INF 385T Data Semantics

Fall 2024 (27805)
Course Information
Course Date and Time: Monday 9-12PM
Location: UTA1.212
Instructor: Ying Ding

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 pave 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- 8/26	Introduction, XML	Course overview	Knowing each other
L2- 9/2	DTD	working on DTD exercise	
L3 – 9/9	XML Schema	work on XML Schema exercise.	Form group for group project
L4 – 9/16	XPath	work on XPath+XQuery exercise	
L5 – 9/23	XQuery	work on XPath+XQuery exercise	
L6 – 9/30	RDF	LLM Tutorial: Langchain I Turtle, work on RDF exercise	**XML Tutorial Due**
Ontology			
L7 – 10/7	RDF Schema	LLM Tutorial: Langchain II work on RDFS exercise,	Group Project (send title and team info to me)
L8 – 10/14	OWL	Protege Lab, finish pizza.owl, work on OWL exercise.doc **Group Project (Mid-term feedback)**	**RDF Tutorial Due**
Jena – RDF Triple Store			
L9 – 10/21	Jena RDF API	Set up Eclipse and set up Apache Jena 4 on Eclipse	**Protégé Tutorial Due,** Group Project
L10 – 10/28	Jena Ontology API	Jena Reasoner API	Group Project
L11 – 11/4	Sparql,	LLM Tutorial: AutoGen I	**Jena RDF API Due**
L12 – 11/11	Jena Sparql API	LLM Tutorial: AutoGen II	Group Project
L13 - 11/18	Jena Examples	Group project	Group Project
11/25	Fall Break		
L14- 12/2	**Final Group Project Presentation**		**Jena Example Due**
L15- 12/9			**Final Group Project Report due**

LLM Tutorial:

- LangChain: https://github.com/gkamradt/langchain-tutorials
- AutoGen: https://microsoft.github.io/autogen/

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 Sparql queries (10%): Use one RDF file and develop 5 Jena Sparql queries (powerpoint slides, code).

Group Project (40%): Self-learning tutorial about Data Semantics or Generative AI (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 or generative AI, for example: 1) prompt engineering; 2) Langchain; 3) Autogen; 4) Graph-Toolformer (https://github.com/jwzhanggy/Graph_Toolformer), 4) graphRAG, 5) introduce new Graph Database (e.g., Neo4J, TigerGraph), 4) Graph visualization, 5) applications 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.

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

Late Work and Making Up Missed Work

Late work without proper excuse is not accepted. Each late work is subject to a 1-point grade reduction for every one hour late.

Absences

Each absence in class without proper excuse will get 1-point grade reduction from your final grade.

Equitable Accommodation

If you want to improve your grade, adding extra work to your assignments can be considered. Please discuss this with your instructor.

Grade Breaks

Grade	Cutoff
A A-	94% 90%
B+ B B-	87% 84% 80%
C+ C C-	77% 74% 70%
D+ D D-	67% 64% 60%
F	<60%

Course Policies and Disclosures

Academic Integrity Expectations

Students who violate University rules on academic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Since such dishonesty harms the individual, all students, and the integrity of the University, policies on academic dishonesty will be strictly enforced. For further information, please visit the Student Conduct and Academic Integrity website at: http://deanofstudents.utexas.edu/conduct.

Getting Help with Technology

Students needing help with technology in this course should contact the ITS Service Desk or email iSchool Helpdesk <help@ischool.utexas.edu> for software, hardware, or other technology available at our iSchool.

Content Warning

Our classroom provides an open space for the critical and civil exchange of ideas. Some readings and other content in this course will include topics that some students may find offensive and/or traumatizing. I'll aim to forewarn students about potentially disturbing content and I ask all students to help to create an atmosphere of mutual respect and sensitivity.

Religious Holy Days

By UT Austin policy, you must notify me of your pending absence as far in advance as possible of the date of observance of a religious holy day. If you must miss a class, an examination, a work assignment,

or a project in order to observe a religious holy day, you will be given an opportunity to complete the missed work within a reasonable time after the absence.

Names and Pronouns

Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. I will gladly honor your request to address you by your chosen name and by the gender pronouns you use. Class rosters are provided to the instructor with the student's chosen (not legal) name, if you have provided one. If you wish to provide or update a chosen name, that can be done easily at this page, and you can add your pronouns to Canvas.

Land Acknowledgment

I would like to acknowledge that we are meeting on the Indigenous lands of Turtle Island, the ancestral name for what now is called North America. Moreover, I would like to acknowledge the Alabama-Coushatta, Caddo, Carrizo/Comecrudo, Coahuiltecan, Comanche, Kickapoo, Lipan Apache, Tonkawa and Ysleta Del Sur 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.

University Resources for Students

SERVICES FOR STUDENTS WITH DISABILITIES (SSD)

The university is committed to creating an accessible and inclusive learning environment consistent with university policy and federal and state law. Please let me know if you experience any barriers to learning so I can work with you to ensure you have equal opportunity to participate fully in this course. If you are a student with a disability, or think you may have a disability, and need accommodations please contact Services for Students with Disabilities (SSD). Please refer to SSD's website for contact and more information: http://diversity.utexas.edu/disability/. If you are already registered with SSD, please deliver your Accommodation Letter to me as early as possible in the semester so we can discuss your approved accommodations and needs in this course.

COUNSELING AND MENTAL HEALTH CENTER (CMHC)

All of us benefit from support during times of struggle. Know you are not alone. If you or anyone you know is experiencing symptoms of stress, anxiety, depression, academic concerns, loneliness, difficulty sleeping, or any other concern impacting your wellbeing – you are strongly encouraged to connect with CMHC. The Counseling and Mental Health Center provides a wide variety of mental health services to all UT students including crisis services, counseling services with immediate support and well-being resources. Additionally, CARE Counselors are located within the academic schools and colleges. These counselors get to know the concerns that are unique to their college's students. For more information on CMHC, visit https://cmhc.utexas.edu or call 512-471-3515.

UNIVERSITY HEALTH SERVICES (UHS)

Your physical health and wellness are a priority. University Health Services is an on-campus high-quality medical facility providing care to all UT students. Services offered by UHS include general medicine, urgent care, a 24/7 nurse advice line, women's health, sports medicine, physical therapy, lab and radiology services, COVID-19 testing and vaccinations and much more. For additional information, visit https://healthyhorns.utexas.edu or call 512-471-4955.

SANGER LEARNING CENTER

Did you know that more than one-third of UT undergraduate students use the Sanger Learning Center each year to improve their academic performance? All students are welcome to take advantage of Sanger Center's classes and workshops, private learning specialist appointments, peer academic coaching, and tutoring for more than 70 courses in 15 different subject areas. For more information, please visit https://ugs.utexas.edu/slc or call 512-471-3614 (JES A332)."

STUDENT EMERGENCY SERVICES (SES)

Student Emergency Services in the Office of the Dean of Students helps students and their families during difficult or emergency situations. Assistance includes outreach, advocacy, intervention, support, and referrals to relevant campus and community resources. If you need to be absent from class due to a family emergency, medical or mental health concern, or academic difficulty due to crisis or an emergency situation, you can work with Student Emergency Services. SES will document your situation and notify your professors. If immediate mental health assistance is needed, call the Counseling and Mental Health Center (CMHC) at 512-471-3515 or you may also contact Bryce Moffett, LCSW (iSchool CARE counselor) at 512-232-2983. Outside CMHC business hours (8a.m.-5p.m., Monday-Friday), contact the CMHC 24/7 Crisis Line at 512-471-2255. Additional information is available at https://deanofstudents.utexas.edu/emergency/ or by calling 512-471-5017.

Important Safety Information

If you have concerns about the safety or behavior of fellow students, TAs or professors, contact BCCAL (the Behavior Concerns and COVID-19 Advice Line) at https://safety.utexas.edu/behavior-concerns-advice-line or by calling 512-232-5050. Confidentiality will be maintained as much as possible, however the university may be required to release some information to appropriate parties.

CLASSROOM SAFETY AND COVID-19

To help preserve our in-person learning environment, the university recommends the following.

- Adhere to university mask guidance. Masks are strongly recommended, but optional, inside university buildings for vaccinated and unvaccinated individuals, except when alone in a private office or single-occupant cubicle.
- Vaccinations are widely available, free and not billed to health insurance. The vaccine will help protect against the transmission of the virus to others and reduce serious symptoms in those who are vaccinated.
- Proactive Community Testing remains an important part of the university's efforts to protect our community. Tests are fast and free.
- We encourage the use of the Protect Texas App each day prior to coming to campus.
- If you develop COVID-19 symptoms or feel sick, stay home and contact the University Health Services' Nurse Advice Line at 512-475-6877. If you need to be absent from class, contact Student Emergency Services and they will notify your professors. In addition, to help understand what to do if you have been had close contact with someone who tested positive for COVID-19, see this University Health Services link.
- Behavior Concerns and COVID-19 Advice Line (BCCAL) remains available as the primary tool to address questions or concerns from the university community about COVID-19.
- Students who test positive should contact BCCAL or self-report (if tested off campus) to University Health Services.
- Visit Protect Texas Together for more information.

TITLE IX DISCLOSURE

Beginning January 1, 2020, TexasSenate Bill 212 requires all employees of Texas universities, including faculty, to report any information to theTitle IX Office regarding sexual harassment, sexual assault, dating violence and stalking that is disclosed to them. Texas law requires that all employees who witness or receive any information of this type (including, but not limited to, writing assignments, class discussions, or one-on-one conversations) must be report it. If you would liketo speak with someone who can provide support or remedies without making an official report to the university, please email advocate@austin.utexas.edu. 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. Although graduate teaching and research assistants are not subject to Texas Senate Bill 212, they are still mandatory reporters under Federal Title IX laws and are required to report a wide range of behaviors we refer to as sexual misconduct, including the types of sexual misconduct covered under Texas Senate Bill 212.The Title IX office has developed supportive ways to respond to a survivor and compiled campus resources to support survivors.

Faculty members and certain staff members are considered "Responsible Employees" or "Mandatory Reporters," which means that they are required to report violations of Title IX to the Title IX Coordinator. I am a Responsible Employee and must report any Title IX-related incidents that are disclosed in writing, discussion, or one-on-one. Before talking with me or with any faculty or staff member about a Title IX-related incident, be sure to ask whether they are a responsible employee. If you want to speak with someone for support or remedies without making an official report to the university, email advocate@austin.utexas.edu For more information about reporting options and resources, visit the Title IX Office or email titleix@austin.utexas.edu.

CAMPUS SAFETY

The following are recommendations regarding emergency evacuation from the Office of Campus Safety and Security, 512-471-5767,

- Students should sign up for Campus Emergency Text Alerts at the page linked above.
- Occupants of buildings on The University of Texas at Austin campus must evacuate buildings when a fire alarm is activated. Alarm activation or announcement requires exiting and assembling outside.
- Familiarize yourself with all exit doors of each classroom and building you may occupy. Remember that the nearest exit door may not be the one you used when entering the building.
- Students requiring assistance in evacuation shall inform their instructor in writing during the first week of class.
- In the event of an evacuation, follow the instruction of faculty or class instructors. Do not re-enter a building unless given instructions by the following: Austin Fire Department, The University of Texas at Austin Police Department, or Fire Prevention Services office.
- For more information, please visit emergency preparedness.