1 Core Semantic Web Technologies

1.1 Introduction to Semantic Web

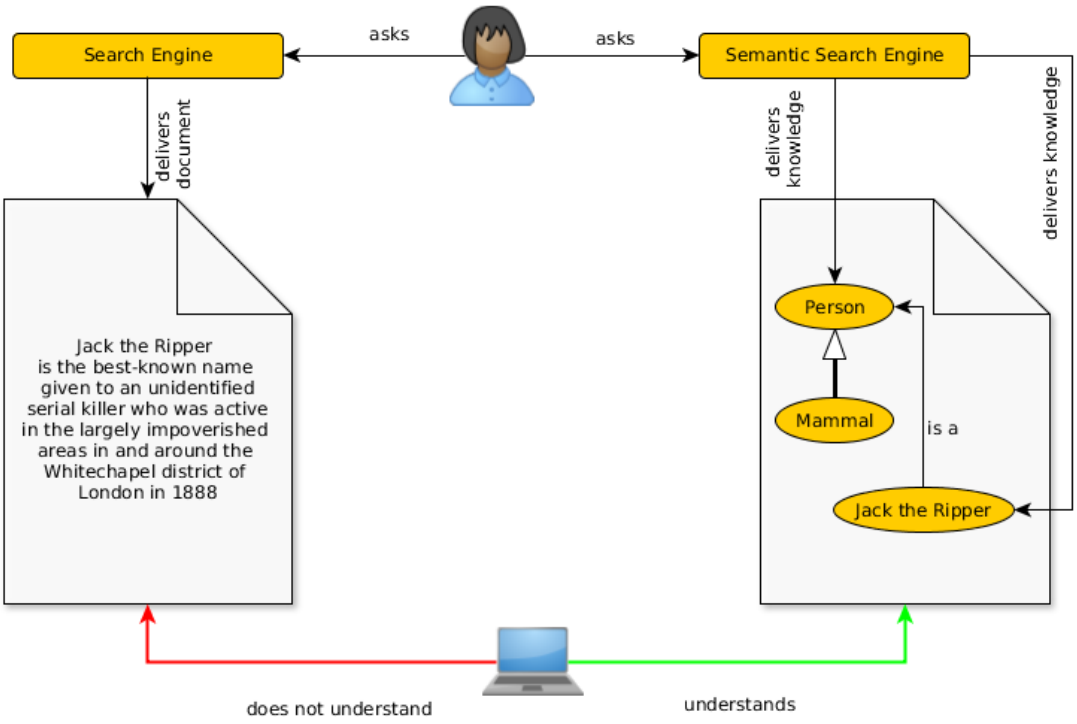
Current Web vs. Semantic Web

- SoA – semistructured HTML or XML data. There is vast amount of search engines like Google, Yahoo, MSN, etc. Many of them are invaluable, but as the engines use just keywords and/or some natural language preprocessing methods, the search results contain lots of irrelevant results that need to be processed manually.

- How to make web search more efficient ?
  - more expressive power for web designers to capture complexities – SW languages (RDF(S), OWL),
  - more efficient search engines to handle SW languages – new inference techniques for these languages,
  - better search engines interfaces – more expressive query languages

- **the amount of (unstructured) data is steadily growing**

Semantic search



Ontologies and Semantic Web

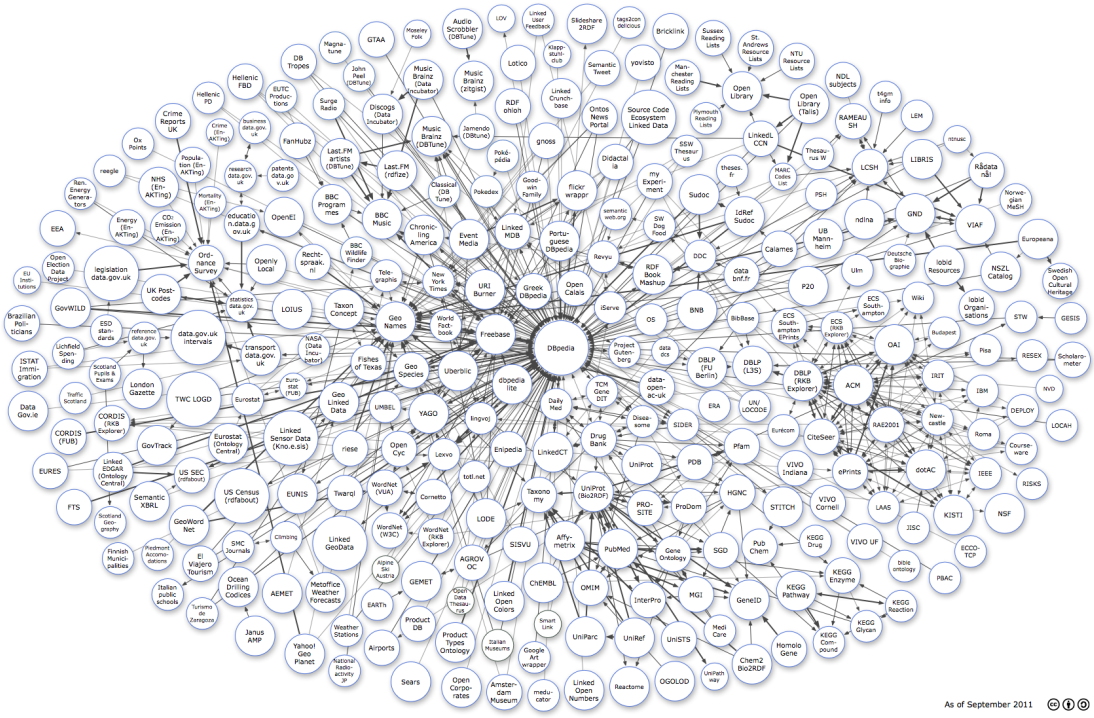
ontology has many definitions, but let's consider it **a formal representation of a complex domain knowledge that is shared with others to ensure intelligent system interoperability,**

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.* (cit. Semantic Web. Tim Berners-Lee, James Hendler and Ora Lassila, Scientific American, 2001)

Idea of Semantic Web

- W3C web page - `http://www.w3.org/2001/sw`
- The data format will be either RDF(S) or OWL,
- Reasoners for RDF(S) can be used for partial derivation in OWL,
- Reasoners for OWL can be used for derivation in RDF(S)

Linked Data



"Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/"

1.1.1 Semantic Web Adopters

Who is Using Semantic Web Technologies

Let's name a few:

- Google – *Knowledge Graph* (although they do not name it Semantic web – `http://semanticweb.com/google-just-hi-jacked-the-semantic-web-vocabulary\_b29092`)
- Microsoft – Satori, `http://research.microsoft.com/en-us/projects/trinity/query.aspx`
- Facebook – Open Graph Protocol `http://ogp.me/`
- BBC – various datasets in RDF – `http://www.bbc.co.uk/developer/technology/apis.html`
- Ordnance Survey – geographic datasets in RDF – `http://data.ordnancesurvey.co.uk`

1.1.2 Semantic Web Principles

Unique Data Identification – URIs

Semantic web speaks about resources.

URI is a unique identifier for adressing web resources in the form

```
<scheme name> : <hier. part> [ ? <query> ] [ # <fragment> ]
```

. HTTP scheme is used typically.

URN a URI with *scheme name* equal to 'urn'; used e.g. in SWRL atom identification,

URL a URI that can be resolved to a content using the protocol (e.g. HTTP),

IRI generalization of URIs allowing non-ascii characters. IRI is the standard identifier for OWL.

Open World Assumption

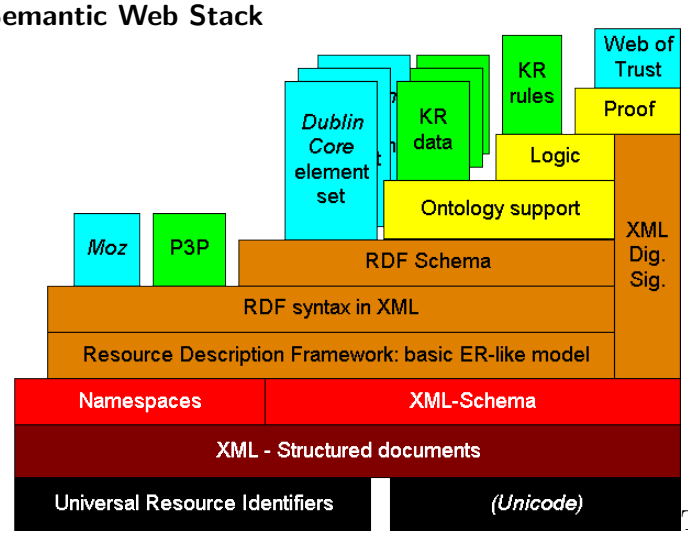
The semantic web inference must take into account that we handle *incomplete knowledge*.

Description

Open world (OWA): Everything that cannot be proven is unknown, Closed world (CWA): Everything that cannot be proven is false.

> Statement : "John is a Man." Query: "Is Jack a Man ?" OWA Answer: "I don't know." CWA Answer: "No."

Semantic Web Stack



Taken from `http://www.w3.org/2000/Talks/0906-xmlweb-tbl/slide9-0.html`, by Tim Berners Lee.