INF 385T Data Semantics

Fall 2023 (28680)
Course Information
Course Date and Time: Tuesday 9:30-12:30PM
Location: 1.502
Instructor: Ying Ding
TA:
Office Hour:
Instructor: TBD, or zoom, or by appointment
TA:

Course Description
The current Web has experienced tremendous changes to connect data, people, and knowledge. There are a couple of exciting efforts trying to bring the Web to its full potential. The Semantic Web is one of them which is heavily embedded in the Artificial Intelligence area with the long-term goal to enhance the human and machine interaction by representing data semantics, integrating data silos, and enabling intelligent search and discovery.

This course aims to provide the basic overview of the Semantic Web in general, and data semantics in particular, and how they can be applied to enhance data integration and knowledge inference. Ontology is the backbone of the Semantic Web. It models the semantics of data and represents them in markup languages proposed by the World Wide Web Consortium (W3C). W3C plays a significant role in directing major efforts at specifying, developing, and deploying standards for sharing information. Semantically enriched data paves the crucial way to facilitate the Web functionality and interoperability.

This course contains three parts: Semantic Web language, RDF graph database (i.e., RDF triple store), and its applications. The fundamental part of the course is the Semantic Web languages. It starts from XML and goes further to RDF and OWL. The RDF graph database part introduces different APIs of Jena and its reasoners. The application part showcases current trends on semantic applications.

Prerequisites
Basic knowledge of HTML and XML is desired.

Course Objectives
This course aims to develop a critical appreciation of semantic technologies as they are currently being developed. At the end of this course, students should be able to:
*   sketch the overall architecture of the Semantic Web.
*   identify the major technologies of the Semantic Web and explain their roles.
*   illustrate the design principles of the Semantic Web by applying the technologies.
*   understand certain limitations of the Semantic Web technologies, and be aware of the kinds of services it can and cannot deliver.

Course aims are achieved through:
*   Lectures covers basic knowledge of the Semantic Web
*   Projects applying semantic technologies to concrete problems of information delivery and use

- Assignments of practicing and utilizing key semantic technologies

Course Materials

This course is developed mainly based on materials from w3schools and W3C

Recommended books for this course:

- Allemang, D., & Hendler, J. (2011). Semantic Web for the working ontologist. 2nd Edition, Morgan&Kaufmann Publisher. [ISBN:978-0-12-385965-5]
- Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. Morgan&Claypool Publisher. (http://linkeddatabook.com/editions/1.0/)
- Daconts, M.C., Orbst, L.J., & Smith, K.T. (2003). The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management. New York: Wiley. [ISBN: 0-471-43257-1]
- Antoniou, G., & van Harmelen, F. (2004). A Semantic Web Primer. Cambridge, MA: MIT Press. [ISBN: 0-262-01210-3]
- Powers, S. (2003). Practical RDF. Sebastopol, CA: O'Reilly. [ISBN: 0-596-00263-7]
- Russell, S., & Norvig, P. (2003). Artificial Intelligence: A Modern Approach. 2nd Edition. Pearson Education.

Software

- Oxygen XML editor
- Protege Ontology editor: available at protege website
- Jena 4
- Instructions on installing Java and Jena on Windows or Mac

Dataset

XML dataset

- https://catalog.data.gov/dataset?res_format=XML
- https://www.nlm.nih.gov/databases/dtd/index.html
- https://www.nlm.nih.gov/bsd/licensee/elements_descriptions.html

RDF dataset

- https://www.bl.uk/collection-metadata/downloads
- https://id.nlm.nih.gov/mesh/
- https://catalog.data.gov/dataset?res_format=RDF&_res_format_limit=0
- https://id.loc.gov/
- https://makg.org/rdf-dumps/

RDF Schema

- https://www.w3.org/2001/sw/BestPractices/WNET/wn-conversion.html

OWL

- https://protegewiki.stanford.edu/wiki/Protege_Ontology_Library

Ontology

- http://www.obofoundry.org/
- https://protegewiki.stanford.edu/wiki/Protege_Ontology_Library
- https://www.bioontology.org/

- https://www.w3.org/wiki/Good_Ontologies
- http://www.loc.gov/standards/premis/ontology/
- https://www.loc.gov/standards/mods/rdf/
- "   Sparql:
- https://id.nlm.nih.gov/mesh/query
- http://data.nobelprize.org/snorql/?_ga=2.257687311.549967076.1624225478-980083378.1624225478
- Running Sparql on Python: https://towardsdatascience.com/how-to-build-your-own-datasets-or-corpora-from-wikipedia-3eb35d78baac

Schedule (Tentative)

Date	Lecture	Lab/Tutorial	Notes
Semantic Web Language			
L1-8/22	Introduction, XML	XML, looking at the examples and write your first XML file.	Knowing each other
L2-8/27	DTD	work on your first DTD example, working on DTD exercise	Form a group
L3 – 9/5	XML Schema	work on your first XML Schema example, work on XML Schema exercise.	Form a group
L4 – 9/12	XPath, XQuery	work on XPath+XQuery exercise	Form a group
L5 – 9/19	RDF	Turtle, work on your RDF example, work on RDF exercise	**XML Tutorial Due**
L6 – 9/26	RDF Schema, OWL	Linked Open Data, work on RDFS exercise, OWL exercise.doc	Group Project (send title and team info to me)
Ontology			
L7 – 10/3	ontology	Protege Lab	**RDF Tutorial Due**
L8 – 10/10	Protégé Lab	Protege Lab, finish pizza.owl; Select one ontology and implement it in Protégé: DublinCore (https://www.dublincore.org/specifications/dublin-core/dcmi-terms/)	Group Project
Jena – RDF Triple Store			
L9 – 10/17	Jena RDF API	Set up Eclipse and set up Apache Jena 4 on Eclipse	**Protégé Tutorial Due**
L10 – 10/24	Jena Ontology API	Jena Reasoner API	Group Project
L11 – 10/31	Sparql, Sparql by Example	Work on your Jena RDF API assignment	Work on your assignment
L12 – 11/7	Jena Sparql API	Work on your Jena RDF API assignment	**Self-learning Tutorial Due**
L13 - 11/14	Jena Examples	Jena Examples	**Jena RDF API Due**

L14-11/28	Data Search	run data search tutorial by yourself	**Jena Example Due**
L15-12/5	**Final Group Project Presentation**		**Final Group Project Report due**

Assignments and Grading

Individual Assignments (50%)

1. XML Tutorial (10%): Please show one XML example, with its schema in D2D and XML Schema, also add 5 queries using XPath or XQuery. Please use Oxygen to show all the XML files are well-formed and validated (powerpoint slides showing XML file, D2D, and XML Schema, XQuery, Oxygen screenshots, and XML/D2D/XML Schema files)

2. RDF Tutorial (10%): Please show one RDF file, with its schema written in RDF Schema, show the validation and generated triples (powerpoint slides showing RDF file, RDF Schema file, screenshots with validation and generated triples, and RDF/RDF Schema files)

3. Protégé Tutorial (10%): Please develop a tutorial on using Protégé to represent one ontology (powerpoint slides showing steps, final OWL file)

4. Jena RDF API (10%): Use RDF API to read one RDF file in and do some basic searches (powerpoint slides, code).

5. Jena Example (10%): Develop a Jena example (powerpoint slides, code).

Group Assignment (15%): Self-learning tutorial about Data Semantics (powerpoint slides, code, and video)

- Instructor will pair two students to form a group
- Your group can select a topic to develop a tutorial about any topics related to data semantics, for example: 1) knowledge graph, 2) graph deep learning (e.g. GNN, GraphTransfomer), 3) introduce new Graph Database (e.g., Neo4J, TigerGraph), 4) Graph visualization, 5) latest development of knowledge graph and ChatGPT, or more.
- Peer learning together: Another team will be assigned to learn the materials you have developed, provide feedback, and grade your tutorial
- The best to way to learn is to be able to teach others: create slides and videos, and upload them to Canvas.

Group Project (25%): Form a group to work on the following topics (select one):

- to develop one tutorial to introduce a graph deep learning method (e.g., explain the method, show the details on how the method can be applied in one graph, and show the results. Examples: node2vec, TransE, GraphSage, Heterogeneous Graph Transformer)
- to develop one tutorial to apply semantic web technologies to solve a specific issue (e.g., library science, healthcare, search/reason on ontology and RDF data)
- to write a survey article on the latest research development of semantic web technologies and its applications (e.g., ISWC, WWW)

- to work on a research article (no need to finish this by the end of the semester, talk to instructor)
- Output:
  - Tutorial: a 3-5 page report (aiming for a workshop paper), powerpoint slides, code, and 5-10mins of video (assuming that you will teach it to other students), presentation
  - Survey/research: a 10-15 page report (aiming for a conference paper), powerpoint slides, 5-10mins of video, presentation

Class discussion, participation, and final presentation (10%)

Grading Scale
This class employs a plus/minus grading system. For more information on this system, please consult the University's *General Information Catalog*. Below is the grade scale the University employs and which will be used in this class:

Meets major requirement		Does not meet requirement	
Grade	Points	Grade	Points
A	≥93.00	C-	70.00-72.99
A-	90.00-92.99	D+	67.00-69.99
B+	87.00-89.99	D	63.00-66.99
B	83.00-86.99	D-	60.00-62.99
B-	80.00-82.99	F	<60.00
C+	77.00-79.99		
C	73.00-76.99		

Land Acknowledgement
We acknowledge that the iSchool sits on indigenous land. We would like to acknowledge that we are meeting on Indigenous land. Moreover, we would like to acknowledge and pay our respects to the Carrizo & Comecrudo, Coahuiltecan, Caddo, Tonkawa, Comanche, Lipan Apache, Alabama-Coushatta, Kickapoo, Tigua Pueblo, and all the American Indian and Indigenous Peoples and communities who have been or have become a part of these lands and territories in Texas, here on Turtle Island.

Title IX
Title IX and Texas SB 212 Reporting requirements: Title IX is a federal law that protects against sex and gender-based discrimination, sexual harassment, sexual assault, unprofessional or inappropriate conduct of a sexual nature, dating/domestic violence and stalking at federally funded educational institutions. UT Austin is committed to fostering a learning and working environment free from discrimination in all its forms. For more information about reporting options and resources, visit http://www.titleix.utexas.edu/, contact the Title IX Office via email at titleix@austin.utexas.edu, or call 512-471-0419.

University Resources for Students
The university has numerous resources for students for your learning, use these to help you succeed in your classes

- The University Writing Center: The University Writing Center offers free, individualized, expert help with writing for any UT student, by appointment or on a drop-in basis. Consultants help students develop strategies to improve their writing. The assistance we provide is intended to foster students' resourcefulness and self-reliance. http://uwc.utexas.edu/
- Counseling and Mental Health Center: Students who are struggling for any reason and who believe that it might impact their performance in the course, please reach out to Bryce Moffett, LCSW (iSchool CARE counselor) at 512-232-2983. If immediate mental health assistance is needed, call the Counseling and Mental Health Center (CMHC) at 512-471-3515. Outside CMHC business hours (8a.m.-5p.m., Monday-Friday), contact the CMHC 24/7 Crisis Line at 512-471-2255
- Basic Needs Security: Any student who faces challenges of affording groceries or accessing sufficient food to eat every day, or who lacks a safe and stable place to live, and believe this may affect their performance in the course, is encouraged to contact the Office of the Dean of Students – Student Emergency Services (SES) for support. Please notify the instructor if you are comfortable doing so and I will make efforts to provide any resources that I may possess to help you navigate issues of food insecurity or residential displacement.
- SES Concerns and Emergencies: http://deanofstudents.utexas.edu/emergency/concernsemergencies.php
- SES Food Pantry: http://deanofstudents.utexas.edu/emergency/pantry.php
- SES Confidential Advocacy and Support: http://deanofstudents.utexas.edu/emergency/advocacysupport.php
- Student Emergency Services: http://deanofstudents.utexas.edu/emergency/
- ITS: Need help with technology? http://www.utexas.edu/its/
- Libraries: Need help searching for information? http://www.lib.utexas.edu/
- Canvas: Canvas help is available 24/7 at https://utexas.instructure.com/courses/633028/pages/student-tutorials
- Behavior Concerns Advice Line (BCAL): If you are worried about someone who is acting differently, you may use the Behavior Concerns Advice Line to discuss by phone your concerns about another individual's behavior. This service is provided through a partnership among the Office of the Dean of Students, the Counseling and Mental Health Center (CMHC), the Employee Assistance Program (EAP), and The University of Texas Police Department (UTPD). Call 512-232-5050 or visit http://www.utexas.edu/safety/bcal.

University of Texas Core Values and Honor Code: The core values of The University of Texas at Austin are learning, discovery, freedom, leadership, individual opportunity, and responsibility. Each member of the university is expected to uphold these values through integrity, honesty, trust, fairness, and respect toward peers and community. As a student of the University of Texas at Austin, I shall abide by πthe core values of the University and uphold academic integrity.

COVID Information

Please visit https://protect.utexas.edu