

Spring Semester 2017
Iowa State University

LING 410X - Language as Data

Course Handbook

Instructor: Sowmya Vajjala

- *Office:* 331 Ross Hall
- *Email:* sowmya@iastate.edu

Course Objectives: This course aims to introduce students to methods of discovering language patterns in text documents and applying them to solve practical text analysis problems in their disciplines. Data of any form (text, numbers, images etc.) is available in large amounts now like never before. Text is one of the major forms of big data and hence text analysis is in huge demand in the information technology industry now. Apart from the technological applications, it is also useful in various disciplines like business intelligence, sociology, psychology and literature to name a few. For example, key word extraction and sentiment analysis are very useful in Business analytics, authorship detection and stylometric analyses are examples applications for literature, studying mental disorders through patient written samples is gaining prominence in clinical psychology. In this background, this course introduces some commonly used methods to work with textual data. After a brief primer in the fundamentals of linguistics and its role in text analysis, the course will introduce the students to writing R scripts (as it is easier to do exploratory analysis and visualization in R without learning a lot of programming principles) to perform text analysis and visualize textual data.

Learning Outcomes After finishing this course, students will know:

1. some common methods for performing automatic text analysis
2. some real-life applications of text analysis
3. how to apply these methods to solve text analysis problems in their domain areas
4. how to visualize textual data using various tools and methods

Pre-requisites: Junior Standing. LING 120 is a preferred but not a mandatory pre-requisite.

Textbooks The primary textbook is: "Text analysis with R for students of literature" by M.J.Jockers. The course will also rely on a wide range of freely accessible online tutorials and videos related to various methods of text analysis. (example: <https://github.com/kbenoit/ITAUR-Short>).

Syllabus - topics covered

1. Introduction
Text analysis, real world applications, usefulness for various disciplines
2. Introduction to Linguistics and the role of linguistic knowledge in solving text analysis problems, with examples
3. Installing R and working with it.
4. Corpus preparation: methods to select, process and clean corpora
5. Keyword and Key-phrase extraction methods
6. Topic modeling and its applications
7. text classification methods and their application for sentiment detection
8. methods of visualizing textual information