GEOG 384: Principles of the Geospatial Web

Tuesdays and Thursdays, 13:05-14:35, Burnside 511

Instructor, Prof. Renee Sieber, renee.sieber@mcgill.ca

Ian Tattersfield, Geospatial Data Officer - Campus Planning and Development, ian.tattersfield@mcgill.ca
Sieber Office Hours Tuesday, Thursday, 2:30-3:30pm, 411 Burnside Hall

TA: Sichen Wan, sichen.wan@mail.mcgill.ca

Introduction
Most often when students think of geographic information systems (GIS), it is a mapmaking tool, a complete yet expensive, desktop-bound software package requiring considerable computational power. We now work in a software "stack" of interoperable and "mashed up" web 2.0 applications, many of which are free and require no more processing power than being able to run a basic web browser. The learning curve has changed as well. Where once a GIS technician entered, managed and visualized data as part of a specific learning curve, these functions are being replaced by "smarter" software and the skillset is shifting to coding by engineers and computer scientists.

With smarter software comes the advent of generative AI and geofoundation models (colloquially known as GPTs or generative pretrained transformers). GPTs promise to write perfect computer code, craft explainable documentation and reason over geographic space (function multi-modally e.g., by integrating coordinates and semantics, automatically processing remotely sensed imagery). GPTs propose to basically cut humans out of the loop.

The nature of geographic data has changed as well. Large and streaming datasets, called big data and are derived from sources such as social media and mobile sensors, are commonplace. This includes facial pattern recognition from autonomous vehicles and CCTV cameras. Entire industries have sprung up to harness the power of interpreting patterns revealed by alphanumeric and image data. These industries also have the capability to influence policy and society as they monetize what has been termed the "ambient awareness" of pervasive space-time data collection. The sheer amount of data with spatial attributes, produced in a single day, could not have been created a mere ten years ago. Whereas the underlying technology has rapidly evolved, our understanding of the information we generate has not kept pace. This calls to question issues of accuracy, rights, and privacy. Should we be able to scrape for user generated content and trust in its validity? Or should we use uncritically utilize crowdsourced asserted data (i.e., location-based sentiments and observations from non-experts called volunteered geographic information {VGI})? The data is far more unstructured and is sourced to a myriad of multimedia (e.g., videos, graphs, and photos). This geospatial data constitutes a departure

from traditional authoritative datasets of latitude, longitude, projection and attributes, and introduces varied data structures, potentially disputed sources and one record of information at a time. GPTs are the newest addition, enabling the mass harvesting of content, regardless of intellectual property, and generation of text, audio and imagery from text prompts. Thus, the user is simultaneously the producer of new forms of spatial data, as well as user and commodity.

In this course we will cover the principles of this paradigm, called the Geospatial Web or Geoweb. Like GIS, the Geoweb has broad applicability that extends well beyond the traditional discipline of geography. Numerous examples can be seen in health provision (e.g., fitness tracking), wildlife and natural resources (e.g., with sensors), criminology (e.g., predictive mapping of likely offenders), and transportation (e.g., Uber, Lyft, as well as docked/dockless vehicles). The Geoweb can be used in retail marketing, political mobilization, tourism development, journalism and the humanities. The Geoweb also can be fully linked to social networking platforms. Perhaps the most lucrative use of the Geoweb, however, has been in location-enabled mobile apps that we use every day. Have you ever used a physical map to navigate space or do you seek directions to a new location via Google or Siri/Alexa?

The course will offer a combination of theory and practice of the Geoweb. In terms of theory, we will cover subjects like VGI and big data and how it changes our ideas about spatial data accuracy. In practical lab sessions students will build Geoweb "apps". An additional goal is not to learn specific skills, but to 'learn how to learn'. Geoweb software changes monthly so students learn of resources and strategies to effectively use emerging technologies and anticipate innovations in geospatial technologies. The course will cover the following topics:

- Exploring the underlying methods of digital earth architectures, including georeferent systems. These architectures underlie most digital earth platforms like Google Maps/ Earth, OpenStreetMap and OpenLayers)
- Exploring political, sociological, economic and legal issues in using VGI and platforms
- Critically analyzing the concept of VGI, for example, spatial data accuracy and uncertainty of heterogeneous data sources (spatial data quality)
- Understanding the infrastructure of the Geoweb, including the geospatial software stack, Application Program Interfaces (APIs), and the Document Object Model (DOM)
- Comparing and contrasting GIS and the Geoweb
- Repurposing geographic digital content (secondary data), for example via web scraping
- Learning underlying concepts of server/cloud geospatial applications
- Understanding issues related to real time streaming data (e.g., changes in sampling and geostatistics)
- Working with location-based services and mobile platforms

- Identifying basic problem-solving requirements for geospatial apps. For proposed solutions, critically evaluating and justifying various existing and emergent geospatial technologies and enabling software stacks
- Explaining, writing prompts, critiquing and generally preparing the student for changes as a result of GPTs.

We also may hear from experts in the field. We usually have guest lectures from representatives of Geoweb firms like Mapbox as well as representatives of traditional GIS firms like ESRI, coordinators of VGI sites like OpenStreetMap, and faculty members working in the Geospatial Web (e.g., researching copyright laws for geospatial data, building foundation models).

Finally, and time willing, we will host some workshops outside class time to assist in learning. Among possible topics are HTML and CSS, the Document Object Model, GIT and Scalable Vector Graphics. Stay tuned for times and locations.

Course Prerequisites
GEOG 201 and COMP 202/GEOG 333 are required (COMP 202 can taken in parallel). OR permission of the instructor.

Books and Other Reading Material
During the course, students are required to read a variety of articles, working papers and other material. These will be posted on the course website (neogeoweb.ca). Students must learn Javascript. This will be done via code academy and W3Schools. If you require additional help in learning material such as KML, web scraping, or Javascript, you may wish to purchase books (e.g., Eloquent JavaScript: A Modern Introduction to Programming). There are other excellent online resources as well—freecodecamp and Mozilla Web Developer. This course assumes you will pick up HTML and CSS; the afore-mentioned sites are excellent resources for those as well.

Evaluation
In-class participation 10% (5% in-class and 5% online)
In class quizzes – 20%
Assignments – 40%
Final Exam – 30%

In-class and online participation: Student preparation and participation as well as performance during class will be assessed for half (5%) of the participation grade. We are looking for quality of contributions over the quantity of contributions. Class participation will be evaluated based on evidence that students have read assigned readings, done exercises and otherwise prepared for class. Students also will be assessed in their ability to thoughtfully and reflectively build on other students' contributions.

The other five percent of the participation grade will be based on online contributions on "Twitter".[1] As with in-class participation, students should not only contribute original content but also react to their colleagues' content. Once again, quality of contributions is preferred over number of contributions.

Assignments: Assignments allow the student to apply lecture material and/or programming to real world cases of geospatial representation on the web. There are five assignments in the course. All assignments are completed in groups of three to five students. The goal is to balance levels of computational experience so every group possesses a similar level. All assignments are graded as a group regardless of individual contribution.

Note 1: The temptation in all assignments is to let the experts or intermediates in the team handle the "hard" technical details. RESIST THIS TEMPTATION because it will hurt you in the quizzes and final lab exam, which are individual-based. As you know from working with GIS or RS, knowing the tech ONLY comes from doing (and occasionally, creatively failing) the tech.

Note 2: With the advent of GPTs, the temptation is to have transformers to write all the code for you. RESIST THIS TEMPTATION because it will hurt you in the quizzes and final lab exam, which are individual-based and in-person. Moreover, learning the logic of coding (like the logic of a natural language) requires practice, for example, how to debug software.

Quizzes: Approximately half of the assignments require the use of Javascript. Quizzes will be announced the class prior to the day the quiz is held. You will be expected to learn the code academy material up to that point (code academy course schedule TBA).

Learning JS: You need to demonstrate your knowledge of JS. To do this, you must email a screenshot from code academy (TBA) by the instructor by the beginning of Week 6. This screenshot will demonstrate that you have completed the required modules. It is possible to waive out of the coding requirement if you know JS. Contact the instructor for details.

Final Exam: The final exam is comprehensive of all material in the course. It is divided into two equally weighted components—a written exam and a lab practical exam. The final exam is *four* hours long with a one-half hour break in between.

Obligatory Statements

In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.

[1] We closely follow the tribulations of "Twitter". Despite tribulations, Twitter/X remains the only realtime social media site also followed by mapping companies. If "Twitter" has a major fail during the course then we will shift to a subreddit.

This does not apply to courses in which acquiring proficiency in a language is one of the objectives.

Conformément à la Charte des droits de l'étudiant de l'Université McGill, chaque étudiant a le droit de soumettre en français ou en anglais tout travail écrit devant être noté (sauf dans le cas des cours dont l'un des objets est la maîtrise d'une langue).

McGill University values academic integrity. Therefore, all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures (see www.mcgill.ca/students/srr/honest/ for more information).

ADDITIONAL STATEMENTS

We endeavor to provide an inclusive learning environment in this course. If you are experiencing barriers to learning in this course, do not hesitate to discuss them with a member of the teaching team and/or the Office for Students with Disabilities, 514-398-6009.

We value intellectual freedom and we also value creating a safe and respectful classroom environment. If you do not feel safe in the classroom by virtue of comments related to your race, gender, sexual orientation, physical ability or for any other reason, please come speak with the course instructors.

Student well-being is a priority for the University. All of our health and wellness resources have been integrated into a single Student Wellness Hub, your one-stop shop for everything related to your physical and mental health. If you need to access services or get more information, visit the Virtual Hub at mcgill.ca/wellness-hub or drop by the Brown Student Services Building (downtown) or Centennial Centre (Macdonald Campus). Within your faculty, you can also connect with your Local Wellness Advisor (to make an appointment, visit mcgill.ca/lwa).

McGill has policies on sustainability, paper use and other initiatives to promote a culture of sustainability at McGill. (See the Office of Sustainability.)

In the event of extraordinary circumstances beyond the University's control, the content and/or evaluation scheme in this course is subject to change.

Lastly, about GPTs: GPTs represent an emerging and evolving technology that holds implications for the geospatial web. We will embrace the use of GPTs in the course.

However, our use of GPTs is likely to change over the course of our learning together, which makes it challenging to definitively scope its acceptable use in our course. To that end

- Actively participate in discussions on the use of GPTs for coding, writing, idea generation and design (we use ChatGPT, i.e., GPT-4 but this includes but is not limited to: Google Bard, Grammarly, Google Translate, DALL E 2, and Midjourney) in class and to explore how and when these tools might be acceptable.
- Acknowledge your use of GPTs in your's and your group's submitted work, for example its use in code generation, validation, documentation, editing, etc.

Misrepresentation of your use of GPT will be considered a breach of this compact.

Syllabus
Introduction Overview of Geoweb and goals of the course. In-class exercise: Work with existing Geoweb platform, Mapbox GL JS. Set up "Twitter" account for online participation Sign up at OpenAI Answer a few questions about your comfort level with various concepts Get set up on the course server Sign up for code academy Sign up for Mapbox account Create a story map https://www.mapbox.com/solutions/interactive-storytelling/
Module 1: History and evolution of geospatial data handling and visualization. Definitions and types of digital earths. Georeferent systems, tiling, and projections. Differences between data structuration in GIS compared to the Geoweb. Introduction to concepts of markup languages (XML, HTML, KML) and list-based ways of storing data. In-class exercise: Create your first dataset in geojson, with one point, line, polygon, extrusion, photo. Assignment 1: Create a storymap with geojson.io
Module 2: Everything You See in Cloud is a Data Source Introduction to web architectures, software stacks. Principles of web harvesting/scraping. Legal issues in "repurposing" data, for example, intellectual property and liability. Example from tourism. Structured and unstructured data. In-class exercise: Work with Unstructured Content Conduct web scraping. Tag content with XML. Create a web page. Build a GPT prompt

Assignment 2: Use Google Spreadsheets and XPath to automatically scrape and then map a popular classified advertising site

Module 3: Fundamentals of Coding in an Era of GPTs

- Variable Typing
- Data Structures
- Control Logic
- Functions
- User Input, Prompts, Alerts and Event Coding
- Debugging (e.g., use of console.log)
- The opportunities & limits of GPTs
- Plagiarism, intellectual property and GPTs

In-class exercise: Write a prompt in ChatGPT. Compare your prompts with others in class. Refine your prompts. Inspect the logic of the results. and insert your own

Module 4: Democratization of Data: Volunteered Geographic Information (VGI), the crowd and Beyond

Concepts in VGI (e.g., citizen sensors, crowdsourcing, and neogeography). Motivations for volunteers to contribute. Common methods to assessing accuracy of VGI. Legal issues underlying VGI (e.g., copyright and intellectual property of using citizen-generated content). Emergence of citizen science, crisis mapping and open data. Use of crowdsourcing for content moderation of GPTs.

In-class exercise: Create and edit your own VGI with Google Forms or other platform.

Assignment 3: Create tracker app with the Mapbox on the browser with Javascript and HTML and utilizing Google spreadsheets as the Internet database

Module 5: Streaming, real time BIG Geospatial Data

Introduction to concepts of big data and data-intensive science. Brief discussion of geosensors, which are a prime source of geolocated data. Challenges to working with big data, like sampling.

In-class exercise: Envision a stream of data

Assignment 4: Create a weather fashion app based on user-chosen locations

Module 6: Geo- and Data Visualization

Principles of cartographic/ geo-visualization and the emergence of data visualization. Exploratory data visualization vs explanatory data visualization. Visualization as storytelling.

In-class exercise: Explore the “maps” of D3JS. Tell a story with a Geoweb platform.

Module 7: Geoweb on Mobile Devices (if time permits)

Concepts of geospatial awareness on devices--location based services (LBS), including location intelligence, vehicle tracking, and RFIDs. Examples from mobile commerce. Social issues in LBS (privacy, surveillance)

In-class exercise: Mobile messaging and mapping of geographic data

Assignment 5: Conduct data journalism using Canvas on jsfiddle or Tableau

TBD: Final exams