Knowledge representation

lecture 2: processing natural language

T. Tammet, TUT

Our plan for NLP and labs

The current plan is to have two labs as parts of one project:

Process English text to:

- Extract knowledge from English to RDF

- Solve ambiguities and answer queries using 1st order logic.

The current lecture is an intro to natural language processing as a bacground for the first lab: extracting potentially ambigous knowledge to RDF.

Kinds of NLP processing goals

There exist highly different methods for NLP, depending on the goals we have:

- Text similarity detection

- Automatic summarization

- Machine translation

- Sentiment analysis

- Speech processing

- Information extraction

- Question answering

- ...

NLP and restricted English

- By **NLP** people normally mean processing „generic" unrestricted English: we will only partially understand the text and we will often misunderstand important parts.

- By **„restricted English"** people normally mean creating and processing a formal and restricted, yet „natural-looking" English, with an exact meaning given to each sentence. Easy to understand, but really hard to write properly

Statistics and reasoning in NLP

- **Statistics and learning for creating probabilistic rules**

  It is very hard to capture and write down „language rules":

  - There are far too many

  - Huge number of exceptions, exceptions to exceptions, ...

  - Everybody has her own version of language, and it changes all the time

- **Logic-based reasoning for disambiguation and knowledge extraction:**

  - Disambiguate the meaning of „she", „this", multi-meaning phrases etc

  - Convert parsed and annotated text to processable database form

  - Add general contextual knowledge

  - Answer questions about the text

Basic steps in NLP

Rough guide to what we have to do for knowledge extraction:

- Named entity recognition (NER)

- Mapping NER-d entities to useful global ID-s

- Annotating text with parts-of-speech, aka POS tagging

- Parsing: many different approaches

- Disambiguation of words and phrases („this", „she", ...)

- Conversion to some database/RDF/logic format

Next steps

We will look into links in the notes file

https://courses.cs.ttu.ee/pages/Useful_NLP_links_and_notes

Starting with Jurafsky/Manning Stanford course:

- NER

  http://spark-public.s3.amazonaws.com/nlp/slides/Information_Extraction_and_Named_Entity_Recognition_v2.pptx

- POS tagging

  http://spark-public.s3.amazonaws.com/nlp/slides/Maxent_PosTagging.pptx

- Parsing

  http://spark-public.s3.amazonaws.com/nlp/slides/Parsing-Intro.pptx

Then

http://www.cogsci.rpi.edu/courses/**logic**andai/documents/LCL.ppt