COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

MULTIMODAL INFORMATION ACCESS & SYNTHESIS
mias

Making Sense of Unstructured Data

Dan Roth

Department of Computer Science

University of Illinois at Urbana-Champaign

September 2014

ACADEMIC ROUNDTABLE @ ANDREESSEN HOROWITZ

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.**

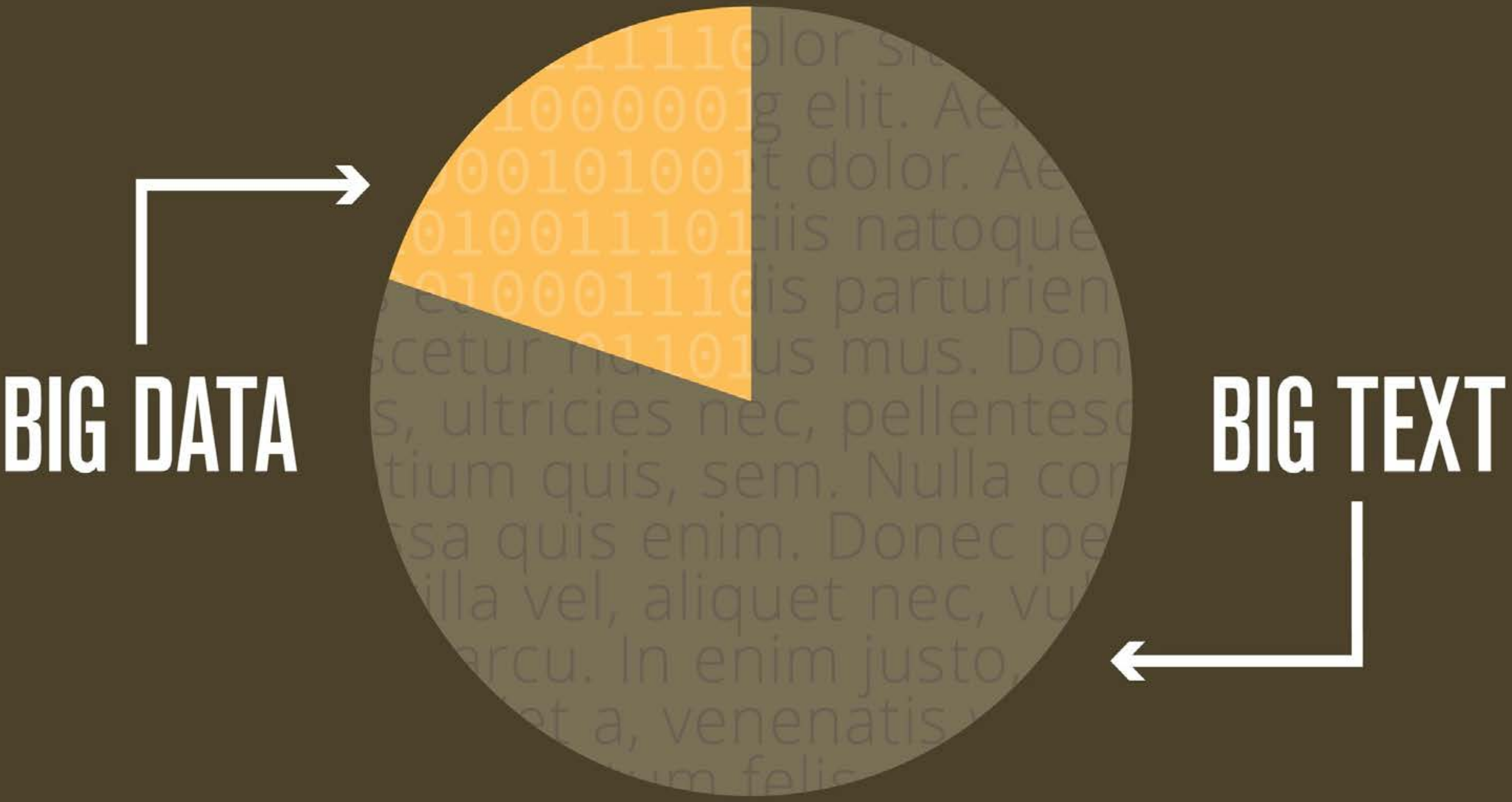
- Develop the theories, algorithms, and tools to enable **transforming raw data** into **useful and understandable information** & integrating it with existing resources
- **[data → meaning] transformation**.

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.**

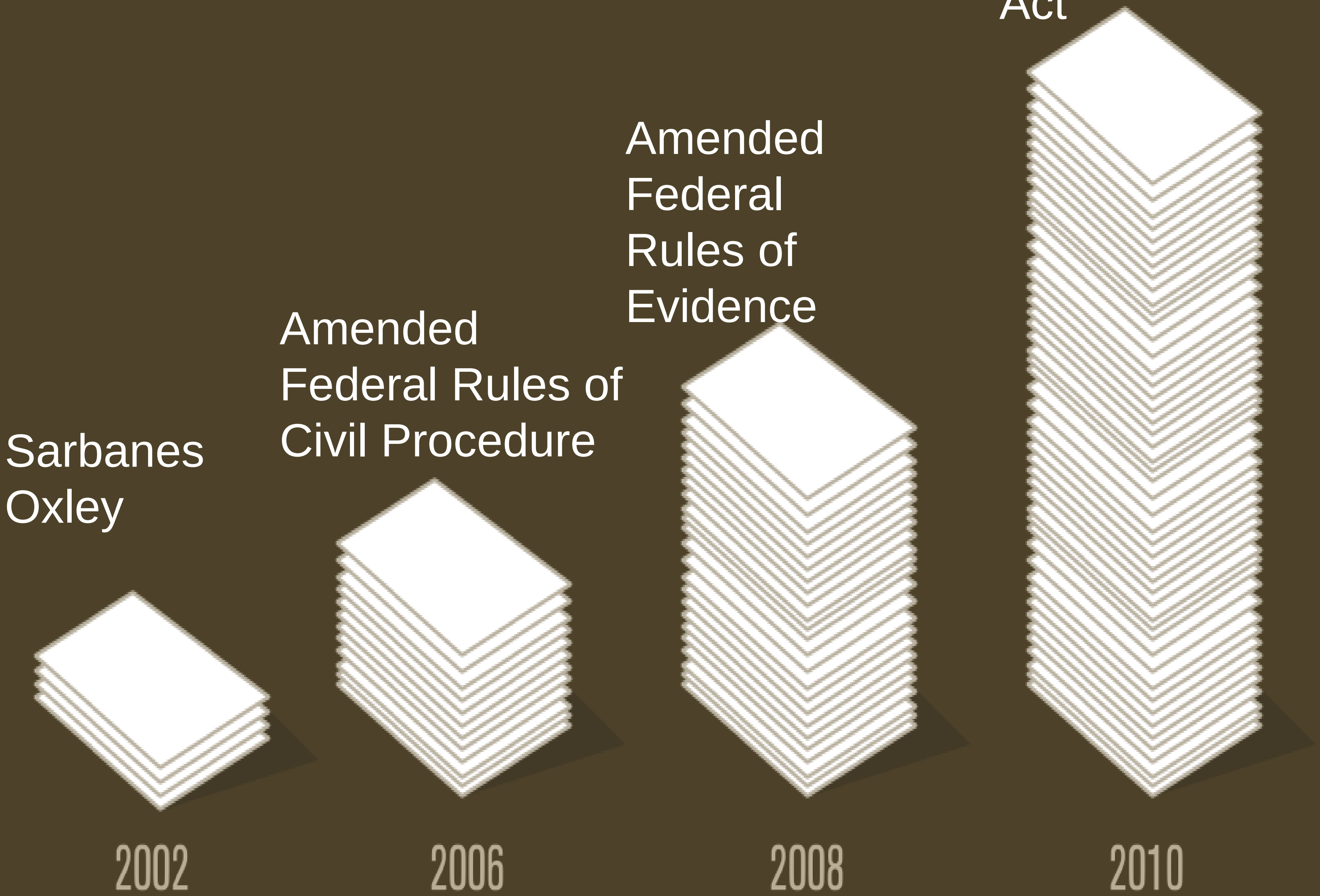
- Develop the theories, algorithms, and tools to enable **transforming raw data** into **useful and understandable information** & integrating it with existing resources
- **[data → meaning] transformation**.
- TODAY: Why is it hard – what we can do….



BIG DATA

BIG TEXT

More than a **million rules,** requiring companies and their boards to understand what their employees are doing and with whom they are communicating.



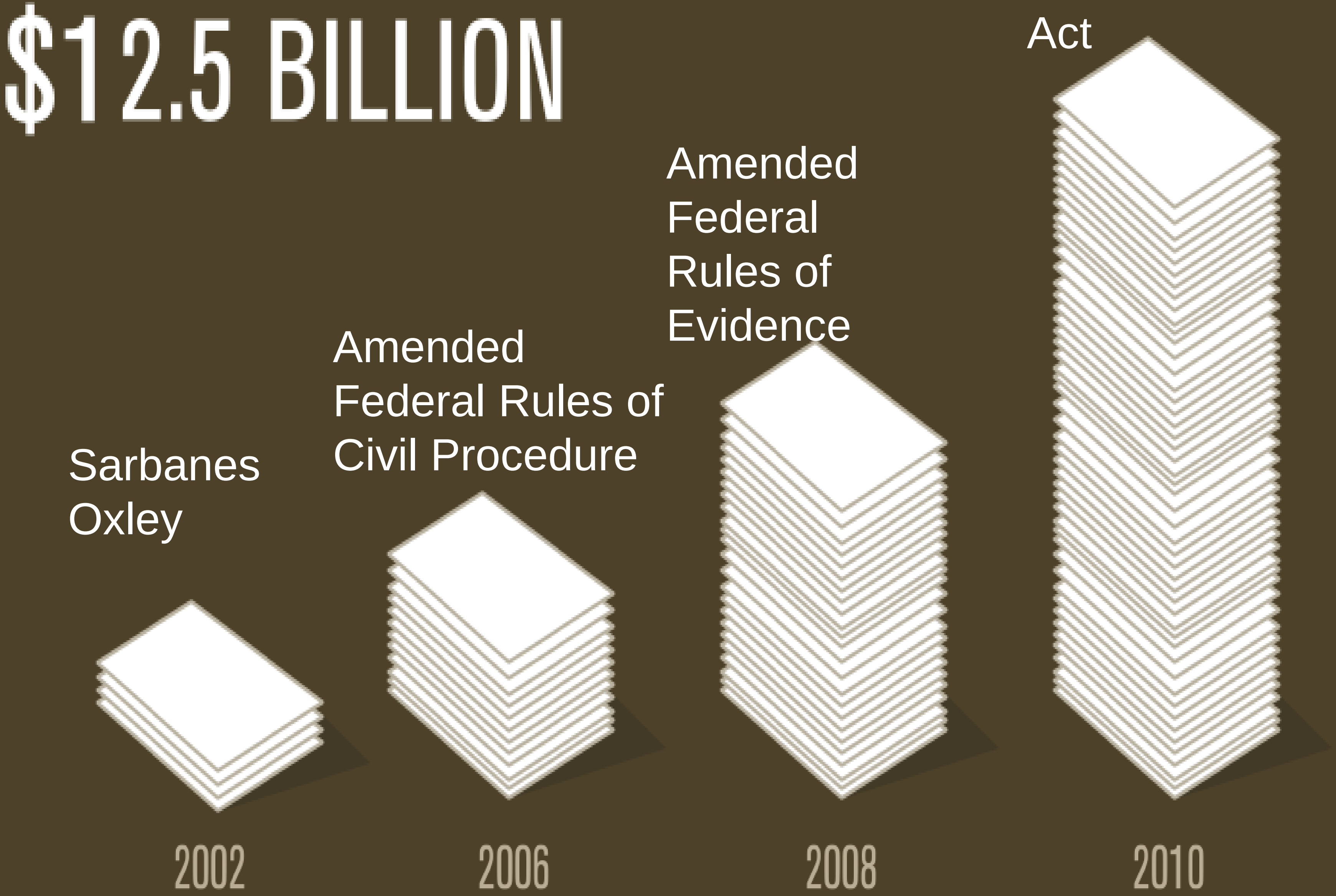
Sarbanes Oxley

Amended Federal Rules of Civil Procedure

Amended Federal Rules of Evidence

Dodd-Frank Act

More than a **million rules,** requiring companies and their boards to understand what their employees are doing and with whom they are communicating.



Sarbanes Oxley

Amended Federal Rules of Civil Procedure

Amended Federal Rules of Evidence

Dodd-Frank Act

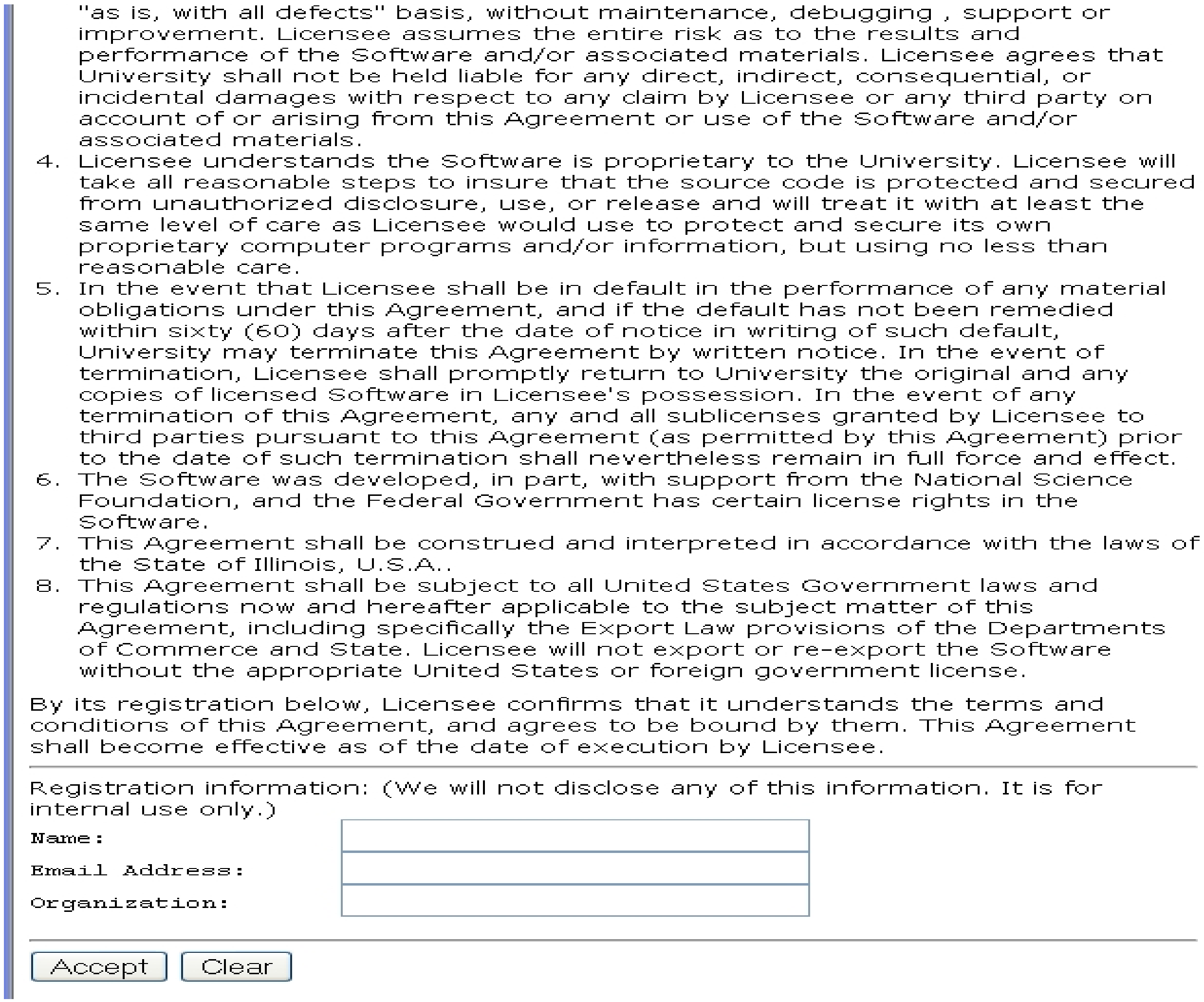
WORLD TEXT

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

2012
2014
2020

KEYWORD
SEARCH

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.

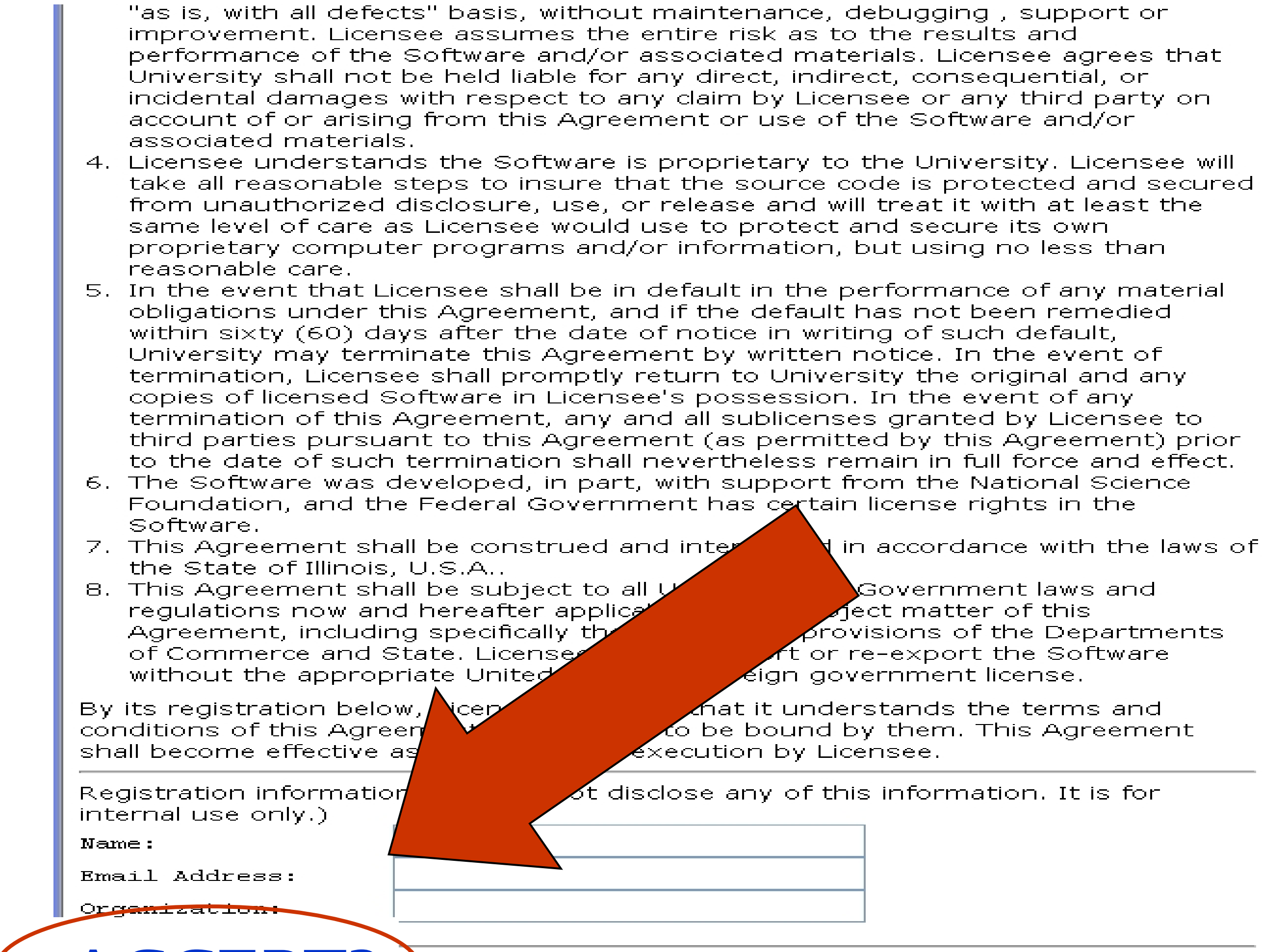
Registration information: (We will not disclose any of this information. It is for internal use only.)

Name:

Email Address:

Organization:

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 inter in accordance with the laws of the State of Illinois, U.S.A..
8. This Agreement shall be subject to all U Government laws and regulations now and hereafter applica ject matter of this Agreement, including specifically th provisions of the Departments of Commerce and State. Licensee rt or re-export the Software without the appropriate United eign government license.

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

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

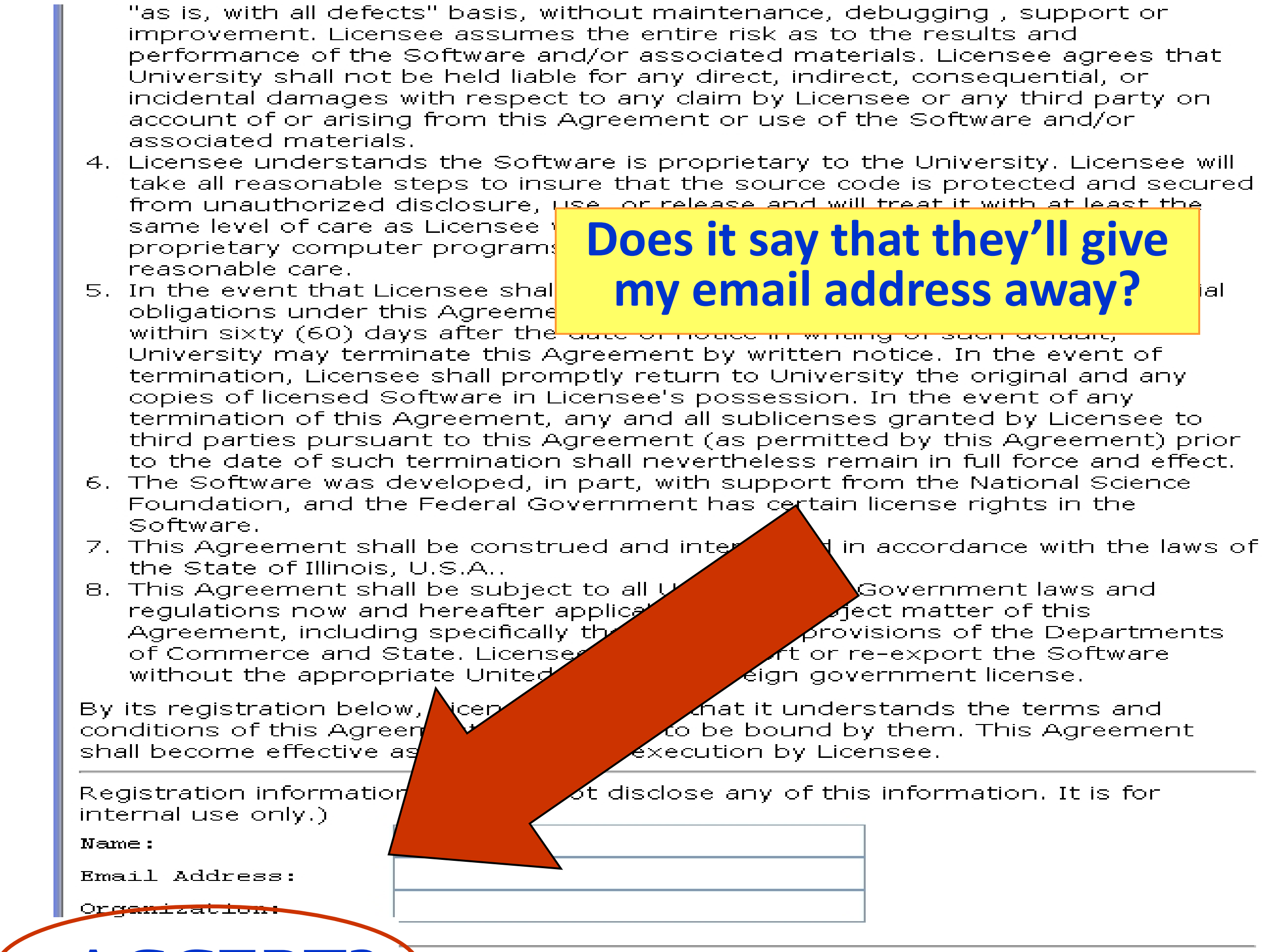
Name:

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
proprietary computer programs
reasonable care.

5. In the event that Licensee shal ial
obligations under this Agreeme
within sixty (60) days after the
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 d in accordance with the laws of
the State of Illinois, U.S.A..

8. This Agreement shall be subject to all U Government laws and
regulations now and hereafter applica ject matter of this
Agreement, including specifically th provisions of the Departments
of Commerce and State. Licensee rt or re-export the Software
without the appropriate United eign government license.

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

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

Name:

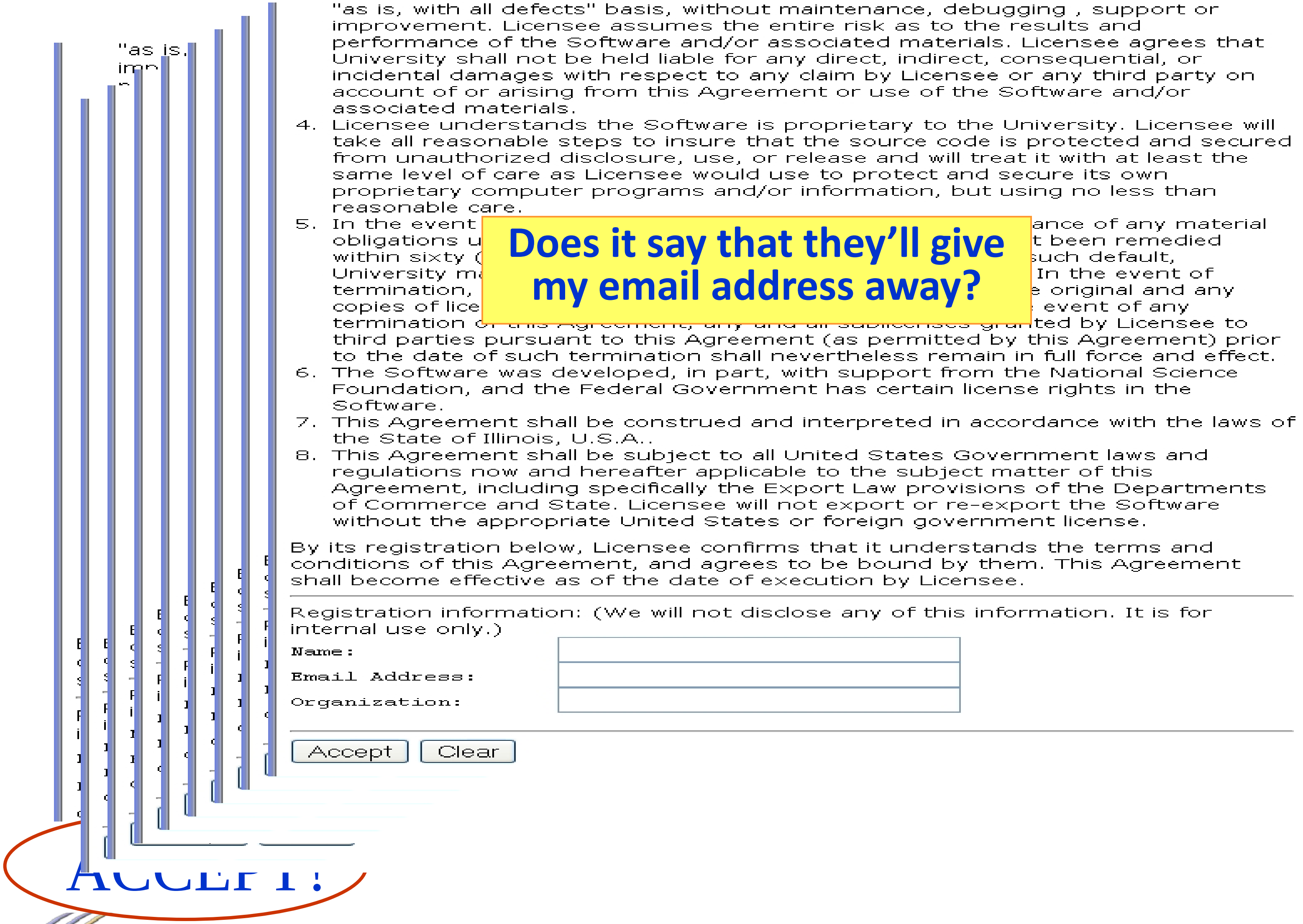
Email Address:

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 m In the event of
termination, e original and any
copies of lice event of any
termination o 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.

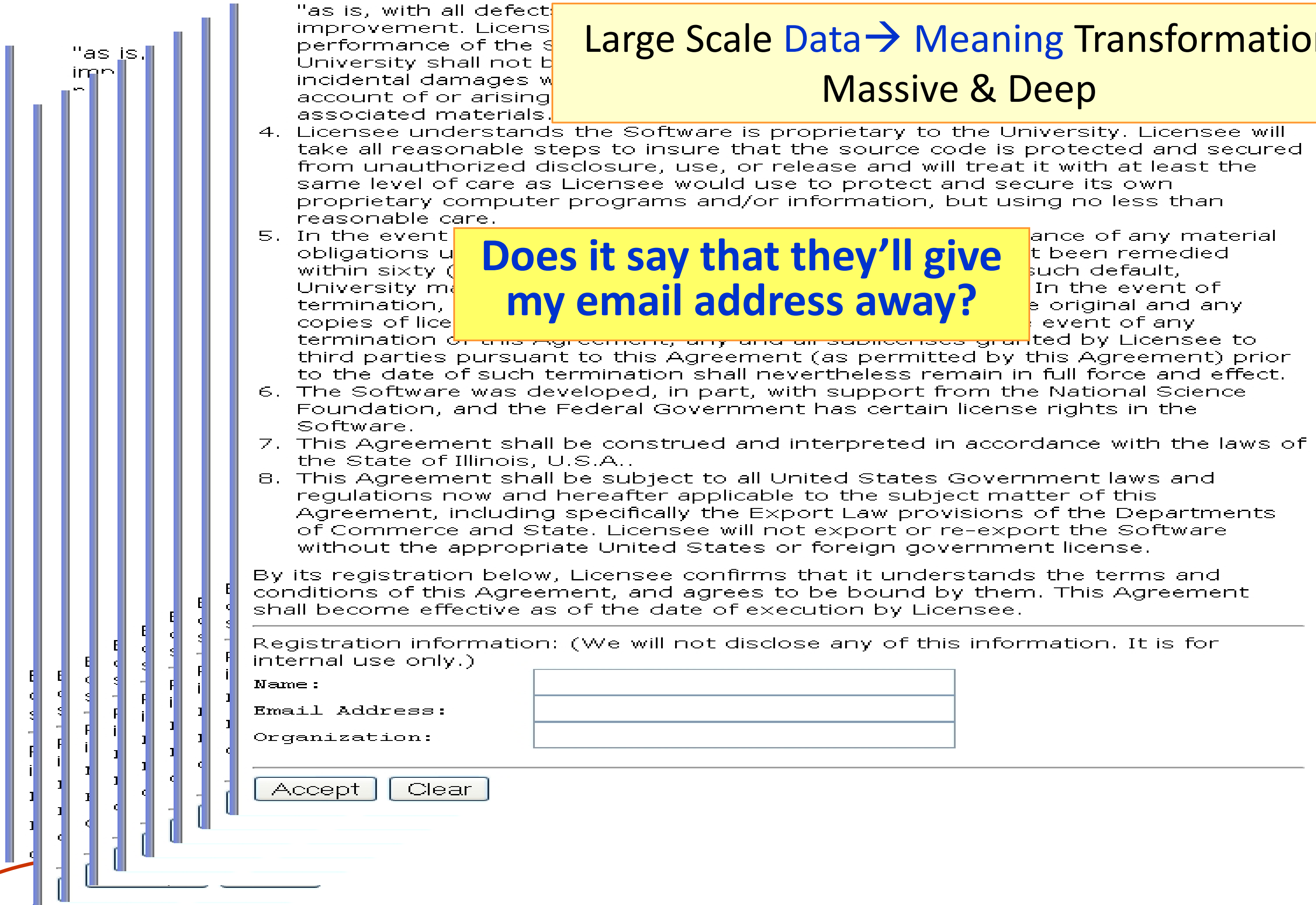
Registration information: (We will not disclose any of this information. It is for internal use only.)

Name:	
Email Address:	
Organization:	

Accept | Clear

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

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 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:

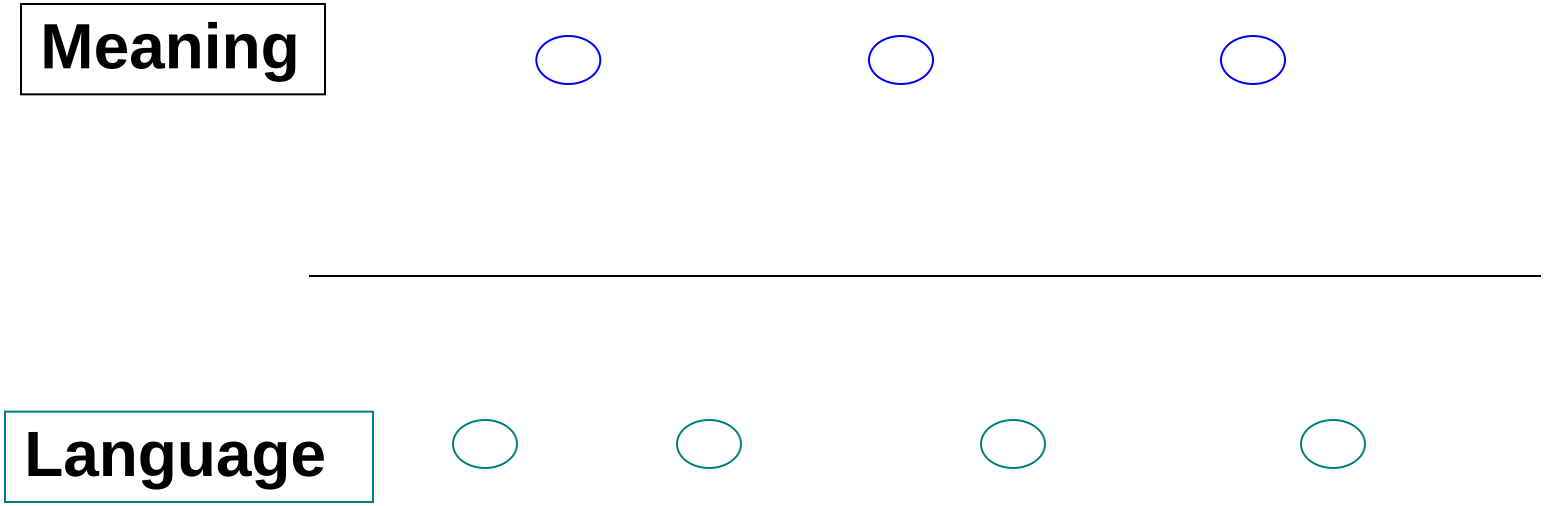
Large Scale Data→ Meaning Transformation

Massive & Deep

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

ACCEPT?

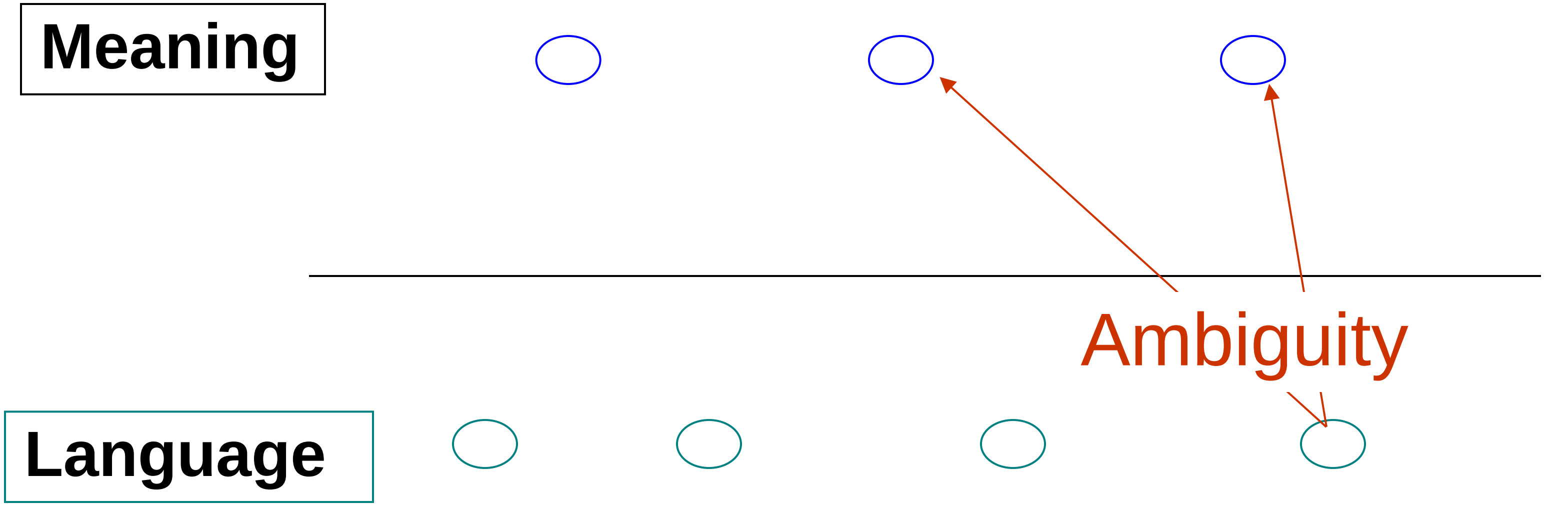
Why is it difficult?



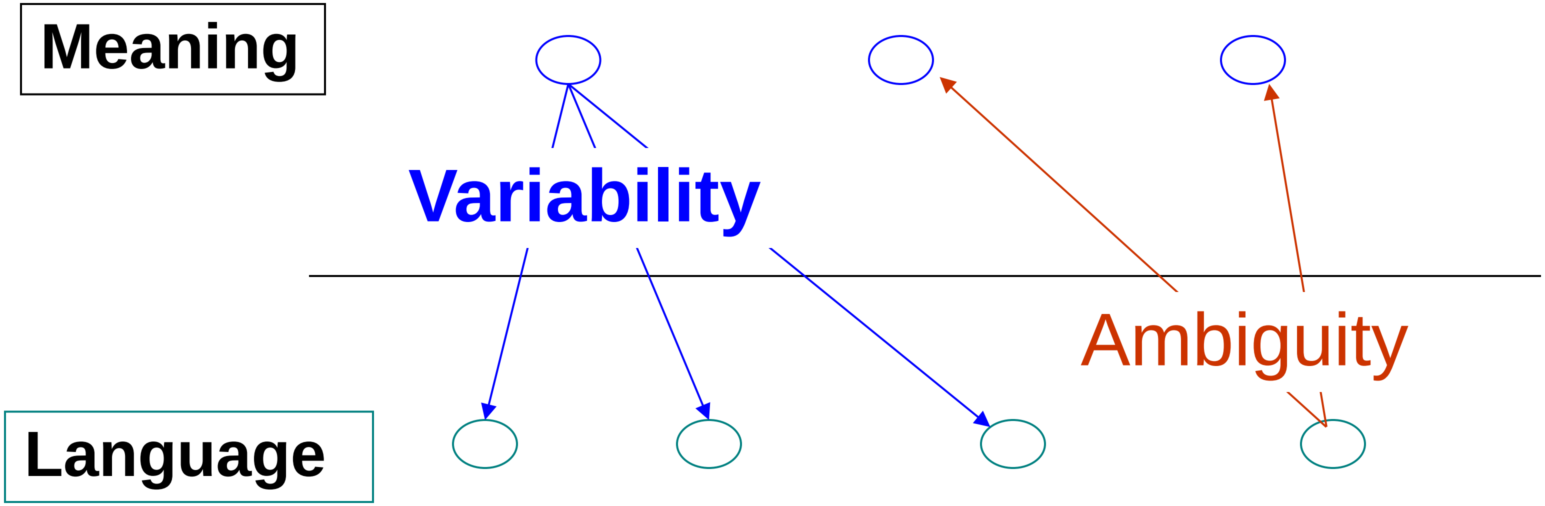
Meaning

Language

Why is it difficult?



Why is it difficult?



Ambiguity

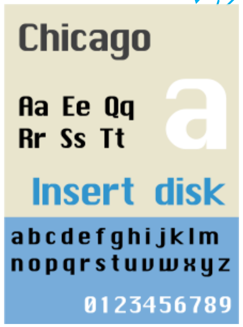
It’s a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the ”N”.	***Chicago*** was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..	***Chicago VIII*** was one of the early 70s-era ***Chicago***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

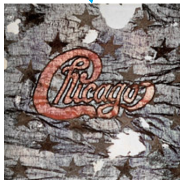
Ambiguity

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***.



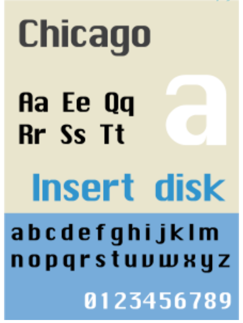


Ambiguity

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..

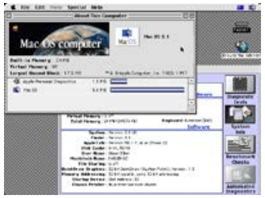
Chicago VIII was one of the early 70s-era Chicago albums to catch my ear, along with Chicago II.

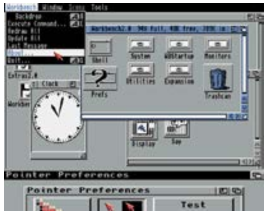

Chicago

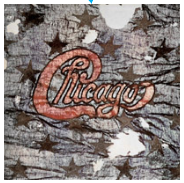
Aa Ee Qq
Rr Ss Tt

Insert disk

abcdefghijklm
nopqrstuvwxyz
0123456789



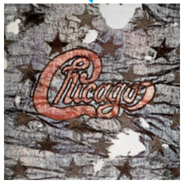




Ambiguity

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***.

Chicago
Aa Ee Qq
Rr Ss Tt
a
Insert disk
abcdefghijklm
nopqrstuvwxyz
0123456789



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.

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.

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.

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.

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.

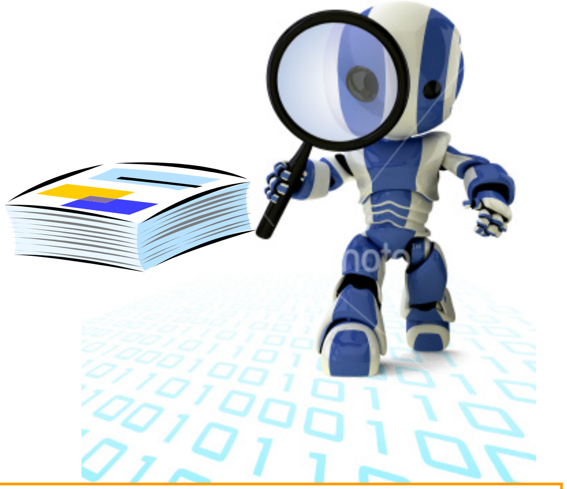
Jim Carpenter worked in a number of important positions. …. As a press liaison for the IRS, he made contacts in the white house.

- 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.

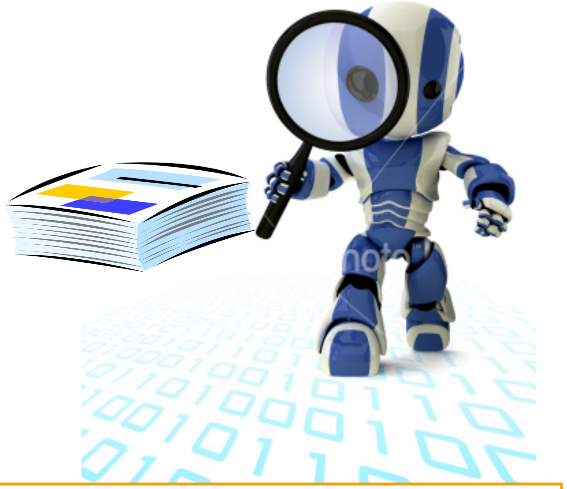
Jim Carpenter worked in a number of important positions. …. As a press liaison for the IRS, he made contacts in the white house.

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:

- ❑ Relations, Entities and Semantic Classes, NOT keywords
- ❑ Bring knowledge from external resources
- ❑ Integrate over large collections of text and DBs
- ❑ Identify, disambiguate and track entities, events, etc.

COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

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 training facility of a large corporation wants to provide new employees easy access to all relevant key concepts, entities (people, techniques, applications) along with relevant projects and background information when they read material about their new job.

What can this give us?

- Moving towards natural language understanding…

- 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?

- Moving towards natural language understanding…

- 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?

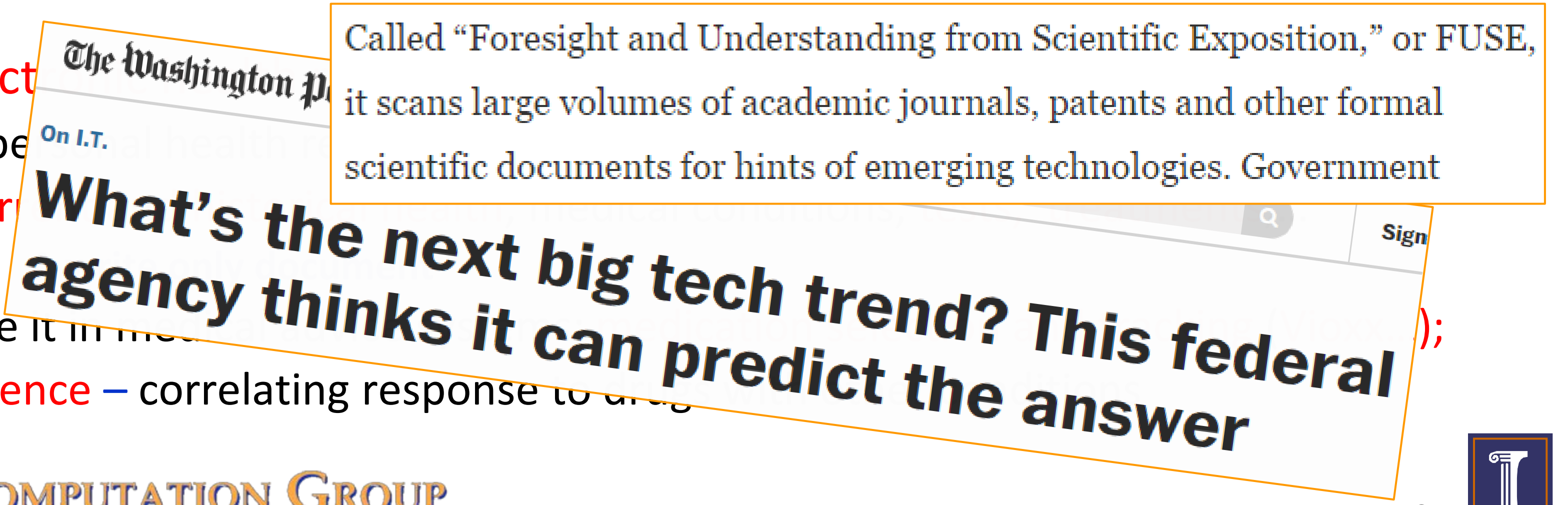
- An electronic health record (EHR):
  - A personal health record in digital format. Includes information relating to:
  - Current and historical health, medical conditions, tests, treatments,…
    - **A write only document**
  - Use it in medical advice systems; medication selection and tracking (Vioxx…);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

What can this give us?

- Moving towards natural language understanding…

- 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 elect
  - A pe
  - Cur
    -
  - Use it in med);
  - Science – correlating response to drug



The Washington Post

On I.T.

What's the next big tech trend? This federal agency thinks it can predict the answer

Called "Foresight and Understanding from Scientific Exposition," or FUSE, it scans large volumes of academic journals, patents and other formal scientific documents for hints of emerging technologies. Government

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 exploiting Indirect Supervision (data abound; not supervised)

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 exploiting Indirect Supervision (data abound; not supervised)

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 exploiting Indirect Supervision (data abound; not supervised)

Well understood; easy to build black box categorizers

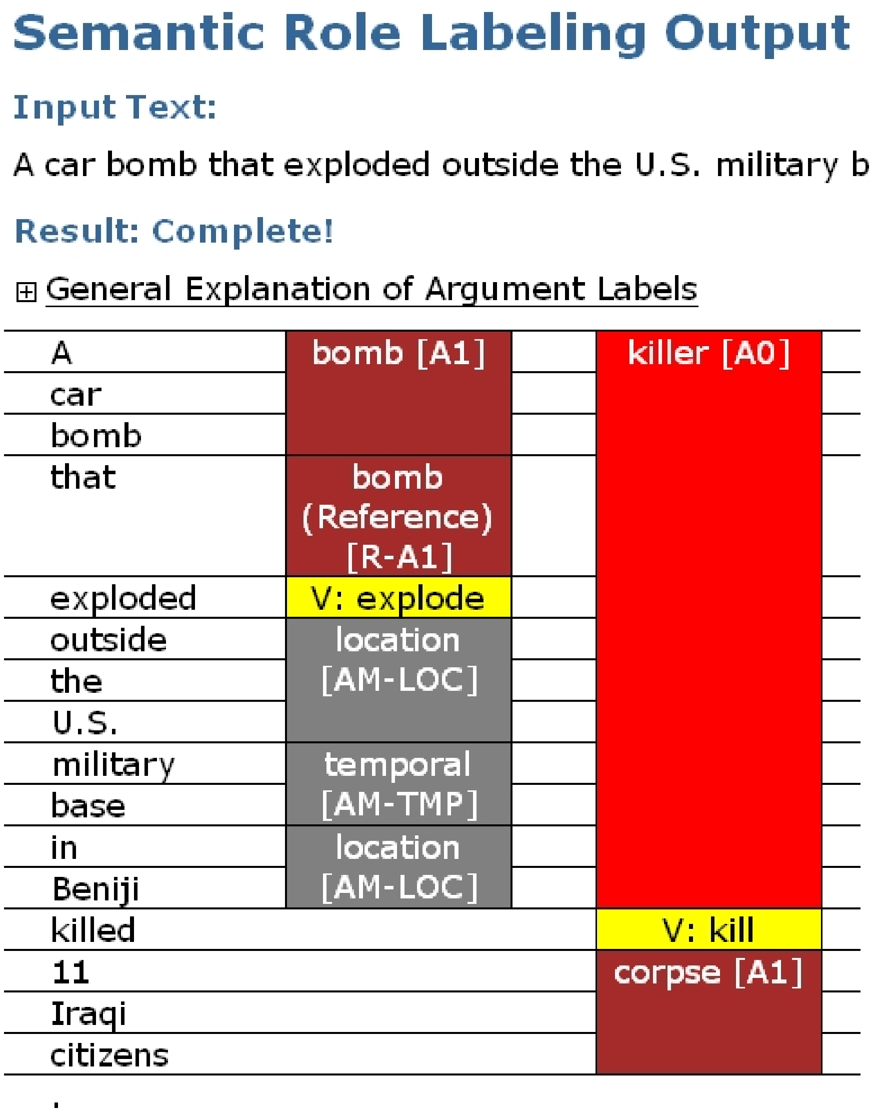
- Inference over learned models as a way to "put things together", introduce domain & task specific knowledge and constraints
  - □ Constrained Conditional Models: formulating inference as ILP

    Learn models; Acquire knowledge/constraints; Make decisions.

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

COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Extracting Relations via Semantic Analysis



Semantic Role Labeling Output

Input Text:

A car bomb that exploded outside the U.S. military base in Beniji killed 11 Iraqi citizens.

Result: Complete!

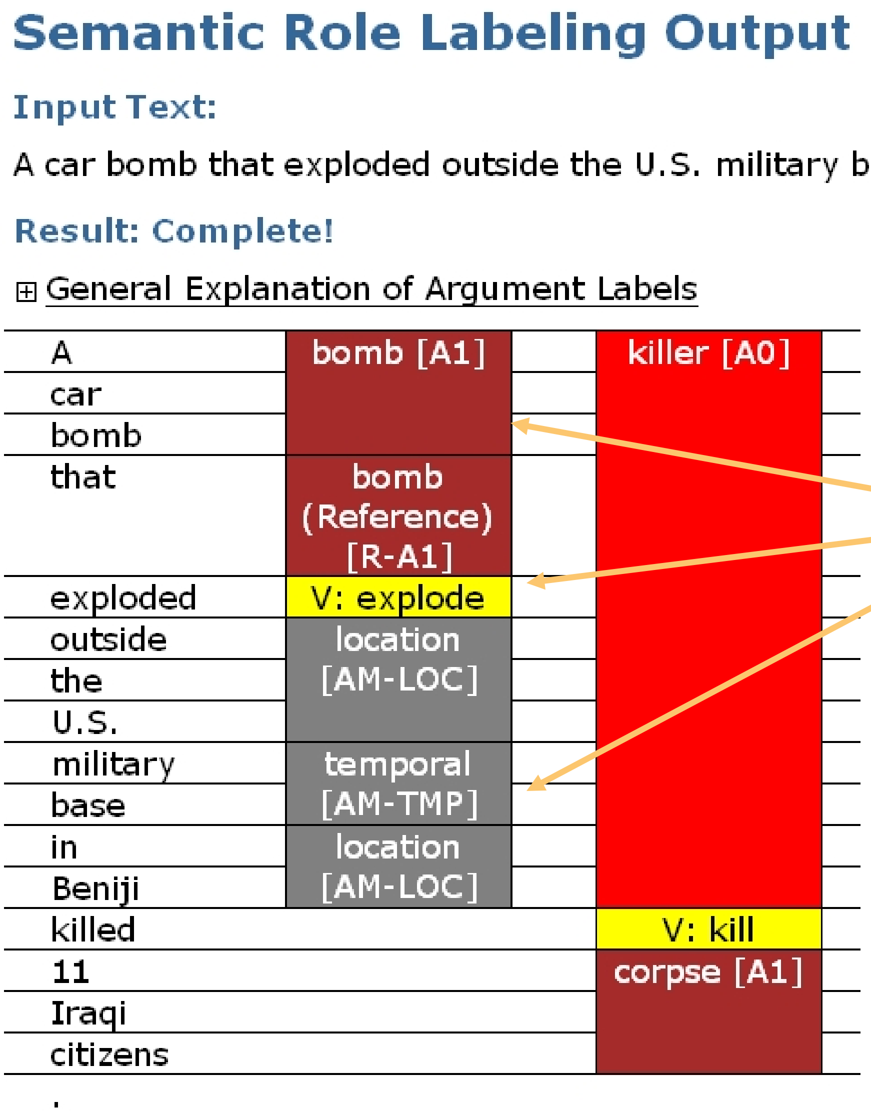
⊞ General Explanation of Argument Labels

A	bomb [A1]	killer [A0]
car		
bomb		
that	bomb (Reference) [R-A1]	
exploded	V: explode	
outside	location [AM-LOC]	
the		
U.S.		
military	temporal [AM-TMP]	
base		
in	location [AM-LOC]	
Beniji		
killed		V: kill
11		corpse [A1]
Iraqi		
citizens		
.		

- Semantic parsing reveals several relations in the sentence along with their arguments.

Extracting Relations via Semantic Analysis

Screen shot from a CCG demo
http://cogcomp.cs.illinois.edu/page/demos



Semantic Role Labeling Output

Input Text:

A car bomb that exploded outside the U.S. military base in Beniji killed 11 Iraqi citizens.

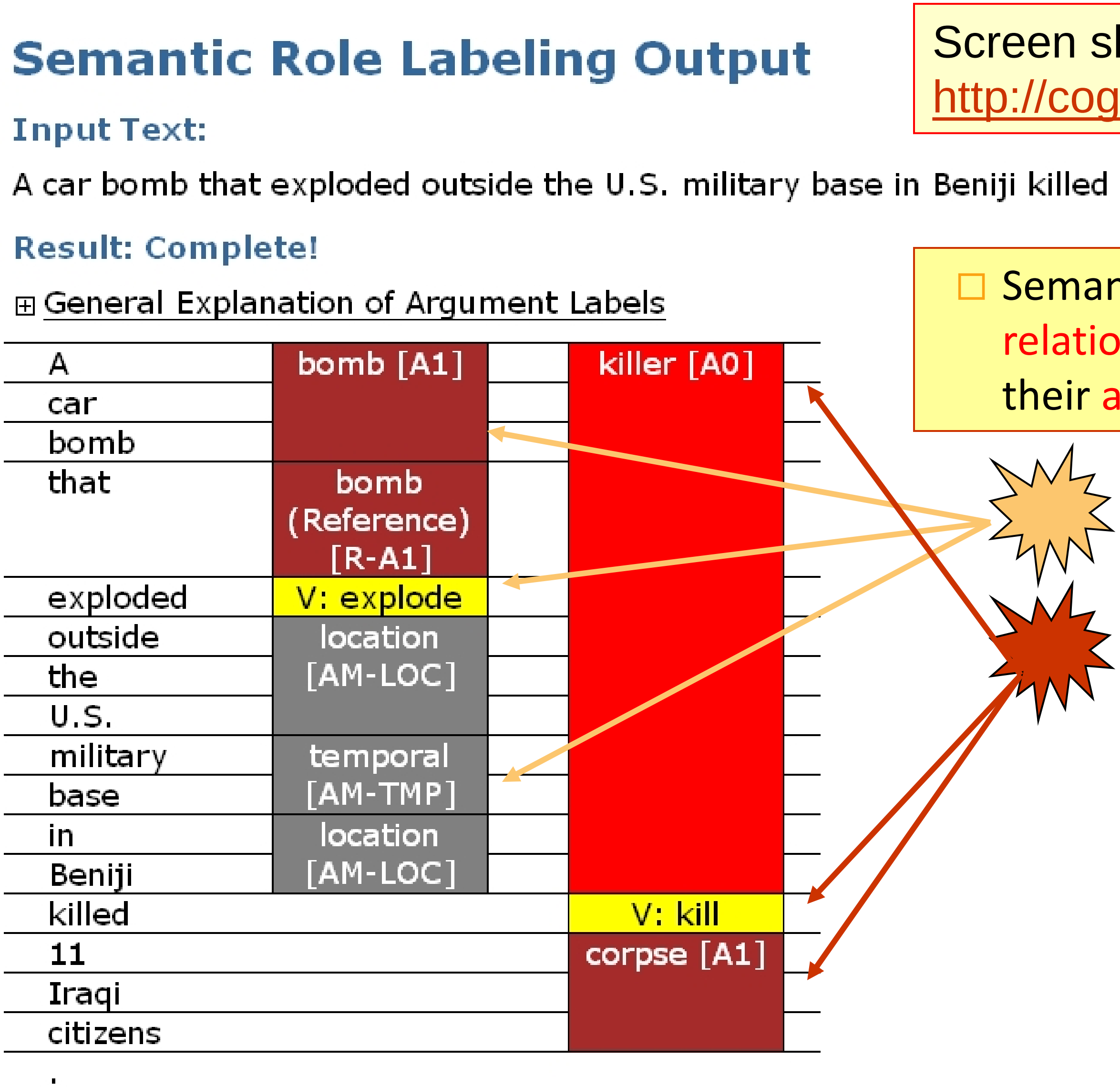
Result: Complete!

⊞ General Explanation of Argument Labels

A	bomb [A1]	killer [A0]
car		
bomb		
that	bomb (Reference) [R-A1]	
exploded	V: explode	
outside	location [AM-LOC]	
the		
U.S.		
military	temporal [AM-TMP]	
base		
in	location [AM-LOC]	
Beniji		
killed		V: kill
11		corpse [A1]
Iraqi		
citizens		
.		

- ☐ Semantic parsing reveals several relations in the sentence along with their arguments.

Extracting Relations via Semantic Analysis



Semantic Role Labeling Output

Input Text:

A car bomb that exploded outside the U.S. military base in Beniji killed 11 Iraqi citizens.

Result: Complete!

⊞ General Explanation of Argument Labels

A	bomb [A1]	killer [A0]
car		
bomb		
that	bomb (Reference) [R-A1]	
exploded	V: explode	
outside	location [AM-LOC]	
the		
U.S.		
military	temporal [AM-TMP]	
base		
in	location [AM-LOC]	
Beniji		
killed		V: kill
11		corpse [A1]
Iraqi		
citizens		
.		

Screen shot from a CCG demo
http://cogcomp.cs.illinois.edu/page/demos

- Semantic parsing reveals several relations in the sentence along with their arguments.

Extended Semantic Role labeling

- Beyond verb predicates
- 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

- Beyond verb predicates
- 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.

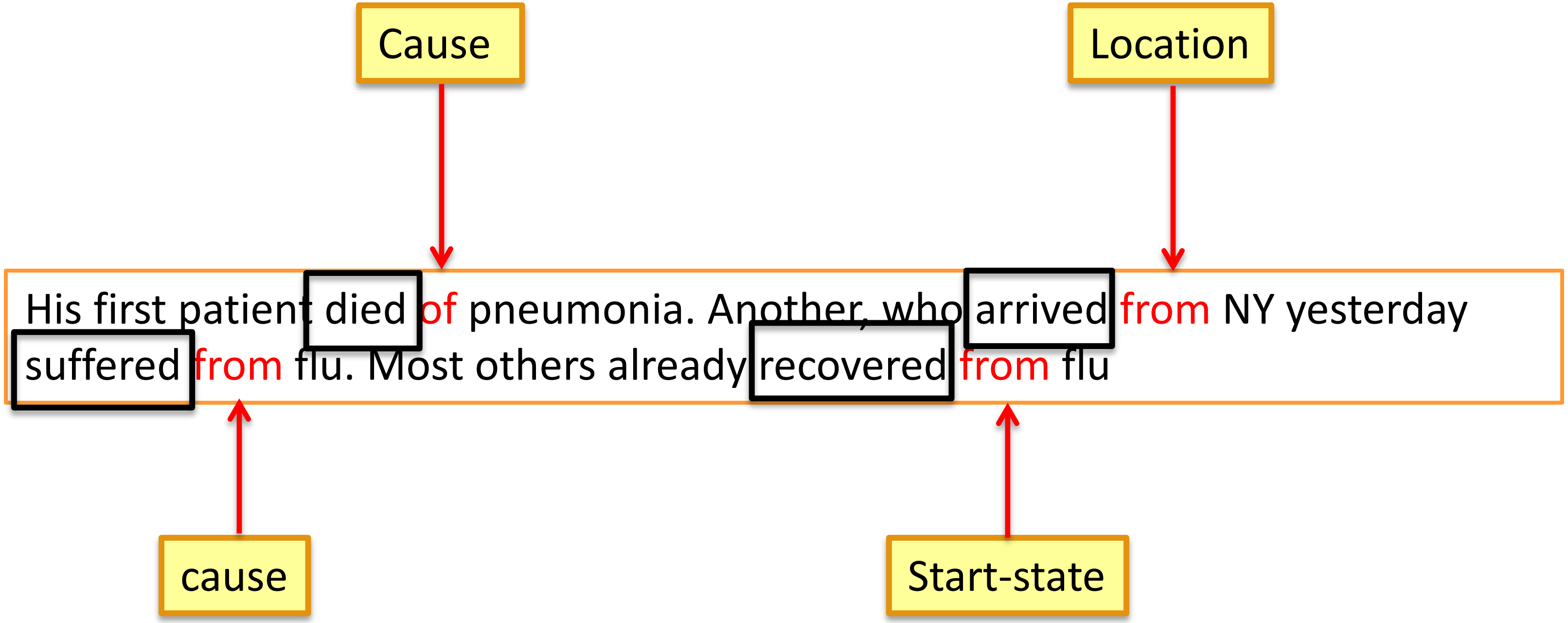
- Beyond verb predicates
- Ambiguity and Variability of Prepositional Relations

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.

- Beyond verb predicates
- 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

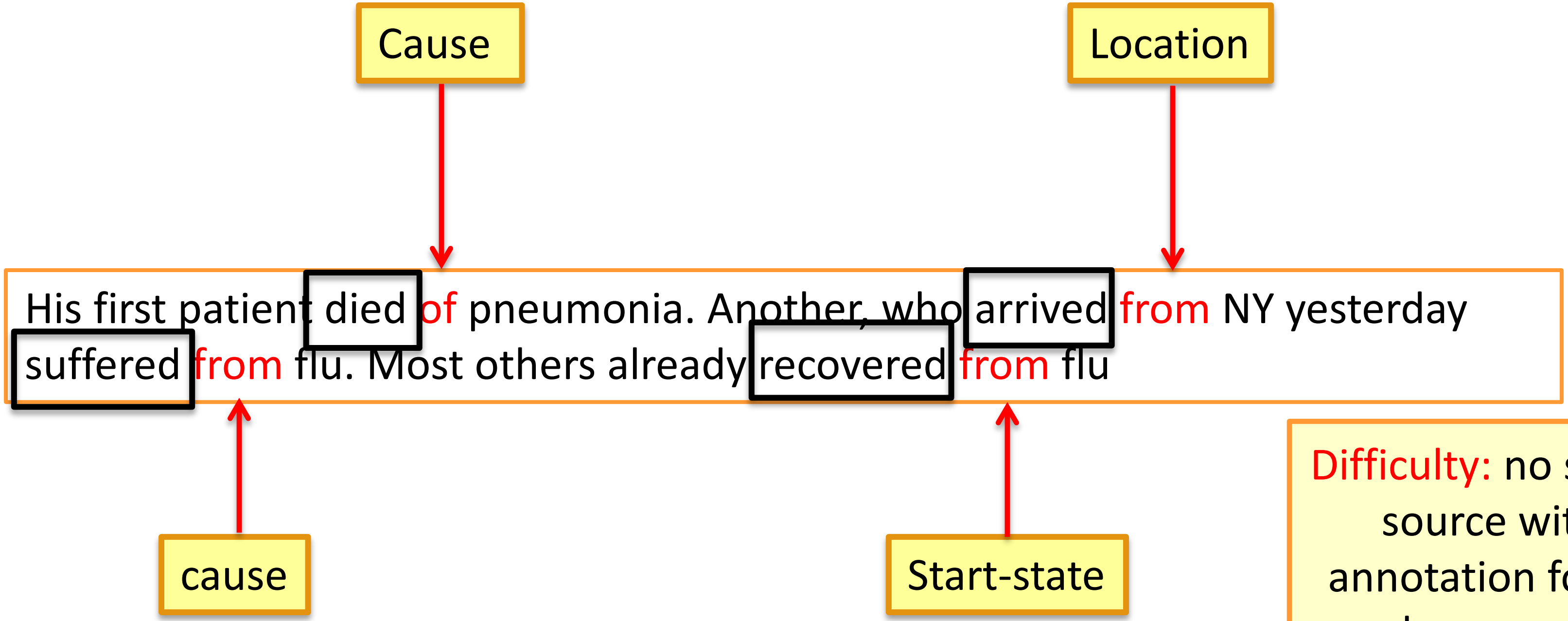
cause

Start-state

Extended Semantic Role labeling

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

- Beyond verb predicates
- 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

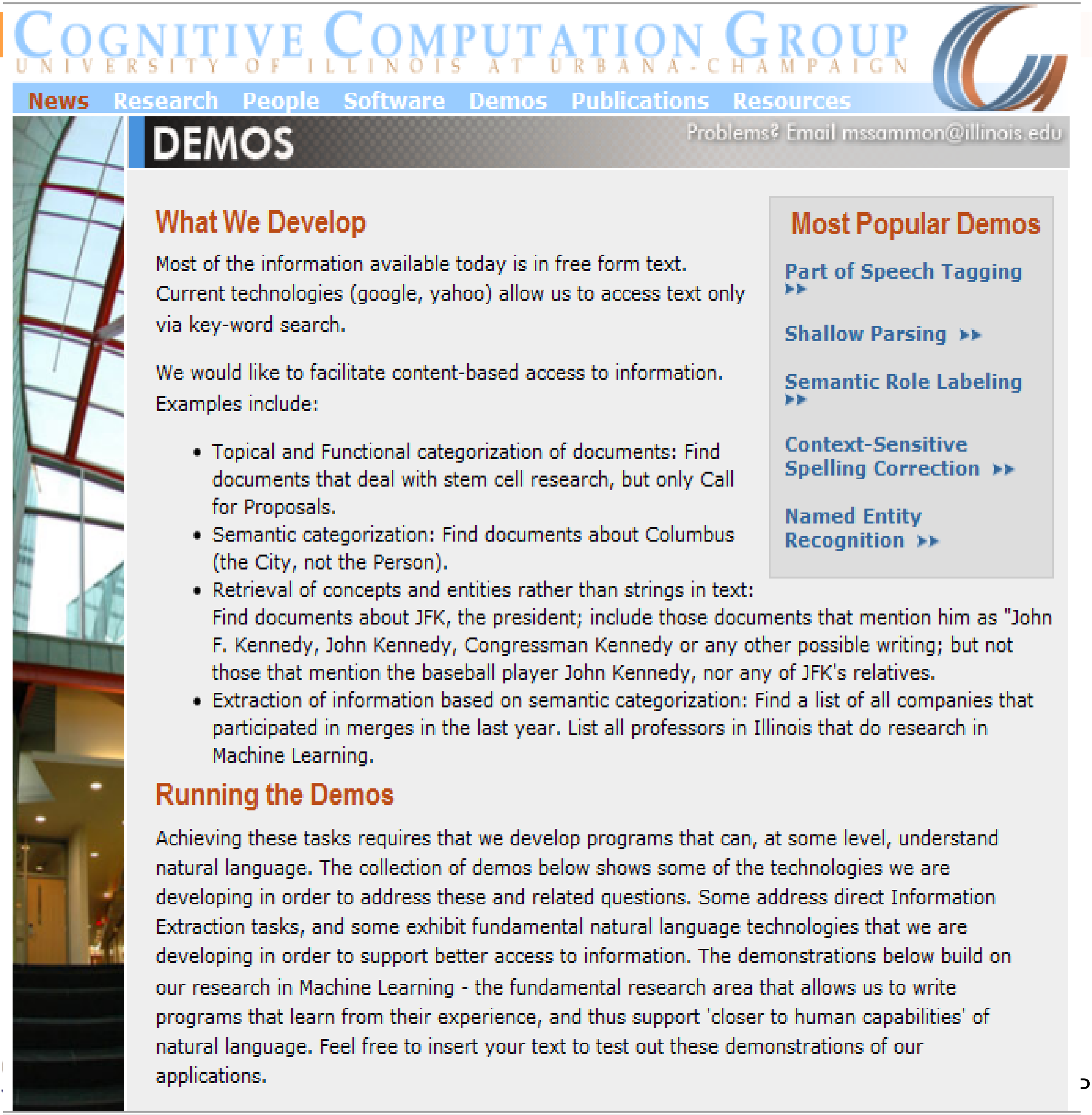
Difficulty: no single source with annotation for all phenomena

Learn models; Acquire knowledge/constraints; Make decisions.

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

Significant Progress in NLP and Information Extraction

Significant Progress in NLP and Information Extraction



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.

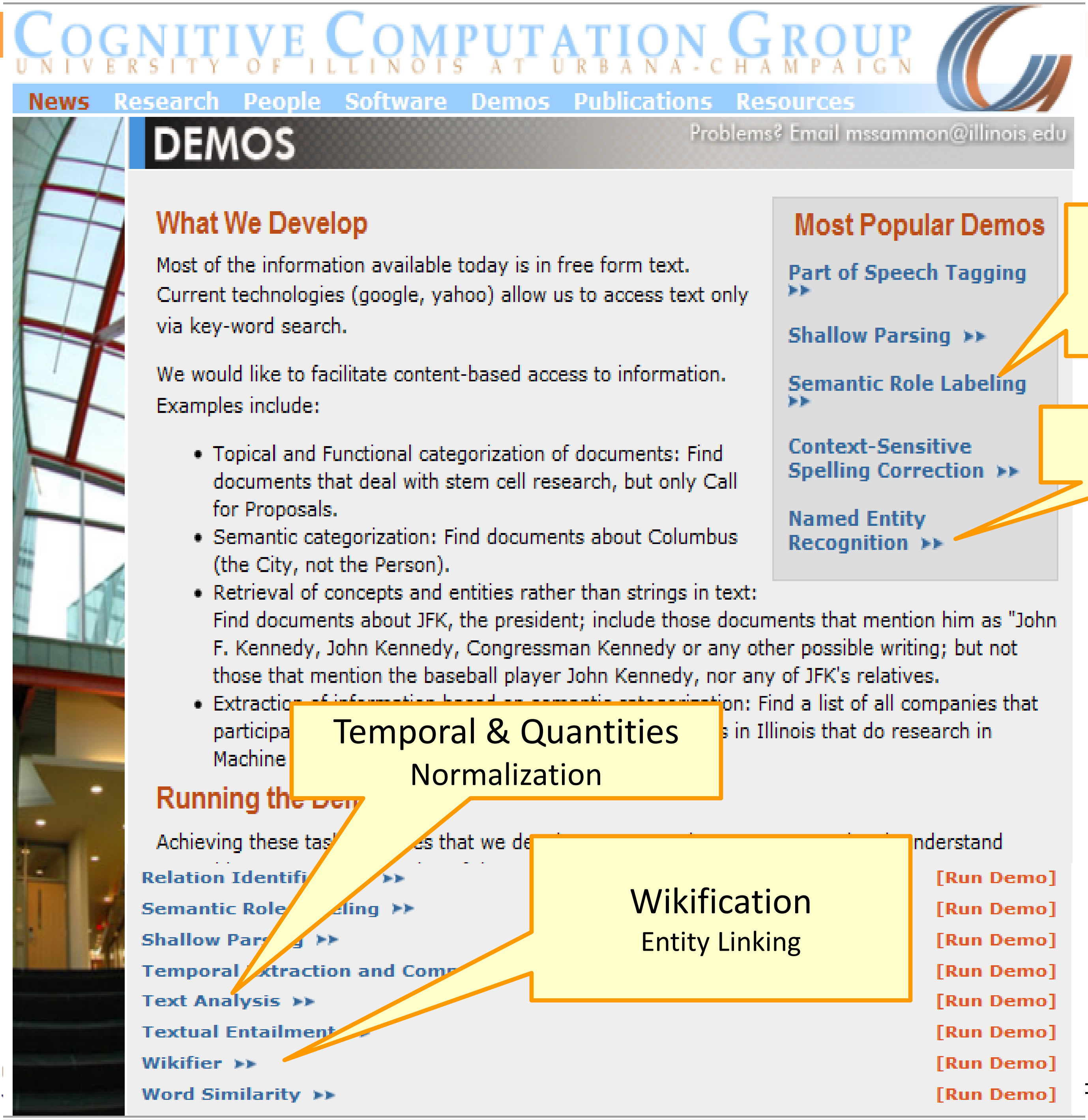
Most Popular Demos

- 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 natural language. The collection of demos below shows some of the technologies we are developing in order to address these and related questions. Some address direct Information Extraction tasks, and some exhibit fundamental natural language technologies that we are developing in order to support better access to information. The demonstrations below build on our research in Machine Learning - the fundamental research area that allows us to write programs that learn from their experience, and thus support 'closer to human capabilities' of natural language. Feel free to insert your text to test out these demonstrations of our applications.

Significant Progress in NLP and Information Extraction



COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

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.
- Extractio on: Find a list of all companies that participa s in Illinois that do research in Machine

Most Popular Demos

Part of Speech Tagging ▸▸

Shallow Parsing ▸▸

Semantic Role Labeling ▸▸

Context-Sensitive Spelling Correction ▸▸

Named Entity Recognition ▸▸

Running the De

Achieving these tas es that we de nderstand

Relation Identifi ▸▸	[Run Demo]
Semantic Role eling ▸▸	[Run Demo]
Shallow Pars g ▸▸	[Run Demo]
Temporal xtraction and Comp	[Run Demo]
Text Analysis ▸▸	[Run Demo]
Textual Entailment	[Run Demo]
Wikifier ▸▸	[Run Demo]
Word Similarity ▸▸	[Run Demo]

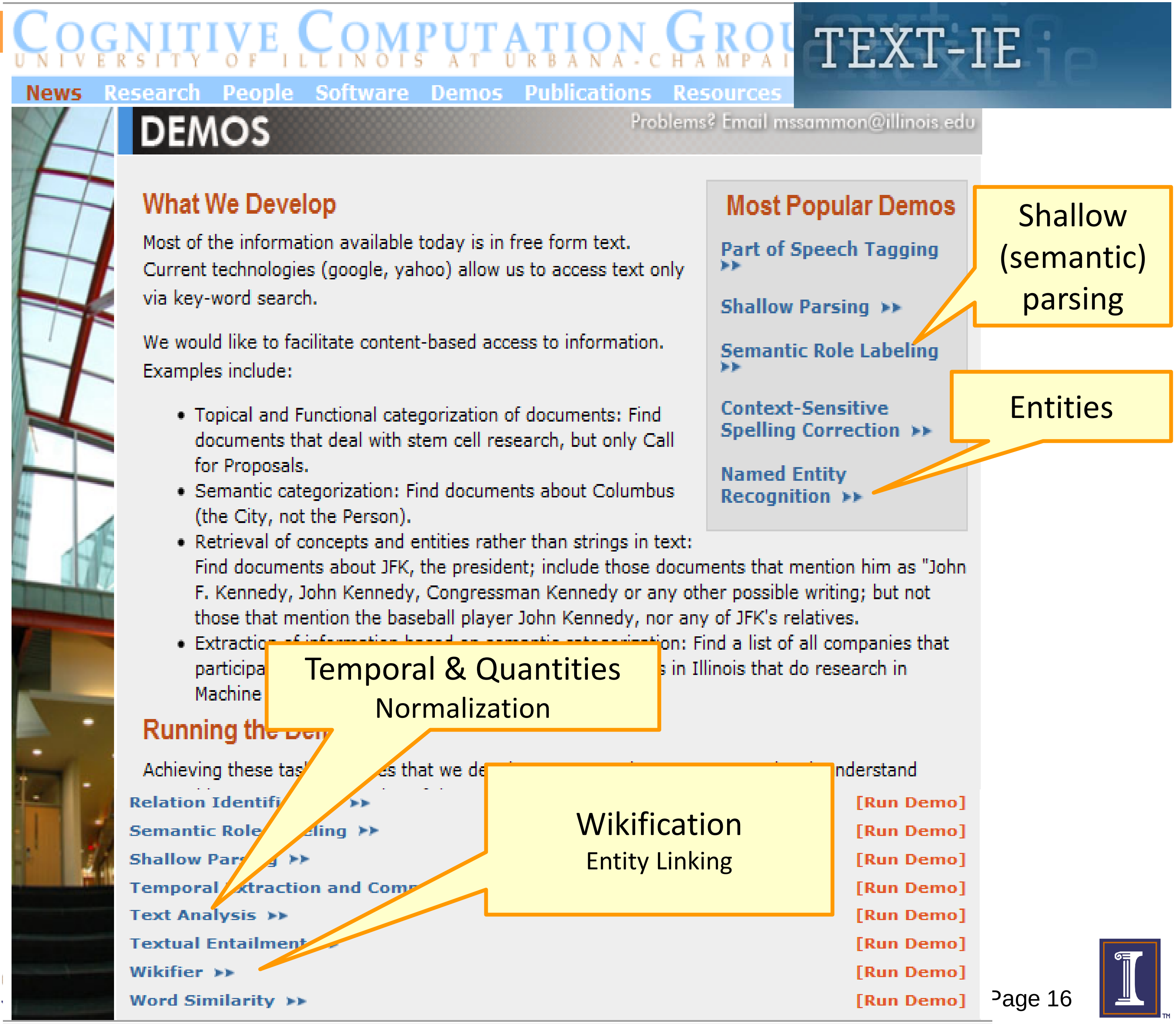
Shallow (semantic) parsing

Entities

Temporal & Quantities
Normalization

Wikification
Entity Linking

Significant Progress in NLP and Information Extraction



Cognitive Computation Group
University of Illinois at Urbana-Champai
TEXT-IE
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.
- Extractio ... on: Find a list of all companies that participa ... s in Illinois that do research in Machine

Running the De

Achieving these tas ... es that we de ... nderstand

- Relation Identifi ▸▸ [Run Demo]
- Semantic Role ...eling ▸▸ [Run Demo]
- Shallow Par...g ▸▸ [Run Demo]
- Temporal ...xtraction and Comp [Run Demo]
- Text Analysis ▸▸ [Run Demo]
- Textual Entailment [Run Demo]
- Wikifier ▸▸ [Run Demo]
- Word Similarity ▸▸ [Run Demo]

Most Popular Demos

- Part of Speech Tagging ▸▸
- Shallow Parsing ▸▸
- Semantic Role Labeling ▸▸
- Context-Sensitive Spelling Correction ▸▸
- Named Entity Recognition ▸▸

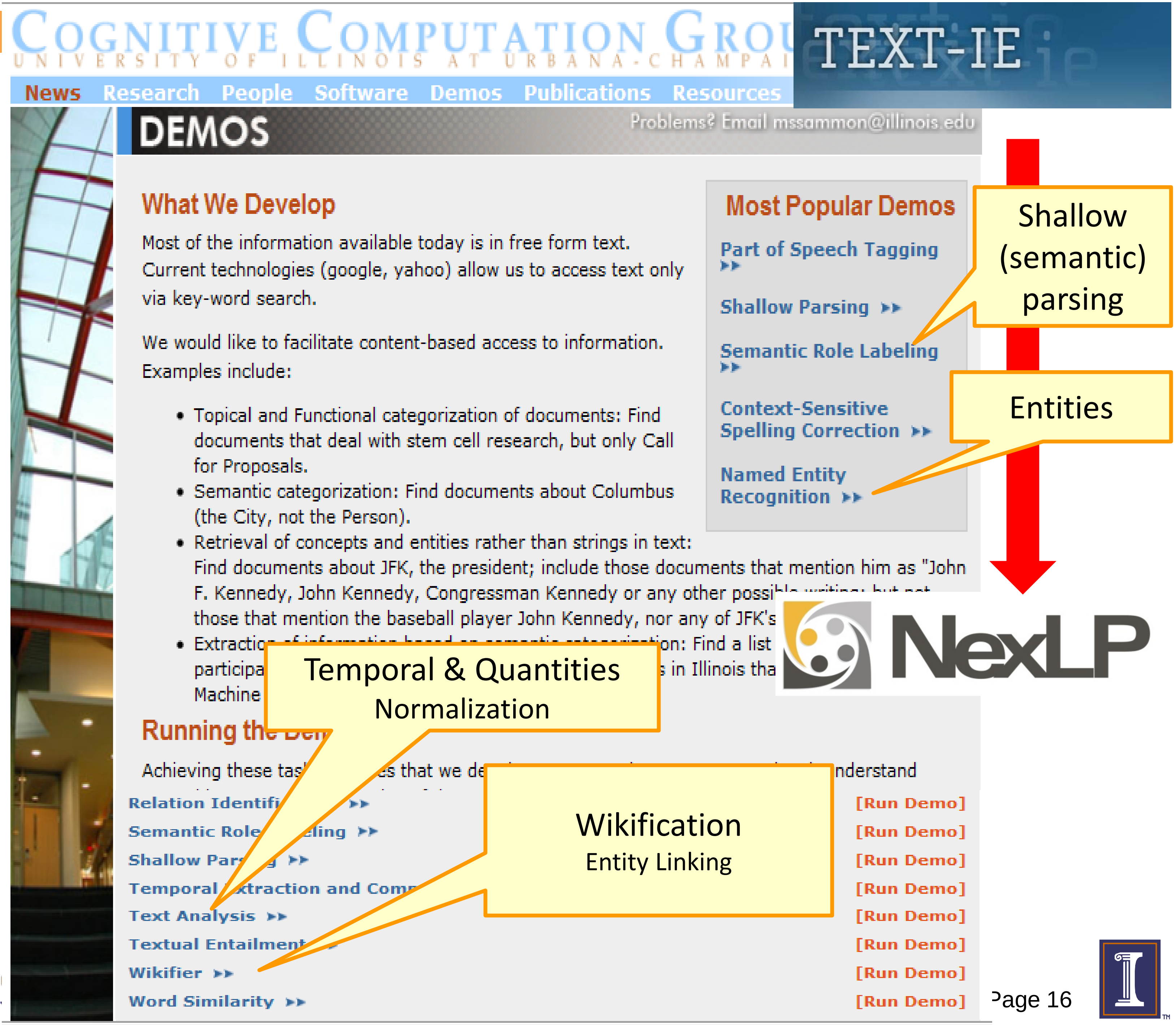
Shallow (semantic) parsing

Entities

Temporal & Quantities
Normalization

Wikification
Entity Linking

Significant Progress in NLP and Information Extraction



COGNITIVE COMPUTATION GROU
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAI
News Research People Software Demos Publications Resources
TEXT-IE

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 possi
those that mention the baseball player John Kennedy, nor any of JFK's
- Extractio on: Find a list
participa s in Illinois tha
Machine

Most Popular Demos

- Part of Speech Tagging ▸▸
- Shallow Parsing ▸▸
- Semantic Role Labeling ▸▸
- Context-Sensitive Spelling Correction ▸▸
- Named Entity Recognition ▸▸

Running the De

Achieving these tas es that we de nderstand

Relation Identifi ▸▸	[Run Demo]
Semantic Role eling ▸▸	[Run Demo]
Shallow Pars g ▸▸	[Run Demo]
Temporal xtraction and Comp	[Run Demo]
Text Analysis ▸▸	[Run Demo]
Textual Entailment	[Run Demo]
Wikifier ▸▸	[Run Demo]
Word Similarity ▸▸	[Run Demo]

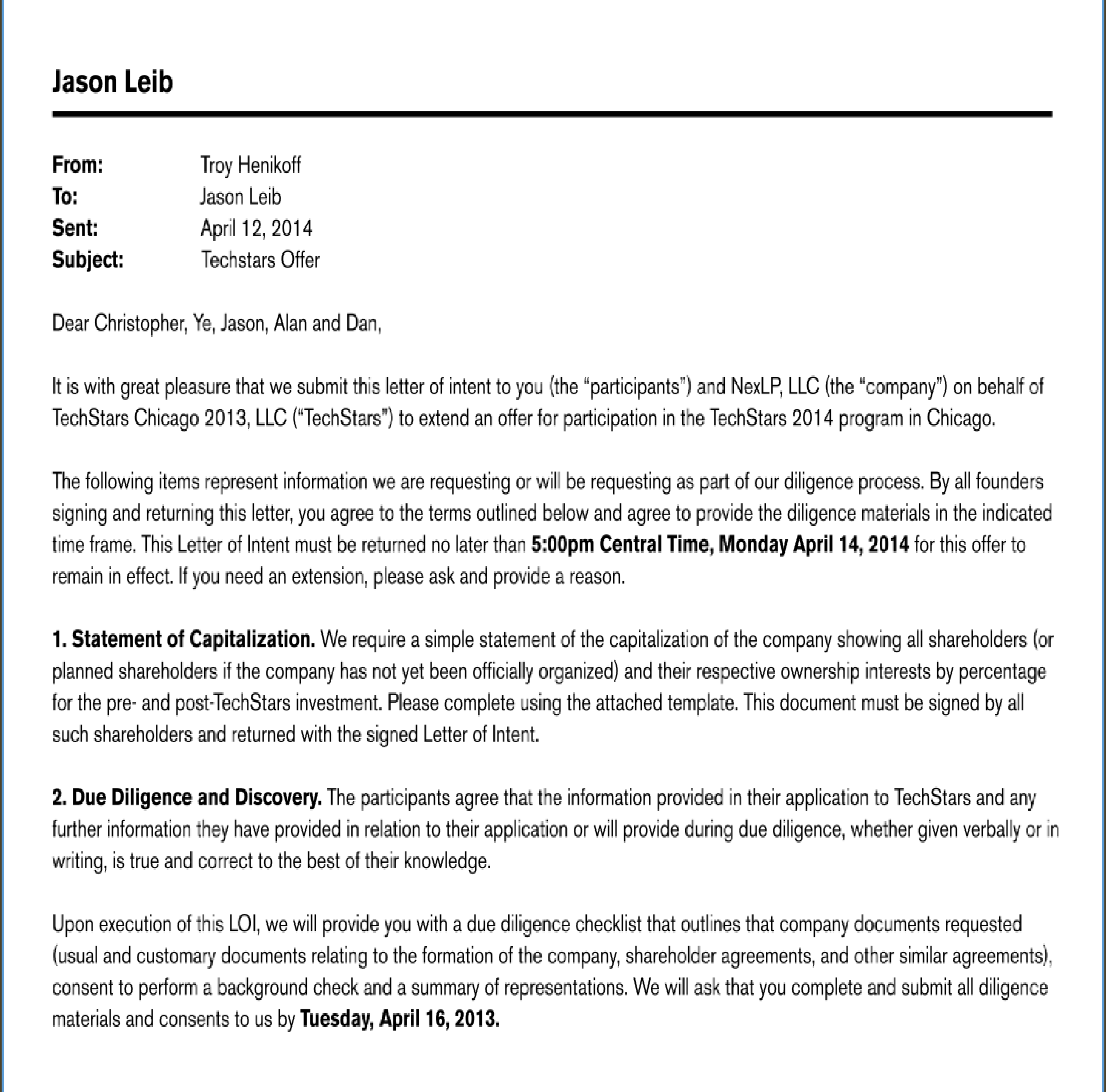
Shallow (semantic) parsing

Entities

Temporal & Quantities Normalization

Wikification Entity Linking

NexLP



Jason Leib

From: Troy Henikoff
To: Jason Leib
Sent: April 12, 2014
Subject: Techstars Offer

Dear Christopher, Ye, Jason, Alan and Dan,

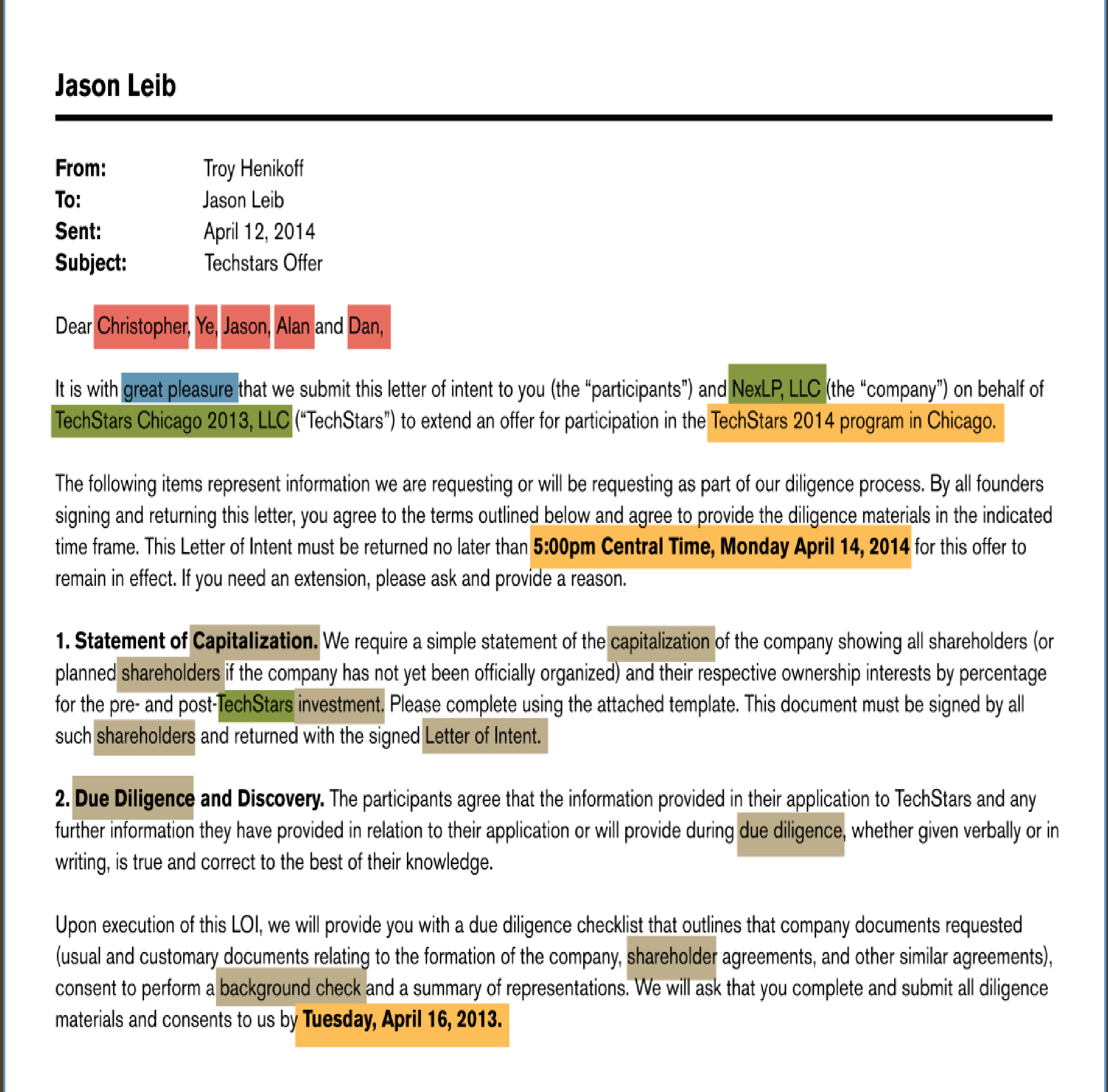
It is with great pleasure that we submit this letter of intent to you (the "participants") and NexLP, LLC (the "company") on behalf of TechStars Chicago 2013, LLC ("TechStars") to extend an offer for participation in the TechStars 2014 program in Chicago.

The following items represent information we are requesting or will be requesting as part of our diligence process. By all founders signing and returning this letter, you agree to the terms outlined below and agree to provide the diligence materials in the indicated time frame. This Letter of Intent must be returned no later than **5:00pm Central Time, Monday April 14, 2014** for this offer to remain in effect. If you need an extension, please ask and provide a reason.

1. Statement of Capitalization. We require a simple statement of the capitalization of the company showing all shareholders (or planned shareholders if the company has not yet been officially organized) and their respective ownership interests by percentage for the pre- and post-TechStars investment. Please complete using the attached template. This document must be signed by all such shareholders and returned with the signed Letter of Intent.

2. Due Diligence and Discovery. The participants agree that the information provided in their application to TechStars and any further information they have provided in relation to their application or will provide during due diligence, whether given verbally or in writing, is true and correct to the best of their knowledge.

Upon execution of this LOI, we will provide you with a due diligence checklist that outlines that company documents requested (usual and customary documents relating to the formation of the company, shareholder agreements, and other similar agreements), consent to perform a background check and a summary of representations. We will ask that you complete and submit all diligence materials and consents to us by **Tuesday, April 16, 2013.**



Jason Leib

From: Troy Henikoff
To: Jason Leib
Sent: April 12, 2014
Subject: Techstars Offer

Dear Christopher, Ye, Jason, Alan and Dan,

It is with great pleasure that we submit this letter of intent to you (the "participants") and NexLP, LLC (the "company") on behalf of TechStars Chicago 2013, LLC ("TechStars") to extend an offer for participation in the TechStars 2014 program in Chicago.

The following items represent information we are requesting or will be requesting as part of our diligence process. By all founders signing and returning this letter, you agree to the terms outlined below and agree to provide the diligence materials in the indicated time frame. This Letter of Intent must be returned no later than **5:00pm Central Time, Monday April 14, 2014** for this offer to remain in effect. If you need an extension, please ask and provide a reason.

1. Statement of Capitalization. We require a simple statement of the capitalization of the company showing all shareholders (or planned shareholders if the company has not yet been officially organized) and their respective ownership interests by percentage for the pre- and post-TechStars investment. Please complete using the attached template. This document must be signed by all such shareholders and returned with the signed Letter of Intent.

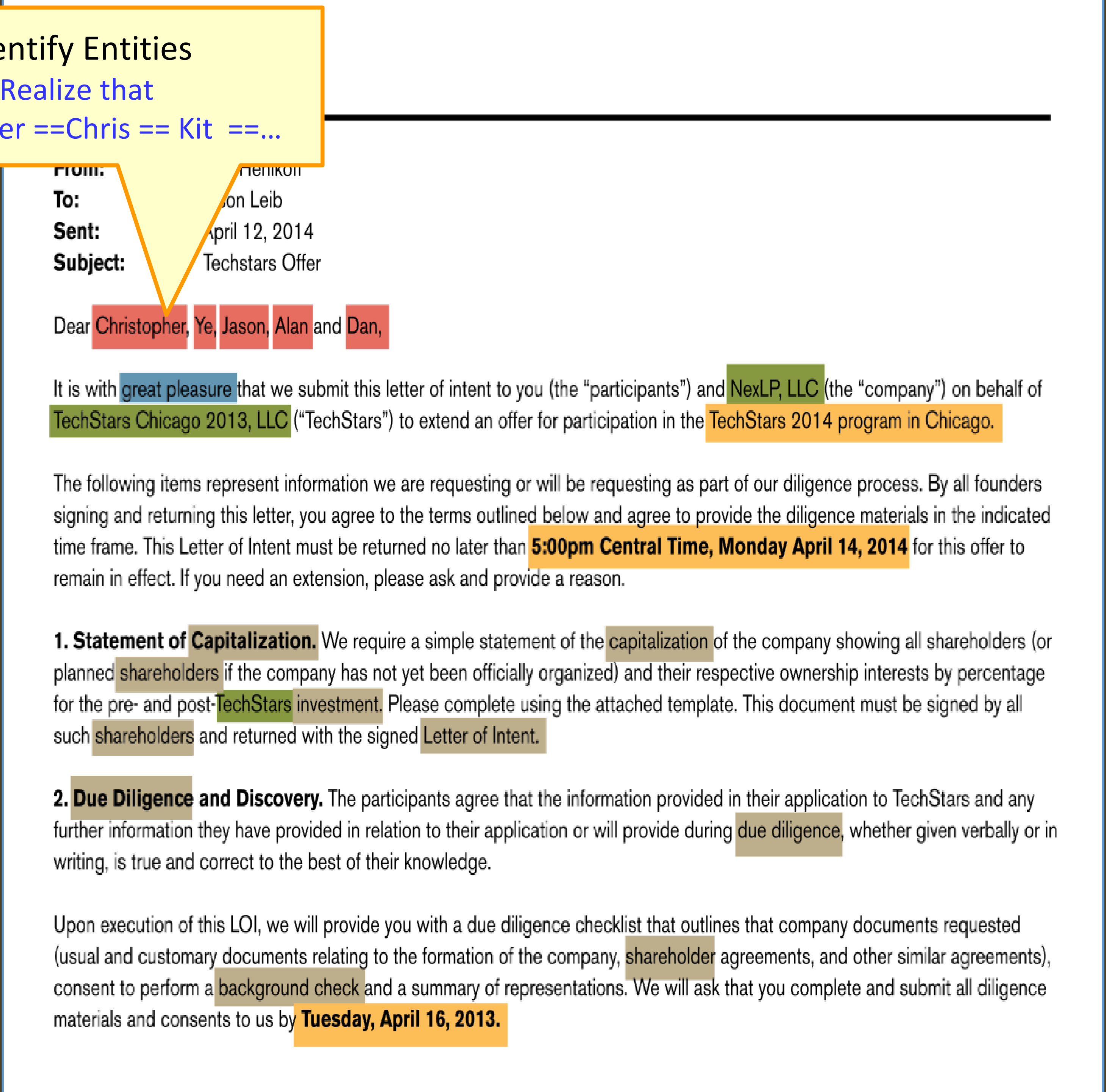
2. Due Diligence and Discovery. The participants agree that the information provided in their application to TechStars and any further information they have provided in relation to their application or will provide during due diligence, whether given verbally or in writing, is true and correct to the best of their knowledge.

Upon execution of this LOI, we will provide you with a due diligence checklist that outlines that company documents requested (usual and customary documents relating to the formation of the company, shareholder agreements, and other similar agreements), consent to perform a background check and a summary of representations. We will ask that you complete and submit all diligence materials and consents to us by **Tuesday, April 16, 2013.**

Identify Entities

Realize that

Christopher ==Chris == Kit ==...



From: Henikoff
To: on Leib
Sent: pril 12, 2014
Subject: Techstars Offer

Dear Christopher, Ye, Jason, Alan and Dan,

It is with great pleasure that we submit this letter of intent to you (the "participants") and NexLP, LLC (the "company") on behalf of TechStars Chicago 2013, LLC ("TechStars") to extend an offer for participation in the TechStars 2014 program in Chicago.

The following items represent information we are requesting or will be requesting as part of our diligence process. By all founders signing and returning this letter, you agree to the terms outlined below and agree to provide the diligence materials in the indicated time frame. This Letter of Intent must be returned no later than **5:00pm Central Time, Monday April 14, 2014** for this offer to remain in effect. If you need an extension, please ask and provide a reason.

1. Statement of Capitalization. We require a simple statement of the capitalization of the company showing all shareholders (or planned shareholders if the company has not yet been officially organized) and their respective ownership interests by percentage for the pre- and post-TechStars investment. Please complete using the attached template. This document must be signed by all such shareholders and returned with the signed Letter of Intent.

2. Due Diligence and Discovery. The participants agree that the information provided in their application to TechStars and any further information they have provided in relation to their application or will provide during due diligence, whether given verbally or in writing, is true and correct to the best of their knowledge.

Upon execution of this LOI, we will provide you with a due diligence checklist that outlines that company documents requested (usual and customary documents relating to the formation of the company, shareholder agreements, and other similar agreements), consent to perform a background check and a summary of representations. We will ask that you complete and submit all diligence materials and consents to us by **Tuesday, April 16, 2013.**

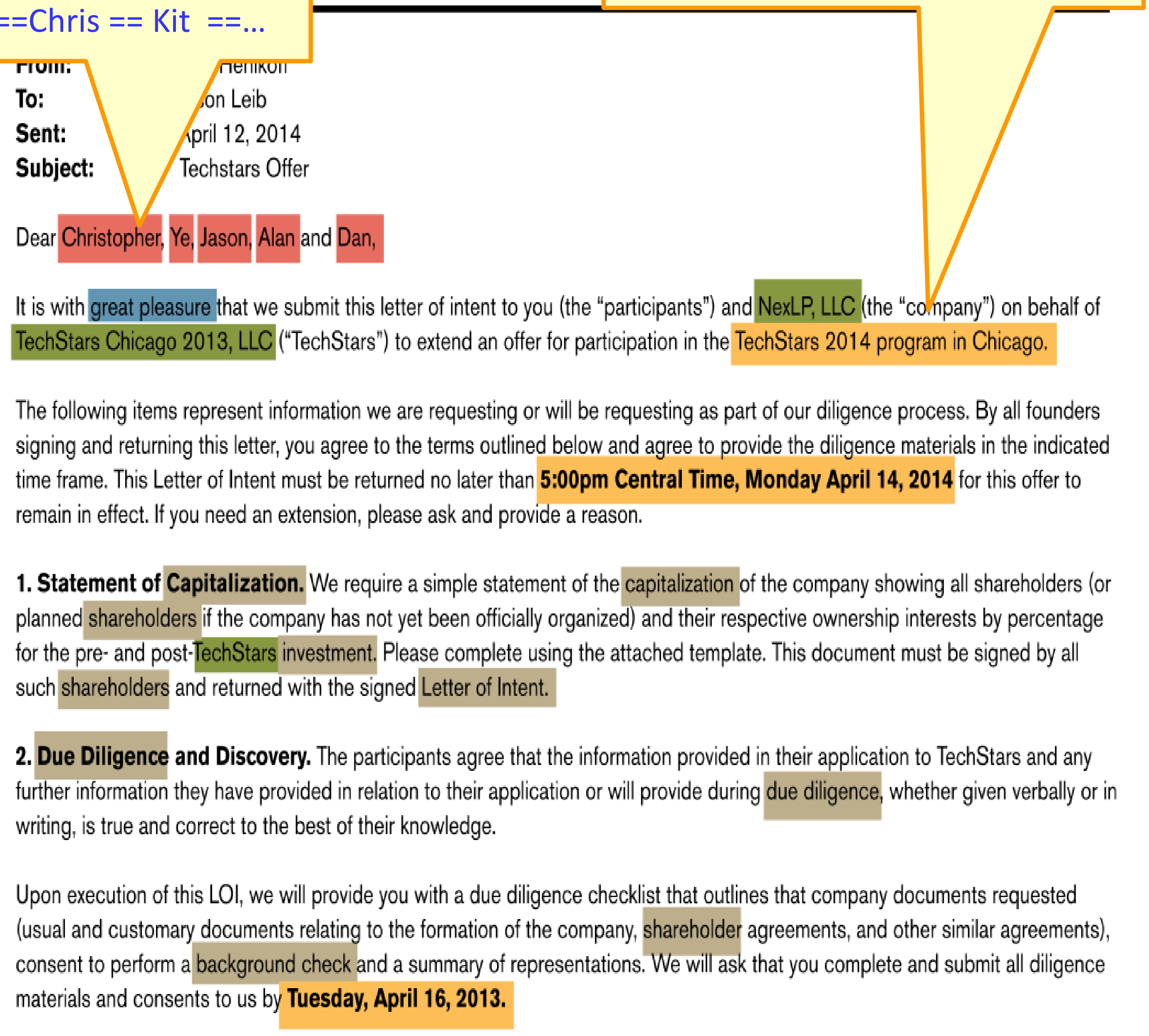
Identify Entities

Realize that

Christopher ==Chris == Kit ==...

Events, Locations, Dates

Normalization



From: Henikoff
To: on Leib
Sent: pril 12, 2014
Subject: Techstars Offer

Dear Christopher, Ye, Jason, Alan and Dan,

It is with great pleasure that we submit this letter of intent to you (the "participants") and NexLP, LLC (the "company") on behalf of TechStars Chicago 2013, LLC ("TechStars") to extend an offer for participation in the TechStars 2014 program in Chicago.

The following items represent information we are requesting or will be requesting as part of our diligence process. By all founders signing and returning this letter, you agree to the terms outlined below and agree to provide the diligence materials in the indicated time frame. This Letter of Intent must be returned no later than **5:00pm Central Time, Monday April 14, 2014** for this offer to remain in effect. If you need an extension, please ask and provide a reason.

1. Statement of Capitalization. We require a simple statement of the capitalization of the company showing all shareholders (or planned shareholders if the company has not yet been officially organized) and their respective ownership interests by percentage for the pre- and post-TechStars investment. Please complete using the attached template. This document must be signed by all such shareholders and returned with the signed Letter of Intent.

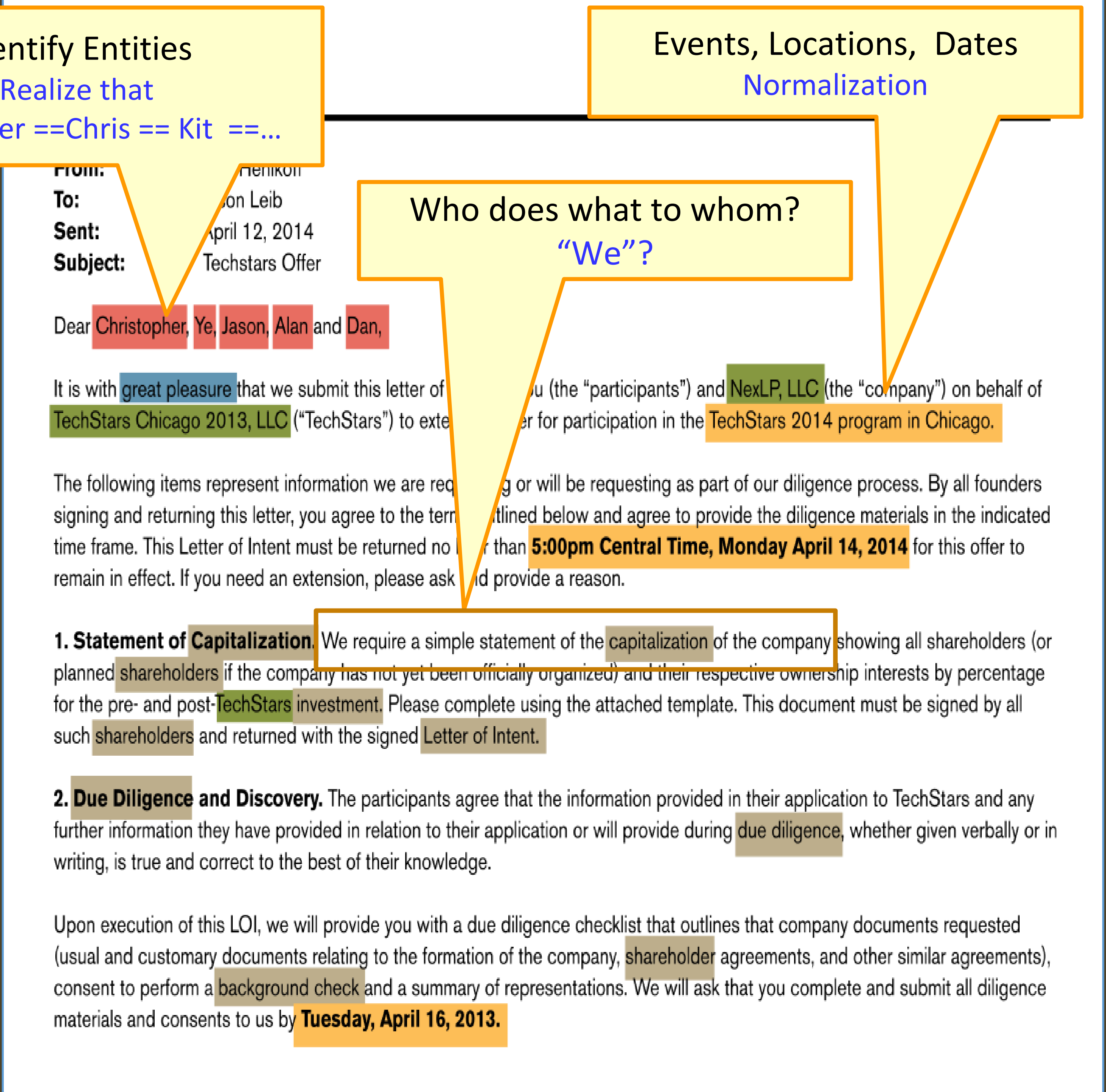
2. Due Diligence and Discovery. The participants agree that the information provided in their application to TechStars and any further information they have provided in relation to their application or will provide during due diligence, whether given verbally or in writing, is true and correct to the best of their knowledge.

Upon execution of this LOI, we will provide you with a due diligence checklist that outlines that company documents requested (usual and customary documents relating to the formation of the company, shareholder agreements, and other similar agreements), consent to perform a background check and a summary of representations. We will ask that you complete and submit all diligence materials and consents to us by **Tuesday, April 16, 2013.**

Identify Entities
Realize that
Christopher ==Chris == Kit ==…

Events, Locations, Dates
Normalization

Who does what to whom?
"We"?



From: Henikoff
To: on Leib
Sent: pril 12, 2014
Subject: Techstars Offer

Dear Christopher, Ye, Jason, Alan and Dan,

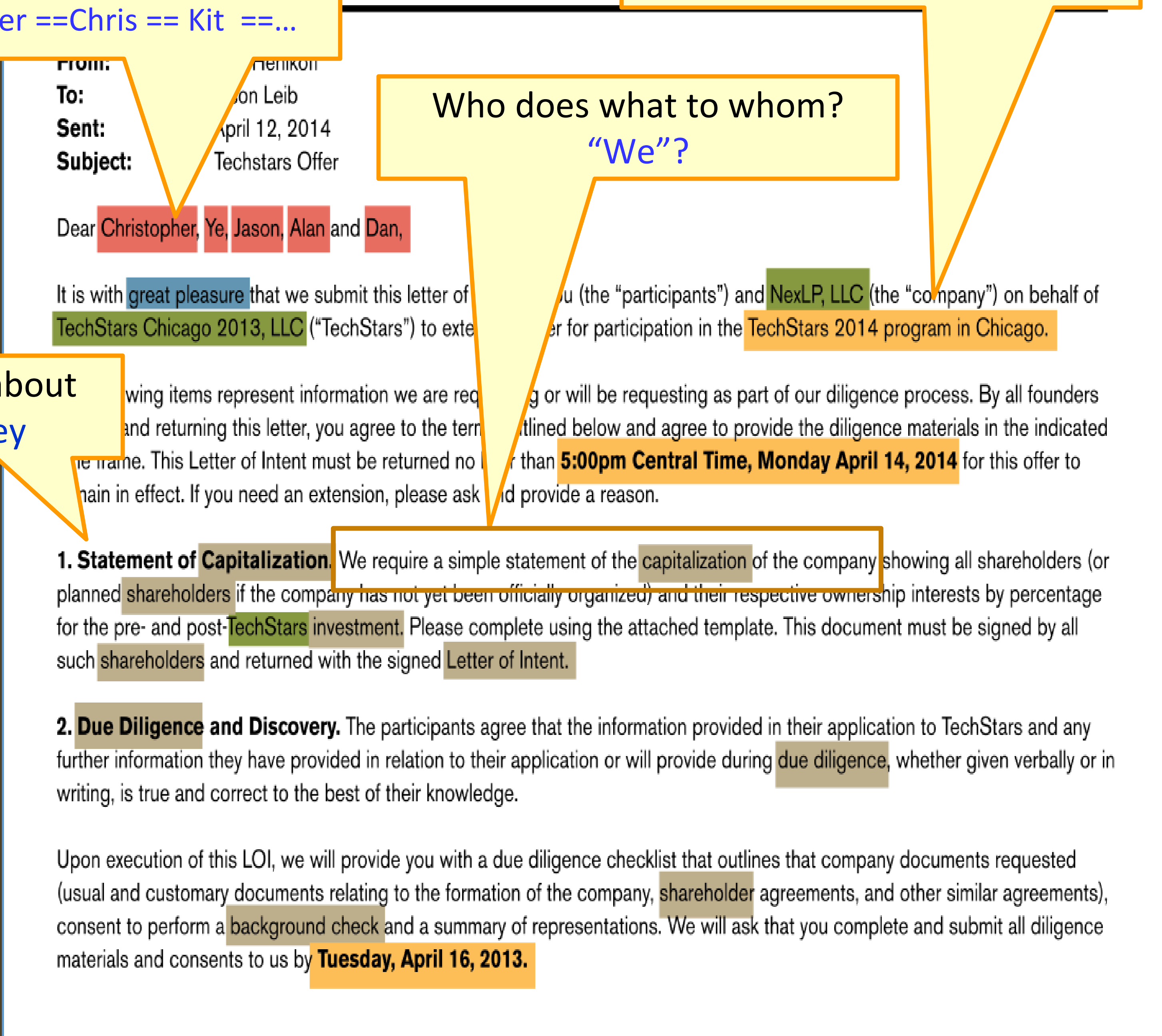
It is with great pleasure that we submit this letter of u (the "participants") and NexLP, LLC (the "company") on behalf of TechStars Chicago 2013, LLC ("TechStars") to exte r for participation in the TechStars 2014 program in Chicago.

The following items represent information we are req g or will be requesting as part of our diligence process. By all founders signing and returning this letter, you agree to the term tlined below and agree to provide the diligence materials in the indicated time frame. This Letter of Intent must be returned no l r than **5:00pm Central Time, Monday April 14, 2014** for this offer to remain in effect. If you need an extension, please ask d provide a reason.

1. Statement of Capitalization. We require a simple statement of the capitalization of the company showing all shareholders (or planned shareholders if the company has not yet been officially organized) and their respective ownership interests by percentage for the pre- and post-TechStars investment. Please complete using the attached template. This document must be signed by all such shareholders and returned with the signed Letter of Intent.

2. Due Diligence and Discovery. The participants agree that the information provided in their application to TechStars and any further information they have provided in relation to their application or will provide during due diligence, whether given verbally or in writing, is true and correct to the best of their knowledge.

Upon execution of this LOI, we will provide you with a due diligence checklist that outlines that company documents requested (usual and customary documents relating to the formation of the company, shareholder agreements, and other similar agreements), consent to perform a background check and a summary of representations. We will ask that you complete and submit all diligence materials and consents to us by **Tuesday, April 16, 2013.**



Identify Entities
Realize that
Christopher ==Chris == Kit ==…

Events, Locations, Dates
Normalization

From: Henikoff
To: on Leib
Sent: pril 12, 2014
Subject: Techstars Offer

Who does what to whom?
"We"?

Dear Christopher, Ye, Jason, Alan and Dan,

It is with great pleasure that we submit this letter of u (the "participants") and NexLP, LLC (the "company") on behalf of TechStars Chicago 2013, LLC ("TechStars") to exte er for participation in the TechStars 2014 program in Chicago.

Talking about
Money

wing items represent information we are req g or will be requesting as part of our diligence process. By all founders and returning this letter, you agree to the ter tlined below and agree to provide the diligence materials in the indicated e frame. This Letter of Intent must be returned no l r than **5:00pm Central Time, Monday April 14, 2014** for this offer to ain in effect. If you need an extension, please ask d provide a reason.

1. Statement of Capitalization. We require a simple statement of the capitalization of the company showing all shareholders (or planned shareholders if the company has not yet been officially organized) and their respective ownership interests by percentage for the pre- and post-TechStars investment. Please complete using the attached template. This document must be signed by all such shareholders and returned with the signed Letter of Intent.

2. Due Diligence and Discovery. The participants agree that the information provided in their application to TechStars and any further information they have provided in relation to their application or will provide during due diligence, whether given verbally or in writing, is true and correct to the best of their knowledge.

Upon execution of this LOI, we will provide you with a due diligence checklist that outlines that company documents requested (usual and customary documents relating to the formation of the company, shareholder agreements, and other similar agreements), consent to perform a background check and a summary of representations. We will ask that you complete and submit all diligence materials and consents to us by **Tuesday, April 16, 2013.**

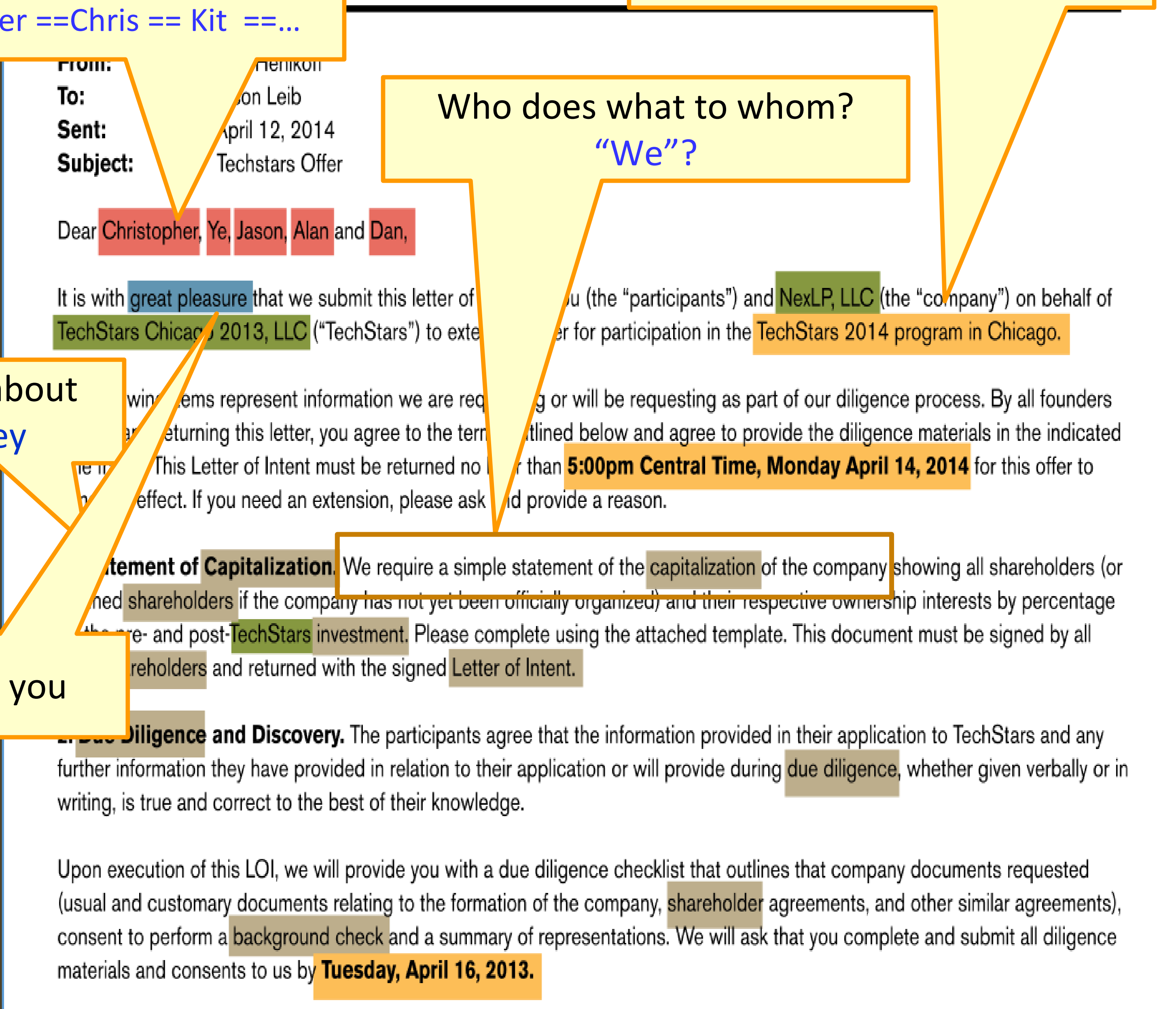
Identify Entities
Realize that
Christopher ==Chris == Kit ==…

Events, Locations, Dates
Normalization

Who does what to whom?
“We”?

Talking about
Money

We like you



From: Henikoff
To: on Leib
Sent: pril 12, 2014
Subject: Techstars Offer

Dear Christopher, Ye, Jason, Alan and Dan,

It is with great pleasure that we submit this letter of u (the “participants”) and NexLP, LLC (the “company”) on behalf of TechStars Chicag 2013, LLC (“TechStars”) to exte r for participation in the TechStars 2014 program in Chicago.

win ems represent information we are req g or will be requesting as part of our diligence process. By all founders eturning this letter, you agree to the ter tlined below and agree to provide the diligence materials in the indicated e i This Letter of Intent must be returned no l r than 5:00pm Central Time, Monday April 14, 2014 for this offer to effect. If you need an extension, please ask d provide a reason.

tement of Capitalization. We require a simple statement of the capitalization of the company showing all shareholders (or ned shareholders if the company has not yet been officially organized) and their respective ownership interests by percentage e- and post-TechStars investment. Please complete using the attached template. This document must be signed by all reholders and returned with the signed Letter of Intent.

Diligence and Discovery. The participants agree that the information provided in their application to TechStars and any further information they have provided in relation to their application or will provide during due diligence, whether given verbally or in writing, is true and correct to the best of their knowledge.

Upon execution of this LOI, we will provide you with a due diligence checklist that outlines that company documents requested (usual and customary documents relating to the formation of the company, shareholder agreements, and other similar agreements), consent to perform a background check and a summary of representations. We will ask that you complete and submit all diligence materials and consents to us by Tuesday, April 16, 2013.

Identify Entities
Realize that
Christopher ==Chris == Kit ==…

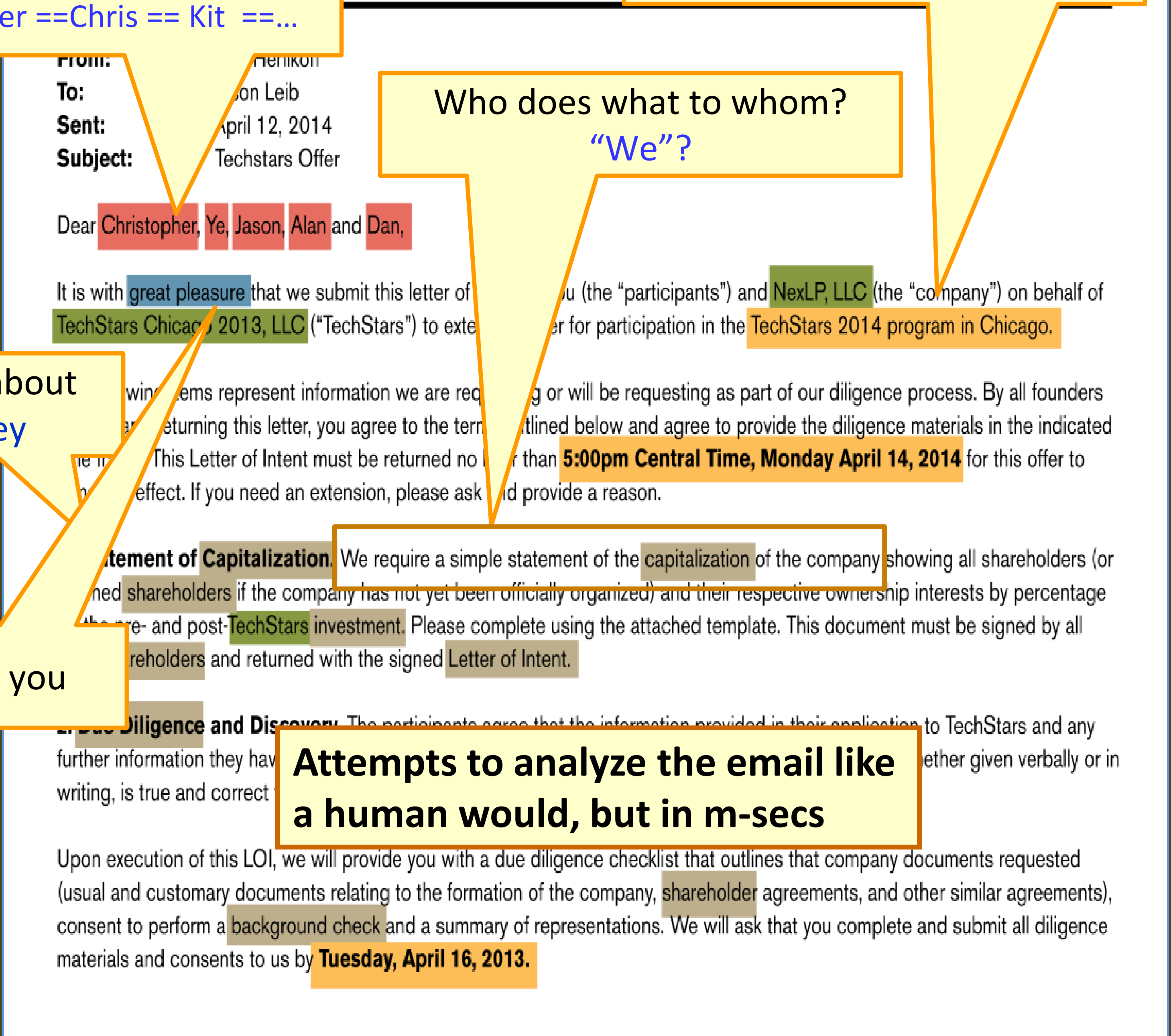
Events, Locations, Dates
Normalization

Who does what to whom?
"We"?

Talking about
Money

We like you

Attempts to analyze the email like a human would, but in m-secs



From: Henikoff
To: on Leib
Sent: pril 12, 2014
Subject: Techstars Offer

Dear Christopher, Ye, Jason, Alan and Dan,

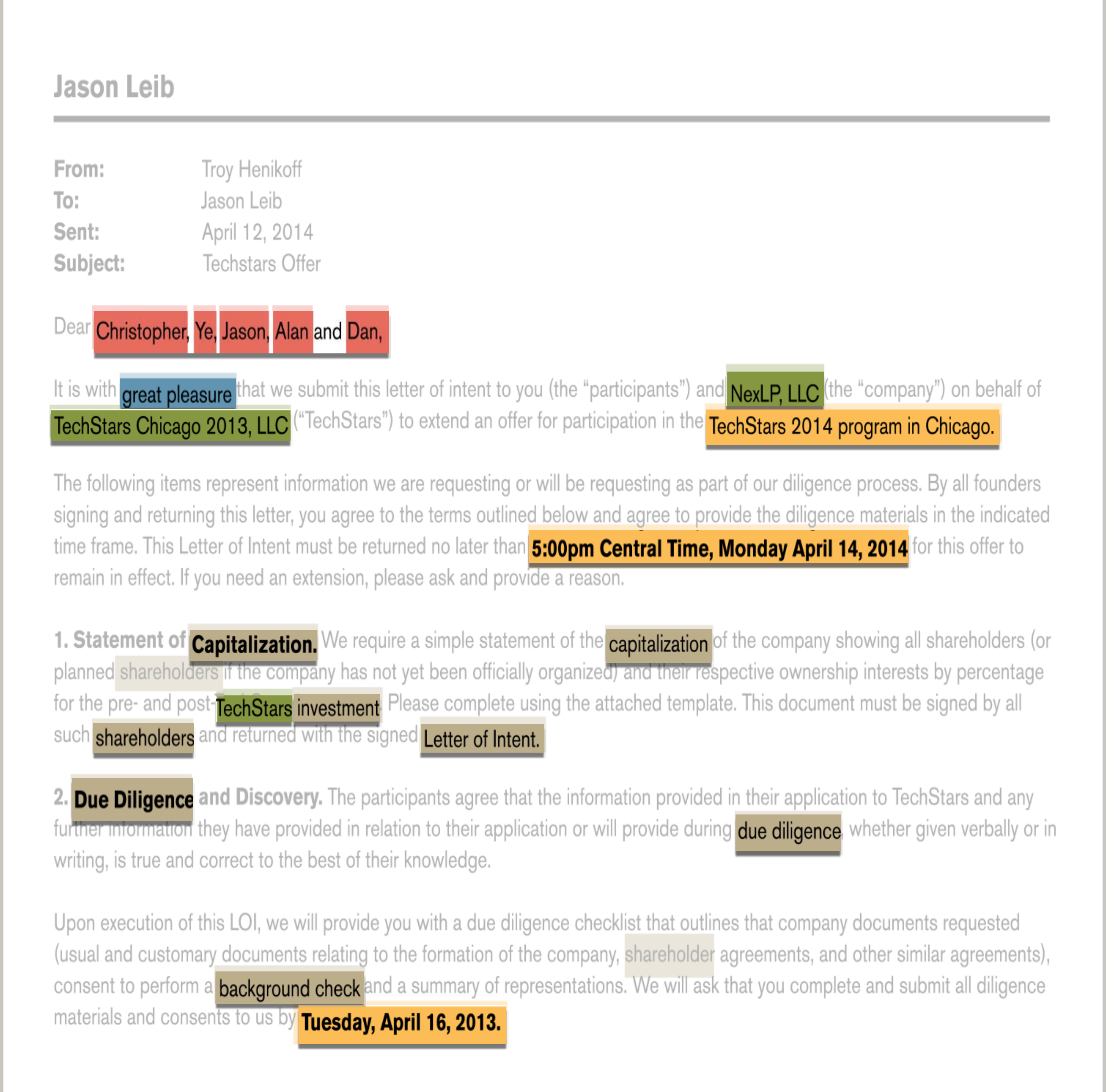
It is with great pleasure that we submit this letter of u (the "participants") and NexLP, LLC (the "company") on behalf of TechStars Chicag 2013, LLC ("TechStars") to exte r for participation in the TechStars 2014 program in Chicago.

win ems represent information we are req g or will be requesting as part of our diligence process. By all founders ar eturning this letter, you agree to the tern tlined below and agree to provide the diligence materials in the indicated e i This Letter of Intent must be returned no l r than **5:00pm Central Time, Monday April 14, 2014** for this offer to effect. If you need an extension, please ask d provide a reason.

tement of Capitalization. We require a simple statement of the capitalization of the company showing all shareholders (or ned shareholders if the company has not yet been officially organized) and their respective ownership interests by percentage the pre- and post-TechStars investment. Please complete using the attached template. This document must be signed by all reholders and returned with the signed Letter of Intent.

2. Due Diligence and Discovery. The participants agree that the information provided in their application to TechStars and any further information they hav ether given verbally or in writing, is true and correct

Upon execution of this LOI, we will provide you with a due diligence checklist that outlines that company documents requested (usual and customary documents relating to the formation of the company, shareholder agreements, and other similar agreements), consent to perform a background check and a summary of representations. We will ask that you complete and submit all diligence materials and consents to us by **Tuesday, April 16, 2013.**



Jason Leib

From: Troy Henikoff
To: Jason Leib
Sent: April 12, 2014
Subject: Techstars Offer

Dear Christopher, Ye, Jason, Alan and Dan,

It is with great pleasure that we submit this letter of intent to you (the "participants") and NexLP, LLC (the "company") on behalf of TechStars Chicago 2013, LLC ("TechStars") to extend an offer for participation in the TechStars 2014 program in Chicago.

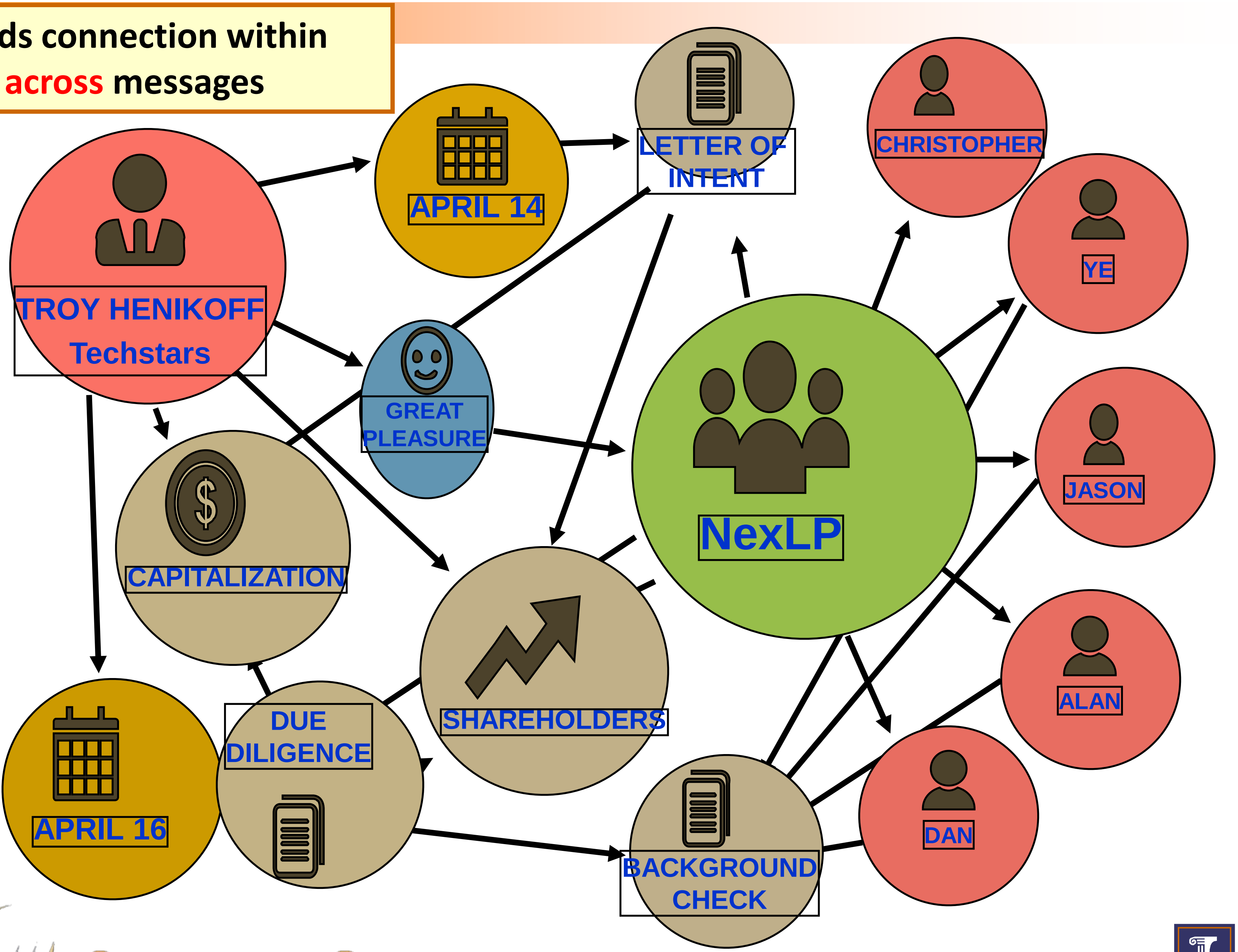
The following items represent information we are requesting or will be requesting as part of our diligence process. By all founders signing and returning this letter, you agree to the terms outlined below and agree to provide the diligence materials in the indicated time frame. This Letter of Intent must be returned no later than **5:00pm Central Time, Monday April 14, 2014** for this offer to remain in effect. If you need an extension, please ask and provide a reason.

1. Statement of Capitalization. We require a simple statement of the capitalization of the company showing all shareholders (or planned shareholders if the company has not yet been officially organized) and their respective ownership interests by percentage for the pre- and post-TechStars investment. Please complete using the attached template. This document must be signed by all such shareholders and returned with the signed Letter of Intent.

2. Due Diligence and Discovery. The participants agree that the information provided in their application to TechStars and any further information they have provided in relation to their application or will provide during due diligence whether given verbally or in writing, is true and correct to the best of their knowledge.

Upon execution of this LOI, we will provide you with a due diligence checklist that outlines that company documents requested (usual and customary documents relating to the formation of the company, shareholder agreements, and other similar agreements), consent to perform a background check and a summary of representations. We will ask that you complete and submit all diligence materials and consents to us by **Tuesday, April 16, 2013.**

Builds connection within and across messages



NexLP CASE STUDY

HENCEWORTH TRADINGS

TEAM OF LAWYERS X 3 WEEKS = $250,000

NexLP CASE STUDY

True Story (de-identified) 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.

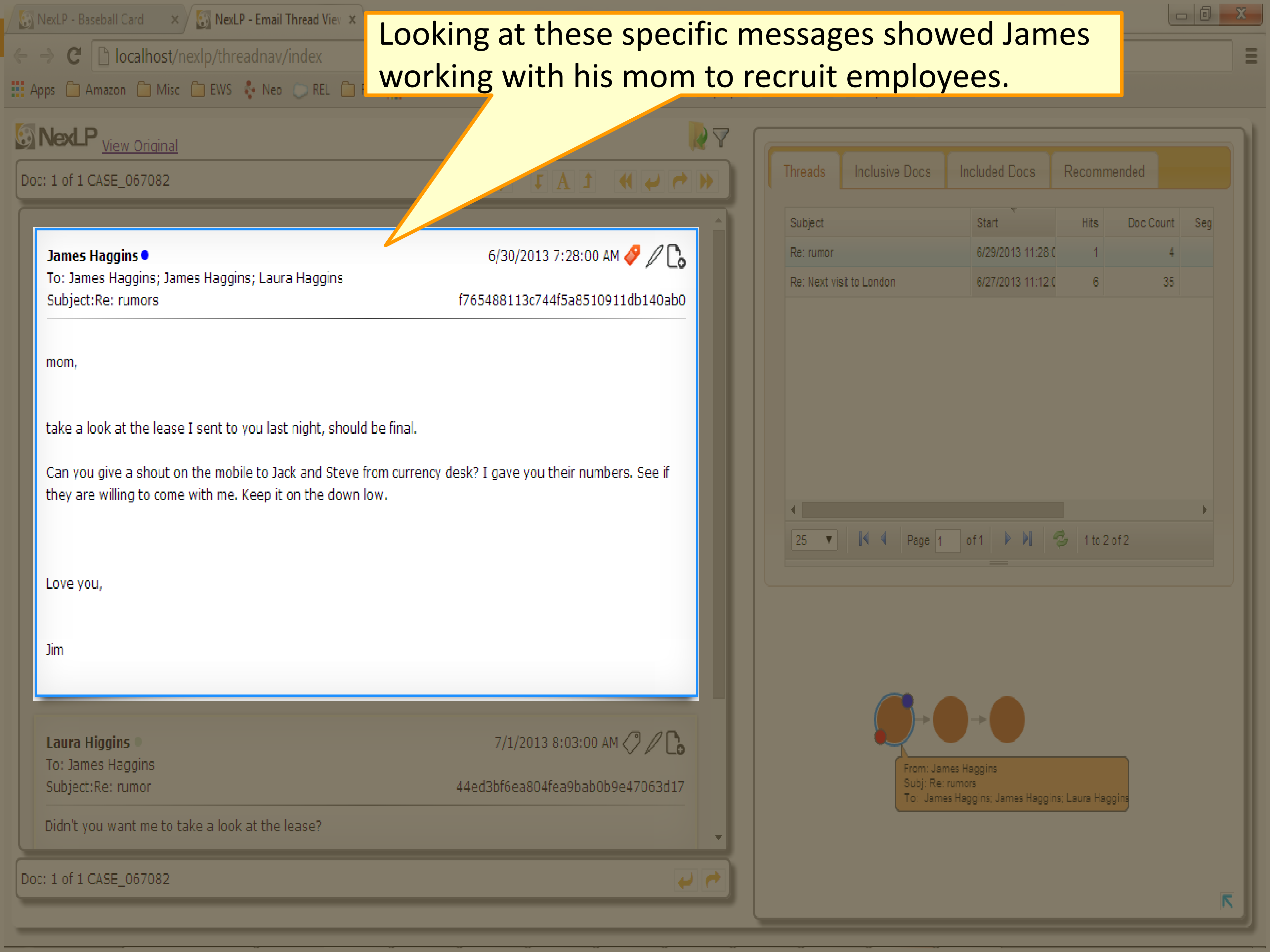
HENCEWORTH TRADINGS

TEAM OF LAWYERS

X 3 WEEKS = **$250,000**

▶

Looking at these specific messages showed James working with his mom to recruit employees.



Doc: 1 of 1 CASE_067082

James Haggins ● 6/30/2013 7:28:00 AM
To: James Haggins; James Haggins; Laura Haggins
Subject:Re: rumors f765488113c744f5a8510911db140ab0

mom,

take a look at the lease I sent to you last night, should be final.

Can you give a shout on the mobile to Jack and Steve from currency desk? I gave you their numbers. See if they are willing to come with me. Keep it on the down low.

Love you,

Jim

Laura Higgins 7/1/2013 8:03:00 AM
To: James Haggins
Subject:Re: rumor 44ed3bf6ea804fea9bab0b9e47063d17

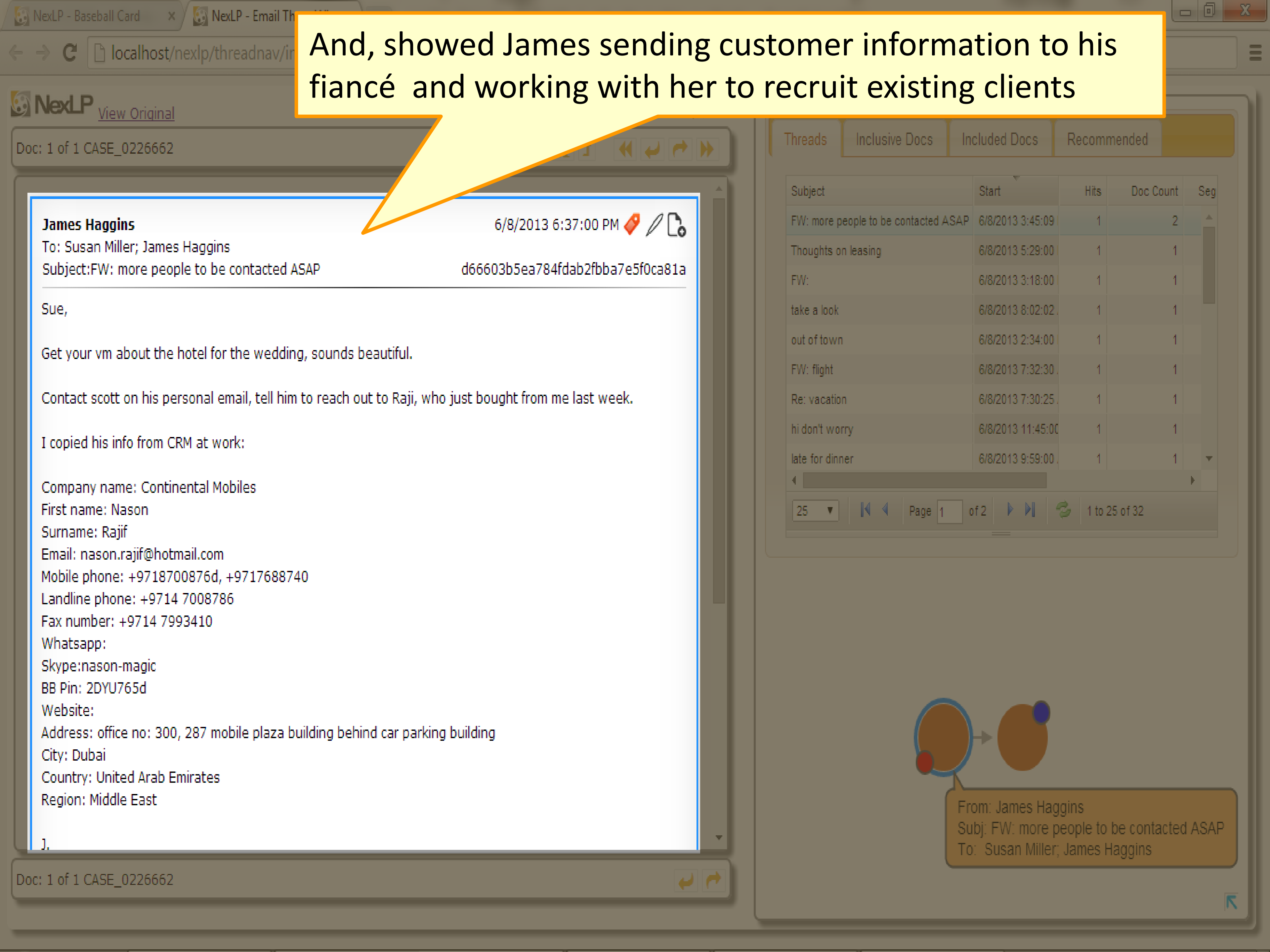
Didn't you want me to take a look at the lease?

Doc: 1 of 1 CASE_067082

Subject	Start	Hits	Doc Count	Seg
Re: rumor	6/29/2013 11:28:(	1	4	
Re: Next visit to London	6/27/2013 11:12:(	6	35	

From: James Haggins
Subj: Re: rumors
To: James Haggins; James Haggins; Laura Haggins

And, showed James sending customer information to his fiancé and working with her to recruit existing clients



Doc: 1 of 1 CASE_0226662

James Haggins 6/8/2013 6:37:00 PM
To: Susan Miller; James Haggins
Subject:FW: more people to be contacted ASAP d66603b5ea784fdab2fbba7e5f0ca81a

Sue,

Get your vm about the hotel for the wedding, sounds beautiful.

Contact scott on his personal email, tell him to reach out to Raji, who just bought from me last week.

I copied his info from CRM at work:

Company name: Continental Mobiles
First name: Nason
Surname: Rajif
Email: nason.rajif@hotmail.com
Mobile phone: +9718700876d, +9717688740
Landline phone: +9714 7008786
Fax number: +9714 7993410
Whatsapp:
Skype:nason-magic
BB Pin: 2DYU765d
Website:
Address: office no: 300, 287 mobile plaza building behind car parking building
City: Dubai
Country: United Arab Emirates
Region: Middle East

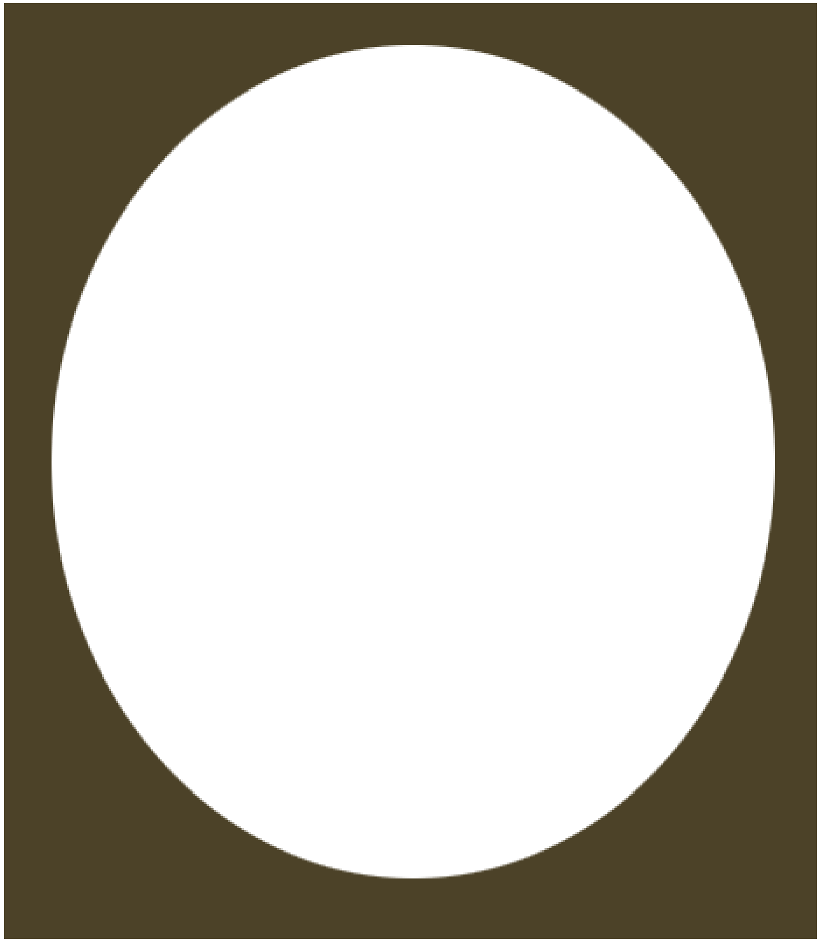
J.

Doc: 1 of 1 CASE_0226662

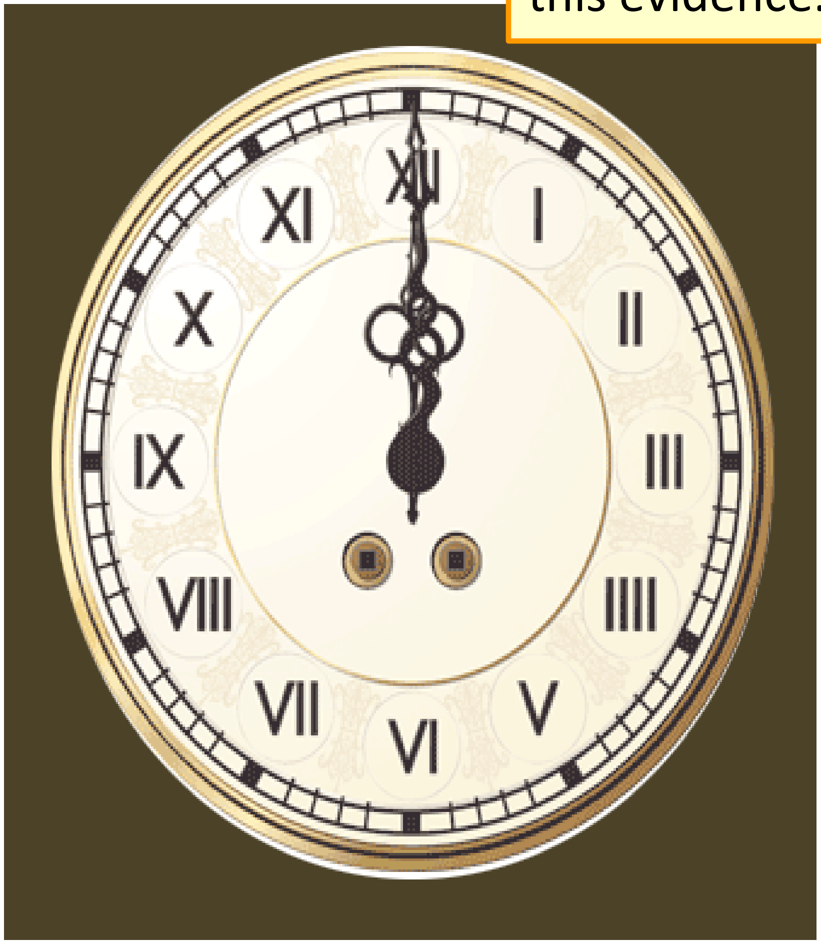
Subject	Start	Hits	Doc Count	Seg
FW: more people to be contacted ASAP	6/8/2013 3:45:09	1	2	
Thoughts on leasing	6/8/2013 5:29:00	1	1	
FW:	6/8/2013 3:18:00	1	1	
take a look	6/8/2013 8:02:02 .	1	1	
out of town	6/8/2013 2:34:00	1	1	
FW: flight	6/8/2013 7:32:30 .	1	1	
Re: vacation	6/8/2013 7:30:25 .	1	1	
hi don't worry	6/8/2013 11:45:00	1	1	
late for dinner	6/8/2013 9:59:00 .	1	1	

25 ▼ Page 1 of 2 1 to 25 of 32

From: James Haggins
Subj: FW: more people to be contacted ASAP
To: Susan Miller; James Haggins



It took 90 minutes to find two key emails in a large collection; a team of lawyers spent 3 weeks on the same collection and could not find this evidence….



8:23 AM
100%
Tuesday
10
Messages
Calendar
Photos
Camera
280
Weather
Clock
Maps
Videos
Notes
Reminders
Stocks
Game Center
Newsstand
iTunes Store
App Store
Passbook
Compass
Settings
NexLP

The goal is to provide realtime notifications – reduce the impact of compliance infractions, potential fraud and even customer issues

3 Compliance Alerts!

Summary: Making Sense of Unstructured Data

- A lot of today’s information is in text
- 80% of data corporations deal with is TEXT
- We are trying to push the level of automatic text understanding

- Very significant progress over the last 10 years or so
  - Mostly using statistical machine learning methods
- The problem isn’t solved – a very active research area
  - We mostly work at a sentence level
  - We make a lot of mistakes
  - We don’t understand events, intention,…
  - We don’t know how to use background knowledge and common sense

Summary: Making Sense of Unstructured Data

- A lot of today’s information is in text
- 80% of data corporations deal with is TEXT
- We are trying to push the level of automatic text understanding

- Very significant progress over the last 10 years or so
  - □ Mostly using statistical machine learning methods
- The problem isn’t solved – a very active research area
  - □ We mostly work at a sentence level
  - □ We make a lot of mistakes
  - □ We don’t understand events, intention,…
  - □ We don’t know how to use background knowledge and common sense
- We still don’t **understand** text. But, we can offer practical solutions and working tools that reliably address a range a problems researchers and corporations are interested in.

Summary: Making Sense of Unstructured Data

Thank you!

- A lot of today’s information is in text
- 80% of data corporations deal with is TEXT
- We are trying to push the level of automatic text understanding

- Very significant progress over the last 10 years or so
  - □ Mostly using statistical machine learning methods
- The problem isn’t solved – a very active research area
  - □ We mostly work at a sentence level
  - □ We make a lot of mistakes
  - □ We don’t understand events, intention,…
  - □ We don’t know how to use background knowledge and common sense
- We still don’t **understand** text. But, we can offer practical solutions and working tools that reliably address a range a problems researchers and corporations are interested in.