

# Semantic MediaWiki and the “SMW family”

Yaron Koren  
Wikimedia Data Summit  
February 4, 2011

# About me

- Based in New York City
- MediaWiki developer and consultant since 2007, with extremely strong focus on Semantic MediaWiki
- Run the MediaWiki consulting company WikiWorks ([wikiworks.com](http://wikiworks.com))
- Run the MediaWiki-based wiki farm Referata ([referata.com](http://referata.com))

# Semantic MediaWiki

A MediaWiki extension that lets you store and query semantic data – data stored in semantic triples.

Initially written by Markus Krötzsch and Denny Vrandečić

Now mostly maintained by Markus Krötzsch and Jeroen De Dauw

<http://semantic-mediawiki.org>

# SMW - example

On the page “Mexico”:

```
[ [Has population::112,468,855] ]
```

In the SMW database table, this is stored as:

<b>Mexico</b>	<b>Has population</b>	<b>112,468,855</b>
(subject)	(predicate)	(object)

# SMW – example part 2

I believe all SMW data should be stored using templates.

So, instead, on “Mexico” page:

```
{{Country
|population=112,468,855
...
}}
```

Template “Country” then contains: “[ [Has population::{{population|}}] ]”

# Displaying SMW data

SMW data can be queried from any page, to produce lists, tables, etc.

<b>Country</b>	<b>Population</b>
...	...
Mexico	112,468,855
...	...

It can also be *exported*, via CSV, RDF, etc.

# Semantic MediaWiki is fairly popular

Over 200 **active** public wikis use it.

Probably around the same number of private wikis.

Thus, in use on over 400 active wikis.

Has a twice-yearly conference: “SMWCon”



It's kind of a big deal



# Some organizations that use SMW:

- 3M
- Audi
- Boeing
- Creative Commons
- IBM
- Johnson & Johnson
- Pfizer
- U.S. Department of Defense
- U.S. Department of Energy
- Wikia

# The SMW “family”

Over 30 additional MediaWiki extensions exist, that rely on SMW and/or enhance its functionality in some way.

# Top 10 most popular SMW-based extensions

(% of SMW-based wikis that use each extension - of the current set of 216 known active, public SMW-based wikis)

1. Semantic Forms - 85%
2. Semantic Result Formats - 33%
3. Semantic Drilldown - 30%
4. Semantic Maps - 25%
5. Semantic Compound Queries - 19%
6. Semantic Google Maps - 7%
7. Semantic Internal Objects - 6%
8. Semantic Forms Inputs - 6%
9. Semantic Tasks - 5%
10. Halo - 3%

# Semantic Forms (85%)

- Mostly written by me
- Lets users define forms for editing template calls within pages (based on my philosophy that SMW data should always be stored in templates)
- A "framework" for page creation and editing
- Lots of features – autocompletion, input validation, file uploads, etc.

# Semantic Result Formats (33%)

- Written by many people, including me and Denny
- A catch-all collection of additional "query formats" for displaying SMW data, including calendars, timelines, charts, etc.

# Semantic Drilldown (30%)

- Mostly written by me
- Provides a drill-down interface for SMW data, filtering on the values of different properties

# Semantic Maps (25%)

- Written by Jeroen De Dauw
- Displays coordinate-based SMW data using Google Maps, Yahoo! Maps and Open Layers

# Semantic Compound Queries (19%)

- Written by me
- Allows for displaying multiple queries together; mostly useful for maps



# Semantic Internal Objects (6%)

- Written by me
- Allows for storing compound/"n-ary"/tabular data within pages

Example: “Grover Cleveland was president of the U.S. from 1885 to 1889, and then again from 1893 to 1897.”

That information can't (easily) be stored on a single page using SMW, but it can using Semantic Internal Objects.

# Potential use of SMW in WMF projects

There's a case to be made to use SMW in every WMF site, except maybe Wiktionary.

On Wikipedia: would eliminate the need for many category and list pages; possibly eliminated inter-wiki data redundancy.

On non-Wikipedia sites, Semantic Forms might make a lot of sense also.

Czech Wikiversity has already requested the usage of SMW + Semantic Drilldown.

# Why not use SMW in WMF projects?

Performance impact is the biggest potential downside.

Performance is mostly related to amount and complexity of queries - querying can be shut off entirely.

Very relevant is a current planned change to SMW: allow for storing data within an **RDF triplestore** instead of the wiki's relational DB - this will speed up queries, and prevent SMW activity from affecting wiki's performance