

Information Quality Work Organization in Wikipedia

BESIKI STVILIA¹, MICHAEL B. TWIDALE², LINDA C. SMITH² AND LES GASSER²

¹*College of Information, Florida State University, 269 Louis Shores Building, Tallahassee, FL 32306-2100, USA (bstvilia@fsu.edu)*

²*Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 501 E. Daniel Street, Champaign, IL 61820, USA; ({twidale, lcsmith, gasser}@uiuc.edu)*

Abstract. The paper provides an understanding of the information quality (IQ) assurance work organization in a large scale open collaborative encyclopedia – Wikipedia. A case study with a content analysis of Wikipedia discussion pages and other process artifacts was used. The study examines how individuals and the community as a whole understand IQ; how they reason about and negotiate the IQ of articles; which IQ assessment criteria are used; how those criteria are defined and evolved over time; which quality problem claims are made; how these claims are addressed and what IQ assurance strategies are used. The study shows that the Wikipedia community takes issues of quality very seriously. Although anyone can participate in editing articles, the results are carefully reviewed and discussed, in ways very similar to open source programming projects. What makes Wikipedia special as a resource is that the quality discussions and processes are strongly connected to the data itself. In most conventional databases these discussions and processes are divorced from the data. We believe it will be very interesting and productive to explore how, inspired by the success of Wikipedia, it might be possible to connect and record quality discussions and processes to the resultant data and so allow various ways for the users of this data to be directly involved in the quality improvement process.

Key words:

Collaborative Quality Control, Collaborative Content Creation, Information Quality, Distributed Collective Practices

1. Introduction

Although collaborative knowledge creation and organization have been in practice since biblical times, with scribes transcribing and at the same time often editing, updating, interpreting or reinterpreting original texts (McArthur 1986), open access large scale public collaborative content creation projects are relatively recent phenomena. The collaborative participation in the creation of the Oxford English Dictionary (Winchester 2003) is the best known example in a pre-computer age. Thousands of volunteers submitted examples of earliest usage from published literature that were collected, verified and incorporated into the mammoth project. The postal service was the major remote collaboration medium. As with more modern widely distributed collaborative activities including much open source software development, the OED project was easy to break down into separate activities. However it required central organization and authority to assemble, assess and reconstitute the final product from the contributions, many of which were inevitably incorrect, obsolete or duplicated. Substantial voluntary participation allowed for the discovery of first published usage of words, a task that is extremely difficult in terms of discovery of candidates, but where verification of candidates is much easier.

Such activities are potentially greatly facilitated by new Internet based content management technologies such as wikis¹. Ward Cunningham developed the first wiki engine and established the first wiki repository in 1995, as well as coining the word wiki². The key characteristic of wiki software is that it allows very low cost collective content creation by using a regular web browser and simple markup language. These features make wiki software a popular choice for knowledge creation projects where minimizing overhead in creating new or editing and accessing already existing content are the priorities. One such project has been the Wikipedia, the world's largest wiki and online encyclopedia, established in 2001.¹ Wikipedia is a community-based encyclopedia that has seen a huge growth both in size and public popularity. As of July 2nd 2006 the English Wikipedia alone had more than 1 million articles. In addition to the English language encyclopedia, Wikipedia has encyclopedia projects in more than 100 languages³ and remained within the top 20 highly used Websites in the world⁴. As a volunteer project Wikipedia needs substantial active participation and contribution from the general public to grow and improve. Therefore, it allows any user with Web access to start a new article or modify the existing ones with the least possible effort and commits and renders the contributions immediately on the user's screen.

To an outsider, the idea of Wikipedia sounds so bizarre and anarchic that it is entirely understandable that there is much skepticism about its quality. How can the entries be any good, can they be trusted, and won't they just get destroyed or maliciously modified by vandals? These are very reasonable questions, and have led to a number of studies (Emigh and Herring, 2005; Giles, 2005) to investigate them. Attention has been given to the quality of the articles, particularly in comparison to equivalents in conventional encyclopedias. These studies appear to indicate that the quality is reasonably good, or at the least, nowhere near as bad as might be expected from such a mechanism. This paper does not try to assess that relative quality directly but rather to understand the collaborative quality control processes used in Wikipedia and how they manage to achieve the kinds of success that others have noted. This paper is part of a larger body of work looking at Information Quality, and how it might be effectively and yet efficiently measured as a way to developing better IQ processes. As a result we have used a range of low cost measures to assess IQ processes, as part of an investigation of the power both of these processes and of the measures used (See Stvilia et al., 2005b).

The Wikipedia community have evolved a series of workflow processes that enable powerful mechanisms for improving information quality, while using relatively simple collaborative technology. It is certainly not the case that the quality control mechanisms were always substantially in place, and have merely been improved somewhat over time. Indeed, unlike many other collaborative activities, there was no pre-existing non-computer process to serve as a template for computerization. We believe that by understanding the IQ processes of the special case of Wikipedia, we can obtain insights that could be applied to other databases, including commercial ones that might exploit some features or variants of collaborative IQ improvement activities by users of those databases.

Wikipedia raises many questions in common with open source software: (1) Why do people bother to contribute? (2) How 'good' is the resultant product (or product-in-time given constant evolution)? (3) Why do people trust it and use it? (4) Why does the project not just disintegrate into anarchy? (5) How is the project organized, and how do the processes change over time? (See Benkler, 2002; Lerner and Tirole, 2004; Crowston and Scozzi, 2004 for reviews of these issues in the case of Free Open Source Software (FOSS)). FOSS has been around since the 1970s with the GNU project and the Free Software Foundation. Large scale, open access collaborative content creation projects like Wikipedia (the world's largest online encyclopedia), however, are a

relatively new phenomenon. The low barrier of entry, and, consequently the large size and the extreme diversity of its editorial community sharply distinguish Wikipedia from any FOSS project.

In this case study we focus on the IQ work organization in the Wikipedia articles. The paper explores how quality issues are discussed by the Wikipedia community, and how by an analysis of this quality and the creation processes used, we can begin to understand why the quality is better than might be expected.

2. Related literature

Neither the idea of “Wikification” of encyclopedia content nor its construction process is new. Before the existence of the Web, when discussing the possible impacts of hypertext technologies on the encyclopedia genre, Smith (1989) predicted that in electronic hypertext-based encyclopedias article sequence will not be linear and multiple paths will be provided; author and reader roles will be blurred and author contributions will be augmented by reader annotations; and article bibliographies will be partially replaced by direct hyperlinks to the source documents. What is new in Wikipedia, however, are the low barriers to participation, the sheer size, speed and geographical distribution of the knowledge construction process, and the ease of accessing this process, all enabled by wiki technology and the Web.

A number of studies looked at the quality of Wikipedia from different perspectives. Lih (2004) studied Wikipedia content construction and use processes from the perspective of participatory journalism. During the 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 most often cited.

Viegas, Wattenberg and Dave (2004) developed a tool - History Flow - to visualize the Wikipedia content evolution using article version histories. Based on the analysis of edit patterns the study 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.

Emigh and Herring (2005) compared two community-based encyclopedias (Wikipedia and Everything2) to the Columbia Encyclopedia on the formality of language used. The formality was assessed based on the frequencies of the parts of speech found to be characteristic of a formal language genre (Biber, 1988; Heylighen and Dewaele, 2002). After analyzing the part of speech frequencies, source and node variables from the total of 49 entries drawn from the encyclopedias and its discussion pages collection, the study concluded that the language of Wikipedia was as formal as that used in Columbia, and more formal than the language of the other community-based encyclopedia – Everything2.

Most of the above studies (with the exception of (Viegas, Wattenberg and Dave, 2004)) limited themselves to an analysis of quantitative features of the Wikipedia content as product and did not examine qualitatively the social context of work organization and communication processes in Wikipedia.

There is a well developed body of research on various aspects of IQ in management science and the database world. Most relevant to this research is the study by Strong, Lee and Wang (1997) who used a qualitative approach to collect and analyze data from a number of organizations. They identified a set of organizational IQ problem types which may arise due to aggregating information created in multiple contexts to support a particular task, or using information created in one context into a different context. However, Strong, Lee and Wang do not address IQ problems caused by *Many to Many* mappings. That is, when information created in many different contexts has to support the needs and IQ requirements of many different activities and perspectives at the same time. These kinds of situations require constant negotiation, compromises or consensus-building similar to what we observe in Wikipedia. Furthermore, there are other aspects of the overall social context of information creation and quality assurance besides task context, like culture, including religious, political and historical beliefs. A good example of cultural differences leading to quality compromise (purposeful degradation) is given by Bowker and Star (1994) who describe Japanese doctors reporting heart disease related deaths as strokes due to negative cultural bias against manual labor. There is also an economic context of IQ decision making. It is a widespread practice to purposefully degrade the quality of a product, including information, give it out for free to attract potential customers,

and then induce them to purchase a version with higher information quality (online images are a typical example).

The effects of these different contexts on IQ decision making often can be revealed by analyzing the instances of IQ negotiations and disputes. According to Strauss (1978) examining negotiation contexts can help the researcher to grasp subtle variations in processes, strategies and roles that otherwise could go unnoticed, and link them consistently and systematically to the main research topic. Fortunately, the discussion pages (a component of Wikipedia associated with articles) and other work coordination artifacts give us access to this kind of data, the analysis of which will be presented in the next sections.

Although sharing many features of open participation with OSS projects (e.g. Sack et al. 2006), Wikipedia has some important differences. Participants do not need to be as technologically sophisticated as the OSS hacker-programmers. Furthermore, participation is immediate – anyone can switch at any time from being a Wikipedia user reading an article, to an editor changing that same article. Individual articles have a relationship with the articles that they link to, but to a large extent, each article can be worked on separately by individuals or sub-groups, allowing a far finer granularity of participation than is possible in most software development. Also, the nature of wiki software enables the tight integration of several people working together on editing the same article, but equally enables conflicts to arise, sometimes leading to edit wars.

There are other socio-technical mechanisms that have been employed to provide a form of collaborative information quality control. Most of these are concerned with the selection of better entries over others, rather than fixing or deleting the low quality entries. Examples include Google's Page Rank algorithm (Brin and Page, 1998), using links as proxies for individual quality votes for a web page, and Slashdot's⁵ moderation process which involves direct participation by certain users to assign a quality score to posted comments. The aggregated results of these scores allow readers to choose to only view posts above a certain quality level.

This paper does not provide a full exploration of all the CSCW aspects of Wikipedia. It would be interesting to see how Wikipedia might be analyzed with the concepts and frameworks used with such great success by CSCW researchers in a host of workplace and other settings. There are clearly major issues of cooperation and conflict, coordination mechanisms, awareness,

articulation, workflows, and evolving processes all mediated to a large extent through a single piece of software and used by a very large heterogeneous community. We focus on the evolution of IQ processes and how they work. However it is worth noting in passing that Wikipedia does manage to address successfully the factors identified by CSCW researchers as important for ensuring success or at least avoiding failure. These include the common artifact criteria of predictability and double level language (Robinson, 1993; Pekkola, 2003) and Grudin's criteria of the importance of addressing disparities in work and benefit, and processes for addressing issues of critical mass, exception handling, unobtrusive accessibility and smooth adoption (Grudin 1994). Additionally there are many mechanisms for allowing the adaptation and evolution of the work processes both for local exceptions and for larger scale evolution of the work processes themselves, by various mechanisms supporting articulation work (Schmidt and Bannon, 1993) and awareness. In particular the wiki features of a history mechanism, notification features and the tight integration of a discussion page with each entry, as well as other spaces for setting policy and norms (and their associated history and discussion pages) play a central role in facilitating these core CSCW factors.

With respect to our focus on information quality, we are not aware of much discussion in the CSCW literature of the collaborative detection and correction of data or process errors, or indeed the degree of focus on errors in general that one sees in other fields such as HCI, human factors, aviation etc. Other than Trepess and Stockman (1999), the only major exception is the work done on studies of programming practice where the seemingly inevitable presence of bugs forces a consideration of detecting and fixing errors. In the rest of the CSCW work practice literature, although there may not be much focus on the occurrence of errors, there is a substantial concern with their avoidance, and with contested interpretation and use of information, including both data and plans, formally encoded, written, or implicit. Problems with the quality and trustworthiness of the available data are widely noted, but the focus is more on detection, judgment, selection, avoidance, or compensation for the presence of data errors rather than on ways that people correct them. For example, Ehrlich and Cash (1999) note how judgments of information quality can be a central part of collaborative work, but they consider assessment rather than improvement activities. The people in their study focused on choosing between sources and how much to trust the sources, rather than being in a position to correct those sources.

3. Research setup

We used the case study method (Yin, 1988) with a number of qualitative and quantitative techniques such as content analysis to gain an understanding of the IQ variable structure, dynamics and organizational issues of IQ assurance in Wikipedia.

The 2005/03/09 and 2006/07/02 dumps of the English Wikipedia⁶ were used. The qualified population sizes were 500,623 and 1,257,152 articles. After removing redirects to other support pages, samples of 1,000 articles were randomly selected. To gain a better understanding of the IQ assessment model used by the community, we collected the Featured Article sets –articles considered as Wikipedia’s best that have been featured⁷ on its front page by the time the dumps were made (236 and 641 accordingly)⁸.

We also harvested the edit histories and discussion pages of all the articles selected. Not all Wikipedia articles have active discussion pages. For instance, in the random sample of 2005/03/09 dump, only 155 articles out of 1,000 had a non-empty discussion page. Some discussion pages are just stubs, only containing automatically inserted template text. To focus on those discussion pages that contained some meaningful content, only pages longer than 100 characters were considered, decreasing the total number to 128 (15%). The situation was substantially different with the Featured set. There, 235 articles out of 236 (99.6%) had discussion pages that were longer than 100 characters. A subset of 30 discussion pages each longer than 100 characters were randomly selected from each of these two sets of March 2005 dump and content analysis (Bailey, 1994) applied to look for IQ problem types and IQ assurance and negotiation patterns.

To identify the motives and incentives that might drive editors to make contributions, we randomly selected and content analyzed 100 profiles/user pages from the pool of the registered Wikipedia users who edited any of the articles from the Random set. To qualify for the analysis, the user profile had to have some content in it beyond the Wikipedia user template/stub.

4. Wikipedia IQ assurance context

This section looks at some of the main components of the IQ assurance context of Wikipedia such as work coordination and support artifacts, roles and processes (see Figure 1). Each article in Wikipedia is accompanied by a discussion page and both the article and discussion page have their update histories recorded. There are also several important processes that shape the IQ profile of Wikipedia. The Article for Deletion (AfD)⁹ and Speedy Deletion¹⁰ processes are intended to enforce the minimum quality requirements for an article to remain in the collection. The Featured article selection¹¹ and Featured status removal¹² processes on the other hand are intended for defining the targeted or desired level of IQ for Wikipedia articles. The last two processes are enacted when the Accuracy¹³ and/or Neutrality – Neutral Point of View (NPOV)¹⁴ - of the article is disputed and needs to be resolved. The following subsections discuss some of these roles, processes and artifacts in more detail.

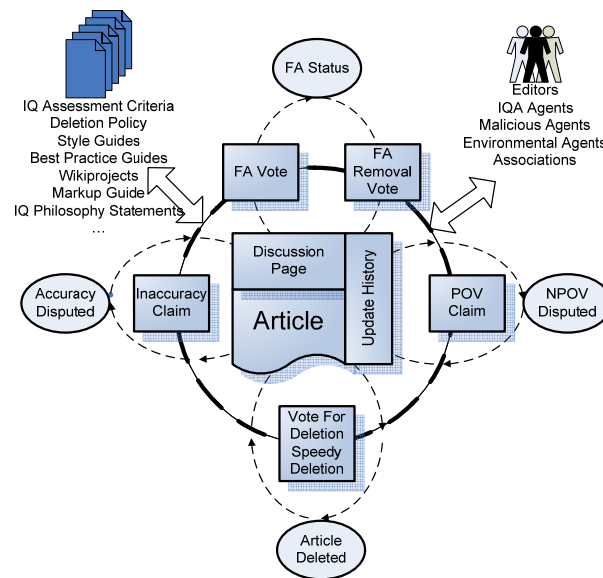


Figure 1: Wikipedia IQ assurance context

4.1 WIKIPEDIA ROLES AND TEAMS

The Wikipedia context is rich with different roles. Trying to understand those roles and the processes they operate in can help us to understand some of the sources of IQ variance in Wikipedia, and consequently identify the sources of relevant metadata for IQ metrics. There are

at least four distinct roles in the Wikipedia content construction process: (1) Editors: agents that contribute/add new content to the article; (2) Information Quality Assurance (IQA) agents, agents that affect article quality: monitor changes made; revert vandalisms; enhance the IQ of the article through minor edits; enhance the IQ of the collection through enforcing IQ norms and criteria across the collection; enhance the IQ of the collection by developing support infrastructure in the form of genre-specific templates and style guides; fostering collaboration and maintaining order in article editorial groups; (3) Malicious agents that purposefully degrade the article quality; (4) Environmental agents that change the representational IQ of articles through changes in the real world states. While mostly degrading IQ, in a few instances Environmental agents may enhance the article's IQ by aligning the real world state with the information contained in an article.

Wikipedia is not a full blown content management system with finely graded user rights/permission management. However, it distinguishes between three groups of accounts: (1) Registered Users – identified and tracked by their login name; (2) Anonymous Users – identified and tracked by the Internet Protocol (IP) address they log on from; (3) Administrators – the same as Registered Users but with special system permissions / privileges. Each of these groups may take different roles at different times. Registered Users can be Editors, IQA or Malicious agents. Likewise, Anonymous Users can act maliciously or can be valuable Editors and IQA agents. Although in theory an Administrator can degrade the article IQ intentionally or unintentionally, one mostly expects “good” edits from them – to play Editor or IQA agent roles.

Wikipedia posts the list of Administrators along with a list of simple agents called Bots. Bots are run by Wikipedia users for automating some simple IQA tasks such as exporting content from public domain databases, spellchecking or identifying vandalized entries based on a curse word list¹⁵. As of March 2005 there were around 431 Wikipedia users with administrative privileges¹⁶ and the community employed around 50 Bots¹⁷. By July 2006 these numbers have more than doubled (~960, ~108).

Thus, the IQ of a Wikipedia article is constructed through a highly dynamic complex interaction among the members of the above three groups engaged in different activities and playing different roles. The footprint or trace of this interaction is logged and can be obtained from the article's Discussion and Edit History objects which we will discuss later.

The analysis of the registered user homepages showed that most of the contributions were motivated by user interests in specific areas often shaped by their organizational and ethnic affiliations, hobbies, professional experiences and beliefs. Still, in quite a few instances the users identified the general desire of fixing IQ problems and enhancing the overall quality of Wikipedia as a motivation for making contributions:

I like to edit randomly if I see something I feel like improving. I like creating new redirects to help people find pages as well as moving pages when appropriate – user X.

My mission on Wikipedia comes from a habit I have in places like bookstores. My habit is that when I see something that needs fixing (books out of order) I fix it. – user Y.

Removing the editor/user distinctions, allowing individuals to self-select themselves for tasks without any prior vetting increases the variance of perceptions of what constitutes good IQ. Many of the editors had strong and often conflicting opinions on what constituted a good encyclopedia article, or an encyclopedia as a whole, and how to achieve it, often expressed on their user pages or in IQ negotiations and votes.

[M]y desire to see Wikipedia being the MOST complete encyclopedia in this world. I plan to do my part to achieve this goal. – user X.

I'm probably older than the overwhelming majority of Wikipedians, I actually remember some of the 1950s popular songs and baseball players I've written articles about. And this, of course, colors my feelings about some of the articles. Not only do I write articles about a lot of older topics, but I try to keep coverage of such topics even by others at a high level. And this has led to some disputes with other Wikipedians who don't seem to think a 50-year-old song, or some other such thing, deserves an article. – user Y.

The notability and encyclopedic-ness of a topic should dictate whether it belongs on Wikipedia, not the quality of the current content. – user Z.

Likeminded Wikipedia members might form associations to achieve desired outcomes in Wikipedia decision and policy making processes. Interestingly, the two most prominent Wikipedia associations – the Inclusionists¹⁸ and the Deletionists¹⁹ associations - have been founded by Wikipedia administrators. While Inclusionists insist that any knowledge is important to at least someone and all knowledge should be included in Wikipedia, Deletionists in general are against including obscure, insignificant and/or short and poorly written articles. Deletionists also advocate the adoption and strict application of specific standards and policies on the types and quality of admissible articles. The statements “Wikipedia is not paper” and “Wikipedia is not Google” sum up the two editorial philosophies.

Similar to the Wikipedia roles, the association space²⁰ and association memberships in Wikipedia are dynamic and informal. The diversity of member IQ preferences and low cost of forming or switching associations may encourage schism in an existing association or evolution of new groups.

“Exclusionism is a form of deletionism, specific to material in articles... The POV exclusionist is one who for seeks to keep material out of the article for sake of a POV, whereas a POV deletionist seeks to remove certain Wikipedia articles altogether.”

The analysis of the Random sample edit histories showed that by March 2005 Wikipedia administrators (~7% of the editor population of the Random sample) had more than a quarter (~26%) of the total number of edits. For the Featured set the administrator share was much higher - ~1.6% of the editors made ~23% of the edits. In the July 2006 Random sample the share of administrator edits was reduced substantially – the administrators (~5%) did only ~14% of the edits. The administrator share still remained high for the Featured articles – 20% of the edits was done by only 0.6% of the editors (see Table 1). In addition, various support pages and constructs, such as templates, have been created and maintained mostly by administrators. Furthermore, the analysis of the IQ votes showed that the member participation in the votes may exhibit a power law pattern – a few editors vote in almost every IQ vote while many vote in a few or do not vote at all. Based on this statistic and the success of the Wikipedia project one may suggest that the presence of heavy contributors or power editors is essential for holding the community together and fostering the application of common IQ standards and conventions. The majority of Wikipedia editors probably have never read the Wikipedia IQ criteria or style guides. Power editors, on the other hand, who in many cases, but not always, are administrators, are familiar with the IQ standards and norms. One may suggest that those power editors are the ones who follow the standards in their edits, promote and enforce their use in the community. We found in the IQ discussions that in certain cases even groups working on FA candidate articles were not familiar with specific FA criteria and had only a general notion of what high information quality in Wikipedia meant.

(Oops, my bad... apparently, a picture is, for some reason, necessary. Again, sorry, I'm new. If that little thing is all that is lacking, I'll keep an eye out for an appropriate picture or two to make into a montage that is appropriate for the article. I'm not promising anything though...)

At the same time, the reduction of the administrator share of edits and editors identified by the samples may point to a potential scalability problem for Wikipedia with regard to quality. It can be suggested that over time, the administrators may get overpowered by the faster growing

general editor population and may not be able to maintain the same level of presence and IQ control. Interestingly, in contrast to the Random samples, the reduction of the administrator edit share for the Featured sets was not that high (see Table 1).

| Measures | March 9, 2005 Dump | | July 2, 2006 Dump | |
|--|--------------------|----------|-------------------|----------|
| | Random | Featured | Random | Featured |
| <i># of edits*</i> | 12,926 | 104,866 | 29,150 | 551,098 |
| <i># of edits by admins</i> | 3,358 | 24,435 | 3,993 | 111,990 |
| <i># of administrators edited the sample articles</i> | 289 | 336 | 552 | 781 |
| <i># of editors</i> | 4,261 | 21,467 | 11,068 | 121,639 |
| <i>article network density (articles as vertices and editors as edges)</i> | 0.04 | 0.90 | 0.03 | 0.98 |

Table 1: Distribution of edits in the samples (* does not include the edits made by bots)

4.2 WORK COORDINATION ARTIFACTS

Wiki software does include an interdependence mechanism. By allowing the disputing sides to obliterate each other's contributions easily, a wiki makes the sides interdependent in achieving their goals and perhaps surprisingly may encourage more consensus building rather than confrontation. In addition, it is well understood in CSCW research that effective articulation of work is an essential factor for successful cooperation (Schmidt and Bannon, 1993). Wikipedia successfully employs a number of work coordination artifacts and processes, which we will analyze next.

4.2.1 DISCUSSION PAGES

A discussion page is an auxiliary wiki object which accompanies a Wikipedia article and, as the name indicates, is intended largely for the purposes of communication among the members of the Wikipedia community when constructing and maintaining the article content. Technically, a discussion page is the same wiki object as an article. Unless locked by Wikipedia administrators it can be updated by anyone. Updates to the article are logged and can be visualized through a history object. The difference between the article and its discussion page lies only in the role assigned to a discussion page in the Wikipedia infrastructure. It is a coordinative artifact

(Schmidt and Simone, 1996), which helps to negotiate and align member perspectives on the content and quality of the article.

Discussion pages are part of Wikipedia's overall work coordination and support architecture, which also includes project management pages called Wikiprojects²¹, style manuals, best practices guides, templates and other work coordination artifacts. Discussion pages are routinely used by IQA agents such as Administrators to communicate different kinds of management information - providing feedback on quality, giving notices and warnings on the article's current status, encouraging cross article communication, and general coordination.

Furthermore, we found that an article's discussion page is often used by those outside the article's contributor community. These outsiders use it for asking the community questions related to the article's topic, and sometimes even soliciting assistance for other Wikipedia articles or projects outside of Wikipedia.

Could anyone who is geologically inclined please have a look at Andes. There is currently a warning saying that the section on geology is pre-plate tectonics, so could anyone who knows about such things please have a squiz.

Traditionally, an encyclopedia article genre is considered to be a concise introduction to a subject in the form of a compressed summary, focusing a reader's attention on the main points and giving references to outside sources for more in-depth information. Following this convention, Wikipedia articles also may not include detailed explanations or descriptions of the concepts and theories mentioned in the article, sometimes making it difficult for a general reader to fully comprehend. However, this kind of information was often explicated in the process of negotiation, reflection on edits made, or when answering questions. Accumulated in discussion pages, such information turned those discussion pages into FAQ style knowledge bases – complements to information presented in articles and a great resource for regular users.

Your responses to my questions have been very informative. If you get the chance, I hope you'll consider incorporating some of this material into the article. Although we're enjoined to be bold in editing, I confess I feel out of my depth here.

Having well developed work articulation artifacts in the form of discussion pages helps in establishing a sense of community and negotiating a merit-based social order. It helps to establish norms and conventions of communication (Orlikowski and Yates, 1994), and to introduce newcomers to those norms and the subject in general. The Featured set had significantly better developed discussion pages than the Random set. Not only were the median lengths of the discussion pages of the Featured articles more than ten times larger, but they were

much better organized and more readable based on their Flesch readability scores (Gunning, 1952) (see Table 2). The presence of a large and well organized footprint in the form of a discussion page can be a sign of the existence of a strong interest and a well organized editorial group around the article. The strong interest, however, may not always be translated into high quality. If there is a consensus within the editorial group, a large size discussion page can be correlated with the high quality of the article's content. Alternatively, if consensus has not been reached, then a large size discussion page can be just a result of longstanding arguments and disputes.

| Article Set | # of Pages | Flesch (higher is better) | Length (characters) | % of Sample |
|-------------------------------|------------|------------------------------|---------------------|-------------|
| <i>Featured (March, 2005)</i> | 235 | 35.9 | 11,841 | 99.6 |
| <i>Featured (July, 2006)</i> | 641 | 43.4 | 17,360 | 100 |
| <i>Random (March, 2005)</i> | 128 | 30.4 | 870 | 12.8 |
| <i>Random (July, 2006)</i> | 210 | 40.9 | 345 | 21 |

Table 2: The median readability measurements of the article discussion pages (empty pages and stubs are not included)

4.2.2 ARTICLE EDIT HISTORIES

An article edit history is a wiki object that contains the log of three-element metadata entries for each instance of the article edit. The edit metadata elements contain the following information: (1) the date and time of the edit; (2) the name of the user who made the edit, or if the user is not logged on with a Wikipedia registered user name, the Internet Protocol (IP) address the edit is made from; (3) optionally, a comment provided by the person doing the edit clarifying the edit purpose. As a result, the history object can be a source of the following meta information about the article: (1) Age; (2) Currency – when the article was last updated; (3) The number of times the article has been edited; (4) The names or IPs of the article editors; (5) The types of the edits, such as reverts (returning the article to an earlier state/version), minor edits, copyediting, etc.

The last two kinds of information may not be complete and unambiguous. It is not necessary for an individual to be logged on or even registered to make an edit. In addition, the same

individual can be registered and make edits with more than one name. One can easily extract the set of unique registered user names and IP addresses for anonymous users from the article's edit history. However, this set may not be mapped one to one into the actual set of the article's editors, and can serve only as its approximation. Likewise, edits are often made without editors filling out comment fields, or filling them out with misleading information. As the conventions of commenting edits are not followed consistently, automatic coding of the comments for analysis cannot be fully accurate either. Nonetheless, these two elements of history entries can still provide valuable information about the social structure and dynamics of the article's content creation (Stvilia et al., 2005b).

4.3 IQ ASSESSMENT CRITERIA AND PROCESSES

The metaphor used in the Wikipedia quality assurance model is of “controlled” selection or evolution. The Wikipedia IQ assurance model has two sets of mechanisms for setting and enforcing the community wide criteria for IQ and controlling the entire IQ profile of the collection:

- Featured article status assignment and removal processes set a target or desired level of quality
- Article deletion and speedy deletion processes enforce a minimum level of quality - the lowest critical boundary of quality for an article to remain in the collection (see Figure 1).

It is important to note, however, that a relatively insignificant number of its articles have been formally evaluated against the Featured Article Criteria²². The number of Featured Articles, the articles whose IQ has been confirmed to be meeting the criteria by community votes, was only 236 (<0.05%) out of the total of 500,623 articles in the Wikipedia's collection as of March 2005. By July 2 2006 this number has grown to 641 articles (~0.05%) that have been already featured on the Wikipedia main page. The share of the Featured Articles in the overall collection, however, has remained the same – ~0.05%. In addition, as our qualitative analysis of the social context of IQ negotiation and control in Wikipedia showed, besides the FA criteria there are a number of additional IQ norms used by the community when evaluating article quality. One such criterion is the norm of an article size not exceeding 32k to ensure a certain level of accessibility over slow dialup modem lines. The information on what percentage of new articles or the

existing articles have been evaluated systematically against the deletion criteria was not accessible to us.

We found, however, that the community did not orient itself on just maintaining the IQ status quo but actively sought the improvement of the overall IQ profile of the collection by deploying various socio-technical mechanisms of IQ assurance and generating an evolutionary pressure by slowly moving up the target and baseline levels of article quality.

4.3.1 FEATURED ARTICLE CRITERIA

Featured Articles²³ (FA) are those declared by Wikipedia's community to be its best. Articles can be nominated as candidates for FA status by individuals or a group. Once nominated, the candidates go through a peer review process to check if they meet the Wikipedia featured article criteria²⁴. According to the history log of the FA directory, the FA process began around April 2002. However, at that time featured article candidates did not go through a peer review process. The directory did not reference any quality assessment criteria except "brilliant prose". As a result those early non-peer reviewed featured articles have been referred to ironically by the current Wikipedia community as "brilliant prose" articles.

We think the following Wikipedia pages are pretty good. This is a selected list--since there are thousands of pages on Wikipedia, we couldn't possibly keep track of *all* the brilliant prose here! But if you come across a particularly impressive page, why not add it to the list as your way of saying "Thanks, good job"?²⁵

It needs to be noted, however, that at the same time Wikipedia already had "the Perfect article" page containing a quite well developed description of what makes a perfect encyclopedia article. It referenced IQ criteria such as clarity, completeness, neutrality, verifiability and compliance to genre.²⁶

According to Marc Pellegrini, the Wikipedia Featured Article Director, the first ad-hoc formal FA quality assessment guideline was developed by the user Eloquence on 7 March 2004²⁷. A separate page defining what can be considered as a featured article was formed by Pellegrini himself on 20 April 2004²⁸ (personal communication, August 1, 2005). Eloquence's guideline included quality criteria - comprehensive, factually accurate, well-written. It also referenced the Perfect article page as an additional quality guide. The page lists eight Featured Article quality assessment criteria: (1) Comprehensive; (2) Accurate and verifiable by including references; (3)

Stable - not changing often; (4) Well-written; (5) Uncontroversial – using neutral language and not having an ongoing edit war; (6) Compliance with Wikipedia standards and project guides; (7) Having appropriate images with acceptable copyright status; and (8) Having appropriate length, using summary style and focusing on the main topic. Figure 2 maps the quality assessment dimensions from the printed encyclopedia quality assessment discussion from (Crawford, 2001) and ones from a generic IQ assessment framework developed earlier (Gasser and Stvilia, 2001) into the FA criteria.

Although the Wikipedia IQ model lists Stable, Uncontroversial and Verifiable as important quality dimensions when assessing an FA candidate's quality, these dimensions do not appear in the Crawford framework. It could be that in the Crawford framework they are taken for granted. The content of a printed encyclopedia article is generally fixed until the next update cycle, which is not the case with Wikipedia where anyone, including malicious agents, can make edits any time. Likewise, the FA criteria do not include the Authority and Currency dimensions. While for multivolume general printed encyclopedia a yearly revision can be a "Herculean and economically infeasible task" (Crawford, 2001), Currency does not seem to be considered as a major quality indicator in Wikipedia where the cost of update is very low and anyone is allowed to do it. Both the Crawford and FA IQ assessment models contain Accuracy as an IQ criterion.

The Wikipedia community does utilize a reputation mechanism, even though it is not formalized in its policies, as found by Viegas, Wattenberg and Dave (2004); some Wikipedia users taking on IQA agent roles said that they used authorship information when monitoring edits made to Wikipedia articles, being more suspicious of edit actions by anonymous or new users than those by users with already established records of valuable contributions. The Wikipedia community also insists on Verifiability to enable peer control of the quality of user edits.

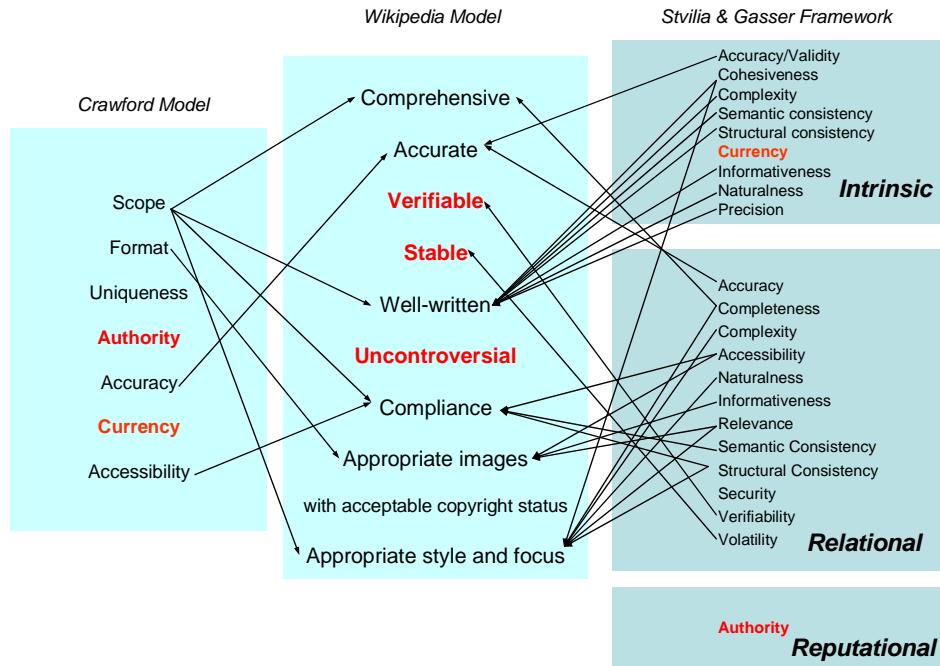


Figure 2: Wikipedia IQ assessment model mapping into the Stvilia & Gasser framework (Stvilia and Gasser, 2001) and the Crawford model (Crawford, 2001)

Wikipedia puts trust not in a single expert author or group, but in the collective knowledge of a large-scale distributed community hoping that, when creating encyclopedias, just as for creating OSS, "given enough eyeballs all bugs are shallow" (Raymond, 1998). Cosley et al. (2005) found that peer oversight was as good as experts in maintaining the quality of an information collection. Hence, one of the IQ measures can be the number of "eyeballs" – the number of distinct editors. Again this is an indirect measure that happens to be easy to measure. The real number of eyeballs is the number of people reading the article. The number of people bothering to make a change is obviously much smaller and probably more interesting and maybe correlating with the real number of eyeballs.

The Stability criterion assessing how frequently the article undergoes significant changes, as well as the requirement of having references, were not included in the earlier versions of the FA candidate assessment guide and were added only in September 2004. To reduce the IQ variation of Wikipedia content and make the IQ assessment process more consistent and systematic the Wikipedia community is rapidly developing sets of style manuals and genre or topic specific

information organization and management pages called WikiProjects. For comparison, the April 2004 version of the FA criteria contains reference only to 5 guides while the May 2005 version lists 11 guides. The July 2006 version references 17 guides, including informal guides from three active participants of the Featured article selection and removal processes. Hence, the differences between the two IQ assessment models are largely caused by the pragmatics of their immediate social context of use. As the Wikipedia social context changes and gets more complex, so does its IQ assurance and assessment infrastructure.

As a part of the continuous IQ feedback process one can also nominate an already featured article for removal of a featured status and have the community vote on it²⁹. According to the Wikipedia logs³⁰ around 120 articles were nominated from July 2004 to May 2005 as candidates for removal and the community voted out 2/3 of them (80 articles).

4.3.2 DELETION POLICY

The community's IQ reasoning is also encoded in the judgments on the pages that are considered as the Wikipedia's worst or not worthy to remain in the collection— deleted articles, stubs, templates, categories and other miscellaneous pages. According to the Wikipedia Deletion Policy (DP) page³¹ several hundred articles are deleted daily at Wikipedia. The reasons for deletion could be *low quality* (absence of any useful or meaningful content, vandalism, redundant entries), *low criticality* (lack of importance or notability), *copyright violations* or content that is *not encyclopedic* such as advertising, self-promotion, original research, or nonWikipedia related user pages.

Pages, including articles, can be nominated for deletion by any Wikipedia user. However, they can be deleted or alternatively restored only by Wikipedia administrators. Once an article is listed at the Articles for Deletion (AfD) (former Votes for Deletion (VfD)) forum, the nomination remains there for a number of days to give the members of the community a chance to vote on it. Decisions on what actions to take (keep, delete, cleanup, merge/redirect or transwiki³²) are made on the basis of consensus³³. A similar procedure applies when someone nominates to undelete an already deleted article³⁴.

To make the weeding process more efficient, however, in certain instances, administrators, guided by the Wikipedia Criteria for Speedy Deletion (CfSD)³⁵, can delete pages without

seeking the Community's consensus at the AfD forum. According to the CfSD, the following kinds of articles can fall under the Speedy Deletion category: (1) very short articles providing little or no context; (2) foreign language articles that already exist at another Wikimedia project; (3) any article whose content consists only of an external link; (4) any article which consists only of attempts to correspond with the person or group named by its title; (5) any article that has been discussed at Articles for Deletion and successfully transwikified; (6) short articles that serve no purpose but to disparage their subject; (7) an article about a real person that does not assert that person's importance or significance. Hence, the criticality/notability of the subject or topic discussed and its appropriateness to the general encyclopedia genre have been two main factors determining whether a particular article needs to be deleted or not. Quality, on the other hand, unless it is unrecoverable nonsense, has not been emphasized much in the DP. It is assumed the quality of new articles will be low at the beginning.

Interestingly, *criticality* or *notability* is not included as a selection factor in the earliest accessible to us version of the DP³⁶. Its deletion criteria are less demanding and precise. The earliest version of DP does not make any reference to the community consensus either, placing the burden of decision making only on three administrators. Thus, similar to what we've observed in the case of the FA criteria, as the Wikipedia collection and community have been growing, the DP criteria have been evolving too, expanding and getting formalized. The AfD and CfSD mechanisms have been used to create an evolutionary pressure on the collection to reduce variance in the behavior of the editorial community, and ultimately, the variance in the IQ of Wikipedia articles.

4.3.3 ACCURACY AND NPOV DISPUTED

The last two processes from Figure 1 – Accuracy Disputed³⁷ and NPOV (Neutral Point of View) Disputed³⁸ are enacted when the *Accuracy* and/or *Neutrality* of the article is disputed and cannot be resolved promptly within the article's editorial group. The person who makes a claim inserts an appropriate tag at the beginning of the article indicating that the article's quality is being challenged. Note that these tags may serve dual purposes – on the one hand, they indicate to a Wikipedia user that there might be a quality problem with the article. On the other hand, it

invites IQA agents, who are often Wikipedia administrators, to come to help the editorial group to resolve the problem.

5. Types of quality problems and activities affected

Quality is ultimately a social construct (Strong et al. 1997). Hence, to evaluate it meaningfully there have to be socially and culturally justified assessment criteria and norms to compare with. This implies the existence of some social order or consensus on what constitutes good IQ, which according to Strauss (1978) is always a negotiated order. Strauss suggests that examining negotiation contexts can help the researcher to grasp subtle variations in processes, strategies and roles that otherwise could go unnoticed, and link them consistently and systematically to the main research topic. In this particular research we were mostly interested in negotiations as a means for identifying the types of IQ problems and dimensions that were considered important by the community. These negotiations were combined in our analysis with other processes such as user self-reported descriptions and justifications of actions taken. This helped us understand the tradeoffs made between those dimensions and their possible effects on the overall content of articles.

Wikipedia is an open system. Hence, the sources of IQ problems in Wikipedia were both changes in *articles*, and changes in their *underlying entities* and *context*. The process of an information object's IQ change can be *passive* or *indirect* caused by changes in the underlying entity and context: culture, socio-technical structures and domain knowledge. In the case of Wikipedia, for instance, it could be changes in the membership of an editorial group, changes in the FA criteria, removal or modification of the articles a given article references to or is referred to by. The context can also be changed *actively* to affect the quality of the information object. New sources can be introduced or existing ones modified with an intention to support or contradict the information presented by the object (Garfinkel, 1967; Gracy, 2002). There can be *active/direct* quality degradation through malicious corruption or removal of the information entity. Quality degradation actions may not be necessarily malicious, however. We observed in Wikipedia how administrators often had to remove edit access to an article – reduce its accessibility – to protect it from greater quality degradation caused by edit wars or frequent

vandalisms. Information also can be abridged (reduced in completeness) to meet the needs of certain audiences or uses, or alternatively the conventions of a certain genre such as an encyclopedia.

The content analysis of 60 discussion pages from the Featured and Random sets identified 12 types of IQ problems the Wikipedia users pointed to (See Table 3). Problem instances were tallied as many times as they occurred in discussions. It is clear from Table 3 that the distribution of IQ problem types claimed or disputed in the discussion pages was not uniform. We use the word ‘claim’ to refer to a user’s complaint about an IQ problem that may not be true or agreed upon by the community. IQ assessments are often relative to a particular community’s cultural and knowledge structures. If the user is not aligned with those structures, his or her claim of the existence of an IQ problem may not be shared by the rest of the community and get rejected.

The analysis showed that often, when claiming or disputing IQ problems, users were mindful of the existence of tradeoffs among various quality dimensions. They sought a balance as a social group among those dimensions through the process of negotiation, logical analysis and sensemaking of their own and each other’s actions (Weick, 1995). We identified the instances of the following tradeoffs from user disputes and self reported reasoning of the current state or retrospective sensemaking of the edits they made in articles: (1) Completeness vs. Accessibility; (2) Accuracy vs. Accessibility; (3) Completeness vs. Cohesiveness; (4) Accessibility vs. Complexity; (5) Completeness vs. Consistency (when the same statement may not be consistently interpreted by different cultural or social groups or institutions); (6) Accessibility vs. Consistency; (7) Completeness vs. Complexity; (8) Volatility vs. Accessibility.

Earlier we suggested that a Criticality or Importance measure of an information object to a given community could be orthogonal to its IQ dimensions when doing selection or making a decision about IQ assurance resource allocation (Gasser and Stvilia, 2003). Examples confirming this proposition were found in the current study. In addition, the community was clearly conscious about constant tradeoffs between quality and cost (Stvilia et al., 2005a).

| Problem Types | # in FA | # in RA | Caused by | Action taken or suggested | Dimensions from the Framework |
|------------------------|----------------|----------------|--|--|--|
| <i>Accessibility</i> | 6 | 3 | <ul style="list-style-type: none"> ▪ Language barrier ▪ Poor organization ▪ Policy restrictions imposed by copyrights, Wikipedia internal policies and automation scripts | reorganize, duplicate, remove, translate, split, join, rearrange | 1. Accessibility |
| <i>Accuracy</i> | 54 | 53 | <ul style="list-style-type: none"> ▪ Typing slips ▪ Low language proficiency ▪ Changes in the real world states ▪ Wording that excluded alternative point of view (POV) ▪ Garbled by software | fix, correct, change, remove, revert, remove exhaustive qualifiers, specify, clarify context, update, provide epistemology, verify, explain; resolve contradictions. | 2. Int. Accuracy 3. Rel. Accuracy |
| <i>Authority</i> | 2 | 0 | <ul style="list-style-type: none"> ▪ Lack of supporting sources ▪ Lack of academic scrutiny of the sources ▪ Known bias of the source ▪ Unfounded generalization | add, replace, remove, reword, qualify | 4. Authority |
| <i>Cohesiveness</i> | 1 | 1 | <ul style="list-style-type: none"> ▪ Loss of focus | restrict, move | 5. Cohesiveness |
| <i>Completeness</i> | 49 | 20 | <ul style="list-style-type: none"> ▪ Existence of multiple perspectives ▪ Unbalanced coverage of different perspectives ▪ Lack of detail ▪ Difference between an encyclopedia article genre and the genre from which the text was imported | add, specify, disambiguate, include, expound, balance, qualify, clarify, integrate | 6. Int. Completeness 7. Rel. Completeness |
| <i>Complexity</i> | 7 | 8 | <ul style="list-style-type: none"> ▪ Low readability ▪ Complex language | replace, rewrite, simplification, move, summarize | 8. Int. Complexity 9. Rel. Complexity |
| <i>Consistency</i> | 13 | 12 | <ul style="list-style-type: none"> ▪ Using different vocabulary for the same concepts within the article or within the collection ▪ Using different structures and styles for the same type of articles ▪ Non-conformance to the suggested style guides ▪ Differences in culture/language semantics ▪ Conflicting reports of factual information ▪ Contradicting or conflicting with a particular cultural or social norm, convention, or standard | reorganize, conform, revert, move, choose most widely used form, vote | 10. Int. Sem. Consistency 11. Int. Strct. Consistency 12. Rel. Sem. Consistency 13. Rel. Strct. Consistency |
| <i>Informativeness</i> | 6 | 4 | <ul style="list-style-type: none"> ▪ Content redundancy | remove, move, revise, cut down | 14. Int. Informativeness 15. Rel. Informativeness |
| <i>Naturalness</i> | 2 | 1 | <ul style="list-style-type: none"> ▪ Obscure language, text does not flow well | edit, rewrite, improve | 16. Int. Naturalness 17. Rel. Naturalness |
| <i>Relevance</i> | 18 | 16 | <ul style="list-style-type: none"> ▪ Adding content that is not relevant or outside of the scope of the article | revert, move, separate, get rid of, remove | 18. Relevance |
| <i>Verifiability</i> | 19 | 12 | <ul style="list-style-type: none"> ▪ Lack of references to original sources ▪ Lack of accessibility of original sources | add, remove, cite, revert, provide, confirm | 19. Verifiability |
| <i>Volatility</i> | 2 | 1 | <ul style="list-style-type: none"> ▪ Lack of stability due to edit wars and vandalisms | avoid, protect | 20. Volatility |

Table 3: IQ problem types, related causal factors and IQ assurance actions taken or suggested (Int. = Intrinsic, Rel. = Relational)

6. Types of activities affected

In earlier research (Gasser and Stvilia, 2001) we suggested that certain types of information-related activities could be more prone to quality problems:

- **Representation-dependent activities:** Whenever activity depends on how well an information repository's content *represents* some external situation, the correspondence between that representation and the underlying reality is a potential locus of IQ problems.
- **Activities that decontextualize information:** Whenever an agent removes information from the context in which it was produced--for example to aggregate raw information from a variety of original sources and integrate it into a focused collection supporting a specific task - the new context may change how information quality is assessed or understood.
- **Stability-dependent activities:** Whenever activity depends upon stable properties of the information, and/or entities and conditions it represents.

We found references to all three activities in the community discussion. In some cases editors even proposed hypothetical information use scenarios to justify their argument or claim:

What I was most concerned about, though, was the "disputed" tag, which normally means there are specific statements in the article that are factually inaccurate, not that there is a missing information. Is there anything in there now (apart from the recent revert) that is wrong enough that we should turn away a student writing a paper on alchemy? And if so, what is it specifically, so we can get it changed?

The mapping of the quality problem types to the activity types shows that Decontextualizing activities may cause most of the IQ problem types. An IQ problem may arise when information brought into an article from an outside source does not match the genre, cultural, or cognitive context of the English Wikipedia's common user. For instance, the community found an article problematic because it was based on a student essay. The essay itself was not of low quality. However, once *decontextualized* and moved into the Wikipedia context it came into a conflict with the Wikipedia NPOV policy as it reflected the author's particular vision or perspective on the topic.

The community discussions had abundance of referrals to *representational* quality problems caused by the lack of completeness and different perspectives on a subject. As we noted earlier changes in the context of information creation and use, or in any of its subcontexts – culture, socio-technical structures – may lead to misalignment, and consequently to a representational quality problem. Wikipedia articles, however, are not created by representatives of a single

culture or social group. From the start, an article creation process can be influenced and shaped by the interplay of different cultural and social perspectives, and the different levels of subject knowledge by its editorial group. The following excerpt presents an interesting example of how these different structures may influence each other consciously or subconsciously:

May a Roman-Catholic cardinal be a woman? "If any cardinal elector is by reason of infirmity confined to his or her room, the.." and "... each cardinal wrote only his or her own name on the ballot..." – User X.

Maybe this is a result of some automatic political-corectness tool? – User Y.

Hence, an article in a stable state reflects a certain consensus among these different variables. If this consensus breaks down, the article can become unstable – an edit war may erupt, or at least quality problem claims of incompleteness, missing perspectives, biases and POV may be voiced.

We found instances when the members of the community expressed their concerns that frequent vandalism or edit wars made some of the articles *unstable*, degraded their utility and drained the community's resources.

I've noticed that there has been a lot of vandalism and reversion today -I'm wondering if this article should be protected somehow (I've forgotten exactly what its called) to prevent the vandals constantly fouling up the article.

However, instability was often caused not by malicious actions but a regular process of having a team of editors working on the article. It, however, does not make much difference to the user who is working on a term paper and wants to reference a definition or the dates of some historical event from the article. The user needs to have some assurance that the definition is correct and/or it will be there intact the next time she or her teacher visits the page. The community recognized that an unstable article could not be a good reference source. Therefore, one of the criteria for an article to achieve an FA status is stability.

We also observed that the persistent records of an article's provenance – its creation and modification processes - proved to be extremely helpful in quality assurance activities such as spotting malicious edits, or assessing the article's quality in general.

His other edits were subtle vandalisms - he changed dates a bit and these changes didn't get caught for very long time. Can someone check whether the change in Ackermann function above is or isn't valid? User X

It's redundant, m is always greater than 0 there. I'll remove it. Thank you for pointing it out. User Y.

Thus, in addition to the earlier identified three types of activities that may often suffer from IQ problems – *Representation*-dependent, *Decontextualizing* and *Stability*-dependent activities – the

rich record of Wikipedia quality control process suggested a fourth type – *Provenance*-dependent activities. Those activities may often fail or become impossible to carry out as the origin of information and the process of its update have not been documented and preserved fully and consistently. Alternatively, as the above example shows, easy access to a well organized, detailed record of the object’s provenance and modification history may lead to the successful outcome of an activity – effective search and evaluation of the information. Spotting vandalism from a particular user in one article led to the quick identification and recovery from the same user’s other malicious edits.

7. Inferring the active IQ model from process data

The IQ claim and discussion logs can be used for inferring Wikipedia’s *active* IQ assessment model and IQ problem structure. Some parts of the model may be formalized into policy artifacts, others may stay informal, but actively used and embedded in the community’s organized activities and actions: discussions, negotiations and votes. For instance, the FA Criteria does not include Authority or Reputation dimensions. However, it was found by earlier studies (Viegas, Wattenberg and Dave, 2004) and corroborated by the findings of this study as well, that the community does take into consideration the editor’s reputation when making IQ judgments. The records of IQ problem claims and votes may also help in inferring specific tradeoffs among different IQ dimensions, and the distributions of IQ problem types. The analysis of the Featured Article Removal Candidate (FARC) discussions suggested that the odds of featured articles losing their status because of *Volatility/Stability* problems were more than twice higher than those because of *Authority* problems (see Table 4). The analysis also identified *Accuracy* and *Completeness* as most frequently encountered IQ problem claims. Looking back at the Wikipedia IQ model again (see Figure 1) and finding there separate processes for addressing *Accuracy disputed* and *NPOV disputed* claims, tells us that the Wikipedia IQ assurance model, in particular the IQ architecture, was indeed quite well aligned with the contemporary IQ problem structure of the collection.

| IQ Problem Types | Status Retained | Status Lost | Total | Retention Probability | Loss Probability |
|---|------------------------|--------------------|--------------|------------------------------|-------------------------|
| <i>Accessibility</i> | 3 | 3 | 6 | 0.50 | 0.50 |
| <i>Accuracy</i> | 6 | 5 | 11 | 0.55 | 0.45 |
| <i>Authority</i> | 2 | 1 | 3 | 0.67 | 0.33 |
| <i>Cohesiveness</i> | 7 | 5 | 12 | 0.58 | 0.42 |
| <i>Completeness</i> | 16 | 49 | 65 | 0.25 | 0.75 |
| <i>Complexity</i> | 11 | 15 | 26 | 0.42 | 0.58 |
| <i>Consistency</i> | 8 | 18 | 26 | 0.31 | 0.69 |
| <i>Informativeness</i> | 4 | 11 | 15 | 0.27 | 0.73 |
| <i>Naturalness</i> | 2 | 9 | 11 | 0.18 | 0.82 |
| <i>Relevance</i> | 3 | 4 | 7 | 0.43 | 0.57 |
| <i>Verifiability</i> | 10 | 28 | 38 | 0.26 | 0.74 |
| <i>Volatility/Stability</i> | 1 | 6 | 7 | 0.14 | 0.86 |
| <i>Using IQ standards retroactively</i> | 5 | 4 | 9 | 0.56 | 0.44 |
| <i>IQ Improvement work done</i> | 14 | 3 | 17 | 0.82 | 0.18 |
| <i>No IQ Improvement work done</i> | 26 | 77 | 103 | 0.25 | 0.75 |

Table 4: Descriptive statistics of IQ problems found in the FARC votes (out of 120 candidates 80 articles lost their FA status)

While the community consistently pointed to the FA quality standards when nominating an article for removal or discussing its IQ, the quality standards themselves kept changing over time. For instance, the requirement of a FA supplying references was only added in September 2004. Consequently, the articles that were well qualified as featured articles under the old IQ requirements did not do so well once the requirements changed. The community realized that having moving IQ standard targets, and applying them retrospectively would make FA status less stable, consequently less attractive, and could discourage article editors to strive for it. The community agreed that the new requirements would not be applicable to the articles that achieved a FA status before the requirement change. At the same time, however, they (at least some of them) understood that not bringing the old articles up to the new standards would increase the quality variance of featured articles, make their quality less predictable, and degrade the overall IQ of the collection.

Keep. When we started to require references, it was clearly said and understood that the requirement would not be retroactive. If we change our minds on that, we need to do so explicitly. – User X

I am going to take the principalled stand here and vote remove. The World War I article is an example of one that critically needs references. As a point of fact, User X, the references requirement was added on Sept 11, 2004, or nearly five months ago. That is more than long enough. If we don't make a stand somewhere it will never happen. Take the pain now for a much greater long term gain for the project. Lets help

eliminate Wikipedia's single greatest weakness. – User Y

Absolutely oppose removal. These attempts at imposing standards retrospectively should stop now. Deadline indeed! User Z

I call nonsense, if we don't hold ALL of our Featured articles to the same standard we invite ourself to a heavy degradation of quality, as we grow bigger our standards grow as well, this is good as it motivates editors to improve articles beyond their current state, up to a higher level.

– User Y

In addition, the community often used nomination for FA status removal to reinvigorate the editorial group of the article.

This article is featured on the main page but has had serious issues for quite a while now.Hopefully its listing here will stimulate improvements.- User X

The community was willing to compromise and allow the article to retain a FA status if they saw a lively interest in the article topic and members willing to put some work in the article to make it meet the IQ standards. In 17 instances out of 120 nominations for removal, editors updated articles and addressed some of the criticism posed in the nominations. In 14 cases out of those 17 the articles were allowed to retain their FA status (see Table 4). Thus, an article which had some IQ improvement work done on it had more than 3 times higher chance to retain FA status than an article which had not – 0.82 vs. 0.25.

I've amended the article to address your objections. - User X

Some of the sections could use a bit of expansion, but this does remove the reason for removal. I will re-list this once-again fine example of Wikipedia prose. :) – User Y

I added some further reading in lieu of knowing what references were actually used. Further I noticed that the none of the objectors had contacted the original author (still an active wikipedian) to comment on the issues, so I did so. - User X

Well done - at least these nominations are triggering improvements to the articles, which is clearly a good thing. - User Y

Some of the other causes of FA status objections besides not meeting the changed IQ standards were actual degradation of article IQ through content deletion or splitting into child articles transforming the original article into a collection of links. Others felt that the article was too long (over an informal limit of 32k) and could create an accessibility problem for users accessing Wikipedia over a regular modem line. In certain cases the community also used length as an indicator of whether an article followed the conventions of the encyclopedia genre (compact summaries of the subject) or not. However, it is important to note that this concern was not a dominant one in the arguments.

It is also horrendously huge (80 KB!), we should not be encouraging such a huge article size by featuring such an unusably long article. It needs to be broken up in discrete digestible bits (NOT another damn series - if you want to write a book then go to Wikibooks (<http://wikibooks.org>!))-

8. Political economy of IQ behavior

In the previous section we identified several types of tradeoffs, including tradeoffs between IQ and Cost, and IQ and Criticality. IQ assurance decisions made based on immediate context pragmatics were not rare in the Wikipedia community. Editors prioritized or optimized their IQ assurance activities based on current events or anniversaries:

Someone needs to update the papal conclave page ...before conclave starts, and that's going to happen very, very soon.

The conclave page is wrong about the methods of choosing a pope ("The election may come:..."). Current law dictates that a secret vote be taken; this is the only way to elect a pope, as dictated by John Paul II.

In instances the community showed an awareness of the risks open content and vandalism might pose to Wikipedia users as well as the benefits the same openness and exposure to the diverse perspectives and knowledge could bring to the article.

Many people access wikipedia without fully understanding what wikipedia is all about (because they arrived via a google search or similar) and so they see the vandalised text and believe that to be accurate.

Remember that featured articles that are displayed on the main page are still works-in-process. It is important that new users can edit them.

Effective resource allocation among different IQ assurance actions was also a concern. They were well aware that each instance of a FARC nomination and vote consumed the community's resources (time) which instead could be spent on fixing the IQ problem itself. Suggestions were often voiced for better aligning IQ assurance activity structures to individual IQ problem types and optimizing IQ assurance resource allocation among different actions of the activity:

These little objections about lead sections, extra references, images, etc., should stop being used as de-feature fodder. If the prose and scope of an article is strong (as it is here), go in and do the modernizing touchup changes yourself. Half the time it will take you barely more effort than starting and defending a nom on this page.

This nomination is also in a terrible condition. You've not made a single edit to the article page, talk page nor specified a single specific problem with the article on this page.

Likewise, we found that editors were conscious of the effectiveness of communication and the use of different coordination artifacts when making changes in articles. For instance, voting or polling was used quite often when editors intended to modify a featured article's content but not for regular articles. An editor notifies the community about his or her intention to change an article's content and states the rationale for the change. If no one objects to the proposed change, the editor makes the change, otherwise a negotiation process starts. If no consensus or compromise is achieved in negotiation, a voting procedure is enacted and the dispute is resolved through a majority of votes. Clearly, this process is more expensive than a straightforward edit clarified only through a subject line.

When the beliefs and contributions of different agents contradicted each other, and could not be reconciled, there were basically two options available: discard all conflicting contributions and deliberately avoid the issue, or present all points of view (POV). Wikipedia articles often served as good examples of the latter scenario, though one could observe the former kind of compromises as well.

I have revised most of the article, basically writing it anew, although I tried to preserve everything that had not been poisoned by inter-Slavonic recriminations. ...I've deliberately left a lot unsaid here, including historical grammar

Why has protected status been removed? The people who wanted Ecnomus mentioned in the first paragraph have not consented with the present version and they will no doubt do their stuff again. – User X

I asked for protection to be removed because I wanted to write something about the battle. It's only one sentence in the introduction, after all, and it's not wholly objectionable, so I think we can live with the Ecnomus enthusiasts. Eventually they will go away and the reference can be removed. - User Y

I see people keep inserting links to that article; although it's now named Papal election of 2005 speculations. I'd prefer the courtesy of explaining to me why I'm wrong in arguing it should not be linked to, but I have no intention of making this a lame edit war. I'll leave it more talkative people to others to sort out the issue

Clearly, there could be an economic argument behind this behavior. Representing alternative POVs in highly contested areas without critical analysis of the content and quality of argument relieves the IQA agent from validation, and some of the negotiation and editing costs. It also gives partial satisfaction or utilities to the disputing parties, motivating them to continue contributing and using Wikipedia. It changes, however, a traditional positivist approach of encyclopedia construction, which assumes that there is always one truth and a certain predictable level of quality, to a constructivist, 'grounded' approach, which assumes that there are always multiple truths and quality, and they change over time. However, along with objective changes caused by the changes in the underlying reality and scientific discoveries, updates and modifications can be motivated by subjective reasons as well. Information can be changed, or reinvented to align with a particular point of view, or achieve a particular outcome. Garfinkel (1967) describes how a decision may come before the information when jurors sought information for justifying an already made decision in retrospect.

We found that editors often signaled their IQ beliefs and philosophies through member home pages. It may serve several purposes. Editors may save time by directing the opponents to their IQ philosophy statements to clarify their behavior. The statement may also serve as an individual coordination artifact helping to provide consistency to the individual editor's IQ assurance actions and behavior – does the editor support the inclusion of school articles in Wikipedia or

not, for instance. The statement may also facilitate searching and identification of likeminded members and forming associations to prevail in various IQ decisionmaking votes and adopt desired policies. Finally, the individual IQ philosophy statements may serve as a basis for the policy statements of such associations.

Interestingly, the formal declaration and differentiation of IQ philosophies through establishing the associations may help in reducing IQ variance in Wikipedia and making it more predictable. Whether creating member IQ behavior policies and guidelines through the synthesis and integration of member IQ beliefs, or just coordinating member IQ votes, the associations may help to reduce the variance of member IQ behavior. At the community level, the rivalry among the associations also may lead to better scrutiny of the existing Wikipedia IQ policies, or the IQ metrics, policies and behaviors advocated by each group and finding points of integration and convergence.

Associations may also contribute to the cost sharing of building an IQ assurance infrastructure. Parts of the competing IQ assurance policies and structures can be reused in the enactment of similar mechanisms at the community level. As a result administrators receive help from the members who might not participate in such activities otherwise.

9. Discussion

If one contrasts the work organization models of Wikipedia and conventional databases, one may notice several important differences (see Figure 3):

- First of all, IQ assessment and the end user feedback processes in traditional databases are detached from the information creation process, while those processes are integrated in Wikipedia.
- In Wikipedia the end user and editor roles are merged. Any Wikipedia user can instantly become an editor. The end user is a part of the article creation process. In traditional databases those roles are mostly distinct.

- The product creation and delivery environments for Wikipedia are the same, and there is no time lag between the creation of the product and its delivery to the user. That also means immediate gratification for the editor, though it applies to both good and bad edits.
- Work coordination in Wikipedia is less formal - the decisions on which role to play and how work is divided are made by individuals themselves. The members of the community can self-identify for jobs in the collaborative work environment based on their interests, knowledge, skills and available resources (time). For instance, one can provide deep knowledge and understanding of the subject while others can contribute good editorial skills.

For uncontroversial, ‘stable’ topics self-selection also ensures that members of editorial groups are substantially well aligned with each other in their interests, backgrounds, and overall understanding of the topics. Hence, self-selection may allow significant savings on selection and coordination costs (Benkler, 2002). It may lead to strongly motivated, successful editorial groups, and consequently to high quality articles. For controversial topics, on the other hand, self-selection may produce a strongly misaligned editorial group. It can lead to conflicts among the editorial group members, continuous edit wars, and may require the use of formal work coordination and control mechanisms. These may include intervention by administrators who enact dispute review and mediation processes, completely disallow or limit and coordinate the types and sources of edits.

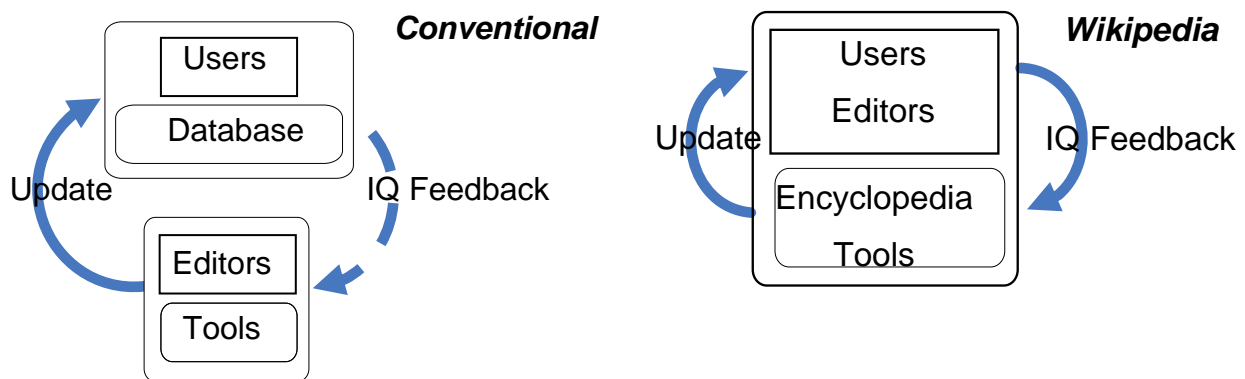


Figure 3: Conventional model vs. Wikipedia IQ assurance model

Merging editor and user roles and having article creation processes visible may cause

problems when the user accesses a ‘raw’ or vandalized image of the article. However, since there is no perfect or permanent IQ, IQ problems matter as much as the time it takes to fix them. From this perspective, the current egalitarian feedback based IQ assurance model of Wikipedia is superior to the model used by traditional encyclopedias where fixing IQ problems may take months if not years (Crawford, 2001).

The model of the Wikipedia quality assurance evolves with the changes in the complexity of the selection task and the use environment. In 2003 Wikipedia had almost no quality assurance guidelines or assessment criteria except for the general “pages that simply will never become encyclopedia articles” and “brilliant prose” statements in the Deletion and FA policies. Within 3 years the Wikipedia community has managed to develop not only a set of formal quality assessment criteria but a whole infrastructure of quality evaluation and promotion mechanisms and guides. New IQ tools and artifacts (bots, Wikiprojects, templates, best practice guides) are continually created to promote consistency in the *content, structure and presentation* of the articles. Indeed, while the article share in the Wikipedia pages went down from 53% in March 2005 to 36% in July 2006, the shares of templates, projects, and especially user discussion pages grew significantly (see Table 5).

| Page Type | March 9, 2005 Dump (945,622 pages) | | July 2, 2006 Dump (3,450,781 pages) | |
|------------------------|------------------------------------|-------------|-------------------------------------|-------------|
| | Count | Percentages | Count | Percentages |
| <i>Articles</i> | 500,623 | 53 | 1,257,152 | 36 |
| <i>Talk/Discussion</i> | 107,569 | 11 | 428,964 | 12 |
| <i>User</i> | 31,357 | 3 | 190,347 | 6 |
| <i>User_Talk</i> | 73,083 | 8 | 702,776 | 20 |
| <i>Project</i> | 18,836 | 2 | 105,284 | 3 |
| <i>Project Talk</i> | 2,870 | 0.3 | 10,474 | 0.3 |
| <i>Template</i> | 8,477 | 1 | 54,376 | 2 |
| <i>Template_Talk</i> | 1,151 | 0.1 | 9,139 | 0.3 |

Table 5: Page counts by type

Likewise, at the individual article level, as the complexity of an IQ management task increases, the local IQ control model and infrastructure get more complex and better organized as well. The records of past quality problem claims, related negotiations and resolutions achieved, if any, get organized (structured, categorized) and enhanced for better *discovery*,

referencing and *comparison*. As Table 2 shows the readability of article discussion pages has increased substantially from March 2005 to July 2006.

10. Conclusion

In this paper we presented a case study of the IQ assurance practices used at the Wikipedia project. The qualitative analysis of the quality of Wikipedia articles helped us to understand the ways in which quality is collaboratively established and improved despite what seems at first glance the seemingly anarchic operation of the project. Featured articles were used as a means of setting a quality standard against which other articles can be compared. This quality standard is not ideal, but it does seem relatively rigorous. As a resource for IQ research, a major advantage is that it requires very little effort to obtain compared to other analytic methods such as blind judging. Additionally we looked at the processes of article creation, and particularly at the article discussion pages as a rich source of qualitative data about what participants in Wikipedia perceive as issues of quality and the processes and tradeoffs that operate in activities to improve quality. The study shows that the Wikipedia community takes issues of quality very seriously. Although anyone can participate in editing articles, the results are carefully reviewed and discussed, in ways very similar to open source programming projects.

We believe that the study of those evolving debates, processes and the IQ assurance model as a whole has useful implications for the improvement of quality in other more conventional databases. The classic problems within information quality are trying to determine what quality is, how it might be measured, and what should be done to improve it (Wang and Strong, 1996; Strong, Lee and Wang, 1997; Wang, 1998). Wikipedia offers a special insight into these problems. The Featured Page feature, Deletion Policy, the evolving criteria of assessment and the debates around attaining and maintaining that status allow us to see how one particular community defines and continually redefines quality, and how they assess it in particular cases. Furthermore, the discussion pages attached to articles reveal how quality issues are discussed and how quality improvements and trade-offs are addressed. What is special about Wikipedia as a resource is that the quality discussions and processes are strongly connected to the data itself. In most conventional databases studied by information quality researchers, these discussions and

processes are divorced from the data, and gaining access to them either for research or even for productive reflection by the practitioners themselves is very difficult. We believe it will be very interesting and productive to explore how, inspired by the success of Wikipedia, it might be possible to connect and record quality discussions and processes to the resultant data and so allow various ways for the users of this data to be directly involved in the quality improvement process (Twidale and Marty, 1999; Twidale and Marty, 2000).

The success of Wikipedia and the Wikipedia community would not be possible without a technological innovation – Wiki software - which enables very low cost, collaborative content creation and effective quality control by using a regular web browser and simple markup language. Using this software Wikipedians not only create the content of Wikipedia, but also evolve, discuss and document their work processes, including those relating to detecting and improving quality. Fortuitously this openness also makes it particularly easy for us as researchers to study a particular kind of collaborative work that is highly distributed and has a particularly substantial focus not just on error detection but also error correction. Hence, the Wikipedia technology is not only influenced by the existing social and cultural structures but the technology itself influences the community and makes new forms of social interaction and cooperation possible. This interplay of the technology and community mechanisms creates a new system of IQ assurance that has proved to be surprisingly successful. Contrary to public perception the Wikipedia IQ assurance model is robust and quite sophisticated. It is proactive and carefully planned to promote continuous IQ improvement.

Notes

- ¹ <http://en.wikipedia.org/wiki/Wiki>
- ² <http://c2.com/cgi/wiki?WikiHistory>
- ³ http://en.wikipedia.org/wiki/Wikipedia:Multilingual_coordination
- ⁴ http://meta.wikimedia.org/wiki/Wikipedia.org_is_more_popular_than...
- ⁵ <http://ask.slashdot.org/faq/editorial.shtml>
- ⁶ <http://download.wikimedia.org/>
- ⁷ Some articles have been awarded a Featured status but have not been actually featured on the Wikipedia main page
- ⁸ http://en.wikipedia.org/wiki/Wikipedia:Featured_articles
- ⁹ <http://en.wikipedia.org/wiki/WP:AFD>
- ¹⁰ http://en.wikipedia.org/wiki/Wikipedia:Criteria_for_speedy_deletion
- ¹¹ http://en.wikipedia.org/wiki/Wikipedia:Featured_article_candidates
- ¹² http://en.wikipedia.org/wiki/Wikipedia:Featured_list_removal_candidates
- ¹³ http://en.wikipedia.org/wiki/Wikipedia:Accuracy_dispute
- ¹⁴ http://en.wikipedia.org/wiki/Wikipedia:NPOV_dispute
- ¹⁵ http://en.wikipedia.org/wiki/Category:Wikipedia_bots
- ¹⁶ <http://en.wikipedia.org/wiki/Wikipedia:Administrators>
- ¹⁷ <http://en.wikipedia.org/wiki/Wikipedia:Bot>
- ¹⁸ http://meta.wikimedia.org/wiki/Association_of_Inclusionist_Wikipedians
- ¹⁹ http://meta.wikimedia.org/wiki/Association_of_Deletionist_Wikipedians
- ²⁰ http://meta.wikimedia.org/wiki/Category:Wikipedian_associations
- ²¹ <http://en.wikipedia.org/wiki/Wikipedia:WikiProject>
- ²² http://en.wikipedia.org/wiki/Wikipedia:What_is_a_featured_article
- ²³ http://en.wikipedia.org/wiki/Wikipedia:Featured_articles
- ²⁴ http://en.wikipedia.org/wiki/Wikipedia:What_is_a_featured_article
- ²⁵ http://en.wikipedia.org/w/index.php?title=Wikipedia:Featured_articles&direction=prev&oldid=47610
- ²⁶ http://en.wikipedia.org/w/index.php?title=Wikipedia:The_perfect_article&oldid=159320
- ²⁷ http://en.wikipedia.org/w/index.php?title=Wikipedia:Featured_article_candidates&diff=2673359&oldid=2673352
- ²⁸ http://en.wikipedia.org/w/index.php?title=Wikipedia:What_is_a_featured_article%3F&oldid=4077849
- ²⁹ http://en.wikipedia.org/wiki/Wikipedia:Featured_article_removal_candidates
- ³⁰ http://en.wikipedia.org/wiki/Wikipedia:Featured_article_removal_candidates/archive
- ³¹ http://en.wikipedia.org/wiki/Wikipedia:Deletion_policy
- ³² transwiki means moving pages between two wikis – for instance, if an article is written in German, it will be transferred to the German Wikipedia (<http://meta.wikimedia.org/wiki/Transwiki>)

³³ http://en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion

³⁴ http://en.wikipedia.org/wiki/Wikipedia:Votes_for_undeletion

³⁵ http://en.wikipedia.org/wiki/Wikipedia:Criteria_for_speedy_deletion

³⁶ http://en.wikipedia.org/w/index.php?title=Wikipedia:Deletion_policy&oldid=1674617

³⁷ http://en.wikipedia.org/wiki/Category:Accuracy_disputes

³⁸ http://en.wikipedia.org/wiki/Category:NPOV_disputes

References

- Bailey, K. (1994): *Methods of social research* (4 ed.). New York, NY: The Free Press.
- Benkler, Y. (2002): Coase's penguin, or, linux and the nature of the firm. *The Yale Law Journal*, 112(3).
- Biber, D. (1988): *Variations across speech and writing*. Cambridge, UK: Cambridge University Press.
- Bowker, G. and S. Star (1994): Knowledge and infrastructure in international information management: problems of classification and coding. In: L. Bud-Frierman (Ed.), *Information Acumen* (pp. 187-216). London: Routledge.
- Brin, S. and L. Page (1998): The anatomy of a large-scale hypertextual web. *Computer Networks and ISDN Systems*, 30(1-7), 107-117.
- Cosley, D., D. Fankowski, S. Kiesler, L. Terveen and J. Riedl (2005): How oversight improves member-maintained communities. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. Portland, OR. 11-20.
- Crawford, H. (2001): Encyclopedias. In: R. Bopp, L. C. Smith (Eds.), *Reference and information services: an introduction* (3 ed.) (pp. 433-459). Englewood, CO: Libraries Unlimited.
- Crowston, K. and B. Scozzi (2004): Coordination practices for bug fixing within FLOSS development teams. In: *Proceedings of the First International Workshop on Computer Supported Activity Coordination (CSAC 2004)*. Porto, Portugal.
- Ehrlich, K. and D. Cash (1999): The invisible world of intermediaries: a cautionary tale. *Computer Supported Cooperative Work*, 8(1-2): 147-167.
- Emigh, W. and S. Herring (2005): Collaborative authoring on the Web: a genre analysis of online encyclopedias. In: *Proceedings of the 39th Hawaii International Conference on System Sciences*.
- Garfinkel, H. (1967): *Studies in ethnomethodology*. Prentice-Hall.
- Gasser, L. and B. Stvilia (2001): *A new framework for information quality* (ISRN UIUCLIS--2001/1+AMAS). Champaign, IL: University of Illinois at Urbana Champaign.
- Gasser, L. and B. Stvilia (2003): Using multi-agent models to understand the quality of organizational information bases over time. In: *Proceedings of the NAACSOS Conference*. Pittsburgh, PA.
- Giles, J. (2005): Internet encyclopaedias go head to head. *Nature*, 438(7070), 900-901.
- Gracy, D. (2002): What you get is not what you see: forgery and the corruption of recordkeeping systems. In: R. Cox, D. Wallace (Eds.), *Archives and the public good : accountability and records in modern society* (pp. 247-264). Westport, CT: Quorum Books.
- Grudin, J. (1994): Groupware and social dynamics: eight challenges for developers. *Commun. ACM* 37(1), 92-105.
- Gunning, R. (1952): *Technique of clear writing*. New York, NY: McGraw-Hill.

- Heylighen, F. and J. Dewaele (2002): Variation in the contextuality of language: an empirical measure. *Foundations of Science*, 6, 293-340.
- Lerner, J. and J. Tirole (2004): *The economics of technology sharing: open source and beyond*. Working Paper 10956, Retrieved Jun 7, 2005, from <http://www.nber.org/papers/w10956>.
- Lih, A. (2004): Wikipedia as participatory journalism: reliable sources? Metrics for evaluating collaborative media as a news resource. In: *Proceedings of 5th International Symposium on Online Journalism*. Austin, TX.
- McArthur, T. (1986): *Worlds of reference: lexicography, learning and language from the clay tablet to the computer*. Cambridge, UK: Cambridge University Press.
- Orlikowski, W. and J. Yates (1994): Genre repertoire: the structuring of communicative practices in organizations. *Administrative Science Quarterly*, 39, 541-574.
- Pekkola, S. (2003): Designed for unanticipated use: common artifacts as design principle for CSCW applications. In: *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work, GROUP 2003*. Sanibel, FL. 359-368.
- Raymond, E. (1998): The cathedral and the bazaar. *First Monday*, 3(3)
- Robinson, M. (1993): Design for unanticipated use In: G. de Michelis, C. Simone, and K. Schmidt (Eds.): *Proceedings of the Third European Conference on Computer Supported Cooperative Work (ECSCW'93)*. Kluwer Academic Publishers, pp. 187-202
- Sack, W., F. Détienne, J. Burkhardt, F. Barcellini, N. Ducheneaut and D. Mahendran (2004): *A methodological framework for socio-cognitive analyses of collaborative design of open source software*. Paper presented at the distributed collective practices workshop in CSCW'04. November 6-10, Chicago, IL.
- Schmidt, K. and L. Bannon (1993): Taking CSCW seriously: Supporting articulation work. *Computer Supported Cooperative Work*, 1(1-2), 7-40.
- Schmidt, K. and C. Simone (1996): Coordination Mechanisms: Towards a Conceptual Foundation of CSCW Systems Design. *Computer Supported Cooperative Work*, 5, 155-200.
- Smith, L. C. (1989): "Wholly new forms of encyclopedias": electronic knowledge in the form of hypertext. In: *Proceedings of the forty-fourth FID Congress*. Helsinki, Finland 245-250.
- Strauss, A. (1978): *Negotiations: varieties, contexts, processes, and social order*. San Francisco, CA: Jossey-Bass.
- Strong, D., Y. Lee, and R. Wang (1997): Data quality in context. *Communications of the ACM*, 40(5), 103-110.
- Stvilia, B., M. B. Twidale, L. Gasser and L. C. Smith (2005a): *Information quality discussions in Wikipedia (ISRN UIUCLIS--2005/2+CSCW)*. Champaign, IL: University of Illinois at Urbana - Champaign.
- Stvilia, B., Twidale, M. B., Smith, L. C., Gasser, L. (2005b): Assessing information quality of a community-based encyclopedia. In: *Proceedings of the International Conference on Information Quality - ICIQ 2005*. Cambridge, MA 442-454.

- Trepess, D. and T. Stockman (1999): A classification and analysis of erroneous actions in Computer Supported Co-Operative Work environment. *Interacting with Computers*, 11 (6) p. 611-622
- Twidale, M. B. and P. Marty (2000): Coping with errors: the importance of process data in robust sociotechnical systems. In: *Proceedings of CSCW'00*. Philadelphia, PA. 269-278.
- Twidale, M. B. and P. Marty (1999): *Investigation of data quality and collaboration*. Champaign, IL: GSLIS, University of Illinois at Urbana - Champaign.
- Viegas, F., M. Wattenberg and K. Dave (2004): Studying cooperation and conflict between authors with history flow visualizations. In: *Proceedings of CHI 2004*. Vienna, Austria 575-582.
- Wales, J. (2005): *Interview*. C-SPAN. Retrieved Oct 20, 2005, from <http://www.q-and-a.org/Transcript/?ProgramID=1042>
- Wang, R. and D. Strong (1996): Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5-35.
- Wang, R. (1998): A product perspective on total data quality management. *Communications of the ACM*, 41(2), 58-65.
- Weick, K. (1995): *Sensemaking in organizations*. Thousand Oaks, CA: Sage.
- Winchester, S. (2003): *The meaning of everything: the story of the Oxford English Dictionary*. Oxford University Press.
- Yin, R. (1988): *Case study research: design and methods*. London, UK: Sage Publications Ltd.