

Course Description

How can digital techniques enhance our textual interpretations? How can we make sense of an increasing number of textual sources in a timely manner? What kind of new interpretative questions can be raised and answered by computer-based text analysis?

In this course, we will apply a set of text analysis techniques to various liberal arts disciplines in order to gain insights on how these tools are used in their specific areas of interest. Examples will include using distant reading on vast literary corpora, investigating authorship of Supreme Court opinions, retracing the history of mobile devices by “reading” 30 years of newspaper, and interpreting Raphael's masterpiece “School of Athens”.

No previous computer programming experience is required. Parallel labs will be provided for leveraging programming skills.

Course Goals

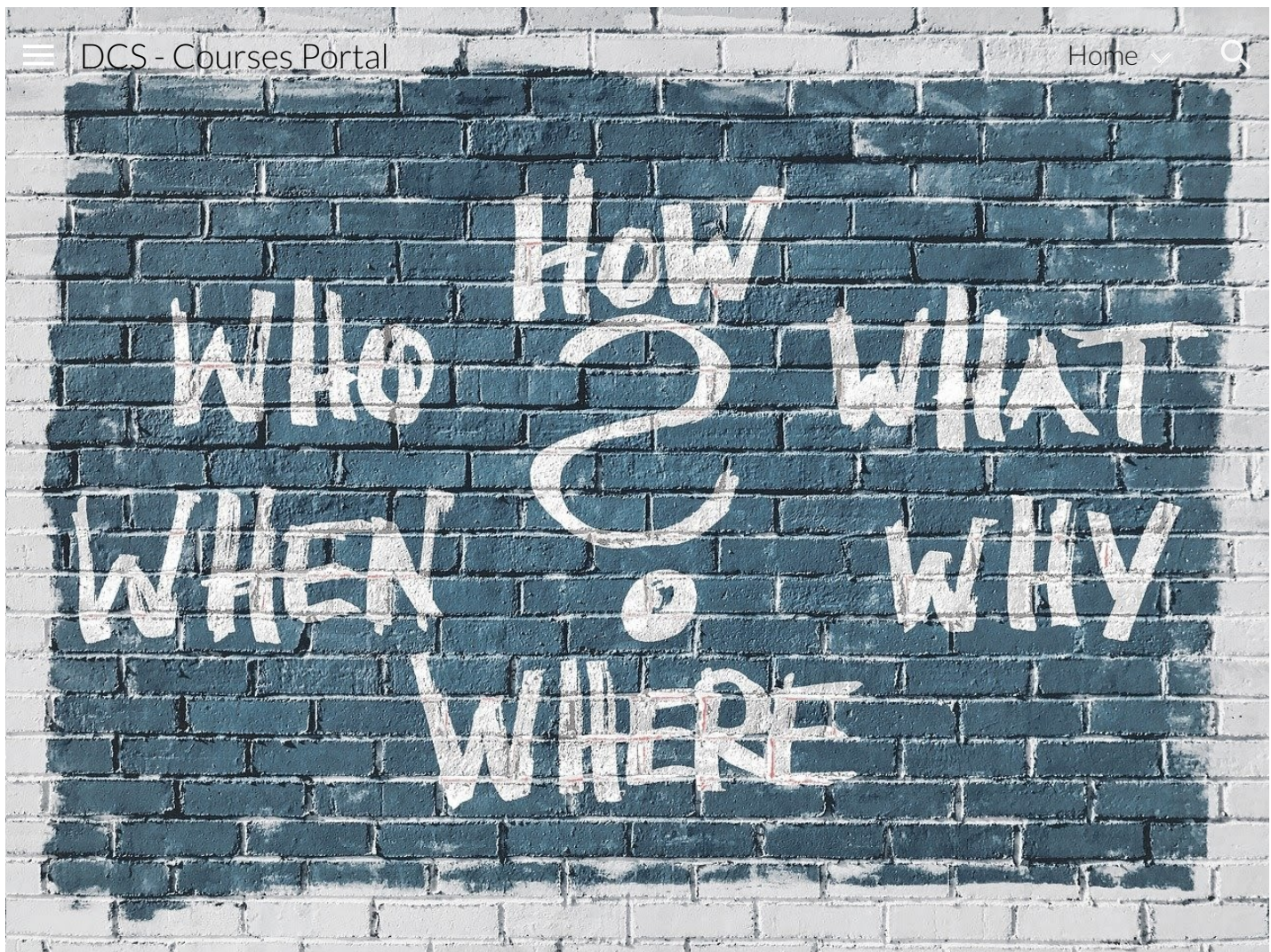
- Acquire and apply a set of techniques to organize, analyze, interpret and present arguments based on digital textual analysis
- Understand how digital text analysis techniques can be integrated into several liberal arts disciplines
- Critically evaluate shortfalls and limitations of each technique
- Extract quantified features of texts in order to augment explanatory capabilities of textual based inquiries



Unit 1 - Process and Methods



[Unit 2 - Clean up and Basic Metrics](#)



Unit 3 - Parsing and Basic NLP



Unit 4 - Clustering and Classifying



[Research Design Form](#)

General Schedule

Sep 13 - 24 - Work on Research Design and fill out the form.

Sep 27- Oct 7 - First Project Meeting to discuss the research design. Please, submit the form one day before your scheduled meeting.

Oct 13 - 22 - Work on Data Gathering and fill out the form.

Oct 18 - 28 - Second Project Meeting to evaluate the project input data. Please, submit the form one day before your scheduled meeting.

Pedagogical Dimensions

1st Dimension - Liberal Arts Context

- a.) Politics and Rhetoric - How inaugural presidential discourses can be analysed through basic digital techniques?
- b.) Literary Studies - Analyze central motifs of literary works, e.g., Sherlock Holmes, Around the World in 80 days, Harry Potter
- c.) History of Philosophy - Assess the dominant interpretation of Raphael's School of Athens.

How Plato and Aristotle works can be compared in regards to their metaphysical and epistemological positions? (Clustering)
- d.) Foreign Language Literary Studies - A critique of clustering techniques applied to Galileo's texts
- e.) Geography - Assessing the drugs quest in the U.S through text analysis of congressional hearings - Topic Modeling
- f.) Government and Legal Studies - Investigate the authorship of Bush vs Gore Supreme Court case - Classification/Machine Learning.

2nd Dimension - Text Analysis phases

1. Corpus Gathering
2. Corpus cleaning and normalization
3. Basic textual metrics
4. Introduction to Natural Language Processing
5. Documents Clustering
6. Documents Classification

Assignments and Evaluations

30% Formative Assessments: Low stakes formative activities that bridge readings/notebooks and in-class discussions.

15% 1st Midterm Exam (Individual)- It will cover theoretical and practical content of the first part of the course. This exam will take place around the last class before the Fall Break.

15% 2nd Midterm Exam (Individual)- It will cover theoretical and practical contents. This exam will take place on the last day of classes.

40% Final Project (Groups of 2 or 3 students)- You will be writing a digital analysis and interpretation that investigates an aspect of your major or minor field of study (or other interests you may have). The grading criteria breakdown is as follows:

10% Corpus Gathering and Preparation

15%: Context and problem definitions

35%: Digital Text Analysis techniques and theoretical concepts applied to the project

25%: Integration of digital analysis to the interpretation

15% Final Presentation

Project Timeline:

Initial Corpus Definition - Meeting during office hours until September 24th

Context and Possible problems to be explored - Submitted by the 1st Midterm exam.

Digital Essay Draft - TBD

Final written submission - TBD

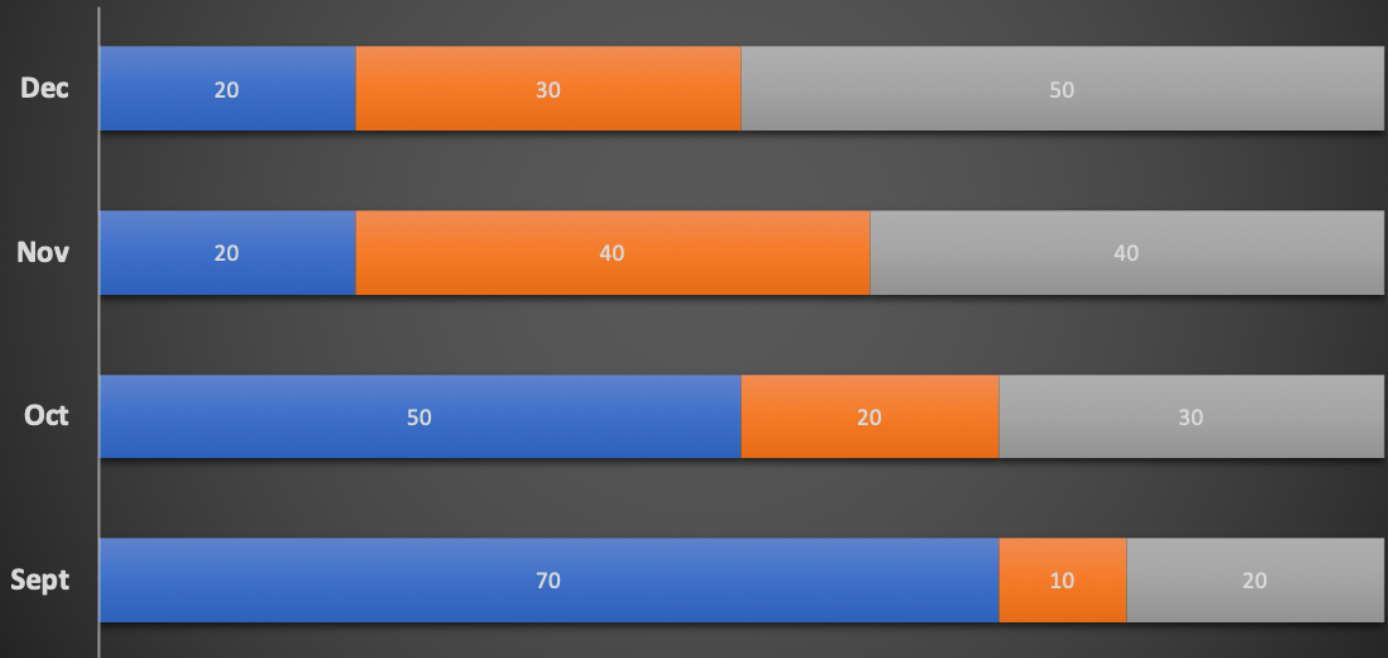
Final project submission - December 15th



Projected Course Workload (by percentage of dedicated time)



Readings Jupyter Notebooks Final Project



Terms and Conditions

Collaboration

One of the principal components of a DCS course is collaboration. However, you should always be clear on what part of the work you hand in is your own, what parts come from other sources, and what parts are collaborative. As a rule of thumb, we distinguish between interacting with another student using any written medium (e.g. pencil and paper, email, looking at their code) and having broad discussions with them. Unless you work with another student in a group, you are not allowed to exchange information through a written medium with him/her or provide answers to problem sets through conversation. This is a zero-tolerance policy.

It is permissible to use materials available from other sources such as the Internet (understanding that you get no credit for using the work of others) as long as: 1) You acknowledge explicitly which aspects of your assignment were taken from other sources and what those sources are. 2) The materials are freely and legally available. 3) The material was not created by a student at Bowdoin as part of this or another course this year or in prior years. To be absolutely clear, if you turn in someone else's work you will not receive credit for it; on the other hand, if you acknowledge it, at least you will not violate the Honor Code. All write-ups, reviews, documentation and other written material must be original and may not be derived from other sources.

Academic Honesty

We assume that every student is abiding by the Code of Academic and Social Conduct to which they agreed upon matriculation at Bowdoin: <http://www.bowdoin.edu/studentaffairs/student-handbook/college-policies/> If you find yourself overwhelmed, unsure about what to do, without materials to use, or otherwise feeling like you have no choice but copying from the web and calling it your own work, contact Professor Nascimento before you make a very bad decision. DO NOT jeopardize your academic career for an assignment in this course. Talk to us first, because if we suspect any kind of cheating or plagiarism we will pursue the matter to the fullest extent allowable by Bowdoin policy.

Religious Holidays



No student is required to take an examination or fulfill any other scheduled course requirement on

“My computer died” and “I only saved to the lab laptop” are not valid excuses for having nothing to show for your work on the day an assignment is due. We are going to discuss and practice responsible management of electronic files so that you are never caught without a very recent copy of your very important work. Having nothing to show on the due date will be a sign that you started the assignment very late.

Accessibility

Bowdoin College is committed to ensuring access to learning opportunities for all students. Students seeking accommodations based on disabilities must register with the Student Accessibility Office. Please discuss any special needs or accommodations with me at the beginning of the semester or as soon as you become aware of your needs; I am eager to work with you to ensure that your approved accommodations are appropriately implemented.