Introduction to the Semantic Web
Presentation Outline

• The Semantic Web vision
• Linked data
• Ontologies
• Applications
Presentation Outline

• The Semantic Web vision
• Linked data
• Ontologies
• Applications
The Web Today

• **Universal resource identifiers (URIs)** to identify documents.
• **The Hypertext Transfer Protocol (HTTP)** to exchange documents between a client and a server.
• **HTML** for marking up information to be presented to human readers through a browser.
• **Search Engines (Google!)** to discover information.
Using Search Engines – Example 1

• Assume that you want to learn about HTTP. What can you do?
• Google it!
Example I (cont’d)
Example I (cont’d)

• So search engines work very well, but are sensitive to vocabulary and have problems associated to **precision** and **recall** (e.g., high recall – low precision or low/no recall).
Search Engines – Example II

• Assume that you want to buy the book “Semantic Web Primer”.
• Google it!
Example II (cont’d)
Example II (cont’d)
Search Engines – Example III

• Assume now that you want to buy:
  – The *cheapest* copy of the book “Semantic Web Primer” or
  – The copy that will *reach you earlier*.

• You can still use Google and your browser. But *how many clicks* do you need?

• How about using some sort of shopbot e.g., [http://www.kelkoo.co.uk/](http://www.kelkoo.co.uk/)?
Towards a Better Web!

• The **Semantic Web** vision articulated in a Scientific American article by Tim Berners-Lee, James Hendler and Ora Lassila (May 2001).
  – “The Semantic Web will bring **structure to the meaningful content of Web pages**, creating an environment where **agents** roaming from page to page readily **carry out sophisticated tasks for users**.”

• You can find the article on various Web sites e.g., [http://www.dcc.uchile.cl/~cgutierrez/cursos/IC/semantic-web.pdf](http://www.dcc.uchile.cl/~cgutierrez/cursos/IC/semantic-web.pdf)

• Notice the words **meaning** (semantics) and **agents** (a role for all of Artificial Intelligence).
How Can we Achieve the Semantic Web? – The Original Vision

• Instead of publishing information to be consumed by humans, publish machine-processable data and metadata using terms/languages that can be understood by machines.

• Build machines (agents) that will search for, query, integrate etc. this data.

• Make sure all agents understand your terms/languages.
The Semantic Web Vision Today

- From the Introduction section of the Semantic Web activity of the W3C
  http://www.w3.org/2001/sw/

  “The Semantic Web is a web of data. There is lots of data we all use every day, and it is not part of the web. I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar?

  Why not? Because we don't have a web of data. Because data is controlled by applications, and each application keeps it to itself.

  The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.”
The Semantic Web Vision Today (cont’d)

• Stressing the growing need for **Web data integration**. Lots of important Web applications demand this e.g., e-science and e-government.

• Stressing the need for **Web standards/languages for expressing shared meaning**. This is important if we want agents that are not handcrafted only for particular tasks to be developed.

• See the paper “The Semantic Web Revisited” by Nigel Shadbolt, Wendy Hall and Tim Berners-Lee at [http://eprints.ecs.soton.ac.uk/12614/1/Semantic_Web_Revisted.pdf](http://eprints.ecs.soton.ac.uk/12614/1/Semantic_Web_Revisted.pdf).
The Semantic Web Layer Cake

User Interface & Applications

Trust

Unifying Logic

Proof

Query: SPARQL

Ontology: OWL

Rule: RIF

RDFS

Data interchange:
RDF

Crypto

XML

URI/IRI
Emphasis of this Course

• The Semantic Web topics that we will cover in this course are:
  – Linked data
    • RDF, RDFS, SPARQL, RDF stores (Sesame).
  – Ontologies
    • Description logics, OWL 2, tools for developing ontologies (Protégé) and reasoners (Pellet, Hermit).
  – Rules
    • SWRL and others
  – Ontology engineering
Outline

• The Semantic Web vision
• Linked data
• Ontologies
• Applications
Open Data

• The open data movement is about making data public so they can be used freely by everyone.

• Lots of such data becoming available today is government data.
http://data.gov
http://data.gov.uk
http://geodata.gov.gr
Linked Open Data

• The goal of this W3C community effort is “to extend the Web with a data commons by publishing various open data sets as RDF on the Web and by setting RDF links between data items from different data sources.”

• The LOD community is developing a set of best practices for achieving this.
The Linked Open Data Cloud
LOD Cloud Data Sources

- US government data
- UK government data
- BBC music database
- Various biological ontologies
- General knowledge ontologies such as DBpedia, YAGO and Cyc
- Various kinds of geographical data e.g., Geonames or OpenStreetMap
- National library catalogs (USA, Germany etc.)
- Scientific publications (DBLP)
- ...

Linked Government Data

[Image of a webpage showing a screenshot of the HM Government website with the title "Opening up government" and a section on "Linked Data". The page includes a search bar and links to various data sets and SPARQL endpoints.]
Building Applications

Knowledge Technologies
Manolis Koubarakis
Linked Government Data (cont’d)
Key Linked Data Technologies

- **URIs**: a generic means to identify entities or concepts in the world
- **HTTP**: a simple, yet universal, mechanism for retrieving resources, or descriptions of resources
- **RDF**: a data model for structuring and linking data that describes things in the world
- **SPARQL**: a query language for querying linked data
Uniform Resource Identifiers

- The Web provides a general form of identifier, called the **Uniform Resource Identifier (URI)**, for identifying (naming) resources on the Web.

- Unlike URLs, URIs are not limited to identifying things that have network locations, or use other computer access mechanisms. A number of different **URI schemes (URI forms)** have been already been developed, and are being used, for various purposes.

- **Examples:**
  - `http:` (Hypertext Transfer Protocol, for Web pages)
  - `mailto:` (email addresses), e.g., `mailto:em@w3.org`
  - `ftp:` (File Transfer Protocol)
  - `urn:` (Uniform Resource Names, intended to be persistent location-independent resource identifiers), e.g., `urn:isbn:0-520-02356-0` (for a book)

- No one person or organization controls who makes URIs or how they can be used. While some URI schemes, such as URL's `http:`, depend on centralized systems such as DNS, other schemes, such as `freenet:`, are **completely decentralized**.
Example

• http://en.wikipedia.org/wiki/Angelina_Jolie

• http://dbpedia.org/page/Angelina_Jolie

• http://www.imdb.com/name/nm0001401/

• http://myfavouriteactors.gr/AngelinaJolie.html (does not exist right now)

• Note the difference between the person Angelina Jolie and the various Web pages for her. It is common to use “hash URIs” to refer to the real person:
  – http://en.wikipedia.org/wiki/Angelina_Jolie#this
Example (cont’d)

http://dbpedia.org/page/Changeling_(film)
- dbpedia-owl:starring: http://dbpedia.org/page/Angelina_Jolie
- dbpedia-owl:director: http://dbpedia.org/page/Clint_Eastwood
Dereferenceable URIs

• Linked data design principles prescribe that URIs should be dereferencable.

• Dereferenceable URIs allow us to obtain a copy or representation of the resource they identify. In this way, we can navigate the Web using e.g., HTTP and obtain useful information.
Problem

• Data sets will typically talk about the same object using different URIs. For example:


http://dbpedia.org/page/Angelina_Jolie
Solution

- Use the **special kind of link** `owl:sameAs` that identifies the resources:

  - [http://dbpedia.org/page/Angelina_Jolie](http://dbpedia.org/page/Angelina_Jolie)
RDF Basics

• The world consists of **resources**.

• Resources are identified by **URIs** and described by simple **properties** and property **values**.

• **Example**: “Angelina Jolie stars in the film Changeling directed by Clint Eastwood.”
RDF Statements

• RDF statements are **triples** of the form subject predicate object.

• Example:

```
<http://dbpedia.org/page/Changeling_(film)>
  <http://dbpedia.org/ontology/starring>
  <http://dbpedia.org/page/Angelina_Jolie> .
```

```
<http://dbpedia.org/page/Changeling_(film)>
  <http://dbpedia.org/ontology/director>
```
Example (cont’d)

• To avoid writing long URIs we can use namespace prefixes:

@prefix dbpedia: http://dbpedia.org/page/ .
@prefix dbpedia-owl: http://dbpedia.org/ontology .

dbpedia:Changeling_(film) dbpedia-owl:director dbpedia:Clint_Eastwood .
RDF Graphs

dbpedia:Changeling_(Film)

dbpedia-owl:starring

dbpedia:Angelina_Jolie

dbpedia-owl:director

dbpedia:Clint_Eastwood
The **RDF Vocabulary Description Language 1.0** (or **RDF Schema** or **RDFS**) is a language that can be used to define the vocabulary (i.e., the terms) to be used in an RDF graph.

The RDF Vocabulary Description Language 1.0 is used to indicate that we are describing specific **kinds** or **classes** of resources, and will use specific **properties** in describing those resources.

The RDF Vocabulary Description Language 1.0 is an **ontology definition language** (a **simple** one, compared with other languages such as OWL; **we can only define taxonomies and do some basic inference about them**).

The RDF Vocabulary Description Language is like a **schema definition language** in the relational or object-oriented data models (hence the alternative name RDF Schema – we will use this name and its shorthand RDFS mostly!).
Example – Class Hierarchy
Example – Property Hierarchy

dbpedia-owl:spouseOf

rdf:subPropertyOf

dbpedia-owl:husbandOf

rdf:subPropertyOf

dbpedia-owl:wifeOf
Standard Vocabularies

• Linked data are written using standard vocabularies i.e., vocabularies that have been agreed by certain communities for describing certain kinds of resources.

• Examples:
  – FOAF
  – Dublin Core
  – schema.org
The Friend of a Friend (FOAF) project has created a standard vocabulary for describing people, the links between them and the things they create and do.

Example

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://en.wikipedia.org/wiki/Angelina_Jolie#this> 
   rdf:type foaf:Person ;
   foaf:name "Angelina Jolie" ;
   foaf:mbox <mailto:ajolie@gmail.com> ;
   foaf:knows [ rdf:type foaf:Person ;
Dublin Core

- Dublin Core is a community effort that has defined a set of metadata attributes (e.g., creator, creation date, title etc.) for **describing a wide range of digital resources**.

- See [http://dublincore.org](http://dublincore.org) for more information.
http://schema.org

• This is a recent initiative by the Big Three (Bing, Google and Yahoo!) to provide a collection of schemas that webmasters can use to markup their pages in ways recognized by major search providers.
http://schema.org/Place

### Thing > Place

Entries that have a somewhat fixed, physical extension.

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>description</td>
<td>Text</td>
<td>A short description of the item.</td>
</tr>
<tr>
<td>image</td>
<td>URL</td>
<td>URL of an image of the item.</td>
</tr>
<tr>
<td>name</td>
<td>Text</td>
<td>The name of the item.</td>
</tr>
<tr>
<td>url</td>
<td>URL</td>
<td>URL of the item.</td>
</tr>
</tbody>
</table>

Properties from `Place`:

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>address</td>
<td>PostalAddress</td>
<td>Physical address of the item.</td>
</tr>
<tr>
<td>aggregateRating</td>
<td>AggregateRating</td>
<td>The overall rating, based on a collection of reviews or ratings, of the item.</td>
</tr>
<tr>
<td>containsPlace</td>
<td>Place</td>
<td>The basic containment relation between places.</td>
</tr>
<tr>
<td>creator</td>
<td>Event</td>
<td>The events held at this place or organization.</td>
</tr>
<tr>
<td>faxNumber</td>
<td>Text</td>
<td>The fax number.</td>
</tr>
<tr>
<td>geo</td>
<td>GeoCoordinates</td>
<td>The geo coordinates of the place.</td>
</tr>
<tr>
<td>interactionCount</td>
<td>Text</td>
<td>A count of a specific user interactions with this item—for example, 20 votes/likes, 5 user comments, or 100 Tweet/Follows. The user interaction type should be one of the sub types of <code>UserInteraction</code>.</td>
</tr>
<tr>
<td>maps</td>
<td>URL</td>
<td>A URL to a map of the place.</td>
</tr>
<tr>
<td>photos</td>
<td>Photograph or ImageObject</td>
<td>Photographs of this place.</td>
</tr>
<tr>
<td>reviews</td>
<td>Review</td>
<td>Review of the item.</td>
</tr>
<tr>
<td>telephoneNumber</td>
<td>Text</td>
<td>The telephone number.</td>
</tr>
</tbody>
</table>

More specific types:

- AdministrativeArea
- Concert
- LandForm
- LandmarksOrHistoricalBuildings
- LocalBusiness
- Residence
- TourismAttraction

Schema Draft Version 0.9
http://schema.org/LocalBusiness
Microdata

• Microdata provide a simple way to **annotate HTML elements with machine readable tags.**

• Microdata can use standardized vocabularies to capture the **semantics** of HTML items.

• See http://www.w3.org/TR/microdata/ for more details.

• Similar (earlier) approaches are RDFa and Microformats.
Example – Original HTML

• The following is some HTML code for describing a local business called “Beachwalk Beachwear and Giftware”.

<h1>Beachwalk Beachwear & Giftware</h1>
A superb collection of fine gifts and clothing to accent your stay in Mexico Beach.
3102 Highway 98
Mexico Beach, FL
Phone: 850-648-4200
Example (cont’d)— HTML With Microdata

<div itemprop itemType="http://schema.org/LocalBusiness">
  <h1><span itemprop="name">Beachwalk Beachwear & Giftware</span></h1>
  <span itemprop="description">A superb collection of fine gifts and clothing to accent your stay in Mexico Beach.</span>
  <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
    <span itemprop="streetAddress">3102 Highway 98</span>,
    <span itemprop="addressLocality">Mexico Beach</span>,
    <span itemprop="addressRegion">FL</span>
  </div>
  <span itemprop="telephone">850-648-4200</span>
</div>
SPARQL

• SPARQL stands for “SPARQL Protocol and RDF Query Language”.

• It is the standard query language for RDF data proposed by the W3C.

• It is based on matching graph patterns against RDF graphs.

• The simplest kind of graph pattern is a triple pattern.
  – A triple pattern is like an RDF triple, but with the option of a variable in the subject, predicate or object positions.
Example Dataset

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://en.wikipedia.org/wiki/Angelina_Jolie#this>
  rdf:type foaf:Person ;
  foaf:name "Angelina Jolie" ;
  foaf:mbox <mailto:ajolie@gmail.com> ;
  foaf:knows [ rdf:type foaf:Person ;
Example SPARQL Query

SELECT ?name
WHERE { ?x foaf:name ?name .
    ?x rdf:type foaf:Person .
}

• Result: "Angelina Jolie"
Presentation Outline

• The Semantic Web vision
• Linked data
• Ontologies
• Applications
Ontologies

- An **ontology** is a formal, explicit, shared specification of a conceptualization of a domain (Gruber, 1993).

- **Conceptualization**: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them. A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose.

- The term **ontology** is borrowed from Philosophy, where ontology is a systematic account of existence (what things exist, how they can be differentiated from each other etc.).

- Today the word ontology is a synonym for a **shared knowledge base**.
Formal Languages for Ontologies

• Ontologies are typically expressed in some formal logic-based language (e.g., first-order logic).

• The literature also offers special formalisms for defining ontologies that contain mainly taxonomic knowledge:
  – Semantic networks
  – Frames
  – Description logics
  – RDF, RDFS and OWL in the Semantic Web.
Formal Languages for Ontologies (cont’d)

• You can think about these formalisms as being object-oriented logics:
  – They have special constructs for representing knowledge about individuals (or objects), categories (or classes) and relationships (or roles).
  – Categories are organized into taxonomies.
  – They have special reasoning methods to deal with these constructs.

• Description logics, RDF, RDFS and OWL 2 are a core part of this course.
• Taxonomies have been used profitably for centuries in various technical fields (biology, medicine, library science etc.).

• Taxonomic information plays a central role in other Computer Science areas e.g., programming languages, databases and software engineering.

• Taxonomies are very important in modern applications: Web information retrieval and integration, knowledge management, e-commerce, e-science, e-government etc.
Example
Very Large Ontologies

• Recently there has been a lot of work on developing very large ontologies that capture various areas of human knowledge and deploying this knowledge in applications such as search engines or question answering.

• Example: Watson, IBM’s question answering system that beat humans in the quiz show Jeopardy (http://www-03.ibm.com/innovation/us/watson/index.html).
Example of Very Large Ontologies

- Cyc
- Wordnet
- DBpedia
- YAGO
- Freebase
- TextRunner
From the Cyc Web site (http://www.cyc.com/):

- The Cyc KB is a formalized representation of a vast quantity of **human common sense knowledge**: facts, rules of thumb, and heuristics for reasoning about the objects and events of everyday life.

- At the present time, the Cyc KB contains nearly **500,000 terms**, including about **15,000 types of relations**, and about **5,000,000 facts** (assertions) relating these terms.

- New assertions are continually added to the KB through a **combination of automated and manual means**.
What does Cyc know?

Domain-Specific Knowledge
(e.g., Healthcare, Computer Security, Command and Control, Mortgage Banking, ...)

Domain-Specific Facts and Data
• Cyc is written in CycL: an extension of FOL with equality, default reasoning, skolemization, and some second-order features.

• The reasoning engine of Cyc uses around 1000 specialized reasoners.

• There is an open source version called OpenCyc.
Wordnet

- From the WordNet Web site (http://wordnet.princeton.edu/):
  - WordNet is a large lexical database of English (around 200,000 of words).
  
  - Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept.

  - Synsets are interlinked by means of conceptual-semantic (hyponym, hypernym, meronym etc.) and lexical relations.

  - WordNet is not really an ontology but parts of it can be made into an ontology containing taxonomic information.
Example

WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: actor
Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) actor, histrion, player, thespian, role player (a theatrical performer)
  - direct hyponym / full hyponym
  - has instance
  - direct hypernym / inherited hypernym / sister term
    - S: (n) performer, performing artist (an entertainer who performs a dramatic or musical work for an audience)
      - derivationally related form
    - S: (n) actor, doer, worker (a person who acts and gets things done) "he's a principal actor in this affair"; "when you want something done get a doer"; "he's a miracle worker"
Example (cont’d)

WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: actor
Search WordNet

Display Options: [Select option to change] Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) actor, histriion, player, thespian, role player (a theatrical performer)
  - direct hyponym / full hyponym
  - has instance
    - S: (n) Allen, Woody Allen, Allen Stewart Konigsberg (United States filmmaker and comic actor (1935-))
    - S: (n) Astaire, Fred Astaire (United States dancer and cinema actor noted for his original and graceful tap dancing (1899-1987))
    - S: (n) Barrymore, Maurice Barrymore, Herbert Blythe (United States actor; husband of Georgiana Emma Barrymore and father of Ethel Barrymore and John Barrymore and Lionel Barrymore (1847-1905))
    - S: (n) Barrymore, John Barrymore (United States actor; son of Maurice Barrymore and Georgiana Barrymore (1878-1954))
    - S: (n) Bogart, Humphrey Bogart, Humphrey DeForest Bogart (United States film actor (1899-1957))
    - S: (n) Booth, John Wilkes Booth (United States actor and assassin of President Lincoln (1838-1865))
    - S: (n) Burbage, Richard Burbage (English actor who was the first to play the leading role in several of Shakespeare’s tragedies (1567-1619))
    - S: (n) Burton, Richard Burton (Welsh film actor who often co-starred with Elizabeth Taylor (1925-1984))
    - S: (n) Cagney, Jimmy Cagney, James Cagney (United States film actor known for his portrayals of tough characters (1899-1988))
    - S: (n) Chevalier, Maurice Chevalier (French actor and cabaret singer (1888-1972))
    - S: (n) Cooper, Gary Cooper, Frank Cooper (United States film actor noted for his portrayals of strong silent heroes (1901-1961))
    - S: (n) Coward, Noel Coward, Sir Noel Pierce Coward (English dramatist and actor and composer noted for his witty and sophisticated comedies
From the DBpedia Web site (http://dbpedia.org/About):

- DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web.
- The DBpedia knowledge base currently describes more than 3.5 million things, out of which 1.67 million are classified in a consistent Ontology, including 364,000 persons, 462,000 places, 99,000 music albums, 54,000 films, 17,000 video games, 148,000 organisations, 169,000 species and 5,200 diseases.
- The DBpedia knowledge base has several advantages over existing knowledge bases: it covers many domains; it represents real community agreement; it automatically evolves as Wikipedia changes, and it is truly multilingual.
DBpedia (cont’d)

• DBpedia, like other ontologies we will see below, is different than Cyc and WordNet. Here the emphasis is on individuals, not general knowledge about various domains.

• DBpedia is written in RDF and it is part of the Linked Open Data effort.
About: Angelina Jolie
An Entity of Type: person, From Named Graph: http://dbpedia.org, With Data Space: dbpedia.org

Angelina Jolie (born Angelina Jolie Voight, June 4, 1975) is an American actress. She has received an Academy Award, two Screen Actors Guild Awards, and three Golden Globe Awards. Jolie promotes humanitarian causes, and is noted for her work with refugees as a Goodwill Ambassador for the United Nations High Commissioner for Refugees (UNHCR).

- **Property**:.fasterxml
doi
event

- **Value**:
  - http://dbpedia.org/page/Angelina_Jolie
  - Knowledge Technologies
  - Manolis Koubarakis

---

http://dbpedia.org/page/Angelina_Jolie
http://dbpedia.org/page/Angelina_Jolie
YAGO

- From the YAGO Web page (http://www.mpi-inf.mpg.de/yagonaga/yago/):
  - YAGO2 is a huge semantic knowledge base, derived from Wikipedia, WordNet and GeoNames. Currently, YAGO2 has knowledge of more than 10 million entities (like persons, organizations, cities, etc.) and contains more than 80 million facts about these entities.
  - In YAGO2, we made an effort to treat time and location data as first-class citizens, extending the basic triple model by special fields for time and location for querying. Also, we took special care to consistently attach temporal and spatial data to all facts where it is semantically meaningful and where time and location can be derived from Wikipedia.
Freebase

• From the Freebase Web site (http://www.freebase.com/):
  – Freebase is an open, Creative Commons licensed repository of structured data of almost 20 million entities.
  – An entity is a single person, place, or thing. Freebase connects entities together as a graph.
• Freebase is owned by Google.
TextRunner

• TextRunner is an **ontology** built by parsing 500,000 Web pages and extracting **relationships among entities** from them.

• TextRunner is part of the research project **KnowItAll** at the University of Washington. This project concentrates on two fundamental questions:
  – How can a computer accumulate a massive body of knowledge?
  – What will Web search engines look like in ten years?
ReVerb Search (Experimental)

NOTE: You may have trouble running ReVerb Search if you are behind a firewall, because it utilizes port 7125. If not seeing the results, try accessing ReVerb from a machine outside your firewall.

Example Queries:
- "Who built the Pyramids?"
- "What did Thomas Edison invent?"
- "What kills bacteria?"
- "What contains antioxidants?"

Filtered Example Queries:
- "What countries are located in Africa?"
- "What foods are grown in what countries?"
- "What sports originated in China?"
- "What chemicals has the FDA approved?"
- "What cities are located in India?"

Search individual fields:
- Argument 1
- Predicates
- Argument 2

Search

To download our Open Information Extractor, click here: http://reverb.cs.washington.edu

Advanced Options

Type selection is based on the FreeBase type scheme. For more info, click here.

Source: Freebase licensed under CC BY
Other content from Wikipedia licensed under the GFDL.
TextRunner (cont’d)

Turing Center KnowItAll Project

ReVerb Search (Experimental)

NOTE: You may have trouble running ReVerb Search if you are behind a firewall, because it utilizes port 7125. If not seeing the results, try accessing ReVerb from a machine outside your firewall.

Example Queries:
- “Who built the Pyramids?”
- “What did Thomas Edison invent?”
- “What kills bacteria?”
- “What contains antioxidants?”

Filtered Example Queries:
- “What countries are located in Africa?”
- “What foods are grown in what countries?”
- “What sports originated in China?”
- “What chemicals has the FDA approved?”
- “What cities are located in India?”

Search individual fields:
Argument 1
Predicate
Argument 2
Angelina Jolie

To download our Open Information Extractor, click here: http://reverb.cs.washington.edu

Advanced Options

Type selection is based on the FreeBase type schema. For more info, click here.

Source Freebase. Licensed under CC BY
Other content from Wikipedia. Licensed under the GPL.
TextRunner (cont’d)

ReVerb Search

Retrieved 6 results for Predicate containing “married” and Argument 2 containing “Angelina Jolie”

Grouping results by argument 2. Group by: and contain

Angelina Jolie (5 results)

Thorton (3), Billy Bob (2), actor (1), gay (1) is married to Angelina Jolie
Brad Pitt (1) will marry Angelina Jolie
The example ontologies we saw have been created using different means:

- Getting a team of trained ontology engineers to organize the ontology and write the axioms (CYC). WordNet was also built by trained experts (lexicographers).
- Importing structured facts from various Web sources possibly with some inference (DBpedia, FreeBase, YAGO2). Manual curation is also possible here (FreeBase, DBpedia).
- Parsing text documents and extracting information from them (TextRunner).

The availability of open sources of knowledge on the Web has fuelled the development of efforts such as DBpedia, FreeBase and YAGO2. We expect this trend to continue in the future!
Building Ontologies (cont’d)

• **Ontology engineering** offers us general methodologies of how to create, maintain and evolve ontologies.

• See the OWL tutorial available at http://www.co-ode.org/resources/tutorials/intro/.
Description Logics

• The origins of DLs lie in research on semantic networks and frames. DLs are languages for describing the **nature and structure of objects**.

• The DL approach to KR was developed in the 80's and 90's in parallel with pure FOL approaches and other languages for structured objects like Telos and F-logic.

• Recently, DLs have been used to provide the **foundations for ontology languages for the Web** e.g., OWL.

• DLs are logics based on **descriptions of concepts or terms**.
The Origins: KL-ONE

Knowledge Technologies
Manolis Koubarakis
Examples

• The set of female doctors.

\[ \text{Female} \cap \text{Doctor} \]

• The set of individuals that have at least 3 children that are male.

\[ (\geq 3 \ \text{hasChild} \ . \ \text{Male}) \]

• The set of individuals such that all their children have graduated from at least one Greek University.

\[ (\forall \text{hasChild} \ . \ (\exists \text{isAlumniOf} \ . \ \text{GreekUniversity})) \]
Examples (cont’d)

• The set of individuals that have at least three children such that all their degrees are from Greek Universities.

\[(\geq 3 \text{hasChild} \land (\forall \text{isAlumniOf} \cdot \text{GreekUniversity}))\]
Examples (cont’d)

• Ann is a female doctor.

\[(Female \cap Doctor)(ANN)\]

• John is a child of Ann.

\[hasChild(ANN,JOHN)\]

• Ann has at least 3 children that are male.

\[(\geq 3 \ hasChild \ . \ Male ) (ANN)\]
Examples (cont’d)

• A woman is a female person.
  \[ Woman \equiv Person \land Female \]

• A parent is a person who has at least one child.
  \[ Parent \equiv Person \land (\exists hasChild . Person) \]
OWL

• OWL 2 is the current version of the Web Ontology Language and a W3C recommendation as of October 2009.

• The previous version of OWL (OWL 1) became a W3C recommendation in 2004.

• All W3C documents about OWL 2 can be found at http://www.w3.org/TR/2009/REC-owl2-overview-20091027/.
The Structure of OWL 2

OWL 2 Ontology

- Ontology Structure
- RDF Graph

Syntax layer
Semantics layer

Direct Semantics
RDF-Based Semantics

correspondence theorem (for DL subset)
Example

SubClassOf(:Student  :Person)

EquivalentClasses (:Man
   ObjectIntersectionOf(:Person  :Male))

EquivalentClasses (:Woman
   ObjectIntersectionOf(:Person  :Female))

EquivalentClasses (:Parent
   ObjectSomeValuesFrom(:hasChild  :Person))
Example (cont’d)

EquivalentClasses (:Teenager
    ObjectIntersectionOf (Person
    DataSomeValuesFrom (:hasAge
    DatatypeRestriction (xsd:integer
        xsd:minExclusive "12"^^xsd:integer
        xsd:maxInclusive "19"^^xsd:integer )))
)
Example (cont’d)

EquivalentClasses(:ChildlessPerson
  ObjectIntersectionOf(:Person
  ObjectComplementOf(ObjectSomeValuesFrom(
  ObjectInverseOf(:hasParent) owl:Thing))))
Example (cont’d)

ClassAssertion(:Person :John)
ClassAssertion(:Male :John)

ClassAssertion(:Person :Mary)
ClassAssertion(:Female :Mary)

ObjectPropertyAssertion(:hasChild :John :Mary)

NegativeObjectPropertyAssertion(:hasChild :John :Tina)

DataPropertyAssertion(:hasAge :Maria "18"^^xsd:integer)
Example (cont’d)

SameIndividual(:John :Jack)
SameIndividual(:John otherOnt:JohnBrown)
SameIndividual(:Mary otherOnt:MaryBrown)

DifferentIndividuals(:John :Bill)
Presentation Outline

- The Semantic Web vision
- Linked data
- Ontologies
- Applications
Applications

- Data integration
- E-government
- E-commerce
- Tourism
- Medicine
- Biology
- ...

Readings

• The papers on the Semantic Web vision mentioned in the presentation.


• The book “Linked Data: Evolving the Web into a Global Data Space” by Tom Heath and Christian Bizer available from http://linkeddatabook.com/editions/1.0/.
Readings (cont’d)

- The W3C Semantic Web Activity web site (http://www.w3.org/2001/sw/). Browse! There are many interesting introductory tutorials. You should definitely see one of the introductory tutorials by Ivan Herman e.g., http://www.w3.org/2010/Talks/0622-SemTech-IH/Tutorial.pdf which gives a very good introduction to the current state of the art in this area.

- The ESWC invited talk of Jim Hendler available at http://videolectures.net/eswc2011_hendler_work/ which looks at what the Semantic Web has achieved so far and deals with various criticisms of the area.
Outline of the Course

• The Resource Description Framework (RDF and RDFS)
• SPARQL: A Query language for RDF and RDFS
• Description logics
• The Web Ontology Language (OWL 2)
• Ontology Engineering
• Rule languages for the Semantic Web
• Other topics can be covered in your presentations.