From Text and Data to Knowledge: The Social Semantic Web in Action

Dr. Mark Greaves  
markg@vulcan.com

Jesse Wang  
jessew@vulcan.com
What is Vulcan?

- **Vulcan is the asset management firm for Paul G. Allen**
  - Vulcan ([www.vulcan.com](http://www.vulcan.com)) creates and advances a variety of world-class endeavors and high impact initiatives that change and improve the way we live, learn, do business, and experience the world.

- **Major Operating Groups**
  - Vulcan Capital
  - Vulcan Sports and Entertainment
    - Seattle Seahawks, Portland Trail Blazers
  - Vulcan Productions
  - Vulcan Real Estate
  - Paul G. Allen Family Foundation
  - Allen Institute for Brain Sciences
  - Vulcan Technology (Stratolaunch, Halo, Marine)
Paul Allen and Bill Gates in 1968
The Founding of Microsoft

[Image of two individuals standing in front of a whiteboard with various diagrams and text on it, including terms like IBM, Apple, TRW, MUSE, Plan, grid, basic graph, and data types.]
Paul Allen’s Concept of the Digital Aristotle

“Digital Aristotle”
- Inspired by the broad science fiction vision of Big AI
- The volume of scientific knowledge has outpaced our ability to manage it
- Need systems to reason and answer questions, rather than simply retrieve 1000s of relevant documents
  - Example: “What are the reaction products if metallic copper is heated strongly with concentrated hydrochloric acid?”

Project Formation
- Initial Workshops conducted in 2001
- Digital Aristotle vision formulated
- Project Halo created in 2002
The Digital Aristotle and Project Halo

**Project Halo** is a staged, long-range research effort by Vulcan Inc. towards the development of a Digital Aristotle.

**Project Halo** envisions two primary classes of application:

1. A digital textbook capable of answering student questions and helping students with their work
2. A research assistant with broad, interdisciplinary skills to help scientists make connections

A Digital Aristotle is a reasoning system capable of answering novel questions and solving advanced problems in a broad range of scientific disciplines.
Where to Begin?

- **AI Grand Challenge**
  - Read a chapter and answer questions in the back of the book (Reddy, 2003)

- **Focus on Science**
  - Focus on science where knowledge is explicitly stated

- **US Advanced Placement Exam**
  - Use this US-based test of human competence as a metric

- **Systems AI**
  - Include the knowledge, knowledge acquisition, reasoning, and question answering pieces
# Project Halo Development Strategy

## Authors
- **Pilot**: KR Expert
- **Phase II**: Single Domain Expert
- **Inquire**: Small Team of Domain and KR Experts
- **DA**: Community of Scientists, Teachers, and KR Experts

## Uses
- **Pilot**: Logic Queries
- **Phase II**: AP Question Answering
- **Inquire**: AP QA General QA Education
- **DA**: AP QA General QA Education Research

## Timeline
- **Pilot**: 2002-2003
- **Phase II**: 2004-2009
- **Inquire**: 2010-2015
- **DA**: 2016-????
What is **Inquire**?

- **Inquire** is a concept for a new kind of electronic textbook
  - Based on Campbell and Reese, *Biology*

- **Inquire** is a standard electronic textbook, but also:
  - Contains a full underlying knowledge base of the book’s contents
  - Answers the reader’s questions and provides tailored instruction
  - Core technology applies to other areas of biology, and other exact sciences (chemistry, physics, geology, etc.)
    - Is less applicable in history, art, languages, philosophy, etc.
  - Should support other languages
Inquire’s Design Goals

An Inquire user can:

- **READ** the textbook contents dynamically, interactively
- **ASK** questions and get explanations on any subject in the book
  - **LEARN** and master the subject through individualized tutoring
  - **CREATE** and explore their own conceptualizations of the material

A Calculator for Biology

- Simple questions are answered from specific knowledge in the text
- Fully embedded in the textbook
Domain experts (biologists) enter knowledge from Campbell’s Biology textbook into a structured, logical, knowledge base. The knowledge base is connected to an electronic version of the textbook, running on an iPad, that enables students to ask questions about the material as they read. Inquire answers a wide range of questions by performing logical inference on the knowledge entered into the knowledge base. Measure retention and understanding for students using Inquire against students using either a paper or electronic version of the textbook.
Inquire Demo

- Digital Textbook iPad Application
  - eBook reader (with standard features)

- Knowledge-based features
  - Glossary pages (from KB)
  - Suggested questions and answers
  - Free-form QA dialog

- AURA Knowledge Server
  - Concept taxonomy from Campbell Biology
  - AURA Knowledge Base (21 chapters)
  - Suggested question generation
  - Reasoning and question answering system
  - AAAI Best Video of 2012

http://www.youtube.com/watch?v=fTiW31MBtfA.
Inquire vs. Siri vs. Watson vs. Search Engines

- **Project Halo’s Inquire**
  - Subject Matter: biology
  - Questions: Question suggestion and simple English queries
  - Answers: Formatted data and textbook content, exact answers

- **Apple’s Siri**
  - Subject Matter: Actions on iPhone, plus Wolfram Alpha searches
  - Questions: Voice commands, short dialogs
  - Answers: Performs tasks, Alpha-formatted pages

- **IBM’s Watson**
  - Subject Matter: General knowledge, trivia, attempting medical
  - Questions: Jepoardy-style “reversed” questions
  - Answers: Single word or short phrase answers

- **Google / Yahoo / Bing / Baidu**
  - Subject Matter: Open domain web knowledge
  - Questions: keywords and search queries
  - Answers: web page listings
<table>
<thead>
<tr>
<th>Knowledge Entry</th>
<th>Question Processing</th>
<th>Question Answering</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task:</strong> Encoding knowledge sentence-by-sentence</td>
<td><strong>Task:</strong> Map a student question to an answerable query</td>
<td><strong>Task:</strong> Generate a useful answer to a student question</td>
</tr>
<tr>
<td>- Define a computer language for capturing knowledge</td>
<td>- English language interpretation</td>
<td>- Derive an answer from the knowledge base</td>
</tr>
<tr>
<td>- Represent background knowledge and explicit knowledge</td>
<td>- Suggest good questions to student</td>
<td>- English generation and media selection</td>
</tr>
<tr>
<td><strong>Systems:</strong> AURA, Semantic Media Wiki (SMW)</td>
<td><strong>Systems:</strong> AURA Question Processor</td>
<td>- Answer presentation and relevance pruning</td>
</tr>
<tr>
<td><strong>Metrics</strong></td>
<td><strong>Metrics</strong></td>
<td><strong>Systems:</strong> AURA, Cyc, Text Analysis</td>
</tr>
<tr>
<td>- Coverage of book at various depths</td>
<td>- Usefulness of suggested questions</td>
<td><strong>Metrics</strong></td>
</tr>
<tr>
<td>- Time per sentence</td>
<td>- Accuracy of free-form question processing</td>
<td>- Answer accuracy</td>
</tr>
<tr>
<td><strong>R&amp;D thrusts</strong></td>
<td><strong>R&amp;D thrusts</strong></td>
<td>- Answer usefulness to student users</td>
</tr>
<tr>
<td>- SMW: Lower authoring costs by crowdsourcing</td>
<td>- Rephrasing</td>
<td>- Time to answer</td>
</tr>
<tr>
<td>- SILK: Capturing more knowledge</td>
<td>- Textual Entailment</td>
<td><strong>R&amp;D thrusts</strong></td>
</tr>
<tr>
<td></td>
<td>- Learning suggested question mappings</td>
<td>- Hybrid system: Answer more questions</td>
</tr>
<tr>
<td></td>
<td>- Question generation from the knowledge base</td>
<td></td>
</tr>
</tbody>
</table>
Knowledge from Campbell Biology in Inquire

- Most-used biology textbook in the US, both high school and college
- Total Core Sentences: 28,878
- 56 Chapters, 262 Sections
  - Avg 4.7 sections/chapter; 515 sentences/chapter; 109 sentences/section
  - Relevant sentences in book (extrapolation from Chaps 6,9,10) 49%; 14,150 sentences

- 750 Core library concepts
- ~6,000 total concepts in taxonomy
  - ~3,000 concepts in Chaps 2-22
- ~65,000 knowledge base assertions
- ~50,000 suggested questions
- 4 hours/sentence authoring time
  - Design, planning, encoding, testing, refining, new relation development
- 60-70 new encoding relations/year
Inquire’s Existing Knowledge Entry Process

Knowledge Entry Group
Delhi, India

- Multiple 3-person teams
- Highly accurate but costly
  - ~4 person-hours per sentence
  - Intensive testing and knowledge refinement

Biology Teams and AI experts represent each sentence

Concept Map Planning → Concept Map Entry → Concept Map Testing

Assess & Correct → KR Review

Senior KR Expert → Senior SME

Content Review
## Knowledge Entry Process

<table>
<thead>
<tr>
<th>Step</th>
<th>Time</th>
<th>Activity</th>
<th>Status Labeling</th>
<th>QA Check</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Determining Relevance</td>
<td>2%</td>
<td>Highlighting, Diagram Analysis</td>
<td>Relevant, Irrelevant (Closed)</td>
<td>QA Check</td>
</tr>
<tr>
<td>2) Reaching Consensus</td>
<td>14%</td>
<td>Universal Truth Authoring, Concept Chosen</td>
<td></td>
<td>QA Check</td>
</tr>
<tr>
<td>3) Encoding Planning</td>
<td>35%</td>
<td>Group Common UTs, ID KR/KE Issues, ID Already Encoded, Write How to Encode</td>
<td>Encoding Complete, KR Issue (Closed)</td>
<td>Pre-Planning, QA Check</td>
</tr>
<tr>
<td>4) Encoding</td>
<td>10%</td>
<td>Encode, File JIRA Issues</td>
<td>Encoding Complete, KE Issue</td>
<td>QA Check</td>
</tr>
<tr>
<td>5) Key Term Review</td>
<td>25%</td>
<td>KR Evaluated by Modeling Expert and Biologist, Encoder Makes Changes</td>
<td></td>
<td>KR Evaluated by Modeling Expert and Biologist</td>
</tr>
<tr>
<td>6) Question-Based Testing</td>
<td>14%</td>
<td>Use Minimal Test Suite, Reasoning JIRA Issues Filed, Encoder Fills KB Gaps</td>
<td></td>
<td>QA Check with Screenshots of “Passing” Comparison and Relationship Questions</td>
</tr>
</tbody>
</table>
# Knowledge Entry Process

## Planning (51%)

<table>
<thead>
<tr>
<th>Step</th>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Determining Relevance</td>
<td>2%</td>
<td>Highlighting, Diagram Analysis</td>
</tr>
<tr>
<td>2) Reaching Consensus</td>
<td>14%</td>
<td>UT Authoring, cMap Chosen</td>
</tr>
<tr>
<td>3) Encoding Planning</td>
<td>35%</td>
<td>Group Common UTs, ID KR/KE Issues, ID Already Encoded, Write How to Encode</td>
</tr>
</tbody>
</table>

## Encoding (10%)

<table>
<thead>
<tr>
<th>Step</th>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>4) Encoding</td>
<td>10%</td>
<td>Encode, File JIRA Issues</td>
</tr>
</tbody>
</table>

## Testing (39%)

<table>
<thead>
<tr>
<th>Step</th>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>5) Key Term Review</td>
<td>25%</td>
<td>KR Evaluated by Modeling Expert and Biologist, Encoder Makes Changes</td>
</tr>
<tr>
<td>6) Question-Based Testing</td>
<td>14%</td>
<td>Use Minimal Test Suite, Reasoning JIRA Issues Filed, Encoder Fills KB Gaps</td>
</tr>
</tbody>
</table>

**Peach = EVS, Light Blue = SRI**
How Can We Improve Knowledge Entry?

- **This is the well-known Knowledge Acquisition Problem**
  - AI authoring is too expensive, too slow, not scalable

- **Three Possible Solutions**
  - **Automatic Machine Parsing/Editing**
    - Not good enough for biology textbook sentences
    - Error rates are too high
    - Need humans in the loop
  
  - **Mechanical Turk Authoring**
    - Biology expertise is difficult to get
    - Mechanical Turk uses individuals, but the Knowledge Entry task appears to require coordination, judgment, discussion, and working together

  - **Social Authoring and Crowdsourcing**
    - Wikipedia showed this could work for text,
    - In 2008, Vulcan started a pilot project to explore social knowledge authoring
Crowdsourcing for Better Knowledge Acquisition
10 YEARS AGO
HUMAN BEING
BEGAN A
GREAT EXPERIMENT
wow. I can change the web.
let’s share and publish knowledge,
to make an [[encyclopedia]]!
Dive into MediaWiki
Success of Wikis

Actual number of articles on en.wikipedia.org (thick blue line) compared with a Gompertz model that leads eventually to a maximum of about 4.4 million articles (thin green line)
A Key Feature of Wiki

Consensus

This distinguishes wikis from other publication tools
Consensus in Wikis Comes from

- **Collaboration**
  - ~17 edits/page on average in Wikipedia (with high variance)
  - Wikipedia’s Neutral Point of View

- **Convention**
  - Users follow customs and conventions to engage with articles effectively
Software Support Makes Wikis Successful

- Trivial to edit by anyone
- Tracking of all changes, one-step rollback
- Every article has a “Talk” page for discussion
- Notification facility allows anyone to “watch” an article
- Sufficient security on pages, logins can be required
- A hierarchy of administrators, gardeners, and editors
- Software Bots recognize certain kinds of vandalism and auto-revert, or recognize articles that need work, and flag them for editors
However, Finding Information Is ...

Wikipedia has articles about...

... all movies with cast, director, budget, running time, gross...
... all German cars with engine size, accelerating data...

Can you find:
Sci-Fi movies made between 2000-2008 that cost less than $10,000 and made $30,000?

Skyscrapers with 50+ floors and built after 2000 in Shanghai (or Chinese cities with 1,000,000+ people)?

Or German(Porsche) cars that accelerate from 0-100km/h in 5 seconds?
Can Search Solve the Problem?

You may create the page "German cars that accelerate from 0-100km/h in 5 seconds", but consider checking the search results below to see whether the topic is already covered.

**Volkswagen Jetta**

{Infobox automobile ...age:Volkswagen Jetta 2.5 sedan 2.jpg|250px]<br />
"Volkswagen Jetta Mk5 2.5 (US)"

122 KB (17,948 words) - 00:45, 13 December 2011

**Volkswagen Jetta** (section Alternative propulsion cars)

produced by German automaker Volkswagen Group for the Volkswagen Passenger Cars ... Electric could accelerate from 0 to 100 km/h in 10 ...

122 KB (17,765 words) - 00:45, 13 December 2011

**Mercedes-Benz W108**

The Mercedes-Benz W108 and W109 were luxury cars built by Mercedes-Benz from 1965 ... Later models: and 0-100km/h in 10 seconds. During summer ...

37 KB (3,972 words) - 05:11, 27 November 2011

**Electric car** (redirect from Electric cars)

short lived interest in electric cars, but in the mid 2000s took ... records was the breaking of the 100 l km/h l abbr on speed barrier, by ...

107 KB (15,592 words) - 02:15, 11 December 2011

**Renault 5**

The Renault 5 (also called the R5) is ... and sold in the US as Le Car, from 1976 to 1986 ... Giving the Phase II a 0–100 km/h time of 7.5 secs In ...

26 KB (3,945 words) - 22:15, 5 December 2011

**Volkswagen Golf** (redirect from Volkswagen Golf Mk. 5)

The Volkswagen Golf is a small family car manufactured by Volkswagen since
List of Articles And Tables, Info-Boxes, etc.
To Find More Info

- All Porsche vehicles made in Germany that accelerate from 1-100 km/h less than 4 seconds
- Sci-Fi movies made between 2000-2008 that cost less than $10M and gross more than $30M
- A map showing where all Mercedes-Benz vehicles are manufactured
- All skyscrapers in China (Japan, Thailand,…) of 50 (40/60/70) floors or more, and built in year 2000 (2001/2002) and after, sorted by built year, floors…, grouped by cities, regions…
What is a Semantic Wiki

- A **wiki** that has an underlying **model of the knowledge** described in its pages.
- To allow users to make their knowledge explicit and formal
- Semantic Web Compatible
Two Perspectives

Wikis for Metadata

Metadata for Wikis
A SIMPLE EXAMPLE: SEMANTIC MOVIES
Basics of Semantic Wikis

- Still a wiki, with regular wiki features
  - Category/Tags, Namespaces, Title, Versioning, ...

- Typed Content (built-ins + user created, e.g. categories)
  - Page/Card, Date, Number, URL/Email, String, ...

- Typed Links (e.g. properties)
  - “capital_of”, “contains”, “born_in”...

- Querying Interface Support
  - E.g. “[[Category:Member]] [[Age::<30]]” (in SMW)
Characteristics of Semantic Wikis

The Web Connects Information
- Search Engines
- Content Portals
- Databases
- File Servers
- "Push" Pub-Sub

Degree of information connectivity
- Artificial Intelligence
- Personal Assistants
- Ontologies
- Taxonomies

Degree of social connectivity
- Semantic Web
  - Connects Knowledge
  - Semantic Webs
  - Group Minds
  - Lifelogs

The Metaweb Connects Intelligence
- Enterprise Minds
- Knowledge Management
- Knowledge Networks
- Smart Marketplaces

Social Software Connects People
- Enterprise Portals
- Marketplaces
- Auctions
- Groupware

The Global Brain
- Decentralized Communities
- Community Portals
- Social Networks
- USENET
- E-mail

Semantic Wikis
- The Relationship Web
- Semantic Weblogs
- Wikis
- Weblogs
- RSS
- IM
- Conferencing
What is the Promise of Semantic Wikis?

- **Semantic Wikis facilitate Consensus over Data**
- Combine low-expressivity data authorship with the best features of traditional wikis
- User-governed, user-maintained, user-defined
- Easy to use as an extension of text authoring
Semantic Wikis are “Schema-Last”
Databases require DBAs and schema design;
Semantic Wikis develop and maintain the schema in the wiki
<table>
<thead>
<tr>
<th>List of Semantic Wikis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AceWiki</td>
</tr>
<tr>
<td>ArtificialMemory</td>
</tr>
<tr>
<td>Wagn - Ruby on Rails-based</td>
</tr>
<tr>
<td>KiWi – Knowledge in a Wiki</td>
</tr>
<tr>
<td>Knoodl – Semantic Collaboration tool and application platform</td>
</tr>
<tr>
<td>Metaweb - the software that powers Freebase</td>
</tr>
<tr>
<td>OntoWiki</td>
</tr>
<tr>
<td>OpenRecord</td>
</tr>
<tr>
<td>PhpWiki</td>
</tr>
<tr>
<td>Semantic MediaWiki - an extension to MediaWiki that turns it into a semantic wiki</td>
</tr>
<tr>
<td>Swirrl - a spreadsheet-based semantic wiki application</td>
</tr>
<tr>
<td>TaOPis - has a semantic wiki subsystem based on Frame logic</td>
</tr>
<tr>
<td>TikiWiki CMS/Groupware integrates Semantic links as a core feature</td>
</tr>
<tr>
<td>zAgile Wikidsmart - semantically enables Confluence</td>
</tr>
</tbody>
</table>
Semantic MediaWiki and SEMANTIC EXTENSIONS
Semantic MediaWiki Software

- Born at AIFB in 2005
  - Open source (GPL), well documented
  - Vulcan kicks off SMW+ in 2007
- Active development
  - Commercial support available
  - World-wide community
- International Conferences
  - Last SMWCon 4/25-27, 2012 in Carlsbad, CA
  - Next SMWCon 10/24-26, 2012 in Koln, Germany

http://semantic-mediawiki.org/
http://smwplus.com (.net)
Beijing University is located in
[[Has location::Beijing]], with
[[Has population::30000|about 30 thousands]]
students.

In page "Property:Has location":
[[Has type::Page]]

In page "Property:Has population":
[[Has type::number]]
Special Properties

- **“Has Type”** is a pre-defined “special” property for meta-data
  - Example: `[[Has type::String]]`

- **“Allowed Values”** is another special property
  - `[[Allows value::Low]],`
  - `[[Allows value::Medium]],`
  - `[[Allows value::High]]`

- In Halo Extensions, there are domain and range support
  - RDFs expressivity
  - Semantic Gardening extension also supports “Cardinality”
Beijing is a city in [[Has country::China]], with population [[Has population::2,200,000]].

[[Category::Cities]]

Categories are used to define classes because they are better for class inheritance.

The Jin Mao Tower (金茂大厦) is an 88-story **landmark supertall skyscraper** in ...

[[Categories: 1998 architecture | Skyscrapers in Shanghai | Hotels in Shanghai | Skyscrapers over 350 meters | Visitor attractions in Shanghai | Landmarks in Shanghai | Skidmore, Owings and Merrill buildings]]

Category:Skyscrapers in China

Category: Skyscrapers by country
Possible Database-style Query over Data

Ex: Skyscrapers in China higher than 50 stories, built before 2000

ASK/SPARQL query target

```json
{
  "#ask:
  [
    [[Category:Skyscrapers]]
    [[Located in::China]]
    [[Floor count::>50]]
    [[Year built::<2000]]
  ]
}
```
Semantic MediaWiki Stack

MediaWiki (XAMPP)

Extension: Semantic MediaWiki

More Extensions and Applications
MediaWiki ➔ SMW+

MediaWiki
- Powerful Wiki engine
- Basic CMS feature set

Semantic MediaWiki
- Core Semantic Wiki engine
- Authoring of explicit knowledge in content
- Basic reasoning capabilities

Halo Extension
- Usability extension to Semantic MediaWiki
- Increases user consensus
- Increases use of semantic data

SMW+
- Shrink wrap suite of open source software products
- Comes with ready to use ontology
- Easy to procure and install
SMW Extensions – Towards Great Apps

Data I/O
- Halo Extensions, Wiki Object Model, Semantic Forms, Notification, …

Query and Browsing
- Semantic Toolbar, Semantic Drilldown, Faceted Search…

Visualization
- Semantic Result Printers, Tree View, Exhibit, Flash charts…

Other useful extensions
- HaloACL, Deployment, Triplestore Connector, Simple Rules…
- Semantic WikiTags and Subversion Integration extensions
- *Linked Data Extension*, with R2R and SILK from F.U.Berlin
Semantic Wikipedia?

- Ultrapedia: An SMW proof of concept built to explore general knowledge acquisition in a wiki
- Domain: German Cars (in Wikipedia)
- Wikipedia merged with the power of a database
- Help Readers and Writers Be More Productive
## Standard View of the Wiki Data

### Statistics

<table>
<thead>
<tr>
<th>Model</th>
<th>Horsepower, Engine</th>
<th>0-60 mph acceleration</th>
<th>Top Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>911 Carrera</td>
<td>320 hp (239 kW) @ 6800 rpm, 3.6L dry- sump</td>
<td>5.9 s</td>
<td>177 mph (285 km/h)</td>
</tr>
<tr>
<td>911 Carrera 4</td>
<td>320 hp (239 kW) @ 6800 rpm, 3.6L dry- sump</td>
<td>5.1 s</td>
<td>177 mph (285 km/h)</td>
</tr>
<tr>
<td>911 Carrera 4S</td>
<td>320 hp (239 kW) @ 6800 rpm, 3.6L dry- sump</td>
<td>5.7 s</td>
<td>174 mph (280 km/h)</td>
</tr>
<tr>
<td>911 Targa</td>
<td>320 hp (239 kW) @ 6800 rpm, 3.6L dry- sump</td>
<td>5.2 s</td>
<td>174 mph (280 km/h)</td>
</tr>
<tr>
<td>911 Turbo</td>
<td>420 hp (313 kW) @ 6000 rpm, 3.6L dry- sump</td>
<td>3.9 s</td>
<td>197 mph (317 km/h)</td>
</tr>
<tr>
<td>911 Turbo S</td>
<td>444 hp (331 kW) @ 6000 rpm, 3.6L dry- sump</td>
<td>3.7 s</td>
<td>197 mph (317 km/h)</td>
</tr>
<tr>
<td>911 GT3</td>
<td>381 hp (284 kW) @ 7400 rpm, 3.6L dry- sump</td>
<td>4.5 s</td>
<td>190 mph (310 km/h)</td>
</tr>
<tr>
<td>911 GT3 RS</td>
<td>381 hp (284 kW) @ 7400 rpm, 3.6L dry- sump</td>
<td>4.5 s</td>
<td>190 mph (310 km/h)</td>
</tr>
<tr>
<td>911 GT2</td>
<td>483 hp (360 kW) @ 5700 rpm, 3.6L dry- sump</td>
<td>4.0 s</td>
<td>198 mph (319 km/h)</td>
</tr>
<tr>
<td>911 Turbo Cabriolet</td>
<td>420 hp (313 kW) @ 6000 rpm, 3.6L dry- sump</td>
<td>4.3 s</td>
<td>190 mph (310 km/h)</td>
</tr>
<tr>
<td>911 Turbo S Cabriolet</td>
<td>444 hp (331 kW) @ 6000 rpm, 3.6L dry- sump</td>
<td>1.3 s</td>
<td>190 mph (310 km/h)</td>
</tr>
<tr>
<td>911 Carrera Cabriolet</td>
<td>300 hp (224 kW) @ 6300 rpm, 3.4L dry- sump</td>
<td>5.2 s</td>
<td>177 mph (285 km/h)</td>
</tr>
<tr>
<td>911 Carrera 4 Cabriolet</td>
<td>300 hp (224 kW) @ 6300 rpm, 3.4L dry- sump</td>
<td>5.4 s</td>
<td>177 mph (285 km/h)</td>
</tr>
<tr>
<td>911 Carrera 4S Cabriolet</td>
<td>320 hp (239 kW) @ 6800 rpm, 3.4L dry- sump</td>
<td>5.3 s</td>
<td>174 mph (280 km/h)</td>
</tr>
</tbody>
</table>

### Police car

## Dynamic View of the Acceleration Data

### Statistics

<table>
<thead>
<tr>
<th>Model</th>
<th>Power (kw)</th>
<th>Top Speed (kmh)</th>
<th>Acceleration</th>
</tr>
</thead>
<tbody>
<tr>
<td>911 Carrera Cabriolet</td>
<td>223</td>
<td>284</td>
<td>5.2</td>
</tr>
<tr>
<td>911 Carrera 4 Cabriolet</td>
<td>223</td>
<td>284</td>
<td>5.4</td>
</tr>
<tr>
<td>911 Carrera 4 S Cabriolet</td>
<td>238</td>
<td>280</td>
<td>5.3</td>
</tr>
<tr>
<td>911 Carrera</td>
<td>238</td>
<td>284</td>
<td>5.9</td>
</tr>
<tr>
<td>911 Carrera 4</td>
<td>238</td>
<td>284</td>
<td>5.1</td>
</tr>
<tr>
<td>911 Carrera 4 S</td>
<td>238</td>
<td>280</td>
<td>5.7</td>
</tr>
<tr>
<td>911 Targa</td>
<td>238</td>
<td>280</td>
<td>5.2</td>
</tr>
<tr>
<td>911 GT3 RS</td>
<td>284</td>
<td>305</td>
<td>4.5</td>
</tr>
<tr>
<td>911 GT3</td>
<td>284</td>
<td>305</td>
<td>4.5</td>
</tr>
<tr>
<td>911 Turbo Cabriolet</td>
<td>313</td>
<td>305</td>
<td>4.3</td>
</tr>
<tr>
<td>911 Turbo</td>
<td>313</td>
<td>317</td>
<td>3.9</td>
</tr>
<tr>
<td>911 Turbo S Cabriolet</td>
<td>331</td>
<td>305</td>
<td>1.3</td>
</tr>
<tr>
<td>911 Turbo S</td>
<td>331</td>
<td>317</td>
<td>3.7</td>
</tr>
<tr>
<td>911 GT2</td>
<td>360</td>
<td>318</td>
<td>4.0</td>
</tr>
</tbody>
</table>
Graph View of the Acceleration Data

Statistics

Model - power(kw) / top speed(kmh) / acceleration

- power(kw)
- top speed(kmh)
- acceleration (*100)

Model:
- 911 Carrera
- 911 Carrera 4
- 911 Carrera 4S
- 911 Carrera Cabriolet
- 911 GT2
- 911 GT3
- 911 Targa
- 911 Turbo
- 911 Turbo Cabriolet
- 911 Turbo S
- 911 Turbo S Cabriolet
Dynamic Mapping and Charting

Porsche Assembly

Where (cities) are Porsche cars made?

- Porsche Cayman
- Porsche Cayman
- Porsche Cayman
- Porsche Boxster
- Porsche Boxster
Information Discovery via Visualization

Mercedes-Benz Assembly

Where (cities) are Mercedes-Benz cars manufactured?

See all comments and rate this table:

<table>
<thead>
<tr>
<th>Automobile</th>
<th>Assembly in</th>
<th>Length(mm)</th>
<th>Width(mm)</th>
<th>Height(mm)</th>
<th>Weight(kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercedes-Benz 500K</td>
<td>Sindelfingen</td>
<td>4269</td>
<td>1778</td>
<td>1612</td>
<td>2700</td>
</tr>
<tr>
<td>Mercedes-Benz A-Class</td>
<td>Rastatt</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mercedes-Benz B-Class</td>
<td>Ciudad Juarez</td>
<td>4269</td>
<td>1778</td>
<td>1612</td>
<td>2700</td>
</tr>
<tr>
<td>Mercedes-Benz B-Class</td>
<td>Rastatt</td>
<td>4269</td>
<td>1778</td>
<td>1612</td>
<td></td>
</tr>
</tbody>
</table>
We Started With a Wiki, Ended with an App

FROM WIKI TO APPLICATION
Build Social Semantic Web Apps

Choose extensions

Design UI: skin, form & templates

Develop schema (ontology)

Fill in data (content)
Social tag-based characterization
Keyword search over tag data
Inconsistent semantics
Easy to engineer

Algorithm-based object characterization
Database-style search
Consistent semantics
Extremely difficult to engineer

Semantic Entertainment Wiki
- Social database-style characterization
- Database search + wiki text search
- Semantic consistency via wiki mechanisms
- Easy to engineer

Increasing technical complexity →
Increasing User Participation ←
Semantic Seahawks Football Wiki
Semantic Entertainment: Query Result ➔ Highlight Reel

- Commercial Look/Feel
- Play-by-play video search
- Highlight reel generation
- Search on crowd-defined patterns (“touchdowns with big hits”)
- Tree-based navigation widget
- Very favorable economics
AGILE APPLICATION DEVELOPMENT

With Semantic MediaWiki+
SMWCon Spring 2012 is on April 25-27, in Carlsbad, California

SMWCon Spring 2012 will include presentations and discussions about state-of-art applications and future development of Semantic MediaWiki, its extensions and more. Participants are invited to present their own work and to discuss experiences and ideas. While the conference will have scheduled sessions, each will encourage free discussion and social interaction, including lots of group time and social events.

Discover SMWCon

- SMWCon Overview
- Organization Committee
- Travel, Venue, Location and Things to do
- Registration

Conference Program

<table>
<thead>
<tr>
<th>Tutorial</th>
<th>Wednesday, April 25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conference</td>
<td>Thursday, April 26</td>
</tr>
<tr>
<td>Lightning talks</td>
<td>All lightning talks</td>
</tr>
</tbody>
</table>

Breakfast for Business

Free: everyone welcome!

SMWCon Spring 2012

Play With Data

@aiwang: With @ai_one's sponsorship and support from @BarbaraStarr, @gwsuperfan, @mmarsh23 etc, we expect record
Who Should Attend
The conference brings together developers, users, and organizations from the Semantic MediaWiki community from around the world. There are tutorials and conference presentations for new and seasoned developers, popular topics for business analysts and researchers, interesting demos for data architects as well as sales managers. New in this SMWCon is a free-to-all breakfast session focused on the application of Semantic MediaWiki technology for business and crowdsourced applications. Read More...

Keynote Talks

Thursday Keynote
26 April 2012 09:15:00 - 10:00:00
Semantics and Wikis, from Genomes to X Games

Smart Cities and Open Data
27 April 2012 08:55:00 - 09:40:00
Emerging trends in Smart Cities and Open Data; a quick overview of the NYCFacets Smart Data Exchange; and a sneak peek at NYCpedia.

Interested in Speaking or Socializing?
As this is a community based event, everyone is encouraged to communicate and tweet #smwcon. Please feel free to submit a proposal, from a regular presentation to a lightning talk, or even just a social event such as Movie screen after dinner Wednesday, or a remote talk (say from Australia or China). Submit your idea now...

Latest news

2012/04/23 10:06:25
We’ll broadcast

2012/04/11 07:00:00
Pre-SMWCon meetup in San Diego

2012/04/03 08:32:28
SMWCon 2012 Event on LinkedIn

2012/04/03 05:19:39
Registration is Live (via EventBrite)
**Personal Info**

**Nick Name:** Jesse Wang  
**Wiki experience:** Developer or Sponsor

**Industry:** Computer Software  
**SMW experience:** Developer or Sponsor

**Organization size:** 100-500  
**Semantic web experience:** Developer or Sponsor

**Location:** 505 5th Ave S., Seattle, WA 98104, US

**LinkedIn:** [http://www.linkedin.com/in/jiaxinwang](http://www.linkedin.com/in/jiaxinwang)

---

**Jesse Wang**

Architect (Knowledge and Productivity Systems) at Vulcan, Inc.  
Greater Seattle Area

[View Profile](http://www.linkedin.com/in/jiaxinwang)

---

**About Me:** Welcome to SMWCon Spring 2012! I am one of the creators and organizers of this conference, and happy to see you checking us out!  
At work, I contribute to [Project Halo](http://www.vulcaninc.com/) in [Vulcan Inc.](http://www.vulcaninc.com/). A big part of my job is to lead the R&D of social semantic web related efforts, including various methods to make Semantic MediaWiki more useful and usable with data, structure, semantics, or knowledge (if there are nuances among them;-) Our near-term goal is to make SMW a viable platform for knowledge acquisition and processing, the foundation of artificial intelligence.

Again, welcome to SMWCon, let's Share Knowledge and Play with Data!

**Twitter:** aiwang
Conference Schedule

07:30:00 - 09:00:00 Breakfast for Business
We network, mingle and discuss real world problems and business applications that relate to semantic wikis or beyond.

Social Events
We'll have people from outside semantic wiki world to come to learn and share their issues, we'll also have semantic mediawiki experts and consultants on site to help answer some questions, and even provide free consulting time. Definitely a great opportunity for communication and even making deals:)

For one who's not attending the conference, please sign on the meetup to help us order breakfast. (WikiSysop)

08:55:00 - 09:40:00 Smart Cities and Open Data
Emerging trends in Smart Cities and Open Data; a quick overview of the NYCFacets Smart Data Exchange; and a sneak peek at NYCpedia. (Joel Natividad, Sami Baig)

09:45:00 - 10:15:00 First report from the Wikidata project
First report from the Wikidata project (Jeroen De Dauw)

10:15:00 - 10:30:00 Coffee Break (Friday Morning)

Social Events
Coffee Break (WikiSysop)

10:30:00 - 11:00:00 SMW in the organization (Part I)
Data Discovery and Semantic MediaWiki (Desiree Gennaro)

11:00:00 - 11:30:00 SMW in the Organization (Part II)
OpenEI: Organizing energy information with SMW (Rmckeel)

11:30:00 - 12:00:00 Cooking with KEF
KEF: Knowledge Encapsulation Framework Learn how PNNL uses KEF to do their knowledge processing (Kfligg, Mikemadison)

12:00:00 - 13:00:00 Lunch (Friday)

Social Events
Lunch (WikiSysop)

13:00:00 - 13:30:00 Velo: a reusable, domain-independent knowledge management framework for modeling and simulation
Click here to submit a new proposal, it can be a lightning talk, or a social event...
From Agile to Quick Agile

MINI-OLYMPIC APPLICATION IN 2 DAYS
Application: Mini-Olympics

Contents

1. Vulcan Mini-Olympic Games for Food Fighting
   - 1.1 Location and Dates
     - 1.1.1 How it Works
   - 2 Power and Strength Games
   - 3 Power and Strength Games
   - 4 Power and Strength Games

Vulcan Mini-Olympic Games for Food Fighting [edit]

Goal

Fund-raising for Northwest Harvest

Games and activities are mostly crowd-sourced and new fun, safe, and easy to measure games can still be accepted. Check out the games below. Each game/activity will have men's and women's winners.

Location and Dates [edit]

Tuesday in Great Hall - Qualifications
<table>
<thead>
<tr>
<th>Name</th>
<th>Image</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Jump Rope Marathon</strong></td>
<td><img src="image_url" alt="Jump Rope" /></td>
<td>Jump rope to reach 1000 and 5000 rounds, and whoever with the fewest times wins.</td>
</tr>
<tr>
<td><strong>Pushup</strong></td>
<td><img src="image_url" alt="Pushup" /></td>
<td>Pushups - loose standard (but does require the chest to be down and knee or belly can’t touch the ground). Who did the most pushups within a time frame (say 3 minutes) is the winner.</td>
</tr>
<tr>
<td><strong>Situps in One Minute</strong></td>
<td><img src="image_url" alt="Situps" /></td>
<td>A complete situp requires your hands behind your head, and your elbow or head must to touch your knee (or thigh). An alternative is to have your hands touch your neck or shoulder, especially if you feel this posture is easier on your neck. Whoever can finish most in a minute wins.</td>
</tr>
<tr>
<td><strong>Stair Jump</strong></td>
<td><img src="image_url" alt="Stair Jump" /></td>
<td>Jump Stairs from 10th floor to 11th floor (in Great Hall). The athlete who achieves this with the fewest jumps wins. In case with a tie, the one with least number of stairs left in the last jump wins. If it is still a tie, the shorter player wins.</td>
</tr>
<tr>
<td><strong>Standing Long Jump</strong></td>
<td><img src="image_url" alt="Standing Long Jump" /></td>
<td>The Standing long jump, also called the Broad Jump, is a common and easy to administer test of explosive leg power. It is one of the fitness tests in the NFL Combine. The standing long jump was also once an event at the Olympic Games, and is also an event in Sports Hall competitions in the UK.</td>
</tr>
</tbody>
</table>
**Type:** Scored Sport

**Category:** Speed and Accuracy

Toss the bean bags (food items) into corresponding animal's mouth. If you hit it right with enough strength, the animal's hat will come up - defined as a successful feed.

From about about 15 feet distance. Each player can have up to 3 attempts (but you can practice as many times as you want when no other players are competing).

**Scoring:**

- Each successful and correct feed (right food for animal) is 10 points.
- Each successful but incorrect feed is 7 points
- Each leftover food item worth 2 points

That is to say, you can intentionally leave a beanbag (not used) to make a difference if you are not sure to score with it.

*Hitting the board hard to cause the hats up does not count as scoring*

---

**Men's Leaderboard**

<table>
<thead>
<tr>
<th>Athlete</th>
<th>Gender</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeffrey Talley</td>
<td>Men</td>
<td>20</td>
</tr>
<tr>
<td>Aron Battle</td>
<td>Men</td>
<td>19</td>
</tr>
<tr>
<td>Myles Brooks</td>
<td>Men</td>
<td>19</td>
</tr>
</tbody>
</table>
# Athlete Info

<table>
<thead>
<tr>
<th>Name:</th>
<th>Department:</th>
<th>Technology</th>
</tr>
</thead>
</table>

### Profile Image:

![Profile Image]

### In Sports:

- Feed the Animals
- Standing Long Jump
- Reaction Ball Catch
- Flying Paper Airplane
- Pushup
- Javelin with Straws

### About me:

Welcome to Vulcan Mini-Olympics! I hope you will enjoy this fund-raising event.

You will learn how to use this software brought to you by Project Halo in Technology R&D! It lets you create a profile, attend a game, and record your result! It also allows you to display the leaders and more.

Please feel free to email me at jessew or call me if you have any questions or need help. And enjoy!

(￣︶￣)

Cheers,
Requirements for Wiki “Developers”

- One need not
  - Write code like a hardcore programmer
  - Design, setup RDBMS or make frequent schema changes
  - Possess knowledge of a senior system admin

- Instead one need
  - Configure the wiki with desired extensions
  - Design and evolve the data model (schema)
  - Design Content
    - Customize templates, forms, styles, skin, etc.
Social Semantic Web App
In Days!
# Effectiveness of SMW as a Platform Choice

<table>
<thead>
<tr>
<th>Packaged Software</th>
<th>SMW+ Suite</th>
<th>Custom Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>☺ Very quick to obtain</td>
<td>☺ Still quick to program</td>
<td>☹ Slow to develop</td>
</tr>
<tr>
<td>☹ Hard to customize</td>
<td>☺ Easy to customize</td>
<td>☹ Extremely flexible</td>
</tr>
<tr>
<td>☹ Expensive</td>
<td>☺ Low-moderate cost</td>
<td>☹ High cost to develop and maintain</td>
</tr>
<tr>
<td>- Microsoft Project</td>
<td>- Mini-Olympics</td>
<td>- .NET Framework</td>
</tr>
<tr>
<td>- Version One</td>
<td>- Scrum Wiki</td>
<td>- J2EE, …</td>
</tr>
<tr>
<td>- Microsoft SharePoint</td>
<td>- RPI map</td>
<td>- Ruby on rails</td>
</tr>
</tbody>
</table>
Can we use SMW to acquire relationship information?
- From data focus (movies) to relationship focus (Neurowiki)
- A Neurowiki user cannot contribute data without also contributing relationships

Experimental goals:
- Create an SMW instance for acquiring both instance and relationship information between data elements
- Integrate SMW with the Berlin LDIF system

Product goals:
- A community-driven collaborative framework for neuroscience data integration
  - Support for community-built neuroscience vocabularies
  - A library of visualizations and simple analytics
- Neuroscientist users can create/modify pages and metadata
  - Scientists can own the evolving knowledge structure
  - Low barrier for scientist to publish data
- Support scientific discovery
Using SMW+ to crowdsource knowledge of mapping and linking

NEUROWIKI DEMO
Reduce costs and increase scale of AURA authoring
- A real, high-quality application of SMW+
- Focus on creating universally-quantified sentences (UTs)
  - Example: “Glycolysis, which occurs in the cytosol, begins the degradation process by breaking glucose into two molecules of a compound called pyruvate.”
    - UT: Glycolysis occurs in Cytosol (concept: Glycolysis)
    - UT: During cellular respiration, glycolysis is the first step (concept Cellular Respiration)
    - UT: During glycolysis, breakdown of 1 glucose results in 2 pyruvate molecules (concept: Glycolysis)
  - Use these UTs as inputs to the AURA authoring process

Use SMW-acquired statements as inputs to the process
- Potential savings of over 50% of AURA authoring time

Use the social semantic web for real AI authoring
AURAwiki Strategy

- Provide a community of authors with everything they need to collaborate on high-quality UT authoring:
  - Top-down Encoding Flow
    - Suggestions and commentary from KE (Knowledge Engineers) and SMEs previous contributions
    - Supported by ontological terms and relationships, questions, guidelines, testing
  - Bottom-up Encoding Flow
    - UT suggestions based upon algorithms (ReVerb by University of Washington)
    - All suggestions validated via Mechanical Turk before use in AURAwiki
  - Build a “Knowledge Integration Center”

- Current Page Designs
  - Main Page
  - User Page
  - Authoring Page

- Will perform initial testing by September 15
Inquire is a new application of AI technology
- Question-answering in textbooks
- Precise, embedded answers over a complex subject matter

Semantic wikis and SMW+ use social and crowd phenomena as a new way to address the knowledge acquisition problem
- Crowdsourcing addresses the cost and update problems
- Validate using Inquire authoring process

From text (textbook sentences) to data (knowledge encodings) to knowledge (use in a QA system): The Social Semantic Web in Action
Thank You

www.vulcan.com
www.projecthalo.com
www.inquireproject.com
www.smwplus.com
Quiz and Homework Scores

P-value from 2 tailed t-tests:

**HW scores**
- Full vs. Ablated: 0.12
- Full vs. textbook: 0.02
- Ablated vs. Text: 0.52

**Quiz scores**
- Full vs. Ablated: 0.002
- Full vs. textbook: 0.05
- Ablated vs. Text: 0.18
Quiz Score Distribution

% of students

A B C D F

0% 10% 20% 30% 40% 50%

inquire

inquire

textbook

inquire

PROJECT HALO