

History 9877a

Digital Research Methods

Fall Term
Wednesdays, 6:30-9:30 pm, Lawson Hall 2270C

William J. Turkel, Professor
Department of History
Office: Lawson Hall 2267
E-mail: wturkel@uwo.ca

Course website:

<http://williamjturkel.net/teaching/history-9877a-digital-research-methods-fall-2015/>

Short Course Description

Historical research now crucially involves the acquisition and use of digital sources. In History 9877A, students 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.

Three Scenarios

Why should you take a course in digital research methods?

1. You've just returned from a whirlwind trip to the archives. On your laptop you have about nine thousand digital photographs of various documents. You could spend the next few years going through the pictures one at a time and typing notes into a word processor. Or you could write a small script to convert each image into readable text and drop the whole batch into a custom search engine. In less than an hour you could be searching for words and phrases anywhere in your primary sources.

2. You discover that the Internet Archive has a collection of eight hundred online texts that are directly related to your research. You could look through the list of titles in your web browser and click on the links one at a time, scanning each to see if it is relevant. Even if you cut-and-paste notes from the sources to a word processor, it will still take you at least a few months to go through the collection. Or you could write a small script to download all of the sources to your own machine and run a clustering program on them. This sorts the texts into folders of closely related documents, then subdivides those by topic. In less than an hour, you would

be able to visualize the contents of the whole collection and focus in on the topics that are of immediate interest to you.

3. You've been working with the written corpus of a historically significant figure. You have the books and essays that he or she wrote, their diary entries and their correspondence with a large number of other individuals. How do you make sense of a lifetime of writing? Can you chart important changes in someone's conceptual world? Spot the emergence of new ideas in the discourse of a community? Map the ever-changing social relations between a network of correspondents?

In this course you will learn to apply techniques that are currently used by fewer than one percent of working historians. 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.

Prerequisites and Workload

There are no prerequisites for the course other than 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.

This course also requires that you spend at least a little 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. 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.

Required Software and Textbook

To get the most out of this class, you will need a Windows, Mac or Linux laptop, which you should bring to every class.

You should purchase a **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).

<http://www.wolfram.com/mathematica/pricing/students.php>

You can purchase the license for the semester (US \$45), for the year (US \$70) or permanently (US \$140). If you are unwilling or unable to purchase the software, please do not take this course.

The textbook for the course will be supplied free of charge.

Evaluation

The techniques that you learn in this class build cumulatively week-by-week. In addition to regular practice, it is essential that you attend every meeting of the class and do the readings carefully.

You will be graded on your participation in class (20%) and on your reflective blogging (80%; see below).

There will be no midterm or final examinations, and no final paper.

Blogging

Every student in the class will have an academic blog and will be required to make weekly posts to it. These entries do not have to be long (300-500 words per week is ample). The use of blogging is to encourage you to engage in 'reflective practice,' that is, to force you to think about your learning and research as you are doing it. It also provides me with feedback for how the course is going. You can use each week's blog entry to talk about what you learned, things that were clear or not, things you would like to know how to do, and so on.

If you do not already have an academic blog, you need to create one. Before the first class you should go to either WordPress or Blogger (not both) and create an account and a blog. If possible, create the blog under your own name; if not, choose something professional sounding. Post an introductory message about yourself and then send me the URL of your blog so that I can add you to the course blogroll for History 9877A.

Wordpress: <http://wordpress.com>

Blogger: <http://www.blogger.com>

Tentative Schedule (Weekly Topics)

This is a preliminary list of topics. Consult the course website for up-to-date information.

- **Analyzing Text.** Frequency analysis. String, list and association data structures. Stopwords. Word cloud visualization. Searching through text. The bag of words representation. Map and pure functions.
- **Pattern Matching.** String patterns. Computable word data. Generating a concordance (keyword in context listing). Capitalized words and phrases. N-gram analysis. Finding one word near another. Sorting and pure functions.
- **Who and What.** Named entity recognition. Computable data about people. Displaying linked information. Identifying probable names. Programmatically searching *WolframAlpha* and *Wikipedia*. NestList.
- **Where and When.** Historical periods and events as entities. Date computations. Timeline visualization. Series of events. Geographical information and visualization.
- **Information Retrieval.** Vector model of documents. Similarity and Relevance. TF-IDF. Text summarization and clustering. Sequence alignments.
- **Harvesting and Processing Webpages.** Parsing URLs and URIs. Batch downloading. Extracting hyperlinks, structured data and text. Indexing.
- **Markup Languages.** Parsing and generating XML and HTML.
- **Structured Data.** Querying databases. Descriptive statistics. Dataset data structure. Selections and transformations. Relational data and normalization.
- **Page Images and PDFs.** Portable document format. Optical character recognition. Assessing OCR quality. Bursting documents. Extracting text and images. Problem of handwritten pages.
- **Still and Moving Images.** Extracting pictures from page images. Image processing. Face recognition. Image identification. Visualizing large sets of images. Photogrammetry. Detecting image tampering.
- **Networks.** Representing and describing networks. Social network analysis Visualization. Social media data.
- **Bibliography and Bibliometrics.** Bibliographic metadata. Citation databases. Citation analysis. Stable sources. APIs.
- **Web Crawling.** Spidering and scraping. Working with WARC files. Networks of named entities. RDF and SPARQL.

ADDITIONAL STATEMENTS

Prerequisites and Antirequisites:

Unless you have either the requisites for this course, as described in the Academic Calendar description of the 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. This decision may not be 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. The Academic Calendar description of each course also indicates which classes are considered antirequisites, i.e., to cover such similar material that students are not permitted to receive academic credit for both courses.

Academic Offences:

Scholastic Offences are taken seriously and students are directed to read the appropriate policy, specifically, the definition of what constitute a Scholastic Offence, at the following Web site:

http://www.uwo.ca/univsec/pdf/academic_policies/appeals/scholastic_discipline_undergrad.pdf

Plagiarism:

Students must write their essays and assignments in their own words. Whenever students take an idea, or a passage from another author, they must acknowledge their debt both by using quotation marks where appropriate and by proper referencing such as footnotes or citations. Plagiarism is a major academic offense (see Scholastic Offence Policy in the Western Academic Calendar).

All required papers may be subject to submission for textual similarity review to the commercial plagiarism detection software under license to the University for the detection of plagiarism. All papers submitted will be included as source documents in the reference database for the purpose of detecting plagiarism of papers subsequently submitted to the system. Use of the service is subject to the licensing agreement, currently between The University of Western Ontario and Turnitin.com (<http://www.turnitin.com>).

The following rules pertain to the acknowledgements necessary in academic papers.

A. In using another writer's words, you must both place the words in quotation marks and acknowledge that the words are those of another writer.

You are plagiarizing if you use a sequence of words, a sentence or a paragraph taken from other writers without acknowledging them to be theirs. Acknowledgement is indicated either by (1) mentioning the author and work from which the words are borrowed in the text of your paper; or by (2) placing a footnote number at the end of the quotation in your text, and including a correspondingly numbered footnote at the bottom of the page (or in a separate reference section at the end of your essay). This footnote should indicate author, title of the work, place and date of Publication and page number. Method (2) given above is usually preferable for academic essays because it provides the reader with more information about your sources and leaves your text uncluttered with parenthetical and tangential references. In either case words taken from another author must be enclosed in quotation marks or set off from your text by single spacing and

indentation in such a way that they cannot be mistaken for your own words. Note that you cannot avoid indicating quotation simply by changing a word or phrase in a sentence or paragraph which is not your own.

B. In adopting other writer's ideas, you must acknowledge that they are theirs.

You are plagiarizing if you adopt, summarize, or paraphrase other writers' trains of argument, ideas or sequences of ideas without acknowledging their authorship according to the method of acknowledgement given in 'A' above. Since the words are your own, they need not be enclosed in quotation marks. Be certain, however, that the words you use are entirely your own; where you must use words or phrases from your source; these should be enclosed in quotation marks, as in 'A' above.

Clearly, it is possible for you to formulate arguments or ideas independently of another writer who has expounded the same ideas, and whom you have not read. Where you got your ideas is the important consideration here. Do not be afraid to present an argument or idea without acknowledgement to another writer, if you have arrived at it entirely independently. Acknowledge it if you have derived it from a source outside your own thinking on the subject.

In short, use of acknowledgements and, when necessary, quotation marks is necessary to distinguish clearly between what is yours and what is not. Since the rules have been explained to you, if you fail to make this distinction, your instructor very likely will do so for you, and they will be forced to regard your omission as intentional literary theft. Plagiarism is a serious offence which may result in a student's receiving an 'F' in a course or, in extreme cases, in their suspension from the University.

Medical Issues:

The University recognizes that a student's ability to meet his/her academic responsibilities may, on occasion, be impaired by medical illness. Please go to https://studentservices.uwo.ca/secure/medical_accommodations_link_for_OOR.pdf to read about the University's policy on medical accommodation. This site provides links to the necessary forms. In the event of illness, you should contact Academic Counselling as soon as possible. The Academic Counsellors will determine, in consultation with the student, whether or not accommodation should be requested. They will subsequently contact the instructors in the relevant courses about the accommodation. Once the instructor has made a decision about whether to grant an accommodation, the student should contact his/her instructors to determine a new due date for tests, assignments, and exams.

SUPPORT SERVICES:

Students who are in emotional/mental distress should refer to [Mental Health@Western](mailto:MentalHealth@Western), <http://www.uwo.ca/uwocom/mentalhealth/> for a complete list of options about how to obtain help.

Please contact the course instructor if you require material in an alternate format or if you require any other arrangements to make this course more accessible to you. You may also wish to contact Services for Students with Disabilities (SSD) at 661-2111 x 82147 for any specific question regarding an accommodation.

If you have any further questions or concerns please contact, Morgan Sheriff, Undergraduate Program Advisor, Department of History, 519-661-2111 x84999 or msherif5@uwo.ca