

## INFO/CS 4300: Language and Information, Spring 2015

### Course Syllabus

- **Time and place:** MW 7:30pm-8:45pm, Upson 215
- **Instructor:** [Prof. Cristian Danescu-Niculescu-Mizil](#)
- **Teaching Assistant:** [Vlad Niculae](#)
- **Course Piazza page:** [piazza.com/cornell/spring2015/cs4300info4300](https://piazza.com/cornell/spring2015/cs4300info4300)
- **Course homepage:** [www.cs.cornell.edu/Courses/cs4300/2015sp/](http://www.cs.cornell.edu/Courses/cs4300/2015sp/)
- **Summary:** How to make sense of the vast amounts of information available on-line, and how to relate it and to the social context in which it appears? This course introduces basic tools for retrieving and analyzing unstructured textual information from the web and social media. Applications include information retrieval, question answering, sentiment analysis and social analysis of text. The coursework will include small projects that play on the interaction between knowledge and social factors.
- **Prerequisites:**
  - INFO 2950 (or CS 2800 and a linear algebra course)
  - INFO 3300 (or a machine learning course)
  - Good Python programming skills and familiarity with iPython notebooks, of which we'll make extensive use.
- **Related courses offered this semester at Cornell:**
  - [CS 4740/5740 Natural Language Processing](#)
  - [CS 4786/5786 Machine Learning for Data Science](#)
  - [CS 4850 Mathematical Foundations for the Information Age](#)

#### Academic Integrity

We will strictly follow Cornell University's policies on academic integrity as outlined in the [Academic Integrity Handbook](#).

Any work submitted by a student in this course for academic credit will be the student's own work. For this course, collaboration is allowed only when it is made explicit in the assignment or project description. In case of doubt, contact the instructor.

### Late submissions and attendance

Late submissions will not be accepted, save for major medical or family events. Attendance is mandatory, as for most lectures there will be no lecture slides.

### Electronic device policy

Notes for this class should be taken on paper. Use of electronic devices such as laptops and tablets will not be permitted during class (with the exception of specific activities). We are not plain evil, we are just following extensive research on the negative effects of in-class laptop use on learning.

### Grading

Grades will be based on:

- participation (in-class or on-line) [10%];
- assignments/homeworks/quizzes [40%];
- midterm project + quiz [20%];
- open ended final project [30%];

### SONA Credits

You can get extra credits for participating in experiments and research studies through [Science Research Participation System](#). You will receive 0.5% extra credit for each 30 minute study (or equivalent), up to a maximum of 2%.

### Textbooks

- Manning, Raghavan, and Schütze. 2008. Introduction to Information Retrieval. Cambridge University Press.
- Jurafsky and Martin. 2009. Speech and Language Processing (2nd Edition). Pearson.

### Course outline

This being the first offering of this version of the course, the schedule and list of topics will be in **constant flux**. Here is a tentative outline:

---

Week	Content
1 W	Intro, Dimensions of information systems, Conversational behavior
2	Types and tokens, Document similarity
3	Boolean retrieval
4	Ranked retrieval, Vector space models

---

Week	Content
5 W	Text classification
6	Review helpfulness, Opinion Mining
7	Project proposals
8	Question answering
9	Project proposals
10	Style, Psycholinguistics
11	Spring Break
12	Trust and credibility
13	Pragmatics, Politeness
14	Final project presentations
15	TBD

---