Language Technology 1: Summary of my lectures

Günter Neumann

Topics in this lecture that I covered

- Shallow Processing
- Named Entity Extraction
- Information Retrieval
- Question Answering

Shallow Processing

- Full parsing vs. light parsing
- Chunk parsing
 - What are chunks?
 - Difference between chunks and constituency
 - What is chinking?
 - Major benefits of chunks

Chunk Parsing

- Major technologies
 - Regular languages/machines
 - Bottom-up strategies
- Finite state cascades
 - Basic idea ?
 - For what kind of natural languages, chunk parsing is suitable ?
- Fastus:
 - Main steps
 - How are templates expressed/processed ?

Divide-and-conquer Chunk Parsing

- Why is standard bottom-up chunk parsing difficult for German ?
- Divide-and-conquer parsing
 - Main idea
 - Basic components
 - Major advantages

Named Entity Recognition

- What are Named Entities (NE) ?
- Why is NE recognition important for QA systems?
- What is NE-coreference ?

NE and Machine Learning

- Why Machine Learning (ML) for NErecognition ?
- Difference between supervised and unsupervised ML approaches for NE
- What is bootstrapping ?
- What are the major difference between the Collins method and Yangarber method ?

Collins method

- Outline of the learning approach
- What are spelling features?
- What are contextual features ?
- How are (spelling,context) pairs identified and extracted ?
- What are seed rules ?
- Why is it important that spelling features and context features are separated ?

Yangarber method

- What are major difference compared to Collins approach ?
- Where do Generalized Names differ from Proper Names ?
- Outline the learning algorithm NOMEN
- What are the seeds in Yangarber's method ?
- What is the difference between positive and negative seeds ?
- How are patterns generated and applied ?
- How is it guaranteed that NOMEN terminates?

Information Retrieval

- Describe the major architecture of an IR system
- What are the major components of an IR system ?
- What are the major difference between information retrieval and information extraction ?
- What is keyword search and what are the problems/restrictions ?
- What are the difference between a fulltext search engine (text corpus) and a web search engine ?

IR

- What is the Vector Space Model in IR ?
- What is TF/IDF ? Formulas! And examples!
- Similarity measure: Inner product
- What is stemming ?
 - What are the differences between stemming and morphological processing ?
 - What is the porter stemmer ?
- Inverted Index: What is this ?
- What is precision, recall, f-measure ?
 - Definitions, interpretation, examples

Question Answering

- What are the major differences between information retrieval and question answering?
- Describe the general QA architecture
 - Figure and description of main components
- What is open-domain question answering ?
- What is cross-lingual QA?
 - Basic idea
 - Major approaches
- What is MRR ? Definition and examples

Web QA: WAG

- Major control flow
- What are NE-indexes and how they are constructed ?
- How are relevant paragraphs determined ?
- In WAG, question/answer pairs are selected from a Quiz book to evaluate it. What are the major differences for using this style of evaluation compared to TREC/CLEF like evaluation forums?
- In WAG, two test scenarios are distinguished.
 - What are they?
 - And why do we have done so?

Web QA: AskMSR

- Major control flow
- How is query rewriting performed ?
- How are answer candidates extracted from Snippets ?
 - Post-processing of snippets
 - Tiling Answers
- Describe the results they achieved for (check paper)
 - TREC data
 - Web data
 - Why are the Web results better than the TERC results ? Explain!