

COMP 791A: Statistical Language Processing

Fall 2006

Outline

Instructor: Leila Kosseim, EV 3-117, kosseim@cse.concordia.ca

Office Hours: Mondays 3:30 to 5pm

Course schedule Mondays, 17:45–20:15, H-619

References

- **Required book** *Foundations of Statistical Natural Language Processing*. Christopher D. Manning and H. Schütze. The MIT Press. 1999.
- **Interesting book** *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Daniel Jurafsky and James H. Martin. Prentice-Hall. 2000.
- Other books and articles will be consulted. Their references will be posted on the course Web site as needed.

Course Objectives The primary objective of this course is to enable you to understand and apply the principles of **robust** methods to natural language processing (NLP) and their applications to manipulate large text collections. The course will offer an introduction to statistical methods used in NLP, will explain the basic theoretical models, and show their practical applications to large collections of texts.

The course will include explanations and discussions of the following topics: introduction to the field, part-of-speech tagging, probabilistic grammars & parsing, N-gram language models & smoothing techniques, word sense disambiguation, information retrieval and text alignment & statistical machine translation.

Web Page Many resources for the course will be available at the URL:
<http://www.cs.concordia.ca/~kosseim/Teaching/COMP791-F06/>.

Evaluation Scheme

1 Assignment	15%
1 Paper presentation	10%
1 Project	30%
2 Exams (2 × 20)	40%
Class participation	5%

Note Although we encourage discussion of topics inside and outside the class, you should be aware of the university's code of conduct as described in the Academic Calendar. In particular, all submitted work must be your own and all references should be properly cited.

Tentative Schedule

<i>Day</i>	<i>Chapter</i>	<i>Material</i>	<i>Special Event</i>
Sept. 11	chap. 1	Introduction to NLP: Overview, history and applications of the field, Zipf's law	
Sept. 18	chap. 3 & 4	Linguistic Essentials & Corpus-Based Work	
Sept. 25	chap. 5	Collocations	Student Presentations
Oct. 2	chap. 6	N-grams Models over Sparse Data	Assignment given Student Presentations
Oct. 9		Thanksgiving - No class	
Oct. 16	chap. 7	Word Sense Disambiguation	Project Description due Student Presentations
Oct. 23	chap. 10	Part-of-Speech Tagging (1)	Exam 1
Oct. 30	chap. 10	Part-of-Speech Tagging (2)	Student Presentations
Nov. 6	chaps. 11&12	Probabilistic Grammars and Parsing	Assignment due Project - Delivery I due Student Presentations
Nov. 13	chaps. 13	Text Alignment and Machine Translation	Student Presentations
Nov. 20	sects. 15.1-15.2	Information Retrieval	Student Presentations
Nov. 27		Catch-Up	Exam 2
Dec. 4		Topic of your choice	Student Presentations Project - Final Delivery due