A Breakdown of Quality Flaws in Wikipedia

Maik Anderka Benno Stein

Bauhaus-Universität Weimar
99421 Weimar, Germany
<first name>.<last name>@uni-weimar.de

ABSTRACT

The online encyclopedia Wikipedia is a successful example of the increasing popularity of user generated content on the Web. Despite its success, Wikipedia is often criticized for containing low-quality information, which is mainly attributed to its core policy of being open for editing by everyone. The identification of low-quality information is an important task since Wikipedia has become the primary source of knowledge for a huge number of people around the world. Previous research on quality assessment in Wikipedia either investigates only small samples of articles, or else focuses on single quality aspects, like accuracy or formality. This paper targets the investigation of *quality flaws*, and presents the first complete breakdown of Wikipedia's quality flaw structure. We conduct an extensive exploratory analysis, which reveals (1) the quality flaws that actually exist, (2) the distribution of flaws in Wikipedia, and (3) the extent of flawed content. An important finding is that more than one in four English Wikipedia articles contains at least one quality flaw, 70% of which concern article verifiability.

Categories and Subject Descriptors: H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing; H.7.5 [Document and Text Processing]: Document Capture— *Document Analysis*

General Terms: Measurement, Experimentation

Keywords: Quality Flaws, Information Quality, Wikipedia, User-generated Content Analysis

1. INTRODUCTION

Since its launch in 2001, the Wikipedia project has become the largest and most prominent collaboratively created reference work on the Web. At the time of this writing, Wikipedia comprises about 21 million articles in more than 280 language editions.[1] The huge amount of public available data and the fact that Wikipedia is collaboratively created solely by volunteers have attracted researchers of several academic disciplines.[2]

[1]List of official Wikipedias: `http://meta.wikimedia.org/wiki/List_of_Wikipedias`.

[2]Academic studies of Wikipedia: `http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_in_academic_studies`.

WebQuality '12, April 16, 2012, Lyon, France

Almost all content of Wikipedia can be edited immediately, by everyone, and with minimum effort. This open editing model and the resulting lack of review mechanisms make clear that one of the most important challenges for Wikipedia pertains to the quality of its content. Wikipedia founder Jimmy Wales announced in a recent interview: "Our goal is to make Wikipedia as high-quality as possible. [Encyclopædia] Britannica or better quality is the goal." [25] The improvement of content quality has also been stated as one of five strategic goals that have been established by the Wikimedia Strategy Task Force.[3] That there is actually low-quality content in Wikipedia, is underlined by the fact that, at the time of writing, less than 0.1% of the English Wikipedia articles are labeled as *featured*, i.e., are considered to be well-written, comprehensive, well-researched, and neutral. This begs the question of what is wrong with the remaining 99.9%. Here, we address this question by pinpointing quality flaws that occur in Wikipedia's articles.

The relevant literature mentions a variety of approaches to automatically assess quality in Wikipedia, but only a few of them target the identification of specific quality flaws. A good deal of the existing research targets the classification task "Is an article featured or not?" Although the developed approaches perform nearly perfectly in distinguishing featured articles from non-featured ones, they cannot (and were not intended to) provide a rationale governing the respects in which an article violates Wikipedia's featured article criteria. Prior work on the identification of quality flaws either investigate only small samples of articles [24] or analyze only a restricted set of flaws [2, 3, 11]. To the best of our knowledge, this paper presents the first comprehensive breakdown of quality flaws in Wikipedia.

1.1 Research Questions and Contributions

We conduct an extensive exploratory analysis of the English Wikipedia to investigate the following research questions:

What kinds of quality flaws exist in Wikipedia? We take advantage of the fact that Wikipedia users who encounter a flaw (but who are either not willing or who do not have the knowledge to fix it) can tag the article with a so-called cleanup tag. We implement an automated approach to extract the existing cleanup tags from Wikipedia, which gives us the set of quality flaws that have so far been tagged. Altogether 388 flaws are identified. We organize the flaws along 12 general flaw types, which reveals the quality flaw structure of Wikipedia.

Where do quality flaws occur? About one third of the Wikipedia pages contain the encyclopedic content, while the remaining pages are used for content organization and user discussions. We investigate the distribution of cleanup tags across Wikipedia's names-

[3]Strategic plan of Wikimedia: `http://strategy.wikimedia.org/wiki/Strategic_Plan/Movement_Priorities`.

paces, which reveals that tagging work in Wikipedia mostly targets the encyclopedic content. Moreover, we investigate the distribution of flaws over 24 top-level Wikipedia topics. Among other things, we find that computer- and belief-related articles have been tagged more frequently than, for instance, geography-related articles.

How to quantify the extent of flawed content? By quantifying the scope of the 388 flaws we find that 307 refer to an article as a whole and 81 to a particular text fragment. Moreover, we analyze the incidence of cleanup tags to quantify the flawed content that have been tagged so far. Our analysis reveals that 27.53% of the English Wikipedia articles are tagged to contain at least one quality flaw. However, due the size and the dynamic nature of Wikipedia, it is more than likely that many flawed articles have not yet been identified. We therefore estimate the actual frequency of a flaw from the ratio of flawed to flawless articles. An interesting finding is that every fourth article is expected to contain the flaw *Unreferenced*, i.e., the article does not cite any references or sources.

Our findings give insights to Wikipedia's quality situation, and this way supports Wikipedia's quality assurance activities by revealing weaknesses with respect to the quality of information.

1.2 Paper Organization

The paper is organized as follows. Section 2 gives some background on Wikipedia's definition of information quality and discusses related work on quality assessment. Section 3 describes our cleanup template mining approach and presents the resulting set of flaws organized along our general flaw types. Section 4 shows the distribution of the flaws over Wikipedia's namespaces and over main topics. Section 5 deals with the quantization of the flaws by means of scope and frequency. Finally Section 6 concludes this paper and gives an outlook on future work.

2. RELATED WORK AND BACKGROUND

Information quality is a general concept that combines several criteria such as accuracy, reliability, timeliness, objectivity, completeness, and relevance [18]. A widely accepted interpretation of information quality is the "fitness for use in a practical application" [26], i.e., the definition of information quality depends on the particular context and use case. In Wikipedia the context is well-defined, namely by the encyclopedic genre. It forms the ground for Wikipedia's core content policies: neutral point of view, no original research, and verifiability. The information quality ideal of the English Wikipedia has been formalized—better: made communicable and quantifiable—within the so-called featured article criteria.[4]

The assessment of information quality is becoming a topic of enormous interest. It stands to reason that incorporating quality metrics into information retrieval approaches can significantly improve the search effectiveness of Web search environments [6, 20, 29, 30]. In recent years, quality assessment of user generated content and social media gained particular interest [1, 5, 10, 17]. A large part of existing research on quality assessment in Wikipedia deals with featured article identification. Therefore, articles are analyzed with respect to the number of edits and editors [15, 28], the mutual dependency between article quality and author authority [13], the number of words [7], the character trigrams distribution [16], and particular combinations of several article features [9, 23]. As already motivated, these approaches provide no indication of the shortcomings of non-featured articles, and hence the practical support for Wikipedia's quality assurance process is marginal.

[4]Featured article criteria: `http://en.wikipedia.org/wiki/Wikipedia:Featured_article_criteria`.

Stvilia et al. [24] were the first who systematically analyzed quality problems in Wikipedia. However, they manually analyzed user discussions on the talk pages of only 60 articles, and hence the study must be considered as too specific. In a previous work we propose the usage of cleanup tags to compile the set of quality flaws that exist in Wikipedia [2]. In the prior work only a specific subset of 70 cleanup tags is investigated. Here, we go one step further and extend our previous work with comprehensive exploratory analyses revealing the quality flaw structure of Wikipedia.

Cleanup tags have been utilized previously for different purposes, however, in contrast to this paper, each of the prior studies target only a single cleanup tag. Gaio et al. [11] investigate the usage and the effectiveness of the cleanup tag *Complex* in SimpleWiki, a relative small language edition of Wikipedia that is written in basic English. In a follow-up study the authors apply similar analyses to the English Wikipedia and investigate the cleanup tag *NPOV* (neutral point of view) [22]. They find among others that editing increases after an article has been tagged. Apic et al. [4] propose an indicator of a country's geopolitical instability based on the number of Wikipedia articles that have a link to the Wikipedia article of the respective country and that have been tagged with the cleanup template *NPOV disputes*.

3. COMPILING QUALITY FLAWS

Cleanup tags are a means to tag flaws in Wikipedia. As shown in Figure 1, they are used to inform readers and editors of specific problems with articles, sections, or certain text fragments. However, there is no single strategy to spot the entire set of all cleanup tags. Cleanup tags are realized based on templates, which are special Wikipedia pages that can be included into other pages. Although templates can be separated from other pages by their namespace (the prefix "Template:" in the page title), there is no dedicated qualifier to separate templates that are used to implement cleanup tags from other templates. A complete manual inspection is infeasible as Wikipedia contains nearly 320 000 different templates. We hence introduce an automatic extraction approach that exploits several sources within Wikipedia containing meta information about cleanup tags (Section 3.2). The extracted cleanup tags give us the set of quality flaws that have been identified and tagged by Wikipedia users. We organize the flaws along a small set of general flaw types, to breakdown the quality flaw situation within Wikipedia (Section 3.3). At first, we describe the data underlying our analyses (Section 3.1).

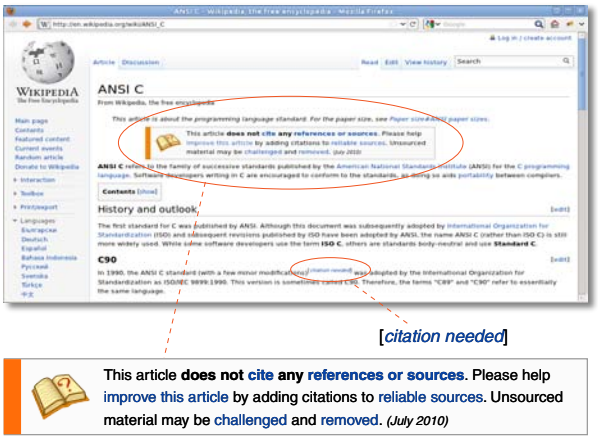

Figure 1: The Wikipedia article "ANSI C" with two cleanup tags. The tag box *Unreferenced* refers to the whole articles and the inline tag *Citation needed* refers to a particular claim.

3.1 Data Base and Preprocessing

To guarantee reproducibility, the analyses in this paper are based on a snapshot instead of investigating Wikipedia up-to-the-minute. Wikipedia snapshots are provided by the Wikimedia Foundation in monthly intervals. We use the English Wikipedia snapshot from January 15, 2011.[5] The English language edition is most appropriate because it is the largest and most popular one. Table 1 summarizes key statistics of the snapshot. A Wikimedia snapshot comprises a complete copy of Wikipedia. The wikitext sources of all pages (and all revisions respectively) are available as a single XML file, which also contains some meta information. Furthermore, several tables of the Wikipedia database are available in the form of SQL dumps, totaling about 40GB. In a preprocessing step, we create a local copy of the Wikipedia database by importing the SQL dumps into a MySQL database. Since we do not target a content analysis, a processing of the XML dumps is not necessary. The local copy of the Wikipedia database allows for efficient analyses, without causing traffic on the Wikimedia servers. Note that all of our analyses can be performed on the original Wikipedia database as well.

Table 1: English Wikipedia snapshot from January 15, 2011.

Number of pages	22 981 145
Number of articles (content pages)	3 557 468
Number of featured articles	3 141
Number of images	847 936
Number of registered users	13 762 201
Number of active users (users who performed an action in the last 30 days)	127 244
Total number of edits	438 200 444
Average number of edits per page	19.07

3.2 Cleanup Tag Mining

We employ a two-step approach to compile the set of cleanup tags: (1) an initial set of cleanup tags is extracted from two meta sources within Wikipedia, and (2) the initial set is further refined by applying several filtering substeps.

Step 1: Extraction The first meta source that we employ is the administration category *Category:Cleanup_templates*, which comprises templates that are used for tagging articles as requiring cleanup. The category also has several subcategories to further organize the cleanup tags by their usage, e.g., inline cleanup templates or cleanup templates for WikiProjects. The page titles of those templates linking to the category or some subcategory are obtained from the local Wikipedia database, using the tables *categorylinks* and *page*, which results in 272 different cleanup tags. The second source that we employ is the meta page *Wikipedia:Template_messages/Cleanup*, which comprises a manually maintained listing of templates that may be used to tag articles as needing cleanup. From a technical point of view, the page is a composition of several pages (transclusion principle). For each of these pages, the content of the revision from the snapshot time is retrieved using the MediaWiki API[6]. A total of 283 different cleanup tags are extracted from the wikitexts of the retrieved pages using regular expressions. Merging the findings from both sources gives 457 different cleanup tags.

Step 2: Refinement A cleanup tag may have several alternative titles linking to it through redirects. For example, the tag *Unreferenced* has the redirects *Unref*, *Noreferences*, and *No refs* among others. We resolve all redirects using the tables *redirect* and *page* of the local Wikipedia database. Moreover, we discard particular subtemplates from the initial set of cleanup tags, namely experimental pages and documentation pages. Experimental pages are identified by the suffixes "/sandbox" and "/testcases" in the page title and are used for testing purposes only. Documentation pages are identified by the suffix "/doc" and provide a template description. We also discard meta-templates, i.e., templates that are solely used either as building blocks inside other templates or to instantiate other templates with a particular parameterization. Meta-templates are derived from the categories *Wikipedia_metatemplates* and *Wikipedia_substituted_templates*. Altogether we collect a set of 388 cleanup tags.

Discussion All documentation pages of the 388 cleanup tags are manually inspected. They give information about purpose, usage, and scope of a template, and our analysis reveals that all tags are indeed related to a particular cleanup task. I.e., each of the 388 cleanup tags defines a particular quality flaw. The flaws are listed in the table in Appendix A (right-most column). Our mining approach does not guarantee completeness though, since the true set of cleanup tags is unknown in general. However, from a quantitative point of view we are confident that we identify the most common cleanup tags, and hence the most important quality flaws.

3.3 Flaw Organization

Several flaws relate to the same type, for instance *Unreferenced* and *Citation needed* both concern article verifiability. To the best of our knowledge, no organization scheme has been proposed before that covers the complete set of quality flaws in Wikipedia. We consider the problem types identified by Stvilia et al. [24] as inappropriate in this respect, as they result from the manual analysis of only 60 Wikipedia talk pages and hence are very specific. Similarly, the set of ten flaw types proposed in our previous work [2] is too specific, as these types target a particular subset of only 70 flaws. At first sight the featured article criteria, may appear as a set of relevant flaw types. However, there are several drawbacks related to this idea: the featured article criteria are not stable, they do not consider technical aspects, and they focus on particularities of high-quality articles. Another idea may be to utilize the organization of the cleanup tags on the meta page *Wikipedia:Template_messages/Cleanup* and consider the section headings as flaw types. However, this organization is mainly intended to serve navigational purposes, and hence the respective sections are very specific. Ultimately, we organize the quality flaws along a newly formed set of 12 general flaw types. Our set is an extension of the ten types proposed in our previous work, and to some extent it covers the featured article criteria, the problem types of Stvilia et al. [24], and the organization of the above mentioned meta page. The table in Appendix A organizes the quality flaws along our flaw types. The labeling is exclusive, i.e., each flaw belongs to exactly one type.

The flaw type *Verifiability* is of particular interest, as the verifiability of information is one of the most important principles of an encyclopedia. The flaws that belong to this type refer to articles that cite no references at all (e.g., *Unreferenced*, *No footnotes*, and *Unreferenced section*), to articles with inadequate and invalid references (e.g., *Refimprove*, *Primary sources*, and *Dead link*), and to unsourced statements within articles (e.g., *Citation needed*, *Who*, and *By whom*). The flaw type *Wiki tech* targets technical aspects of an article, including categorization issues (e.g., *Uncategorized*, *Uncategorized stub*, and *Cat improve*), syntactical problems (e.g., *Cleanup-HTML*), connectivity in terms of Wikipedia-internal links

[5] Wikimedia downloads: `http://download.wikimedia.org`.
[6] MediaWiki API: `http://www.mediawiki.org/wiki/API`.

(e.g., *Orphan*, *Wikify*, and *Dead end*), and ambiguity of links (e.g., *Dn* and *Dablinks*). The flaw type *General cleanup* groups those cleanup tags that either list several flaws in a single tag (e.g., *Multiple issues* and *Expertsubject-multiple*) or merely state that some cleanup is required at all but provide no further information (e.g., *Cleanup* and *Expert subject*). The flaws that belong to the flaw type *Expand* state that particular sections are under-represented or that certain information is missing. The flaw type *Unwanted content* comprises flaws that refer to content that is either not appropriate for an encyclopedia (e.g., *Notability*, *Advert*, and *Original research*) or that is better suited for a different project of the Wikimedia Foundation (e.g., *Copy to Wikiquote* and *Copy to Wikibooks*). The flaw type *Style of writing* targets stylistic issues related to grammar, style, cohesion, tone, and spelling. Most of these issues are described in Wikipedia's manual of style.[7] A fundamental principle of Wikipedia is that articles should be written from a neutral point of view, i.e., representing all significant views unbiased and without opinions. The respective flaws are organized under the flaw type *Neutrality*. The flaws under the type *Merge* refer to articles that deal with a similar subject and hence should be combined. Flaws that focus on a particular topic are organized under the flaw type *Cleanup of specific subjects*. For instance, the flaw *Plot* states that an article's plot summary may be too long or excessively detailed, which may only apply to certain articles describing films or novels for instance. The flaw type *Structure* addresses the articles' organization into sections as well as the length of the sections. For example, an article is expected to have a lead section that summarizes its content. Flaws that address the currency and the lifespan of information are organized under the flaw type *Time-sensitive*. The flaw type *Miscellaneous* comprises flaws that are very specific and that occur relatively infrequently.

4. ANALYZING THE DISTRIBUTION OF QUALITY FLAWS

We analyze the distribution of cleanup tags in Wikipedia and this way reveal where quality flaws are likely to occur. In particular, we investigate the distribution over Wikipedia's namespaces, to see what kind of pages are being tagged (Section 4.1), and the distribution over Wikipedia's top-level topics, to see what flaws occur in which topics (Section 4.2).

4.1 Distribution over Namespaces

The MediaWiki software provides the concept of namespaces as a means to organize the pages from a technical point of view. At the date of the snapshot, the English Wikipedia uses ten basic namespaces, each with a corresponding talk namespace.[8] Table 2 (leftmost column) shows the namespaces organized along three content types: encyclopedia, organization, and community. The namespace "Main" contains the encyclopedic content of Wikipedia. A special subset of this namespace is called *articles*. A page is considered as an article if it belongs to the namespace "Main", is not a redirect page, and contains at least one wiki link.[9] Articles are also called proper content pages. The second content type comprises namespaces that contain meta-pages. These pages are used to organize the encyclopedic content ("Portal", "Category", "Book"), provide policies and support information ("Wikipedia", "Help"), describe non-textual content ("File"), and handle technical and administrative stuff ("Template", "MediaWiki"). The third content type comprises the talk namespaces and the namespace "User". These namespaces contain pages that relate to the community of Wikipedia authors. Each basic namespace has an associated talk namespace, indicated by the suffix "talk". An exception is the talk namespace "Talk" that is associated with the basic namespace "Main". A talk namespace contains the discussion pages for the regular pages of the respective basic namespace. Registered users have the possibility to maintain personal pages, which belong to namespace "User".

Table 2: For Wikipedia's namespaces: the total number of pages, the number of pages that have been tagged with at least one cleanup tag, and the ratio of tagged pages. The namespaces are grouped into three content types: encyclopedia, organization, and community. The namespace "Main" is divided into articles, redirects, and pages that do not contain any wiki link.

Namespace		Pages	Tagged pages	Ratio in %
Encyclopedia				
Main	Articles	3 557 468	979 299	27.53
	Redirects	4 721 190	24	< 0.01
	No wiki link	561	15	2.67
Organization				
File		875 833	32 184	3.67
Wikipedia		607 854	2 321	0.38
Template		329 282	544	0.17
Portal		95 645	217	0.23
Category		666 154	119	0.02
Help		822	15	1.82
Book		2 042	2	0.10
MediaWiki		1 401	0	0.00
Community				
Talk		3 901 638	121 717	3.12
User		1 118 456	8 586	0.77
User talk		6 377 234	3 366	0.05
Wikipedia talk		84 996	733	0.86
Template talk		110 237	193	0.18
Category talk		414 626	44	0.01
File talk		97 162	29	0.03
Portal talk		15 365	12	0.08
MediaWiki talk		829	5	0.60
Help talk		408	3	0.74
Book talk		1 942	0	0.00
		∑ 22 981 145	∑ 1 149 428	Ø 5.00

Table 2 shows the distribution of pages and tagged pages over the namespaces. The snapshot comprises a total of 22 981 145 pages, of which 1 149 428 have been tagged with at least one cleanup tag. The largest namespace is "Main", which comprises 8 279 219 pages (36.03%), including 3 557 468 articles. The namespaces in the content types organization and community comprise 11.22% and 52.75% of the pages respectively. The vast majority (82.23%) of the tagged pages belong to the namespace "Main", and nearly all tagged pages in this namespace are articles. Altogether, 27.53% of the articles have been tagged. Another 10.59% of the tagged pages belong to the namespace "Talk". The majority (84.06%) of these pages are associated with articles, and 92.19% of the articles have a talk page. Note that it is unclear whether a cleanup tag on a talk page refers to the content of the talk page itself or to the content of the associated page. It is a subject of ongoing discussion in the Wikipedia community whether cleanup tags should be placed on articles or on the respective talk pages. The same applies for the namespace "File". A considerable number of file description

[7] Wikipedia article style guide: `http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style`.

[8] There are also two virtual namespaces: "Media" and "Special".

[9] We use the "automatic definition" of an article, which is also used in the MediaWiki software; see: `http://en.wikipedia.org/wiki/Wikipedia:What_is_an_article?`

pages have been tagged, and the cleanup tags might refer to the files themselves or to the descriptions. The remaining namespaces contain relative few tagged pages (less than 1.5%).

Finally, it can be said that cleanup tagging work in Wikipedia mostly targets the encyclopedic content, and articles in particular. The remaining analyses in this paper are restricted to the set of articles because we are particularly interested in quality flaws that occur in proper content pages, as these are the pages that are mostly viewed by typical Wikipedia users.

4.2 Distribution over Topics

We utilize the category system of Wikipedia to derive a set of main topics. The category system is a directed cyclic graph with *Category:Contents* as the (virtual) root. A subcategory of the root category is *Category:Articles*, which is intended as a starting point for browsing categories that contain only articles. This category has two subcategories: (1) *Category:Fundamental categories*, whose subcategories represent fundamental areas of human knowledge, and (2) *Category:Main topic classifications*, whose subcategories are organized thematically and reflect more detailed fields of knowledge. We use the 24 direct subcategories of *Category:Main topic classifications* as top-level-topics, as shown in the leftmost column of Table 3.

To identify the top-level topics of an article, we traverse the category graph bottom-up starting from the categories that are associated with the article. The traversal ends if one path reaches a top-level topic. Note that a single article may have multiple top-level topics, this happens if several paths reach different top-level topics by traversing the same number of categories. The majority (64.12%) of articles have one top-level-topic.

Table 3 shows the distribution of articles as well as of tagged articles over the topics. A good deal of articles belong to the topic "Chronology". These articles describe time periods such as years (e.g., *1895 BC*), days of the year (e.g., *July 4*), and decades (e.g., *70s*). A large number of articles also belong to the topics "People" and "Geography". These articles mainly describe individuals (e.g., *Albert Einstein*) and places (e.g., *Badwater, California*) respectively. The largest proportions of tagged articles are in the topics "Computers" (48.51%) and "Belief" (46.33%). This does not mean that these articles are more likely to be flawed, but rather that the respective topics are more common and hence these articles are more likely to be evaluated with respect to flaws. Moreover, the articles in controversial topics such as "Belief" are more likely to be challenged than those with more agreement such as "Geography".

Table 3: For Wikipedia's top-level topics: the total number of articles, the number of articles that have been tagged with at least one cleanup tag, and the ratio of tagged articles. The rows are ordered by the number of tagged articles.

Topic	Articles	Tagged articles	Ratio, %
Chronology	905 090	237 122	26.20
People	760 687	223 318	29.36
Culture	444 210	158 946	35.78
Geography	726 938	144 157	19.83
Society	415 890	141 319	33.98
Humanities	223 688	79 201	35.41
History	263 038	74 258	28.23
Technology	200 066	71 680	35.83
Arts	192 230	65 828	34.24
Life	208 657	61 124	29.29
Business	149 514	59 794	39.99
Politics	185 226	54 618	29.49
Applied sciences	109 688	40 475	36.90
Nature	154 804	37 044	23.93
Education	95 939	35 431	36.93
Environment	131 091	29 632	22.60
Health	77 010	28 371	36.84
Science	113 521	28 041	24.70
Language	76 882	23 520	30.59
Computers	46 897	22 748	48.51
Agriculture	89 846	20 275	22.57
Law	49 101	17 654	35.95
Belief	23 084	10 695	46.33
Mathematics	23 499	5 974	25.42

5. MEASURING THE EXTENT OF QUALITY FLAWS

In this section, we investigate the scope of the quality flaws (Section 5.1), and analyze the incidence of cleanup tags to quantify the flaws' frequencies (Section 5.2).

5.1 Flaw Scope

The quality flaws differ by their scope. Some flaws refer to the whole article (e.g., *Unreferenced*), others to a certain section (e.g., *Expand section*), and sill others to particular claims (e.g., *Citation needed*) or links (e.g., *Dead link*). Here, we distinguish two scopes: article and inline. The former refers to an article as a whole (or to a certain section), whereas the latter refers to a particular text fragment within an article.

The scope of a flaw is quantified by the kind of cleanup tags that defines the flaw. A cleanup tag is either a tag box or an inline tag, as shown in Figure 1. Tag boxes are placed at the top of an article (or at the top of a section) and hence their scope refers to the whole article (or section). By contrast, inline tags are placed within the text after the sentence, claim, or word they refer to. From a technical point of view both kinds of cleanup tags can be distinguished by the respective meta-templates that are used to implement the tag (e.g., *Ambox* or *Fix*).

The table in Appendix A shows the scope for each quality flaw. From the 388 flaws, 307 are article flaws an 81 are inline flaws. The majority (80.02%) of tagged articles are tagged with an article flaw. In general, inline flaws are more specific than article flaws. For instance, the article flaw *Unreferenced* states that the article does not cite any references or sources. By contrast, the inline flaw *Citation needed* gives a direct indication about a claim that needs to be referenced. Consequentially, it is easier for a human corrector to fix inline flaws. However, some flaws refer to the whole article per definition and hence it is not appropriate to use inline tags for these flaws; consider for instance the flaws that belong to the flaw type *Structure*.

5.2 Flaw Frequency

The table in Appendix A shows for each single quality flaw its absolute frequency, i.e., the number of articles that have been tagged with this flaw. The five most frequent flaws are *Unreferenced*, *Citation needed*, *Orphan*, *Dead link*, and *Refimprove*. 83.21% of the tagged articles suffer from these flaws. The most frequent one is *Unreferenced*; it occurs in 273 230 articles, which corresponds to a fraction of 7.7% of all articles. The table in Appendix A also shows the absolute and relative frequencies for the 12 quality flaw types. The majority of the tagged articles (70.7%) have been tagged with a flaw that concerns the articles' verifiability; 19.46% of all articles are tagged with a particular verifiability flaw. Note that the frequencies of the individual flaws belonging to the same type do not sum up to the absolute frequency of the respective type. This

is due to the fact that some articles are tagged with multiple flaws (multi-labeling). The number of flaws per article differs from 1 to 17; Figure 2 shows the distribution. The majority (74.95%) of the tagged articles are tagged with one flaw.

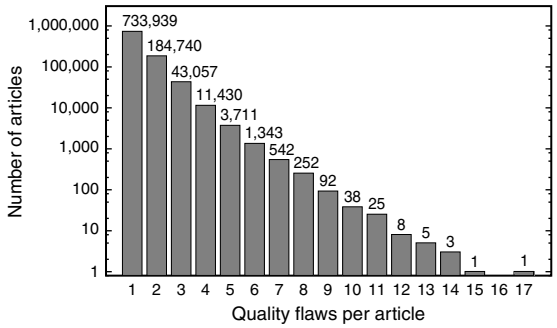

Figure 2: Number of articles over the number of different quality flaws that are tagged in a single article.

Altogether, 979 299 articles (27.53%) are tagged with at least one quality flaw. The number of tagged articles, however, underestimates the actual frequencies of the quality flaws since due to the size and the few control mechanisms in Wikipedia it is more than likely that many flawed articles are not yet identified. Stated formally: Let D be the set of the 3 557 468 Wikipedia articles and let $D^- \subset D$ be the 979 299 tagged articles, see Figure 3. We have no information about the remaining articles $D \setminus D^-$; these articles are either flawless or have not yet been evaluated. The same applies to each single flaw $f_i \in F$, where F denotes the 388 quality flaws; it is unclear whether the articles in $D^- \setminus D_i^-$ either do not contain f_i, or if they have not been evaluated yet with respect to f_i, where $D_i^- \subset D^-$ denotes the articles that have been tagged with f_i. In order to estimate the actual frequency of a flaw f_i we make two assumptions:

1. each article in D^- is tagged completely, i.e. with all flaws that it contains (Closed World Assumption), and
2. the distribution of f_i in D^- is identical to the distribution of f_i in D.

Based on these assumptions, we estimate the actual frequency of a quality flaw f_i by the ratio of articles in D_i^- and articles in D^-. The ratios are shown in the table in Appendix A. For example, the ratio of the quality flaw *Unreferenced* is about 1:3 (273 230 : 979 299). In other words, about every fourth article is expected to contain this flaw.

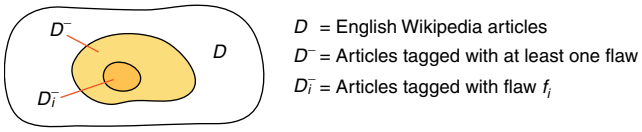

Figure 3: Sets of Wikipedia articles distinguished in this paper.

6. CONCLUSION AND OUTLOOK

This paper gives a breakdown of all quality flaws in the English Wikipedia that were tagged by users in the form of cleanup tags. The analysis is based on the Wikipedia snapshot from January 15, 2011, and we consider it as the first complete compilation of this kind. The key contributions of this paper as well as the respective benefits and applications can be summarized as follows:

- We propose an automatic mining approach to compile an overview of cleanup tags that actually exist in Wikipedia. Wikipedia users currently spend a lot of their time trying to compile such an overview manually, which is only partly successful. Our approach automates this task, relieving users who then hence can spend their time on other jobs. Our approach can be used to generate an up-to-date overview of cleanup tags, whenever a Wikipedia user encounters some flaw and needs to find the respective cleanup tag.

- Wikipedia is often criticized for containing low-quality information, but up to now there is no comprehensive analysis that gives empirical evidence. We close this gap: our quality flaw breakdown and the organization along general flaw types shows the quality flaw structure of Wikipedia. Our compilation will be useful for the development of future quality assurance strategies, e.g., mechanisms and policies to make editors avoid certain types of flaws. In particular, the breakdown shows the types of flaws a user may encounter when searching information in Wikipedia.

- We analyze the distribution of quality flaws in Wikipedia's namespaces as well as in 24 top-level topics. The benefits for the Wikipedia community are twofold: 1. the distribution over namespaces reveals the potential for making cleanup tagging more efficient, e.g., by adjusting the tagging policies to allow cleanup tags either in articles or in associated talk pages. 2. the analysis shows what topics are likely to contain flaws—which is valuable information for a large number of WikiProjects associated with the respective topic.

- We quantify the extent of flawed content in Wikipedia by two measures: flaw scope and flaw frequency. Thus we are the first who give empirical evidence for the amount of low-quality content in Wikipedia. Our analysis reveals that 27.52% of the English Wikipedia articles contain at least one quality flaw, whereas 19.46% have been tagged with a flaw that concerns the articles' verifiability, which is one of the most important principles of an encyclopedia. The actual frequency of the flaws is even higher, it is expected that one out of four articles does not cite any references or sources.

Our findings are relevant for all people who use Wikipedia, including authors, readers, researchers, and data analysts. Moreover, our findings can be beneficial for a variety of information retrieval and machine learning approaches that utilize (possibly flawed) knowledge from Wikipedia; see e.g., [8, 12, 19, 21, 27].

With respect to future work, our quality flaw breakdown forms the basis for an automatic flaw detection. In our current research we use the tagged Wikipedia articles to develop machine learning approaches that predict quality flaws of untagged documents [3]. Our current research also targets the evolution of quality flaws in Wikipedia and the investigation of flawed content over time.

APPENDIX A. Overview of Quality Flaws

The following table shows the 388 quality flaws organized by 12 flaw types. The number of flaws that belong to a particular type is given in parentheses after the type name. The percentage of tagged articles relates to the set of all articles. The scope distinguishes article flaws that refer to the whole article and inline flaws that refer to a certain text fragment, indicated by "a" and "i" respectively. The ratio 1:n (flawed articles : flawless articles) corresponds to the estimated actual frequency of a flaw. Due to space limitations the ratio is only given for $n < 500$.

Flaw type	Tagged articles		Flaws Flaw name (Scope, Tagged articles, Ratio)
Verifiability (66)	692 241	19.46%	Unreferenced (a, 273 230, 1:3), Citation needed (i, 192 295, 1:5), Dead link (i, 92 603, 1:10), Refimprove (a, 89 686, 1:10), BLP sources (a, 36 406, 1:26), No footnotes (a, 26 920, 1:36), Primary sources (a, 21 836, 1:44), BLP unsourced (a, 11 805, 1:82), Unreferenced section (a, 10 686, 1:91), Who (i, 7 819, 1:125), One source (a, 6 406, 1:152), More footnotes (a, 6 122, 1:159), Citations missing (a, 4 962, 1:197), BLP IMDB refimprove (a, 3 335, 1:293), Dubious (i, 3 030, 1:323), Ibid (a, 2 933, 1:333), By whom (i, 2 705, 1:361), Verify source (i, 2 698, 1:362), Ref improve section (a, 2 304, 1:424), Verify credibility (i, 1 936), Citation style (a, 1 927), Cleanup-link rot (a, 1 706), Page needed (i, 1 628), Failed verification (i, 1 427), Volume needed (i, 1 309), Disputed (a, 1 097), Which? (i, 1 056), Specify (i, 813), Cite quote (i, 618), Self-published (a, 601), Unreliable sources (a, 327), Self-published inline (i, 294), Cite check (a, 271), Full (i, 254), Disputed-section (a, 253), Whom? (i, 242), Better source (i, 213), ISBN (a, 212), Citation broken (i, 201), Attribution needed (i, 198), Request quotation (i, 184), Nonspecific (i, 163), Page numbers needed (a, 162), Religious text primary (a, 149), Crystal (a, 83), BLP sources section (a, 67), Citations broken (a, 67), Unreliable medical source (i, 43), Year missing (i, 38), Chronology citation needed (i, 37), Biblio (a, 31), Copyvio link (i, 23), Citation needed (lead) (i, 18), Third-party (a, 12), Date missing (i, 11), Citation not found (i, 5), Too many references (a, 5), Third-party-inline (i, 5), Title missing (i, 4), Author missing (i, 4), Medical citation needed (i, 2), Author incomplete (i, 1), geographical imbalance (i, 0), Title incomplete (i, 0), Translate quote (i, 0), Page numbers improve (a, 0)
Wiki tech (21)	194 649	5.47%	Orphan (a, 166 933, 1:5), Wikify (a, 14 333, 1:68), Dn (i, 8 460, 1:115), Uncategorized (a, 6 252, 1:156), Uncategorized stub (a, 4 121, 1:237), Cat improve (a, 1 261), Dablinks (a, 238), Dead end (a, 116), Overlinked (a, 41), Dead link header (a, 25), Incoming links (a, 14), Newinfobox (a, 6), Cleanup-HTML (a, 4), More-specific-links (a, 1), Recategorize (a, 1), Broken (a, 1), Disambig-cleanup (a, 0), Missing fields (a, 0), Category relevant? (a, 0), Category unsourced (a, 0), Cleanup-infobox (a, 0)
General cleanup (18)	71 401	2.01%	Multiple issues (a, 34 845, 1:28), Cleanup (a, 28 244, 1:34), Expert-subject (a, 4 292, 1:228), Cleanup-rewrite (a, 2 667, 1:367), Lead rewrite (a, 667), Expert-subject-multiple (a, 574), Cleanup-reorganize (a, 476), Expert-verify (a, 233), Cleanup-list (a, 145), Further reading cleanup (a, 86), Prune (a, 36), MOS (a, 35), Cleanup-remainder (a, 24), Cleanup section (a, 22), Remove-section (a, 21), Cleanup AfD (a, 21), Spacing (a, 19), Refactor (a, 1)
Expand (11)	64 450	1.81%	Empty section (a, 46 184, 1:21), Expand section (a, 15 594, 1:62), Expand Spanish (a, 2 579, 1:379), Incomplete (a, 1 136), Year needed (i, 442), Missing information (a, 192), Generalize (a, 116), Data missing (i, 84), Incomplete table (a, 32), Generalize-section (a, 19), List years (a, 8)
Unwanted content (33)	51 130	1.44%	Notability (a, 32 396, 1:30), Advert (a, 7 186, 1:136), Original research (a, 6 630, 1:147), Or (i, 2 254, 1:434), External links (a, 1 794), Howto (a, 357), NOT (a, 337), Syn (i, 208), Importance-section (a, 207), Dicdef (a, 158), Cleanup-spam (a, 119), Non-free (a, 117), Copy to Wikiquote (a, 117), TWCleanup (a, 110), Obituary (a, 66), Relevance note (i, 47), Copy to Wikisource (a, 30), Not English (a, 26), Copy to Wikibooks (a, 20), Copy to Wiktionary (a, 18), TWCleanup2 (a, 17), Cleanup-articletitle (a, 15), Importance-inline (i, 12), Spam link (i, 6), Copy to Wikiversity (a, 6), Schedule (a, 6), Copy to Wikibooks Cookbook (a, 3), Copy to Wikimedia Commons (a, 2), Contact information (i, 2), Now Commons (a, 0), ShadowsCommons (a, 0), Move to userspace (a, 0), Copy to Meta (a, 0)
Style of writing (66)	42 972	1.21%	Clarify (i, 10 289, 1:95), Trivia (a, 4 778, 1:204), Inappropriate tone (a, 4 696, 1:208), When (i, 4 575, 1:214), Context (a, 4 464, 1:219), Copy edit (a, 2 847, 1:343), In-universe (a, 2 597, 1:377), Essay-like (a, 1 641), Prose (a, 1 638), Confusing (a, 1 281), Vague (i, 1 138), In popular culture (a, 570), Rough translation (a, 385), Over detailed (a, 324), Technical (a, 314), Example farm (a, 306), Review (a, 291), Cleanup-jargon (a, 280), Quote farm (a, 236), Quantify (i, 222), Story (a, 220), Off-topic (a, 201), Cleanup-tense (a, 170), Magazine (a, 144), Travel guide (a, 142), Buzzword (a, 122), Contradict-other (a, 120), Inappropriate person (a, 116), Copy edit-section (a, 113), Contradict (a, 92), Misleading (a, 75), Elucidate (i, 72), Ambiguous (i, 66), Jargon-statement (i, 54), Incoherent (a, 45), Repetition (a, 43), Too many see alsos (a, 41), Editorial (a, 40), Examples (i, 38), Abbreviations (a, 38), Manual (a, 36), Example needed (i, 35), Unclear section (a, 33), Time-context (a, 29), Contradiction-inline (i, 27), Expand acronym (i, 25), Textbook (a, 24), Inconsistent (i, 18), Debate (a, 17), Definition (a, 16), Context-inline (i, 10), Directory (a, 10), Term paper (a, 10), Incoherent-topic (a, 9), Bio-context (a, 9), Contradict-other-multiple (a, 7), Context needed (i, 5), Awkward (i, 4), Pro and con list (a, 4), Colloquial (a, 2), Off-topic-inline (i, 1), Specific time (i, 1), Over explained (i, 1), Capitalization (a, 1), Too abstract (a, 1), Debate-section (a, 0)
Neutrality (35)	18 023	0.51%	POV (a, 5 327, 1:183), COI (a, 2 856, 1:342), Globalize (a, 2 824, 1:346), Peacock (a, 1 182), POV-check (a, 1 119), POV-section (a, 1 023), Weasel (a, 851), Weasel-inline (i, 716), News release (a, 516), POV-statement (i, 437), Says who (i, 401), Fanpov (a, 395), Autobiography (a, 373), Unbalanced (a, 323), Criticism section (a, 273), Peacock term (a, 233), Why? (i, 233), Recentism (a, 220), Undue (a, 154), NPOV language (a, 89), Lopsided (i, 76), Puffery (a, 55), POV-title (a, 36), Editorializing? (i, 32), Geographical imbalance (a, 32), Coat rack (a, 31), POV-lead (a, 26), News release section (a, 20), Cherry picked (a, 10), ASF (i, 7), Mission (a, 6), Cleanup-weighted (a, 3), Criticism title (a, 1), Strawman (a, 0), Cleanup-weighted-section (a, 0)
Merge (6)	15 251	0.43%	Merge to (a, 7 093, 1:138), Merge (a, 4 480, 1:218), Merge from (a, 3 738, 1:261), Merging (a, 1), Merged-to (a, 0), Merged-from (a, 0)
Cleanup of specific subjects (60)	7 474	0.21%	Plot (a, 2 469, 1:396), Like resume (a, 799), Cleanup FJ biography (a, 753), Cleanup-school (a, 353), Famous (a, 329), Famous players (a, 315), Game cleanup (a, 261), In-universe/Television (a, 242), In-universe/Dungeons & Dragons (a, 242), Mileposts (a, 188), In-universe/Three Kingdoms (a, 181), In-universe/Comics (a, 161), All plot (a, 157), Game guide (a, 144), Cleanup-biography (a, 135), Album ratings prose (a, 107), CIA (a, 94), Local (a, 78), Fiction (a, 77), USRD-wrongdir (a, 75), In-universe/Literature (a, 73), Episode (i, 63), In-universe/Anime and manga (a, 62), Cleanup Congress Bio (a, 59), Cleanup-university (a, 54), Cleanup-tracklist (a, 48), Alumni (a, 45), Where is it (a, 44), In-universe/Star Trek (a, 36), ToLCleanup (a, 30), In-universe/Video game (a, 28), In-universe/Star Wars (a, 25), Fictionrefs (a, 21), Cleanup-London (a, 20), Animals cleanup (a, 17), In-universe/Film (a, 14), In-universe/Sopranos (a, 10), No plot (a, 10), In-universe/Tolkien (a, 9), In-universe/Transformers (a, 9), Aero-table (a, 8), Cleanup-comics (a, 8), Cleanup-chartable (a, 7), Cleanup-ICHD (a, 6), Cleanup-GM (a, 6), In-universe/Warhammer (a, 6), Cleanup-book (a, 5), In-universe/Discworld (a, 4), NCBI taxonomy (a, 4), Season needed (i, 4), Book-fiction (a, 4), Nonfiction (a, 3), Religion primary (a, 3), Symbolism (a, 2), Film-fiction (a, 1), Hadith authenticity (a, 0), Ship infobox request (a, 0), Kmposts (a, 0), Single infobox request (a, 0), AnimalsTaxobox (a, 0)
Structure (16)	7 280	0.20%	Lead too short (a, 4 144, 1:236), Lead missing (a, 1 615), Sections (a, 852), Very long (a, 378), Lead too long (a, 186), Longish (a, 37), Condense (a, 36), Sub-sections (a, 36), Cleanup-combine (a, 27), Summarize section (a, 18), Inadequate lead (a, 17), Section-diffuse (a, 9), Summary style (a, 2), Verbose (i, 0), Section-sort (a, 0), Too-many-boxes (a, 0)
Time-sensitive (6)	6 185	0.17%	Update (a, 4 348, 1:225), Update after (i, 1 051), Out of date (a, 809), Recently revised (a, 39), Unclear date (a, 13), Anachronism (a, 5)
Miscellaneous (50)	2 208	0.06%	Cleanup-laundry (a, 531), Split (a, 260), Split section (a, 255), Cleanup-gallery (a, 240), Copypaste (a, 110), Split-apart (a, 75), Translation WIP (a, 72), Cleanup-translation (a, 63), Sync (a, 59), Duplication (a, 59), ORList (a, 57), MOSLOW (a, 50), TBD (i, 48), List to table (a, 45), Need-IPA (i, 41), Cleanup-IPA (a, 30), Too many photos (a, 29), List dispute (a, 23), Cleanup-images (a, 21), Split sections (a, 20), Close paraphrasing (a, 18), Disputed-list (a, 15), Create-list (a, 14), Summarize (a, 14), Reqmap (a, 11), Bad summary (a, 8), Metricate (a, 8), NFimageoveruse (a, 7), No definition (a, 6), Cleanup split (a, 6), Pronunciation needed (i, 6), Icon-issues (a, 5), Formula missing descriptions (a, 4), Cleanup-colors (a, 3), Infobox requested (a, 3), TranslatePassage (a, 2), Split dab (a, 1), Reqscreenshot (a, 1), Dubious conversion (i, 1), Foreign (a, 1), Inline need translation (i, 1), Integrate (a, 1), Bad unit conversions (a, 1), External links-inline (i, 0), Convert to SVG and copy to Wikimedia Commons (a, 0), Image requested (a, 0), Reqdiagram (a, 0), Romanization needed (i, 0), Repair coord (a, 0), Translated page (a, 0)
Total over all types	979 187	27.52%	

7. REFERENCES

[1] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding high-quality content in social media. In *Proceedings of the conference on Web search and Web data mining (WSDM'08)*, pages 183–194, 2008. ACM.

[2] M. Anderka, B. Stein, and N. Lipka. Towards automatic quality assurance in Wikipedia. In *Proceedings of the 20th conference on World Wide Web (WWW'11)*, pages 5–6, 2011. ACM.

[3] M. Anderka, B. Stein, and N. Lipka. Detection of text quality flaws as a one-class classification problem. In *Proceedings of the 20th ACM conference on information and knowledge management (CIKM'11)*, pages 2313–2316, 2011. ACM.

[4] G. Apic, M. Betts, and R. Russell. Content disputes in Wikipedia reflect geopolitical instability. *PLoS ONE*, 6(6), 2011.

[5] R. Baeza-Yates. User generated content: how good is it? In *Proceedings of the 3rd workshop on information credibility on the Web (WICOW'09)*, pages 1–2, 2009. ACM.

[6] M. Bendersky, W. Croft, and Y. Diao. Quality-biased ranking of Web documents. In *Proceedings of the 4th ACM conference on Web search and data mining (WSDM'11)*, pages 95–104, 2011. ACM.

[7] J. Blumenstock. Size matters: word count as a measure of quality on Wikipedia. In *Proceedings of the 17th conference on World Wide Web (WWW'08)*, pages 1095–1096, 2008. ACM.

[8] D. Carmel, H. Roitman, and N. Zwerdling. Enhancing cluster labeling using Wikipedia. In *Proceedings of the 32rd ACM SIGIR conference on research and development in information retrieval (SIGIR'08)*, pages 139–146, 2008. ACM.

[9] D. Dalip, M. Gonçalves, M. Cristo, and P. Calado. Automatic quality assessment of content created collaboratively by Web communities: a case study of Wikipedia. In *Proceedings of the 9th ACM/IEEE-CS joint conference on digital libraries (JCDL'09)*, pages 295–304, 2009. ACM.

[10] F. Figueiredo, F. Belém, H. Pinto, J. Almeida, M. Gonçalves, D. Fernandes, E. Moura, and M. Cristo. Evidence of quality of textual features on the Web 2.0. In *Proceedings of the 18th ACM conference on information and knowledge management (CIKM'09)*, pages 909–918, 2009. ACM.

[11] L. Gaio, M. den Besten, A. Rossi, and J. Dalle. Wikibugs: using template messages in open content collections. In *Proceedings of the 5th symposium on wikis and open collaboration (WikiSym'09)*, pages 14:1–14:7, 2009. ACM.

[12] T. Gottron, M. Anderka, and B. Stein. Insights into explicit semantic analysis. In *Proceedings of the 20th ACM conference on information and knowledge management (CIKM'11)*, pages 1961–1964, 2011. ACM.

[13] M. Hu, E. Lim, A. Sun, H. Lauw, and B. Vuong. Measuring article quality in Wikipedia: models and evaluation. In *Proceedings of the 16th ACM conference on information and knowledge management (CIKM'07)*, pages 243–252, 2007. ACM.

[14] A. Kittur, B. Suh, B. A. Pendleton, and E. H. Chi. He says, she says: conflict and coordination in Wikipedia. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '07)*, pages 453–462, 2007. ACM.

[15] A. Lih. Wikipedia as participatory journalism: reliable sources? Metrics for evaluating collaborative media as a news resource. In *Proceedings of the 5th symposium on online journalism*, pages 16–17, 2004.

[16] N. Lipka and B. Stein. Identifying featured articles in Wikipedia: writing style matters. In *Proceedings of the 19th conference on World Wide Web (WWW'10)*, pages 1147–1148, 2010. ACM.

[17] Y. Lu, P. Tsaparas, A. Ntoulas, and L. Polanyi. Exploiting social context for review quality prediction. In *Proceedings of the 19th conference on World Wide Web (WWW'10)*, pages 691–700, 2010. ACM.

[18] S. Madnick, R. Wang, Y. Lee, and H. Zhu. Overview and framework for data and information quality research. *Journal of data and information quality*, 1(1):1–22, 2009.

[19] D. Milne and I. H. Witten. Learning to link with Wikipedia. In *Proceedings of the 17th ACM conference on information and knowledge management (CIKM'08)*, pages 509–518, 2008. ACM.

[20] A. Pirkola and T. Talvensaari. A topic-specific Web search system focusing on quality pages. In *Proceedings of the 14th European conference on research and advanced technology for digital libraries (ECDL'10)*, pages 490–493, 2010. Springer.

[21] M. Potthast, B. Stein, and M. Anderka. A Wikipedia-based multilingual retrieval model. In *Proceedings of the 30th European conference on IR research (ECIR'08)*, pages 522–530, 2008. Springer.

[22] A. Rossi, L. Gaio, M. den Besten, and J. Dalle. Coordination and division of labor in open content communities: the role of template messages in Wikipedia. In *Proceedings of the 38th annual Hawaii conference on system sciences (HICSS'10)*, pages 1–10, 2010. IEEE CS.

[23] B. Stvilia, M. Twidale, L. Smith, and L. Gasser. Assessing information quality of a community-based encyclopedia. In *Proceedings of the conference on information quality (ICIQ'05)*, pages 442–454, 2005. MIT.

[24] B. Stvilia, M. Twidale, L. Smith, and L. Gasser. Information quality work organization in Wikipedia. *Journal of the American society for information science and technology*, 59(6):983–1001, 2008.

[25] The Wall Street Journal. Jimmy Wales on Wikipedia quality and tips for contributors. November 2009. URL: http://blogs.wsj.com/digits/2009/11/06/jimmy-wales-on-wikipedia-quality-and-tips-for-contributors.

[26] R. Wang and D. Strong. Beyond accuracy: what data quality means to data consumers. *Journal of management information systems*, 12(4):5–33, 1996.

[27] P. Wang and C. Domeniconi. Building semantic kernels for text classification using Wikipedia. In *Proceedings of the 14th ACM SIGKDD conference on knowledge discovery and data mining (KDD'08)*, pages 713–721, 2008. ACM.

[28] D. Wilkinson and B. Huberman. Cooperation and quality in Wikipedia. In *Proceedings of the 3rd symposium on wikis and open collaboration (WikiSym'07)*, pages 157–164, 2007. ACM.

[29] Y. Zhou and W. Croft. Document quality models for web ad hoc retrieval. In *Proceedings of the 14th ACM conference on information and knowledge management (CIKM'05)*, pages 331–332, 2005. ACM.

[30] X. Zhu and S. Gauch. Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web. In *Proceedings of the 23rd ACM SIGIR conference on research and development in information retrieval (SIGIR'00)*, pages 288–295, 2000. ACM.