Introduction to Knowledge Graphs and the Semantic Web
Questions

• What are Knowledge Graphs
• What is the Semantic Web?
• How are they related?
• How are they being used today?
• What can we expect in the future?
Web is our greatest knowledge source
But it has limitations
Designed for people, not machines
Designed for people, not machines
- Content is mostly text, spoken language, images and videos
- Easy for people to understand
- But hard for machines

Machines need access to this knowledge too
Vannevar Bush envisioned a hypertext/IR system in 1945.
Access is primarily via information retrieval
• Key-word queries → ranked document list
• We still need to read the documents or watch the videos
• We often want an answer to a question: where is the Census Big Data Day event
And so do our machines and apps

Vannevar Bush envisioned a hypertext/IR system in 1945
We need to add knowledge graphs
We need to add knowledge graphs

• High quality semi-structured information about entities and relations
• Represented and accessed via Web standards
• Easily integrated, fused and reasoned with
State of the Art?

Google is a good example, but Microsoft, IBM, Apple and Facebook all have similar capabilities

- 2010 Google acquired MediaWeb and its Freebase KB
- 2014: Freebase: 1.2B facts about 43M entities
- 2015+: Google knowledge graph, updated by text IE

DBpedia open source RDF KB is another

- 800M facts about 4.6M subjects from English Wikipedia, data available in 21 other languages
- Helps integrate 90B facts from 1000 RDF datasets in the linked data cloud
Ask: When was Tom Sawyer written?

The Adventures of Tom Sawyer - Published: 1876

Novel by Mark Twain

- Originally published: 1876
- Author: Mark Twain
- Text: The Adventures of Tom Sawyer at Wikisource
- Cover artist: Created by Mark Twain
- Characters: Tom Sawyer, Huckleberry Finn, Becky Thatcher, Aunt Polly, Joe Harper, Sid Sawyer
- Genres: Bildungsroman, Picaresque, Fiction, Satire, Folklore, Children's literature
- Followed by: Wuthering Heights, The Prince and the Pauper
Many commercial recipe sites on Web
Most recipe sites embed **semantic data** about their recipes in an RDF-compatible form using terms from the **schema.org** ontology.

Search engines read and use this data to better understand the semantics of the page content.
Conversational Bots

Voice-driven conversational systems like Amazon Echo and Google Home use knowledge graphs to help understand our requests.
Where does the knowledge come from?

• Knowledge graphs like *DBpedia* and *Freebase* started with *Wikipedia* data encoded in custom ontologies

• Semantic Web technologies are an open source way to encode the knowledge

• They are and will continue to evolve

• Current: extract data from text documents, e.g., articles, newswire, social media, etc.
Who invented the Web?
Who invented the Web?
Semantic Web Origin

Tim Berners-Lee’s original 1989 proposal described a web of relationships among named objects unifying many information management tasks.

Capsule history

- Guha’s MCF (~94)
- XML+MCF=>RDF (~96)
- RDF+OO=>RDFS (~99)
- RDFS+KR=>DAML+OIL (00)
- W3C’s SW activity (01)
- W3C’s OWL (03)
- ...

http://www.w3.org/History/1989/proposal.html
W3C’s Semantic Web Goals

Focus on machine consumption:

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

Why is this hard?

after Frank van Harmelen and Jim Hendler
What this looks like to a machine...

...after Frank van Harmelen and Jim Hendler
OK, so HTML is not helpful

Maybe we can tell the machine what the different parts of the text represent?

<table>
<thead>
<tr>
<th>title</th>
<th>speaker</th>
<th>time</th>
<th>location</th>
<th>abstract</th>
<th>biosketch</th>
<th>host</th>
</tr>
</thead>
</table>

- **Title**: Χομμουνιστικός Συνέδριο Συνεδρίου Διεθνών Μετατροπής το Μοντέλο-Αγώνα, Real-Time Ναυαρχείονα Προφητείας Ακτη Μαρίων Ελλάδας

- **Speaker**: Μαρίων Δομεταράκα

- **Time**: 1:00-2:00 ωρών, Φεβρουάριος 14, 2002

- **Location**: Συνεδρία Συνεδρίου Διεθνών Μετατροπής το Μοντέλο-Αγώνα, Real-Time Ναυαρχείονα Προφητείας Ακτη Μαρίων Ελλάδας

- **Abstract**: Η ευρετηρία των ανθρώπινων δεδομένων έχει ανακηρυχθεί μετά από πολλά χρόνια με τη σειρά της συνέχειας που ξεκινάει από την εποχή της ανακάλυψης των ανθρώπινων δεδομένων. Η ευρετηρία των ανθρώπινων δεδομένων προσφέρει έναν απαραίτητο άξονα της μεταφοράς των γνώσεων Ελλάδας και της Βαλτιμόρης, με αποτέλεσμα να καταστούν σημαντικότερη και πιο διάσπαρτη σε μια εποχή ιστορικής μεταφοράς. Η ευρετηρία των ανθρώπινων δεδομένων επιβραβεύεται με έναν έμβλημα για την ευεργεσία της επιστημονικής μελέτης και την επικοινωνία με τους μελέτηδες των ανθρώπινων δεδομένων. Η ευρετηρία των ανθρώπινων δεδομένων προσφέρει μία πληροφορική σύνολο που ευγενίζεται από την επιστημονική δραστηριότητα και την ευεργεσία της επιστημονικής μελέτης.

- **Biosketch**: Μαρίων Δομεταράκα είναι οπτικός στην ΥΜΕΘ. Είναι έμπειρος σε διάφορες θέσεις διεύθυνσης και διαχείρισης προγραμμάτων για μεταφορά και διαχείριση αποθεμάτων και μεταφορά αποθεμάτων σε διαφορετικές σχέσεις. Η διεύθυνση και η διαχείριση της Προγραμματικής Εκπαίδευσης της ΥΜΕΘ είναι κλειδιά για την επιτυχία της επιτυχίας της ΥΜΕΘ και της Ακτής Eλλάδας.
XML fans propose creating a XML tag set to use for each application.

For talks, we can choose <title>, <speaker>, etc.
XML ≠ machine accessible meaning

But, to your machine, the tags still look like this....

The tag names carry no meaning.

XML DTDs and Schemas have little or no semantics.

after Frank van Harmelen and Jim Hendler
XML Schema helps

XML Schemas provide a simple mechanism to define shared vocabularies.

after Frank van Harmelen and Jim Hendler
But there are many schemas

after Frank van Harmelen and Jim Hendler
There’s no way to relate schema.

Either manually or automatically.

XML Schema is weak on semantics.
An Ontology level is needed

XML Ontology 256

Ontologies add
• Structure
• Constraints
• Mappings

XML Ontology 1

XML Ontology 42

We need a way to define ontologies in XML
So we can relate them
So machines can understand (to some degree) their meaning
Semantic Web

Use Semantic Web Technology to publish shared data & knowledge

Semantic web technologies allow machines to share data and knowledge using common web language and protocols.

~ 1997

Semantic Web beginning
Semantic Web => Linked Open Data

Use Semantic Web Technology to publish shared data & knowledge

Data is inter-linked to support integration and fusion of knowledge

LOD beginning
The node in the center is DBpedia

http://dbpedia.org/page/University_of_Maryland,_Baltimore_County
Semantic Web => Linked Open Data

Use Semantic Web Technology
to publish shared data &
knowledge

Data is inter-linked to support integration and fusion of knowledge

LOD growing
Semantic Web => Linked Open Data

Use Semantic Web Technology
to publish shared data &
knowledge

Data is inter-
linked to support inte-
gration and fusion of knowledge

... and growing
Linked Open Data

Use Semantic Web Technology to publish shared data & knowledge

LOD is the new Cyc: a common source of background knowledge

Data is inter-linked to support integration and fusion of knowledge

2010...growing faster
Linked Open Data

Use Semantic Web Technology to publish shared data & knowledge

LOD is the new Cyc: a common source of background knowledge

Data is inter-linked to support integration and fusion of knowledge

2011: 31B facts in 295 datasets interlinked by 504M assertions on ckan.net
Semantic Web: 1, 2, 3

Languages typically divided into three parts:

1. Syntax: legal forms that make up the sentences in a language

2. Semantics: mapping of sentences to meaning (perhaps truth theoretic)

3. Pragmatics: everything else (how to do things with language, knowledge of world, etc.)
1: Syntax

• **URIs** denote classes, properties, objects, relations
  - http://live.dbpedia.org/resource/Alan_Turing
  - http://schema.org/Person
  - http://www.w3.org/1999/02/22-rdf-syntax-ns#type

• Use strings for literals

• Use **triples** to make statements
  - `dbpedia:Alan_Turing  rdfs:type  schema:Person .`
  - “Alan Turing is a Person”

*URI = Uniform Resource Identifier*
Semantics maps URIs to the things they denote in “the world”.

Some of this is in your mind or in how you write your program.

Meaning of some URIs allow *inference*:
- parent relation is *inverse* of the child relation
- schema:parent owl:inverse schema:child
3: Pragmatics

• Semantics is more than just about truth (statements that assert things)
• Must account for commands, requests, questions, context, etc.
  - Some handled by Web protocols (GET, POST)
  - Some by special protocols (e.g., SPARQL queries)
  - Some by having reference KBs of the world (e.g., DBpedia) to help identify common entities
Where are we

• The W3C version of the open semantic web has been growing steadily

• Languages and standards are being used in
  - BBC uses RDF to make up much of its content online
  - Google and Facebook detect AND MAKE USE OF (some) RDF embedded in html pages
  - Google, Yahoo, Microsoft and Yandex formed Schema.org to develop useful vocabularies
  - Data.gov has many datasets in RDF
Wikipedia data in RDF

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbpedia-owl:almaMater</td>
<td>dbpedia:King's_College_Cambridge</td>
</tr>
<tr>
<td></td>
<td>dbpedia:Princeton_University</td>
</tr>
<tr>
<td>dbpedia-owl:award</td>
<td>dbpedia:Royal_Society</td>
</tr>
<tr>
<td></td>
<td>dbpedia:Order_of_the_British_Empire</td>
</tr>
<tr>
<td></td>
<td>dbpedia:Fellow_of_the_Royal_Society</td>
</tr>
<tr>
<td></td>
<td>dbpedia:Officer_of_the_Order_of_the_British_Empire</td>
</tr>
<tr>
<td>dbpedia-owl:birthDate</td>
<td>1912-06-23 (xsd:date)</td>
</tr>
<tr>
<td></td>
<td>1912-06-23 (xsd:date)</td>
</tr>
<tr>
<td>dbpedia-owl:birthName</td>
<td>Alan_Mathison_Turing</td>
</tr>
<tr>
<td>dbpedia-owl:birthPlace</td>
<td>dbpedia:Paddington</td>
</tr>
<tr>
<td></td>
<td>dbpedia:Maida_Vale</td>
</tr>
<tr>
<td>dbpedia-owl:deathDate</td>
<td>1954-06-07 (xsd:date)</td>
</tr>
<tr>
<td>dbpedia-owl:doctoralStudent</td>
<td>dbpedia:Robin_Gandy</td>
</tr>
<tr>
<td>dbpedia-owl:doctoralAdvisor</td>
<td>dbpedia:Alonzo_Church</td>
</tr>
<tr>
<td>dbpedia-owl:field</td>
<td>dbpedia:Computer_science</td>
</tr>
<tr>
<td></td>
<td>dbpedia:Mathematics</td>
</tr>
<tr>
<td></td>
<td>dbpedia:Cryptanalysis</td>
</tr>
<tr>
<td>dbpedia-owl:individualisedPnd</td>
<td>118802976</td>
</tr>
<tr>
<td>dbpedia-owl:knownFor</td>
<td>dbpedia:Turing_machine</td>
</tr>
<tr>
<td></td>
<td>dbpedia:Cryptanalysis_of_the_Enigma</td>
</tr>
<tr>
<td></td>
<td>dbpedia:Automatic_Computing_Engine</td>
</tr>
<tr>
<td></td>
<td>dbpedia:Turing_test</td>
</tr>
</tbody>
</table>

dbpedia:Alan_Turing dbpedia-owl:doctoralAdvisor dbpedia:Alonzo_Church.
**Wikidata**

- **Wikidata** aims to create an rdf-like KG that can be read/edited by humans & machines
  - Wikimedia project started in April 2012
- Wikidata clients use the repository, e.g., to populate Web pages or Wikipedia infoboxes
- Based on ideas from [Semantic MediaWiki](https://semanticmedia.org) and [Freebase](https://freebase.com)
Semantic Media Wiki

Semantic MediaWiki (SMW) is a free, open-source extension to MediaWiki – the wiki software that powers Wikipedia – that lets you store and query data within the wiki's pages. Semantic MediaWiki is also a full-fledged framework, in conjunction with many spinoff extensions, that can turn a wiki into a powerful and flexible "collaborative database". All data created within SMW can easily be published via the Semantic Web, allowing other systems to use this data seamlessly.

More about Semantic MediaWiki
- Introduction to SMW
- FAQ
- Talks and publications
- Testimonials

Wiki of the Month - January 2013

Triple A is a site that provides IT architecture information for educational organizations. It supports MBO-scholen (Dutch secondary schools) wanting to innovate educational processes.

Installation
- Semantic MediaWiki 1.8
- Administrator manual
- Installation
- Configuration
- Related extensions

User community
- Getting support
- SMW Community Wiki
- List of SMW-using wikis
- SMWCon, the Semantic MediaWiki Conference

The SMW project
- About the project
- Version history
- Development roadmap
- Programmer's guide

News
- RSS
- Atom

10 January 2013: Survey: What do you think of SMW?
Freebase

“An entity graph of people, places and things, built by a community that loves open data”

Acquired by Google in 2010
Google Knowledge Graph

Google’s slogan for the knowledge graph: “things, not strings”
Who wrote Tom Sawyer?
Facebook Open Graph

Annotate your web pages in RDFa

=> object in the FB graph

Introduction

The Open Graph protocol enables any web page to become a rich object in a social graph. For instance, this is used on Facebook to allow any web page to have the same functionality as any other object on Facebook.

While many different technologies and schemas exist and could be combined together, there isn’t a single technology which provides enough information to richly represent any web page within the social graph. The Open Graph protocol builds on these existing technologies and gives developers one thing to implement. Developer simplicity is a key goal of the Open Graph protocol which has informed many of the technical design decisions.

Basic Metadata

To turn your web pages into graph objects, you need to add basic metadata to your page. We’ve based the initial version of the protocol on RDFa which means that you’ll place additional <meta> tags in the <head> of your web page. The four required properties for every page are:

- og:title - The title of your object as it should appear within the graph, e.g., "The Rock".
- og:type - The type of your object, e.g., "video.movie". Depending on the type you specify, other properties may also
Apple’s SIRI

SIRI needs lots of semantic data about entities in the world

SIRI engineers from AI/SW community

speech => text => entities => task
IBM’s Watson

IBM used Semantic Web technology and data in Watson, see http://bit.ly/X44alE
Schema.org

A collection of useful ontologies

Embed in HTML using RDFa to make machine understand-able statements

---

**Thing > Person**

A person (alive, dead, undead, or fictional).

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>additionalType</td>
<td>URL</td>
<td>An additional type for the item, typically used for adding more specific types from external vocabularies in microdata syntax. This is a relationship between something and a class that the thing is in. In RDFa syntax, it is better to use the native RDFa syntax - the 'typeof' attribute - for multiple types. Schema.org tools may have only weaker understanding of extra types, in particular those defined externally.</td>
</tr>
<tr>
<td>alternateName</td>
<td>Text</td>
<td>An alias for the item.</td>
</tr>
<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>sameAs</td>
<td>URL</td>
<td>URL of a reference Web page that unambiguously indicates the item's identity. E.g. the URL of the item's Wikipedia page, Freebase page, or official website.</td>
</tr>
<tr>
<td>url</td>
<td>URL</td>
<td>URL of the item.</td>
</tr>
</tbody>
</table>

Properties from Person

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>additionalName</td>
<td>Text</td>
<td>An additional name for a Person, can be used for a middle name.</td>
</tr>
<tr>
<td>address</td>
<td>PostalAddress</td>
<td>Physical address of the item.</td>
</tr>
<tr>
<td>affiliation</td>
<td>Organization</td>
<td>An organization that this person is affiliated with. For example, a school/university, a club, or a team.</td>
</tr>
<tr>
<td>alumnusOf</td>
<td>EducationalOrganization</td>
<td>An educational organizations that the person is an alumni of.</td>
</tr>
</tbody>
</table>
Summary

• Web has made us smarter by sharing information and knowledge as text, audio and images
• Machines should also be able to use the Web to publish & retrieve information & knowledge
• Human forms of knowledge are hard for machines to understand and generate
• The Semantic Web is a collection of languages, ontologies, software tools, services and KBs that are designed to support machines