

Introduction to Text Analytics

CS 5970

Spring 2016 Syllabus

Instructor: Dr. Christan Grant

Email: cgrant@ou.edu

Office: DEH 234

Office Hours: [Google Hangout](#) on

Every **Tues** from 12:00-1:00 PM CST

Every **Thurs** from 5:00-6:00 PM CST

or by appointment

Teaching Assistant: Supreetha Akella

Email: supreetha.akella@ou.edu

TA Office Hours: by appointment

Note: When emailing the professor or TA *always* include **CS5970** in the subject line.

Prerequisites

Students must have a working understanding of statistics and data structures, in addition to a select set of software skills. The prerequisite courses/skills are listed here:

- **Statistics:** MATH 4743 or MATH 4753 or ISE 3293 or ISE 5013
- **Data Structures/Discrete Structures:** CS 5005 or CS 2413 and CS 2813
- **Software skills:** Students should be well versed in Java and/or C++ and also familiar with at least one scripting language such as Python. Students should also be comfortable working with the Linux command line.

Undergraduate students with a 3.5 GPA or higher may enroll with permission from the instructor.

Course Description

Text Analytics are the methods and techniques used to extract useful knowledge from text to support decision making. This field includes a collection of research from the natural language processing, databases, data mining, and machine learning communities. The aim of this course is to be a primer for text analytics theory and practice. After taking this course, students will have an understanding of how to independently obtain, parse, and analyze textual information for organizations that want to extract valuable insights.

Topics discussed in the course include: obtaining data sets, understanding data formats, duplicate detection, cleaning data sets, tagging, indexing and search, evaluating algorithms, classification, clustering, topic modeling and entity resolution. Time permitting we may discuss advanced topic such as relation extraction, slot filling, knowledge graphs, knowledge base construction, the semantic web, question answering or other cutting edge topics.

Learning Management System

All course resources, activities and projects will be hosted on www.janux.ou.edu

Course Materials

Required Materials

Croft, W. Bruce, Donald Metzler, and Trevor Strohman. *Search engines: Information retrieval in practice*. Reading: Addison-Wesley, 2010.

CHAPTERS 3.7, 4, 5, 7.1, 8.4, 9.1, 9.2

Manning, Christopher D., and Hinrich Schütze. *Foundations of statistical natural language processing*. MIT press, 1999.

CHAPTERS 1, 3, 4, 5, 8.1, 14, 15.1.2

Supplementary Materials

Additional materials will be provided to you on the Janux platform.

Grading Policy

All coursework must be uploaded to the Janux platform by the due date and time indicated on the assignment unless otherwise announced. Ensure your familiarity with the Janux system in advance, and allot a sufficient amount of time for submission and any technical difficulties that may arise. For late assignments, 10% will be deducted from the grade every 24 hours after the official due date/time. All grades are final one week after grading. A specific naming convention will be required for submissions, where X is the assignment/project title: **X_lastname_firstname.pdf**.

Percentages of course grading requirements are as follows:

	Percentages
Quizzes	25
Assignments	25
Course Projects	35
Participation	15
Grading Scale: A=[90, Inf), B=[80,90), C=[70,80), D=[60,70), F=[0,60)	

Quizzes

Quizzes will be assigned regularly based on each course module to evaluate student progress on each learning objective. The exact point scale of each activity will be available through the JANUX platform.

Assignments

The assignments will consist of a variety of written and programming problems. All assignments must be uploaded to the Janux platform by the due date and time provided on the assignment. For each assignment please read the submission requirements outlined in the instructions.

Course Projects

The course projects will be developed over the course of the semester. Information such as instructions, guidelines and due dates will be posted on the Janux platform. Be sure to check the "Project" page online early and often. Make sure you are aware of the deadlines and requirements for each project.

Participation

Since this is a fully online class, you are expected to view all videos and complete all reading assignments and practice activities. Also, you will be expected to participate fully in online discussions by posting initial responses *by Fridays at noon* and replying to other students' posts *by the following Monday at noon*. Participation is evaluated and counts toward your final grade.

Communication

Janux is a social platform that enables peer-to-peer communication. We encourage you to interact with your classmates on course relevant topics. The instructor and teaching assistant will also participate in online discussions. In addition, the instructor will hold regular office hours using the Google Hangouts video feature, which you can access through any Gmail account. The Hangout can be accessed during office hours: <https://bit.ly/1ZiLbEH>

Extra content or examples may be provided through additional videos posted to YouTube.

Note: When emailing the professor or TA *always* include **CS5970 in the subject line.**

Reasonable Accommodation

Any student in this course who has a disability that may prevent him or her from fully demonstrating his or her abilities should contact me personally as soon as possible, so that we can discuss accommodations necessary to ensure full participation and facilitate your educational opportunities.

Adjustment for Pregnancy/Childbirth Related Issues

Should you need modifications or adjustments to your course requirements because of documented pregnancy-related or child birth related issues, please contact me as soon as possible to discuss. Generally, modifications will be made where medically necessary and similar in scope to accommodations based temporary disability. Please see <http://www.ou.edu/eoo/faqs/pregnancy-faqs.html> for commonly asked questions.

Title IX Resources

For any concerns regarding gender-based discrimination, sexual harassment, sexual misconduct, stalking, or intimate partner violence, the University offers a variety of resources, including advocates on-call 24.7, counseling services, mutual no contact orders, scheduling adjustments and disciplinary sanctions against the perpetrator. Please contact the Sexual Misconduct Office 405-325-2215 (8-5) or the Sexual Assault Response Team 405-615-0013 (24.7) to learn more or to report an incident.

Technical Support

The instructor and teaching assistant will not be able to help with issues related to the Janux platform. Requests for platform assistance should be directed to the Next Thought Help Center at www.janux.ou.edu. For OU IT support, please phone (405) 325-HELP.

Week	Unit	Topic	Notes	Due
19-Jan	0.	Course Orientation Review and Remediation		
25-Jan	1.	Understanding Text		
1-Feb	2.	Data Formats		
8-Feb	3.	Cleaning Data Sets		
15-Feb	4.	Duplicate Detection		
22-Feb	5.	Tagging Text		
29-Feb	6.	Indexing and Search		
7-Mar	7.	Evaluating Algorithms		
14-Mar		Spring Break	Spring Break	
21-Mar	0.	PROJECT ONE	Project One Specs released on 21-Mar	
28-Mar	8.	Classification	Continue working on Project One	
4-Apr	9.	Clustering		Project One due 4-Apr
11-Apr	10.	Entity Resolution		
18-Apr	11. 0.	Research: Advanced Topics FINAL PROJECT	Adv. Topics Specs released 18-Apr Final Project Specs released 18-Apr	
25-Apr	11. 0.	Presentation: Advanced Topics FINAL PROJECT	Final Project	*Final Project Checkpoint 1 due 25-Apr
2-May	11. 0.	Response/Critique: Advanced Topics FINAL PROJECT	Continue working on Final Project	* Adv. Topic Presentation due 2-May
9-May		Advanced Topic Response/Critique Due FINAL PROJECT DUE		*Adv. Topic Response due 9-May *Final Project due 9-May

* **Content:** New units will open on Mondays at 12:01 pm.

** **Discussion boards:** Initial posts are due on Fridays by 12:00 pm (noon).

Responses/replies are due before the following Monday by 12:00pm (noon)

*****Quizzes/Assignments:** Due dates will be listed on the Janux platform.

This syllabus is subject to change. Students are responsible for any changes/additions to this syllabus announced during the semester.