WEB OF DATA

Jelena Jovanović

Email: jeljov@gmail.com

Web: http://jelenajovanovic.net

Web of Documents - issues

The focus is on content presentation for human consumption -> markup comprises:

* rendering information (e.g., font size and color)

* hyper-links to related content

The semantics of the content is accessible to humans but not (easily) to computers

Web of Documents - issues

The conference dates

WWW 2015 conference will take place in Florence, Italy

Papers submission info

Important Dates

Research track

Abstract submission	November 3
Full Paper submission	November 10
Acceptance notification	January 17

Follow us on Facebook

WWW 2015

WWW 2015
November 3 at 5:43am

Please remember that the abstracts deadline is tonight

Plenty of information available to humans

Web of Documents - issues

Information available to programs

Web of Documents - issues

Comprehensible to machines, as much as the following menu is comprehensible to large majority of the Western hemisphere:

Web of Documents - issues

<u>The overall idea</u>: make the content of the Web 'legible' to computers, by presenting it in the language they 'understand'

Today's Web – Web of Documents

- Designed for: direct human consumption

- Primary objects: documents (including multi-media)

- Degree of structure in objects: fairly low

- Links: between documents (or sub-parts of)

- Semantics of content and links: implicit

- Analogy: a global file system

Web of Documents - challenges

- Challenge #1: Efficient data integration

  – Provide users with information by gathering (and processing) data from different, often heterogenous sources

  – Examples:

    • show me all blog posts written by my colleagues that are relevant for the study I'm currently working on

    • which music albums of the group whose song I've just heard were highly rated by my friends who share my music taste?

Web of Documents - challenges

- Challenge #2: Advanced data-driven services

  – Example: service for recommendation of ho(s)tels

    • Show me ho(s)tels in San Francisco that were highly rated by people who share my travel interests and style?

      [make use of data from as many Web sources as possible]

Web of Documents - challenges

- Chalenge #3: Adaptive RSS feeds

  – from arbitrary source, filtered by custom set of criteria

  – E.g., show me blog posts that were written by my favorite bloggers and provoked community reaction (i.e., received 10+ comments)

- Challenge #4: Ad-hoc, dynamic mash-ups

  – E.g., FreebaseMaps

Web of Data

- **Main features:**

  – Data (on the Web) is structured and interlinked

  – The semantics of data and links are made explicit

  – Allows for performing complex queries over multiple sources

  – The vision of the Web as a gigantic global database

Web of Data – example of a complex search

- Which European city has a substantial collection of Caravaggio's work?

  - ...and has direct flights from my home town?

  - ...with an airline that is rated good or excellent?

  - ...by me? ...by my friends?

Image source:
http://ceo.ca/caravaggio-and-the-myth-of-the-boring-painter/

Web of Data

- Designed for: humans served by machines

- Primary objects: resources (or descriptions of resources)

- Links: between resources

- Degree of structure in (descriptions of) things: high

- Semantics of content and links: explicit

- Analogy: a global database

Web of Docs vs. Web of Data

	Web of Documents	Web of Data
Designed for	Human consumption	Humans served by computer programs
Primary objects	Documents (including multimedia)	Resources* (and descriptions of resources)
Links between	Documents	Resources
Degree of structure	Fairly low	High
Semantics of content and links	Implicit	Explicit
Analogy	A global file system	A global database

* E.g., a resource is Belgrade, and its description is DBpedia entry for Belgrade

Web of Data

Recommendation:

Video *What is Linked Data?* available at:

http://www.youtube.com/watch?v=4x_xzT5eF5Q

Why it might be relevant?

It will help you better understand the concept of Data Web and linking data on the Web

Web of Data – challenge

- Restrictions on data access; data kept 'behind the walls'

  - High barriers to access to and integration of data from disparate sources (apps/websites)

  - Full data access only within the bounds of and with the approval of individual Web sites/applications

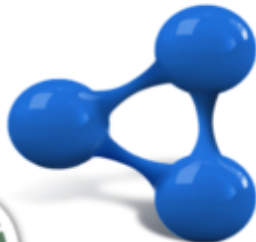 - The technology is available, (business/political) interests are the impediment

Web of **Open** Data

- Open Data initiative

  – world-wide initiative aimed at making the data on Web open, i.e., publicly available

  – the initiative spans a wide range od domains
    - government, economics, culture, art, education,…

Web of Open Data

- Open Data initiave

  – In 2012 the city of New York introduced a law ("Local Law 11 of 2012") mandating that

  > "*all internal city data that can be made public has to be made public*"

  In addition, this law states that

  > "*The city's open data has to be published in machine-readable format*"

Semantic Web

- The next step in the evolution of Web of Data

- The 'Intelligent' Web

  - Reasoning over data integrated from multiple sources

  - Web-scale recommendation systems

  - Intelligent agents doing search and discovery
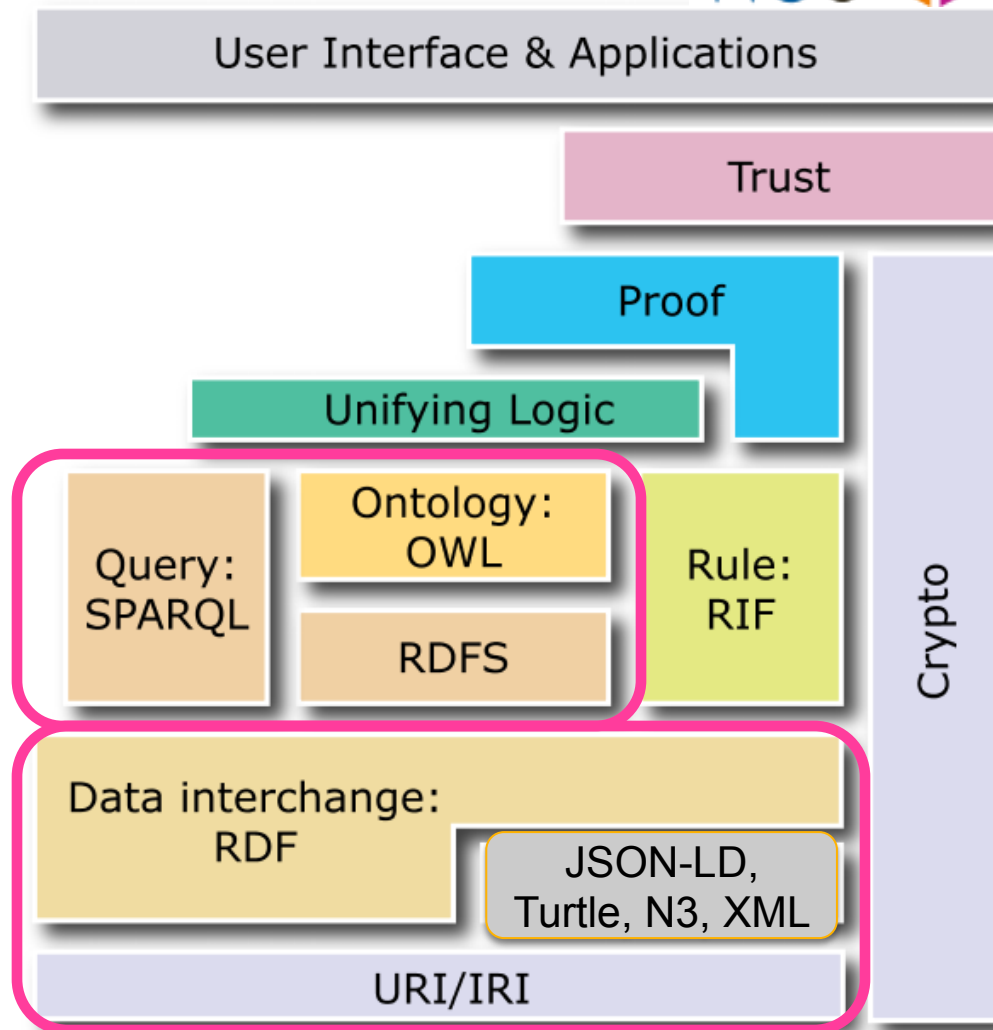
  - …

Scientific American, May 2001:

The Semantic Web in 6 minutes

Open Technologies and standards for the Web of Data / Semantic Web

Semantic Web Technology stack

| User Interface & Applications |
| Trust |
| Proof |
| Unifying Logic |
| Query: SPARQL / Ontology: OWL / RDFS / Rule: RIF / Crypto |
| Data interchange: RDF / JSON-LD, Turtle, N3, XML |
| URI/IRI |

Suggested readings

- **Semantics in the Real World: Where to Begin?**
  - a set of four articles that
    - starts with the characteristics a use case should have to be a good target for semantic technology application,
    - continues to provide surveys of semantic technology applications on the Web and in the enterprise

- Examples of benefits offered by Semantic Web technologies
  - Why Most Job Sites Miss the Boat (blog post at LinkedIn)
  - Startup Pitch: Sépage sees semantic as the answer
  - Sports are the Semantic Focus In Britain at the BBC…
  - Interview: Oracle on Data on the Web
  - Poderopedia Uses Public Data to Take on Powerful Interests in Chile

(Anonymous) questionnaire for your critique, comments, suggestions:

http://goo.gl/cqdp3l