

A revised version of this paper has been submitted to ICKM05

INFORMATION QUALITY DISCUSSIONS IN WIKIPEDIA

Besiki Stvilia, Michael B. Twidale, Les Gasser, Linda C. Smith

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

Abstract.

We examine the Information Quality aspects of Wikipedia. By a study of the discussion pages and other process-oriented pages within the Wikipedia project, it is possible to determine the information quality dimensions that participants in the editing process care about, how they talk about them, what tradeoffs they make between these dimensions and how the quality assessment and improvement process operates. This analysis helps in understanding how high quality is maintained in a project where anyone may participate with no prior vetting. It also carries implications for improving the quality of more conventional datasets.

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 [McArthur86], open access large scale public collaborative content creation projects are relatively recent phenomena. They are enabled 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 April 21st 2005 the English Wikipedia boasted more than 500,000 articles and daily usage in October 2004 was 6 million pages³. As a volunteer project Wikipedia needs 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.

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 [Lerner04][Benkler02][Crowston04] for reviews of these issues in the case of OSS). In this paper,

¹ <http://en.wikipedia.org/wiki/Wiki>

² <http://c2.com/cgi/wiki?WikiHistory>

³ <http://en.wikipedia.org/wikistats/EN/TablesWikipediaEN.htm>

we focus on the quality of the information in the Wikipedia articles. Given the very open approach to participation in Wikipedia and contrasting it with conventional multi-authored encyclopedias [Collison66] that use careful invitations and rigorous editorial review *prior* to publication, it might be suspected that the quality of the information in the Wikipedia articles would be very low. With such an open approach, surely the contributions of individuals would be highly variable and really could not be trusted to be accurate or complete. This 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. We believe that such an analysis has much to teach us about how information quality can be improved in more conventional data sources as well.

The paper presents results from our preliminary empirical studies of information quality (IQ) of the Wikipedia articles. We present a number of qualitative and quantitative characterizations of a random sample of the Wikipedia articles discussion pages. Based on this analysis we identify ten information quality problem types encountered by the Wikipedia community, the types of user information activities that might be affected by those problems and the processes of quality assurance and negotiation enacted by the community. Finally, we discuss some possible implications of the patterns of communication and work organization exhibited by the community on the quality of the encyclopedia content.

1.1 Overview of Approach

We start with a brief review of the background of the current research and the related past research on Wikipedia. Section 2 introduces the general context of IQ assurance in Wikipedia and the main unit of the analysis – Article Discussion Pages. The section also briefly reviews the research design and methodology of this study. Section 3 looks at how IQ problems are identified and treated by the community. Section 4 discusses some of the patterns of work articulation and negotiation documented in the discussion pages and their possible impacts on the IQ of Wikipedia content. We conclude the paper with some wider implications of the work and future research directions.

1.2 Background and Related Research

Wikipedia is an encyclopedia, drawing on the rich tradition of the genre of the encyclopedia. The conventions and forms of reference texts – dictionaries, encyclopedias and others– have been evolving over thousands of years starting from the clay tablets of the ancient Sumer [McArthur86]. [Smith89] citing [Collison66] states that by the end of the 19th century the genre of the encyclopedia was already well defined with almost universally accepted principles of its form:

- (a) written in the language of the country in which it was published;
- (b) contents arranged in alphabetical order;
- (c) articles of any substance written by specialists;
- (d) subject specialists employed either wholly or part-time as subeditors;
- (e) inclusion of living people's biographies;
- (f) inclusion of illustrations, maps, plans, etc.;
- (g) provision of bibliographies appended to the longer articles;
- (h) provision of an analytical index of people and places and minor subjects;
- (i) provision for the publication of supplements to bring the main work up-to-date;

(j) provision of numerous and adequate cross references in the text.

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, [Smith89] 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. As a result we can consider the suitability of using quality criteria developed for assessing conventional encyclopedias to assess Wikipedia.

Libraries have been playing a crucial role in developing reference genres in general. They have served not only as resources to scholars developing reference texts, but being the main consumers of the genre, they also contributed significantly to the development of the policies and norms of its assessment and mediation based on typified information use activities. Crawford identifies 3 categories of questions that can be answered best using encyclopedias: (1) ready reference questions – “what is?” and “where can I find?”; (2) general background information questions; (3) pre-research information leading to more targeted and detailed sources [Crawford01]. In addition, she proposed the following general dimensions when evaluating encyclopedia quality: (1) Scope (Purpose, Subject Coverage, Audience, Arrangement and Style); (2) Format; (3) Uniqueness; (3) Authority; (4) Accuracy (Accuracy and Reliability, Objectivity); (5) Currency, and (6) Accessibility (Indexing). Two other dimensions – Relevance to user needs and Cost – were listed as selection criteria for a particular reference source.

Clearly, not meeting user community requirements at any of the above dimensions can lead to IQ problems. The authors of this paper earlier developed an IQ assessment framework consisting of 22 dimensions divided into 3 categories (see Appendix). The framework was applied to evaluate the IQ of another type of reference object – Dublin Core metadata records [Stvilia04][Shreeves05]. While the purposes of an encyclopedia article and a catalog record are somewhat different (providing a compact representation of subject knowledge and providing a compact description of the attributes of another object, respectively), as representational objects these two genres clearly share a number of potential IQ problem types and related dimensions (see Figure 1).

In addition to comparing Wikipedia articles and those of conventional encyclopedias, it is also possible to compare articles within Wikipedia, particularly those that have been denoted as being of particularly high quality or particularly problematic (as described in the next section) to see if there is any difference in their creation processes that have led to this quality variation.

There have been a number of studies recently that looked at the quality of Wikipedia from different perspectives. [Lih04] studied Wikipedia content construction and use processes from the perspective of participatory journalism. In addition to providing a rather comprehensive account of the Wikipedia project history, the author analyzed the change in the quality of Wikipedia articles before and after they had been cited in the press. A combination of two metrics – the total number of edits (Rigor) and the total number of unique editors (Diversity) were used as indirect measures for assessing the quality of articles. As a benchmark the study used the median values of the above metrics (61 and 36.5) which were calculated based on a set of nodes obtained from the mappings of the 333 subjects in the Dorling Kindersley e.encyclopedia to Wikipedia. According to the study, during the 14 months of observation (from January 2003 through March 2004) 113 Wikipedia articles were cited in the press. The analysis of the histories of these articles showed that number of edits and distinct editors of the articles increased substantially and the number of articles staying above the benchmark values had more than doubled after those

citations were made. The most important observation of the study, however, was identifying the types or genres of the cited article which were mostly related to current events, slang and colloquial terminology.

[Viegas04] 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.

In May 2003 the smallest mean and median revert times were for obscene edits (mean: 1.8 days, median: 1.7 minutes) and the largest revert times were shown for complete deletions (mean: 22.3 days, median: 90.4 minutes).

[Emigh05] 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 [Biber88][Heylighen02]. 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 [Viegas04]) 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 [Strong97] 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, [Strong97] 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 (religious beliefs and various kinds of biases) for instance. A good example of cultural differences leading to quality compromise (purposeful degradation) is given in [Bowker94] where they 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, and 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 [Strauss78] 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) give us access to this kind of data, the analysis of which will be presented in the next sections.

2. RESEARCH DESIGN AND METHODOLOGY

This section looks at some of the components of the IQ assurance context of Wikipedia such as support artifacts, roles and processes. It also briefly reviews the research design and methodology of the study.

2.1 Wikipedia Roles

Wikipedia contributors participate from all over the world and with a rich variety of backgrounds. We can try to understand their contributions to quality by considering different roles. Activities fitting in these roles can also be performed by certain programs, so we use the term “agent” to refer to either people or programs. Indeed, at the time of writing this article the Wikipedia community employed around 50 automatic quality maintenance scripts called bots⁴. We have identified four types of agents: User agents that benefit from the information contained in Wikipedia; Information Quality Assurance (IQA) agents that contribute to the system by adding new content and/or maintaining and enhancing the existing articles; Malicious agents that degrade article quality by knowingly deleting valid content and/or inserting invalid entries; and, finally, Environmental agents that represent temporal changes in the real world states and human knowledge stock that make encyclopedia articles outdated or invalid.

The main group of IQA agents are Wikipedia Administrators⁵. As of April 2005 there were 431 Wikipedia users with administrative privileges. Extrapolation from the edit histories of a random sample of English Wikipedia articles suggest that Administrators comprise around 6% of Wikipedia’s contributor population and provide around 21% of the total number of edits.

Obviously the same agent may assume more than one role. IQA agents can be Users of the Wikipedia content and a maliciously acting agent might have made valuable contributions in the past or might do so in the future. All these lead to a host of different kinds of activities and interactions among different agents and highly dynamic socially negotiated and constructed information quality of the encyclopedia articles.

2.2 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 [Schmidt96] which helps to negotiate and align member perspectives on the content and quality of the article.

Discussion pages are part of Wikipedia’s overall support architecture, which also includes WikiProject⁶ resource pages, style manuals, best practices guides and other work coordination artifacts. Discussion pages are routinely used by IQA agents such as Administrators to

⁴ <http://en.wikipedia.org/wiki/Wikipedia:Bot>

⁵ <http://en.wikipedia.org/wiki/Wikipedia:Administrators>

⁶ <http://en.wikipedia.org/wiki/Wikiproject>

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.

2.3 Featured Articles

Featured Articles⁷ (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 criteria⁸. 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.

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"?⁹

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.

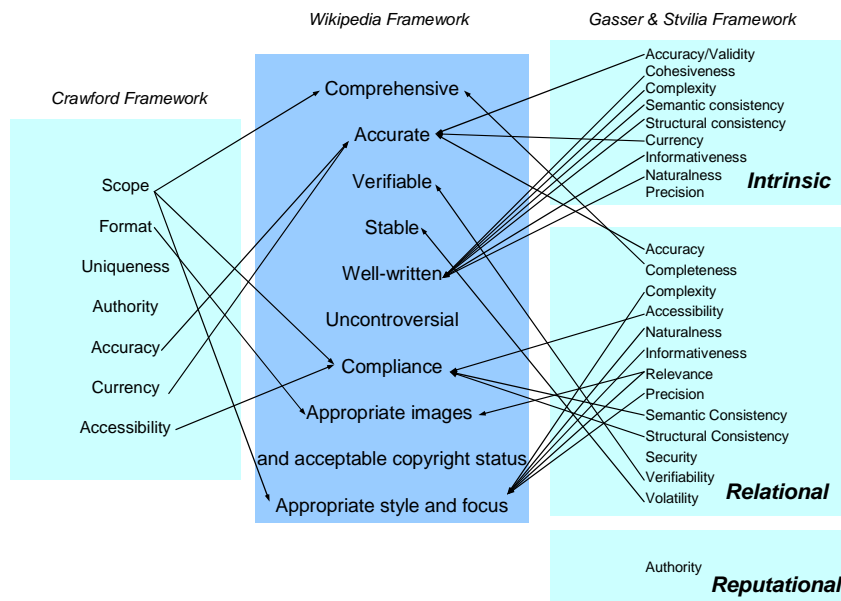


Figure 1 - Mapping between the Crawford, Wikipedia and Gasser & Stvilla IQ assessment frameworks

It is not clear exactly when the first formal quality assessment guideline was developed. According to the logs Wikipedia has a separate page defining what can be considered as a featured article starting from 20 April 2004. The current version of the page lists eight Featured

⁷ http://en.wikipedia.org/wiki/Wikipedia:Featured_articles⁷

⁸ http://en.wikipedia.org/wiki/Wikipedia:What_is_a_featured_article

⁹ http://en.wikipedia.org/w/index.php?title=Wikipedia:Featured_articles&direction=prev&oldid=47610

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 1 maps the quality assessment dimensions from the earlier mentioned printed encyclopedia quality assessment discussion from [Crawford01] and the ones proposed by [Gasser01] into the FA criteria.

Although the Wikipedia IQ framework lists Stable and Uncontroversial 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 or Reputation dimensions. Wikipedia puts trust not in a single expert author or group, but in the collective knowledge of a large-scale distributed community hoping that "given enough eyeballs all bugs are shallow" [Raymond98]. However the Wikipedia community does utilize a reputation mechanism, even though it is not formalized in its policies, as found by [Viegas04]; some Wikipedia users taking on editorial roles (as per our IQA role, see Section 2.1) 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. The recent study by [Cosley05] found that peer oversight was as good as experts in maintaining the quality of an information collection. Hence, the differences between the two IQ frameworks are largely caused by the pragmatics of their immediate social context of use.

The Stability criterion 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-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 current version lists 11 guides. Some of the relatively new additions were a verifiability guide, guidelines for controversial articles, and the guide to good image captions. Thus, as Wikipedia content evolves and gets refined 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 it¹⁰. According to the Wikipedia logs¹¹ 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). An analysis of the removal negotiations and votes data is given in section 3.1.

2.4 Methods

The current study is a part of a larger project which uses a method of case study [Yin88] with a number of qualitative and quantitative techniques such as content analysis, statistical experimentation and multi-agent simulation to gain understanding of the IQ variable structure, dynamics and organizational issues of IQ assurance in large-scale community-based open source information collections such as Wikipedia.

¹⁰ http://en.wikipedia.org/wiki/Wikipedia:Featured_article_removal_candidates

¹¹ http://en.wikipedia.org/wiki/Wikipedia:Featured_article_removal_candidates/archive

We wanted to analyze the IQ dynamics of the Wikipedia content over a period of time using three dumps¹² of the English Wikipedia database made on 2005/03/09, 2005/04/06 and 2005/04/21. After excluding entries that were redirects to other articles, the qualified population size of the 2005/03/09 dump was 500,623 articles, from which 1,000 articles were randomly selected. It was found that 153 of these articles were actually stubs¹³ (articles that were simply templates and contained little content). The sample was further reduced by 6 (0.7%) to 841 articles because of article deletions and redirections that took place between 2005/03/09 and 2005/04/21 (43 days). Out of those six article entries, four became redirects to other entries due to article merger and name change. The remaining two entries were simply removed from the article namespace. Hence, we ended up with a sample of 841 articles. In addition, to gain a better understanding of the qualitative structures and dynamics of IQ and the patterns of interactions among the different kinds of IQ agents, we harvested three special categories of articles identified as such by the Wikipedia community:

(1) Accuracy Disputed - articles whose accuracy is publicly disputed (234 articles as of 04/21/2005)¹⁴;

(2) NPOV Disputed - articles which are claimed to be biased – failing to conform to the Wikipedia norm of conveying a Neutral Point of View (NPOV) (520 articles as of 04/21/2005)¹⁵;

(3) Featured –articles that have been considered as the Wikipedia’s best and have been featured on its front page (236 articles as of 04/21/2005)¹⁶.

We also harvested the edit histories and discussion pages of all the articles selected. Not all Wikipedia articles have active discussion pages. In the Random sample, only 155 articles out of the total 841 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. Hence, for the purposes of the current study, we randomly selected 30 discussion pages that were longer than 100 characters from each of these two sets (the Featured and Random sets) and analyzed their content for IQ problem types and IQ assurance and negotiation patterns using the technique of content analysis [Bailey94a].

3 TYPES OF QUALITY PROBLEM INCIDENTS

In this section we explore how certain features of Wikipedia can be exploited to enable an assessment and comparison of information quality between certain articles, and how certain records of the process of article revision can be used to understand how issues of quality are discussed and addressed by the community.

3.1 Talking about Quality – Discussion Pages

Quality is often defined as “fitness for use” [Juran92]. It is ultimately a social construct [Strong97][Stvilia04]. To evaluate IQ meaningfully one has to have socially and culturally justified assessment criteria and norms to compare with. This implies the existence of some social consensus or order which according to [Strauss78] is always a negotiated order. Strauss suggests

¹² <http://download.wikimedia.org/>

¹³ http://en.wikipedia.org/wiki/Wikipedia:Perfect_stub_article

¹⁴ http://en.wikipedia.org/wiki/Category:Accuracy_disputes

¹⁵ http://en.wikipedia.org/w/index.php?title=Category:NPOV_disputes

¹⁶ http://en.wikipedia.org/wiki/Wikipedia:Featured_articles

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. Recently, [Sandusky04] used negotiation theory for analyzing software problem management processes in an open source software bug management system and understanding coordination mechanisms of collaborative work. In a similar study [Twidale05] examined a number of FOSS bug report systems to understand how usability bugs are discussed and solutions negotiated. In this particular research we were mostly interested in negotiations along with other processes such as user self-reported descriptions and justifications of actions taken as the means for identifying the types of IQ problems and dimensions that were considered important by the community, along with the kinds of tradeoffs made among those dimensions and their possible effects on the overall content of articles.

[Strong97] defines data quality problems as “difficulty encountered on one or more quality dimensions that renders data completely or largely unfit for use”. The content analysis of 60 discussion pages from the Featured and Random sets identified 10 types of IQ problems the Wikipedia users pointed to (See Table 1). Problem instances were tallied as many times as they occurred in discussions.

Problem Types	No. of Occurrences (Featured)	No. of Occurrences (Random)	Caused by	Action taken or suggested
Accessibility	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
Accuracy	57	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) ▪ Differences in culture/language semantics ▪ Garbled by software ▪ Conflicting reports of factual information 	Choose most widely used form, vote, fix, correct, change, remove, revert, remove exhaustive qualifiers, specify, clarify context, update, provide epistemology, verify, explain,
Authority	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
Completeness	49	20	<ul style="list-style-type: none"> ▪ Existence of multiple perspectives ▪ Unbalanced coverage of different perspectives ▪ Difference between an encyclopedia article genre and the genre from which the text was imported 	add, specify, disambiguate, include, expound, balance, qualify, clarify, integrate
Complexity	7	8	<ul style="list-style-type: none"> ▪ Low readability ▪ Complex language 	Replace, rewrite, simplification, move, summarize
Consistency	13	12	<ul style="list-style-type: none"> ▪ Using different vocabulary for the same concepts within the article or 	reorganize, conform, revert, move

			<ul style="list-style-type: none"> ▪ within the collection ▪ Using different structures and styles for the same type of articles ▪ Non-conformance to the suggested style guides 	
Informativeness	6	4	Content redundancy	Remove, move, revise, cut down
Relevance	18	16	Adding content that is not relevant or outside of the scope of the article	revert, move, separate, get rid of, remove
Verifiability	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
Volatility	2	1	Lack of stability due to edit wars and vandalisms	avoid, protect

Table 1: Problem types, related causal factors and IQ assurance action taken or suggested.

It is clear from Table 1 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 most frequently encountered IQ problem category in both the Featured and Random sets was *Accuracy*. It included incidents when the legitimacy of information was questioned on the basis of some reference source, such as invalid or less accurate use of vocabulary

Careen means violent side-to-side motion, rocking of ships, or laying of ships on their side for repair as ww says. Career means rushing forward at a great rate. In US English careen is also used synonymously with career (frowned on by guardian-of-the-language types, accepted by less stern dictionaries); in UK English it isn't. It would be perverse to insist on using a word in a sense that's used only in US English when there's a perfectly good word that reads much better in other forms of the language as well, so I've changed it back to career.

or factual inaccuracies.

The physical circumference of earth is about 40,000 km or 25,000 miles, so 8,000 miles is roughly a third, not halfway around the world.

Multiple kinds of factors caused *Accuracy* problem claims. Many of them were related to cultural differences, as in the first example. New discoveries and subsequent changes in the subject knowledge might invalidate a reference record for the subject. Alternatively, it could lead to false inaccuracy claims if the user’s stock of knowledge on the subject was not up to date.

4200 years or 2300 years before it comes back? -User X
 Just to note this is now answered in the article - its period was 4200 years but is now 2380 due to gravitational interaction with Jupiter. – User Y

Accessibility IQ problems refer to the instances when claims were made that reference information was not sufficiently accessible. This could happen for a number of reasons. Important information could be hidden deep in an article and difficult to retrieve, such as in the cases below:

As for other encyclopedias, the Encyclopaedia Britannica, for example, includes dates and places of birth and death in brackets after the name in biographies. People check encyclopedia entries for precisely this kind of information: it shouldn't be buried in the text forcing them to search for it, and I don't see how adding the day and month clutters an introduction any more than the year does.

Be specific. Don't just point to a web site where the reader has to spend half a day sorting through irrelevant articles

Information could also be less accessible if it was in a language that common users of the English Wikipedia could not speak and read:

I removed the characters because I think they don't add anything to the article. This is an English-language encyclopedia, so we can expect few people to read Japanese characters. It is better to provide transliteration instead because that will be accessible to all readers.

The other frequently observed cause of an *Accessibility* IQ problem was copyright protection. The community was meticulous in examining posted content, especially photos, for copyright violation. If such instances are found, the material is removed immediately.

An *Authority* problem arose when the authority or reputation of information was questioned:

Shouldn't we note that the "study of authenticity" comes from the Institute of Historical Review, an organization dedicated to denying the Holocaust? I think including it as a link without noting that context might be a little sneaky.

Another most frequently encountered problem type was a *Completeness* IQ problem. The analysis showed that the majority of them occurred due to a missing perspective on the topic:

Big problem right now is the material on the Heroic Age is not linked extensively from the article, which is very heavily focused/biased on the desidaemoniac Hesiod/Homeric Hymns perspective. We seriously need more on the atheistic world of the Iliad and Odyssey, the blasphemies of Thebes, the bizarre crimes of Crete and Argos, etc. before the article itself can be an honest starting point for the whole subject-matter.

Often, the claims of biased coverage or lack of Neutral Point of View (NPOV) were raised when content from a different genre was pasted in a Wikipedia article. Indeed, Wikipedia edit guides clearly point out that Wikipedia content should not be original research¹⁷. In this particular case the content of a student paper was decontextualized and placed into an encyclopedia article without applying necessary changes to conform to the norms of the new context - an encyclopedia genre:

This text is really well written and informative, but I suspect that in it's current state, is more of an essay, rather than an NPOV article. ...it claims to be objective, and in some degree it is, however, this type of writing style, asserts an opinion, or some analysis, but it is debatable whether it is a fact. I think a slight rewording may be beneficial (for example: "we should" is POV), this analysis of alchemy may be dominant today, but may change in a 100 years like it changed a 100 years ago. I will try to reword these paragraphs a bit.

The existence of a *Completeness* problem was claimed when the user felt that the article did not provide enough detail on the topic:

Could someone add the details of the copyright status of the KJV in Great Britain? It is mentioned that it has "special status", but I would like to know more.

Adding too much detail, on the other hand, could lead to *Complexity* problem claims and compromise the goal of a general purpose encyclopedia article – that is, to provide a starting research point in the form of a topic summary:

this section either needs much more work or much simplification ... the U.S. Geological Survey's excellent on line publication This Dynamic Earth provides more than is ever likely to be found here. Perhaps a brief descriptive definition and reference to several such publications would be better

Semantic Consistency problems were claimed when the same concepts and meanings were not conveyed with the same vocabulary or statements within the same article or across the collection. Likewise the *Structural Consistency* IQ problem type refers to quality accidents when different structure, format or precision was used to represent the same elements in the text.

Perhaps this is an odd question, but how exactly do we reference some of the claims we've made in this article? For instance, can we cite the game for all of the course information? Is that original research?

The *Informativeness* problem refers to the cases when the user claimed that the article contained redundant and/or duplicate information:

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

A *Relevance* problem was claimed when the user considered particular information irrelevant to the overall topic of the article:

¹⁷ http://en.wikipedia.org/wiki/Original_research

You removed the image of long nails that I uploaded. Why? – User X

Because I don't quite see the relevance to the article - why did you add it? There are certainly pictures that are much more suitable for the article – User Y

A *Verifiability* problem arose when the user did not feel that information could be promptly and inexpensively verified due to the lack of references and links to original sources:

A complete lack of reference to actual sources. Currently the section reads like an essay, an example of what wikipedia is not. Wikipedia is a secondary source. Addition of actual notable sources holding the POV noted would fix this problem, and the one above.

And, finally, a *Volatility* problem was pointed out when the user felt that frequent vandalisms and edit wars made the article content unstable and not reliable:

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.

The analysis showed that often, when claiming or disputing IQ problems, users were mindful about 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 [Weick95]. 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

This article is rather huge as it is but it set up in a such a way that this could be minimized. Moving that detail to the separate article and leaving a good summary of that article here (several paragraphs or maybe a few short sub-sections), would serve users who just want the summary and those that want the detail about that aspect. This makes the article more useful to more people.

2) Accuracy vs. Accessibility

I removed the characters because I think they don't add anything to the article. This is an English-language encyclopedia, so we can expect few people to read Japanese characters. It is better to provide transliteration instead because that will be accessible to all readers.

The stanza from King Haraldr is in the spelling given in Gordon's Old Norse textbook. I avoided using the hooked o character because it has issues in a number of fonts and browsers; my understanding is that ø is the canonical replacement, though I may be wrong. Thanks for the edits; I borrowed the line from Hávamál from the Auden translation, and I thought it looked funny.

3) Completeness vs. Relevance

I come down on the side of preserving the article's focus on the original punk scene, with brief descriptions of ongoing manifestations of punk and the many punk-influenced movements, including links to more extensive articles on those. I think a narrower focus makes for a clearer article and better history.

4) Accessibility vs. Complexity

People check encyclopedia entries for precisely this kind of information: it shouldn't be buried in the text forcing them to search for it, and I don't see how adding the day and month clutters an introduction any more than the year does.

5) Completeness vs. Appropriateness (Relational Accuracy)

Okay, maybe not the same degree, but same idea - just because it is "offensive" to one group doesn't mean it should be removed if it is relevant.

6) Accessibility vs. Consistency

We really need the caption, otherwise the reader just sees a smoke cloud, not exactly informative in the encyclopedic way - User X .

It's not omitted -- it appears when you hover the mouse over the picture. And the reason I changed it is because this (as the article stands now) is the standard battlebox format (you can find the whole writeup at Wikipedia:WikiProject Battles) - User Y.

7) Completeness vs. Complexity

The problem comes in that this is a WP article and is not supposed to be comprehensive, thus leading to the odd misleading bit a few of which you picked up on. To the extent that some of all this should be in the article, I suggest that you have at it, as I'm likely to leave something out by inadvertence, since 'everyone knows that'. You are likely to serve the Average Reader in this instance better than I.

8) Volatility vs. Accessibility

No they should not!!! We have to remember that this still a wiki and we should not abandon that philosophy. 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. If it gets vandalized we block the vandals, and IF NOTHING ELSE HELPS we can block the page temporarily. Don't let the vandals win!

In [Stvilia04] 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 decision about IQ assurance resource allocation. Examples confirming this proposition were found in the current study as well.

This mediocre article is Featured simply because its subject is interesting and important. The treatment here is not really up to par yet.

In addition, the community was clearly conscious about constant tradeoffs between quality and cost:

Disadvantages of the new "category" feature is that a) you can't list things in alphabetical order b) you can't list things that don't have a page yet (redlinks). Still, it is far too impractical to maintain this page by hand.

3.2 Reevaluating IQ criteria - Featured Article Status Removal Process

Another good source for learning how the Wikipedia community reasons about IQ, what dimensions they consider important and how their notion of 'high quality' changes over time are the Featured Article Removal Candidates (FARC) pages.

As we mentioned earlier one can nominate an already featured article for the removal of a featured status and have the community vote on it¹⁸. To make removal decision-making more consistent, the community has a procedure in place the main points of which are: (a) a removal candidate not meeting the above mentioned FA quality criteria; and (b) communicating quality concerns to the article's contributor community through the article's discussion page and giving them some time to address the concerns before placing the article for removal vote. This section provides some of the findings of the qualitative and quantitative analysis of 120 vote and negotiation instances carried out from July 2004 to May 2005 and documented in the Wikipedia logs.

IQ Problem Types	Status Retained	Status Lost	Total
1. Accessibility	3	3	6
2. Accuracy	6	5	11
3. Authority	2	1	3
4. Completeness	16	49	65
5. Complexity	11	15	26
6. Consistency	8	18	26
7. Informativeness	4	11	15
8. Relevance	3	4	7
9. Verifiability	10	28	38
10. Volatility	1	6	7
<i>Using IQ standards retroactively</i>	5	4	9
<i>IQ Improvement work done</i>	14	3	17

Table 2: Descriptive statistics of IQ problems found in the FARC instances (120 instances).

¹⁸ http://en.wikipedia.org/wiki/Wikipedia:Featured_article_removal_candidates

Table 2 shows that the *Completeness* problem was the most often identified reason for revoking FA status. The other frequently occurring problem type was *Verifiability* which mostly referred to the absence of references in articles.

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 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 principled 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. 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 nominations for removal were matched by editors updating articles and addressing some of the criticism posed in the nominations. As a result, in 14 cases out of those 17 the articles were allowed to retain their FA status (see Table 2).

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>!))-

4. ANALYSIS

Even though there is no straightforward answer to what makes a high quality article, the comparison of the discussion pages of the Featured and Random set articles points to certain factors that might have helped some articles to achieve a featured status. The main distinguishing factor of the Featured articles from the articles in the Random set is a presence of a small core group of users which is relatively homogeneous in terms of sharing social norms of cooperation, including communication protocols. Indeed, the median *Diversity Rate* (*# of Unique Users / Total # of Edits*) measure of the Featured set is 0.4 vs. 0.7 of the Random sample, pointing to more homogeneity in the Featured Article editor groups.

It has been asserted that group cooperation is driven by interdependence in having work done [Schmidt93]. 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. Furthermore, there is an additional mechanism that promotes cooperation. The community has to reach and maintain a consensus on an article being of high quality to have that article nominated for a featured status. Although in certain cases the goal of achieving featured status was a driving force behind the community's effort to cooperate and maintain a consensus, in the majority of the sample this goal was not stated explicitly.

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 and Kincaid readability scores [Gunning52] (see Table 3). It is well understood in CSCW research that effective articulation of work is an essential factor for successful cooperation [Schmidt93]. 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 [Orlikowski94], and to introduce newcomers to those norms and the subject in general. Additionally, it was suggested by [Scacchi04] when discussing FOSS development practices that a FOSS system cannot be sustained unless its developer community reaches a critical mass and they coevolve together. 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.

Article Set	Sample size	Flesch	Kincaid	Length (characters)
Featured	235	35.9	12.5	11,841
Random Sample	128	30.4	13.1	870
Total	363365	34.7	12.6	5,725

Table 3: The median readability measurements of the discussion pages.

We found that voting or polling was used quite often and effectively in the Featured set when users intended to make changes in article content or resolve disputes and disagreements. The protocol is simple. The user notifies the community about his or her intention to change the article content and states the rationale for the change. If no one objects to the proposed change, the user 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 may not always work that smoothly and from time to time an edit war may erupt. But successful communities manage to put a stop to edit wars by educating and/or

persuading fighting sides to resolve their disagreements through a community dispute resolution process:

Can we remove the hideous infobox? For one thing it's redundant; for another, why is that quote more representative than any other and who gets to choose it? - User X
 * Support. I agree for the reasons you've stated. There's no need for it, it doesn't really add anything except by killing some of the whitespace, and the quote seems arbitrary. --User Y
 I removed it but User Z added it back. I have removed it again. User V
 - Though I don't like it, I think we ought to try and talk about it before just removing and reverting. User Y
 @ Was it talked about before adding it? User V
 # Does it really matter? --User Y

Table 1 shows that the share of *Completeness* problem claims was much higher for the Featured than for the Random Sample. One may argue that these claims helped the articles in the Featured set to become more comprehensive and better quality. The median length of Featured set was more than 20 times larger than for the Random set. The readability scores of Featured articles were also better than those of the Random set (See Table 4). However, all those could be an effect of a third factor or variable – Age. A featured article on average was 3 times older than a random one. Furthermore, it could not be identified with certainty when these claims were made - before the article had been featured on the main page or after. Suggestions that featuring on the main page could invite both ‘bad’ (malicious) and ‘good’ edits were found in the discussion pages. As noted earlier, [Lih04] found that citing Wikipedia articles in the main press had a significant impact on the number of edits for those articles. Further qualitative and quantitative analysis is needed to find out whether obtaining a featured status had a similar effect or not, and whether the relatively high average quality of a featured article was due to its maturity and age or better communication practices and work articulation of its editor community.

Statistics	Featured	Random
Flesch	36	27
Kincaid	12	13
Images	5	0
Article size (in # of characters)	24,708	1,358
Age (in # of days)	1,153	385
Num. of Edits	257	8

Table 4. Article statistics (medians)

We found that discussion pages were also used by people to ask a question about a specific part of the article they did not understand, or to inquire about missing details. In some cases a question might not be related to the current content of the article at all, and instead might be seeking an expert opinion on another article or simply on the subject in general, outside of the Wikipedia context.

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 starting point 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. We found that following this convention, Wikipedia articles too might not include detailed explanations or descriptions of the concepts and theories mentioned in the article, making it difficult for a layman to 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.

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 could serve 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 is an economic argument behind it. 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. [Garfinkel67] describes how a decision may come before the information when jurors seek information for justifying an already made decision in retrospect.

We already mentioned earlier several types of tradeoffs identified in the discussion pages, 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. Users 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 a number of 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.

Earlier, based on the results of a multi-agent simulation we suggested that certain types of information-intensive organizational activities could be more prone to quality problems:

- **Representation-dependent activities:** Whenever organizational 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.
- **Relative stability of activities and information:** Whenever information quality depends upon stable properties of the information, its context.
- **Provenance dependent activities:** Whenever information quality depends on the provenance and mediation record of the information [Gasser03].

We found references to all four activities in the community discussion. In some cases users 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 the majority of quality problem types are related to Decontextualizing activities when information brought into an article from an outside source does not match the genre or the cultural or cognitive context of the Wikipedia's common user.

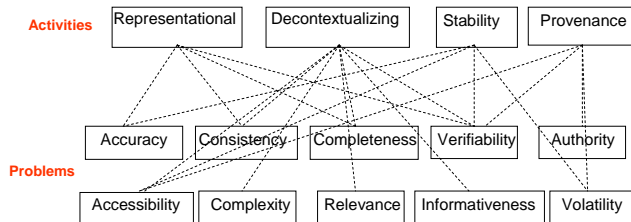


Figure 2: IQ Problem Type – Activity Type Mapping

We believe that a study of these evolving debates and processes 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 [Strong97][Wang96][Wang98]. Wikipedia offers a special insight into these problems. The Featured Page feature, 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 [Twidale99][Twidale 00].

5. CONCLUSION

This initial analysis of the quality of Wikipedia articles helps us to understand the ways in which quality is established and improved despite what seems at first glance the seemingly anarchic operation of the project. Featured articles are 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 have 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 anticipate that subsequent investigations, applying various quantitative and qualitative techniques, can help clarify how quality is understood by participants, the relative importance they give to it, the mechanisms that they initiate and subsequently involve to improve it, and the consequences for those who use Wikipedia as a resource. Finally, these participatory mechanisms, linking discussion of quality and quality maintenance processes directly with the data itself can serve as a useful inspiration for improving conventional datasets.

APPENDIX

	Dimension	Definition
<i>Intrinsic</i>	1. Accuracy/Validity	the extent to which information is legitimate or valid according to some stable reference source such as a dictionary, and/or set of domain constraints and norms (soundness)
	2. Cohesiveness	the extent to which the content of an object is focused on one topic
	3. Complexity	the extent of cognitive complexity of an information object measured by some index/indices
	4. Semantic consistency	the extent of consistency of using the same values (vocabulary control) and elements for conveying the same concepts and meanings in an information object. This also includes the extent of semantic consistency among the different or the same components of the object
	5. Structural consistency	the extent to which similar attributes or elements of an information object are consistently represented with the same structure, format and precision
	6. Currency	the age of an information object
	7. Informativeness / redundancy	the amount of information contained in an information object. At the content level it is measured as a ratio of the size of the informative content (measured in word terms which are stemmed and stopped) to the overall size of an information object. At the schema level it is measured as a ratio of the number of unique elements over the total number of elements in the object
	8. Naturalness	the extent to which an information object's model/schema and content are expressed by conventional, typified terms and forms according to some general purpose reference source

	9. Precision / Completeness	the granularity or precision of an information object's model or content values according to some general purpose IS-A ontology such as WordNet
Relational/Contextual	10. Accuracy	the degree to which an information object correctly represents another information object or a process in the context of a particular activity
	11. Complexity	the degree of cognitive complexity of an information object relative to a particular activity
	12. Accessibility	Accessibility of information relative to a particular activity (speed, ease of locating and obtaining)
	13. Naturalness	the degree to which an information object's model and content are semantically close to the objects, states or processes they represent in the context of a particular activity (measured against the activity/community specific ontology)
	14. Informativeness / Redundancy	the extent to which the information is new or informative in the context of a particular activity/community
	15. Relevance (aboutness)	the extent to which information is applicable and helpful/applicable in a given activity
	16. Precision / Completeness	the extent to which an information object matches the precision and completeness needed in the context of a given activity
	17. Security	the extent of protection of information from harm
	18. Semantic consistency	the extent of consistency of using the same values (vocabulary control) and elements required or suggested by some external standards and recommended practice guides for conveying the same concepts and meanings in an information object
	19. Structural consistency	the extent to which similar attributes or elements of an information object are consistently represented with the same structure, format and precision required or suggested by some external standards and recommended practice guides
	20. Verifiability	the extent to which the correctness of information is verifiable and/or provable
21. Volatility	the amount of time the information remains valid	
Reputational	22. Authority	the degree of reputation of an information object in a given community

Table 5. IQ Assessment Framework (categories and dimensions) [Gasser01]

REFERENCES

- [BAILEY94A] Bailey, K. (1994). *Methods of social research* (4 ed.). New York, NY: The Free Press.
- [BENKLER02] Benkler, Y. (2002). Coase's penguin, or, Linux and The Nature of the Firm. *The Yale Law Journal*, 112(3).
- [BIBER88] Biber, D. (1988). *Variations across speech and writing*. Cambridge, UK: Cambridge University Press.
- [BOWKER94] Bowker, G., Star, S. (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.
- [COLLISON66] Collison, R. (1966). *Encyclopaedias: their history throughout the ages* (2 ed.). New York, NY: Harper.
- [COSLEY05] Cosley, D., Fankowski, D., Kiesler, S., Terveen, L., Riedl, J. (2005). How oversight improves member-maintained communities. In: *Proceeding of the SIGCHI conference on Human factors in computing systems*. Portland, OR 11-20.

- [CRAWFORD01] 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.
- [CROWSTON04] Crowston, K., Scozzi, B. (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.
- [EMIGH05] Emigh, W., Herring, S. (2005). Collaborative authoring on the Web: a genre analysis of online encyclopedias. In: *Proceedings of the 39th Hawaii International Conference on System Sciences*.
- [GARFINKEL67] Garfinkel, H. (1967). *Studies in ethnomethodology*. Prentice-Hall Inc.
- [GASSER01] Gasser, L., Stvilia, B. (2001). A new framework for information quality. Technical report ISRN UIUCLIS--2001/1+AMAS. Champaign, IL: University of Illinois at Urbana Champaign.
- [GASSER03] Gasser, L., Stvilia, B. (2003). Using multi-agent models to understand the quality of organizational information bases over time. In: *Proceedings of the NAACSOS Conference*. Pittsburgh, PA.
- [GUNNING52] Gunning, R. (1952). *Technique of clear writing*. McGraw-Hill.
- [HEYLIGHEN02] Heylighen, F., Dewaele, J. (2002). Variation in the contextuality of language: an empirical measure. *Foundations of Science*, 6, 293-340.
- [JURAN92] Juran, J. (1992). *Juran on quality by design*. New York, NY: The Free Press.
- [LERNER04] Lerner, J., Tirole, J. (2004). The economics of technology sharing: open source and beyond. Working Paper 10956. Retrieved Jun 7, 2005, from <http://www.nber.org/papers/w10956>.
- [LIH04] 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.
- [MCARTHUR86] McArthur, T. (1986). *Worlds of reference: lexicography, learning and language from the clay tablet to the computer*. Cambridge, UK: Cambridge University Press.
- [ORLIKOWSKI94] Orlikowski, W., Yates, J. (1994). Genre repertoire: the structuring of communicative practices in organizations. *Administrative Science Quarterly*, 39, 541-574.
- [RAYMOND98] Raymond, E. (1998). The cathedral and the bazaar. *First Monday*, 3(3)
- [SANDUSKY04] Sandusky, R., Gasser, L., Ripoche, G. (2004). How negotiation shapes coordination in distributed software problem management. In: *Distributed collective practice: Building new directions for infrastructural studies workshop of the CSCW 2004 conference*. Chicago, IL.
- [SCACCHI04] Scacchi, W. (2004). Free/Open Source Software development practices in the computer game community. *IEEE Software*, 21(1).
- [SCHMIDT93] Schmidt, K., Bannon, L. (1993). Taking CSCW seriously: Supporting articulation work. *Computer Supported Cooperative Work*, 1(1-2), 7-40.
- [SCHMIDT96] Schmidt, K., Simone, C. (1996). Coordination Mechanisms: Towards a Conceptual Foundation of CSCW Systems Design. *Computer Supported Cooperative Work*, 5, 155-200.
- [SHREEVES05] Shreeves, S., Knutson, E., Stvilia, B., Palmer, C., Twidale, M., Cole, T. (2005). Is 'quality' metadata 'shareable' metadata? The implications of local metadata practices for federated collections. In: *Proceedings of the Association of College and Research Libraries (ACRL) 12th National Conference*. Minneapolis, MN.
- [SMITH89] 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.
- [STRAUSS78] Strauss, A. (1978). *Negotiations: varieties, contexts, processes, and social order*. San Francisco, CA: Jossey-Bass.

- [STRONG97] Strong, D., Lee, Y., Wang, R. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103-110.
- [STVILIA04] Stvilia, B., Gasser, L., Twidale, M., Shreeves, S., Cole, T. (2004). Metadata quality for federated collections. In: *Proceedings of ICIQ04 - 9th International Conference on Information Quality*. Boston, MA 111-125.
- [TWIDALE99] Twidale, M., Marty, P. (1999). Investigation of data quality and collaboration. Champaign, IL: GSLIS, University of Illinois at Urbana - Champaign.
- [TWIDALE00] Twidale, M.B. & Marty, P.F. (2000). Coping with errors: the importance of process data in robust sociotechnical systems. *Proceedings, CSCW'00*, Philadelphia, 269-278.
- [TWIDALE05] Twidale, M., Nichols, D. (2005). Exploring usability discussion in open source development. In: *Proceedings of Thirty-Eighth Annual Hawaii International Conference on System Sciences, HICSS-38, Track 7*. 198c.
- [VIEGAS04] Viegas, F., Wattenberg, M., Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. In: *Proceedings of CHI 2004*. Vienna, Austria 575-582.
- [WANG96] Wang, R., Strong, D. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5-35.
- [WANG98] Wang, R. (1998). A product perspective on Total Data Quality Management. *CACM*, 41(2), 58-65.
- [WEICK95] Weick, K. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.
- [YIN88] Yin, R. (1988). *Case study research: design and methods*. London, UK: Sage Publications Ltd.