Graduate School of Library and Information Science
Spring 2016

Ontology Development

(LIS590)

Section OD Monday, 5:30 PM–7:30 PM Room 341, LIS Building

Instructor

David Dubin
Office: LIS 330
Office Hours: Tuesdays 1:00–4:00pm, and by appointment (face to face is your instructor's preferred contact method).
Phone: 217–244–3275 (217–BIG–EARL)
Email: ddubin
Web: http://people.lis.illinois.edu/~ddubin/

Required Texts

Tom Heath and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space* (Morgan and Claypool Publishers, 2011).

Dean Allemang and James Hendler. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL* (Morgan Kaufmann, 2011).

Natalya F. Noy and Deborah L. McGuinness. *Ontology Development 101: A Guide to Creating Your First Ontology* (Stanford Knowledge Systems Laboratory, 2001).

Library Resources

Library resources and information are available at the following online addresses:

- University of Illinois Library: http://www.library.illinois.edu/
- Library and Information Science Virtual Library: http://www.library.illinois.edu/lsx/
- Journal and Article Locator: http://search.grainger.uiuc.edu/linker/
- LIS Librarian (email: lislib@library.illinois.edu, phone: 217-333-3804)

Course Description

An introduction to formal ontology focusing on development and implementation issues and contemporary ontology software tools and languages. The ontology editor Protege will be used throughout and the representation of ontologies in W3C semantic web languages RDF(S) and OWL will be emphasized.

This Syllabus

The official syllabus for this course is the SGML version at http://tinyurl.com/gum6ynj. Expressions of the syllabus in other formats are derived from the SGML version. The current SGML version should be consulted to resolve any inconsistencies among other renditions.

Statement of Inclusion

The following expression is adopted from the Chancellor's Commitment Statement of November 2012:

As the state's premier public university, the University of Illinois at Urbana-Champaign's core mission is to serve the interests of the diverse people of the state of Illinois and beyond. The institution thus values inclusion and a pluralistic learning and research environment, one which we respect the varied perspectives and lived experiences of a diverse community and global workforce. We support diversity of world views, histories, and cultural knowledge across a range of social groups including race, ethnicity, gender identity, sexual orientation, abilities, economic class, religion, and their intersections.

Inclusiveness as a topic in LIS590-OD

The models and theories we cover in this class frame the way we understand resources, our professional roles with respect to their stewardship, and the access and preservation problems to which our efforts are directed. Under the right circumstances they can help make our services more inclusive, as we've seen, for example, with the digital encoding of writing systems and attention to web content accessibility issues. But knowledge organization systems can also marginalize, as reforms of library subject and medical classifications have shown. The topics for our class meetings are selected with the aim of encouraging reflection and discussion, but those choices are not socially or culturally neutral. Where we may be missing opportunities to align the class content and conduct with the values expressed in the inclusion statement, students are encouraged to call classmate and instructor attention to those issues.

Disability Statement

To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES you may visit 1207 S. Oak St., Champaign, call 217-333-4603 (V/TTY), or e-mail a message to disability@uiuc.edu.

Basis for Grading and Evaluation

The most important standards for success in a class like this one are the educational goals that students bring to the class. Your instructor hopes that the activities, assignments, and presentations planned for this semester will be instrumental in your achieving the goals you set for yourself. Each exercise and assignment has been selected to provide an experience that will foster your own learning. Do not think that grading and evaluative feedback are meant as assesments of your success or failure in the class: they are provided as an incentive to engage with the material to the best of your ability, and as a diagnostic, to find you're not getting the benefits that the assignments should provide.

Final grades will be calculated as follows:

- Graded homework assignments 80%
- Class Participation: 20%

Evaluative and constructive feedback

Students are entitled to both evaluative and constructive feedback on the assignments. Evaluative feedback reports how well a completed assignment satisfied the requirements for a grade. Constructive feedback provides more detailed criticism of the work, and suggestions for improvement.

It is the policy in this class to provide evaluative feedback privately to students when an assignment is returned after grading. However, constructive feedback will be provided in a public forum (discussions in class or on message boards)

On Adapting the Work of Others

Criteria for grading homework assignments include (but are not limited to) creativity and the amount of original work demonstrated in the assignment. However, students are permitted to use and adapt the work of others, provided that the following guidelines are followed:

- Use of other people's material must not infringe the copyright of the original author, nor violate the terms of any licensing agreement. Know and respect the principles of fair use with respect to copyrighted material.
- Students must scrupulously attribute the original source and author of whatever material has been adapted for the assignment. Summarize (e.g. using source code comments) the changes or adaptations that have been made. Make plain how much of the assignment represents original work.

Submitting Assignments to the Instructor

All assignments must be submitted in machine readable form. The instructors will discuss detailed requirements for file naming, packaging, and submission for each assignment.

Graded Assignments

There will be six graded assignments:

- Conforming FOAF profile in Turtle/N3 (due February 15).
- Serving linked data for humans and non-humans (due March 7).
- Domain analysis and proposal in natural language (due March 28).
- Class, subclass, and property definitions in OWL (due April 11).
- OWL Revision 1: Restrictions and inference (due April 25).
- OWL Revision 2, integration with other class vocabularies (due May 9).

Conforming FOAF profile in Turtle/N3

This assignment is due February 15.

Create a syntactically valid RDF description of yourself using the FOAF vocabulary, and any other RDF vocabularies that you decide are useful for this assignment. Select a URI for yourself that you are authorized to create and administer, such as one in a web-accessible space where you are able to put resources online.

Express the description in the turtle subset of the N3 RDF serialization syntax, and use a validating parser of your choice to confirm that what you have written is conforming RDF. Submit the description along with a short description of any interesting issues or difficulties you encountered and which parser or validation service you used to check the syntax.

Evaluation criteria for this assignment are based only on technical and syntactic correctness, but your instructor will try to provide feedback on semantic and social issues where appropriate.

Serving linked data for humans and non-humans

This assignment is due March 7.

Publish the RDF description of yourself from assignment 1 at a web-accessible address, and make the URI denoting yourself work according to linked data principles:

- An RDF validator directed to the URI should receive a syntactically correct RDF description.
- A Mozilla-compatible browser directed to the URI should receive an HTML page describing you.

You can change the URI denoting yourself if you need to. Submit a short written report on this assignment (1 page) that includes the URI.

Domain analysis and proposal in natural language

This assignment is due March 28.

Prepare a written analysis of the domain you have chosen that includes:

- Scope of the proposed area (1-2 pages)
- Relationships to definitions in current GSLIS vocabulary (including things that should be there but aren't)
- Existing published vocabularies that are candidates for extension (FOAF, bibo, schema.org, etc.)
- Suggestive RDF illustrations using the example.com domain.
- Ideas for classes and properties.

Class, subclass, and property definitions in OWL

This assignment is due April 11.

Publish an OWL vocabulary based on your domain analysis as linked data. URIs for any entity in your schema should resolve by default to the URL for your RDF file in either turtle or rdf/xml. You need not provide an html counterpart for humans, but do use annotation properties like label and comment to document your definitions.

- Import the GSLIS ontology and any others on which you wish to base subclasses or subproperties.
- Define range and domain constraints for the properties.
- Define individual instances for testing purposes, and create property assertions for them that exercise each of your properties at least once.

Confirm that your ontology can be loaded via a URI into Protege. Submit the RDF source file and post the URI for your vocabulary to the class forum.

OWL Revision 1: Restrictions and inference

This assignment is due April 25.

In the last assignment you defined individual instances as part of your vocabulary to illustrate your property and class definitions. For this assignment, propose an information management strategy for your vocabulary that distinguishes between facts that should be asserted and facts that should be deduced. For example, in the current GSLIS ontology facts concerning job appointments are asserted (appointee, appointing organization, job title), but status as a GSLIS faculty member is deduced. Prepare and submit a written explanation of your strategy. It should be between one half page and one page in length.

Revise your vocabulary and your instance illustrations so that when facts are asserted in accordance with your strategy, the appropriate deduced facts are entailed by your class and property definitions (just as with the illustrations included in gslis.ttl). I will confirm the deductions using the Hermit plugin that accompanies our Protege installation.

At least one of your class definitions should employ either an owl:allValuesFrom (Allemang and Hendler, page 227) or an owl:someValuesFrom (Allemang and Hendler, page 225) restriction. The remaining deductions can be accomplished via as many or as few of the following as you wish:

- `owl:allValuesFrom`
- `owl:someValuesFrom`
- `owl:hasValue`
- `rdfs:subClassOf`
- `rdfs:subPropertyOf`
- `owl:inverseOf`
- `rdfs:range`
- `rdfs:domain`

Post an announcement of your revision to the discussion forum, and submit a copy of the revised vocabulary. The same validity and linked data requirements apply as with Assignment 4: URIs for entities in your schema should resolve to the URL for your vocabulary's RDF, but you need not provide an html counterpart for humans. One additional requirement is that in addition to being syntactically valid, your vocabulary must be logically consistent. That is to say, your declarations should not cause the Hermit reasoner to report that your model is inconsistent.

OWL Revision 2: integration with other class vocabularies

This assignment is due May 9.

Make one final revision to your OWL vocabulary, and document it with 1-2 pages of natural language submitted along with the RDF file. Make at least two changes or additions that make your definitions better integrated with one or more classmates' vocabularies. Aim for one change (to your vocabulary) that accommodates the needs of someone else, and one change (to your vocabulary) that makes it easier or possible for you to use a classmate's vocabulary. Include at least two illustrative RDF statements: one that uses instances of your classmate's OWL class and one that uses a property imported from your classmate's vocabulary.

Class Participation

The class participation grade is based on consistent attendance, contribution to in-class and/or online discussions, and providing assistance to classmates outside of class. Please alert the instructor if a classmate has been of help to you outside of class.

Semester Outline

Part I: **Introduction to the class** *January 25*

Ontologies, Linked Open Data

Readings: Syllabus

Part II: **Descriptions and the RDF Model** *February 1*

RDF graphs and serialization

Readings: Allemang and Hendler chapters 1–3

Part III: **RDF Serialization** *February 8*

Readings: RDF Primer; Git documentation

Part IV: **Linked Data** *February 15*

Readings: Heath and Bizer

Part V: **Application Architecture 1** *February 22*

Readings: Best Practice Recipes for Publishing RDF Vocabularies; .htaccess syntax guide

Part VI: **Application Architecture 2** *February 29*

Readings: Allemang and Hendler chapters 4–6

Part VII: **Vocabulary schemas** *March 7*

RDF Schema

Readings: Allemang and Hendler chapters 7–8

Part VIII: **RDFS vs. SKOS** *March 14*

Readings: Allemang and Hendler chapters 9–10; SKOS primer

Part IX: **Spring Break: no meeting** *March 21*

Part X: **Using Protege** *March 28*

Readings: Protege Desktop User Documentation

Part XI: **OWL Part 1** *April 4*

Readings: Allemang and Hendler, chapter 11; Noy and McGuineness

Part XII: **OWL Part 2** *April 11*

Readings: Allemang and Hendler, chapters 12–13

Part XIII: **Cooperation and interoperability** *April 18*

Readings: Allemang and Hendler, chapter 14

Part XIV: **Expressive power, tractable computations, decidability** *April 25*

Readings: Allemang and Hendler, chapter 15

Part XV: **Topical Wrapup** *May 2*

Readings: Allemang and Hendler, chapter 16