

CSE 454
Advanced Internet Systems
Features for Relation Extraction

 Dan Weld

Preprocessed Data Files
 Each line corresponds to a sentence. "John likes eating sausage."

tokens	after tokenization	John likes eating sausage.
--------	--------------------	----------------------------

Preprocessed Data Files
 Each line corresponds to a sentence. "John likes eating sausage."

tokens	after tokenization	John likes eating sausage.
pos	Part-of-Speech tags	John/NNP likes/VBZ eating/VBG sausage/NN ./.

Grade School: "9 parts of speech in English"

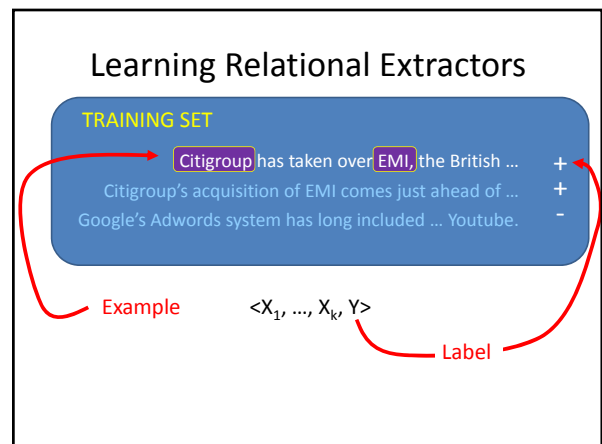
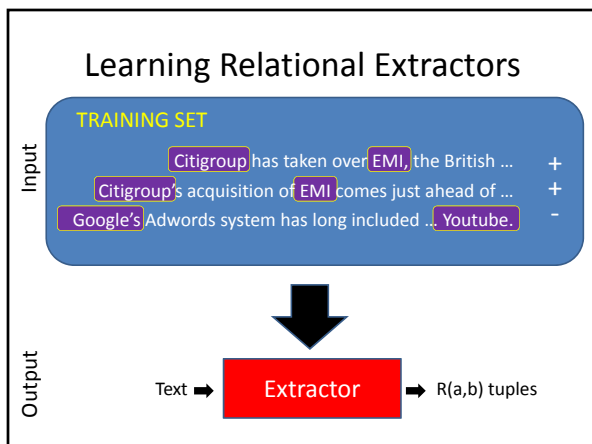
- Noun
- Pronoun
- Verb
- Adverb
- Article
- Conjunction
- Adjective
- Interjection
- Preposition

But: plurals, possessive, case, tense, aspect,

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential there
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNN	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PP	Prepositional phrase
19.	PRP	Pronoun personal
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	To
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBS	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WPS	Possessive wh-pronoun
36.	WRB	Wh-adverb

Preprocessed Data Files
 Each line corresponds to a sentence. "John likes eating sausage."

tokens	after tokenization	John likes eating sausage. .
pos	Part-of-Speech tags	John/NNP likes/VBZ eating/VBG sausage/NN ./.
ner	Named Entities	<u>John</u> likes eating sausage.



Features

Citigroup has taken over **EMI**, the British ...

$X_i =$

- NER tag of Arg1
- NER tag of Arg2
- Does word-53 (acquire) appear in span?
 - Consider all words?
 - Just use verbs & prepositions?
- Does bigram-199 (take over) appear in span?
- Trigrams?

Outside the Span

Birthplace Relation

Dan had lunch in **Boston**

Returning to his birthplace, **Dan** had lunch in **Boston**

Dan had lunch in **Boston**, his birthplace.

Proximity

Birthplace Relation

Dan, who was very tired from deadlines and cranky because of problems with his boss, was born in **Boston**

```

(ROOT
 (S
  (NP (NNP Dan))
  (. .))
  (SBAR
   (NP wh))
  (S
   (TP (TBD was)
      (SADP (SB wh)) (P (IN from))
      (NP (NNB deadlines)
          (CC and)
          (NP (NNB cranky))))
      (PP (IN because) (NP (NP (NNB problems)
          (PP (IN with)
              (NP (PP (NP his) (NN boss)))))))
      (. .))
      (TP (TBD was)
          (TP (TBD born)
              (PP (IN in)
                  (NP (NNP Boston))))
          (. .)))
  )
 )
 )
 
```

Typed dependencies, collapsed

- nsubj(tired-6, Dan-1)
- nsubjpass(born-19, Dan-1)
- cop(tired-6, was-4)
- advmob(tired-6, very-5)
- rcmod(Dan-1, tired-6)
- prep_from(tired-6, deadlines-8)
- prep_from(tired-6, cranky-10)
- conj_and(deadlines-8, cranky-10)
- prep_because_of(tired-6, problems-13)
- poss(boss-16, his-15)
- prep_with(problems-13, boss-16)
- auxpass(born-19, was-18)
- root(ROOT-0, born-19)
- prep_in(born-19, Boston-21)

Proximity

Birthplace Relation

Dan, who was very tired from deadlines and cranky because of problems with his boss, was born in **Boston**

Typed dependencies, collapsed

```

nsubj(tired-6, Dan-1)
nsubjpass(born-19, Dan-1)
cop(tired-6, was-4)
advmob(tired-6, very-5)
rcmod(Dan-1, tired-6)
prep_from(tired-6, deadlines-8)
prep_from(tired-6, cranky-10)
conj_and(deadlines-8, cranky-10)
prep_because_of(tired-6, problems-13)
poss(boss-16, his-15)
prep_with(problems-13, boss-16)
auxpass(born-19, was-18)
root(ROOT-0, born-19)
prep_in(born-19, Boston-21)
 
```

Proximity

Birthplace Relation

Dan, who was very tired from deadlines and a screaming baby, was born in **Boston**

Parsing Ambiguity

Temporal Expression Examples

Expression	Value in Timex Format
December 8, 2012	2012-12-08
Friday	2012-12-07
today	2012-12-08
1993	1993
the 1990's	199X
midnight, December 8, 2012	2012-12-08T00:00:00
5pm	2012-12-08T17:00
the previous day	2012-12-07
last October	2011-10
last autumn	2011-FA
last week	2012-W48
Thursday evening	2012-12-06TEV
three months ago	2012.09

Reference Date = December 8, 2012

Slide from Dan Roth, Heng Ji, Taylor Cassidy, Quang Do TIE Tutorial 19

Temporal Expression Extraction

- Rule-based (Stratgen and Gertz, 2010; Chang and Manning, 2012; Do et al., 2012)
- Machine Learning
 - Risk Minimization Model (Boguraev and Ando, 2005)
 - Conditional Random Fields (Ahn et al., 2005; UzZaman and Allen, 2010)
- State-of-the-art: about 95% F-measure for extraction and 85% F-measure for normalization

Slide from Dan Roth, Heng Ji, Taylor Cassidy, Quang Do TIE Tutorial 20

Ordering events in discourse

- (1) John entered the room at 5:00pm.
- (2) It was pitch black.
- (3) It had been three days since he'd slept.

Slide from Dan Roth, Heng Ji, Taylor Cassidy, Quang Do TIE Tutorial 21

Ordering events in time

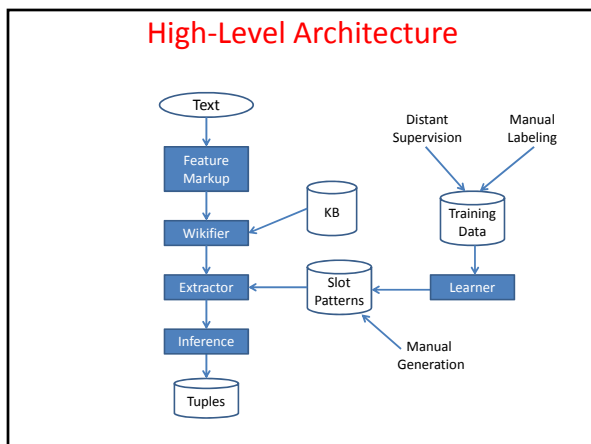
- Speech (S), Event (E), & Reference (R) time (Reichenbach, 1947)

Sentence	Tense	Order
John wins the game	Present	E,R,S
John won the game	Simple Past	E,R<S
John had won the game	Perfective Past	E<R<S
John has won the game	Present Perfect	E<S,R
John will win the game	Future	S<E,R
Etc...	Etc...	Etc...

- **Tense:** relates R and S; **Gr. Aspect:** relates R and E
- R associated with *temporal anaphora* (Partee 1984)
- Order events by comparing R across sentences
- By the time Boris noticed his blunder, John had (already) won the game

See Michaelis (2006) for a good explanation of tense and grammatical aspect

Slide from Dan Roth, Heng Ji, Taylor Cassidy, Quang Do TIE Tutorial 22



Teams

- Named Entity Linking (1)
- Time (1)
- Distant Supervision (1)
- InstaRead (1)
- Relation-Specific (3-5)

Slide from Dan Roth, Heng Ji, Taylor Cassidy, Quang Do TIE Tutorial