Efficient Natural Language Processing
Seminar WS 2011 / 2012

Session 1, Wednesday October 26, 2011
(Introduction, Organization, Topic assignment)

Prof. Dr. Hannah Bast

Chair for Algorithms and Data Structures

Department of Computer Science

University of Freiburg

Topic of this seminar   1/2

- **Natural Language Processing (NLP)**
  - Do not just treat text as a string of characters, but try to "understand" it … up to a certain level
- **NLP is needed by lots of applications**
  - In particular, by a new kind of semantic search we are currently developing at our chair   show Broccoli demo
- **But NLP is computationally expensive**
  - Advanced methods take seconds / minutes for a single sentence
  - The English Wikipedia alone has ~ 200 million sentences
  - Even at 1 second / sentence ⇨ 6 years
  - This raises the central (for this seminar) question:

    What level of understanding can we gain in limited time … say 10 – 50 milliseconds / sentence ?

Topic of this seminar   2/2

- Your task in this seminar in some more detail
  - NLP naturally splits into various subfields → Elmar's slides
    POS tagging, Chunking, Parsing, Question Answering, …
  - Various techniques / implementations for each
  - **Task 1:** Get an overview of existing techniques + code
  - **Task 2:** Evaluate performance of some state-of-the-art code
  - **Task 3:** Identify the bottlenecks wrt speed

Organization of the Seminar

- One or two presentations per session
  - Today: introduction of topics + topic assignment
  - After today, we have 12 sessions left
  - Next session is on November 16 (**three** weeks from now)
  - Time schedule for your presentation: next slide

Time schedule for your presentation

- **3 weeks before your presentation**
  - Start collecting material and make a plan of what you want to talk about
- **2 weeks before your presentation**
  - Meet with us (Hannah + Elmar) and present your plan; **please do not waste our time** by coming unprepared
  - In the week that follows, work out all the necessary details and play around (extensively) with software
  - Prepare an outline of your presentation
- **1 week before your presentation**
  - Meet with us again, and present your findings and the outline of your presentation (tentative slides)
  - In the week that follows, finish your work and the presentation

Your presentation

- **Guidelines**

  - You have 30 minutes for your talk + 30 minutes dicussion

  - Use slides in PPT or PDF

  - Your talk will be recorded

Challenges

- **Research**
  - You have to collect yourself interesting and relevant material, we only provide the general topic and give some pointers

- **Understand**
  - Get a decent overview over your topic and understand what you will be talking about

- **Presentation**
  - Present your material in an interesting manner, don't forget that you have an audience

- **Insight**
  - To identify the bottlenecks wrt speed it's not enough to just run the code, you have to understand what goes on inside

Our role

- **We will help you, don't worry**
  - In your first meeting with us (two weeks before your talk) we will help you focus on a good selection of material for your talk
  - We will also give you feedback and advice on the structure and contents of your slides
  - And, of coure, we try to help when you have problems understanding something
  - However, the **initiative** has to come from you !