Introduction to Semantic Web

Dr. Knarig Arabshian
Knarig.arabshian@hofstra.edu

Who Am I?

Dr. Knarig Arabshian

kuh-NAH-rig uh-RAB-shyun

How to address me:
1) Dr. Arabshian
2) Professor Arabshian
3) Professor

Who Am I?

- Background:
  - Grew up in Middle Village, Queens
  - Discovered Computer Science in an AP High School Class and never went back
  - NYC educated: ➔
    - B.S from Queens College, CUNY
    - M.S., PhD from Columbia University
  - 5+ years as Member of Technical Staff at Bell Laboratories first in Antwerp, Belgium and then Murray Hill, NJ
  - Adjunct Professor at Columbia University

Who Am I?

- Research Interests
  - Semantic Web
  - Service Discovery
  - Context-aware Computing
  - Distributed Systems
- Leisurely Interests:
  - Playing piano
  - Classical music
  - Literature
  - NYC Culture

Who are you?

- Name
- Where are you from?
- Why Semantic Web?
- Fun Fact

What is this class about?

Semantic Web

- Ontology Languages: RDF, RDFS, OWL
- Ontology Design and Management using the Protege editor
- Ontology Reasoning with Pellet
- Ontology Querying with SPARQL
- Ontology Programming with Jena API
- Current Applications of the Semantic Web

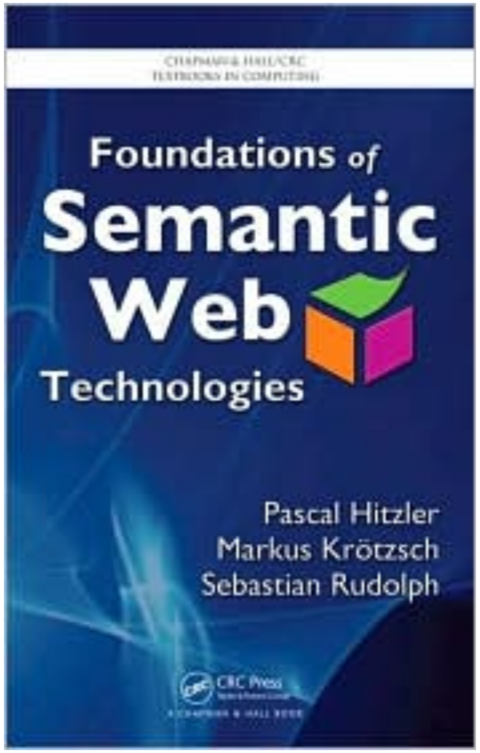
Prerequisites

Prerequisites:

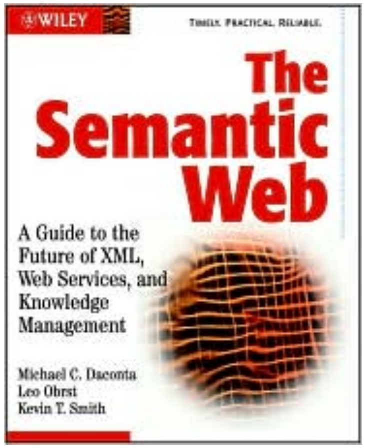
- Fluent in Java programming
- Knowledge of data structures and algorithms
- Database systems helps

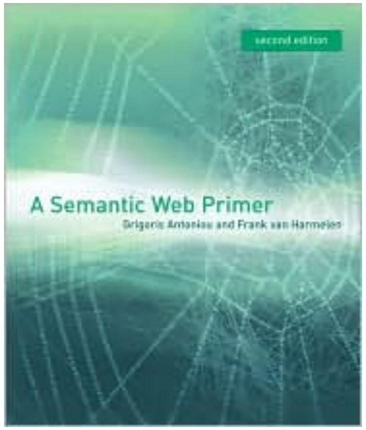
Textbooks

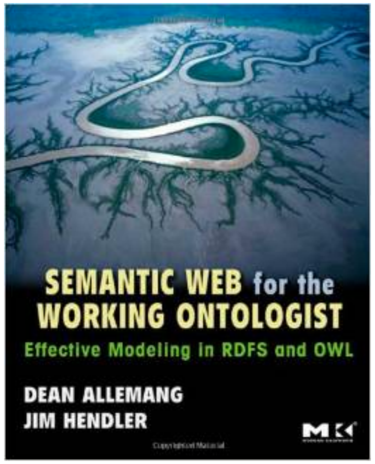
Mandatory



Optional







Course Administration

- Instructor: Dr. Knarig Arabshian
  - Email: knarig.arabshian@**hofstra.edu**
  - **Office:** Adams 115
  - Office Hrs: T/Th 12:40 – 2:10 PM
- Website: http://www.cs.hofstra.edu/~knarig/SemanticWeb

Grading Information

- Grade determinates
  - Attendance ~5%
    - "Thinking Question" given first 10 minutes of each class
  - Quizzes (2) (Closed Book) ~10%
  - Assignments(approx 5) ~40%
  - Midterm Exam (Open Book) ~20%
    - Wednesday, March12
  - Final Exam (Closed Book) ~25%
    - (TBA)
- Grades will be posted on Blackboard

Procedure for Grade Change Request

- Email me an explanation as to why the grade was incorrect, with specific details
- If your explanation is correct, I will re-grade
- You have up to THREE DAYS to do this

Grading Policy

- Letter Grades Determined By

A = 91 to 100
A- = 85 to 90

B+ = 80 to 84
B = 75 to 79
B- = 70 to 74

C+ = 65 to 69
C = 60 to 64
C- = 55 to 59

D = 45 to 54,
F = below 45.

Late Penalty

- All Work Due in Blackboard
- Late Penalty: up to THREE days (exceptions only in case of sickness)

1 – day* late	25%
2 – day late	50%
3 – day late	75%
4 – day late	100%

* Day – Business Day (Holiday Excluded, Sick Day with Documents excluded)

Attendance

- Required
- Responsible for all course materials

Cheating

- Is not tolerated on any level
- All assignments, unless specified, are to be done on your own
- I will occasionally pass your code through a program called MoSS (Measure of Software Similarity) which detects software plagiarism
- If caught once, you will receive a grade of zero
- If caught a second time, you will be reported to the Chair
- Third strike: you will fail the class

Homework

- Must be done on your own unless otherwise specified
- Submitted before class begins
- Homework submission via blackboard
- 4-5 Homework Assignments in two parts:
  - Written problems
  - Programming Assignments

Contact

- Best way to contact me is via email knarig.arabshian@hofstra.edu
- Start your email's subject with the string [SemWeb]; otherwise, I will most likely not see it

Resources

- Check the website frequently for relevant links on tutorials, updated syllabus, lectures and assignments
  - http://www.cs.hofstra.edu/~knarig/SemanticWeb
  - Will use this more than courseworks
- Tools:
  - Eclipse IDE
  - Protege Editor
  - Jena API

Tentative Schedule

- Week 1:Introduction & XML
- Week 2: RDF
- Week 3:FreeBase, DBPedia, RDF Schema, OWL
- Week 4: Ontology Theory: OWL & Ontology Engineering/Protege Editor
- Week 5: Ontology Engineering/Protege Editor
- Week 6:OWL Formal Syntax
- Week 7: Sem Web Programming
- Week 8: Midterm Review/Midterm
- Week 9: OWL Reasoning Examples
- Week 10: Querying with Ontologies, SPARQL
- Weeks 11-12: Semantic Web Services
- Week 13: Ontology Rules
- Week 14-15: Conjunctive Queries and Semantic Markup
- Week 16: Linked Data

Definitions of Semantic Web

- Webopedia
  - An extension of the current Web that provides an easier way to find, share, reuse and combine information.
- Wikipedia
  - a group of methods and technologies to allow machines to understand the meaning – or "semantics" – of information on the WWW
- Tim Berners-Lee
  - a web of data that can be processed directly and indirectly by machines

Today's Web

- Most of today's Web content is suitable for human consumption
  - Even Web content that is generated automatically from databases is usually presented without the original structural information found in databases
- Typical Web uses today
  - seeking and making use of information (blogs, news, readers, web search)
  - social networks (facebook, linkedin)
  - reviewing catalogs of online stores and ordering products by filling out forms (e-commerce)

Problem With Today's Web

- Meaning of Web content is not machine-accessible
  - lack of semantics
- Web content is heterogeneous with little or no structure
- Data is not easily shared between web content providers
- Humans are left with the work of gathering information from various websites

Problems with Keyword Search Engines

- High recall, low precision
- Results are highly sensitive to vocabulary/syntax
- Results are single Web pages
- Human involvement is necessary to interpret and combine results
- Results of Web searches are not readily accessible by other software tools

Semantic Web Vision

- Web information has exact meaning
  - Initial Steps:
    - Google's Knowledge Graph performs semantic search
    - Wolfram Alpha search engine
    - Example
- Web information can be processed by computers
- Computers can integrate information from the web

Knowledge Management

- Knowledge management concerns itself with acquiring, accessing, and maintaining knowledge within an organization
- Key activity of large businesses: internal knowledge as an intellectual asset
- It is particularly important for international, geographically dispersed organizations
- Most information is currently available in a weakly structured form (e.g. text, audio, video)

Semantic Web Enabled Knowledge Management

- Knowledge will be organized in conceptual spaces according to its meaning.
- Automated tools for maintenance and knowledge discovery
- Semantic query answering
- Query answering over several documents
- Defining who may view certain parts of information (even parts of documents) will be possible.

The Semantic Web Impact: E-Commerce

- A typical scenario: user visits one or several online shops, browses their offers, selects and orders products. a
- Ideally humans would visit all, or all major online stores; but too time consuming
- Shopbots are a useful tool but have limitations
  - Reprogrammed every time there is a change
  - Relies on textual analysis

Semantic Web Enabled E-Commerce

- Software agents that can interpret the product information and the terms of service.
  - Pricing and product information, delivery and privacy policies will be interpreted and compared to the user requirements.
- Information about the reputation of shops
- Sophisticated shopping agents will be able to conduct automated negotiations

Semantic Web Impact

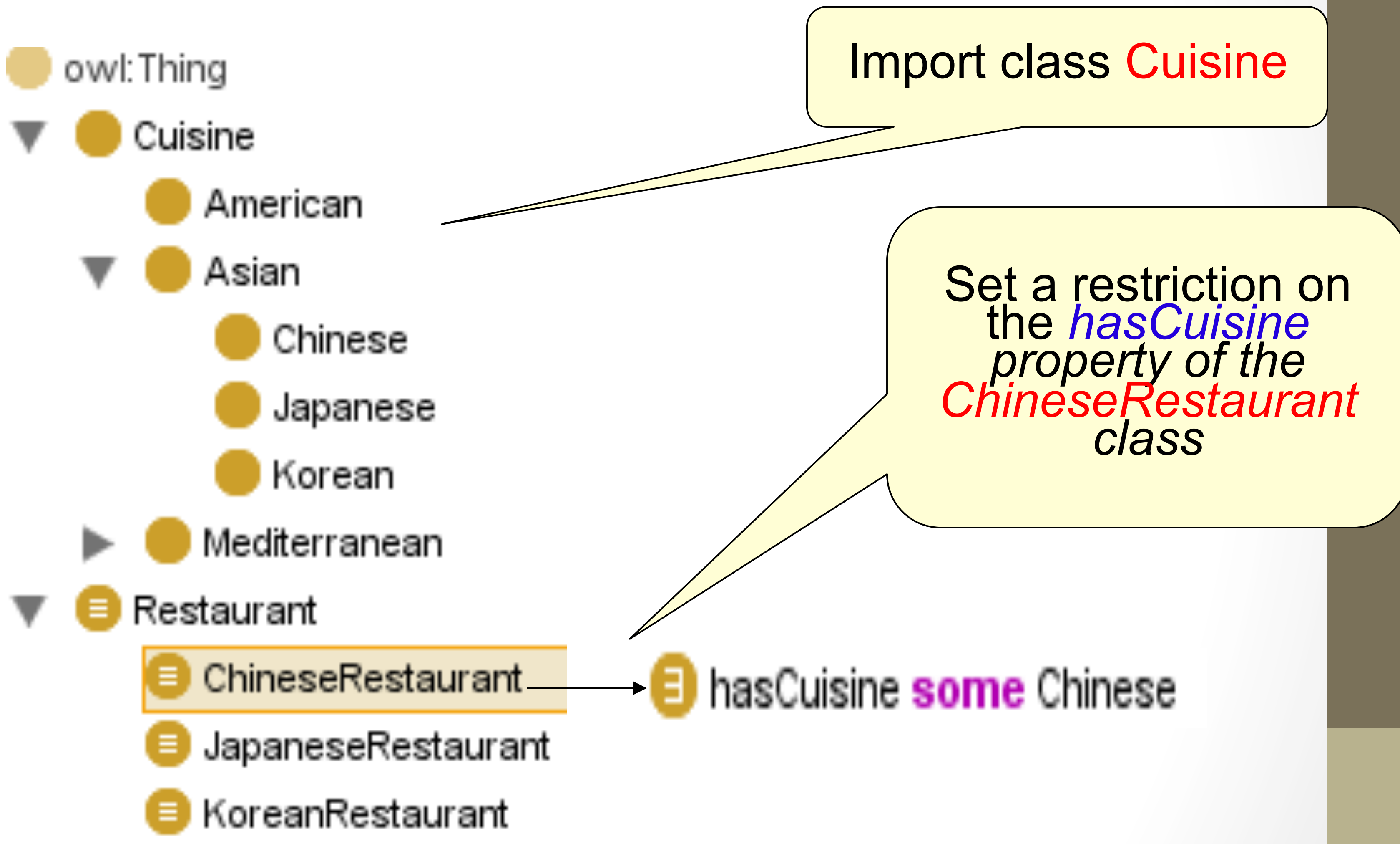
- Wikis
- Social networks
- Tags
- User Profiles

Ontology

- Knowledge Representation using Ontologies
- Definition:
  - Wikipedia: is a formal representation of knowledge as a set of concepts within a domain, and the relationships between those concepts. It is used to reason about the entities within that domain, and may be used to describe the domain.
  - Webopedia: computer-based resources that represent agreed domain semantics...relatively generic knowledge that can be reused by different kinds of applications or tasks.
  - Gruber: A formal, explicit specification of a shared conceptualization

Ontology

Example Restaurant Ontology



Ontology

Classification

Since Chinese cuisine has non-disjoint siblings Japanese and Korean then also conclude that these are *similar* to Chinese

Conclude that QueryRestaurant is equivalent to ChineseRestaurant

Ontologies and Relational Databases

- Similarities
  - Both use a model to identify common classes and properties
  - ER model can be seen as a simple hierarchical ontology
- Differences
  - Ontologies are broader in scope (rules, incomplete knowledge)
  - Ontologies provide a way for automated reasoning to occur in order to discover new relationships between entities

Ramez Elmasri, Shamkant B. Navathe: Fundamentals of Database Systems, 3rd Edition. Benjamin/Cummings 2000

Discussion: How to build a model?

- Categorizing books
- Think about the different ways they can be classified
- Create a hierarchical classification for it
- Put as much information as needed

Quest for Semantics

Three main goals of the Semantic Web:

- *Building models*: quest for describing the world in abstract terms to allow for an easier understanding of complex reality
- *Computing with knowledge*: constructing reasoning machines that can draw meaningful conclusions from encoded knowledge
- *Exchanging Information*: the transmission of complex information resources among computers that allows us to distribute, interlink, and reconcile knowledge on a global scale

Hitzler, et al., Foundations of Semantic Web Technolgoies

Building Models

- Model: simplified description of certain aspects of reality, use for understanding, structuring, or predicting parts of the real world
- History of scientific modeling
  - Plato (429-347BC)
    - What is reality?
    - Which things can be said to exist?
    - What is the true nature of things?
    - First major contribution to philosophical field of *ontology*
- Ontology in computer science
  - Description of knowledge about a domain of interest, the core of which is a machine-processable specification with a formally defined meaning

Hitzler, et al., Foundations of Semantic Web Technolgoies

Building Models

- Taxonomy: hierarchical classification
  - Linnaean taxonomy: classifies all life forms
  - WHO's International Classification of Diseases
  - Dewey Decimal Classification: ordering books in a library
- Non-hierarchical classifications
  - Periodic table of chemical elements
  - Thesaurus

Calculating with Knowledge

Syllogism:

All A are B.

All B are C.

All A are C.

Domain-independent rules provide template-like ways for inferring knowledge

Calculating with Knowledge

- Goal of AI: build machines exhibiting human intelligence
- Amount of knowledge for basic AI applications is overwhelming. Transforming human knowledge to machine-processable form is difficult
- Inference techniques became too slow for medium or large-scale tasks
- Consequently: research focused on restricted domains
  - Expert systems, rule-based systems for highly structured areas

Exchanging Information

- Internet
  - Packet-switching developed by Baran, Davies and Kleinrock
  - Splitting transmission into small “packets” and transmitted individually
  - ARPANET first packet-switching network in 1969
- Applications
  - E-mail, Usenet
  - HTML, HTTP
  - Wikis, blogs, social networks, tagging

Syntax vs Semantics

- Communication
  - Different modes of communication (speech, writing, smoke signals)
- Sharing data can be broken down into two problems
  - Syntactic sharing problem
    - Finding a common medium for communication
  - Semantic sharing problem
    - Finding a mutual encoding of concepts within a common medium

Hitzler, et al., Foundations of Semantic Web Technolgoies