

# SEMANTIC CLUSTERING OF QUESTIONS

RESEARCH REPORT, 2<sup>ND</sup> SEMESTER

Cristina Groapă

# Problem statement

2

- Part of the Smart Presentation project
- Efficient management of audience feedback
- Question clustering:
  - ▣ Suggest similar asked questions
  - ▣ Group all questions according to topic
- Important: **real-time process**

# Specificity

3

- **Specificity = Information Content**
- E.g. {*collie, sheepdog*} vs. {*go, be*}
  
- Evaluation:
  - ▣ Taxonomy depth
  - ▣ Corpus-based  $TF - IDF(term) = \frac{D}{N}$
  
- Combine with measures of semantic similarity for better results

# Semantic Similarity Measures

4

## □ Path-based

- Leacock-Chodorow:  $sim_{LCh}(c_1, c_2) = -\log \frac{len(c_1, c_2)}{2D}$

## □ IC-based

- Resnik:  $sim_{Res}(c_1, c_2) = IC(lcs(c_1, c_2))$

## □ Semantic Relatedness

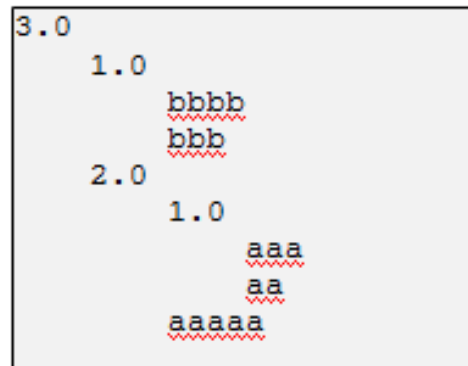
- Hirst-and-St. Onge:  $rel_{HStO}(c_1, c_2) = C - len(c_1, c_2) - k * d$

# NLP Tools

5

- Stanford CoreNLP

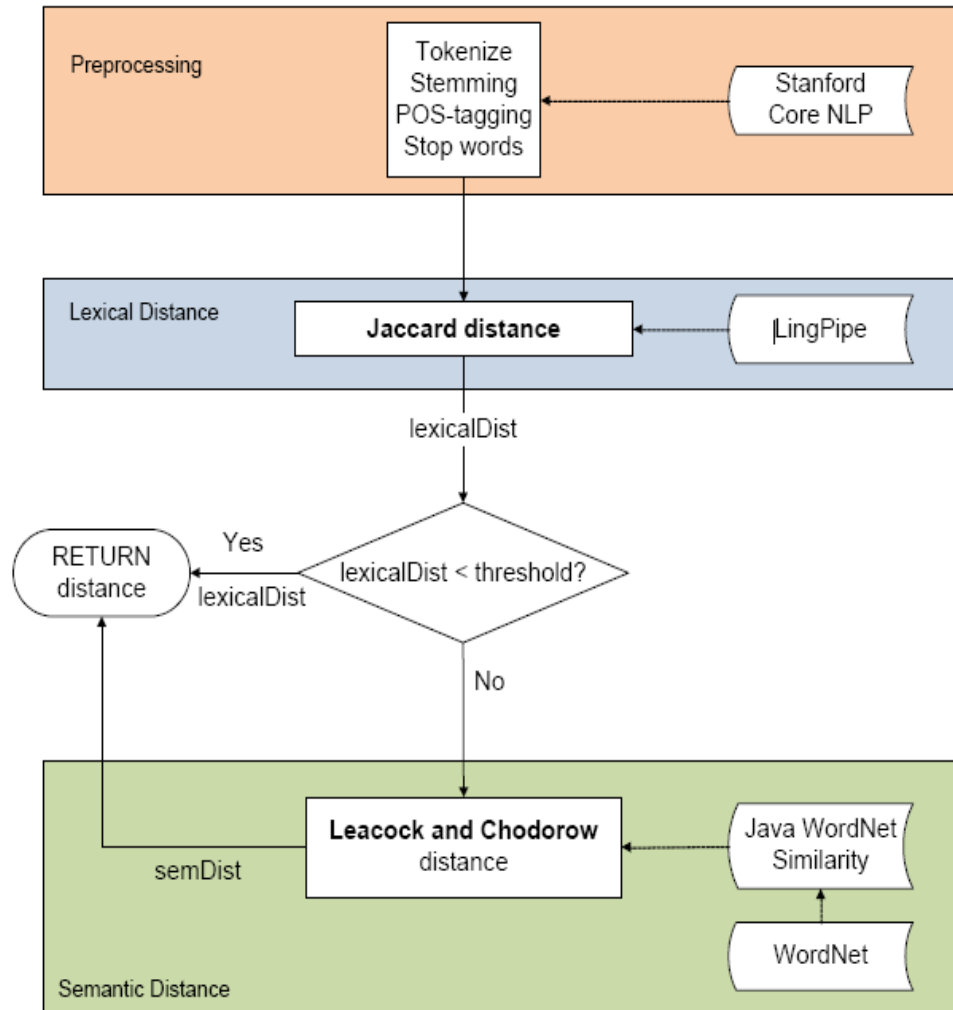
- LingPipe



- Java WordNet::Similarity

# Implementation

6



# Results

7

- **143 questions ~ 8 min** (dualCore 2GHz processor, 3GB RAM)

- Good:

0.816817225518873

0.6911711540417835

What color is a poison arrow frog ?

What is the gestation period for a cat ?

What do bats eat ?

- Bad:

0.6354185724342598

What person 's head is on a dime ?

How much does the human adult female brain weigh ?

# Results (2)

8

## □ Good and bad:

0.8113508830299405

0.7777011702584125

0.6478291732270708

Who was elected president of South Africa in 1994 ?

Which president was unmarried ?

Who was the first governor of Alaska ?

Where is Prince Edward Island ?

0.8623704751655283

0.8166644890005959

Who was Abraham Lincoln ?

When did John F. Kennedy get elected as President ?

Who was the first female United States Representative ?



# Future work

9

- Test on real data
- Increase weight on NERs compared to common nouns
- Introduce specificity
- Word Sense Disambiguation

# Thank you

10

- Questions?