

Information Extraction

Referat + Hausarbeit

CIS, LMU München
Winter Semester 2022-2023

Prof. Dr. Alexander Fraser, CIS

Information Extraction – Reminder

- Vorlesung
 - Learn the basics of Information Extraction (IE), **Klausur – only on the Vorlesung!**
- Seminar
 - Deeper understanding of IE topics
 - Each student who wants a Schein will have to make a presentation on IE
 - New: 3 (sub-)presentations on a single topic, each are 9 minutes (LaTeX, PowerPoint, Keynote)
 - THIS MAY CHANGE A LITTLE AS I MAKE THE SCHEDULE!
 - If so, I will tell you this next time in the Vorlesung
- Hausarbeit
 - 4 page "Ausarbeitung" (an essay/prose version of the material in the slides), **due 3 weeks after the Referat**
 - **One Hausarbeit per student, submitted separately, per email!**

Why this Seminar (not an Übung)?

- Develop competence in carrying out a literature review, writing and presentation
- Has similarities to the Bachelorarbeit you will do next semester
- Good practice for the Masters, there are many seminars
- Note: Getting a good grade here will be useful for the 2,50 average requirement for the Masters
- Learn by observing what other students do well, but also not so well

Topics

- Topic will be presented in roughly the same order as the related topics are discussed in the Vorlesung
- To understand the topics fully requires you to do a literature search
 - There will usually be one article (or maybe two) which you find is the key source for your presentation
 - For some topics, a suggestion will be made on the slide
 - If the sources you use are not standard peer-reviewed scientific articles, YOU MUST SEND ME AN EMAIL 2 WEEKS BEFORE YOUR REFERAT to ask permission
 - If a paper is behind a paywall, try to use the E-Media service of the LMU library (using your LMU Kennung):
 - <https://www.ub.uni-muenchen.de/e-medien-der-ub/index.html>
- All students will present at least one paper (!)

Referat

- Tentatively (MAY CHANGE!):
 - 3 presentations, each is 9 minutes. 15 minutes for the advisor to ask questions, a few more minutes for discussion
- The first student will present the problem, the motivation and a single paper
 - The first presentation starts with what the overall problem is, and why it is interesting to solve it (motivation!)
 - It is often useful to present an example and refer to it several times
- The second student will present one or two papers on different approaches to the problem
- The third student will present the most recent paper and an analysis (brief comparison of the different approaches) and a conclusion
 - Don't forget to address the disadvantages of the approaches as well as the advantages
 - Be aware that advantages tend to be what the original authors focused on!

Important tips

- **List references and recommend further reading!**
- **Number your slides (useful in discussion)!**
- **The three students working on a single topic need to coordinate! Have one outline clearly indicating where the transitions between students are**
 - **Show this at the start of each of the sub-presentations**
- **IMPORTANT: practice the talk in the group, and give each other feedback to improve the talk**

Language

- If you do the slides in English, then presentation in English (and Hausarbeit in English)
- If you do the slides in German, then presentation in German (and Hausarbeit in German)
- You must specify the presentation language when you specify topics, I will use this in scheduling the topics
- Each set of three topics is in a single language!

References I

- Please use a standard bibliographic format for your references
- This includes authors, date, title, venue, like this:
- Academic Journal
 - Alexander Fraser, Helmut Schmid, Richard Farkas, Renjing Wang, Hinrich Schuetze (2013). Knowledge Sources for Constituent Parsing of German, a Morphologically Rich and Less-Configurational Language. *Computational Linguistics*, 39(1), pages 57-85.
- Academic Conference
 - Alexander Fraser, Marion Weller, Aoife Cahill, Fabienne Cap (2012). Modeling Inflection and Word-Formation in SMT. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 664-674, Avignon, France, April.

References II

- In the Hausarbeit, use ***inline* citations**:
 - "As shown by Fraser et al. (2012), the moon does not consist of cheese"
 - "We build upon previous work (Fraser and Marcu 2007; Fraser et al. 2012) by ..."
 - Sometimes it is also appropriate to include a page number (and you ***must*** include a page number for a quote or graphic)
- Please do not use numbered citations like:
 - DO NOT USE: "As shown by [1], ..."
 - DO NOT USE: footnotes containing the citations
 - Numbered citations are useful to save space, otherwise quite annoying

References III

- If you use graphics (or quotes) from a research paper, MAKE SURE THESE ARE CITED ON THE *SAME SLIDE* IN YOUR PRESENTATION!
 - These should be cited in the Hausarbeit in the caption of the graphic
 - Please include a page number so I can find the graphic quickly
- Web pages should also use a standard bibliographic format, particularly including the date when they were downloaded
- I am not allowing Wikipedia as a primary source
 - I no longer believe that Wikipedia is reliable, for most articles there is simply not enough review (mistakes, PR agencies trying to sell particular ideas anonymously, etc.)
 - Wikipedia can be useful for background, but please don't cite Wikipedia pages!
- You also cannot use student work (not peer-reviewed by people with PhDs) as a primary source
 - If in doubt, email me!

Last Tips

- Please check that all laptops being used can actually project with the **projector in the seminar room**
- Rehearse the talk so that you know it really ends after 9 minutes each. I will cut you off shortly after this time limit!
- PLEASE DO NOT FORGET THE SLIDE NUMBERS!

Schedule

- Today: Seminar + Vorlesung
- Next week: Vorlesung only (Seminar is cancelled)
- Following week: Topics in Seminar + Vorlesung

- Questions?

- Thank you for your attention!