Semantic Wikipedia – Checking the Premises[1]

Rainer Hammwöhner

Institut für Medien-, Informations- und Kulturwissenschaft
Universität Regensburg
Universitätsstraße
93040 Regensburg
rainer.hammwoehner@sprachlit.uni-regensburg.de

Abstract: Enhancing Wikipedia by means of semantic representations seems to be a promising issue. From a formal or technical point of view there are no major obstacles in the way. Nevertheless, a close look at Wikipedia, its structure and contents reveals that some questions have to be answered in advance. This paper will deal with these questions and present some first results based on empirical findings.

1 Introduction

Up to now Wikipedia has accumulated an enormous wealth of information by the effort of an open community of volunteers. This information however is semi-structured at best and therefore imposes restrictions on automatic processing. Automatic processing of Wikipedia contents is desirable for a couple of reasons, e.g.:

- Enhanced information services can improve the utility of Wikipedia itself. Implicit knowledge scattered over separated parts of the corpus can be brought together and made explicit.
- Consistency of the corpus can be enforced by autonomous agents operating on semantic representations.
- Information extracted from Wikipedia can be used in other contexts.

There are several approaches to this task, but two very general types may be distinguished:

---

[1] A short version of this paper is included in: Sören Auer, Christian Bizer, Claudia Müller, Anna V. Zhdanova (Eds.): The Social Semantic Web 2007, Proceedings of the 1st Conference on Social Semantic Web (CSSW), September 26-28, 2007, Leipzig, Germany. LNI 113 GI 2007, pp. 173-178.

- Information is extracted from Wikipedia by interpretation of existing explicitly defined structures [AL07]. The main sources of information are templates embedded within Wikipedia's articles. The resulting knowledge is represented in terms of a formal language and may be subject to viewing and querying via the OntoWiki software [ADR06].

- The syntax of the mark-up language of the MediaWiki software is enhanced in order to allow for link typing and attribute assignment [Vö06]. A process of information extraction and representation will again lead to formal representations that may be employed by inference processes. A necessary prerequisite of this approach is an extension to the MediaWiki software that is the technical core of Wikipedia [KVV06].

According to [Vö06] the following key elements are necessary to achieve the intended semantic annotation of Wikipedia's articles:

- *categories* classify articles according to their content,

- *types* express the meaning of links connecting Wikipedia's articles and

- *attributes* capture atomic properties related to the contents of an article.

Categories are the only of these devices being already in use and ready for evaluation. Thus the notion of categorizing Wikipedia's articles will play a crucial role within the theoretical and practical considerations of this paper.

2 The Premises

Introducing at least one of the approaches mentioned above will be of major consequence to the users of Wikipedia. New information services will be available on the one hand and the authoring process will be more demanding on the other hand. The latter applies at least for the second approach. The success of this project is bound to some central premises that should be made explicit and checked before the effort of large scale implementation is to be taken.

P1 Technical feasibility: Prototypes for both of the approaches have been implemented.

P2 Formal soundness: The proposed semantic representations are based on rigidly defined structures. However, there is some lack of clarity about the further use of typed links. As far as no terminological reasoning is intended, no problems should arise.

P3 Reliability of results: Recent studies have attested Wikipedia's convenient average quality [Ha07a, Ha07b, Wi07]. However, Wikipedia articles of abysmal quality can be found easily. The user of Wikipedia needs the competence to distinguish reliable from erroneous information. Semantic operations on Wikipedia should not accumulate errors and must not blur the user's view by hiding the sources of errors. It is not quite clear, whether this criterion is met by the proposed approaches.

P4 Reliability of the authoring process: The first approach does not impose additional tasks on the author. No new problems should arise here. The second approach relies heavily on the proper assignment of link types and categories by the user. The author can decide which and how many link types or categories to use. He can select from predefined denominators or enter new link types and categories at his will. Obviously, problems can arise out of the inconsistent and ambiguous use of type and category identifiers. [Vö06, section 4.1] infer from the seemingly unproblematic use of the category system that a consistent use of a link type system is to be expected too. This conclusion is problematic simply, because there is no empirical evidence of a proper use of the category system at all. It is the major objective of this paper to present some observations which are relevant to this issue.

P5 Multi-lingual system: Approaches to realizing a Semantic Wikipedia should consider that Wikipedia is a multi-lingual information base. At least an interlingual mapping mechanism for link types and attributes corresponding to interlingual category mapping should be developed.

P7 Usability: All efforts in enhancing Wikipedia by innovative information services will be futile unless they are integrated within an environment devoted to strict usability criteria. This applies for the authors and information seekers as well.

The list introduced above may not be complete. But the relevance of the mentioned premises does not seem to be questionable. ***P4*** occupies a key position since a fundamental question is involved here. Usable interfaces may be revamped, formal systems can be redesigned, but the competence of a large user community can be adjusted only in the long run. Thus ***P4*** may be the decisive criterion in the choice between more or less demanding approaches to a Semantic Wikipedia.

3 Some observations on Wikipedia's category system

The category system of Wikipedia is intended to provide an additional navigation structure on the set of articles [Wi07a]. It is not used as a device of query support. The proper assignment of categories is defined by a set of rules of thumb [Wi07a]. [Vo06] provides a comprehensive overview on structure and use of Wikipedia's category system.

In the following we will present some results from an explorative study focused on two questions:

1. Is Wikipedia's category system a thesaurus?
2. What kind of quality issues can be observed concerning category assignment?

Both questions are discussed with respect to the demands of semantic interpretation.

3.1 Tool and data source

The data which are presented in the following are derived from the following samples:

- Random samples of ~1000 texts each from the English, German, French, and Italian parts of Wikipedia.
- Complete downloads of all excellent articles written in the languages mentioned above.
- Samples from the category systems of the English, German, French, and Italian Wikipedia.

The data extraction and evaluation was performed by a tool developed by the author. It is capable of:

- Downloading samples of Wikipedia articles which are selected from previously defined lists (e.g. excellent articles), at random using Wikipedias random function, by random walk starting from a pre-given seed, or by crawling the Wikipedia web.
- Evaluating the accessed Wikipedia articles with respect to text length (no. of words), link density, no. of versions and authors, categories etc. These data may similarly be obtained for the talk and user pages as well.
- Processing data from the English, German, French, Polish, Dutch, Italian, Spanish, Portuguese, Swedish and Danish part of Wikipedia. Adding an additional language implies approximately 1 hour of work – notwithstanding that some evaluations like word counting can be used with alphabetic writing systems only.
- Data may be exported for statistical interpretation (csv-format).

3.2 Is Wikipedia's category system a thesaurus?

This question was firstly brought up by [Vo06]. This paper compares Wikipedia's category system to thesauri (MeSH: Medical Subject Headings), hierarchical classifications (Dewey Decimal Classification) and folksonomies (del.icio.us). [Vo06] comes to the conclusion, that Wikipedia's category system is a thesaurus, since the requirements of ISO 2788 are met:

- The equivalence relation connecting synonymous terms may be represented using redirects.

- The hierarchical relation between broader and narrower terms is expressed by the category ⇒ subcategory relation.

- Associations between related terms are represented by hyperlinks.

Obviously the mark-up language of Wikipedia is capable of expressing thesaurus structures. The question, however, is, whether the existing category systems *are* thesauri. [Vo06] further elaborates his conclusions by comparing excerpts from the MeSH thesaurus and from the English Wikipedia. The presented structures are reasonably similar. But counter examples may be found easily at least within the English Wikipedia:

> *categories ⇒ fundamental ⇒ thought ⇒* ***knowledge*** *⇒ academia ⇒ academic institutions ⇒ school counseling ⇒ personal development ⇒ personal finance ⇒ microeconomics ⇒ information, knowledge and uncertainty ⇒ information ⇒* ***knowledge*** *⇒ nature ⇒ life ⇒ death ⇒ extinction ⇒ fossils ⇒ dinosaurs*[2]

This illustrative example demonstrates:

- The existence of cycles (*knowledge*) within the category ⇒ subcategory relation is conform to Wikipedia's rule set [Wi07a], but not to ISO 2788 since the resulting structure is no hierarchy.

- The category ⇒ subcategory relation does not lead generally from broader to narrower terms, but in many cases to related terms.

Thus, the category ⇒ subcategory relation may not be considered as a transitive relation representing terminological subordination. As a consequence there is no support of terminological reasoning by the English category system. Even retrieval support, e.g. by spreading activation, may lead to unwanted results, if the terminology is as weakly structured as the example suggests. The same criticism is valid for the French Wikipedia as well. The category systems of the Italian and German Wikipedia are quite different in structure. They contain a few cycles only, their hierarchy has a considerably lower depth (s. table 1). This applies to the maximal descriptor level (first value) and the longest observed path within the hierarchy (value in brackets) as well. A substantial difference between both of the depth values indicates a lack of balance within the category system.

[2] Observed: 6.07.2007

The data presented above are derived from the following samples: two bilingual samples of de-en (size 152) and de-it (size 169) were chosen at random using interlingua links. A sample of 134 French articles was added to the latter one, once more using interlingua links. The basic categories describing these articles were sampled as well as all of their superordinate categories. It can be seen, that sample size has some influence on the number of basic categories, less influence on the total number of categories and no impact on the depth of hierarchy and number of cycles. It can be assumed, that deep category systems are error prone. Authors will have difficulties to get an overview on the overall structure since the number of paths to the top category shows exponential growths behaviour.

	articles	basic categories	all categories	depth	superord. per cat. (median)	cycles
de (en)	152	366	1740	10 (15)	2	4
de (fr,it)	169	394	1816	10 (15)	2	4
en	152	581	6274	14 (156)	2	493
it	167	321	1091	12 (15)	2	7
fr	134	360	3116	14 (83)	2	424

Table 1: Basic features of category systems

An additional example will illustrate the pitfalls of big category hierarchies in Wikipedia. It shows the longest path within the category ⇒ subcategory multi-hierarchy as found in the sample of the English Wikipedia:

digital revolution ⇒ cryptography ⇒ application of cryptography ⇒ authentication methods ⇒ personal identification ⇒ biometrics ⇒ physical anthropology ⇒ human evolution ⇒ evolutionary psychology ⇒ memetics → anticipatory thinking ⇒ strategic management ⇒ product management ⇒ product development ⇒ design ⇒ built environment ⇒ architecture ⇒ architecture and engineering occupations ⇒ building engineering ⇒ building materials ⇒ metals ⇒ alloys ⇒ copper alloys ⇒ bronze ⇒ bronze age ⇒ ancient near east → ancient near eastern religions → ancient semitic religions ⇒ Abrahamic religions ⇒ Judaism → messianism ⇒ Jesus ⇒ doctrines and teachings of Jesus → nonviolence → peace ⇒ peace churches ⇒ anabaptism ⇒ amish ⇒ simple living ⇒ environmentalism ⇒ environmental ethics ⇒ extinction ⇒ extinct species ⇒ extinct animals ⇒ prehistoric animals ⇒ mesozoic animals ⇒ cynodonts ⇒ mammals⇒ primates ⇒ apes ⇒ humans ⇒ anthropology ⇒ prehistory ⇒ archaeology ⇒ periods and stages in archaeology ⇒ ancient history ⇒ ancient mysteries ⇒ astrology → ~~astrological factors~~ → classical elements → earth⇒ earth sciences ⇒ environmental science ⇒ environment ⇒ urban studies and planning ⇒ transportation ⇒ travel ⇒ tourism ⇒ cultural heritage ⇒ cultural history ⇒ cultural movements ⇒ art genres ⇒ graphic design ⇒ printing ⇒ books → fiction ⇒ fictional ⇒ fictional abilities ⇒ superhuman powers ⇒ psychic powers → prediction ⇒ futurology ⇒ population ⇒ demography ⇒ ethnicity ⇒ ethnicity in politics ⇒ anti-national sentiment ⇒ prejudices ⇒ bias ⇒ appearance ⇒ aesthetics ⇒ arts ⇒ visual arts → communication design ⇒ mass media ⇒ media by format ⇒ digital media ⇒ software ⇒ software engineering ⇒ software testing ⇒ formal methods ⇒ semantics ⇒ lexical semantics ⇒ vocabulary ⇒ terminology ⇒ philosophical terminology ⇒ ontology → reality ⇒ alternate reality ⇒ mental health ⇒ psychology ⇒ branches of psychology ⇒ social psychology ⇒ personal development ⇒ personal finance ⇒ microeconomics ⇒ household behavior and family economics ⇒ consumer theory ⇒ goods ⇒ manufactured goods ⇒ computer hardware ⇒ computer storage ⇒ computer data ⇒ data management ⇒ data collection ⇒ scientific observation ⇒ measurement ⇒ probability and statistics ⇒ statistics ⇒ statistical mechanics ⇒ specific models ⇒ economics models → economic systems ⇒

socialism ⇒ labor ⇒ social programs ⇒ healthcare ⇒ health promotion ⇒ determinants of health ⇒ health effectors ⇒ prevention ⇒ security ⇒ national security ⇒ public safety ⇒ emergency management ⇒ disasters ⇒ economic disasters ⇒ economic problems ⇒ social inequality ⇒ socioeconomics ⇒ development ⇒ economic ⇒ devolopment ⇒ poverty

This example was extracted from the English Wikipedia at 15th of June and verified at the 7th of August 2007. In the meantime one category and 10 category ⇒ subcategory relations have been deleted (→). Some of these deletions lead to a simplification of the overall structure: some others were caused by the insertion of additional hierarchy levels. It is an open question, which effects will result from the volatility of the category system as observed in this example.

These findings, however, have to be confirmed using bigger samples or the complete data set. It would be desirable to develop diagnostic tools which could identify problematic category inclusions. One promising approach is the comparison of category systems from various Wikipedias. If a category ⇒ subcategory inclusion is present in more than one Wikipedia, it is likely to be valid. If a category ⇒ subcategory pair occurs in one Wikipedia only, it can be invalid or culture specific as well.

3.3 Some data on category assignments

Tables 2 and 3 show results of a basic quantitative evaluation of the sample. Table 2 presents the number of categories per Wikipedia article. It can be seen easily that featured articles get up to twice as much categories than non-featured ones and that the English version of Wikipedia exceeds the other ones by far in the use of categories.

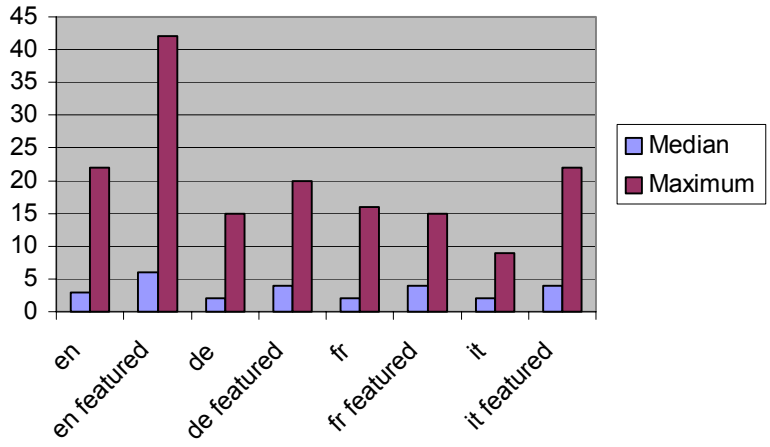

Table 2 No. of categories assigned to an article

Table 3 depicts the number of categories that are used to categorize 1,000 articles. Once more the English version of Wikipedia presents higher values. Seemingly language specific styles of categorization have developed. What has been noticed in the context of one language cannot be transferred easily to another part of Wikipedia. What may be surprising at the first glance is the lack of data reduction achieved by categorization. There are lots and lots of very sparsely populated categories. This is due to the high variety of purposes pursued by categorization of Wikipedia articles.

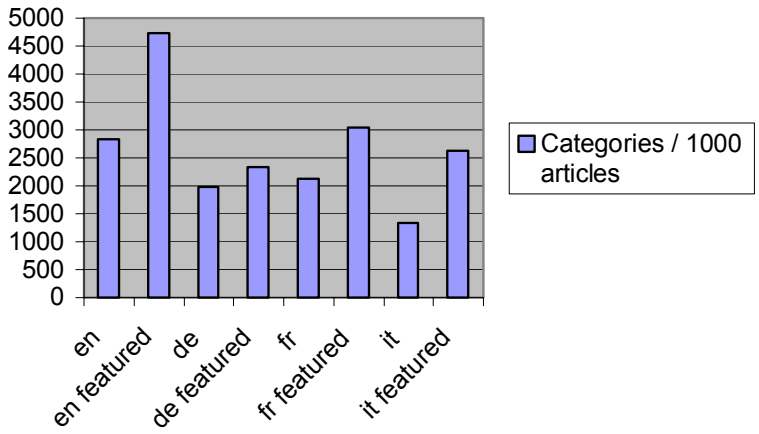

Table 2 Number of different categories per 1000 articles

3.4 Types of categories

The following, probably not exhaustive list of category types can be derived from every medium size sample of Wikipedia articles. According to the type of category articles are classified according to:

1. **Topic**: Lemmas, which cover a special – perhaps scientific – topic, are classified to a broader thematic field – e.g. *wildlife* is covered by categories like *biology* or *ecology*. This kind of categories is most similar to text-descriptors as employed in information retrieval.

2. **Feature of entity**: Lemmas, which describe objects are classified according to specific features of these entities. Which features are chosen is up to the decision of the author or fixed in conventional rules. The description of persons by categories, for instance, is described by a couple of Wikipedia articles, e.g. [Wi07d,e] – e.g. *1749 births*, *1832 deaths*. Similar rules apply to geographic entities – city, counties, states etc.

3. **Media of article**: If special media are employed within an article, this article is assigned to the appropriate category, e.g. *spoken article*.

4. **Quality of article**: Special categories are devoted to the quality of articles. A few categories group articles with high quality – e.g. *featured article* – in most cases specific quality problems are disclosed – *all articles lacking sources*, *articles which may contain original research*.

5. **Protection status of article**: Some articles may not be edited at all or not by anonymous authors. Categories like *protected* or *semi-protected* indicate this status.

3.5 Distribution of basic categories in the corpus

A simple approach to get an impression of the use of categories is extracting the topmost used categories from the corpus. Not surprisingly no categories of type 2 will show up. They are too specific to appear more than once or twice even in a large corpus – e.g. *1872 births* or *Visitor attractions in Hampshire*. Type 3 categories show the same distribution in all samples. Major differences exist in the use of type 5 and especially type 4 categories. 16 of the 20 most used categories from the English sample are of type 4. The featured articles still show 7 of 20 categories from type 4 – among them 15 uses of *factual verification needed*. This casts some shadow on the quality process of Wikipedia. Either the categories are no longer appropriate or, perhaps, the articles should not be considered as excellent. The Italian Wikipedia makes heavy use of type 4 categories but for a different goal. In this case all articles that are considered as excellent in any other language of Wikipedia are grouped in special categories. In the French and German part of Wikipedia type 4 categories are of little or no importance.

Some rather subtle differences show up when the use of categories of type 1 and 2 is looked at (see table 3). Obviously the German categories tend to be more general (*Mann*, *Frau*) than the English ones (*Knights of the garter*).

en featured (1364)	de featured (1044)	fr featured (350)	it featured (299)
Living people	Mann	Histoire économique	Biografie
Atlantic hurricanes	Deutscher	Macédoine antique	BioBot
American film actors	US-Amerikaner	Site archéologique de Grèce	Comuni italiani
Knights of the garter	Frau	Peintre	Gruppi musicali statunitensi
Grammy award winners	Autor	Histoire du français	
Pacific Ocean theater of World War II	Literatur (Deutsch)	Esclavage	
Video games developed in Japan	Literatur (20. Jahrhundert)	Patrimoine du XIXe siècle	
American films	Geschichte von Frankfurt am Main	Mammifère (nom vernaculaire)	
Battles involving the United States	Fleischfressende Pflanze	Terme japonais	
	Millionenstadt	Économie des membres de l'OMC	
	Berliner Geschichte	Histoire	
	Gotisches Bauwerk		

Pretenders to the throne of the kingdom of France (Plantagenet) Battles involving Japan	Tiere NASA 1944 Brite Einzelsprache Literarisches Werk	contemporaine de la Grèce Histoire du monde indien Index égyptologique Bombardement	

Table 3 Categories of type 1 and 2 from the 20 topmost used categories from the sample of featured articles

3.6 Problematic or erroneous use of categories

Major problems in the automatic processing of category structures may arise from the erroneous assignment of categories. From the data some types of problems or errors may be derived.

1. **Wrong assignment:** The lemma does not belong to that category. *Definite clause grammars*, for instance, are neither formal languages nor does the respective article [Wi07g] deal with formal languages. Thus, this article should not belong to a category named *formal languages*.

2. **Unmotivated assignment:** The assignment to a category implies a proposition which is not mentioned in the article and may even be controversial. A former version of the article on *William Shakespeare* [Wi07h] was assigned to the category *influences on Sigmund Freud* without mentioning *Freud* at all.

3. **Broad Categories:** Very broad categories – like *Frau* and *Mann* in German Wikipedia – accumulate thousands of articles. Thus, they are not useful for browsing.

4. **Narrow Categories:** Very narrow categories – like *Ancient Greek slaves* – hold only few or no entries. Their use for browsing purposes is limited, too.

5. **Missing Category:** English Wikipedia has taken a deep interest in death causes. If a person is not categorized according to its death cause, should this article be considered as incomplete?

6. **Inconsistent category depth:** Some phenomena are categorized up to a very great detail, others are not. *James Dean* [Wi07k] can be found in the categories *Entertainers who died in a road accident* and *California road accident deaths*. But there is no category like *Death by Spanish Flu*, which would group celebrities like the Austrian painter *Egon Schiele* [Wi07j] and the former *First Lady Rose Cleveland* [Wi07i].

The impact of these errors to further information processing is quite different. Wrong or unmotivated category assignment leads to bad retrieval results and, perhaps, to erroneous inferences. Broad or narrow categories are of little use, when browsing the category system. Inconsistent category assignment leads to unpredictable system behaviour with respect to information retrieval and inferencing.

3.6 Discussion

The phenomena described above were observed in the context of an empirical study not finished yet. It seems to be clear however, that language specific differences in category use exist. Some few cases of categorization could be identified as obviously erroneous. In most cases the question is not, whether a category assignment is correct or not, but whether it is appropriate. This can be answered only in the context of a given context. There is however no such privileged point of view. Wikipedia as an encyclopaedia on the one hand must be open to literarily all questions. On the other hand there was – at least as far German Wikipedia is concerned – no broad consensus from the very beginning, how to implement and to use the category system [Wi07f]. There may be an intuition in the public, what an encyclopaedia is about and how it should work. But there is no such intuition about proper knowledge organization. As a consequence an enormous amount of categories has been defined by a huge effort of the community. A complete overview over that structure is not feasible any more. A recently performed study [Ha07b] did not indicate any positive effect of the category system on Wikipedia's usability. This study however was comparatively small. It should be enhanced by further studies with significant user participation.

4 What does this mean to Semantic Wikipedia

This study has illustrated that Wikipedia's category system is not obviously a sound base for the development of a more demanding semantic system. The proliferation of the category system indicates what may happen to a link type system that may freely be extended by the user. This aspect is of crucial importance since evaluation of link typing had controversial results even in more controlled settings [Ma91]. As a consequence more empirical studies on category assignment are needed in order to understand the unfolding of the rather different category systems within the German and Italian Wikipedia on one side and the French and English Wikipedia on the other. Various settings – for instance with open and closed link type systems – should be considered before modifications at the existing encyclopaedia are brought into effect.

Approaches to Semantic Wikipedia – or other Web 2.0 applications – should be presented and planned on use cases derived from real world scenarios. This requires user participation and empirical studies. The introduction of these new and demanding instruments should perhaps not be done the Wiki-way only – publish a tool and look, what happens. The data corpus at least of Wikipedia is too big to play around with it.

Nevertheless, the introduction of more semantic features into Wikipedia has lots of promising aspects, too. The category system can be relieved from alien tasks like fact representation. The problem of redundant assignment of categories and subcategories [Wi07b] to Wikipedia articles can be solved by simple inference processes in combination with appropriate presentation tools. These are just examples of the positive effects that can be achieved by Web 2.0 techniques. Furthermore, the technical soundness and good performance of the existing prototypes promises that experiments may be carried out with reasonable effort.

The task is promising and demanding. It will require additional effort in the field of formal representations – to allow quality estimations – innovation of user interfaces and extensive user studies.

References

[AL07] Auer, S.; Lehmann, J.: What have Innsbruck and Leipzig in common? Extracting Semantics from Wiki Content. accepted at ESWC 2007. http://www.informatik.uni-leipzig.de/~auer/publication/ExtractingSemantics.pdf, cited 20.05.2007

[ADR06] Auer, S.; Dietzold, S.; Riechert, T.: OntoWiki - A Tool for Social, Semantic Collaboration. In I. Cruz et al. (Eds.): Proceedings of 5th International Semantic Web Conference, Nov 5th-9th, Athens, GA, USA, LNCS 4273, pp. 736-749, 2006. Springer-Verlag Berlin Heidelberg 2006. http://www.informatik.uni-leipzig.de/~auer/publication/ontowiki.pdf, cited 20.05.2007

[Ha07a] Hammwöhner, R. et.al.: Qualität der Wikipedia. Eine vergleichende Studie. In Oßwald, A.; Stempfhuber, M; Wolff, C. (eds.) Open Innovation. Proc. 10th Int. Symposium on Information Science in Cologne. UVK, 2007, pp. 77-90.

[Ha07b] Hammwöhner, R.: Qualitätsaspekte der Wikipedia. In: Stegbauer, C.; Schmidt, J.; Schönberger, K. (eds): Wikis: Diskurse, Theorien und Anwendungen, Sonderausgabe von kommunikation @ gesellschaft, Jg. 8, 2007, Online-Publication: http://www.soz.uni-frankfurt.de/K.G/B3_2007_Hammwoehner.pdf

[ISO86] ISO 2788: 1986: Guidelines for the establishment and development of monolingual thesauri.

[KVV06] Krötzsch, M.; Vrandecic, D.; Völkel, M.: Semantic Mediawiki. In Proc. 5th Int. Semantic Web Conf. (ISWC06). http://korrekt.org/papers/KroetzschVrandecicVoelkel_ISWC2006.pdf, cited 20.05.2007

[Ma91] Marshall, C.C. et.al.: Aquanet: a hypertext tool to hold your knowledge in place. In *Proc. Hypertext' 91, San Antonio,* S. 261-275, New York, 1991. ACM.

[Vö06] Völkel, M. et.al.: Semantic Wikipedia. In Proc. 15th Int. Conf. on World Wide Web, WWW 2006, Edinburgh, Scotland, May 23-26, 2006. http://www.aifb.uni-karlsruhe.de/WBS/hha/papers/SemanticWikipedia.pdf, cited 20.05.2007

[Vo06] Voss, J.: Collaborative thesaurus tagging the Wikipedia way. (v2; 2006-04-27; http://arxiv.org/abs/cs.IR/0604036) – [[Wikimetrics]] research papers, volume 1, issue 1 (cited 0.6.07.2007).

[Wi07] Wiegand, D.: Entdeckungsreise, Digitale Enzyklopädien erklären die Welt, c't, Magazin für Computer und Technik, Nr. 6, 2007, S. 136-145.

[Wi07a] Wikipedia:Categorization. In Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Wikipedia:Category, cited 27.05.2007.

[Wi07b] Wikipedia: Categorization and subcategories. In Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Wikipedia:Categorization_and_subcategories, cited 27.05.2007.

[Wi07c] Wikipedia:Categories, lists, and series boxes. In Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Wikipedia:Categories%2C_lists%2C_and_series_boxes, cited 27.05.2007.

[Wi07d] Wikipedia: Categorization of people. In Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Wikipedia:Categorization_of_people, cited 27.05.2007.

[Wi07e] Wikipedia: Categorization/Gender, race and sexuality. In Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Wikipedia:Categorization/Gender%2C_race_and_sexuality, cited 27.05.2007.

[Wi07f] Wikipedia Diskussion: Kategorien. In Wikipedia, The Free Encyclopedia. http://de.wikipedia.org/w/index.php?title=Wikipedia_Diskussion:Kategorien/Archiv1, http://de.wikipedia.org/w/index.php?title=Wikipedia_Diskussion:Kategorien/Archiv2, http://de.wikipedia.org/w/index.php?title=Wikipedia_Diskussion:Kategorien/Archiv3, cited 27.05.2007

[Wi07g] Definite clause grammar. In Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Definite_clause_grammar, cited 28.05.2007.

[Wi07h] William Shakespeare. In Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/w/index.php?title=William_Shakespeare&oldid=133058761, cited 28.05.2007.

[Wi07i] Rose Cleveland. In Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Rose_Cleveland, cited 28.05.2007.

[Wi07j] Egon Schiele. In Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/Egon_Schiele, cited 28.05.2007.

[Wi07k] James Dean. In Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/wiki/James_Dean, cited 28.05.2007.