A SURVEY ON RELATION EXTRACTION

Nguyen Bach & Sameer Badaskar

Language Technologies Institute

Carnegie Mellon University

Introduction

- Structuring the information on the web
- Involves annotating the unstructured text with
  - Entities
  - Relations between entities
- Extracting semantic relations between entities in text

Entity-Relations

- Example 1: "<u>Bill Gates</u> works at <u>Microsoft Inc</u>."
  - *Person-Affiliation*(Bill Gates, Microsoft Inc)
- Example 2: *Located-In*(CMU, Pittsburgh)
- Higher order relations
  - *Protein-Organism-Location*
- Entity tuple: entities are bound in a relation
  - $r(e_1, e_2, \ldots, e_n)$

Applications

- Question Answering: Ravichandran & Hovy (2002)
  - Extracting entities and relational patterns for answering factoid questions (Example: "When was Gandhi born ?" amounts to looking for *Born-In*(Gandhi, ??) in the relational database)
- Mining bio-medical texts
  - Protein binding relations useful for drug discovery
  - Detection of cancerous genes ("Gene X with mutation Y leads to malignancy Z")

Evaluation

- Datasets
  - Automatic Content Extraction (ACE) http://www.nist.gov/speech/tests/ace/index.htm
  - Message Understanding Conference (MUC-7) http://www.ldc.upenn.edu
- Supervised Approaches
  - Relation extraction as a classification task.
  - Precision, Recall and F1
- Semi-supervised Approaches
  - Bootstrapping based approaches result in the discovery of large number of patterns and relations.
  - Approximate value of precision computed by drawing a random sample and manually checking for actual relations

Outline

- Supervised approaches
  - Feature based
  - Kernel based
  - Concerns
- Semi-supervised approaches
  - Bootstrapping
  - DIPRE, Snowball, KnowItAll, TextRunner
- Higher-order relation extraction

Supervised Approaches (1)

- Formulate the problem as a classification problem (in a discriminative framework)
- Given a set of +ve and –ve training examples
- Sentence : $S = w_1 w_2 \ldots e_1 \ldots w_i \ldots e_2 \ldots w_{n-1} w_n$

$$f_R(T(S)) = \begin{cases} +1 & \text{If } e_1 \text{ and } e_2 \text{ are related by } R \\ -1 & \text{Otherwise} \end{cases}$$

Supervised Approaches (2)

- $f_R(.)$ can be a discriminative classifier
  - SVM, Voted Perceptron, Log-linear model …
  - Can also be a multiclass classifier!

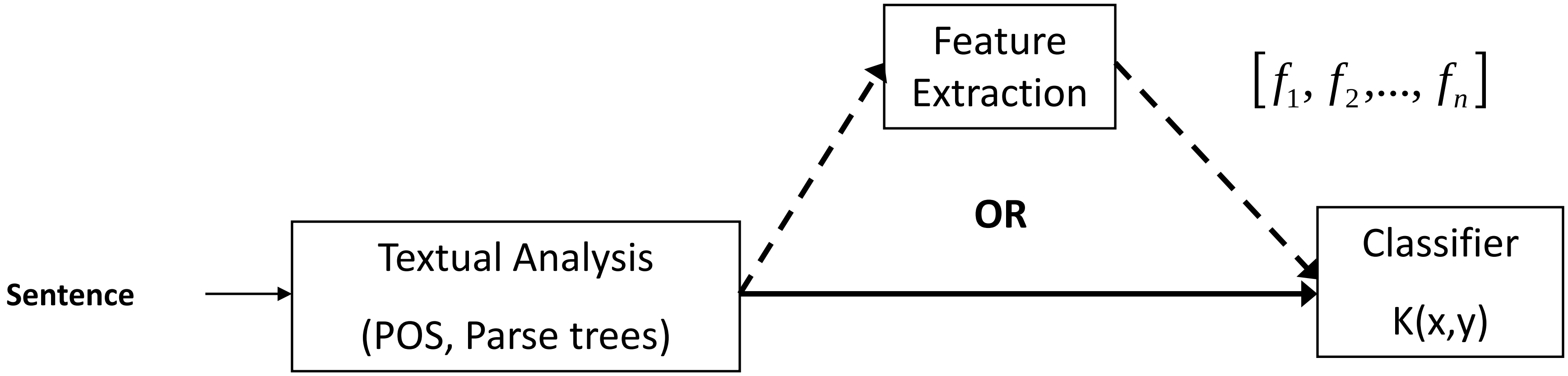
- $T(S)$ can be
  - A set of features extracted from the sentence
  - A structured representation of the sentence (labeled sequence, parse trees)

Supervised Approaches (3)

- Features
  - Define the feature set
  - Similarity metrics like cosine distance can be used

- Structured Representations
  - Need to define the similarity metric (Kernel)
  - Kernel similarity is integral to classifiers like SVMs.

Supervised Approaches (4)



- We'll come back to K(x,y) a bit later

Features

- Khambhatla (2004), Zhou et. al. (2005)
- Given a sentence
  1. Perform Textual Analysis (POS, Parsing, NER)
  2. Extract
     - Words between and including entities
     - Types of entities (person, location, etc)
     - Number of entities between the two entities, whether both entities belong to same chunk
     - # words separating the two entities
     - Path between the two entities in a parse tree

Features

- Textual Analysis involves POS tagging, dependency parsing etc.
- What forms a good set of features ?
- Choice of features guided by intuition and experiments.
- Alternative is to use the structural representations and **define an appropriate similarity metric** for the classifier!

Kernels

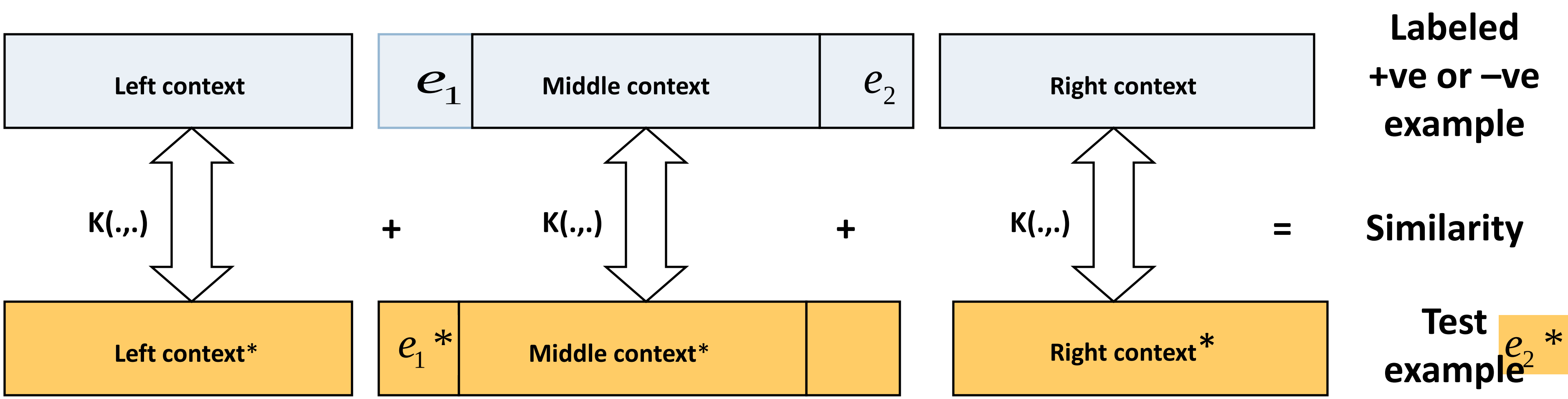
Homework #5

We were almost there!!!

- Kernel K(x,y) defines similarity between objects x and y implicitly in a higher dimensional space
- (x,y) can be
  - Strings: similarity $\propto$ number of common substrings (or subsequences) between x and y
  - Example: *sim*(**cat**, **ca**n**t**) > *sim*(**cat**, **c**ont**a**c**t**)
  - Excellent introduction to string kernels in Lodhi et. al. (2002)
- Extend string kernels to word sequences and parse trees for relation extraction

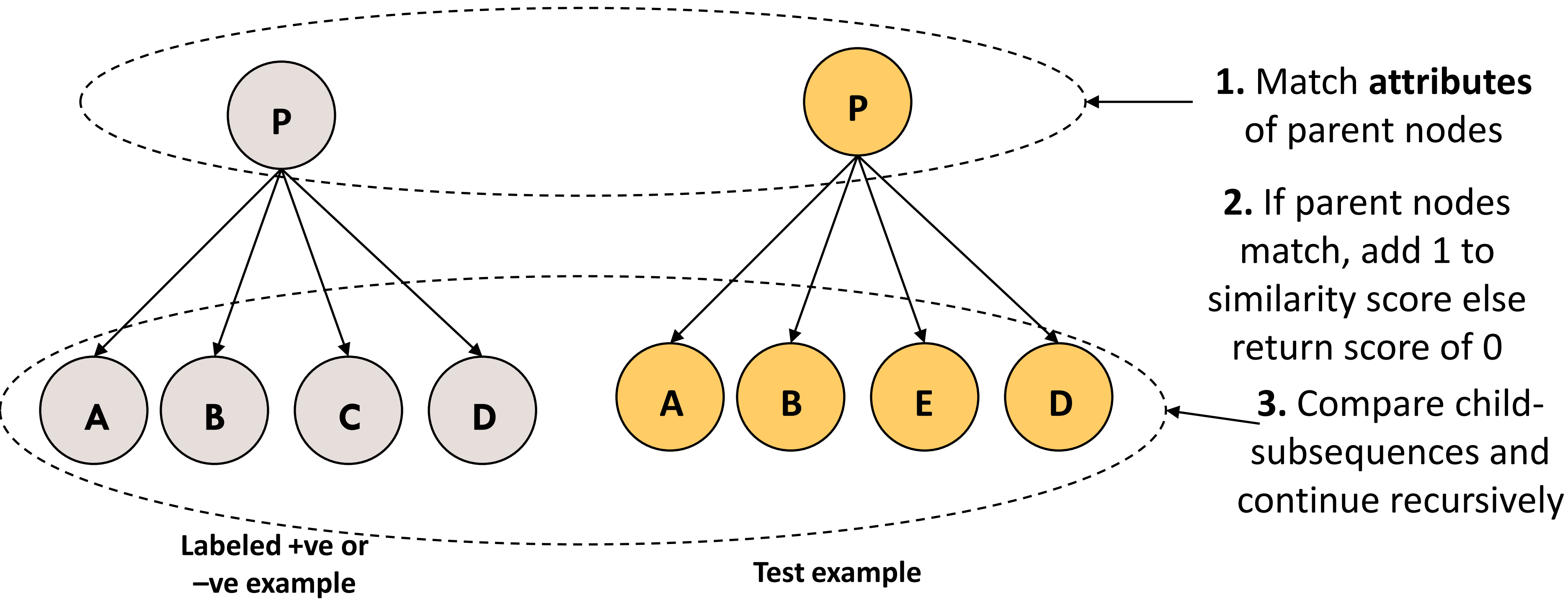
Kernels (Word Subsequences)

- **Word context** around entities can be indicator of a relation - Bunescu & Mooney (2005a)



- Each word is augmented with its POS, Generalized POS, chunk tag (NP, VP, etc), entity type (Person, Organization, none)

Kernels (Trees)



- **Similarity computed by counting the number of common subtrees**
- Attributes (??), Complexity (polynomial)

Kernels (Trees)

- **Tree kernels differ over types of trees used and attributes of nodes**
- Zelenko et. al. (2003)
  - Use shallow parse trees. Each node contains
    - Entity-Role (Person, Organization, Location, None)
    - Text it subsumes
    - Chunk tag (NP, VP etc)
  - Tasks: organization-location, person-affiliation detection
  - Tree kernel with SVM improves over feature based SVM for both tasks (**F1** 7% and 3% respectively)

- Culotta & Sorensen (2004)
  - Use dependency trees. Node attributes are
    - Word, POS, Generalized POS, Chunk tag, Entity type, Entity-level, Relation argument

- These tree kernels are *rigid* – attributes of nodes must match exactly!

Kernels

- Bunescu & Mooney (2005b)
  - Sufficient to use only the **shortest path between entities** in a dependency tree.
  - Each word in shortest path augmented with POS, Generalized POS, Entity type etc…
  - Structure of the dependency path is also encoded
  - Performs the best among all kernels

Kernels Vs Features

	Feature set Definition	Computational Complexity
Feature based Methods	Required to define a feature-set to be extracted after textual analysis. Good features arrived at by experimentation	Relatively lower
Kernel Methods	No need to define a feature-set. Similarity computed over a much larger feature space implicitly.	Relatively higher

Supervised Approaches (Concerns)

- Perform well but difficult to extend to new relation-types for want of labeled data

- Difficult to extend to higher order relations

- Textual analysis like POS tagging, shallow parsing, dependency parsing is a pre-requisite. This stage is prone to errors.

Semi-supervised Approaches

So far …

- Formulate relation extraction as a supervised classification task.
- Focused on feature-based and kernel methods
- We now focus on relation extraction with semi-supervised approaches
  - Rationale
  - DIPRE
  - Snowball
  - KnowItAll & TextRunner
  - Comparison

Rationales in Relation Extraction

- EBay was originally founded by Pierre Omidyar.
  - Founder (Pierre Omidyar, EBay)

- Ernest Hemingway was born in Oak Park-Illinois.
  - Born_in (Ernest Hemingway, Oak Park-Illinois)

- Read a short biography of Charles Dickens the great English literature novelist author of Oliver Twist, A Christmas carol.
  - Author_of (Charles Dickens, Oliver Twist)
  - Author_of (Charles Dickens, A Christmas carol)

- "Redundancy" : context of entities
- "Redundancy" is often sufficient to determine relations

DIPRE (Brin, 1998)

- Relation of interest : (author, book)
- DIPRE's algorithm:
  - Given a small seed set of (author, book) pairs

1. Use the seed examples to label some data.
2. Induces patterns from the labeled data.
3. Apply the patterns to unlabeled data to get new set of (author,book) pairs, and add to the seed set.
4. Return to step 1, and iterate until convergence criteria is reached

Seed: (Arthur Conan Doyle, The Adventures of Sherlock Holmes)

A Web crawler finds all documents contain the pair.

…

Read The Adventures of Sherlock Holmes by Arthur Conan Doyle online or in you email

…

Extract **tuple**:

[0, Arthur Conan Doyle, The Adventures of Sherlock Holmes, Read, online or, by]

A tuple of 6 elements: [order, author, book, prefix, suffix, middle]

order = 1 if the author string occurs before the book string, = 0 otherwise

prefix and *suffix* are strings contain the 10 characters occurring to the left/right of the match

middle is the string occurring between the author and book

…

know that Sir Arthur Conan Doyle wrote The Adventures of Sherlock Holmes, in 1892

…

Extract tuple:

[1, Arthur Conan Doyle, The Adventures of Sherlock Holmes, now that Sir, in 1892, wrote]

…

When Sir Arthur Conan Doyle wrote the adventures of Sherlock Holmes in 1892 he was high

…

Extract tuple:

[1, Arthur Conan Doyle, The Adventures of Sherlock Holmes, When Sir, in 1892 he, wrote]

Extracted list of tuples:

[0, Arthur Conan Doyle, The Adventures of Sherlock Holmes, Read, online or, by]

[1, Arthur Conan Doyle, The Adventures of Sherlock Holmes, now that Sir, in 1892, wrote]

[1, Arthur Conan Doyle, The Adventures of Sherlock Holmes, When Sir, in 1892 he, wrote]

…

Group tuples by matching *order* and *middle* and induce patterns

Induce patterns from group of tuples:

[longest-common-suffix of prefix strings, author, middle, book, longest-common-prefix of suffix strings]

Pattern:

[Sir, Arthur Conan Doyle, wrote, The Adventures of Sherlock Holmes, in 1892]

Pattern with wild card expression:

[Sir, .*?, wrote, .*?, in 1892]

Use the wild card patterns **[Sir, .*?, wrote, .*?, in 1892]**

search the Web to find more documents

…

Sir Arthur Conan Doyle **wrote** Speckled Band **in 1892**, that is around 62 years apart which would make the stories

…

Extract new relations:

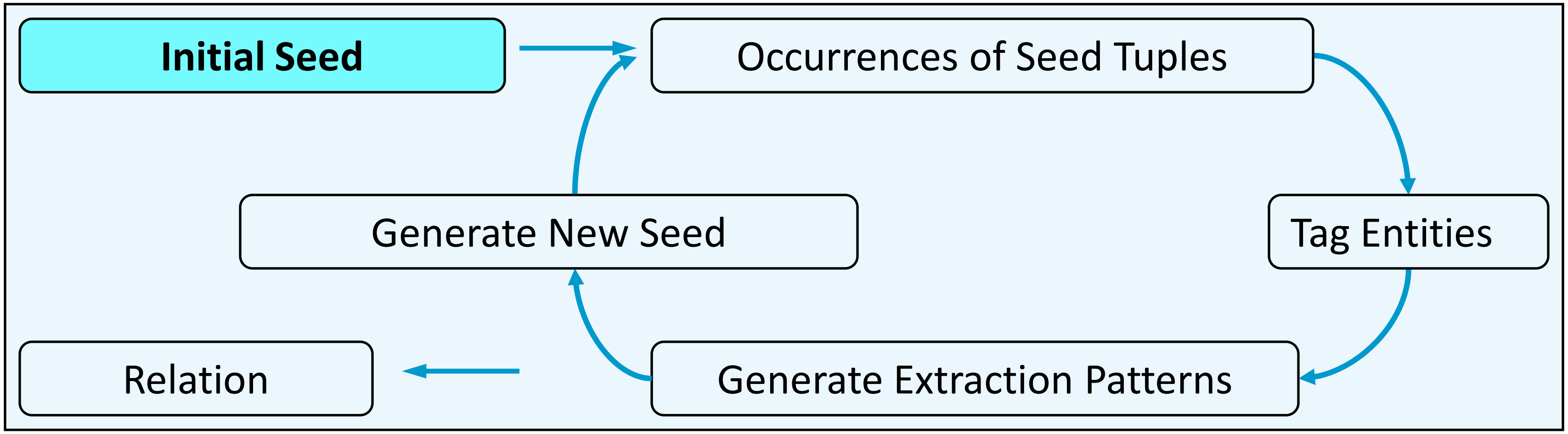
(Arthur Conan Doyle, Speckled Band)

Repeat the algorithm with the new relation.

Snowball (Agichtein & Gravano, 2000)

- Architecture: similar to DIPRE; relation (organization, location)

ORGANIZATION	***LOCATION***
MICROSOFT	REDMOND
IBM	ARMONK
BOEING	SEATTLE
INTEL	SANTA CLARA



Agichtein, 2000

Snowball (Agichtein & Gravano, 2000)

- Tuples: [author, book, prefix, suffix, middle]
  - represented in feature vectors, each token is associated with a term weight
- Group tuples by a similarity function

$$Match(tuple_i, tuple_j) = (prefix_i.prefix_j) + (suffix_i.suffix_j) + (middle_i.middle_j)$$

  - Higher similarity: tuples share common terms
- Induce patterns:
  - A pattern is a centroid vector tuple of a group
  - Assign pattern confidence score

KnowItAll (Etzioni et al. 2005)

- An autonomous, domain-independent system that extracts facts from the Web.
- The primary focus of the system is on extracting entities (unary predicates).
- The input to KnowItAll is a set of entity classes to be extracted, such as “city”, “scientist”, “movie”, etc., and the output is a list of entities extracted from the Web.

KnowItAll (Etzioni et al. 2005)

- Uses only the generic hand written patterns. The patterns are based on a general Noun Phrase (NP) chunker.

- Example patterns
  - NP1 "such as" NPList2
    - … including ***cities*** such as ***Birmingham, Montgomery, Mobile, Huntsville*** …
    - … publisher of ***books*** such as ***Gilgamesh, Big Tree, the Last Little Cat*** …
  - NP1 "and other" NP2
  - NP1 "including" NPList2
  - NP1 "is a" NP2
  - NP1 "is the" NP2 "of" NP3
  - "the" NP1 "of" NP2 "is" NP3

  …

TextRunner (Banko et al. 2007)

- DIPRE, Snowball, KnowItAll: relation types are predefined. TextRunner discovers relations automatically

- Extract Triple representing binary relation **(Arg1, Relation, Arg2)** from sentence.

EBay was originally founded by Pierre Omidyar.

EBay was originally founded by Pierre Omidyar.

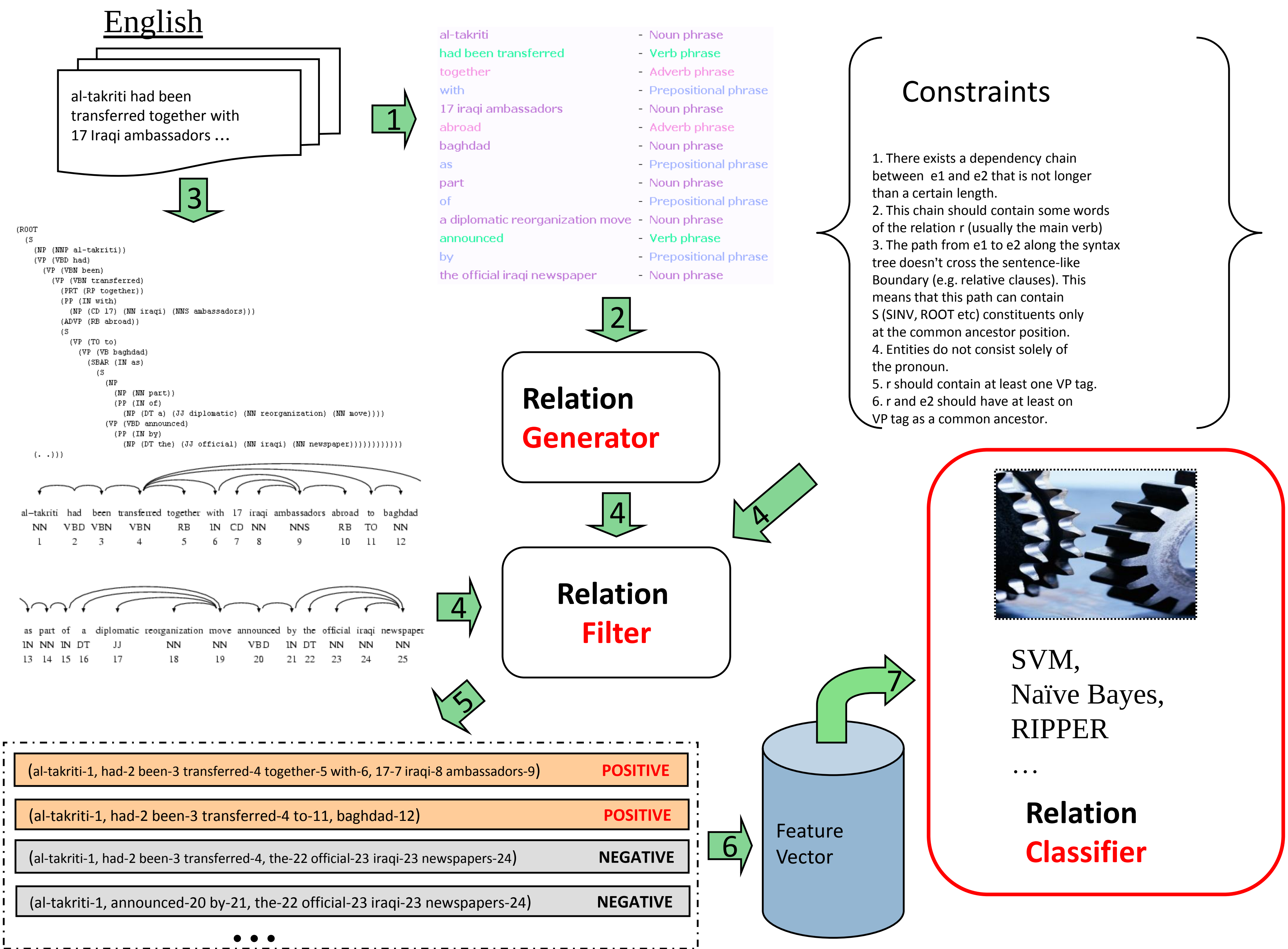
(Ebay, Founded by, Pierre Omidyar)

TextRunner (Banko et al. 2007)

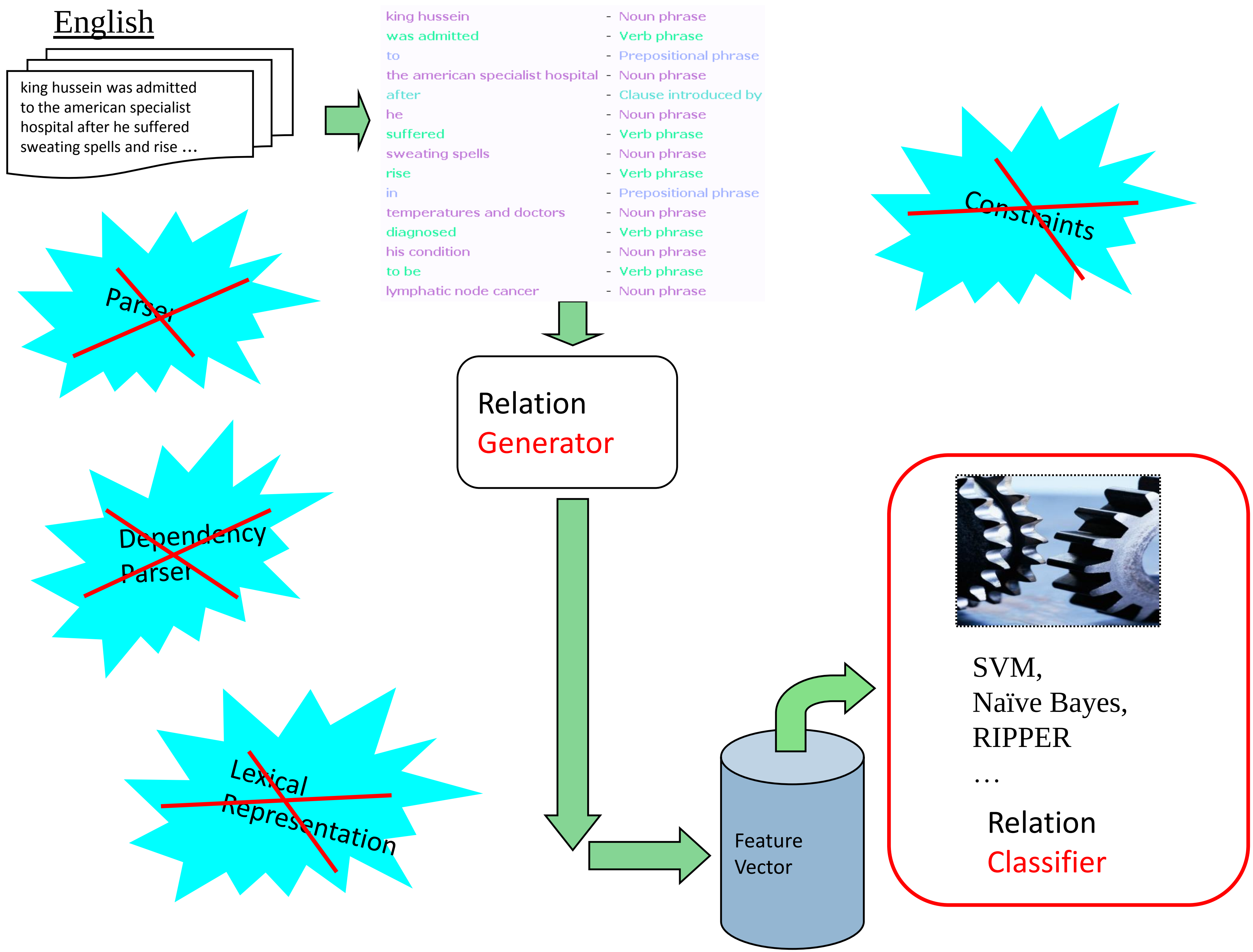
3 main components

1. **Self-Supervised Learner**: automatically labels +/- examples & learns an extractor
2. **Single-Pass Extractor:** single pass over corpus, identifying relations in each sentence
3. **Redundancy-based Assesor:** assign a probability to each retained relations based on a probabilistic model of redundancy in text introduced in based on (Downey et al. 2005)

Etzioni, 2007



English
al-takriti had been transferred together with 17 Iraqi ambassadors ...
1
al-takriti - Noun phrase
had been transferred - Verb phrase
together - Adverb phrase
with - Prepositional phrase
17 iraqi ambassadors - Noun phrase
abroad - Adverb phrase
baghdad - Noun phrase
as - Prepositional phrase
part - Noun phrase
of - Prepositional phrase
a diplomatic reorganization move - Noun phrase
announced - Verb phrase
by - Prepositional phrase
the official iraqi newspaper - Noun phrase
Constraints
1. There exists a dependency chain between e1 and e2 that is not longer than a certain length.
2. This chain should contain some words of the relation r (usually the main verb)
3. The path from e1 to e2 along the syntax tree doesn't cross the sentence-like Boundary (e.g. relative clauses). This means that this path can contain S (SINV, ROOT etc) constituents only at the common ancestor position.
4. Entities do not consist solely of the pronoun.
5. r should contain at least one VP tag.
6. r and e2 should have at least on VP tag as a common ancestor.
3
(ROOT
(S
(NP (NNP al-takriti))
(VP (VBD had)
(VP (VBN been)
(VP (VBN transferred)
(PRT (RP together))
(PP (IN with)
(NP (CD 17) (NN iraqi) (NNS ambassadors)))
(ADVP (RB abroad))
(S
(VP (TO to)
(VP (VB baghdad)
(SBAR (IN as)
(S
(NP
(NP (NN part))
(PP (IN of)
(NP (DT a) (JJ diplomatic) (NN reorganization) (NN move))))
(VP (VBD announced)
(PP (IN by)
(NP (DT the) (JJ official) (NN iraqi) (NN newspaper)))))))))))
(. .)))
2
Relation Generator
4
4
al-takriti had been transferred together with 17 iraqi ambassadors abroad to baghdad
NN VBD VBN VBN RB IN CD NN NNS RB TO NN
1 2 3 4 5 6 7 8 9 10 11 12
as part of a diplomatic reorganization move announced by the official iraqi newspaper
IN NN IN DT JJ NN NN VBD IN DT NN NN NN
13 14 15 16 17 18 19 20 21 22 23 24 25
4
Relation Filter
SVM,
Naïve Bayes,
RIPPER
...
Relation Classifier
5
7
(al-takriti-1, had-2 been-3 transferred-4 together-5 with-6, 17-7 iraqi-8 ambassadors-9) POSITIVE
(al-takriti-1, had-2 been-3 transferred-4 to-11, baghdad-12) POSITIVE
(al-takriti-1, had-2 been-3 transferred-4, the-22 official-23 iraqi-23 newspapers-24) NEGATIVE
(al-takriti-1, announced-20 by-21, the-22 official-23 iraqi-23 newspapers-24) NEGATIVE
• • •
6
Feature Vector


English
king hussein was admitted to the american specialist hospital after he suffered sweating spells and rise …
king hussein - Noun phrase
was admitted - Verb phrase
to - Prepositional phrase
the american specialist hospital - Noun phrase
after - Clause introduced by
he - Noun phrase
suffered - Verb phrase
sweating spells - Noun phrase
rise - Verb phrase
in - Prepositional phrase
temperatures and doctors - Noun phrase
diagnosed - Verb phrase
his condition - Noun phrase
to be - Verb phrase
lymphatic node cancer - Noun phrase
Constraints
Parser
Dependency Parser
Lexical Representation
Relation Generator
Feature Vector
SVM,
Naïve Bayes,
RIPPER
…
Relation Classifier

Comparison

	DIPRE	Snowball	KnowItAll	TextRunner
Initial seed	Yes	Yes	Yes	No
Predefined relation	Yes	Yes	Yes	No
External NLP tools	No	Yes: NER	Yes: NP chunker	Yes: dependency parser, NP chunker
Relation types	Binary	Binary	Unary/Binary	Binary
Language dependent	No	Yes	Yes	Yes
Classifier	Exact pattern matching	Matching with similarity function	Naïve Bayes classifier	Self-supervised binary classifier
Input parameters	2	9	>= 4	N/A

Higher-order Relation Extraction

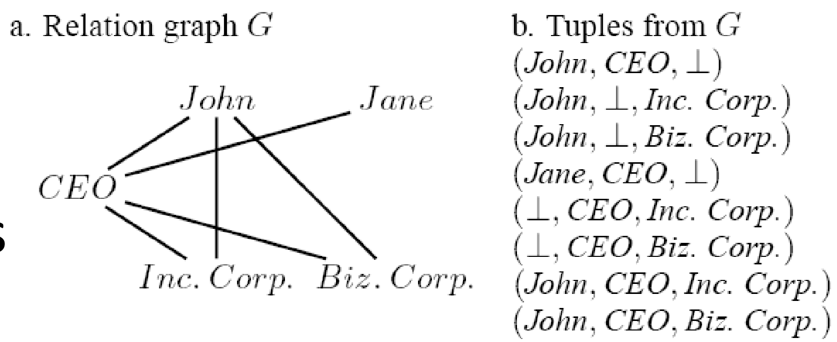
Higher-order Relations

- So far, reviewed methods focus on binary relations
- It is not straightforward to adapt to higher-order relation types.
- $(e_1, e_2, \ldots, e_n)$: each e_i has a type t_i
- Ternary relation: T= (people, job, company)
  - "John Smith is the CEO at Inc. Corp"
    - (John Smith, CEO, Inc. Corp)

McDonald et al. 2005

- Factoring higher-order relations into a set of binary relations

- Use a classifier to extract binary relations
- Create entities graph
- Reconstruct higher-order relations by finding maximal cliques



Conclusion

- Supervised approaches
  - Feature-based and kernel methods
- Semi-supervised approaches
  - Bootstrapping
- Higher-order relation extraction
- Applications
  - Question-Answering
  - Mining biomedical text
- Evaluation

THANK YOU

Feedback: nbach@cs.cmu.edu & sbadaska@cs.cmu.edu

Available Toolkits

- Parser
  - Stanford parser: syntax and dependency parser (Java)
  - MST parser: dependency parser (Java)
  - Collins parser: syntax parser (C++) ; Dan Bikel duplicates in Java.
  - Charniak parser: syntax parser (C++)
- English NP chunker
  - OpenNLP: Java
  - GATE: Java
  - Ramshaw&Marcus: Java
- Named Entities Recognizer
  - Stanford NER: Java
  - MinorThird: Java (from William Cohen's group at CMU)
  - OpenNLP
  - GATE
- Tree Kernels in SVM-light

References

- Abney, S. (2004). Understanding the yarowsky algorithm. *Comput. Linguist. (pp. 365–395). Cambridge,* MA, USA: MIT Press.
- Agichtein, E., & Gravano, L. (2000). Snowball: Extracting relations from large plain-text collections. *Proceedings of the Fifth ACM International Conference on Digital Libraries.*
- Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., & Etzioni, O. (2007). Open information extraction from the web. *IJCAI '07: Proceedings of the 20th International Joint Conference on Artificial Intelligence. Hyderabad, India.*
- Bikel, D. M., Schwartz, R. L., & Weischedel, R. M. (1999). An algorithm that learns what's in a name. *Machine Learning, 34, 211–231.*
- Blum, A., & Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. COLT: *Proceedings of the Workshop on Computational Learning Theory, Morgan Kaufmann Publishers* (pp. 92–100).
- Brin, S. (1998). Extracting patterns and relations from the world wide web. *WebDB Workshop at 6th International Conference on Extending Database Technology, EDBT '98.*
- Bunescu, R. C., & Mooney, R. J. (2005a). A shortest path dependency kernel for relation extraction. *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (pp. 724–731). Vancouver, British Columbia, Canada:* Association for Computational Linguistics.
- Bunescu, R. C., & Mooney, R. J. (2005b). Subsequence kernels for relation extraction. *Neural Information Processing Systems, NIPS 2005, Vancouver, British Columbia, Canada.*
- Culotta, A., McCallum, A., & Betz, J. (2006). Integrating probabilistic extraction models and data mining to discover relations and patterns in text. *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (pp. 296–303). New York, New York: Association for Computational* Linguistics.
- Culotta, A., & Sorensen, J. (2004). Dependency tree kernels for relation extraction. *ACL '04: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (p. 423).* Morristown, NJ, USA: Association for Computational Linguistics.
- Downey, D., Etzioni, O., & Soderland, S. (2005). A probabilistic model of redundancy in information extraction. *IJCAI (pp. 1034–1041).*
- Etzioni, O., Cafarella, M., Downey, D., Popescu, A. M., Shaked, T., Soderland, S., Weld, D. S., & Yates, A. (2005). Unsupervised Named-Entity Extraction from theWeb: An Experimental Study. *Artificial Intelligence (pp. 191–134).*
- Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating non-local information into information extraction systems by gibbs sampling. *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (pp. 363–370). Morristown, NJ, USA: Association* for Computational Linguistics.

References

- Grishman, R., & Sundheim, B. (1996). Message understanding conference - 6: A brief history. *Proceedings of the 16th conference on Computational Linguistics (pp. 466–471).*
- GuoDong, Z., Jian, S., Jie, Z., & Min, Z. (2002). Exploring various knowledge in relation extraction. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (pp. 419–444).*
- Kambhatla, N. (2004). Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. *Proceedings of the ACL 2004.*
- Liu, Y., Shi, Z., & Sarkar, A. (2007). Exploiting rich syntactic information for relationship extraction from biomedical articles. *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers (pp. 97–100). Rochester, New York: Association for Computational Linguistics.*
- Lodhi, H., Saunders, C., Shawe-Taylor, J., & Cristianini, N. (2002). Text classification using string kernels. *Journal of Machine Learning Research (pp. 419–444).*
- McDonald, R. (2004). Extracting relations from unstructured text. *UPenn CIS Technical Report.*
- McDonald, R., Pereira, F., Kulick, S., Winters, S., Jin, Y., & White, P. (2005). Simple algorithms for complex relation extraction with applications to biomedical ie. *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (pp. 491–498). Ann Arbor,* Michigan.
- Nguyen, D. P., Matsuo, Y., & Ishizuka, M. (2007). Subtree mining for relation extraction from Wikipedia. *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers (pp. 125–* 128). Rochester, New York: Association for Computational Linguistics.
- NIST (2007). The ace 2007 (ace07) evaluation plan. *http://www.nist.gov/speech/tests/ace/ace07/doc/ace07-evalplan.v1.3a.pdf.*
- PubMed (2007). Medline. *PubMed Home, http://www.ncbi.nlm.nih.gov/sites/entrez.*
- Ravichandran, D., & Hovy, E. (2002). Learning surface text patterns for a question answering system. *In proceedings of the ACL Conference.*
- Yarowsky, D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. *Proceedings of the 33rd conference on Association for Computational Linguistics (pp. 189–196). NJ,* USA.
- Zelenko, D., Aone, C., & Richardella, A. (2003). Kernel methods for relation extraction. *Journal of Machine Learning Research.*
- Zhao, S., & Grishman, R. (2005). Extracting relations with integrated information using kernel methods. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (pp. 419–426).