History 4816A
Digital Research Methods
Fall 2023

N.B. It is anticipated that this course will not be offered in 2024-25.

Instructor: **Professor William J. Turkel**

Email: **wturkel@uwo.ca**

This is a **draft** syllabus. Please see the course site on OWL for a final version.

Course Description

In this course students will learn how historical content is produced, presented, and published online; how to find and evaluate digital primary and secondary sources; and how to use computational techniques to work with digital resources. No previous background in the subject area is required.

Prerequisite(s):

Registration in any academic program at the second-year level or above.

Unless you have either the prerequisites for this course or written special permission from your Dean to enroll in it, you may be removed from this course and it will be deleted from your record. The decision may not appealed. You will receive no adjustment to your fees in the event that you are dropped from a course for failing to have the necessary prerequisites.

Antirequisite(s): [List any antirequisites here.]
Digital Humanities 2130A/B, Digital Humanities 3902F/G, the former History 2816F/G.

Course Syllabus

Research of all kinds now crucially involves the acquisition and use of digital sources, both primary and secondary. In this course, you will learn to find, harvest, manage, excerpt, cluster and analyze digital materials throughout the research process, from initial exploratory forays through the production of an electronic article or monograph which is ready to submit for publication. The digital sources that we work with include texts, images, maps, code, social media, spreadsheets, databases and linked open data. The techniques that we will employ are currently used by relatively few humanists and social scientists—although they have become much more crucial in an age of pandemic lockdowns. Computation won't magically do your research for you, but it will make you much more efficient. You can focus on close reading, interpretation, and writing, and use machines to help you find, summarize, organize, and visualize sources.

Prior Experience

The only thing that you need to do well in this course is a willingness to learn new things and the perseverance to keep working when you're confused or when you realize that you could spend a lifetime learning about the topics and technologies that we will cover in class, and still not master them all. Students will come into the course with very different levels of experience and expertise. Some, probably most, will be familiar only with the rudiments of computer and internet use. A few may already be skilled programmers.

Workload

This course requires that you spend at least a bit of time each day (say 20-30 minutes) practicing your new skills. It's a lot like learning a new language, learning to play a musical instrument, or going to the gym. At first it is going to be hard but be patient with yourself and ask a lot of questions. With daily practice, you will soon find ways to do your research and coursework faster and more efficiently. If you can't commit to regular practice, however, you should probably not take this course. The techniques that you learn in this class build cumulatively week-by-week, and there are several evenly spaced assignments of equal value.

Learning Outcomes

At the conclusion of this course, you will

- Be aware of a wide variety of different kinds of digital sources and strategies for making use of them computationally

- Be familiar with a range of research approaches in digital humanities, computational social science and related disciplines

- Be acquainted with methodological applications of artificial intelligence

- Be more familiar with using computer code as a medium of expression

Methods of Evaluation

Students must submit <u>all</u> course assignments to pass the course.

Assignment	Worth	Handed Out	Due Date	Zero Date
01	20%	Week 01	Week 02	Week 03
02	20%	Week 03	Week 04	Week 05
03	20%	Week 05	Week 06	Week 07
04	20%	Week 08	Week 09	Week 10
05	20%	Week 10	Week 11	Week 12

Each assignment has a suggested due date and, a week later, a zero date. If you hand in your assignment before the zero date, you will receive full credit for your work. After the zero date, it will be worth nothing.

If you are unable to meet a course requirement due to illness or other serious circumstances, you must provide valid medical or other supporting documentation to Academic Counseling as soon as possible.

Course Materials
To get the most out of this class, you will need a Windows, Mac or Linux laptop or desktop. A tablet or netbook will probably not be sufficient. If you have any concerns, email me.

You will need a **student desktop license** for Wolfram Research's *Mathematica* software. (Don't let the name scare you, you won't need any particular training in mathematics to do well in the class). In the past, this software has sometimes been available for free for Western students. Check here to see if that is the case:

https://www.wolfram.com/siteinfo/

If it is not freely available this term, you will have to purchase a license. A one semester license is about $90 CAD. You can purchase the software here

https://www.wolfram.com/mathematica/pricing/students/

If you are unable or unwilling to purchase the software, please do not take the course.

The required text and screencasts for the course are freely available online:

William J. Turkel. *Digital Research Methods with Mathematica*, 2nd rev ed. (2020). https://williamjturkel.net/digital-research-methods-with-mathematica/

Course Schedule and Readings
- **Week 01.**

    - **Introduction to *Mathematica*.** Interacting with notebooks

- o **Reading Code**. Word frequency, word clouds, and stopwords.

- **Week 02.**

  - o **Computable Knowledge**. Entities, tables, timelines, and maps.

  - o **Text Content**. Mathematica notebooks and expressions, strings, and natural language processing.

- **Week 03.**

  - o **Data Structures**. Lists, associations, and datasets.

  - o **Reusing Code**. Defining and developing functions, keyword in context (KWIC).

- **Week 04.**

  - o **Networks**. Metadata, matrices, and social network analysis.

  - o **Indexing and Searching**. Pattern matching, topic classification and term distribution.

- **Week 05.**

  - o **Geospatial Analysis**. Geographic information: raster, vector, and attribute data.

  - o **Images**. Computer vision, face detection, feature extraction and image mining.

- **Week 06.**

  - o **Page Images**. Optical character recognition (OCR), figure extraction and classification.

  - o **Crawling**. Browser automation, batch downloading, web archives and WARC files.

- **Week 07.**

  - o **Linked Open Data**. Resource description framework (RDF), SPARQL queries and endpoints, JSON-LD.

  - o **Markup Languages**. Scraping and parsing, XML, really simple syndication (RSS) and text encoding initiative (TEI).

- **Week 08.**

- o **Studying Societies**. Computational social science, search data, social media, and social networks.

  - o **Extracting Keywords**. Information retrieval, term frequency-inverse document frequency (TF-IDF) and rapid automatic keyword extraction (RAKE).

- **Week 09.**

  - o **Word and Document Vectors**. Feature extraction, dimension reduction, word embeddings and global vectors.

  - o **Citations**. References, web services, bibliographic linked open data, and citation networks.

- **Week 10.**

  - o **Natural Language**. Multilingual analysis, computational linguistics, and sentiment analysis.

  - o **Web Services**. Entity networks, publication search, dashboards, manipulating JSON.

- **Week 11.**

  - o **Databases**. Parts, selections and transformations, computations and querying, relations.

  - o **Measuring Images**. Photogrammetry, georectification, handwriting and facial 3D reconstruction.

- **Week 12.**

  - o **Machine Learning**. Unsupervised clustering, classify, predict, and transfer learning.

Additional Statements

Please review the Department of History's shared policies and statements for all undergraduate courses at: https://history.uwo.ca/undergraduate/program_module_information/policies.html for important information regarding accessibility options, make-up exams, medical accommodations, health and wellness, academic integrity, plagiarism, and more.