SMW
Managing Data in MediaWiki

Markus Krötzsch

1) SMW Project
2) Computing Laboratory, University of Oxford
3) Institute AIFB, Karlsruhe Institute of Technology

Wikimania 2010
SMW – Semantic MediaWiki

MediaWiki for collaborating on

- Text
- Rich text and media
- Data
SMW – Semantic MediaWiki

MediaWiki for collaborating on

- Text
- Rich text and media
- Data

That’s where SMW comes in
Part I: What SMW Does
Organising Data in SMW

Basic idea:
Assign property-value pairs to wiki pages
Organising Data in SMW

Basic idea:
Assign property-value pairs to wiki pages

<table>
<thead>
<tr>
<th>Page</th>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gdansk</td>
<td>Population</td>
<td>455,830</td>
</tr>
<tr>
<td>Wikimania 2010</td>
<td>Starts at</td>
<td>July 9th 2010</td>
</tr>
<tr>
<td>Markus</td>
<td>Lives in</td>
<td>Oxford</td>
</tr>
</tbody>
</table>
Why?

1) Searching & querying data in the wiki

2) Sharing data with other sites & apps
Example: AIFB Web Portal

- http://www.aifb.kit.edu/
- Topic: People, Events, Publications
Bearbeiten von Denny Vrandecic

{{Mitarbeiter
|Vorname=Denny
|Nachname=Vrandecic
|Akademischer Titel=Dipl.-Inf.
|Kürzel=dv
|Forschungsgruppe=Wissensmanagement
|Stellung=Wissenschaftlicher Mitarbeiter
|Ehemaliger=False
|Telefon=+49 721 608 6592
|Fax=+49 721 608 6580
|Email=denny(at)aifb.uni-karlsruhe.de
|Raum=262
|Bild=dv.png
|Publikationen anzeigen=True
|Vorträge anzeigen=True
|Organisation=AIFB, KIT
}}

Bitte beachte, dass alle Beiträge zu AIFBportal von anderen Mitwirkenden bearbeitet, geändert oder gelöscht werden können. Reiche hier keine Texte ein, falls du nicht willst, dass diese ohne Einschränkung geändert werden können.

Du bestätigst hiermit auch, dass du diese Texte selbst geschrieben hast oder diese von einer gemeinfreien Quelle kopiert hast (siehe AIFBportal-Urheberrecht für weitere Details). ÜBERTRAGEN OHNE GENEHMIGUNG KEINE URHEBERRECHTLICH GESCHÜTZTEN INHALTE!

Zusammenfassung:
<table>
<thead>
<tr>
<th>Eigenschaft</th>
<th>Wert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vorname</td>
<td>Denny</td>
</tr>
<tr>
<td>Nachname</td>
<td>Vrandecic</td>
</tr>
<tr>
<td>Akademischer Titel</td>
<td>Dipl.-Inf.</td>
</tr>
<tr>
<td>Kürzel</td>
<td>dvr</td>
</tr>
<tr>
<td>Forschungsgruppe</td>
<td>Wissensmanagement</td>
</tr>
<tr>
<td>Stellung</td>
<td>Wissenschaftlicher Mitarbeiter</td>
</tr>
<tr>
<td>Telefon</td>
<td>+49 721 608 6592</td>
</tr>
<tr>
<td>Mobile</td>
<td></td>
</tr>
<tr>
<td>Fax</td>
<td>+49 721 608 6580</td>
</tr>
<tr>
<td>Email</td>
<td>denny(at)aifb.uni-karlsruhe.de</td>
</tr>
<tr>
<td>Raum</td>
<td>262</td>
</tr>
</tbody>
</table>
Personen

Professoren

Andreas Oberweis
Email: oberweis(at)aifb.uni-karlsruhe.de
Tel.: +49 721 608 4516
Raum: 2A-19

Hartmut Schmeck
Email: hartmut.schmeck(at)kit.edu
Tel.: +49 721 608 4242
Raum: 2C-16

Detlef Seese
Email: detlef.seese(at)kit.edu
Tel.: +49 721 608 6037
Raum: 2C-01

Wolffried Stucky
Email: stucky(at)aifb.uni-karlsruhe.de
Tel.: +49 721 608 3812
Raum: 2A-01

Rudi Studer
Email: studer(at)aifb.uni-karlsruhe.de
Tel.: +49 721 608 3923
Raum: 216

Stefan Tai
Email: tai(at)aifb.uni-karlsruhe.de
Example: SNPedia

- http://www.snpedia.com/
- Topic: SNPs (DNA pieces), diseases, ...
Rs1234

This page represents an arbitrary SNP. The main text body of a SNP page summarizes the associations between SNP variants and a range of human traits (body height, disease risk, response to drugs, even behavior). The science that supports these associations (peer-reviewed publications with citations) is indicated by links to PubMed abstracts in the form of PMID citations.

Technically, there was once a valid rs1234, but it has long since been merged into a different identifier, due to the natural changes of that system.

For this reason, on SNPedia, we will use this SNP to explain how to use many key SNPedia features.

To the upper right you will see a box with these fields:
- is a : snp. This tells you that you are looking at a page about a snp. You may sometimes see the words gene or genotype.
- is mentioned by : This will show you what other pages on SNPedia are related to this snp.
- dbSNP rs1234 : This will take you to the the graphs of population diversity on its ncbi dbSNP page.
- hapmap rs1234 : overlay snpedia on the hapmap genome.
- ensembl rs1234 : Takes you to the snp report at ensembl.
- google : Automatic google search.
- 23andMe : Links to the 23andMe.com web site where registered users can find their own genotypes for this SNP.

The colored box indicates the frequency of each genotype in 4 HapMap populations. CEU is European so the box indicates that 10% of Europeans are (A, A) colored brown, 60% are (A, C) colored green, and 30% are (C, C) colored blue. The Chinese (CHB) and Japanese (JPT) are mainly (C, C). Click on the question mark for more information.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1234(A, A)</td>
<td>summary of something interesting</td>
</tr>
<tr>
<td>rs1234(A, C)</td>
<td>summary of something less interesting</td>
</tr>
<tr>
<td>rs1234(C, C)</td>
<td>normal</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Magnitude</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rs1234(A, A)</td>
<td></td>
<td>summary of something interesting</td>
</tr>
<tr>
<td>Rs1234(A, C)</td>
<td></td>
<td>summary of something less</td>
</tr>
</tbody>
</table>
**Promethease Report**

**Version:** 0.1.93

**Generated:** 2010-07-09 01:36

**Infile:** insiders/inputs/ genome_Misha_Angrist_20090303050118.txt.gz, insiders/inputs/ genome_Misha_Angrist_NAVigenics.txt.gz, insiders/inputs/pgp/GM21677.bed.gz

8594 genotypes annotated

**Reference Population CEU**

<table>
<thead>
<tr>
<th>Snps</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>gs1068</strong></td>
<td>Mitochondrial haplogroup V is believed to have originated approximately 12,000 years before present, possibly in Iberia. It is found with particularly high concentrations in the Netherlands, the Croatian Islands, Scotland and northern Scandinavia, as well as the Basque people and somewhat higher among the isolated Basques in nearby Cantabria. <strong>See Also:</strong> <em>13001813</em> <em>rs35788393</em> <em>13001464</em>* Links: <strong>Ian Logan's mtDNA Website</strong> <a href="https://en.wikipedia.org/wiki/Mitochondrial_DNA">Wikipedia mtDNA</a></td>
</tr>
<tr>
<td><strong>gs144</strong></td>
<td>Male</td>
</tr>
<tr>
<td><strong>gs115</strong></td>
<td>Paternal haplogroup R1a1 R1a arose 15,000 years ago in the vicinity of Ukraine. Today it is found at high frequency (more than 40%) from the Czech Republic across to the Altai Mountains in Siberia and south throughout Central Asia. rs3908 is the SNP that distinguishes R1a1 from R1a.</td>
</tr>
<tr>
<td><strong>gs140</strong></td>
<td>NAT2 slow metabolizer source Most non-Scandinavian Caucasians and approximately half of African-Americans are slow metabolizers. This variation is important because of its primary role in the deactivation of many chemicals in the body's environment, including those produced by caffeine and cigarettes as well as aromatic amine and hydrazine drugs used medicinally. In general, slow metabolizers have higher rates of certain types of cancer and are more susceptible to side effects from chemicals metabolized by NAT2.</td>
</tr>
<tr>
<td><strong>gs151</strong></td>
<td>CYP2C19 intermediate Metabolizer. Your body breaks down some medicines at a slightly slower than normal rate (which is represented by gs150). Individuals with gs152 genotypes have even slower metabolism.</td>
</tr>
<tr>
<td><strong>gs153</strong></td>
<td>CYP2C19 ultra-rapid metabolizer. Your body breaks down (ie detoxifies) certain medicines more quickly. It is possible that you may need a higher than average dosage of the medicines metabolized by CYP2C19 in order for them to be effective. *some hormones (such as estrogens and progesterone) *anti-epilotics (such as diazepam, phenytoin, and phenobarbitone) *anti-depressants (such as amitriptylne and clomipramine) *Clopidogrel Plavix Plavix Nexium Prevacid</td>
</tr>
</tbody>
</table>
| **gs157** | CYP1A2 slow metabolizer. The same amount of caffeine will tend to have more stimulating effect on slow metabolizers than on fast metabolizers.
Popular

- rs1815739 sprinters vs endurance athletes
- rs4481887 “asparagus anosmia”, the inability to smell the methanethiol produced after eating asparagus
- rs7412 and rs429358 can raise the risk of Alzheimer's disease by more than 10x
- rs6152 can influence baldness
- rs333 resistance to HIV
- rs1800497 in a dopamine receptor may influence the sense of pleasure
- rs1805007 determines red hair and sensitivity to anesthetics
- rs9939609 triggers obesity and type-2 diabetes
- rs662799 prevents weight gain from high fat diets
- rs7495174 green eye color and rs12913832 for blue eye color
- rs7903146 in 3% of the population greatly increases the risk of type-2 diabetes
- rs12255372 linked to type-2 diabetes and breast cancer
- rs1799971 makes alcohol cravings stronger
- rs17822931 determines earwax
- rs4680 varied cognitive effects
- rs1333049 coronary heart disease
- rs1801133 folate metabolism and several cancers
- rs1051730 and rs3750344 nicotine dependence
- rs3057 perfect musical pitch
- rs4988235 lactose intolerance

View all snps
Further Examples

- WikiTaaable: Smart recipe suggestions
- Chickipedia: censored
- IT resource management & ticketing
- Innovation management (various companies)
- ... and many more
- ... essentially any Wiki with “Infoboxes”
Part II: How SMW Works
Editing Data

Typical (but not the only) wiki syntax:

```
[[property::value]]
```

<table>
<thead>
<tr>
<th>Wiki page</th>
<th>Wikitext</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gdansk</td>
<td>[[population::455,830]]</td>
</tr>
<tr>
<td>Wikimania 2010</td>
<td>[[starts at::July 9 2010]]</td>
</tr>
<tr>
<td>Markus</td>
<td>[[lives in::Oxford]]</td>
</tr>
</tbody>
</table>
Properties

Properties in the wiki ...

- ... can be introduced as needed
- ... have their own wiki pages
- ... may be given a datatype

Examples: Capital of, Population, Start date, Name, Homepage, ...
Datatypes in the wiki ...

- ... affect how data values are displayed
- ... influence searching and browsing
- ... are usually not user-defined

Examples: Number, Text, Page, Date, URL, ...
Inline Queries

Basic idea:
Allow dynamic views on wiki data to be embedded into pages.
Aspects of an inline query:

1) Which pages constitute the result?
2) Values for which properties are displayed per result?
3) How should all this be formatted?
Inline Queries

Example:
1) Show all countries in Africa
2) each with their population and area
3) within a table.
Inline Queries

User input:

1) SMW query language
2) Additional printout statements
3) Formatting parameters
Inline Queries

Example:

```{}
{{#ask: [[Category:Country]]
   [[located in::Africa]]
   | ?Population
   | ?Area
   | format=table
}}
```
## Inline Queries

### List of African countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Population</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>32,854,000</td>
<td>2,381,740,000,000 m²</td>
</tr>
<tr>
<td>Angola</td>
<td>15,941,000</td>
<td>1,246,700,000,000 m²</td>
</tr>
<tr>
<td>Benin</td>
<td>8,439,000</td>
<td>112,620,000,000 m²</td>
</tr>
<tr>
<td>Botswana</td>
<td>1,765,000</td>
<td>600,370,000,000 m²</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>13,228,000</td>
<td>274,200,000,000 m²</td>
</tr>
<tr>
<td>Burundi</td>
<td>7,548,000</td>
<td>27,830,000,000 m²</td>
</tr>
<tr>
<td>Cameroon</td>
<td>16,322,000</td>
<td>475,440,000,000 m²</td>
</tr>
<tr>
<td>Cape Verde</td>
<td>507,000</td>
<td>4,033,000,000 m²</td>
</tr>
<tr>
<td>Central African Republic</td>
<td>4,038,000</td>
<td>622,984,000,000 m²</td>
</tr>
<tr>
<td>Chad</td>
<td>10,748,000</td>
<td>1,128,000,000,000 m²</td>
</tr>
</tbody>
</table>
Part III: Reflexion and Outlook
Lessons Learnt

1) “Semantics” tends to be over-mystified

2) Structuring data can be a challenge

3) “There is no free lunch.”
Beyond SMW Core

Active user community

• Some hundreds of sites known using it
• World-wide usage (20-50 languages)
• Public mailing lists, forum, bugtracker

Active developer community

• Various independent extensions
• Teams at ontoprise and KIT
• Integrated into MediaWiki community
SMWCon 2010

5th International SMW Meetup

Sept 18–19, 2010
Amsterdam

http://semantic-mediawiki.org/wiki/SMWCon
Outlook

SMW: basic data (or metadata) management for MediaWiki

Future goals for SMW Core:

- Stability (including performance)
- Code quality
- Extendibility

Stay tuned for more cool stuff
SMW: http://semantic-mediawiki.org
Copyright Markus Krötzsch

Terms of Licensing

Creative Commons Attribution

http://creativecommons.org/licenses/by/3.0/de/deed.en

All redistributions and reuses of this work must include the following:

“Based on slides by Markus Krötzsch, http://korrekt.org”

Image copyrights

Icons by http://everaldo.com, released under the terms of GPL