



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

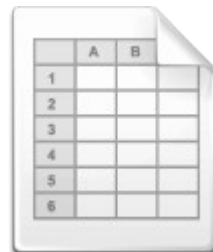
- Text



- 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

| Page | Property | Value |
|----------------|------------|---------------|
| Gdansk | Population | 455,830 |
| Wikimania 2010 | Starts at | July 9th 2010 |
| Markus | Lives in | Oxford |

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

A screenshot of the AIFB Web Portal homepage. The page features a header with the KIT logo and navigation links. A search bar is located below the header. The main content area is divided into several sections: a welcome message, a list of professors, and two news items. The right sidebar contains sections for 'Neuigkeiten' (News), 'Veranstaltungen' (Events), and 'Shortcuts'. The footer includes the text 'KIT - Universität des Landes Baden-Württemberg und nationales Forschungszentrum in der Helmholtz-Gemeinschaft' and 'NACH OBEN'.

- ▀ Einstellungen
- ▀ Spezialseiten
- ▀ Bearbeiten
- ▀ Versionen/Autoren
- ▀ Löschen
- ▀ Verschieben
- ▀ Aktualisieren

AIFB

Latest News

2009-11-20

[2nd trend congress net economy in Karlsruhe](#)

[more...](#)



Events

2009-11-18

[Large-Scale Pattern-Based Information Extraction from the World Wide Web](#)

2009-11-12

[Inaugural Event KIT Focus COMMputation](#)

[more...](#)



[Faculty of Economics and Business Engineering](#)

Shortcuts

[Theses](#)
[Contact](#)
[Verein AIK](#)
[Intern](#)

▸ Home

▸ Teaching and Exams

▾ People

Professors
 Administration
 Scientific Staff
 Honorary Professors
Lecturers
 Guests and External Staff
 Technics
 Trainees
 Student Assistants
 Former Members

▸ Research Groups

▸ Publications

Events

▸ Projects

▸ Cooperations

Jobs

Contact



Dipl.-Inf. Denny Vrandeic

Research Associate

[vCard](#)

Phone: +49 721 608 6592

Telefax: +49 721 608 6580

Email: [denny\(at\)aifb.uni-karlsruhe.de](mailto:denny(at)aifb.uni-karlsruhe.de)

Research Group: [Knowledge Management](#)

Room: [262](#) (11.40)

Publications

[Publications](#)

Talks

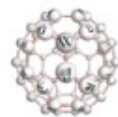
[Talks](#)

Active Projects



Active

<http://www.active-project.eu>



SMW

<http://semantic-mediawiki.org>



STI2

<http://www.sti2.org>

Activities

- ▀ Einstellungen
- ▀ Spezialseiten
- ▀ mit Formular bearbeiten
- ▀ Bearbeiten
- ▀ Versionen/Autoren
- ▀ Löschen
- ▀ Verschieben
- ▀ Aktualisieren

AIFB

[Home](#)
[Lehre und Prüfung](#)
[Personen](#)

Professoren
 Organisation
**Wissenschaftl.
 Mitarbeiter**
 Honorarprofessoren
 Lehrbeauftragte
 Gäste und Externe
 Technik
 Auszubildende
 Hilfskräfte
 Ehemalige

[Forschungsgruppen](#)
[Veröffentlichungen](#)
[Veranstaltungen](#)
[Projekte](#)
[Kooperationen](#)
[Stellenmarkt](#)
[Kontakt](#)

Bearbeiten von Denny Vrandecic



```

{ {Mitarbeiter
|Vorname=Denny
|Nachname=Vrandecic
|Akademischer Titel=Dipl.-Inf.
|Kürzel=dvr
|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=dvr.jpeg
|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). **ÜBERTRAGE OHNE GENEHMIGUNG KEINE URHEBERRECHTLICH GESCHÜTZTEN INHALTE!**

 Zusammenfassung:

Neuigkeiten

20. November 2009

[2. Trendkongress net economy in Karlsruhe](#)
[mehr...](#)


Veranstaltungen

18. November 2009

[Large-Scale Pattern-Based Information Extraction from the World Wide Web](#)

12. November 2009

[Eröffungsveranstaltung KIT-Schwerpunkt COMMputation](#)
[mehr...](#)


[Fakultät für
Wirtschaftswissenschaften](#)

Shortcuts

[Abschlussarbeiten](#)
[Kontakt](#)
[Verein AIK](#)
[Intern](#)

- ▀ Einstellungen
- ▀ Spezialseiten
- ▀ mit Formular bearbeiten
- ▀ Bearbeiten
- ▀ Versionen/Autoren
- ▀ Löschen
- ▀ Verschieben
- ▀ Aktualisieren

AIFB

Neuigkeiten

20. November 2009

[2. Trendkongress net economy in Karlsruhe](#)

[mehr...](#)



Veranstaltungen

18. November 2009

[Large-Scale Pattern-Based Information Extraction from the World Wide Web](#)

12. November 2009

[Eröffnungsveranstaltung KIT-Schwerpunkt COMMputation](#)

[mehr...](#)



Fakultät für
Wirtschaftswissenschaften

Shortcuts

[Abschlussarbeiten](#)
[Kontakt](#)
[Verein AIK](#)
[Intern](#)

▸ Home

▸ Lehre und Prüfung

▾ **Personen**

Professoren
 Organisation
**Wissenschaftl.
 Mitarbeiter**
 Honorarprofessoren
 Lehrbeauftragte
 Gäste und Externe
 Technik
 Auszubildende
 Hilfskräfte
 Ehemalige

▸ Forschungsgruppen

▸ Veröffentlichungen

Veranstaltungen

▸ Projekte

▸ Kooperationen

Stellenmarkt

Kontakt

Bearbeite Mitarbeiter: Denny Vrandecic

Vorname:

[Hilfe](#) (öffnet neue Seite)

Nachname:

[Hilfe](#) (öffnet neue Seite)

Akademischer Titel:

[Hilfe](#) (öffnet neue Seite)

Kürzel:

[Hilfe](#) (öffnet neue Seite)

Forschungsgruppe:

[Hilfe](#) (öffnet neue Seite)

Stellung:

[Hilfe](#) (öffnet neue Seite)

Ehemaliger:

[Hilfe](#) (öffnet neue Seite)

Telefon:

[Hilfe](#) (öffnet neue Seite)

Mobile:

[Hilfe](#) (öffnet neue Seite)

Fax:

[Hilfe](#) (öffnet neue Seite)

Email:

[Hilfe](#) (öffnet neue Seite)

Raum:

[Hilfe](#) (öffnet neue Seite)

- ▀ Einstellungen
- ▀ Spezialseiten
- ▀ Bearbeiten
- ▀ Versionen/Autoren
- ▀ Löschen
- ▀ Verschieben
- ▀ Aktualisieren

AIFB

Neuigkeiten

20. November 2009

[2. Trendkongress net economy in Karlsruhe](#)

[mehr...](#)



Veranstaltungen

18. November 2009

[Large-Scale Pattern-Based Information Extraction from the World Wide Web](#)

12. November 2009

[Eröffnungsveranstaltung KIT-Schwerpunkt COMMPutation](#)

[mehr...](#)



[Fakultät für Wirtschaftswissenschaften](#)

Shortcuts

[Abschlussarbeiten](#)
[Kontakt](#)
[Verein AIK](#)
[Intern](#)

▸ Home

▸ Lehre und Prüfung

▾ Personen

Professoren
 Organisation
 Wissenschaftl. Mitarbeiter
 Honorarprofessoren
 Lehrbeauftragte
 Gäste und Externe
 Technik
 Auszubildende
 Hilfskräfte
 Ehemalige

▸ Forschungsgruppen

▸ Veröffentlichungen

Veranstaltungen

▸ Projekte

▸ Kooperationen

Stellenmarkt

Kontakt

Personen

Professoren



Andreas Oberweis

Email: [oberweis\(at\)aifb.uni-karlsruhe.de](mailto:oberweis(at)aifb.uni-karlsruhe.de)

Tel.: +49 721 608 4516

Raum: [1A-19](#)



Hartmut Schmeck

Email: [hartmut.schmeck\(at\)kit.edu](mailto:hartmut.schmeck(at)kit.edu)

Tel.: +49 721 608 4242

Raum: [2C-16](#)



Detlef Seese

Email: [detlef.seese\(at\)kit.edu](mailto:detlef.seese(at)kit.edu)

Tel.: +49 721 608 6037

Raum: [2C-01](#)



Wolffried Stucky

Email: [stucky\(at\)aifb.uni-karlsruhe.de](mailto:stucky(at)aifb.uni-karlsruhe.de)

Tel.: +49 721 608 3812

Raum: [2A-01](#)



Rudi Studer

Email: [studer\(at\)aifb.uni-karlsruhe.de](mailto:studer(at)aifb.uni-karlsruhe.de)

Tel.: +49 721 608 3923

Raum: [216](#)



Stefan Tai

Email: [tai\(at\)aifb.uni-karlsruhe.de](mailto:tai(at)aifb.uni-karlsruhe.de)

Example: SNPedia



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

page discussion edit history

July 19-20 Washington, DC let the FDA know you want to protect your right to learn about your DNA

SNPedia SNPedia

SNPedia is a wiki investigating human genetics. We share information about the effects of variations in DNA, citing peer-reviewed scientific publications. It is used by [Promethease](#) to analyze and help explain your DNA.

Help! [\[edit\]](#)

- look at the example [rs1234](#)
- learn more about [SNPs](#)
- browse
 - [genes](#)
 - [genomes](#)
 - [genosets](#)
 - [medicines](#)
 - [medical conditions](#)

Popular [\[edit\]](#)

- [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

July 19-20 Washington, DC let the FDA know your views on access to direct to consumer dna testing

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](#) popuations. 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](#).

| is a | snp |
|-----------------------------|---------------------------------------|
| is | mentioned by |
| dbSNP | rs1234 |
| hapmap | rs1234 |
| hgdp | rs1234 |
| ensembl | rs1234 |
| gopubmed | rs1234 |
| scholar | rs1234 |
| google | rs1234 |
| pharmgkb | rs1234 |
| hgvsbaseg2p | rs1234 |
| medrefsnp | rs1234 |
| 23andMe | rs1234 |
| SNP Nexus | <input type="text" value="rs1234"/> |
| Chromosome | 3 |
| Orientation | plus |
| Position | 123615654 |
| Genotype | Effect |
| rs1234(A;A) | summary of something interesting |
| rs1234(A;C) | summary of something less interesting |
| rs1234(C;C) | normal |

| Genotypes | Magnitude | Summary |
|-----------------------------|-----------|---------------------------------------|
| Rs1234(A;A) | | summary of something interesting |
| Rs1234(A;C) | | summary of something less interesting |

navigation

- SNPedia
- FAQ
- Blog
- Recent changes
- Random page

search

toolbox

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link
- Browse properties

Choose the DNA test that's right for you.



Fill in your family tree.

Ancestry Edition, \$399 [Learn more](#)

[Buy Now](#)



Take charge of your health.

Health Edition, \$429 [Learn more](#)

[Buy Now](#)



Choose to have it all.

23andMe Complete, \$499

[Buy Now](#)

Find a disease or trait that we cover:

Select a Disease or Trait

Popular Topics:

- Type 2 Diabetes
- Rheumatoid Arthritis
- Psoriasis
- Breast Cancer
- Colorectal Cancer
- Prostate Cancer
- Celiac Disease
- Crohn's Disease
- Restless Legs Syndrome
- Age-related Macular Degeneration
- Parkinson's Disease
- Coumadin® / Warfarin Sensitivity

News and Press



23andMe Makes New Discoveries in Genetics Using Novel, Web-based, Participant-driven Methods

June 24, 2010



Sergey Brin's Search for a Parkinson's Cure

June 22, 2010



23andMe Enlists Informed Medical Decisions to Make Genetic Counseling

Scientific Resources and Principles

Physician Resources

- [Read our open letter to the medical community](#)

Our Science

- [Scientific Advisory Board](#)
- [Read our open letter to the scientific community](#)

Insurance, Privacy and Genetic Discrimination

- [Learn how the Genetic Non-Discrimination Act protects your genetic privacy](#)

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/gpg/GM21677.bed.gz

8594 genotypes annotated

Reference Population CEU

[Help](#) interpreting your re

[Chat](#)

[Show Everything](#)

These seem to be the most interesting snps.

[\(hide\)](#)

[gs1068](#) Magnitude: 5 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 Pasiegos in nearby Cantabria. ==See Also== *i3001813 *rs35788393 *i3001464 ==Links== *Ian Logan's mtDNA Website *Wikipedia mtDNA

[gs144](#) Magnitude: 4 Male.

[gs115](#) Magnitude: 3 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.

[gs140](#) Magnitude: 3 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.

[gs151](#) Magnitude: 3 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.

[gs153](#) Magnitude: 3 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-epileptics (such as diazepam, phenytoin, and phenobarbitone), *anti-depressants (such as amitriptyline and clomipramine) *Clopidogrel Plavix Prilosec Nexium Prevacid

[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]]

| Wiki page | Wikitext |
|----------------|----------------------------|
| Gdansk | [[population::455,830]] |
| Wikimania 2010 | [[starts at::July 9 2010]] |
| Markus | [[lives in::Oxford]] |

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



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.

Inline Queries



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

| ☒ | ☒ Population | ☒ Area |
|--------------------------|--------------|----------------------------------|
| Algeria | 32,854,000 | 2,381,740,000,000 m ² |
| Angola | 15,941,000 | 1,246,700,000,000 m ² |
| Benin | 8,439,000 | 112,620,000,000 m ² |
| Botswana | 1,765,000 | 600,370,000,000 m ² |
| Burkina Faso | 13,228,000 | 274,200,000,000 m ² |
| Burundi | 7,548,000 | 27,830,000,000 m ² |
| Cameroon | 16,322,000 | 475,440,000,000 m ² |
| Cape Verde | 507,000 | 4,033,000,000 m ² |
| Central African Republic | 4,038,000 | 622,984,000,000 m ² |
| Chad | 9,719,000 | 1,284,000,000,000 m ² |

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