



The Evolving Semantic Web: From Military Technology to Venture Capital

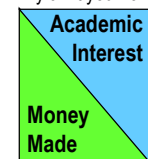
Dr. Mark Greaves
Vulcan Inc.
markg@vulcan.com

Thinking about Revolution

■ How do you measure Impact?

- Academic: Citation counts, dissertations written, journals started, Turing awards...
- Industrial R&D: Conference attendance, prototypes built, early adopters
- Commercial Corporate Success:
 - How many people have had their lives changed by the technology?
 - How much are people willing to pay to use the technology?
 - How much money is being made by companies selling the technology?
 - How many people are living on a beach somewhere because of the profits from the technology?

If you're so smart,
why ain't you rich?



■ Venture Capital is the US engine for creating these companies

- Venture Capital is always seeking new technologies that can drive the next generation of Internet businesses
- Driven by the fusion of an idea and an entrepreneur/team
- Acceptance of risk/failure; high payoffs justify high risks (and high startup mortality rates)

■ You can measure a revolution by the number of people whose lives change

- Don't be afraid of commercialization, because it is one way to measure your impact
- Some group of people are going to get wealthy by changing the world with semantic technology. You might as well try to be one of them.



Talk Outline: The Evolving Semantic Web

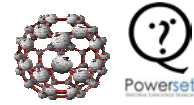
- **US Semantic Web R&D**

- DARPA's DAML Program



- **Semantic Web Evolution to 2008**

- Three Generations of Semantic Dreams
- Markets and Companies



- **The Fourth Generation**

- A Scalable Revolution



3

Talk Outline: The Evolving Semantic Web

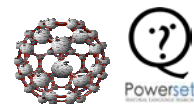
- ➔ ■ **US Semantic Web R&D**

- DARPA's DAML Program



- **Semantic Web Evolution to 2008**

- Three Generations of Semantic Dreams
- Markets and Companies



- **The Fourth Generation**

- A Scalable Revolution



4

Preliminaries on US Computer Science Funding

■ Major US Computer Science Funding Organizations

- US National Science Foundation – mainly individual professors
- Department of Energy Office of Science – large-scale scientific computing
- Department of Defense – DARPA, AFRL, ARL, ONR

■ DARPA = Defense Advanced Research Projects Agency



- Long Range R&D Organization of the US Department of Defense
 - Established 1958 as a US response to the Soviet launch of Sputnik
 - Pursues high-risk, high-payoff basic and applied research with military application
- Chartered to Prevent Technological Surprise
 - Funds work that is a counterpoint to traditional thinking and approaches
 - Noteworthy programs include VELA HOTEL, M-16, HAVE BLUE (F-116), TACIT BLUE (B-2), Sea Shadow, GPS, ARPANET, TEAL RAIN (Global Hawk), AMBER (Predator)
- Recent budgets are \$3-3.5B/year

■ DARPA is the major source for large-scale AI research funding in the US

■ Lightweight organizational model

- “120 Program Managers with a common travel agent”
- No dedicated facilities beyond simple office space
- Program Managers have a greater degree of operational control than EU Program Officers
- <http://www.darpa.mil> has current programs, solicitations, lists, areas of interest

5



At the End of the 90s: Traditional KR and the Google Property

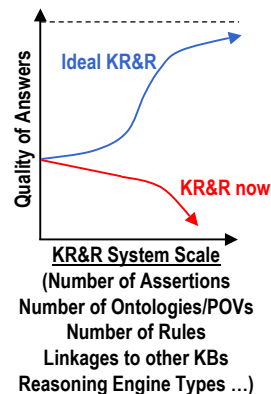
■ We seek KR systems that have the “Google Property:” *they get (much) better as they get bigger*

- Google PageRank™ yields better relevance judgments as it indexes more pages
- Current KR&R systems have the antithesis of this property

■ So what are the components of a scalable KR&R system?

- Distributed, robust, reliable infrastructure
- Multiple linked ontologies and points of view
 - Single ontologies are feasible only at the program/agency level
 - Multiple authors and overlapping data sources
 - Private and public knowledge
- Mixture of deep and shallow knowledge
- Tractable reasoning algorithms
- Tolerant KB – you are typically doing open-world reasoning (no NAF), things go away, contradiction is present, data is incomplete and dirty, computing must be resource-aware, surveying the KB is not possible
- (Relatively) easy for non-KE’s to author, validate, and maintain

KR&R Goals



Scalable KR&R Systems should look just like the Web!!

6



The Beginnings of the US Semantic Web: DARPA's DAML Program

Problem:

Computers cannot process most of the information stored on web pages

Solution:

Augment the web to link machine-readable knowledge to web pages

- Extend RDF with Description Logic
- Use a frame-based language design
- Create the first fully distributed web-scale knowledge base out of networks of hyperlinked facts and data

Approach:

Design a family of new web languages

- Basic knowledge representation (OWL)
- Reasoning (SWRL, OWL/P, OWL/T)
- Process representation (OWLS)

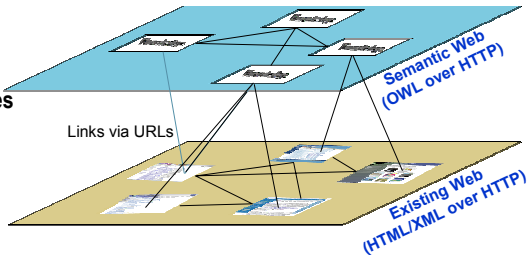
Build definition and markup tools

Link new knowledge to existing web page elements

Test design approach with operational pilots in the US DoD

Partner with parallel EU efforts to standardize the new web languages

Computers require explicit knowledge to reason with web pages



People use implicit knowledge to reason with web pages

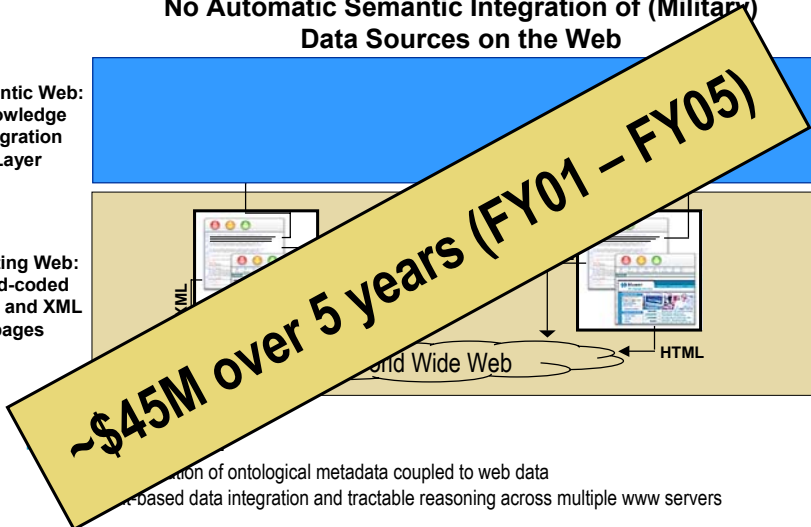


DAML Operational Problem

No Automatic Semantic Integration of (Military) Data Sources on the Web

Semantic Web: Knowledge Integration Layer

Existing Web: Hand-coded HTML and XML pages



Integration of ontological metadata coupled to web data
 Data-based data integration and tractable reasoning across multiple www servers

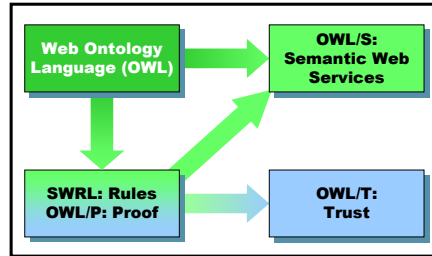
- Early semantic web pilots with various members of the US DoD
- Coordination and data sharing groups within the US DoD



DAML Program Elements

- **Web Ontology Language (OWL) (2/10/04)**
 - – Enables knowledge representation and tractable inference in a web standard format
 - – Based on Description Logics and RDF
- **OWL Reasoning Languages**
 - – SWRL and SWRL-FOL: Supports business rules, policies, and linking between distinct OWL ontologies
 - OWL/P Proof Language: Allows software components to exchange chains of reasoning
 - OWL/T Trust Language: Represents trust that OWL and SWRL inferences are valid
- **Semantic Web Services (OWL/S)**
 - – Allows discovery, matching, and execution of web services based on action descriptions
 - – Unifies semantic data models (OWL) with process models (Agent) and shows how to dynamically compose web services
- **OWL Tools**
 - – www.semwebcentral.org and www.daml.org
- **Several US DoD pilots and prototypes**

DAML Program Technical Flow



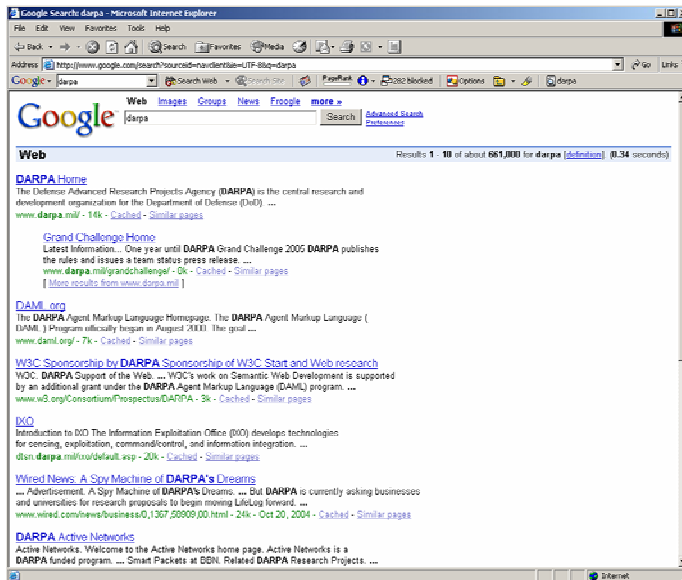
- Completed standards process
- Started standards process
- Unfinished

Each DAML Program Element includes specifications, software tools, coordination teams, and use cases



9

Another Measure of Impact



- #2 →
- #3 →

Google "darpa" on 10/21/04



10

The Semantic Web in 2008

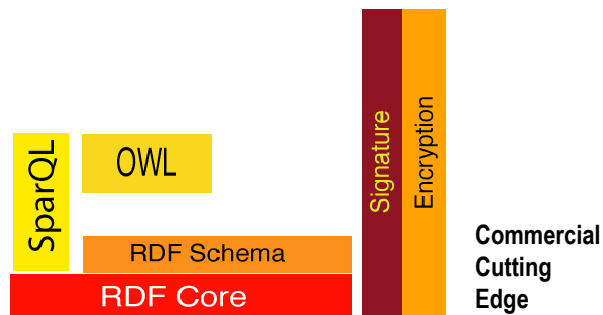


“The Famous Semantic Web Technology Stack”

11



The Semantic Web in 2008

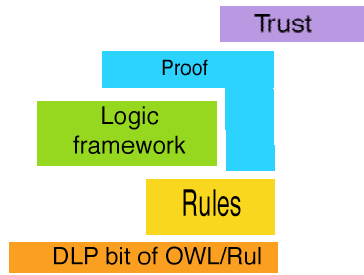


“The Famous Semantic Web Technology Stack”

12



The Semantic Web in 2008



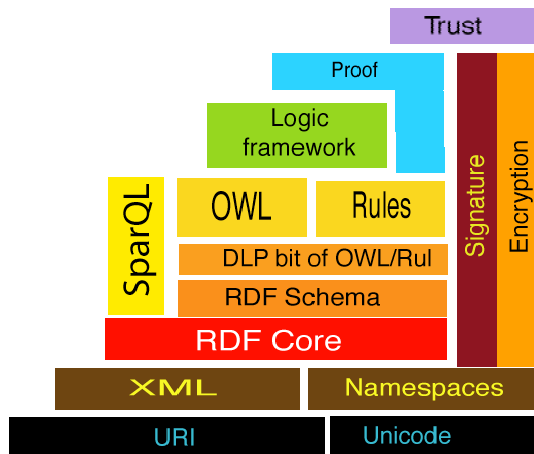
Active Research
and Standards
Activity

“The Famous Semantic Web Technology Stack”

13



The Semantic Web in 2008



Active Research
and Standards
Activity

Commercial
Cutting
Edge

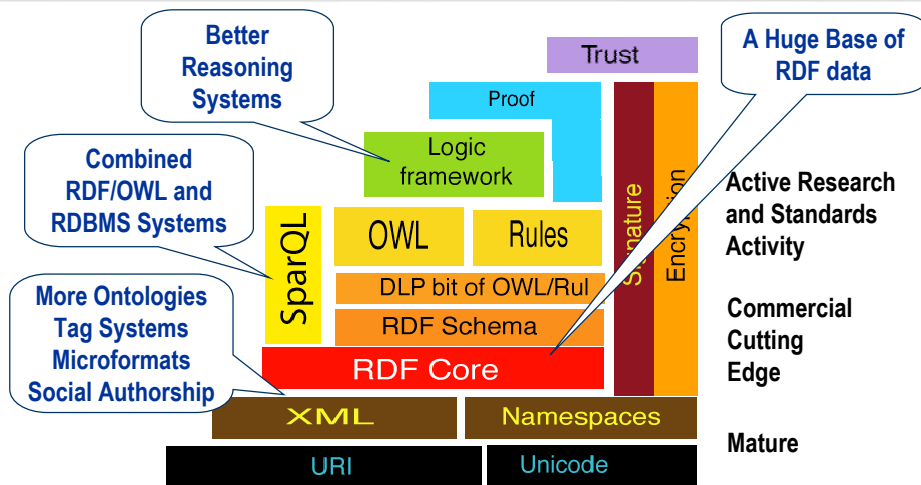
Mature

“The Famous Semantic Web Technology Stack”

14



Completing the Semantic Web Picture



Other Technologies Impact the Semantic Web

15



State of Semantic Web Work in the US

- **DAML finished in 2005, with no followons**
 - NIH (Protégé, NCBO), NSF, some small DoD funding
 - PAL/CALO had a small semantic piece which is ending
- **But... leading-edge Venture Capital moved in**
 - Vulcan, Crosslink, In-Q-Tel, Benchmark, Intel Capital...
- **An emerging commercialization ecosystem**
 - Startup: Radar, Metaweb, Evri, AdaptiveBlue, RealTravel...
 - Midsized: Monitor, Thetus, Metatomix, Franz, Saltlux, Ontotext...
 - Large: Yahoo!, Oracle, IBM, HP, Microsoft, Thomson/Reuters...
- **Emphasis is mostly *Database* dimension of Semweb**
 - RDBMS scale and orientation, powerful analytics for Business Intelligence
 - Centralized workflows for ontology definition and management
 - Use cases surrounding data integration
 - Emerging microformats and structured blogging (e.g., Twine)
 - ... But mainly enterprise data description concerns



16



State of Semantic Web Work in the EU

■ Continuing Large Public-Sector Investments

- Framework 6 (2002-6) – More than €100M in several different programs
- Framework 7 (2007-13) – ~€1B/year for information and communications technologies
 - €100M in 2007/8 and €70M in 2009/10 for Digital Libraries
 - Semantics is also heavily present as a general systems technology



■ Two Dedicated Multi-site R&D Institutes

- Semantic Technology Institute International
- DERI: 100+ people and the world leader in research
- A strong and growing cadre of graduate students



■ Emphasis on the *Social* and *Web* Dimensions of Semweb

- Web-scale, social networks, simple scalable imperfect inference
- Ontology and data dynamics, imperfections, versioning
- Semantically-boosted collaboration with limited knowledge engineer involvement
- A base of socially-curated semantic data

Clear technical leadership but a lack of commercialization



17

Talk Outline: The Evolving Semantic Web

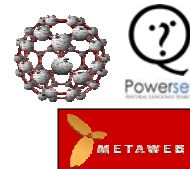
■ US Semantic Web R&D

- DARPA's DAML Program



■ Semantic Web Evolution to 2008

- Three Generations of Semantic Dreams
- Markets and Companies



■ The Fourth Generation

- A Scalable Revolution



18

Evolving Conceptions for the Semantic Web

Initial Semantic Web Conception*

- **Semantic markup would be tightly associated with individual web pages**
 - “Translate the Web for machines”
 - RDFa shows this is still a powerful vision
- **Core problem is labeling free-text web pages with a (pre-defined) ontology markup vocabulary**
 - Entity extraction and other lightweight NLP
 - Document segmentation technologies
 - Manual annotation
- **Need an all-encompassing ontology or set of logically compatible ontologies**
- **Small number of knowledge engineers do semantic annotation because the modeling problems are so hard**
 - Knowledge engineers rarely get markup right because they aren't domain experts

* By most people but not TBL

The Semantic Web in 2008

- **The Web is a publishing platform for formal knowledge as well as pages**
 - Semantic data doesn't have to be associated with HTML web text (just a URI)
 - Huge numbers of knowledge publishers
 - Simple RDF and owl:sameAs links
- **Core problem is maintaining a set of evolving and partial agreements on semantic models and labels**
 - Consensus is a human social problem
 - There will be massive numbers of overlapping ontologies and class hierarchies
 - Hard problem is cost-effectively maintaining semantic models and labeling data
- **Supplemental semantics is carried in the free-text web**

19



First Generation Semantic Web Applications

Semantically-Boosted Search and Classification

- **A really old problem type**
 - Semantics as the keystone technology for unstructured Information Retrieval
 - Requires powerful NLP and document interpretation systems
 - Often also requires powerful semantic representations (e.g., events or causality)
 - Can use semantic web KR but usually augments it
- **Market Segments and Players**
 - Enterprise Document Management (EDM) and search systems
 - Documentum, Autonomy, Convera, FAST (bought by Microsoft for \$1.2B)...
 - Email autoclassifiers and inbox managers
 - Web question answering
 - Hakia, Powerset, Answers.com, TextDigger, TrueKnowledge...
 - Cycorp
- **What are the issues with a VC bet in this space?**
 - Still waiting for a compelling match between technical capability and business need
 - Statistical methods are surprisingly good (e.g., Latent Semantic Indexing)
 - Verticals (esp. health care) have seen some success
 - Semantic processing is only a small differentiator in these markets – you have to be great at nonsemantic queries, data import, crawling, storage, performance...

20



First Generation Example – Powerset

- **Natural language consumer search**
 - Web crawling, keyword indexing, relevance ranking
 - High performance for web-scale commercialization
 - Parsing of web page text with Xerox PARC's XLE system
 - Question answering with Wikipedia text
 - Questions like "What did Microsoft acquire in 2006?" or "What did Steve Jobs say about the iPod?"
 - Performance is in the eye of the beholder
 - Approaches Google in keyword search relevance performance
 - Plans to expand to multilingual and larger fragments of the web
 - Barney Pell (CEO) keynoted at ISWC 2007
- **Powerset's semantic knowledge is a superset of semantic web KR**
 - But Powerset used semantic web resources as a basic KB
- **Bought by Microsoft in June for (rumored, approximate) USD\$100M**
 - Why wasn't the valuation higher?
 - Fits with Microsoft strategy of search verticals
 - Reaction to Yahoo! Acquisition defeat



21



Second Generation Semantic Web Applications

Strategic Enterprise Information Technology

- **An only slightly newer problem type**
 - Exploitation of mainly structured enterprise data (RDBMS, Spreadsheets, files)
 - SOA integration, Enterprise Information Integration, Enterprise Application Integration
 - Backwards to Data Management to reduce cost of managing, migrating, integrating
 - Forwards to Business Process Management
 - Support for unified query, analytics, and application access
- **Markets Segments and Players**
 - Gardner estimates that EII software and services alone is \$14B/year, with 40% growth over 5 years
 - Very complex market space includes EAI, Entity Analytics, MDM, BI, BPM, CPM...
 - Huge entrenched players (IBM, SAP, Oracle...) and consulting shops (Accenture, Logica...)
 - Well-understood ROI and distribution models, well-heeled customers
- **What are the issues with a VC bet in this space?**
 - Fundamental problem is (as always) elucidating semantics from legacy systems, not in KR
 - Pure Semantic technology companies tend to be unsophisticated about large data
 - Tends to be IT sale (not Line-of-Business sale)

22



Semantic Submarkets in Strategic IT

	SDM	SBM	EII	BI	CPM	SOA	MDM	BPM
SDM	=	overlaps	is part of	is part of	is part of	is part of	is part of	is part of
SBM	overlaps	=		is part of	is part of			helps
EII	requires		=					
BI	requires	requires	requires	=				
CPM	requires	requires	requires	requires	=			
SOA	requires					=	is part of	is part of
MDM	requires		requires	helps	helps	requires	=	
BPM	requires					requires		=

From Automata Inc.

Observations

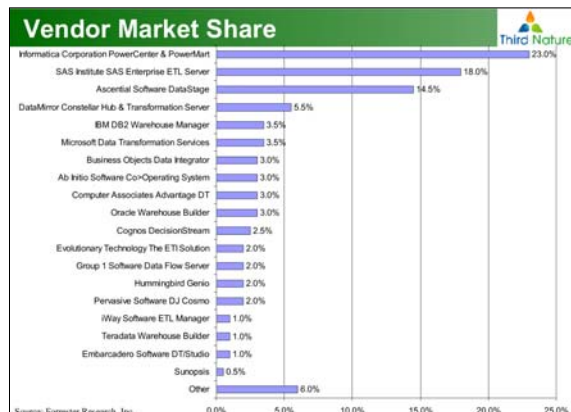
- Wholesale reinvention is an impossible sale
- Semantic Data Modeling is the critical core where we can get traction, but Semweb technology itself isn't a product
- EII costs are 90% modeling
- Semantic BI & CPM need Semantic Business Models
- BI requires and MDM is just beyond EII

Acronyms

SDM = Semantic Data Modeling
 SBM = Semantic Business Modeling
 EII = Enterprise Information Integration
 BI = Business Intelligence
 CPM = Corporate Performance Mgmt
 SOA = Service Oriented Architecture
 MDM = Master Data Mgmt
 BPM = Business Process Mgmt

23

~\$2B Market for EII Software



- Services are ~5x software; ~65% is "custom" work
- Good news is that the market is large
- Bad news is that it is not friendly to new technology startups

24



Third Generation Semantic Web Applications

Web 2.0 and the Socio-Semantic Web

- **A new problem type**
 - “Semantic Web should allow people to have a better online experience” – Alex Iskold, CEO of AdaptiveBlue
 - Enhance the human activities of content creation, publishing, linking my data to other data, socializing, forming community, purchasing satisfying things, browsing, etc.
 - Improve the effectiveness of advertising
- **Market Segments and Players**
 - Mashup systems and consumer-oriented semantic web services (Drupal, Ning, ...)
 - Semantic enhancements to blogs and wikis (Zemanta, Salzburgresearch, Ontoprise, Radar, ...)
 - Semantics in Social Networking (MySpace RDF service and microformats, Facebook, etc.)
- **What are the issues with a VC bet in this space?**
 - If we don't have semantic convergence, then semantics isn't a differentiator
 - No one really knows the design principles that cause some Web 2.0 sites to be successful and others to never get traction



25

Web 2.0 and the Semantic Web: Sibling Rivalry

- **Similar Birthdays**
 - “Web 2.0” coined in a conference brainstorming session between Tim O'Reilly and MediaLive in early 2004
 - Semantic Web RDF/OWL recommendations passed by W3C in Feb 2004
- **Similar Technical Inspiration**
 - Web 2.0 depends on an implicit lightweight semantic agreement between its elements (primarily *users and developers*)
 - Semantic Web depends on an explicit shared semantic model between its elements (primarily *computers*)
- **Similar Goals**
 - Web 2.0 is the next evolution of the web from a *user's* point of view
 - Semantic Web is the next evolution of the web from a *computer's* point of view

How Have These Two Siblings Grown Up?



26

Web 2.0 and the Semantic Web

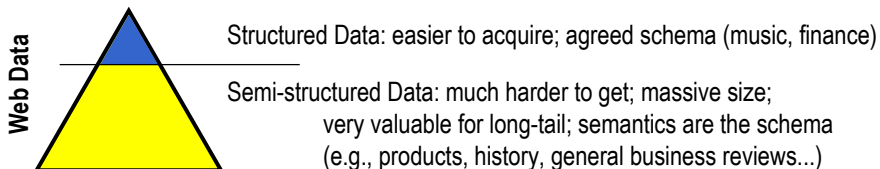
- **The first major revolution in IT that was not driven by business**
 - The basic human need to publish, to interact, to socialize
 - A simple, quick, dirty, user-oriented version of an SOA
 - Driven by free software (sometimes hosted) and open, user-contributed data
- **Semantic Web in the context of the Web 2.0 revolution**
 - *Web 2.0 and Semantics*
 - How do semantic web technologies match up with the semantic demands of Web 2.0 applications?
 - *Semantics and Web Ecosystems*
 - Web 2.0 applications often strive to build participatory ecosystems of content that is supplied and curated by their users. Can Web 2.0 users effectively create, maintain, map between, and use RDF/OWL content in a way that reinforces the ecosystem?
 - *Semantic Web in Practice*
 - Does semantic web technology enable the cost-effective creation of Web 2.0 applications that are simple, scalable, and compelling for a targeted user community?
- **“Semantic Web 2.0” in *IEEE Intelligent Systems* 22:2 (Mar 07)**

27



Semantic Web and Web 2.0 Ecosystems

- **Ecosystem benefits depend on a large amount of data**



- **Semantic authoring is more complex than Web 2.0 authoring**
 - Time-to-market issues for startups that use semantics
 - Need zero-training, easy-to-implement, tolerant of mistakes
 - Sufficient user agreement/reuse on ontology fragments, terms, and maintenance
 - Lack of user incentive to provide semantic markup
 - Attribution, visibility from others, use by others
 - Annotation tools often require investments of time and judgment



Several exciting products are in this space

28



Third Generation Example: Geocommons mashup

■ How do Web 2.0 technologies currently share meaning?

- Tags, distinct social networks, group wikis
- Explicit developer agreement on REST and SOAP and XQuery parameters
- Mashup data fusion algorithms

■ Example: www.geocommons.com

- "The Hippest Places to Live in San Francisco"
- Data sets used
 - San Francisco municipal neighborhood boundaries
 - Crime index by census tract
 - Home median age and density
 - Scraped ratings and locations of San Francisco bars and clubs
 - Occupation by census tract (techies and artists)
 - Commute mode (Female motorcycle ridership)



29

Third Generation Example: Semantic Wikis

■ Wikis are tools for *Publication* and *Consensus*

■ MediaWiki (software for Wikipedia, Wikimedia, Wikibooks, etc.)

- Most successful Wiki software
 - High performance: 10K pages/sec served, scalability demonstrated
 - LAMP web server architecture, GPL license
- Publication: simple distributed authoring model
 - Wikipedia: >2M articles, >180M edits, 750K media files, #7 most popular web site in April
- Consensus achieved by global editing and rollback
 - Fixpoint hypothesis (2:1 discussion/content ratio), consensus is not static
 - Gardener/admin role for contentious cases



■ Semantic Wikis apply the wiki idea to basic (typically RDFS) structured information

- Authoring includes instances, data types, vocabularies, classes
- Natural language text for explanations
- Automatic list generation from structured data, basic analytics
- Reuse of wiki knowledge

Semantic Wiki Hypotheses:

- (1) Significant interesting non-RDBMS Semantic Data can be collected cheaply
- (2) Wiki mechanisms can be used to maintain consensus on vocabularies and classes

30

Third Generation Example: Metaweb and Freebase



freebase[™]
alpha

Keyword search Freebase Search

Home My Profile Types Developers Help Welcome back, mgreaves. Not you? Sign out

San Francisco, California

Discuss "San Francisco, California" Show Empty Fields



Types: US City/Town (Location), Film Location (Film), City/Town (Location), US County (Location), Location (Location)

Also known as: San Francisco, CA, SF, CA, Frisco, San Francisco

Population: 739,426 - 2005

State: California

Geolocation: -122.4183 - 37.775

Page History
Created by Metaweb Oct 22, 2006 11:53am
Last edited by danm 6 days ago at 7:33pm

State
California

Featured in Films
Vertigo
Scorpio
Bullitt

Contains
Mission District, San Francisco, California

Contained By
SF Bay Area
California

Description
The City and County of San Francisco is the fourth most populous city in California and the fourteenth-most populous in the United States, with a 2005 population of 739,426 [2] It is located on the tip of the San Francisco Peninsula and is the focal point of the San Francisco Bay Area. San Francisco is the second most densely populated major city[3] in the United States [4] In 1776, the Spanish settled the tip of the San Francisco peninsula, establishing a fort at the Golden Gate and a mission named for Francis of Assisi. The California Gold

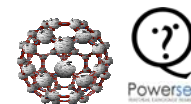
- Massive amounts of almanac-style RDF data (Creative Commons license) that is commonly available
- Social authoring tools and wiki-style consensus
- Public data outsourcing model for long-tail startups

31



Talk Outline: The Evolving Semantic Web

- **US Semantic Web R&D**
 - DARPA's DAML Program
- **Semantic Web Evolution to 2008**
 - Three Generations of Semantic Dreams
 - Markets and Companies



- ➔
- **The Fourth Generation**
 - A Scalable Revolution



32



Fourth Generation Semantic Web

The Web of Data meets the Future Internet

■ A problem of scale

- The number of Internet devices is starting to explode (again!)
 - Mobile devices, embedded systems, and sensors
 - Many of these involve natural Semantic Web applications
- Gartner (May 2007, Report G00148725):
 - "By 2012, 70% of public Web pages will have some level of semantic markup, 20% will use more extensive Semantic Web-based ontologies"
- Can Semantic Web technologies work at web scales? (see Frank van Harmelan's work)
 - Sindice (www.sindice.com) is now indexing >10B triples/microformats
 - 20% of 30 billion pages @ 1000 triples per page = **6 trillion triples**
 - 30 billion and 1000 are underestimates

■ What are the issues with a VC bet in this space?

- Does the Semantic Web have the Google Property?
- Can we exploit billions of triples, microformats, ontologies, rules, and services
 - Are Semantic Web systems deployable on parallel web architectures, friendly to out-of-core algorithms, and compatible with giant databases?
- Is there a scaling limit to useful, profitable Semantic Web implementations?

33



Fourth Generation Example: DBpedia

■ Mine Wikipedia for assertions

- Scrape Wikipedia Factboxes
 - ~23M triples
- High-confidence shallow English parsing
- Category assertions

■ DBpedia 3.0 dataset

- ~2M things, ~220M triples
 - 80K persons, 293K places, 62K music albums, 36K films, 489K links to images, 2.7M links to relevant external web pages, 2.1M links into RDF datasets
- Classifications via Wikipedia categories and WordNet synsets
- One of the largest broad knowledge bases in the world

■ Simple queries over extracted data

- Public SPARQL endpoint
- "Sitcoms set in NYC"
- "Soccer players from team with stadium with >40000 seats, who were born in a country with more than 10M inhabitants"

```
1 ((Infobox Town AT |
2   name = Innsbruck |
3   image_csa = InnsbruckWappen.png |
4   image_ssp = Karte-tirol-1.png |
5   state = [[Tyrol]]) |
6   regtok = [[Statutory city]] |
7   population = 117,342 |
8   population_as_of = 2006 |
9   pop_dens = 1,119 |
10  area = 104.91 |
11  elevation = 574 |
12  lat_deg = 47 |
13  lat_min = 56 |
14  lat_max = 2 |
15  lon_deg = 11 |
16  lon_min = 23 |
17  lon_max = 2 |
18  postal_code = 6010-6080 |
19  area_code = 0512 |
20  licence = 1 |
21  mayor = Hilde Zach |
22  website = http://innsbruck.at |
23  )}
```

Innsbruck	
Country	Austria
State	Tyrol
Administrative region	Statutory city
Population	117,342 (2006)
Area	104.91 km²
Population density	1,119/km²
Elevation	574 m
Coordinates	47°18′N 11°23′E﻿ / ﻿47.300°N 11.383°E﻿ / 47.300; 11.383
Postal code	6010-6080
Area code	0512
License plate code	I
Mayor	Hilde Zach
Website	www.innsbruck.at



UNIVERSITÄT LEIPZIG



34



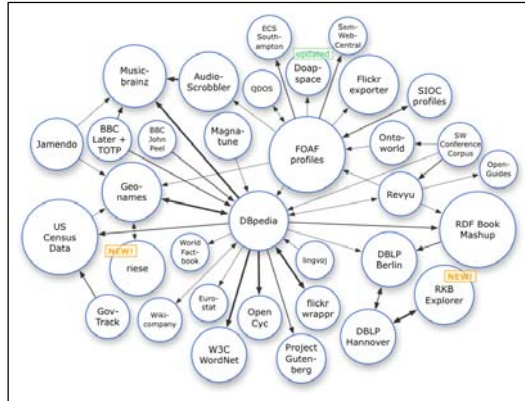
Fourth Generation Example: Linking Open Data

Goals

- Create a single, simple access mechanism for web RDF data
- Build a data commons by making open data sources available on the Web as RDF
- Set RDF links between data items from different data sources

Total LOD dataset

- ~2B triples, and ~3B RDF links
- Growing all the time (ex: 3B Eurostat triples)
- Database linkage means that LOD will soon be impossible to count except via order of magnitude



35



Semantic Dynamism at Web Scale

Semantics are always changing

- Per minute, there are:
 - 100 edits in Wikipedia (144K/day)
 - 200 tags in del.icio.us (288K/day)
 - 270 image uploads to flickr (388K/day)
 - 1100 blog entries (1.6M/day)
- Will the Semantic Web be less dynamic?



There is no “right ontology”

- Ontologies are abstractions
 - Different applications lead to different ontologies
 - Ontology authors make design choices all the time
- Google Base: >100K schemas
- “Ontologies = Politics”



Intentionally false material (Spam)

- Lesson of the HTML <META> tag

36

Material from Denny Vrandečić, AIFB



Fourth Generation Application: The Large Knowledge Collider?



- **EC Framework 7 Program**
 - Lead partners: Univ. Innsbruck and Vrije University Amsterdam, plus 12 partners
- **Goals of LarKC – Scaling to Infinity**
 - Give up soundness & completeness
 - Combine reasoning/retrieval and search
 - Want to trade off answer quality and answer timeliness
 - Heavy emphasis on probability, decision theory, anytime algorithms
- **Reasoning pipeline**
 - Plugin architecture, with sampling
 - Explicit cost models
- **Public releases of LarKC platform, with APIs**
- **Encourage participation through Thinking@home**
 - Kind of like SETI@Home

37



Fourth Generation Application: The Large Knowledge Collider?



- **EC Framework 7 Program**
 - Lead partners: Univ. Innsbruck and Vrije
- **Reasoning pipeline**
 - Plugin architecture, with sampling
- **The real money in semantics will be made in apps/tools that exploit web-scale data**
 - The cost of semantic data creation is going to zero
 - The size of semantic data is going to web-scale
- ***If LarKC is successful, this could be as big as PageRank!***
 - Heavy emphasis on probability, decision theory, anytime algorithms
 - Kind of like SETI@Home

38



Summing Up: Out of the Lab and Into the World

- The Semantic Web is a transformational idea
- Massive opportunities are out there, especially in 3rd and 4th-generation semantic web applications
- When you get a great idea, change the world with it

**Thank You, Good Luck, and
Remember My Email Address**