A Semantic Web Primer

In partnership with
History & Perspective

“We might hope to see the finances of the Union as clear and intelligible as a merchant's books, so that every member of Congress and every man of any mind in the Union should be able to comprehend them, to investigate abuses, and consequently to control them.”

Thomas Jefferson, 1802
History & Perspective

Web 1.0
- Read Only
- Visual
- Human Presentation

Web 2.0
- Interactive
- Social
- Applications

Web 3.0
“Web of Data”
Graph Theory

Konigsburg, Prussia, 1735

Problem: Traverse each bridge once and only once in touring the city

Solution: Zero or two nodes of odd degree
Web 3.0 – The GGG

Facebook Graph Search
Edamam
RelFinder
Linked Life Data
SpaceTime
Diseasome
iDrove.it
Google Gets It

Google Knowledge Graph

Geosearch Demo
The Resource Description Framework (RDF)

- **http://www.example.org/~joe/contact.rdf#joesmith**
  - "Subject" resource (joesmith in RDF)

- **http://xmlns.com/foaf/0.1/homepage**
  - "Predicate" resource (FOAF relationship)

- **http://www.example.org/~joe/**
  - "Object" resource (joe's home page)

Can be a literal...

URI:
- URL
- URN
Story in RDBMS

Concept: Harrison has a dog friend named Layla whose favorite activity is chasing a ball.
Tables, Not Objects

If libraries were like relational databases
Concept: Harrison has a dog friend named Layla whose favorite activity is chasing a ball.
## Atomic Concepts

### RDBMS

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Type</th>
<th>Legs</th>
<th>Fur</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>Harrison</td>
<td>Human</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>002</td>
<td>Layla</td>
<td>Dog</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>003</td>
<td>MyBall</td>
<td>Ball</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

### OODBMS

**Concept (dog):**
- Name
- Favorite Activity

**Instance (Layla):**
- ID1: 002
- ID2: 003
Reuse

Time to Implementation Diminishes with each Phase & Domain
Power of Linking

Metcalf's Law

$N \cdot (N-1) \over 2$

Reed's Law

$2^N - N - 1$
WWW Contrast

Diagram showing the contrast between a traditional web server and a server that uses RDF and a triple store. The diagram illustrates the flow of data between a web server and HTML documents, compared to the use of RDF and a triple store, with entities linked using "href" and "foaf:knows" connections.
Quantitative

B-Tree Index: $M \times O(\log_2 N)$

Graph: $1 \times O(\log_2 N)$ – plus pointer traversal

Ref: Michel Domenjoud, blog.octo.com
Qualitative

Horizontal Scale (Complexity)

Vertical Scale (Size)


The “Ontology”

**Taxonomy**: Organizational hierarchy based upon similarities in properties

**Ontology**: Set of representational primitives (classes, attributes & relationships) with which to *model* a domain of knowledge or discourse*

*Encyclopedia of Database Systems, L. Liu et al, 2009*
## Summary

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Web 2.0</th>
<th>Web 3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unification</td>
<td>Join</td>
<td>Built-in</td>
</tr>
<tr>
<td>Retrieval</td>
<td>Keyword indexing</td>
<td>Pointer traversal</td>
</tr>
<tr>
<td>Query</td>
<td>Keyword matching</td>
<td>Inference</td>
</tr>
<tr>
<td>Citizens</td>
<td>Documents</td>
<td>Data</td>
</tr>
<tr>
<td>Polysemy &amp; Synonymy</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Construct</td>
<td>Table</td>
<td>Triple (how)</td>
</tr>
<tr>
<td>Descriptor</td>
<td>URL - where</td>
<td>URI = URL + URN (what)</td>
</tr>
<tr>
<td>Data</td>
<td>Aggregated</td>
<td>Connected</td>
</tr>
<tr>
<td>Flexibility</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

---

The diagram illustrates relationships between dynamic and meaning aspects.
Ian Robinson, Neo4J
Basic (very) Architecture

Structure-er

Unstructured Data

Inference and rendering engine

Triple-store

Web Server
The “Tower”
Not a new idea...
But It Is Growing

Ref: lod-cloud.net
Testimonials

- As is generally the case, infrastructure technologies are accepted through gradual adoption. Web 3.0 is already in action, just not on a wide enough scale to change people’s perceptions and business strategies. When acceptance reaches a critical point, it would suddenly become state of the art, triggering an avalanche of rapid adoption. We can expect this to happen within the next two years. *Boris Motusic, United Experts, May 2012,*

  While traditional, relational data warehousing and federation approaches can scale well and are effective for many core data storage and access requirements, such approaches often fail when facing the dynamic changes and the inherent complexity of data integration requirements for Healthcare / Life Sciences (HCLS) research. Semantic integration methods assure coherence, harmonize synonyms and different terminologies, and provide an extensible, flexible data integration platform and interactive knowledge base for relevant network analysis. *IO Informatics*

  This inference capability makes both the journalist tagging and the triple store powered SPARQL queries simpler and indeed quicker than a traditional SQL approach. Dynamic aggregations based on inferred statements increase the quality and breadth of content across the site. The RDF triple approach also facilitates agile modeling, whereas traditional relational schema modeling is less flexible and also increases query complexity. *BBC*

- Because there haven't been graph models in the market-leading relational databases, the graph approach has been mostly overlooked by corporate users. But it shouldn't be ignored anymore, he added: Graphs provide "a good representation of the data structure with which you're dealing." *Robin Bloor, the Bloor Group, July 2013*

  We believe that Graph Search is part of a trend that is much bigger than Facebook, and more widespread than search. Facebook is tapping into a fundamentally new way to exploit the information that exists in all the world’s databases. In this post, we will look at the Facebook announcement from a different angle, that of connected data: a growing trend that is on the verge of changing how companies large and small understand their data. *Andreas Kollegger, Jan 2013*

- The Web as we know it today will not be the Web as we know it tomorrow. The Web of today is oriented towards the universal accessibility of files (e.g. web pages, images). The Web of today can be thought of as a large-scale, distributed file system. The Web of tomorrow will encode any datum (e.g. strings, integers, dates). The Web of tomorrow can be thought of as a large-scale, distributed data base. *Marko Rodriguez, Los Alamos, 2009*

- • The construction and structure of graphs or networks is the key to understanding the complex world around us. Albert Laszlo, “Linked: The New Science of Networks”

  Today, most data is being managed in traditional relational databases. These databases have trouble scaling to accommodate the increase in data volume, variety, and complexity. It is also difficult and expensive for these relational databases to provide answers to clinical questions in a timely manner. These relational databases are not a viable solution for solving this new category of data challenges. *KnowMED web site.*
Drivers for a Paradigm Shift

80% of data being generated is unstructured. *

- Hobbled analytics
- Poor collaboration
- Opacity
- TCO
- Performance

*Dennis McCatherty, Baseline magazine
Barriers

- Financial
- Practical
- Technical
- Making the business case
Enablers

- Necessity - W3C Use Cases
- Tools
- Accelerators & Incubators
- Awareness
- Money
John Snelson describes how to query over RDF and documents together, using SPARQL and XQuery, in this sneak-peek into MarkLogic’s future semantics capabilities. *MarkLogic World Conf, 9 Apr 2013*
Community & Regulatory

- DTC – Data Transparency Coalition (Data Transparency 2013, Sep 11-12)
- UDI – Unique Device Identification (FDA)
- ACA – Affordable Care Act (administrative simplification clause)
- FITA – Foundation for Information Technology Accessibility
- LEI – Legal Entity Identifier (mandate)
- FIBO – Financial Industry Business Ontology
- ACT – Advisory Committee on Transparency (Sunlight Foundation)
- Schema.org
- DATA – Digital Accountability and Transparency Act
- CMS – Centers for Medicaid and Medicare Services
- W3C eGovernment
- Executive Order

Specifically, this Memorandum requires agencies to collect or create information in a way that supports downstream information processing and dissemination activities. This includes using machine readable and open formats, data standards, and common core and extensible metadata for all new information creation and collection efforts. OMB Memorandum, M-13-13, May 9, 2013
Pew Study, May 2010

**Expert responses (N=371)**

- Agree: 52%
- Disagree: 38%
- Did not respond: 10%

**Total sample (N=895)**

- Agree: 47%
- Disagree: 41%
- Did not respond: 12%

**Agree:** By 2020, the semantic web envisioned by Tim Berners-Lee will not be as fully effective as its creators hoped and average users will not have noticed much of a difference.

**Disagree:** By 2020, the semantic web envisioned by Tim Berners-Lee and his allies will have been achieved to a significant degree and have clearly made a difference to the average internet users.
LOD vs Graphs

Seal of approval
Unintended consequences (of the good kind)

- Other domains and business use cases
  - Supply chain
  - Philanthropy
  - Social
  - Investor relations
  - Word search
  - Real Estate
  - Arts & humanities
  - E-Commerce ("perfect capitalism")
  - Education

- Other benefits
  - SEO - “The Knowledge Graph does away with the inherent ambiguity of conventional search at Google's search box”
  - Accessible community driven content
  - Unleash the full power of “big data”
  - Unanticipated reuse

¹The Physics of the Future by Michu Kaku
²Google Semantic Search by David Amerland
The Way Forward

- Get the word out
- Build POC's
- Nurture community and industry involvement
- Innovate!
- Facilitate ontology development
- LOD Goes Global
Further Reading

- Use Cases
- Linked Data e-book
- Semantic Community
- Web Science Trust
- STI International
- Schema.org
- semanticweb.com
- World Wide Web Consortium
- Depository Trust and Clearing Corporation
- Freebase
- Jena
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDF</td>
<td>Resource Description Framework, a way of representing information on the web (ie, a set of statements), successor to XML, provides the ontology syntax (framework) for describing resources</td>
</tr>
<tr>
<td>OWL</td>
<td>Web Ontology Language - Standard framework for describing ontologies, based upon XML for the low-level syntax and RDF/RDFS for the mid-level syntax</td>
</tr>
<tr>
<td>RDFS</td>
<td>RDF Schema, recommended structure for RDF resources</td>
</tr>
<tr>
<td>R2RML</td>
<td>Relational to RDF Mapping Language - W3C sanctioned language for expressing customized mappings from relational databases to RDF data sets, D2RQ is the same thing in open source (not as capable)</td>
</tr>
<tr>
<td>Open/Closed World</td>
<td>Open: Any statement not explicitly known to be true or false is considered unknown (OWL). Closed: any statement that is not known to be true is false (SQL).</td>
</tr>
<tr>
<td>Concept</td>
<td>Description logic synonym for “Class”</td>
</tr>
<tr>
<td>Jena</td>
<td>Apache foundation PPMC (Podling Project Management Committee) to develop a semantic app framework (Java API)</td>
</tr>
<tr>
<td>RIF</td>
<td>Rule Interchange Format – W3C recommendation</td>
</tr>
<tr>
<td>SPARQL</td>
<td>SPARQL Protocol and RDF Query Language – query language for RDF data stores, an “endpoint” being a resource that can accept such queries and return results</td>
</tr>
<tr>
<td>SWRL</td>
<td>Semantic Web Rule Language – combines OWL &amp; RML</td>
</tr>
<tr>
<td>FIBO</td>
<td>Financial Industry Business Ontology</td>
</tr>
<tr>
<td>SBVR</td>
<td>Semantics of Business Vocabulary and Rules - OMG developed human readable business modeling language</td>
</tr>
<tr>
<td>OMG</td>
<td>Object Management Group - Nonprofit industry consortium dedicated to data interoperability (ex. UML, MDA, CORBA)</td>
</tr>
<tr>
<td>OASIS</td>
<td>Organization for the Advancement of Structured Information Standards - Standards body for web services, e-commerce, supply-chain and many other domains (including OpenDocument), focus on XML</td>
</tr>
</tbody>
</table>
FacetApp SmartData Platform

Harness Reusable Objects for **Fast**, Scalable Model-driven Enterprise Apps

Configure enterprise applications in a model vs. programming low level source code.

Copyright 2013 FacetApp LLC, Confidential and Proprietary
SmartData Breakthrough

Focus is where it should be – on maximizing ROI
S&P Findings.

"FacetApp provided conservatively, a cost savings of 30% and depending on the situation could have returned 70% based on our experience last month."

- Gene Mishchenko
- Lead Info Architect, S&P Ratings Services
FacetApp API

Interfaces (PhP / AJAX)

Business Logic (Ontology, F-Logic Rules)

SPARQL / Ontostudio

Inference Engine

RDF Triples

Pages (Screens): DataSourceFields

hasInstanceLabel

DataSource > hasModelName: String

hasmodelName

DataModel Element: Concept

OS and Jena APIs were insufficient for closing the gaps. We created our own API using PhP.

PhP (language) + Inference engine are critical tools that enabled us to build the FacetNow API that closes the 2 interface gaps above.
Thank You

Tony Dobaj (doe-bye)
apdobaj@facetapp.com
720-277-2052

FacetApp
Models, not coding