# Using WordNet glosses to refine Google queries

Jan Nemrava

Department of Information and Knowledge Engineering,
University of Economics, Prague

nemrava@vse.cz



### The Present

- Web search results affected by synonymy
- Inexperienced users inserting too general queries
- Ambiguity of inserted queries

- Goal: Discover synonymy in proper nouns
- Pankow inspiration

### Some of solutions

- Offer query refinement
  - Zoom in
  - Zoom out
    - Ask Jeeves (ask.com)
- Term suggestions
  - Not grammatically and meaningfully perfect
    - Google Suggest Beta, Centrum, Seznam (http://www.google.com/webhp?complete=1&hl=en)
- Offer arranged results
  - According to synonymy class
  - Vivisimo



# Discovering synonymy

- Information Sources
  - WordNet
  - monothetic clustering
  - Google API
  - NLP tools
  - Hearst Patterns



### WordNet and NLP

- WordNet contains 150,000 words organized in over 115,000 synsets
- Each word has a short description gloss
- Key idea is to check if nouns in glosses contain the meaning (i.e. superclass) of the word.

- NLP
  - Stop words list and POS tagger



### Hearst Patterns

 lexico-syntactic patterns indicating the existence of class/subclass relation in unstructured data sources

- $NP_0$  such as  $NP_1$ ,  $NP_2$ , . . ,  $NP_{n-1}$  (and | or)  $NP_n$
- such  $NP_0$  as  $NP_1$ ,  $NP_2$ , . .,  $NP_{n-1}$  (and | or)  $NP_n$
- $NP_1$ ,  $NP_2$ ,...,  $NP_{n-1}$  (and | or) other  $NP_0$
- $NP_0$  (including—especially)  $NP_1$ ,  $NP_2$ , . . ,  $NP_{n-1}$  (and | or)  $NP_n$
- and very common "NP<sub>i</sub> is a NP<sub>0</sub>"

# Technique

- get all synonyms and their glosses for given word from WordNet
- Process all glosses with NLP tools to get candidate nouns
- Apply "is a" pattern on each of candidate nouns and count the number of results
- Validate the results with "and other" pattern
- Compare the results and give the most probable hypernym



# Extracting "candidate nouns"

- Considering the proper noun PLUTO
  - 3 meanings in WordNet (planet, God, cartoon) Glosses
  - a small planet and the farthest known planet from the sun; has the most elliptical orbit of all the planets
  - (Greek mythology) the god of the underworld in ancient mythology; brother of Zeus and husband of Persephone
  - a cartoon character created by Walt Disney



# Extracting "candidate nouns"

- Considering the proper noun PLUTO
  - 3 meanings in WordNet (planet, God, cartoon)

### Glosses

- a small planet and the farthest known planet from the sun; has the most elliptical orbit of all the planets
- (Greek mythology) the god of the underworld in ancient mythology;
   brother of Zeus and husband of Persephone
- a cartoon character created by Walt Disney
- Candidate nouns
- planet;sun;orbit;planets;
- Greek;god;underworld;mythology;brother;Zeus;husband;Persephone;
- cartoon;character;Walt;Disney;

# Extracting "candidate nouns"

- Considering the proper noun PLUTO
  - 3 meanings in WordNet (planet, God, cartoon)

### Glosses

- a small <u>planet</u> and the farthest known <u>planet</u> from the <u>sun</u>; has the most elliptical <u>orbit</u> of all the <u>planets</u>
- (Greek mythology) the <u>god</u> of the <u>underworld</u> in ancient <u>mythology</u>;
   brother of Zeus and <u>husband</u> of <u>Persephone</u>
- a <u>cartoon</u> character created by Walt Disney
- Candidate nouns
- planet;sun;orbit;planets;
- Greek;god;underworld;mythology;brother;Zeus;husband;Persephone;
- cartoon;character;Walt;Disney;

# Applying Two Patterns

- "Pluto is a planet" (1550), "Pluto is planet" (145)
- "Pluto is a sun" (2), "Pluto is sun" (0)
- "Pluto is a orbit" (0), "Pluto is orbit" (1)
- "Pluto is a planets" (0), "Pluto is planets" (0)

For each of 16 candidate nouns create similar pattern and search for it

### Validation

- "Pluto and other planets" (57)
- "Pluto and other planet" (0)
- "Pluto and other suns" (0)
- "Pluto and other sun" (0)
- Problem with adjectives
- Plural problem

**Candidate nouns** 

planet;sun;orbit;planets;



# Applying Two Patterns

- "Pluto is a planet" (1550), "Pluto is planet" (145)
- "Pluto is a sun" (2), "Pluto is sun" (0)
- "Pluto is a orbit" (0), "Pluto is orbit" (1)
- "Pluto is a planets" (0), "Pluto is planets" (0)

For each of 16 candidate nouns create similar pattern and search for it

### Validation

- "Pluto and other planets" (57)
- "Pluto and other planet" (0)
- "Pluto and other suns" (0)
- "Pluto and other sun" (0)
- Problem with adjectives
- Plural problem



### Results

Test set: 50 nouns from travel, space and zodiac

Table 1. Overall precision

Total number of words in list	50	(100%)
Words listed in WordNet	48	(96%)
Correct	39	(78%)
- completely correct	31	(62%)
- partially correct	8	(16%)
Wrong	9	$(18\%)_{\rm stor}$

## Results

### Bad results

Table 2. Statistics of wrongly discovered terms

Number of wrong instances	17	(100%)
Both patterns wrong	7	(41%)
"is a" correct, "and other" wrong	4	(23%)
"is a" wrong, "and other" correct	6	(35%)





Web Images Groups News Froogle Maps more »

Pluto

Google Search

specify the context:



Pluto as a planet

### Hades Pluto God at the Underworld

Plain Plain see the Horsen name for the Greek and Hades. Myo: In Direct and Horsen mythology, the Glyx was the principal over in the unservants..... welleruneshades htm - 6k - Gached - Saniar pages

### Pluto planet and Greek god of weath

Plain Greek god of weeth, rated the descendensarif of myth (arth plane) ... Plate was the god of the lower world in Greek mothelogy whose name meetil ...
www.wordsources.inforplace.hant Tis Cached Similar space.

Plato Plato Clubb Symbol Plato, God of the Underweld, a fine rule of Scarpin ... The Book of Plato by Steven Longet, Cale Admings Role: The book is ... www.catecstralegy.com/plute.html 15k Cached Similar saces

Pluto
Pluto Pluto socities guilafitie Desc. When someone died frey randied down to
the thirdworld have they had no creative bear or the loved ...
some goods down to elisiphen any pluto him. Th. (2020) 12:00:02.00022



### Disney's <u>Muto</u> Classic Shorts Series on tog Cartoon Database Episode quite for Plain Cartoon Episode Guide Compile Filmocraphy and information.

The Big Cartoon DataBase is the ubmate searchable list and episods www.bodb.com/cartoons/Not\_Disney\_Studies/Shorts/Flute/ - Similar pages

### Disney Shorts Mouseworks "A Pluto Cartoon" (Third Season)

Information on the Phate educts from Michay's Mance Works.
disneyshars heareone netime use works iphose Print - 1th - Gazzed - Garnier pages.

Partry Pirate, A Plate Carteon, Paracycling (see Gody's Extreme Sports : Paracycling) ... Platrice Cris-Puplets. A Plate the Pup Carroon ... diarregations.tocrasine.not/index/plront - 18k - Cachod - Similar super

### Pluto - Disney Characters History

Plate, scke/s fatirul pel deg Plute starred in 48 of its own cartoons, but also appeared sharp with Mickey Macses and Daniel Duck in many of their contours ... disney go comhochlarchwesicharacterstandardigitate/place intril. Rik. Cached. Smiler pages.



### Pluto as a God

Hartes Plute God of the Lindenword
Plute Plute Was the Roman name for the Greek god Hades. Styx: in Greek and Roman mythology, the Styrt was the principal over in the underworld. ... using not hades then - lik - Dacrad - Senior pages

Pluto, planet and Greek god of weath.

Pluto, Creek god of weath, used the date underwork of myth jointh planet ...

Pluto wee the god of the lower world in Greek mythology whose rame meant ... www.wordsources.info/plute.html 7h - Coched - Dimilar pages

Plate. Plate Gliph Symbol. Plate. God of the Undersofts, is the ruler of Scerpio. ... The Book of Plate by Stown Forest. Cafe Astrology Note: This book is .... now columnicating complete land - 15t - Cachad - Senio comme

Pluto Pluto sate the god of the Dead When someone died, they to redied down to the Underworld. First, they had to cross the Area of the Dead .... now guydr demance alujahomaniphoto htm - Jik - <u>Cached</u> - <u>Similar pages</u>



### Drawbacks and Future

### Drawbacks

- No proximity search in Google
- Slow response of Google API
- Lot of queries to discover one meaning

### Future:

- Add another validation step
- Discover collocations in results
- Discover Named Entities



### Questions?

Thank you for your attention

