Wikipedia as a Distributed Problem-Solving Network

Matthijs den Besten, Max Loubser and Jean-Michel Dalle

Introduction

Wikipedia is a large-scale collaborative project where people collaborate with the aim to put together an encyclopaedia that is free for everyone to read and edit. The collaboration has produced, amongst others, an English version with over 2 million articles, a German version with more than 600,000 and a Simple English version of over 20,000 articles. At the same time, Wikipedia has become extremely visible and more often than not its pages appear first in the result list of a Google search. What we want to do here is to take a closer look at the organization of Wikipedia, to look at the problems it faces and the way it addresses them, and, ultimately, to assess its potential for the future: Will Wikipedia become an even more important source of information that it already is or will its efforts wither under the weight of vandalism and zealotry?

At first sight, Wikipedia is a clear success. In 2005, *Nature* published the results of an attempt to assess the quality of Wikipedia. This attracted much attention. For what was found was that the upstart Wikipedia had a similar number of errors as the Encyclopaedia Britannica (Giles 2005). Britannica was compared to Wikipedia for quality as measured by experts in the area for each article sampled. Only 42 articles were sampled, chosen “to have a reasonable spread of Biology, Physics, Chemistry, Environment and Others” and to “have a good mix of People, Things, Events, Ideas / Processes and Places”. The study was criticised by the Britannica organisation for the way the texts were compared and suffers from the weakness of a small sample from a collection of millions, but it was influential in establishing that the distributed collaboration on Wikipedia can perform with success comparable to traditional methods of encyclopaedia production.

In terms of gross production, Wikipedia seems healthy too: content grew exponentially from 2002 to 2005 (Voß 2005). However, this is only a rough approximation of performance and does not take into account the quality of the content: As Denning, Horning, Parnas, and Weinstein (2005) note, a proper assessment of the performance and quality of Wiki projects would require attention to a whole range of factors.

With tremendous growth, performance is affected by how well maintenance and organisational work is carried out and how well this non-article work can be scaled. Reagle (2004) examines the way the Wikipedia community develops a neutral point of view in articles and the norms that govern meta-discussions, such as those that take place on article discussion pages. Kittur, Suh, Pendleton, and Chi (2007) measure directly the impact of these meta-discussions on the performance of the project. They found that the proportion of work on non-article content (such as discussion and fixing

vandalism) is increasing in Wikipedia, demonstrating that more overhead is required as the collection grows and casting doubt over its sustainability.

The increasing demand for maintenance work includes the need to undo the actions of those with malicious intent. In this regard Wikipedia appears to be performing satisfactorily. Even in the early stages of development, it was clear that malicious edits were removed quickly: Ciffolilli (2003) argues that it is the nature of the Wiki that makes it easy for friendly contributors to "clean up" vandalism. Later work by Viegas, Wattenberg, and Dave (2004) showed that "obscene" edits from their sample had a median lifetime of only 1.7 minutes.

And so, doubts remain over the long-term viability of Wikipedia as a reliable source of information (Duguid, 2006; Lanier, 2006). From the outside, vandalism and well-meaning ignorance are a continuous threat to article quality and from the inside, the dominance of difficult people with lots of time on their hands could easily become a danger of equal if not larger size. In the end, all depends on how well Wikipedia as a community manages to address these threats. Below, we will first review what we know about the way Wikipedia functions today. After that, we will examine the kinds of studies that have been done and could be done to better understand how Wikipedia functions in the future. Finally, we will present initial insights from a small-scale study in which we surveyed the tiny part of the Wikipedia universe occupied by Simple Wikipedia as an illustration of the kinds of insights that could be gained by an analysis of Wikipedia as a distributed problem-solving network at large.

Background

To put Wikipedia in perspective, it helps to consider the general framework of knowledge-work that was formulated by Mateos-Garcia and Steinmueller (2003). The organization of collective production of information goods, in their view, depends on four kinds of actors: There is, of course, the author, who creates the "information good component". In addition, there are editors "examining the internal consistency and completeness of the information good component that is to be integrated"; there is a system integrator to "take account of the immediately adjacent information components in order to preserve the integrity of the entire system"; and finally there is an overall co-ordinator or conductor who takes responsibility for the project as a whole. Mateos-Garcia and Steinmueller (2003, p. 7) distinguish between two types of good production: vertical production (systems) and horizontal efforts (collections). "Vertical information good production provides the strongest motives for collaboration because the possibilities of sharing in the externalities generated by other participants are matched by incentives to edit and integrate contributions due to the cumulative dependency of contributions," while "horizontal efforts, involving complementary dependence, are neither so well developed nor so likely to lure participants from the opportunities offered by self-publication and looser affiliation." In this lies the crux of the problem for Wikipedia: Voluntary management of the development of collections like Wikipedia is much harder to achieve than voluntary management of the development of systems like Linux as the direct benefit to the individual volunteer is so much lower. Collections like Wikipedia 'suffer' from a relative abundance of authors - the much smaller number of editors that

take on the roles of system integrator and co-ordinator struggle to cope and often need assistance from automated agents.

History

Wikipedia was started in 2001 after Nupedia, an earlier attempt to create an open content encyclopaedia, had failed (Sanger, 2005). Nupedia had aimed to be a highly reliable, peer-reviewed resource. Everyone could submit articles, but before the articles were accepted they had to go through a seven-step review process. In practice, this process proved to be too burdensome and part of the success of Wikipedia is due to the ease with which people can participate. However, as Larry Sanger (2005) points out, there were important factors that played a role as well:

> Wikipedia started with a handful of people, many from Nupedia. The influence of Nupedians was, I think, pretty important early on; […] All of these people, and several other Nupedia borrowings, had a good understanding of the requirements of good encyclopedia articles, and they were good writers and very smart. The direction that Wikipedia ought to go in was pretty obvious to myself and them, in terms of what sort of content we wanted. But what we did not have worked out in advance was how the community should be organized, and (not surprisingly) that turned out to be the thorniest problem. But the facts that the project started with these good people, and that we were able to adopt, explain, and promote good habits and policies to newer people, partly accounts for why the project was able to develop a robust, functional community and eventually to succeed.

Without the former Nupedia staff there to assume the roles of system integrator and co-ordinator, the take-off of Wikipedia would probably have been much harder to achieve. As time passes, administrators, members of the Wikipedia community that help with maintenance, seem to get burdened with more and more responsibilities. In addition, scripts known as bots appeared and took over an increasing pay-load for typical system-integrator tasks like the establishment and maintenance of links between Wikipedia project in different languages, and other repetitive edits that would be extremely tedious to do manually. Occasionally bots are used for automatic content generation, but this is generally frowned upon.

Technology

In terms of a technological platform, Wikipedia and its sister projects all employ the MediaWiki framework to structure their collections. However, there is a layer of practices that has evolved on top of that. The details of these practices are different from project to project and to some extent even within sub-communities in a project.

In the case of the English Wikipedia, the largest and first project to use the MediaWiki framework, most users arrive at the project website for encyclopaedic information, but can then go on to become authors in the collaboration. Below is a description of the workings of the platform.

The MediaWiki platform consists of a database and a user interface accessible through a web browser using HTTP. The interface provides anyone accessing it with basic document creation and management functions, but grants certain permissions only to registered users. Changes made to the database through the web interface are

immediately visible to other users and a full history of all transactions is kept: this history is also visible to all users.

The basic Wiki functions are page creation and page editing. A user can click an 'edit' link on any article, taking the user to an edit page where the contents of the article are displayed in Wiki markup. Wiki markup simplifies basic formatting tasks such as hyperlinking, sectioning and bullet points, making knowledge of HTML unnecessary. After changes are made to a page in Wiki markup, a preview of how the page will display is shown before the changes are saved and committed to the database. An entry for the edit is created in the page history and the changes are associated with the user's name or IP address.

The page development history can be viewed by any user and changes can be undone, by clicking the 'undo' link which appears next to an edit in the history list.

Users can discuss the work on a page by using the associated 'Talk' page. Talk pages are Wiki pages that are edited just like the article pages and are linked to from the main article page. Talk pages are named by preceding the article page name with 'Talk:'.

Organization

On Wikipedia the MediaWiki software allows for three main types of users: unregistered users, registered users and administrators. Each type of user has a different set of privileges and normally also slightly different, but overlapping roles.

Registered users have provided basic information and a username to the system and perform their work in a 'logged in' state. Registering users allows for the contributions of a single person to be tracked, should they wish to log in before they contribute. Registered users have user home pages and associated 'user talk' pages. User home pages can be used by registered users to provide information about themselves and 'user talk' pages can be used by others to leave messages for a user. Much of the communication between contributors takes place using 'user talk' pages, as a more personally directed, yet still publicly visible alternative to the article discussion page.

Administrators, also known as admins or sysops, are registered users with authority for special maintenance tasks. Wikipedia practice is "to grant administrator status to anyone who has been an active and regular Wikipedia contributor for at least a few months, is familiar with and respects Wikipedia policy, and who has gained the trust of the community." As of October 2007, there are about 1350 Wikipedia administrators in the main English Wikipedia. Administrators have the ability to delete (remove) pages, protect articles from being changed and block certain users from making changes to the project.

Administrators can intervene in the collaboration process by protecting (sometimes referred to as 'locking') an article in two ways: Full protection and semi-protection. Under full protection only administrators can make changes to an article. Under semi-protection only registered users with accounts older than 4 days can make changes to an article. The primary purpose of this power is to stop vandalism to articles, but protection is also used to intervene in an 'edit war', where two or more users continually change each other's contributions. Logs of all administrator actions are kept.

Registered users can 'watch' pages that they are interested in – watching a page will provide updates on changes made to the page without giving anyone control over the page. Administrators can see a list of pages that are not on anyone's watchlist.

The table below summarises the different actions that are possible to the Wikipedia users of MediaWiki.

Table 1 Range of Possible Interventions in Wikipedia

Action	Unregistered user	Registered user	Administrator
Edit page	Yes	Yes	Yes
Create page	Yes	Yes	Yes
Edit semi-protected page	No	Yes	Yes
Edit full protected page	No	No	Yes
Move (rename) a page	No	Yes if > 4days	Yes
Block a user	No	No	Yes
Delete a page	No	No	Yes
Delete history - view the history of a deleted page or a user's deleted contributions	No	No	Yes
Protect a page	No	No	Yes
Rollback – admin feature available on the 'User contribution' page when an edit is the last edit made to that page (marked with 'top'). The last contribution(s) of the last user to edit the page can be removed by clicking the rollback link.	No	No	Yes
Undelete - view deleted versions, undelete a previously deleted page, or undelete specific revisions of a deleted page	No	No	Yes

A wide range of Wikipedia specific practices and guidelines have been collaboratively developed alongside the encyclopaedia itself. These are stored in the Wikipedia database with page names that start with 'Wikipedia:'. Developing the policy and process documents require a large amount of non-article work. Kittur, Chi, Pendleton, Suh (2007) found that the proportion of this indirect work is increasing in Wikipedia and some of the pages relating to community practice have developed to large manuals, such as the Manual of Style or detailed policy documents such as Neutral Point of View.

Kittur, Chi, Pendleton, Suh, and Mytkowicz (2007) examine the role of "elite" users versus the role of the "common" user, testing whether Wikipedia success is due to a "wisdom of crowds" effect: Is most content created by a core group (admins, experienced registered users), or by participants on the periphery of the community? Using the same computational methods as Kittur et al. (2007), with a July 2006 dump of the entire archive, they find that administrator edits, as a proportion of all the edits in Wikipedia, has declined since mid 2003. At the same time edits by non-registered users have increased. They observe that this "rise of the crowd" is not due to a decrease in administrator activity, but was driven by a growth in the population of low-edit Wikipedia contributors.

A similar finding about the importance of low-edit, anonymous users is made by Anthony, Smith, and Williamson (2007). Using theory from studies of collective action,

they hypothesise that registered users will make higher quality contributions: "zealots" driven by the incentive of gaining reputation in the Wikipedia community. However, from a sample of 7058 users from the French and Dutch versions of Wikipedia, they find that anonymous users make higher quality contributions as measured in the retention of their work in articles. They ascribe this to anonymous "good samaritans" who make a small number of high quality contributions, but conclude that the zealots play at least as important a role.

Approach

An online effort like Wikipedia lends itself to all kinds of research. In our approach we take advantage of the detailed data on the editing that are logged and made available by MediaWiki. On the basis of the various measurements that such data allow, we can construct econometric models, which in turn could potentially help to detect trends and address problems in future.

Method

The performance of Wikipedia as a distributed problem solving can be assessed in a variety of ways. Participation in Wikipedia (editing and consuming articles, discussing content) is probably a very good way to get an impression of the organization of its community and of the effectiveness of this organization in producing and maintaining a collection of articles. However, such participation would need to be augmented by interviews or surveys to see whether the impressions are reflected in the experience of a wider group of people. Findings can also be quantitatively corroborated with a wide variety of measurements and detailed models using the information on the individual contributions and the editing of those contributions that is kept in the Wikipedia archives.

Qualitative methods can be used to examine the social context of work organisation and communication processes specific to Wikipedia: Viegas, Wattenberg, and Dave (2004) used interviews with Wikipedia contributors to corroborate their findings that reputation plays a role in the longevity of contributions to an article. Reagle (2005) has focused on ethnographic methods to study the role of individual leaders in Wikipedia and how they shape the direction of the articles they are involved with. Bryant, Forte, and Bruckman (2005) used qualitative data from interviews with nine "active, committed" Wikipedia community members and analysed their behaviour using activity theory, tracking how experienced "Wikipedians" assume their roles in the community.

Data

Mediawiki archives the edit history of the articles in the Wikipedia collection. This edit history contains the text of each revision of each article together with formatting instructions and information on the identity of the editor, the time of the edit, and possible comments that the editor added when he or she committed the edit. Thus, we have three sources of information: Text, tags, and meta-data. Each of these sources lends itself to particular measurements:

For text, simple measurements are the numbers of words and lines in versions of the articles, for instance. Through a combination of those measurements we can arrive at more complex, composite, metrics such as readability indicators. Further down the line, we can take an even larger variety of measurements and apply statistical techniques like principal component analysis to arrive at indicators of formality like Emigh and Herring (2005) did. Undoubtedly much more text mining could be done.

Tags, used to mark-up the text, give all kinds of interesting hints. Not only do they allow us to infer the structure of the article in terms of sections and subsections, they also allow us to infer the link-structure in between articles and indicate references to the Internet outside Wikipedia. However, in Wikipedia, there is a third and perhaps more interesting function of tags: Special tags called templates can be inserted to signal that there is a problem with the current text that needs to be addressed. To name a few, an article can be tagged with the label “unsimple”, “controversial”, or “stub” and subsequent readers will then know that someone feels that the article is not readable enough, too biased, or too small, respectively.

Meta-data, make it possible to measure the timing and distribution of labour as they convey the identity of the editor, the time of the edits, and the type of page the edit was made on.

In a first attempt to evaluate article quality using meta-data, Lih (2003) studied Wikipedia content construction and use processes from the perspective of participatory journalism. During 14 months of observation from January 2003 through March 2004 113 Wikipedia articles were cited in the press, with current events, slang and colloquial terminology being cited most often. The quality of the cited articles was then analysed in terms of the total number of edits (‘rigor’) and the total number of unique editors (‘diversity’), both of these available in the meta-data. A relationship between press citation and quality was established using these two metrics - articles with more press mentions had higher scores for rigor and diversity.

Emigh and Herring (2005) used article text as data source and compared writing style in Wikipedia with the Columbia Encyclopaedia. They counted the frequencies of the parts of speech (POS) known to be characteristic of a formal language genre for a sample of 49 Wikipedia articles and compared them to the frequencies of the same POS calculated for the Columbia Encyclopaedia. Based on these measurements they concluded that the language of Wikipedia articles is as formal as the language of the printed encyclopaedia, demonstrating that features of article text can be used as indicators for quality and performance.

Viegas, Wattenberg, and Dave (2004) drew on meta- and text data and developed a method of visualising article edit histories they call history flow. They visualised the development of 70 articles in this way. They compared their findings with statistics on an early (May 2003) version of the entire Wikipedia, containing 130,596 article pages and 11000 meta pages, but their findings are largely based on trends they observed in the history of the small sample, citing computational restrictions and the initial growth period of Wikipedia as problems for analysing the whole collection. Based on the analysis of edit patterns revealed by the history flow, they identified five types of active article quality degradation or vandalism: (1) Mass deletion: removing most or all of an article; (2) Offensive copy: inserting slurs and offensive words; (3) Phony copy: insertion of text unrelated to the page topic; (4) Phony redirect: redirecting to unrelated, often offensive material; (5) Idiosyncratic copy: adding related but biased and/or inflammatory content. The visual representation of the available data is useful for manually assessing the

relationship between quality and history for a small number of articles, but is limited by the need for careful human study of the resulting graphs.

Stvilia, Twidale, Smith, and Gasser (2005) searched for more scalable, automatic information quality measures from almost all the available sources of information in the MediaWiki archives. They worked from the assumption that article quality in Wikipedia is largely based on the use and edit distributions, as Lih (2003) did. They developed 19 information quality measures and grouped them into a set of seven information quality metrics. They used their metrics to compare community determined quality labels (the set of 'featured' articles) with their automatic quality measures, using a random sample of articles as a control. Their random sample size was 834 articles, compared to a featured article sample of 236 from an April 2005 dump of Wikipedia. They also performed a longitudinal analysis with three historical reference points sampled from a 43 day period. Using the Density Based Clustering algorithm they tested the accuracy of the metrics in determining whether an article is in the random or featured set. 14% of the random sample and 9% of the featured sample was incorrectly classified. Using a decision tree classification algorithm correct classification rates of 90% for the featured sample and 98% for the random sample were achieved, indicating that measuring elements of article structure, text, or meta-data for a relatively small subset of articles shows promise.

Although Emigh and Herring (2005), Stvilia et al. (2005), Viegas et al. (2004) and Anthony et al. (2007) all use samples of articles or samples of users in their measurements, the most accurate picture of the workings of Wikipedia will be from use of all the available data. Unfortunately, when not working with a sample, managing the large size of the full archive (now over 800 gigabytes for the English Wikipedia) is problematic. Kittur, Chi, Pendleton, Suh, and Mytkowicz (2007) and Kittur, Suh, Pendleton, and Chi (2007) were the first to address issues of performance on the entire archive of the English Wikipedia (others such as Voß (2005), used the smaller archives from other languages). Their work was replicated and extended to the non-English Wikipedias by Ortega and Barahona (2007), who developed a tool called WikiXRay to download and analyse entire Wiki databases.

Objectively determining article quality and dealing with the large amount of available data raises methodological questions: Which of the many endogenous or exogenous indicators should be used? Lih (2003) used proxy indicators from the meta-data: the number of edits and the number or different editors. Den Besten and Dalle (2007) use a direct and endogenous indicator from text data, readability scores, as a measure of quality. Stvilia, Twidale, Smith, and Gasser (2005) use a mixture of exogenous indicators such as the number of inlinks, editors and reverts with endogenous ones such as readability scores, number of images and information noise (The Appendix contains a table of metrics uses in different works). The comprehensive record of participation makes Wikipedia unique as a data source. Quantitative studies of Wikipedia in its more mature stages, after 2003, have begun to make it possible to derive trends, that may apply in other similar distributed problem solving efforts where the data may not be as detailed. Emigh and Herring (2005) and Kittur et al. (2007) have already compared Wikipedia to other collaborative projects such as del.icio.us.

Focus

We see the performance of Wikipedia as a distributed problem solving effort in which what counts is the effectiveness of the community to locate and deal with a series of

sub-problems. These sub-problems range from making sure that there are no spelling mistakes in articles, and the recovery from defacing caused by vandalism, to making sure that articles are consistent and adhere to standards and the management of an article's life cycle. Many of the sub-problems that Wikipedia faces are self-identified. There are scripts to address minor problems like spelling mistakes and correct interlinking across articles; there are squads of Wikipedia-users who scour recently edited pages for signs of vandalism, and, most importantly, templates are used to label articles which have deficiencies, such as a lack of readability, that cannot be easily remedied. It is those kinds of problems in which we are most interested and from this perspective the performance of Wikipedia is reflected in the collection-wide trend that emerges from continuous efforts to identify and triage articles that are problematic and from the associated efforts to improve the articles by a self-selected set of editors. Table 2 provides a schematic outline of this idea.

Table 2 Tracing Performance Throughout Wikipedia

	Aggregate Trend	**Triage Template**	**Article Treatment**
Quantity	No. articles, words/article	Stub	Expand
Quality	Readability; Bias	Unsimple, Regional	Rewrite
Effort	Edits; Interventions	Featured	Calibrate

Measuring the performance of Wikipedia from this perspective has two tiers. The first tier is to measure any general trends that we are able to tease out of the aggregate of articles. These are trends like the growth in the number of articles, and number of contributors, but also trends like overall changes in readability and changes in the ratio of "steering" interventions to "constructive" edits. These performance measures are quantitative and use textual and meta-data and allow us to assess the sustainability of Wikipedia practices. The second tier of our approach, in contrast, is to trace the processes of triage and treatment that have been put in place by the community to improve certain aspects of individual articles. In this tier we use tag and template data. For example, labelling certain pages with the "unsimple" template can be seen as an attempt to address a general trend towards increasing complexity of pages. The identification of pages that require attention is a process of triage and the subsequent rewriting or change can be seen as the treatment.

Even when explicit triage and treatment affects only a limited number of articles, we see this process as crucial for the performance of Wikipedia as a whole. Compared to the other types of editorial intervention that are available to the Wikipedia core, triage and treatment is most likely to establish some sort of best practice in article writing. The alternatives, adding a comment on an article's or editor's discussion page or locking an article so that it can only be edited by a limited group of people, are either too specific and invisible to the community as a whole in the case of commenting or too blunt and heavy-handed in the case of locking. Each article is likely to go through phases of development in which a large diversity of contributions is required and integration in which adherence to standards is what counts. What triage and treatment do is to make these phase-transitions or regime-changes visible and explicit.

Preliminary Results: Simple Wikipedia

In order to test the viability of our approach, we did a pilot study on Simple Wikipedia. Simple Wikipedia is a spin-off of Wikipedia that was initiated in 2003 because people felt that many articles in Wikipedia were too hard to read, due to jargon, formality, or for other reasons – especially for children and non-native speakers of English. Simple Wikipedia, was the hope, would be the place where people go to look for an easily readable descriptions of topics. At the same time, contributors to Simple Wikipedia would commit to the ideals of this sub-project or at least adhere to an editorial policy that calls for greater readability when contributing descriptions of topics.

Simple Wikipedia is an ideal candidate for a study of the performance of distributed problem solving networks in the way that we envisage such a study. For, Simple Wikipedia is a project for which we can determine relatively easily how well it adheres to its goals, zoom in on efforts that are made to address specific problems, and assess the result of these efforts. First of all, with less than 20 000 articles in its collection, Simple Wikipedia is a relatively small encyclopedia project. Consequently, no extraordinary computational resources are needed to extract the project archive and analyze its contents. Besides, Simple Wikipedia is a project centered around a very specific goal, readability. What's more, readability is something that can be measured. So, here we have a project that we can assess on its own terms. Moreover, Simple Wikipedia, as a separate project, was able to experiment with and implement its own editorial policies and managerial policies specifically to help attain its goal of simplicity. Most importantly, Simple Wikipedia has come to rely on the tag "unsimple" – more recently called "complex" – to single out articles which do not meet its standards and need to be improved.

Data

The archive of Simple Wikipedia is available from the web at downloads.wikimedia.org. Our analysis here is based on the archive of July 2007, which contains the revision history of over 25 000 articles and around 27 000 pages of a different type such as discussion pages and user-pages where regular contributors present themselves. For each edit on an article, the archive lists the ip-address or user-name and user-id of the editor, the time of edit, comments made by the editor, and the actual text resulting from the edit. The text is marked up with tags to identify structural elements like sections and sub-sections and tags of a different type to identify labels – also called templates – that are applied to the text. For instance, an article that is considered to be hard to read will contain the string "{{unsimple}}" in the raw text. Slightly harder to determine is the status of an editor. We can easily distinguish between editors who are known to the system and editors who contribute anonymously, as the latter are identified by their user-id while for the former only the ip-address is listed. Bots – scripts that carry out small repetitive edits such as spell checks and interlinking of articles – usually have a user-name that ends with "bot". In addition, the user-id of these bots is listed as a user belonging to a special group in the auxiliary user-group table, as is the user-id of users with special rights known as administrators. The readability of an article is determined by computing the Flesch readability score of the article's text with help of the GNU Style package. This score is a function of the number of syllables per word and the number of words per sentence (Flesch, 1979). More precisely, the formula 'score = 206.835 – 84.6*syllables/words – 1.1015*words/sentences' yields a number that is usually between

0 and 100 and between 60 and 70 for standard English texts. This Flesch reading easy formula, which has been elaborated on the basis of school texts by Flesch in 1948, has been very popular, especially in the US, as a measure of plain English. Its popularity rests on the fact that the formula is easy to compute, yet often accurate. Even work-processing programs like Word often provide the score as part of their statistics.

Trends

On the basis of the archive, it is possible to reconstruct the history of Simple Wikipedia. For each article in the archive we know when it was first introduced and for every month that Simple Wikipedia existed, we can find the versions of the articles in the archive that existed in that month and we can count their number and properties like their overall readability. Figure 5 in the appendix shows the result of such a reconstruction for Simple Wikipedia from January 2003 until December 2007. The figure shows an upward sloping line indicating the total number of articles in the collection for each month and a downward sloping line indicating the overall readability of the articles in the collection for each month. Note that the continuous growth in the number of articles in Simple Wikipedia is at least partially due to the fact that articles that have been removed from the collection and are not currently available anymore, do not appear in the most recent archive either. Even so, the growth in the number of articles over time is impressive. This rate of growth may be a factor that explains why the readability indicated by the Flesch readability score shows a gradual decline: As the size of articles in Simple Wikipedia grew, it became more unwieldy and editors faced an ever harder task to maintain the standards of readability, is one interpretation that suggests itself. Looking closely, we can distinguish a phase of substantial decline in readability in the first half of 2004 followed by a more stable phase in the second half of 2004. It was in the second half of 2004 that the practice of tagging articles with the label “unsimple” first appeared. It might be that the slow down in the decline in readability could be attributed to this. With a readability index of over 70, Simple Wikipedia still scores very well, given that standard English falls between 60 and 70. Still, the fact that readability continues to decline should be worrying.

Figure 6 tries to shed light on who has been editing articles in Simple Wikipedia from 2004 until 2006. Most dominantly, it shows the growing importance of bots – indicating that automated and semi-automated scripts are relied upon more and more to keep the articles consistent. At the same time, the proportion of edits done by people who do not identify themselves to the system decreases and the proportion of edits done by people who are registered increases relative to them. Whatever causes the gradual decline in readability, it is probably not the hordes of outsiders in the case of Simple Wikipedia.

A final issue is the number of articles that have been identified as “unsimple”. This turns out to be a relatively small number: Less than 200 between 2004 and 2006. We do find however that there is a clear improvement in readability on average following the appearance of the label. Another interesting finding is that the act of tagging seems to be carried out by users with a special administrator status most of the time and hardly by anonymous outsiders (cf. Den Besten and Dalle 2007).

Triage

Despite the overall trend towards a decline in readability and the limited number of articles that have been labeled with “unsimple”, applying such a label does seem to have

the desired effect in several cases. Take for instance the article on propaganda. Figure 7 depicts the evolution of the readability over time of this page. Readability is based on the Flesch reading easy score and time is given as the number of days relative to the first occurrence of the tag “unsimple” on the page. In addition, the dotted vertical lines indicate the period during which the page was labelled with the tag “unsimple” and the shapes of the points correspond to the different contributors to the page.

The first thing to notice is the dramatic effect that the tag “unsimple” seems to have on the readability of the page. Before the tag first appears, the readability score of the page hovers around the border between a standard and a fairly difficult level of readability. When the tag is removed, 100 days later, the readability has jumped to the other side of the spectrum, now bordering between a standard and a fairly easy level of readability. Looking at the shapes of the points, a history emerges in which someone wrote the initial page to which others then added a bit without changing much in terms of readability. At some point, day 0, a new guy comes along and points out that the page is too difficult to read for an article in the Simple Wikipedia collection. The tag “unsimple” then attracts the attention of someone else who actively starts working to improve the readability of the page. As soon as the problem has been remedied, the tag “unsimple” is removed and regular maintenance is resumed. More specifically the history that this Figure represents is as follows (http://simple.wikipedia.org/w/index.php?title=Propaganda&action=history) : The first version of the article on propaganda appeared on 1 March 2004 and was written by Randywombat. In the year that followed we see five edits by non-identified users and two by identified users. There are also three edits by bots, but these do nothing more than adding so-called inter-wiki links to similar pages in other languages. In November 2005, the user Heroismic, a high-school junior (according to his/her user-page) who has not edited the page before, comes along and labels the article “unsimple”. In the following months, we see a concerted effort by the users Eptalon, according to his/her user-page a Simple Wiki administrator from Central Europe, and AmanitaMuscaria to improve the page. On 22 February 2006 Eptalon decides that the page has become readable enough and removes the label “unsimple”.

Figure 8 shows the development of readability for few more articles. In this figure, the readability of the articles is represented by a ‘o’ while the article has the label “unsimple” and a ‘+’ otherwise. Sometimes readability increases but no one bothers to remove the “unsimple” label. In other cases, the label leads to some improvement in readability initially but contributors fall back on bad habits later. And, sometimes, the label seems to be ignored altogether. The interesting question here is not just how likely it is that this kind of “treatment” by labelling succeeds sometimes, but more why it succeeds in some cases and not in others. In a paper that is currently in preparation, we carry out a survival analysis on several variables describing the articles’ state at the point of triage and subsequent edits during the treatment to find out exactly this (Dalle et al, in preparation).

Conclusion

The quality and sustainability of Wikipedia is a topic in which scholars of all kinds are getting more and more interested. In this paper, we have provided a review of recent studies of Wikipedia and we have proposed a new approach to study its performance

that takes advantage of the way in which this encyclopedia project is organized. In particular, we propose to look for signs within the Wikipedia archive that specific shortcomings of articles are identified and then assess how well these shortcomings are addressed. We have illustrated our approach with a pilot study of Simple Wikipedia, a spin-off of the main Wikipedia that focuses on articles that are easy to read. We found that while the editors of this collection have a difficult time in keeping articles simple, the label "unsimple" that they employ to single out articles that are particularly problematic still provides a good entry-point in studying how this particular community manages to keep its encyclopedia readable.

With the Simple Wikipedia study as a "proof of concept", the time is now ripe to move to other, larger, Wikipedia projects and extend the study to indicators of shortcomings like "controversial" and "stub".

References

D. Anthony, S. W. Smith, and T. Williamson. The quality of open source production: Zealots and good samaritans in the case of wikipedia. Technical report, Darmouth College, Hannover, NH, April 2007.

S. L. Bryant, A. Forte, and A. Bruckman. Becoming wikipedian: transformation of participation in a collaborative online encyclopedia. In *GROUP '05: Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, pages 1–10, New York, NY, USA, 2005. ACM.

L. Buriol, C. Castillo, D. Donato, S. Leonardi, and S. Millozzi. Temporal Analysis of the Wikigraph. *Web Intelligence, 2006. WI 2006. IEEE/WIC/ACM International Conference*, pages 45–51, 2006.

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

J.-M. Dalle, F. Galia, and M. den Besten. Wikibugs – Treatment of defects in Wikipedia. *In Preparation.*

M. den Besten and J.-M. Dalle. Collaborative maintenance of collections - keep it simple: a companion for simple wikipedia? In *Proceedings of EURAM 2007*, May 2007.

P. Denning, J. Horning, D. Parnas, and L. Weinstein. Wikipedia risks. *Communications of the ACM*, 48(12):152–152, 2005.

P. Duguid. Limits of self-organization: Peer production and laws of quality. *First Monday*, 11(10), Oct. 2006.

W. Emigh and S. C. Herring. Collaborative authoring on the web: A genre analysis of online encyclopedias. In *HICSS '05: Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 4*, page 99.1, Washington, DC, USA, 2005. IEEE Computer Society.

R. Flesch. *How to Write Plain English*. Harper and Row, New York, NY, 1979.

J. Giles. Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901, 2005.

J. Grudin. Groupware and social dynamics: eight challenges for developers. *Communications of the ACM*, 37(1):92–105, 1994.

A. Kittur, E. H. Chi, B. A. Pendleton, B. Suh, and T. Mytkowicz. Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie. *Online:* http://www.viktoria.se/altchi/submissions/submission_edchi_1.pdf *(last retrieval May 18, 2007)*, 2007.

A. Kittur, B. Suh, B. A. Pendleton, and E. H. Chi. He says, she says: conflict and coordination in wikipedia. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 453–462, New York, NY, USA, 2007. ACM Press.

A. Krowne. Building a digital library the commons-based peer production way. *D-Lib Magazine*, 9(10), Oct. 2003.

A. Krowne and A. Bazaz. Authority models for collaborative authoring. *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference*, pages 18–24, 2004.

A. Lih. Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource. *Nature*, 2004, 2003.

J. Mateos Garcia and W. E. Steinmueller. Applying the open source development model to knowledge work. INK Open Source Research Working Paper 2, SPRU - Science and Technolgy Policy Research, University of Sussex, UK, January 2003.

A. Neus. Managing information quality in virtual communities of practice. In E. Pierce and R. Katz-Haas, editors, *Proceedings of the 6th International Conference on Information Quality*, 2001.

F. Ortega and J. M. G. Barahona. Quantitative analysis of thewikipedia community of users. In *WikiSym '07: Proceedings of the 2007 international symposium on Wikis*, pages 75–86, New York, NY, USA, 2007. ACM.

S. Pekkola. Designed for unanticipated use: common artefacts as design principle for CSCW applications. *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work*, pages 359–368, 2003.

J. Reagle. A Case of Mutual Aid: Wikipedia, Politeness, and Perspective Taking. *Online:* http://reagle.org/joseph/2004/agree/wikip-agree.html *(last retrieval May 10, 2007)*, 2004.

J. Reagle. Do as I do: leadership in the Wikipedia. *Online:* http://reagle.org/joseph/2005/ethno/leadership.html *(last retrieval October 15, 2007)*, 2005.

M. Robinson. Design for unanticipated use... *Proceedings of the Third Conference on Computer-Supported Cooperative Work*, pages 187–202, 1993.

C. Roth. Viable wikis: struggle for life in the wikisphere. In *WikiSym '07: Proceedings of the 2007 International Symposium on Wikis*, pages 119–124, New York, NY, USA, 2007. ACM.

L. Sanger. The early history of Nupedia and Wikipedia: A memoir. In *Open Sources 2.0*. O'Reilly, 2005.

L. Smith. Wholly new forms of encyclopedias: electronic knowledge in the form of hypertext. *Proceedings of the forty-fourth FID Congress. Helsinki, Finland*, pages 245–250, 1989.

B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser. Assessing information quality of a community-based encyclopedia. In *Proceedings of the 2005 International Conference on Information Quality (MIT IQ Conference)*, 2005.

F. Viégas, M. Wattenberg, and K. Dave. Studying cooperation and conflict between authors with history flow visualizations. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 575–582, 2004.

J. Voß. Measuring Wikipedia. In *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics*, 2005.

V. Zlatic, M. Božičevic, H. Štefančic, and M. Domazet. Wikipedias: Collaborative web-based encyclopedias as complex networks. *Physical Review E*, 74(1):16115, 2006.

Appendix

Table 3 Wikipedia measures by publication

Study	Measures
Lih 2003	Edits per article
	Unique editors per article
	Average article size over time
Viegas, Wattenberg, and Dave 2004	Article length over time
	Mass deletions
	Survival time of edits
Stvilia, Twidale, Smith, Gasser 2005	Num. of Anonymous User Edits
	Total Num. of Edits
	Num. of Registered User Edits
	Num.of Unique Editors
	Article length (in # of characters)
	Currency (a)
	Num. of Internal Links
	Num. of Reverts
	History Num. of External Links
	Article Median Revert Time
	Num. of Internal Broken Links
	Connectivity (b)
	Num. of Images
	Article Age
	Diversity (c)
	Information Noise(content) (d)
	Flesch readability
	Kincaid readability
	Article Admin. Edit Share
den Besten and Dalle 2007	Flesch readability
Kittur, Suh, Pendleton, and Chi 2007	Indirect work (f)
	Direct work
	Reverts

	Vandalism fixes
	Revisions
	Page length
	Unique editors
	Unique editors / revisions
	Links from other articles
	Links to other articles
	Anonymous edits (#, %)
	Administrator edits (#, %)
	Minor edits (#, %)
	Reverts (#, by unique editors)
Kittur, Chi, Pendelton, Suh, and Mytkowicz 2007	Percentage of total edits made by admins
	Number of edits per month made by admins
	Percentage of total edits made by bots
	Average number of edits per user per month
	Number of words added and removed per edit
Anthony, Smith, and Williamson 2007	Retention rate of contributions
	Number of contributions per author
	Registered / unregistered status of author
	Article size
	Contribution size

Figure 1 Wikipedia and its sister projects

Figure 2 Protection log for article 'Burt Reynolds'

Logs

Protection log | User: | Title: Burt Reynolds | Go

View (previous 50) (next 50) (20 | 50 | 100 | 250 | 500)

- 20:12, 25 September 2007 AndonicO (Talk | contribs) protected Burt Reynolds *(Edit war. [edit=sysop:move=sysop])*
- 21:11, 20 September 2007 JzG (Talk | contribs) unprotected Burt Reynolds *(Blocking warrior instead)*
- 21:09, 20 September 2007 JzG (Talk | contribs) protected Burt Reynolds *(unbelievably lame [edit=sysop:move=sysop] (expires 21:09, September 21, 2007 (UTC)))*

View (previous 50) (next 50) (20 | 50 | 100 | 250 | 500)

Figure 3 Wikipedia page history

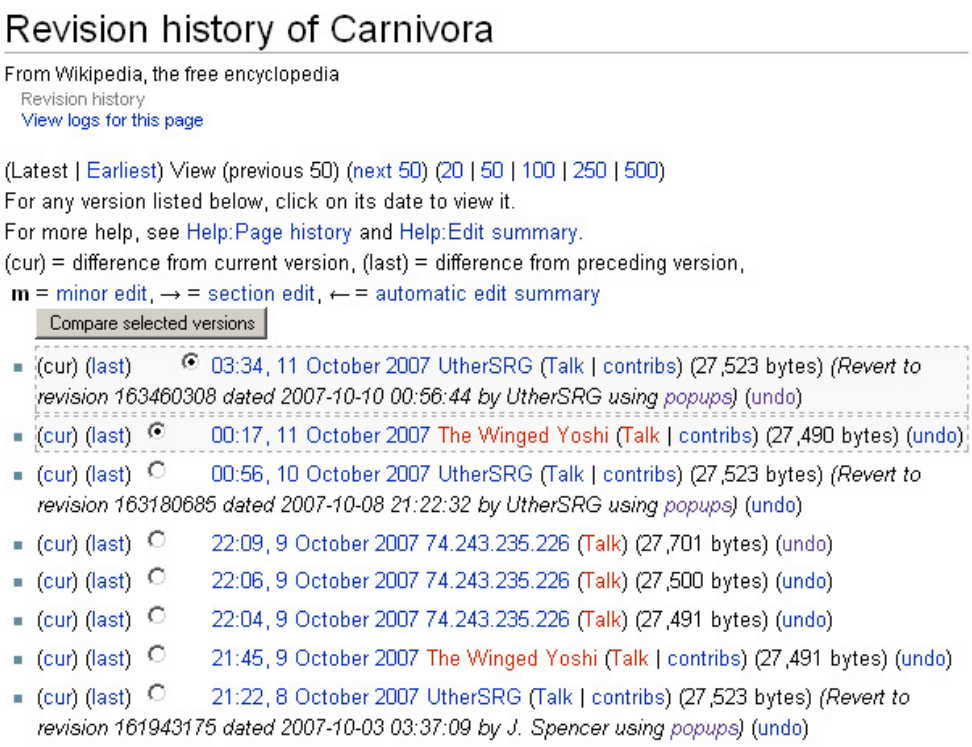

Figure 4 Page heading, showing locked status with an icon and links to edit, discussion and history

Figure 5 Performance at Project Level – Simple Wikipedia

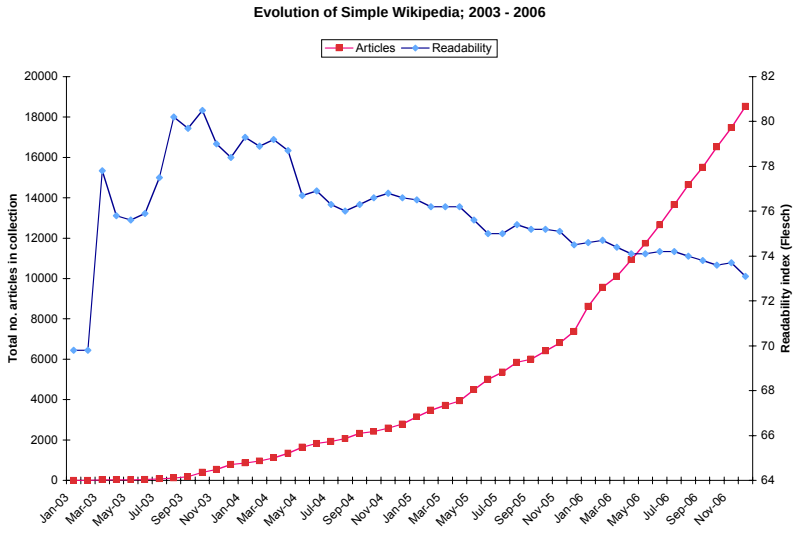

Figure 6 Variation in editor background over time

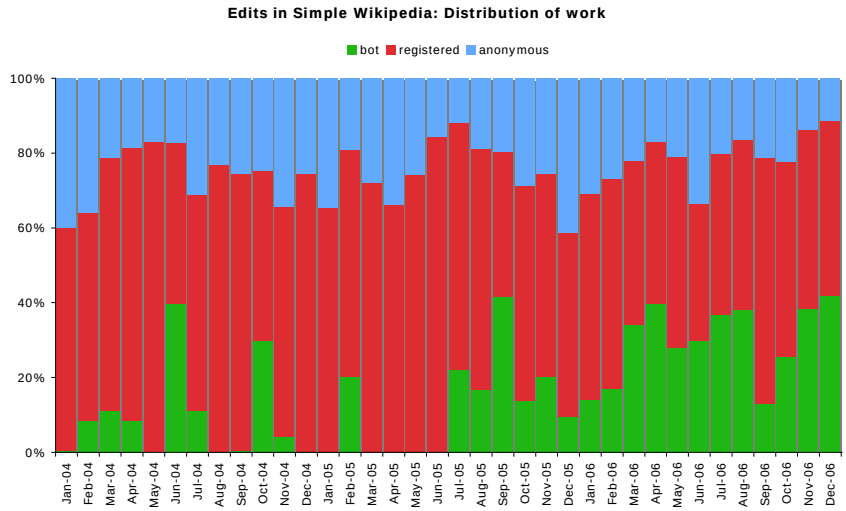

Figure 7 Treatment of "Propaganda" in Simple Wikipedia

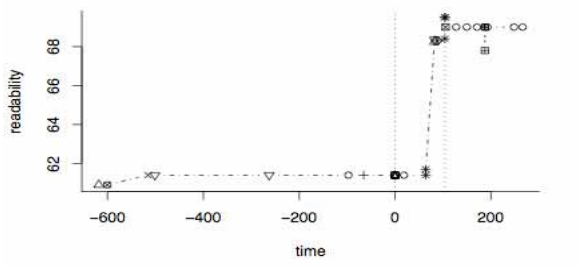

Figure 8 Development of Readability in Selected Articles