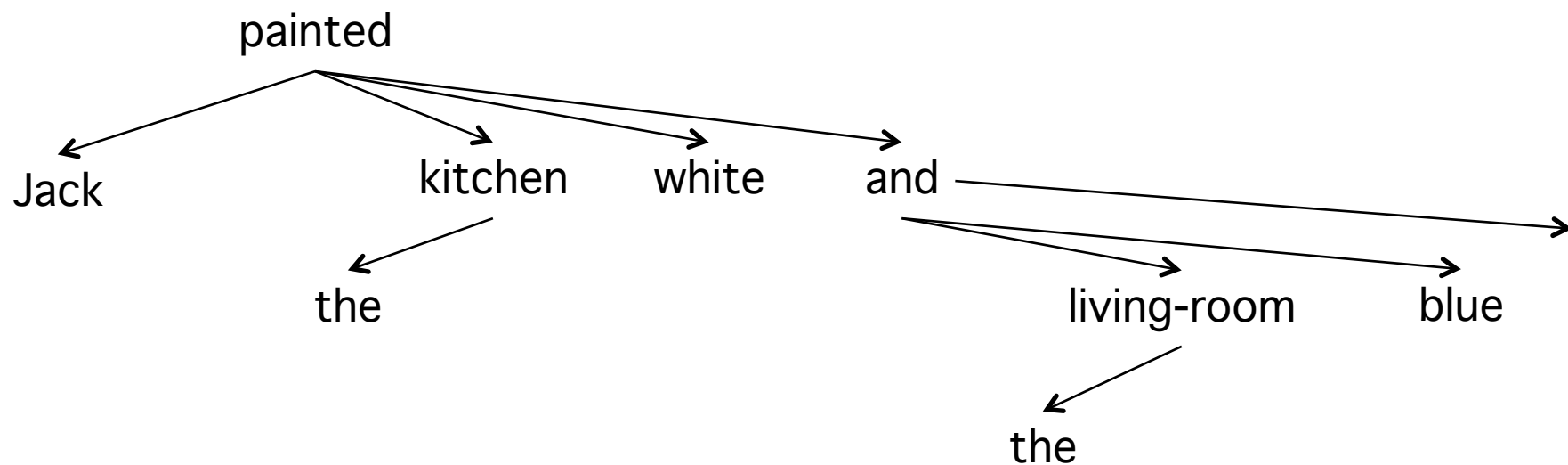


**I see a bird.**

Original în

<http://www.ling.helsinki.fi/~tapanain/dg/doc/TR-1/node1.html#SECTION00010000000000000000>

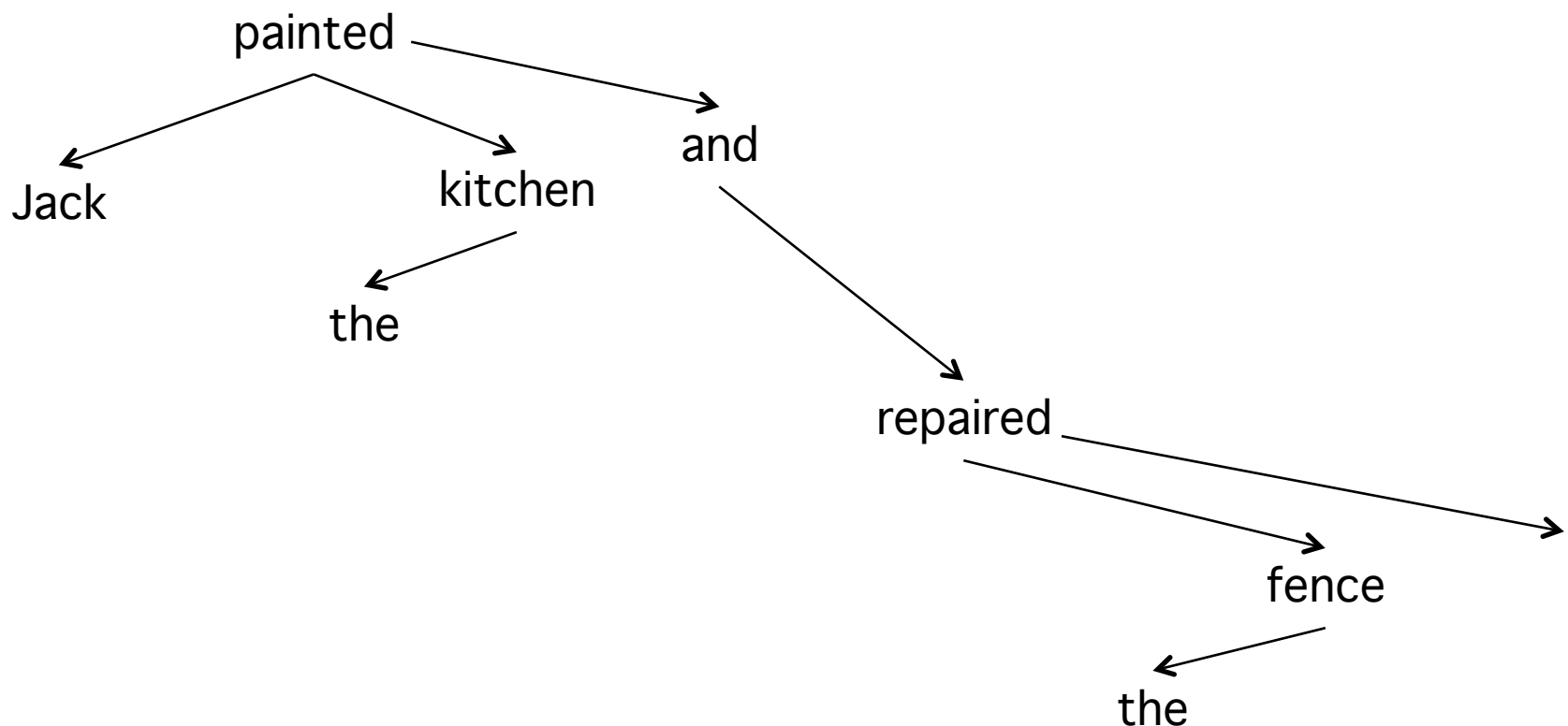
# Relații funcționale în FDG



**Jack painted the kitchen white and the living-room blue.**

Original în <http://www.ling.helsinki.fi/~tapanain/dg/doc/TR-1/node38.html#fig973>

# Relații funcționale în FDG



**Jack painted the kitchen and repaired the fence.**

Original în <http://www.ling.helsinki.fi/~tapanain/dg/doc/TR-1/node38.html#fig973>

# Parsare

- *Top-down* versus *bottom-up*
- *Depth-first* versus *breadth-first*
- Algoritmul CYK
- Diagrame active
- Tomita...

# *Top-down* versus *bottom-up* parsing

## ***Top-down***

- se fixează un scop (structura bănuită a reprezenta fraza)
- scopul curent se descompune succesiv în constituienți (subscopuri)
- se repetă acest proces până se ajunge la terminali, moment în care se face confruntarea acestora cu cei aflați în șirul de intrare



proces direcționat de scop  
(*goal-driven*)

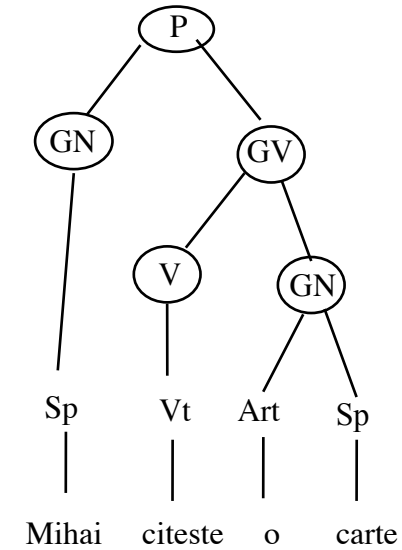
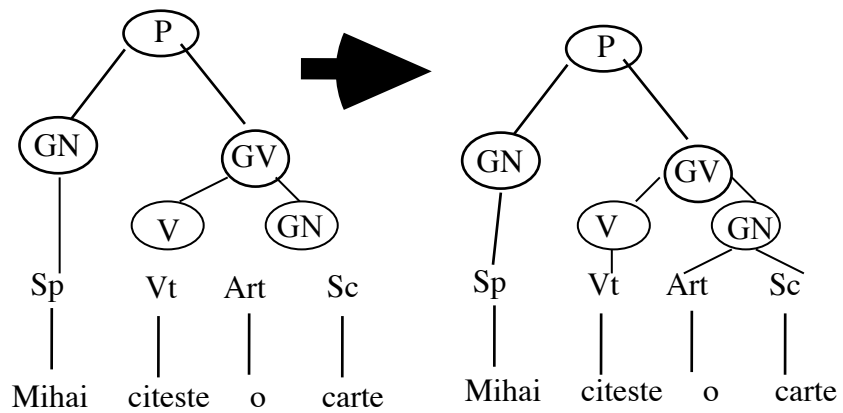
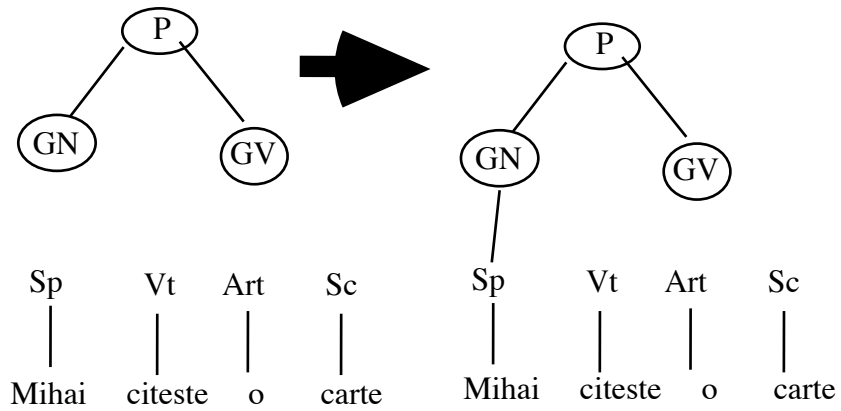
## ***Bottom-up***

- cuvintele se înlocuiesc prin categoriile lor
- acestea se combină în constituienți conform părților drepte ale regulilor
- procesul de combinare se continuă în toate modurile posibile până când, eventual, se reușește obținerea unei singure structuri



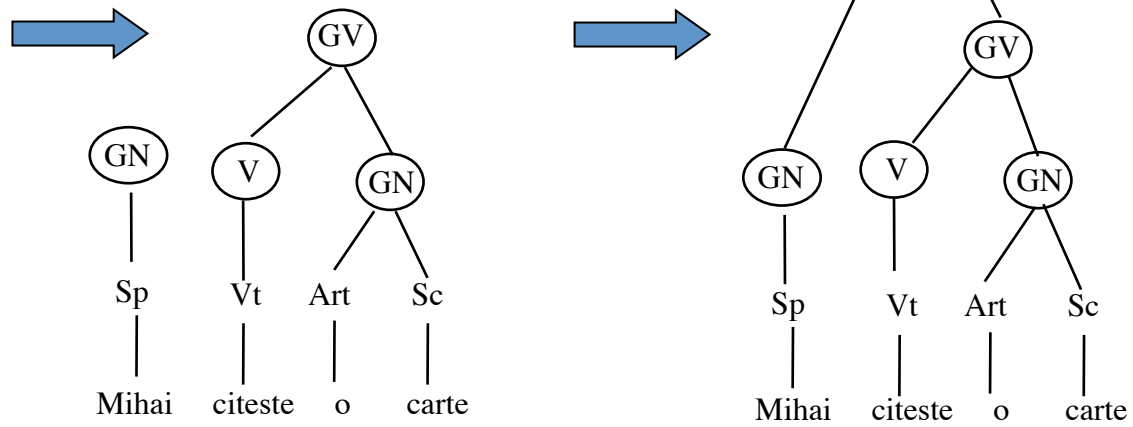
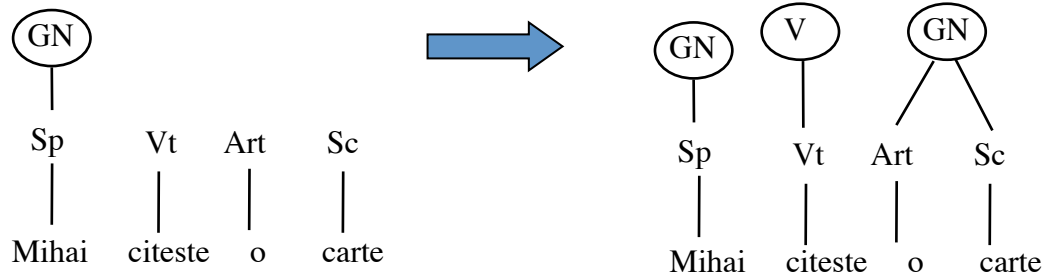
proces direcționat de date  
(*data-driven*)

## TOP-DOWN



**După Dan Tufiş**

## BOTTOM-UP



**După Dan Tufiş**



# *Depth-first* versus *breadth-first* parsing

## *Depth-first*

- notează toate opțiunile atunci când există mai multe alternative
- alege una și urmează-o
- dacă eșuează întoarce-te și alege altă cale
- repetă acest proces până obții analiza (succes) sau până la epuizarea tuturor căilor (eșec)



proces care se pretează la  
investigarea soluțiilor una după alta

## *Breadth-first*

- notează toate opțiunile atunci când există mai multe alternative
- baleiază pe rând toate aceste căi
- abandonează toate căile care eșuează dar continuă cu restul
- repetă acest proces până obții analiza (succes) sau până la epuizarea tuturor căilor (eșec)



proces care se pretează la  
investigarea în paralel a soluțiilor

# Tehnologiile limbajului scris

- Analiza și înțelegerea limbajului
  - Prelucrări semantice și de discurs
    - dezambiguizare semantică ➔ sensurile cuvintelor
    - determinarea rolurilor semantice ale verbelor
    - structura retorică a discursului și dialogului
    - rezoluția anaforelor

## Cuvintele își precizează sensul în context

- *Ion se prinse în horă cu o fată cu cosițe lungi.*
- *Când fată iapa ta?*
- *Mă mai dau o dată pe pârtia roșie.*
- *I-am dat una peste mână.*
- *Maria a dat cartea înapoi.*
- *M-am scos...*
- *Mi-am scos măseaua de minte.*

# Roluri semantice

*Carla bought the computer from Sally for \$100.*



*Sally sold the computer to Carla for \$100.*

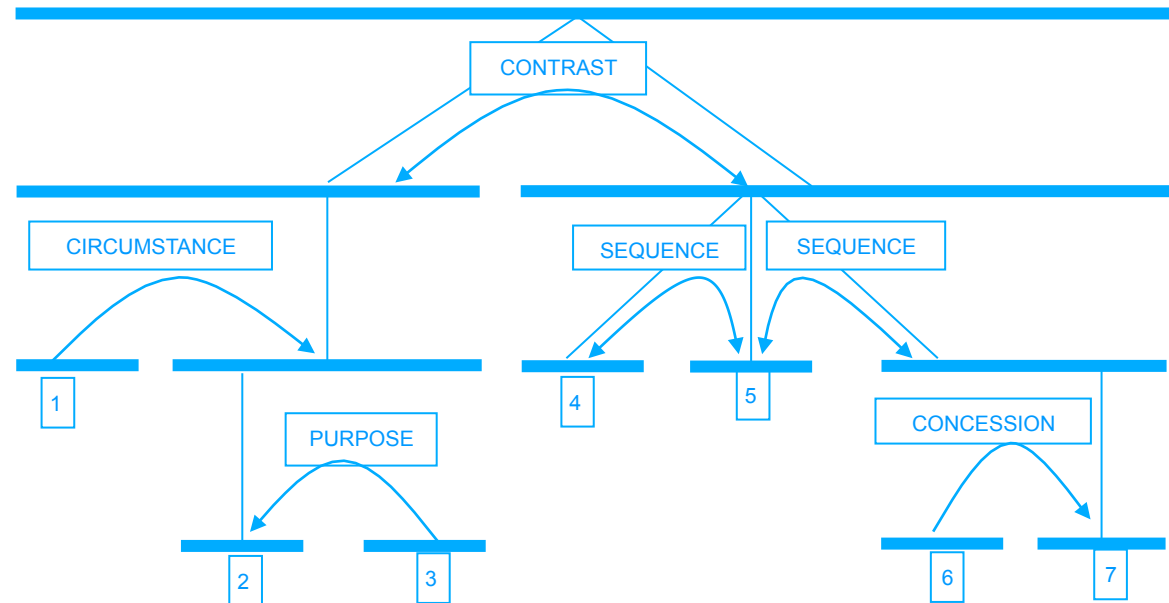
# Structura retorică a discursului

1. *Cînd a ajuns Apolo în muntele Parnas,*
2. *dihania uriașă s-a avîntat spre dînsul,*
3. *dornică să-l ucidă.*
4. *Dar zeul și-a întins arcul.*
5. *A tras prima săgeată.*
6. *Erau doar patru zile de cînd văzuse lumea,*
7. *și întîia lui săgeată a și nimerit monstrul.*

din Legendele Olimpului, de Alexandru Mitru

# Structura retorică a discursului

1. *Cînd a ajuns Apolo în muntele Parnas,*
2. *dihania uriașă s-a avîntat spre dînsul,*
3. *dornică să-l ucidă.*
4. *Dar zeul și-a întins arcul.*
5. *A tras prima săgeată.*
6. *Erau doar patru zile de cînd văzuse lumea,*
7. *și întâia lui săgeată a și nimerit monstrul.*



# Rezoluția anaforei

*If the baby does not thrive on raw milk, boil it.*

Hobbs, 1997

# Tehnologiile limbajului scris

- Aplicații complexe
  - rezumare automată
  - clasificarea documentelor
    - cuvinte și expresii cheie
  - interogări asupra textelor - *question-answering*
    - pe domenii închise/deschise
  - implicații logice – *textual entailment*
  - traducere automată
  - analiza sentimentelor
  - analize temporale



# Rezumare:

## un extras din “Legendele Olimpului”

*Zeul - după aceea - a coborât în lume și a cutreierat prin țări nenumărate, căutându-și loc de templu. Și, tot umblînd pe drumuri, Apolo a ajuns în niște locuri stranii, stîncoase, mohorîte. Pasămite sosise în muntele Parnas. Aici sînt niște vîrfuri de piatră cenușie, ce parcă stau să cadă într-un abis de smoală clocotitor de vuet. Șuvoaiele de aer - un aer cald și umed - țîșnesc neconținut și vorbele rostite se-ntorc iute-n ecouri, sunînd din stîncă-n stîncă.*

*Pădurea-i misterioasă. Frunzișul ei șoptește. Un izvor își prăvale apa sa cristalină spre valea-nnegurată sărind peste pripoare. Și-n munte-i o strîmtoare afundă, mocirloasă, unde traia atunci - precum spunea legenda - un șarpe numit Piton. Trupul acestui monstru era lung cît un munte, cu-o sută de grumazuri și o sută de guri, care azvîrleau flăcări, duhori înecăcioase.*

*Duhorile acestea înțesau tot văzduhul și oameni, fiare, gîze, păsări sau alte ființe, dacă treceau pe acolo, se stingeau otrăvite. Șarpele se născuse dintr-o mocirlă verde. Mocirla rămăsese din apele vărsate acolo, prin potopul orînduit de Zeus.*

*Piton primise-n taină poruncă de la Hera să-l pîndească pe Apolo, cînd va trece prin munte, și să-i răpună viața. Hera-l ura pe fiul cel nou născut al Letei, pentru că soțul său, prea puternicul Zeus, ținea mai mult la dînsul decît la fiii ei: Hefaistos și Ares.*

*Cînd a ajuns Apolo în muntele Parnas, dihania uriașă s-a avîntat spre dînsul, dornică să-l ucidă. Dar zeul și-a întins arcul. A tras prima săgeată. Erau doar patru zile de cînd văzuse lumea, și întia lui săgeată a și nimerit monstrul.*

*E drept că și săgeata fusese făurită de fierarul din Lemnos, Hefaistos, schilodul, cu mare meșteșug. Cum i-a pătruns săgeata în carnea mucedă, șarpele a simțit că i-a venit sfîrșitul. A încercat să scape și să se furișeze într-o pesteră adîncă. Dar n-a mai izbutit...*

*Chinuit de dureri, Piton s-a zvîrcolit prin verzile mocirle în care viețuia. A sărit în nisipul cel galben de pe maluri și a lunecat iară, cu șuiere cumplite, în putreda baltoacă, în vreme ce veninul i se scurgea din boturi.*

*Monstrul s-a zvîrcolit tot mai încet în apă. Ochii i s-au lăsat. S-au cufundat pe încetul, de tot, în întuneric. Așa s-a prăpădit. Iar trupul său lungit pe zeci și zeci de stînji, i s-a uscat cu timpul, sub razele de soare...*

*Și locurile acestea unde-a răpus Apolo pe fiorosul monstru au fost numite Delfi.*

*În Delfi se vor pune temelii unui templu și-unui vestit oracol, a hotărît Apolo. Oamenii vor putea să-ntrebe și să afle răspuns la întrebări, prin preuteasa mea, Pitia, înțeleaptă. Ea se va așeza pe un trepid de aur, învăluit în pielea lui Piton cel răpus, deasupra unei grote adînci și fumegînde, ce duce pînă-n țara lui Hades, mohorîtul... Prin ea voi da răspunsuri. Dar oamenii datori sînt să-mi dăruiască aur și scule prețioase și zeci de bogății, răsplată pentru mine și preoții din templu... Mai poruncesc, de-asemeni, să aibă loc, în cinstea luptei mele cu monstrul, jocuri, serbări, concursuri, și cîntece, și danțuri... Să fie veselie, iar eu să fiu slăvit...*

*După aceea zeul, pătât încă de sînge - de sîngele lui Piton - a pornit să se spele, să-și purifice trupul, în micul rîu Tempe.*

# Rezumare:

## un extras din “Legendele Olimpului”

*Zeul - după aceea - a coborât în lume și a cutreierat prin țări nenumărate, căutându-și loc de templu. Și, tot umblînd pe drumuri, Apolo a ajuns în niște locuri stranii, stîncose, mohorîte. Pasămite sosise în muntele Parnas. Aici sînt niște vîrfuri de piatră cenușie, ce parcă stau să cadă într-un abis de smoală clocotitor de vuet. Șuvoaiele de aer - un aer cald și umed - fîșnesc neconținut și vorbele rostite se-ntorc iute-n ecouri, sunînd din stîncă-n stîncă.*

*Pădurea-i misterioasă. Frunzișul ei șoptește. Un izvor își prăvale apa sa cristalină spre valea-nnegurată sărind peste pripoare. Și-n munte-i o strîmtoare afundă, mocirloasă, unde traia atunci - precum spunea legenda - un șarpe numit Piton. Trupul acestui monstru era lung cît un munte, cu-o sută de grumazuri și o sută de guri, care azvîrleau flăcări, duhori înecăcioase.*

*Duhorile acestea înțesau tot văzduhul și oameni, fiare, gîze, păsări sau alte ființe, dacă treceau pe acolo, se stingeau otrăvite. Șarpele se născuse dintr-o mocirlă verde. Mocirla rămăsese din apele vărsate acolo, prin potopul orînduit de Zeus.*

*Piton promise-n taină poruncă de la Hera să-l pîndească pe Apolo, cînd va trece prin munte, și să-i răpună viața. Hera-l ura pe fiul cel nou născut al Letei, pentru că soțul său, prea puternicul Zeus, ținea mai mult la dînsul decît la fiii ei: [Hefaistos](#) și Ares.*

*Cînd a ajuns Apolo în muntele Parnas, dihania uriașă s-a avîntat spre dînsul, dornică să-l ucidă. Dar zeul și-a întins arcul. A tras prima săgeată. Erau doar patru zile de cînd văzuse lumea, și întia lui săgeată a și nimerit monstrul.*

*E drept că și săgeata fusese făurită de fierarul din Lemnos, Hefaistos, schilodul, cu mare meșteșug. Cum i-a pătruns săgeata în carnea mucedă, șarpele a simțit că i-a venit sfîrșitul. A încercat să scape și să se furișeze într-o pesteră adîncă. Dar n-a mai izbutit...*

*Chinuit de dureri, Piton s-a zvîrcolit prin verzile mocirle în care viețuia. A sărit în nisipul cel galben de pe maluri și a lunecat iară, cu șuiere cumplite, în putreda baltoacă, în vreme ce veninul i se scurgea din boturi.*

*Monstrul s-a zvîrcolit tot mai încet în apă. Ochii i s-au lăsat. S-au cufundat pe încetul, de tot, în întuneric. Așa s-a prăpădit. Iar trupul său lungit pe zeci și zeci de stînji, i s-a uscat cu timpul, sub razele de soare...*

*Și locurile acestea unde-a răpus Apolo pe fiorosul monstru au fost numite Delfi.*

*În Delfi se vor pune temelii unui templu și-unui vestit oracol, a hotărît Apolo. Oamenii vor putea să-ntrebe și să afle răspuns la întrebări, prin preuteasa mea, Pitia, înțeleaptă. Ea se va așeza pe un trepid de aur, învăluit în pielea lui Piton cel răpus, deasupra unei grote adînci și fumegînde, ce duce pînă-n țara lui Hades, mohorîtul... Prin ea voi da răspunsuri. Dar oamenii datori sînt să-mi dăruiască aur și scule prețioase și zeci de bogății, răsplată pentru mine și preoții din templu... Mai poruncesc, de-asemeni, să aibă loc, în cinstea luptei mele cu monstrul, jocuri, serbări, concursuri, și cîntece, și danțuri... Să fie veselie, iar eu să fiu slăvit...*

*După aceea zeul, pătât încă de sînge - de sîngele lui Piton - a pornit să se spele, să-și purifice trupul, în micul rîu Tempe.*

# Rezumare:

## un extras din “Legendele Olimpului”

*Zeul - după aceea - a coborât în lume și a cutreierat prin țări nenumărate, căutându-și loc de templu. Și, tot umblînd pe drumuri, Apolo a ajuns în niște locuri stranii, stîncoase, mohorîte. Pasămite sosise în muntele Parnas. Aici sînt niște vîrfuri de piatră cenușie, ce parcă stau să cadă într-un abis de smoală clocotitor de vuet. Șuvoaiele de aer - un aer cald și umed - țîșnesc neconținut și vorbele rostite se-ntorc iute-n ecouri, sunînd din stîncă-n stîncă.*

*Pădurea-i misterioasă. Frunzișul ei șoptește. Un izvor își prăvale apa sa cristalină spre valea-nnegurată sărind peste pripoare. Și-n munte-i o strîmtoare afundă, mocirloasă, unde traia atunci - precum spunea legenda - un șarpe numit Piton. Trupul acestui monstru era lung cît un munte, cu-o sută de grumazuri și o sută de guri, care azvîrleau flăcări, duhori înecăcioase.*

*Duhorile acestea înțesau tot văzduhul și oameni, fiare, gîze, păsări sau alte ființe, dacă treceau pe acolo, se stingeau otrăvite. Șarpele se născuse dintr-o mocirlă verde. Mocirla rămăsese din apele vărsate acolo, prin potopul orînduit de Zeus.*

*Piton promise-n taină poruncă de la Hera să-l pîndească pe Apolo, cînd va trece prin munte, și să-i răpună viața. Hera-l ura pe fiul cel nou născut al Letei, pentru că soțul său, prea puternicul Zeus, ținea mai mult la dînsul decît la fiii ei: Hefaistos și Ares.*

*Cînd a ajuns Apolo în muntele Parnas, dihania uriașă s-a avîntat spre dînsul, dornică să-l ucidă. Dar zeul și-a întins arcul. A tras prima săgeată. Erau doar patru zile de cînd văzuse lumea, și întia lui săgeată a și nimerit monstrul.*

*E drept că și săgeata fusese făurită de fierarul din Lemnos, [Hefaistos](#), [schilodul](#), cu mare meșteșug. Cum i-a pătruns săgeata în carnea mucedă, șarpele a simțit că i-a venit sfîrșitul. A încercat să scape și să se furișeze într-o pesteră adîncă. Dar n-a mai izbutit...*

*Chinuit de dureri, Piton s-a zvîrcolit prin verzile mocirle în care viețuia. A sărit în nisipul cel galben de pe maluri și a lunecat iară, cu șuiere cumplite, în putreda baltoacă, în vreme ce veninul i se scurgea din boturi.*

*Monstrul s-a zvîrcolit tot mai încet în apă. Ochii i s-au lăsat. S-au cufundat pe încetul, de tot, în întuneric. Așa s-a prăpădit. Iar trupul său lungit pe zeci și zeci de stînjeni, i s-a uscat cu timpul, sub razele de soare...*

*Și locurile acestea unde-a răpus Apolo pe fiorosul monstru au fost numite Delfi.*

*În Delfi se vor pune temelii unui templu și-unui vestit oracol, a hotărît Apolo. Oamenii vor putea să-ntrebe și să afle răspuns la întrebări, prin preuteasa mea, Pitia, înțeleapta. Ea se va așeza pe un trepied de aur, învăluit în pielea lui Piton cel răpus, deasupra unei grote adînci și fumegînde, ce duce pînă-n țara lui Hades, mohorîtul... Prin ea voi da răspunsuri. Dar oamenii datori sînt să-mi dăruiască aur și scule prețioase și zeci de bogății, răsplată pentru mine și preoții din templu... Mai poruncesc, de-asemeni, să aibă loc, în cinstea luptei mele cu monstrul, jocuri, serbări, concursuri, și cîntece, și danțuri... Să fie veselie, iar eu să fiu slăvit...*

*După aceea zeul, pătât încă de sînge - de sîngele lui Piton - a pornit să se spele, să-și purifice trupul, în micul rîu Tempe.*

# Un rezumat obișnuit

*Zeul - după aceea - a coborât în lume și a cutreierat prin țări nenumărate, căutându-și loc de templu. Și, tot umblînd pe drumuri, Apolo a ajuns în niște locuri stranii, stîncoase, mohorîte. Pasămite sosise în muntele Parnas*

*Și-n munte-i o strîmtoare afundă, mocirloasă, unde traia atuncea - precum spunea legenda - un șarpe numit Piton.*

*Piton primise-n taină poruncă de la Hera să-l pîndească pe Apolo, cînd va trece prin munte, și să-i răpună viața. Cînd a ajuns Apolo în muntele Parnas, dihania uriașă s-a avîntat spre dînsul, dornică să-l ucidă. Dar zeul și-a întins arcul. A tras prima săgeată. Erau doar patru zile de cînd văzuse lumea, și întîia lui săgeată a și nimerit monstrul.*

*Monstrul s-a zvîrcolit tot mai încet în apă... Așa s-a prăpădit.*

*După aceea zeul, pătat încă de sînge - de sîngele lui Piton - a pornit să se spele, să-și purifice trupul, în micul rîu Tempe.*

# Un rezumat focalizat pe prima apariție a lui Hefaistos

*Zeul - după aceea - a coborât în lume și a cutreierat prin țări nenumărate. Apolo a ajuns în niște locuri stranii, stîncoase, mohorîte. Pasămite sosise în muntele Parnas.*

*Și-n munte-i o strîmtoare afundă, mocirloasă, unde traia atuncea un șarpe numit Piton.*

*Piton primise-n taină poruncă de la Hera să-l pîndească pe Apolo, cînd va trece prin munte, și să-i răpună viața. Hera-l ura pe fiul cel nou născut al Letei, pentru că soțul său, prea puternicul Zeus, ținea mai mult la dînsul decît la fiii ei: **Hefaistos** și Ares...*

*Cînd a ajuns Apolo în muntele Parnas, dihania uriașă s-a avîntat spre dînsul, dornică să-l ucidă. Dar zeul și-a întins arcul. A tras prima săgeată. Erau doar patru zile de cînd văzuse lumea, și întîia lui săgeată a și nimerit monstrul.*

*Monstrul s-a zvîrcolit tot mai încet în apă... Așa s-a prăpădit.*

*După aceea zeul, pătat încă de sînge - de sîngele lui Piton - a pornit să se spele, să-și purifice trupul, în micul rîu Tempe.*

# Un rezumat focalizat pe a doua apariție a lui Hefaistos

*Zeul - după aceea - a coborât în lume și a cutreierat prin țări nenumărate. Apolo a ajuns în niște locuri stranii, stîncoase, mohorîte. Pasămite sosise în muntele Parnas.*

*Și-n munte-i o strîmtoare afundă, mocirloasă, unde traia atuncea un șarpe numit Piton.*

*Piton primise-n taină poruncă de la Hera să-l pîndească pe Apolo, cînd va trece prin munte, și să-i răpună viața.*

*... dihania uriasă s-a avîntat spre dînsul, dornică să-l ucidă. Dar zeul și-a întins arcul. A tras prima săgeată. ... și întâia lui săgeată a și nimerit monstrul.*

*E drept că și săgeata fusese făurită de fierarul din Lemnos, [Hefaistos](#), [schilodul](#), cu mare meșteșug...*

*Monstrul s-a zvîrcolit tot mai încet în apă... Așa s-a prăpădit.*

*După aceea zeul, pătat încă de sînge - de sîngele lui Piton - a pornit să se spele, să-și purifice trupul, în micul rîu Tempe.*

# Limbaajul în corelație cu alte modalități

- Reprezentarea timpului și a spațiului
- Text și imagine
- Vorbire și gest
- Mișcările faciale și recunoașterea vorbirii
- Mișcările faciale și sinteza vorbirii
- Semantic Web

# Question Answering - Introduction

- Question Answering (QA) can be defined as *the task which takes a question in natural language and produces one or more ranked answers from a collection of documents*
- The QA research area has emerged as a result of a monolingual English QA track being introduced at **TREC** (Text Retrieval and Evaluation Conference: <http://trec.nist.gov/>)



# Question Answering - Modules

- QA systems normally adhere to the pipeline architecture composed of three main modules (Harabagiu and Moldovan, 2003):
  - **question analysis** – *the results are keywords, answer and question type, focus*
  - **paragraph retrieval** - *the results are a set of relevant candidate paragraphs/sentences from the document collection*
  - **answer extraction** – *the results are a set of candidate answers ranked using likelihood measures*

# QA – Question Type

- Harabagiu and Moldovan, 2003:
  - **Factoid** – *“Who discovered the oxygen?”, “When did Hawaii become a state?” or “What football team won the World Cup in 1992?”*
  - **List** – *“What countries export oil?” or “What are the regions preferred by the Americans for holidays?”.*
  - **Definition** – *“What is a quasar?” or “What is a question-answering system?”*
- *How, Why, hypothetical, semantically constrained, polar (Yes/No) and cross-lingual questions*

# QA – Answer Type

- **Person** - "What", "Who", "Whom", "With who"
- **Location** (City, Country, and Region) - "What state/city", "From where", "Where"
- **Organization** - "Who produced", "Who made"
- **Temporal** (Date and Year) – "When"
- **Measure** (Length, Surface and Other) – "How many/much"
- **Count** - "How many/much"
- **Yes/No** – "Did you fear that?", "Are you blue?"

# QA – Search collection

- Local collections, internal organization documents, newspapers, Internet
- **Closed-domain** - deals with questions from a specific domain (medical, baseball, etc.). Can exploit domain-specific knowledge (ontologies, rules, disambiguation)
- **Open-domain** – general question about anything. Can use general knowledge about the world, wikipediae, etc.

# Examples of QA systems

- **BASEBALL** (Green, 1963) - answer questions about baseball games
- **LUNAR** (Woods, 1977) – geological analysis of rocks returned by the Apollo moon missions
- **IURES** (Cristea, Tufiş, Mihaiescu, 1985) – ICI repository of programs, medicine, geography

# QA - Systems of today

- Powerset: <http://www.powerset.com/> (<http://www.bing.com/>)
- Assimov the chat bot: <http://talkingrobot.org/b/>
- AnswerBus: <http://www.answerbus.com/index.shtml>
- NSIR: <http://tangra.si.umich.edu/clair/NSIR/html/nsir.cgi>
- START (The first question answering system):  
<http://start.csail.mit.edu/>



Asimov

#### OUR CHATBOTS

[Asimov the robot](#)

[Data the android](#)

[Trinity](#)

[More chatbots](#)

#### ABOUT CHATBOTS

[Home.](#)

We create chatbots.

[The Present.](#)

About todays chatbots.

[The Future.](#)

The future of chatbots.

[Get a Chatbot.](#)

Get a chatbot for your website.

#### ABOUT ASIMOV

[Help on Asimov](#)

[Demo transcript](#)

[Asimov FAQ](#)

Hello Adrian.  
My name is Asimov.

> **Hello Asimov. How are you?**

Hello Adrian.

I am fine, thank you.

I am always the same. I don't have feelings. I have no feelings or moods.

NSIR - Question Answering System - Mozilla Firefox

File Edit View History Bookmarks Tools Help

NSIR - Question Answering Syst... +

clair.si.umich.edu/clair/NSIR/html/nsir.cgi?question=Who+is+Tom+Cruise+married+to%3F&search\_engine=Google\_snippet&n\_documents=50i ☆ Google

Most Visited Getting Started Latest Headlines LITORAL 2010 ROMANI... Adevarul

# NSIR

## Question Answering System

[:: Home](#) [:: Help](#) [:: About NSIR](#) [:: Contact Us](#) ::

Sample Questions

1. [What year did Wilt Chamberlain score 100 points?](#)
2. [How many chromosomes does a human zygote have?](#)
3. [Who is Tom Cruise married to?](#)
4. [What is South Africa's total population?](#)
5. [What is Canada's most populous city?](#)
6. [Which country exports the most tea?](#)
7. [What state has the most Indians?](#)
8. [George Bush purchased a small interest in which baseball team?](#)
9. [What county is Modesto, California in?](#)
10. [Who is the author of the book, 'The Iron Lady: A Biography of Margaret Thatcher'?](#)

Type your question below

or choose from [TREC-8](#) [TREC-9](#) [TREC-2001](#) [TREC-2002](#)

Who is Tom Cruise married to?

Search Engine / Index: [help](#)

No. of Hits: [help](#)

Answer Length: [help](#)

No. of Answers: [help](#)

Answer [help](#) (for eval; use ";" to delimit multiple answers)

Cached results: [help](#)

Debug information: [help](#)

Google-snippet (faster but less accurate) ▼

50 ▼

Phrases ▼

10 ▼

Nicole Kidman optional

☒ Check to use cache if there is cached result

☐ Check to see the detailed scores

Get Answer

This is a PERSON type question

Ranked Answers (click to see context)

Top Web Documents Retrieved By Google

1)

[\(What are these numbers?\)](#) FARR 1 FHS 1 TRDR 1 TTRDR 1 PREC 1 MRR 1 TRWR 1 NTRWR 1



# AnswerBus

Ask

Type in your question! We only want to offer the best possible answers to you.

## Possible Answers

**Mas d'Augustine** Luxury Bed & Breakfast In The South Of France [www.masdaugustine.com](http://www.masdaugustine.com)

**Tichy Racing Camps** The best all inclusive summer ski race camps in North America [www.tichyracing.com](http://www.tichyracing.com)

**Teen Climbing Camp** Ages 13 - 18, Intro to Advanced AMGA and PCIA Accredited [climbacadia.org/](http://climbacadia.org/)



AdChoices

**Bread Recipes** Free e-Cook Book Explains Secrets to Making Bread - View Now [TotalRecipeSearch.com](http://TotalRecipeSearch.com)

**Horizon Summer Camps** Learn English in Unique Interaction Residential Camps w/ Irish Children [www.horizonireland.com](http://www.horizonireland.com)

**Portable Living Units** Portable Rig Camps, Residential, Commercial (Full Catering avail.) [maps.google.ca](http://maps.google.ca)



AdChoices

### [How do i make french toast with only Egg And bit of milk and bread.?](#)

...my first time **making french toast** can someone maybe put it step by step on **how to make** it and also **I** don't Have Vanilla something... is ( Sugar, **Milk**, **Bread**, and **eggs**, and butter) that's it so **how do i make** it like that Please help...

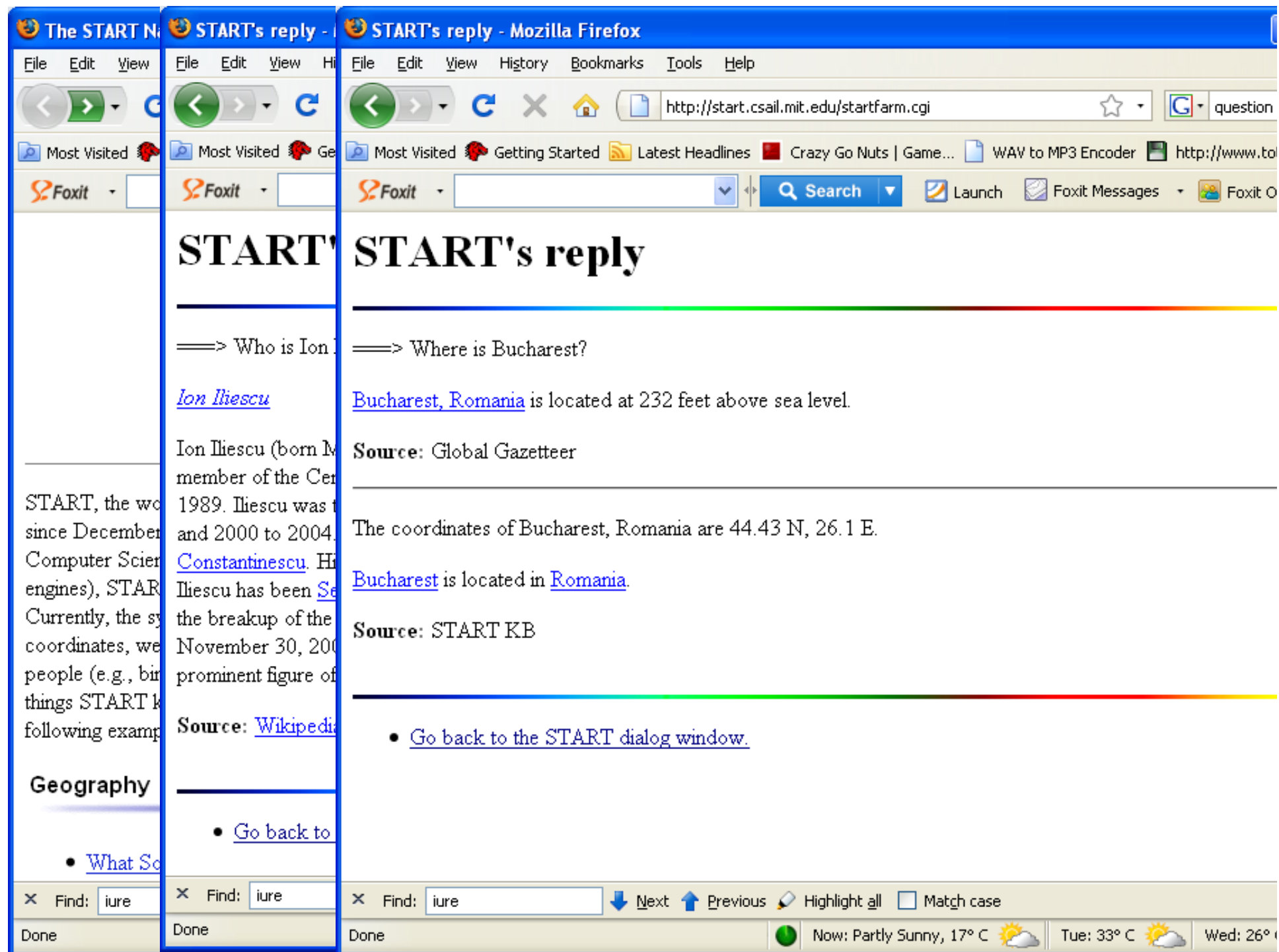
### [How many calories are on a piece of French toast?](#)

**I'm doing** a school project to see **how many calories I** eat a day and **how much**...this is the **only thing I** cant find!? Its normal white **bread** dipped in **eggs** and a skim **milk** and cooked on a pan **with a bit of** butter.. my mom **makes** it so **I** don't know...

### [Did I give a good answer to this question??](#)

...with lots of butter **Milk** is really good but **make** sure its 2% or less Lunch...**make** sure again its on wheat **bread** and try to add other ...terrible, cut them out of your diet or at least don...have if its good Cereal and **toast** is good like **I** said above if its...and you don't load it with lots of butter What... you getting? **Eggs** are good 2 but **how** are you cooking...up at 5:45 to exercise, **do** you have no time to exercise...a week Pancakes, **French Toast**, and Waffles...have to have them **only** have a little **bit** Dinner Subway is good...

## More Resources from Wiki and Web



# QA - Competitions

- **CLEF** (Cross Language Evaluation Forum) started in 2000 - <http://www.clef-campaign.org/>  
European languages in both monolingual and cross-language contexts
  - Coordination: Istituto di Scienza e Tecnologie dell'Informazione, Pisa, Italy
- **TREC** (Text REtrieval Conference) - started in 1992 <http://trec.nist.gov/>
  - National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, USA

-<test-set>

-<topic t\_id="1" t\_name="AIDS">

-<reading-test r\_id="1">

-<doc d\_id="1">

Annie Lennox: De ce sunt activist împotriva HIV/SIDA Vă voi spune povestea transformării mele într-un activist împotriva HIV/SIDA. Iar acesta este numele campaniei mele, Campania SING [Cântă]. În noie fundația sa împotriva HIV/SIDA. Iar 46664 este numărul pe care îl avea Mandela în închisoarea din Robben Island. Aici sunt cu Youssou N'Dour, pe scenă, distrându-mă de minune. A doua zi, toți artiștii au f fața fostei sale celule de închisoare. Puteți vedea gratiile de la fereastră aici. A fost un eveniment plin de semnificații pentru noi toți. În acel moment Mandela a spus presei mondiale că are loc aproape un genoc mai vulnerabile victime, erau femeile și copiii. Asta a avut un impact uriaș asupra minții mele, deoarece sunt femeie și sunt mamă și nu mi-am dat seama că pandemia HIV/SIDA afecta direct femeile în acest fe ceva despre care trebuie să vorbesc. Trebuie să servesc această cauză." Și astfel, în consecință, am participat la fiecare eveniment al fundației 46664 la care am putut lua parte și am organizat conferințe de pr respect pentru extraordinara, incredibilă muncă pe care a făcut-o. Toată lumea îl respectă pe Nelson Mandela. Toată lumea îl idolatrizează pe Nelson Mandela. Dar știu ei despre ce se întâmplă în Africa de Sud și aș spune oamenilor ce se întâmplă acolo, ar fi șocați. Am fost foarte, foarte norocoasă, câțiva ani mai târziu, să-l întâlnesc pe Zackie Achmat, fondatorul Treatment Action Campaign, un activist incredibil. I Vă spune că sunt solidară cu oamenii care au HIV, oamenii care trăiesc cu HIV. Și într-un fel, din cauza stigmatului, prin purtarea acestui tricou, eu spun "Da, putem discuta despre această problemă." "Nu tre acestei organizații incredibile. Este o campanie ce se bazează pe comunitate cu 80% dintre membri fiind femei, majoritatea fiind infectate cu HIV. Lucrează pe teren. Au un impact extraordinar asupra oamenilo extraordinar ceea ce fac. Și da, campania mea SING a sprijinit Treatment Action Campaign în sensul că am încercat să conștientizez oamenii și am încercat să strâng fonduri. O mulțime din banii pe care am i încă continuă să o facă, în Africa de Sud. Așadar aceasta este campania mea SING. Campania SING este de fapt formată din mine și încă trei sau patru oameni minunați care mă ajută și mă susțin. Am călător Norvegia, primind un cec consistent, cântând la Hong Kong, încercând să-i conving pe oameni să doneze bani. În Johannesburg am avut ocazia să cânt în fața unei audiențe majoritar albe din clasa mijlocie ca oamenii au tendința să o evite, pentru că sunt plictisiți, și nu știu de fapt care sunt soluțiile. Aaron Motsoaledi, ministrul sănătății aflat în funcție, a participat la acel concert, și am avut ocazia să-l întâlnesc, iar devenit ulterior o reprezentantă a Scoției și a HIV. Și le-am prezentat experiențele mele în încercarea de a-i face conștienți de ce se întâmplă. Și din nou, în Edinburgh, împreună cu minunatul cor al copiilor afr afectată de virusul SIDA. Aici sunt în New York cu Michel Sidibe. El e directorul UNAIDS. Sunt foarte onorată de faptul că Michel m-a invitat doar cu câteva luni în urmă să devin ambasador UNAIDS. Și înt lume este că am vrea să vedem eliminarea propriu-zisă a transmiterii virusului de la mamă la copil până în anul 2015. Este un obiectiv foarte ambițios, dar credem că poate fi atins cu voință politică. Este posib suntem foarte încrezătoare, pentru că știm că această tânăra este tratată astfel încât să i se prelungească viața pentru a avea grijă de bebelușul pe care îl va naște. Iar bebelușul său va fi tratat cu PMTCT, ceea modalitățile de intervenție pentru stoparea pandemiei SIDA. Aș dori să închei spunându-vă povestea despre Avelile. Ea este Avelile. Merge cu mine oriunde mă duc. Spun povestea ei tuturor, pentru că ea reprez complementare SIDA. Avelile avea virusul. S-a născut cu virusul. Și aici are 7 ani, cântărind nu mai mult decât un bebeluș de 1 an. În acest moment al vieții sale, ea suferă de SIDA în plină dezvoltare și are pr Doctorii și asistentele erau fenomenali. Au trecut-o pe un regim nutritiv foarte special și au avut mare grijă de ea. Și atunci când am plecat de la spital nu știam -- pentru că am filmat atunci povestea ei -- nu ști foarte marcați de această experiență directă, de acest copil, știți, acea poveste. Cinci luni mai târziu ne-am întors în Africa de Sud pentru a o reîntâlni pe Avelile. Și devin -- firele de păr de pe -- nu știu dacă p care a avut loc. Nu este extraordinar? (Aplauze) Aceste aplauze sunt de fapt pentru doctorii și asistentele din spitalul în care au avut grijă de Avelile. Și înțeleg că apreciați acest tip de transformare. Așadar, aș d la hrană bună și asistență medicală de calitate și dacă credeți că Obiectivele de Dezvoltare ale Mileniului, în mod special al cincelea și al șaselea, ar trebui urmărite absolut de toate guvernele din toată lumea -- n din sală. Vă mulțumesc mult. (Aplauze)

</doc>

-<q q\_id="1">

-<q\_str>

Ce eveniment a determinat-o pe Annie Lennox să se angajeze în lupta contra SIDA?

</q\_str>

<answer a\_id="1">detenția lui Nelson Mandela la Robben Island</answer>

-<answer a\_id="2">

ceremonia de închidere a fundației lui Nelson Mandela

</answer>

<answer a\_id="3">întâlnirea cu Youssou N'Dour</answer>

<answer a\_id="4">segregarea rasială din Africa de Sud</answer>

-<answer a\_id="5">

conferința lui Nelson Mandela pentru presa mondială

</answer>

</q>

-<q q\_id="2">

<q\_str>Cine este fondatorul campaniei SING?</q\_str>

<answer a\_id="1">Nelson Mandela</answer>

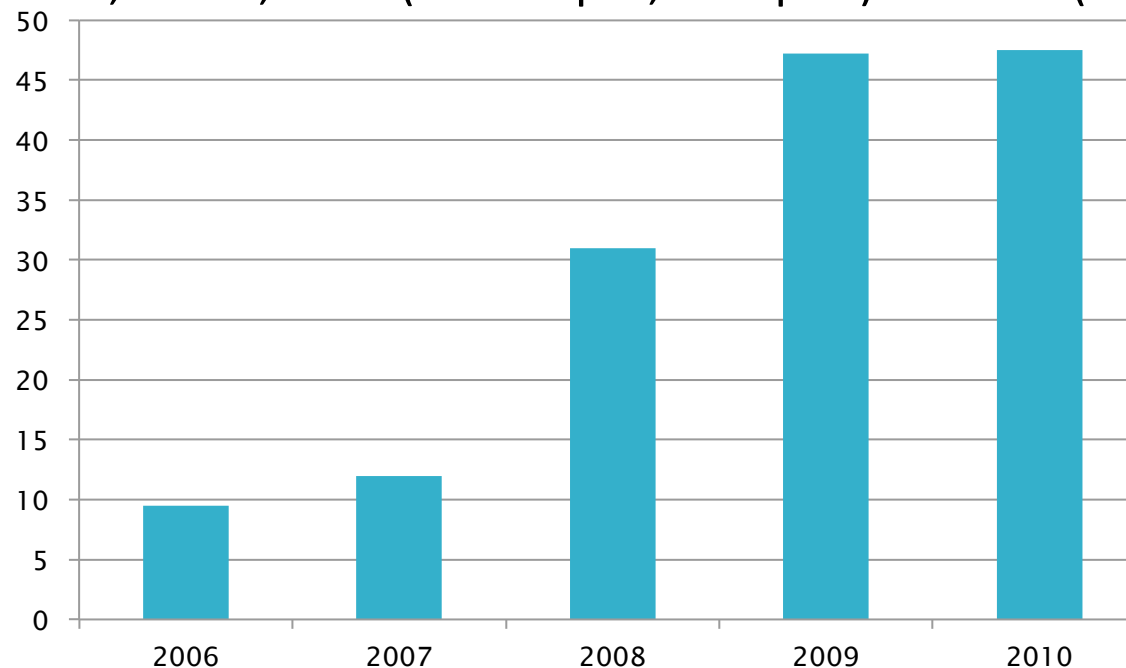
<answer a\_id="2">Youssou N'Dour</answer>

<answer a\_id="3">Michel Sidibe</answer>

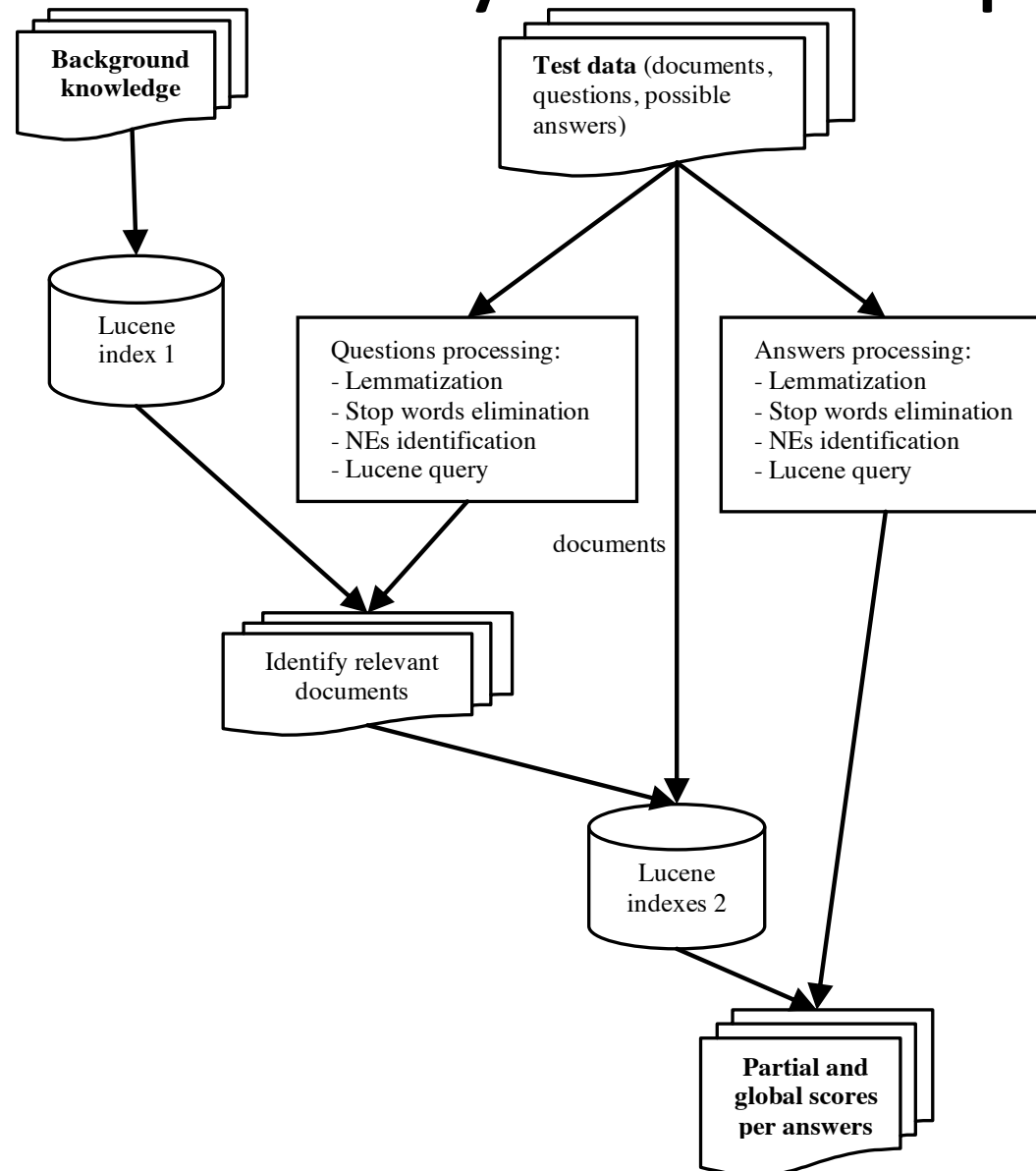
<answer a\_id="4">Zackie Achmat</answer>

# UAIC System – CLEF 2011

- Our group participated at CLEF exercises since 2006:
  - 2006 – Ro–En (English collection) – 9.47% right answers
  - 2007 – Ro–Ro (Romanian Wikipedia) – 12 %
  - 2008 – Ro–Ro (Romanian Wikipedia) – 31 %
  - 2009 – Ro–Ro, En–En (JRC-Acquis) – 47.2 % (48.6%)
  - 2010 – Ro–Ro, En–En, Fr–Fr (JRC-Acquis, Europarl) – 47.5% (42.5%, 27 %)

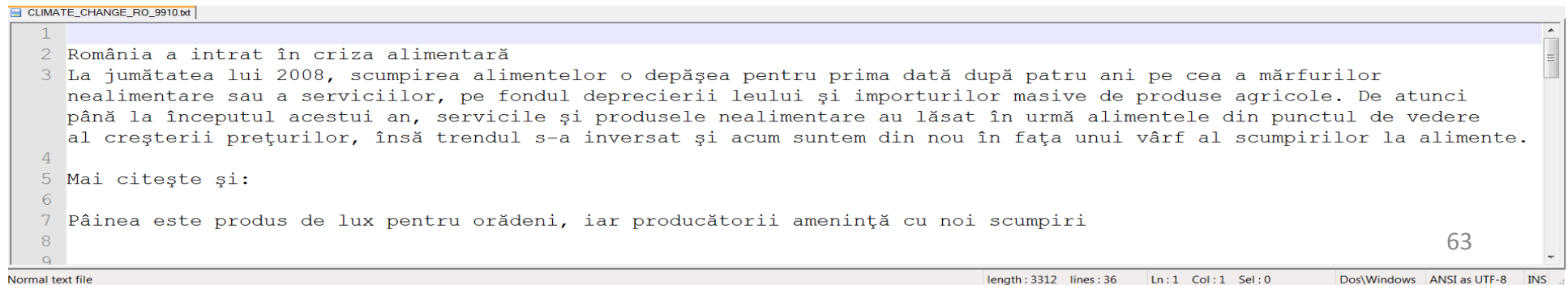


# The UAIC system components



# Background knowledge indexing

- The Romanian background knowledge has 161,279 documents in text format
  - 25,033 correspond to the AIDS topic
  - 51,130 to Climate Change topic
  - 85,116 to Music and Society topic
- The indexing component considers the name of the file and the text from it => **Lucene index 1**



```
CLIMATE_CHANGE_RO_9910.txt
1
2 România a intrat în criza alimentară
3 La jumătatea lui 2008, scumpirea alimentelor o depășea pentru prima dată după patru ani pe cea a mărfurilor
  nealimentare sau a serviciilor, pe fondul deprecierei leului și importurilor masive de produse agricole. De atunci
  până la începutul acestui an, serviciile și produsele nealimentare au lăsat în urmă alimentele din punctul de vedere
  al creșterii prețurilor, însă trendul s-a inversat și acum suntem din nou în fața unui vârf al scumpirilor la alimente.
4
5 Mai citește și:
6
7 Pâinea este produs de lux pentru orădeni, iar producătorii amenință cu noi scumpiri
8
9
```

63

Normal text file length: 3312 lines: 36 Ln: 1 Col: 1 Sel: 0 Dos\Windows ANSI as UTF-8 INS

# Test data processing

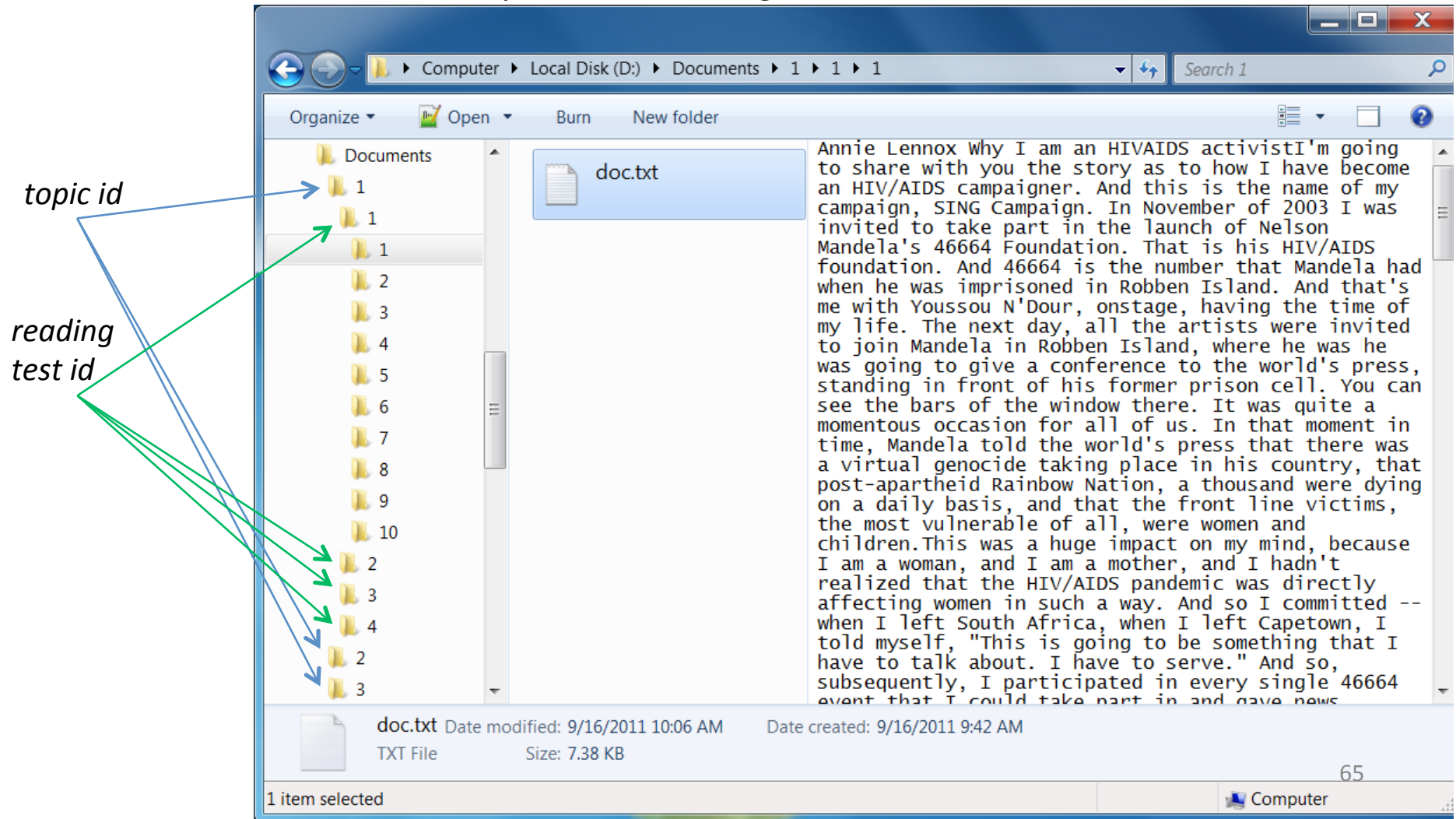
- Test data was an XML file with 12 test documents
  - 4 documents for each of the three topics (12 in total)
  - 10 questions for each document (120 in total)
  - 5 possible answers for each question (600 in total)
- *Test data processing* involved 3 operations:
  - extracting documents
  - processing questions
  - processing possible answers

```
en_test.xml
2 <test-set>
3   <topic t_id="1" t_name="AIDS">
4     <reading-test r_id="1">
5       <doc d_id="1">
6         Annie Lennox Why I am an HIV/AIDS activist ...
7       </doc>
8       <q q_id="1">
9         <q_str>What event caused Annie Lennox to commit herself to the fight against AIDS?</q_str>
10        <answer a_id="1">the imprisonment of Nelson Mandela at Robben Island</answer>
11        <answer a_id="2">the closing ceremony of Nelson Mandela's Foundation</answer>
12        <answer a_id="3">the meeting with Youssou N'Dour</answer>
13        <answer a_id="4">the racial segregation in South Africa</answer>
14        <answer a_id="5">Nelson Mandela's conference to the world press</answer>
15      </q>
16      <q q_id="2">
17        <q_str>Who is the founder of the SING campaign?</q_str>
```



# Test data processing – Extracting documents

- The content of `<doc>` => `<topic id>\<reading test id>\1..10`



# Test data processing – Processing questions

- Stop words elimination
- Lemmatization
- Named Entity identification
- Lucene query building

```
<topic t_id="1" t_name="AIDS" r_id="2" q_id="3">  
<string>Which African country did Bono Vox visit?</string>  
<focus>country</focus>  
<verb>visit </verb>  
<noun></noun>  
<adj></adj>  
<nameEntities>African Bono Vox</nameEntities>  
<query>African^3 country Bono^3 Vox^3 visit</query>  
<questionType>FACTOID</questionType>  
<answerType>OTHER</answerType>
```

# Test data processing – Processing possible answers

- Similar to processing questions +
- We **use ontology** (Iftene and Balahur, 2008) for elimination of possible answers with low probability to be final answer (relation [is\_located\_in])
- *In which European cities has Annie Lennox performed?*
- We eliminate from the list of possible answers the answers with non-European cities (we replace non-European cities with the value XXXXX)

---

```
<answer a_id='1'>London^3 Paris^3</answer>
<answer a_id='2'>XXXXX^3</answer>
<answer a_id='3'>XXXXX^3</answer>
<answer a_id='4'>Edinburgh^3 Oslo^3</answer>
<answer a_id='5'>XXXXX^3</answer>
```

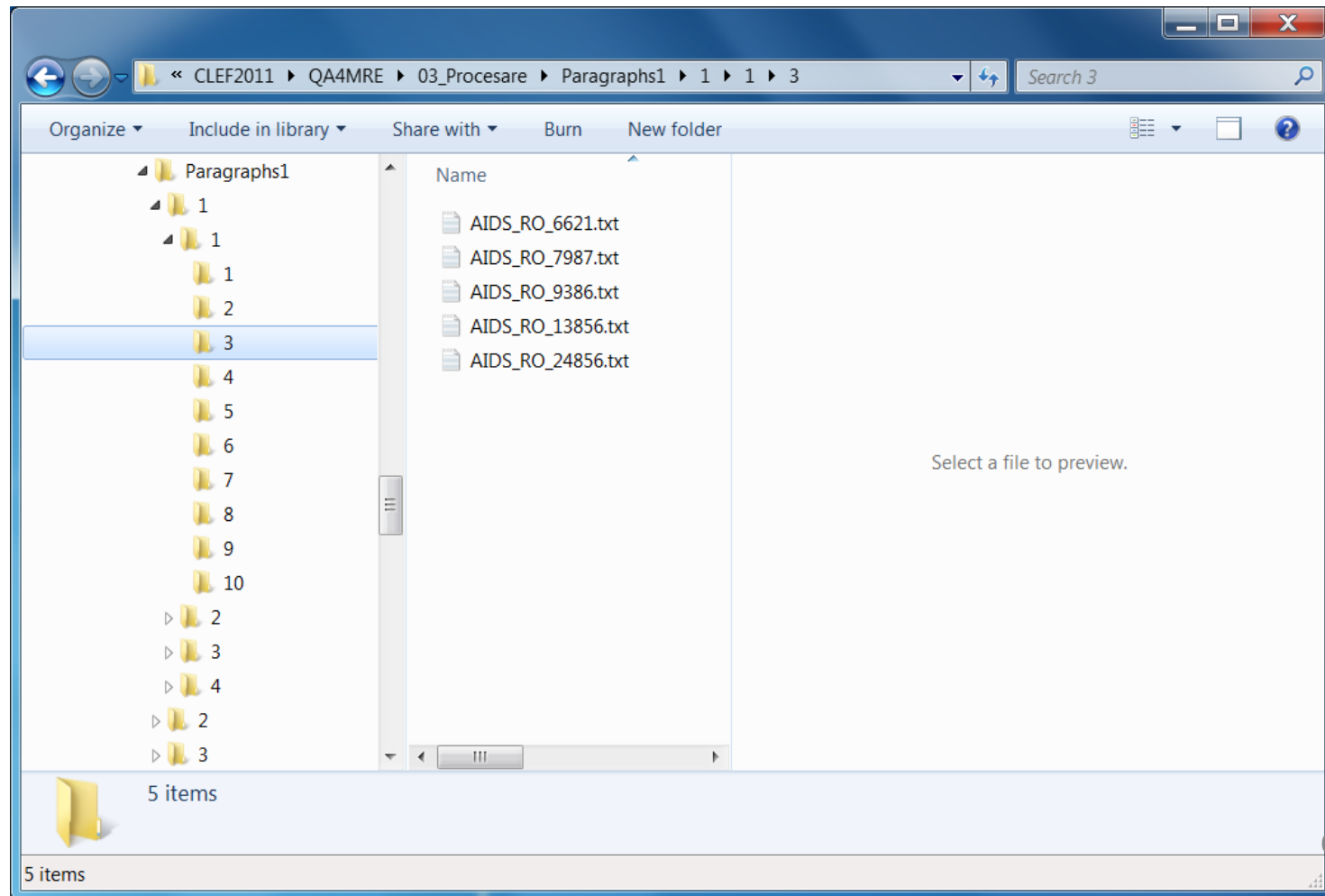
# Information Retrieval on Background Knowledge

- We used Lucene to retrieve, for every question, the relevant documents from the background knowledge
- The result of this step is a list of documents  $d$  for every query  $q$ , with associated values
  - $Score1(d, q)$  - the relevance score for a document  $d$  when we search the background knowledge with the Lucene query associated to question  $q$

```
<question q_id="1" r_id="1" t_id="1">
  <paragraph doc_id="AIDS_RO_23626.txt" score1="0.70500904"/>
  <paragraph doc_id="AIDS_RO_13800.txt" score1="0.42047265"/>
  <paragraph doc_id="AIDS_RO_9102.txt" score1="0.42047265"/>
</question>
<question q_id="2" r_id="1" t_id="1">
  <paragraph doc_id="AIDS_RO_6084.txt" score1="0.35075542"/>
  <paragraph doc_id="AIDS_RO_13556.txt" score1="0.29256588"/>
  <paragraph doc_id="AIDS_RO_13557.txt" score1="0.29256588"/>
</question>
```

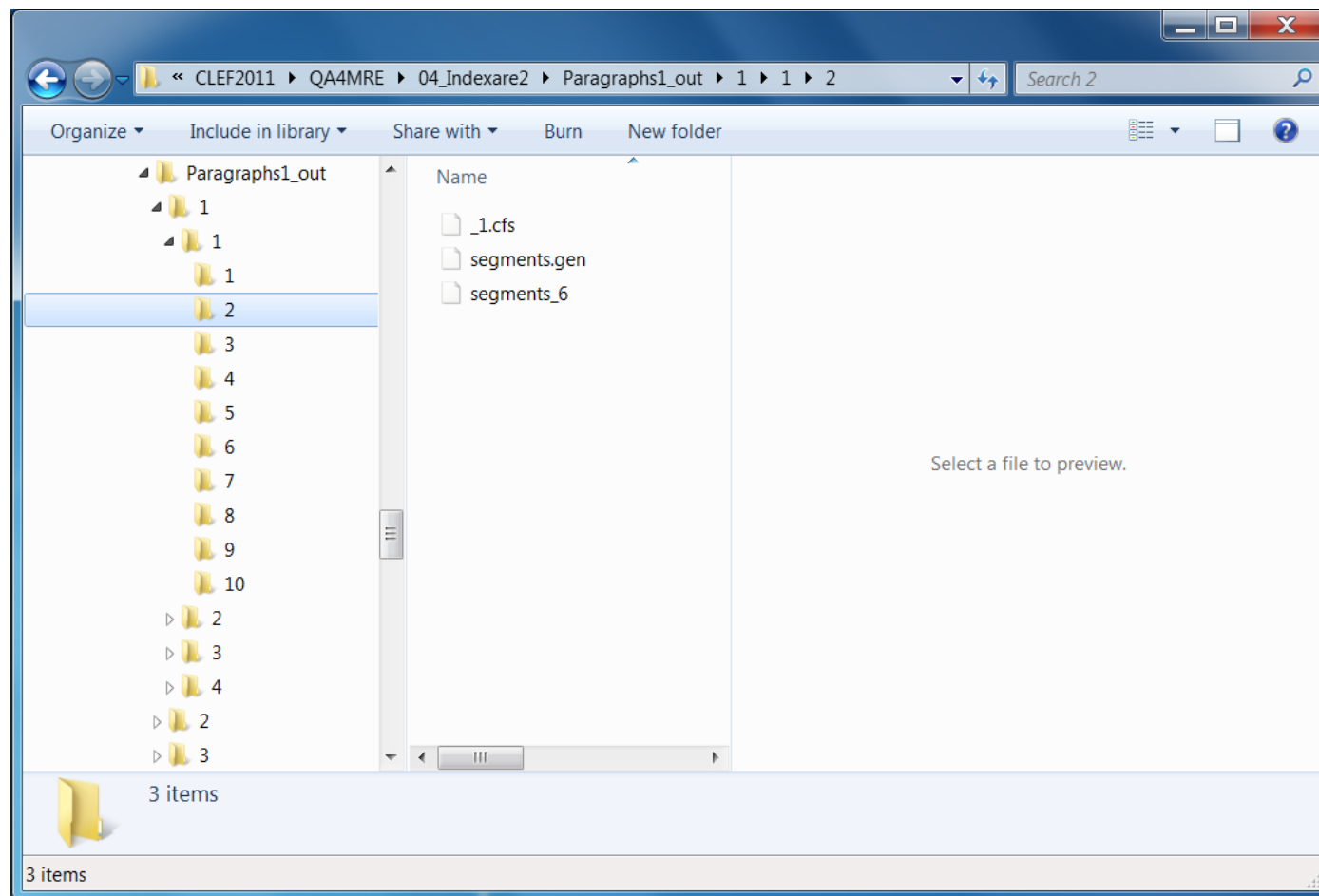
# Information Retrieval – Save relevant documents

- Relevant files are copied in a relative path *<topic id>\<reading test id>\<question id>*



# Indexing relevant documents for questions

- For every question we index relevant documents returned by Lucene at the previous step and relevant documents saved from the initial test file



# Searching using relevant documents for questions

- Then in every index, we performed searches using Lucene queries associated to possible answers
- For every answer, we obtained a list of documents with Lucene relevance scores
- $Score2(d, a)$  is the relevance score for document  $d$  when we search with the Lucene query associated to the answer  $a$

```
<question q_id="1" r_id="1" t_id="1" a_id="1">
  <paragraph doc_id="doc.txt" score2="0.11814499"/>
</question>
<question q_id="1" r_id="1" t_id="1" a_id="2">
  <paragraph doc_id="doc.txt" score2="0.068914026"/>
</question>
<question q_id="1" r_id="1" t_id="1" a_id="3">
  <paragraph doc_id="doc.txt" score2="0.018048882"/>
</question>
<question q_id="1" r_id="1" t_id="1" a_id="4">
  <paragraph doc_id="doc.txt" score2="0.04478345"/>
</question>
<question q_id="1" r_id="1" t_id="1" a_id="5">
  <paragraph doc_id="doc.txt" score2="0.1216592"/>
  <paragraph doc_id="AIDS_RO_13800.txt" score2="0.0015860073"/>
</question>
```

# Identifying the most probable answer

- We combine *Score1* and *Score2*

$$Score(a) = \sum_{d \in \text{Relevant\_docs for } q} Score1(d, q) \times Score2(d, a)$$

- In the end, we consider the answer with the highest value as being the most probable answer
- We submitted different classes of runs, based on thresholds used to consider a NOA response:
  - for the Ro-Ro task, we used three thresholds (0, 0.2, 0.5)
  - for the En-En task, we consider one threshold (0.2)



# Results 1

- Results of UAIC's runs at question answering level

	<b>Ro-Ro</b>			<b>En-En</b>
answered right	30	11	19	25
answered wrong	85	19	43	47
total answered	115	30	62	72
unanswered right	0	19	11	12
unanswered wrong	0	66	42	34
unanswered empty	5	5	5	2
total unanswered	5	90	58	48
<b>Overall accuracy</b>	<b>0.25</b>	<b>0.09</b>	<b>0.16</b>	<b>0.21</b>
<b>C@1 measure</b>	<b>0.26</b>	<b>0.16</b>	<b>0.23</b>	<b>0.29</b>

# Results 2

- Results of UAIC's runs at reading test level

	<b>RO-RO</b>			<b>EN-EN</b>
Topic1 median	0.10	0.00	0.07	0.23
Topic2 median	0.40	0.00	0.29	0.31
Topic3 median	0.30	0.32	0.33	0.36
<b>Overall median</b>	<b>0.20</b>	<b>0.00</b>	<b>0.16</b>	<b>0.31</b>
Topic1 mean	0.10	0.04	0.08	0.25
Topic2 mean	0.39	0.08	0.26	0.27
Topic3 mean	0.29	0.30	0.31	0.32
<b>Overall mean</b>	<b>0.26</b>	<b>0.14</b>	<b>0.22</b>	<b>0.28</b>

# Error Analysis (1)

- One of the most common error sources arises from our attempt to take into account all of the supporting snippets (Sum) that our information retrieval procedure returns => **Possible future solution (Max or Avg)**
- When two candidates have identical scores we choose the first candidate => **Possible future solution**: use the question focus, and perform an additional step of determining the distance between each candidate and the focus

## Error Analysis (2)

- Another problem appears when the top scoring snippet is obtained for an entity name that has the highest Tf/Idf value => **Possible future solution:** use the question focus
- For the En-En task, a problem was that we did not use background information.
- Numbers are also a major cause of errors, mainly because they can be written either with letters or with digits

# Requirements (Team: max 2 persons, Deadline: 1 March)

- 1) Create a consumer service for the following Web Service:  
<http://instrumente.infoiasi.ro/WebQuestionAnswering/>
- Links: <http://jax-ws.java.net/articles/jaxws-netbeans/>
- 2) Lucene: Use the following archive in order to index and to search in a collections of texts <http://thor.info.uaic.ro/~adiftene/Scoala/2012/APLN/Resurse/>

# Links

- **Yes–no question:** [http://en.wikipedia.org/wiki/Yes%E2%80%93no\\_question](http://en.wikipedia.org/wiki/Yes%E2%80%93no_question)
- **Question Answering:** [http://en.wikipedia.org/wiki/Question\\_answering](http://en.wikipedia.org/wiki/Question_answering)
- **Information Extraction:** [http://en.wikipedia.org/wiki/Information\\_extraction](http://en.wikipedia.org/wiki/Information_extraction)
- **Information Retrieval:** [http://en.wikipedia.org/wiki/Information\\_retrieval](http://en.wikipedia.org/wiki/Information_retrieval)
- **Lecture 13: Evaluation: Precision and Recall**  
<http://courses.washington.edu/ling473/Lecture13.pdf>
- **Precision and Recall of Five Search Engines for Retrieval of Scholarly Information in the Field of Biotechnology:** <http://www.webology.org/2005/v2n2/a12.html>