

SI 570 - Semantics-based Knowledge Description and Organization

Term: Winter, 2019
Instructor: Graham Hukill
Office: Phone and email only
Phone: 206.375.2061
Email: gshukill@umich.edu
Office Hours (phone, [online chat](#)):
Thursdays, 7-9pm

Meeting Time: Mondays, 5:30-8:30pm
Location: NQ 2185
Credits: 3
Prerequisite: SI 502

Table of Contents

- [Description](#)
- [Objectives](#)
- [Accommodations for Students with Disabilities](#)
- [Readings](#)
- [Software](#)
- [Assignments](#)
 - [In-Class Labs](#)
 - [Final Project](#)
- [Grading](#)
- [Course Schedule](#)

Description

The spectacular growth in the volume and variety of online content in all forms - text, images, movies, data, etc. - has raised new challenges for information management, organization, description, and curation. Many of the assumptions underlying traditional models for organizing information are inappropriate for the distributed, dynamic, and highly heterogeneous Internet context. Almost two decades of research and development of automated crawling, indexing, and querying techniques have remarkably improved the organization and discovery of massive amounts of networked textual data. Nevertheless, these techniques have limited effectiveness when applied across the multiple genre of resources (e.g., images, video, scientific data) and when “information or knowledge discovery”, rather than “data discovery” is the goal. In response, a wide variety of knowledge organizations, institutions, and companies are increasingly turning towards semantic-based technologies and tools to impose some “order over chaos” on the Internet information context. These techniques permit the expression of information and knowledge models, their component entities, the agents and activities involved in their creation and management, and the relationships among those entities. The result is an information environment that is both responsive to human inquiry and discovery and available to machine-based algorithmic methods for information analysis, mining, and knowledge extraction.

This course will cover the foundations, tools, and application of semantic-based techniques for organizing networked information. Because networked information touches all aspects of our lives and underlies all aspects of the iSchool curriculum, this course should appeal to all students in the MSI program. Content of the course will mix theoretical underpinnings with a strong technical component, in the end giving students the foundation to deploy semantic technologies in multiple practical contexts.

Objectives

- Identify and understand the importance of semantic tools and techniques for managing human- and machine-usable knowledge in the contemporary networked information context.
- Understand and attain proficiency with foundations of information/object/data modeling.
- Understand and attain proficiency with the XML technology stack (e.g., syntax, schema, namespaces) and its utility and appropriateness for representing and manipulating semi-structured information.
- Understand the role, utility, and limitations of metadata in the networked information context.
- Understand the role of vocabularies and ontologies for structuring information and inferring knowledge.
- Understand and attain proficiency with the semantic web technology stack (e.g, RDF, RDFs, SPARQL, OWL, SKOS) and its utility for expressing semantics.
- Understand the technologies for disseminating semantics in the networked information context including linked data, RDFa, Microdata and JSON markup.

Accommodations for Students with Disabilities

The University Faculty Senate (SACUA) in 2006 endorsed the following language for inclusion on course syllabi: If you think you need an accommodation for a disability, please let me know at your earliest convenience. Some aspects of this course, the assignments, the in-class activities, and the way we teach may be modified to facilitate your participation and progress. As soon as you make me aware of your needs, we can work with the Office of Services for Students with Disabilities (SSD) to help us determine appropriate accommodations. SSD (734-763-3000; <https://ssd.umich.edu/>) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. I will treat any information you provide as private and confidential.

Readings

We will be using three primary texts for the semester. We will cover most of the material in these three books and supplement them with additional readings and tutorials, all of which are available for free online.

- **TDO** - R. J. Glushko, (2016). *The Discipline of Organizing, 4th professional edition*. Sebastipool, CA: O'Reilly, (online from mlibrary or purchase from <http://disciplineoforganizing.org.>)

- available online: <https://mirlyn.lib.umich.edu/Record/014939124>
- **SWWO** - D. Allemang, J. A. Hendler (2011). *Semantic web for the working ontologist: modeling in RDF, RDFS and OWL, 2nd edition*. Amsterdam, Boston: Morgan Kaufmann Publishers/Elsevier. (available online from mlibrary or for purchase from Amazon)
 - available online: <https://mirlyn.lib.umich.edu/Record/012701879>
- **DR** – W. Kent, S. Hoberman (2012). *Data and Reality: A Timeless Perspective on Perceiving and Managing Information in Our Imprecise World, 3rd Edition*. Third edition. Westfield, NJ: Technics Publications, LLC. (available online from mlibrary or for purchase from Amazon)
 - available online: <https://mirlyn.lib.umich.edu/Record/012122377>

Students are expected to have completed the readings in time for the class session with which they are associated. They are expected to be ready to contribute to discussions about the readings and to ask questions about parts of the readings they may not understand. The instructor reserves the right to “cold call” on students during the class session to evaluate comprehension of the readings. These discussions and questions will be a factor in each student’s engagement grade.

Software

We will be using the following software tools throughout the semester:

- Violet UML editor (<http://sourceforge.net/projects/violet>)
- Oxygen XML editor (<http://www.oxygenxml.com>)
- Protégé ontology editor (<http://protege.stanford.edu>)

Details on obtaining this software will be given in class.

Assignments

Assignments for this course fall under two main types: in-class labs and a term-long, final project. You can find them [here on the Assignments page in Canvas](#).

In-Class Labs

Most weeks we will have an in-class exercise / lab related to the topic at hand. These will be worth a significant part of your grade. Time will be allotted in the second half of each class for the week. This will often be enough time to complete the lab, but in instances where it is not, the lab will be due by the following class meeting (one week). Where students are not able to attend class, completing these labs / exercises can be negotiated with the instructor on a case-by-case basis.

Final Project - Sears Catalog Data Challenge

More information about this final project can be found on the [Canvas assignment page](#).

Grading

Grading for this course will emphasize engagement and exploration of topics, as opposed to a mastery of syntax or formatting. While there is an expectation of accurateness and attention to detail for in-class work and the final project, and these things do matter in the domain of

structured data and information, grading will more heavily weight demonstrated intuition and insight into when and how to apply these concepts.

| Assignment | Points |
|---|------------|
| In-class labs / exercises (x10 @ 5pts) | 50 |
| Participation | 10 |
| Final project <ul style="list-style-type: none">project proposal (5pts)project check-in (5pts)project deliverables (30) | 40 |
| Total | 100 |

Course Schedule

Week 1 (January, 14th) - Course Introduction

- Course logistics and review of the syllabus
- Foundations and assumptions
- Semantics and their role in the information environment
- Lab: [Basic Data Modeling](#)
- Readings
 - **TDO**: Ch. 1-3
 - **DR**: Ch. 1,3-6

Week 2 (January, 21st) - No Class

- Happy Martin Luther King Jr. day!

Week 3 (January, 28th) - Introduction to data modeling with the Unified Modeling Language (UML)

- Unified Modeling Language (UML)
- Object variance
 - Functional Requirements for Bibliographic Records (FRBR)
- Lab: [UML diagram](#)
- Readings
 - **TDO**: Ch. 4
 - UML Class Diagram Tutorial (Lucidchart), <https://www.youtube.com/watch?v=UI6lqHOVHic&t=320s>
 - UML Class Diagrams (Dr. Mike Murphy), <https://www.youtube.com/watch?v=5Q0ZM6ODBhw>

Week 4 (February, 4th) - From data to information

- Semistructured data
- Data encoding
 - XML
 - JSON
- Lab: [UML and XML](#)

- Readings
 - Robert J. Glushko and Tim McGrath, Document Engineering, Chapter 2, "XML Foundations" ([available online from mLibrary](#))
 - Text Encoding Initiative. A Gentle Introduction to XML, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>
 - Introducing JSON, <http://json.org>

Week 5 (February, 11th) - Meta-modeling

- **Assignment Due: Final Project, Project Proposal**
- Expressing constraints in a knowledge domain
- Schema languages
- Relax NG
- Lab: [Relax NG Schema](#)
- Readings
 - Van der Vlist, E. (2001). Comparing XML Schema Languages. Retrieved from <https://web.archive.org/web/20161101121702/https://www.xml.com/lpt/a/884/>
 - The Organization for the Advancement of Structured Information Standards [OASIS]. (2001). RELAX NG Tutorial. Retrieved from <http://www.relaxng.org/tutorial-20011203.html>
 - McDonough, J. (2008). Structural Metadata and the Social Limitation of Interoperability: A Sociotechnical View of XML and Digital Library Standards Development. In Basilage The Markup Conference 2008. [doi:10.4242/BalisageVol1.McDonough01](https://doi.org/10.4242/BalisageVol1.McDonough01)

Week 6 (February, 18th) - Metadata: annotating data with semistructured information

- Vocabularies
- Namespaces
- Metadata Encoding and Transmission Standard (METS)
- International Image Interoperability Framework (IIIF)
- Lab: [METS analysis and validation](#)
- Readings
 - **TDO:** Ch 5.
 - Lois Chan and Marcia Zeng, "Metadata Interoperability and Standardization: A Study of Methodology Part I. Achieving Interoperability at the Schema Level", D-Lib Magazine, 12(6), June 2006 <http://www.dlib.org/dlib/june06/chan/06chan.html> ([Canvas link](#))
 - NISO. (2004). Understanding Metadata. Bethesda, MD: NISO Press. ([Canvas link](#))

Week 7 (February, 25th) - Transforming semistructured data

- Crosswalking between metadata formats
- Xpath and XSLT
- Lab: [XSLT](#)
- Readings
 - David Birnbaum, XSLT basics, https://web.archive.org/web/20130127180348/http://dh.obdurodon.org/xslt_basics.html

- W3Schools, XSLT tutorial, https://www.w3schools.com/xml/xsl_intro.asp

Spring Break (March, 4th) - No Class

- Enjoy your break!

Week 8 (March, 11th) - Controlling vocabularies

- Authorities
- Taxonomies
- Thesauri
- Ontologies
- Readings
 - **TDO**: Ch.7, 8.

Week 9 (March, 18th) - Expressing semantics: RDF

- Resource Description Format (RDF)
- **Assignment Due: Final Project, Data Structures and Preliminary Models**
- Lab: [RDF \(part 1\)](#)
- Readings
 - **TDO**: Ch.6.
 - **SWWO**: Ch. 1-3

Week 10 (March, 25th) - Meta-modeling semantics: RDF schema (RDFS)

- RDF Schema (RDFS)
- Simple Knowledge Organization (SKOS)
- Linked Data
- Lab: [RDF \(part 2\)](#)
- Readings
 - **SWWO**: Ch. 6
 - Miles, A., Matthews, B. & Wilson, M., Brickley, D. (2005). SKOS Core: Simple knowledge organisation for the Web. <http://dcpapers.dublincore.org/pubs/article/view/798/794>
 - Miles, A. & Bechhofer, S. (2009). SKOS Simple Knowledge Organization System Reference: W3C Recommendation 18 August 2009. Retrieved November 19, 2017, from <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>

Week 11 (April, 1st) - Ontologies

- Ontology theory and practice
- Ontology deployment: Web Ontology Language (OWL)
- Ontology tools (Protégé)
- Lab: [SKOS Vocabulary](#)
- Readings
 - **SWWO**: Ch. 9, 10
 - Simperl, E, "Reusing ontologies on the Semantic Web: A feasibility study", *Data & Knowledge Engineering* Volume 68, Issue 10, October 2009, Pages 905-925.
 - M. Horridge, Protégé OWL Tutorial, http://mowl-power.cs.man.ac.uk/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf

- Gruber, T. What is an Ontology?: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
- N. F. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," Palo Alta, CA, 2001

Week 12 (April, 8th) - Information to knowledge

- **Guest Lecture and Lab: Dr. Allen Flynn**
- Introduction to reasoning and knowledge inference
- Inferring knowledge from ontologies
- Lab: [exploring Wikidata with SPARQL](#)
- Readings
 - **SWWO**: Ch. 5
 - OWL Reasoning Examples, <http://owl.man.ac.uk/2003/why/latest/>

Week 13 (April, 15th) - Semantic technologies in libraries, museums, and archives

- SPARQL
- Linked Open Data (LOD)
- Compound document models
- Readings
 - Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology, 1(1), 1–136. <http://linkeddatatbook.com/editions/1.0/> (Chap. 1-2, skim Chap. 3)
 - Lagoze C, Van de Sompel H, Johnston P, et al. (2008) ORE User Guide - Primer. Available from: <http://www.openarchives.org/ore/1.0/primer>.

Week 14 (April, 22nd) - Wrap-up and Final Project Presentations

- Final project presentations
- Course Wrap-Up
- **Assignment Due: Final Project, Presentations**
- **Assignment Due: Final Project, Report and Deliverables**