

UTK
Geography
Department
Fall 2017

Geography 509: Geospatial Semantics

Welcome to GEOG 509 – Geospatial Semantics. My name is Yingjie Hu, and I'm your instructor of this course. In today's Big Data age, many datasets present a natural association between geographic locations and unstructured textual descriptions. Examples of such data include geotagged social media data (e.g., geotagged Tweets), geotagged Web pages (e.g., geotagged Wikipedia pages), reviews about restaurants (or more generally, Points of Interest), and general human descriptions about places. In this course, you will learn how to make sense of these unstructured geo-text data by integrating geospatial analysis and natural language processing (NLP) techniques.

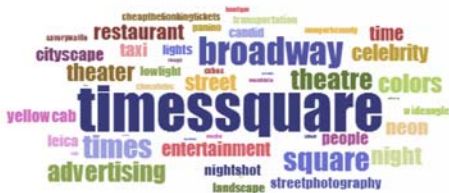


Course Information

- **Course number:** GEOG 509 003 (CRN: 51313)
- **Credit hours:** This course is 3 credit hours (graduate and undergraduate) consisting of one 75-minute lecture and one 75-minute lab each week.
- **Lecture location:** Burchfiel Geography Building (BGB), Room 405
- **Lecture time:** 2:10 PM - 3:25 PM (T/R)
- **Instructor:** Yingjie Hu, Ph.D.
- **Instructor Office hours:** 1:15 PM – 2:15 PM (M/W)
- **Instructor Office:** BGB 206A
- **Class email:** utkgis2@gmail.com

Course Description

GEOG 509 will cover the basic techniques for mining and analyzing geo-text data. The goal is to discover geospatial and semantic knowledge and to support decision making. We will start from the emergence of geo-text data in the context of big data, continue with basic and relatively more advanced techniques for mining geo-text data, and finish with basic visualization methods to represent the discovered knowledge. You will learn concepts about geo-text data as well as hands-on skills on analyzing such data using software packages and tools, such as R and Stanford NLP toolbox.



Central Learning Objectives

Students who have successfully completed the course should be able to:

- Know the basic concepts of mining geo-text data.
- Know a number of sources for collecting geo-text data.
- Use existing software tools to perform basic NLP.
- Use existing software tools to perform basic geospatial analyses.
- Visualize the discovered knowledge as maps and data plots.

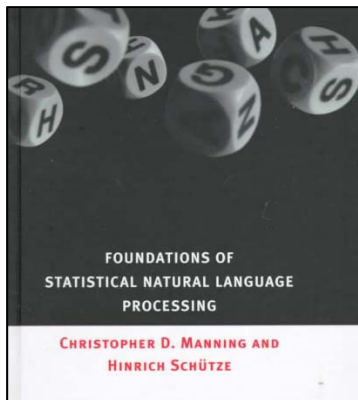
Degree Learning Objectives (for Geography Majors)

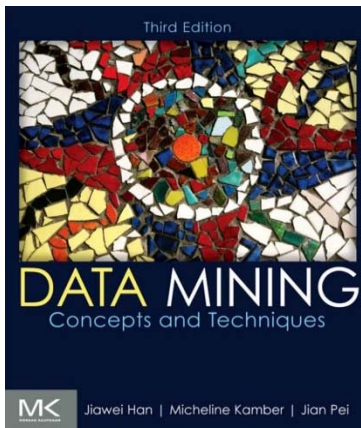
Upon successful completion of this course, students should be able to:

- Apply current innovations in geographic information science (GIS) to spatially analyzing social and environmental issues.
- Demonstrate competency in conducting geographic research design, data collection, and quantitative/qualitative analysis.
- Apply knowledge and skills in one or more specialty areas offered by the department to address specific geographical problems.
- Students can identify job opportunities in the public, private, and academic sectors. Additionally, students will be prepared to explore these opportunities.

Textbooks (Optional)

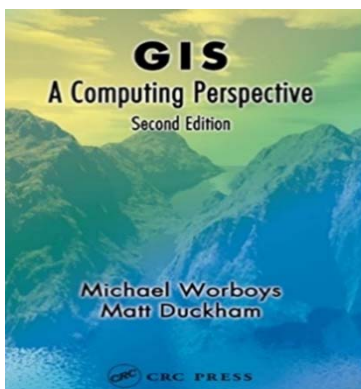
There is no single textbook that systematically introduces geospatial semantics (I hope to write one some year later). Accordingly, we will use book chapters here and there, and some chapters are from the three textbooks as below.





- Manning, C. D., & Schütze, H. (1999). Foundations of statistical natural language processing, Cambridge: MIT press. (I will refer to this book as **NLP** later)
- Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier. (**DM**)
- Worboys, M. F., & Duckham, M. (2004). GIS: a computing perspective. CRC press. (**GIS**)

Notes on Reading Assignments: The textbooks will be used as general guidelines for this course. Lectures and labs will NOT follow the textbooks exactly. Additional materials other than those covered in the textbooks will be discussed.



Course Website

The website of this course is available at Canvas (<https://utk.instructure.com/>). The lecture notes and other course materials will be uploaded by the instructor.

Prerequisites and Skills

You do not need to have prior knowledge in NLP and GIS to take this course. We will start step by step. However, you should feel comfortable working with computing tools, quantitative methods, and some programming. In addition, you should also be able to:

- Search on the Internet for data and information relevant to the course.
- Download files from Canvas, handle zip/unzip the file for compressed file (.zip., .tar. etc) using compression software (7-Zip, Win-zip)
- Communicate with the instructor in person and through emails
- Use Microsoft Word, Latex, or any other text processing tool for writing lab and final project reports
- Be willing to learn new software tools

Additional requirement: At least one thumb drives (>4GB) to store your lab and project work.

Tentative Course Schedule*

***Note:** this is a tentative course schedule. We may need to adjust it depending on our actual progress.

Week (Month, Dates)	Lecture Topics	Reading Chapters	Lab Sections
Week 1 (Aug. 24)	- Course overview and policy - The emergence of geo-text data	NLP Chapter 1	No Lab
Week 2 (Aug. 29, 31)	- Why mine geo-text data? - Semantics vs. syntax	DM Chapter 1	Lab1: Getting Familiar with Java and Eclipse
Week 3 (Sep. 5, 7)	- Geospatial semantics - Expert driven or data driven?	Hu (2017)	Lab2: Working with Twitter API; Retrieving and mapping geotagged Tweets
Week 4 (Sep. 12, 14)	- Implicit geo-text data - Place name recognition and disambiguation	NLP Chapter 7 Hu (2014)	Lab3: Extracting place names from texts using Stanford NLP kit
Week 5 (Sep. 19, 21)	- An intro to gazetteer - Linking names to coordinates	Hill (2000) Goodchild & Hill (2008)	Lab4: Adding coordinates to place names using DBpedia Spotlight
Week 6 (Sep. 26, 28)	- Explicit geo-text data - Geospatial clustering and segmentation	GIS Chapter 3 DM Chapter 10	Lab5: Getting familiar with R; Geospatial clustering using R
Week 7 (Oct. 3, 5)	- TF-IDF - Finding representative words for geographic areas	Information retrieval Chapter 6 (goo.gl/119CNY)	Lab6: Working with geotagged Flickr data
Week 8 (Oct. 10, 12)	- Examining data from space, time, and semantics perspectives	Hu (2015)	Lab7: Analyzing geotagged Twitter data
Week 9 (Oct. 17, 19)	- Topic modeling: LDA - Text first vs. geo first	Blei (2012)	Lab8: Topic modeling on travel blog data
Week 10 (Oct. 24, 26)	- The emotions of places - Basic Sentiment analysis	Liu (2012)	Lab9: Sentiment analysis on neighborhood review data



Week 11 (Oct. 31, Nov. 2)	- Word cloud visualization - Geo visualization - Network visualization	Lecture notes	Lab10: Visualizing data using R and Gephi
Week 12 (Nov. 7, 9)	- No class or labs - Prepare for your project proposal		Prepare for your final project proposal
Week 13 (Nov. 14, 16)	- Final project proposal presentation		Working on your final project
Week 14 (Nov. 21)	- Working on your final project		* No lab on Nov. 23 (Thanksgiving)
Week 15 (Nov. 28, 30)	- Final project presentation (I) - Final project presentation (II) - Course evaluation		* No lab on Dec. 1 st

* Classes end on Dec. 5.

* **This course does not have a final exam. However, each student should write a final project report that is publishable on Pursuit (<http://trace.tennessee.edu/pursuit/>).** The report should follow the submission guideline of Pursuit (unless you plan to submit to another peer-reviewed journal or conferences), and should be at least 10 pages (12 fonts, double space). **The final report is due Dec. 14th 11:59 pm.**

How will your final grade be calculated?

- Lab Assignments (45% for undergrads; 42% for grads)

There are 10 lab assignments in total. Each assignment is due in one week. **All assignments must be turned in on time. A penalty of 20% of points per day will be applied towards assignments that are turned in after the deadline.** Drop boxes will be made available on Canvas for the submission of lab reports.

- Final Project Presentation (15%)

Each student will make a presentation on his/her final project on the scheduled time.

- Final Project Report (35%)

Each student should write a final project report. The report should follow the submission guideline of Pursuit (unless you plan to submit to another outlet), and should be at least 10 pages (12 fonts, double space). The final report is due Dec. 14th 11:59 pm. **A penalty of 20% of points per day will be applied towards assignments that are turned in after the deadline. No report will be accepted 5 days after the deadline.**



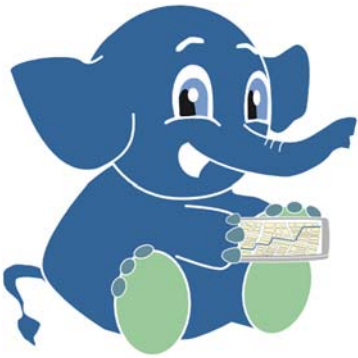
- Attendance (5%)

There will be two random checks on the class attendances, and each is counted as 2.5%.

- Literature Review (3%; graduates only)

Graduate students are required to do a comprehensive literature review related to the final project report he/she writes. The literature review should be at least 1000 words, and should be submitted as a separate document.

Letter Grades



The grade scale is set as follows:

Undergraduate Students:

A	93.0% or higher
A-	between 90.0% and 92.9%
B+	between 87.0% and 89.9%
B	between 83.0% and 86.9%
B-	between 80.0% and 82.9%
C+	between 77.0% and 79.9%
C	between 73.0% and 76.9%
C-	between 70.0% and 72.9%
D+	between 67.0% and 69.9%
D	between 63.0% and 66.9%

D-	between 60.0% and 62.9%
F	59.9% or lower

Graduate Students:

A	93.0% or higher
B+	between 87.0% and 92.9%
B	between 83.0% and 86.9%
C+	between 77.0% and 82.9%
C	between 73.0% and 76.9%
D	between 63.0% and 72.9%
F	62.9% or lower

Note: The instructor reserves the right to move this scale downward, but not upward to better reflect their assessment of the class's work as a whole.

Additional Information

1. Students with disabilities: If you need course adaptations or accommodations because of a documented disability, please contact the Office of Disability Services at 2227 Dunford Hall (telephone/TTY: 865-974-6087; e-mail: ods@utk.edu). This will ensure that you are properly registered for services. It is the student's responsibility to alert the instructor of documented special needs.

2. Policy for make-up exams and quizzes: No make-up exams or quizzes will be allowed except for legitimate reasons, as follows:

- **For serious illness,** the student must submit a doctor's statement.
- **For attending academic conferences,** the student must inform the instructor **2 weeks before the exam** and should submit a letter from a faculty member stating the importance of the conference.



- **For death in the family**, the student should submit a statement written by a parent or a faculty member to prove that the unfortunate event is true.

3. Extra credits: The Department of Geography does not allow individual students to perform extra credit work in order to increase their grade. Any extra credit options must be made available to the entire class.

4. Cell phones, headphones and other electronic devices: Please turn off cell phones during the class time. Headphones and other electronic devices are not permitted in class.

5. Academic dishonesty: Cheating, plagiarism, and other forms of academic dishonesty will not be tolerated.

6. My goal as your instructor: My goal is to make this course as enjoyable and beneficial to you as possible. I'm not here to fail you or to give you a hard time. If you have any questions about anything I've covered in the lecture, please ask me during or after class, or come to see me in my office hours, or simply send an email. The time to seek help from me is before, not after, a test. Don't wait until the last week of classes to ask for help. The earlier you seek help, the better!

7. Instructor's expectation and class environment: It is important to recognize that the classroom is an environment that requires respect for all participants. Therefore, students are expected to conduct themselves in a considerate manner. All participants in the class must respect the classroom environment by being on time, avoiding loud talking and chatting, and refraining from reading non-class material. Since the class is held in a computer lab, no food or drinks are allowed inside.

8. The number of hours you should plan for this course each week. You should plan between 10 - 14 hours each week for this course.