NLP

Introduction to NLP

Administrative

Structure of the Course

- Four major parts:
    - Linguistic, mathematical, and computational background
    - Computational models of morphology, syntax, semantics, discourse, pragmatics
    - Core NLP technology: parsing, part of speech tagging, text generation, etc.
    - Applications: text classification, machine translation, information extraction, etc.
- Three major goals:
    - Learn the basic principles and theoretical issues underlying natural language processing
    - Learn techniques and tools used to develop practical, robust systems that can understand text and communicate with users in one or more languages
    - Gain insight into some open research problems in natural language

Available Books

- **Speech and Language Processing**
  - Daniel Jurafsky and James Martin
  - http://www.cs.colorado.edu/~martin/slp.html

- **Foundations of Statistical Natural Language Processing**
  - Chris Manning and Hinrich Schütze
  - http://nlp.stanford.edu/fsnlp/

- **Natural Language Understanding**
  - James Allen

Courses at Other Places
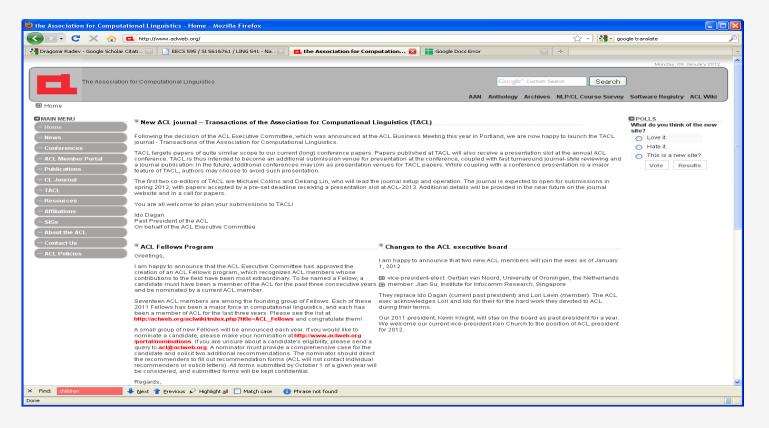
- Brick–and–Mortar
  - Johns Hopkins University (Jason Eisner)
  - Cornell University (Lillian Lee)
  - Stanford University (Chris Manning)
  - U. Maryland (Hal Daumé)
  - Berkeley (Dan Klein)
  - U. Texas (Ray Mooney)
- Coursera
  - Manning/Jurafsky (2012, survey)
  - Michael Collins (2013, more advanced)

The Association for Computational Linguistics (ACL)

The Alphabet Soup

- NLP (Natural Language Processing)
- CL (Computational Linguistics)
- IR (Information Retrieval)
- SP (Speech Processing)
- HLT (Human Language Technology)
- NLE (Natural Language Engineering)
- ML (Machine Learning)

Research in NLP

- <u>Conferences</u>:
  - ACL/NAACL, EMNLP, SIGIR, AAAI/IJCAI, Coling, HLT, EACL/NAACL, AMTA/MT Summit, ICSLP/Eurospeech
- <u>Journals</u>:
  - Computational Linguistics, TACL, Natural Language Engineering, Information Retrieval, Information Processing and Management, ACM Transactions on Information Systems, ACM TALIP, ACM TSLP
- <u>University centers</u>:
  - Berkeley, Columbia, Stanford, CMU, JHU, Brown, UMass, MIT, UPenn, USC/ISI, Illinois, Michigan, UW, Maryland, etc.
  - Toronto, Edinburgh, Cambridge, Sheffield, Saarland, Trento, Prague, QCRI, NUS, and many others
- <u>Industrial research sites</u>:
  - Google, MSR, Yahoo!, IBM, SRI, BBN, MITRE, AT&T Labs
- The ACL Anthology
  - http://www.aclweb.org/anthology
- The ACL Anthology Network (AAN)
  - http://clair.eecs.umich.edu/aan/index.php

NLP