CSCI 599 – Geospatial Data Integration

Professors Craig Knoblock & Yao-Yi Chiang

TA: Bo Wu

Spring 2014

Room: VKC 252

Monday and Wednesdays 2-3:20pm

1. Introduction

This class will cover the theoretical foundations, methods, techniques, and software systems for geospatial data integration. This includes the latest research in a variety of topics that are central to spatial computing, including the geospatial semantic web, geospatial linked data, spatial data mining, geocoding, document linking, location-based services, volunteered geographic-information, feature extraction, layer registration and alignment, and geospatial mashups. Students will also gain a deep understanding and hands-on experience in the software for spatial computing, including geographic information systems (e.g., ArcGIS), online GIS (e.g., Google Earth), semantic web tools, and spatial databases through a combination of homework and projects. Students will learn about the wide variety of geospatial data and services available, including how to find relevant data and transform it as needed so that it can be used for solving specific problems.

2. Course Structure

The course will be taught using a lecture format where the instructor will present the core topics and the students will participate and then give lectures on some of the related topics. There are weekly quizzes to ensure that students keep up with the material and readings. In the first half of the course there are also weekly homework assignments to give students first hand experience with the wide variety of software and systems that can be used for spatial computing. In the second half of the courses, students will form teams and propose and conduct a class project that will give them more depth in one or more course topics. The class will encourage student participation with ample discussion time for reviewing readings, homework, quizzes, and other course material.

3. Course Outline / Readings

The course will be organized around the following topics and the accompanying reading assignments.

Week 1, Jan 13: Introduction to Class (Professor Knoblock)

Jan 15: Geographic Information Systems (Professor Chiang)

Brief introductions with a discussion of class goals, projects, technology, plans, and expectations. Presentation of the basics of geographic information systems with a focus on hands-on use of ArcGIS so that students become familiar with the basic capabilities of such systems.

Readings:

1. Clemmer G (2010) The GIS 20 Essential Skills, ESRI Press, Redlands, CA

Week 2, Jan 22: GIS Data Basics (Professor Knoblock)

These lectures will cover the basics of spatial data, including representations of spatial data, coordinate systems, datums, projections, etc.

Readings:

1. Getting Started with Geographic Information Systems, Keith C. Clarke, Prentice Hall, 2010, Chapters 2 & 3.

Week 3, Jan 27: Online GIS (Professor Knoblock)

Presentation and hands on training with online GIS software with a particular focus on GoogleEarth and Bing Maps.

Readings:

1. Google (2013) Google Earth Tutorials, http://www.google.com/earth/outreach/tutorials/all.html
2. Microsoft (2013) Bing Maps Videos, http://www.microsoft.com/maps/developers/videos.aspx

Week 4, Feb 3: Spatial Databases and Streaming Spatial Data (Professor Chiang)

A discussion of the capabilities of spatial database systems. This topic will include hands-on use of the Postgres PostGIS Spatial Database.

Reading:

Mon:

1. Chapter 1 in Spatial Databases: A Tour (Prentice Hall 2003, ISBN 0-13-017480-7) (you can use the chapter draft on the author's website http://www.spatial.cs.umn.edu/Book/)
2. Chapter 1 in Spatial Database Systems: Design, Implementation and Project Management by Yeung and Hall, 2007, Springer. DOI: 10.1007/1-4020-5392-4_1 (you can find the PDF version of this book on USC Library website)
3. Object-Oriented Representation of Environmental Phenomena: Is Everything Best Represented as an Object? (Bian, 2007)

Wed:

1. Chapter 3 in Spatial Databases: A Tour (Prentice Hall 2003, ISBN 0-13-017480-7)
2. Finish the OpenGeo Tutorial on PostGIS: http://workshops.boundlessgeo.com/postgis-intro/

Week 5, Feb 10: Geospatial Semantic Web (Professor Knoblock)

A discussion of the methods for representing and reasoning about geospatial data using the infrastructure of the Semantic Web. Students will get hands-on experience in using tools for creating and using geospatial semantic data.

Readings:

1. RDF: http://www.w3.org/TR/rdf11-primer/
2. SPARQL: http://www.w3.org/TR/sparql11-query/ (don't worry about the various advanced features, just the basics)
3. GeoSPARQL: http://www.semantic-web-journal.net/sites/default/files/swj176_3.pdf

Week 6, Feb 17: Geospatial Semantic Web (cont.)(Professor Knoblock)

Readings:

1. Design and Development of Linked Data from The National Map, E. Lynn User and Dalia Varanka, Semantic Web Journal, http://www.semantic-web-journal.net/content/design-and-development-linked-data-national-map

Week 7, Feb 24: Geospatial Linked Data (Professor Knoblock)

A discussion of the research and techniques for creating and using geospatial linked data.

Readings:

1. **Watch this video before you come to class:** Tim Berners-Lee TED Talk on Linked Data: https://www.ted.com/talks/tim_berners_lee_on_the_next_web.html.
2. **Chapters 1 and 2** from Tom Heath and Christian Bizer (2011) *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.
3. Rahul Parundekar, Craig A. Knoblock and Jose´ Luis Ambite (2010), Aligning Ontologies of Geospatial Linked Data, http://www.isi.edu/integration/papers/parundekar10-lstd.pdf
4. Exploring the Geospatial Semantic Web with DBpedia Mobile Paper

Optional Readings:

1. Krzysztof Janowicz, Simon Scheider, Todd Pehle, and Glen Hart (2012) Geospatial Semantics and Linked Spatiotemporal Data – Past, Present, and Future, Semantic Web – Interoperability, Usability, Applicability, IOS Press, http://www.semantic-web-journal.net/content/geospatial-semantics-and-linked-spatiotemporal-data-—-past-present-and-future

2. M. Koubarakis, K. Kyzirakos, M. Karpathiotakis, Ch. Nikolaou, M. Sioutis, G. Garbis, K. Bereta (2012) Introduction in stRDF and stSPARQL, http://www.strabon.di.uoa.gr/files/stSPARQL_tutorial.pdf

Week 8, Mar 3: Linking Text to Location (Professor Knoblock)

A discussion of the various approaches for linking textual information to geographic locations.

Required Readings:

1. Craig A. Knoblock. Reduce data overload. Earth Imaging Journal, pages 28–30, March/April 2012. http://www.isi.edu/integration/papers/knoblock12-eij.pdf
2. **STEWARD: Architecture of a spatio-textual search engine.** M. D. Lieberman, H. Samet, J. Sankaranarayanan, and J. Sperling. In *Proceedings of the 15th ACM International Symposium on Geographic Information Systems (ACM GIS'07)*, pages 186-193, Seattle, WA, November 2007. http://www.umiacs.umd.edu/~hjs/pubs/steward-acmgis2007.pdf
3. **Web-a-where: Geotagging Web Content.** Amitay E., Har'El N., Sivan R., Soffer A. (2004). ACM **SIGIR** 2004. http://einat.webir.org/SIGIR_2004_GeoMiner_p273-amitay.pdf
4. **A confidence-based framework for disambiguating geographic terms.** E. Rauch, M. Bukatin, and K. Baker. In Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References, pages 50-54, Edmonton, CA, May 2003. http://portal.acm.org/citation.cfm?id=1119402
5. **Geospatial Mapping and Navigation of the Web** Kevin S. McCurley, Proceedings of the World Wide Web Conference, 2001. http://www10.org/cdrom/papers/pdf/p278.pdf

Optional Readings:

1. **Determining the spatial reader scopes of news sources using local lexicons.** G. Quercini, H. Samet, J. Sankaranarayanan, M. D. Lieberman, In A. El Abbadi, D. Agrawal, M. Mokbel, and P. Zhang, editors, Proceedings of the 18th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pages 43-52, San Jose, CA, November 2010. http://www.cs.umd.edu/~hjs/pubs/acmgis10-gianluca.pdf
2. **Geotagging: Using proximity, sibling, and prominence clues to understand comma groups.** M. D. Lieberman, H. Samet, J. Sankaranarayanan In R. Purves, C. Jones, and P. Clough, editors, Proceedings of 6th Workshop on Geographic Information Retrieval, Zurich, Switzerland, February 2010. http://www.cs.umd.edu/~hjs/pubs/commagroup-GIR.pdf

Week 9, March 10: Extracting Features from Raster Maps (Professor Chiang)

A discussion of the methods for extracting features from scanned raster maps.

Readings:

1. **Harvesting Geographic Features from Heterogeneous Raster Maps**, Y.-Y. Chiang, Ph.D. Thesis, Department of Computer Science, University of Southern California. Chapter 2, pages 12-57
   Thesis
2. **Integrated text and line-art extraction from a topographic map**, L. Li, G. Nagy, A. Samal, S. C. Seth, and Y. Xu. International Journal of Document Analysis and Recognition, 2(4):177-185, 2000.
   Paper
3. **Reviving legacy population maps with object-oriented image processing techniques**, N. Kerle and J. de Leeuw. IEEE Transactions on Geoscience and Remote Sensing, 47(7):2392-2402, 2009. Paper
4. **Toponym recognition in scanned color topo- graphic maps**, J. Pouderoux, J. C. Gonzato, A. Pereira, and P. Guitton. In Proceedings of the Ninth International Conference on Document Analysis and Recog- nition, volume 1, pages 531–535, Sept. 2007. Paper
5. **Colors of the past: color image segmentation in historical topographic maps based on homogeneity**, S. Leyk and R. Boesch. GeoInformatica, 14(1):1-21, 2010. Paper

Week 10, Mar 24: Registering and Aligning Geospatial Layers (Professor Knoblock)

A discussion of techniques for automatically aligning various geospatial layters, including both vector and raster layers.

Readings:

1. **Automatically conflating road vector data with orthoimagery**, Ching-Chien Chen, Craig A. Knoblock, and Cyrus Shahabi. *Geoinformatica*, 10(4):495--530, December 2006.
   Paper
2. **Automatically and Accurately Conflating Raster Maps with Orthoimagery,** Chen, C.; Knoblock, C. A.; and Shahabi, C. 2008. *Geoinformatica*, 12(3):377--410. Paper
3. **Automatic alignment of large-scale aerial rasters to road-maps. X. Wu, R. Carceroni**, H. Fang, S. Zelinka, and A. Kirmse. , In Proceedings of the 15th ACM International Symposium on Advances in geographic information systems, pages 1–8, 2007. Paper
4. **Image registration methods: a survey**, B. Zitova, Image and Vision Computing, Vol. 21, No. 11, 2003, pp. 977-1000. Paper

Week 11, Mar 31: Geocoding (Linking Addresses to Location) (Professor Chiang)

A discussion of the methods and approaches to linking addresses to geographic locations.

Readings:

1. **Exploiting Online Sources to Accurately Geocode Addresses.** Bakshi, R., C.A. Knoblock, and S. Thakkar, 2004, In D. Pfoser, I. F. Cruz, and M. Ronthaler (Eds.), ACM-GIS '04: Proceedings of the 12th ACM International Symposium on Advances in Geographic Information Systems, Washington DC, USA, November, 2004, 194–203.
   Paper

2. **Improving geocode accuracy with candidate selection criteria.** Goldberg, D. W., Cockburn, M. G. (2010). Transactions in GIS. Vol. 14 (S1), pp. 129-146.
   Paper
3. **Toward Quantitative Geocode Accuracy Metrics.** Goldberg, D. W., Wilson, J. P., Cockburn M. G. (2010) In Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. pp. 329-332 Leicester, UK.
   Paper
4. **From text to geographic coordinates: The current state of geocoding. Journal of the Urban and Regional Information Systems Association.** Goldberg, D. W., Knoblock, C. A., Wilson, J. P. (2007). Vol. 19 (1), pp. 33-46.
   Paper
5. **A comparison of address point, parcel and street geocoding techniques**, Paul A. Zandbergen, Computers, Environment and Urban Systems 32 (2008) 214–232
   Paper
6. **A Flexible Addressing System for Approximate Geocoding**, Davis et al.
   Paper

Week 12, Apr 7: Geospatial Mashups (Professor Knoblock)

An introduction to the research, tools, and techniques for building online integrated applications with geospatial data. The focus in this section is on the ability to rapidly compose new applications from available sources and services.

Readings:

1. Jeffrey Wong and Jason I. Hong. 2007. **Making mashups with marmite: towards end-user programming for the web**. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07). ACM, New York, NY, USA, 1435-1444. http://doi.acm.org/10.1145/1240624.1240842
2. Rob Ennals, Eric Brewer, Minos Garofalakis, Michael Shadle, and Prashant Gandhi. 2007. **Intel Mash Maker: Join the web**. SIGMOD Rec. 36, 4 (December 2007), 27-33. http://doi.acm.org/10.1145/1361348.1361355
3. Shubham Gupta and Craig A. Knoblock. **Building geospatial mashups to visualize information for crisis management**, In Proceedings of the 7th International Conference on Information Systems for Crisis Response and Management, 2010.
   Paper

Week 12, Apr 9: Volunteered Geographic Information (VGI) (Professor Knoblock)

Discuss the recent developments in volunteered geographic information (VGI) including the widely used sources, techniques for crowd-sourcing data, and attempts to evaluate the quality of VGI data.

Readings:

1. Bin Jiang, (2012) **Volunteered Geographic Information and Computational Geography: New Perspectives**, http://arxiv.org/pdf/1212.0941v1

2. Michael F. Goodchild, Linna Li, **Assuring the quality of volunteered geographic information**, Spatial Statistics, Volume 1, May 2012, Pages 110-120, http://www.sciencedirect.com/science/article/pii/S2211675312000097

Week 13, Apr 14: Spatial Data Mining and Reasoning (Professor Chiang)

An introduction to some of the techniques for spatial data mining and reasoning.

Readings:

1. Shashi Shekhar, Pusheng Zhang, Yan Huang, and Ranga Raju Vatsavai, **Trends in Spatial Data Mining, Data Mining**: Next Generation Challenges and Future Directions", Hillol Kargupta and Anupam Joshi(editors), AAAI/MIT Press, 2003 PDF
2. **A Framework for Integrating and Reasoning about Geospatial Data.** Gupta, S., and Knoblock, C. A. 2010.In *Extended Abstracts of the Sixth International Conference on Geographic Information Science (GIScience)*. Paper
3. **A Constraint Satisfaction Approach to Geospatial Reasoning.** Michalowski, M., and Knoblock, C. A. 2005. In *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*.
   Paper
4. **Information Fusion for Feature Extraction and the Development of Geospatial Information.** Michael A. O'Brien and John M. Irvine
   Paper
5. **Merging of Heterogeneous Data for Emergency Mapping: Data Integration or Data Fusion.** Florin Savopol and Costas Armenakis
   Paper

Week 14, Apr 21: Location-based Services and Privacy (Professor Knoblock)

Discuss the various features of successful modeling applications, including the need for authenticity (i.e. the evaluation of the model relative to the real system), parsimony (i.e. the desirability of keeping things simple and avoiding unnecessary complications), transparency (i.e. the need for clear documentation and user-friendly organization of both the model and the documentation), and patience (i.e. the fact that it takes time to construct and/or implement a model).

Readings:

1. Location-based services. Iris A. Junglas and Richard T. Watson. 2008. Commun. ACM 51, 3 (March 2008), 65-69. Paper
2. AK Dey, G Abowd (2000) Towards a better understanding of context and context-awareness, CHI 2000 Workshop on The What, Who, Where, When, and How of Context-Awareness
   http://scholar.google.com/citations?view_op=view_citation&hl=en&user=ydA8Q5AAAAAJ&citation_for_view=ydA8Q5AAAAAJ:u5HHmVD_uO8C
3. G Myles, A Friday, N Davies (2003), **Preserving privacy** in **environments** with **location-based applications**, Pervasive Computing, IEEE http://eprints.lancs.ac.uk/12211/2/b1056.pdf

Week 15, Apr 28: Final Presentations (Students)

Students will present their team projects, summarizing their results and what they learned from their projects.

4. Statement on Academic Integrity

USC seeks to maintain an optimal learning environment. General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that individual work will be submitted unless otherwise allowed by an instructor, and the obligations both to protect one's own academic work from misuse by others as well as to avoid using another's work as one's own. All students are expected to understand and abide by these principles. Scampus, the Student Guidebook, contains the Student Conduct Code in Section 11.00, while the recommended sanctions can be found at http://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions/. Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty. The review process can be found at http://www.usc.edu/student-affairs/SJACS/.

5. Academic Accommodations

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. More information about academic accommodations based on a disability can be found at http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered as early in the semester as possible. DSP is located in STU 301 and is open from 8:30 a.m. to 5:00 p.m., Monday to Friday. The phone number for DSP is 213-740-0776.

6. Emergency Preparedness/Course Continuity in a Crisis

In case of a declared emergency if travel to campus is not feasible, USC executive leadership will announce an electronic way for instructors to teach students in their residence halls or homes using a combination of Blackboard, teleconferencing, and other technologies.

7. Course Personnel

Instructors:

Professor Craig A. Knoblock
Information Sciences Institute (ISI 922)
Spatial Sciences Institute (AHF B55D)
310-448-8786
knoblock@usc.edu

Professor Yao-Yi Chiang
Spatial Sciences Institute (AHF B55C)
213-740-7618
yaoyic@usc.edu

8. Course Requirements and Grading Scheme

Students must prepare a lecture, participate in a team project, participate in class discussion, take weekly quizzes, and turn in weekly homework assignments.

Class Presentation (10%): Students will conduct a seminar on a topic determined in consultation with the instructor. Students will be expected to become an expert on that topic and present a short lecture of 30-45 minutes on the topic.

Quizzes (30%): There will be weekly quizzes on the lectures and readings from the previous week. There is no final, so this is the assessment of how well the students have learned the material.

Homeworks (20%): Students will be assigned weekly homework during the first half of the course.

Team Project (40%): In the second half of the course, students will work in teams on projects determined in consultation with the instructor. The team will propose their own projects based on the topics covered in class.