Papyrus to Pixels, spring '24: Syllabus

Course syllabus

All course material (including an on-line version of this syllabus) is available on the course web site: https://neelsmith.github.io/papyrus_to_pixels/.

Embedded in the text of this syllabus are sentences identifying a two-word phrase summarizing one theme of the course. Please note that as part of your class preparation for Thursday, Jan, 25, you will be required to add a file with this phrase to your personal folder on the course Google drive.

Overview

"Papyrus to Pixels" is being offered in collaboration with Biology 199, "Change Through Time." Throughout the semester, our Tuesday-Thursday classes will include a mix of joint meetings of the two courses (held in O'Neil 101) and separate meetings (in Fenwick 420). Together, the two courses will explore shared ideas about how the material that biologists and textual scholars study changes or evolves. Both Charles Darwin's understanding of evolution and nineteenth-century scholars studying the history of texts shared the idea of a "family tree," for exmaple, and viewed the variation that appeared over time as a series of branches, each forking from a single source. Today biologists and textual scholars are exploring more complex relationships. The genealogies and descent of both biological species and of texts show networks of connections not limited to a rigid tree structure.

"Papyrus to Pixels" will introduce some standard approaches to modeling and comparing digital corpora of texts, and consider how we can apply methods from disciplines like computational biology to reconstruct and interpret their transmission. Students will gain practical experience using computational methods to address important questions for anyone interested in understanding a digital corpus. How do we identify literary style? How do we isolate texts or sections of texts with particular subject matter? How do we recognize relations among many texts, or texts in different languages?

Contact information

Instructor: Neel Smith. **Email:** at `holycross.edu`, user name `nsmith`

Regular class meeting times: TTh 2:00-3:15. Check the course schedule to see if we are meeting in O'Neil 101 or Fenwick 420.

Regular drop-in hours (S24): Wednesday 10:00-12:00, Thursday, 10:30-12:00, Friday 10:00-12:00, in Fenwick 415; or available anytime by appointment.

Manuscripts, Inscriptions and Documents Club: Friday, 2:00 pm - ?, fourth floor of Fenwick Hall.

Please wear masks when we meet in Fenwick 420, and in my office. Masks are optional when we meet in O'Neil 101.

On-line course material and technical requisites

On-line resources

1. On-line course material is available on the course web site. You will also have access to a course Google drive, where we may post copies of copyrighted reading material.
2. You will have access to a personal folder on the course Google drive. You will submit assignments by dropping a file into your individual folder. (Other students do not have access to your individual folder.)
3. When you use your Holy Cross Google login, you can have free access to a specialized version of ChatGPT trained to generate code in the Julia language, called AskAI. AskAI is available online here.

Any supplementary hard-copy material will be distributed in class. There is no textbook, and no Canvas site for CLAS 199-S05.

Technical requisites

As part of your lab assignments, you will learn how to write code in the Julia language. You will want to install Julia on your personal computer (you can follow directions on the course web site), or speak with Prof. Smith about access to the Classics Department research lab, where machines have Julia installed.

Goals

Shared goals with both sections:

- work collaboratively

- cultivate habits of test-driven thinking:

    - hypotheses are wrong until proven otherwise
    - to know if a test works, you must first fail it. That's why the first word of our embedded theme is "fail."
    - iteratively improve

- develop a reproducible research project from an initial question to implementation, including

    - explicit license for reuse

- – source material and analytical methods identified
- · oral presentation
- · written presentation: one source, multiple formats

Specific objectives in CLAS 199-S05

- · read texts from URLs or local files either as a single string or as a series of sections (chapters in a book, lines in a poem)
- · tokenize a text and construct frequency histograms
- · plot histograms and analyze frequencies in relation to Zipf's Law
- · extract features like named entities from a corpus
- · identify significant terms with metrics like TF-IDF
- · select an appropriate data model for features in Julia, and organize feature data as graphs, matrices, and dictionaries

What to expect

There are no formal prerequisites for this course: it is intended to be accessible to any Holy Cross student. In class work, and especially in an extended course project, you will have opportunities to contribute to the course by drawing on your own experiences and interests.

The course emphasizes hands-on work, and you are expected to be able to use a computer to follow this basic sequence of tasks on your personal computer, or the computer you are using in a lab:

1. Given a link to a text file on the internet, download the file and save it to the computer.
2. Find the saved file on the computer, and open it with a text editor.

Before our first hands-on class session on Thursday, Jan. 25, please make sure that you are comfortable doing this. If you have questions or need help, you can get help from Educational Design and Digital Media Services.

Tips

Expect to be wrong. You will try things in this course you have never tried before; you will not always succeed on your first try. That's great! Every time you're wrong is an opportunity to learn something, and that's our highest course goal.

Test to find out if you're right. Make a habit of asking ahead of time how you will determine if your solution to a problem is correct. Create a test that you can apply to your work. Apply the test *before* you attempt to solve a problem, and *fail* the test. *Then* try to solve the problem. If your solution passes your test, you can be confident your solution is valid.

Rewrite. Writing well always requires rewriting. This is true of your prose, and of your code. While we have to set practical deadlines when we stop rewriting, you should never expect to write something once. That's why the second word of our embedded course theme is "forward."

Fail early. Give yourself time to be wrong. You won't be able to learn from our failures and rewrite your work if you are trying something for the first time just before a deadline.

Break problems down into smaller, simpler pieces. It is sometimes said that problem decomposition is the essence of computer science, but the ability to analyze complex problems in terms of smaller parts is central to scholarly thinking more generally. For example, if you wanted to compare the rhetorical style of Abraham Lincoln's Gettysburg address with the keynote speech by Edward Everett at the Gettysburg dedication (as you will in your first lab exercise this semester), you would need to break down the idea of style into specific features. If one feature was length of sentences, you would need to break that down in to the problems of isolating sentences, and measuring their length.

Course requirements and grades

You determine your own course grade by satisfactorily completing a number of assignments in these categories:

1. four lab assignments, completed in small groups
2. a contribution to a collection of course resources
3. a longer research project completed in multiple stages
4. class attendance

All assignments will be graded satisfactory/unsatisfactory according to an explicit specification of requirements: if you complete *all* requirements, the assignment is satisfactory. If you complete an assignment and submit it on time, but it fails to satisfy all requirements, you will have opportunities to revise the assignment to receive full credit for it.

The deadline for all written submissions (such as lab notebooks) is noon on the day of the due date.

Small-group lab assignments

In the first part of the course, you will complete four lab exercises that you will work on in teams. Each exercise is worth two points. You earn one point if the *completed* assignment is submitted by the due date. You earn a second point if *all* the specifications for the assignment are satisfied by the final deadline for revisions.

The total points you can earn in this category is 8.

Contribution to a collection of course resources

In the course of developing a longer research project, you will share some part of your experience with the rest of the class. You can earn one point for successfully proposing an idea to share, and a second point for submitting a satisfactory contribution.

The total points you can earn in this category is 2.

Longer research project

In the last part of the course, you will develop an original research project in five steps. Each step is worth two points, calculated just like the previous lab assignments: you earn one point if the step is submitted by the due date, and a second point if all specifications for that step are satisfied by the final deadline for revisions.

Attendance

Being prepared and attending class is an essential part of the learning process. We have 27 scheduled class meetings. Occasionally, your class preparation instructions on the daily schedule will require you to submit some preparatory work prior to class. To receive credit for a class meeting, you must submit any required preparatory work on time and attend class. (See the section of the syllabus on "Policies" for guidance on excused absences.) You may therefore receive credit for up to 27 class meetings, which earn you up to four points in the class attendance category.

The following table summarizes how your class attendance record translates to points for your course grade.

Class meetings attended	Points earned
26+	4
25	3
24	2
23	1
< 23	0

The total points you can earn in this category is 4.

Determination of final course grade

Your course grade will be recorded by reading the following table. Each column records points in one of the five grade categories. In each row, the numbers represent the *minimum* number of points required in that category for the grade in that row. For example, if you get 0 points in the category "Resource" (contribution to a collection of course resources), the highest grade you could earn is a C-. If you earn the maximum of 8 points for lab assignments, 10 points for a course project, and 2 points for contribution to course resources, but only 2 points for attendance, your course grade is a B-.

Course grade	Lab assignments (8)	Project (10)	Resource (2)	Attendance (4)
A	8	10	2	4
A-	7	10	2	4
B+	7	9	2	3
B	6	8	2	3
B-	6	8	2	2

Course grade	Lab assignments (8)	Project (10)	Resource (2)	Attendance (4)
C+	6	7	1	2
C	5	7	1	1
C-	4	6	0	1
D	3	6	0	1

Policies

Class attendance

Work in class is an essential component of the course, and regular attendance is required. (See the course grading rubric in this syllabus for how class attendance contributes to your final course grade.) Excused absences do not count against your course grade. If you know of scheduled conflicts with class meeting times, please speak with me as soon as possible to make arrangements ahead of time for how best to make up work for an excused absence.

If you are sick or have symptoms of a possible communicable disease such as COVID or flu, on the other hand, please do not come to class. You should get in touch with me by email as soon as possible to make arrangements for following up on an unplanned excused absence.

Masking and testing

In January, 2024, COVID infections are again on the rise. It is impossible to foresee how the spread of infections will develop this semester, and we continue to teach and learn in ways we did not anticipate. I will monitor the spread of the virus and re-evaluate our masking practices periodically throughout the semester.

Because even a "mild" case can have serious consequences for at-risk populations, as the semester begins, medical-grade or better masks are required for class meetings in Fenwick 420, and for drop-in hours in my office (Fenwick 415) until further notice.

Masking is uncomfortable for us all, but feeling unsafe creates a poor learning environment so please do your part. Please continue to proactively monitor, test, and isolate at the onset of COVID-19 symptoms and close contact with any individuals with known infections. We will continue to closely monitor viral surges and current College recommendations in regards to COVID-19 and any other infectious diseases that might emerge as we move through the semester.

Diversity and Inclusivity

The study of texts and other cultural material belongs to everyone: there are no special privileges or claims of ownership. This course has no rerequisites. The diversity of experience and perspectives that we collectively bring to the class is a resource, a strength and a shared benefit for all of us.

Accommodations for disabilities

Any student who feels the need for accommodation based on the impact of a disability should contact the Office of Disability Services to discuss support services available. The office can be reached by calling 508 793-3693 or by visiting Hogan Campus Center, room 215A.

If you are already registered with Disability Services, please let me know as soon as possible, so that I can take account of this in planning for tests or other course activities.

Academic integrity

You should be familiar with the College's policy on Academic Integrity posted at https://catalog.holycross.edu/requirements-policies/academic-policies/#academicintegritytext