Measuring Wikipedia

Jakob Voß

jakob.voss@nichtich.de
Humboldt-University of Berlin, Institute for library science

Abstract
Wikipedia, an international project that uses Wiki software to collaboratively create an encyclopaedia, is becoming more and more popular. Everyone can directly edit articles and every edit is recorded. The version history of all articles is freely available and allows a multitude of examinations. This paper gives an overview on Wikipedia research. Wikipedia's fundamental components, i.e. articles, authors, edits, and links, as well as content and quality are analysed. Possibilities of research are explored including examples and first results. Several characteristics that are found in Wikipedia, such as exponential growth and scale-free networks are already known in other context. However the Wiki architecture also possesses some intrinsic specialities. General trends are measured that are typical for all Wikipedias but vary between languages in detail.

Introduction

Wikipedia is an international online project which attempts to create a free encyclopaedia in multiple languages. Using Wiki software, thousand of volunteers have collaboratively and successfully edited articles. Within three years, the world's largest Open Content project has achieved more than 1.500.000 articles, outnumbering all other encyclopaedias. Still there is little research about the project or on Wikis at all.[1] Most reflection on Wikipedia is raised within the community of its contributors. This paper gives a first overview and outlook on Wikipedia research. Wikipedia is particularly interesting to cybermetric research because of the richness of phenomena and because its data is fully accessible. One can analyse a wide variety of topics that range from author collaboration to the structure and growth of information with little effort for data collection and with high comparability.

A short history of Wikipedia

Wikipedia is one of the largest instances of a Wiki and one of the 100 most popular Websites worldwide.[2] It uses Wiki software, a type of collaborative software that was invented first by Ward Cunningham in 1995 (Cunningham & Leuf, 2001). He created a simple tool for knowledge management and decided to name it WikiWikiWeb using the Hawaiian term "wiki" for "quick" and with allusion to the WWW. Briefly a Wiki is a collection of hypertext documents that can directly be edited by anyone. Every edit is recorded and thus can be retraced by any other user. Each version of a document is available in its revision history and can be compared to other versions.

Wikipedia originated from the Nupedia project. Nupedia was founded by Jimmy Wales who wanted to create an online encyclopaedia licensed under the GNU Free Documentation License. In January 2001 Wikipedia was started as a side project to allow collaboration on articles prior to entering the lengthy peer review process. Soon it grew faster and attracted more authors than Nupedia which was closed in 2002. In addition to the original English Wikipedia Wikipedias in other languages soon were launched. In March 2005 there are 195 registered languages, 60 languages have more than 1.000 and 21 languages more than 10.000 articles. The English Wikipedia is the largest one, followed by German, Japanese and French. In June 2003 the Wikimedia Foundation was founded as an independent institution. It is aimed at developing and maintaining open content, Wiki-based projects and to provide the full contents of these projects to the public free of charge. Meanwhile Wikimedia hosts also Wiki projects to create dictionaries, textbooks, quotes, media etc. In 2004 a German non-profit association was founded to assist the project locally. Associations in other countries follow. Their main goals are collecting donations, promotion of Wikipedia and social activities within the

---

[1] See http://de.wikipedia.org/wiki/Wikipedia:Wikipedistik/Bibliographie (retrieved April 10, 2005) on Wikipedia and http://www.public.iastate.edu/~CYBERSTACKS/WikiBib.htm (January 30, 2005) on Wikis.
[2] http://www.alexa.com/data/details/traffic_details?y=t&url=Wikipedia.org, retrieved April 7, 2005

community of "Wikipedians" – the name that around 10.000 regular editing contributors call themselves.

Wikipedia Research

While Wikipedia is mentioned more and more often in the press, there is currently little scientific research exclusively about it. Ciffolilli (2003) describes the type of Wikipedia's community, processes of reputation and reasons for its success. Viégas, Wattenberg and Kushal (2003) introduced a method for visualising edit histories of Wiki pages called the History Flow and found some collaboration patterns. Lih (2004) analyzed citation of Wikipedia articles in the press and the ratio between number of edits and unique editors. Miller (2005) deals with the blurring border between reader and author. By far most articles consist of introductions, reviews and comments such as its first mention in Nature (Mitch, 2003) or detailed descriptions (Voss & Danowski, 2004). Several masters theses are under way.[3]

Wiki research in general is likewise in the beginnings. Important papers after the book of Cunningham and Leuf (2001) are Aronsson (2002) and Mattison (2003). There is more stimulus in research on social software for instance weblogs (Kumar et al., 2004), and in research on collaborative writing at large as well as on the Wiki community itself. Further progress is expected at the first Wikimania conference in August, 2005[4] and the symposium on Wikis in October, 2005.[5]

All content of Wikipedia is licensed under the GNU Free Documentation License (Free Software Foundation, 2002) or a similar license. The license postulates that when you publish or distribute the content you have to provide a "transparent copy" in a machine-readable format whose specification is available to the general public. Except for some personal data one can download SQL-dumps of Wikipedia's database. The entire dump of texts in all languages is more than 50 gigabyte compressed (15 for just current revisions). The current versions of each English Wikipedia's articles without images are 585 megabyte and 285 megabyte for the German version.[6] Erik Zachte has written the Wikistat Perl script that creates extensive statistics out of the database dumps and some server logfiles.[7] In addition one can extract data out of the live Wikipedia using the Python Wikipedia robot framework.[8] Furthermore there are several smaller statistics distributed on different Wikipedia pages.

In the following the SQL-dumps of the German, Japanese, Danish and Croatian Wikipedia and some other languages of January 7, 2005 will be used, as well Wikistat statistics of May 9, 2004. By number of articles German and Japanese are the second and third largest Wikipedias while Danish and Croatian are examples of middle-size and small-size. The English Wikipedia was not used because in October 2002 around 36,000 gazetteer entries about towns and cities in the USA were created automatically which still biases some statistics. Authors per article and articles per author statistics are based on the database dump of September 1, 2004 that was used to create the first Wikipedia-CD.

Structure of Wikipedia

All Wikipedias are hosted by Wikimedia Foundation using MediaWiki – a GPL licensed Wiki engine that is developed especially for Wikimedia projects but also used elsewhere. There are many suggestions for features though the main challenge is to deal with Wikipedia's enormous growth. After a look at growth, Wikipedia's fundamental components will be analysed followed by a discussion of content and quality.

Growth

Soon after its start Wikipedia began to grow exponentially from about 2002.

---

[3] http://meta.wikimedia.org/wiki/Research, retrieved April 7, 2005
[4] http://wikimania.wikimedia.org, retrieved April 10, 2005
[5] http://www.wikisym.org, retrieved April 7, 2005
[6] http://download.wikipedia.org/, retrieved March 9, 2005
[7] http://en.wikipedia.org/wikistats/EN/Sitemap.htm, retrieved April 10, 2005
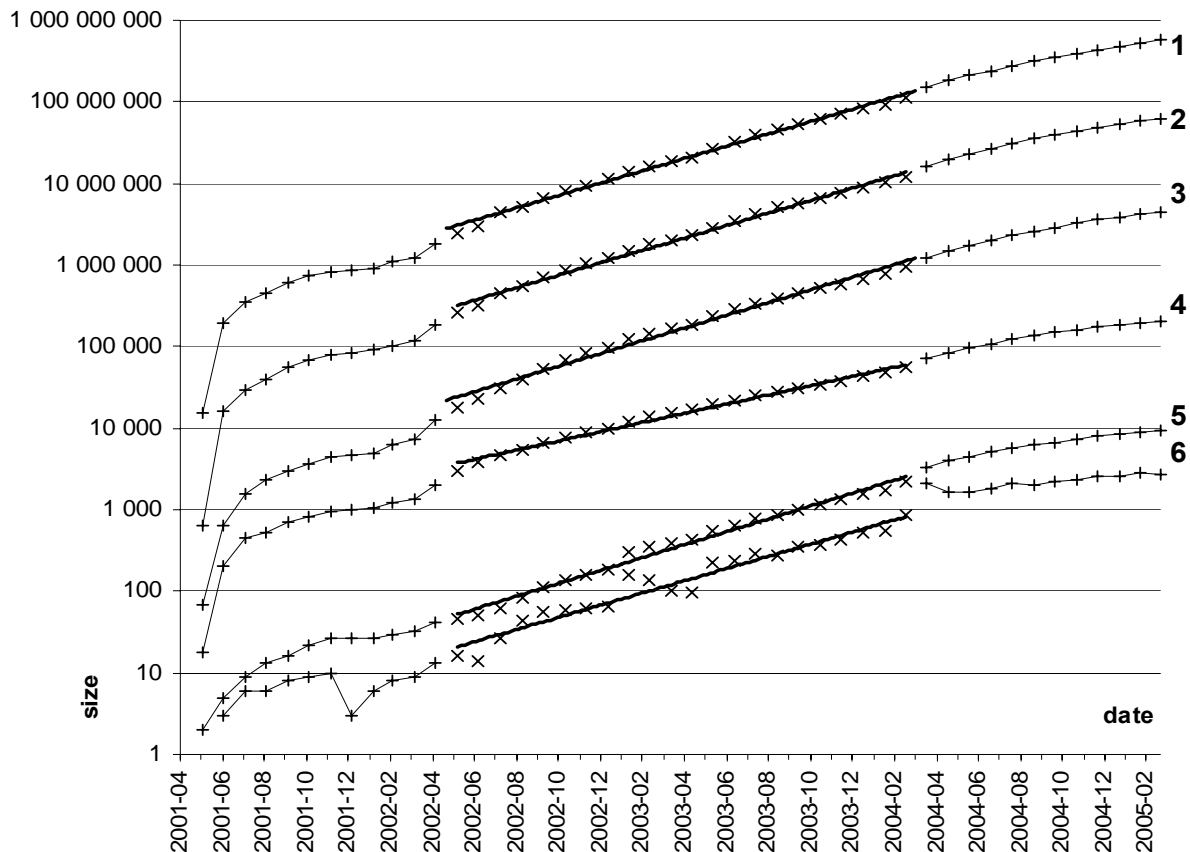[8] http://pywikipediabot.sourceforge.net/, retrieved April 7, 2005

Figure 1: Growth of German Wikipedia between May 2001 and February 2005

Figure 1 shows six fundamental metrics of Wikipedia's growth. These are:

1. Database size (combined size of all articles including redirects in bytes)
2. Total number of words (excluding redirects and special markup)
3. Total number of internal links (excluding redirects and stubs)
4. Number of articles (at least contain one internal link)
5. Number of active Wikipedians (contributed 5 times or more in a given month)
6. Number of very active Wikipedians (contributed 100 times or more in a given month)

You can divide three phases. After the first, linear phase exponential growth started around April 2002 when 10 active Wikipedians and 2.000 articles were exceeded. Until February 2005 all metrics increased similarly with around 18 percent increase per month. Only the number of articles increased slower with 13.8 percent. Around March 2005 the growth abruptly changed to linear growth but with a big jump in total growth at the same time. A similar effect can only be observed at the Japanese Wikipedia around November 2003. All other Wikipedias are in the first (linear) or second (exponential) phase – including English that only dropped increase rate around March 2005. Probably the third phase (linear) at the German Wikipedia is caused by technical limits and may only be a transitional state. At regular intervals Wikipedia becomes slow because software and hardware cannot cope with the exploding number of users. Obviously Wikipedia cannot grow infinitely but at the moment there is no evidence that it will not continue growing exponentially. De Solla Price (1963) gives some forms of reaction when exponential growth nears its natural boundaries. In one case the subject changes and the measure becomes absurd. In another case new techniques create new phases of exponential growth. In any case Wikipedia will have to change its nature in some way in the future. An understanding of its growth process can be useful in doing so.

Generally all measures but number of articles increase in the same way. That means article's size, number of links and numbers of active Wikipedians per article continuously rise. This effect is steady in different languages with different expansion rates. German is one of the faster growing

Wikipedias with 15.6% word number increase per article and month in 2004 (Japanese: 16%, Danish: 7.2%, Croatian: 18.6%, English: 7.8%).

Articles

Every article in Wikipedia is referenced via unique name. You access it with an URL like http://en.wikipedia.org/wiki/name-of-article where the subdomain (here: 'en') mostly corresponds to the particular ISO-639 language code. For synonyms *Redirects* that directly forward to another article can be created. Due to the fact that there is a special Wiki syntax which can be learned quickly, articles or single chapter can be directly edited without knowledge of HTML. Extensions allow timelines, hieroglyphs and formulas in LaTeX. Easy graph drawing, map generation and music markup is also being developed.

For each article there is a *Talk* page for discussions. Articles on the project itself belong to a special *Wikipedia* namespace. Uploaded image or other media files can be described at pages in the *Media* namespace. Each logged-in user has a *User*-page where he can introduce himself. Talk pages of user pages are essential for personal communication within the project. Furthermore there is the *Template* namespace for text modules that can be used in multiple articles, the *Category* namespace for categories that can be assigned to articles and the *MediaWiki* and *Help* namespace for localisation and documentation of the software.

Comparing namespace sizes is a simple method for getting an overview of a Wikipedia structure. Figure 2 provides an example. The percentage of all pages (no redirects) except for those in MediaWiki and Help namespace is compared. Talk pages are added to their corresponding page but for articles in the default namespace. Unsurprisingly normal articles make up the majority of pages with 60 to 80 percent. Together with talk and media pages they amount around 90 percent. Categories and user pages are used on a varying amount. German and Japanese Wikipedia have similar structures but there are less normal articles and more media, talk and other pages in the German Wikipedia.
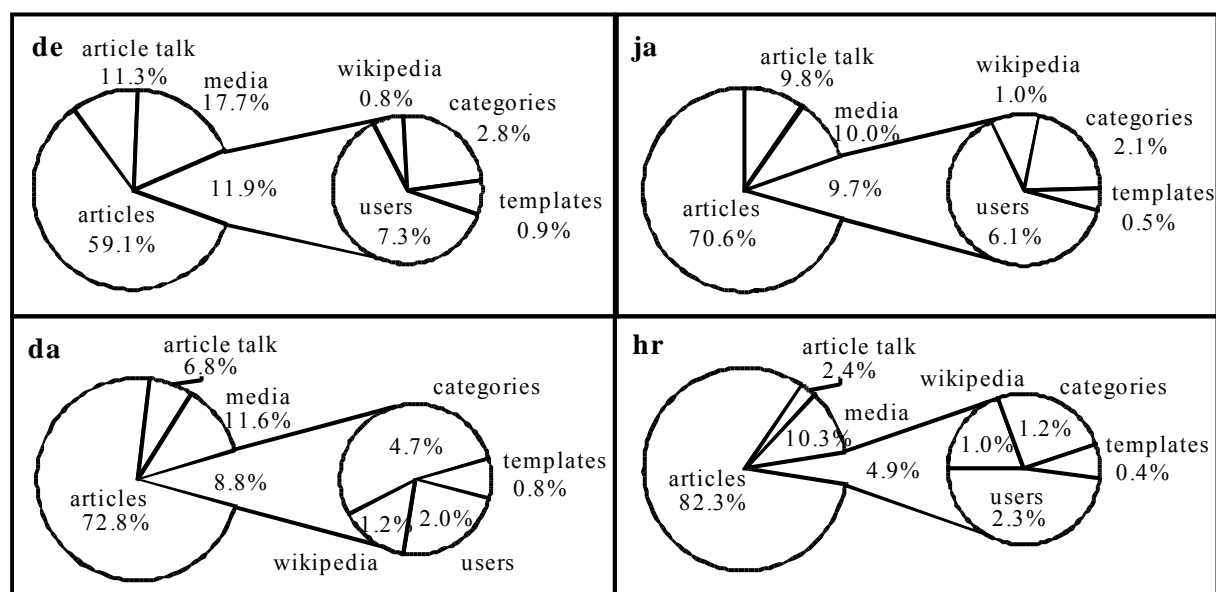

Figure 2: Namespace sizes of German (de), Japanese (ja), Danish (da) and Croatian (hr)

Table 1 contains some values that show the diversity of different Wikipedias. In German there is a talk page for around 19% of all articles, so it is probably more usual to comment on articles. The number of user talk pages by user pages in the Japanese Wikipedia is remarkable high because there are more than twice as much of the former than of the latter. Supposable Japanese Wikipedians better like notifying and discussing with each other then presenting themselves. A statistical comparison of all Wikipedias and the change of namespace fractions in time should give more precise results as well as the number of edits in each namespace.

Table 1: Some comparisons based on namespace counting

	German (de)	Japanese (ja)	Danish (da)	Croatian (hr)
number of articles	188,408	93,561	22,513	5,118
talk pages by articles	0.192	0.139	0.93	0.29
redirects by articles	0.495	0.424	0.355	0.911
user talk pages by user pages	0.94	2.51	0.88	0.74

Each article should only deal with one concept so article titles form a controlled vocabulary. In addition to redirects for synonyms there are disambiguation pages for homonyms. A disambiguation page lists different articles that might reside under the same title to let the reader choose between. Up to now there is no special software support for disambiguation pages, so you get no feedback if you accidentally link to a disambiguation page. More or less one could detect disambiguation pages automatically. Ideally disambiguation pages could serve as dictionary entries on words or phrases while normal articles cover topics. Unfortunately many disambiguation pages also contain some content that meant to be part of a normal article. In further research disambiguation pages have to be treated separately. Other kind of special articles in the default namespaces are list pages that only list a number of other articles. Most Wikipedias also have heavily linked articles for years, day month, decades, etc. which constitute an almanac.

Approximately article sizes are lognormal distributed. In the beginning, when a Wikipedia is small, it also contains more small articles ("stubs") with dropping fraction (Figure 3). Lognormal distributions are known for many natural processes where the growth over a time step is a normal distributed random factor independent of the current size (Crow & Shimizu,1988). Overall Wikipedia articles may grow in percentage what leads to a lognormal-like distribution but probably the perfect article size is also influenced by its topic.
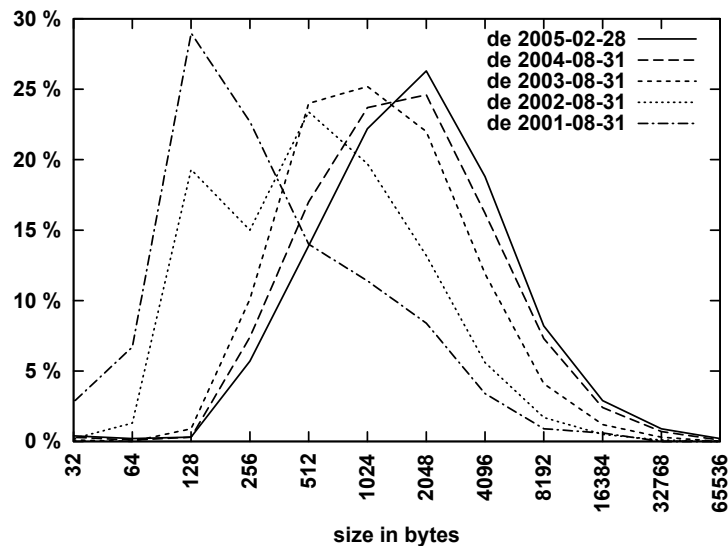

Figure 3: article sizes distribution in German Wikipedia.

Because of the skewed distribution one should not use the mean but the median as an average measure for values like number of words and links per article. The median article size grows linearly with different speed depending on the language. Between January 2003 and December 2004 it steadily grew around 28.7 (German), 11.5 (Japanese) and 14.1 (Danish) bytes per month achieving 1,190 (German), 459 (Japanese) and 366 bytes (Danish). The Croatian Wikipedia's average article size was not yet growing significantly.

Authors

Since everyone is invited to contribute, Wiki articles can have a large number of authors. Meanwhile multiauthorship is also the norm in scientific papers. It has increased markedly in recent years and differs between journals and subject area but is more or less normally distributed around 3 and 4 authors (Vuèkoviæ-Dekiæ, 2000; Regalado, 1995). The number of distinct authors per Wikipedia article – a number that Lih (2004) called diversity – follows a power law distribution. Figure 4 shows the distribution of authors per articles based on the data of the first Wikipedia CD-ROM (Directmedia, 2004). It contains all articles of the German Wikipedia on September 1, 2004. For articles with between 5 and 40 distinct authors the number follows a power-law with γ ≈ 2.7. Some special articles have more authors. 'September 2003', 'Gerhard Schröder', 'März 2004' and '2004' were leading with more than 100 contributors each. Almost half of the articles (47.9%) have less than 5 distinct authors and almost a quarter (27.6%) of all articles in the German Wikipedia of September 1, 2004 had only been edited by one logged-in user. Note that anonymous edits are omitted in this calculation. The number of anonymous edits varies by language between 10% (Italian) and 44% (Japanese). A deeper investigation may also factor in anonymous edits, number or size of edits by one author and edits on discussion pages. There is also special interest in whether there is any correlation with quality, size, age and other metrics. Generally you can say that multiauthorship in Wikipedia begins with 4 or 5 users and follows a power-law.
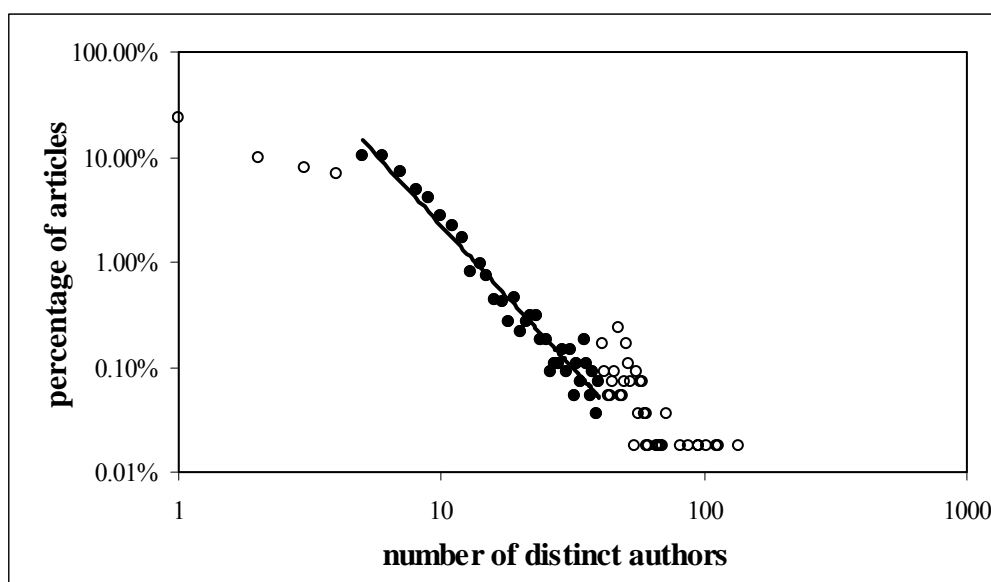

Figure 4: Distinct authors per article (γ ≈ 2.7)

The same data is used in figure 5 to show the number of distinct edited articles per author. A third of them have only contributed in one article and only 20% has ever touched more than 16 articles. This includes tiny edits like spelling corrections as well as "real" authorship on articles. It is to be expected that there is a similar distribution of edits per author. To get more precise values, edits of bots (software agents to batch article edits) have to be removed and successive edits of one author within a short period of time have to be counted as one edit. Lotka's Law has numerously been determined in scientific publishing. It can be found in other areas for instance open source development (Newby, Greenberg and Jones, 2003) and seems to apply also for Wikipedians as shown in figure 6. Based on the fact that the more experience you have in editing articles the more articles you will edit, this is not surprising.
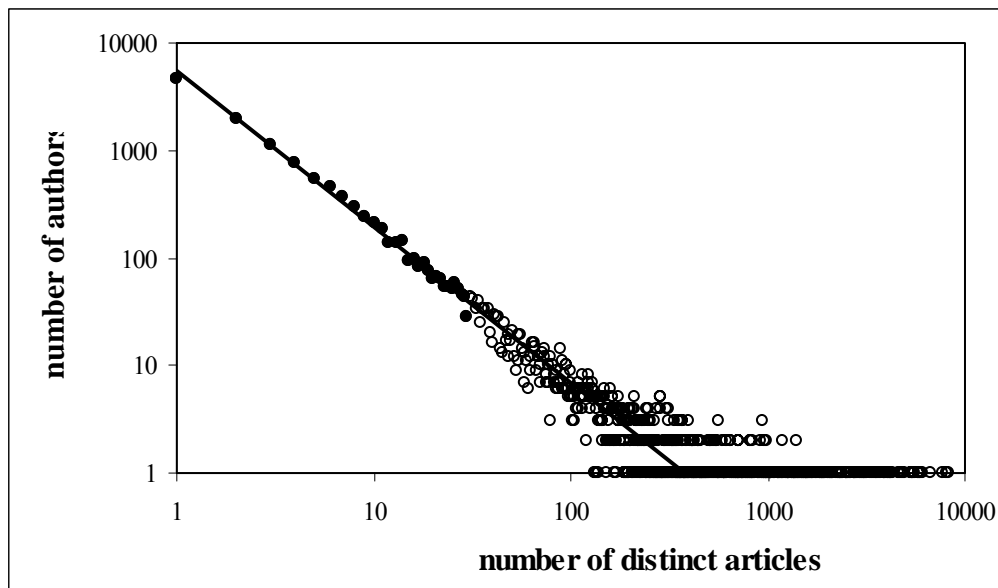
Figure 5: Number of distinct articles edited per author (γ ≈ 1.5)
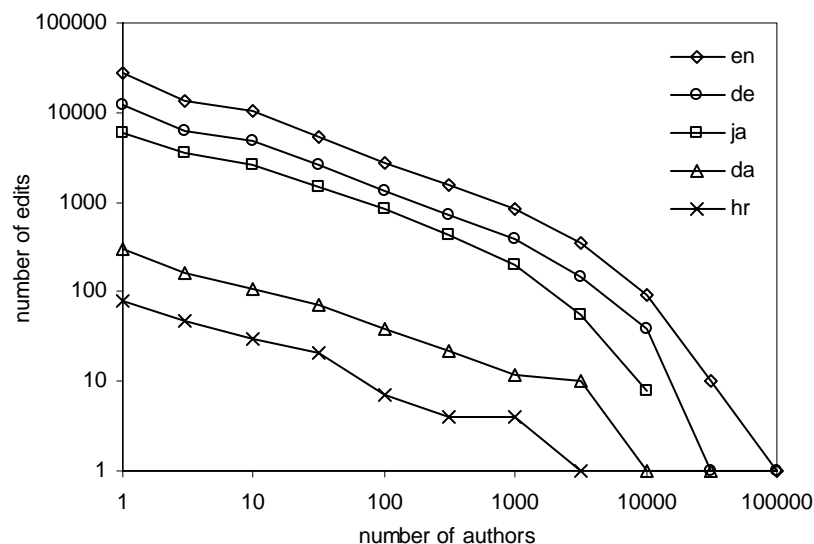
Figure 6: Edits per author for five Wikipedias as counted by Wikistat April 9, 2005 (γ ≈ 0.5)

The total number of authors is difficult to measure because of multiple accounts and anonymous edits. Wikistat counts all accounts as "contributors" and all accounts with 5 or more contributions in one month as "active Wikipedians". Obviously the latter is more significant. There is large interest in sociology of Wikipedians within the community. Most likely one can determine subcommunities. Wales (2004) mentioned some role models he is going to observe in detail. The Wiki concept highly facilitates division of work. Some users concentrate on correcting language and grammar, some on linking articles and some on arbitrating in a dispute. Like other online media Wikipedia also attracts vandals (users intentionally writing nonsense) and trolls (users provoking for sake of discussions) but the possibility to restore any article version decreases their ability to do harm. Reliable Wikipedians get admin status so they can restore articles more quickly, delete and undelete articles and ban users if needed. In the community there is heavy effort to reach consensus in such decisions. A public survey in the German Wikipedia collected ages of 284 users between 12 and 77 years and resulted in a mean of 31 and median of 29 years for December 2004.[9] A scientific survey on motivation of Wikipedians

---

[9] http://de.wikipedia/Wikipedia:Altersumfrage/Auswertung, retrieved January 24, 2005

will be published soon.[10] Details of the community structure can be determined by social network analysis. Links between actors are detectable by frequent co-authorship on articles, sequent edits, discussion on talk pages and so on. One can easily create user profiles from the list of a user's contributions but depersonalisation is absolutely recommended because of privacy.

Edits

Whenever someone changes an article, his edit is recorded and gets listed in the article's version history where one can highlight differences between selected versions. Users can add articles to their watch list to get notified on changes. All changes are listed at the recent changes, an important place to observe new contributors and suspect edits. Some structures – for instance two authors who revert each others edits ("edit war") – can be detected automatically. Averaged there are 16 edits per minute in the English Wikipedia with daily oscillation (6.6 in the German version). Percentage of anonymous edits varies by language (English: 22%, German: 26%, Japanese: 42%, French 15%, Croation 16% till end of 2004).

To measure the amount of an edit, the shared information distance between two versions can be used. It was used first to calculate the similarity between sequences of genomes (Li, Ma & Vitany, 2003) and also applied to other mutating sequences, for instance chain letters (Bennet, Li & Ma, 2003). The shared information distance between two strings is calculated with their Kolmogorov complexities (the minimal number of bits needed to encode a string) and the complexity of the string that results in concatenating them. Compression techniques give sufficient approximation. Using a compression function z the similarity between two strings a and b can be calculated as $d(a,b) = [\ z(a) + z(b)\ ] / z(ab) - 1$. For small texts you should also factor in the tiny difference between $z(a)$ and $z(aa)$. A more subtle application of shared information distance can help in determining main authors and detect how parts of articles have been moved to another article (although it won't detect semantic changes and translation).

Linking structure

Wikipedia articles and links between them form a network of concepts. Networks have extensively been studied in social sciences, graph theory and network research (Newmann, 2003). Many real-world networks are mostly found to have a very right skewed degree distribution with power laws in their tails that means $p(k) \sim k^{\gamma}$ for the number $p(k)$ of nodes with degree k and some constant exponent γ. Networks with power-law degree distribution are also called scale-free networks. Link distribution for existing articles of all namespaces in the German Wikipedia is shown in figure 7. The number of articles with less than 5 outgoing links is lower because articles that do link to other articles will be enlarged or deleted soon.
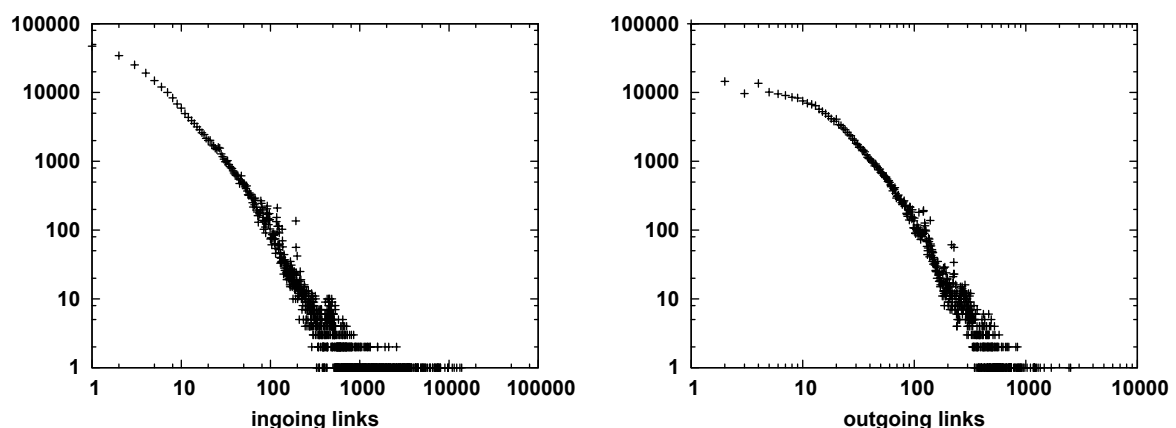

Figure 7: Distribution of articles per ingoing and outgoing links in the German Wikipedia

---

[10] http://wy2x05.psychologie.uni-wuerzburg.de/ao/research/wikipedia.php, retrieved April 10,2005

Models of network growth that explain scale-free degree distribution mainly base on *preferential attachment* (Barabási & Albert, 1999). Heavily linked nodes get linked more likely. The principle was introduced to network growth by Price (1976). Naranan (1970) explains how power law distribution arises from items that grow exponentially in number and in size. However the evolution of links between Wikipedia articles is more complex.

Links can be removed and added any time and links may point to pages that do not exist yet. Links to such a page ("broken links") are marked red to encourage people to create missing articles. Figure 8 shows a log-log diagram of the number of wanted pages in the German Wikipedia, number of broken links pointing to them and a visual power law fit with $\gamma \approx 3$. Pages with more than 17 incoming links (less then 1 percent with 190 of 19791) are omitted because they are special cases for instance links in templates that are copied in multiple articles automatically. The clean power law can be explained by the exponential growth of links and the assumptions that the more a not-existing article is linked, the more likely someone will create it. Most languages have similar values but one can also find some exceptions and differences between small and large Wikis.
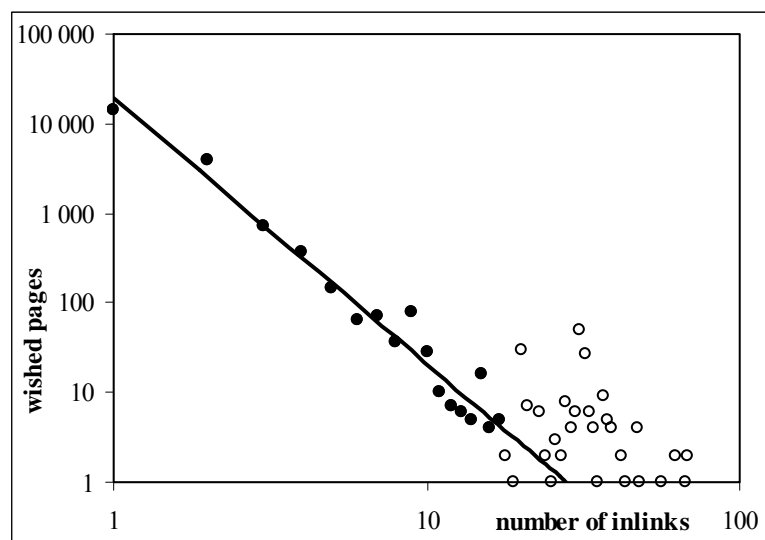

Figure 8: Distribution of broken links for all namespaces in German Wikipedia

Content

Content structure of Wikipedia articles can be identified using samples or using categories. Wikipedia's category system is a special form of social tagging (Mathes, 2004) related to classification and subject indexing. But the "Folksonomy" that is created in this process is still very inhomogeneous. In several Wikipedias there is a variety of special pages for organisation and navigation, for instance subject portals, lists, articles for single years, month and dates, templates, and navigation bars. There are also attempts to add more semantic markup. In German Wikipedia there are special templates to judge articles, to mark links to external databases and for collecting metadata on people. Beside such special pages, every article links to its most related topics. First tests on deriving thesaurus structures out of the linking between Wikipedia articles showed interesting results. Detailed comparisons with existing thesauruses and semantic networks and linguistic methods are even more promising. Bellomi and Bonato (2004) are doing research in this area. They applied HITS (Hyperlink-Induced Topic Search) algorithm to Wikipedia to find lexical authorities.

Quality

Richard Stallman's open source statement "given enough eyeballs all bugs are shallow" partly fits for Wikipedia too. The more people read an article the more errors are emended. But one can hardly be sure how many qualified people have read an article and how many errors remain. Edit history and user contributions are auxiliary clues but very time-consuming to review. Lih (2004) used the number of edits and the number of unique editors as a very simple and arguable approximation for quality. Wikipedia has been evaluated several times by journalist with very good results and Wikipedia articles

are regularly cited in press. Strongest critics one can find within the community itself. Many methods to improve quality – for instance rating or stricter policy on what should belong into Wikipedia – are discussed but in the end the openness is the most important feature the community does not want to give up. In some Wikipedias there are proceedings to selected articles which then are reviewed and improved until they get "featured article" status. Their number increases linearly around 20 to 30 articles a month.[11] WikiReaders[12] are another approach: an editor (or group) is accountable to select articles from Wikipedia on a certain topic to print or publish them as PDF.

Viégas, Wattenberg and Kushal (2004) have been cited as proof that vandalism gets removed quickly. This is true for obvious nonsense but other false statements may remain longer. Several people did simple tests with anonymously vandaling articles[13] but most interesting questions regarding vandalism and amateurism remain open. The overall quality of Wikipedia mainly depends on your definition of (information) quality and the time scale you are interested in. Wikipedia still is a young and rapidly changing project and it's not even clear if it is an encyclopaedia at all or if it will become an entirely new form of a knowledge base.

Conclusions

This paper presented a first overview on Wikipedia research. After a linear phase, Wikipedias grow exponentially with different rates per language. Generally the growth rate of the number of articles is smaller than other ratios. Different Wikipedias can be compared with indicators such as their growth rates, namespaces and article types. Article sizes are lognormal distributed with a linear growing median. For articles with 5 or more distinct authors its number follows a power law, so does the number of distinct articles per author. More detailed methods for measuring edits exist. Because every Wikipedia article covers a single concept and links to related ones, you can derive thesaurus-like structures out of the network of articles. The network is scale-free on ingoing links, outgoing links and broken links. There are many possibilities for deeper investigation, especially about quality and content. Like Wikipedia itself, research about it is at the very beginning but very promising.

References

Aronsson, L. (2002). Operation of a Large Scale General Purpose Wiki Website. In Carvalho, J., Hübler, A. & Baptista A. (Eds.) Proceedings of the 6th International ICCC/IFIP Conference on Electronic Publishing. Berlin: Verlag für Wissenschaft und Forschung

Barabási, A.-L. & Albert, R. (1999): Emergence of scaling in random networks. *Science* 286, 509-512.

Bellomi F. & Bonato R. (2004). Lexical Authorities in an Encyclopedic Corpus: a Case Study with Wikipedia. Retrieved April 10, 2005 from http://www.fran.it/blog/2005/01/lexical-authorities-in-encyclopedic.html.

Bennett, C., Li, M. & Ma, B. (2003). Chain letters and evolutionary histories. *Scientific American*, 288(6), 76-81.

Ciffolilli, A. (2003). *Phantom* authority, self–selective recruitment and retention of members in virtual communities: The case of Wikipedia. *First Monday*, 8(12)

Crow, E. L. & Shimizu, K. (1988). *Lognormal distributions: Theory and application*. New York : Marcel Dekker.

Cunningham, W. & Leuf, B. (2001): *The Wiki Way. Quick Collaboration on the Web*. Reading, Mass.: Addison-Wesley.

Directmedia (2004). Wikipedia - Die freie Enzyklopädie : Ausgabe Herbst 2004. Berlin: Directmedia Publishing

Free Software Foundation (2002). GNU Free Documentation License Version 1.2. Retrieved April 10, 2005 from: http://www.gnu.org/copyleft/fdl.html.

Kumar R., Novak J., Raghavan P. & Tomkins A. (2004). Structure and evolution of Blogspace. *Communications of the ACM*, 47(12), 35-39.

Li, M., Chen, X., Ma, L. & Vitany P. (2003). The Similarity Metric. *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*

Lih, A. (2004). Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource. Symposium. *Proceedings of the International Symposium on Online Journalism 2004* (in press).

---

[11] http://de.Wikipedia.org/wiki/Wikipedia:Statistik_der_exzellenten_Artikel, retrieved April 10, 2005

[12] http://en.wikipedia.org/wiki/Wikipedia:WikiReader, retrieved April 10, 2005

[13] http://de.wikipedia.org/wiki/Wikipedia:Wikipedistik/Vandalismusanfälligkeit, retrieved April 10, 2005

Mathes, A. (2004). Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Retrieved April 10, 2005 from http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html.

Mattison, D. (2003). QuickiWiki, sWiki, tWiki, zWiki, and the plone wars: Wiki as PJM and collaborative content tool. In: Searcher:Magazine for Database Professionals 11(4) , 32-41

Miller, Nora (2005). Wikipedia and The Disappearing "Author". *ETC: A Review of General Semantics*, 62(1), 37-40

Mitch, L. (2003). The People's Encyclopedia. *Nature*, 301(5), 1299.

Naranan, S. (1970). Bradford's Law of Bibliography Science: an Interpretation. *Nature* 227(8), 631-632.

Newby, G. B., Greenberg, J. & Jones, P. (2003). Open source software development and Lotkas Law: Bibliometric patterns in programming. *JASIST*, 54(2), 1169-1178.

Newmann, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review* 45(2), 167-256.

Price, D. J. de S. (1963). *Little Science - Big Science.* NY: Columbia University Press

Price, D. J. de S. (1976). A general theory of bibliometric and other cumulative advantage processes. *JASIST* 27, 292-306.

Regalado, A. (1995). Multiauthor papers on the rise. *Science*, 268(25), 785-786.

Viégas, F., Wattenberg, M. & Kushal, D. (2004) Studying Cooperation and Conflict between Authors with History Flow Visualization. In *Proceedings of the 2004 conference on Human factors in computing systems.* New York: ACM

Voss, J. & Danowski, P. (2004). Bibliothek, Information und Dokumentation in der Wikipedia. *Information : Wissenschaft und Praxis*, 8, 457-462

Vuèkoviæ-Dekiæ, L. (2000). Multiauthorship in three oncologic scientific journals. *Archive of Oncology*, 8(3), 109-110.

Wales, J. (2004). Wikipedia Sociographics. Talk at the 21C3-Conference, December 27, 2004, Berlin