MLT Language technology resources (LT2304), Fall 2012, course paper

In addition to the take-home exam, there is an obligatory course paper assignment for this course, consisting of a written part and an oral presentation. In the course paper, you are expected to treat some aspect of LT resources in somewhat more depth than I expect from an answer to a take-home exam question. Since there is some overlap of the proposed themes for the course paper and the take-home exam questions, you will be required not to choose a take-home exam question which covers the same ground as your course paper.

The course paper should adhere to the normal requirements on scientific writing, but it should be accessible to someone outside the field represented by the paper, i.e., you should strive for a somewhat more general (popular) account. The idea is that the paper should give a brief overview of some aspect of LT resources, similar in scope and ambition to a section in a research paper describing previous relevant research. Scientific writing requires that (non-self-evident) claims be substantiated, normally by references to other scientific works. The text should still be in your own words, however, and direct quotes should make up at the most a small part of the text. You are expected to go beyond the course literature in order to write the course paper.

The course paper is an individual assignment. It should be about 2500 words (approximately 5 pages) in length, plus references and appendices. Format the paper using one of the style files available at <http://www.coling2012-iitb.org/call_for_papers.php> (scroll down beyond the instructions for authors to find the links). The deadline for submitting the written course paper is 22nd October 2011. Please send it by email to me <lars.borin@svenska.gu.se> as a pdf-format attachment by this date (NOTE: no MS Word or LibreOffice/OpenOffice documents, please!). The oral presentations are scheduled for the 30th October 2012. Instructions for the presentations are posted separately.

Below are some proposals for aspects of LT resources that you may want to explore in your course papers. This is by no means an exhaustive list, and you are very welcome to propose your own topic. In all cases, you will need to clear your topic with me before you start working on the paper. The deadline for deciding on your topic is 4th October 2012 at 12.00 noon. Send an email to me <lars.borin@svenska.gu.se>.

- representativeness of LT resources
- LT resources and linguistics vs. LT resources and language technology
- the limits of automated LT resource acquisition
- the pros and cons of LT resource standardization and harmonization
- IPR, reproducibility of research results, and the web as corpus
- LT resources for ‘lesser used languages’/‘low-density languages’
- LT resources in X (X = “e-science” [e.g., “linguistics”, “history”, “political science”], “language learning”, “alternative and augmented communication”, “language checking”, “information access”, “machine translation”, etc.)

---

1 You are not required to write an abstract or specify keywords. The main purpose of using the COLING stylesheet is for specifying how the main text and references should be formatted.