Natural Language Processing DSECL ZG565

Dr. Chetana Gavankar, Ph.D,
IIT Bombay-Monash University Australia
Chetana.gavankar@pilani.bits-pilani.ac.in

BITS Pilani
Pilani Campus

innovate achieve lead

BITS Pilani
Pilani Campus

Session 14: Semantic Web Ontology
Date – 16th March 2024

These slides are prepared by the instructor, with grateful acknowledgement of Prof. Philip Cimiano , Johanna Völker , Paul Buitelaar and many others who made their course materials freely available online.

Session Content

- Introduction
- Semantic Web and Ontology
- Ontology Languages
- Ontology Engineering
- State of the Art
  - Ontology Learning
  - Knowledge Graphs
  - Linked open data

A Brief History of Web Search

- 1991: Tim Berners-Lee “invents” the World Wide Web
- First Web search engines:
  - Archie: Query file names by regular expressions
  - Architext/Excite: Full text search, simple ranking (1993)
- Until 1998, web search meant information retrieval
- 1998: Google was founded

Problem:
Computers don't understand Meaning

"My mouse is broken. I need a new one"

The Syntactic Web

[Hendler & Miller 02]

The Syntactic Web is…

- A hypermedia, a digital library
  - A library of documents called (web pages) interconnected by a hypermedia of links
- A database, an application platform
  - A common portal to applications accessible through web pages, and presenting their results as web pages
- A platform for multimedia
  - BBC Radio anywhere in the world! Terminator 3 trailers!
- A naming scheme
  - Unique identity for those documents

A place where computers do the presentation (easy) and people do the linking and interpreting (hard).

Why not get computers to do more of the hard work?

[Goble 03]

Is it possible to get answers for these using syntactic web?

- Complex queries involving background knowledge
  - Find information about "animals that use sonar but are not either bats or dolphins"
- Locating information in data repositories
  - Results of human genome experiments

e.g., Barn Owl

What is the Problem?

Consider a typical web page:

- Markup consists of:
  - rendering information (e.g., font size and colour)
  - Hyper-links to related content
- Semantic content is accessible to humans but not (easily) to computers…

[Davies, 03]

What information can we see…

WWW2002

The eleventh international world wide web conference

Sheraton waikiki hotel

Honolulu, hawaii, USA

7-11 may 2002

1 location 5 days learn interact

Registered participants coming from

australia, canada, chile denmark, france, germany, ghana, hong kong, india, ireland, italy, japan, malta, new zealand, the netherlands, norway, singapore, switzerland, the united kingdom, the united states, vietnam, zaire

Register now

On the 7th May Honolulu will provide the backdrop of the eleventh international world wide web conference. This prestigious event …

Speakers confirmed

Tim berners-lee

Tim is the well known inventor of the Web, …

Ian Foster

Ian is the pioneer of the Grid, the next generation internet …

What information can a machine see…

♈♈♈🗎🗀🗀🗎
❄♒♏ ♏●♏❖♏■♦♒ ♓■♦♏❒■♋♦♓□■♋● ⬩□❒●♎ ⬩♓♎♏ ⬩♏♌ ♍□■♐♏❒♏■♍♏
💧♒♏❒♋♦□■ ⬩♋♓&♓&♓ ♒□♦♏●
📌□■□●◆●◆📭 ♒♋⬩♋♓♓📭 ✝💧✌
📠📭🗁🗁 ❍♋♎ 🗎🗀🗀🗎
🗁 ●□♍♋♦♓□■ 🗄 ♎♋♎⬩ ●♏♋❒■ ♓■♦♏❒♋♍♦
☼♏♑♓⬩♦♏❒♏♎ □♋❒♦♓♍♓□♋■♦⬩ ♍□❍♓■♑ ♐❒□❍
♋◆⬩♦❒♋●♓♋📭 ♍♋■♋♎♋📭 ♍♒♓●♏ ♎♏■❍♋❒&📭 ♐❒♋■♍♏📭 ♑♏❒❍♋■♎📭
♑♒♋■♋📭 ♒□■♑ &□■♑📭 ♓■♎♓♋📭 ♓❒♏●♋■♎📭 ♓♦♋●♎📭 ♈♋□♋■
📭 ❍♋●♦♋📭 ■♏⬩ ⌘♏♋●♋■♎📭 ♦♒♏ ■♏♦♒♏❒●♋■♎⬩📭 ■□❒⬩♋♎📭 ⬩♓■
♑♋□□❒♏📭 ⬩⬩♓♦⌘♏❒●♋■♎📭 ♦♒♏ ◆■♓♦♏♎ &♓■♑♎□❍📭 ♦♒♏ ◆■♓♦♏
♎ ⬩♦♋♦♏⬩📭 ❖♓♏♦■♋❍📭 ⌘♋♓❒♏
☼♏♑♓⬩♦♏❒ ■□⬩
🏳■ ♦♒♏ 🖮♦♒ 💣✳♋♎ 📌□■□●◆●◆ ⬩♓●● □❒□❖♓♎♏ ♦♒♏ ♌♋♍&♎❒□□ □↗
♦♒♏ ♏●♏❖♏■♦♒ ♓■♦♏❒■♋♦♓□■♋● ⬩□❒●♎ ⬩♓♎♏ ⬩♏♌ ♍□■♐♏❒♏■♍♏
📭 ❄♒♓⬩ □❒♏⬩♦♓♑♓□◆⬩ ♏❖♏■♦ ⑤
💧□♏♋&♏❒⬩ ♍□■♐♓❒❍♏♎
❄♓❍ ♌♏❒■♏❒⬩📭●♏♏
❄♓❍ ♓⬩ ♦♒♏ ⬩♏●● &■□⬩■ ♓■❖♏■♦□❒ □↗ ♦♒♏ ♈♏♌📭 ⑤
✋♋■ ☞□⬩♦♏❒
✋♋■ ♓⬩ ♦♒♏ □♓□■♏♏❒ □↗ ♦♒♏ ☝❒♓♎📭 ♦♒♏ ■♏☒♦ ♑♏■♏❒♋♦♓□■ ♓
■♦♏❒■♏♦ ⑤

Data on the Web

- The data on the Web increases every second
  - government data, health related data, general knowledge, company information, flight information, restaurants,…
- More and more applications rely on the availability of that data

But… data are often in isolation, “silos”

Photo credit Alex (ajagendorf25), Flickr

Web data issues

Many challenges in data management

- Heterogeneous data
- Implicit semantics
- No background knowledge
- Data quality difficult to measure
- Proprietary schemas
- Unstructured data
- Ambiguity
- Multilinguality
- Various units and currencies

Imagine…

- A “Web” where
  - documents are available for download on the Internet
  - but there would be no hyperlinks among them

And the problem *is* real…

CoCoDat - Collation of Cortical Data - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.cocom

CoCoMac DATABASES ORT EXAMPLES

CoCoDat: Collation of Cortical [s
microcircuitry] Data

CoCoDat is a microcircuitry database that
published experimental reports. The data
and cellular compartment), as well as the

- Morphology
- Firing properties
- Ionic currents
- Ionic conductances
- Synaptic currents
- Connectivity

The database is available for download u
data tables but also a Search Board with
manual or automatic relaxation of the sea

- Brain region
- Layer
- Neuron type

http://www.cocomac.org/cocodat/catalyzer/inc

Cell Centered Database - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://ccdb.ucsd.edu/sand/main?event=gallery&action=show&dpl=y

red database

Cell Centered Database™
National Center for Microscopy and Imaging Research
Gallery

Data | Search | Gallery | Dictionary | Publications | MyCCDB | Data Download | Contact us | Help

2D image | Reconstruction | Segmentation | Animation

23
071202tjd
P1188
Brain Maps
Tom Deerinck
BioRad RTS 2000MP
Multiphoton
montage of cerebellar cortex triple labeled for Hoescht stain (blue), IP3 receptor (green) and GFAP (red)

logged out

NeuronDB = Thalamic relay neuron - Overview (A) () - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://senselab.med.yale.edu

senselab

NeuronDB

Back

Thalamic relay neuron

Mode: **Overview** Data/Search plus Connectivity plus Classical References/Notes Models

Region: Distal equivalent dendrite | Middle equivalent dendrite | Proximal equivalent dendrite | Soma | Axon hillock | Axon fiber | Axon terminal | All Compartments

Properties: Receptors Channels Transmitters **All Properties**

Interoperation: Gene and Chromosome Experimental Data (neurodatabase.org) Microscopy Data (CCDB)

Neuron type: principal
Organism: Vertebrates

1. Equivalent dendrite — Show other
2. Distal equivalent dendrite — Show other
3. Middle equivalent dendrite — Show other
4. Proximal equivalent dendrite — Show other
5. Soma — Show other

Done

logged out

Data on the Web is not enough…

- We need a proper infrastructure for a real *Web of Data*
  - data is available on the Web
    - accessible via standard Web technologies
  - data are interlinked over the Web
  - ie, data can be *integrated* over the Web
- This is where Semantic Web technologies come in

The Web of Data

- A vision of the Web as a huge database
- Can we query the Web as a database? **What were the most popular songs when Obama was elected?**

SELECT ?song, COUNT(?vote) AS ?likes FROM WorldWideWeb WHERE (?vote inFavourOf ?song) AND (?vote madeAt ?date) AND (Obama electedAt ?date) ORDER BY ?likes DESC

A Web of Data unleashes new applications

We start with a example of a book...

A simplified bookstore data (dataset “A”)

ID	Author	Title	Publisher	Year
ISBN 0-00-6511409-X	id_xyz	The Glass Palace	id_qpr	2000

ID	Name	Homepage
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

ID	Publisher’s name	City
id_qpr	Harper Collins	London

1st: export your data as a set of *relations*

Some notes on the exporting the data

- Relations form a graph
  - the nodes refer to the “real” data or contain some literal
  - graph is represented in machine using semantic web technologies

Some notes on the exporting the data

- Data export does not necessarily mean physical conversion of the data
  - relations can be generated on-the-fly at query time
    - via SQL “bridges”
    - scraping HTML pages
    - extracting data from Excel sheets
    - etc.
- One can export part of the data

Same book in French…

Another bookstore data (dataset “F”)

	A	B	C	D
1	**ID**	**Titre**	**Traducteur**	**Original**
2	ISBN 2020286682	Le Palais des Miroirs	$A12$	ISBN 0-00-6511409-X
3				
4				
5				
6	**ID**	**Auteur**		
7	ISBN 0-00-6511409-X	$A11$		
8				
9				
10	**Nom**			
11	Ghosh, Amitav			
12	Besse, Christianne			

2nd: export your second set of data

3rd: start merging your data

3rd: start merging your data (cont)

3rd: start merging your data

Start making queries…

- User of data “F” can now ask queries like:
  - “give me the title of the original”
    - well, … « donnes-moi le titre de l’original »
- This information is not in the dataset “F”…
- …but can be retrieved by merging with dataset “A”!

However, more can be achieved…

- We “feel” that a:author and f:auteur should be the same
- But an automatic merge doest not know that!
- Let us add some extra information to the merged data:
  - a:author same as f:auteur
  - both identify a “Person”
  - a term that a community may have already defined:
    - a “Person” is uniquely identified by his/her name and, say, homepage
    - it can be used as a “category” for certain type of resources

3rd revisited: use the extra knowledge

Start making richer queries!

- User of dataset “F” can now query:
  - “donnes-moi la page d’accueil de l’auteur de l’original”
    - well… “give me the home page of the original’s ‘auteur’”
- The information is not in datasets “F” or “A”…
- …but was made available by:
  - merging datasets “A” and datasets “F”
  - adding three simple extra statements as an extra “glue”

Combine with different datasets

- Using, e.g., the “Person”, the dataset can be combined with other sources
- For example, data in Wikipedia can be extracted using dedicated tools
  - e.g., the “dbpedia” project can extract the “infobox” information from Wikipedia already…

Merge with Wikipedia data

Merge with Wikipedia data

Merge with Wikipedia data

It could become even more powerful

- We could add extra knowledge to the merged datasets
  - e.g., a full classification of various types of library data
  - geographical information etc
- This is where ontologies come in
  - ontologies sets can be relatively simple and small, or huge, or anything in between…
- Even more powerful queries can be asked as a result

What did we do?

Applications

Manipulate
Query
…

Data represented in abstract format

Map,
Expose,
…

Data in various formats

So where is the Semantic Web?

- The Semantic Web provides technologies to make such integration possible!

The Semantic Web Vision

"… the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes ,but for automation, integration and reuse of data across various applications"

http://www.w3.org/sw/

Semantic Web

Tim Berner Lee's Vision:

- Web as a means of collaboration for people
- Web as a means of collaboration for machines

Semantic Web is a web of data that machines can "understand" too.

Semantic Web – New Users

Need to Add "Semantics"

- Use Ontologies to specify meaning of annotations
  - Ontologies provide a vocabulary of terms
  - New terms can be formed by combining existing ones
  - Meaning (semantics) of such terms is formally specified
  - Can also specify relationships between terms in multiple ontologies

Ontology Definition

unambiguous terminology definitions

conceptual model of a domain (ontological theory)

Formal, explicit specification of a shared conceptualization

[Gruber93]

machine-readability with computational semantics

commonly accepted understanding

Structure of an Ontology

Ontologies typically have two distinct components:

Names for important concepts in the domain

- Elephant is a concept whose members are a kind of animal
- Herbivore is a concept whose members are exactly those animals who eat only plants or parts of plants
- Adult_Elephant is a concept whose members are exactly those elephants whose age is greater than 20 years

- Background knowledge/constraints on the domain
  - Adult_Elephants weigh at least 2,000 kg
  - All Elephants are either African_Elephants or Indian_Elephants
  - No individual can be both a Herbivore and a Carnivore

Ontology Example

Concept
conceptual entity of the domain

Attribute
property of a concept

Relation
relationship between concepts or properti

Axiom
coherent description between Concepts / Properties s via logical expressions

holds(Professor, Lecture) $\Rightarrow$ Lecture.topic $\in$ Professor.researchField

Ontology Elements

- Concepts (classes) + their hierarchy
- Concept properties (slots/attributes)
- Property restrictions (type, cardinality, domain)
- Relations between concepts (disjoint, equality)
- Instances

Ontology Languages

•For the purpose of formalizing ontologies, variants of first-order logic with standard model-theoretic semantics are used

- RDF (Resource Description Framework)
  - Specifies relationship between data
- RDFS(Resource Description Framework Schema)
  - Specifies relationship between schema
- OWL (Web Ontology Language)
  - Specifies more complex relationship between schema based on description logics

Semantic Web Layers

URI/IRI
Universal Resource Identifier
Internationalized Resource Identifier

XML
eXtended Markup Language

RDF
Resource Description Framework

RDFS
RDF Schema

OWL
Web Ontology Language

SPARQL
Simple Protocol and RDF Query Language

RDF and RDFS

- RDF stands for

Resource Description Framework

- is a W3C standard, which provides tool to describe Web resources
- provides interoperability between applications that exchange machine-understandable information

RDF and RDFS

- RDFS extends RDF with “schema vocabulary”, e.g.:
  - Class, Property
  - type, subClassOf, subPropertyOf
  - range, domain

The RDF Data Model

- Statements are

<subject, predicate, object> triples:

```
<Ian,hasColleague,Uli>
```

- Can be represented as a graph:

  - Statements describe properties of resources
  - A resource is any object that can be pointed to by a URI:
    - a document, a picture, a paragraph on the Web;
    - http://www.cs.man.ac.uk/index.html
    - isbn://5031-4444-3333
    - …
  - Properties themselves are also resources (URIs)

Linking Statements

- The subject of one statement can be the object of another
- Such collections of statements form a directed, labelled graph

RDF Syntax

- ***Subject*** of an RDF statement is a resource
- ***Predicate*** of an RDF statement is a property of a resource
- ***Object*** of an RDF statement is the value of a property of a resource

RDF Example

- *Ora Lassila is the creator of the resource http://www.w3.org/Home/Lassila*

```
<rdf:RDF>
  <rdf:Description about=
      "http://www.w3.org/Home/Lassila">
    <s:Creator>Ora Lassila</s:Creator>
  </rdf:Description>
</rdf:RDF>
```

RDFS Examples

- RDF Schema terms (just a few examples):
  - Class
  - Property
  - type
  - subClassOf
  - range
  - domain
- These terms are the RDF Schema building blocks (constructors) used to create vocabularies:

```
<Person,type,Class>
<hasColleague,type,Property>
<Professor,subClassOf,Person>
<Carole,type,Professor>
<hasColleague,range,Person>
<hasColleague,domain,Person>
```

From RDF to OWL

- OWL is a language for defining Web Ontologies and their associated Knowledge Bases

Description Logics (DLs)

- A key feature of OWL is its basis in Description Logics
- DL is family of logic-based knowledge representation formalisms that have a formal semantics based on first-order logic (FOL).

Description Logics (Ontology)

- **TBox T: defining terminology of application domain**
  - Inclusion assertion on concept :$C \sqsubseteq D$
    - Pericardium $\sqsubseteq$ Tissue $\sqcap$ □ part-of.Heart
  - Inclusion assertion on roles: $R \sqsubseteq S$
    - Part-of $\sqsubseteq$ has-location
- **ABox A: stating facts about a specific "world"**
  - membership assertion: C(a) or R(a,b)
    - HappyMan(Bob), HasChild(Bob, Mary)

OWL Example

- There are two types of animals, Male and Female.

```
<rdfs:Class rdf:ID="Male">
  <rdfs:subClassOf rdf:resource="#Animal"/>
</rdfs:Class>
```

- The subClassOf element asserts that its subject - Male - is a subclass of its object -- the resource identified by #Animal.

```
<rdfs:Class rdf:ID="Female">
  <rdfs:subClassOf rdf:resource="#Animal"/>
  <owl:disjointWith rdf:resource="#Male"/>
</rdfs:Class>
```

- Some animals are Female, too, but nothing can be both Male and Female (in this ontology) because these two classes are disjoint (using the disjointWith tag).

OWL Example in Protégé

- Class
  - Person superclass
  - Man, Woman subclasses
- Properties
  - isWifeOf, isHusbandOf
- Property characteristics, restrictions
  - inverseOf
  - domain
  - range
  - Cardinality
- Class expressions
  - disjointWith

OWL Example in Protégé

MyOntology Protégé 2.0 beta (file:/C:/ellisr/ontology/MyOntology.pprj, OWL files)

Project Edit Window OWL Help

OWLClasses | Properties | Forms | Individuals | Ontology

Class Hierarchy

- :THING
  - Person
    - Woman
    - Man

Superclasses

- Person

Woman (type=owl:Class)

Name | Labels | SameAs | Different

Woman

Documentation

Annotations

Property	Value

Properties at Class

Name	Type	Cardinality	Other Facets
isWifeOf	Instance	single	classes={Man}

Restrictions

Property	Restriction	Filler

Definition

Disjoint classes

- Man

OWL Example in Protégé

MyOntology Protégé 2.0 beta (file:/C:/ellisr/ontology/MyOntology.pprj, OWL files)

Project Edit Window OWL Help

OWLClasses | Properties | Forms | Individuals | Ontology

Properties

- isHusbandOf
- isWifeOf

isHusbandOf (type=owl:ObjectProperty)

Name | Labels | SameAs | DifferentFrom

isHusbandOf

Documentation

Annotations

Property	Value

Cardinality

required at least

multiple at most 1

HasValue

Equivalent

Range

Instance

Some Values From

Domain defined

Classes ⊔

Woman

Domain ⊔

Man

Inverse Property

isWifeOf

Symmetric

Transitive

AnnotationProperty

InverseFunctional

Ontology Languages Summary

- RDF is a flexible data model for Semantic Web
- RDF Schema provides simple inference capability
- OWL allows more expressive representation of knowledge but is hard to scale to Web data
- Semantic technologies have been adopted by major companies such as Google, Yahoo and Facebook

Application Areas of Ontologies

- Information Retrieval
  - As a tool for intelligent search instead of keyword matching
  - Cross Language Information Retrieval
  - Improve recall by query expansion through the synonymy
  - Improve precision through Word Sense Disambiguation (identification of the relevant meaning of a word in a given context among all its possible meanings)
- Information Integration
  - Seamless integration of information from different websites and databases
- Knowledge Engineering and Management
  - As a knowledge management tools for selective semantic access (meaning oriented access)
- Natural Language Processing
  - Better machine translation
  - Queries using natural language

Semantic Search

Research in Life Sciences

As a researcher in pharmaceutical industry, I want to plan new experiments more efficiently.
I want to know what's already available.
I'm interested in former experiments where

- certain genes were tested
- under specific treatment conditions
- in a target therapeutic area
- with help from categorisation systems like 'disease hierarchies'

→ Linking Structured to Unstructured Data and to Industry Knowledge Graphs

UniProt, ChEMBL

Experiments Documentation

MeSH

DrugBank

Data and text mining of electronic health records

https://www.slideshare.net/larsjuhljensen/data-and-text-mining-of-electronic-health-records-55027555

Covid 19 datasets

Sentiment Analysis

- **Voices**: drive actionable intelligence from member voices…

Ontology Engineering

- Collaborative design of ontologies
- Ontology reuse
  - Fusion or Merging: Process of unifying various ontologies
  - Composition or Integration: Process of selecting parts of certain ontologies
- Engineering Paradigms
  - Mathematical, Cognitive, Linguistic, Computer Science, Domain specialist

Reusing Ontologies

State of art

- Ontology Learning
- Knowledge graph
- Linked open data

Ontology Learning

Automated generation of ontologies from natural language text including:

- Term extraction:
  - Mostly run a part-of-speech tagger over the domain corpus used for the ontology learning task and identifying terms over manually constructed ad hoc patterns
- Taxonomy induction:
  - Concept learning should provide an intentional definition of the concept, an extensional set of concept instances, and a set of linguistic realizations for the concept.
  - Population of existing concepts with instances is typically referred to as **Ontology Population.**
- Relation learning:
  - Concept Hierarchy – set of taxonomic relationships among classes
  - Relation extraction is related to the problem of acquiring selection restrictions for verb arguments

Knowledge is the key

Knowledge Graph

- **ontology + data = knowledge graph**
- Information extraction techniques for performing named entity recognition. Another subfield which has gained much interest from the community is keywords extraction.
- A **knowledge graph** is a way of storing data that resulted from an information extraction task. Many basic implementations of knowledge graphs make use of a concept we call **triple**, that is a set of three items(a subject, a predicate and an object)

Example

London is the capital of England.
Westminster is located in London.

After some basic processing which we will see later, we would 2 triples like this:
(London, be capital, England),
(Westminster, locate, London)

Knowledge graph representation

Building Knowledge Graph

- Identifying the entities and the relation between them is not a difficult task for us. However, **manually building a knowledge graph is not scalable.**
- Nobody is going to go through thousands of documents and extract all the entities and the relations between them!
- **To build a knowledge graph from the text, it is important to make our machine understand natural language**.
- This can be done by using NLP techniques such as sentence segmentation, dependency parsing, parts of speech tagging, and entity recognition

Do machines understand user intent? Do they have enough context?

1. Intent recognition
2. Entity linking
3. Background knowledge

Car	Loadspace	Max no. of seats
KIA Sorento	550 litres	7
Peugeot 5008	823 litres	7
BMW X3	550 litres	5

poolparty

Artificial "Intelligence"

Background knowledge is key

Support complex Q&A:

Which cities located in the Commonwealth of Nations have a population of more than 2 mio. people?

Avoid illogical answers:

Knowledge is the key

- **Semantic descriptions** of **entities** and their **relationships**
- Uses a **knowledge representation formalism** (Focus here: RDF, RDF-Schema, OWL)

- **Entities**: real world objects (things, places, people) and abstract concepts (genres, religions, professions)
- **Relationships**: graph-based data model where relationships are first-class
- **Semantic descriptions**: types and properties with a well-defined meaning (e.g. through an ontology)
- Possibly axiomatic knowledge (e.g. rules) to support automated reasoning

Knowledge Graph

Applications

- Semantic Search
- Analytics
- Knowledge Sharing
- Question Answering
- Dashboards
- Knowledge Management

Algorithms

- Inferencing
- Entity Recognition
- Text Understanding
- Machine Learning
- Disambiguation
- Recommendations

Knowledge Graph

- Entities
- Relationships
- Semantic Descriptions

Data Sources

Data Transformation, Integration
Natural Language Processing

Knowledge Graphs for **Data Integration & Analytics**

Metadata enrichment, linked data, text mining, entity-centric search, agile reporting

Knowledge Graph

https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/

Google Knowledge Graph

Knowledge Graph Adaption

Largest change in market cap by company (2009 to 31 March 2018)

	Company name	Location	Industry	Change in market cap 2009-2018 ($bn)	Market cap 2018 ($bn)
1	Apple	United States	Technology	757	851
2	Amazon.Com	United States	Consumer Services	670	701
3	Alphabet	United States	Technology	609	719
4	Microsoft Corp	United States	Technology	540	703
5	Tencent Holdings	China	Technology	483	496
6	Facebook	United States	Technology	383(1)	464
7	Berkshire Hathaway	United States	Financial	358	492
8	Alibaba	China	Consumer Services	302(1)	470
9	JPMorgan Chase	United States	Financials	275	375
10	Bank of America	United States	Financials	263	307

Known knowledge graph builders

Operator of Taobao and KG builder

Known KG builders

Linked Open Data Project

- Goal: “expose” open datasets in RDF
- Set RDF links among the data items from different datasets
- Set up, if possible, query endpoints

Example data source: DBpedia

- DBpedia is a community effort to
  - extract structured (“infobox”) information from Wikipedia
  - provide a query endpoint to the dataset
  - interlink the DBpedia dataset with other datasets on the Web

Freie Universität Berlin

UNIVERSITÄT LEIPZIG

OPENLINK SOFTWARE

Extracting structured data from Wikipedia

```
@prefix dbpedia <http://dbpedia.org/resource/>.
@prefix dbterm  <http://dbpedia.org/property/>.

dbpedia:Amsterdam
  dbterm:officialName "Amsterdam" ;
  dbterm:longd "4” ;
  dbterm:longm "53" ;
  dbterm:longs "32” ;
  dbterm:leaderName dbpedia:Lodewijk_Asscher ;
  ...
  dbterm:areaTotalKm "219" ;
  ...
dbpedia:ABN_AMRO
  dbterm:location dbpedia:Amsterdam ;
  ...
```

Amsterdam

The Keizersgracht at dusk

Location of Amsterdam

Coordinates: 52°22'23"N 4°53'32"E

Country	Netherlands
Province	North Holland
Government	
- Type	Municipality
- Mayor	Job Cohen[1] (PvdA)
- Aldermen	Lodewijk Asscher Carolien Gehrels Tjeerd Herrema Maarten van Poelgeest Marijke Vos
- Secretary	Erik Gerritsen
Area [2][3]	
- City	219 km² (84.6 sq mi)
- Land	166 km² (64.1 sq mi)
- Water	53 km² (20.5 sq mi)
- Urban	1,003 km² (387.3 sq mi)
- Metro	1,815 km² (700.8 sq mi)
Elevation [4]	2 m (7 ft)
Population (1 October 2008)[5][6]	
- City	755,269
- Density	4,459/km² (11,548.8/sq mi)
- Urban	1,364,422
- Metro	2,158,372
- Demonym	Amsterdammer
Time zone	CET (UTC+1)
- Summer (DST)	CEST (UTC+2)
Postcodes	1011 – 1109
Area code(s)	020

Website: www.amsterdam.nl

Automatic links among open datasets

```
<http://dbpedia.org/resource/Amsterdam>
  owl:sameAs <http://rdf.freebase.com/ns/...> ;
  owl:sameAs <http://sws.geonames.org/2759793> ;
  ...
```

```
<http://sws.geonames.org/2759793>
  owl:sameAs <http://dbpedia.org/resource/Amsterdam>
  wgs84_pos:lat "52.3666667" ;
  wgs84_pos:long "4.8833333";
  geo:inCountry <http://www.geonames.org/countries/#NL> ;
 ...
```

Processors can switch automatically from one to the other…

The LOD “cloud”

Wikipedia Related
Other
proven...
sandra...
Web Sc...
Schema...
WordLi...
Linked...
Gemein...
xxxxx
ICANE
Italia...
Learni...
DataGo...
Europe...
URIBurner
Copyri...
The Co...
mEduca...
WarSampo
EU: fi...
WordLi...
Linked...
Open E...
Freebase
OpenCyc
Linked...
Judaic...
WebIsALOD
Images...
Serend...
ISPRA ...
Suface...
Projet...
DBpedi...
UMBEL ...
YAGO
DBpedi...
DBpedi...
Nobel ...
DBkWik
ECCO-T...
Whisky...
Geogra...
Linked...
The Or...
DBpedi...
GeoEcu...
Linked...
EventKG
iServe...
RDFLic...
SmartL...
Source...
DBpedia
Lexvo.org
Alpine...
Eniped...
associ...
Linked...
Brazil...
LinkLi...
Open D...
Muninn...
DBpedi...
DBpedi...
Ruben ...
Ruben ...
NPM
Temple...
Semant...
dbpedi...
OpenCa...
Instan...
World ...
JRC-Na...
Produc...
Linked...
Open L...
W3C
Uberbl...
interv...
Influe...
vulner...
Framester
Google...
EPA-FRS
QBOAir...
DBpedi...

Lots of Tools (*not* an exhaustive list!)

- Categories:
  - Triple Stores
  - Inference engines
  - Converters
  - Search engines
  - Middleware
  - CMS
  - Semantic Web browsers
  - Development environments
  - Semantic Wikis
  - …

- Some names:
  - Jena, AllegroGraph, Mulgara, Sesame, flickurl, …
  - TopBraid Suite, Virtuoso environment, Falcon, Drupal 7, Redland, Pellet, …
  - Disco, Oracle 11g, RacerPro, IODT, Ontobroker, OWLIM, Talis Platform, …
  - RDF Gateway, RDFLib, Open Anzo, DartGrid, Zitgist, Ontotext, Protégé, …
  - Thetus publisher, SemanticWorks, SWI-Prolog, RDFStore…
  - …

References

- Berners-Lee, Hendler, Lassila (May 17, 2001). The Semantic Web. Scientific American Magazine. @ http://bit.ly/2z2LZIE
- Berners-Lee. A roadmap to the Semantic Web. Tim Berners-Lee's Design Issues @ http://bit.ly/2z29fGP
- Berners-Lee. Linked Data. Tim Berners-Lee's Design Issues @ http://bit.ly/21MR3Zt
- Heath, Bizer. Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology @ http://bit.ly/2xlZKou
- https://wiki.dbpedia.org/services-resources/ontology
- https://www.emse.fr/~zimmermann/Teaching/SemWeb/semwebintro.pdf
- https://www.youtube.com/watch?v=C9M7N979pIU
- https://enterprise-knowledge.com/whats-the-difference-between-an-ontology-and-a-knowledge-graph/
- https://www.w3.org/Consortium/facts
- https://pages.semanticscholar.org/coronavirus-research

References

- https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/
- Google Knowledge Graph
- Building Knowledge graph
- https://www.youtube.com/watch?v=lUc0woFX16M
- https://www2.slideshare.net/jeffpan_sw/linked-data-and-knowledge-graphs-constructing-and-understanding-knowledge-graphs?qid=ab5124c7-a15a-4f77-9fc9-914feac00a6a&v=&b=&from_search=1
- https://www2.slideshare.net/phaase/getting-started-with-knowledge-graphs?from_action=save
- https://lod-cloud.net/
- https://www.w3.org/TR/2012/REC-owl2-syntax-20121211/