COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

MULTIMODAL INFORMATION ACCESS & SYNTHESIS mias

Learning and Inference for Natural Language Understanding

Dan Roth

Department of Computer Science

University of Illinois at Urbana-Champaign

April 2015

Boston University

Making Sense of Unstructured Data

Dan Roth

Department of Computer Science

University of Illinois at Urbana-Champaign

April 2015

Boston University

Data Science: Making Sense of (Unstructured) Data

- Most of the data today is unstructured
  - Text, Images, Sensory Data
  - It's not only **BIG,** it's **COMPLEX & Heterogeneous**

- **Challenge:** How to *understand* what the data says?
  - How to deal with the huge amount of unstructured data as if it was organized in a database with a *known* schema.
  - **Organize, access, analyze and synthesize unstructured data.**

Data Science: Making Sense of (Unstructured) Data

- Most of the data today is unstructured
  - Text, Images, Sensory Data
  - It's not only **BIG,** it's **COMPLEX & Heterogeneous**

- **Challenge:** How to *understand* what the data says?
  - How to deal with the huge amount of unstructured data as if it was organized in a database with a *known* schema.
  - **Organize, access, analyze and synthesize unstructured data.**

- Theories, algorithms, and tools to enable **transforming raw data** into **useful and understandable information** & integrating it with existing resources

A view on Extracting Meaning from Unstructured Text

"as is, with all defects" basis, without maintenance, debugging , support or improvement. Licensee assumes the entire risk as to the results and performance of the Software and/or associated materials. Licensee agrees that University shall not be held liable for any direct, indirect, consequential, or incidental damages with respect to any claim by Licensee or any third party on account of or arising from this Agreement or use of the Software and/or associated materials.

4. Licensee understands the Software is proprietary to the University. Licensee will take all reasonable steps to insure that the source code is protected and secured from unauthorized disclosure, use, or release and will treat it with at least the same level of care as Licensee would use to protect and secure its own proprietary computer programs and/or information, but using no less than reasonable care.
5. In the event that Licensee shall be in default in the performance of any material obligations under this Agreement, and if the default has not been remedied within sixty (60) days after the date of notice in writing of such default, University may terminate this Agreement by written notice. In the event of termination, Licensee shall promptly return to University the original and any copies of licensed Software in Licensee's possession. In the event of any termination of this Agreement, any and all sublicenses granted by Licensee to third parties pursuant to this Agreement (as permitted by this Agreement) prior to the date of such termination shall nevertheless remain in full force and effect.
6. The Software was developed, in part, with support from the National Science Foundation, and the Federal Government has certain license rights in the Software.
7. This Agreement shall be construed and interpreted in accordance with the laws of the State of Illinois, U.S.A..
8. This Agreement shall be subject to all United States Government laws and regulations now and hereafter applicable to the subject matter of this Agreement, including specifically the Export Law provisions of the Departments of Commerce and State. Licensee will not export or re-export the Software without the appropriate United States or foreign government license.

By its registration below, Licensee confirms that it understands the terms and conditions of this Agreement, and agrees to be bound by them. This Agreement shall become effective as of the date of execution by Licensee.

Registration information: (We will not disclose any of this information. It is for internal use only.)

Name:

Email Address:

Organization:

Accept | Clear

A view on Extracting Meaning from Unstructured Text

"as is, with all defects" basis, without maintenance, debugging , support or improvement. Licensee assumes the entire risk as to the results and performance of the Software and/or associated materials. Licensee agrees that University shall not be held liable for any direct, indirect, consequential, or incidental damages with respect to any claim by Licensee or any third party on account of or arising from this Agreement or use of the Software and/or associated materials.

4. Licensee understands the Software is proprietary to the University. Licensee will take all reasonable steps to insure that the source code is protected and secured from unauthorized disclosure, use, or release and will treat it with at least the same level of care as Licensee would use to protect and secure its own proprietary computer programs and/or information, but using no less than reasonable care.
5. In the event that Licensee shall be in default in the performance of any material obligations under this Agreement, and if the default has not been remedied within sixty (60) days after the date of notice in writing of such default, University may terminate this Agreement by written notice. In the event of termination, Licensee shall promptly return to University the original and any copies of licensed Software in Licensee's possession. In the event of any termination of this Agreement, any and all sublicenses granted by Licensee to third parties pursuant to this Agreement (as permitted by this Agreement) prior to the date of such termination shall nevertheless remain in full force and effect.
6. The Software was developed, in part, with support from the National Science Foundation, and the Federal Government has certain license rights in the Software.
7. This Agreement shall be construed and inte[illegible] in accordance with the laws of the State of Illinois, U.S.A..
8. This Agreement shall be subject to all U[illegible] Government laws and regulations now and hereafter applica[illegible]ject matter of this Agreement, including specifically th[illegible] provisions of the Departments of Commerce and State. Licensee[illegible]rt or re-export the Software without the appropriate United[illegible]eign government license.

By its registration below, Licen[illegible]hat it understands the terms and conditions of this Agreem[illegible] to be bound by them. This Agreement shall become effective as[illegible] execution by Licensee.

Registration information[illegible]ot disclose any of this information. It is for internal use only.)

Name:

Email Address:

Organization:

ACCEPT?

A view on Extracting Meaning from Unstructured Text

"as is, with all defects" basis, without maintenance, debugging , support or improvement. Licensee assumes the entire risk as to the results and performance of the Software and/or associated materials. Licensee agrees that University shall not be held liable for any direct, indirect, consequential, or incidental damages with respect to any claim by Licensee or any third party on account of or arising from this Agreement or use of the Software and/or associated materials.

4. Licensee understands the Software is proprietary to the University. Licensee will take all reasonable steps to insure that the source code is protected and secured from unauthorized disclosure, use, or release and will treat it with at least the same level of care as Licensee w[illegible] proprietary computer programs [illegible] reasonable care.
5. In the event that Licensee shal[illegible] ial obligations under this Agreeme[illegible] within sixty (60) days after the [illegible] University may terminate this Agreement by written notice. In the event of termination, Licensee shall promptly return to University the original and any copies of licensed Software in Licensee's possession. In the event of any termination of this Agreement, any and all sublicenses granted by Licensee to third parties pursuant to this Agreement (as permitted by this Agreement) prior to the date of such termination shall nevertheless remain in full force and effect.
6. The Software was developed, in part, with support from the National Science Foundation, and the Federal Government has ce[illegible]tain license rights in the Software.
7. This Agreement shall be construed and inte[illegible] in accordance with the laws of the State of Illinois, U.S.A..
8. This Agreement shall be subject to all U[illegible] Government laws and regulations now and hereafter applica[illegible]ject matter of this Agreement, including specifically th[illegible] provisions of the Departments of Commerce and State. Licensee[illegible]rt or re-export the Software without the appropriate United[illegible]eign government license.

By its registration below, Licen[illegible] that it understands the terms and conditions of this Agreem[illegible] to be bound by them. This Agreement shall become effective as[illegible] execution by Licensee.

Registration information[illegible]ot disclose any of this information. It is for internal use only.)

Name:

Email Address:

Organization:

Does it say that they'll give my email address away?

ACCEPT?

A view on Extracting Meaning from Unstructured Text

"as is, with all defects" basis, without maintenance, debugging , support or improvement. Licensee assumes the entire risk as to the results and performance of the Software and/or associated materials. Licensee agrees that University shall not be held liable for any direct, indirect, consequential, or incidental damages with respect to any claim by Licensee or any third party on account of or arising from this Agreement or use of the Software and/or associated materials.

4. Licensee understands the Software is proprietary to the University. Licensee will take all reasonable steps to insure that the source code is protected and secured from unauthorized disclosure, use, or release and will treat it with at least the same level of care as Licensee would use to protect and secure its own proprietary computer programs and/or information, but using no less than reasonable care.
5. In the event ... ance of any material obligations u... t been remedied within sixty (... such default, University ma... In the event of termination, ... e original and any copies of lice... event of any termination of this Agreement, any and all sublicenses granted by Licensee to third parties pursuant to this Agreement (as permitted by this Agreement) prior to the date of such termination shall nevertheless remain in full force and effect.
6. The Software was developed, in part, with support from the National Science Foundation, and the Federal Government has certain license rights in the Software.
7. This Agreement shall be construed and interpreted in accordance with the laws of the State of Illinois, U.S.A..
8. This Agreement shall be subject to all United States Government laws and regulations now and hereafter applicable to the subject matter of this Agreement, including specifically the Export Law provisions of the Departments of Commerce and State. Licensee will not export or re-export the Software without the appropriate United States or foreign government license.

By its registration below, Licensee confirms that it understands the terms and conditions of this Agreement, and agrees to be bound by them. This Agreement shall become effective as of the date of execution by Licensee.

Registration information: (We will not disclose any of this information. It is for internal use only.)

Name:

Email Address:

Organization:

Accept | Clear

Does it say that they'll give my email address away?

ACCEPT?

A view on Extracting Meaning from Unstructured Text

"as is, with all defect
improvement. Licens
performance of the S
University shall not b
incidental damages w
account of or arising
associated materials.

4. Licensee understands the Software is proprietary to the University. Licensee will take all reasonable steps to insure that the source code is protected and secured from unauthorized disclosure, use, or release and will treat it with at least the same level of care as Licensee would use to protect and secure its own proprietary computer programs and/or information, but using no less than reasonable care.
5. In the event ... ance of any material obligations u... t been remedied within sixty (... such default, University m... In the event of termination, ... e original and any copies of lice... event of any termination ... ted by Licensee to third parties pursuant to this Agreement (as permitted by this Agreement) prior to the date of such termination shall nevertheless remain in full force and effect.
6. The Software was developed, in part, with support from the National Science Foundation, and the Federal Government has certain license rights in the Software.
7. This Agreement shall be construed and interpreted in accordance with the laws of the State of Illinois, U.S.A..
8. This Agreement shall be subject to all United States Government laws and regulations now and hereafter applicable to the subject matter of this Agreement, including specifically the Export Law provisions of the Departments of Commerce and State. Licensee will not export or re-export the Software without the appropriate United States or foreign government license.

By its registration below, Licensee confirms that it understands the terms and conditions of this Agreement, and agrees to be bound by them. This Agreement shall become effective as of the date of execution by Licensee.

Registration information: (We will not disclose any of this information. It is for internal use only.)

Name:

Email Address:

Organization:

Accept | Clear

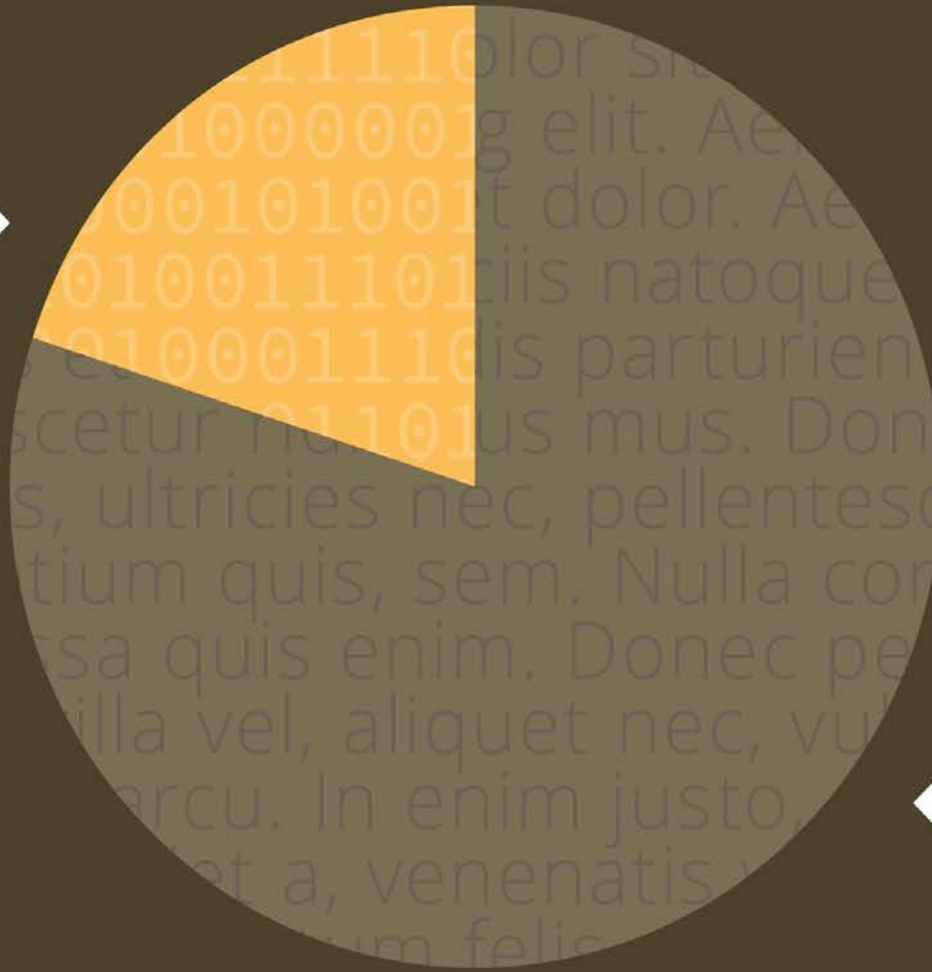
Large Scale Data → Meaning Transformation

Massive & Deep

Does it say that they'll give my email address away?

ACCEPT?

BIG DATA



BIG TEXT

90% of the world's text has been created in the last 2 years, and there will be a 50-fold increase by 2020.

WORLD TEXT

2012 2014 2020

90% of the world’s text has been created in the last 2 years, and there will be a 50-fold increase by 2020.

WORLD TEXT

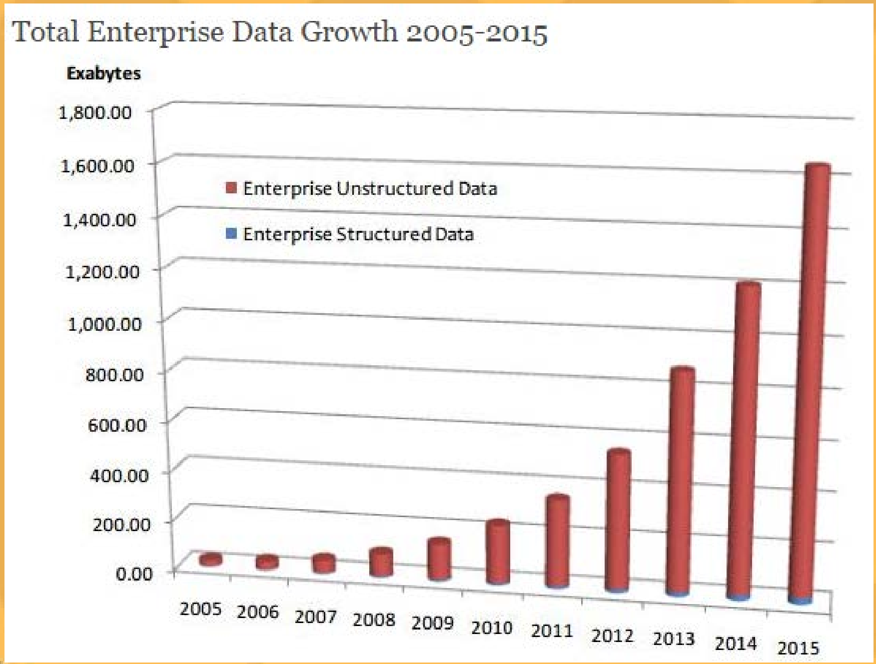
- Scientific Articles
- Medical Records
- Education
- Business
- Social Media

http://www.dataversity.net/the-growth-of-unstructured-data-what-are-we-going-to-do-with-all-those-zettabytes/
http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/
http://www.datasciencecentral.com/profiles/blogs/structured-vs-unstructured-data-the-rise-of-data-anarchy

90% of the world's text has been created in the last 2 years, and there will be a 50-fold increase by 2020.

WORLD TEXT

Scientific Articles
Medical Records
Education
Business
Social Media

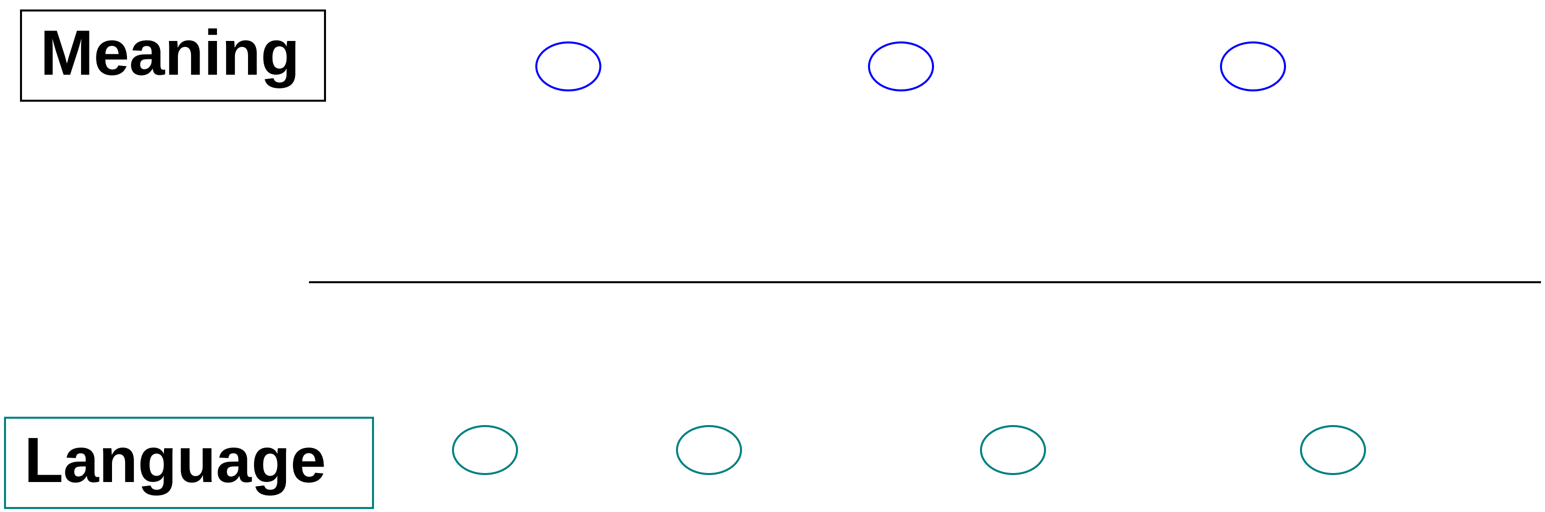


2012 2014 2020

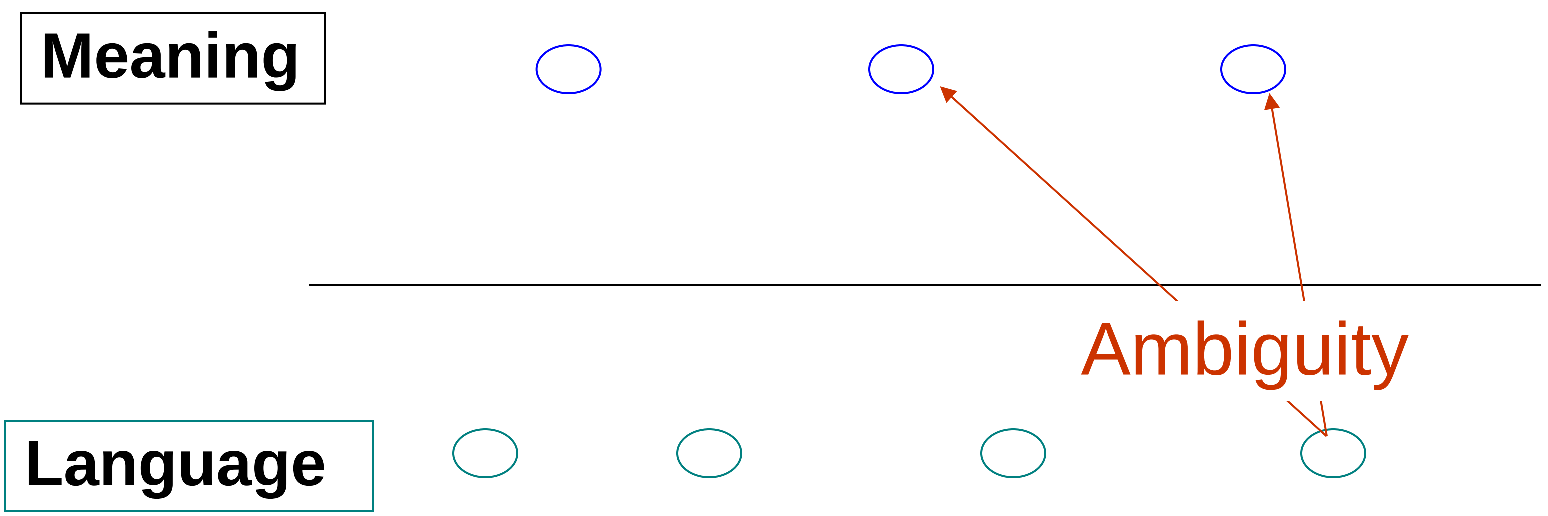
http://www.dataversity.net/the-growth-of-unstructured-data-what-are-we-going-to-do-with-all-those-zettabytes/
http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/
http://www.datasciencecentral.com/profiles/blogs/structured-vs-unstructured-data-the-rise-of-data-anarchy


KEYWORD
SEARCH

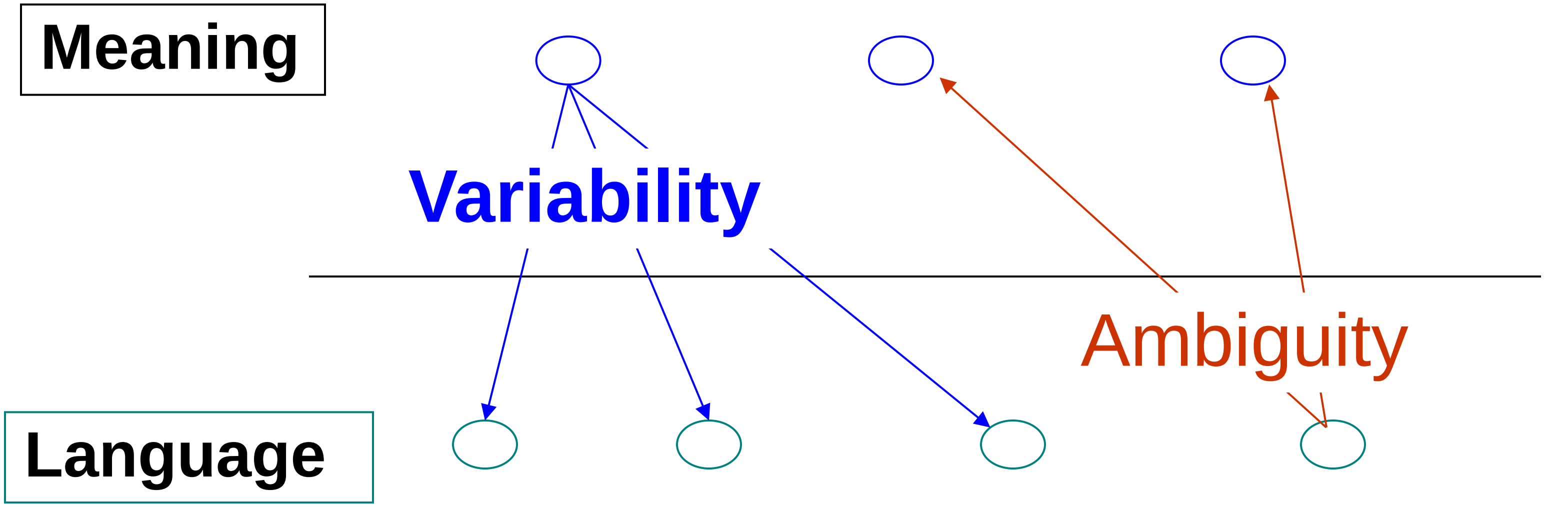
Why is it difficult?



Why is it difficult?



Why is it difficult?

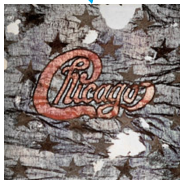


Ambiguity

It's a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the "N".	***Chicago*** was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..	***Chicago VIII*** was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.

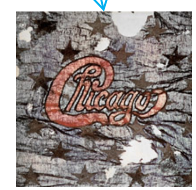
Ambiguity

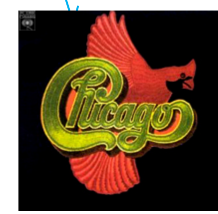
It's a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the "N".	***Chicago*** was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..	***Chicago VIII*** was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.

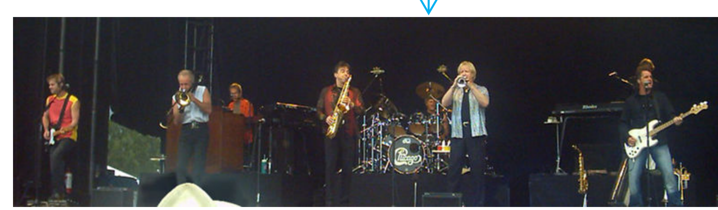


Ambiguity

It's a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the "N".	***Chicago*** was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..	***Chicago VIII*** was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.

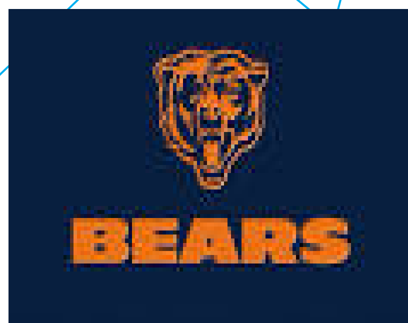


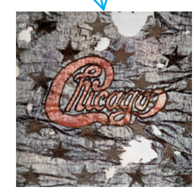


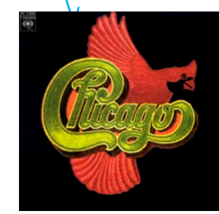


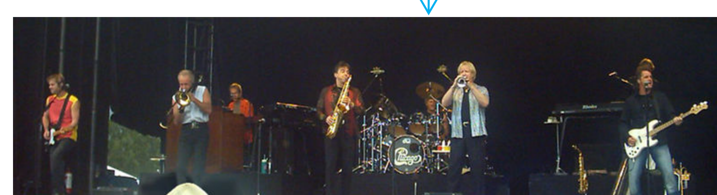
Ambiguity

It's a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the "N".	***Chicago*** was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..	***Chicago VIII*** was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.

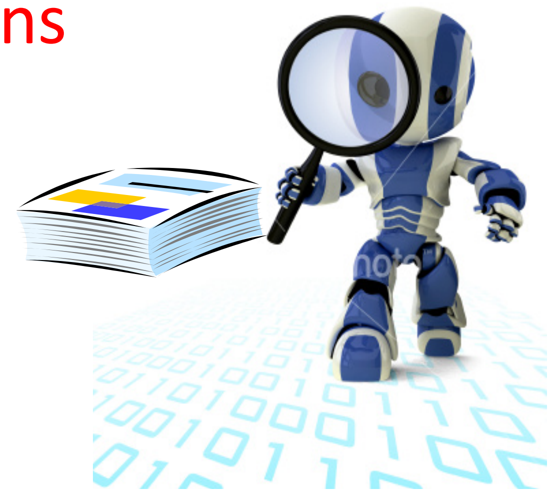








Variability in Natural Language Expressions



Variability in Natural Language Expressions

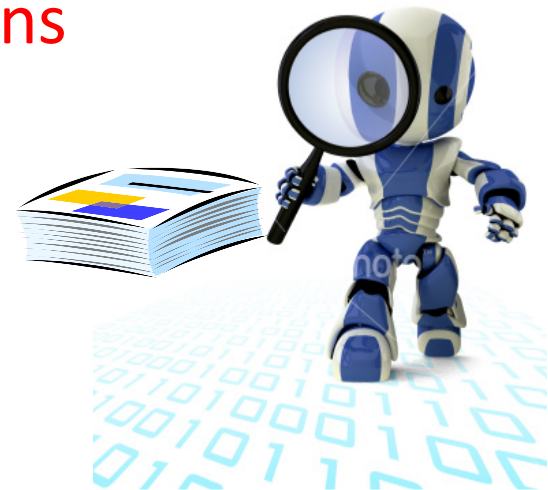
Determine if Jim Carpenter works for the government

Variability in Natural Language Expressions

Determine if Jim Carpenter works for the government

Jim Carpenter works for the U.S. Government.

Variability in Natural Language Expressions



Determine if Jim Carpenter works for the government

Jim Carpenter works for the U.S. Government.

The American government employed Jim Carpenter.

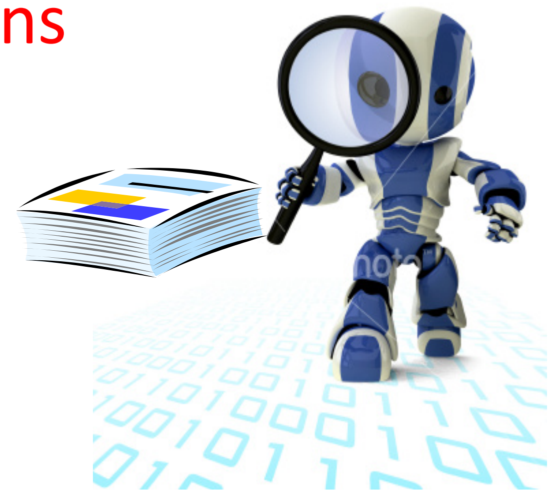
Jim Carpenter was fired by the US Government.

Jim Carpenter worked in a number of important positions.
…. As a press liaison for the IRS, he made contacts in the white house.

Russian interior minister Yevgeny Topolov met yesterday with his US counterpart, Jim Carpenter.

Former US Secretary of Defense Jim Carpenter spoke today…

Variability in Natural Language Expressions



- Determine if Jim Carpenter works for the government

Jim Carpenter works for the U.S. Government.

The American government employed Jim Carpenter.

Jim Carpenter was fired by the US Government.

Jim Carpenter worked in a number of important positions.
…. As a press liaison for the IRS, he made contacts in the white house.

- Russian interior minister Yevgeny Topolov met yesterday with his US counterpart, Jim Carpenter.

Former US Secretary of Defense Jim Carpenter spoke today…

Variability in Natural Language Expressions

- Determine if Jim Carpenter works for the government

Jim Carpenter works for the U.S. Government.

The American government employed Jim Carpenter.

Jim Carpenter was fired by the US Government.

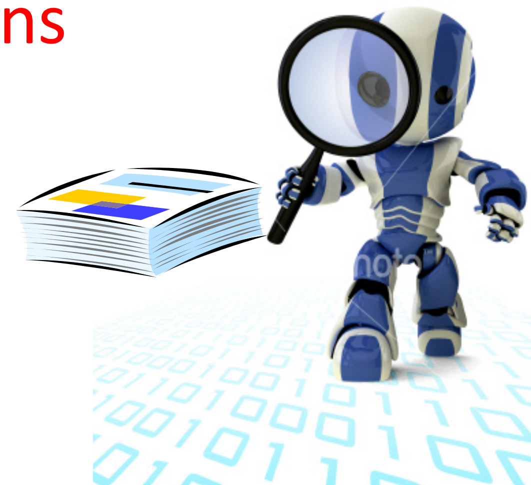
Jim Carpenter worked in a number of important positions.
   …. As a press liaison for the IRS, he made contacts in the white house.

- Russian interior minister Yevgeny Topolov met yesterday with his US counterpart, Jim Carpenter.

Former US Secretary of Defense Jim Carpenter spoke today…

Standard techniques cannot deal with the variability of expressing meaning nor with the ambiguity of interpretation

Variability in Natural Language Expressions



- Determine if Jim Carpenter works for the government

Jim Carpenter works for the U.S. Government.

The American government employed Jim Carpenter.

Jim Carpenter was fired by the US Government.

Jim Carpenter worked in a number of important positions. …. As a press liaison for the IRS, he made contacts in the white house.

- Russian interior minister Yevgeny Topolov met yesterday with his US counterpart, Jim Carpenter.

Former US Secretary of Defense Jim Carpenter spoke today…

Standard techniques cannot deal with the variability of expressing meaning nor with the ambiguity of interpretation

Needs:

- Understanding Relations, Entities and Semantic Classes
- Acquiring knowledge from external resources; representing knowledge
- Identifying, disambiguating & tracking entities, events, etc.
- Time, quantities, processes…

What can this give us?

What can this give us?

- Moving towards natural language understanding…

What can this give us?

- Moving towards natural language understanding…

- A law office wants to get the list of all people that were mentioned in email correspondence with the office.
  - For each name, determine whether is was mentioned adversarially or not.

What can this give us?

- Moving towards natural language understanding…

- A law office wants to get the list of all people that were mentioned in email correspondence with the office.
  - For each name, determine whether is was mentioned adversarially or not.

- A political scientist studies Climate Change and its effect on Societal instability. He wants to identify all events related to demonstrations, protests, parades, analyze them (who, when, where, why) and generate a timeline and a causality chain.

What can this give us?

- Moving towards natural language understanding…
- A law office wants to get the list of all people that were mentioned in email correspondence with the office.
  - For each name, determine whether is was mentioned adversarially or not.
- A political scientist studies Climate Change and its effect on Societal instability. He wants to identify all events related to demonstrations, protests, parades, analyze them (who, when, where, why) and generate a timeline and a causality chain.
- English as a Second Language (ESL): Most people that write English today are non-native speakers. Nevertheless, the only assistance we can given them is spelling correction against a fixed, large dictionary….

What can this give us (Cont.) ?

- Compliance & E-Discovery: A trading company had half of their sales team leave to start a rival company. The CEO wanted proof they stole company information and broke their employee covenants.
  - Ideally, know about it before it happens
- An analyst in a financial institution sends company A information about company B
  - Mistakenly? Deliberately?

What can this give us (Cont.) ?

- Compliance & E-Discovery: A trading company had half of their sales team leave to start a rival company. The CEO wanted proof they stole company information and broke their employee covenants.
  - Ideally, know about it before it happens
- An analyst in a financial institution sends company A information about company B
  - Mistakenly? Deliberately?

- An electronic health record (EHR):
  - A personal health record in digital format. Includes information relating to:
  - Current and historical health, medical conditions, tests, treatments,…
    - **A write only document**
  - Use it in medical advice systems; medication selection and tracking (Vioxx…);
  - Better care: ISU patients summary average 110 pages….
  - Science – correlating response to drugs with other conditions

Machine Learning + Inference based NLP

- It's difficult to program predicates of interest due to
  - Ambiguity (everything has multiple meanings)
  - Variability (everything you want to say you can say in many ways)
- Models are based on Statistical Machine Learning & Inference

Machine Learning + Inference based NLP

- It's difficult to program predicates of interest due to
  - Ambiguity (everything has multiple meanings)
  - Variability (everything you want to say you can say in many ways)
- Models are based on Statistical Machine Learning & Inference

Research Focus:

Machine Learning + Inference based NLP

- It's difficult to program predicates of interest due to
  - Ambiguity (everything has multiple meanings)
  - Variability (everything you want to say you can say in many ways)
- Models are based on Statistical Machine Learning & Inference

Research Focus:

- Modeling and learning algorithms for different phenomena
  - Classification models
  - Structured models
  - Learning protocols that exploit Indirect Supervision

Machine Learning + Inference based NLP

- It's difficult to program predicates of interest due to
  - Ambiguity (everything has multiple meanings)
  - Variability (everything you want to say you can say in many ways)
- Models are based on Statistical Machine Learning & Inference

Research Focus:

- Modeling and learning algorithms for different phenomena
  - Classification models
  - Structured models
  - Learning protocols that exploit Indirect Supervision

Well understood; easy to build black box categorizers

Machine Learning + Inference based NLP

- It's difficult to program predicates of interest due to
  - Ambiguity (everything has multiple meanings)
  - Variability (everything you want to say you can say in many ways)
- Models are based on Statistical Machine Learning & Inference

Research Focus:

- Modeling and learning algorithms for different phenomena
  - Classification models
  - Structured models
  - Learning protocols that exploit Indirect Supervision
- Inference: make decisions that account for domain & task specific knowledge
  - Constrained Conditional Models: formulating inference as ILP

Well understood; easy to build black box categorizers

Learn models; Acquire knowledge/constraints; Make decisions.

$$\boxed{\operatorname*{argmax}_y \boldsymbol{\lambda} \cdot F(x,y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})}$$

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

Comprehension

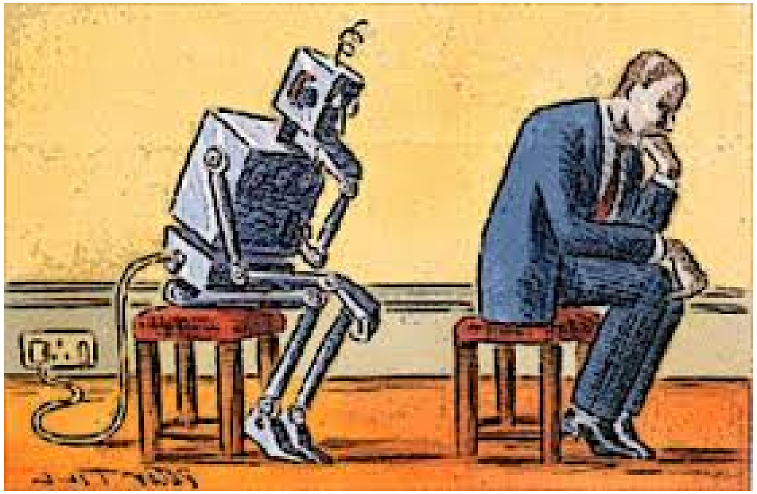
(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

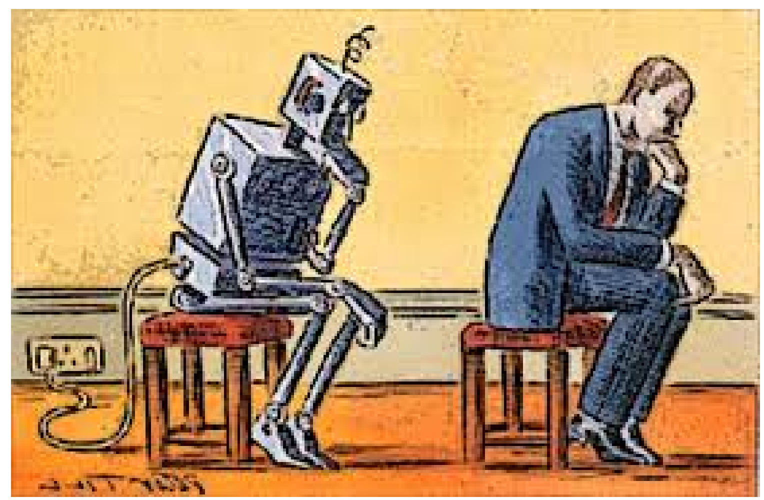
This is an Inference Problem

What is Needed?

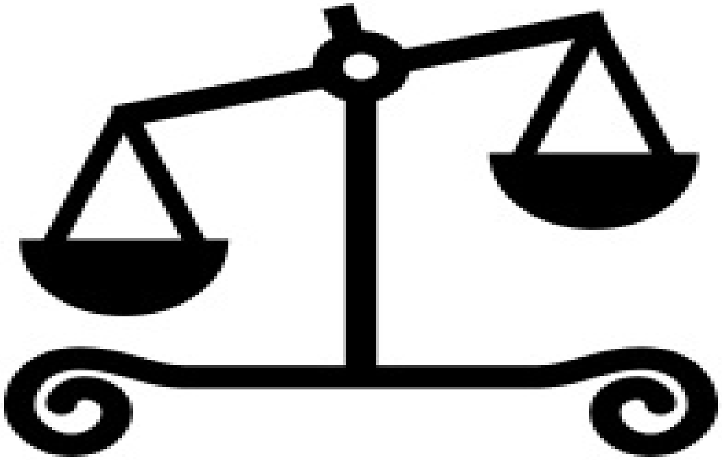
What is Needed?

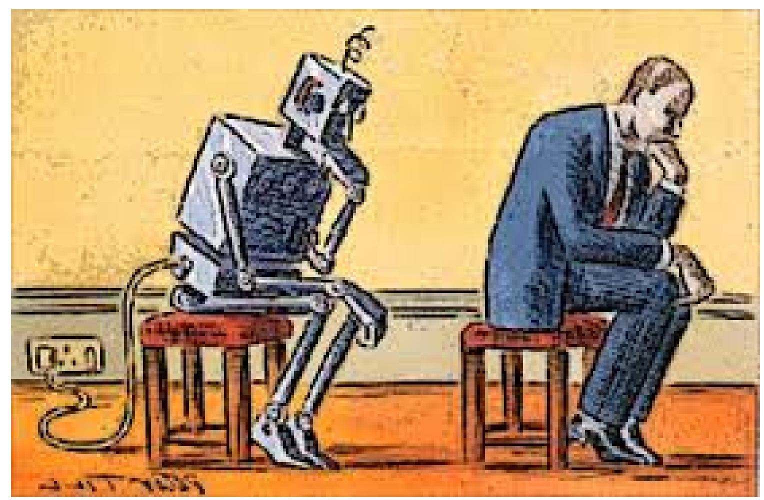


What is Needed?



What is Needed?





Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem

Learning and Inference

- Natural language understanding decisions are global decisions in which several local decisions play a role, but there are mutual dependencies on their outcome.

Learning and Inference

- Natural language understanding decisions are global decisions in which several local decisions play a role, but there are mutual dependencies on their outcome.

  - We need to think about:
    - **(Learned) models for different sub-problems**
    - **Reasoning with knowledge relating sub-problems**
    - **Knowledge that may appear only at evaluation time**

Learning and Inference

- Natural language understanding decisions are global decisions in which several local decisions play a role, but there are mutual dependencies on their outcome.
  - We need to think about:
    - **(Learned) models for different sub-problems**
    - **Reasoning with knowledge relating sub-problems**
    - **Knowledge that may appear only at evaluation time**
- Goal: Incorporate models' information, along with knowledge (constraints) in making coherent decisions
  - Decisions that respect the local models as well as domain & context specific knowledge/constraints.

Learning and Inference

- Natural language understanding decisions are global decisions in which several local decisions play a role, but there are mutual dependencies on their outcome.

Natural Language Interpretation is an Inference Problem that is best thought of as a knowledge constrained optimization problem, done on top of multiple statistically learned models.

  - We need to think about:
    - **(Learned) models for different sub-problems**
    - **Reasoning with knowledge relating sub-problems**
    - **Knowledge that may appear only at evaluation time**
- Goal: Incorporate models' information, along with knowledge (constraints) in making coherent decisions
  - Decisions that respect the local models as well as domain & context specific knowledge/constraints.

Learning and Inference

- Natural language understanding decisions are global decisions in which several local decisions play a role, but there are mutual dependencies on their outcome.

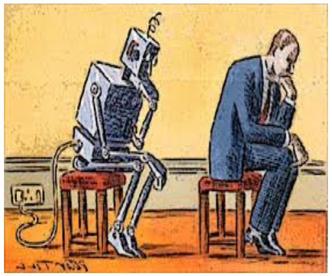
Natural Language Interpretation is an Inference Problem that is best thought of as a knowledge constrained optimization problem, done on top of multiple statistically learned models.

  - We need to think about:
    - **(Learned) models for different sub-problems**
    - **Reasoning with knowledge relating sub-problems**
    - **Knowledge that may appear only at evaluation time**
- Goal: Incorporate models' information, along with knowledge (constraints) in making coherent decisions
  - Decisions that respect the local models as well as domain & context specific knowledge/constraints.

Many forms of Inference; a lot boil down to determining best assignment

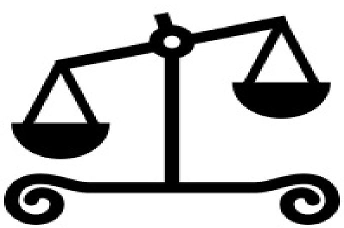
Technical Outline

- Knowledge and Inference
  - Combining the soft with the logical/declarative nature of Natural Language
    - **Constrained Conditional Models: A formulation for global inference with knowledge modeled as expressive structural constraints**
    - **Some examples**



- Cycles of Knowledge
  - Grounding for/using Knowledge

- Learning with Indirect Supervision
  - Response Based Learning: learning from the world's feedback



- Scaling Up: Amortized Inference
  - Can the k-th inference problem be cheaper than the 1st?

Constrained Conditional Models

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

Constrained Conditional Models

$$\underset{y}{\operatorname{argmax}} \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

Constrained Conditional Models

$$\underset{y}{\operatorname{argmax}} \, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

Weight Vector for "local" models

COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Constrained Conditional Models

$$\underset{y}{\operatorname{argmax}} \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

Constrained Conditional Models

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

(Soft) constraints component

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

Constrained Conditional Models

Penalty for violating the constraint.

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

(Soft) constraints component

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a "legal" assignment

Constrained Conditional Models

$$\operatorname*{argmax}_{y} \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

Penalty for violating the constraint.

(Soft) constraints component

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a "legal" assignment

How to solve?

This is an Integer Linear Program

Solving using ILP packages gives an exact solution.

Cutting Planes, Dual Decomposition & other search techniques are possible

How to train?

Training is learning the objective function

Decouple? Decompose?

How to exploit the structure to minimize supervision?

Examples: CCM Formulations

$$\operatorname*{argmax}_{y} \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

Examples: CCM Formulations

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

CCMs can be viewed as a general interface to easily combine declarative domain knowledge with data driven statistical models

Examples: CCM Formulations

$$\operatorname*{argmax}_{y} \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

CCMs can be viewed as a general interface to easily combine declarative domain knowledge with data driven statistical models

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Examples: CCM Formulations

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

CCMs can be viewed as a general interface to easily combine declarative domain knowledge with data driven statistical models

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Sequential Prediction

HMM/CRF based:

Argmax $\sum \lambda_{ij} x_{ij}$

Linguistics Constraints

Cannot have both A states and B states in an output sequence.

Examples: CCM Formulations

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

CCMs can be viewed as a general interface to easily combine declarative domain knowledge with data driven statistical models

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Sentence Compression/Summarization:

Language Model based:

Argmax $\sum \lambda_{ijk} x_{ijk}$

Linguistics Constraints

If a modifier chosen, include its head
If verb is chosen, include its arguments

Examples: CCM Formulations

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x,y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

CCMs can be viewed as a general interface to easily combine declarative domain knowledge with data driven statistical models

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Sentence Compression/Summarization:

Language Model based:

Argmax $\sum \lambda_{ijk} x_{ijk}$

Linguistics Constraints

If a modifier chosen, include its head
If verb is chosen, include its arguments

Examples: CCM Formulations

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

CCMs can be viewed as a general interface to easily combine declarative domain knowledge with data driven statistical models

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Constrained Conditional Models Allow:

- Learning a simple model (or multiple; or pipelines)
- Make decisions with a more complex model
- Accomplished by directly incorporating constraints to bias/re-rank global decisions composed of simpler models' decisions
- More sophisticated algorithmic approaches exist to bias the output [CoDL: Cheng et. al 07,12; PR: Ganchev et. al. 10; DecL, UEM: Samdani et. al 12]

Semantic Role Labeling (SRL)

I *left* my pearls to my daughter in my will .

[I]$_{A0}$ *left* [my pearls]$_{A1}$ [to my daughter]$_{A2}$ [in my will]$_{AM\text{-}LOC}$.

- ***A0*** Leaver
- ***A1*** Things left
- ***A2*** Benefactor
- ***AM-LOC*** Location

I *left* my pearls to my daughter in my will .

Semantic Role Labeling (SRL)

Archetypical Information Extraction Problem: E.g., Concept Identification and Typing, Event Identification, etc.

I *left* my pearls to my daughter in my will .

[I]$_{A0}$ *left* [my pearls]$_{A1}$ [to my daughter]$_{A2}$ [in my will]$_{AM-LOC}$.

- ***A0*** Leaver
- ***A1*** Things left
- ***A2*** Benefactor
- ***AM-LOC*** Location

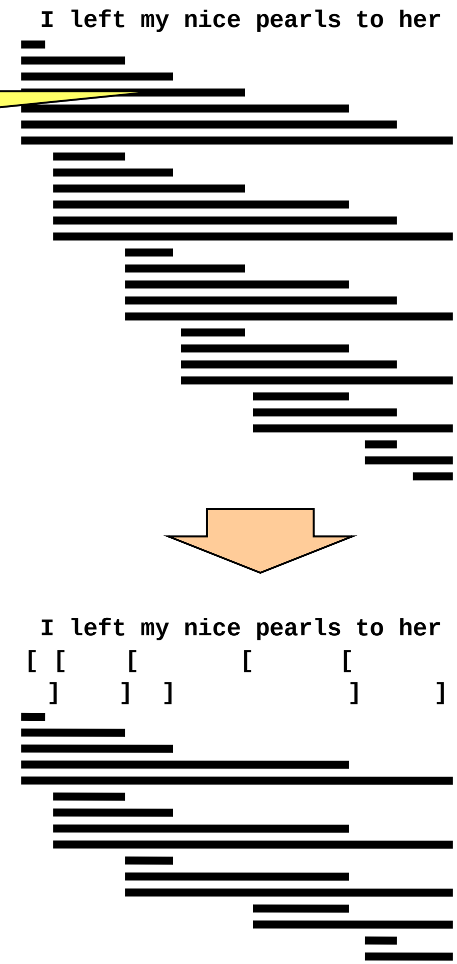
I *left* my pearls to my daughter in my will .

Algorithmic Approach

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output

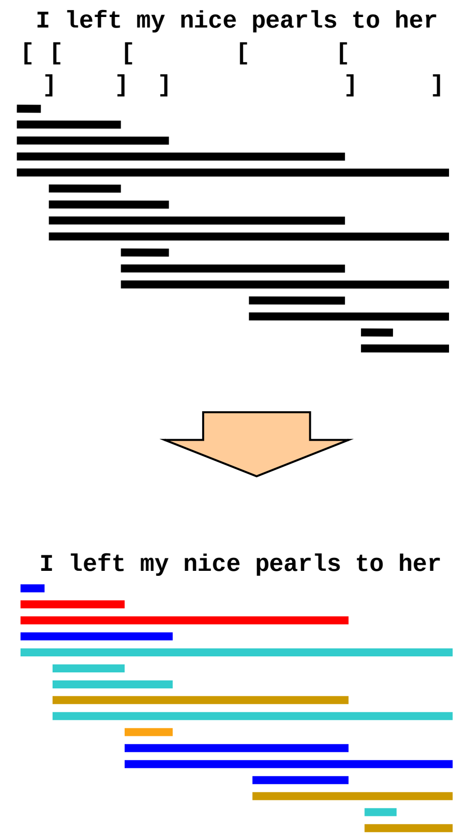
Algorithmic Approach

- **Identify** argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- **Classify** argument candidates
  - Argument Classifier
    - **Multi-class classification**
- **Inference**
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output



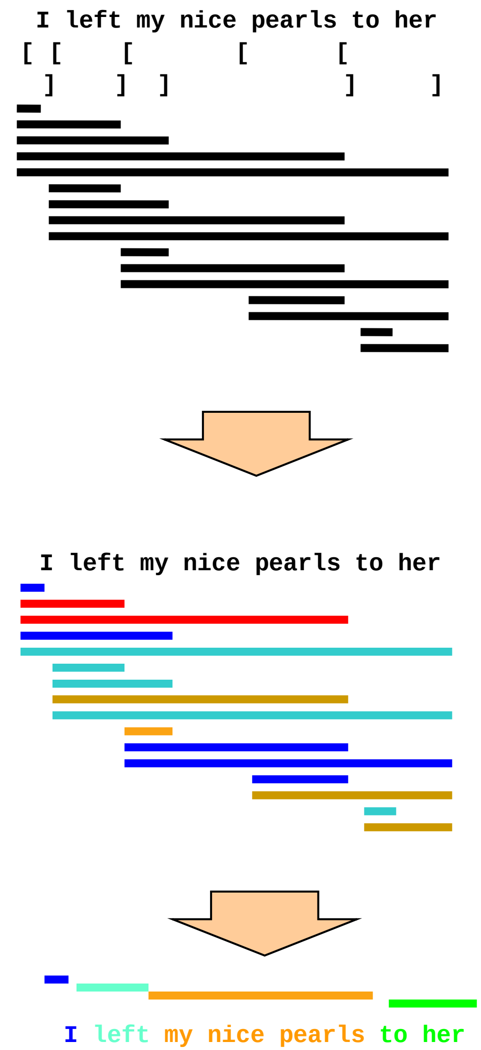
Algorithmic Approach

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output



Algorithmic Approach

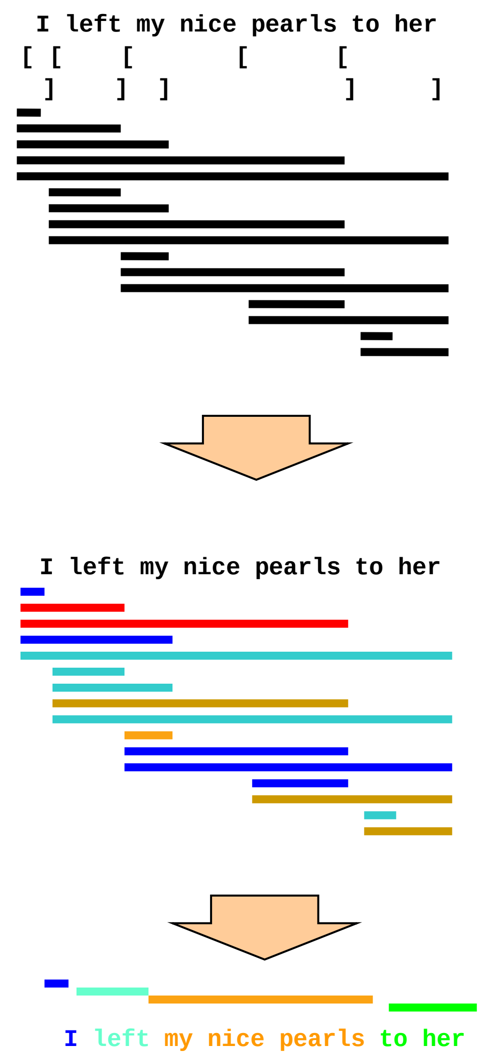
- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output



Algorithmic Approach

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference
  - Use the estimated probability distribution given by the argument classifier
  - Use st ic constraints
  - Infer t utput

One inference problem for each verb predicate.



Cognitive Computation Group
University of Illinois at Urbana-Champaign

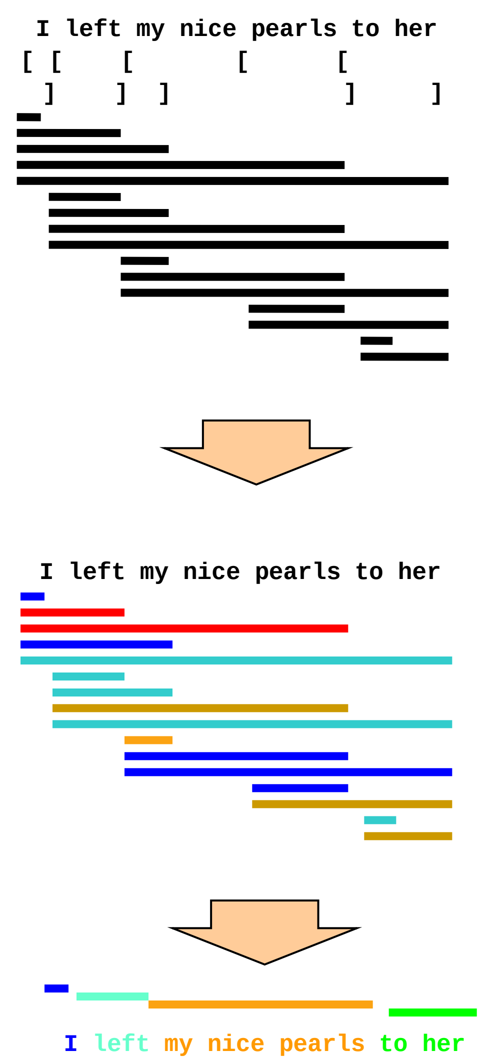
Algorithmic Approach

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference

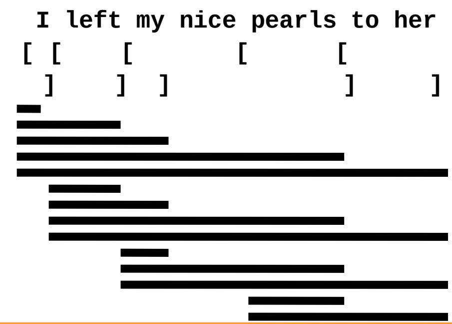
$\text{argmax} \sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}$

Subject to:

- One label per argument: $\sum_t y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,....



Algorithmic Approach



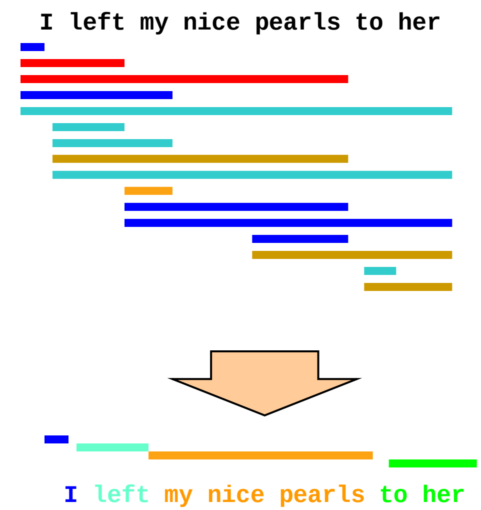
- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference

Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

argmax $\sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}$

Subject to:

- One label per argument: $\sum_t y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,….



Algorithmic Approach

- Identify argument candid
  - Pruning [Xue&Palmer, EM
  - Argument Identifier
    - **Binary classification**
- Classify argument candid
  - Argument Classifier
    - **Multi-class classification**
- Inference

No duplicate argument classes $\forall i, \sum_{y\in\mathcal{Y}} 1_{\{y_i=y\}} = 1$

Unique labels

$$\forall y\in\mathcal{Y}, \sum_{i=0}^{n-1} 1_{\{y_i=y\}} \le 1$$

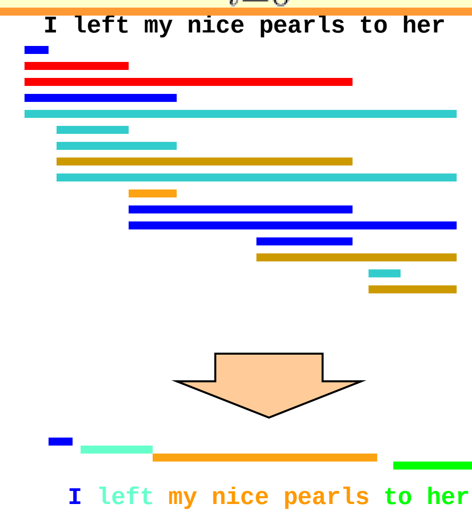
$$\forall y\in\mathcal{Y}_R, \sum_{i=0}^{n-1} 1_{\{y_i=y=\text{“R-Ax”}\}} \le \sum_{i=0}^{n-1} 1_{\{y_i=\text{“Ax”}\}}$$

$$\forall j, y\in\mathcal{Y}_C, 1_{\{y_j=y=\text{“C-Ax”}\}} \le \sum_{i=0}^{j} 1_{\{y_i=\text{“Ax”}\}}$$

argmax $\sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}$

Subject to:

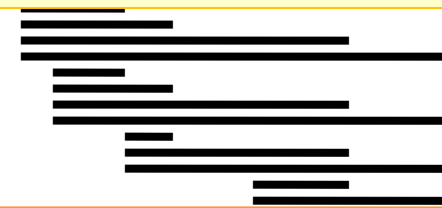
- One label per argument: $\sum_{t} y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,....



Algorithmic Approach

Learning Based Java: allows a developer to encode constraints in First Order Logic; these are compiled into linear inequalities automatically.

- **Identify** argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- **Classify** argument candidates
  - Argument Classifier
    - **Multi-class classification**
- **Inference**

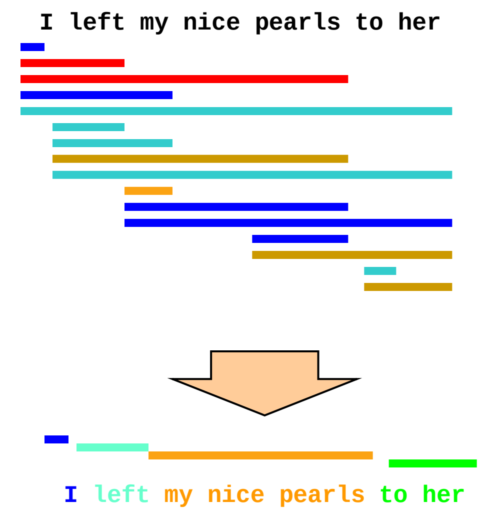


Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

argmax $\sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}$

Subject to:

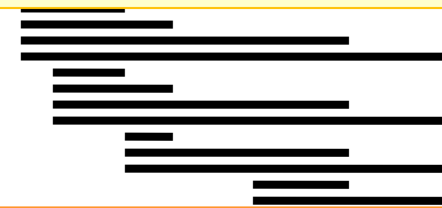
- One label per argument: $\sum_{t} y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,….



Algorithmic Approach

Learning Based Java: allows a developer to encode constraints in First Order Logic; these are compiled into linear inequalities automatically.

- **Identify** argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- **Classify** argument candidates
  - Argument Classifier
    - **Multi-class classification**
- **Inference**

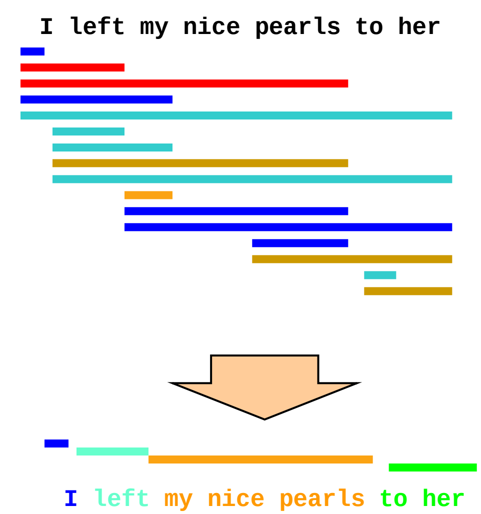


Variable $y^{a,t}$ indicates whether candidate argument a is assigned a label t.
$c^{a,t}$ is the corresponding model score

$$\text{argmax} \sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}$$

Subject to:

- One label per argument: $\sum_t y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,….



Use the **pipeline architecture's simplicity** while **maintaining uncertainty**: keep probability distributions over decisions & use global inference at decision time.

Verb SRL is not Sufficient

- *John, a fast-rising politician, slept on the train to Chicago.*
- **Verb Predicate: sleep**

Verb SRL is not Sufficient

- *John, a fast-rising politician, slept on the train to Chicago*.
- **Verb Predicate: sleep**
  - Sleeper: John, a fast-rising politician
  - Location: on the train to Chicago

Verb SRL is not Sufficient

- *John, a fast-rising politician, slept on the train to Chicago.*
- **Verb Predicate: sleep**
  - Sleeper: John, a fast-rising politician
  - Location: on the train to Chicago
- **Who was John?**

Verb SRL is not Sufficient

- *John, a fast-rising politician, slept on the train to Chicago.*
- **Verb Predicate: sleep**
  - Sleeper: John, a fast-rising politician
  - Location: on the train to Chicago
- **Who was John?**
  - Relation: Apposition (comma)
  - John, a fast-rising politician

Verb SRL is not Sufficient

- *John, a fast-rising politician, slept on the train to Chicago.*
- **Verb Predicate: sleep**
  - Sleeper: John, a fast-rising politician
  - Location: on the train to Chicago
- **Who was John?**
  - Relation: Apposition (comma)
  - John, a fast-rising politician
- **What was John's destination?**

Verb SRL is not Sufficient

- *John, a fast-rising politician, slept on the train to Chicago.*
- **Verb Predicate: sleep**
  - Sleeper: John, a fast-rising politician
  - Location: on the train to Chicago
- **Who was John?**
  - Relation: Apposition (comma)
  - John, a fast-rising politician
- **What was John's destination?**
  - Relation: Destination (preposition)
  - train to Chicago

Extended Semantic Role Labeling

- Improved sentence level analysis; dealing with more phenomena

BEIRUT, Lebanon — Lebanon's main opposition group called for widespread protests on Sunday in the wake of a powerful bomb attack for which it blamed Syria, posing a challenge to a shaky coalition government that is led by pro-Syrian factions and intensifying fears that Syria's civil war is spilling over into this country.

BEIRUT, Lebanon — Lebanon's main opposition group called for widespread protests on Sunday in the wake of a powerful bomb attack for which it blamed Syria, posing a challenge to a shaky coalition government that is led by pro-Syrian factions and intensifying fears that Syria's civil war is spilling over into this country.

[Beirut] is in [Lebanon].

[Lebanon] has a main opposition group.

[Lebanon's main opposition group] called for [widespread protests] [on Sunday].

There was [a powerful bomb attack].

[Lebanon's main opposition group] blamed [Syria].

[Pro-Syrian factions] lead [a shaky coalition government]

[Syria] has a [civil war].

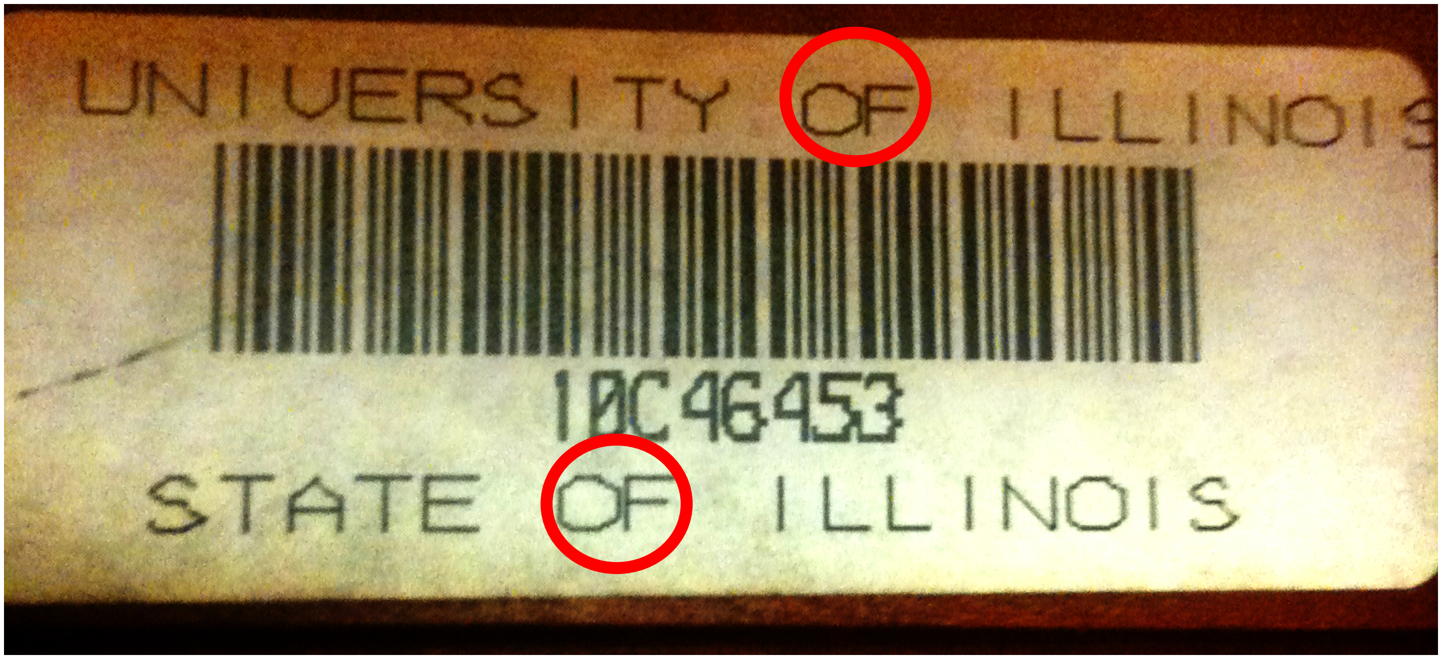
[Someone] fears that [Syria's civil war is spilling over into this country].

...

Sentence level analysis may be influenced by other sentences

Examples of Preposition Relations

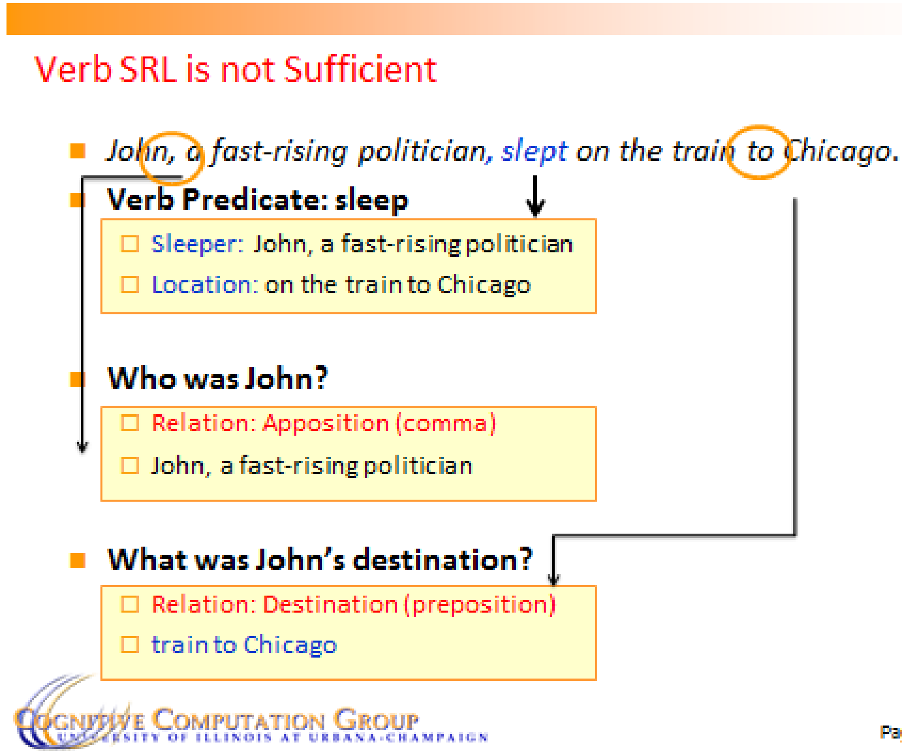
Queen of England



City of Chicago

Computational Questions

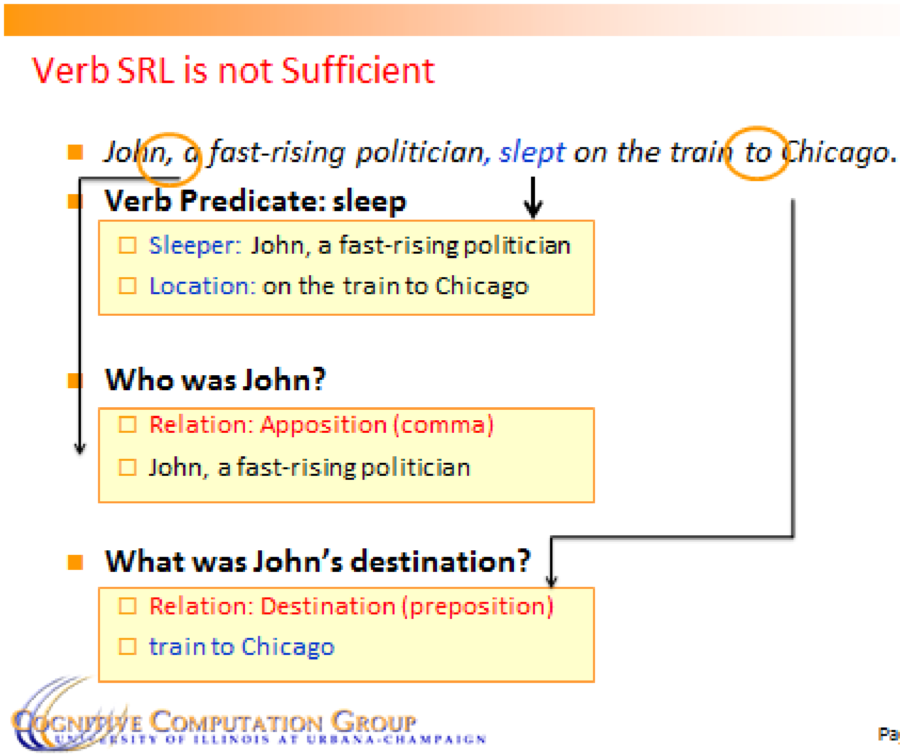
- Predict the preposition relations
  - [EMNLP, ’11]
- Identify the relation’s arguments
  - [Trans. Of ACL, ‘13]



Computational Questions

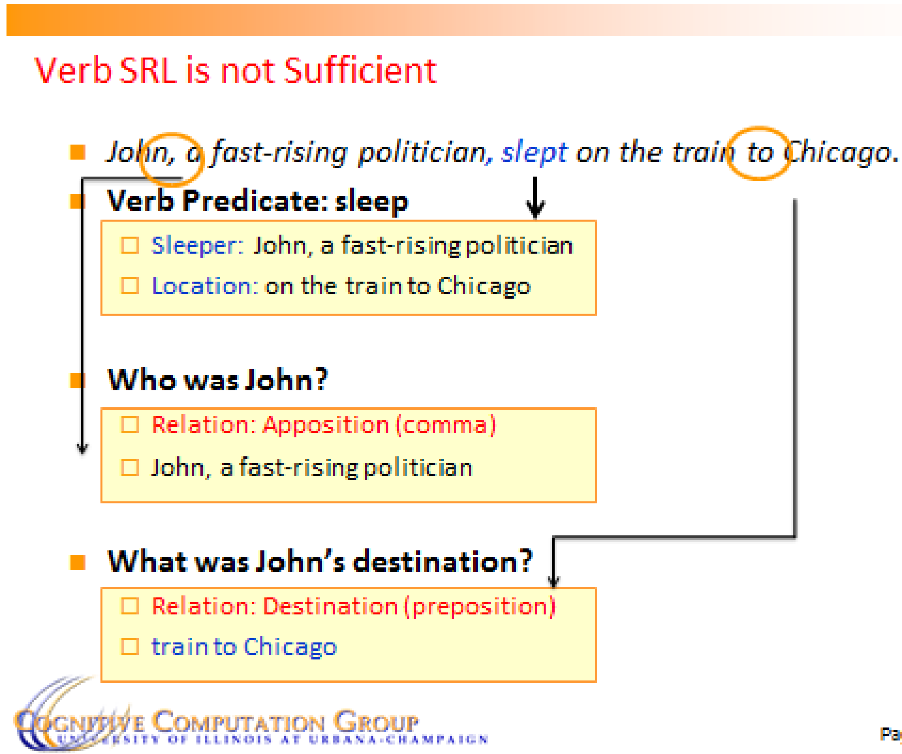
- Predict the preposition relations
  - [EMNLP, ’11]
- Identify the relation’s arguments
  - [Trans. Of ACL, ‘13]

- Very little supervised data
  - per phenomena
- Minimal annotation
  - only at the predicate level



Computational Questions

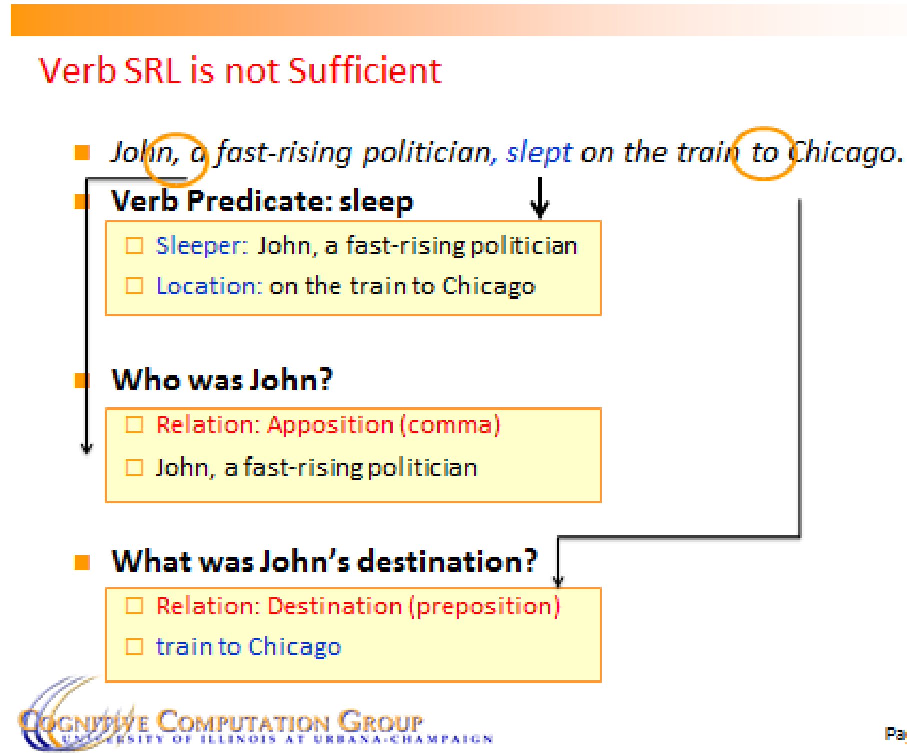
- Predict the preposition relations
  - [EMNLP, '11]
- Identify the relation's arguments
  - [Trans. Of ACL, '13]



- Very little supervised data
  - per phenomena
- Minimal annotation
  - only at the predicate level
- The Learning & Inference paradigm exploits two principles:
  - Coherency among multiple phenomena
  - Constraining latent structures (relating observed and latent variables)

Computational Questions

- Predict the preposition relations
  - [EMNLP, ’11]
- Identify the relation’s arguments
  - [Trans. Of ACL, ‘13]



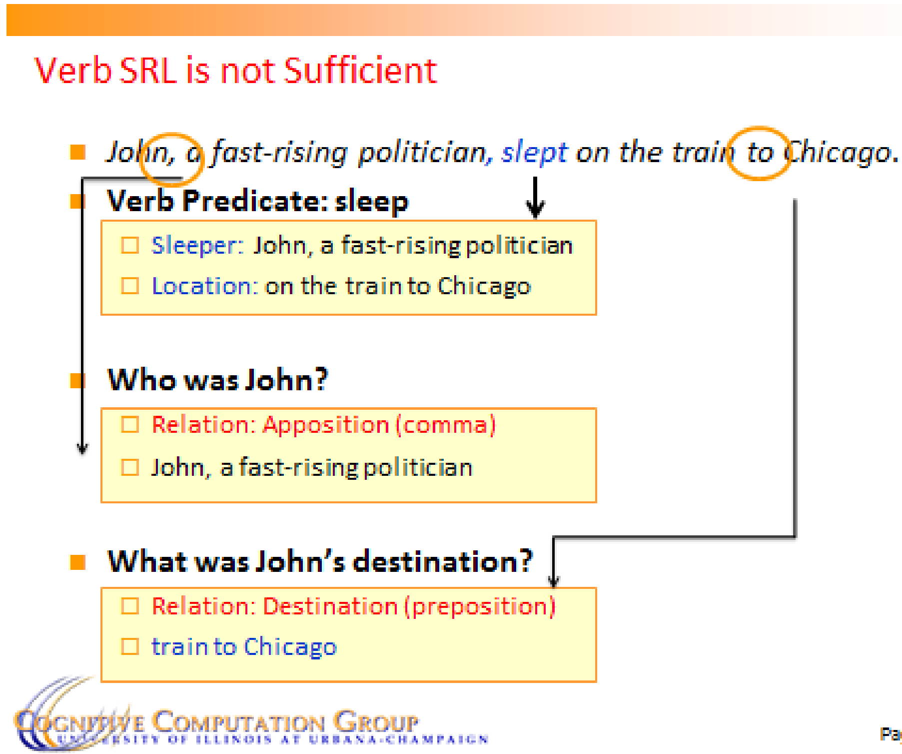
- Very little supervised data
  - per phenomena
- Minimal annotation
  - only at the predicate level
- The Learning & Inference paradigm exploits two principles:
  - Coherency among multiple phenomena
  - Constraining latent structures (relating observed and latent variables)

Input & relation

Argument & their types

Computational Questions

- Predict the preposition relations
  - [EMNLP, '11]
- Identify the relation's arguments
  - [Trans. Of ACL, '13]



- Very little supervised data
  - per phenomena
- Minimal annotation
  - only at the predicate level
- The Learning & Inference paradigm exploits two principles:
  - Coherency among multiple phenomena
  - Constraining latent structures (relating observed and latent variables)
  - Skip

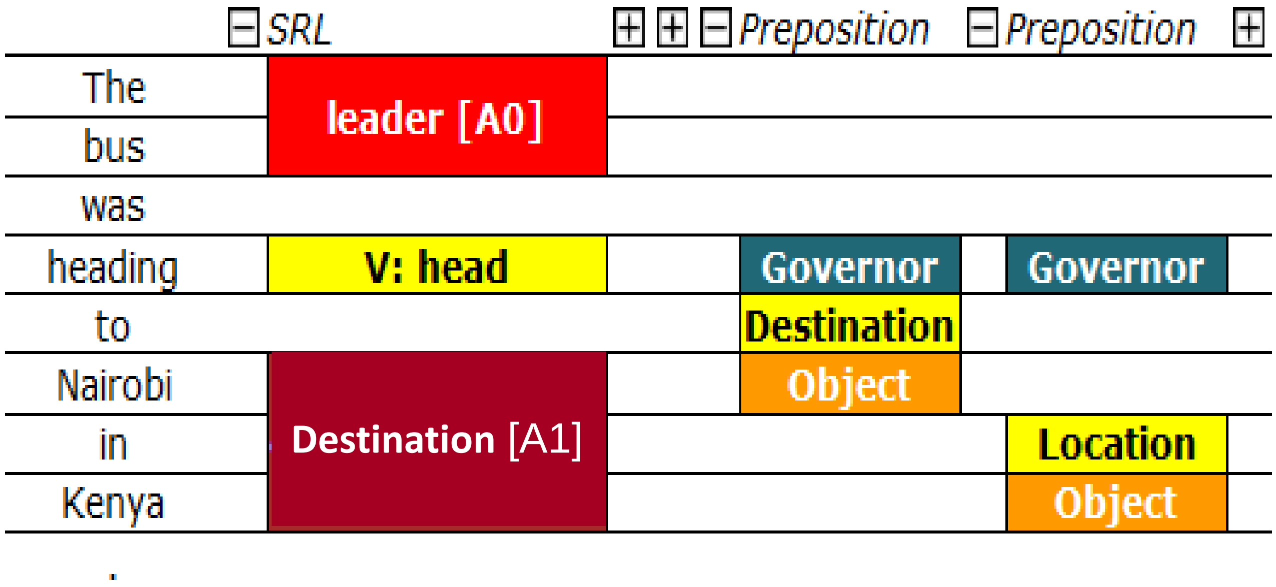
Input & relation

Argument & their types

Extended SRL [Demo]

	SRL	Preposition	Preposition
The	leader [A0]		
bus			
was			
heading	V: head	Governor	Governor
to		Destination	
Nairobi	Destination [A1]	Object	
in			Location
Kenya			Object
.			

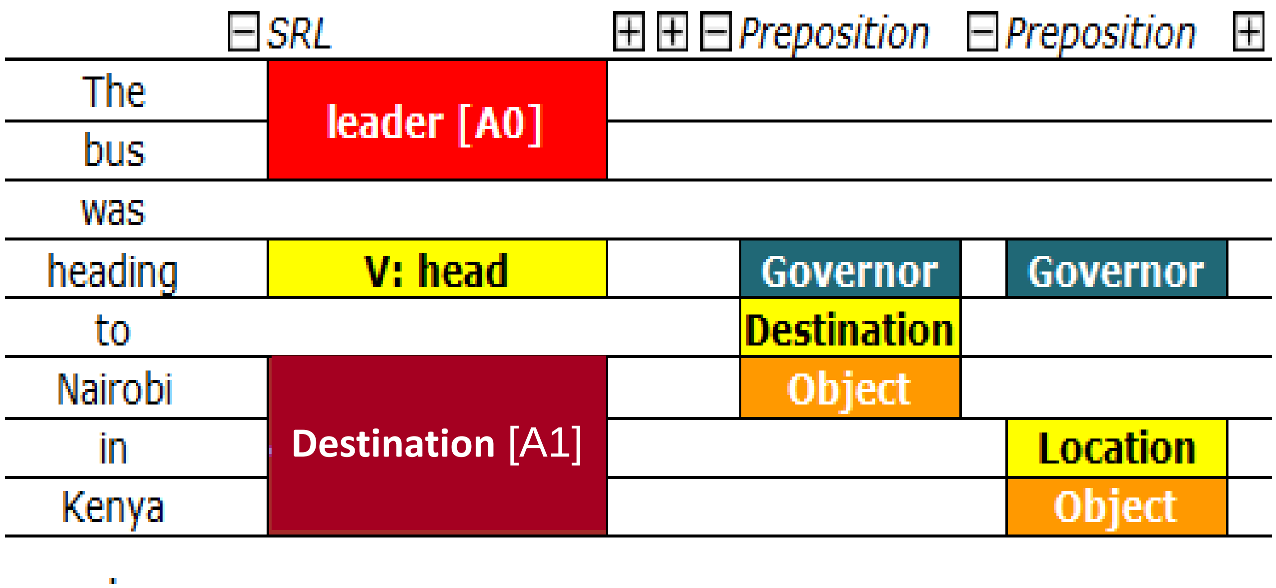
Extended SRL [Demo]



Joint inference over phenomena specific models to enforce consistency

Models trained with latent structure: senses, types, arguments

Extended SRL [Demo]



Joint inference over phenomena specific models to enforce consistency

Models trained with latent structure: senses, types, arguments

- More to do with other relations, discourse phenomena,…

Constrained Conditional Models—ILP Formulations

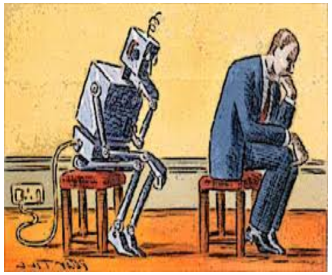
- **Have been shown useful in the context of many NLP problems**
- [Roth&Yih, 04,07: Entities and Relations; Punyakanok et. al: SRL …]
  - Summarization; Co-reference; Information & Relation Extraction; Event Identifications and causality ; Transliteration; Textual Entailment; Knowledge Acquisition; Sentiments; Temporal Reasoning, Parsing,…
- Some theoretical work on training paradigms [Punyakanok et. al., 05 more; Constraints Driven Learning, PR, Constrained EM…]
- Some work on Inference, mostly approximations, bringing back ideas on Lagrangian relaxation, etc.

Constrained Conditional Models—ILP Formulations

- **Have been shown useful in the context of many NLP problems**
- [Roth&Yih, 04,07: Entities and Relations; Punyakanok et. al: SRL …]
  - Summarization; Co-reference; Information & Relation Extraction; Event Identifications and causality ; Transliteration; Textual Entailment; Knowledge Acquisition; Sentiments; Temporal Reasoning, Parsing,…
- Some theoretical work on training paradigms [Punyakanok et. al., 05 more; Constraints Driven Learning, PR, Constrained EM…]
- Some work on Inference, mostly approximations, bringing back ideas on Lagrangian relaxation, etc.
- Good summary and description of training paradigms: [Chang, Ratinov & Roth, Machine Learning Journal 2012]
- **Summary of work & a bibliography: http://L2R.cs.uiuc.edu/tutorials.html**

Outline

- Knowledge and Inference
  - Combining the soft with the logical/declarative nature of Natural Language
    - **Constrained Conditional Models: A formulation for global inference with knowledge modeled as expressive structural constraints**
    - **Some examples**



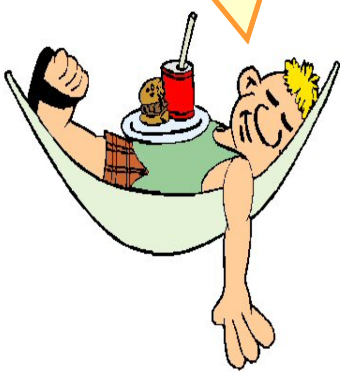
- Cycles of Knowledge
  - Grounding for/using Knowledge

- Learning with Indirect Supervision
  - Response Based Learning: learning from the world's feedback

- Scaling Up: Amortized Inference
  - Can the k-th inference problem be cheaper than the 1st?

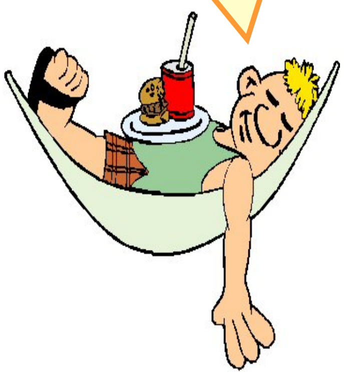
Understanding Language Requires Supervision

Can I get a coffee with lots of sugar and no milk



Understanding Language Requires Supervision

Can I get a coffee with lots of sugar and no milk

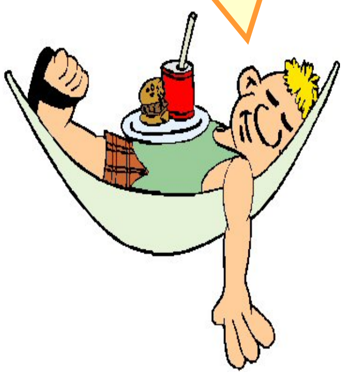


Semantic Parser

MAKE(COFFEE,SUGAR=YES,MILK=NO)

Understanding Language Requires Supervision

Can I get a coffee with lots of sugar and no milk



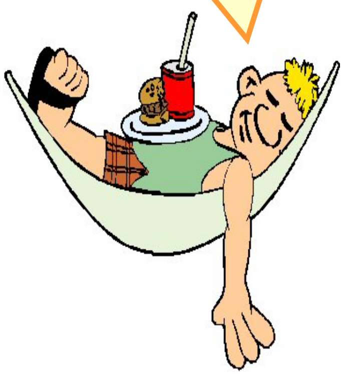
Semantic Parser

MAKE(COFFEE,SUGAR=YES,MILK=NO)

- How to recover meaning from text?

Understanding Language Requires Supervision

Can I get a coffee with lots of sugar and no milk



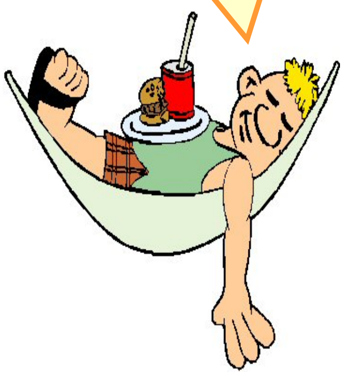
Semantic Parser

MAKE(COFFEE,SUGAR=YES,MILK=NO)

- How to recover meaning from text?
- Standard "example based" ML: annotate text with meaning representation
  - Teacher needs deep understanding of the learning agent ; not scalable.

Understanding Language Requires Supervision

Can I get a coffee with lots of sugar and no milk

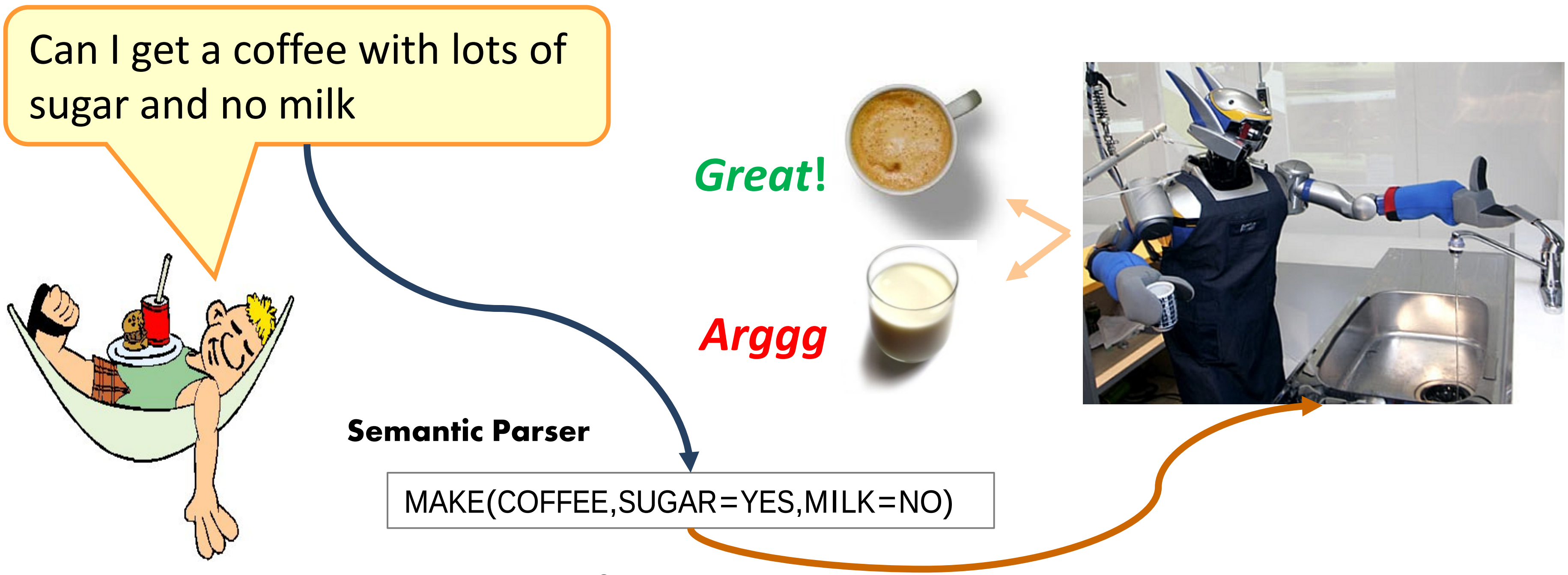


Semantic Parser

MAKE(COFFEE,SUGAR=YES,MILK=NO)

- How to recover meaning from text?
- Standard “example based” ML: annotate text with meaning representation
  - Teacher needs deep understanding of the learning agent ; not scalable.
- Response Driven Learning: Exploit indirect signals in the interaction between the learner and the teacher/environment

Understanding Language Requires Supervision

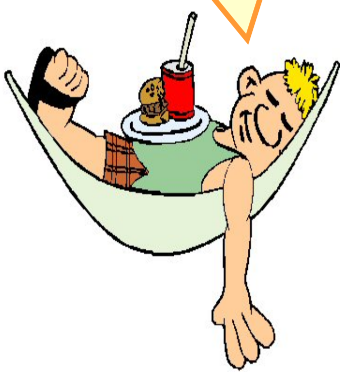


- How to recover meaning from text?
- Standard "example based" ML: annotate text with meaning representation
  - Teacher needs deep understanding of the learning agent ; not scalable.
- Response Driven Learning: Exploit indirect signals in the interaction between the learner and the teacher/environment

Understanding Language Requires Supervision

Can we rely on this interaction to provide supervision (and eventually, recover meaning) ?

Can I get a coffee with lots of sugar and no milk



Great!

Arggg

Semantic Parser

MAKE(COFFEE,SUGAR=YES,MILK=NO)

- How to recover meaning from text?
- Standard "example based" ML: annotate text with meaning representation
  - Teacher needs deep understanding of the learning agent ; not scalable.
- Response Driven Learning: Exploit indirect signals in the interaction between the learner and the teacher/environment

Response Based Learning

- We want to learn a model that transforms a natural language sentence to some meaning representation.

English Sentence → **Model** → Meaning Representation

- Instead of training with (Sentence, Meaning Representation) pairs

Response Based Learning

- We want to learn a model that transforms a natural language sentence to some meaning representation.

English Sentence → **Model** → Meaning Representation

- Instead of training with (Sentence, Meaning Representation) pairs

- Think about some simple derivatives of the models outputs,
  - Supervise the derivative [verifier] (easy!) and
  - Propagate it to learn the complex, structured, transformation model

Scenario I: Freecell with Response Based Learning

- We want to learn a model to transform a natural language sentence to some meaning representation.

English Sentence → **Model** → Meaning Representation

Scenario I: Freecell with Response Based Learning

- We want to learn a model to transform a natural language sentence to some meaning representation.

English Sentence → **Model** → Meaning Representation

A top card can be moved to the tableau if it has a different color than the color of the top tableau card, and the cards have successive values.

Scenario I: Freecell with Response Based Learning

- We want to learn a model to transform a natural language sentence to some meaning representation.

English Sentence → **Model** → Meaning Representation

A top card can be moved to the tableau if it has a different color than the color of the top tableau card, and the cards have successive values.	Move (a1,a2) top(a1,x1) card(a1) tableau(a2) top(x2,a2) color(a1,x3) color(x2,x4) not-equal(x3,x4) value(a1,x5) value(x2,x6) successor(x5,x6)

Scenario I: Freecell with Response Based Learning

- We want to learn a model to transform a natural language sentence to some meaning representation.

English Sentence → **Model** → Meaning Representation

A top card can be moved to the tableau if it has a different color than the color of the top tableau card, and the cards have successive values.

Move (a1,a2) top(a1,x1) card(a1) tableau(a2) top(x2,a2) color(a1,x3) color(x2,x4) not-equal(x3,x4) value(a1,x5) value(x2,x6) successor(x5,x6)

- Simple derivatives of the models outputs
  - Supervise the derivative and
  - Propagate it to learn the transformation model

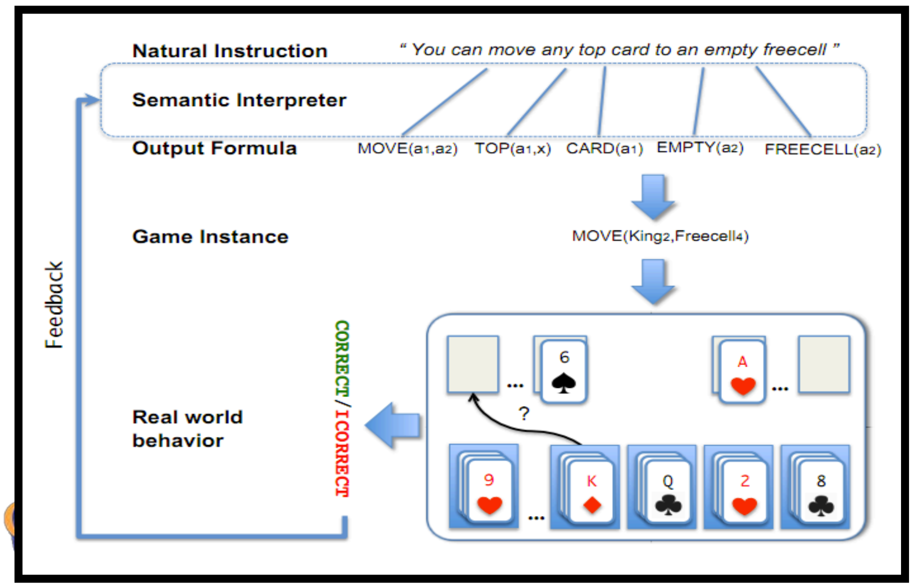
Scenario I: Freecell with Response Based Learning

- We want to learn a model to transform a natural language sentence to some meaning representation.

English Sentence → **Model** → Meaning Representation

A top card can be moved to the tableau if it has a different color than the color of the top tableau card, and the cards have successive values.

Move (a1,a2) top(a1,x1) card(a1) tableau(a2) top(x2,a2) color(a1,x3) color(x2,x4) not-equal(x3,x4) value(a1,x5) value(x2,x6) successor(x5,x6)



← Play Freecell (solitaire)

- Simple derivatives of the models outputs
  - Supervise the derivative and
  - Propagate it to learn the transformation model

Scenario II: Geoquery with Response based Learning

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

Scenario II: Geoquery with Response based Learning

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

What is the largest state that borders NY?

Scenario II: Geoquery with Response based Learning

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

What is the largest state that borders NY? → largest(state(next_to(const(NY))))

Scenario II: Geoquery with Response based Learning

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

What is the largest state that borders NY?

largest(state(next_to(const(NY))))

- Query a GeoQuery Database. ←
- Simple derivatives of the models outputs

Scenario II: Geoquery with Response based Learning

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

What is the largest state that borders NY?

largest(state(next_to(const(NY))))

- Query a GeoQuery Database. ← - Simple derivatives of the models outputs

- "Guess" a semantic parse. **Is [DB response == Expected response] ?**
  - **Expected**: Pennsylvania **DB Returns**: Pennsylvania → **Positive Response**
  - **Expected**: Pennsylvania **DB Returns**: NYC, or ???? → **Negative Response**

Response Based Learning: Using a Simple Feedback

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

- Instead of training with (Sentence, Meaning Representation) pairs
- Think about some simple derivatives of the models outputs,
  - Supervise the derivative (easy!) and
  - Propagate it to learn the complex, structured, transformation model

Response Based Learning: Using a Simple Feedback

- We want to learn a model to transform a natural language sentence to some formal representation.

English Sentence → **Model** → Meaning Representation

- Instead of training with (Sentence, Meaning Representation) pairs
- Think about some simple derivatives of the models outputs,
  - Supervise the derivative (easy!) and
  - Propagate it to learn the complex, structured, transformation model

LEARNING:

- Train a structured predictor (semantic parse) with this binary supervision
  - Many challenges: e.g., how to make a better use of a negative response?
- Learning with a constrained latent representation, making used of CCM inference, exploiting knowledge on the structure of the meaning representation.

Geoquery: Response based Competitive with Supervised

Clarke, Goldwasser, Chang, Roth CoNLL'10; Goldwasser, Roth IJCAI'11, MLJ'14

Algorithm	Training Accuracy	Testing Accuracy	# Training Examples
NoLearn	22	--	-
Supervised	--	86.07	**600 structs.**

NoLearn :Initialization point

Supervised : Trained with annotated data

Response based Learning is gathering momentum:

- Liang, M.I. Jordan, D. Klein, Learning Dependency-Based Compositional Semantics, ACL'11.
- Berant et-al ' Semantic Parsing on Freebase from Question-Answer Pairs, EMNLP'13

Supervised: Y.-W. Wong and R. Mooney. Learning synchronous grammars for semantic parsing with lambda calculus. ACL'07

Geoquery: Response based Competitive with Supervised

Clarke, Goldwasser, Chang, Roth CoNLL'10; Goldwasser, Roth IJCAI'11, MLJ'14

Algorithm	Training Accuracy	Testing Accuracy	# Training Examples
NoLearn	22	--	-
Response-based (2010)	82.4	73.2	250 answers
Liang et-al 2011	--	78.9	250 answers
Response-based (2012)	**86.8**	**81.6**	250 answers
Supervised	--	86.07	**600 structs**.

NoLearn :Initialization point

Supervised : Trained with annotated data

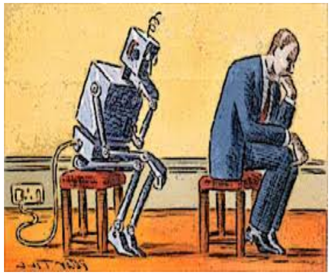
Response based Learning is gathering momentum:

- Liang, M.I. Jordan, D. Klein, Learning Dependency-Based Compositional Semantics, ACL'11.
- Berant et-al ' Semantic Parsing on Freebase from Question-Answer Pairs, EMNLP'13

Supervised: Y.-W. Wong and R. Mooney. Learning synchronous grammars for semantic parsing with lambda calculus. ACL'07

Outline

- Knowledge and Inference
  - Combining the soft with the logical/declarative nature of Natural Language
    - **Constrained Conditional Models: A formulation for global inference with knowledge modeled as expressive structural constraints**
    - **Some examples**



- Cycles of Knowledge
  - Grounding for/using Knowledge

- Learning with Indirect Supervision
  - Response Based Learning: learning from the world's feedback
- Scaling Up: Amortized Inference
  - Can the k-th inference problem be cheaper than the 1st?

Amortized ILP Structured Output Inference

- Imagine that you already solved many structured output inference problems
  - Co-reference resolution; Semantic Role Labeling; Parsing citations; Summarization; dependency parsing; image segmentation,…
  - Your solution method doesn’t matter either

Amortized ILP Structured Output Inference

- Imagine that you already solved many structured output inference problems
  - Co-reference resolution; Semantic Role Labeling; Parsing citations; Summarization; dependency parsing; image segmentation,…
  - Your solution method doesn't matter either
- **How can we exploit this fact to save inference cost**?

After solving **n** inference problems, can we make the (**n+1**)th one faster?

Amortized ILP Structured Output Inference

- Imagine that you already solved many structured output inference problems
  - Co-reference resolution; Semantic Role Labeling; Parsing citations; Summarization; dependency parsing; image segmentation,…
  - Your solution method doesn't matter either
- **How can we exploit this fact to save inference cost**?

> After solving **n** inference problems, can we make the (**n+1**)th one faster?

- We will show how to do it when your problem is formulated as a 0-1 LP, Max **cx**

$$A\mathbf{x} \leq \mathbf{b}$$

Amortized ILP Structured Output Inference

- Imagine that you already solved many structured output inference problems
  - Co-reference resolution; Semantic Role Labeling; Parsing citations; Summarization; dependency parsing; image segmentation,…
  - Your solution method doesn't matter either
- **How can we exploit this fact to save inference cost**?

> After solving **n** inference problems, can we make the (**n+1**)th one faster?

- We will show how to do it when your problem is formulated as a 0-1 LP, Max $\mathbf{cx}$

$$A\mathbf{x} \leq \mathbf{b}$$

> - Very general: All discrete MAP problems can be formulated as 0-1 LPs
> - We only care about inference formulation, not algorithmic solution

Inference for BIG TEXT

- In NLP, we typically don’t solve a single inference problem.
- We solve one or more per sentence.
- Beyond improving the inference algorithm, what can be done?

Inference for BIG TEXT

- In NLP, we typically don't solve a single inference problem.
- We solve one or more per sentence.
- Beyond improving the inference algorithm, what can be done?

S1	S2	POS
He	She	PRP
is	is	VBZ
reading	watching	VBG
a	a	DT
book	movie	NN

S1 & S2 look very different but their output structures are the same

The inference outcomes are the same

Inference for BIG TEXT

- In NLP, we typically don't solve a single inference problem.
- We solve one or more per sentence.
- Beyond improving the inference algorithm, what can be done?

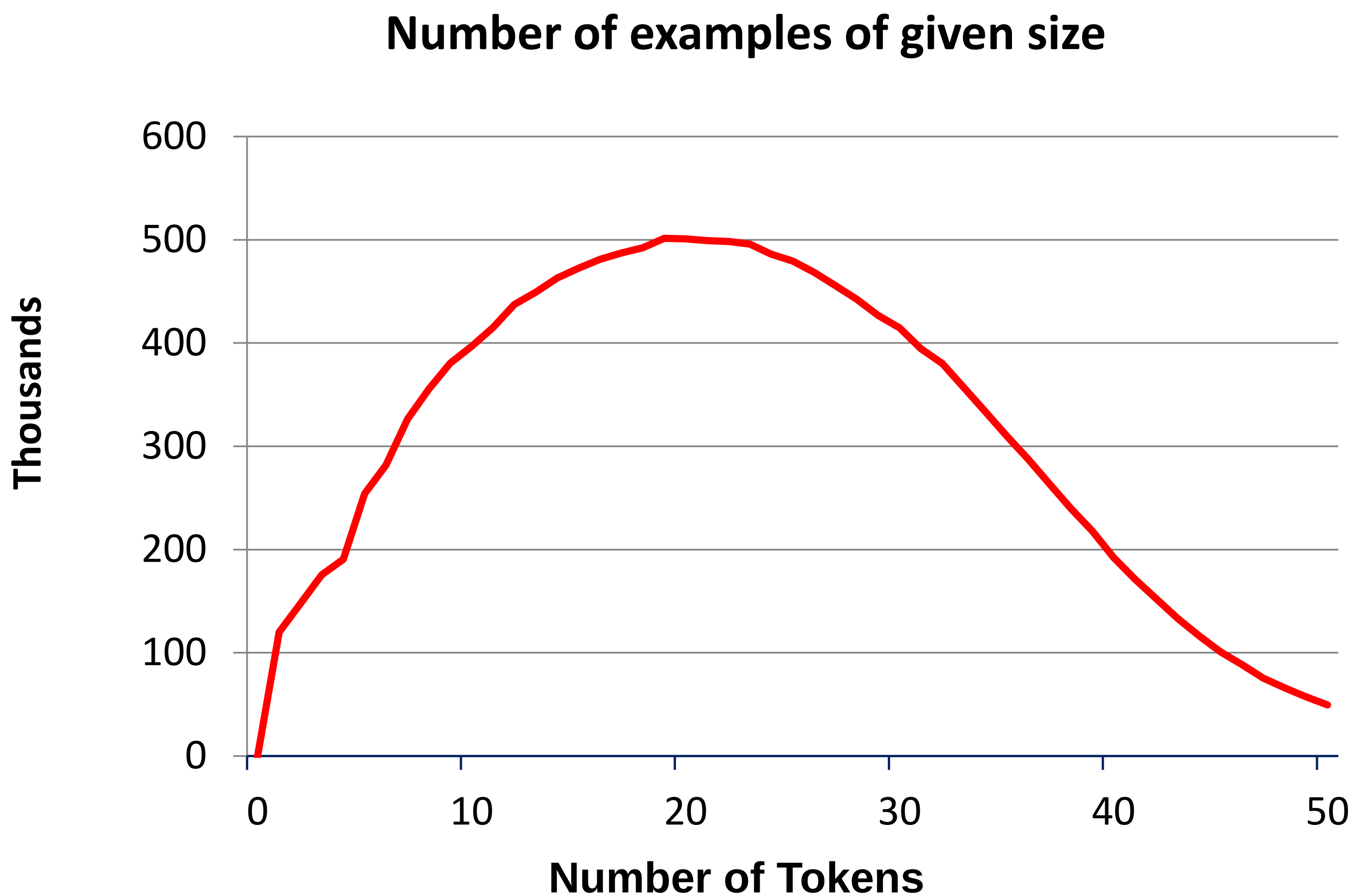
S1	S2	POS
He	She	PRP
is	is	VBZ
reading	watching	VBG
a	a	DT
book	movie	NN

S1 & S2 look very different but their output structures are the same

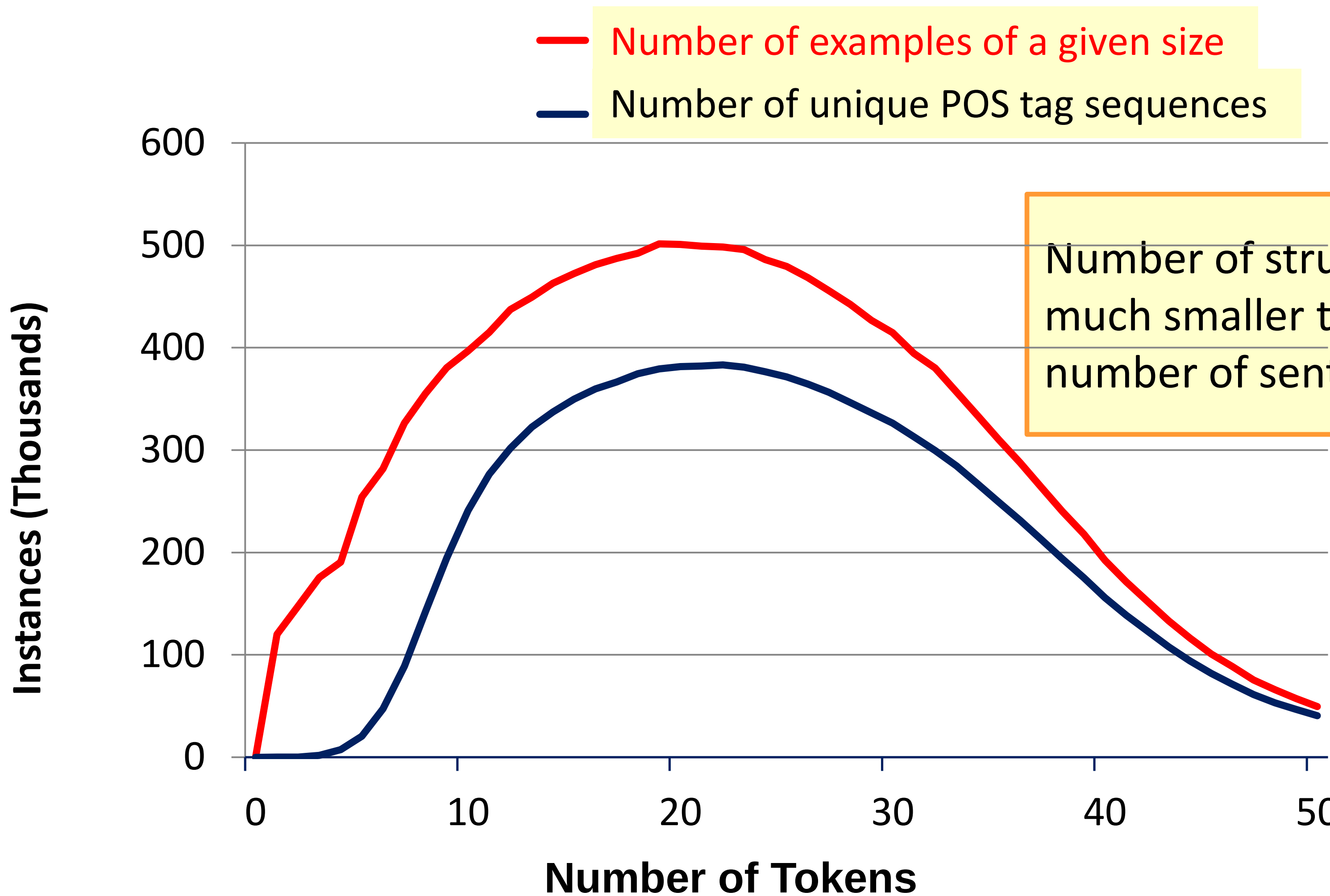
The inference outcomes are the same

After inferring the POS structure for S1,
Can we speed up inference for S2 ?

The Hope: POS Tagging on Gigaword

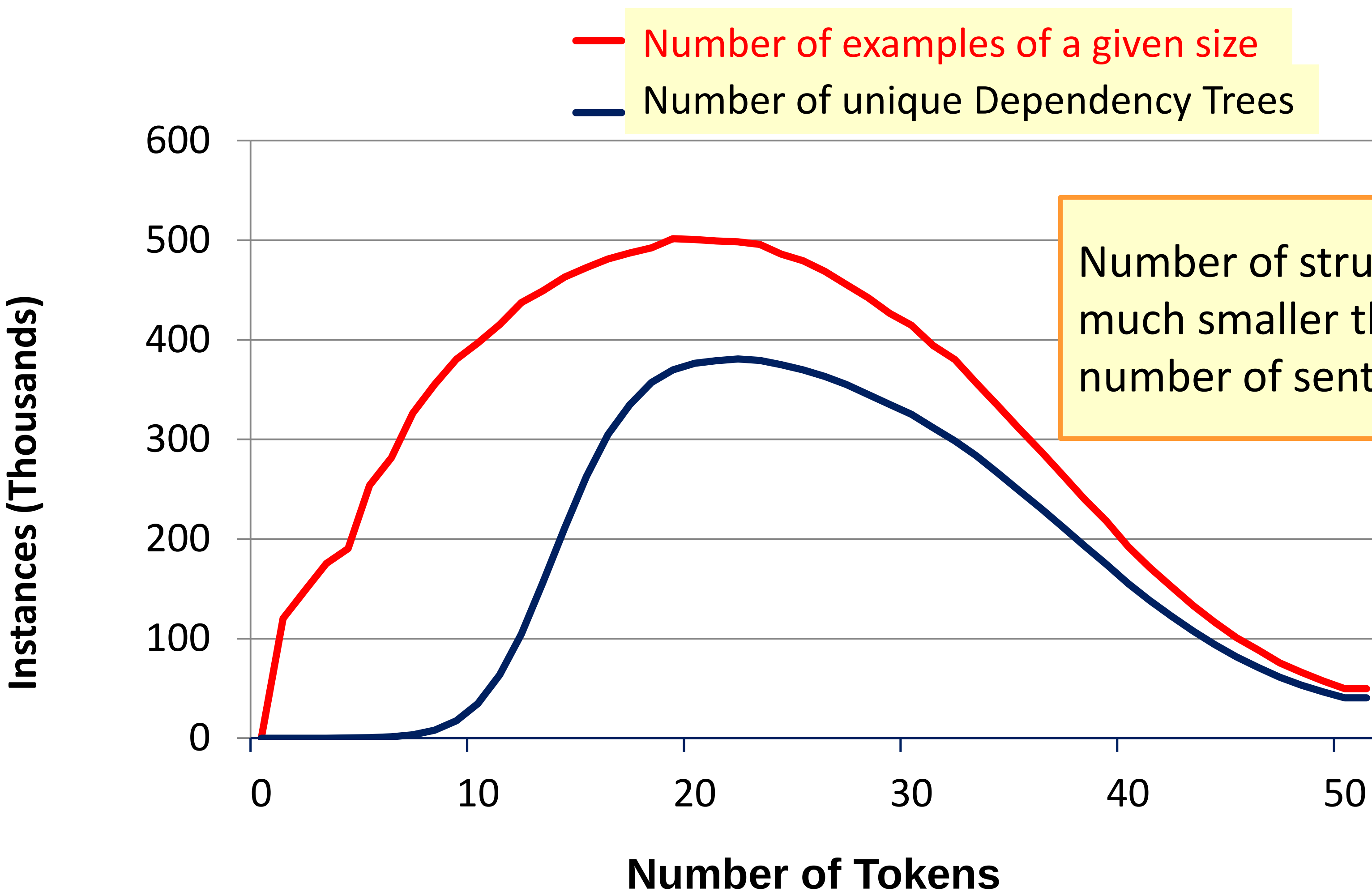


The Hope: POS Tagging on Gigaword



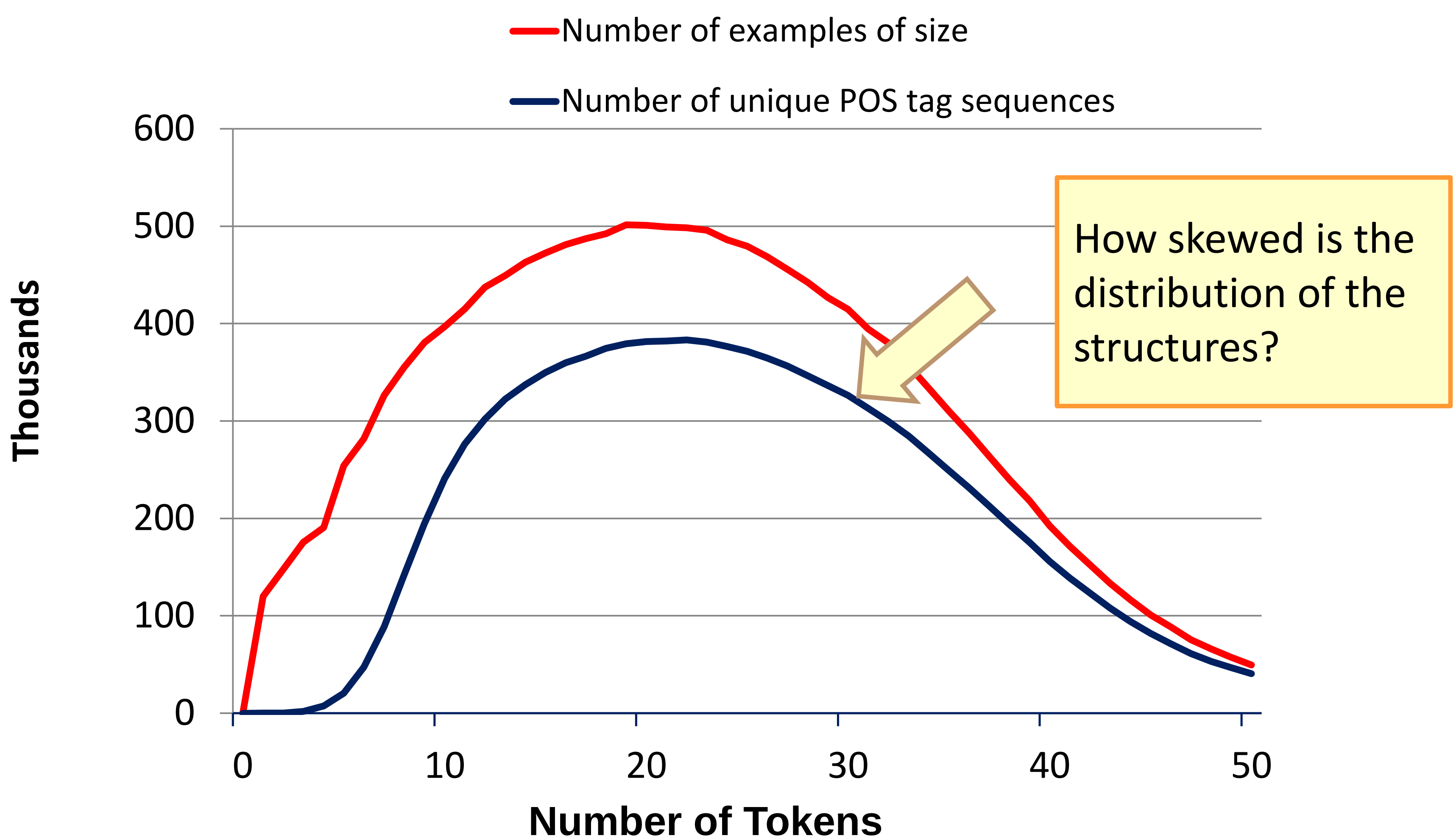
Number of structures is much smaller than the number of sentences

The Hope: Dependency Parsing on Gigaword

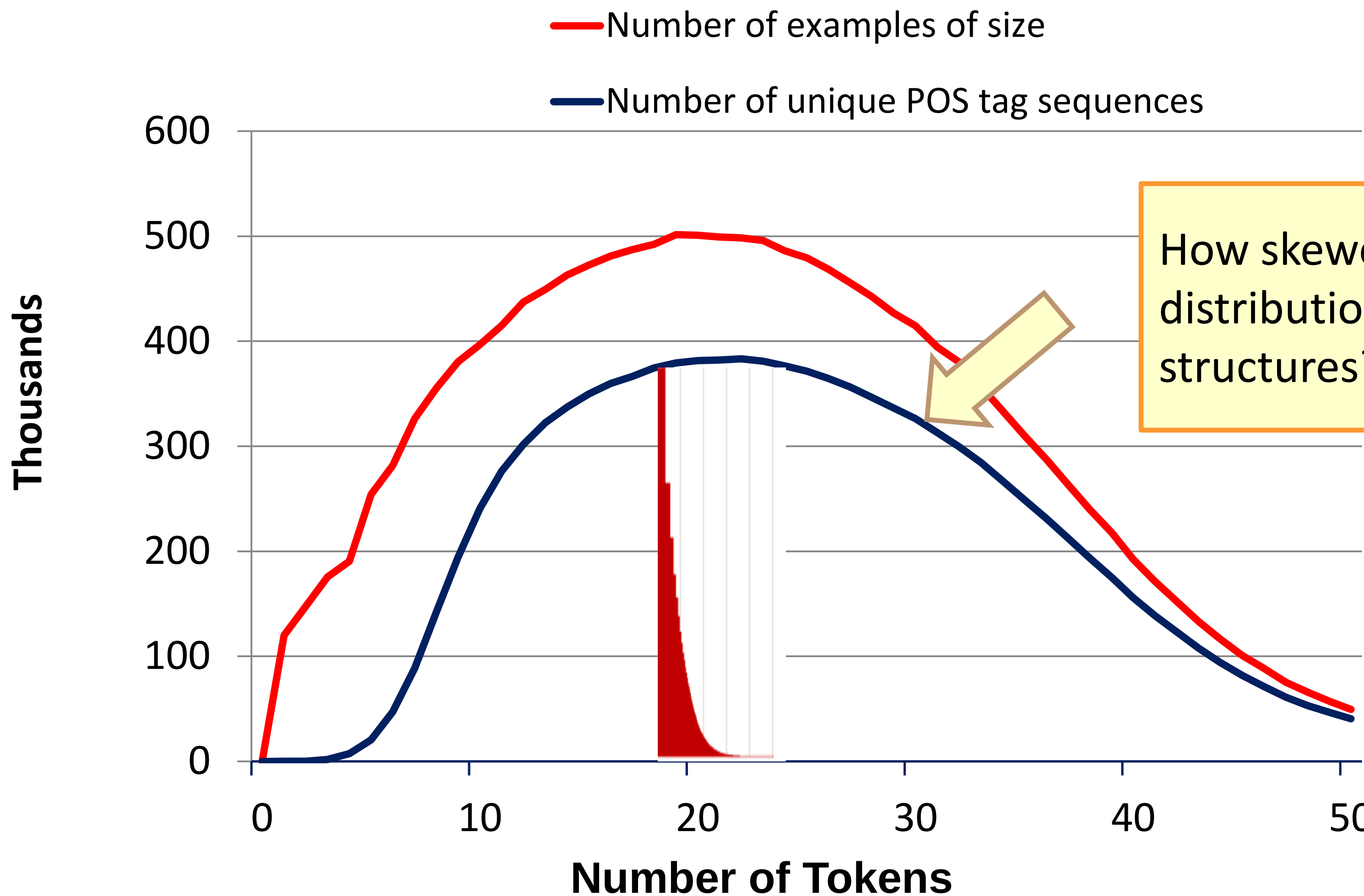


Number of structures is much smaller than the number of sentences

POS Tagging on Gigaword

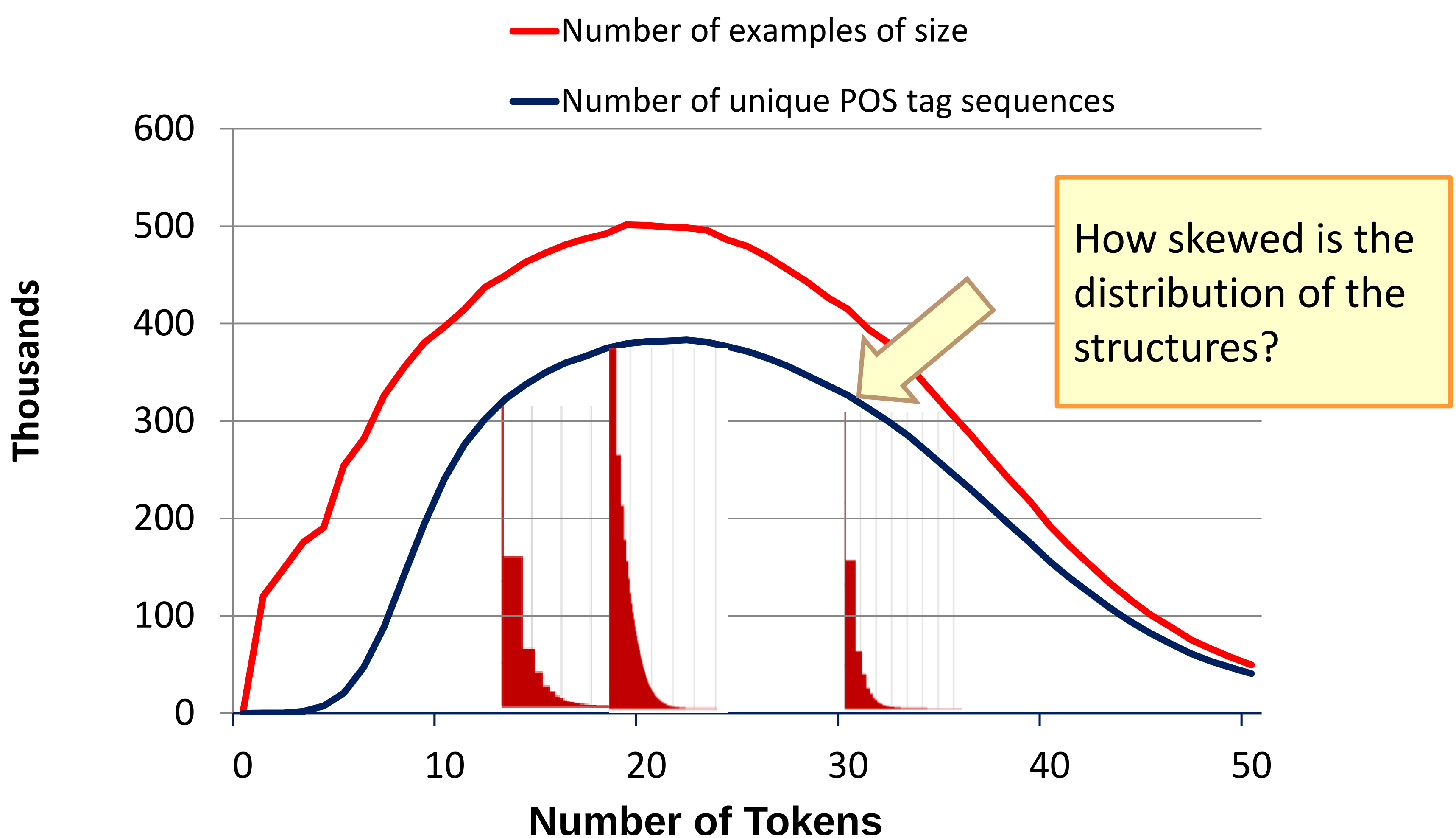


POS Tagging on Gigaword

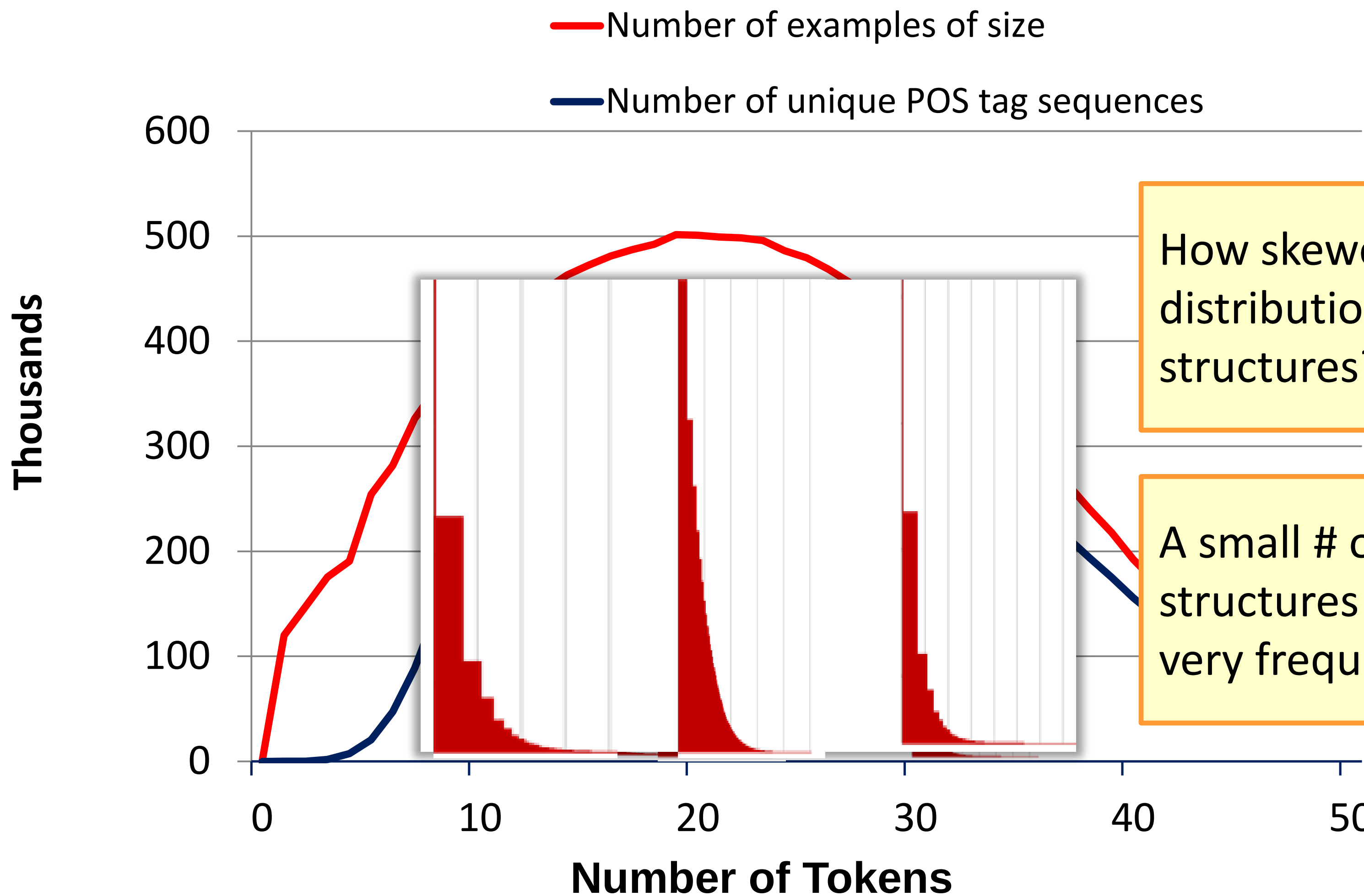


How skewed is the distribution of the structures?

POS Tagging on Gigaword



POS Tagging on Gigaword



How skewed is the distribution of the structures?

A small # of structures occur very frequently

Amortized ILP Inference

- These statistics show that many different instances are mapped into identical inference outcomes.
  - Pigeon Hole Principle
- **How can we exploit this fact to save inference cost over the life time of the agent?** ?

Amortized ILP Inference

- These statistics show that many different instances are mapped into identical inference outcomes.
  - Pigeon Hole Principle
- **How can we exploit this fact to save inference cost over the life time of the agent?** ?

We give conditions on the objective functions
(for all objectives with the same # or variables and same feasible set),
under which the solution of a new problem Q is the same as the one of P (which we already cached)

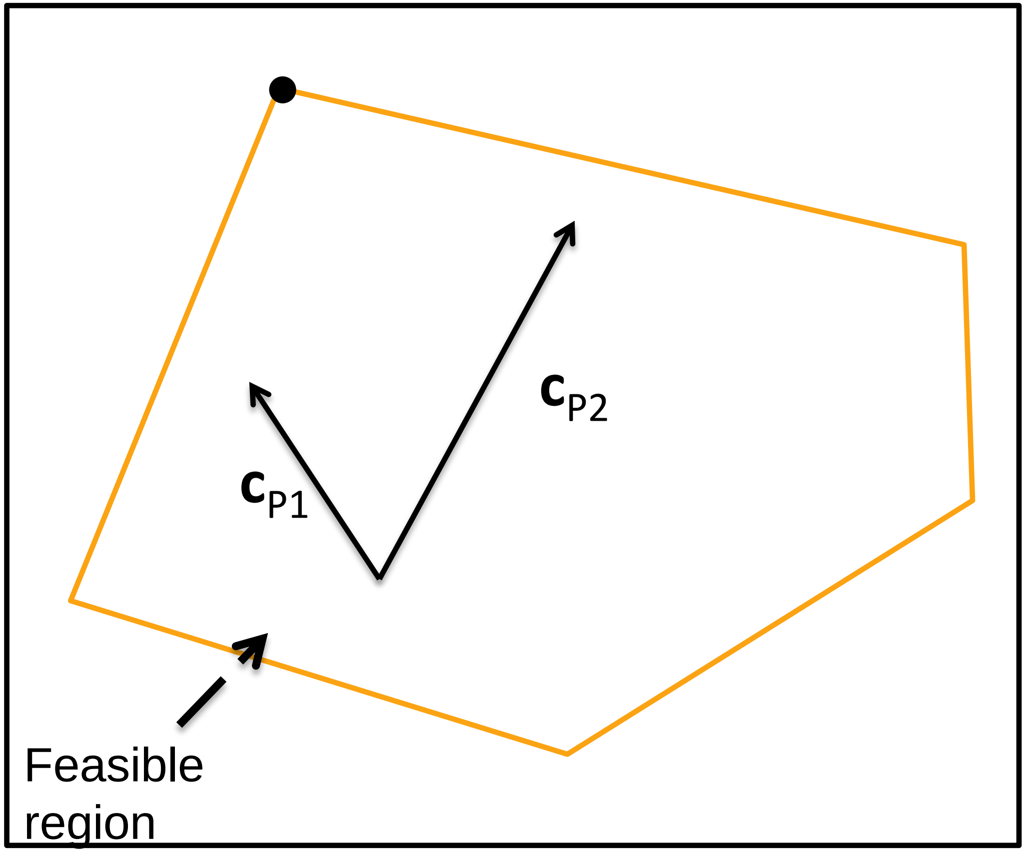
Amortized ILP Inference

We argue here that the inference formulation provides a new level of abstraction.

- These statistics show that many different instances are mapped into identical inference outcomes.
  - Pigeon Hole Principle
- **How can we exploit this fact to save inference cost over the life time of the agent?** ?

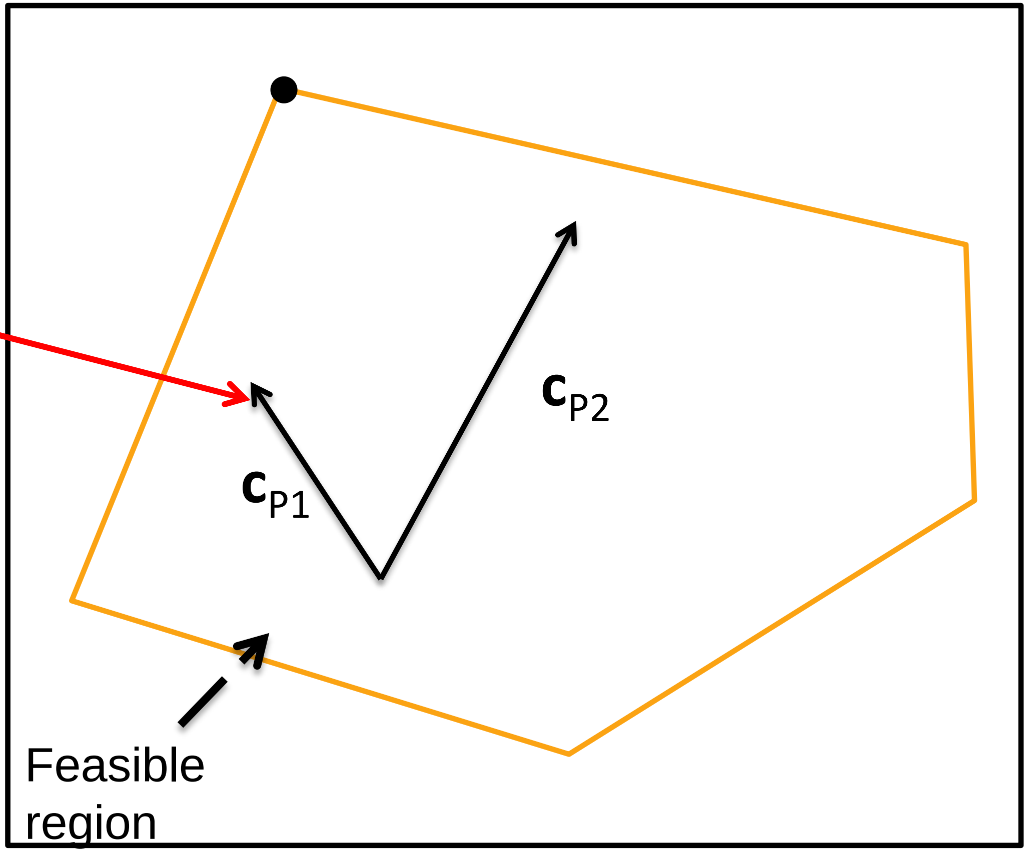
We give conditions on the objective functions
(for all objectives with the same # or variables and same feasible set),
under which the solution of a new problem Q is the same as the one of P (which we already cached)

Theorem II (Geometric Interpretation)



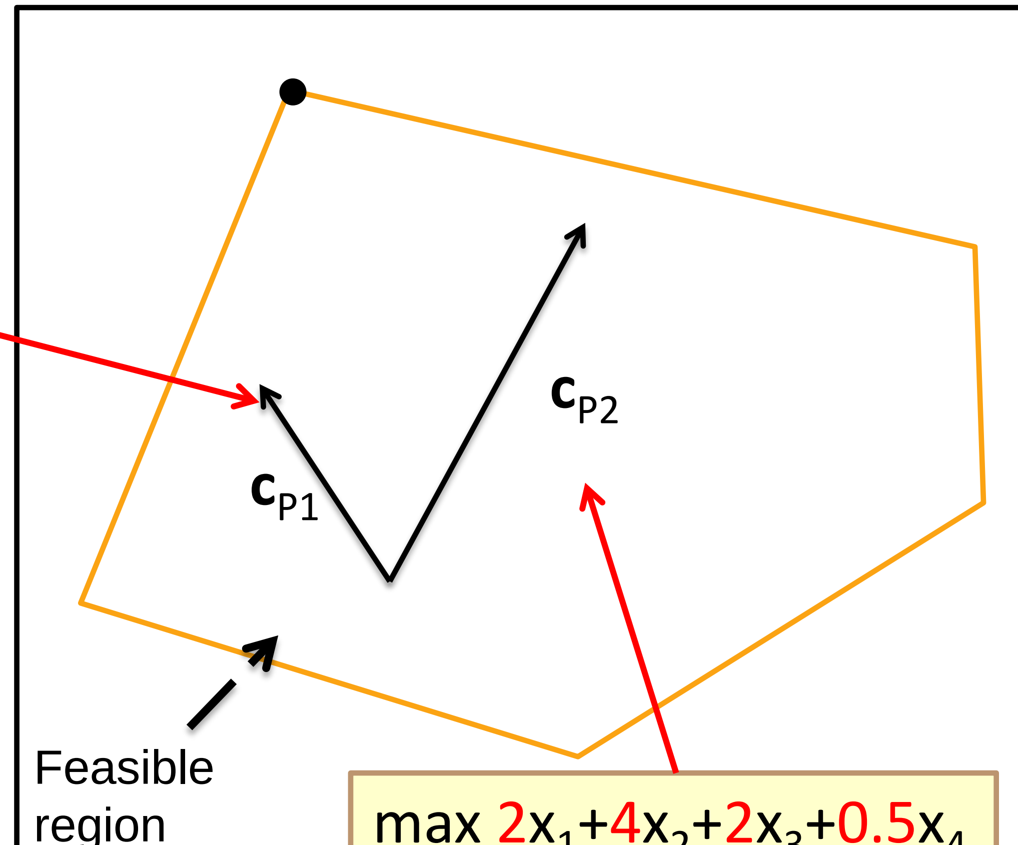
Theorem II (Geometric Interpretation)

$$\max\ 2x_1+3x_2+2x_3+1x_4$$
$$x_1 + x_2 \leq 1$$
$$x_3 + x_4 \leq 1$$



Theorem II (Geometric Interpretation)

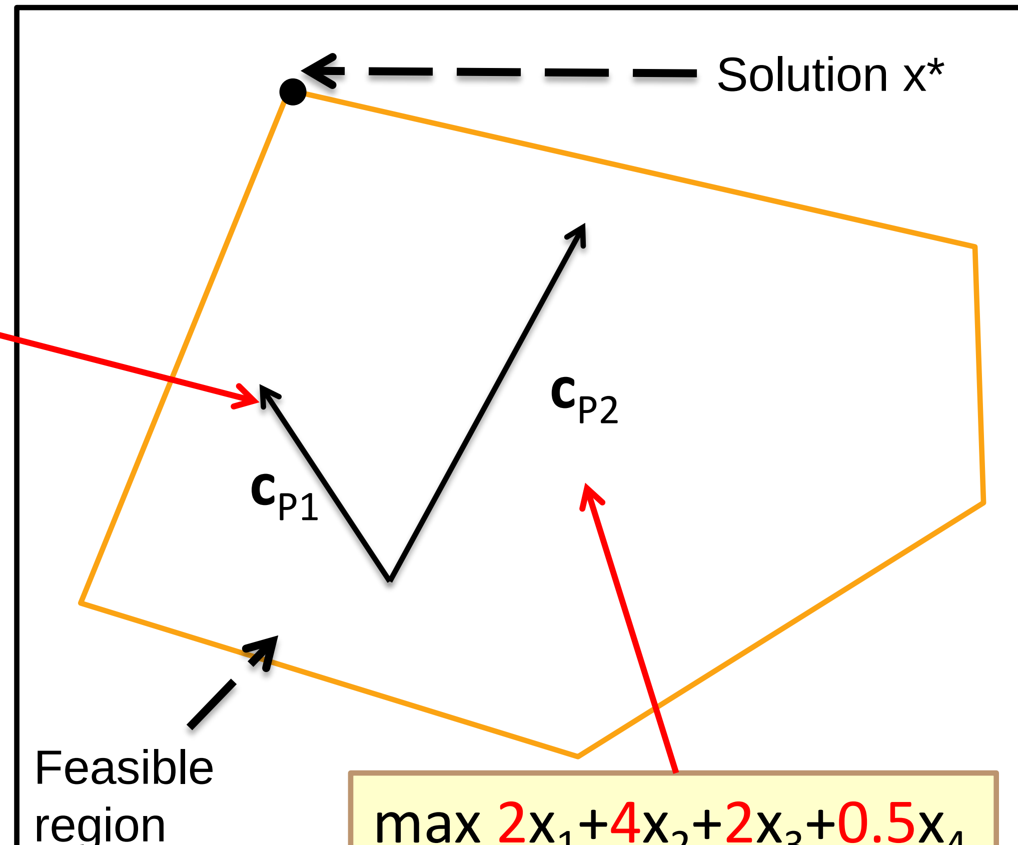
$$\max\ 2x_1+3x_2+2x_3+1x_4$$
$$x_1 + x_2 \leq 1$$
$$x_3 + x_4 \leq 1$$



$$\max\ 2x_1+4x_2+2x_3+0.5x_4$$
$$x_1 + x_2 \leq 1$$
$$x_3 + x_4 \leq 1$$

Theorem II (Geometric Interpretation)

$$\max\ 2x_1+3x_2+2x_3+1x_4$$
$$x_1 + x_2 \le 1$$
$$x_3 + x_4 \le 1$$

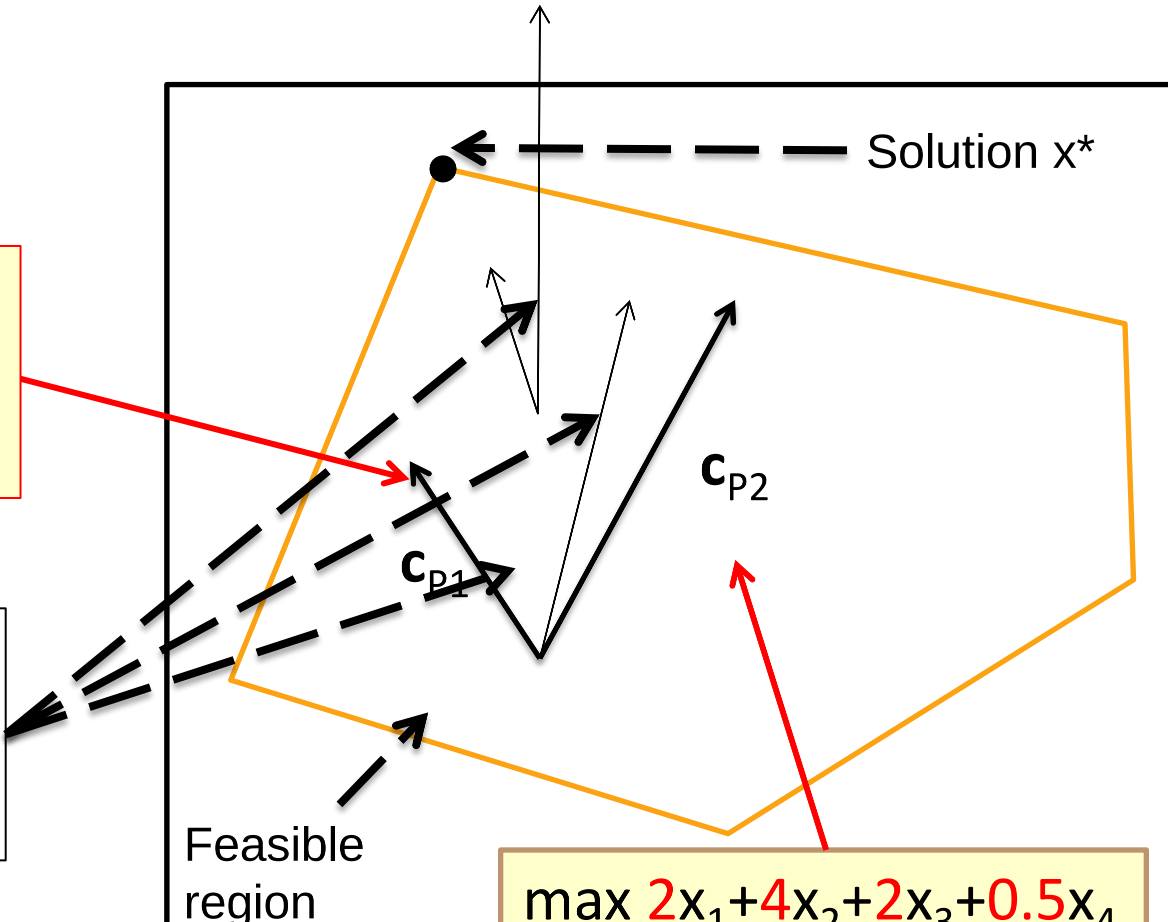


$$\max\ 2x_1+4x_2+2x_3+0.5x_4$$
$$x_1 + x_2 \le 1$$
$$x_3 + x_4 \le 1$$

Theorem II (Geometric Interpretation)

$$\max\ 2x_1+3x_2+2x_3+1x_4$$
$$x_1 + x_2 \leq 1$$
$$x_3 + x_4 \leq 1$$

ILPs corresponding to all these objective vectors will share the same maximizer *for this feasible region*

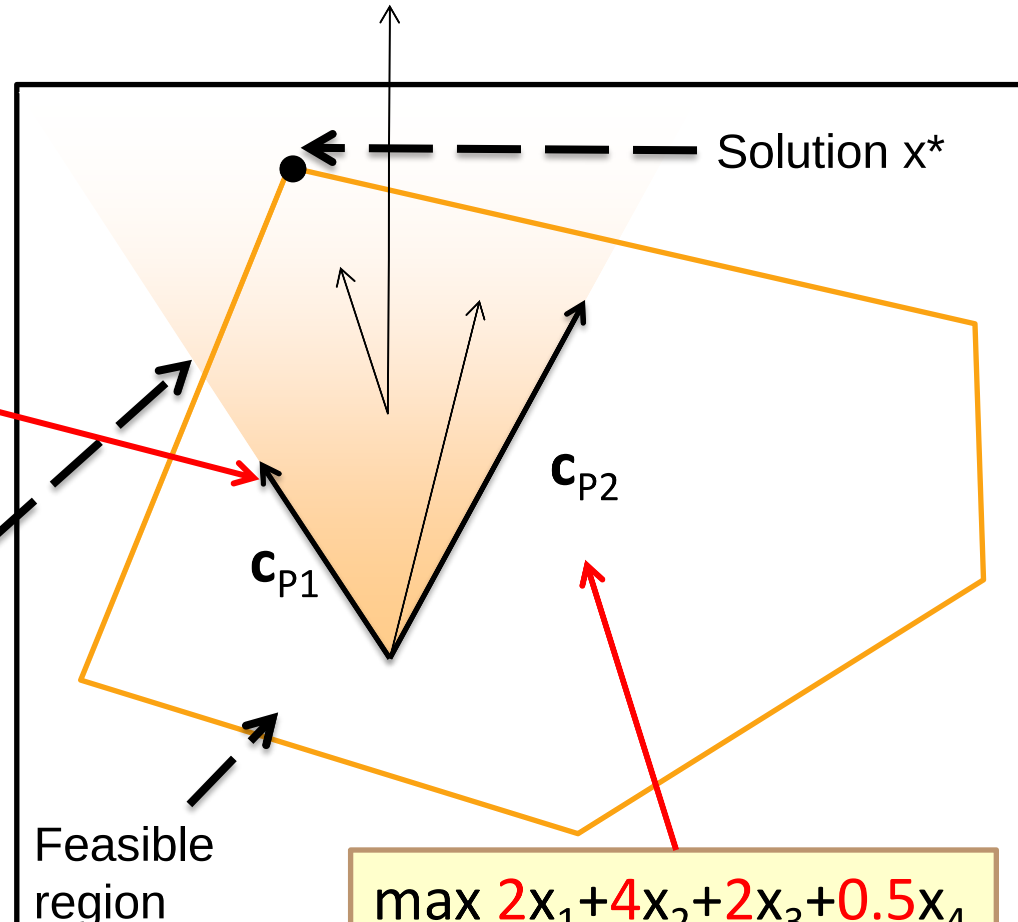


$$\max\ 2x_1+4x_2+2x_3+0.5x_4$$
$$x_1 + x_2 \leq 1$$
$$x_3 + x_4 \leq 1$$

Theorem II (Geometric Interpretation)

$$\max \; 2x_1 + 3x_2 + 2x_3 + 1x_4$$
$$x_1 + x_2 \le 1$$
$$x_3 + x_4 \le 1$$

All ILPs in the ***cone*** will share the maximizer



$$\max \; 2x_1 + 4x_2 + 2x_3 + 0.5x_4$$
$$x_1 + x_2 \le 1$$
$$x_3 + x_4 \le 1$$

COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Amortized Inference Experiments

- Setup
  - Verb semantic role labeling; Entity and Relations
  - Speedup & Accuracy are measured over WSJ test set (Section 23) and Test of E & R
  - Baseline: solving ILPs using the Gurobi solver.

Amortized Inference Experiments

- Setup
  - Verb semantic role labeling; Entity and Relations
  - Speedup & Accuracy are measured over WSJ test set (Section 23) and Test of E & R
  - Baseline: solving ILPs using the Gurobi solver.
- For amortization
  - Cache 250,000 inference problems (objective, solution) from Gigaword
  - For each problem in test set either call the inference engine or re-use a solution from the cache, if our theorems hold.

Amortized Inference Experiments

- Setup
  - Verb semantic role labeling; Entity and Relations
  - Speedup & Accuracy are measured over WSJ test set (Section 23) and Test of E & R
  - Baseline: solving ILPs using the Gurobi solver.

- For amortization
  - Cache 250,000 inference problems (objective, solution) from Gigaword
  - For each problem in test set either call the inference engine or re-use a solution from the cache, if our theorems hold.

No training data is needed for this method.
Once you have a model, you can generate a large cache that will be then used to save you time at evaluation time.

Speedup & Accuracy

$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$

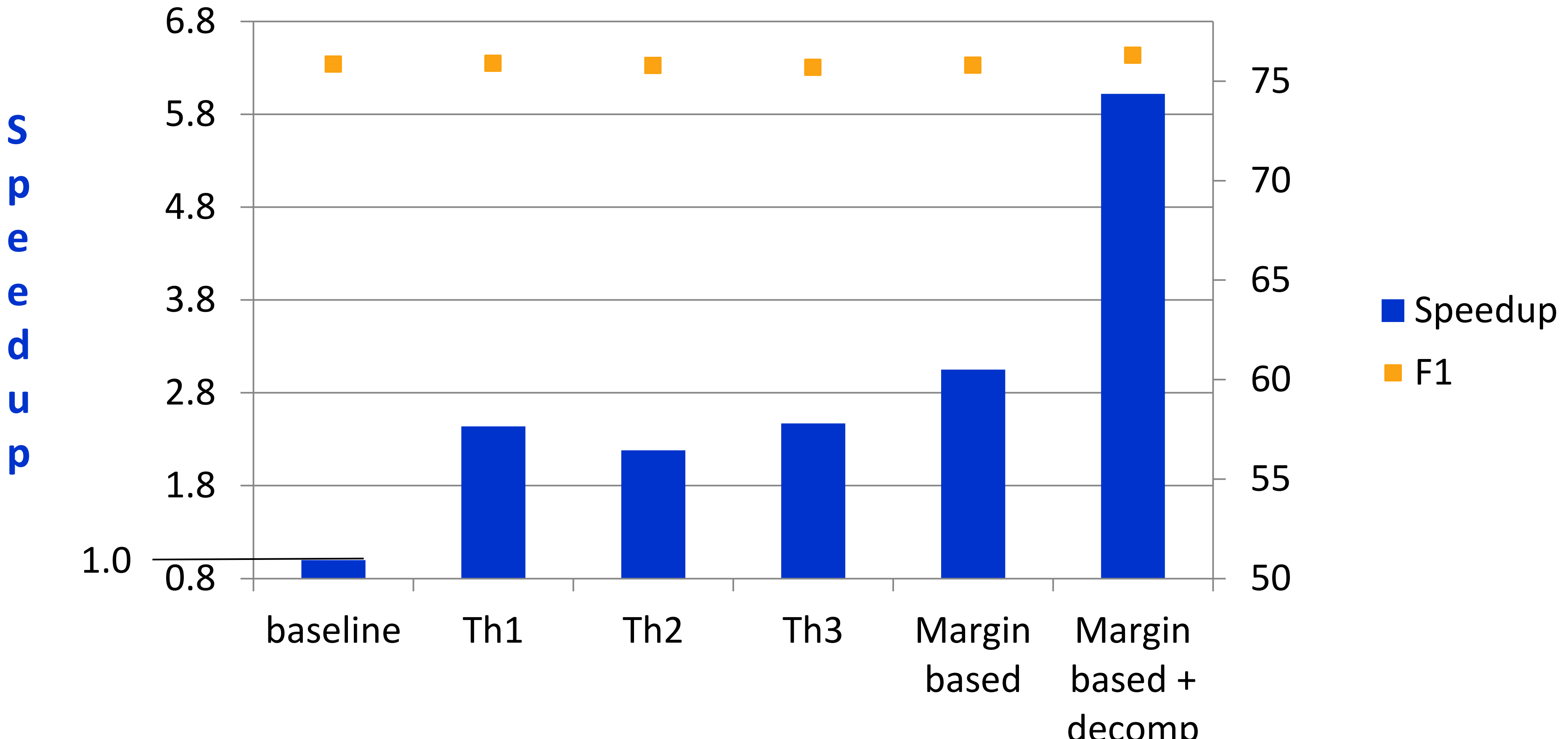
S
p
e
e
d
u
p

Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

By decomposing the objective function, building on the fact that "smaller structures" are more redundant, it is possible to get even better results.

$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$

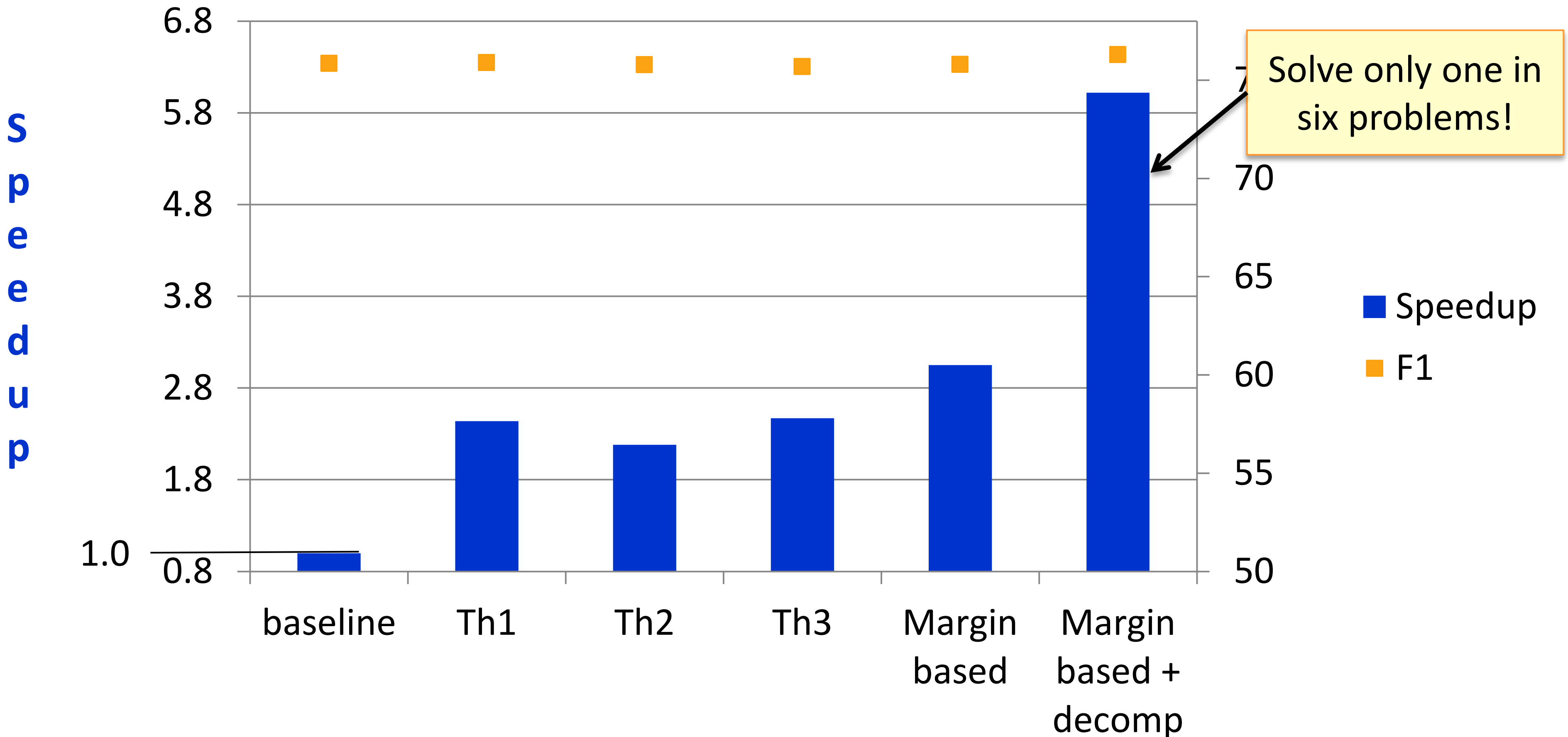


Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

By decomposing the objective function, building on the fact that "smaller structures" are more redundant, it is possible to get even better results.

$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$

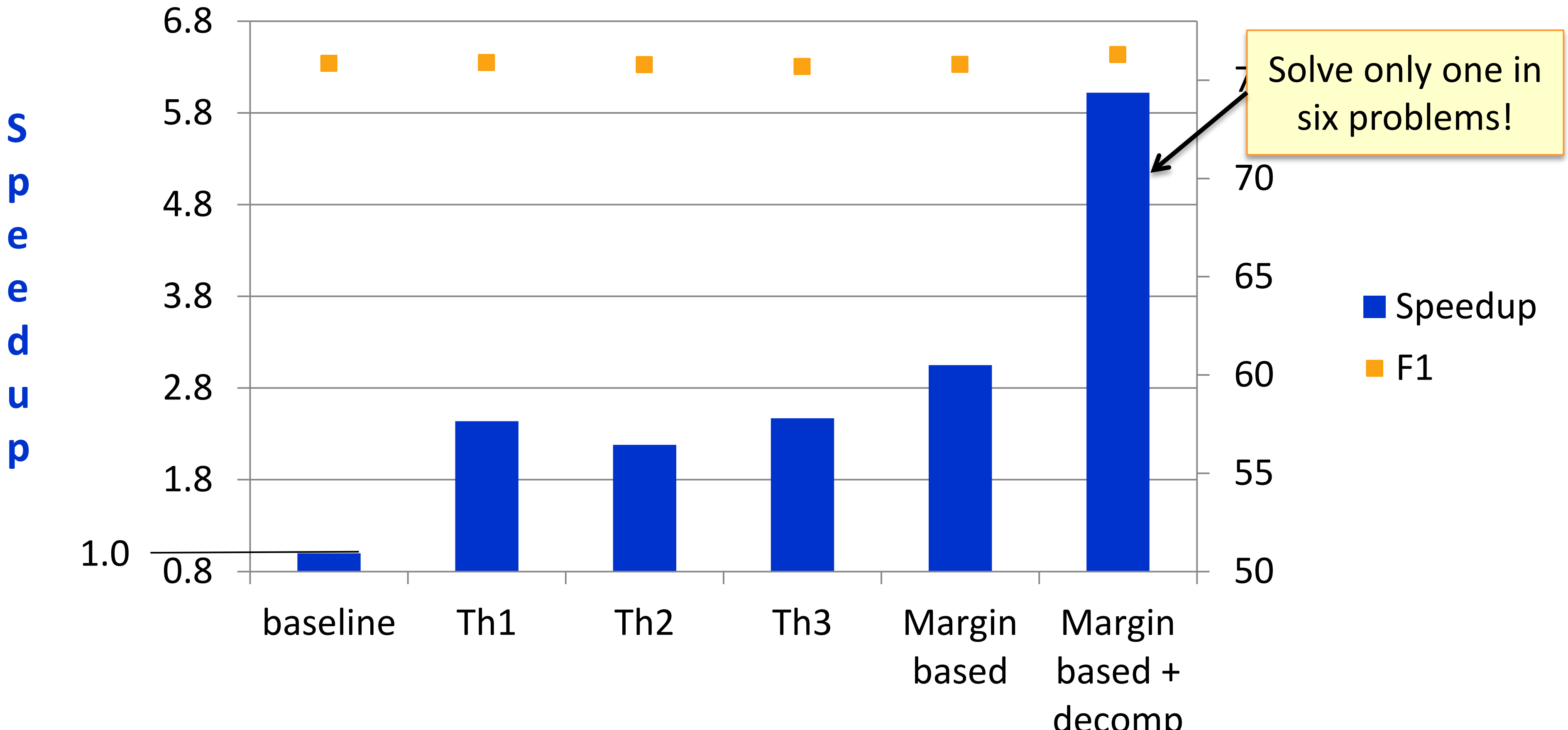


Amortization schemes [EMNLP'12, ACL'13]

Speedup & Accuracy

The results show that, indeed, the inference formulation provides a new level of abstraction that can be exploited to re-use solutions

$$\text{Speedup} = \frac{\text{number of inference calls without amortization}}{\text{number of inference calls with amortization}}$$



Amortization schemes [EMNLP'12, ACL'13]

Where are we?

Significant Progress in NLP and Information Extraction

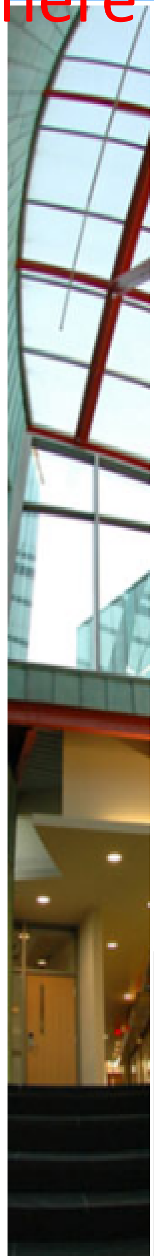
Where are we?

COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

News Research People Software Demos Publications Resources

DEMOS

Problems? Email mssammon@illinois.edu



What We Develop

Most of the information available today is in free form text. Current technologies (google, yahoo) allow us to access text only via key-word search.

We would like to facilitate content-based access to information. Examples include:

- Topical and Functional categorization of documents: Find documents that deal with stem cell research, but only Call for Proposals.
- Semantic categorization: Find documents about Columbus (the City, not the Person).
- Retrieval of concepts and entities rather than strings in text: Find documents about JFK, the president; include those documents that mention him as "John F. Kennedy, John Kennedy, Congressman Kennedy or any other possible writing; but not those that mention the baseball player John Kennedy, nor any of JFK's relatives.
- Extraction of information based on semantic categorization: Find a list of all companies that participated in merges in the last year. List all professors in Illinois that do research in Machine Learning.

Most Popular Demos

Part of Speech Tagging ▸▸

Shallow Parsing ▸▸

Semantic Role Labeling ▸▸

Context-Sensitive Spelling Correction ▸▸

Named Entity Recognition ▸▸

Running the Demos

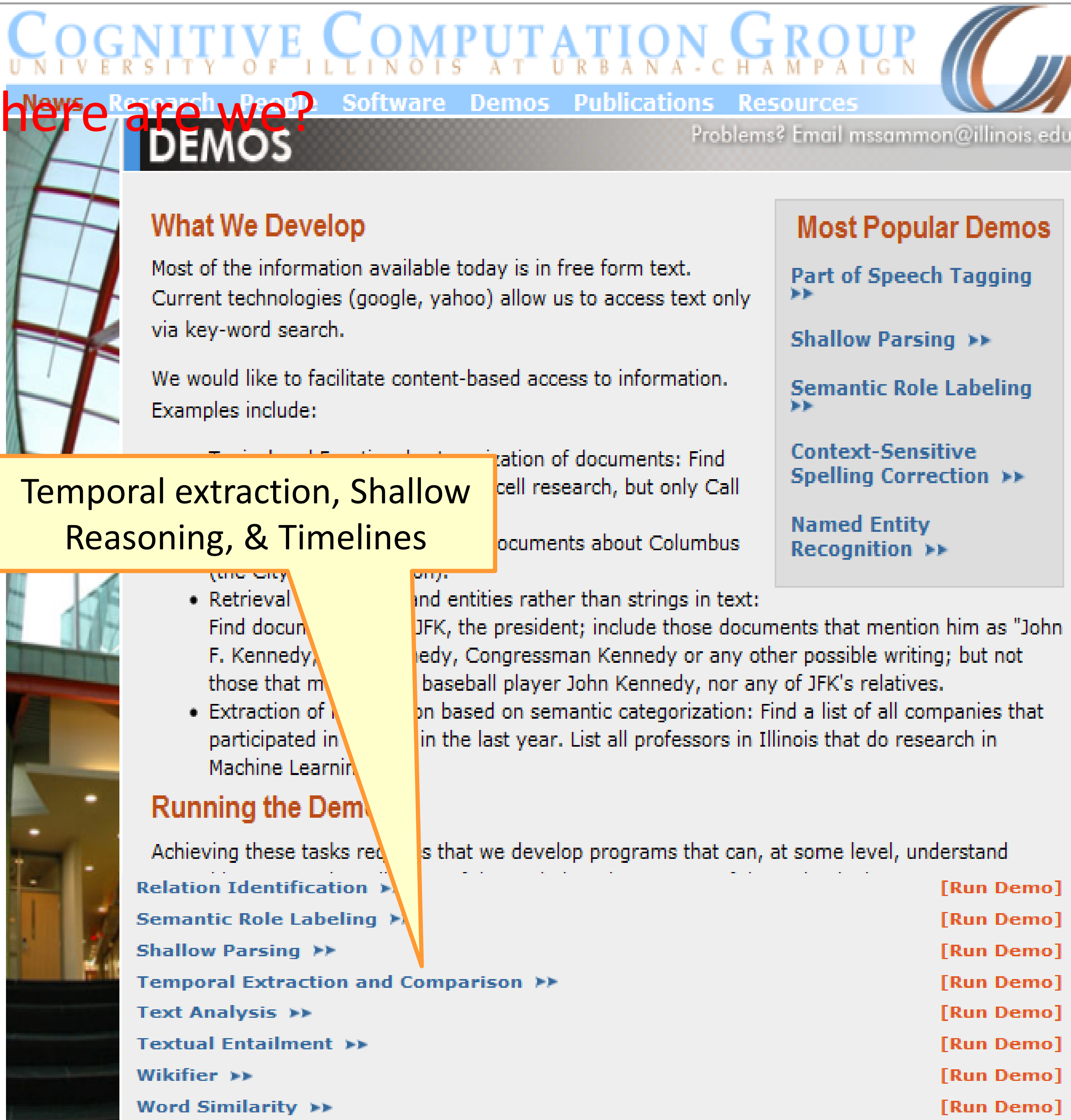
Achieving these tasks requires that we develop programs that can, at some level, understand

Relation Identification ▸▸	[Run Demo]
Semantic Role Labeling ▸▸	[Run Demo]
Shallow Parsing ▸▸	[Run Demo]
Temporal Extraction and Comparison ▸▸	[Run Demo]
Text Analysis ▸▸	[Run Demo]
Textual Entailment ▸▸	[Run Demo]
Wikifier ▸▸	[Run Demo]
Word Similarity ▸▸	[Run Demo]

COGNITIVE
UNIVERSITY

Significant Progress in NLP and Information Extraction

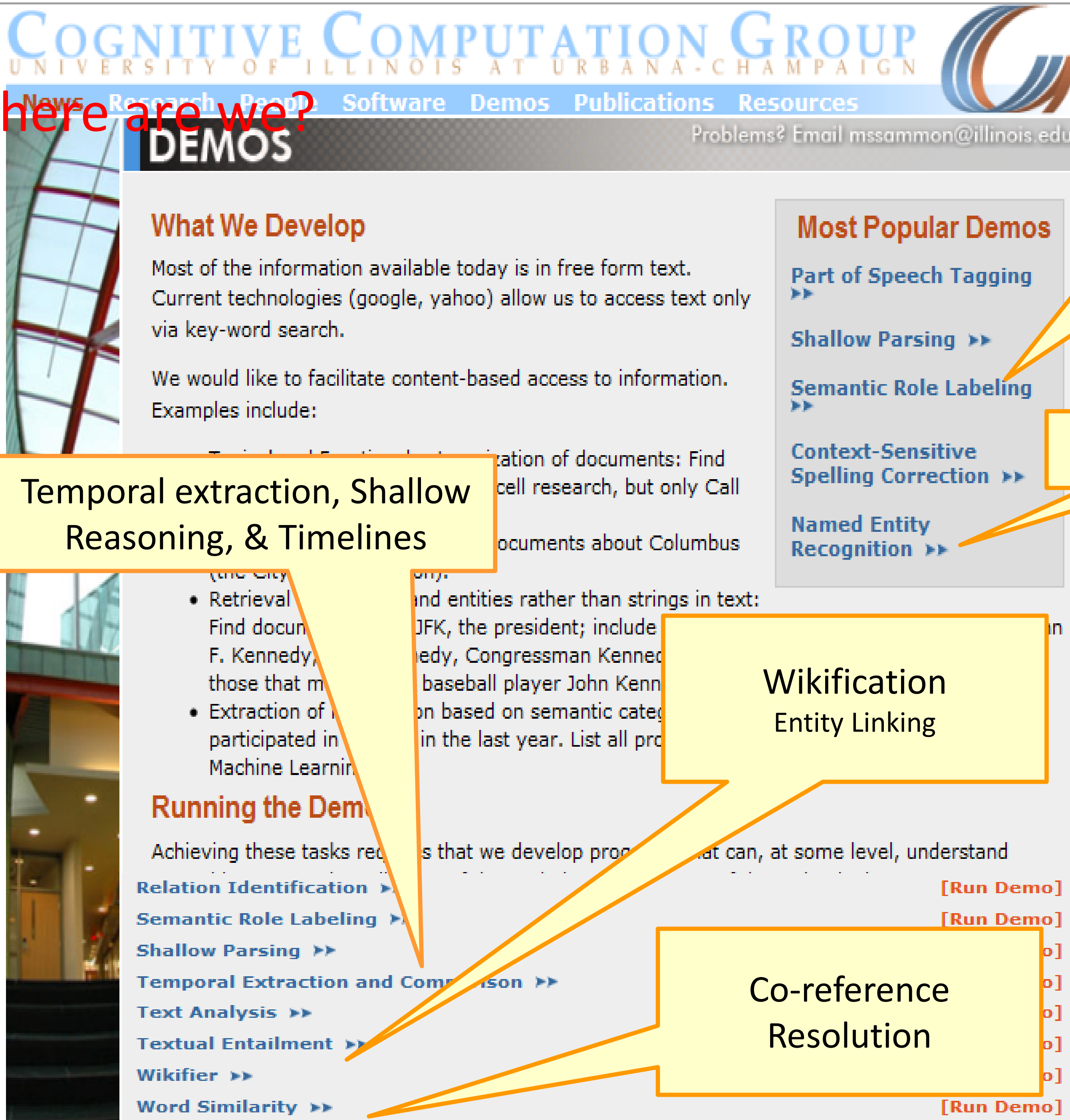
Where are we?



Temporal extraction, Shallow Reasoning, & Timelines

Significant Progress in NLP and Information Extraction

Where are we?



Shallow (semantic) parsing

Entities

Temporal extraction, Shallow Reasoning, & Timelines

Wikification
Entity Linking

Co-reference Resolution

Analyzing Electronic Health Records

The patient is a 65 year old female with post thoracotomy syndrome that occurred on the site of her thoracotomy incision .

She had a thoracic aortic aneurysm repaired in the past and subsequently developed neuropathic pain at the incision site .

She is currently on Vicodin , one to two tablets every four hours p.r.n. , Fentanyl patch 25 mcg an hour , change of patch every 72 hours , Elavil 50 mgq .h.s. , Neurontin 600 mg p.o. t.i.d. with still what she reports as stabbing left-sided chest pain that can be as severe as a 7/10.

She has failed conservative therapy and is admitted for a spinal cord stimulator trial .

Analyzing Electronic Health Records

Identify Important Mentions

[The patient] is a 65 year old female with [post thoracotomy syndrome] [that] occurred on the site of [[her] thoracotomy incision] .

[She] had [a thoracic aortic aneurysm] repaired in the past and subsequently developed [neuropathic pain] at [the incision site] .

[She] is currently on [Vicodin] , one to two tablets every four hours p.r.n. , [Fentanyl patch] 25 mcg an hour , change of patch every 72 hours , [Elavil] 50 mgq .h.s. , [Neurontin] 600 mg p.o. t.i.d. with still what [she] reports as [stabbing left-sided chest pain] [that] can be as severe as a 7/10.

[She] has failed [conservative therapy] and is admitted for [a spinal cord stimulator trial] .

Analyzing Electronic Health Records

Red : Problems
Green : Treatments
Purple : Tests
Blue : People

Identify Concept Types

[The patient] is a 65 year old female with [post thoracotomy syndrome] [that] occurred on the site of [[her] thoracotomy incision] .

[She] had [a thoracic aortic aneurysm] repaired in the past and subsequently developed [neuropathic pain] at [the incision site] .

[She] is currently on [Vicodin] , one to two tablets every four hours p.r.n. , [Fentanyl patch] 25 mcg an hour , change of patch every 72 hours , [Elavil] 50 mgq .h.s. , [Neurontin] 600 mg p.o. t.i.d. with still what [she] reports as [stabbing left-sided chest pain] [that] can be as severe as a 7/10.

[She] has failed [conservative therapy] and is admitted for [a spinal cord stimulator trial] .

Analyzing Electronic Health Records

Coreference Resolution

[The patient] is a 65 year old female with [post thoracotomy syndrome] [that] occurred on the site of [[her] thoracotomy incision] .

[She] had [a thoracic aortic aneurysm] repaired in the past and subsequently developed [neuropathic pain] at [the incision site] .

[She] is currently on [Vicodin] , one to two tablets every four hours p.r.n. , [Fentanyl patch] 25 mcg an hour , change of patch every 72 hours , [Elavil] 50 mgq .h.s. , [Neurontin] 600 mg p.o. t.i.d. with still what [she] reports as [stabbing left-sided chest pain] [that] can be as severe as a 7/10.

[She] has failed [conservative therapy] and is admitted for [a spinal cord stimulator trial] .

Analyzing Electronic Health Records

Other needs: temporal recognition & reasoning, relations, quantities, etc.

Coreference Resolution

[The patient] is a 65 year old female with [post thoracotomy syndrome] [that] occurred on the site of [[her] thoracotomy incision] .

[She] had [a thoracic aortic aneurysm] repaired in the past and subsequently developed [neuropathic pain] at [the incision site] .

[She] is currently on [Vicodin] , one to two tablets every four hours p.r.n. , [Fentanyl patch] 25 mcg an hour , change of patch every 72 hours , [Elavil] 50 mgq .h.s. , [Neurontin] 600 mg p.o. t.i.d. with still what [she] reports as [stabbing left-sided chest pain] [that] can be as severe as a 7/10.

[She] has failed [conservative therapy] and is admitted for [a spinal cord stimulator trial] .

Extended Semantic Role labeling

His first patient died of pneumonia. Another, who arrived from NY yesterday suffered from flu. Most others already recovered from flu

Extended Semantic Role labeling

His first patient died of pneumonia. Another, who arrived from NY yesterday suffered from flu. Most others already recovered from flu

Extended Semantic Role labeling

- Ambiguity and Variability of Prepositional Relations

His first patient died of pneumonia. Another, who arrived from NY yesterday suffered from flu. Most others already recovered from flu

Extended Semantic Role labeling

Verb Predicates, Noun predicates, prepositions, each dictates some relations, which have to cohere.

- Ambiguity and Variability of Prepositional Relations

His first patient died of pneumonia. Another, who arrived from NY yesterday suffered from flu. Most others already recovered from flu

Extended Semantic Role labeling

Verb Predicates, Noun predicates, prepositions, each dictates some relations, which have to cohere.

- Ambiguity and Variability of Prepositional Relations

Cause

Location

His first patient died of pneumonia. Another, who arrived from NY yesterday suffered from flu. Most others already recovered from flu

cause

Start-state

Extended Semantic Role labeling

Verb Predicates, Noun predicates, prepositions, each dictates some relations, which have to cohere.

- Ambiguity and Variability of Prepositional Relations

Cause

Location

His first patient died of pneumonia. Another, who arrived from NY yesterday suffered from flu. Most others already recovered from flu

cause

Start-state

Learn models; Acquire knowledge/constraints; Make decisions.

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

Extended Semantic Role labeling

Verb Predicates, Noun predicates, prepositions, each dictates some relations, which have to cohere.

- Ambiguity and Variability of Prepositional Relations

Cause

Location

His first patient died of pneumonia. Another, who arrived from NY yesterday suffered from flu. Most others already recovered from flu

cause

Start-state

Difficulty: no single source with annotation for all phenomena

Learn models; Acquire knowledge/constraints; Make decisions.

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

Constrained Conditional Models

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

Penalty for violating the constraint.

(Soft) constraints component

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a "legal" assignment

Constrained Conditional Models

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x,y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

Penalty for violating the constraint.

(Soft) constraints component

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a "legal" assignment

A collection of probabilistic models

Coreference: pairwise classifier between mentions

Concepts: a model that determines boundaries for important phrases.

Relations: Per-relation classifier

Knowledge as Constraints

Doctor cannot co-ref with a patient.

Consistency with KB resources

Consistency across relation types

Legitimacy of relations

Multiple Clinical and Scientific Applications

- Clinical Decisions:
  - "Please show me the reports of all patients who had headache that was not cured by Aspirin."
    - **Concept Recognition; Relation Identification (Problem, Treatment)**
  - "Please show me the reports of all patients who have had myocardial infarction (heart attack) more than once."
    - **Coreference Resolution**
- Identification of sensitive data (Privacy Reasons)
  - HIV Data, Drug Abuse, Family Abuse, Genetic Information
    - **Concept Recognition, Relations Recognition (drug, drug abuse), coreference resolution (multiple incidents, same people)**

Multiple Clinical and Scientific Applications

- Clinical Decisions:
  - "Please show me the reports of all patients who had headache that was not cured by Aspirin."
    - **Concept Recognition; Relation Identification (Problem, Treatment)**
  - "Please show me the reports of all patients who have had myocardial infarction (heart attack) more than once."
    - **Coreference Resolution**
- Identification of sensitive data (Privacy Reasons)
  - HIV Data, Drug Abuse, Family Abuse, Genetic Information
    - **Concept Recognition, Relations Recognition (drug, drug abuse), coreference resolution (multiple incidents, same people)**
- Generating summaries for patients
- Creating automatic reminders of medications
- Studying development and identification of diseases

Events

The police arrested AAA because he killed BBB two days after Christmas

Events

A "Kill" Event

The police arrested AAA because he killed BBB two days after Christmas

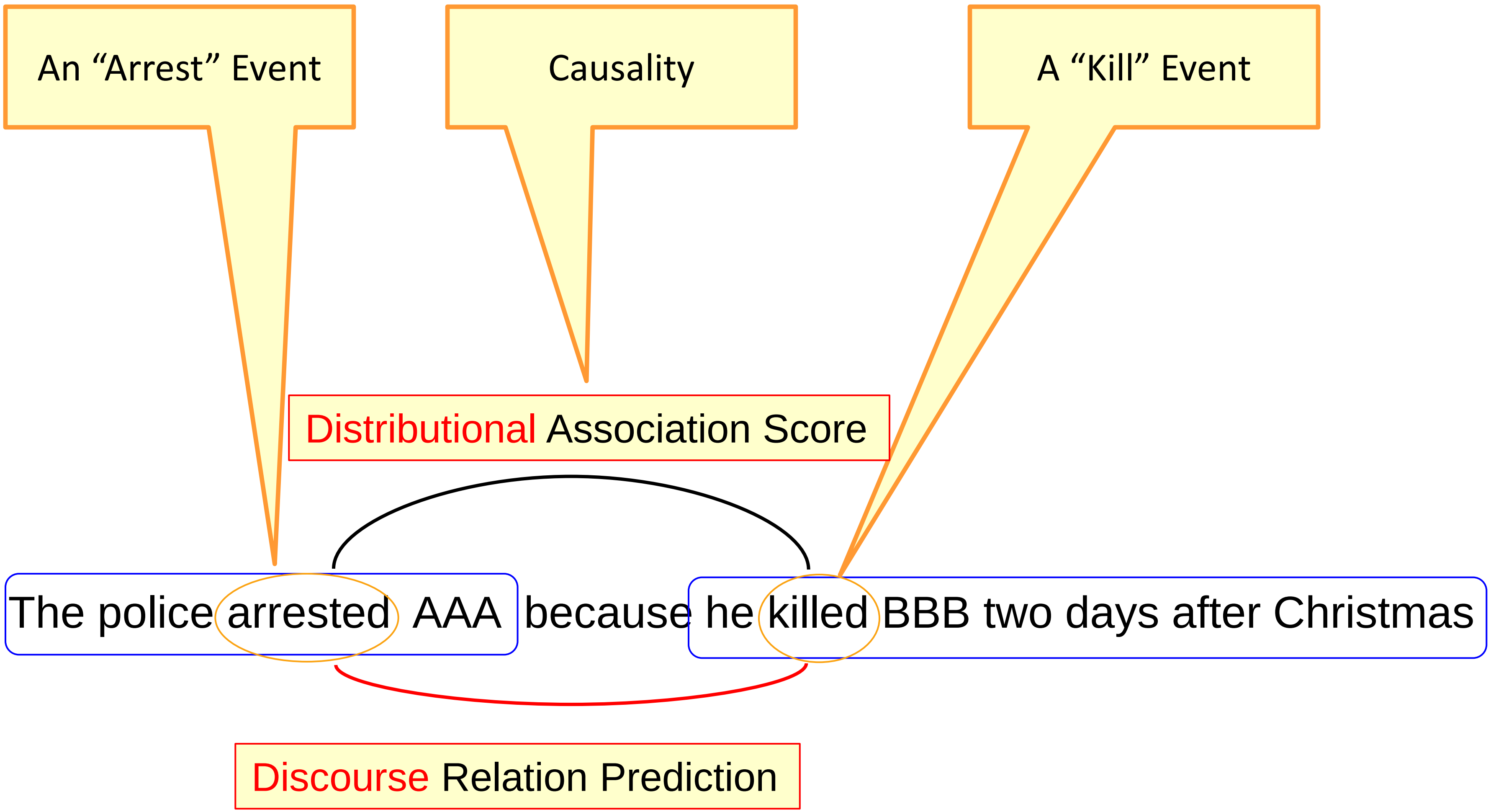
Events

An "Arrest" Event

A "Kill" Event

The police arrested AAA because he killed BBB two days after Christmas

Events


An “Arrest” Event
Causality
A “Kill” Event
Distributional Association Score
The police arrested AAA because he killed BBB two days after Christmas
Discourse Relation Prediction

Events

An “Arrest” Event

Causality

A “Kill” Event

Temporal

Distributional Association Score

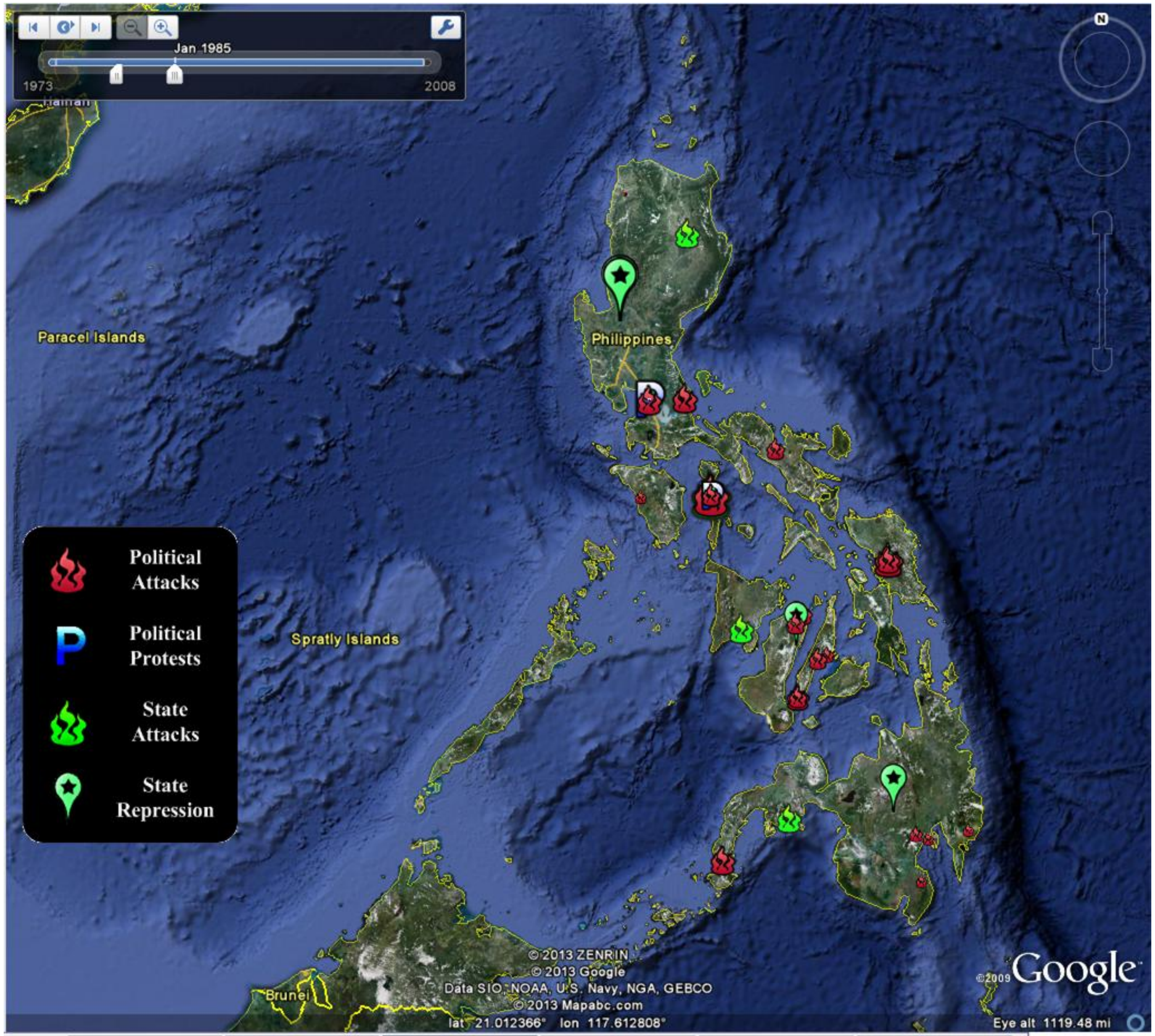
The police arrested AAA because he killed BBB two days after Christmas

Discourse Relation Prediction

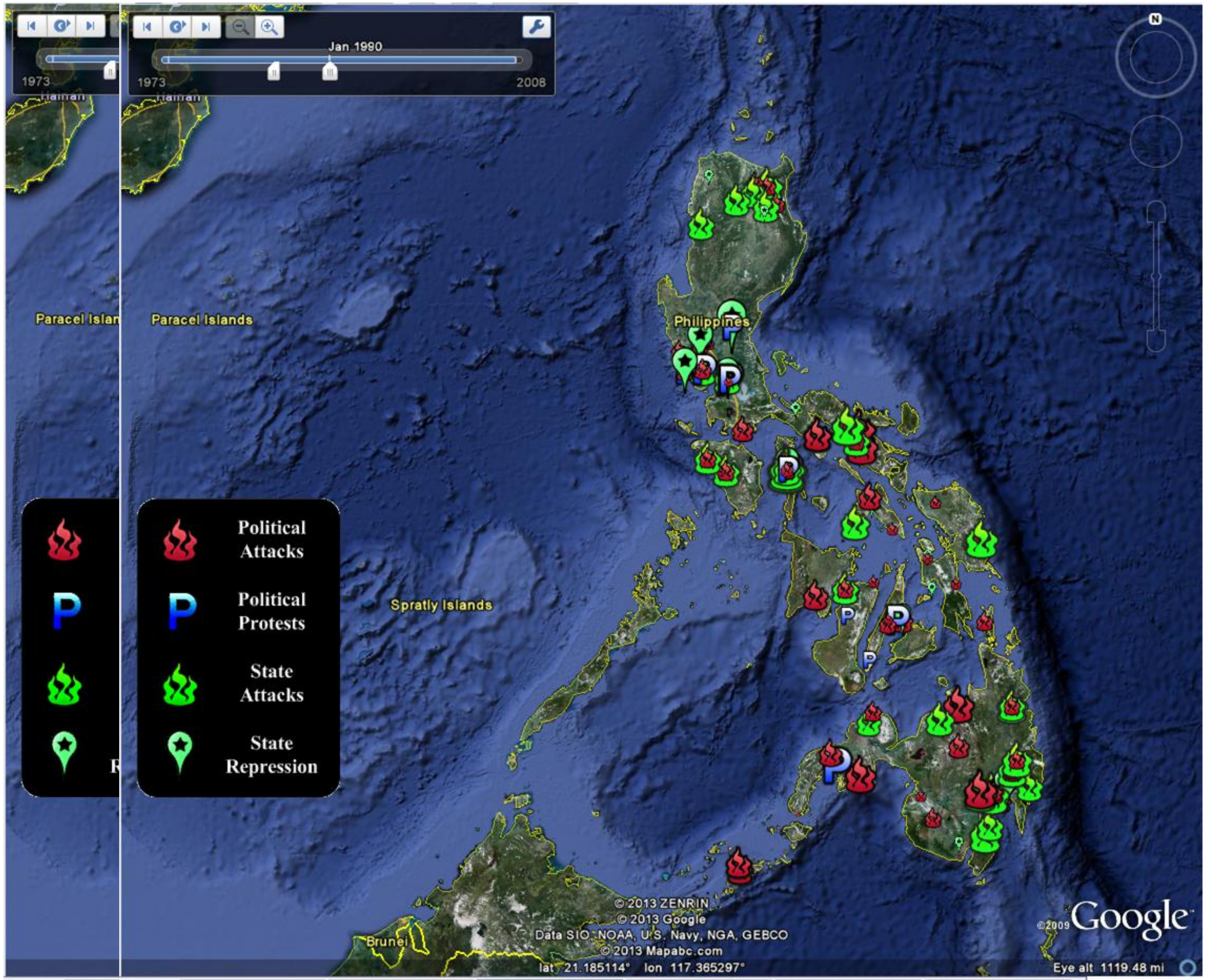
Cognitive Computation Group
University of Illinois at Urbana-Champaign

Social, Political and Economic Event Database (SPEED)

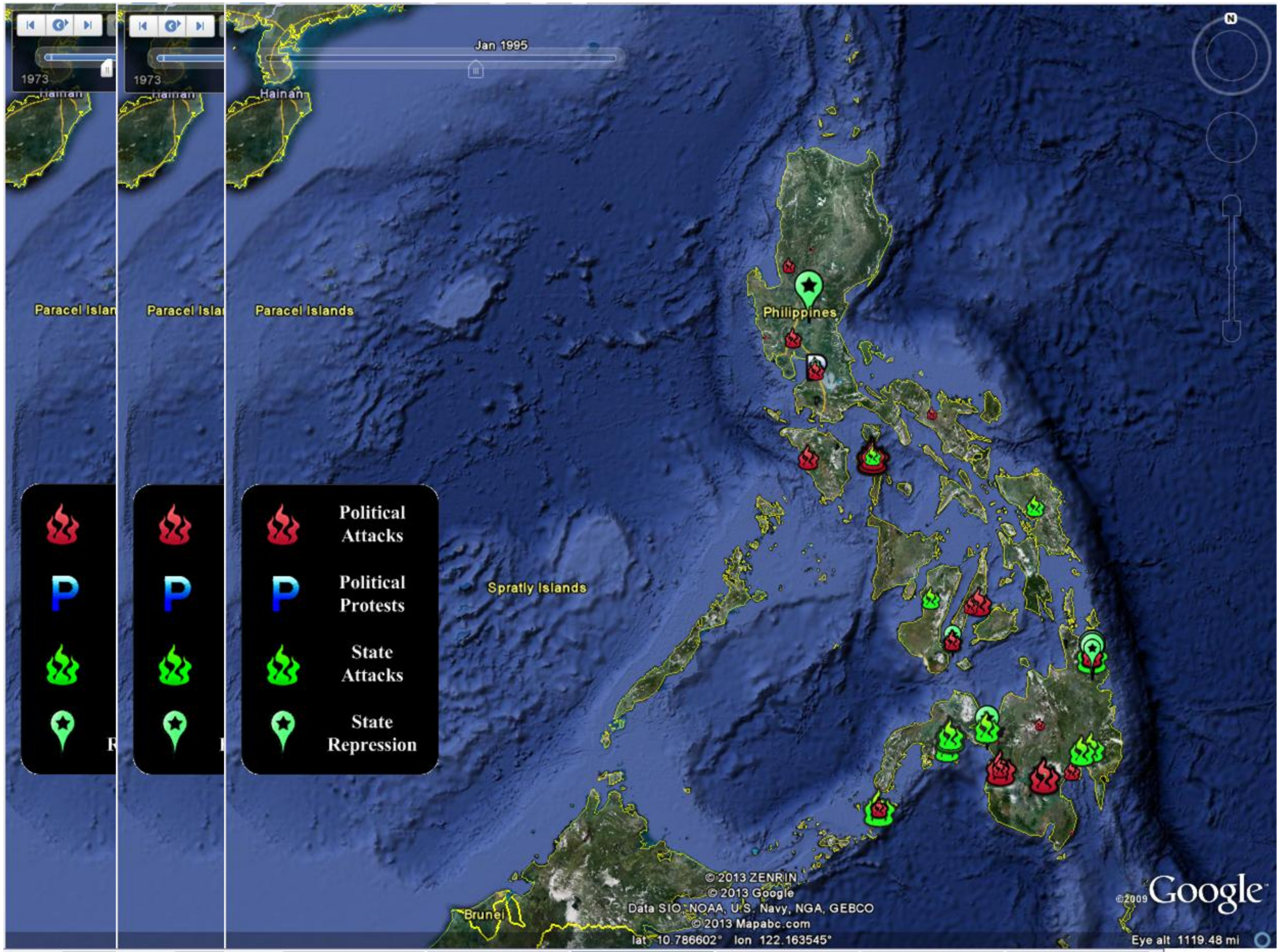
Social, Political and Economic Event Database (SPEED)



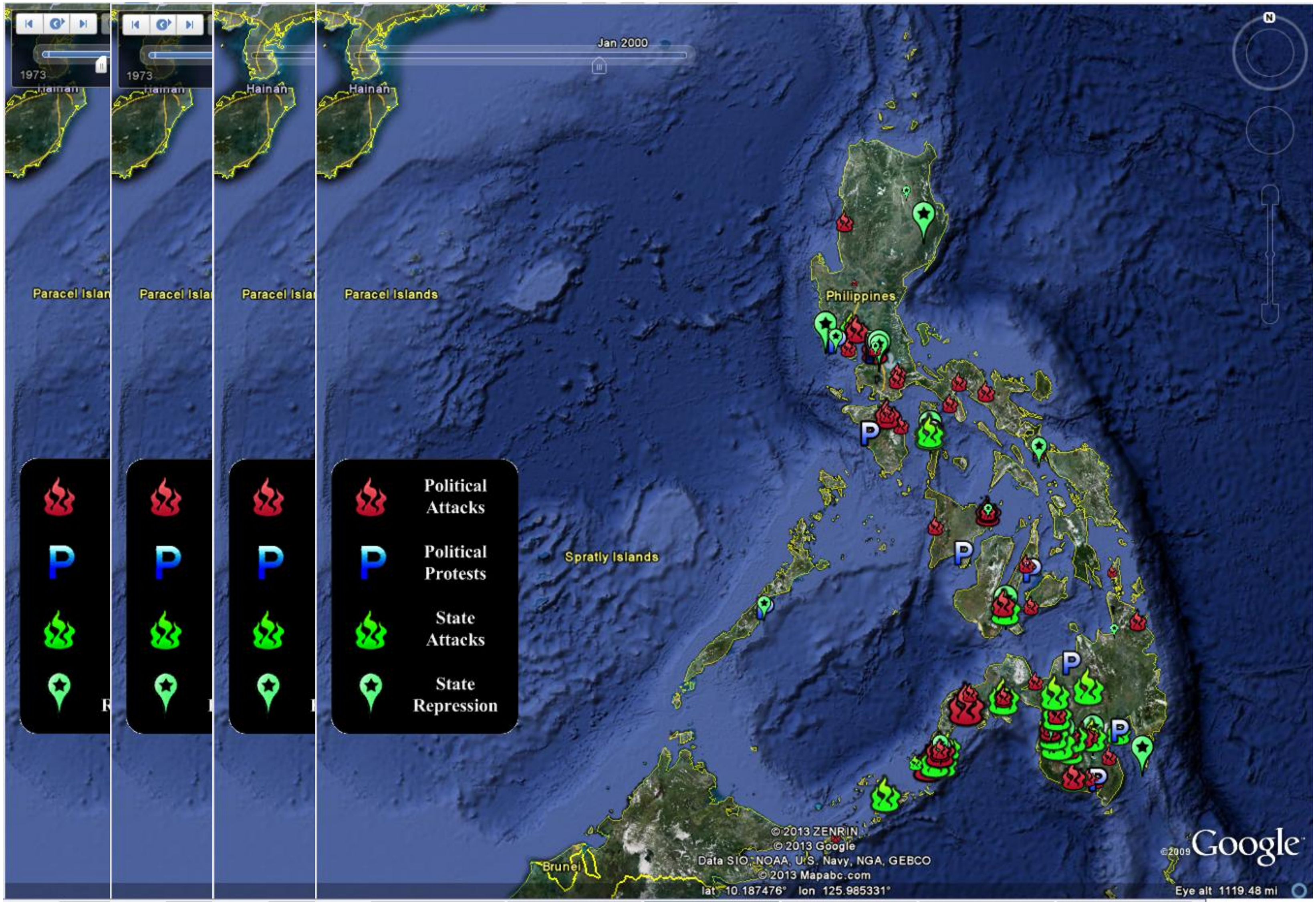
Social, Political and Economic Event Database (SPEED)



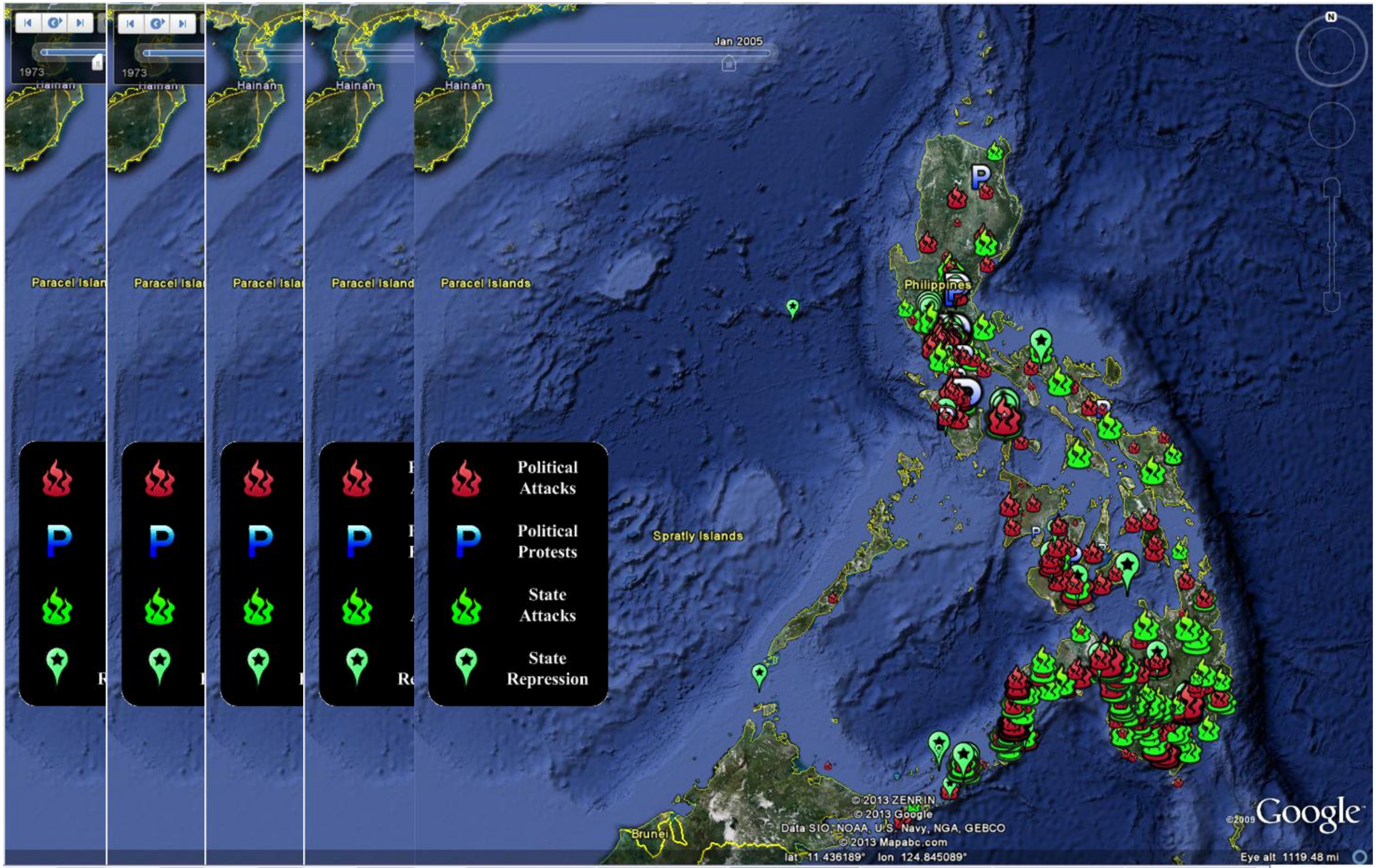
Social, Political and Economic Event Database (SPEED)



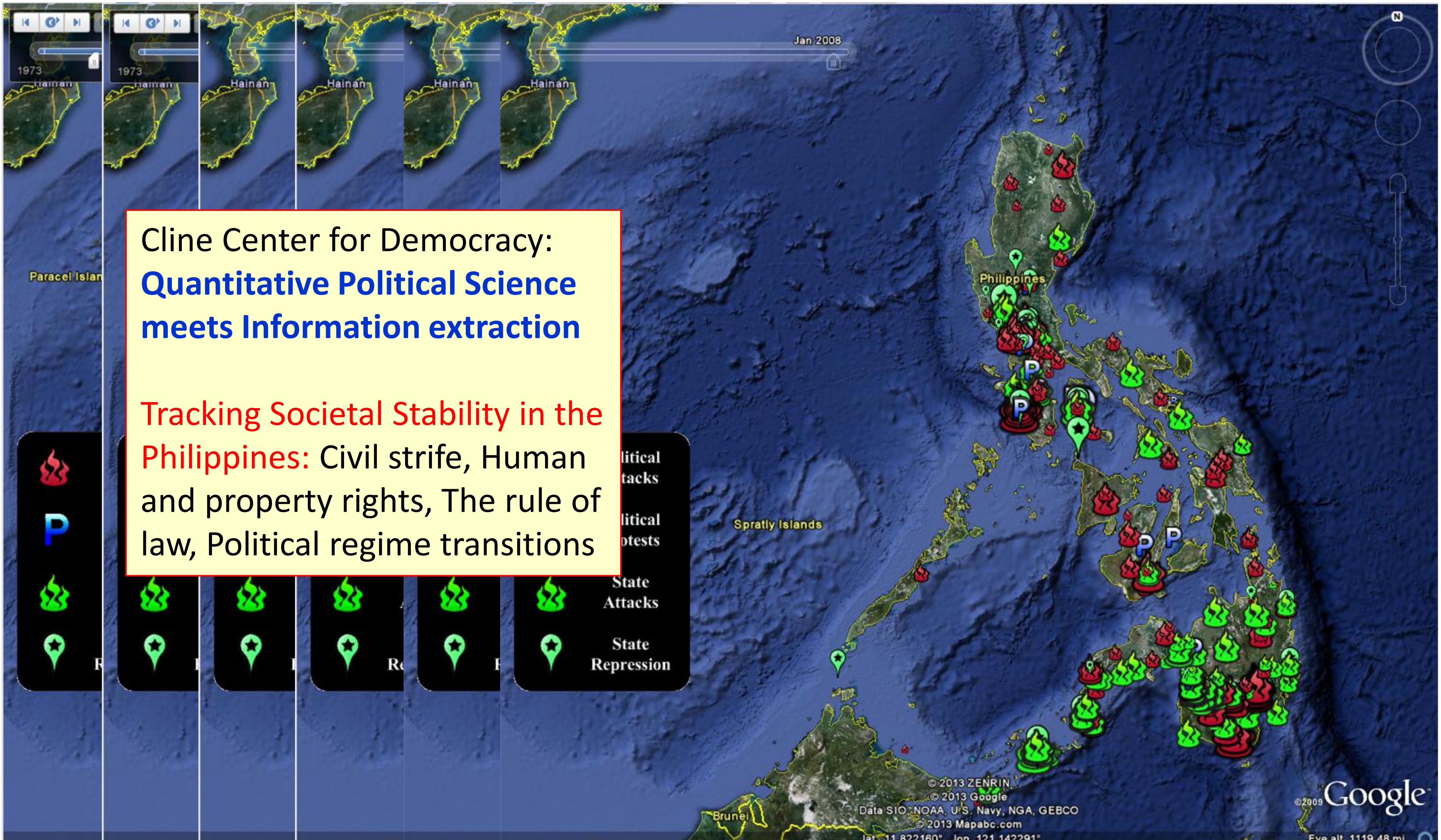
Social, Political and Economic Event Database (SPEED)



Social, Political and Economic Event Database (SPEED)



Social, Political and Economic Event Database (SPEED)



Cline Center for Democracy:
Quantitative Political Science meets Information extraction

Tracking Societal Stability in the Philippines: Civil strife, Human and property rights, The rule of law, Political regime transitions

Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem

Natural Language Understanding

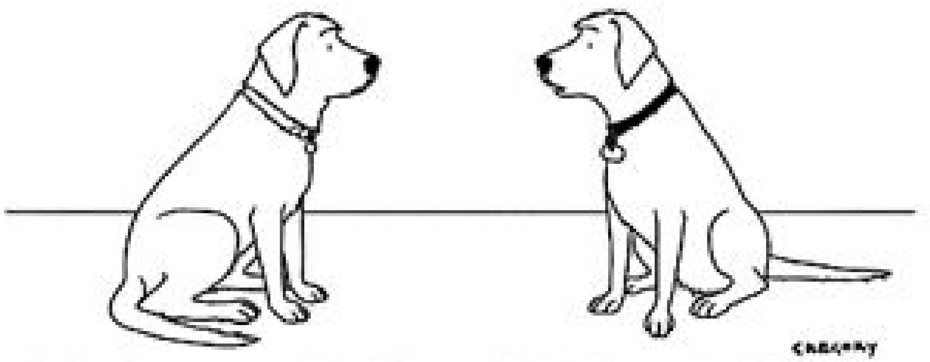
- Much research into [data → meaning] attempts to tell us what a document *says* with some level of certainty
  - Why is it difficult to do?
  - What can we do today?
    - How?
  - What can we expect to do?

Natural Language Understanding

- Much research into [data → meaning] attempts to tell us what a document *says* with some level of certainty
  - Why is it difficult to do?
  - What can we do today?
    - How?
  - What can we expect to do?
- But what should we believe, and who should we trust?

Knowing what to Believe

- The advent of the Information Age and the Web
  - Overwhelming quantity of information
  - But uncertain quality.



"I had my own blog for a while, but I decided to go back to just pointless, incessant barking."

- **Collaborative media**
  - Blogs
  - Wikis
  - Tweets
  - Message boards

- **Established media are losing market share**
  - Reduced fact-checking

Distributed Trust

- Sources may provide conflicting information or mutually reinforcing information.
  - Mistakenly or for a reason

Distributed Trust

- Sources may provide conflicting information or mutually reinforcing information.
  - Mistakenly or for a reason


The Truth-O-Meter Says:

Abortion services are "well over 90 percent of what Planned Parenthood does."

Jon Kyl on Friday, April 8th, 2011 in a Senate floor speech

Jon Kyl says abortion services are “well over 90 percent of what Planned Parenthood does”

As the government inched toward a shutdown on April 8, 2011, Sen. Jon Kyl, R-Ariz., gave a speech on the Senate floor to respond to Democratic charges that the major sticking point in the negotiations was a disagreement over Planned Parenthood.

Distributed Trust

- Sources may provide conflicting information or mutually reinforcing information.
  - Mistakenly or for a reason

Distributed Trust

- Sources may provide conflicting information or mutually reinforcing information.
  - Mistakenly or for a reason
  - Not feasible for human to read it all
  - **A computational trust system can be our proxy**
  - Ideally, assign the trust judgments the user would
  - The user may be another system
    - **A question answering system; A navigation system; A news aggregator**
    - **A warning system**

Emergency Situations

Emergency Situations

- A distributed data stream needs to be monitored

- All Data streams have Natural Language Content
  - Internet activity
    - **chat rooms, forums, search activity, twitter and cell phones**
  - Traffic reports; 911 calls and other emergency reports
  - Network activity, power grid reports, networks reports, security systems, banking
  - Media coverage

- Often, stories appear on tweeter before they break the news
- But, a lot of conflicting information, possibly misleading and deceiving

Medical Domain: Many support groups and medical forums

Medical Domain: Many support groups and medical forums

HealthBoards
HEALTH MESSAGE BOARDS

HOME | MESSAGE BOARDS | HEALTH GUIDE | JOIN FOR FREE

SEARCH Go

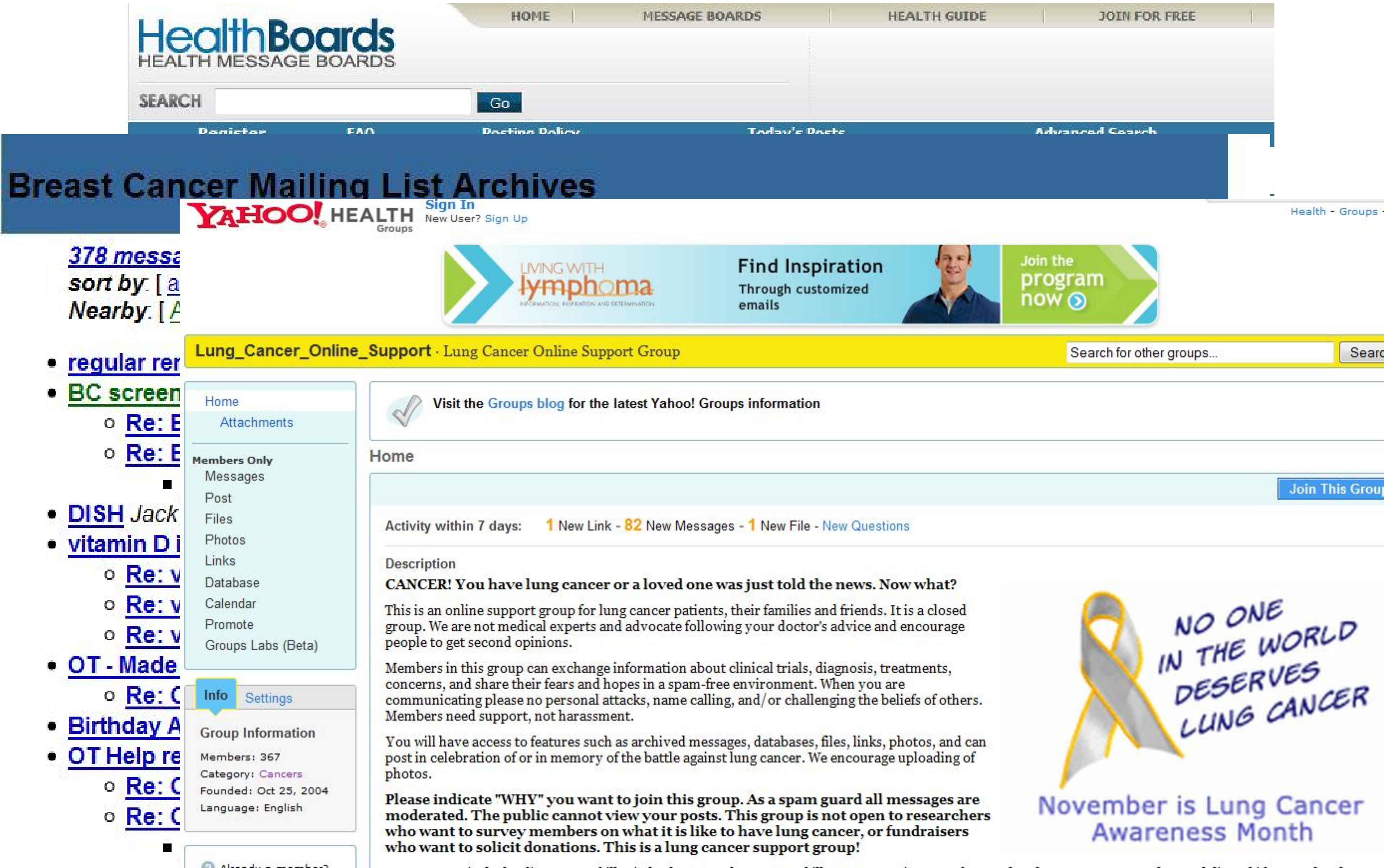
Register | FAQ | Posting Policy | Today's Posts | Advanced Search

Breast Cancer Mailing List Archives

378 messages: ***Starting*** Thu Jan 01 2009 - 15:13:39 EST, ***Ending*** Sat Jan 31 2009 - 15:32:46 EST
sort by: [author] [date] [subject]
Nearby: [About this archive]

- **regular reminder: FBCL donations (please delete if not interested; *NOT* required for participation on this list!)**
- **BC screening** *Jack And Diane (Sat Jan 31 2009 - 08:53:55 EST)*
  - **Re: BC screening** *Marlyne Rohan (Sat Jan 31 2009 - 13:02:59 EST)*
  - **Re: BC screening** *Maria Wetzel (Sat Jan 31 2009 - 14:48:00 EST)*
    - **Re: BC screening** *M. Manning (Sat Jan 31 2009 - 15:19:50 EST)*
- **DISH** *Jack And Diane (Sat Jan 31 2009 - 08:21:08 EST)*
- **vitamin D information and testing/how my husband is** *Jack And Diane (Sat Jan 31 2009 - 08:17:26 EST)*
  - **Re: vitamin D information and testing/how my husband is** *Hilde Horvath (Wed Jan 28 2009 - 17:22:49 EST)*
  - **Re: vitamin D information and testing/how my husband is** *Jean Brugger (Wed Jan 28 2009 - 19:17:49 EST)*
  - **Re: vitamin D information and testing/how my husband is** *Holly Anderson (Fri Jan 30 2009 - 19:40:42 EST)*
- **OT - Made my head ache!** *Norma Steele (Sat Jan 31 2009 - 02:03:07 EST)*
  - **Re: OT - Made my head ache!** *Hilde Horvath (Sat Jan 31 2009 - 07:29:41 EST)*
- **Birthday Alert for Tomorrow (31st)** *Sarah Webster-Eastman (Fri Jan 30 2009 - 14:39:23 EST)*
- **OT Help request** *Jacqueline (Fri Jan 30 2009 - 14:01:54 EST)*
  - **Re: OT Help request** *maria roseb (Fri Jan 30 2009 - 16:14:17 EST)*
  - **Re: OT Help request** *Kaye N (Fri Jan 30 2009 - 18:16:58 EST)*
    - **Re: OT Help request** *M. Manning (Fri Jan 30 2009 - 18:30:56 EST)*

Medical Domain: Many support groups and medical forums



Medical Domain: Many support groups and medical forums

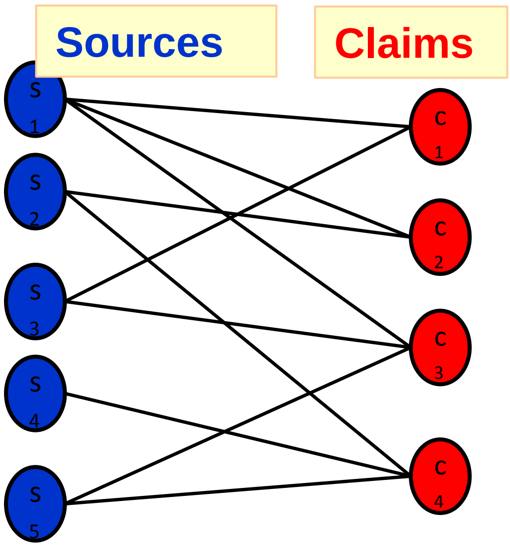
Medical Domain: Many support groups and medical forums

Hundreds of Thousands of people get their medical information from the internet

- **Best treatment for…..**
- **Side effects of….**
- **But, some users have an agenda,… pharmaceutical companies…**

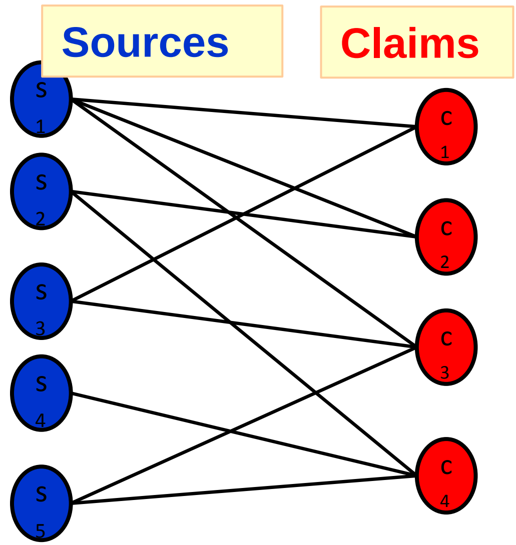
Trustworthiness

- Given:
  - Multiple content sources
  - Some target relations ("facts")
    - **E.g. [disease, treatments], [treatments, side-effects]**
  - Prior beliefs & background knowledge



Trustworthiness

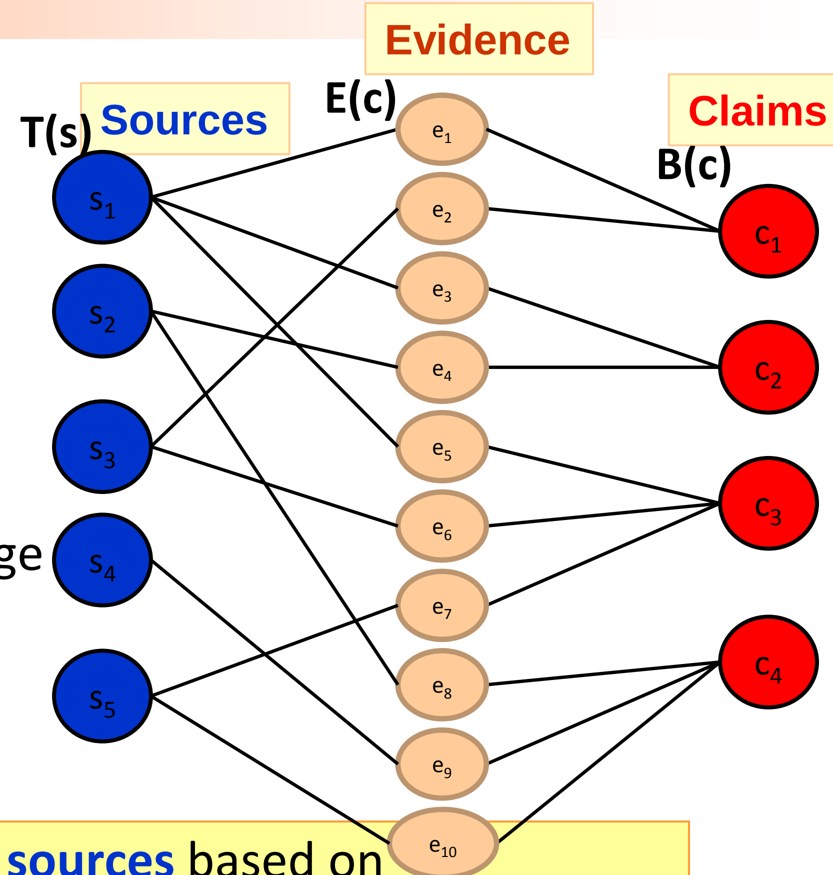
- Given:
  - Multiple content sources
  - Some target relations ("facts")
    - **E.g. [disease, treatments], [treatments, side-effects]**
  - Prior beliefs & background knowledge



- Our goal is to:
  - Score trustworthiness of **claims** and **sources** based on
    - **Support across multiple (trusted) sources**
    - **Source characteristics:**
      - reputation, interest-group (commercial / govt. backed / public interest), verifiability of information (cited info)
    - **Prior Beliefs and Background knowledge**
    - **Understanding content**

Trustworthiness

- Given:
  - Multiple content sources
  - Some target relations (“facts”)
    - **E.g.** **[disease, treatments], [treatments, side-effects]**
  - Prior beliefs & background knowledge



- Our goal is to:
  - Score trustworthiness of **claims** and **sources** based on
    - **Support across multiple (trusted) sources**
    - **Source characteristics:**
      - reputation, interest-group (commercial / govt. backed / public interest), verifiability of information (cited info)
    - **Prior Beliefs and Background knowledge**
    - **Understanding content**

Summary: Making Sense of Unstructured Data

- A lot of today's information is in text
- Making sense of unstructured data
  - Automatic text understanding (Natural Language Processing) is essential to supporting better access, analysis, and synthesis of data
  - Discussed a unified Learning and Inference approach that has had large impact on our ability to move forward in this direction.
  - Very active research area – the problem isn't solve yet…
  - But we can offer practical solutions that reliably address a range a problems.
- Trustworthiness of information
  - Comes up in the context of **social (and "standard" media)**, but also in the context of using sensory information
- Very broad applications, with huge societal impact.

Summary: Making Sense of Unstructured Data

Check out our tools, demos, LBJava, CCM tutorial,…

- A lot of today's information is in text
- Making sense of unstructured data
  - Automatic text understanding (Natural Language Processing) is essential to supporting better access, analysis, and synthesis of data
  - Discussed a unified Learning and Inference approach that has had large impact on our ability to move forward in this direction.
  - Very active research area – the problem isn't solve yet…
  - But we can offer practical solutions that reliably address a range a problems.
- Trustworthiness of information
  - Comes up in the context of **social (and "standard" media)**, but also in the context of using sensory information
- Very broad applications, with huge societal impact.

Thank you!

Summary: Making Sense of Unstructured Data

Check out our tools, demos, LBJava, CCM tutorial,…

- A lot of today's information is in text
- Making sense of unstructured data
  - Automatic text understanding (Natural Language Processing) is essential to supporting better access, analysis, and synthesis of data
  - Discussed a unified Learning and Inference approach that has had large impact on our ability to move forward in this direction.
  - Very active research area – the problem isn't solve yet…
  - But we can offer practical solutions that reliably address a range a problems.
- Trustworthiness of information
  - Comes up in the context of **social (and "standard" media)**, but also in the context of using sensory information
- Very broad applications, with huge societal impact.