Wikipedia is not the sum of all human knowledge: do we need a wiki for open data?

Finn Årup Nielsen

Lundbeck Foundation Center for Integrated Molecular Brain Imaging
at
Department of Informatics and Mathematical Modelling
Technical University of Denmark
and
Neurobiology Research Unit,
Copenhagen University Hospital Rigshospitalet

July 5, 2010

Abstract

"Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That's what we're doing" is an oft-repeated quote of Jimmy Wales. However, in cases of knowledge representation of neuroscience information I have found that the encyclopedia genre used by Wikipedia is not well suited, and it has been the reason why I have started a MediaWiki-based wiki: The Brede Wiki. There are two issues with science data on Wikipedia: The first is notability and the second issue that of data representation. In Wikipedia a scientific paper can rarely reach a sufficient level of notability to maintain its own page, and a large portion of scientific results can simply not be represented in Wikipedia. Numerical results from individual research papers are of interest to database for the scientific community, and therefore the Brede Wiki has a page for each scientific paper. On that page the bibliographic information as well as experiment data from the paper are stored. Templates provide a means for storing data in a structured way. Wikipedia use templates primarily for formating, so that field values might be obfuscated with formating code such as wikilinks and multi-level templates, making it difficult to do extraction and process the data further. In the Brede Wiki with its simple templates the extraction is complete, so all structured data is added to a SQL database. Semantic MediaWiki provides means for simple database operation within a wiki. A wiki with this extension enabled would also help to organize scientific open data better, and provide a platform for data sharing.

“The sum of all human knowledge. That’s what we’re doing.”

The problem

"I am considering adding articles for individual published scientific articles to Wikipedia."

All negative comments:

— "I would oppose efforts to write articles about the vast majority of scientific articles. While there are articles that are notable enough to warrant having a page on wikipedia, very few are. What you should instead do is use those sci. articles as sources to expand the pages of the relevant topic." — Gaëtan Landry/Headbomb

— "1000s of PR articles are released each year that sink without a trace (a couple of mine did that), you'd never satisfy basic notability criteria for most." — Cameron Scott

— "The only papers I can think of that have their own articles are the Alpher-Bethe-Gamow paper and An Exceptionally Simple Theory of Everything. In both cases there is independent third party coverage of the paper, which is pretty much essential for establishing notability". — Dragons flight

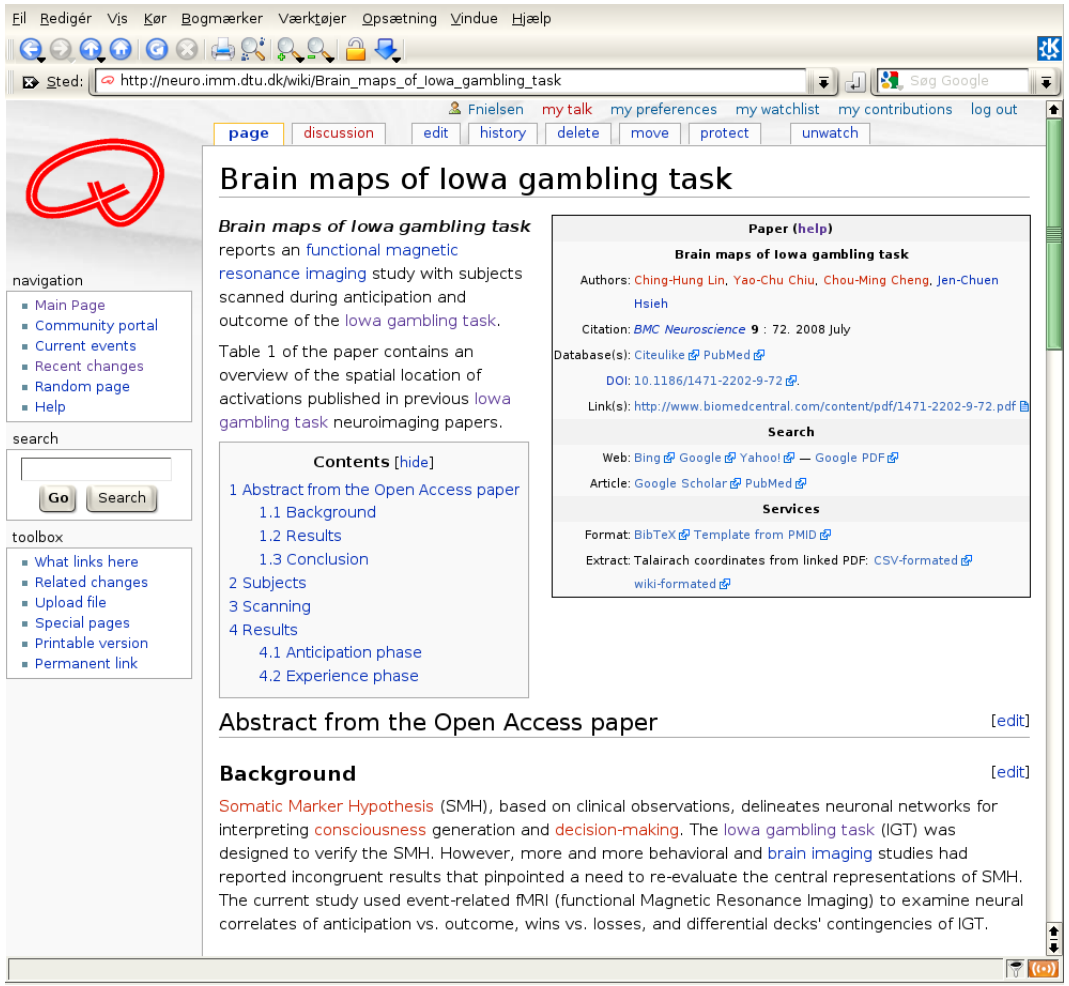
— "I agree that the above attempt to write an article on a scientific paper is taking the wrong approach" — Carcharoth

"My main problem would probably be the justification for these kinds of articles in terms of notability and encyclopedia-worthiness to potential deletionists."

— `Wikipedia:Village_pump_(proposals)/Archive_40&oldid=286707567`

Category:Journal_articles and subcategories have 57 pages.

My own wiki: Brede Wiki



One page for each scientific articles.

One page for each author (notable as well as unnotable)

Bibliographic information encoded in a MediaWiki template rendered as an infobox

If CC-by: Actual content included and wiki-linked

Description of details and comments, related papers, links to meta-analyses.

Simple templates

```
{{Paper
 | title = Brain maps of Iowa gambling task
 | author1 = Ching-Hung Lin | author2 = Yao-Chu Chiu
 | author3 = Chou-Ming Cheng  | author4 = Jen-Chuen Hsieh
 | journal = BMC Neuroscience
 | volume = 9  | pages = 72  | year = 2008  | month = July
 | doi = 10.1186/1471-2202-9-72
 | citeulike = 3046312  | pmid = 18655719
 | url1 = http://www.biomedcentral.com/content/pdf/1471-2202-9-72.pdf
}}
```

Simple use of templates: No nesting, no wiki-markup.

Template definition takes care of wiki-markup and creates wiki-link

Easy machine extraction of items

Other bibliographic projects and proposal

Possible features: Bibliographic information, quotations/notes management, article authoring system, two-way citation tracking.

Proposed projects on Meta-wiki: *WikiTextrose*, *Wikicite*, *Wikicat*

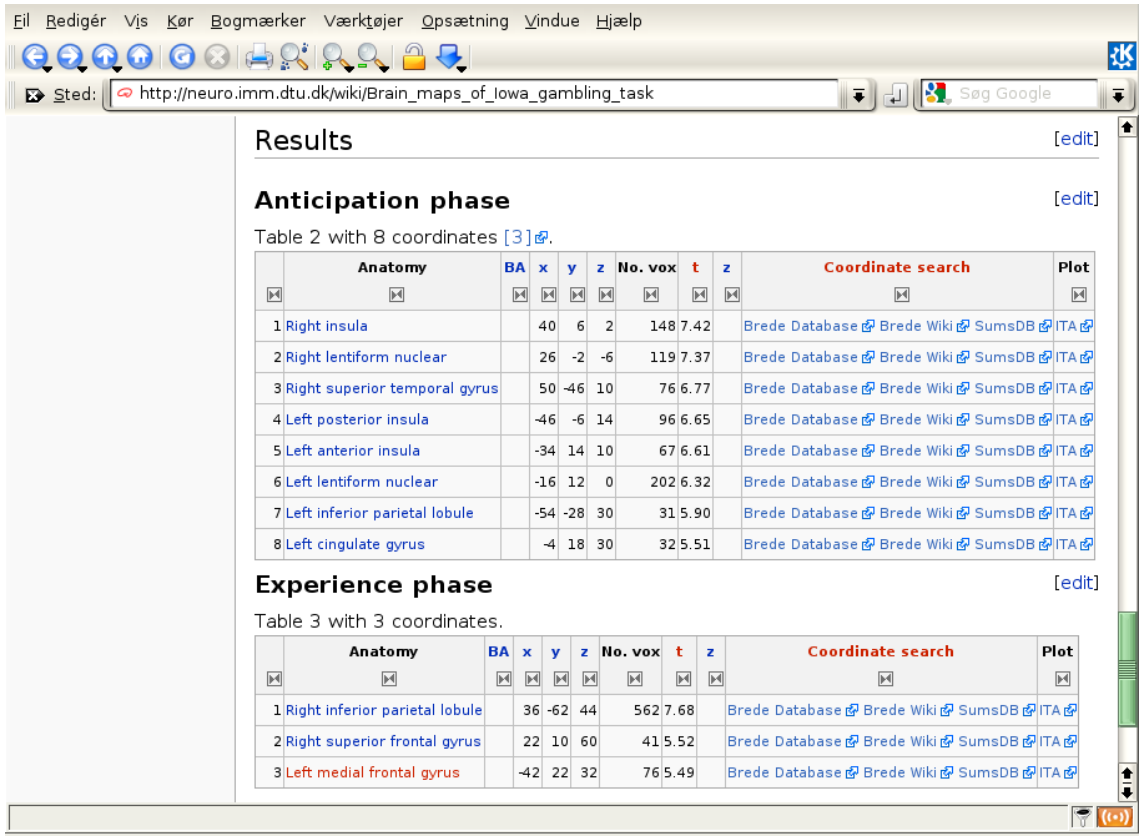
Strategy proposals: “BibTeX database and Bibliography namespace” and “Building a database of all books ever published”

“[. . .] a centralized wiki that contains citation information that other wikis can then reference using something like a {{cite}} template or a simple link.” — Brian Mingus on *wiki research list* 21 june 2010 about “WikiPapers”

Sunir Shah’s *Bibdex*, *AcaWiki* (“AcaWiki is like a ”Wikipedia for academic research””)

(Social reference management: *Mendeley*, *BibSonomy*, *Connotea*, . . .)

Numerical results in the Brede Wiki



Fil Redigér Vis Kør Bogmærker Værktøjer Opsætning Vindue Hjælp

Sted: http://neuro.imm.dtu.dk/wiki/Brain_maps_of_Iowa_gambling_task

Søg Google

Results [edit]

Anticipation phase [edit]

Table 2 with 8 coordinates [3].

	Anatomy	BA	x	y	z	No. vox	t	z	Coordinate search	Plot
1	Right insula		40	6	2	148	7.42		Brede Database Brede Wiki SumsDB	ITA
2	Right lentiform nuclear		26	-2	-6	119	7.37		Brede Database Brede Wiki SumsDB	ITA
3	Right superior temporal gyrus		50	-46	10	76	6.77		Brede Database Brede Wiki SumsDB	ITA
4	Left posterior insula		-46	-6	14	96	6.65		Brede Database Brede Wiki SumsDB	ITA
5	Left anterior insula		-34	14	10	67	6.61		Brede Database Brede Wiki SumsDB	ITA
6	Left lentiform nuclear		-16	12	0	202	6.32		Brede Database Brede Wiki SumsDB	ITA
7	Left inferior parietal lobule		-54	-28	30	31	5.90		Brede Database Brede Wiki SumsDB	ITA
8	Left cingulate gyrus		-4	18	30	32	5.51		Brede Database Brede Wiki SumsDB	ITA

Experience phase [edit]

Table 3 with 3 coordinates.

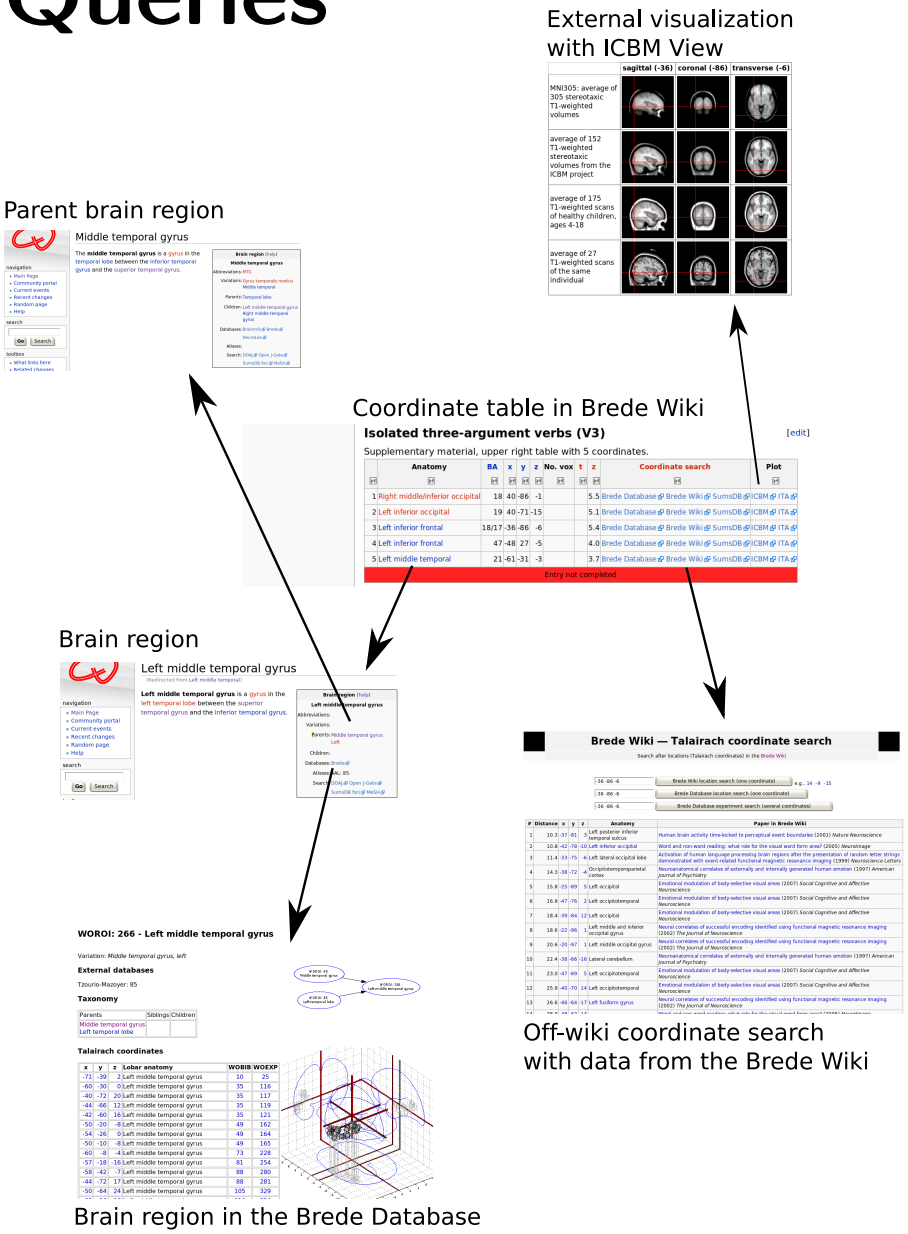
	Anatomy	BA	x	y	z	No. vox	t	z	Coordinate search	Plot
1	Right inferior parietal lobule		36	-62	44	562	7.68		Brede Database Brede Wiki SumsDB	ITA
2	Right superior frontal gyrus		22	10	60	41	5.52		Brede Database Brede Wiki SumsDB	ITA
3	Left medial frontal gyrus		-42	22	32	76	5.49		Brede Database Brede Wiki SumsDB	ITA

Numerical results from individual research papers are of interest to database

Values are stored in MediaWiki templates in the Brede Wiki

```
=== Anticipation phase ===
Table 2 with 8 coordinates
[http://www.biomedcentral.com/1471-2202/9/72/table/T2].
{{Talairach coordinates begin}}
{{Talairach coordinate
 | m = 1
 | n = 1
 | anatomy = Right insula
 | x = 40
 | y = 6
 | z = 2
 | number_of_voxels = 148
 | t_value = 7.42
 | p_value_voxel_corrected = 0.008
 | p_value_voxel_uncorrected = <0.0001
 | space = MNI
}}
{{Talairach coordinate
 | m = 1
 | n = 2
 | anatomy = Right lentiform nuclear
 | x = 26
 | y = -2
 | z = -6
 | number_of_voxels = 119
```

Queries



With numerical data from scientific articles: Meta-analysis, visualizations, specialized search.

Queries are possible, but not within the wiki

For Brede Wiki: Query on nearby brain coordinates with the SQLite file and an off-wiki script.

All data in templates of the wiki written to a database (Nielsen, 2009)

Other structured data projects and proposals

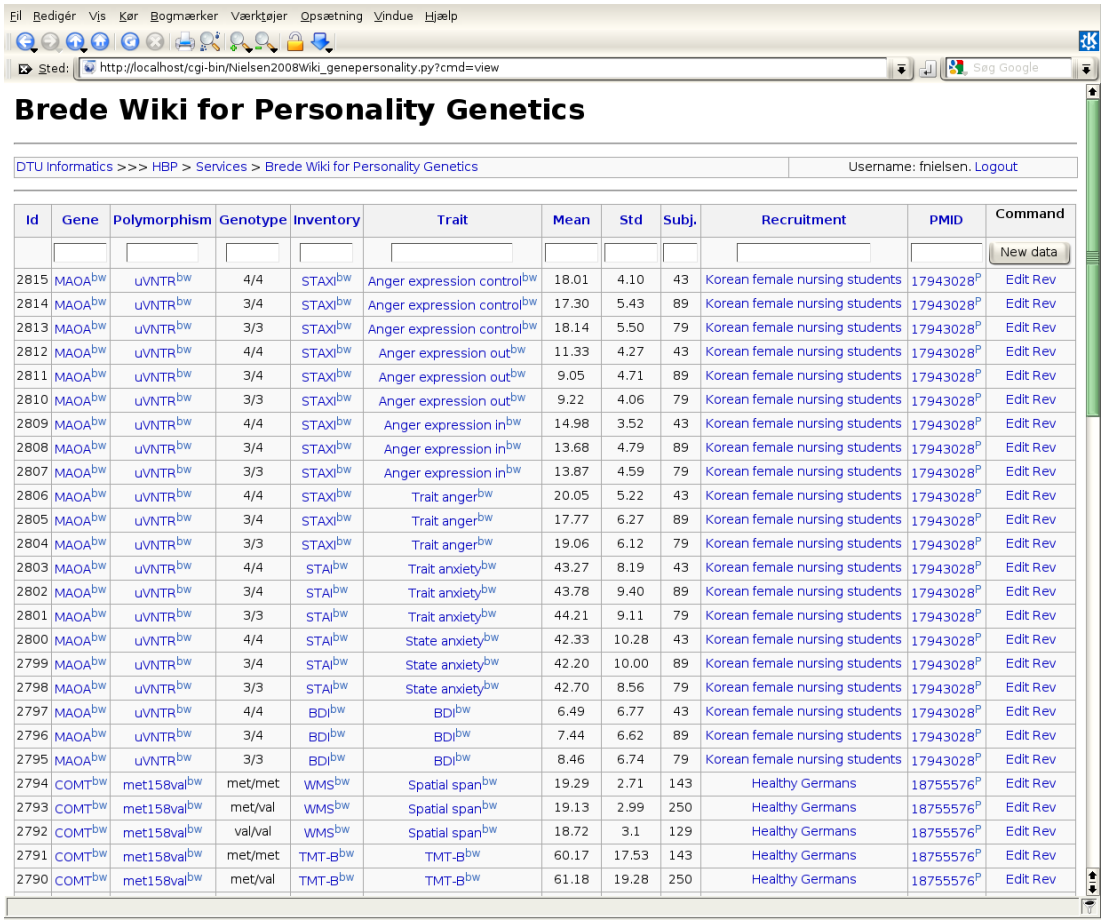
Strategy proposals: “Structured Data”, “Data.wikimedia.org” and “Category:Proposals for data-related features”. Category:Wikidata

Semantic MediaWiki, Freebase, OmegaWiki, Wiki-Data, . . .

Extraction from Wikipedia: DBpedia (Auer et al., 2008)

Wikis with programming (Anslow and Riehle, 2008)

Highly structured wiki



Fil Redigér Vis Kør Bogmærker Værktøjer Opsætning Vindue Hjælp

Sted: http://localhost/cgi-bin/Nielsen2008Wiki_genepersonality.py?cmd=view

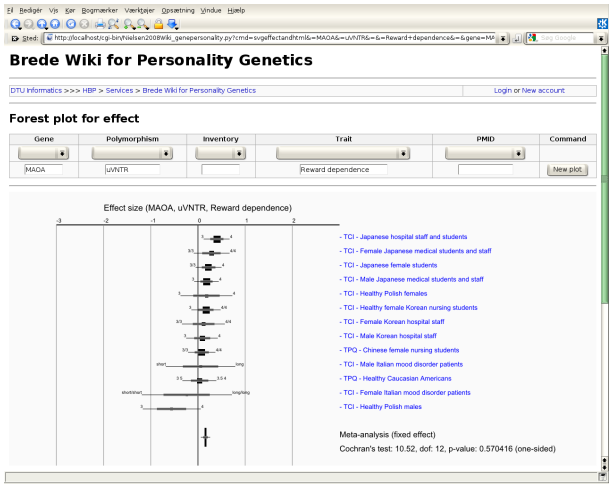
Brede Wiki for Personality Genetics

DTU Informatics >>> HBP > Services > Brede Wiki for Personality Genetics — Username: fnielsen. Logout

Id	Gene	Polymorphism	Genotype	Inventory	Trait	Mean	Std	Subj.	Recruitment	PMID	Command
											New data
2815	MAOA^{bw}	uVNTR^{bw}	4/4	STAXI^{bw}	Anger expression control^{bw}	18.01	4.10	43	Korean female nursing students	17943028^{P}	Edit Rev
2814	MAOA^{bw}	uVNTR^{bw}	3/4	STAXI^{bw}	Anger expression control^{bw}	17.30	5.43	89	Korean female nursing students	17943028^{P}	Edit Rev
2813	MAOA^{bw}	uVNTR^{bw}	3/3	STAXI^{bw}	Anger expression control^{bw}	18.14	5.50	79	Korean female nursing students	17943028^{P}	Edit Rev
2812	MAOA^{bw}	uVNTR^{bw}	4/4	STAXI^{bw}	Anger expression out^{bw}	11.33	4.27	43	Korean female nursing students	17943028^{P}	Edit Rev
2811	MAOA^{bw}	uVNTR^{bw}	3/4	STAXI^{bw}	Anger expression out^{bw}	9.05	4.71	89	Korean female nursing students	17943028^{P}	Edit Rev
2810	MAOA^{bw}	uVNTR^{bw}	3/3	STAXI^{bw}	Anger expression out^{bw}	9.22	4.06	79	Korean female nursing students	17943028^{P}	Edit Rev
2809	MAOA^{bw}	uVNTR^{bw}	4/4	STAXI^{bw}	Anger expression in^{bw}	14.98	3.52	43	Korean female nursing students	17943028^{P}	Edit Rev
2808	MAOA^{bw}	uVNTR^{bw}	3/4	STAXI^{bw}	Anger expression in^{bw}	13.68	4.79	89	Korean female nursing students	17943028^{P}	Edit Rev
2807	MAOA^{bw}	uVNTR^{bw}	3/3	STAXI^{bw}	Anger expression in^{bw}	13.87	4.59	79	Korean female nursing students	17943028^{P}	Edit Rev
2806	MAOA^{bw}	uVNTR^{bw}	4/4	STAXI^{bw}	Trait anger^{bw}	20.05	5.22	43	Korean female nursing students	17943028^{P}	Edit Rev
2805	MAOA^{bw}	uVNTR^{bw}	3/4	STAXI^{bw}	Trait anger^{bw}	17.77	6.27	89	Korean female nursing students	17943028^{P}	Edit Rev
2804	MAOA^{bw}	uVNTR^{bw}	3/3	STAXI^{bw}	Trait anger^{bw}	19.06	6.12	79	Korean female nursing students	17943028^{P}	Edit Rev
2803	MAOA^{bw}	uVNTR^{bw}	4/4	STAI^{bw}	Trait anxiety^{bw}	43.27	8.19	43	Korean female nursing students	17943028^{P}	Edit Rev
2802	MAOA^{bw}	uVNTR^{bw}	3/4	STAI^{bw}	Trait anxiety^{bw}	43.78	9.40	89	Korean female nursing students	17943028^{P}	Edit Rev
2801	MAOA^{bw}	uVNTR^{bw}	3/3	STAI^{bw}	Trait anxiety^{bw}	44.21	9.11	79	Korean female nursing students	17943028^{P}	Edit Rev
2800	MAOA^{bw}	uVNTR^{bw}	4/4	STAI^{bw}	State anxiety^{bw}	42.33	10.28	43	Korean female nursing students	17943028^{P}	Edit Rev
2799	MAOA^{bw}	uVNTR^{bw}	3/4	STAI^{bw}	State anxiety^{bw}	42.20	10.00	89	Korean female nursing students	17943028^{P}	Edit Rev
2798	MAOA^{bw}	uVNTR^{bw}	3/3	STAI^{bw}	State anxiety^{bw}	42.70	8.56	79	Korean female nursing students	17943028^{P}	Edit Rev
2797	MAOA^{bw}	uVNTR^{bw}	4/4	BDI^{bw}	BDI^{bw}	6.49	6.77	43	Korean female nursing students	17943028^{P}	Edit Rev
2796	MAOA^{bw}	uVNTR^{bw}	3/4	BDI^{bw}	BDI^{bw}	7.44	6.62	89	Korean female nursing students	17943028^{P}	Edit Rev
2795	MAOA^{bw}	uVNTR^{bw}	3/3	BDI^{bw}	BDI^{bw}	8.46	6.74	79	Korean female nursing students	17943028^{P}	Edit Rev
2794	COMT^{bw}	met158val^{bw}	met/met	WMS^{bw}	Spatial span^{bw}	19.29	2.71	143	Healthy Germans	18755576^{P}	Edit Rev
2793	COMT^{bw}	met158val^{bw}	met/val	WMS^{bw}	Spatial span^{bw}	19.13	2.99	250	Healthy Germans	18755576^{P}	Edit Rev
2792	COMT^{bw}	met158val^{bw}	val/val	WMS^{bw}	Spatial span^{bw}	18.72	3.1	129	Healthy Germans	18755576^{P}	Edit Rev
2791	COMT^{bw}	met158val^{bw}	met/met	TMT-B^{bw}	TMT-B^{bw}	60.17	17.53	143	Healthy Germans	18755576^{P}	Edit Rev
2790	COMT^{bw}	met158val^{bw}	met/val	TMT-B^{bw}	TMT-B^{bw}	61.18	19.28	250	Healthy Germans	18755576^{P}	Edit Rev

"Fielded wiki": One GUI and database field to each value.

Statistical computation of combined effects and visualization, export to Brede Wiki



Brede Wiki for Personality Genetics

Forest plot for effect

Effect size (MAOA, uVNTR, Reward dependence)

Meta-analysis (fixed effect)

Cochran's test: 10.52, dof: 12, p-value: 0.570416 (one-sided)

Summary

The encyclopedia genre is not suitable to store “all human knowledge”.

A wiki for articles (scientific as well as other) with the metadata and data is useful.

MediaWiki templates can be used to store data.

With simple templates all data can be extracted.

Wikis with open data in science is interesting.

Thanks!

References

Anslow, C. and Riehle, D. (2008). Toward end-user programming with wikis. ACM.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2008). DBpedia: A nucleus for a web of open data. In *The Semantic Web*, volume 4825 of *Lecture Notes in Computer Science*, pages 722–735, Heidelberg/Berlin. Springer. Description of a system that extracts information from the templates in Wikipedia, processes and presents them in various ways. Some of the methods and services they use are MySQL, Virtuoso, OpenCyc, GeoNames, Freebase, SPARQL and SNORQL. The system is available from http://DBpedia.org.

Nielsen, F. Å. (2009). Brede Wiki: Neuroscience data structured in a wiki. In Lange, C., Schaffert, S., Skaf-Molli, H., and Völkel, M., editors, *Proceedings of the Fourth Workshop on Semantic Wikis — The Semantic Wiki Web*, volume 464 of *CEUR Workshop Proceedings*, pages 129–133, Aachen, Germany. RWTH Aachen University. http://ceur-ws.org/Vol-464/paper-09.pdf.