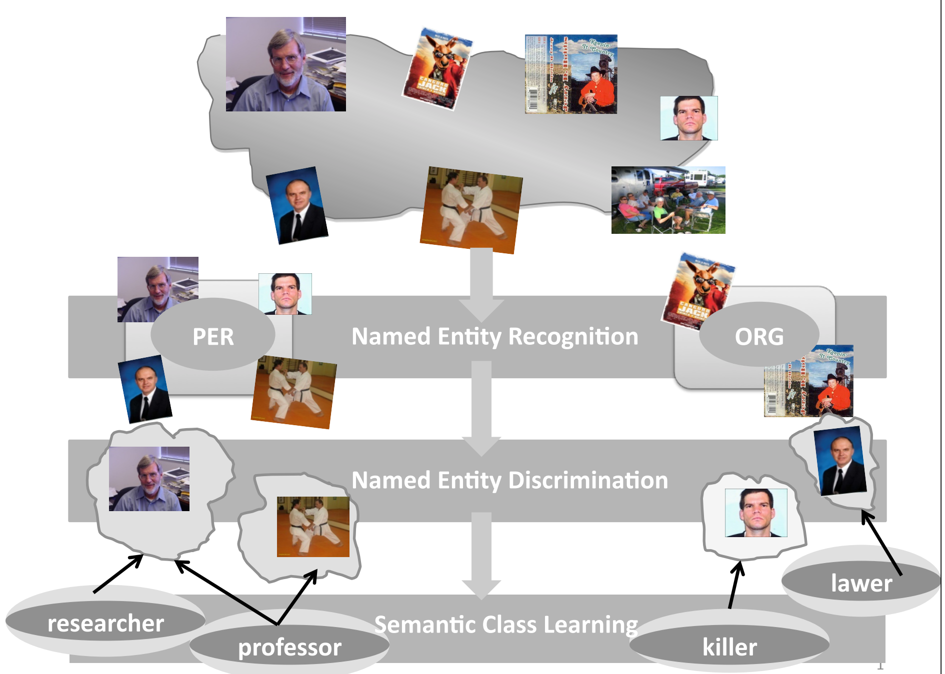
CS544: Semantic Class Learning

April 8, 2010

Zornitsa Kozareva
USC/ISI
Marina del Rey, CA
kozareva@isi.edu
www.isi.edu/~kozareva



Semantic Class Learning: Objectives

- Given a class and an instance, learn automatically with minimum supervision new instances, classes and the ISA relations among them.

- Examples:
  - *class_name*: Nobel prize winners
  - *instances*: Albert Einstein, Max Plank …

  - *class_name*: former Russian federation states
  - *instances*: Georgia, Urkaine, Lithuania …

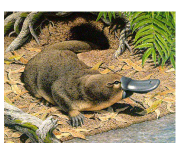
2

Why Semantic Class Learning (1)

- It is valuable for current NLP applications.
- Question Answering:
  – How are Max Planck, Angela Merkel, Jim Gray and Dalai Lama related?

all four have doctoral degrees from German universities

- Information Retrieval:
  – mammals that lay eggs



platypus

3

Why Semantic Class Leaning (2)

- WordNet has limited coverage
  - many instances and classes are missing
  - knowledge does not cover all domains

Ex. if you are interested in extracting:

- *all names of US presidents, you will notice that the name of Barack Obama is not present*
- *Chinese, French, Italian presidents, you will notice that these classes and their instances are not listed at all*
- *the amount of information present for animals vs. people is different*

4

Why Semantic Class Leaning (3)

- Even the biggest knowledge repository must be constantly updated, over time instances of a class may change

Ex. Presidents of a Country

- Barack Obama (2009-present)
- George Bush (2001-2009)

Country Names

- Czechoslovakia (1918-1992)
- Spain

5

Characteristics

- Semantic classes are diverse:
  - closed
    - small (names of countries, states, planets)
    - large (names of diseases, cities)
  - open

    Ex. singers, movie titles
- Users might not know sample instance of a class
- An instance can belong to multiple classes

  Ex. orange the *fruit* vs. orange the *color*

6

Challenge

- The relevant information is scattered across different sources
- Automatic knowledge acquisition is necessary
- How does one evaluate precision and recall for the harvested information?
  - currently no repository that contains all the information

7

Lexico-Syntactic Patterns (Hearst 92)

(S1) `Agar is a substance prepared from a mixture of red algae, such as Gelidium, for laboratory or industrial use.`

(1a) NP_0 *such as* NP_1 $\{, NP_2 \ldots ,$ *(and | or)* $NP_i\}$ $\quad i \geq 1$

are such that they imply

(1b) *for all* NP_i, $i \geq 1$, *hyponym*(NP_i, NP_0)

Thus from sentence (S1) we conclude

hyponym("Gelidium", "red algae").

Examples are adapted from Marti Hearst

Lexico-Syntactic Patterns (Hearst 92)

(2) *such NP as {NP ,}* {(or | and)} NP*

`... works by such authors as Herrick, Goldsmith, and Shakespeare.`

$\Longrightarrow$ *hyponym("author", "Herrick"),*
hyponym("author", "Goldsmith"),
hyponym("author", "Shakespeare")

(3) *NP {, NP}* {,} or other NP*

`Bruises, ..., broken bones or other injuries ...`

$\Longrightarrow$ *hyponym("bruise", "injury"),*
hyponym("broken bone", "injury")

Examples are adapted from Marti Hearst

Properties

- A good pattern
  - should occur frequently in text
  - should (nearly) always suggest the relation of interest
  - should be recognizable with little pre-encoded knowledge.

Examples are adapted from Marti Hearst

Examples

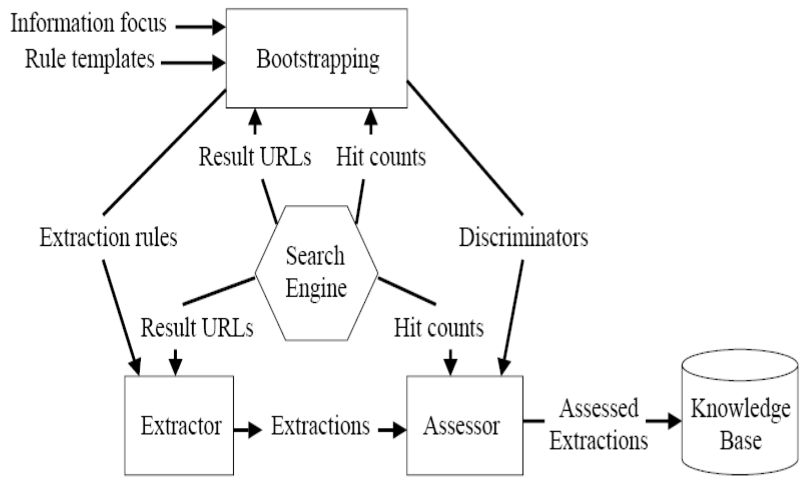
- **Cities** such as **Boston**, **Los Angeles**, and **Seattle**…”

(“**C** such as **NP1**, **NP2**, and **NP3**”) => IS-A(each(head(**NP**)), **C**)

- Detailed information for several **countries** such as **maps**

- I listen to pretty much all music but prefer **country** such as **Garth Brooks**

KnowItAll Architecture (Etzioni et al.05)



Learning Cities

- Input:
  - search query:
    - *"city; town"*, *"cities; towns"*
  - extraction rules (Hearst 92):
    - <*class2*> such as <NPList>
    - <NP> is a <*class1*>
    - <*class2*> including <NPList>

- Generate extraction queries for search engine:
  - "*cities* such as"
  - "is a *town*"
  - "*towns* including"

13

Learning Cities

- Submit extraction queries to Google and collect the returned snippets:

Central Highlands Council - Welcome - Enjoy the historic buildings ...
Enjoy historic buildings and friendly **towns including Bothwell**, Hamilton, Gretna and Ellendale to name a few. Fish at great fishing spots.
www.centralhighlands.tas.gov.au/ - Cached - Similar

Wichita, Kansas RE/MAX Agent serving Wichita and surrounding **towns** ...
Wichita, Kansas RE/MAX realtor serving Wichita, Goddard, Maize, Bentley, Halstead, Sedgwick, Park City, Valley Center, Bel Aire, Andover, Derby, Rose Hill, ...
www.wichitarealestate4you.net/ - Cached

Public Health And Poor-Law Medical Services
towns, including London,. 6,l44 births and 5,l67 deaths were registered during the week ending Saturday,. July 25th. The annual rate of mortality ...
www.jstor.org/stable/20236873

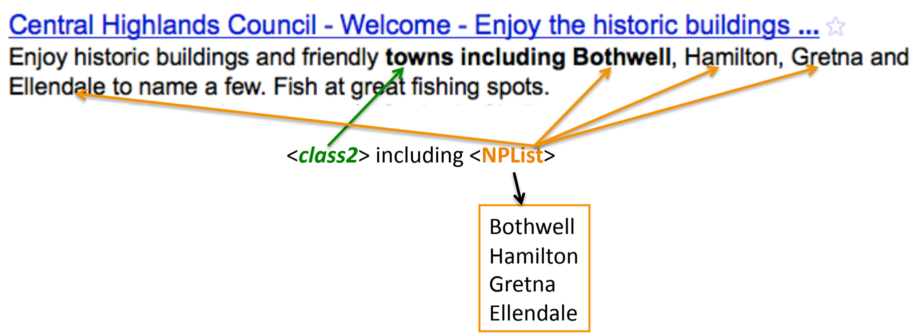
John D. Williams, M.D., B.Sc.Edin., Honorary Gynæcologist To The ...
by JWB - 1901
towns, including London,. 6561 births and 3674 deaths were registered during the week ending Saturday last, May 25th. The annual rate of mortality ...
www.jstor.org/stable/20268562

Sanitary and meteorological notes
annually of 21"2 in twenty-eight large English **towns (including London**, in which the rate was 19"7), 30"8 in the sixteen chief towns of Ireland, ...
www.springerlink.com/index/30401P77HV34488X.pdf

14

Extracting City Names

- Pull all ***candidate*** city names from the snippets using extraction rules



15

Assessing Candidates

- Generate *discriminators* (from rules and user input):
  - cities such as <*Candidate*>
  - <*Candidate*> is a town
  - <*Candidate*> is a city
  - towns including <*Candidate*>

- Generate *discriminator queries* (from discriminators and candidates):
  - cities such as *London*
  - *London* is a town
  - *London* is a city
  - towns including *London*

16

Assessing Candidates

- Evaluate each *candidate* with each *discriminator query* and compute PMI *as:*

$$PMI(Cnd, Disc) = \frac{|\, Hits(Disc + Cnd)\, |}{|\, Hits(Cnd)}$$

$$PMI(London, city) = \frac{Hits(city \quad London)}{Hits(London)} = \frac{8{,}590{,}000}{533{,}000{,}000} = 0.0161$$

$$PMI(Avocado, city) = \frac{Hits(city \quad Avocado)}{Hits(Avocado)} = \frac{5{,}980}{8{,}320{,}000} = 0.000718$$

$$PMI(London, city) >> PMI(Avocado, city)$$

17

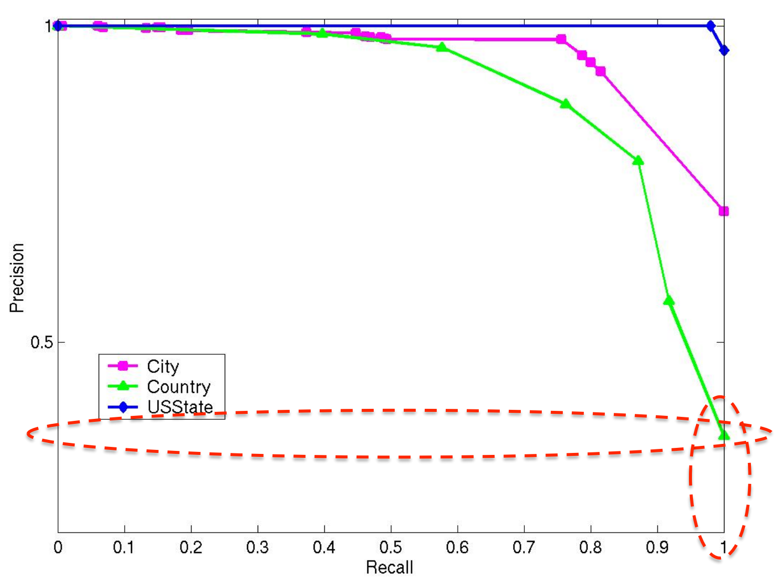
Assessing Candidates

- Train NaïveBayes classifier using PMI as features
- Training set contains positive and negative instances of the class
  - choose ***n*** candidates
  - compute average PMI, take ***m*** candidates with highest average PMI as positive examples and ***m*** candidates with lowest average PMI as negative examples
  - select ***k*** best discriminators tested on ***m***
- Evaluate all candidates on ***k*** discriminators

18

This slide was adopted from Oren Etzioni

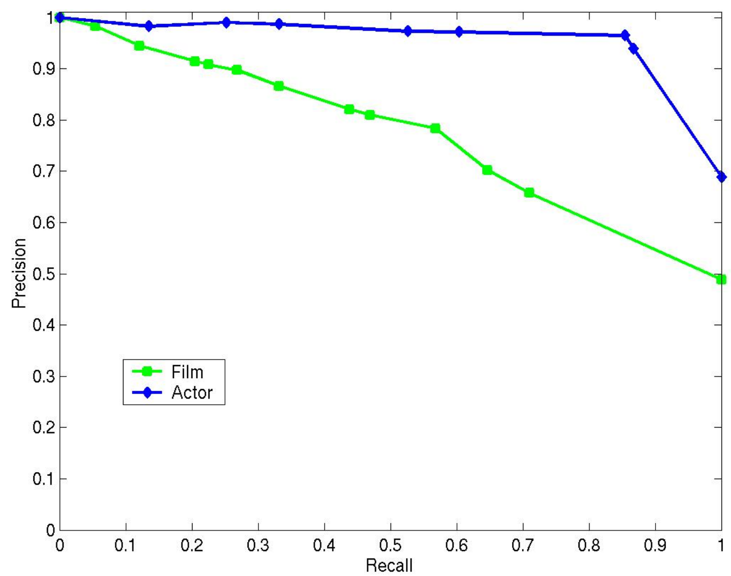
Results for City, Country and US State extraction



19

This slide was adopted from Oren Etzioni

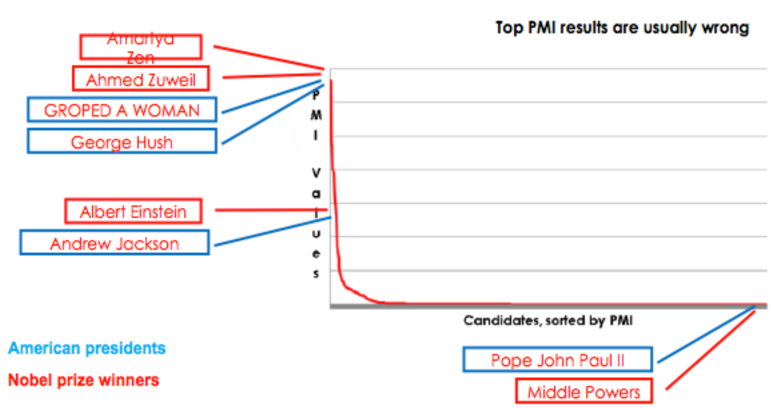
Results for Actor and Film Extraction



20

This slide was adopted from Luka Bradesko

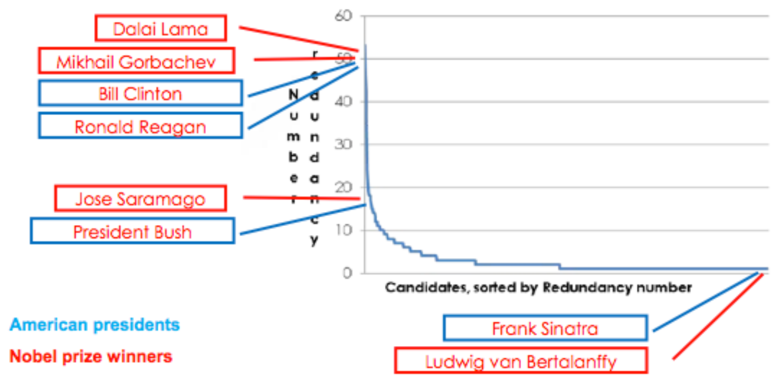
Bradesko's implementation of KnowItAll



- Top PMI features are not always useful
- An extractor with high PMI can harvest wrong candidate examples

21

Bradesko’s suggestion



- Look for redundancy of candidates rather than PMI

22

Results for Nobel Prize Winners and American Presidents

KnowItAll PMI based		
	Precision	Recall
Nobel Winner	83.7	53.4
American President	66.0	81.4

Bradesko Redundancy based (first 100,35)	
Precision	Recall
100	12.7
90	65

23

Next …

- How to choose synonyms for class expansion? (this can be tricky even for humans)
- How many seed examples are necessary to learn the instances of a class?
- How to eliminate ambiguous examples?
- Can we improve precision/recall?
- How well does the method scale?

24