



# NORTHWEST

MISSOURI STATE UNIVERSITY

MARYVILLE | KANSAS CITY

School of Computer Science and Information Systems

44-520:Web Mining(3 hrs.)

Spring 2023

**Instructor:** Dr. Nathan Eloë

**Office hrs:** TR 9:00–11:00, 1:00–2:00; W 9:00–10:00

**Phone:** 660.562.0832

**Email:** [nathane@nwmissouri.edu](mailto:nathane@nwmissouri.edu)

**Office:** CH2270

**Prerequisites:** There is no textbook for this course. Course slides and additional materials will be made available on the course website. Additional material will be presented in class.

### Textbook and Supplementary Materials:

**Textbook:** There is no required textbook for this course; however, much of the course content is motivated by *Mining the Social Web* by Matthew A. Russell (O'Reilly books). While portions of the book are already out of date (API versions and social networks that no longer exist) a good amount of the content is relevant.

**Required Software:** Anaconda Python (<https://www.anaconda.com/download/>)

Students are required to have a GitHub account for this course. The student's school email address must be associated with the account, but is not required to be the primary email on the account.

**Course Description:** The Web, the largest publicly accessible data source in the world, has grown at an unprecedented pace during the past decade. The rapidly changing news and events make the Internet a non-stationary environment, which in turn generates rapidly changing data. This course examines the unique characteristics of Web from a data mining perspective. It provides a review of data mining foundations and covers basic concepts of text mining, information retrieval, web search and data mining applications on social media.

**Student Learning Outcomes:** *In addition, after completing this course, a student should be able to:*

- retrieve data from the World Wide Web
- transform unstructured data into a usable format and extract information from this data
- perform analysis on extracted data to draw conclusions.
- use tools to perform Natural Language Processing on a corpus of data

**Instructional Methods:** Class time will consist primarily of lecture, hands on project work, and student discussion about course topics, with time allotted for individual or group work on projects and homework.

**Graded Course Requirements:** This course is primarily a project based course. Students will be graded on their implementation and presentation of their results as well as the conclusions they can draw from their projects.

**Projects:** Students will explore various aspects of the Internet through programming projects. Students will have some latitude in determining what question they want to answer using the techniques and resources made available in this class. Graduate students will be required to do an additional project during the semester.

**Posters:** Students may be asked to design and present posters about the work they do in this course.

**Final Exam:** The final exam activity will be given during the time specified by the registrar at <https://www.nwmissouri.edu/registrar/finals.htm>.

**Quizzes and Worksheets:** You may be assigned exercises to do on your own time. At any time, the instructor may choose to give a quiz in lieu of collecting an assigned worksheet.

**Grading Scale:** The grading scale is the standard 10 point grading scale, but may be revised in the students' favor at the instructor's discretion.

>90%	A
80%-89.9%	B
70%-79.9%	C
60%-69.9%	D
<60%	F

**Late Policy:** Any work submitted within 24 hours of the deadline will receive half of the credit that would have been received on a timely submission. Work submitted after 24 hours of the deadline will not be accepted for a grade.

The instructor may give extensions in extenuating circumstances; extensions should be discussed with the instructor before the deadline. Extensions will not be granted after the deadline.

**Tentative Course Outline/Major Topics Studied:**

- **(Re)Introduction to Python**
  - Python Basics (using Jupyter/iPython)
  - Functional Python
  - Visualization
- **Web APIs: Twitter**
  - Python API Modules
  - Frequency Analysis
  - Lexical Diversity
  - Visualization of Data/Results
- **Graphs and Connections**
  - Basic graph theory concepts
- **Web Scraping and Natural Language Processing**
  - Natural Language Processing
  - Automatic Summarization
- **Clustering and Recommender Systems**
- **Mining GitHub**

**Attendance:** Students are expected to attend all classes as specified in the course syllabi for each course. Exams, homework due, quizzes given, or in-class assignments completed during your absence may only be handed in at a later date if the absence is excused. Excused absences include attendance at a university sponsored event (documented with an excuse signed by the university sponsor prior to the event) or by circumstances considered adequately extenuating by the course instructor. It is the responsibility of the student to promptly notify his or her instructor when unable to attend class. In order for an absence to be excused, the student must email the instructor to prior to class time. Please refer to the university policy on attendance at <https://www.nwmissouri.edu/policies/academics/Attendance.pdf>

**Final Exams:** If an emergency occurs that prevents the administration of a course scheduled final examination, the final course grades will be calculated based on the work in the course completed to that point in time and the faculty member's considered judgment. Final exams will not be rescheduled, and a grade of "I" will not be given as a result of an institutional cancellation of a final examination. This final exam policy does not apply to online courses.

**Administrative Drop:** An instructor may request the Office of the Registrar delete a student from a course roster if the student has not met the prerequisite for the course as stated in the catalog, or as a result of non-attendance in the course.

**University Communications:** Students are expected to use their Northwest student email account for any electronic correspondence within the university. Students are also strongly advised to check their email and CatPAWS accounts on a regular basis.

**Disclaimer:** Course schedule is subject to change and you will be responsible for abiding by any such changes. Your instructor will notify you of any changes.

**Academic Integrity Policy:** The students, faculty, and staff at Northwest endeavor to sustain an environment that values honesty in academic work, that acknowledges the authorized aid provided by and intellectual contributions of others, and that enables equitable student evaluation. Please refer to Northwest Missouri State University's Academic Integrity Policy at <http://www.nwmissouri.edu/policies/academics/Academic-Integrity.pdf>

**Special Accommodations Policy:** Northwest Missouri State University complies with Section 504 of the Rehabilitation Act of 1973 and the Americans with Disabilities Act of 1990 [ADA] and the ADA Amendments Act of 2008 [ADAAA]. If a student has a disability that qualifies under the ADA/ADAAA and requires accommodations, they should contact the Office for Equity and Accessibility for information on appropriate policies and procedures at 660.562.1639, or [pjp@nwmissouri.edu](mailto:pjp@nwmissouri.edu) / [ADA@nwmissouri.edu](mailto:ADA@nwmissouri.edu). For the university policy on disability accommodation refer to <http://www.nwmissouri.edu/policies/student/Disability-Accommodation.pdf>.

**Non-discrimination and Anti-harassment Policy:** Northwest Missouri State University is committed to maintaining an environment for all faculty, staff, students, and third parties that is free of illegal discrimination and harassment. Please refer to the Non-Discrimination and Anti-Harassment Policy at <http://www.nwmissouri.edu/diversity/titlevi.htm>.

**Family Education Rights and Privacy Act (FERPA) Policy:** Family Educational Rights and Privacy Act of 1974, as amended (commonly known as the Buckley Amendment), is a federal law which provides that colleges and universities will maintain the confidentiality of student education records. Please refer to the Family Educational Rights and Privacy Act (FERPA) Policy at <http://www.nwmissouri.edu/policies/academics/Family-Educational-Rights-and-Privacy-Act.pdf>

**Disclaimer:** Course schedule is subject to change and you will be responsible for abiding by any such changes. Your instructor will notify you of any changes.