595: Natural Language Processing

Syllabus

Instructor
Rada Mihalcea, BBB 3769, mihalcea@umich.edu

GSI/IAs
RJ Duquette (duquetr@umich.edu)
Yiting Shen (yiting@umich.edu)
Harry Zhang (zharry@umich.edu)

Meeting times
Wed 1:30-4:30pm, DOW 1013

Office hours
Instructor: Tue 3-4:00pm, Beyster 3769
GSI/IAs: Mon 1:30-2:30pm, Beyster 1695 (Yiting)
         Wed 11:00-12:30pm, Beyster 1695 (RJ)
         Thu 4:00-6:00pm, Beyster 1695 (Yiting)
         Fri 11:30-1:00pm, Beyster 1695 (Harry)

Course description
This course provides an introduction to the theory and practice of natural language processing (NLP) - the creation of computer programs that can understand, generate, and learn natural language. The course will cover the three major subfields of NLP: syntax, semantics, and pragmatics. The course will introduce both knowledge-based and statistical approaches to NLP, illustrate the use of NLP techniques and tools in a variety of application areas, and provide insight into many open research problems.

Website
Most course materials are made available on Canvas, and are considered required reading. A detailed schedule, including lecture topics, assignment due dates and exam dates, will also be available on Canvas.

Forum
We will be using Piazza to host a course forum. You are required to read this forum regularly; it is the venue we will use for important course announcements and assignment clarifications. In addition, it will be a significant source of help and hints on the assignments. We do not answer technical questions via email. In order to save everyone time, we want all students to have the benefit of seeing each question and its answer, so please use the forum. Please do not post your own solutions, project code, test cases, or output to the forum. Also, please search the forum before posting to avoid questions that have already been asked and answered.
Contact
Please direct all technical questions to our Piazza forums. For other questions, you can reach us through private posts on Piazza, or by email.

List of Topics

<table>
<thead>
<tr>
<th>Brief overview of English linguistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text preprocessing</td>
</tr>
<tr>
<td>Language models</td>
</tr>
<tr>
<td>Word embeddings</td>
</tr>
<tr>
<td>Deep learning for NLP</td>
</tr>
<tr>
<td>Part of speech tagging</td>
</tr>
<tr>
<td>Syntactic parsing</td>
</tr>
<tr>
<td>Lexical semantics</td>
</tr>
<tr>
<td>Word sense disambiguation</td>
</tr>
<tr>
<td>Subjectivity and sentiment analysis</td>
</tr>
<tr>
<td>Information extraction</td>
</tr>
<tr>
<td>Automatic summarization</td>
</tr>
<tr>
<td>Question answering</td>
</tr>
<tr>
<td>Machine translation</td>
</tr>
<tr>
<td>Natural language processing for social sciences</td>
</tr>
<tr>
<td>Other special topics</td>
</tr>
</tbody>
</table>

Required/recommended readings
(required) Speech and Language Processing, by D. Jurafsky and R. Martin (2nd edition)
(recommended) Foundations of Statistical Natural Language Processing, by C. Manning and H. Schutze
Additional readings will be assigned throughout the semester.

Programming Environment
For this course, you must have a CAEN account. Anyone enrolled in EECS 595/498 is eligible for one, but it takes a little while to get it set up. You will have to write your programs in Python. You are free to develop your programs on any platform you like, but they must run on
the CAEN platform. We will grade your programs in the CAEN Linux environment and they must run correctly in this environment.

**Course work and grading**

Programming assignments: 30%
Exam I (Oct. 10): 25%
Exam II (Nov. 28): 25%
Project: 20%

There will be four programming assignments throughout the semester. Each assignment will specify the material to be turned in. Assignments are due by 11:59pm on the due date. Assignments may be turned in up to 3 days late, with a penalty of 10% for each day late. No credit will be given after 3 days. No exceptions.

**Accommodations for Students with Disabilities**

If you think you need an accommodation for a disability, please let your instructor know at the beginning of the semester or at least 2 weeks in advance. As soon as you make us aware of your needs, we can work with the Services for Students with Disabilities (SSD) office to help us determine appropriate academic accommodations. SSD (734-763-3000; http://ssd.umich.edu) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide is private and confidential and will be treated as such.