Removing Disambiguation from Search Text Patterns

Presented By
Sarita Mehta ,Karan Seth, Rahul Dureja, Salim Ali Khan
(in no particular order)

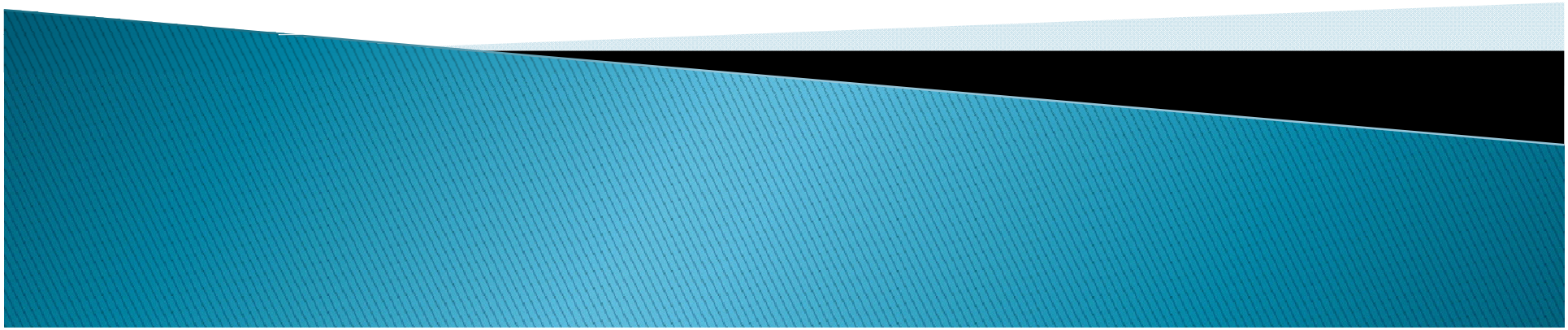


Table of Contents

- What makes our work interesting?
- Goal/Objective
- Resources Used
- How we did it?
- Results (Demonstration)
- Conclusion
- Future Work

Table of Contents

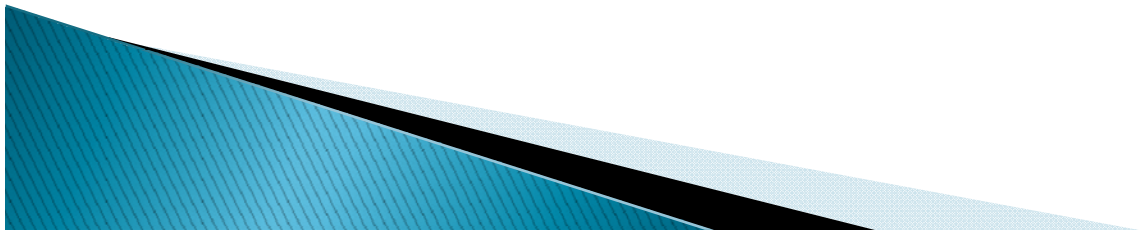
What makes our work interesting?

Word Sense Disambiguation

- Identifying the sense of a ‘word’ in a ‘sentence’ or ‘context’.
- Is an open problem of natural language processing.

I went fishing for some sea bass.

The bass line of the song is too weak.



Going Back in Time…

- WSD goes back to 1940 during the days of Machine Translation. (one of the oldest problems in computational linguistics).
- Warren Weaver in his memorandum of translation first introduced this problem.
- In 1970s WSD systems were largely rule based or hand coded.
- In 1980s Oxford Advanced Learner's Dictionary replaced the hand coding.
- In 1990s, statistical revolution swept through computational linguistics. WSD became a paradigm problem.
- The 2000s saw supervised techniques reach a plateau in accuracy, and so attention has shifted to coarser-grained senses, domain adaptation, semi-supervised and unsupervised corpus-based systems

Too Many Dictionaries

- A major blockade that any algorithm which aims to solve the WSD problem is dealing with different dictionaries.
- Every Dictionary gives different meanings to the same word. As an example the lexical dictionary WordNet and Roget's Thesaurus give different meanings for words.
- Here we use the WordNet lexical dictionary. We will tell you why.

Table of Contents

- What makes the work interesting
- Goal/Objective
- Resources Used
- How we did it?
- Results (Demonstration)
- Conclusion
- Future Work

Goal/Objective

- Removing disambiguation from search text patterns on a domain specific Knowledge Base. (We use a University Ontology as a Knowledge base)
- Interpreting what the user wants, not just what the user types.

I need all courses offered in fall by Prof X.
It doesn't mean show me all courses.

Table of Contents

- What makes our work interesting
- Goal/Objective
- Resources Used
- How we did it?
- Results (Demonstration)
- Conclusion
- Future Work

Resources Used

LUBM University Ontology

- Developed to facilitate the evaluation of Knowledge Bases in a standard and systematic way.
- Knowledge Base with respect to extensional queries over a large data set that commits to a single realistic ontology.
- It consists of university domain ontology, customizable and repeatable synthetic data, a set of test queries, and several performance metrics

Resources Used (contd)

WordNet

- Large lexical database of English, developed under the direction of George A. Miller.
- Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept.
- Synsets are interlinked by means of conceptual-semantic and lexical relations.

Table of Contents

- What makes our work interesting
- Goal/Objective
- Resources Used
- How we did it?
  - The Knowledge Base
  - The Schema
  - The Process
- Results (Demonstration)
- Conclusion
- Future Work

How we did it?

- Used the LUBM University ontology (http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl) which contains all aspects such as classes, relations, sub-classes, sub-classes, etc.
- A parser was coded from scratch which parsed the University Ontology and loaded the data from the ontology in a MySQL database.
- The above formulates our Knowledge Base or just KB.

The Knowledge Base

- The KB constitutes of the following,

Classes (i.e. all the classes of the ontology)

Relation (which defines object property of the classes)

Elements (which are the instances of the classes)

The Schema

- Our Knowledge base is a relational database which contains table which contain all properties of a relational database (Primary Keys, Foreign Keys, Indexes, etc).
- The detailed explanation of the tables is given below:

Document

The Process

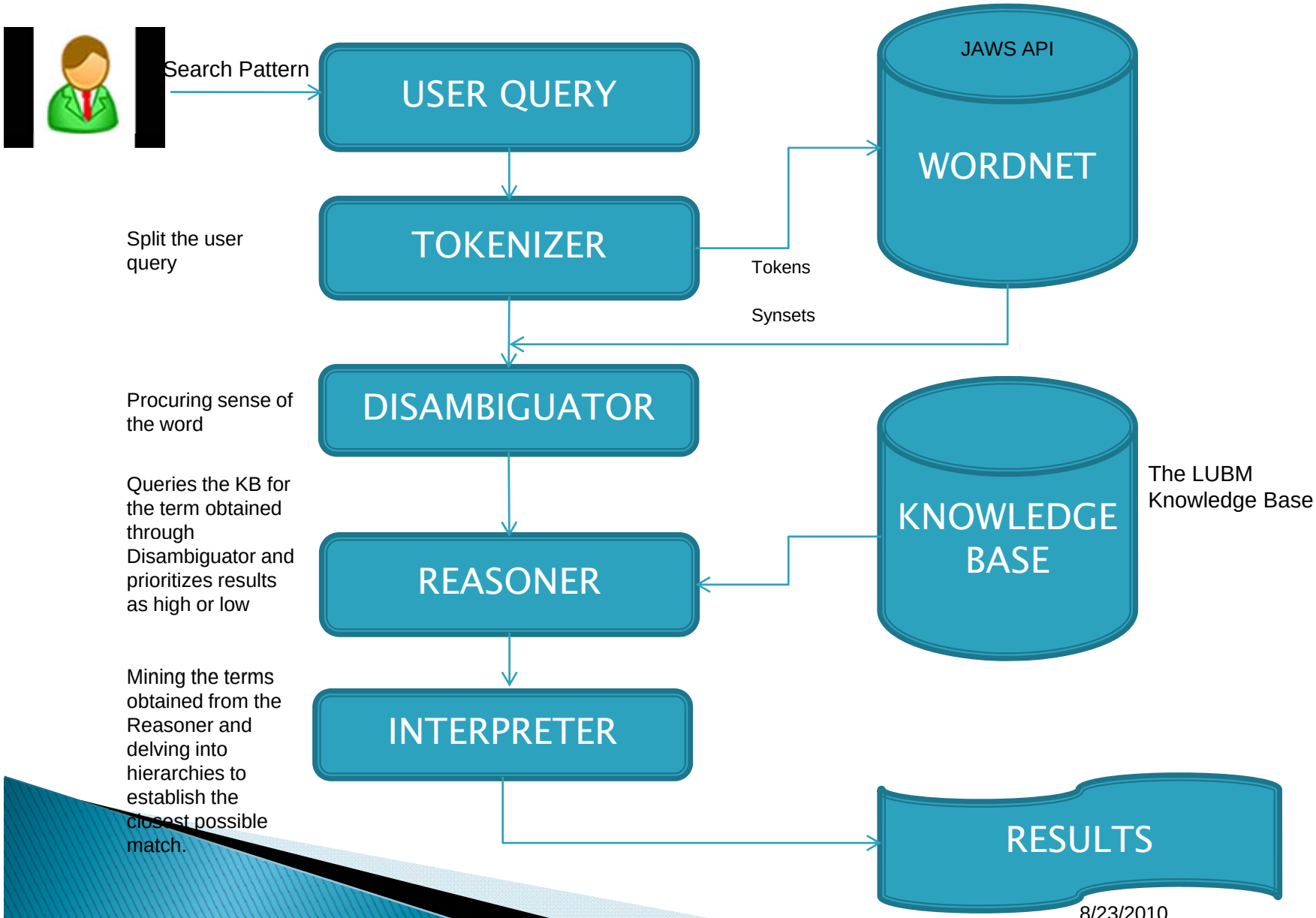

Search Pattern
USER QUERY
TOKENIZER
DISAMBIGUATOR
REASONER
INTERPRETER
RESULTS
JAWS API
WORDNET
KNOWLEDGE BASE
Split the user query
Tokens
Synsets
Procuring sense of the word
Queries the KB for the term obtained through Disambiguator and prioritizes results as high or low
Mining the terms obtained from the Reasoner and delving into hierarchies to establish the closest possible match.
The LUBM Knowledge Base

Table of Contents

- What makes our work interesting
- Goal/Objective
- Resources Used
- How we did it?
- Results (Demonstration)
- Conclusion
- Future Work

Results (Demonstration)

All talk and no demo makes tall claims. So here it goes…

Demo screenshot

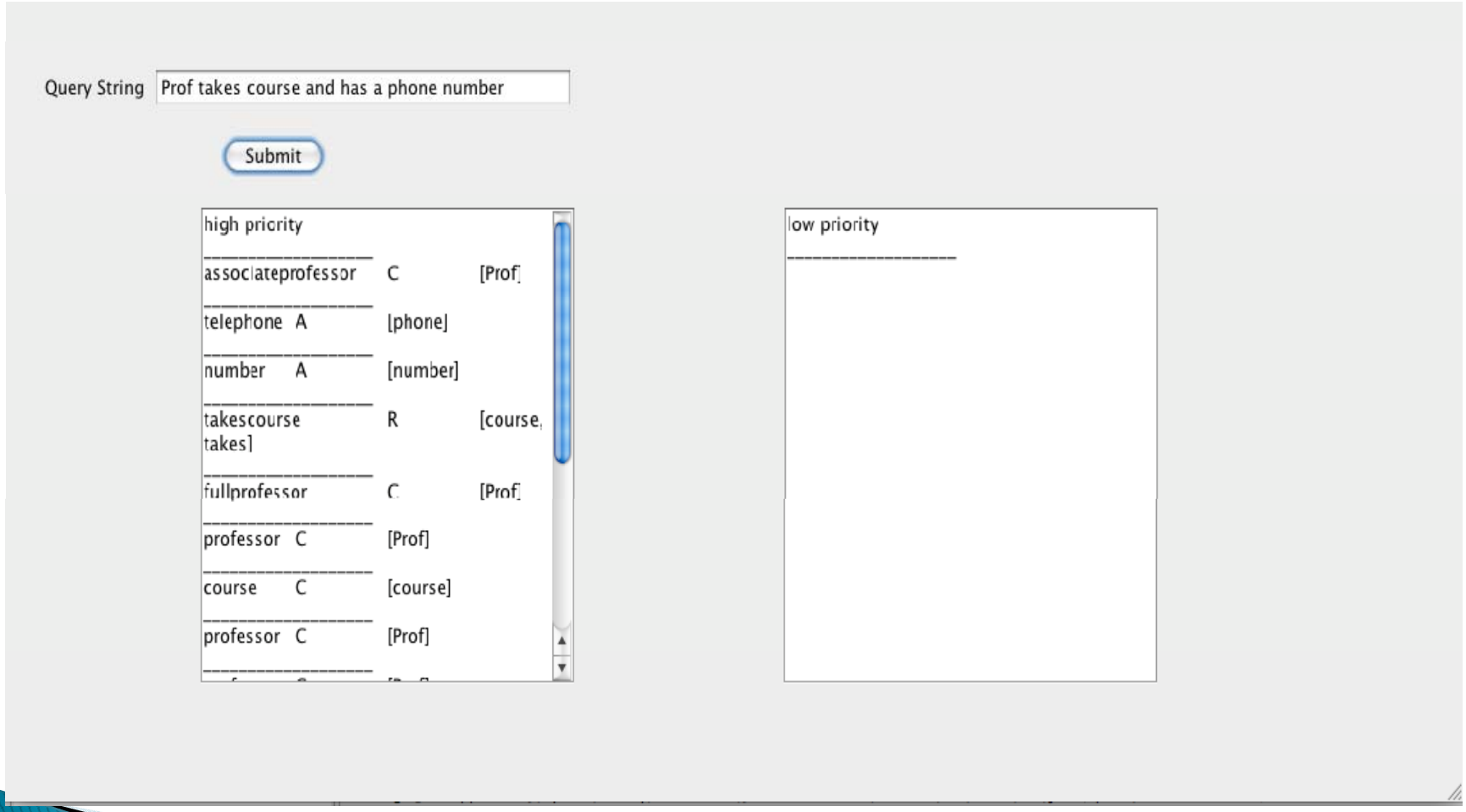


Table of Contents

- What makes our work interesting
- Goal/Objective
- Resources Used
- How we did it?
- Results (Demonstration)
- Conclusion
- Future Work

Conclusion

- Multiple ontologies could share same keywords/tokens.
- Disambiguation allows sense interpretation of the context.
- Sense helps us to mine down through large amounts of irrelevant data.
- This also increases the percentage of relevant results by adding semantics.

Table of Contents

- What makes our work interesting
- Goal/Objective
- Resources Used
- How we did it?
- Results (Demonstration)
- Conclusion
- Future Work

Future Work

- Betterment of disambiguation algorithm.
- Setting up the link between the Reasoner and Interpreter.
- Making the interpreter intelligent enough to understand the users' need.
- Further exploring the power of WordNet by using concepts like Hypernyms, Hyponyms.
- We plan to make use of JWSL (Java WordNet Similarity Library) to better our disambiguation algorithm.
- Used the noun sense for the words and plan to use adjectives and verb sense going ahead.
- We aim at satisfying the 14 queries posed by LUBM so that our system satisfies the benchmark

A Special Thanks

- Prof. Dennis McLeod
- Sang Su Lee
- Giuseppe Pirro (JWSL)
- Brett Spell (JAWS)
- Miller, Fellbaum, Tengi (WordNet)
- LUBM (Leigh University)

Merci!!!

Thank You

Danke!!!

شكرا لك

Gracias

धन्यवाद