

IS289 Final Project

Project Statement	1
Methods and Rationale	2
Design and Technical Specifications	3
Sustainability and Rights Statement	4
Project Summary	5
Workflow	6
Appendices	8
Core Competencies	8

Project Statement

This project is an investigation into the courses and students comprising the Digital Humanities minor, determining the core competencies through textual mining of the existing course list and extracting narratives from students who are currently are enrolled or have received the minor.

The UCLA Digital Humanities program prepares students to work in the emerging technologies impacting research and pedagogy in the humanities, providing them with knowledge about the tools, methods, and theoretical issues. The project aims to tell a narrative of current and past DH minors, indicating what majors they are coming from and what courses fulfill program competencies. This project is in coordination with the Digital Humanities department at UCLA.

Methods and Rationale

Research in this project is informed by doing textual analysis on registrar data, survey data, and course descriptions and syllabi. For determining which courses satisfy core competencies as determined by DH faculty (see appendix), textual mining is performed in order to align textual phrases to the competencies. Once courses have been tagged with this metadata, an online interface which allows users to select courses based on their interests and needs will be developed. Furthermore, branching logic will provide students with a method for planning out their course schedule as they navigate the minor. Data from past and current students of the program will assist in determining which courses should be offered in the program.

This information will also provide DH administrators with an overhead snapshot of the program and the paths that students take as they navigate the minor. It will also provide administrators with an understanding of how the minor compares with other minors at UCLA as well as at other schools that offer a similar degree. Administrators will also be able to answer which majors are more likely to take DH minors and have a better idea of what students are pursuing after completing their degree.

To convey this information, this project presents a number of graphs showing how the minor has compared to other UCLA Humanities minors over time. There are also charts depicting which majors feed into the DH minor. Furthermore, there will be a search function that allows students to recall courses comprising the minor and how they relate to others through the core competencies of the program. Students will also be able to search by courses based on metadata tags, noting which courses satisfy the requirements necessary for the program. The

findings and features of the project will ultimately be hosted on the UCLA Digital Humanities website (<https://dh.ucla.edu/>) while the staging of the site will take place on a Google Site (<https://sites.google.com/g.ucla.edu/is289/>).

Design and Technical Specifications

The staging for the Digital Humanities project will take place on a Google Site. A Google Site was chosen for a number of reasons: the data was already hosted in various Google documents, particularly Google Sheets; UCLA has a contract with Google allowing any or all assets to be shared with only with UCLA Google users if necessary; Google documents and Sites integrate easily for quick presentation of data and concept. If necessary, outside sources can be embedded within the Google Site easily using iframes.

In order to identify which courses satisfy the core competencies as defined by the DH faculty, textual mining has been done on the course descriptions via Google Sheets. A regexmatch function searches the course descriptions for various text strings that map to competencies. If a string is identified, a tag is placed with that course. For example, to ascertain which courses deal with “computer science”, the following formula is used to apply the tag “computer science” to the course:

```
=IF(REGEXMATCH(lower(I2), "programming|computer science|computer engineering|java|c[++]|python"), "computer science", "")
```

From here, the Google FILTER function is used in order to query courses against core competencies. As a proof of concept, a spreadsheet with the courses delineated by their

general topics has been embedded in the Google Site so that users can browse courses based on topics extracted from the course description. An example of the query function is as follows:

```
=IF(A2="history",FILTER(tags!B:B,tags!J:J=A2),IF(A2="archives",FILTER(tags!B:B,tags!K:K=A2),IF(A2="intellectual property",FILTER(tags!B:B,tags!L:L=A2),IF(A2="religion",FILTER(tags!B:B,tags!M:M=A2),IF(A2="computer science",FILTER(tags!B:B,tags!N:N=A2),IF(A2="geographic information systems",FILTER(tags!B:B,tags!O:O=A2),IF(A2="economics",FILTER(tags!B:B,tags!P:P=A2),IF(A2="information studies",FILTER(tags!B:B,tags!Q:Q=A2))))))))))
```

Some graphs have also been included on the site detailing how the DH minor compares to other minors in the UCLA Humanities division. The graphs show the number of DH minors awarded over the last 5 years as well as the trend of DH minors awarded over the last 5 years. This information assists those interested in the minor to have greater context within the wider UCLA context.

Sustainability and Rights Statement

This project is presently hosted on Google infrastructure. The University of California has a system wide contract with Google Apps for Education

(<https://softwarecentral.ucla.edu/google-apps>). This project will continue to use this

infrastructure for as long as UC has this agreement with Google. If this agreement were to ever be terminated, all of the data in the project can be downloaded as a .csv or in Microsoft Excel format. While some of the functionality may be lost, many of the graphs and functions should be compatible with the Microsoft suite.

All data in this project is owned by the UCLA DH division and is shared with their student employees. This project is commissioned by the DH division in order to gather an understanding of where the program stands both academically and among alumni. Some data that contains personal information from the UCLA Registrar is not made available to any individual outside the DH staff and is contained in a FileMaker database behind a protective login. When and if these data will be used for telling the story of the DH program, all Personally Identifiable Information (PII) will be obscured according to UCLA policy.

Project Summary

This project is currently in its infancy, having only begun in November 2018. In its current state, a complete dataset has not been obtained. This project is currently waiting for these large data elements: registrar data on students currently enrolled or who have completed the minor; survey responses on current or previously enrolled students; syllabi for all DH course electives. These data will fill in currently missing gaps such as which majors are taking the DH minor, what students are doing as alumni of the program, and what courses are popular among students who go through the program.

From the data that currently exists, there are few conclusions that can be extracted with much confidence. For example, mapping courses currently listed as electives in the minor to core competencies has proven difficult based solely on the course descriptions alone. As almost all courses and all electives are external to the program, courses are not designed with DH squarely in mind. This results in a number of courses that are not necessarily designed with the DH core competencies in mind. However in order to perform textual mining on a corpus of texts

under established parameters, standardized language would be useful for pairing courses with competencies. As of the current data on hand, this process is not tenable.

To perform the textual analysis necessary to map courses with core competencies, a larger and more tailored set of course descriptions and learning outcomes needs to be acquired. When this set is acquired and the textual mining complete, a branching course decision scenario can be developed to serve potential and current students. This aspect of the project will require the implementation of a relational database, not unlike the spreadsheet in this project, that will serve as backend for branching scenario tool.

Workflow

1. Acquire datasets from the UCLA Digital Humanities program. Most data in the program is formatted as a CSV or has been captured in Google sheets. Registrar data exists in a FileMaker database.
2. Determine what DH needs with the data. A number of meetings with the Vice Chair of the DH program helped outline the needs, namely craft narratives of current and past DH students by examining their course history; determine a course branching structure by establishing which courses align with core competencies of the program.
3. Once user needs have been established, data needs to be configured to support the project objectives. The following steps were taken to perform text mining on the course descriptions:
 - a. Map which key words or phrases relate to various topic categories. Once a thesaurus is developed, a regexmatch function was applied to the course

description column pairing words or phrases found in the text to designated topics. An example function is as follows:

i. `=IF(REGEXMATCH(lower(I2), "programming|computer science|computer engineering|java|c[+]|python"), "computer science", "")`

b. Once the text has been mined and metadata tags applied to the course, the dataset can be filtered by topic tag. A data validation drop down was created to allow a user to select a topic tag and query a list of courses that satisfied that topic. An example function is as follows:

i. `=IF(A2="history",FILTER(tags!B:B,tags!J:J=A2),IF(A2="archives",FILTER(tags!B:B,tags!K:K=A2))`

c. Once these tag queries have been built, the spreadsheet can be embedded or linked from a website so that users can perform these queries.

4. Place the query function on the website via embed code or iframe.

IS289

topic tag: Computer Science.

To query by topic, open the spreadsheet in a new tab (click the button in the upper right) and click the drop down arrow under Select Topic.

Select Topic	Courses	Course Description
computer science	Freshman Computer Science Seminar	Introduction to department resources and principal topics and key ideas
	Great Ideas in Computer Science	Broad coverage for liberal arts and social sciences students of compute
	Introduction to Computing for Social Sciences and Humanities	Lecture, three hours; discussion, two hours. No prior programming know
	Introduction to Programming	Lecture, three hours; discussion, two hours; laboratory, eight hours. No
	Intermediate Programming	Lecture, three hours; discussion, two hours; laboratory, eight hours. En
	Advanced Programming	Lecture, three hours; discussion, two hours; laboratory, eight hours. En
	Python with Applications	Lecture, three hours; discussion, two hours; laboratory, eight hours. En
	Principles of Java Language with Applications	Lecture, three hours; discussion, two hours; laboratory, eight hours. En
	Introduction to Programming for Internet	Lecture, three hours; discussion, two hours; laboratory, eight hours. En
	Introduction to Electronic Literature	We might begin by asking, what is not electronic literature today? Rath
	World Media Systems	Designed for juniors/seniors. Global analysis of internal and external br
	Geographic Information Systems (restricted to Geography minors)	Introduction to fundamental concepts and architecture of programming c

query tags

The UCLA Digital Humanities Minor is in the process of mapping its course electives to core competencies. This process will involve more textual mining similar to the above

Appendices

Core Competencies

The following is a list of core competencies and skills as derived by the Digital Humanity faculty. The list is incomplete but serves to show the competencies important to the DH Minor and the tools that are important for students to gain while they are in the program.

Competencies

- Management of digital records
- Data curation
- Data modelling
- Data cleaning
- Legacy data preservation and cross-walks
- Knowledge of open linked data specifications and practices
- Data formats
- Data structures
- Metadata standards
- Web development
- Database production
- Topic Modelling
- Data mining
- Information visualization
- Interface design
- Server management and basic linux command line for terminal data work
- Intellectual property/fair use
- Ethical data use and design.

Tools

- HTML
- CSS
- Omeka or other content management system
- Zotero (general bibliographical tool)
- Excel
- Open Refine
- Javascript
- Python

- Mallet
- SQL
- NVivo
- Access
- Tableau
- XML/XSL
- Gephi
- EAD
- Oxygen
- Wordpress/Drupal
- CollectionSpace
- ArchiveSpace / Archivists ToolKit