Wikipedia research and tools: Review and comments

Finn Årup Nielsen

January 24, 2019

Abstract

I here give an overview of Wikipedia and wiki research and tools. Well over 1,000 reports have been published in the field and there exist dedicated scientific meetings for Wikipedia research. It is not possible to give a complete review of all material published. This overview serves to describe some key areas of research.¹

1 Introduction

Wikipedia has attracted researchers from a wide range of diciplines—phycisist, computer scientists, librarians, etc.—examining the online encyclopedia from a several different perspectives and using it in a variety of contexts.

Broadly, Wikipedia research falls into four categories:

- 1. Research that examines Wikipedia
- 2. Research that uses information from Wikipedia
- 3. Research that explores technical extensions to Wikipedia
- 4. Research that is using Wikipedia as a resource for communication

Research that examines Wikipedia look on how the encyclopedia evolves, how the users interact with each other, how much they contribute and most of this kind of research is not interested in the content per se. Vandalism in Wikipedia presents usually not a problem: It will just be one more aspect to investigate. Work in this area may benefit Wikipedia as it could help answering what rules the site should operate under, e.g., how beneficial is open editing with no user registration. One representative publication in this category is Jakob Voss 2005 article Measuring Wikipedia. One humorous quote is "Wikipedia cannot work in theory, but does in practice": The major topic in this line of

research is to find the answer to the question "why does it work at all?"

Research using information from Wikipedia will typically hope for the correctness and perhaps completeness of the information in Wikipedia. Many aspects of Wikipedia can be used, not just the raw text, but the links between components: Language links, categories and information in templates provide structured content that can be used in a variety of other applications, such as natural language processing and translation tools. Large-scale efforts extract structured information from Wikipedia and link the data and connect it as *Linked Data*. The papers with description of *DBpedia*.² represent examples on this line of research.

It is odd to write a Science 1.0 article about a Web 2.0 phenomenon: An interested researcher may already find good collaborative written articles about Wikipedia research on Wikipedia itself, see Table 1. These articles may have more complete and updated lists of published scientific work on Wikipedia, and much research-like reporting on Wikipedia of relatively good quality occurs outside ordinary academic channels, — on web-pages and blogs, and several non-academic organization have produced reports from large surveys, e.g., Pew Re-

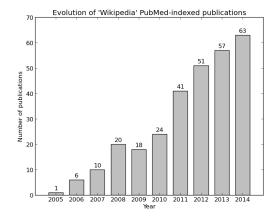


Figure 1: Number of scientific articles returned from the PubMed bibliographic database with a query on 'Wikipedia'.

 $^{^{\}rm i} {\rm The}$ present version is a working paper and continues to be expanded and revised. Revision: 1.622

Wikipedia article	Description
m:Research:Index	Primary entry point for Wikimedia research
en:Wikipedia	Main article about the encyclopedia
en:Reliability of Wikipedia	English Wikipedia article about an aspect of Wikipedia
en:Criticism of Wikipedia	
en:Academic studies about Wikipedia	
en:User:Moudy83/conference papers	Long list of Wikipedia conference papers
$en: User: No September/The_No September_Admin_Project$	
en:Wikipedia:Academic studies of Wikipedia	Comprehensive list of studies on Wikipedia
en:Wikipedia:Ethically researching Wikipedia	
en:Wikipedia:Modelling Wikipedia's growth	Specific results on the growth of Wikipedia
en:Wikipedia:Notability (academics)	Notability guideline for academics
en:Wikipedia:Researching Wikipedia	Discusses quantitatively measures and links to various statistics
en:Wikipedia:Survey_(disambiguation)	
en:Wikipedia:Wikipedia as an academic source	List of papers
en:Wikipedia:Wikipedia in research	Essay
en:Wikipedia:WikiProject Wikidemia	
en:Wikipedia:WikiProject Countering systemic bias	
en:Wikipedia:WikiProject Vandalism studies	Studies of damaging edits
m:Research	List resources for wiki research and researchers
m:Wiki Research Bibliography	Bibliography of scholar and science articles
m:Wikimedia Foundation Research Goals	
m:Research:data	Overview of Wikipedia-related data
en.wikiversity.org/wiki/Portal:Wikimedia Studies	
s:Wikimedia-pedia	Overview of research questions

Table 1: Wikimedia articles related to Wikipedia research. Some of these articles are in the main namespace, while others require the Wikipedia: namespace prefix, while others (m: prefixed) are on the meta wiki (meta.wikimedia.org).

search Center's Internet & American Life Project^{3,4} or Wikimedia Foundation and its chapters.⁵

Wikipedia research continues to grow and now there are many thousands of research articles. In July 2010 Google Scholar claimed to return 196,000 articles when Queried about Wikipedia, while Pubmed returns 55. In May 2012 I found 114 articles in PubMed, see also Figure 1 on page 1.

A few researchers have examined the development of research literature on Wikipedia through time. Han-Teng Liao reporting on his blog in 2012 found the number of theses from major Chinesespeaking regions to peak in 2009.⁶

Several researchers have already reviewed the Wikipedia and/or wiki research literature with varying degree of depth and breadth: A short 2009 report found 1'000 articles,⁷ and identifying 400 peer-reviewed articles a short 2009 paper summarized some of these articles.⁸ Another 2009 focused

entire on the use of Wikipedia-derived data, e.g., for natural language processing and ontology building. On the other hand Nicolas Jullien's 86-page long working paper¹⁰ focused on, e.g., editor motivation, process, contributor roles and quality. A 2012 version of the present paper¹¹ together with a number of other initial papers by Montreal and Finnish research^{7,8,12,13} were evolved into a 138-page working paper. ¹⁴ This lengthy paper was later split into several peer-reviewed journal papers. ^{15,16}

Why is Wikipedia research of interest?

Why does Wikipedia attract researchers? The popularity of the phenomenon probably attracts many, and most Wikipedia articles makes sense for the 'common researcher' in contrast to, say, bioinformatic databases that typically require ex-

WMF data sets	Description
Wikimedia Downloads	Database backup dumps and static HTML dumps
Page view statistics ('raw')	Desktop raw page view statistics
Page view statistics ('all')	Full all page view statistics
Repackaged page view statistics	Page view statistics in a more compressed format.
English Wikipedia pageviews by second ¹⁷	Total page view statistics from the English Wikipedia by timestamp collected in March and April 2015
Wikipedia Clickstream	Referer-resource pairs from the request log
Mediacounts	Requests of media files on upload.wikimedia.org
Third-party data sets	Description
Wikipedia XML Corpus ¹⁸	Annotated in XML.
DeepDive Open Datasets (WIKI)	Natural language processing-annotated sentences from Wikipedia
Scholarly article citations in Wikipedia	Citation from Wikipedia articles to journal articles identified by PubMed, PubMed Central or DOI identifier.
Structured citations in the English Wikipedia	Citation metadata from the English Wikipedia dump.
enwiki-pageviews2007-2016	Pageview statistics for the English Wikipedia collected by Alex Druk in connection with www.wikipediatrends.com.
WikiCite data dumps	Dump of the bibliographic data in Wikidata.

Table 2: Wikipedia data sets.

pert biomedical knowledge. Compared to Free and Open Source Software Wikipedia does not require familiarity with software. 19

The openness of the project and easy availability of data also makes Wikipedia of interest. Typically, research on the Web requires crawling many sites, whereas each complete language version of Wikipedia lies available for download as a compressed XML file ready for use and analysis. Other Web 2.0 large-scale data sets, e.g., from Facebook, may simply not be available for researchers, unless they get exceptional access to the data of the private company. The MediaWiki software also makes Wikipedia available available through the many-faceted API. The *Toolserver* facility, that was used by many Wikipedia programmers and some researchers, even enabled direct database queries. The Toolserver closed, but a similar service, Wikimedia Tool Labs, took over its functionality.

Multiple language editions make it possible to explore areas of language translation. Although the text of Wikipedia is not sentence aligned the large number of languages covered makes Wikipedia unique. With Wikidata multilingual labels for each

concept can easily be obtained.

The availability of the revision history enables dynamic studies of content and contributors. In this aspect Wikipedia is not unique as studies of free and open source software based in public code repositories also make similar research possible.

The structured information in Wikipedia provided through categories and MediaWiki templates may also help researchers, and the sister project Wikidata provides an even more structured resource.

Researchers have also noted difficulties doing research with Wikipedia: As it is constantly changing a researcher may write about Wikipedia and later find it to be untrue, and the vast scale makes it difficult to do studies on the entire Wikipedia. 20

Tools and data sets

dumps.wikimedia.org provides compressed XML files of complete Wikipedias and their sister projects. This data contain the raw wiki markup along with metadata. Independent researchers have converted the wiki markup to an XML format, so that, e.g., links, lists and paragraphs are indicated

with XML tags rather than with wiki markup. They refer to this processed data as the "Wikipedia XML Corpus", ¹⁸ an one set of data from 2006 Wikipedias are available from http://www-connex.lip6.fr/~denoyer/wikipediaXML/.

There are a number of other datasets, e.g., dumps.wikimedia.org/other/articlefeedback/ has article feedback data and https://frdata.wikimedia.org/ has data Wikimedia Foundation fundraising data. The MediaWiki API enables specialized queries, e.g., to get backlinks, blocks or use of templates, and to return the data in different formats such as XML and JSON.

Wikimedia makes usage statistics available to a certain extent, though full logs are not released in the open due to (the readers') pri-Simple page view counts are available vacy. from dumps.wikimedia.org/other/pagecounts-These statistics was originally collected by database engineer and former Wikimedia trustee Domas Mituzas (and distributed from dammit.lt/wikistats/). Further processing of this data is presented at http://stats.grok.se. JSON formatted data are available from stats.grok.se/json/. In September 2014 Andrew West discovered that these services underreported, — they did not count mobile views. Given the rise in mobile traffic the August 2014 the number of page views for the English Wikipedia view was underestimated considerable, — with about a third.²¹ New complete statistics are made available from https://dumps.wikimedia.org/other/pagecountsall-sites/. This includes, apart from page views statistics to the default sites and to the mobile sites, also the Wikipedia Zero views.

With enabling a click tracking extension Media-Wiki administrators can track users' navigation around the wiki. During Wikipedia Usability Initiative Wikimedia enabled the extension for beta testing.

There are several derived data sets. The Wikilinks data set provides links from entities on webpages to English Wikipedia articles. It comprises around 40 million mentions and 3 million entities.^{22,23}

perlwikipedia, as the name implies, works with Perl and recent change patrolling program has been using it. Pywikipediabot is a Python-based collection of tools for bot programming on Wikipedia and other MediaWikis. It can, e.g., create interlanguage links. WikiXRay is another software tool written in Python and R which may download and process data from the wikimedia site for generating graphics and data files with quantitative results. Markus Krötzsch's Java library Wikidata Toolkit can download and parse Wikidata dump files.

The command-line program wikipedia2text downloads a specified Wikipedia article and formats it for display on the command-line, while WikipediaFS makes raw text Wikipedia articles available under the Linux filesystem so a Wikipedia articles looks like an 'ordinary' file.

Edits on Wikipedias are relayed to a IRC server. Researcher may grap this real-time data, e.g.: In his dynamically updated webpage from 2011, *Wikistream*, Ed Summers, presents an aggregated continously updated list with links to edited articles across the major language versions of Wikipedia.

The Wikimedia Toolserver was a platform for hosting various software tools written and used by Wikimedia editors. Programmers could apply for an account. It ran several specialized Web services, e.g., a user edit counter, and CatScan which enabled searching categories recursively. randombio of mzmcbride could sample randomly among bibliographies of living persons. See Table 3 on page 5 for other Web services. The Toolserver has now been replaced by Wikimedia Tool Labs and many of the Toolserver tools have been migrated there, e.g., catscan is now running from https://tools.wmflabs.org/catscan2/catscan2.php.

Research on 'human subjects'

Several researchers have mentioned the problem of gaining access to users for surveys or interviews. The randomization tool on Wikipedia allows an unbiased sampling among registered editors. On the English Wikipedia the randomization tool has the address http://en.wikipedia.org/wiki/Special:Random/User. After sampling a set of editors researchers can contact the them by adding notes on the user talk pages or by emailing the user. However, editors may simply ignore the request, regard it as 'survey spamming', or simply have left the project. Emailing may also be hindered as the emailing facility in MediaWiki as users opt-in on the email contact ability.

Interview- or survey-based research may require the approval of the institutional review board as well as a recruitment form signed by the subject participating in the survey. Another issue raised in a blog post by Heather Ford in 2013 is whether to "onymize", pseudonomize or anonymize, the research subjects. ²⁴ In Wikipedia research, as well as in Internet research generally, anonymization of a quote is difficult as the present day advanced Internet search engine usually has no problem tracking down text fragments. With the attribution inherent in the GPL and Creative Commons licences users may even insist on being attributed, — to quote Ford's experience: ²⁴

Name	Developer	Description
Article revision statistics	soxred93	Detailed overview of edits of a page, interactive graphs for edits over time, article size, and top page editors
Wikipedia page history statistics	aka	Detailed overview of edits of a page with graphs for edits over time, and page editors
X!'s Edit Counter	Several	Edit count for user with graphs over time
Edit Summary calculator	soxred93	Statistics about the edit summary with graph
Contributors	Daniel Kinzler	Ranked list of contributors for a page
Recent change statistics	aka	Summarize recent edits (on German Wikipedia)
Revvis	Finn Årup Nielsen	Sequential collaboration network visualization
Watcher	mzmcbride	Displays the number of users watching a page
Quarry	YuviPanda	Web-based SQL queries to Wikimedia databases

Table 3: The no longer functioning Toolserver and related Web services. See a large list at http://www.mediawiki.org/wiki/Toolserver/List_of_Tools.

I had thought that this was the right thing to do: to anonymize the data, thus protecting the subjects. But the 'subject' was angry that he had been quoted 'without attribution'. And he was right. If I was really interested in protecting the privacy of my subjects, why would I quote his sentence when anyone could probably Google it and find out who wrote it.

Books

A number of longer works describe Wikipedia and related topics from different aspects. New media journalist Andrew Lih's book *The Wikipedia Revolution*²⁵ recounts the history of Wikipedia and its controversies. *How Wikipedia works*²⁶ by Wikipedians focuses on editing, policy and the community, while *MediaWiki*²⁷ and *Working with MediaWiki*²⁸ describes the MediaWiki software. The first book, *The Wiki Way*, presents the most central ideas of the wiki.²⁹ *Critical Point of View: A Wikipedia Reader* is a edited book discussing a number of issues.³⁰ There are several other books published.^{31–33}

Scientific meetings

Several dedicated scientific meeting centers around wikis and Wikipedia. The ACM affiliated meeting WikiSym presents results in all areas of wiki research. Since 2006 the SemWiki workshop has presented results in semantic wiki research. The research community around that meeting also interacts at the semanticweb.org site, — itself a semantic wiki. The WikiAI workshop concentrates on the interface between artificial intelligence with ma-

chine learning, computational linguistics on the one side and wiki and other collaborative-built knowledge bases on the other side. Workshop on Collaboratively Constructed Semantic Resources focuses on social semantics and social natural language processing with several contribution centered on the Wikipedia corpus. The 2009 workshop The People's Web Meets NLP: Collaboratively Constructed Semantic Resources also featured Wikipedia research content.

The community meeting Wikimania focuses on Wikipedia and its sister Wikimedia foundation operated projects. Apart from community-related topics the meeting usually has a good deal of research-oriented material presented. The organizers publish no formel proceedings, but the Wikimania site has usually some description of the contributions, and Wikimedia Commons make videos from recent meetings available.

Other communication channels

WikiSym mailing list wiki-research has little activity, while Wikimedia mailing lists wiki-research-l and wiki-tech-l are fairly active.

Wikimedia Foundation-employed data analyst Erik Zachte makes charts, tables and comma-separated values files available from stats.wikimedia.org of large-scale analyses of all Wikimedia projects. His blog is available from www.infodisiac.com. The web service 'Vital Signs' running from Wikimedia Labs displays interactive charts of total pageviews and other metrics over all Wikimedia projects and all languages.

The newsletter *The Signpost* (Wikipedia Signpost) discusses different matters of Wikimedia projects. More critical is the forum

Wikipedia Review and individuals have composed sites with critical commentaries, see, e.g., www.wikipedia-watch.org. Wikipedia Weekly is a podcast with episodes from 2006 to presently Since July 2011 the meta-wiki of Wikimedia has published the monthly Wikimedia Research Newsletter (WRN) available from meta.wikimedia.org/wiki/Research:Newsletter. It focuses on new Wikimedia-related research. Wikimedia Foundation-employed Dario Taraborelli often writes entries, but individual researchers make substantial contributions. The newsletter is also aggregated into a single volume for each year. The 2012edition was 95 pages long. $^{34}\,$ The Twitter and Identi.ca user WikiResearch posts links to new research on Wikipedia.

2 Examining Wikipedia

Wikipedia has been examined in a number of ways, both quantitative and qualitative, and with an analysis of the full data set and just a small part.

What is Wikipedia?

The most basic question in the study of Wikipedia asks "what is Wikipedia"? Wikipedia itself reports in its first sentence in October 2013 "Wikipedia is a collaboratively edited, multilingual, free Internet encyclopedia supported by the non-profit Wikimedia Foundation." ii In the initial version of the Wikipedia article from November 2001 Larry Sanger defined Wikipedia as "the name of an open content, WikiWiki encyclopedia found at http://www.wikipedia.com/, as well as of its supporting, very active encyclopedia-building project." iii Note that the recent quote defines Wikipedia as a work rather than as an organization, the difference between "free" and "open", and between "collaboratively edited" and the more technical term "WikiWiki". Wikidata presently (2013) claims it as an instance of a "wiki", an "internet encyclopedia" and a "Wikimedia project" and the English description reads "free online encyclopedia that anyone can edit" while the German description use the word project. iv Interestingly, a study on the German Wikipedia community found that interviewees would see a German focus on the end product, while claiming the English-speaking community focused on the process citing a German Wikipedian: "We are not here because we want to use the wiki

Whereas a definition of Wikipedia as an online encyclopedia would be from a reader's point of view, regarding Wikipedia as a collaborative writing application (CWA) would also regard it from an editor's point of view. In a review of health CWAs researchers would map CWA depending on use patterns: virtual communities (patients, e.g., *Dealing with Autism*), professional communities (e.g., *RadiologyWiki*) and Science 2.0 (e.g., *OpenWetWare*). The researchers would place Wikipedia at the very center of the map, see Figure 2.36

Researchers have presented other definitions: Benkler and Nissenbaum refered to Wikipedia as a "project" and mentioned it as an example of commons-based peer production.³⁷ Konieczny would call Wikipedia an "online organization[...]", an "online communit[y]" and "related to several social movements, including the Free and Open Source Software Movement, Open Publishing Movement, and Free Culture Movement".³⁸ Such definitions focus not so much on the end product, but rather the process, the project and the group of people that led to the result.

Yet other researchers regard Wikipedia as a work beyond the encyclopedia, e.g., as a semi-structured knowledge source which can be used as a basis to derive explicit facts, ³⁹ or a "continuously edited globally contributed working draft of history". ⁴⁰ Wikipedia might also be regarded as part of the Web 2.0 or social media.

The answer to the question "what is Wikipedia" directs the researcher in his/her study. If the researcher thinks Wikipedia is an encyclopedia then the researcher likely focuses on the content and most naturally compares Wikipedia against other reference works, — printed or online. If Wikipedia is a semi-structured knowledge source then Wikipedia should instead be compared to works such as WordNet and Semantic Web ontologies.

Quality

Several formel studies of the quality of Wikipedia have been performed. Such studies typically select a sample of Wikipedia articles and "manually" read and judge the quality, sometimes in comparison with other encyclopedias or other resources. The quality may be rated on several dimensions:

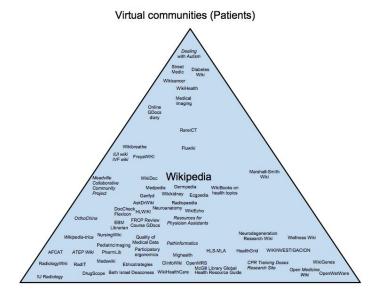
and have fun with it, but we want to have an encyclopedia which is bigger and better than any encyclopedia that has been there before [...]." 35

ii "Wikipedia" (oldid=576265305)

iii "Wikipedia" (oldid=331655534)

ivhttps://www.wikidata.org/wiki/Q52

 $^{^{\}rm v}{\rm See}$ the overviews on the English Wikipedia 'en:Wikipedia:External peer review' (oldid=169010073) and 'en:Reliability of Wikipedia' (oldid=179243736).



Professional communities

Science 2.0 (Researchers)

Figure 2: Map of collaborative writing applications with Wikipedia at the center. From Archambault et al., 2013, CC-BY.³⁶

Accuracy (no factual errors), coverage, bias, conciseness, readability, up-to-dateness, usable/suitable and whether the articles are well-illustrated and well-sourced. For Wikipedia's 'featured articles' Wikipedia has the following quality dimensions: comprehensive, accurate, verifiable, stable, well-written, uncontroversial, compliance, appropriate images, appropriate style and focus, while Stvilia et al. in their study of Wikipedia discussions worked with 10 different dimensions. 41

There have been many quality studies in the health/medicine/drug domain: In a 2013 review of collaborative writing applications in health care researchers identified 25 papers reporting on the quality of information in such systems, with 24 of them evaluating Wikipedia,³⁶ and researchers continues to study health-related information quality in Wikipedia.

Overall quality

In the perhaps most widely referenced investigation the science journalists of Nature collected 42 science articles from Wikipedia and Encyclopædia Britannica and let blinded experts evaluate them. The comparison between the two encyclopedia showed that the articles of Wikipedia contained the most factual errors, omissions and misleading statements, — but surpricingly not particularly many more than Encyclopædia Britannica:

162 against 123. Both encyclopedias contained 4 "serious errors, such as misinterpretations of of important concepts". The Nature study was not itself peer-reviewed.

In the 2006-published study Roy Rosenzweig examined several quality aspects of American history articles in English Wikipedia compared against Encarta and American National Biography Online.²⁰ He found the essays on the United States history to have inaccurate descriptions and with incomplete coverage. He attributed this to "broad synthetic writing is not easily done collaboratively" and found biographies of historical figures to "offer a more favorable terrain for Wikipedia since biography is always an area of popular historical interest". Rosenzweig then examined bibliographies of American historical figures in Encarta and the 18,000 entries American National Biography Online for comparison against Wikipedia. Of 52 examined people listed in American National Biography Online about half were listed in Wikipedia and onefifth in Encarta. He found the American National Biography Online had more details with about four times as many words as Wikipedia. He noted a bias in coverage between articles on Isaac Asimov, President Woodrow Wilson and Lyndon LaRouche judging the American National Biography Online to give a more proportionate coverage. He went on to examine factual errors. In 25 Wikipedia articles he found clear cut errors in four, and 3 articles with

Topic	Comparisons	Articles	Evaluation
Science ⁴²	Britannica	42	Blinded experts
Biographies of American historical figures 20	Encarta, American National Biography Online	52/25	The author
Pop culture, current affairs, science 43	_	3 broad topics	3 librarians
Surgical procedures ^{44,45}	_	30	Experts
$General^{46,47}$	Brockhaus	50	Research institute
Drug information ⁴⁸	Medscape Drug Reference	80 questions	The authors
$\begin{array}{ll} {\rm Medical} & {\rm students} & {\rm information}^{49,50} \\ \end{array}$	AccessMedicine, eMedicine, Up-ToDate	3	Blinded experts
Cancer information ^{51,52}	US National Cancer Institute's Physician Data Query (PDQ)	10	Medically trained personnel
Osteosarcoma ⁵³	NCI patient and professional site	20 questions	3 independent observers
Mental disorders ⁵⁴	13 websites	10 topics	3 experts
Medication information ⁵⁵	Manufacturer's package insert	20 drugs	Four drug information residency- trained pharmacists
Orthognathic surgery ⁵⁶	24 other websites	The topic	Scoring against "DISCERN"
Nephrology ⁵⁷		95 ICD-10 codes	Counting references, readability index computation
$\mathrm{General}^{58}$	_	134	96 experts
Medical conditions ^{59–63}	Peer-review literature	10	Pairs of medicine residents or rotat- ing interns
Prescription drugs ⁶⁴	_	22	The authors
Drugs^{65}	Text books	100	The authors(?)
Health, medicine, nutrition ⁶⁶	WebMD, Mayo Clinic	92 statements	Raters/authors
Breast reconstruction ⁶⁷	9 web sites	1	Readability in- dex computation, authors rating

Table 4: Selection of Wikipedia quality studies. See also the Multimedia Appendix 3 of the Archambault 2013 review. 36

factual errors among 10 examined for Encarta and 1 article with errors in American National Biography Online. Rosenzweig found Wikipedia "more anecdotal and colorful than professional history" and focus on topics with recent public controversy, and concluded "Wikipedia, then, beats Encarta but not American National Biography Online in coverage and roughly matches Encara in accuracy", and further noted American National Biography Online has richer contextualization and easily outdistances Wikipedia on "persuasive analysis and interpretations, and clear and engaging prose".

Another early quality of information study, a peer-reviewed one from 2006, looked on the credibility of Wikipedia and its articles.⁶⁸ However, this

study was not a comparison and not blinded.

In December 2007 Stern performed an examination of the German Wikipedia and the on-line edition of the German encyclopedia Brockhaus. 46,47 This weekly magasin had asked an independent research institute, Wissenschaflicher Informationsdienst Köln, to evaluate 50 articles in relation to criteria of correctness, completeness, up-to-dateness and comprehensibility. In 43 cases the Wikipedia article was evaluated as better than Brockhaus', and Wikipedia got the best grade average. 69,70 Wikipedia was regarded as better in up-to-dateness and—perhaps surpricingly—in correctness, while Brockhaus scored better in completeness. Furthermore some Wikipedia articles were regarded as too

complicated for the lav reader.

In the summer of 2009 Danish newspaper Berlingske Tidende made a small informal comparison between the Danish Wikipedia and the larger (in terms of number of articles) expert-written *Den Store Danske* online encyclopedia. Overall the Danish Wikipedia came slightly ahead due to its many links, typically longer articles and more frequent updates, even considering the background of the authority of *Den Store Danske*.⁷¹ The author also noted the quicker and more precise searching in Wikipedia.

Among the many health-related quality of information studies $^{44,\,45,\,48-57,\,59,\,60,\,64-67,\,72}$ is a study from 2007 where medical doctors reported on surgical information in Wikipedia. 44 Identifying 39 common surgical procedures the researchers could find 35 corresponding Wikipedia articles with all of them judged to be without overt errors. The researchers could recommend 30 of the articles for patients (22 without reservations), but also found that 13 articles omitted risks associated with the surgical procedure. 45

Other researchers examined Wikipedia August 2009 cancer information and US National Cancer Institute's Physician Data Query (PDQ).^{51,52} They found that Wikipedia had similar accuracy and depth when compared against the professionally-edited PDQ, but they also found that Wikipedia had lower readability as measured with the Flesch–Kincaid readability test.

Considering scope, completeness, and accuracy of information for osteosarcoma on April 2009 English Wikipedia compared against patient and professional sites of US National Cancer Institute (NCI) 3 independent observers scored the answers to 20 questions on a 3 point scale. Wikipedia scored lower compared to the two NCI versions, though the statistical test only showed significant difference against the NCI professional version. ⁵³

Three psychologists with relevant expertice examined 10 topics in mental disorders across 14 websites with respect to accuracy, up-to-dateness, coverage, referencing and readability. Among the websites, beside Wikipedia, were NIMH, WebMD and Mayo Clinic and among the topics examined were "childhood onset of psychosis" and "gambling and depression". Wikipedia scored high ("generally rated higher") on accuracy, up-to-dateness and referencing, while low on readability.

While numerous studies have examined the quality of Wikipedia, one finds far fewer studies of the qualities of other wikis, e.g., the Wikipedia sister projects. In 2008 a lexicography study would claim that Wiktionary had a poor quality, while at the same time noting the comparative studies favor-

able to Wikipedia.⁷³ One of the few studies to compare Wiktionary with other language resourses examined the German Wiktionary with GermaNet and OpenThesaurus and found that the scope of the three resourses varied depending on which variable they look at, e.g., Wiktionary had the hightest number of word senses but the lowest number of synonyms.⁷⁴

Factual errors

Trusting specific facts on Wikipedia is questionable, as there might be typos, intentional or unintentional errors, biased presentation, or hoaxes, e.g., a misinformation on Wikipedia propagated to the orbitury for composer Maurice Jarre on *The Guardian* web-site. As a research experiment a student had entered made-up Jarre quotes in Wikipedia immediately after Jarre's death. An obituary writer working under a tight deadline picked up this information though it stayed in Wikipedia for only 25 hours. The hoax was only revealed after the student contacted the publishers. Similar vandalism that spreads to obituaries happened for Norman Wisdom.

Cautionary notes have been cast for the open wiki-model in cases where potentially hazardous procedures are described.⁷⁸ Especially chemical and medical procedures and compounds may call for complete and accurate description. For medical drug information Kevin Clauson and his co-authors compared Wikipedia and Medscape Drug Reference (MDR), a free online "traditionally edited" database. 48 They found that Wikipedia could answer fewer drug information questions, e.g., about dosage, contraindications and administration. In the evaluated sample Wikipedia had no factual errors but a higher rate of omissions compared to MDR. The authors could also find a marked improvement in the entries of Wikipedia over a just 90 days period. The study went on to mainstream media with headlines such as "Wikipedia often omits important drug information" and even "Why Wikipedia Is Wrong When It Comes To Prescription Medicine". As noted by Wikipedians on a discussion page the study did not mention the fact, that one of the Wikipedia manual of styles explicitly encourages Wikipedia authors not to include dosage information with the present wording of "Do not include dose and titration information except when they are notable or necessary for the discussion in the article."vi Thus in one of the 8 examined question categories the omissions on the

 $^{^{\}mathrm{vi}}\mathrm{http://en.wikipedia.org/wiki/Wikipedia:MEDMOS,}\ 252051701$

part of Wikipedia comes as an intention by consensus.

On a small comparison study on medical information with just 3 topics blinded experts found some factual errors in Wikipedia, — around the level of medical online resources UpToDate and eMedicine. AccessMedicine were found to have no factual errors among its 3 articles examined. 49,50

Coverage

One kind of critique often carried forth is that Wikipedia tends to have a emphasis on topics in pop culture, — the critique following the template "there are more entries for [a pop culture phenomenon] than for Shakespeare." vii Is there a bias in the topical coverage of Wikipedia? Are there any other bias in coverage, e.g., with respect to gender and nationality?

Studies on topical coverage in Wikipedia often examine the number of Wikipedia articles within a given subject area and compare that number to associated numbers in works or databases from governments, well-established companies or other organizations, which then acts as a reference, ^{79–84,86} see Table 5. In 2005 Altmann could write that "[m]edical Informatics is not represented sufficiently since a number of important topics is missing". He had compared the English Wikipedia to 57 terms he found in "Handbook of Medical Informatics". ⁷⁹

Looking at outbound scientific citations in the English 2007 Wikipedia I found astronomy and astrophysics articles rather much cited compared to Journal Citation Reports from Thomson Scientific, but generally an overall agreement. So Journal of Biological Chemistry got undercited but that changed after automated mass-insertion of genetic information. One peculiarity with the sample occured for Australia botany journals. A Wikipedia project had produced a number of well-sourced articles on Banksia some reaching featured article status. The citation from these Wikipedia articles would skew the statistics.

By sampling 3000 articles from the English 2006 Wikipedia and categorizing them against the Library of Congress categories Halavais and Lackaff found categories such as social sciences, philosophy, medicine and law underrepresented in Wikipedia compared to statistics from *Books in Print*.⁸² The two latter categories had, however, on average a comparably large article size. They identified science, music, naval and, e.g., geography as overrepresented, with music probably benefitting from

fans contributions and other categories from the mass-insertion of material from public data sources such as United States Census. The two investigators could also find missing articles in the 2006 Wikipedia, when compared to three specialized encyclopedias in linguistics, poetry and physics. Halavais and Lackaff also noted some peculiarities in Wikipedia, e.g., extensive list of arms in the military category, comics fans to some extent driving the creation of articles in the fine art category and voluminous commentary on the Harry Potter series in the literature category.

For twentieth century philosophers Elvebakk compared Wikipedia against two online peer-review resources, The Stanford Encyclopaedia of Philosophy and the Internet Encyclopedia of Philosophy, with respect to coverage of gender, nationality and discipline. She concluded that Wikipedia did not represent "the field of philosophy in a way that is fundamentally different from more traditional resources" in 2008. 83 Wikipedia had far more articles about the philosophers than the two other resources and only some minor differences in fractions, such as a smaller fraction of German and French philosophers.

In a study on the efficiency of Web resources for identifying medical information for clinical questions Wikipedia failed to give an answer in a little above a third of the cases, while Web search engines, especially Google, were much more efficient. However, Wikipedia was more efficient than medical sites such as UpToDate and eMedicine in terms of failed searches and number of links visited, and the 'end site' that most often provided the ultimate answer from a Google search was Wikipedia.⁹¹ In another 2008 medical coverage study researchers found over 80% of ICD-9 and ICD-10 diagnostic codes in gastroenterology covered by Wikipedia.⁸⁴ A similar study for nephrology found 70.5% of ICD-10 codes represented in August 2012.⁵⁷ In another life science coverage study, researchers constructed a semi-automated program for matching LOINC database parts with Wikipedia articles. Of the 1705 parts they examined in October 2007 they found 1299 complete matches in Wikipedia with their semi-automated method, 15 partial matches and a further 15 matches from manual search, i.e., 1329 corresponding to 78%.86 They concluded that "Wikipedia contains a surprisingly large amount of scientific and medical data"

A 2008 study compared the number of words in sets of Wikipedia articles with the year associated with the articles and found that articles associated with recent years tended to be longer, i.e., recency was to a certain extent a predictor for coverage: The length of year articles between 1900 and 2008

 $^{^{\}rm vii}$ "Why are there more Wikipedia entries for Doctor Who than there are for Shakespeare?" 90

Topic	Comparisons	Result
Medical informatics ⁷⁹	Handbook of Medical Informatics	"Medical Informatics topics are not very well represented in the Wikipedia currently [2004/2005]"
Scientific citations ^{80,81}	Thomson Scientific Journal Citation Reports	Generally good correlation with scientific paper citations. Astronomy and banksia somewhat overcited. Dependence on bots.
General topics, physics, linguistics, poetry ⁸²	Library of Congress categories, Encyclopedia of Linguistics, New Princeton Encyclopedia of Po- etry and Poetics, Encyclopedia of Physics	82% (physics), $79%$ (linguistics) and $63%$ (poetry) coverage
Twentieth century philosophers 83	The Stanford Encyclopaedia of Philosophy and the Internet Encyclopedia of Philosophy	Wikipedia had 534 philosophers covered while the other two had 60 and 49, respectively
Gastroenterology ⁸⁴	ICD-9 and ICD-10 codes	83% coverage
Philosophers ⁸⁵	Facts extracted from A History of Western Philosophy, A History of Western Philosophy, The Oxford Companion to Philosophy and The Columbia History of Western Phi- losophy	52% coverage
Medical terminology ⁸⁶	LOINC database	78% coverage
General topics	_	30% obtained 'good' or 'excellent' marks
US gubernatorial candidates and elections ⁸⁷	Number of real world candidates and elections	93% candidate coverage, $11100%$ election coverage
Women ⁸⁸	National Women's History Project	23/174 or $77/268$ missing
$Nephrology^{57}$	ICD-10 codes	70.5% coverage
Scientists ⁸⁹	Thompson Reuter list	22%– $48%$ coverage
Drugs ⁶⁵	Pharmacology text books	83.8% (German), 93.1% (English)

Table 5: Selection of Wikipedia coverage studies.

and the year as a predictor variable had a Spearman correlation coefficient on 0.79. The results were not homogeneous as the length associated with articles for Time's person of the year had a correlation of zero with the year. Academy award winning films and "artist with #1 song" had correlation between the two: 0.47 and 0.30, respectively. The authors of the study also examined other sets of articles in Wikipedia and the correlation with column inches in Micropaedia of the Encyclopædia Britannica, country population and company revenue. The correlations were 0.26, 0.55 and 0.49, respectively. In their comparison with 100 articles from Micropædia they found that 14 of them had no Wikipedia entry, e.g., "Russian Association of Proletariat", "League for the Independence of Vietnam" and "urethane". 92

Bill Wedemeyer presented the quality of scientific articles on the English Wikipedia on Wikimania 2008 as he and his students had examined the coverage based on several data sets. On a cross-

section of 446 articles randomly and blindly sampled from *Encyclopædia Britannica* Wikipedia articles lacked entries for 15, e.g., "Bushmann's carnival", "Samarkand rug" and "Catherine East". All of 192 random geographical articles from Britannica had corresponding articles in Wikipedia. Of 800 core scientific topics selected from biochemistry and cell biology text books 799 could be found in Wikipedia. He concluded that science is better covered than general topics and that Wikipedia covers nearly all encyclopedic topics. 93

Kittur, Chi and Suh developed an algorithm that would assign a topic distribution over the top-level categories to each Wikipedia article. After evaluating the algorithm on a human labeled data set they examined the English Wikipedia and found that 'Culture and the arts' and 'People and self' as the most represented categories. Between the 2006 and 2008 data set they found that 'Natural and physical sciences' and 'Culture and the arts' categories grew the most. By combining the algorithm

with a method for determining degree of conflict of each article⁹⁵ they could determine that 'Religion' and 'Philosophy' stood out as the most contentious topics.

A case of bias in coverage with an individual Wikipedia article reached mainstream media. A user flagged the article on Kate Middleton's wedding grown for deletion. The flagging and the ensuing debate about the notability of the dress was seen a symptom of the 'gender gap' of Wikipedia. Jimmy Wales argued with a "strong keep" that "I believe that our systemic bias caused by being a predominantly male geek community is worth some reflection in this context" and pointed out that Wikipedia in contrast has "over 100 articles on different Linux distributions, some of them quite obscure" and with "virtually no impact on the broader culture, but we think that's perfectly fine." 96 Parallel to the media focus on the gender imbalance among contributors, 97 a couple of studies have examined on the possible biased representation of women on Wikipedia. $^{88,\,98-101}$

Reagle and Rhue have reported on the female proportion in biographic resources for persons born after 1909: 28.7% (Gale Biographical Resource Center) and 24.5% (Wilson's Current Biography Illustrated), but also as low as 15% (American National Biography Online). Other lists of notable persons yield percentage on 10% (The Atlantic top 100 most influential figures in American history) and 12% (Chambers Biographical Dictionary). For the English Wikipedia the researchers found 16% and after a similar analysis of Encyclopædia Britannica the researchers concluded "Wikipedia and Britannica roughly follow the biases of existing works".88 In 2011 Gregory Kohs would report a higher number on 19% for the female proportion, — this was for a random sample of 500 living people biographed on Wikipedia. ⁹⁸ Reagle and Rhue also compared biographic article lengths in Wikipedia and Encyclopædia Britannica with respect to gender and found no consistent bias in either female or male direction.⁸⁸ Wikidata, where a property of an entity may indicate the gender, can scale up the analysis and make it multilingual to several hundred thousand persons, even over a million: Using Wikidata's 'sex'/'gender' property and its language links Max Klein compared the sex ratio across Wikipedia language versions finding (among the big Wikipedias) the most equal rate on the Chinese Wikipedia yet still well below 25% female, while the English Wikipedia had a female percentage on 17.55%. A reference file, the Virtual International Authority File (VIAF) data, gave a 24.35% female rate in Klein's study. 99 In 2015 Max Klein published a blog post with a more detailed

reporting of the results of his study performed together with Piotr Konieczny. Among variables in relation to gender they considered celebrity status finding that "recorded females [in Wikipedia] are more likely to be celebrities". 100 Also in 2015 Magnus Manske would publish a blog post with his Wikidata analysis comparing gender representation grouped by centuries, region and country and when he compared Wikidata against VIAF and Oxford dictionary of National Biography he found Wikidata had a more equal representation of males and females. 102 Later that year he used Wikidata to compare gender representation across Wikipedia language-versions with respect to 'Getty Union List of Artist' and the 'Rijksbureau voor Kunsthistorische Documentatie' database now finding a clear male bias for almost all Wikipedias for both lists of artists. 103 Manske was prompted by Jane Darnell who has made several analyses of the gender gap. viii

The studies find an increasing ratio of female representation over time (e.g., date of birth). Why does this increase appear? Is it because women have increased their position in society? Han-Teng Liao put forward an alternative 'male gaze' hypothesis, where the increase comes through a gender-interest bias, e.g., young males interested in female celebrities such as porn actresses. ¹⁰⁴

Tools associated with Wikidata can make coverage estimation across Wikipedias trivial. The Mix'n'match lists entries from several external databases and display matches with Wikidata items, e.g., it can list the members of the European Parliament based on information from its homepage http://www.europarl.europa.eu/meps/ together with the Wikidata item. Statistics can then show that (in the case with this database) all members are matched to Wikidata items, but, e.g., only 293 members, corresponding to around 8\%, have a Danish Wikipedia article, while the English Wikipedia has a coverage of over 50% with 2020 articles for parliament members (as of October 2014). The European Parliament list is only one of several. Examples of other catalogues are Oxford Dictionary of National Biography, BBC Your Paintings and Hamburgische Biografie.

Wikipedians with a particular interest in coverage organize themselves in the *WikiProject Missing encyclopedic articles*, a project where the main goal "is to ensure that Wikipedia has a corresponding article for every article in every other general purpose encyclopedia available". The project lists quite a number of reference works and other resources for benchmarking Wikipedia coverage. Ex-

 $^{{}^{\}rm viii} {\rm https://commons.wikimedia.org/wiki/Category:Jane_Darnell.}$

amples are *Encyclopædia Britannica* 1911, 2007 *Macropædia* and The Encyclopedia of Robberies, Heists and Capers.

Editor and researcher of Wikipedia, Emilio J. Rodríguez-Posada, has attempted to estimate the number of "notable articles needed to cover all human knowledge". ix As of January 2014 the number stands on 96 million. It is made up of, e.g., the number of species, in one source estimated to be 8.7 million (eukaryotes). The around 14 million entities on Wikidata (as of January 2014) makes up around 15% of 96 million, yet the 96 million does not include, e.g., the majority of the large number of chemical compounds described. Chemical Abstracts Service announced in 2011 that they had reached the 60 millionth entry in their registry. 106

Limitations in coverage due to Wikipedia policy of notability has inspired other web-sites: *Deletion-pedia* records deleted pages from Wikipedia in a MediaWiki run site with no editing possible, and *Obscuropedia* is a wiki for non-notable topics not covered by Wikipedia.

Up-to-dateness

Several quality comparison studies examine the upto-dateness and find that Wikipedia compares well in this aspect, ⁴⁸, ⁵⁴, ⁶⁹, ⁷⁰, ⁹³ although not equivo-cal. ⁶⁴ In the comparison between Wikipedia and Medscape the researchers found four factual errors in Medscape among 80 articles examined. Two of these occured due to lack of timely updates. No factual errors occured in Wikipedia. ⁴⁸ The Wedemeyer study found that Wikipedia was much better up to date than *Encyclopædia Britannica*. ⁹³

In a study on twentieth century philosophers Wikipedia had far more articles on philosophers born after the Second World War than two other online encyclopedias The Stanford Encyclopedia of Philosophy and The Internet Encyclopedia of Philosophy. 83

The Danish Wikipedia has a large number of bibliographies copied more or less unedited from two old reference works with expired copyright: Dansk biografisk Leksikon and Salmonsens Konversationsleksikon. The age of the works affects the language and viewpoint of the Wikipedia articles. 107

Information on the death of a TV host, Tim Russert, came out in Wikipedia before news organizations published it. The author of the Wikipedia entry came inside a news organization.¹⁰⁸

Medical drugs may have associated safety alerts, e.g., United States Food and Drug Administration issues Drug Safety Communications with warnings and recommendations on medical drug use. When FDA issues these communications Wikipedians should incorporate them in the articles about the drug for timely information to patients and physicians. A study on these communications from 2011 and 2012 showed that Wikipedians do not update the information satisfactorily: For 22 prescription drug articles researcher found that Wikipedians had not incorporated specific FDA communications in 36% of the articles when they examined the articles more than a year after the FDA communications.⁶⁴

Sources and links

Many studies and tools examine the outbound references to sources that Wikipedia uses^{80,81,93,101} or the inbound links that comes from documents to Wikipedia. Often the count of sources are used as a feature in studies of article quality.^{93,101} Wedemeyer's study looked on the references of Wikipedia articles. They found that most developed articles had sufficient references comparable to a scientific review article, but some articles, even two featured, had insufficient referencing.⁹³

Ed Summers' Linkypedia Web service available from http://linkypedia.info/ makes statistics available online on which Wikipedia articles links to specific webpages on selected websites. As of 2012 statistics was mainly available for selected GLAMs. It enable, e.g., British Museum to see that their page on ancient Greece had the highest number of inlinks from Wikipedia: 39 as of February 2012; that Wikipedia articles in the category "BBC Radio 4 programmes" linked much to their website; that the "Hoxne Hoard" article had no less than 27 links to their website; and that the total number of Wikipedia links to the British Museum website was 2'673 from 1'209 pages.

Another Web service of Ed Summers, wikitweets, displays microposts from Twitter that link to Wikipedia. The Web service runs in real-time from wikitweets.herokuapp.com and tweets with excerpt of the linked Wikipedia articles are stored in machine readable JSON format at the Internet Archive (archive.org/details/wikitweets).

Readers clicks the links in Wikipedia articles to outside sources to a considerable extent. In 2014 a CrossRef statistics reported that Wikipedia was the "8th largest referrer of CrossRef DOIs". 109 Web services from CrossRef allow easy identification of which Wikipedia articles cite a scientific article based on DOI information as well as a real-time citation events updates.*

ixhttps://en.wikipedia.org/wiki/Wikipedia:ALL

^{*}See, e.g., http://det.labs.crossref.org/works and

How Wikipedia is used as a source has also been described. In the media business, e.g., *Philadelphia Inquirer* instructs journalist never to use Wikipedia "to verify facts or to augment information in a story" and one reporter has been cited for "there is no way for me to verify the information without fact-checking, in which case it isn't really saving me any time." Other news organizations allow occationally citation of Wikipedia as a source, e.g., *Los Angeles Times*. ¹¹⁰ An analysis of press mentioning (cited, quoted or referred) of early Wikipedia found, e.g., that Daily Telegraph Online accounted for roughly a third of all citations. ¹¹¹ The site consistently referred to Wikipedia for further reading and background information in sidebars.

Genre and style

Researchers have also investigated other content topics besides quality. In one study researchers examined 15 Wikipedia articles and their corresponding talk page and compared them with other online knowledge resource: Everything2 and Columbia Encyclopedia. They specifically looked on the formality of the language by counting words indicative of formality or informality, such as contraction, personal pronouns and common noun-formative suffixes. With factor analysis they found that the style Wikipedia articles is close to that of the Columbia Encyclopedia. ¹¹²

Lexical analysis was featured in a study on Wikipedia biased representation wrt. gender. The study found that, e.g.,the word 'divorce' appear with a higher rate in articles about women compared to articles about men. 101

The genre may also evolve as editors extends and change the articles. 113

Accessibility

Lopes and Carriço examined 100 Wikipedia and 265 non-Wikipedia Web articles cited by Wikipedia. They looked for their level of accessibility, i.e., to which extent the fulfilled the Web Content Accessibility Guidelines of the World Wide Web Consortium designed 'to make Web content accessible to people with disabilities'. The authors found that Wikipedia articles on average scored better than the Web articles they cited. They further argued that the discrepancy between the accessibilities could lower the credibility of Wikipedia. It is not so odd that Wikipedia scores well in accessebility since HTML mark is automatically contructed from wiki-markup, and the

software can be programed to ensure that, e.g., the 'alt' field of the 'img' HTML tag is automatically set.

Use of Wikipedia in court

The supreme court of India used Wikipedia for the definition of the word 'laptop', 116 and several American courts have used Wikipedia in their rulings, 117 e.g., Connecticut Supreme Court cited Wikipedia for the number of gay Congressmen. 118 These cases are not singular: In February 2007 Washington Post noted that courts cited Wikipedia four times as often as Encyclopædia Britannica, 119 and in 2008 Murley ran a database search and found "1516 articles and 223 opinions that had cited to Wikipedia articles" in the Westlaw's and ALLCASES databases. 120 The English Wikipedia maintains incomplete lists of mostly English language court cases using Wikipedia as a source: Wikipedia: Wikipedia as a court source and Wikipedia: Wikipedia in judicial opinions. A couple of other language versions of Wikipedia have similar lists for cases in their respective languages.

Three lengthy papers examine the judiciary use of Wikipedia and discuss the controversy of using Wikipedia as an authority. ^{121–123} They find the first references to appear in 2004, peaking in 2007 and a decrease towards the end of their examined periods, — November 2007 for Breinholt ¹²¹ and 2008 for Stoddard. ¹²² Breinholt classifies the different uses of Wikipedia into four categories:

- 1. Wikipedia as a dictionary. Wikipedia used, e.g., to answer what "candystriper" means.
- Wikipedia as a source of evidence. In the most perilous use of Wikipedia judges rely on Wikipedia for evidence with one example being a judge relying on Wikipedia for whether Interstate-20 US highway does or does not extend from California.
- Wikipedia as a rhetorical tool. Harmless use of Wikipedia, e.g., for literary allusions.
- 4. Judiciary commentary about Wikipedia. One among the few cases involved a judge cautioning against citing Wikipedia in an appellant brief.

In some cases Wikipedia may be the only reference available for definitions of words, e.g., at one point Google returned only Wikipedia for a 'define' query on "candystriper". Should one entirely ignore Wikipedia? I my opinion Wikipedia articles can be used for definitions provided that the definition of the provided that the provided that the definition of the provided that the provided

http://events.labs.crossref.org/events/types/WikipediaCitationinition has been overlooked by many readers and

editors and that many reliable editors have over time edited the article, — so we may regard it as a consensus definition. For establishing that 'many reliable editors have edited' one would need to examine the revision history and possibly the discussion page and its associated revision page. This process might require an expert Wikipedian.

Cases in Wikipedia and Wikimedia Commons may give rise to legal discussions. The copyright status around the so-called Monkey selfie has probably been the the most widely discussed. The essay 'Final exam for wikilawyers' by the Wikipedian Newyorkbrad sets up number of interesting questions from fictional and real-world cases.

Size across languages

Why does the language editions of Wikipedia differ in size? If it is often pointed out that the number of speakers of a language is a good indicator for the size of Wikipedia in that language, 124 then why is the Norwegian Wikipedia larger than the Danish? And why was the Esperanto larger than the Arabic until 2011?xi

Morten Rask analyzed 11 Wikipedia language editions with respect to creation date of Wikipedia, number of speakers of the language, Human Development Index, Internet users, Wikipedia contributors and edits per article and found a number of correlations between these variables, e.g., the Internet penetration and level of human developement was correlated to the number of contributors: 125 Wikipedia contributors are rich and e-dy. Explaining the quality of the German Wikipedia Sue Gardner put forth related factors, saying: "Germany is a wealthy country. People are well educated. People have good broadband access. So the conditions for editing Wikipedia are there." 126 Other variables that may affect the Wikipedia size of different language edition are culture of volunteering, willingness to translate (from other language Wikipedia) and problems with non-latin characters. Among the reasons for the relatively small size of the Korean and Chinese Wikipedias Shim and Yang suggested the competition faced by Wikipedia from other knowledge-sharing Web services: Korean question/answering site Jisik iN and Chinese online encyclopedia Baidu Baike. 127 ther factor Andrew Lih and Sue Gardner would also mention the ability to meet face-to-face due to German-speakers geographically location in a relatively small area and the German Verein cul $ture.^{25, 126}$

The Arabic Wikipedia has had a relatively small size compared to the number of speakers. The low attendance for a Wikipedia event in Egypt was blamed on 'general lack of awareness of the importance of the issue' and 'culture of volunteer work'. ¹²⁸ Arabic users may choose to write in English because they find it easier to communicate in that language due to keyboard compatibility problems and to bring their words to a wider audience. ¹²⁹ Users of the Internet may also be hindered by low cabel capacity in some areas as has been the case in East Africa. ¹³⁰

One obvious factor for the size of a Wikipedia comes from the willingness of the community to let bots automatically create articles. Letting bots create articles on each species may generate many hundreds of thousands articles.¹³¹ The unwillingness to let bots roam and the elimination of stub articles and a focus on quality compared to quantity in the German Wikipedia³⁵ may explain the why the its article-to-speaker ratio is (as of February 2014) quite lower thatn the Dutch and Swedish.

Network analysis, matrix factorizations and other operations

Modern network analysis has come up with a number of new notions, e.g., small world networks, ¹³² the power law of scale-free networks, ¹³³ PageRank^{134, 135} and hubs and authority. ¹³⁶ Formula with algorithms have been put forward that quantatively characterize the concepts and they have been applied to a diverse set of networks, e.g., the network of movie actors, power grid, neural network and the world wide web. Wikipedia researchers have also examined the quantative characteristics for the networks inherent in Wikipedia. Among the many networks characteristics reported for a variety of Wikipedia derived networks are PageRank and Kleinberg's HITS or other eigenvalue-base measures, ^{137, 138} small world coefficient, with cluster-coefficient and average shortest path, 138-141 the size of the 'bow tie' or giant components, 138, 139, 141, 142 power law coefficients, $^{1,138,139,141-144}$ h-index, 141 reciprocity, 139,141 assortativity coefficients, 138,139,141 triade significance profil. 139 and acceleration. 145

Networks can be represented in matrices, thus matrices can also be constructed from content and metadata in Wikipedia articles. Mathematical operations can be performed on the matrices to examine aspects of Wikipedia or to test computational algorithms on large-scale data. Wray Buntine built a matrix from the **within-wiki links** be-

 $^{^{\}mathrm{xi}}$ compare http://stats.wikimedia.org/EN/TablesWikipedia Weem 500'000 pages of the English 2005 Wikipedia with http://stats.wikimedia.org/EN/TablesWikipedia AR.htm and used a discrete version of the hubs and au-

thority algorithm to find topics in Wikipedia, 146 e.g., one topic would display the Wikipedia articles "Scientific classification" and "Animal" as the top authorities and "Arterial hypertension" and "List of biology topics" as the top hubs. xii other approach builds a normalized matrix from the adjacency matrix of the within-wiki links between the articles. By augmenting the normalized adjacency matrix with an extra term the so-called Google matrix can be formed. The first eigenvector associated with the Google matrix determines the PageRank of an article. 135 The adjacency matrix may be transposed, normalized and augmented. Its first eigenvector may be found to yield what has been called the *CheiRank*, ¹⁴⁷ — a concept closely linked to Kleinberg's idea of hubs in his HITS algorithm¹³⁶ and the analysis of Buntine. PageRank and "CheiRank" may be combined to form what has been called the 2DRank. 136 In the family of these analysis Bellomi and Bonato wound in an analysis from 2005 find, e.g., United States, France, Russia, World War II, Jesus and George W. Bush on the top of the HITS authority and PageRank lists. 137 Longitudinal analysis of these measures have also been performed. 148

Barcelona Media researchers performed wikilink network analyses of only the biographic articles of Wikipedia, — but across 15 different language versions. They found that PageRank for the English Wikipedia put America presidents highest on the ranked list. Comparing betweenness centrality across language versions showed overlaps among the top five most central persons with American Presidents often in the top five, as well as singers and World War II figures. However, cross-cultural differences also existed, e.g., with Chinese leaders central in the Chinese Wikipedia, Latin American revolutionaries central in the Spanish and Catalan Wikipedias, and Pope John Paul II central in the Polish and Italian Wikipedia. A gender gap existed as only three women appeared among the top five central persons across the 15 languages: Elizabeth II, Marilyn Monroe and Margaret Thatcher. 149

Yet another wikilink analysis restricted the analysis to Western philosophers—only analyzing 330 pages and 3706 links. They used the *Prefuse* toolkit for network visualization and also employed graph simplification under the name "Strongest Link Paths". With the influence network constructed from "influenced by" and "influenced" fields of the infoboxes on the philosophers Wikipedia pages they showed that Kant had influenced the largest number of other philosophers listed in Wikipedia. ¹⁵⁰

Instead of working from the links, the words of

a Wikipedia articles may also be used as features in the construction of a matrix, so the resulting matrix is a **document-term** matrix. A decomposition of such a matrix is often termed latent semantic analysis, particularly if singular value decomposition is the decomposition method. For assessing the performance of newly developed algorithms Řehůřek constructed a document-term matrix from the entire English Wikipedia with the resulting size of 100,000 times 3,199,665 corresponding to a truncated vocabulary on 100,000 words and almost 3.2 million Wikipedia articles. ¹⁵¹ Analyses of document-term matrices can form the bases for content-based filtering methods and recommender systems, e.g., a system applied Latent Dirichlet Allocation on the 4'635 sized 2007 Wikipedia Selection for Schools^{xiii} dataset, potentially providing an alternative method for school children to navigate this corpus. 152

The **category network** forms another dataset. Chris Harrison has made large "clusterball" visualizations of the three levels of Wikipedia category pages. Another group of his Wikipedia visualizations, WikiViz, display millions of vertices.

Several researchers have considered the graph/matrix formed when editors co-edits articles, i.e., the articles-by-editors matrix or co-authorship **network**. 138, 153–156 In an 2006 analysis of the articles-by-editors matrix constructed from of the Danish Wikipedia using non-negative matrix factorization on the (12774×3149) -sized matrix I found co-edited pattern on: Danish municipalities, Nordic mythology, Danish governments, article discussions, churches, science computers, pop culture, years, Scania, sport, countries and cities, plants and animals. As the edits are distributed unequal with respect to number of edits, both over articles and users, the results from the matrix factorization depended on the normalization of the matrix. With an unsuitable normalization identified co-edit cluster may simply reflect the contribution of single prolific author. With suitable normalization the factorization shows that articles and editors may be characterized by their co-editing patterns. ¹⁵⁴ Another work on the article-by-editor network used an algorithm to identify maximal bicliques, showing that some of the dense clusters of articles and editors were related to controversial topics (Scientology, creationism) or projects (WikiProjectElements for elements in the periodic table). 155 Laniado et al. reported a large number of network metrics for the co-authorship network over time and topics after the main contributors for each article had been identified with a method by Adler et al., ¹³⁸

 $^{^{\}rm xii}{\rm Model}$ parameters for topic "ID# 3"

 $^{^{\}mathrm{xiii}}\mathrm{See}$ http://schools-wikipedia.org

e.g., even after the thinning of the co-authorship network they found that over 96% of the contributors are connected in the giant component and with removal of 5000 contributors with the highest betweenness centrality the authors 85.9% author were still connected in the component. The analysis of the co-editing pattern has also been combined with geographic resolving of editors using natural language processing of user pages. ¹⁵⁶

A sequential collaboration network is a network where users are represented as nodes and links between one editor and another are present if the second editor edits an article right after the first editor edited the article. 157 The network is a directed graph and may be said to be a social network, as the adjecency matrix is a user-by-user matrix, albeit the editing may only be a form form of indirect collaboration. Some researchers call it the "article trajectory". 158 Several different patterns might occure when the sequential collaboration network is visualized for different articles. Iba et al. name "snake" "wheel" and "star" as "egobooster" networks. 157 An example of a sequential collaboration network is displayed in Figure 3 for an article of a Danish Politician on the Danish Wikipedia. Other applications of sequential collaboration networks have appeared for an online service on multiple business-related articles that adds color-coding based on sentiment analysis. 159, 160

When networks are formed from the **reply network** of article discussion/talk pages or the activity on the user talk pages (**user talk network** and **wall network**) the networks better represent the 'real' social interaction among users in contrast to indirect stigmergic collaboration seen in the network formed from articles co-authorship. Laniado et al. construct such networks and identify the presence of *welcomers*, — the users and bots writing a welcome message on new users' talk pages, as well as reports on the user associativity. They also show that long chains of replies between users on article talk pages appear for articles such as "Intellligence design", "Gaza War" and "Brack Obama". 141

Article feedback

In September 2010 Wikimedia launched an experimental article feedback tool letting users evaluate articles as well-sourced, neutral, complete or readable on a five-point scale. Results of an analysis presented by Howie Fung of 1,470 ratings across 289 articles showed that registered users tended to give lower ratings than unregistered users. The maximum score came as the most frequent given score among unregistered users on all four dimensions. Registered users chose the minimum score

as the most frequent on the dimensions of well-sourcedness and completeness, while they rated neutrality and readability higher.

Hoax content

Bloggers have made experiments with Wikipedia adding factual errors to observed how fast they get corrected — if at all. 161, 162 How long the hoax survive may depend on how obvious it is. P.D. Magnus made vandalisms anonymously and from various IP addresses across separate philosophy articles, e.g., the edit 'Kant's poetry was much admired, and handwritten manuscripts circulated among his friends and associates' in the 'Immanuel Kant' article. After 48 hours 18 of 36 of this kind of vandalism remained. 163, 164 Danish computer science student Jens Roland had his short article about the fictitious municipality Æblerød surviving 20 months on the Danish Wikipedia and translated into several other language versions. 165 The scam was supported by external web-sites with information that looked genuine. For the cases of the bibliographies on John Seigenthaler¹⁶⁶ and Bertrand Meyer¹⁶⁷ that drew much media attention the vandalism remained for 132 and four days, respectively. In both these cases the false statements made could possible have let to criminal charges against the author. The English Wikipedia maintains a list of Wikipedia hoaxes (the page Wikipedia:List of hoaxes on Wikipedia) showing that complete hoax articles may persist for many years.

Perhaps the most elaborate single incident concerned the article for the fictitious "Bicholim conflict", which persisted for well over 5 years. It was even rewarded with "Good Article" status, casting serious doubts on the article reviewing process on the English Wikipedia. This article was supported by citations to fictitious books. These kinds of hoaxes resemble the kind you also find in ordinary reference works, e.g., "Dag Henrik Esrum-Hellerup" — a fictious entry in a music reference work.

T. Mills Kelly teaches the course *Lying About the Past* at George Mason University where students fabricate histories and add them to Wikipedia. In 2008 the tale about Edward Owens survived until Kelly announced the hoax at the end of the semester, while a 2012 experiment on the fabricated history of serial killer Joe Scafe survived 26 minutes after an announcement on Reddit. ¹⁶⁹

Most of the hoaxes seem to arise from humor or critical examination. When prolific and barnstarawarded editors insert false information it may require a major cleanup, such as in the case with user Legolas2186. On the English Wikipedia he

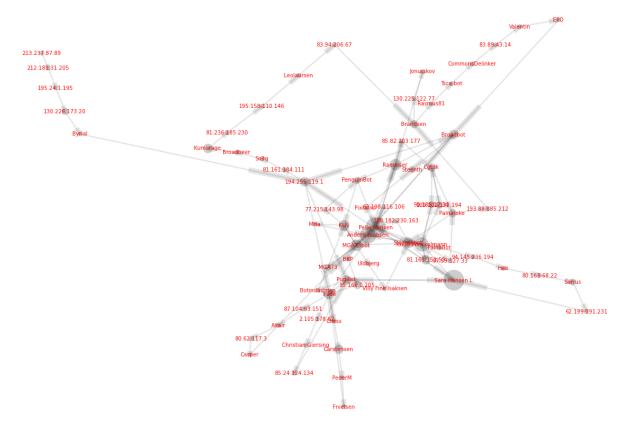


Figure 3: Sequential collaboration network¹⁵⁷ for Danish politician (Ellen Trane Nørby) on the Danish Wikipedia generated with my revvis online tool on the Toolserver. The size of the nodes are determined by the number of edits of the user. The article begins with the 213 IP address creating the article (upper left corner of plot) followed by an edit by the 212 IP address. At one point a Danish parliament IP (194.255.119.1) edits several times.

was blocked indefinitely after inserting false statements that was supported by citations to invented articles. ¹⁷⁰ His case seems not to originate in humor or critique. Should such a case be explained as a kind of addiction to barnstars and edit counts?

Vandalism

For obvious vandalism, where large part of an article is deleted, a 2004 study showed that it typically only took a couple of minutes before an article gets reconstructed. With data up to 2006 Reid Priedhorsky and his co-authors found that 5% "damaged" edits among 57.6 million revisions 42% of article damages were reverted within one estimated page view. Scientists have observed a lower rate of vandalism on specialized scientific topics: The edits on Wikipedia articles related to the WikiProject RNA and the Rfam database reach a vandalism rate of slightly over 1% (actually "possible vandalism"). 173

In his ethnographic study Lorenzen observed the Wikipedia: Vandalism in Progress page in two time

periods in October and November 2005 to see how vandals were dealt with.¹⁷⁴ During the time periods he noticed hundreds of reports of vandalism, while not all were reported vandalism were in fact vandalism. He reported 16 false reports and 39 user bans in the time period. He also discussed the issue of subtle vandalism escaping detection.

Among the temporal aspects examined in a 2006 study is the number of reverts through time from 2002 to the beginning of 2006. The researchers found that the number of reverts, including fast reverts have risen almost monotonically from below 1% to over 6%, and "that may signal an increasing amount of vandalism per page". An exception on the monotonicity was at the introduction of the 3-revert rule established in November 2004 which almost halved the so-called double-reverts. 175

The communal 'recent change patrol' watch over the recent changes in Wikipedia on an entirely voluntary basis and edits or deletes vandalism. Wikipedians have contructed many tools, with names such as WikiGuard and WikiMonitor, for monitoring and semi-automated editing and

vandalism reversion. On http://www.wpcvn.com Dmitry Chichkov constructs a dynamic web page that displays an overview of recent changes along with extra useful information such as editors 'karma'. More advanced tools does reversion au-An example is ClueBot. 176 tomatically. early tools were mostly rule-based applying simple heuristics, but vandalism detection may also be viewed as a machine learning problem where the task is to classify an edit as a vandalism or not. 177-180 Geiger and Ribes give a narrative of the reverting and blocking of a vandal using a combination of assisted editing tools and bots showing how multiple Wikipedia editors as vandal fighters use Huggle and Twinkle and how the tools interact together with ClueBot over a fifteen minute period. 181 They also show the role of user talk pages in signaling warning messages to the user and to other vandal fighters and their tools. The presently operating ClueBot NG bot automatically reverts much simple vandalism very fast. Semi-automated anti-vandalism tool STiki have been reported to have a median age of reversion on approximately 4.25 hours. 179

Dmitry Chichkov has also reported descriptive statistics on the most reverted articles. XiV He has made statistics sorted on revert ratio and filtered for number of revisions larger than one thousand available from wpcvn.com/enwiki-20100130.most.reverted.txt. Sexual and vulgar articles tend to get high on the list, e.g., 69, Nipple, Urine, Butt and Pussy, but also Geometry, Bubonic plague and Italian Renaissance for some reason.

Based on categories by Viégas et al. ¹⁷¹ Priedhorsky et al. worked with the following 7 categories (features) of Wikipedia damage/vandalism: Misinformation, mass delete, partial delete, offensive, spam, nonsense and other. ¹⁷² Having three independent judges and using a majority vote procedure they found 53% of damages with the nonsense feature, 28% with the offensive feature, 20% misinformation, 14% partial delete and the rest of the features occuring in less than 10% of the damages. In their discussion they noted that mass deletion is easy to detect while the more often occuring feature misinformation "may be the most pernicious form of damage".

Biased editing

In biased editing editors diverge from the neutral point of view of Wikipedia, — one of its cornerstones. It may be associated with conflict of interest. Whereas graffiti vandalism is relatively easy to

spot, biased editing comes usually from competent editors and biased edits may be harder to judge. To boost their reputation organizations or individuals may make such edits or they may pay others, e.g. public relations companies, to do it: Either deleting negative information or add positive. Law Professor Eric Goldman predicted in a 2005 blog post that in five years Wikipedia would fail due to "gamers" and "marketeers". 182, 183 In 2013 Wikipedia is still fairly neutral, but there has been several cases of biased edits.

One early reported case from 2007, that exemplifies the problem, had Microsoft offering to pay blogger Rick Jelliffe to change Wikipedia articles on open-source document standards and the Microsoft rival format. The company was sure that the articles contained inaccuracies. ¹⁸⁴ Controversial edits from an IP address associated with podcasting entrepreneur Adam Curry in the "podcasting" Wikipedia article have also been noted. ^{172, 185}

The investigation of biased edits became considerably easier when American PhD Student Virgil Griffith created and in August 2007 opened the web-based tool called Wikiscanner. It would merge IP address information with Wikipedia anonymous edit information. 186 The tool pinpointed edits from organizations, which were exposed in mainstream media. Some of these edits were harmless spelling correction, others graffiti vandalism, yet others removal of negative information. Examples of the latter category are: Edits from the Vatican on Sinn Fein leader Gerry Adams removed links to news stories that alleged a link to murders in 1971 and edits from voting machine supplier Diebold removed unfavorable information about political funding and alleged rigging of the 2000 United States election. 187 BBC edits rewrote a part on political correctness of the 'Criticism of the BBC' article. 188 The Wikiscanner has even revealed edits from the public relations firm Hill & Knowlton to the Politics of the Maldives Wikipedia articles. The firm was employed in 2004 by the former President to improve the image of the country. 189 The edits removed words such as "torture", "propaganda" and "censorship". After exposure of Wikipedia articles the Dutch justice ministry would temporarily block its 30.000 employers from using Wikipedia at work. 190 A Wikiscanner, constructed by Peter Brodersen, is also available for the Danish Wikipedia. 191 An extension of Wikiscanner using trademark and Ip2location databases would compute link distance between pages and categories and were reported to detect "the majority of COI edits". 192 Biased edit stories continues to pop up in mainstream media from time to time. In 2013, e.g., a user from City of Melbourne edited in the Occupy

 $^{^{\}rm xiv}{\rm Code}$ for Chichkov's method is available at http://code.google.com/p/pymwdat/

Melbourne article,¹⁹³ and the staff of Danish politicians were found to edit Danish Wikipedia articles on their politicians. As troublesome Wikipedia content Andrew Lih would also point to university articles with "glowing accolades, trumpeting selective accomplishments in the opening paragraphs as the most significant, noteworthy events, even if taken from obscure sources." ¹⁹⁴

In 2011 a French court ruled that a company had to pay its competitor company damages after someone from the former company had edited the *Micropaiement* article on Wikipedia. ¹⁹⁵

A user may create multiple accounts, and if used for deception Wikipedia editors speak of 'sockpuppets'. By their multiplicity sockpuppet accounts can create a sense of consensus, bias polls and work around Wikipedia's three-revert-rule for fighting other editors, — an instance of so-called 'gaming the system'. Through text analysis, looking for repeated spelling errors and idiosyncracies in sentence construction, socket puppets may be identified. 196 One sockpuppet was exposed after another user examined the irrelevant references in an article about a company of questionable notability. In the ensuing deletion discussion, where several other "users" defended the company article, sufficient suspicion arose to warrant a sockpuppet investigation, that would eventually uncover a network of 232 confirmed sockpuppet accounts. These accounts predominantly edited in articles about companies and persons being promotional in nature. 197

The Wikiganda, a student project mentored by Virgil Griffith and at one point available from www.wikiwatcher.com, made use of automated text analysis to detect biased edits. The opinion mining and sentiment analysis technique used a lexicon of over 20.000 words from General Inquirer and Wiebe wordlists so each revision could get a 'Propaganda Score' and labeled as negative, positive or 'vague' propaganda. Also using the WikiTrust system and evaluating against 200 manually labeled revisions the system showed a precision/recall performance on 52%/63%. ¹⁹⁸ Our similar online system focused on business-related Wikipedia edits. ¹⁶⁰

A special case of biased editing happened with Gibraltar and high profile Wikipedian Roger Bamkin. Bamkin did not change content per se, but rather promoted Gibraltar-related content in the 'Did You Know' section of the main page. ¹⁹⁹

Wikipedia has the lengthy "Conflict of interest editing on Wikipedia" article, and a Facebook group called *Corporate Representatives for Ethical Wikipedia Engagement* (CREWE), established in the beginning of 2012, points to articles and members discuss the issues around conflict of interest edits. Read also the description of the Face-

book group by DiStaso²⁰⁰ and the Wikipedia article about the Facebook group. Chartered Institute of Public Relations has established best practice guidance for public relation professions, and David King from a firm offering Wikipedia consulting argues for "transparent community collaboration".²⁰¹ Several companies offer paid editing, with one even claiming to have Wikipedia administrators among the editors that can help customers with changing Wikipedia. The journalist Simon Owen gained contact with one of their clients, who said that "they paid between \$500 and \$1,000 to have the page created, then an additional \$50 a month afterwards for 'monitoring'".¹⁹⁷

Does demographics Wikipedia the of contributors—the prototype being the young male procrastinating graduate students—bias the content of Wikipedia? In 2013 American writer Amanda Filipacchi²⁰² and editor Elissa Schappell²⁰³ noticed that the English Wikipedia subcategorized the "American Novelists" category into "American Women Novelists" but not "American Male Novelists". Filipachi would call this "Wikipedia's Sexism", and one could hold the view that relegating women to a subcategory would lower the visibility of female writers. The case would be noted by several media outlets, xv and lengthy discussions ensued on Wikipedia. Apart from the discussion of sexism, Wikipedia contributors also discussed more generally about subcategorization, particularly about subcategorization based on ethnicity, gender, religion and sexuality. Before the "American Women Novelists" case there had, e.g., been discussions about the category "Male actors by nationality" (and its subcategories) in 2012. The result of the discussion was to keep the subcategorization, and as of July 2013, e.g., the "Danish male actors" (of November 2012^{xvii}) persists but has been applied far less widely than "Danish actresses" (also of November 2012xviii). Wikipedia maintains guidelines for categorization by ethnicity, gender, religion, or sexuality and a version reads: "Do not create categories that are a cross-section of a topic with an ethnicity, gender, religion, or sexual orientation, unless these characteristics

xvCategory talk:American novelists on the English Wikipedia. Look under "mentioned by multiple media organizations:"

xviCategory:Male actors by nationality on the page Wikipedia:Categories for discussion/Log/2012 November 22 in the English Wikipedia.

 $^{{\}bf xviii}$ Category: Danish actresses: Revision history on the English Wikipedia

xixWikipedia:Categorization/Ethnicity, gender, religion and sexuality on the English Wikipedia.

are relevant to the topic." and "In almost all cases, gendered/ethnic/sexuality/religion-based categories should be non-diffusing, meaning that membership in the category should not remove membership from the non-gendered/non-ethnic/etc parent category." The end of the American novelists subcategorization controversy is as of July 2013 that no individual pages are grouped there and writers are all moved to either "American women novelist" or "American male novelist".

Authorship

A page will almost always have multiple authors. The revision history records each author contributions, but the format of the revision history makes it not trivial to determine who contributed what and the most, since text may be reformulated, moved, deleted and reintroduced.

To get an overview of the edits the convenient program history flow takes the revision history as input and visualizes the entire history of an article with colorings determined by author. 171 Another related tool—WikiDashboard²⁰⁴—also generates a visualization of the edit activity of each Wikipedia page. It embeds the generated plot in a proxy copy of the Wikipedia article showing the amount of edits of each author through time for the given article. Yet others have presented tools with examples of color coding of text depending on authorship – so-called *blaming* tools. XXX Author color coding ('Einfärben') operates in full in the German competing encyclopedia Wikiweise. Yet another tool, WikiBlame, helps in identifying versions of a Wikipedia article where a phrase occurs. The MediaWiki extension Contributors just counts the number of edits.

These approaches do not reveal if editors just copied text from other articles in Wikipedia. The detection presents a challenge that other researchers also ignore: For a full examination of the originality of an entry one would need to go through the entire revision history up to the date of the entry of all articles in Wikipedia. The same problem arises with translations: An entry may be novel across language versions or 'just' a translation.

A 'good' deletion, e.g., of vandalism or poorly written entries, increases the value of a Wikipedia article, so to which extent should a deletion count in the rank of authorships? In user reputation mod-

eling deletion by administrators are considered to represent good deletions. ²⁰⁵

Authorship determination has importance for redistribution of Wikipedia: Editors release their Wikipedia contributions under a share-alike license and this license requires that everyone copying or distributing Wikipedia lists 'at least five of the principal authors of the Document'. A strict interpretation of this requirement call for a list of Wikipedia editors, e.g., in copies of Wikipedia articles on other web-sites. A quick look makes it evident that redistribution web-sites do not usually honour this strict interpretation.

Dynamic social network analysis may be used to examine editor contribution and role. Network visualization may identify "egoboosters" that relentlessly checks and edits an article. 157

User contributions

Wikipedia authors contribute with varying amount of work. A relatively few number of editors make the great percentage of the number of edits, e.g., Jimmy Wales reported that 'half the edits by logged in users belong to just 2.5% of logged in users' based on a count from December $2004.^{206}$ A study based on a 2007 dump of Wikipedia found that 46% of registered users made only one edit. ²⁰⁷ In 2008 Ortega et al. would quantify the contributor inequality for registered users with the Gini coefficient across several Wikipedias: It was between 0.924 and $0.964.^{208}$

Are the many edits of elite editors of major value? Perhaps the elite editors are just fixing, e.g., minor comma errors. Based on a small investigation of a few Wikipedia articles Aaron Swartz argued that the occational contributor is actually the one that make substantive edits. Swartz had looked on the "Alan Alda" article and results from "several more randomly-selected articles" showed the same pattern. ²⁰⁹

A Minnesota 2007 study considered the "impact of an edit" taking the number of page views into account with a metric the authors called *persistent word view* (PWV),¹⁷² and after some discussions on the Wikipedia research mailing list in November 2008 Alain Désilets summarized with the following assertions:^{xxi}

- Most edits done by a small core
- But, most of the text created by the long tail
- However, most of the text that people actually read, was created by the small core

xx User 'Jah' on the German Wikipedia has made a Perl script for color coding available at http://de.wikipedia.org/wiki/Benutzer:Jah/Hauptautoren. and others display examples: http://mormegil.info/wp/blame/AFC_Ajax.htm and http://hewgill.com/~greg/wikiblame/AFC_Ajax.html.

 $^{^{\}bf xxi}{\rm http://lists.wikimedia.org/pipermail/wiki-research-l/2008-November/000697.html}$



Figure 4: The program history flow can visualize the evolution of an article and easily give an overview when major changes occur: The visualization shows an anonymous user making a major extension of the Danish Wikipedia article about Bobby Fischer on April 20th 2005. However, 'Zorrobot' makes the most changes. This Norwegian program automatically updates links to Bobby Fischer articles in other language versions of Wikipedia.

The Minnesota study had found that by July 2006 more than 40% of the persistent word views where made by users in the top $0.1\%.^{172}$

Counting of anonymous, bot and sock puppet editing as well as the name spaces investigated may confound such conclusions: Are anonymous edits infrequent editors or a registered user not logged in? Is a specific IP-number a network proxy merging multiple authors? Are some of the edits actually automatic edits from an unregistered bot?

Bots and edits via assisted editing tools make up a considerable part of the total number of edits. In 2007 Stuart Geiger found that the number of edits of this type is larger than the number of edits made by anonymous editors. On one particular page such edits made up 75% of the total edits, ^{181,210} see also page 37.

Even in dedicated teaching wikis where students are given writing assignments the level of contribution may be highly variable. $^{211,\,212}$

User characteristics

A Pew Research Center's Internet & American Life Project survey among adult Americans based on telephone interviews found close to 80% were Internet users. Among these Internet users education level, age and home Internet connection type were the strongest predictors for Wikipedia use. Household income and race/ethnicity were also predictors for Wikipedia use, while gender was not a strong predictor: 56% of male and 50% females Internet users used Wikipedia.⁴

In April 2011 the Wikimedia Foundation surveyed Wikipedia editors with the results published in a 75-page report and they reported several user characteristics for the over 5'000 respondents:²¹³ One question is whether one can claim that Wikipedia is built by mostly non-experts or in the words of Andrew Lih 'a bunch of nobodies'. For the highest level of education 8% survey respondents reported a PhD while 18% reported a Master. A large portion of the editors feel proficient with computers with 36% reporting that they program &

create their own applications. In an interview study among 32 contributors to health-related Wikipedia articles 15 (47%) worked in a health-related fields and mainly as clinicians. 214

Wikipedia contributors have a sizable gender gap. Quite a number of surveys have addressed the issue: A 2015 blog post listed 12 surveys.²¹⁵ Among the surveys is for instance the Wikipedia Survey published in 2010. It found that only 13% of contributors were female, ²¹⁶ Another survey, the Wikimedia Foundation editor survey of 2011, found 9% of the respondents to be females.²¹³ The Media-Wiki software allows users with an account to set their gender in the preferences. Robin Peperman's Toolserver application 'User preference statistics' aggregated the statistics across users, e.g., on the Danish Wikipedia it showed that in February 2012 3'153 users declared themselves male and 677 female, i.e. 82.32% and 15.68%, respectively. Lam et al. used the user preference setting statistics on the English Wikipedia and also userboxes on user pages specifying gender. They found ratios in a similar range: Around 13% for the userbox method and around 16% for the preference setting method.²¹⁷ Furthermore, they found that the ratio has remained more or less constant in the time window they looked at (2006–2011). The reason surveys have reported somewhat lower ratios than the user preference setting statistics may be due to survey non-response bias, and Hill and Shaw argue that the survey ratios should be around 25% higher.²¹⁸ One sees the approximately 85-to-15 percent ratio in public thought-leadership forums as monitored by The OpEd Project.²¹⁹ However, the community on free and open source software (FOSS, sometimes Free/Libre and Open Source Software, FLOSS) has a much large gender gap: a survey found only 1.1% females in the sample. ²²⁰ Various reasons for the gender gap can be put forth, e.g., that the Wikipedia editing is somewhat technical and that women are 'scared away' because of hostile discussions in the online community of Wikipedians. Indeed the Wikimedia Foundation editor survey of 2011 found that, e.g., 4% of the responding female editors reported 'I was stalked online'.²¹³

For the WikiWomenCamp May 2012 Laura Hale and others summarized the involvement of women in the different Wikimedia projects across countries, see WikiWomenCamp/FAQ/Perspectives article on the Meta-wiki.

A 2007 study provided evidence that young patients with schizotypal personality disorder would use the Internet for social interaction significant more than controls.²²¹ Does Wikipedians have a special personality type? An Israeli study sought

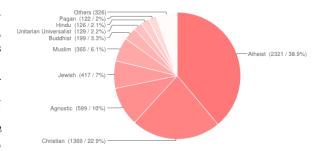


Figure 5: Religions of Wikipedians as analyzed by WikiChecker based on the subcategories of the Wikipedians Category. © WikiChecker.com. GFDL.

to answer this question subjecting 139 Wikipedians and non-Wikipedia to a personality question-naire. Page 1222 In the study Wikipedians scored lower on 'Agreeableness' and higher on 'Openness'. Scores on 'Extroversion' and 'Conscientiousness' personality dimensions depended on the sex of the subject. Based on previous research they hypothesized that Wikipedians would score lower on extroversion, but their results indicated that only female Wikipedians would score lower. The authors wondered why Wikipedians scored lower on agreeableness, suggesting that contribution to Wikipedia as an apparent prosocial behavior, links to egocentric motives such as "personal expression, raising self-confidence, and group identification". Page 222

One of the major dichotomies of user characteristics with respect to Wikipedia editing behavior is concerning inclusionists and exclusionists/deletionist, — inclusionists being contributors believing "that Wikipedia should contain pretty much anything, as long as it's factual and verifiable", ²⁵ while exclusionists being more stringent on what go into the encyclopedia.

With the use of infoboxes or explicit categorization on their user page users may manifest that they belong in one or more user categories. There is a sizable network of user categories, for the English Wikipedia, see the subcategories from the top at Category:Wikipedians. Some users explicitly declare themselves as, e.g., inclusionists or Hindu Wikipedians. The www.wikichecker.com website made analysis results available of some of the user categories across a number of language versions of Wikipedia. Among the results one finds that of 5'973 users manifesting their beliefs in total on the English Wikipedia 7% declare themselves to be Jewish and 38.9% to be Atheists (see Figure 5), xxii while 20.1% of 4,433 manifesting their editing phi-

xxiihttp://www.wikichecker.com/religion/, February 2014.

losophy declare themselves to be inclusionists. xxiii

Geography

An IP number can be resolved to a geographical location ("geo lookup"). Wikimedia data analyst Erik Zachte used information from server logs, converted it to a geographical coordinates and constructed a dynamic visualization of the edit pattern through time with world map background. xxiv There are websites with similar dynamic visual-László Kozma's Wikipedia Vision runs ization. from http://www.lkozma.net/wpv/. Wikipedia Recent Changes Map was built by Stephen La-Porte and Mahmoud Hashemi and available from http://rcmap.hatnote.com in May 2013. It captured edits from unregistered users on different language versions of Wikipedia in real-time and displayed them on a map. While these Web services focused on recent changes edits Federico Scrinzi, Paolo Massa and Maurizio Napolitano's Wiki Trip website would display geo-resolved unregistered edits based on data from single Wikipedia articles through time with an OpenStreetMap map as background.

With access to page view data, Erik Zachte construted the WiViVi service with a map of page view with respect to Wikipedia language version. xxv

Students from the Technical University of Denmark constructed a Web service that would download the blocked IP addresses information from Wikipedia (or any other MediaWiki installation) and after a geo lookup render the geographical distribution of the blocked IPs with a temporally static heat map visualization.

By examining the user page with natural language processing techniques researchers determined the country associated with a user. They further examined geotagged English Wikipedia articles associated with the Middle East and North Africa (MENA) region and found that Westernbased editors formed the majority of contributors to the MENA articles. Another investigation of the geotagged articles showed that across 44 language versions of Wikipedia more than half of the 3'336'473 geotagged articles had a geotag inside a small circle emcompassing all of western Europe,

visualized/

southern Scandinavia and parts of Eastern Europe, i.e., showing a geographically uneven coverage of Wikipedia. 223

Organization and social aspects

How is Wikipedia organized? In a critical news comment from 2008 Seth Finkelstein characterized it as "a poorly-run bureaucracy with the group dynamics of a cult", an "oligarchy" and having "an elaborate hierarchical structure which is infested with cliques and factional conflicts". 224 Researchers describe Wikipedia and its relation to concepts such as "community", "social movement", "benevolent dictator" and "network sociality". 38, 225, 226 Several studies refer to a "core group of dedicated volunteers" or "critical mass", and Wikipedia has been regarded as governed by a so-called "benevolent dictator" or "constitutional monarch" (Jimmy Wales). 19, 35, 225, 227 The new information and communication technologies have been regarded as creating "network sociality" rather than a "community", — not based on a common narrative and eroding enduring relationships.²²⁶

Papers compare the Wikipedia development model with FOSS development. 228, 229 In his classic text on FOSS development, The Cathedral and the Bazaar, 230 Eric Raymond came up with the concept of bazaar-type development as opposed to cathedral-type development, with bazaar development characterized by an open development process making early and frequent releases of the developed code. Wagner has compared Wikipedia work with the bazaar approach and argues that Wikipedia fit many of the features of bazaar-style development.²²⁸ On the other hand Magrassi argues that FOSS requires more top-down coordination than Wikipedia because that software needs to be coherent while Wikipedia can have low coherence: A Wikipedia article may still be good even if other articles are of low quality or non-existent. In major FOSS systems one may see a hierarchical structure with a "benevolent dictator" or "core developers" on the top, followed by "co-developers" and ordinary developers, active users and passive users.^{229,231} For the "development" of an article on Wikipedia this is not necessary.

Wikipedia may meet face-to-face^{25,35} and Wikipedians can coordinate work in WikiProjects. However, most of the construction process of Wikipedia takes the form of an *indirect collaboration*, where individual members in a group make small changes to a shared structure inspiring other to improve it even further, — to paraphrase Peter Miller.²³² Indeed researchers have frequently in-

 $^{^{\}rm xxiii}{\rm http://www.wikichecker.com/editing-philosophy/,}$ February 2014.

 $^{^{\}rm xxiv}{\rm The}$ visualization was available from http://stats.wikimedia.org/wikimedia/animations/requests/-AnimationEditsOneDayWp.html and a blog post had background information http://infodisiac.com/blog/2011/05/wikipedia-edits-

xxvhttps://stats.wikimedia.org/wikimedia/animations/pageviews/wivivi.html

voked Pierre-Paul Grassé's concept of stigmergyoriginating in the study of animal behavior—and applied it on the Wikipedia process. 25, 194, 232-235 Mark Elliott would argue that "collaboration in small groups (roughly 2-25) relies upon social negotiation to evolve and guide its process and creative output" while "collaboration in large groups (roughly 25-n) is dependent upon stigmergy" admitting that in the stigmergic collaborative context of Wikipedia social negotiation can take place (e.g., article discussion, email, Internet relay chat), but take a secondary role. He also sees stigmergic wiki collaboration as distinct from "co-authoring", where his idea of "co-authoring" consists of social negotiation in the creative gestation period.²³⁴

Wikipedia has seen an increasing institutionalisation with Wikimedia Foundation, Arbitration Committee and Association of Members' Advocates.¹⁹

The Wikimedia Foundation has the primary role of running the computer that serves Wikipedia and its sister projects. Other organizations may also affect Wikipedia. The transition from GFDL to CC licence involved the Wikimedia Foundation, the Free Software Foundation (FSF), Wikimediawide voting and Wikimedia Foundation Board of Trustees decision. FSF also acts as a license stew $ard.^{236}$

Other issues that the research literature has discussed are privacy²³⁷ and legal aspects (copyleft license). 236

Popularity

A number of companies, such as comScore and Alexa, collect usage statistics from Web users, and such statistics allows Wikipedia to be compared to other Web sites. The September 2006 traffic got Wikipedia Sites listed as number sixth on com-Score's worldwide ranking,²³⁸ and in October it broke into U.S. comScore top ten.²³⁹ Alexa put wikipedia.org as number 7 in traffic rank in September 2009. xxvi stats.wikimedia.org/reportcard reports the temporal statistics from comScore statistics, and it, e.g., shows that throughout most of the first half of 2010 Wikipedia had 350 million unique visitors and around 8 milliard page requests. In January 2016, Wikimedia Foundation chosed to stop using the comScore statistics as they believed the data were "no longer fully representative" of their traffic. $^{\tt xxvii}$

Pew Research Center's Internet & American Life

 $^{\rm xxvi} \rm http://www.alexa.com/siteinfo/wikipedia.org$ continuesly Wikimedia. record page and comment on the Alexa statistics:

Project has conducted surveys on Wikipedia use in 2007 and 2010. In May 2010 42% of adult Americans used Wikipedia. This was up from 25% in February 2007.⁴ At 24% Wikipedia is by far the most used educational and reference site. In 2007 Yahoo Answers was second (4%), while, e.g., Google Scholar was on 1%.3 In 2010 use of Wikipedia among Internet users was more popular (with 53%) than instant messaging (47%) and rating a product, service or person (32%), but less popular than using social network sites (61%) or watching online videos (66%).⁴

Another way of capturing statistics related to usage is through the Google Trends Web service (http://www.google.com/trends) that displays search volume for the Google search engine through the years based on a query, see Figure 7. Users use "wikipedia" and "wiki" to refer to Wikipedia, but "wiki" may also be used to refer to other wikis related to, e.g., Minecraft, Naruto and Skyrim. There have been steady declines in the "wikipedia" search volume since the middle of 2010 and for "wiki" since september 2011. The declines are not necessary an indication of a decline of the use of search engine to access information on Wikipedia or other wikis. Given that Wikipedia articles rank high in search engine results, one may speculate that Internet users have come to be expect this and not finding it necessary to explicitly mention Wikipedia in the search query. The Google Trends has also be used to collected data related to individual Wikipedia articles. Gene Wiki researchers correlated Google Trends and Wikipedia article views for genes.²⁴⁰

An analysis of the Arabic blogosphere identified the English Wikipedia as the second most often linked to site from Arabic blogs and only surpassed by YouTube. After Al Jezeera, BBC and Flickr the Arabic Wikipedia ranked sixth in received citations. 241

Several researchers have investigated the Internet search ranking of the articles of Wikipedia. $^{242-245}$ Using search engine optimization techniques two researchers investigated the Google ranking of the English Wikipedia for health topics. The queries were 1726 keywords from an index of the American MedlinePlus, 966 keywords from a NHS Direct Online index and 1173 keywords from an American index of rare diseases (U.S. National Organization of Rare Diseases) and compared Wikipedia to .gov domains, MedlinePlus, Medscape, NHS Direct Online and a number of other domains. They found the English Wikipedia as the Web site with most top rankings.²⁴²

Wikipedia's Internet search engine ranking on inhttp://meta.wikimedia.org/wiki/Wikipedia.org_is_more_populalividual queries may depend on the type of query xxviihttps://meta.wikimedia.org/wiki/ComScore/Announceme@whether it is navigational, transactional or infor-

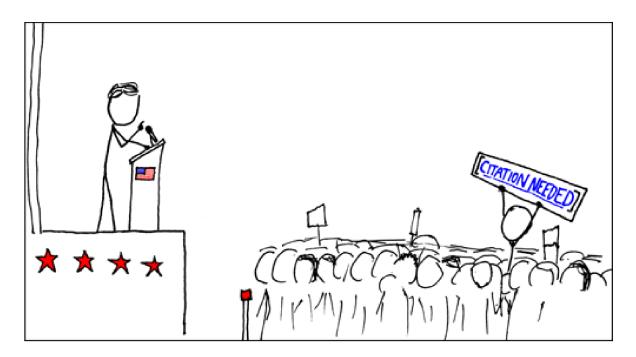


Figure 6: With the rise of its popularity Wikipedia finds its way into comics strips. Here the "Citation needed" template in Wikipedia used in Randall Munroe's XKCD. © Randall Munroe. CC-BY.

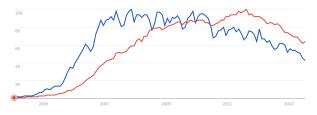


Figure 7: Google Trends for the query "wikipedia, wiki". "Wikipedia" and "wiki" are the blue and the red curve, respectively.

mational) and the number of words in the query. One study reported in 2012 found that single word informational queries (example: "lyrics") placed Wikipedia on page one of Google for 80–90% of the queries, while, e.g., 3-words transactional queries (example: "boots for sale"), only placed Wikipedia on page one in under 10% of the cases. Navigational queries may yield even lower Wikipedia ranking. ²⁴⁴ By not distinguishing between types of queries, not using "real" queries, but random nouns, another study had found Wikipedia on 99% of Google's page one. ²⁴⁵

Based on "customer" interviews *The American Customer Satisfaction Index* (ACSI) has ranked Internet social media in 2010 to 2013 with Wikipedia ahead of YouTube and Facebook. In this category Wikipedia has maintained the highest score through the years 2010–2013 among the 8 web-

sites analyzed. However, these social media sites have generally been ranked lower than most Internet portals and search engine, such as Google, Bing and Yahoo!, especially for the years 2011 and 2012. Wikipedia has consistently maintained its score on around 77–78, and as the ACSI of Internet portals and search engines dropped in 2013 (perhaps as a result of the leaks of Edward Snowden?), Wikipedia would score higher on the index than these websites. In the more general category "E-Business" only FOXNews.com would have a higher score than Wikipedia for the year 2013. 246, 247

A 2012 survey found that almost 50% of 173 Danish medical doctors used Wikipedia search for information related to their profession. $^{248,\,249}$ The survey recruited subjects via the online network DoctorsOnly which might have biased the results.

Detailed analysis of the individual articles of the English Wikipedia shows that bursts of popularity on individual pages occur during Super Bowl halftime entertainment and around the death of a famous subject, e.g., Whitney Houston, Amy Winehouse and Steve Jobs. Other causes of increased view of particular pages are the so-called Google Doodles, denial of service attacks on particular pages, second screen and Slashdot effects. ²⁵⁰ Page views on individual articles are highly skewed, see Figure 8. ^{250,251}

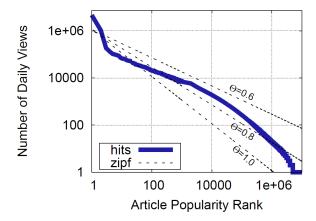


Figure 8: Distribution of page views in the English Wikipedia.^{250,251} © Andrew G. West. CC-BY-SA.

Economy

How is the economics of Wikipedia? What is its economical value? How much would users be willing to pay for it? What would it cost to reconstruct Wikipedia? In 2013 these issues were considered in a short review by Band and Gerafi, that concluded that the economic value of Wikipedia was in the range of tens of milliards and consumer benefits hundreds of milliards, ²⁵² see Table 6. One study included in the review estimated the market value by comparing Wikipedia to other Internet companies in terms of reputation and unique visitors. This economic triangulation with LinkedIn, Twitter and Facebook put the market value of Wikipedia to between \$10 and \$30 milliard (United States Dollars). Other methods estimate the economic value of Wikipedia between \$6.6 and \$1'080 milliard. The extreme maximum value comes from considering a Wikipedia reader session as a 'research question' and using a value of a 'research question' on \$45 from the National Network of Libraries of Medicines. Given the popularity of entertainment and pop cultural phenomenons on Wikipedia²⁵³ and that, e.g., students report using Wikipedia for 'entertainment or idle reading' it seems questionable that the majority of Wikipedia sessions should be regarded as valuable as a 45dollars research question.

Estimates of replacement costs may use estimate of labor hours. Geiger and Halfaker has put one such estimate forward for data on the English Wikipedia up until April 2012 reporting a value of a bit over 40 million hours. Extrapolating to all Wikipedias they reported a number on approximately 103 million labor hours.²⁵⁵ Another Internet phenomenon may provide comparison: Korean pop artist PSY's 4:12 minute xxviiiPrograms:Evaluation portal/Library/Edit-a-thons

Milliard USD	Description
10-30	Market value from comparison
	with other social media sites.
21 - 340	Market value based on in-
	come from potential fee paided
	by reader subscribers times
	revenue-value ratio
8.8 – 86	Market value based on possible
	advertising revenue
6.6	Replacement cost based on a
	\$300 per article cost
10.25	Replacement cost based on esti-
	mates of labor-hours and a \$50
	per hour cost
$0.63/\mathrm{year}$	Estimate for cost with full-time
	paid writers employed to up-
	date Wikipedia
16.9 - 80	Annual consumer value based
	on received benefit that could
	have had a fee
54 - 1'080	Annual consumer value based
	on comparison with librarian
	research question cost

Table 6: Economic value of Wikipedia based on an 2012 overview by Band and Gerafi.²⁵² 1 milliard = 1'000 million.

YouTube video Gangnam Style reached 2 milliard views in May 2014 corresponding to 140 million hours, thus the hours used on building all the Wikipedias compare roughly with time spend on watching the (as of 2014) most popular YouTube video. For the double-digit milliard market value estimate from advertising it is worth remembering that just a small note about the potential of advertisement ("Bomis might well start selling ads on Wikipedia sometime within the next few months, [...]") spawned the Spanish Fork in $2002.^{25}$

Edit-a-thons, where (unpaid) editors come together writing Wikipedia articles, may have a variety of goals, such as writing new articles, extending or discussing existing, socializing among Wikipedians, recruiting new editors or increasing general awareness. In 2013 Sarah Stierch from the Wikimedia Foundation surveyed a number of these edita-thons. xxviii Some of the surveyed program leaders tracked budget and donations for the event. Stierch found an average edit-a-thon budget on around \$360, and for the 5 events where she had sufficient data she could compute a 'production cost' of around \$17 for each 'printed page' (equivalent of 1'500 characters). English Wikipedia articles had

an average of 3'655 bytes per article in January 2010. xxix Other language versions usually have a lower byte count, so a rough estimate of an average Wikipedia article is between 1 and 2 printed pages. With \$25 per Wikipedia article and 30.8 million Wikipedia articles as the official January 2014 count the replacement cost is only \$770 million with unpaid 'edit-a-thon' writers, — considerably lower than any of the values presented by Band and Gerafi. In 2014 Stierch also became the source for the price of a 'Wikipedia page for individual': At oDesk she reported her personal writing service at $$300,^{256}$ — a value that corresponds with the article charge assumed by Band and Gerafi when they computed the estimated replacement cost of all of Wikipedias. 252,257

The Wikimedia Foundation gets donations throughout the year, and it makes some of the donation data available, e.g., from the Web server frdata.wikimedia.org. The donation data has not caught a major share of researchers' attention, — if any at all.

Why do people edit?

What motivates Wikipedians? Wikipedia itself has a section in the "Wikipedia community" English Wikipedia article discussing the motivation of volunteers giving pointers to research on the issue. Much of the research in this area is based on qualitatively analyzed interviews of Wikipedians. ^{20,258,259} However, the area has also seen mathematical modeling being employed on edit histories and survey data. ^{260,261} Others have conducted 'field' experiments on Wikipedians or relied on events for natural experiments. ^{262,263}

Some Wikipedians mention in interviews that they started editing Wikipedia because they discovered errors or omissions in an article that they knew something about.²⁵⁸ Does this means that the more correct and complete Wikipedia becomes the more difficult it will be to attract new editors?

The attraction of Wikipedia is in part due to the low barrier for entry, ^{258,264} also refered to as low opportunity cost. ¹⁹ Introducing required user registration and 'sighted' or 'flagged revisions' would heightened the barrier and possibly result in less recruitment of new users. The wiki markup could form a barrier for entry, and Wikimedians have hoped that a WYSIWYG editor would increase take-up of new contributors, and indeed a rational for developing of the of the so-called VisualEditor stated: "The decline in new contributor growth is the single most serious challenge facing the Wikime-

dia movement. Removing the avoidable technical impediments associated with Wikimedia's editing interface is a necessary pre-condition for increasing the number of Wikimedia contributor". 265

But why do Wikipedia contributors keep on editing? The same question may be asked for free and open source software programmers working for free. Some researchers view Wikipedia as a public good where the contributors "devote substantial amounts of time with no expectation of direct compensation" and a Wikipedian's participation seems "irrationally altruistic". ²²⁵

Self-interest may lay behind some contribution: A software program can solve a specific problem for its creator, and a Wikipedia entry can organize information relevant for the editor to remember. But Wikipedia editors not only work on information directly relevant for themselves. Surveys among Free and Open Source developers mention skill development as the most important factor for beginning and continuing. Tonce the skills are developed the software showcases the skills for potential employers, so the contribution can be seen as a form of self-marketing on the job market. For Wikinews, Jimmy Wales noted that contributors recognized it a site, where they could learn to become a journalist. 124

It is unclear if Wikipedia contribution has the same skill signaling value as free software development. Andrew George writes that the potential is "simply nonexistent". 225 Nonexistent is not entirely true: A few Wikipedians have gained temporary employment as Wikipedian in Residence. Such positions typically require thorough experience with Wikipedia. It is possible that experienced Wikipedians may also look forward to a job with paid editing. The controversial company Wiki-PR boasts of having a "network of established Wikipedia editors and admins" available to edit Wikipedia on behalf of companies.²⁶⁶ A few people are professionally employed to work on Wikipedia. On the Danish Wikipedia articles on runic stones have been made by professionals as part of a job. In her survey among 1'284 public relations/communications professionals DiStaso found that 31% of the respondents had edited in their client or their company Wikipedia page.²⁰⁰ It may be worth to note that now company paid contributions form the majority of work carried out on the FOSS Linux kernel: Only between 13.3% and 16.3% of Linux kernel changes come from contributors without corporate affiliation, ²⁶⁷ thus there is a marked difference between FOSS and Wikipedia labor market.²⁶⁸

After interviewing 22 volunteer Wikipedians Forte and Bruckman likened the incentive system of Wikipedia to that of the scientific community,

 $^{{}^{\}mathbf{x}\mathbf{x}\mathbf{i}\mathbf{x}}\mathbf{h}\mathbf{t}\mathbf{t}\mathbf{p}://\mathbf{s}\mathbf{t}\mathbf{a}\mathbf{t}\mathbf{s}.\mathbf{w}\mathbf{i}\mathbf{k}\mathbf{i}\mathbf{m}\mathbf{e}\mathbf{d}\mathbf{i}\mathbf{a}.\mathbf{o}\mathbf{r}\mathbf{g}/\mathbf{E}\mathbf{N}/\mathbf{S}\mathbf{i}\mathbf{t}\mathbf{e}\mathbf{m}\mathbf{a}\mathbf{p}.\mathbf{h}\mathbf{t}\mathbf{m}.$

Motivation

Low barrier 19,258,264

Wanted to correct errors 213,258

It solves a problem

Self-expression³⁵

Self-efficacy, sense of personal achievement $^{261,\,264,\,268}$

Skill development, self-education, future employment, career^{19, 20, 37, 213, 264, 269, 270}

Social status, peer recognition, reputation, visibility, enhancement 19,35,225,262,264,268,270,271

Peer response, get reactions 35

 $\begin{array}{ll} Community & identification/involvement/social \\ relations ^{35,37,225,264,268,270} \end{array}$

 ${\it Ideological}, \ {\it Weltverbesserungs antrieb}^{35,\,270}$

Altruism, values, volunteering^{213,225,268–270}

Money/part of job^{256}

Creative pleasure, joy, 'flow', fun^{19, 20, 37, 213, 225, 269, 270}

Protection, reducing guilt and loneliness $^{270}\,$

Obsession, encyclopedic urge, wikipediholism 235,272

Table 7: Explanations (motivations) for wiki work.

where the "cycle of credit" is the most important aspect in the incentive system and prestige is earned over long periods of time.²⁷³ Attribution of authorship may on the surface seems to be a difference between the science world and Wikipedia (authors of Wikipedia articles are not displayed prominently on the page). However, Forte and Bruckman find that through the Wikipedia article editing history contributors recognize one another and that Wikipedians also often claim "ownership" of articles. Another comparison of Wikipedia and academic writing was much more critical: Seth Finkelstein commented on Wikipedia prestige and strongly critized Wikipedia claiming that "it fundamentally runs by an extremely deceptive sort of social promise. It functions by selling the heavy contributors on the dream, the illusion, that it'll give them the prestige of an academic ('writing an encyclopedia')". 274

Yang and Li attempted to contact a random sample of 2'000 users on the English Wikipedia. Receiving 219 valid responses to their questionnaire they used structural equation modeling to describe knowledge sharing behavior in terms of intrinsic motivation, extrinsic motivation, external self-concept and internal self-concept. They found that internal self-concept-based motivation was the most important factor for the knowledge sharing behavior, and this factor was associated with questions such as "I consider myself a self-motivated person" and "I like to share knowledge which gives me a sense of personal achievement". On the other hand intrinsic motivation was found to rarely motivate. This factor was associated with questions such as "I enjoy sharing my knowledge with others" and "Sharing my knowledge with others gives me pleasure". 261 Other researchers have also distinguished between intrinsic and extrinsic motivations. 19 Müller-Seitz and Reger mentions low opportunity costs, reputation among peers, future career benefits, learning and development and contributions from the community as extrinsic motivations while creative pleasure, altruism, sense of belonging to the community and anarchic ideas as intrinsic.

Andrew George proposes the state of 'flow' as an intrinsic motivation for Wikipedians, altruism, community identification and peer recognition. ²²⁵ Based on interviews Rosenzweig noted that many Wikipedians had a passion for self-education. ²⁰

Barnstar awards are a form of peer recognition in Wikipedia, and a few studies have analyzed this element with respect to motivation.^{262, 275, 276} In the experiment of Restivo and van de Rijt reported in 2012 the researchers split 200 productive Wikipedians in two groups and awarded a barnstar to the

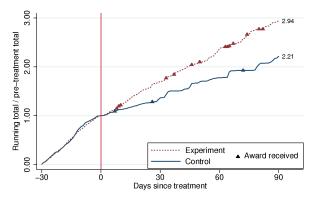


Figure 9: Effect of barnstar awards on productivity of Wikipedians: Figure 1 from Restivo and van de Rijt (2012). Experiment is median productivity of 100 barnstar awarded Wikipedians, while "control" is matched Wikipedians that received no awards from the experimentors. Third-party awards are displayed with triangles. © Restivo and van de Rijt. CC-BY.

Wikipedians in one of the groups ("experiment" in Figure 9) and none in the other group ("control"). The productivity decreased in both groups, but the productivity was sustained better in the group with awarded Wikipedians, see Figure 9.²⁶² Benjamin Mako Hill and Aaron Shaw would also see an effect of barnstars summarizing their findings with: "We show barnstars can encourage editors to edit more and for longer periods of time, but that this effect depends on the degree to which the recipient chooses to use the award as a way to show off to other editors". ²⁷⁶ In his classic work on FOSS development, The Cathedral and the Bazaar, Eric Raymond also discusses motivation for voluntary work and mentions the concept of "egoboo", a shortform of ego-boosting. He argues that science fiction fandom "has long explicitly recognized 'egoboo' (ego-boosting, or the enhancement of one's reputation among other fans) as the basic drive behind volunteer activity." ²³⁰ The existence of public edit counter tools xxx as well as pages such as Wikipedia:List of Wikipedians by number of edits allow Wikipedians to 'egoboost' via their 'edit count'.

The notion of addiction in Wikipedia contribution (colloquially 'wikipediholic', 'wikiholic' or 'wikipediholism') has also been noted. ^{235, 277, 278} New York-based Wikipedia researcher Dan Cosley even claimed "It's clearly like crack for some people". ²⁷⁷ Finnish researchers proposed gateway theory, originating in the study of drug use, as a frame-

work for explaning online participation in FLOSS development and Wikipedia contribution (cautioning that they do not claim that "online participation, at large, would qualify as addiction").²⁷² Contrasting with antecedent-based explanations, the gateway theory-based argument notes that the initial online participation occurs 'due to chance' without prior planning.

Evolutionary biological explanations of religion may involve the idea of religious rituals as reliable costly signals between adherents for social commitment. Should the edit count partially be explained as a hard to fake, costly signal between the adherents of the 'Wikipedianistic' religion?

Why do people leave?

Studies of why editors leave Wikipedia are much less performed than studies on why editors start out. In 2009 Wikimedia Foundation staff and volunteers developed an online survey and emailed it to 10'000 contributors. The included contributors should have had between 20-99 edits and no edits in the past three months. The survey conducted in the beginning of 2010 had a response rate on 12% from 1238 editors. Around half left due to personal reasons (e.g., other commitments), another half left due to "bad terms". The researchers split the "bad terms" into reasons related to complexity (e.g., writing a encyclopedia is difficult) and community (e.g., other editors being rude). Among the high scoring statements for leaving were "my work kept being undone" and "several editors were too stubborn and/or difficult to work with". 280 Eric Goldman speculates that Wikipedia has a particular vulnerability life changes among its contributors, because they tend to be young, unmarried and childless.²⁶⁸ The Wikimedia Foundation editor survey of April 2011 found 'less time' with 37% as the most reported reason for becoming less active on Wikipedia, compared with, e.g., 7% for 'conflict with other editors'.²¹³

Based on large-scale quantitative analysis of, e.g., the editing pattern and its temporal evolution the researchers concluded that Wikipedia had a slowing growth up to 2009. They explained this phenomenon with "limited opportunities in making novel contributions" (the low-hanging fruits are already taken) and "increased patterns of conflict and dominance due to the consequences of the increasingly limited opportunities". An increasing 'xenophobia', where seasoned Wikipedia editors under constant spam and vandalism threat develop a 'revert first' mentality and away from 'assume good faith', may make it harder for contribution from unsophisticated users to stick. Perhaps this

xxx See Wikipedia: WikiProject edit counters on the English Wikipedia.

aspect affects the retention of new contributors.

Contributor may also leave because of project forking, 282 examples being Wikitravel and the so-called Spanish Fork of Wikipedia. 25

Edit patterns

Several studies analyze and models the edit patterns of users and articles. 278, 283, 284

One claim for the number of new edits to an article within a time interval comes from Wilkinson and Huberman, who hypothesized that the number of new edits in the time interval " $\Delta n(t)$ is on average proportional to the total number of previous edits" with some amout of fluctuation 283,284

$$\Delta n(t) = [a + \xi(t)]n(t), \tag{1}$$

where a is a constant representing the average rate of edits and $\xi(t)$ is the fluctuation. Via the central limit theorem they arrive at the lognormal distribution for the number of edits

$$P[n(t)] = \frac{1}{n\sqrt{2\pi}\sqrt{s^2t}} \exp\left[-\frac{(\log n - at)^2}{2(s^2t)}\right],$$
 (2)

where $\mu=at$ and $\sigma^2=s^2t$, i.e., the mean and variation of this lognormal distribution it linearly related to the age t of the article. They show on a robot-cleaned dataset from the English Wikipedia that the distribution of the number of edits follows the model well except for a few events, e.g., one is the edits of US town articles. It is fairly surpricing that the number of edits can be modeled with such a simple model that only contains Wikipedia-intrinsic variables. No external forcing variables, e.g., number of editors, 'importance' or 'notability' of article topic, appear in the equation. Note that the data set only covers the period where Wikipedia had exponential growth and making it how well the model fits newer data.

The overall temporal evolution of the number of articles and editors are perhaps the most import statistics of Wikipedia. It is regularly computed by Wikimedia Foundation data analyst Erik Zachte and reported on the Wikimedia Statistics homepage http://stats.wikimedia.org/. stant worry for Wikimedia is the statistics of the number of active editors and the editor retention. The decaying number of active editors on the English Wikipedia has been apparent for a number of years and given rise to the so-called 'oh shit' graph, see Figure 10, which shows the number of active editors to peak in 2007 and then decaying. The number has continued to decrease and was in the beginning of 2014 around 30'000 according to the Erik Zachte data analysis.

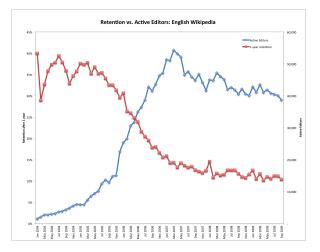


Figure 10: The so-called 'oh shit' graph: English Wikipedia: Retention Rate vs. Active Editors 2004-2009 with a decay number of active editors. From Wikimedia Commons by user Howief.

Wikipedians have reported mathematical models of growth of the number of articles on the English Wikipedia and made them available from Wikipedia: Modelling Wikipedia's growth. Between 2003 and 2007 an exponential model fitted the data well.²⁸⁵ A logistic model estimated from data available in 2008 indicated that Wikipedia would reach a maximum around 3 to 3.5 million articles, a result reported by Suh et al. in $2009.^{281}$ As the logistic model does not model the data well at least after 2010 other models have been suggested and a seasonal variable have been included in the model. A model involving the Gompertz function, $y(t) = ae^{be^{ct}}$, was reported to project a maximum of about 4.4 million article, — a number which the English Wikipedia has surpassed. Suh et al. suggested a hypothetical Lotka-Volterra population growth where the growth was not bounded by a constant but limited by a linear function of time arguing "there is a general sense that the stock of knowledge in the world is also growing." ²⁸¹

Why does it work?

Why does Wikipedia not succumb to vandalism and opinionated content? There are other open fora on the Internet: The Usenet and emails. The Usenet is like Wikipedia a public space, but from a start among 'geeks' it has been used for pornography and pirated content. Another collaborative knowledge building project, Open Directory Project, also relied on free labor, but was in 2006 declared to be 'effectively worthless' by Eric Goldman. 183

In 2011 Benjamin Mako Hill presented his research about Wikipedia and other crowd-sourced

online encyclopedia, Interpedia, The Distributed Encyclopedia, Everything 2, h2g2, Aaron Swatch's The Info Network, Nupedia and GNUpedia, sought to answer what distinguished the successful Wikipedia from the failed or less successful projects. Hill noted that Wikipedia offered low transaction cost in participation and initial focus on substantive content rather than technology.²⁸⁷ Two of the earliest comment on the Wikipedia phenomenon pointed to the low transaction cost as a reason for the Wikipedia success. 264,288 Philippe Aigrain noted low transaction costs as an essential factor for success, — not only for Wikipedia but for other "open information communities", e.g., Slashdot.²⁸⁸ Aigrain would see transaction costs as having a number of aspects apart from (possible) monetary costs: cognitive, information, privacy, uncertainty and locking-in costs, with, e.g., the cost of "navigating the transaction management layers" as a information cost.²⁸⁸ Aigrain also argued for two other factors contributing to the Wikipedia success: a clear vision and mechanisms to fight hostile contributions. He sees the statement on neutrality of point of view as a clear vision and the revision control system as the key mechanism to counter hostile or noisy contributions, 288 and indeed it seems very difficult to imaging how Wikipedia could work without its revision control system where contributors can restore 'blanked' articles. There are likely a large number of important technical elements in the MediaWiki software that lower transaction costs for 'good' contributors and heightened it for 'bad': CAPTCHAs, 'autoconfirmed' user level, user blocking, IP range blocking, blacklists, new pages patrolling and watchlists. The flagged revisions of the German Wikipedia might be the reason for its high quality. The ability of Wikipedians to communicate beyond the stigmergic article space collaboration, e.g., via discussion pages, is presumably important, — at least they are almost always used for high quality 'featured articles'.⁴¹

As a newer development of a Wikipedia challenger, Google launched the *Knol* service in 2008, and while contributors could monitize their content Google announced the closing of the service in 2011. Although a quantitative study could find differencies in the content of articles of Google Knol and Wikipedia, e.g., with respect to the number of references, ^{289,290} it seems unclear if we can blame such differences for Google Knol's demise. One commentor noted that "Google didn't seem to allocate many resources to the project, and it certainly didn't put much emphasis behind it from a press standpoint". ²⁹¹ Before Knol and with likely inspiration from Wikipedia Microsoft unsuccessfully opened for user contribution to their encyclopedia

Encarta. Andrew Lih noted "it was not timely, open, social, or free." and a "half-hearted implementation". 25

I have not found mentioned that Wikipedia is a excellent example of hypertext. Much quality text on the Web is not linked very much, e.g., a typical New York Times news article has typically no links in the body text. Web scientific articles rarely has links. Google Knol had typically no hypertext links besides table of content-type or external references. Even MediaWiki-based Scholarpedia has few intrawiki links. URLs in Danish encyclopedia Den Store Danske and Google Knol are/were not well-predictable. Links on the ordinary Web based on non-wiki technology do not display the availability of the linked page. The URLs of Wikipedia articles are well-predictable, likely reducing navigational costs.

The creation of Wikipedia is a decentral process without economic incentives, where individual contributions get aggregated to a collective result. Certain Web-games have the same character: In the Web-based market game Hollywood Stock Exchange the goal is to predict the box office of a movie by buying and selling virtual shares. Four weeks after the opening the Hollywood shares get the toy money in accordance with the earnings of the movie. The price of the stock then becomes a prediction on the real earnings of the movie. Artificial market games perform surprisingly well in predictions, e.g., Hollywood Stock Exchange could predict Oscar winners better than individual experts.^{292,293} Apparently the prediction markets attracts well-informed and well-motivated players whose mutual trade creates information, even though the participants does not gain any money. There is not real trade in Wikipedia, but Wikipedia forms a decentral knowledge aggregation, like the knowledge markets.

The 'power' of the collective has often been invoked and sometimes under the name Linus' Law, that Eric Raymond coined in The Cathedral and the Bazaar to characterize an element in FOSS development: "Given a large enough beta-tester and co-developer base, almost every problem will be characterized quickly and the fix obvious to someone," 230 i.e., openness gives robust results. The concept has also been applied on Wikipedia to explain its success.²⁹⁴ Greenstein translates the 'law' to the Wikipedia context with: "Enough attention to such details [facts and commonly accepted knowledge presented on Wikipedia elicits the relevant objective information and many viewers then correct errors as needed." and "[g]iven enough cheap storage, all viewpoints can be represented." ²⁹⁴ He also notes its difficulties with costly

verification and subjective information where Linus' Law might not work.

3 Using Wikipedia

Users meet not only Wikipedia as readers of the Wikipedia site. Other web-sites utilize Wikipedia content to enhance their own site, and the rich structured annotations with links, tags and templates allow researchers and developers to use Wikipedia as a knowledge base. Especially researchers in information retrieval and automatic text processing benefit from Wikipedia. In this kind of research WordNet has stood as the foremost open computer-based lexical database for the English Language. Some research compares or combines WordNet and Wikipedia. The review Mining meaning from Wikipedia from 2009 gives an overview of studies using Wikipedia data in natural language processing, information retrieval and extraction as well as ontology building.⁹

Other wikis besides Wikipedia provide data for this kind of research. Wiktionary, the dictionary-like sister project, supplements Wikipedia, and several studies have used information from Wiktionary. ^{295–298}

Information retrieval

A program uses Wikipedia articles as references to detect so-called "content holes". Content holes are neglected or missed subtopics in a text. The researchers regarded user generated content from discussion threads and compared that to the sections found in Wikipedia articles by extracted keywords such as proper nouns and numbers. ^{299–301}

Web search queries occuring with high frequency often directly relates to Wikipedia pages. Within a large sample of web queries 38% matched exactly a title of an article in Wikipedia. The content and context of the matched Wikipedia page can then be used to automatically expand the query, e.g., Wikipedia pages can form an intermediate between a user query and a collection of books being searched. Researchers have used the redirects in Wikipedia to form a thesaurus for use in document retrieval systems with automated query expansion. 303, 304

Researchers also use images from Wikipedia in information retrieval systems research. 305,306 Researchers used part of the French Wikipedia for the demonstration of a combined text and image browsing system with similiarity computation. 305

Thesaurus construction

The comprehensive Rouget's International Thesaurus 307 lists both the word agency as well as $Federal\ Bureau\ of\ Investigation$. None of these two articles link to the other directly. Several groups attempt to make systems that will automatically determine the relatedness of such words and phrases based on mining data in Wikipedia or other sources, e.g., WordNet: "synonym search" or "semantic relatedness". $^{140,298,303,308-310}$

Chesley et al. used Wiktionary to expand a labeled list of adjectives. ²⁹⁶ For text sentiment analysis they needed the polarity and objectivity of adjectives. They already had a list with a limited number of words manually labeled in this respect. By lookup in Wiktionary they could assign polarity and objectivity to yet unlabeled words by examining the definition of the adjective. The approach also handled words with negated definition. ²⁹⁶ Zesch and Gurevych used the Wikipedia article and category graph in their semantic relatedness method. ¹⁴⁰

Named entity recognition

One of the basic tasks of information extraction, named entity recognition (NER), deals with identifying named entities such as personal names, names of organizations or genes from freeform text. NER often relies on a machine learning algorithm and an annotated dictionary (gazetteer). Serveral researchers have used Wikipedia for NER or more general keyword/keyphrase/"concept" extraction. 304, 310–315

French Exalead's Wikifier is a Web service that will take a document (either as a URL or a text) and serve it back as an HTML page and link terms within the document to relevant Wikipedia articles. With named entity recognition it detects people, places or organizations. CC-BY-NC-SA-licensed Accurate Online Disambiguation of Named Entities (AIDA) software for online named entity extraction with disambiguation uses Wikipedia indirectly through YAGO2. A online demo version is available

A game development system, that starts with a newspaper article, identifies country and notable person based on information in Wikipedia. The system furthermore gauges the sentiment towards the person by using posts from Twitter in sentiment analysis. 316

Translation

Wikipedia can act as a resource in machine translation and multilinguage systems by using the language link that connects different language versions of Wikipedia. Before the Wikidata project interlanguage links was part of the article page of the MediaWiki site through a special markup that usually appeared at the bottom of the page. Daniel Kinzler's WikiWord system extracts lexical and semantic information from Wikipedia to form a multilingual thesaurus. 306,317 In an extension to the system called WikiPics he use it in combination with images from Wikimedia Commons for multilingual image retrieval.

It seems obvious that the multilingual Wikidata will become a valuable and easier accessible resource for future translation studies relying of Wikimedia data. Researchers will no longer need to crawl multiple Wikipedias to establish naming variations (through redirects) and extract interwiki language links. Instead all basic language links is readily available in Wikidata in machine readable format. Arun Ganesh ('Planemad'), one of first to explore Wikidata-based translation, built a prototype with a map of India with labels for the states of India retrieved from Wikidata: The Wiki Atlas. It would allow the user to switch between, e.g., Hindi, English and Zhongwen representations of state names. xxxi In 2014 Denny Vrandečić and Google would take this approach a step further by publishing the qLabel Javascript library, enabling a website developer an easy method to provide Wikidata-based multilingual website content. xxxii

The Danish company GrammarSoft provide online translation of Wikipedias from one language to another. It has translated the Swedish Wikipedia to Danish and made it available from http://dan.wikitrans.net/ complete with hyperlinks, while an Esperanto-English version is available from http://wikitrans.net/. These system uses Constraint Grammer extended to handle the 'heavily layoutted' text of Wikipedia. 318,319

Even though one language version of a Wikipedia articles may be a translation of a Wikipedia article in another language, the wikitext is not readily sentence-aligned. Thereby a potentially valuable resource for training statistical machine translation systems is lost. Parts of the documentation for MediaWiki and some other Wikimedia resources are sentence aligned.

Ontology construction and categories

Social websites may allow users to put free-form tags on content. The relationship between the combined set of tags has been term folksonomy.³²⁰ Users tag Wikipedia articles with categories and the users can organize the categories in a hierarchy (actually a directed acyclic graph, usually, but not necessarily a hierarchy). While editors may use free-form categories, the categories that survive are the ones which have support in consensus. Jakob Voß distinguishes between collaborative tagging (e.g., used the del.icios.us website), classification (e.g., Dewey Decimal Classification) and thesaurus indexing (e.g., Medical Subject Headings, MeSH, where records/categories may have multiple broader terms), and he regards the Wikipedia category system as a collaborative thasaurus.³²¹ The categories together, with categorized articles, have been used for ontology learning in a semiautomated system. 322

Another way of using Wikipedia begins with an already established knowledgebase and then attempts to match and extract data from Wikipedia for further population of the knowledgebase. Friedlin and McDonald would start with records in the medical terminology database LOINC and with a semi-automated program attempt to find corresponding articles in Wikipedia, so the introductory Wikipedia text could be used for a helping description text for the LOINC records. Reagle and Rhue have described an application of a personal name matching method between multiple biographical works—including Wikipedia—as well as gender determination method in a study of gender bias. 88

Two Technical University of Denmark students, Magnús Sigurdsson and Søren Christian Halling, used the part of Wikipedia dealing with musical groups for the Internet-based search engine, MuZeeker and could group search results according to Wikipedia categories.³²³ A further related application ran on Nokia mobil telephones. **xxxiii*

Studies on Wikipedia also uses the Wikipedia category network as a form a ground truth to assign top categories to Wikipedia articles. ⁹⁴, ¹⁴¹ Altough the category network is directed the researcher consider the undirected network to avoid disconnected categories, though weighting the graph according to right/wrong direction help categorization. ¹⁴¹

Databasing the structured content

Several groups have extracted information from the categories and templates of Wikipedia and built

xxxihttp://4thmain.github.io/projects/hacks/wikiatlas.html.

xxxiihttp://googleknowledge.github.io/qlabel/.

xxxiii MuZeeker

databases. The YAGO system extracted data from Wikipedia and combined it with WordNet. ^{39,324} Their aim is ontology building with facts such as "Elvis" "is a" "rock singer" and "Elvis" "has the nationality" "American". They note that the semistructured data of Wikipedia is not directly useful, e.g., the category tree has "Grammy Award" as a supercategory of "Elvis" as the rock singer had won that award, but it would be wrong to infer "Elvis" "is a" "Grammy Award". This is in contrast to WordNet which has a cleaner graph. When reported in 2008 YAGO had 1.7 million entities and 15 million facts. ³⁹ The number of entities may be compared to the number of terms in Gene Ontology that I in 2013 counted to around 40'000.

The largest effort is probably DBpedia² with extraction of Wikipedia templates and the information made available at http://dbpedia.org/ BBC uses among others DBpedia for linking documents across their web site.³²⁵

With the advent of Wikidata, direct extraction of data from the categories and infoboxes of Wikipedia may still be relevant, but it seems that Wikidata will be a much stronger base.

Trend spotting and prediction

Just prior to the announcement of the choice of the United States vice presidential candidant Sarah Palin's Wikipedia article saw a high editing activity. The activity was higher than for other candidates, thus giving an indication for the choice of Palin. In another case Paula Broadwell's Wikipedia page contained the short statement "Petraeus is reportedly one of her many conquests" for several months before the secret relationship became public knowledge and led to an FBI investigation and the resignation of the CIA director. Cyber intelligence systems can use this kind of information for predicting events prior to it becoming public knowledge.

Public Web tools monitor the editing pattern and presents analyzed results based on queries: Craig Wood's *Wikirage* Web-site ranks Wikipedia articles based on edit activity measured not only as the total number of edits, but also, e.g., the number of undos and unique authors. This tool is useful to identify edit conflicts which may generates lots of edits. *WikiChecker* also generates statistics over users and individual articles as well as list highly edited pages. A similar Web-site by Dennis Yurichevxxxiv shows articles created within the last month, last week and last 24 hours ranked by the number of contributing editors. Such a Web

service puts emphasis on trending topics that have not before been in the spotlight. One example is Atifete Jahjaga rising from relative obscurity to President of Kosovo. Thomas Steiner and his coworkers present yet another web service for realtime trend spotting: Wikipedia Live Monitor monitors Wikipedia edits across 42 different language version and detect concurrent edit spikes to report breaking news, e.g., the 2013 Russian meteor event. 328 Steiner has presented a related web service, the Wikipedia Natural Disaster Monitor, focusing on the natural disaster articles on Wikipedia for crisis response. 329

Sérgio Nunes' WikiChanges creates an on-the-fly graph of the edit activity through time of one or two Wikipedia articles. One example pointed to compares 'Barack Obama' and 'Hillary Rodham Clinton' articles and shows most edits for the winner of the United States primary election. Inspired by Maximillian Laumeister's Listen to Bicoin, Stephen LaPorte and Mahmoud Hashemi built the Web service Listen to Wikipedia which converted Wikipedia recent change feed to an online dynamic visualization with sound. Circles would appear and bells would sound each time an edit was made.

The Wikimedia Foundation did for some time not make statistics on page views available. However, in December 2007 board member Domas Mituzas announced the availability of hourly statistics for each page from http://dammit.lt/wikistats/. Such data makes it easy to spot trends by simple counting. The user "Henrik" has presented a useful interactive Web service that renders the statistics in monthly histograms from http://stats.grok.se/. Showing topics of public interests — and disinterest — the user "Melancholie" pointed to an example: "Hurricane Gustav" and "Bihar flood", both natural disasters of 2008 but one in rich USA, the other in an impoverished state of India. The former reached several hundred thousand page views in August and September while the latter only a few thousands.

Based on the viewing statistics Ed Summers made the Web service Wikitrends available in 2012 from http://inkdroid.org/wikitrends/ showing the top 25 most viewed Wikipedia articles for the past two hours. Since 2008 Johan Gunnarsson has run a similar Web service, also called Wikitrends, first from the Toolserver and later on on the Wikimedia Tool Labs servers with up- and down-trends based on day, week, month or year and across many language versions of Wikipedia. Furthermore, Andrew West makes weekly reports of the 5000 most visited pages on the English Wikipedia available on one of his user subpages on Wikipedia. Weekly annotated

xxxivhttp://unit1.conus.info:8080/en.wikipedia.stats/

versions of the Top 25 articles from the West list are made available in the Wikipedia:TOP25 article. West together with user Milowent have also reported on page view spikes, see page 26.

Working from stats.grok.se June and January 2008 statistics and examing health-related topics with probable seasonal effect—such as frostbite, hypothermia, hyperthermia and sunburn-Laurent and Vickers found a clear effect in the page views:²⁴² On average the ratio was around 3 between June and January page views for these topics. They also analyzed the page view statistics of three articles describing melamin, salmonella and ricin. These examples were associated with official health alerts in 2008, and page view statistics also shows a marked increase correlating with the announcements. In 2014 a company would publish an analysis of English Wikipedia articles with health care information reporting correlation between article views and related prescriptions and unit sales of medication.³³⁰ Another interesting application would examine how well Wikipedia activity could be used to predict movie box office success.³³¹ Not only the viewing statistics was used as a feature in the prediction, but also the number of edits, the number of users editing and "collaborative rigor". Keegan did a related small longitudinal study with prediction of 2014 Academy Awards winners based on the number of edits and the number of unique editors in 2014. He predicted 3 out or 5 correctly. ³³² Yet another analysis of the utility of Wikipedia page view statistics for prediction of real-life events showed mixed results for the result of political elections in Germany, Iran and United Kingdom. 333

Like Wikipedia, the Twitter Web service provides an API for returning content in structured format.³³⁴ The Wikipedia services for trend spotting may resemble some of the many Web services for Twitter postings, and programmers have presented systems combining Wikipedia and Twitter. Tom Scott implemented a Twitter bot with Wikiscanner-like functionality: A bot monitoring Wikipedia changes would track Wikipedia edits from United Kingdom Parliament IP-adresses and output changes to a specific Twitter account (@parliamentedits). Inspired from this system Ed Summers implemented a system for the United States Parliament (@congressedits) and released his code, xxxv enabling the monitoring of several other parliaments, e.g., Sweden with @Riksdag-WikiEdit, Denmark with @FTingetWikiEdit and Canada with @gccaedits. Danish Wikipedian Ole Palnatoke Andersen would collect these Wikipedia monitoring Twitter accounts on his list

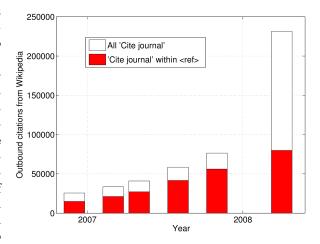


Figure 11: Growth of the number of structured journal citations. Plotted with data obtained for citation clustering.⁸¹

https://twitter.com/palnatoke/lists/wikiedit with the number of list members reaching 48 in October 2014.

Altmetrics

Altmetrics is short for alternative metrics and is the study and use of measures for scholarly impact based on activity in social media. Mendeley and CiteULike are useful services, but also Twitter, blogs and Wikipedia have been examined as sources for altmetrics, ³³⁵ and there have been a few studies on altmetrics with Wikipedia. In a survey among 71 bibliometrics researchers 33.8% stated that "mentions of or links to your work in Wikipedia" had the "potential for article or author evaluation". For the statement "article about you on Wikipedia" 26.8% confirmed its potential. ³³⁵

My analysis of outbound links to scientific journals from Wikipedia (using the cite journal template) found that the correlation with the de facto standard for scholarly impact measure was reasonable. 80,81 Bot-generated content can change the number of citations considerable. Whereas typical citations uses the ref tag for in-text citations, the Protein Box Bot would add citations to the end of the article without using the ref tag. When the bot created and amended thousands of articles in 2008 there was a large increase in structured citations using the cite journal template, but without using the ref tag, see Figure 11. Indeed the results indicate that the bot in 2008 was the creator of the majority of the journal citations in the English Wikipedia. American researcher carried out a similar analysis with the 2010 English Wikipedia dump, but instead based the citation extraction on

xxxvhttps://github.com/edsu/anon

the PubMed identifier (PMID) or the Digital Object Identifier (DOI). By analyzing the full revision history of the English Wikipedia they identified a trend for journal articles to be cited faster in later years compared to earlier years.³³⁶

The verbose journal clustering results from my journal citation data mining is available from http://neuro.compute.dtu.dk/services/wikipedia/citejournalminer.html. On the English Wikipedia itself the WikiProject Academic Journals keeps bot-updated detailed statistics of journal citations with the overview page at Wikipedia:WikiProject Academic Journals/Journals cited by Wikipedia.

Analyzing the matrix formed by intrawiki Wikipedia linking with the PageRank and CheiRank methods Toulouse and Hungarian researchers would examine the ranking of the top 100 universities over time and compare it to Shanghai university ranking. They found the PageRank of the top 10 university to increase over time, and, e.g., Harvard and Columbia at the top. 148

As researchers themselves can edit pages and insert mentions or links to their own work (as I have done) Wikipedia citations are susceptible to manipulation of the altmetrics measure. Notability criteria, deletionists and general collaboration among contributors may keep a bound on this problem. Also on the downside of using Wikipedia for altmetrics is the poor coverage of scientists on Wikipedia.⁸⁹

The Altmetrics.com website features Wikipedia as one of altmetrics sources for paper-level scientometrics. It tracks Wikipedia citations to journal articles by date and user. xxxvi

Geotagged content

Some Wikipedia articles have markup with geographical coordinates. These coordinates can be extracted and used, e.g., with rendered maps as, e.g., in Google Earth and Danish Findvej.dk. Wikidata with its structured content including geotagging can provide a plentora of information for enriching maps, e.g., OpenStreetMap user *ubahnverleih* combined Wikidata's entities with geotags and coat of arms images to render coat of arms on map, see http://osm.lyrk.de/wappen/.

The augmented reality platform Wikitude and its smartphone world browser uses location-based Wikipedia content such that Wikipedia article excerpts are overlaid on the camera image. The Wikipedia article shown is selected based on GPS and compass information in the mobile phone. Another augmented reality browser, *Layar*, embeds

Wikipedia landmarks.

Other uses

WikiTrivia developed by Alex Ksikes and Philipp Lenssen uses the introductory paragraph from Wikipedia for an online quiz. With the title words blacked out from the paragraph the task is to guess the title from the paragraph. The quiz participant chooses a category, such as politics or animals, before the game begins. The online quiz WhoKnows? also uses information from Wikipedia, but indirectly through DBpedia.³³⁷

Researchers have used the edit histories of Wikipedia articles in a study of the performance of system for sharing mutable data. 338

4 Extending Wikipedia

The original Web protocol of Tim Berners-Lee has always had the ability to both read and write on the Internet. However, in the first decade of the Web the writing cabability was little used. Some Web-based applications enabled collaborative workspaces, e.g., the BSCW system, that let users upload files.³³⁹ Another effort, WebDAV, would standardize a protocol for distributed authoring on the Web. 340 Wikis begin with the work of Ward Cunningham, with many of the early technical ideas described in the book *The Wiki Way*:²⁹ The ability of users to edit and create new pages, the easy way to make page link between pages, the free structure of wikis, the notion of a sandbox, page locking, camel-case (wikiwords), the basic edit conventions (the wiki markup) and revision control.

Researchers and developers have suggested many technical extensions for MediaWiki, the Wikipedia software: Software tools that automatically create new content or modify existing, e.g., to fight vandalism, tools that help users (readers and editors), e.g., get an overview or provide the content in another form than the usual way, and extensions with additions to the data representation and interfaces to handle structured content.

Wikipedia editing by bots of assisted editing tools has increased over time. When Stuart Geiger examined the fraction of bot and assisted editing in 2009 he found 16.33% of all edits performed by a bot and assisted editing programs used in around 12.16% of all edits. On a particular page used for vandal fighting, Wikipedia: Administrator intervention against vandalism (AIV), edits by tools are predominant with around 75% of the edits. ^{181,210}

xxxviSee, e.g., 422759 on www.altmetrics.com.

First class	Other class	Features	Performance	Ref.
Featured	Clean-up	Computed trust	82%/84%	341
Featured	Random	Word count	96%	342
_	_	30 features	98%	342
Featured	Nominated, but rejected	12 features		343

Table 8: Automated article classification. The performance column may indicate different types of performance, e.g., accuracy.

Automated quality tools

There have been several suggestions and implementations of tools for quality assurance in wikis and Wikipedia particularly, and number of tools are in operation. The tools that operate automatically on Wikipedia are usually refered to as bots and may perform mondane tasks, such as between-language Wikipedia version links (between-language Wikipedia links are now taken over by Wikidata) before the advent of Wikidata, add dates to templates, include titles from cited external pages, etc.³⁴⁴ The most important bots are probably the anti-vandalism tools (vandal fighter bots) on Wikipedia. These kinds of bots have been operating since user Tawker's Squidward began in 2006.345This one and the initial ClueBot operated with simple text pattern matching.¹⁷⁶ Christopher Breneman's and Cobi Carter's ClueBot NG, now operating on the English Wikipedia, leaves a major impact by rejecting 40,000 vandalisms a month (mid-2012 number).³⁴⁵ It monitors the IRC stream of changes and uses a trained machine learning algorithm (an artificial neural network) to classify and it will then revert the edit if it is sufficiently probable that the edit is a vandalism. Training set data for ClueBot NG's machine learning is obtained through a webbased interface at review.cluebot.cluenet.org where Wikipedians can label data.

Besides the ClueBot NG, other anti-vandalism and article quality prediction systems using machine learning or other complex algorithms have been suggested. 178, 205, 341–343, 346–351 Some of these systems do not necessarily focus on vandalism detection, but on more general quality aspects, e.g., classifying "featured" articles from other kinds of articles, see Table 8. Reported in 2006 one system used dynamic Bayesian networks and the BUGS software to compute the "trust" of English Wikipedia articles based the trust of its previous revision, the user role and the amount of insertion and deletions with these variables linked up in a Markov chain. 341 The researchers used beta distributions setting priors based on the status of the

user, regarding administrators as the most trustworthy and then in decreasing order: registered users, anonymous users and blocked users. The researchers in this particular study restricted themselves to 50 featured articles, 50 "clean-up" articles and 768 normal articles in geography for the computation of trust. It is unclear whether the method using BUGS would scale to full Wikipedia application. Besides computing trust the researchers also used the trust values to predict the article class. On a test set of 200 new articles with 48805 revisions they could report predicting 82% featured and 84% clean-up article correctly.

An anti-vandalism approach from 2008 used bagof-words and a naïve Bayes classifier as well as 'probabilistic sequency modeling', though could not report better than ClueBot results. Another approach used features such as 'edits per user' and 'size ratio' and obtained quite good results. The UC Santa Cruz WikiTrust, presented at the Web at wikitrust.soe.ucsc.edu, has an associated API available, which may report the author of a word, its "trust" and the revision where the word was inserted. This API is, e.g., used by STiki. WikiTrust extracts a number of features and uses a classifier from the Weka toolkit. The systems enables individual words in Wikipedia articles to be colored according to computed 'trust'. 349

In contrast to the complex algorithm approaches, it has been found that the word count of an article performs surprisingly well as a predictor for article quality. 342, 352 Blumenstock suggested to make a cut at 2'000 words between featured and random articles and found a error rate on around 96% on a corpus of 1'554 featured and 9'513 randomly selected articles.³⁴² The researcher of this study also tried a number of other more complex features, e.g., readability indices, citation count, category count and sentence count, and a range of machine learning algorithms, increasing the accuracy to 98%. When featured articles are compared to other selections of articles rather than random ones then qualityassociated features may change. Using an analysis of the 2007 Virginia Tech Massacre article for hypothesis generation Gerald Kane set out to examine which features predicted whether a nominated article would become a featured article or would be rejected as such. Using 12 different features from 188 Wikipedia articles and binary logistic regression analysis he found that reference count and depth of top contributor experience signaled quality while higher word count and breadth of top contributor experience as well as, e.g., percentage of anonymous contributors would make it more likely that the nominated article would be rejected as featured article.³⁴³

By considering good-quality "tokens" as those edits inserted by a user that are present after the intervention of an admin Javanmardi, Lopes and Baldi reported in 2010 about the performance of models for user reputations.²⁰⁵ They examined four different user roles: Admins, "vandals", "good users" and blocked users. Looking at the entire revision history of the English Wikipedia and utilizing a diff algorithm and MD5 checksums they were able to classify between admins and vandals with an area under the ROC curve value of around 97.5%. In one of their models the speed of the deletion affected the reputation of the user negatively and in their most elaborate model the reputation of the user that deleted the content also affected the reputation of the user inserting the content.

In 2010 a Wikipedia vandalism classification competition was held, where 9 systems competed. The competition was based on a data set with 32'452 edits on 28'468 different English Wikipedia articles annotated with crowd-sourcing through Amazon's Mechanical Turk. 355 A plentora of different features were suggested, 354 see also Figure 9 for a few. The system by Mola Velasco won with WikiTrust coming in 2nd or 3rd (depending on the measure used for evaluation). 354 Mola Velasco used a broad range of features and for the classification he used algorithms from the Weka framework.³⁵¹ However, the organizers of the competition could report (with some measures) better performance than any of the individual classifiers when the classifiers were combined into an ensemble classifier.³⁵⁴ The competition the following year would also include German and Spanish Wikipedia edits. 356 The winning entry used 65 features. Among the presently best performing algorithms are apparently a system using 66 features and Lasso for estimation.³⁵⁷

To avoid that good edits automatically get reverted (false positives) the bots usually operate with a high threshold leaving false negatives. This problem has left room for the semi-automated vandal fighter tool of Andrew West called STiki. The original system hooked on to the IRC and the

Wikipedia API. With feature extraction and machine learning applying support vector regression running on a server, edits would be scored for vandalism and high-scoring edits would be presented to a client where a Wikipedian could indicate whether the edit was a vandalism or not. ^{179,180} The Wikipedians using the system would need to download the Java client program. The newest version of STiki uses information from ClueBot NG and WikiTrust.

An different approach, Wikibu, seeks to indicate the reliability of a Wikipedia article by embedding the Wikipedia articles on a page that also shows the number of visitors, number of editors, number of links and sources of the Wikipedia article. A version for the German Wikipedia is available from http://www.wikibu.ch.

Other efforts focus on approval of article through peer review. There are already mechanism for peer review on Wikipedia itself (e.g., the notion of 'featured article'), but also combinations with external systems have been suggeted.³⁵⁸ One example was *Veropedia*, that had over 5000 checked Wikipediabased articles. In July 2010 the site was unavailable.

Back in 2006 Bertrand Meyer called for a certification of an entry and a refereeing process and believed that a rating system for authors and entries is inevitable. 167

Automatic creation of content

Andrew Lih has called the start of the masscreation of content via bots "the most controversial move in the history of Wikipedia": ²⁵ In the autumn of 2002 Derek Ramsey began to add articles on US counties to the English Wikipedia based on information extracted from public United States Census data. The automatically constructed text, corresponding to a couple of thousands of articles, he manually added to Wikipedia. ter this addition he started with automated creation of articles for 33'832 US cities completing the task in less than a week in October 2002. As the English Wikipedia had just over 50'000 articles before Ramsey began, he quickly became responsible for the creation of 40% of the articles on the English Wikipedia.²⁵ Longitudinal studies show the effect of Ramsey's work. 283, 284 Before Rambot's activity a few other bots operated on Wikipedia creating content on a more limited scale: The 'Wikipedia: History of Wikipedia bots' English Wikipedia page lists, e.g., import of a glossary of telecommunication terms in February 2002 from Federal Standard 1037C. Bots can still make a major impact on the size of a Wikipedia: Sverker Jo-

Edit feature	Description/assumption		
Edit time of day	180		
Edit day-of-week	180		
Registration time	Time since registration of editor ¹⁸⁰		
Last edit time	Time since last edit of article ¹⁸⁰		
Vandalized time	Time since editor last vandalized ¹⁸⁰		
Article reputation	180		
Editor reputation	180		
Categorical reputation	180		
Geographical reputation	180		
Comment length	Length of the revision comment ¹⁸⁰		
Article features	Description/assumption		
Number of edits/rigor	More edits are better. ^{283,284} "[] more editing cycles on an article provides for a deeper treatment of the subject or more scrutiny of the content." ¹¹¹		
Number of editors/diversity	More edits are better. ^{283,284} "With more editors, there are more voices and different points of view for a given subject" ¹¹¹		
Word count	The length of an article. Long articles are better. 342,352		
Semantic stability	Articles with semantically stable revisions are better. ³⁵³		

Table 9: Quality features of an edit or article. See also the many more features used in the Wikipedia vandalism detection competition.³⁵⁴

hansson's bot operating on Swedish, Waray-Waray and Cebuano Wikipedias has constructed 2.7 million articles xxxvii as of February 2014. The large number of articles for species has made the Swedish Wikipedia one of the largest Wikipedias.¹³¹

Scientists have also automatically added information to Wikipedia by formatting data in already existing scientific databases for inclusion in Wikipedia. In the molecular biology project, Gene Wiki, the Protein Box Bot has since August 2007 added several thousand articles on genes, 240,359 with automated construction of an infobox, freetext summary and relevant publication aggregated from Entrez Gene and a gene at las database. $^{360}\,$ In a similar project, the WikiProject RNA started out with the Rfam database containing information about RNA and added the information to Wikipedia, creating over 600 new articles. After the addition of content to Wikipedia the information flow was turned around and the (possibly extended) content of Wikipedia was used for the annotation of the Rfam database by daily monitoring Wikipedia for changes.³⁶¹ Since November 2008 the Rfambot has been working with updates and conversion of infoboxes in Wikipedia articles related to the project.

Tools may automatically create on-the-fly text from the structured information in Wikidata. Magnus Manske's Reasonator Web service xxxviii shows one of the first examples, see Figure 12 for the generated page of Johann Sebastian Bach. A related on-the-fly construction of Wikidata semantic information is provided by Metaphacts with https://wikidata.metaphacts.com.

Stuart Geiger has made a longitudinal overview of overall bot and assisted editing activity on the English Wikipedia for an around two month period in the beginning of 2009. Assisted editing amounted to approximately 12% of all edits. 181, 210 Most of these edits seem not to be related to creation of content but rather vandal fighting, but bots can have a major impact on some areas of Wikipedia. With longitudial examinination of the number of scientific citations in Wikipedia I found a big jump from 2007 to 2008 due to content with citations added by the *Protein Box Bot*. 81

The WikiOpener MediaWiki extension creates a framework for querying external databases such that their material can be merged in the wiki. 362

In May 2014 Magnus Manske introduced gamifi-

xxxviiWikipedia Statistics. Bot article creations only: http://stats.wikimedia.org/EN/BotActivityMatrix-Creates.htm

xxxviiihttp://tools.wmflabs.org/reasonator/



Figure 12: Magnus Manske's Reasonator Web service with on-the-fly generation from data in Wikidata. The text in the grey field is template text filled with data from Wikidata.

cation of Wikidata entry, letting Wikidata editors set values of properties (e.g., the gender of a human or whether an item 'is a' human) based on the introductory paragraphs of Wikipedia pages. xxxix The interface would present buttons for the user to press and the Web service would perform the actual Wikidata editing behind the scene.³⁶³ Google developers constructed another semi-automated Wikidata entry tool, Primary Sources Tool, initially used to move data from Google's Freebase to Wikidata. Via a gadget the tool would suggest new claims by modifying the Wikidata edit interface and giving the user a choice to either accept or reject the suggested claim. In January 2016 users had performed about 90,000 approval or rejection actions.364

Automatic creation of links

In a procedure what may be called automated link discovery tools suggest intrawiki links from a word in a Wikipedia article to an appropriate Wikipedia article. $^{365-367}$

At one point Wikipedia user Nickj ran the Link Suggester and Can We Link It. Another user ran a link discover bot from the account Anchor Link Bot. The DPL $bot^{\mathbf{xl}}$ does not suggest new links but notify (human) editors about links they have added, that points to disambiguation articles and may need to be resolved.

Automatic links to external Web pages can also be suggested. Kaptein, Sedyukov and Kamps reported that 45% of all Wikipedia articles have an "External link" section. They built a system that would predict these external links. By using the ClueWeb category B data consisting of 50 million English Web pages, anchor text index, document priors (length, anchor text length and URL class priors), document smoothing and Krovetz stemmer they could reach 0.68 in performance measured with the so-called *Mean Reciprocal Rank* (MRR). By furthermore using the social bookmarking site *Delicious* they could improve MRR to 0.71.

A more elaborate wiki content linking has been demonstrated with the *IntelliGenWiki* system, which incorporates natural language processing for extraction of, e.g., enzymes and organisms terms from wiki text. The extracted data is added to a Semantic Web system allowing the wiki so show embedded results of semantic queries with a Semantic MediaWiki extension.³⁶⁹ Their system has also been used to generate an index from wiki content³⁷⁰

Work suggestion

A standard installation of the MediaWiki software can generate lists for work suggestions. MediaWiki users find these links on the page 'Special:SpecialPages' and examples are 'dead-end pages', 'wanted categories' and 'wanted pages'.

The Wikidata-based service Wiki ShootMe! lists geographical items with missing images on Wikidata based on a query coordinate given by the user. A Wikipedian can use it to identify photo opportunities in his/her nearby area. The MediaWiki software embeds similar functionality with the 'Special:Nearby' page that lists nearby pages and associated images if the use shares his/her estimated browser geolocation.

Semantic wikis

The framework referred to as the 'Semantic Web' structures data.³⁷¹ Essentially, a subject-verbobject triple data structure represents data, where the triple data may be represented in RDF. In the middle of the 00s several research groups proposed systems that merged the Semantic Web idea with wiki technology^{372–376} with some of the systems based on extensions around the MediaWiki software.^{377–379} Researchers compared 10 different semantic wikis in a 2008 paper and ended up selecting the Semantic MediaWiki system for their knowledgebase.³⁸⁰

xxxixhttp://tools.wmflabs.org/wikidata-game/ xlhttps://en.wikipedia.org/wiki/User:DPL_bot

In the German Semantic MediaWiki (SMW) the wiki page determines the subject of the triple structure, while the wikilink sets the object. The verb of the triple structure comes as an element in an double colon extension to the MediaWiki link syntax, so that, e.g., [[has capital::Berlin]] on the Germany page will result in the triple structure (Germany, has capital, Berlin). In the Japanese Semantic MediaWiki³⁷⁷ the suggested syntax would be [[Term:Berlin|has capital]]. The German SMW comes with query facility, so user may formulate semantic on-the-fly queries across all the semantic data of the wiki.

Though not implemented in Wikipedia nor its sister projects, the SMW extension has proven valuable in many other contexts, and a large number of extension on top of SMW have come into existence, ²⁸ e.g., the Semantic Result Forms extension allows user to let the semantic query return the result in a large number of formats, both textoriented and in plots of different kinds. The SemanticHistory extension can enable semantic queries on the wiki edit history. ³⁸¹

The Semantic Web does not directly allow for specification of n-ary relations beyond the triplet to capture a structure such as (Germany, has capital, Bonn, in year, 1990). However, an n-ary relation can be broken down into multiple triplets. The BOWiki extends Semantic MediaWiki so n-ary relations can be represented with the wiki syntax. Extensions for Semantic MediaWiki also allows for representation of n-ary objects.

Wikidata

Within the MediaWiki framework the 2012 Wikidata proposalxli discussed the issues and presented a mockup with field/value editing. This system became operational in October 2012, and major bot activity soon made it the largest Wikimedia wiki in terms of pages (or 'items' in the parlance of Wikidata) and in terms of editing activity with about 90% of the edits made by bots. Inspired from the Semantic Web and semantic wikis the page of Wikidata acts as the subject, while the verb is referred to a 'property' and the object will refer to another Wikidata page or another type of value. Wikidata allows n-ary relations through so-called qualifiers. Limited support for units, such as heights and mass, came in September 2015.

The query language on the base installation of Wikidata was initially limited with no means to build advanced queries like the ones seen in SPARQL or Freebase's MQL. How-

ever, a few third-party services soon appeared with Magnus Manske as the primary driver behind the tools. His Wikidata Query editor at http://wdq.wmflabs.org/wdq/ allowed for formulation of queries such as "Places in the U.S. that are named after Francis of Assisi", and his autolist tool at https://tools.wmflabs.org/autolist/ would display a list of retrieved pages. Kingsley Idehen's OpenLink Software company also provided a Virtuoso-based SPARQL endpoint for querying Wikidata together with other Semantic Web data from the *Linked Open Data* cloud. xlii Metaphacts and University of Chile also made early public SPARQL endpoints with Wikidata data available.³⁸⁶ Finally, in September 2015 Wikimedia Foundation announced its SPARQL endpoint, the Wikidata Query Service, running from http://query.wikidata.org. "List of countries ordered by the number of their cities with female mayor" was an example of possible queries.

With its qualifiers and references the Wikidata data structure does not map directly to the triple structure of the Semantic Web and SPARQL tools. However, various approaches—reifications—can convert Wikidata data to the triple format. Benchmark profiling with four different approaches and five different SPARQL engines found that the so-called *n*-ary relation model enabled the expression of property paths.³⁸⁷

When Wikidata is coupled with natural language question parsing powerful open-domain question-answering (QA) systems can be constructed. Students from École Normale Supérieure de Lyon constructed one of the first online QA systems with Wikidata data, the Platypus system running from http://askplatyp.us.. It was capable of answering questions such as "Who is the prime minister of France?" and "Who was member of the Beatles?".

Other structured content

Several systems exist that extending wikis with table and database-like functionality beyond the Semantic Web framework, e.g., DynaTable, shiii that defines structured data in a separate name space and can display the structured data in tables on wiki pages with the use of special tags. Other similar extensions with varying degree of maturity are WikiDB and DataTable extensions. The relatively simple TemplateTable extension extracts keys and values from templates used across the wiki and generates a table with columns according to key and each row corresponding to a template instantiation. No filtering is possible. The

xli http://www.mediawiki.org/wiki/Wikidata

^{*}liihttp://lod.openlinksw.com/sparql.

xliiihttp://sourceforge.net/projects/wikidynatable/

Brede Wiki keeps its MediaWiki template data in a sufficiently simple format for complete databasing the template content by extraction from XML dumps, enabling complex SQL data mining queries to be made off-wiki. ³⁹⁰ The Brede Wiki also keeps table-like data in simple comma-separated values format on wikipages for meta-analysis. ³⁹¹ The Simple Table extension formats the CSV data.

None of these systems enjoy as large a success as the semantic wikis and Wikidata.

Form-based editing

Standard wikis have only one input field for the content of the Wiki. Semantic Web-oriented wiki engines, such as OntoWiki, present edit interfaces with one input field for each property of a page, thus each pages will typically have many forms. The Swiki system can associate forms to wiki pages, ³⁹² and the template mechanism in the CURE wiki engine allows the editor to define the interface for editing a multiple input form as well as the presentation of the content. ³⁹³

An extension for the MediaWiki software brings support for editor-defined forms, used in conjunction with Semantic MediaWiki and the template functionality. As with Semantic MediaWiki the form extension is not enabled in Wikipedia, but is used on a number of SMW-enabled wikis ensuring a consistent input of data.

With Wikidata Wikipedians got the opportunity to make form-based edits on language links and structured content as usually found in infoboxes. Wikidata form-based input helps the user by suggestions for autocompletion.

Markup

Wikis have a simple markup language. Unfortunately the markup have diverged slightly so the markup is not entirely compatible between wiki engines, e.g., MediaWiki does not support CamelCase linking. WikiCreole (or just Creole) attempts to standardize markup and some wikis have adopted it, 394 see http://www.wikicreole.org/wiki/Creole1.0. The WikiCreole language has also been formalized in ENBF. 395, 396

The difficulties with representing MediaWiki wikitext so standard context-free grammars may parse it means that creating tools for WYSIWYG editing is difficult. Software developers within the Wikimedia Foundation project Parsoid implement a node.js program that may represent MediaWiki wikitext as HTML5, and convert the HTML5 back

to wikitext. The project should not only make enable good visual editing in MediaWiki-based wikis, but also increase rendering performance by avoiding conversion overhead. 397

Text-oriented formatting languages, such as IATEX and HTML, can have fairly dense markup obscuring the text ifself. Documentation languages focus on readability while maintaining formatting capabilities, e.g., AsciiDoc, Markdown and reStructuredText. These language typically differ from wiki markup, though, e.g., GitHub Flavored Markdown has URL autolinking like typical wiki markup.

Extended authoring

The interface and the wiki markup may not appeal to potential contributors with low level of technical ability. If they are knowledgeable contributors, Wikipedia have lost the benefit of their knowledge. In 2008 the Wikimedia Foundation received a grant from the U.S.-based Stanton Foundation for a project to make the editing facility easier.
xliv Years later in 2013 the Wikimedia Foundation rolled out the "WYSIWYG-like" editor VisualEditor on Wikipedia giving contributors the choice between usually raw wiki markup and WYSIWYG-like editing. Editing defaulting to VisualEditor generated controversy among Wikipedians, and English Wikipedia administrators manage to revert the decision making the VisualEditor an opt-in option. $^{\rm 398}$

Som wikifarm companies, such as wetpaint and PBwiki, have WYSIWYG functionality. MediaWiki-based Wikia wikifarm has WYSIWYG support. On Wikipedia user Cacycle's wikEd Javascript in-browser text editor adds enhanced text processing.

The Firefox web-browser plugin Zotero formats bibliographic references in the style of Wikipedia citation templates. ProveIt from Georgia Tech (http://proveit.cc.gatech.edu/) provides a structured interface for editing references in MediaWikis via a popup window.

Digitizing publications

Wikimedia Foundation project Wikisource collects electronic versions of free content publications. Project Gutenberg (http://www.gutenberg.org/) started by Michael Hart has a longer history going back to 1971 and had by 1992 digitized many works. 399 Lars Aronsson's Project Runeberg is a Scandinavian effort inspired by Project Gutenberg.

xlivWikipedia to become more user-friendly for new volunteer writers

To solve conflicts about differences between editions Aronsson started in 1998 to incorporate facsimiles with optical character recognition (OCR) and collaborative proof-reading. Aronsson's Runeberg approach has now been adopted by Wikisource. 400 MediaWiki still forms the basis for Wikisource, but enables the *Proofread Page* MediaWiki extension and has a 'Page' namespace where a facsimile page is presented next to the wiki edit field containing OCR'ed text from PDF or djvu files for collaborative proof-reading. The approach splits the content and presentation on different wikipages and a full book may be split across a number of pages, so tools using Wikisource as a corpus may need to implement a not straightforward procedure for assembling of the actual texts and discarding of wiki markup.⁴⁰¹

Facsimiles of paper encyclopedias may have multiple articles on one page or one article spanning multiple pages, so facsimiles present two structures: the page-oriented structure and the chapter/article-oriented structure. In Wikisource and Runeberg proof-reading is page-oriented while the presentation of proof-read books is chapter-oriented in Wikisource.

Geographical extension

Wikipedia lets users markup a page with geographical coordinates via the use of MediaWiki templates, — a simple approach to so-called *volunteered geographic information* (VGI). Other wikioriented systems present a map that users draw on while also providing descriptions of the geographical coordinates. ⁴⁰² One such system is WikiMapia (http://wikimapia.org) where users can annotate with retangles and polygons on a background of satellite images from Google Maps. The CC-BY-NC-SA licensed data are exposed through an API.

One of the most (if not the most) prominent VGI efforts is OpenStreetMap (OSM), www.openstreetmap.org, that also provides complete maps under share-alike license. Founded by Steve Coast in July 2004, users may add, e.g., roads and nodes with key-value attributes with the help of a background satellit image and GPS tracks. Rendering may be through the Mapnik open source library. OSM content may be integrated with Wikipedia, so articles with geographical coordinates display OSM maps, e.g., through the OpenStreetMap SimpleMap MediaWiki extension.

Another geographically oriented wiki is Wiki-Crimes (http://www.wikicrimes.org) that seeks to map crimes collaboratively.

Extending browsing

A number of tools exists for reformating Wikipedia content for specialized devices. Jesse David Hollington has reviewed applications for the iPhone, including the Wikipanion with features such as cache and offline reading. xlv

Wikipedia Diver is a Firefox plugin. It logs clicks between Wikipedia pages to a database and displays the browsing path graphically. The Wikipedia Beautifier Chrome extension presents a cleaner view of the text with automatic hyphenation and various adjustments to the Wikipedia interface. Google Firefox add-on shows relevant Wikipedia articles next to Google search results. In Mikipedia articles of other browser plugins changes or enhance the presentation of Wikipedia in the browser.

Graphic extensions

The image-oriented browsing program Indy-wikixlviii enables a user to search Wikipedia independently of an ordinary Web browser. The program displays the ten most related images prominantly in its browser interface while also displaying the text and links of the article. A similar system is the *Qwiki* website that display images from Wikipedia and other sources in a multimedia environment.

The Java program WikiStory contructs interactive time lines based on Wikipedia material. The Web application History Viz displays events related to a queried person on a timeline. Apart from this visualization the system also features a Java applet graph visualization of Wikipedia pages. 404 The system relied on algorithms for categorization of Wikipedia articles into persons, places or organizations. 405 Yet another timeline visualizer is Navino Evans' histropedia served from histropedia.com. 406 Timeline visualization features also as part of the Reasonator Wikidata presentation Web service.

Wikis can be extended with on-the-fly creation of graphs, 407 and there exists an extension for Media-Wiki that use the GraphViz software, 408, 409 and, e.g., the Brede Wiki uses the tool for automatic creation of visualizations of intrawiki links between brain regions, topics and organizations from information defined in MediaWiki templates. The intrawiki links between Wikipedia articles form a directed graph. Together with a path finding algo-

xlviPhone Gems: Wikipedia Apps.
xlvihttps://github.com/scotchi/wikipediabeautifier/wiki/Wikipedia-Beautifier.
xlviihttps://addons.mozilla.org/enUS/firefox/addon/googlepedia/.
xlviiihttp://indywiki.sourceforge.net/

rithm Erika Arnold's wikiGraph Web service can find and visualize a path between to user-specified Wikipedia articles using the dump and the Neo4j graph database.xlix

The 'Copernicus' system makes an attempt on a 3D wiki: 410-412 A two-layer interface presents the Wikipedia article in a transparent foreground, while the background presents a 3D model related to the Wikipedia article. The user can trigger predefined camera movements and adjust the transparency.

Video extensions

Multimedia authors have added animated images and small video clips in the Ogg Theora format in Wikipedia. Other editors can in principle edit these clips with external video editing software, but with far less ease of use as when editing wikitext. The company Kaltura have worked on a Open Source technology to better support peer production of video content. In the beginning of 2008 the company announced experiments with this collaborative video platform together with the Wikimedia Foundation. As with other material Wikipedia already enables collaborative tagging and discussion of videos through category markup and discussion pages on Wikimedia Commons.

Real-time editing

When two editors edit the same time on a Media-Wiki article and each one saves his/her edits then the editor with the last save will experience an edit conflict that needs to be resolved manually. In this aspect Media-Wiki differs not from standard revision control systems used in software development. Several systems enable real-time browser-based collaborative editing, and extended Etherpad will allow for some wiki functionality. Igor Kofman's Hackpad script for live-editing Media-Wiki pages requires only editing of a user page for installation.

Distributed and disconnected Wikipedia

Wikipedia can be downloaded and read from the local copy. Groups of editors have checked a selected number of articles and distributed them on DVDs, see www.wikipediaondvd.com and Schoolswikipedia.org. The offline multimedia reader *Kiwix* makes Wikipedia available offline. The simple small AAA battery-powered WikiReader comes with

touch screen and 3 million topics from Wikipedia. An even simpler device, Humane Reader, connects to TV and could include an offline version version of Wikipedia. The One Laptop Per Child (OLPC) project features the *Wikibrowser* with a selection from the Spanish and English Wikipedia, while the XOWA Java tool enables the setup of local offline copies of Wikipedias and their sister projects as well as their images. The setup of local offline copies of Wikipedias and their sister projects as well as their images.

Offline Wikipedia deployment may be important in areas with no Internet, and researchers have studied the use of Wikipedia in offline OLPC in this context. In a field study Ethiopia students indicated that "browsing schools books and offline Wikipedia page" was the second most favorited activity on the OLPC. 416

Versioning in Wikipedia is centralized as in the client-server software revision control systems Concurrent Versions System (CVS) and Subversion (SVN). Newer software revision control systems, such as Git, may be distributed. There have been several suggestions and developments for distributed wikis that merged the wiki idea with the distributed revision control system idea. Levitation would convert a Wikipedia database dump to Git repositories. In Ward Cunningham's Smallest Federated Wiki a user will fork a wiki when he edits. The owner of the original edit may then merge that edit back to the original wiki. 417 Some of his thoughts on federating wikis go all the way back to a 1997 memorandom.⁴¹⁸ Another distributed wiki system is research prototype Concerto (http://concerto.xwiki.com), that was built for the XWiki wiki engine as a peer-to-peer replicated wiki with an aim for data consistency between the replicated wikis. 419

Distributed wikis may store content that diverges from neutral point of view: What has been called "every point of view". 420 The question on whether distributed wikis (Wikipedias) might also help for an inclusionist approach to specialized knowledge, have also been posed. 421

Wiki and programming

There are different ways to combine programming and wikis. The extension framework of MediaWiki allows developer to make additional programs that use resources either on the wiki server or on remote servers. Other approaches enables end-user programming for editors, e.g., one system allows editors to write database queries for a SAP system. 422 A MediaWiki extension integrates output from the R statistical computing environment.

Wikis that allow for inclusion of different applications, e.g., calendars, may be termed 'application

xlixhttp://wikigraph.erikaarnold.com/

 $^{{}^{1}}http://hackpad.posterous.com/live-editing-mediawiki-with-hackpad}$

wikis'.

The MediaWiki software has a simple 'programming language' for templates, which features, e.g., an 'if' sentence. However, the capabilities are limited and the syntax quite obscure with excessive number of curly braces. In 2012 Wikimedia began serious experiments with enabling the Lua programming language on test MediaWiki installations. A successful application would fundamentally change the way that MediaWiki users write templates and make way for more advanced programming. In the first part of 2013 Lua was enabled on the Wikipedias.

A programming language embedded in a wiki could provide means to format data, e.g., commaseparated values into tables, and could be used to make small scale numerical computations, —functions that now needs to be programmed either in a MediaWiki extension or in an external web script.³⁹¹

The PyPedia website www.pypedia.com combines MediaWiki and the Python programming language. With each wikipage containing a Python function or a Python class PyPedia allows Internet users to write and execute code in the browser or execute the code on the wiki in a local installation of Python. 423

5 Using Wikipedia and other wikis in research and education

As a Web-based collaborative environment Wikipedia and other wikis offer researchers and students on all levels a means for communicating, — reading and writing about the topic at interest, e.g., wikis can offer Ph.D. Students an environment for literature research where a wiki page contains information about a specific research paper: its summary, research questions, methods, results and comments. ³⁹³

In Ten simple rules for editing Wikipedia Logan et al. guide scientists how to write on Wikipedia. 424 For example they suggest scientists to register an account for privacy, security and reputation building as well as to gain access to the "watchlist" feature. They suggest that scientist should "avoid shameless self-promotion" by not writing their biography page on Wikipedia (let others do that).

Attitudes towards Wikipedia

A Wikimedia Foundation survey has found researchers to be generally quit positive towards

Wikipedia: Over 90% of 1743 self-selected respondents were 'very favorable' or 'somewhat favorable'. Among Public Library of Science (PLoS) authors the result was 96%. Other results showed that 68% answered 'yes, on a large scale' to the question 'would you be in favor of effects to invite scientist to add or improve Wikipedia articles'. Such results are very positive for Wikipedia, but may be biased due to the self-selection of respondents and that the publisher web site with initial reference to the survey was Open Access.

In 2010 Dooley surveyed 105 university faculty members and found that 54.4% would rank Wikipedia to be moderately to very credible, 26.6% would said it had some credibility, although not much and 20% that it had "no credibility". 426

Individual researchers have called out for collegues to contribute to Wikipedia, e.g., "Scientists who receive public or charitable funding should [...] seize the opportunity to make sure that Wikipedia articles are understandable, scientifically accurate, well sourced and up-to-date." ⁴²⁷ In 2011 Alexander Bond called out for ornithologists to appropriate Wikipedia "as a teaching and outreach tool", e.g., "[P]rofessors can replace essays and reports assigned to students with the creation or improvement of a taxonomic Wikipedia entry" and recommend the WikiProject Birds page as a starting for ornithologists. 428 Writing for students and lawyers that would read Wikipedia, Diane Murley concluded that "Wikipedia can be a great research tool if used appropriately" and that "it is a good tool to use to teach researchers about the necessity of evaluating sources". 120

Some scientist contributing to Wikipedia have voiced concern about the priorities of Wikipedia. Antony Williams asked in a blog post in 2011 "Why are pornstars more notable than scientists on Wikipedia?" noting that scientist biographies he had written had been flagged for notability, even though one of the biographed scientists had over 300 publication and an h-index of 27. Williams' own Conflict of Interest in the case had been questioned. Direct Wikipedia contributions from scientists, as well as other experts, are not automatically regarded as appropriate. Wikipedia is not interested in what experts know, but rather from where the experts know their knowledge, and that the source has "institutional approval".

Rosenzweig argued that professional historians can even learn from the open and democratic production and distribution model of Wikipedia, and he asked why scholarly journals are behind paywalls and sponsored *American National Biography Online* is available only to libraries at a high cost.²⁰ He thought that historians should join in writing

history in Wikipedia, but also noted issues with the ban on no original research and "difficult people" that drive experts away as hinderances for professional historians writing on Wikipedia. He called for systems for digitizing historial documents and for open-source textbooks, — puzzlingly not mentioning the Wikisource (with oldest mainpage edits in November 2003) and Wikibooks (with oldest mainpage edits in May 2005) efforts.

Rosenzweig Echoing $_{
m in}$ another domain, Wikipedian and healthcare professional James Heilman encouraged healthcare professionals to get involved in Wikipedia stating among his reasons that "Wikipedia has taught me critical reading, which has made me better equipped to deal with less reliable sources of information such as pharmaceutical representatives." ⁴²⁹ The positive attitude towards Wikipedia in the healthcare domain has led to a closer collaboration between Wikipedians and members of the Cochrane Collaboration for strengthening the dissemination of sythesized ${\rm research.}^{430}$

Use of Wikipedia

Several studies have used surveys to examine the use of Wikipedia among university faculty⁴²⁶ and students.^{93,254,431,432} These studies may aim to answer how much and for what purpose they use Wikipedia. Apart from the direct academic and cognitive purposes of obtaining knowledge and satisfying curiosity other motivations for using Wikipedia may be hypothesized:⁴³¹ emotional, personal integrity, social and tension-release needs.

The Wedemeyer study asked students to evaluate Wikipedia biochemistry articles. One third responded that they never use Wikipedia. Among the remaining two thirds 12% used Wikipedia as their primary source and 31% used their textbook and Wikipedia equally. The remaining 57% used Wikipedia only as a supplement. The majority of the students preferred Wikipedia to the textbook. 93

Sook Lim performed a survey among undergraduate students, where 134 out of 409 responded. Page 134 and 2010 she found that all students had used Wikipedia, but that the use was not necessarily for academic purposes alone as students would also use Wikipedia for "entertainment and idle reading" and nonacademic personal information interests. That was especially the case for the male students that also had a higher perception of information quality and belief in the Wikipedia project. Another study reported in 2011 found that 47% of 186 medical students who recently had completed psychiatric clinical clerkship used Wikipedia as one of the pri-

mary sources for preparing for psychiatry exams. Question books (88%) and the peer-reviewed website Up-to-Date (59%) were more frequently used, but textbooks (10%) less used. Among the students using Wikipedia 84% also used question books.⁴³²

In Dooley's 2010 study among university faculty 45 of 105 respondents said they used Wikipedia in their teaching and/or research, 40 occationally, and 20 respondents said they never used Wikipedia for teaching/research. 426

In a paper discussing how students could use Wikipedia Jeff Meahre suggests that students peruse the discussion pages to "see the process of knowledge creation" and show why citations are important. 433

In a study on high school students' writing assignment for on a wiki the researchers found that they would use Wikipedia even though they were aware of its limitation as a source. 434

Citing Wikipedia

Many users would say that Wikipedia works well for background reading and doubt that you can use Wikipedia as a source. In the beginning of 2007 a department at the Middlebury College would hold students responsible for using Wikipedia as a source after a batch of students had used erroneous information on Wikipedia about topics in the history of Japan (Shimabara Rebellion and Ogyu Sorai). A35, A36 Media reports implied that the department of Neil Waters, the teacher of the class, 'was at war with Wikipedia itself'. However, Waters himself actually told students 'that Wikipedia is a fine place to search for a paper topic or begin the research process'. The policy adopted by the department was:

- "Students are responsible for the accuracy of information they provide, and they cannot point to Wikipedia or any similar source that may appear in the future to escape the consequences of errors,
- 2. Wikipedia is not an acceptable citation, even though it may lead one to a citable source."

At Lycoming College a teacher in history would outright ban Wikipedia from student's bibliographies giving a grade of zero if any such citation appeared. 437

Waters' policy is in line with the opinion of the Wikimedia Foundation. However, in the end of 2007 Jimmy Wales said that he saw no problem in younger students using Wikipedia as a reference, and that it should be used as a stepping stone to other sources. 438

It is fairly easy for a teacher to point to correct information on Wikipedia by the use of permanent links to a specific version of an article. Teacher may link to the version s/he has reviewed, so that students may feel comfortable in using the information on Wikipedia and use it as a supplement to textbooks.

Citing Wikipedia will be a problem if administrators delete the page, e.g., due to notability. Both the present and the older versions of the article will be inaccessible to the editor and reader.

Regardless of the problems with citing Wikipedia research papers do indeed cite Wikipedia. 426, 439-441 Dooley examined 250 research reports published in 2009 and in the beginning of 2010 that returned on a query on Wikipedia from a search on Academic OneFile electronic database: 249 of the papers used Wikipedia as a source of scholarly information. In 27 of the papers Wikipedia was the main topic and 62 had brief mentions of Wikipedia. 426 In a similar citation study Brazzeal would find 370 chemistry article from 2005 to 2009 citing Wikipedia.⁴⁴¹ The author had searched among 312 chemistry journals from the websites of the three publishers Elsevier, Springer and the American Chemical Society. Of these 312 journals 147 had at least one citation to Wikipedia. 9% percent of the articles had a citation to numerical values of physical and chemical properties in Wikipedia. In a 2008 study in the area of humanities Lisa Spiro retrieved 167 results between 2002 and 2008 citing Wikipedia, while a corresponding number for Encyclopædia Britannica was 152. Well over half of the Wikipedia citation was "straight citations" without commentary about it. 439 Using the ISI Web of Science service Noruzi would find 263 citations to Wikipedia over a 6 year period, — far more than a corresponding 10 citation to Encyclopædia Britannica. 440 The English, Frensh and Chinese Wikipedias themselves maintain (presumably incomplete) lists of research papers citing Wikipedia as a source: Wikipedia:Wikipedia as an academic source.

The number of citations to Wikipedia from research papers is not large and perhaps authors check information in Wikipedia before citing it only using Wikipedia as a proxy, so one may ask: Has any research paper actually cited information that was wrong, as, e.g., the case in news media with the hoax for the Maurice Jarre biography?

Publishing research on wikis and Wikipedia

The "no original research" (NOR) policy on Wikipedia means that researchers cannot di-

rectly use Wikipedia as primary venue to publish their findings: In "Ten Simple Rules for Editing Wikipedia" Wikipedian scientists would write: "Perhaps most important for scientists is the appreciation that Wikipedia is not a publisher of original thought or research", 424 or as historian Roy Rosenzweig expresses it NOR "means that you cannot offer a startling new interpretation of Warren Harding based on newly uncovered sources." Researchers are of course welcomed, but should cite the relevant work with pointers to institutionally approved sources, e.g., scientific journal articles.

From a scientist point-of-view the problem with writing in Wikipedia is that of "academic reward". Writing a Wikipedia article has not traditionally given the same amount of academic reward as writing a peer-reviewed scientific article.

For better integration of Wikipedia and scientific writing—both handling the problems of NOR and academic reward—journals have come up with the scheme of joint publication in a scientific journal and Wikipedia. In 2008 RNA Biology became probably the first journal to venture into this approach, requiring authors, that submitted a paper to a section in the journal, to also submit a Wikipedia page summarizing the work. The first article became ASurvey of Nematode SmY RNAs. 442 The journal article would undergo full peer-review and thus act as a source for the Wikipedia content. Another case came in 2012 where the scientific journal *PLoS* Computational Biology allowed researcher to submit educational "topic articles" that were meant to be both published in the journal as well as included in Wikipedia. 443 The first such article was "Circular Permutations in Proteins". $^{444}\,$

It is interesting to see whether Wikipedia articles can compete (and even outcompete) the traditional scientific review article.

Special science wikis

Wikipedia has inspired numerous other science wikis. Bioinformatics alone has already quite a large number of wikis. Several biomedical researcher have suggested dedicated wikis for their field since around 2006. 390, 391, 445–449

The reason why scientists choose to operate their own wiki instead of adding information to Wikipedia is probably due to one or more of four issues:

- The control and ownership of data (an issue particularly important for companies),
- the addition of specialized interfaces and data,

- the problem of notability in Wikipedia, i.e., certain parts of the research information is viewed to be too specialized for a general encyclopedia,
- the control of authorship

Several wikis are up and running handling research information: A Leipzig group extended the MediaWiki software incorporating semantic wiki functionality for gene function (http://bowiki.net). 383,450 Other groups have created wikis for laboratory protokols (OpenWet-Ware), 451, 452 single nucleotide polymorphisms (the SNPedia)⁴⁵³ and biological pathways (WikiPathways).454 Other Internet-based biomedical databases have added community annotation, e.g., $\mathrm{CutDB.}^{455}$ CHD wiki for congenital heart defects and YTPdb for classification of the veast membrane transporter use the WikiOpener MediaWiki extension to include material from bioinformatics Web databases in the wiki.³⁶² Other bioinformatics wikis are ArrayWiki, 456 BioData Mining, EcoliWiki, Open Notebook Science, PDBWiki, 457 PiQSi, Proteopedia, 458 TOPSAN, WikiGenes, 459 and WikiProteins. 446, 460, 461 A large share of the scientific wikis uses the MediaWiki software, and often with extensions for handling structured data, e.g., OmegaWiki for WikiProteins⁴⁶⁰ or Semantic MediaWiki for SNPedia.⁴⁵³

The Rfam database for RNA families is an example of a combination of scientific expert curated database and wiki. The summary information of Rfam was moved to Wikipedia and now the free-form summary information in Rfam is included from Wikipedia with the result that "[a]ll of the family annotations are now greatly improved". ³⁶¹ The Rfam database edit button leads to the Wikipedia edit page. Rfam-specialized data in Rfam "remain curated by Rfam's specialized data in Rfam "remain curated by Rfam's specialist annotators". With the added information in Wikipedia around 15% of web traffic to the Rfam database was in 2010 driven by Wikipedia. ¹⁷³

The problem of notability was the reason I started the Brede Wiki. Wiki pages on this wiki may each describe an individual research articles and usually such articles will not be deemed notable enough for Wikipedia as a general encyclopedia. A specific idea of Brede Wiki was also to maintain structured information in a machine readable format for data mining. 390, 391, 449 Related wikis summarizing academic work are Acawiki, WikiPapers⁴⁶² and WikiLit, 14 all using Semantic Media-Wiki.

Scholarpedia is a MediaWiki-based science wiki especially strong on computational neuroscience

and related areas and has strong control over authorship. Operated with Eugene M. Izhikevich as Editor-in-Chief it lets invited experts write specific articles and let other experts review the article. He has had the ability to attract Nobel prize winners as authors. At least one of the articles are archieved in the PubMed Central and reference in PubMed under Scholarpedia J..463 With such strong authorship control Scholarpedia has been low on hyperlinks and seems to have little collaboration, thus it lacks some of the fundamental characteristics of a wiki. In the end of year 2011 Izhikevich changed the editing system so that "any user can nominate himself to write an article on any topic of his expertise" although the user had to gain "sponsorship" from existing Scholarpedia curators. li

Other wikis also control the authorship: Ganfyd and AskDrWiki limit the editing to registered medical doctors, ⁴⁵² and the CC-BY-SA MediaWikibased Mediapedia similarly controls the authorship by requiring editors to be either approved physicians or scientist with doctoral-degree in biomedical specialities. ⁴⁶⁴ Finally the general encyclopedia Citizendium does not allow anonymity.

Censorship

In 2008 Bryan Derksen uploaded images of Rorschach test inkblots to the Commons, and in June 2009 James Heilman added the complete set of images to the Rorschach test article on the English Wikipedia. Along with the images came most common response to the each inkblots. Psychologists administring the test angrily held that publishing such information on Wikipedia would jeopardize this psychological test, and American Psychological Association executive director of science Steve J. Breckler said "We wouldn't be in favor of putting the plates out where anyone can get hold of them". 465

Some amount of censoring may occure for certain 'dangerous' information, such as the manufactoring of acetone peroxide or details of Zippe-type centrifuge and gas centrifuge.

Personality rights may vary between countries, weighing differently the border between free speech and the right to private life. The Wikimedia blog reported that a German university professor brought suit against the Wikimedia Foundation. He felt that the German Wikipedia violated his right of personality, particularly with respect the mentioning of his membership of student associations. The report did not mention why the professor felt this way. To me the membership seems

 $^{^{\}rm li}{\rm Scholar pedia}$ change announced on the ${\it Connectionist}$ mailing list.

entirely innocuous. However, the mentioning of it could be disputed. Denmark treat personal information, e.g., concerning race, ethnicity, religious, philosophical or union affiliations as well as health, sex and gender in a special way. Data cannot usually be registered and distributed without consent, and the mentioning of his membership could expose his religious affiliation. However, in the case with German professor the Wikimedia blog stated that "the information at issue was both accurate and freely available on several other websites under Asche's authorization", and the Wikimedia Foundation won the case. 466

Scientists on Wikipedia

Elvebakk in her analysis of online philosophy resources speculated that philosophers have added Wikipedia entries for themselves as a form of self-promotion. 83

Carl Hewitt, a retired professor from MIT, came into the spotlight at the end of 2007 after edits on the Wikipedia. He has a long career within programming, but though he has made significant contribution within the area, several other Wikipedians have regarded his edits as self-promotion. Among others, he have added description of his research (the so-called actor model) in physics articles, where others thought they did not belong The disagreement led to long discussions and a case for the Arbitration Committee, which decided to ban him.

The case shows an example on how difficult it may be to write from a neutral point of view if there is a conflict of interest and one is deeply engaged in a field of research. However, the guidelines of Wikipedia allow specialists to add their own research, if has background in a reliable publication and follows other rules set by Wikipedia. lii

I have myself added a few references to articles I have authored. Universities expect researchers to make their work more widely known, and extending Wikipedia is one way to spread research both fellow researchers as well as ordinary information seekers.

Researchers may be quite attentive to what is written on Wikipedia about themselves: In a 2011 Nature poll, recruiting 840 researchers through email and social web sites, 19% responded that they once per week or more often check Wikipedia for citations to themselves or their own work. The corresponding number for citation-counting sites (such as ISI Web of Science) and search engines were 30% and 38%, respectively. The same poll reported that

9% of the researchers had edited or inserted a reference to their work on Wikipedia within the last 12 months and around 3% has edited their own Wikipedia biography. 468,469

Wikipedia and wikis as a teaching tool

Although educators generally warn against using Wikipedia as a authoritive resource, some promote Wikipedia as an example of a digital environment with rich opportunities for collaborative writing and for considering "new roles and relationships for writers". 470 In a widespread Wikipedia engagement teachers give writing assignment to students in Wikipedia editing, 471 e.g., at the University of East Anglia students of International Relation should edit and write article about the Middleeast. 472 They were rated based on their improvements of eight articles and should write one of their one. The often very controversial topic in this area demanded that the students were able to balance the different opinions and support with sources. Similar assignments were given in history in Lycoming College, 437 "Exploring the Past" course at the University of Baltimore⁴⁷³ and environmental history at the University of Washington-Bothell. 474 Some teachers chooses to use dedicated wikis, such as CoWeb (a technology that goes back to 1998)^{211,212} or an installation of MediaWiki.⁴³⁴

Apart from Wikipedia writing assignments, teachers may also give assignments in Wikipedia article review, adding references to an already existing article, translating or copy editing. 471, 475, 476 Yet another exercise has students monitoring the changes made to the article they have contributed to and post their findings. 437

Students benefit from the Wikipedia writing assignments by having their work exposed not only to the teacher but to the entire world.⁴⁷¹ Students also mention gaining skill of doing thorough research as a benefit.⁴⁷⁵ Stronger research base is also mentioned by teachers. 437 In wikis, interaction between peer students may help to improve their writing and in a open wiki the students may become aware of the audience 212 and contemplate on the value of authority.⁴³⁷ Wikis may also make the writing process more visible though students revise extensively offline.²¹² Furthermore, students using wikis may get a better attitude towards collaboration and higher grades.²¹¹ Teachers mentions other advantages such as student improvements in computer literacy, Internet resource critique and collaborative work preparation. 476 Student reaction to the assignment may initially be worry and anxiety and later reactions may be irritation, pride and indignation.⁴³⁷

 $^{^{\}rm lii} \rm See}$ "Wikipedia: Conflict of interest" (388727443) and "Wikipedia: No original research" (388842869).

In a graduate seminar on plant–animal interactions participants assessed the quality and content of ecology content on Wikipedia. They found Wikipedia generally limited in depth and breadth and with too few citations. Then they proceeded to edit Wikipedia in their domain and found the process "straightforward and efficient, particularly once we learned the protocol for proposing and implementing changes". ⁴⁷⁷

Apart from having the students learn the topic and the teacher having a means to evaluate the students, Wikipedia assignments may also help to construct learning material when no good textbook fully covers the area. 476

For smaller Wikipedias the work of the students on Wikipedia may not catch sufficient attention from other Wikipedians. In a small study on the Danish Wikipedia historian Bekker-Nielsen monitored 17 Wikipedia articles in 380 days after his students had created the initial articles, e.g., for Demetrios Poliorketes and Carausius. The highest number of edits an article received in the 380 days was 26. Of the total number of edits only 6% were content-related. Bekker-Nielsen also pointed to one worsening edit by a Wikipedian as well as wrong references in one article created by the student and not removed by a Wikipedian before Bekker-Nielsen's study was made public. 107

Among the problems faced in Wikipedia assignments are, e.g., student username being censored and students banned because of repeated insertion of copyrighted material. 437 If students work on controversial and highly visible topics the article may suffer vandalism and it may be difficult to "disentangle" the contribution of the student. Combination of little notability and poor writing may result in the student article gets deleted. In a writing assignment on history a teacher found that the vast majority of articles created by students were deleted. 473 Modern educational software systems often use automated plagiarizing detection software. If the student submit directly to Wikipedia then the teacher may have difficulty in evaluating the extend (if any) of plagiarism, and as Wikipedia articles get copied to numerous sites (and possible into the plagiarizing detection software) it may quickly become difficult to track down the origin of a text. Teachers can get around the plagiarizing detection problem by having the students submit their Wikipedia assignment to the teacher before any Wikipedia edits are made. 437

Descriptions of courses that used Wikipedia as a teaching tool appear on the page *Wikipedia:School and university projects.*⁴⁷¹ In 2011 Klaus Wannemacher would review these projects. He found 132 projects on the English Wikipedia and

a number of others on other language versions of Wikipedia. The earliest dated all the way back to 2002. The field associated with most projects were humanities followed by social science, engineering, medical and natural sciences. He noted teacher issues: making student understand the implication of the license under which they would publish their work, choosing Wikipedia text and the construction of warm-up tasks (creating an account, basic editing, wiki syntax experimentation and categorization). Aims mentioned were increase in student motivation and knowledge, learning to do collaborative writing and gaining the "rigour and balance in writing encyclopeaedic articles" with neutral point of view, theoretical analysis of Wikipedia and propaedeutic: making research, editing and bibliographic processing. Some projects attempted to reach "Good Article" Wikipedia status or better while also failures occurred: In one case only 7 of 70 produced articles survived.

Robert Cummings' 2009 book Lazy Virtues: Teaching Writing in the Age of Wikipedia³³ explores teaching writing in relation to Wikipedia. The book was reviewed by 3 Wikipedias in the 27 April 2009 Wikipedia Signpost and its issues discussed in the podcast Wikivoices episode 41.

Using wikis for course communication

Teachers may use wikis for asyncronous communications in a course. As an encyclopedia Wikipedia should not contain information about specific courses or discussion between individual teachers and students. Another Wikimedia Foundation project, *Wikiversity*, serves the purpose of building a resource for teaching and learning. A quotation from the page Wikiversity: Scope regarding teaching reads:

"The Wikiversity should become a premier source for accurate teaching materials — bibliographies, lesson plans, syllabi, effective animations, well written papers, ideas for in class and at home exercises, discussion topics, course work, etc. — as well as offer a clearing house for new and emerging teaching methodologies and technologies of learning. Teachers making use of Wikiversity materials are encouraged to comment on and adapt those materials in a continual effort to refine and improve Wikiversity's resources."

Researchers have reported experiences of developing and delivering a course through Wikiversity. The 10-week course Composing Free and Open Online Educational Resources communicated weekly

readings, quasi–formal assignments and feedback through Wikiversity. The course also used video conferencing and blogs and included some 70 participates from all over the world, though "towards the end of the course less than two dozen individuals were participating with any level of consistency". ⁴⁸⁰

Wikiversity falls within the open education movement. Outside Wikimedia are other open education wikis, such as the WikiEducator (wikieducator.org). The ideas behind these efforts have been summed up in the *Cape Town Open Education Declaration* (www.capetowndeclaration.org).

Textbooks

Wikimedia Foundation project *Wikibooks* provides a site where editors may write free books collaboratively, e.g., open textbooks. Though a number of books have been declared 'featured books' the success has been mixed. Commentors have pointed to the difficulty in modularization of the text as compared to an encyclopedia as a reason for lack of major success. ⁴⁸¹

A 2008 survey among 80 Wikibook contributors (based on 1'500 contacted) found that the respondents were mostly younger than 25 years (58%) and male (97.5%). 482 Development of books are usually associated with high educational level, but the survey found that 39% only had a high school degree or lower. About one in four reported frustration with their wikibook project. 42% indicated that the book could not be completed. 482

Israeli researchers have reported on the development of an academic textbook on information systems using wiki technology. 483 Faculty, undergraduate and graduate students participated during the course of two years using the MediaWiki software. The authors argues that the wiki textbooks leads to empowerment of both teachers and students over commercial publishers. About 1'200 students from 20 different classes and three different universities have used the text book. The wikitextbook was developed from an original outdated e-textbook consisting of 225 subsections that were converted to wikipages. The students read and augmented the pages and created new ones, so the number of pages would approximately double. The researchers found that for some classes the wiki contribution of students was associated with higher grades on average. They also found a long-tailed distribution of edit activity among the students. Interestingly, one of the successful Wikibooks, the one on the LaTeX markup language, was also seeded from an established work. Computer scientist Andrew Roberts had written an well-structured online tutorial called *Getting to Grips with LaTeX* available from http://www.andy-roberts.net/writing/latex. Wikibookians contacted Roberts, who gave permission to the use of his content on Wikibooks. This book is now one of the prime resources on LaTeX reaching top 5 in my search on Google and Bing with the query "latex". So is the initial seeding the problem with Wikibooks? Would a good initial seed establish a modularization that Wikibooks contributors can use as a base for a successful work?

6 Open questions

What are the open questions in wiki and Wikipedia research?

The big equation?

Is it possible to setup an equation for when a wiki will evolve successfully? Or how successful it is? What variables are part of such an equation? Are money and gender part of the equation?

The Wilkinson-Huberman equation (see page 31) is surpricing. Why would the number of edits on an ensemble of articles only depend on the age of the articles and a "random" fluctuation? What is this "random" fluctuation? Why are there no explicit external variables in the equation? Surely, an edit from a user may result in that edit showing up on the "my watchlist" of another user and possibly motivate that other user to make a new edit. In this case the edits are driven by a Wikipediaintrinic mechanism, but intuitatively should edits not primarily be driven by notability of the subject? The Wilkinson-Huberman equation models the collective of an ensemble of articles of similar age, meaning that the model may have room for the notability of the subjects within the lognormal distribution. The modeling rested on Wikipedia data up to November 2006 in the period of Wikipedia large growth. Given the decline in the number of editors, does the model still hold? In the same $\mathrm{work}^{283,\,284}$ the two researcher show that quality articles are associated with more edits, so do quality articles attract more edits or do many edits results in a quality article?

Why are some Wikipedias bigger than others? The usual initial answer is that the size relates to the number of speakers of the language. This does not explain all variation, and Morten Rask introduced further features to explain the size variations: Human Development Index (HDI), Internet penetration and days in operations. These extensions can apparently not explain some size variations, e.g., the Norwegian bokmål Wikipedia had as of

January 2014 over 400'000 articles while the Danish well below 200'000 and the Swedish over 1.6 million. The differences in sizes of Nordic Wikipedias are mirrored in the size of Wikimedia Chapters: Wikimedia Norway has had almost 100 paying members of the organization, while the Danish chapter has had around 20 members. While differences in days of operation, language speakers, HDI and Internet penetration can to some extent explain the Wikipedia numbers it does not explain all. The bot activity with automated creation of articles on the Swedish Wikipedia explains part of its large size, but why is the Norwegian bigger than the Danish? Faroese Islands and Greenland share almost the same population, yet the Greenlandic Wikipedia shows very little activity with some 1'600 articles, while the Faroese Wikipedia shows more activity with over 7'000 articles. The activity of the Faroese Wikipedia seems to be related to the dedicated activity of a single individual.

What can explain the large difference in number of edits contributors make on Wikipedia? For employed knowledge workers, such as scientists and computer programmers, studies have reported individual variability in productivity with ratios in the tens, 484-486 as well as reporting heavy-tailed distributions of productivity. 484 Although these numbers are large, the individual differences in Wikipedia contributor output dwarf them with orders of magnitude. The usual Wikipedia studies focusing on contributor output do not measure productivity directly (as work per time unit), but rather the overall output, so the time used for Wikipedia contribution could possibly explain the huge differences. A report on Justin Knapp, the first Wikipedian to reach one million edits, said "he has spent 10 hours a day reading and adding to posts". 487 Is there a difference in productivity among Wikipedians? Studies on programmer productivity suggest experience, motivation and intelligence and "a general factor of programmer proficiency" as explanations for the large differences, 485, 486 while William Shockley in his study on scientists considers an 'organizational hypothesis', mental combinatorial ability and a hurdle model with interacting mental factors. 484 The organizational hypothesis (dismissed by Shockley for scientists), where heads of laboratories get coauthorships, is not relevant for Wikipedians as contributors cannot become coauthors on edits. Wikimedia's suggestion for an "Attrition Pipeline" with "impediments to editing at each stage of editor lifecycle" would resemble the Shockley hurdle model, if there are individual differences in contributor persistency in overcoming each impediment, e.g., a struggle with wiki syntax or dealing with policy

complexities. Other explanations for the observed variation build on the idea of 'preferential attachment', 'success-breeds-success' or 'cumulative advantage'. If this explanation is the case then what is the mechanism of feedback? The experiment with barnstars (see page 29) produced only a relatively small variability in productivity between groups that did and did not receive a social reward. The Kaggle prediction competition Wikipedia's Participation Challenge in 2011 sought "to understand what factors determine editing behavior". In this competition participants should predict the future number of edits a contributor would make. One of successfull competitors used only the edit history for prediction based on the features: number of edits in recent time periods, number of edited articles in recent time periods and time between first and last edit. 488 Another system used over 40 features but did not yield better results. 489 A deeper explanation for these results lacks: "You edit because you have edited" seems not a satisfactory explanation.

Can Wikipedia withstand contributions from professional public relations professionals and search engine optimization companies? Can the evolution of conflict-of-interests edits be quantified? In 2005 and again in 2006 Eric Goldman predicted that Wikipedia would fail within five years. 182, 183 Yet again in 2010 he would point to the problems faced by the Open Directory Project as a user-generated content site becoming overwhelmed by relentless spam, providing a cautionary tale to Wikipedia.²⁶⁸ Seth Finkelstein would in 2008 write that he "sometimes remind people that ideological communes tend to end badly, too often with a few manipulative leaders extensively profiting at the expense of a mass of followers who lose everything". 224 Yet in 2015 Wikipedia still stands. Still in 2013 reporting from a major sockpuppert investigation journalist Simon Owen gave a pessimistic outlook: "But while reporting this article I couldn't help comparing the sockpuppet discovery to a large drug bust—perhaps it might take out a major kingpin, but at the end of the day it's a relatively minor victory in what is an otherwise losing war on drugs." 197 Furthermore, in the spring of 2015 Wikipedian and researcher Piotr Konieczny op-eded his growning concerns about "advertisements masquerading as articles" arguing that "Wikipedia has been gravitating towards a vehicle for business and product promotion for too long" and called out for a major cleanup drive of promotional articles which he with a napkin estimate put to 300,000 on the English Wikipedia, i.e., 3% of the total number of articles.⁴⁹⁰ The Sony email leaks from November 2014 cast some light

on companies Wikipedia practices, showing a Sony employee apparently updating Wikipedia articles on the CEO and the company encouraging creation of articles about their movies. 491

In 2013 Jimmy Wales pointed to what he saw as two threats for Wikipedia's ongoing success: authoritarian governments censoring the Internet and "clumsy legislation driven by special interests", such as the Stop Online Piracy Act. 492 Are these threats serious?

What relation should there be between academics and Wikipedia?

What relation should there be between academics and Wikipedia? Should academics be encouraged to add their work on Wikipedia and rewarded or would a large-scale involvement of academics deteriorate Wikipedia because academics in search of employment and grants would tend to make biased conflict-of-interest contributions with overstating the importance of their own work? Does formel academic contributions short circuit the open collaborative hobby-oriented nature of Wikipedia editing? Should Wikipedia develop a method whereby experts can claim an article version as expert verified?

Since 2012 Wikiversity has incorporated a few peer-reviewed articles, e.g., "Diagram of the pathways of human steroidogenesis", "Reference ranges for estradiol, progesterone, luteinizing hormone and follicle-stimulating hormone during the menstrual cycle" and the initial study "Average weight of a conventional teaspoon made of metal" ⁴⁹³ all written by Mikael Häggström. On Wikipedia's WikiProject Medicine talk page editors have discussed whether such articles can act as reliable sources for Wikipedia. liii The discussants would not fully embrace and recognize Wikiversity as reliable sources, — with its present form of peer-review. It is a question whether wiki-based publishing can gain traction and sufficient recognition, so that such a Wikiversity-based approach is a viable option for career-oriented scientists. Should such an approach be supported?

Why are not all scientific articles cited or mentioned on Wikipedia? Is it because the topics that the scientific articles cover are not yet dealt with in depth by Wikipedia, or is it because the topics of the majority of scientific articles are not notable enough and should be left out? Is much of science simply not encyclopedia-worthy? The *Open Access Reader* is a project initiated in 2014 aiming to

systematically ensure coverage of 'significant' published open access research in Wikipedia. How feasible is such a project given the large number of scientific articles published each year?

Writing about scientific topics in Wikipedia one comes to question the validity of the claims in many scientific articles. Science itself has had discussions about the issue sometimes referred to as the 'reproducibility crisis' in experimental sciences. 494 Ioannidis claims from theoretical considerations and a number of assumptions that "most published research findings are false", 495 — ironically a claim itself critized as false. 496-498 However, he has done empirical research finding that a considerable part of highly-cited biomedical does not replicate in subsequent studies. 499 Wikipedia already has a warning in the content guideline Wikipedia:Identifying reliable sources against relying on individual studies: "Isolated studies are usually considered tentative and may change in the light of further academic research." Is Wikipedia more accurate than the average scientific article? If so, can the Wikipedia process be transferred to the scientific process? Can Wikipedia and similar tools do anything about the 'reproducibility crisis' in experimental sciences?

Although citations and universities may be ranked to some degree based on citations in links in Wikipedia, 80,148 scientists have poor coverage on Wikipedia. 89 Should we caution against using Wikipedia for altmetrics? And if Wikipedia altmetrics becomes established would that have a deterioating effect of the quality of Wikipedia? How do we ensure that Wikipedia continuously reflects the knowledge in the academic literature?

Small scientific wikis, independent on large organizations such as the Wikimedia Foundation, may run into sustainability problems like other online scientific databases: The wikis are created and operated not seldomly by research groups on timelimited research funding, and if the funding stops the continued operation of the wiki is jeopardized. Open wikis may require daily attention which is not available from researchers moved on to new projects. A few scientific wikis have stopped operation, e.g., in January PDBWiki produced the message "PDBWiki is offline as of January 14th, 2014" vi Also WikiProtein seems no longer available. It is an open question whether Wikipedia and its sister projects, particularly now with Wikidata, can provide a better framework for sustainable scientific wiki work compared to separate wikis. Surely, if applying Wikidata instead of a locally

liii"Nomination" of steroidogenesis article in Wikiversity to be used as reference section on "Wikipedia talk:WikiProject Medicine", 22:58, 15 April 2014, oldid 604356183.

livhttps://meta.wikimedia.org/wiki/OpenAccessReader.
lvWikipedia:Identifying reliable sources, oldid = 596070076, English Wikipedia.

lvihttp://pdbwiki.org/

controlled wiki the scientist's control of authorship is mostly lost but the task of system and wiki administration is considerably reduced. It is unclear whether Wikidata sufficiently supports the many varied data representations found in scientific wikis and whether external tools can sufficiently present the Wikidata data in the ways its presently done on specialized scientific wikis. Furthermore, will notability issues hinder the representation of scientific data on Wikimedia sites?

What can Wikipedia say about the world and can it affect the world?

To what extent can Wikipedia be used for for prediction of world events such as movie boxoffice, election results, etc.? Does Wikipedia have a positive societal value (whatever that is) or is it merely a Friday bar argument settling site? How appropriate is the two digit—or even four digit—value in milliard US Dollars mentioned by Band and Gerafi²⁵² for the consumer value or positive externalities?

Does Wikipedia affect the world and to what extent? The Chinese censorship of the Chinese Wikipedia⁵⁰⁰ and the Iranian censorship of some of the Persian Wikipedia articles⁵⁰¹ could indicate that some authorities are afraid of the impact of Wikipedia information. To what extent does political Wikipedia articles affect voting?

Noam Cohen attributed the Microsoft Encarta demise to the competion from the free Wikipedia. The introduction of Encarta had earlier — in the 1990s — impacted the sale of the competiting <code>Encyclopædia Britannica</code>: <code>Encyclopædia Britannica</code> hard copy book sales fell by 50% over a couple of years, affecting an annual revenue on several hundreds of millions US Dollars. $^{503}, ^{504}$

Do students get better by using Wikipedia? Researchers have carried out a field study with a controlled experiment using OLPC computers, that among the applications had an offline copy of Among several hundreds Ethiopian Wikipedia. children the researchers found that children with no laptop got higher grades in mathematics. They also found that school engagement was no different among the two groups with and without an OLPC device. Could this indicate that children are put at a disadvantage by reading Wikipedia? This is not what the study says, because the researchers also examined abstract reasoning in the children and reported that the results indicated that children with laptops outperformed children without. 416 Other technologies have been promoted in learning, but although heralded as a new age for learning, they may not fulfill the promises to revolutionize education. A study on iPad reading in Year 6 pupils concluded "[a]lthough the students participating in the study reported an increase in engagement when using the iPad, there was not a corresponding rise in achievement." 505 Comparisons of the scores from the Programme for International Student Assessment (PISA) in Denmark with pupils grouped according to whether they had access to tablet computers in school or not showed a lower average score for those with tablet access.⁵⁰⁶ Such studies call into question whether Wikipedia furthers education, and whether teachers should encourage students to use Wikipedia at all. Given that at least the spikes in Wikipedia page views are related to pop culture phenomenons, such as celebrity deaths and Super Bowl halftime entertainers, 250 are students diverged away from learning 'useful knowledge' to reading pop culture when they engage with Wikipedia?

Structured data?

In January 2013 the Wikidata project was launched and it grew quickly. In terms of pages it quickly surpassed the English Wikipedia. A few papers have described the system in overview, ^{384,385} but otherwise very little research have been done on this young project. What is the authorship and readership of Wikidata? Are bots the dominating force on Wikidata or do humans have a say? What types of data can humans contribute with? Do we see a highly scewed editing pattern where a few prolific bots make the majority of the work?

Will we see 'ontology wars' on Wikidata where editors cannot agree on properties? One example of property discussions is how humans/persons should be described. Initially, the property 'GND-type' (German National Library) stating 'person' was used but later this property was deprecated for labeling persons as persons.

What about the multi-lingual nature of the Wikidata project? Is the regimentional description of items across languages 'dangerous'? Does an ontology uniform across languages prohibits cultural diversity? Is the predominantly English discussions on Wikidata a problem for non-English users? Before the launch of Wikidata Mark Graham raised concerns that the "highly significant and hugely important" changes brought by Wikidata "have worrying connotations for the diversity of knowledge" on Wikipedia. He believed that "[i]t is important that different communities are able to create and reproduce different truths and worldviews," exemplifying the problem with the population of Israel: Should it include occupied and contested territories?⁵⁰⁷ Wikidata—at least partly—counters Graham's worry: each property (e.g., 'population')

may have multiple values, and qualifiers associated with Wikidata values can distinguish between the scope of each claim. Denny Vrandečić, project director of Wikidata, explained: "We do not expect the editors to agree on the population of Israel, but we do expect them to agree on what specific sources claim about the population of Israel." lvii However, the question remains whether the Wikidata system provides sufficient flexibility to capture, e.g., the slight interlanguage differences in some concepts, and when there are differences does the definition fall back to one centered in the anglocentric world-view? Take the Germanic concept of 'Hochschule'/'højskole'/'högskola'. The English Wikipedia has a separate article about the German(ic) concept of 'Hochschule', and Wikidata has an item for the concept. However, that concept dangles in Wikidata, since the German Wikipedia 'Hochschule' article leads to another Wikidata item associated with the concept of a 'higher education organization'. Another potential problem could be whether certain values for a property should exist or not. Take the borders of Israel as an example: The property 'shares border with' presently lists Syria, Jordan, Egypt and Lebanon, and not the State of Palestine. The international recognition of the State of Palestine as a state varies between countries, so according to one view the State of Palestine should be listed, while the opposing view would hold it should not. A suitable qualier ('as recognized by'?) could possibly resolve it. Another issue arises for almost similar concepts which can be linked by the informal interwiki Wikipedia links with no problem, but where the semantic description will be difficult in a multilingual environment. Take the classic scientific article The magical number seven plus or minus two: some limits on our capacity for processing information. In the English and French Wikipedias it has its own article, while the German Wikipedia chooses to make an article on the psychological concept described in the article ("Millersche Zahl"): Whether this item has an author or not and a date of publication depends on whether one regard it as a publication or a concept. A pure semantic approach would split such cases, increasing babelization.

I have experienced a few instances of vandalism on Wikidata: French footballer Anthony Martail called 'Rien' (French: nothing, actually a Norwegian lake) and American basketball player Michael Jordan called 'insect'. 'Called' here means attributed the GND type of 'Rien' or 'insect', — neither standard GND types. Will vandalism be a problem on Wikidata? On Wikipedia a bot with

machine learning-based detection from vandalism features operate. Would automatic vandalism detection be possible on Wikidata?

Can Wikidata describe everything? What kinds of data can Wikidata not conveniently describe? Wikisource, Wikivoyage and Wikimedia Commons got relatively unhindered their language links represented in Wikidata and according to the Wikidata development plan for 2014+lviii Wikiquote, Wikinews, Wikibooks and Wikiversity are schedule for inclusion in Wikidata. However, Wiktionary was as of January 2014 not scheduled for Wikidata inclusion. It is unclear if the item/property system of Wikidata is an appropriate representation for words and how lexemes, lemmas, forms and senses should most easily be represented with the Wikidata data model.

How advanced queries can be made with Wikidata data? Magnus Manske's AutoList tool can already now carry out on-the-fly queries like "all poets who lived in 1982". But can such queries continue to be carried out effectively, and will Wikidata generally be able to scale? For example, how will Wikidata cope with several thousand claims per item? It may be worth to remember that Wikipedia articles seldomly reach past 200-300 kB because articles get split into subarticles, e.g., "Barack Obama" splits into "Family of Barack Obama" and "Illinois Senate career of Barack Obama" etc. This sharding technique seems not to be readily possible with Wikidata. Dynamic Wikidata-based translation in its present form, e.g., through the qLabel Javascript library, can result in multiple requests to Wikidata servers for just a single page view on a third-party website. If successful, will Wikidatabased translation results in unmanageable load on Wikidata?

Acknowledgment

Thanks to Daniel Kinzler, Torsten Zesch, Felipe Ortega, Piotr Konieczny, Claudia Koltzenburg and James Heilman for pointing to references and tools Thanks also to Chitu Okoli, Mohamad Mehdi, Mostafa Mesgari and Arto Lanamäki with whom I am writing systematic reviews about Wikipedia research.¹⁴

References

[1] Jakob Voß. Measuring Wikipedia. In Proceedings International Conference of the International So-

lvii See comments to Mark Graham's article.

lviiihttps://www.wikidata.org/wiki/Wikidata:Development_plan.

ciety for Scientometrics and Informetrics: 10th, 2005.

[2] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. DBpedia: A nucleus for a web of open data. In The Semantic Web, volume 4825 of Lecture Notes in Computer Science, pages 722–735, Heidelberg/Berlin, 2007. Springer.

Annotation: Description of a system that extracts information from the templates in Wikipedia, processes and presents them in various ways. Some of the methods and services they use are MySQL, Virtuoso, Open-Cyc, GeoNames, Freebase, SPARQL and SNORQL. The system is available from http://DBpedia.org

[3] Lee Rainer and Bill Tancer. Data memo. Report, Pew Research Center's Internet & American Life Project, April 2007.

Annotation: Reports the result of a survey on Wikipedia use among American adults.

[4] Kathryn Zickuhr and Lee Rainie. Wikipedia, past and present. Report, Pew Research Center's Internet & American Life Project, Washington, D.C., January 2011.

Annotation: Presents the results of a survey on Americans use of Wikipedia. It is based on telephone interviews conducted in the spring 2010. It is found that 53% of Internet users used Wikipedia, corresponding to 42% of adult Americans.

- [5] Seth Schaafsma and Selina Kroesemeijer. Getting to know the grassroots. Technical report, Vereniging Wikimedia Nederland, July 2013.
- [6] Han-Teng Liao. Growth of academic interest in Wikipedia from major Chinese-speaking regions: Has it peaked?. Internet, February 2012.

Annotation: Short blogpost plotting the number of theses form major Chines-speaking regions as a function of year.

[7] Chitu Okoli and Kira Schabram. Protocol for a systematic literature review of research on Wikipedia. In Proceedings of the International ACM Conference on Management of Emergent Digital EcoSystems, New York, NY, USA, 2009. Association for Computing Machinery.

Annotation: Short article that sets up a framework for systematic review of Wikipedia research identifying roughly 1,000 articles.

[8] Chitu Okoli, Kira Schabram, and Bilal Abdul Kader. From the academy to the wiki: practical applications of scholarly research on Wikipedia. Wikimania, 2009.

> Annotation: Short review of peerreview Wikipedia research. The researchers identified over 400 academic papers. A few of these 400 are briefly summarized.

- [9] Olena Medelyan, David Milne, Catherine Legg, and Ian H. Witten. Mining meaning from Wikipedia. *International Journal of Human-Computer Studies*, 67(9):716–754, September 2009
- [10] Nicolas Jullien. What we know about Wikipedia: A review of the literature analyzing the project(s). ArXiv, May 2012.

Annotation: A review of Wikipedia research.

- [11] Finn Årup Nielsen. Wikipedia research and tools: Review and comments. SSRN, February 2012 2012.
- [12] Chitu Okoli. A brief review of studies of Wikipedia in peer-reviewed journals. In Digital Society, 2009. ICDS '09. Third International Conference on, pages 155–160. IEEE, 2009.
- [13] Arto Lanamäki, Chitu Okoli, Mohamad Mehdi, and Mostafa Mesgari. Protocol for systematic mapping of Wikipedia studies. In Timo Leino, editor, Proceedings of IRIS 2011, number 15 in TUCS Lecture Notes, pages 420–433. University of Turku, October 2011.
- [14] Chitu Okoli, Mohamad Mehdi, Mostafa Mesgari, Finn Årup Nielsen, and Arto Lanamäki. The people's encyclopedia under the gaze of the sages: A systematic review of scholarly research on Wikipedia. Social Science Research Network, October 2012.

Annotation: Systematic review on research on Wikipedia focusing on peer-reviewed journal articles and theses published up until June 2011 and included some important conference papers. The review covers all aspects of Wikipedia research, including uses of Wikipedia-derived data.

- [15] Mostafa Mesgari, Chitu Okoli, Mohamad Mehdi, Finn Årup Nielsen, and Arto Lanamäki. "The sum of all human knowledge": A systematic review of scholarly research on the content of Wikpedia. Journal of the Association for Information Science and Technology, 66(2):219–245, 2015.
- [16] Chitu Okoli, Mohamad Mehdi, Mostafa Mesgari, Finn Årup Nielsen, and Arto Lanamäki. Wikipedia in the eyes of its beholders: a systematic review of scholarly research on Wikipedia

readers and readership. Journal of the Association for Information Science and Technology, 65(12):2381–2403, 2014.

Annotation: Systematic review on scientific research on Wikipedia and its readership

[17] Oliver Keyes. English Wikipedia pageviews by second. figshare, April 2015.

Annotation: Data set with page view statistics from the English Wikipedia collected in March and April 2015.

[18] Ludovic Denoyer and Patrick Gallinari. The Wikipedia XML corpus. ACM SIGIR Forum, 40(1):64–69, June 2006.

Annotation: Describes a dataset with Wikipedia text represented in XML.

- [19] Gordon Müller-Seitz and Guido Reger. 'Wikipedia, the free encyclopedia' as a role model? lessons for open innovation from an exploratory examination of the supposedly democratic-anarchic nature of Wikipedia. International Journal of Technology Management, 32(1):73–88, 2010.
- [20] Roy Rosenzweig. Can history be open source? Wikipedia and the future of the past. Journal of American History, 93(1):117-146, June 2006.

Annotation: Discuss several aspects of history on the English Wikipedia and how professional historians should regard that wiki. The author also make a quality assessment of a Amerian history articles on Wikipedia and compare them against Encarta and American National Biography Online.

[21] The ed17 and Tony1. Wikipedia's traffic statistics understated by nearly one-third. Wikipedia Signpost, September 2014.

Annotation: Blog post on Wikipedia page view statistics discovered to be wrong.

- [22] Sameer Singh, Amarnag Subramanya, Fernando Pereira, and Andrew McCallum. Wikilinks: A large-scale cross-document coreference corpus labeled via links to Wikipedia. Technical report, University of Massachusetts, October 2012.
- [23] Sameer Singh, Amarnag Subramanya, Fernando Pereira, and Andrew McCallum. Large-scale cross-document coreference using distributed inference and hierarchical models. In *Human Lan*guage Technologies. Association for Computational Linguistics, 2011.
- [24] Heather Ford. Onymous, pseudonymous, neither or both?. Ethnography Matters, June 2013.

Annotation: Blog post that 'explores the complications of attribution and identification in online research'.

[25] Andrew Lih. The Wikipedia revolution: How a bunch of nobodies created the world's greatest encyclopedia. Hyperion, March 2009.

Annotation: Book on various aspect of Wikipedia and related phenomenons: Usenet, Nupedia, wikis, bots, Seigenthaler incident, Essjay controvery, Microsoft Encarta, etc.

- [26] Phoebe Ayers, Charles Matthews, and Ben Yates. How Wikipedia works: An how you can be a part of it. No Starch Press, September 2008.
- [27] Daniela J. Barrett. Mediawiki. O'Reilly, 2008.
- [28] Yaron Koren. Working with MediaWiki. Wiki-Works Press, November 2012.

Annotation: Book about MediaWiki and Semantic MediaWiki.

- [29] Bo Leuf and Ward Cunningham. The wiki way: Quick collaboration on the web. Addison-Wesley, Boston, April 2001.
- [30] Geert Lovink and Nathaniel Tkacz, editors. Critical point of view: A Wikipedia reader, volume 7 of INC Reader. Institute of Network Cultures, Amsterdam, The Netherlands, 2011.
- [31] John Broughton. Wikipedia: The missing manual. O'Reilly Media, 2008.
- [32] Andrew Dalby. The world and Wikipedia: How we are editing reality. Siduri Books, 2009.
- [33] Robert E. Cummings. Lazy virtues: Teaching writing in the age of Wikipedia. Vanderbilt University Press, 2009.
- [34] Aaron Shaw, Amir E. Aharoni, Angelika Adam, Bence Damokos, Benjamin Mako Hill, Daniel Mietchen, Dario Taraborelli, Diederik van Liere, Evan Rosen, Heather Ford, Jodi Schneider, Giovanni Luca Ciampaglia, Lambiam, Nicolas Jullien, Oren Bochman, Phoebe Ayers, Piotr Konieczny, Adam Hyland, Sage Ross, Steven Walling, Taha Yasseri, and Tilman Bayer. Wikimedia Research Newsletter, volume 2. 2012.

Annotation: Aggregation of the Wikimedia Research Newsletter for the year 2012 written by various researchers.

[35] Rikke Frank Jørgensen. Making sense of the German Wikipedia community. MedieKultur, 28(53):101–117, 2012.

Annotation: Describe the German Wikipedia based on qualitative interviews with seven members of the Berlin Wikipedia community.

- [36] Patrick M. Archambault, Tom H. van de Belt, Francisco J. Grajales III, Marjan J. Faber, Craig E. Kuziemsky, Susie Gagnon, Andrea Bilodeau, Simon Rioux, Willianne L.D.M. Nelen, Marie-Pierre Gagnon, Alexis F. Turgeon, Karine Aubin, Irving Gold, Julien Poitras, Gunther Eysenbach, Jan A.M. Kremer, and France Légaré. Wikis and collaborative writing applications in health care: A scoping review. Journal of Medical Internet Research, 15(10):e210, 2013.
- [37] Yochai Benkler and Helen Nissenbaum. Commons-based peer production and virtue. The Journal of Political Philosophy, 14(4):394–419, 2006.

Annotation: Discuss commonsbased peer production exemplified with free and open source software, SETI@home, NASA Clickworkers, Wikipedia and Slashdot.

- [38] Piotr Konieczny. Wikipedia: community or social movement. *Interface*, 1(2):212–232, November 2009.
- [39] Fabian M. Suchanek and Gerhard Weikum. YAGO: a large ontology from Wikipedia and WordNet. *Journal of Web Semantics*, 6(3):203–217, 2008.
- [40] Andrew Lih. How Wikipedia solved the knowledge gap. TEDx-American University, April 2014.

Annotation: Talk on Wikipedia at TEDx-American University.

[41] Besiki Stvilia, Michael B. Twidale, Les Gasser, and Linda C. Smith. Information quality discussions in Wikipedia. In Suliman Hawamdeh, editor, Knowledge Management. Nurturing Culture, Innovation, and Technology. Proceedings of the 2005 International Conference on Knowledge Management, pages 101–113, Singapore, October 2005. World Scientific.

Annotation: An analysis with respect to information quality of a sample of the discussion pages on Wikipedia. The analysis is based on the their own information quality assessment model, and they provide example quotations from the pages. They conclude that 'the Wikipedia community takes issues of quality very seriously'.

[42] Jim Giles. Internet encyclopaedias go head to head. Nature, 438(7070):900–901, December 2005.

Annotation: Report of a comparison of the accuracy in Wikipedia and Encyclopedia Britannica

[43] Barry X. Miller, Karl Helicher, and Teresa Berry. I want my Wikipedia. Library Journal, 6, April 2006

Annotation: Report on three experts informally reviewing popular culture, current affairs and science articles in Wikipedia. Popular culture reviewer calls it "the people's encyclopedia"; current affairs reviewer "was pleased by Wikipedia's objective presentation of controversial subjects", but cautioned "a healthy degree of skepticism and skill at winnowing fact from opinion is required". Lastly, the science reviewer found it difficult to characterize because of the variation, noting "flaws" and "the writing is not exceptional", but "good content abounds".

- [44] Lara Devgan, Neil Powe, Brittony Blakey, and Martin Makary. Wiki-surgery? internal validity of Wikipedia as a medical and surgical reference. *Journal of the American College of Surgeons*, 205(3, supplement):S76–S77, September 2007.
- [45] John Gever. Wikipedia information on surgical procedures generally accurate. DocGuide.com, October 2007.

Annotation: Summary of the research by Lara Devgan et al. "Wiki-Surgery? Internal Validity of Wikipedia as a Medical and Surgical Reference".

- [46] Wikipedia schlägt Brockhaus. Stern, December 2007.
- [47] K. C. Jones. German Wikipedia outranks traditional encyclopedia's online version. *Information-Week*, December 2007.
- [48] Kevin A. Clauson, Hyla H. Polen, Maged N. Kamel, and Joan H. Dzenowagis. Scope, completeness, and accuracy of drug information in Wikipedia. The Annals of Pharmatherapy, 42, December 2008.

Annotation: Examines Wikipedia as a drug reference with a comparison against Medscape Drug Reference. Wikipedia has more omissions but no factual errors among 80 drug-related questions/answers, e.g., for administration, contraindications and issues around pregnancy/lactation. They also find that Wikipedia has no dosage information, but this is not surprising given that Wikipedia explicit encourage authors not to add this information. Four factual errors were found in Medscape. Two were conflicting information in different

parts of the text while the remaining two were due to lack of timely update.

[49] Michael P. Pender, Kaye E. Lasserre, Lisa M. Kruesi, Christopher Del Mar, and Satyamurthy Anuradha. Putting Wikipedia to the test: a case study. In The Special Libraries Association Annual Conference, June 2008.

Annotation: A small blinded comparison of 3 Wikipedia articles for medical student information against AccessMedicine, eMedicine and UpTo-Date online resources. Wikipedia was found to be unsuitable for medical students.

[50] Michael P. Pender, Kaye E. Lasserre, Christopher Del Mar, Lisa Kruesi, and Satyamurthy Anuradha. Is Wikipedia unsuitable as a clinical information resource for medical students? *Medical Teacher*, 31:1094–1098, 2009.

Annotation: The same study as "Putting Wikipedia to the test: a case study".

- [51] Malolan S. Rajagopalan, Vineet K. Khanna, Yaacov Leiter, Meghan Stott, Timothy N. Showalter, Adam P. Dicker, and Yaacov R. Lawrence. Patient-oriented cancer information on the internet: a comparison of Wikipedia and a professionally maintained database. *Journal of Oncology Practice*, 7(5):319–323, September 2011.
- [52] M. S. Rajagopalan, V. Khanna, M. Scott, Y. Leiter, T. N. Showalter, A. Dicker, and Y. R. Lawrence. Accuracy of cancer information on the internet: A comparison of a wiki with a professionally maintained database. *Journal of Clinical* Oncology, 28:7s, 2010. Supplement abstract 6058.

Annotation: A short abstract reporting that Wikipedia had similar accuracy and depth compared to the professionally edited information in the National Cancer Institute's Physician Data Query (PDQ), but Wikipedia was less readable as evaluated with the Flesch–Kincaid readability test.

- [53] Andreas Leithner, Werner Maurer-Ertl, Mathias Glehr, Joerg Friesenbichler, Katharina Leithner, and Reinhard Windhager. Wikipedia and osteosarcoma: a trustworthy patients' information. Journal of the American Medical Informatics Association, 17(4):373–374, July-August 2010.
- [54] N. J. Reavley, A. J. Mackinnon, A. J. Morgan AJ, M. Alvarez-Jimenez, S. E. Hetrick SE, E. Killackey, B. Nelson, R. Purcell, M. B. Yap, and A. F. Jorm. Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources. Psychological medicine, 42(8):1753–1762, December 2011.

Annotation: Study on the quality of information on Wikipedia about mental disorders in terms of accuracy, up-to-dateness, coverage, references and readbility with comparison against 13 other websites. 10 topics were rated the three psychologists with relevant expertise. Wikipedia was generally rated higher than the other websites.

- [55] Stacey M. Lavsa, Shelby L. Corman, Colleen M. Culley, and Tara L. Pummer. Reliability of Wikipedia as a medication information source for pharmacy students. *Currents in Pharmacy Teaching and Learning*, 3(2):154–158, April 2011.
- [56] T. Aldairy, S. Laverick, and G. T. McIntyre. Orthognathic surgery: is patient information on the Internet valid. European Journal of Orthodontics, 34(4):466–469, August 2012.
- [57] Garry R. Thomas, Lawson Eng, Jacob F. de Wolff, and Samir C. Grover. An evaluation of Wikipedia as a resource for patient education in nephrology. Seminar in Dialysis, 2013.

Annotation: An analysis of the quality of Wikipedia articles in the neuphrology area. They looked at comprehensiveness, reliability and readability.

[58] Olavi Koistinen. World's largest study on Wikipedia: Better than its reputation. Helsinski Times, December 2013.

Annotation: News articles about an expert evaluation of 134 articles on the Finish Wikipedia on the dimensions lack of errors, coverage and balance, sourcing, topicality, neutrality and clarity.

[59] Robert T. Hasty, Ryan C. Garbalosa, Vincenzo A. Barbato, Pedro J. Jr Valdes, David W. Powers, Emmanuel Hernandez, Jones S. John, Gabriel Suciu, Farheen Qureshi, Matei Popa-Radu, Sergio San Jose, Nathaniel Drexler, Rohan Patankar, Jose R. Paz, Christopher W. King, Hilary N. Gerber, Michael G. Valladares, and Alyaz A. Somji. Wikipedia vs peer-reviewed medical literature for information about the 10 most costly medical conditions. The Journal of the American Osteopathic Association, 114(5):368–373, May 2014.

Annotation: A study with comparison of assertions in Wikimedia articles on medical conditions with assertions in the peer-reviewed literature.

[60] Lori Fitterling. Wikipedia: Proceed with caution. The Journal of the American Osteopathic Association, 114(5):334–335, 2014.

Annotation: Editorial about study on Wikipedia quality for medicinal conditions.

- [61] Joy H. Fraser and Norman Temple. Wikipedia vs peer-reviewed medical literature for information about the 10 most costly medical conditions-I. The Journal of the American Osteopathic Association, 114(10):761, October 2014.
- [62] Jonathan Leo and Jeffrey R. Lacasse. Wikipedia vs peer-reviewed medical literature for information about the 10 most costly medical conditions-II. The Journal of the American Osteopathic Association, 114(10):761-764, October 2014.
- [63] George S. Chen and Yi Xiong. Wikipedia vs peerreviewed medical literature for information about the 10 most costly medical conditions-III. The Journal of the American Osteopathic Association, 114(10):764–765, October 2014.
- [64] Thomas J. Hwang, Florence T. Bourgeois, and John D. Seeger. Drug safety in the digital age. The New England Journal of Medicine, 370:2460– 2462, 2014.

Annotation: Analysis of 22 Wikipedia articles on prescription drugs and whether they are updated in relation to new drug-safety communications from the U.S. Food and Drug Administration. 36% articles remained unchanged after more than a year.

[65] Jona Kräenbring, Tika Monzon Penza, Joanna Gutmann, Susanne Muehlich, Oliver Zolk, Leszek Wojnowski, Renke Maas, Stefan Engelhardt, and Antonio Sarikas. Accuracy and completeness of drug information in Wikipedia: a comparison with standard textbooks of pharmacology. PLOS ONE, September 2014.

Annotation: Analysis of the quality of drug information in the German and English Wikipedias. Accuracy, completeness, readability, number of references, revisions and editors were considered.

- [66] Behdin Nowrouzi, Basem Gohar, Camille Smith, Behnam Nowrouzi-Kia, Rubaiya Khan, Alicia McDougall, Martyna Garbaczewska, Shalini Sivathasan, Keith Brewster, and Lorraine Carter. An examination of health, medical and nutritional information on the internet: a comparative study of wikipedia, webmd and the mayo clinic websites. The International Journal of Communication and Health, (6), 2015.
- [67] Christina R. Vargas, Neelesh A. Kantak, Danielle J. Chuang, Pieter G. Koolen, and Bernard T. Lee. Assessment of online patient materials for breast reconstruction. *Journal of Surgical Research*, May 2015.

[68] Thomas Chesney. An empirical examination of Wikipedia's credibility. First Monday, 11(11), 2006

Annotation: A study where 258 researcher were asked to evaluate a Wikipedia article. 69 researcher responded and evaluated either an article within their own field of research or a random article. The evaluation focused on credibility.

- [69] Wikipedia schlägt die Profis. Spiegel Online, December 2007.
- [70] Torsten Kleinz. Reputation und Wikipedia: von Einfluss, Anerkennung und Vergleichen. heise online, December 2007.
- [71] Lars Henrik Aagaard. Hvem er bedst?. Berlingske Tidende, July 2009.

Annotation: Danish newspaper article comparing eight articles in the Danish Wikipedia with eight articles in Den Store Danske encyclopedia, with Wikipedia performing slightly better.

- [72] Christina R. Vargas, Joseph A. Ricci, Danielle J. Chuang, and Bernard T. Lee. Online patient resources for liposuction: A comparative analysis of readability. *Annals of Plastic Surgery*, February 2015.
- [73] Henning Bergenholtz. Hvilke krav skal en eordbog opfylde for at kunne anvendes som brugbart værktøj?. DF Revy, 31(5):12–14, 2008.

Annotation: Discussion of electronic reference works.

[74] Christian M. Meyer and Iryna Gurevych. Worth its weight in gold or yet another resource a comparative study of Wiktionary, OpenThesaurus and GermaNet. In A. Gelbukh, editor, CI-CLing, volume 6008 of Lecture Notes in Computer Science, pages 38–49, Berlin-Heidelberg, 2010. Springer-Verlag.

Annotation: A comparative study of German Wiktionary, OpenThesaurus and GermaNet.

- [75] Patrick O'Connor. Maurice Jarre. Guardian, March 2009.
- [76] Siobhain Butterworth. The readers' editor on ... web hoaxes and the pitfalls of quick journalism. The Guardian, May 2009. Accessed 4 May 2009.
- [77] 'Monkey'. 'Wikipedia vandals' strike again in Norman Wisdom obits. Media Monkey, Guardian, October 2010.
- [78] Stephen Caddick. Wiki and other ways to share learning online. *Nature*, 442:744, August 2006.

- [79] Udo Altmann. Representation of medical informatics in the Wikipedia and its perspectives. In R. Engelbrecht et al., editors, Connecting Medical Informatics and Bio-Informatics, volume 116 of Studies in Health Technology and Informatics, pages 755–760. IOS Press, 2005.
- [80] Finn Årup Nielsen. Scientific citations in Wikipedia. First Monday, 12(8), August 2007.

Annotation: Statistics on the outbound scientific citation from Wikipedia with good correlation to the Journal Citation Reports from Thomson Scientific.

[81] Finn Årup Nielsen. Clustering of scientific citations in Wikipedia. In Wikimania, 2008.

Annotation: Study of the citations in Wikipedia to scientific journals. Wikipedia articles and journals are biclustered and a longitudinal analysis of the number of citations is performed.

[82] Alexander Halavaisa and Derek Lackaff. An analysis of topical coverage of Wikipedia. Journal of Computer-Mediated Communication, 13(2):429–440, January 2008.

> Annotation: Examination of the coverage of Wikipedia by two methods: First method entails sampling 3000 Wikipedia articles and manually categorizing them against the Library of Congress category system, then comparing them with the number of published books within the categories. This method shows that, e.g., the literature category is underrepresented in Wikipedia, explained by the lack of fiction in Wikipedia. Other categories underrepresented are, e.g., medicine, philosophy and law, whereas science and some areas with systematice massinserted articles, such as naval and geography, are well represented. Also fans tend to add, e.g., popular music groups and literature description articles. The length of medicine and law article are large compared to the rest of the articles. The second method matched Wikipedia articles with specialized encyclopedia in physics, linguistic and poetry and found missing articles in Wikipedia.

[83] Beate Elvebakk. Philosophy democratized? a comparison between Wikipedia and two other Web-based philosophy resources. First Monday, 13(2), November 2008.

Annotation: A comparison of Stanford Encyclopedia of Philosophy, Internet Encyclopedia of Philosophy

and Wikipedia of twentieth century philosophers. Wikipedia had by far the most philosophers, and Wikipedia had also a large number of more recently born philosophers. Besides this the three encyclopedia did not differ greatly in coverage with respect to gender, disciplines and nationality, though Wikipedia had a smaller fraction of German and French but a higher fraction of other Europeans, and Wikipedia has also categorizes some in the spiritual disciplines as philosophers.

[84] Kasia Czarnecka-Kujawa, Rupert Abdalian, and Samir C. Grover. The quality of open access and open source internet material in gastroenterology: is Wikipedia appropriate for knowledge transfer to patients. Gastroenterology, 134(4, Supplement 1):A-325-A-326, April 2008.

Annotation: Study on the comprehensiveness of Wikipedia in ICD-9 and ICD-10 diagnostic codes in gastroenterology. Over 80% were covered. Readability is also considered.

- [85] George Bragues. Wiki-philosophizing in a marketplace of ideas: evaluating Wikipedia's entries on seven great minds. *MediaTropes eJournal*, 2(1):117–158, 2009.
- [86] Jeff Friedlin and Clement J. McDonald. An evaluation of medical knowledge contained in Wikipedia and its use in the LOINC database. Journal of the American Medical Informatics Association, 17(3):283–287, May 2010.

Annotation: Describes a system for finding articles on Wikipedia that maps to parts in the LOINC database.

- [87] Adam R. Brown. Wikipedia as a data source for political scientists: Accuracy and completeness of coverage. PS: Political Science & Politics, 44(2):339–343, April 2011.
- [88] Joseph Michael Reagle, Jr. and Lauren Rhue. Gender bias in Wikipedia and Britannica. International Journal of Communication, 5:1138– 1158, 2011.

Annotation: Study on the possible gender bias in the the content of Wikipedia with examination of several thousand biographic articles in comparison to Britannica and other resources.

[89] Anna Samoilenko and Taha Yasseri. The distorted mirror of Wikipedia: a quantitative analysis of Wikipedia coverage of academics. EPJ Data Science, 3:1, 2013.

Annotation: Analysis of the coverage of scientists on Wikipedia. Wikipedia presence is compared with data from Scopus.

[90] PBS. 10 questions for Wikipedia founder Jimmy Wales. PNS NewsHour: The Rundown, July 2012

Annotation: Show news article with interview of Jimmy Wales.

[91] Pamela T. Johnson, Jennifer K. Chen, John Eng, Martin A. Makary, and Elliot K. Fishman. A comparison of World Wide Web resources for identifying medical information. *Academic Ra-diology*, 15(9):1165–1172, September 2008.

Annotation: A study with 135 medical students answering clinical questions with the help of Web resource. Google and other search engine were most effective. Sites such as Wikipedia, eMedicine and MDConsult less, though they sometimes provided the ultimate answer, and Wikipedia did that the most.

[92] Cindy Royal and Deepina Kapila. What's on Wikipedia, and what's not...? assessing completeness of information. Social Science Computer Review, 27(1):138–148, February 2009.

Annotation: Examines the length of sets of Wikipedia articles and compares them with a number of other variables: With year, Encyclopedia Britanica, country population and company revenue.

[93] Bill Wedemeyer. The quality of scientific articles on the English Wikipedia. Video, 2008. Presentation at Wikimania in Alexandria, Egypt.

Annotation: A study on a range of quality of scientific articles on the English Wikipedia along a number of dimensions, e.g., coverage, referencing, length, user perception.

[94] Aniket Kittur, Ed H. Chi, and Bongwon Suh. What's in Wikipedia? mapping topics and conflicts using socially annotated category structure. In CHI 2009, pages 1509–1512, New York, NY, USA, 2009. ACM.

Annotation: Describes an algorithm for determine the topic distribution of a Wikipedia article over top-level categories based on the Wikipedia category graph, and evaluate it against human labeled data. By applying the algorithm to the 2006 and 2008 data sets of the English Wikipedia the topic coverage over time is descriped, and with combining the method with a degree

of conflict page labeling the most contentious topics are found.

- [95] Aniket Kittur, Bongwon Suh, Bryan A. Pendleton, and Ed H. Chi. He says, she says: Conflict and coordination in Wikipedia. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 453–462, New York, NY, USA, April 2007. ACM.
- [96] Torie Bosch. How Kate Middleton's wedding gown demonstrates Wikipedia's woman problem. Slate, July 2012.
- [97] Stine Eckert and Linda Steiner. (re)triggering backlash: responses to news of wikipedia's gender gap. Journal of Communication Inquiry, 37(4):284–303, 2013.

Annotation: Analysis of how the gender gap of Wikipedia contributors was represented and discussed in media.

[98] Gregory Kohs. Wikipedia biographies favor men. examiner.com, January 2011.

Annotation: Small report on the proportion of males in Wikipedia biographies about living people.

[99] Max Klein. Sex ratios in Wikidata, Wikipedias, and VIAF. hangingtogether.org, May 2013.

> Annotation: Report of an analysis of the gender ratio on Wikipedias by using Wikidata material.

[100] Max Klein. Preliminary results from WIGI, the Wikipedia Gender Inequality Index. Notconfusing.com, January 2015.

Annotation: Blog post reporting a study with Piotr Konieczny on gender representation in Wikipedia.

- [101] Claudia Wagner, David Garcia, Mohsen Jadidi, and Markus Strohmaier. It's a man's Wikipedia? assessing gender inequality in an online encyclopedia. In International AAAI Conference on Weblogs and Social Media, North America, 2015.
- [102] Magnus Manske. Red vs. blue. The Whelming, January 2015.

Annotation: Blog post with an analysis of the representation of gender in Wikidata.

[103] Magnus Manske. Sex and artists. The Whelming, March 2015.

Annotation: Blog post with a Wikidata-based analysis of gender bias in Wikipedias.

[104] Han-Teng Liao. preelminary results from the Wikipedia Gender Inequality Index project - comments welcome. Wiki-research-l mailing list, January 2015.

ANNOTATION:

[105] Camilo Mora, Derek P. Tittensor, Sina Adl, Alastair G. B. Simpson, and Boris Worm. How many species are there on earth and in the ocean? PLoS Biology, 9(8):e1001127, August 2011.

Annotation: Estimates the number of species.

[106] American Chemical Society. CAS REG-ISTRY(sm) keeps pace with rapid growth of chemical research, registers 60 millionth substance. Chemical Abstracts Service Media Releases, May 2011.

Annotation: Press release stating that the chemical compound database has reached its 60 millionth entry.

[107] Tønnes Bekker-Nielsen. Historie på Wikipedia. Noter, 188:48–52, March 2011.

Annotation: Describes the quality of history articles on the Danish Wikipedia. A university teacher monitored articles created by his students as they were edited by other Wikipedians.

[108] Jon Fine. Sad news about Tim Russert broken by Wikipedia?. Fine on Media, June 2008. Blog on Business Week.

Annotation: Short comment on Wikipedia publishing news before it becomes published by general media.

- [109] Geoffrey Bilder. Many metrics. such data. wow.. CrossTech, 2014.
- [110] Donna Shaw. Wikipedia in the newsroom. American Journalism Review, 30(1):40–45, February/-March 2008.

Annotation: Describes by a few examples how Wikipedia is used as a source in news media and other domains. The article notes that editor at Philadelphia Inquirer warned journalist never to use Wikipedia "to verify facts or to augment information in a story". It mentions the WikiScanner and the case of Seigenthaler. A number of different people in the news business are interviewed expressing different opinion on the use of Wikipedia in news media, and studies by Nature and Roy Rosenzweig are described briefly.

[111] Andrew Lih. Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource. In 5th International Symposium on Online Journalism. The University of Texas at Austin, April 2004.

Annotation: A study examining the inbound citations to Wikipedia from press articles published in 2003 and parts of 2004. It displays "diversity" (number of individual editors of an article) against "rigor" (number of edits for an article) in some Wikipedia articles and the study regard these as indicators of quality. World War II, Islam and Astronomy scored high on these dimensions. The indicators of article quality increased after press citation for some Wikipedia articles.

[112] William Emigh and Susan C. Herring. Collaborative authoring on the web: A genre analysis of online encyclopedias. In Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05), page 99.1, Washington, DC, USA, 2005. IEEE Computer Society.

Annotation: Reports on a genre analysis of Wikipedia compared with Everything2 and Columbia Encyclopedia. A quantitative analysis of the formality of the text is performed by counting words related to formality or informality. They find that the style of Wikipedia articles is close to that of Columbia Encyclopedia.

[113] Malcolm Clark, Ian Ruthven, and Patrik O'Brian Holt. The evolution of genre in Wikipedia. Journal for Language Technology and Computational Linguistics, 24(1):1–22, 2009.

Annotation: A non-quantitative, descriptive study on a few Wikipedia articles in terms of structural form (i.e., genre) and its evolution as the articles are extended.

[114] Rui Lopes and Luís Carriço. On the credibility of Wikipedia: an accessibility perspective. In Second Workshop on Information Credibility on the Web (WICOW 2008), New York, 2008. ACM.

Annotation: A quantitative test of conformity with the W3C's Web Content Accessibility Guidelines of 100 Wikipedia articles and 265 articles on the Web referenced by Wikipedia. The 100 Wikipedia articles had on average a better score of accessibility.

- [115] Wendy Chisholm, Gregg Vanderheiden, and Ian Jacobs. Web content accessibility guidelines 1.0. W3c recommendation, The World Wide Web Consortium, March 1999.
- [116] S. B. Sinha and Harjit Singh Bedi. Commr. Of Customs, Bangalore vs. m/s ACER India Pvt. Ltd. The Judgement Information System, December 2007. Appeal (civil) 2321 of 2007.

Annotation: A case in the Supreme Court of India where Wikipedia is used for the definition of 'laptop'.

- [117] Noam Cohen. Courts turn to Wikipedia, but selectively. The New York Times, January 2007.
- [118] Isaac Arnsdorf. Judge uses wiki in case. Yale Daily News, October 2008.
- [119] Cass R. Sunstein. A brave new wikiworld. Washington Post, February 2007.
- [120] Diane Murley. In defense of Wikipedia. Law Library Journal, 100(3):593-599, 2008.

Annotation: Describes Wikipedia as a research tool for students and lawyers.

- [121] Jeff Breinholt. The wikipediazation of the American judiciary. The NEFA Foundation, January 2008
- [122] Morgan Michelle Stoddard. Judicial citation to Wikipedia in published federal court opinions. Master's thesis, School of Information and Library Science of the University of North Carolina at Chapel Hill, April 2009.
- [123] Lee F. Peoples. The citation of Wikipedia in judicial opinions. SSRN, 2009–2010.
- [124] Jens Ramskov. Wikipedia-grundlægger: Wikiopdateringer må ikke blive en gammelmandshobby. Ingeniøren, December 2013.

Annotation: Danish article with an interview of Jimmy Wales as the founder of Wikipedia.

- [125] Morten Rask. The reach and richness of Wikipedia: Is wikinomics only for rich countries. First Monday, 13(6), June 2008.
- [126] Brooke Gladstone. 10 years of Wikipedia. On The Media, January 2011.

Annotation: Interview with Executive Director of the Wikimedia Foundation Sue Gardner.

- [127] J. P. Shim and Jeongwon Yang. Why is Wikipedia not more widely accepted in Korea and China? factors affecting knowledge-sharing adoption. *De*cision Line, 40(2):12–15, March 2009.
- [128] Tamim Elyan. Arabic content in Wikipedia very weak, users to blame. Daily News Egypt, September 2008. Assessed 2010-07-28.
- [129] Noam Cohen. A Wikipedian challenge: Convincing Arabic speakers to write in Arabic. The New York Times, July 2008.

Annotation: A news article on the issue of why the Arabic Wikipedia is relatively small.

[130] Xan Rice. Internet: Last piece of fibre-optic jigsaw falls into place as cable links east Africa to grid. guardian.co.uk, August 2008. Annotation: News article about a new submarine cabel for communication for the poorly connected east

[131] Lennart Guldbrandsson. Swedish Wikipedia surpasses 1 million articles with aid of article creation bot. Wikimedia blog, June 2013.

Annotation: Blog article reporting about the work of a computer program, a bot, creating many articles on species on the Swedish Wikipedia.

[132] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. Nature, 393(6684):397–498, June 1998.

Annotation: Describes so-called 'small-world' networks which can be constructed as interpolations between regular (lattice) and random networks. They exhibits short averages paths while still being clustered. Examples are given by film actors, power grid and the neural network of a worm together with dynamic networks, e.g, multiplayer prisoner's dilemma.

- [133] Albert-Lászlo-Barabási and Réka Albert. Emergence of scaling in random networks. Science, 286:509–512, October 1999.
- [134] Gabriel Pinski and Francis Narin. Citation influence for journal aggregates of scientific publications: Theory with application to literature of physics. *Information Processing & Management*, 12:297–312, 1976.
- [135] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999.
- [136] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, September 1999.
- [137] Francesco Bellomi and Roberto Bonato. Network analysis of Wikipedia. In Proceedings of Wikimania 2005 — The First International Wikimedia Conference, 2005.

Annotation: Describes results of application of PageRank and Kleinberg's HITS algorithm on the English Wikipedia corpus. "United States" scored highest in both. Entries related to religion scored high PageRank.

[138] David Laniado and Riccardo Tasso. Coauthorship 2.0: patterns of collaboration in Wikipedia. In Proceedings of the 22nd ACM conference on Hypertext and Hypermedia. ACM, June 2011. Annotation: Network analysis of the coauthorship network in English Wikipedia, where the main contributors to a Wikipedia article is identified with a method of Adler et al. and various network metrics are extracted across time and topics.

[139] V. Zlatić, M. Božičević, H. Štefančić, and M. Domazet. Wikipedias: Collaborative web-based encyclopedias as complex networks. *Physical Review E*, 74(1):016115, July 2006.

Annotation: Network analysis of intrawiki links on a number of different language versions of Wikipedia and report a range of network characteristics

[140] Torsten Zesch and Iryna Gurevych. Analysis of the Wikipedia category graph for NLP applications. In TextGraphs-2: Graph-Based Algorithms for Natural Language Processing. Proceedings of the Workshop, pages 1–8, New Brunswick, New Jersey, USA, April 2007. The Association for Computational Linguistics.

Annotation: Description of the use of the Wikipedia category graph for determining semantic relatedness. Several different similarity and distance measures on the graph are examined on several human labeled datasets from the German Wikipedia. Graph characteristics (average shortest path, cluster coefficient and power law exponent) are also shown.

[141] David Laniado, Riccardo Tasso, Yana Volkovich, and Andreas Kaltenbrunner. When the Wikipedians talk: network and tree structure of wikipedia discussion pages. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. AAAI, July 2011.

Annotation: Network analysis of the user interaction on the article and user talk pages of the English Wikipedia.

[142] A. Capocci, V. D. P. Servedio, F. Colaiori, L. S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli. Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Physical Review E*, 74:036116, September 2006.

Annotation: Network analysis of the intrawiki links on the English Wikipedia.

[143] Todd Holloway, Miran Božičević, and Katy Börner. Analyzing and visualizing the semantic coverage of Wikipedia and its authors. *Complexity*, 12(3):30–40, January 2007.

Annotation: Analysis of various aspects of early Wikipedia, especially the number of articles and categories.

- [144] A. Capocci, F. Rao, and G. Caldarelli. Taxonomy and clustering in collaborative systems: The case of the on-line encyclopedia Wikipedia. EPL, 81(2):28006, January 2008.
- [145] David M. D. Smith, Jukka-Pekka Onnela, and Neil F. Johnson. Accelerating networks. New Journal of Physics, 9(181), 2007.

Annotation: Discuss accelerating networks with network analysis of the number of nodes and links, exemplifying it on three different language versions of Wikipedia.

[146] Wray Buntine. Static ranking of web pages, and related ideas. In Michel Beigbeder and Wai Gen Yee, editors, Open Source Web Information Retrieval, pages 23–26, October 2005.

Annotation: A multivariate analysis of the wikilink graph of the English 2005 Wikipedia with around 500'000 pages with a discrete version of the hubs and authority algorithm.

[147] A. O. Zhirov, O. V. Zhirov, and D. L. Shepelyansky. Two-dimensional ranking of Wikipedia articles. arXiv, September 2010.

Annotation: Examines the PageRank and the 'reverse' PageRank (CheiRank) for links within Wikipedia as well as a combined rank they call 2DRank.

[148] Young-Ho Eom, Klaus M. Frahm, András Benczúr, and Dima L. Shepelyansky. Time evolution of wikipedia network ranking. The European Physical Journal B, 86(12):492, December 2013.

Annotation: Analysis of intrawiki links on Wikipedia across several years.

[149] Pablo Aragon, David Laniado, Andreas Kaltenbrunner, and Yana Volkovich. Biographical social networks on Wikipedia: a cross-cultural study of links that made history. In 8th International Symposium on Wikis and Open Collaboration, 2012.

> Annotation: Network analysis of the bibliographic pages across the 15 largest Wikipedias.

[150] Sofia J. Athenikos and Xia Lin. The WikiPhil Portal: visualizing meaningful philosophical connections. Journal of the Chicago Colloquium on Digital Humanities and Computer Science, 1(1), 2009. Annotation: Network analysis and visualization of the link network among Western philosphers described on Wikipedia.

[151] Radim Řehůřek. Fast and faster: a comparison of two streamed matrix decomposition algorithms. In NIPS Workshop on Low-Rank Methods for Large-Scale Machine Learning, 2010.

Annotation: Describes algorithms for large-scale eigen decomposition and applies them on a large document-term matrix constructed from the English Wikipedia.

[152] Choochart Haruechaiyasak and Chaianun Damrongrat. Article recommendation based on topic model for Wikipedia selection for schools. In Digital Libraries: Universal and Ubiquitous Access to Information, volume 5362 of Lecture Notes in Computer Science, pages 339–342, Berlin, 2008. Springer.

Annotation: Brief article describing topic mining with latent Dirichlet allocation using LingPipe on Wikipedia text as available in the Wikipedia for Schools dataset.

[153] Robert P. Biuk-Aghai. Visualizing co-authorship networks in online Wikipedia. In International Symposium on Communications and Information Technologies, 2006. ISCIT '06, pages 737–742, 2006.

Annotation: Visualization of relationships between Wikipedia pages based on co-authorship patterns.

[154] Finn Årup Nielsen. Wiki(pedia) and neuroinformatics. Workshop on Wikipedia Research 2006, August 2006.

Annotation: Slides about neuroinformatics and data clustering of Wikipedia content.

[155] Rut Jesus, Martin Schwartz, and Sune Lehmann. Bipartite networks of Wikipedia's articles and authors: a meso-level approach. In Proceedings of the 5th International Symposium on Wikis and Open Collaboration, page 5. Association for Computing Machinery, 2009.

> Annotation: Study with an application of an algorithm for identification of maximal bicliques on the network of articles and editors on the English Wikipedia.

[156] Bernie Hogan, Mark Graham, and Ahmed Medhat Mohamed. The vocal minority: local self-representation and co-editing on Wikipedia in the Middle East and North Africa. 2012.

Annotation: Analysis of geotagged Wikipedia articles about topics in the with Middle East and North Africa (MENA) region and established the co-editing social network. By examining the user page they tried to determine if the user was from a specific country. They found that most editors were from the West rather than MENA countries.

[157] Takashi Iba, Keiichi Nemoto, Bernd Peters, and Peter A. Gloor. Analyzing the creative editing behavior of Wikipedia editors: through dynamic social network analysis. Procedia - Social and Behavioral Sciences, 2(4):6441–6456, 2010.

Annotation: Describes a social network analysis by looking at coauthor graph in Wikipedia from sequential editing.

- [158] Brian Keegan, Darren Gergle, and Noshir Contractor. Staying in the loop: structure and dynamics of Wikipedia's breaking news collaborations. In WikiSym 2012, 2012.
- [159] Finn Årup Nielsen. Sequential collaboration network with sentiment coloring. In NetSci, 2013. Abstract.

Annotation: Short description of a network visualization of edits in Wikipedia including sentiment analysis of the edits.

[160] Finn Årup Nielsen, Michael Etter, and Lars Kai Hansen. Real-time monitoring of sentiment in business related Wikipedia articles. April 2013. Submitted.

> Annotation: Description of an online service with sentiment analysis of Wikipedia edits for monitoring business-related Wikipedia articles and display of the results with network visualizations.

- [161] Alex Halavais. The Isuzu experiment. a thaumaturgical compendium, August 2004. Blog.
- [162] How authoritative is Wikipedia. Dispatches from the Frozen North, September 2004. Blog.
- [163] P. D. Magnus. Fibs in the Wikipedia, December 2007.
- [164] P. D. Magnus. Early response to false claims in Wikipedia. *First Monday*, 13(9), September 2008.

Annotation: Report on a study that injected false fact in Wikipedia. Of 36 errors only 15 was removed within 48 hours.

[165] Jens Lenler. Verdens største leksikon: ...og største fejlrisiko. Politiken, page 1, December 2006. Kultur section. Annotation: Danish newspaper article about the Æblerød Wikipedia hoay

- [166] John Seigenthaler. A false Wikipedia 'biography'. USATODAY.com, 2005.
- [167] Bertrand Meyer. Defense and illustration of Wikipedia. EiffelWorld Column, January 2006.

Annotation: A generally positive comment on Wikipedia with some anecdotal evidence, especially about the vandalism on the authors bibliography.

[168] Kevin Morris. After a half-decade, massive Wikipedia hoax finally exposed. Daily Dot, January 2013.

Annotation: News report on a Wikipedia hoax.

[169] Yoni Appelbaum. How the professor who fooled wikipedia got caught by reddit. The Atlantic, May 2012.

Annotation: Reports on a teacher at that lets undergraduates fabricate history as part of a course and add the information to Wikipedia, e.g., the Edward Owens hoax.

[170] Kevin Morris. How vandals are destroying Wikipedia from the inside. The Daily Dot, January 2013.

Annotation: News report on the Legolas2186 hoax case in Wikipedia.

- [171] Fernanda B. Viégas, Marting Wattenberg, and Kushal Dave. Studying cooperation and conflict between authors with history flow visualizations. In Elizabeth Dykstra-Erickson and Manfred Tscheligi, editors, Proceedings of the SIGCHI conference on Human factors in computing systems, pages 575–582, New York, NY, USA, 2004. ACM.
- [172] Reid Priedhorsky, Jilin Chen, Shyong (Tony) K. Lam, Katherine Panciera, Loren Terveen, and John Riedl. Creating, destroying, and restoring value in Wikipedia. In Proceedings of the 2007 International ACM Conference on Supporting Group Work, pages 259–268, New York, NY, USA, 2007. ACM.
- [173] Paul P. Gardner, Jennifer Daub, John Tate, Benjamin L. Moore, Isabelle H. Osuch, Sam Griffiths-Jones, Robert D. Finn, Eric P. Nawrocki, Diana L. Kolbe, Sean R. Eddy, and Alex Bateman. Rfam: Wikipedia, clans and the "decimal" release. Nucleic Acids Research, 39:D141–D145, November 2010.

Annotation: Update on the Rfam database including its connection to Wikipedia.

- [174] Michael Lorenzen. Vandals, administrators, and sockpuppets, oh my! an ethnographic study of Wikipedia's handling of problem behavior. *MLA forum*, 5(2), December 2006.
- [175] Luciana S. Buriol, Carlos Castillo, Debora Donato, Stefano Leonardi, and Stefano Millozzi. Temporal analysis of the wikigraph. In Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, pages 45–51, Washington, DC, USA, 2006. IEEE Computer Society.
- [176] Jacobi Carter. ClueBot and vandalism on Wikipedia, 2007.
- [177] Martin Potthast, Benno Stein, and Robert Gerling. Automatic vandalism detection in Wikipedia. In Advances in Information Retrieval, volume 4956 of Lecture Notes in Computer Science, pages 663–668, Berlin/Heidelberg, 2008. Springer.

Annotation: Reports a machine learning approach for vandalism detection in Wikipedia where features such as 'upper case ratio', 'size ratio' and 'edits per user' are used. The system performed well compared to rule-based vandalism detection bots on Wikipedia.

[178] Koen Smets, Bart Goethals, and Brigitte Verdonk. Automatic vandalism detection in Wikipedia: Towards a machine learning approach. Benelearn08, The Annual Belgian-Dutch Machine Learning Conference, July 2008.

Annotation: Short abstract that reports an machine learning approach to vandalism detection in Wikipedia. With naive Bayes, bag-of-words and probabilistic sequence modeling results better than the rule-based Clue-Bot was not obtained.

[179] Andrew G. West, Sampath Kannan, and Insup Lee. Spatio-temporal analysis of Wikipedia metadata and the STiki anti-vandalism tool. In Proceedings of the 6th International Symposium on Wikis and Open Collaboration, page 18, New York, NY, USA, 2010. ACM.

Annotation: Describes the STiki system for semi-automated vandalism detection on Wikipedia using metadata from the revisions.

[180] Andrew G. West, Sampath Kannan, and Insup Lee. STiki: an anti-vandalism tool for Wikipedia using spatio-temporal analysis of revision metadata. In Proceedings of the 6th International Symposium on Wikis and Open Collaboration, New York, NY, USA, 2010. ACM. Annotation: More or less the same paper as 'Spatio-temporal analysis of Wikipedia metadata and the STiki anti-vandalism tool'

[181] R. Stuart Geiger and David Ribes. The work of sustaining order in Wikipedia: the banning of a vandal. In Proceedings of the 2010 ACM conference on Computer supported cooperative work, pages 117–126, New York, NY, USA, 2010. Association for Computing Machinery.

Annotation: Describes the technical systems around Wikipedia that helps in vandal fighting, describes the Huggle assisted editing tools and gives a narrative about reverting and blocking a vandal.

- [182] Eric Goldman. Wikipedia will fail within 5 years. Technology & Marketing Law Blog, December 2005.
- [183] Thomas Claburn. Law professor predicts Wikipedia's demise. Information Week, December 2006.

Annotation: News story about a blog post from Professor Eric Goldman, that predicted the failure of Wikipedia to counter a massive attack by marketeers.

[184] Brian Bergstein. Microsoft offers cash for Wikipedia edit. Washingtonpost.com, January 2007.

Annotation: News article about Microsoft offering to pay blogger Rick Jelliffe to change Wikipedia articles as the company was sure that it contained inaccuracies about an open-source document standard and the Microsoft rival format.

[185] Roger Cadenhead. Adam Curry caught in sticky wiki. Internet, December 2005.

Annotation: Blog on how an a person from an IP address associated with podcasting entrepreneur Adam Curry has edited in the "Podcasting" Wikipedia article.

[186] Emily Biuso. Wikiscanning. New York Times, December 2007.

Annotation: News article about Virgil Griffith's Wikipedia mining tool Wikiscanner.

- [187] Jonathan Fildes. Wikipedia 'shows CIA page edits'. BBC NEWS, August 2007.
- [188] Brian Brady. BBC staff rewrote Wikipedia pages to water down criticism. The Independent, December 2007.

[189] Maryam Omidi. PR firm accused of whitewashing Maldives entry on Wikipedia. Minivan News, September 2009.

Annotation: News story about a biased Wikipedia edits originating from a PR firm employed by the Maldives.

- [190] Associated Press. Ministry bans Wikipedia editing. Guardian Unlimited, November 2007.
- [191] Asger Westh. Wikipedia: Den elektroniske slagmark. epn.dk, January 2008.

Annotation: A news article about a version of the WikiScanner for the Danish Wikipedia.

[192] Daniel Erenrich. Wikiscanner: Automated conflict of interest detection of anonymous Wikipedia edits. In Student-Faculty Programs: 2008 Abstract Book, page 22. California Institute of Technology, 2008.

Annotation: Student abstract about the Wikiscanner, that detects conflict of interest edits in Wikipedia also using Ip2location and USPTO trademark databases and computing link distance between pages and categories.

- [193] Cory Doctorow. Official City of Melbourne IP address used for biased edits to Wikipedia page for Occupy Melbourne prior to local election. Boing Boing, February 2013.
- [194] Andrew Lih. Long live Wikipedia? Sustainable volunteerism and the future of crowd-sourced knowledge. In John Hartley, Jean Burgess, and Axel Bruns, editors, A Companion to New Media Dynamics, pages 185+. Blackwell publishing, first edition, 2013.
- [195] Figaro. Condamné pour une correction Wikipedia. Figaro, July 2011.
- [196] Ryan McGrady. Gaming against the greater good. First Monday, 14(2), February 2009.

Annotation: A discussion of authority, rhetoric on Wikipedia and wikilawyering where Wikipedia policy is misused by editors. The report of an interview with a single editor shows the problem with abuse of policy in a case of an sock puppet account.

[197] Simon Owen. The battle to destroy Wikipedia's biggest sockpuppet army. The Daily Dot, October 2013.

Annotation: Technical news article about a sockpuppet investigation and paid editing on Wikipedia.

[198] Rishi Chandy. Wikiganda: Identifying propaganda through text analysis. Caltech Undergraduate Research Journal, 9(1):6–11, Winter 2008–2009.

Annotation: Description of a opinion mining system for Wikipedia edits.

- [199] Violet Blue. Corruption in wikiland? Paid PR scandal erupts at Wikipedia. CNET News, September 2012.
- [200] Marcia W. DiStaso. Measuring public relations Wikipedia engagement: how bright is the rule?. Public Relations Journal, 6(2), 2012.

Annotation: Reports on a survey among 1'284 public relations/communications professionals about their perceptions and practice with Wikipedia.

- [201] David King. Ethical Wikipedia strategies for brands. CW Bulletin, 2012.
- [202] Amanda Filipacchi. Wikipedia's sexism toward female novelists. The New York Times, April 2013.

Annotation: Op-ed on Wikipedia subcategorizing female but not male American Novelists.

[203] Emma Gray. Women novelists Wikipedia: Female authors absent from site's 'American Novelists' page?. The Huffington Post, April 2013.

Annotation: News article on Wikipedia subcategorizing female but not male American Novelists.

[204] Bongwon Suh, Ed H. Chi, Aniket Kittur, and Bryan A. Pendleton. Lifting the veil: Improving accountability and social transparency in Wikipedia with WikiDashboard. In CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, pages 1037–1040, New York, NY, USA, 2008. ACM.

Annotation: Description of the WikiDashboard web-site that makes a visualization of author contributions of a Wikipedia article and embeds it in a proxy copy of the Wikipedia article. Visualization may also be generated with respect to a user. Social transparency in relation to Wikipedia is also discussed.

[205] Sara Javanmardi, Cristina Lopes, and Pierre Baldi. Modeling user reputation in wikis. Statistical Analysis and Data Mining, 3(2):126–139, April 2010.

> Annotation: Describes a model for time-depedent user reputation on wikis and applies it on the entire revision history of the English Wikipedia.

[206] Jimmy Wales. Wikipedia, emergence, and the wisdom of crowds. wikipedia-l mailing list, May 2005. Accessed 2009 February 8.

- [207] B. Thomas Adler, Luca de Alfaro, Ian Pye, and Vishwanath Raman. Measuring author contribution to the Wikipedia. In Proceedings of WikiSym 2008. ACM, 2008.
- [208] Felipe Ortega, Jesús M. González-Barahona, and Gregorio Robles. On the inequality of contributions to Wikipedia. In Proceedings of the 41st Hawaiian International Conference on System Sciences (HICSS-2008), 2008.

Annotation: Computation of the Gini coefficient for author contributions for the top ten largest Wikipedias. They find the Gini coefficients to be from 0.92 to 0.97.

[209] Aaron Swartz. Who writes Wikipedia. Rav Thought, September 2006. Blog.

Annotation: A report on a small investigation on the different amount of contribution authors make to Wikipedia. Contrary to previous thought the examination of a few Wikipedia articles shows that essential content was mostly written by occasional contributors.

[210] R. Stuart Geiger. The social roles of bots and assisted editing programs. In Proceedings of the 5th International Symposium on Wikis and Open Collaboration, page 30, New York, NY, USA, 2009. Association for Computing Machinery.

Annotation: A study on the amount of edits on the English Wikipedia in 2009 made by bots or through assisted editing tools. Combined they make up 28.49% of the edits, — a number slightly larger than the number of edits performed by anonymous users.

[211] Jochen Rick, Mark Guzdial, Karen Carroll, Lissa Holloway-Attaway, and Brandy Walker. Collaborative learning at low cost: CoWeb use in English composition. In Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community, pages 435–442. International Society of the Learning Sciences, 2002.

Annotation: Describes experience with the CoWeb wiki used for teaching English. Students using the wiki had better attitude towards collaboration and got higher grades compared to another online environment.

[212] Andrea Forte and Amy Bruckman. From Wikipedia to the classroom: exploring online publication and learning. In *Proceedings of the 7th international conference on Learning sciences*, pages 182–188. International Society of the Learning Sciences, 2006.

Annotation: Describes experience with having fresman-level college students write and interact in a wiki where the students can comment of each others articles.

[213] Wikimedia Foundation. Wikipedia editors study: Results from the editor survey, april 2011. Wikimedia Commons, August 2011.

Annotation: Report on a survey among several thousand Wikipedia editors.

[214] Nuša Farič and Henry W. W. Potts. Motivations for contributing to health-related articles on Wikipedia: An interview study. In *Medicine* 2.0'13, September, 2013.

Annotation: Report on an interview study on 32 Wikipedians contributing to Wikipedia health-related articles.

- [215] Tilman Bayer. How many women edit wikipedia?. Wikimedia Blog, April 2015.
- [216] Ruediger Glott, Philipp Schmidt, and Rishab Ghosh. Wikipedia survey — overview of results. Report, UNU-MERIT, United Nations University, Maastricht, Netherlands, March 2010.
- [217] Shyong (Tony) K. Lam, Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R. Musicant, Loren Terveen, and John Riedl. Wp:clubhouse? an exploration of Wikipedia's gender imbalance. In Proceedings of the 7th International Symposium on Wikis and Open Collaboration, New York, NY, USA, 2011. Association for Computing Machinery.
- [218] Benjamin Mako Hill and Aaron Shaw. The Wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PloS one*, 8(6):e65782, 2013.
- [219] Noam Cohen. Define gender gap? look up Wikipedia's contributor list. The New York Times, January 2011.

Annotation: Reports on the gender gap among Wikipedia contributors and discusses its influence on the topic emphasis on Wikipedia.

[220] Rishab A. Ghosh, Ruediger Glott, Bernhard Krieger, and Gregorio Robles. FLOSS. deliverable D18: Final report. part 4: Survey of developers. International Institute of Infonomics, University of Maastricht, The Netherlands, June 2002.

Annotation: Reports on a survey of FLOSS developers.

[221] Vijay A. Mittal, Kevin D. Tessner, and Elaine f. Walker. Elevated social internet use and schizotypal personality disorder in adolescents. *Schizophrenia Research*, 94(1-3):50–57, August 2007.

[222] Yair Amichai-Hamburger, Naama Lamdan, Rinat Madiel, and Tsahi Hayat. Personality characteristics of wikipedia members. *CyberPsychology & Behavior*, 11(6):679–681, December 2008.

Annotation: Application of the Big Five Inventory and Real-Me personality questionnaires to 139 Wikipedia and non-Wikipedia users. The recruitment was based on targeting posting of links. Wikipedians scored lower on agreeableness and higher on openness. Differences in extroversion and conscientiousness depended on the sex of the subject.

[223] Ralph Straumann and Mark Graham. The geographically uneven coverage of Wikipedia. Information Geographies at the Oxford Internet Institute, 2014.

> Annotation: Analysis of the geotagged articles in 44 different language versions of Wikipedia.

- [224] Seth Finkelstein. Wikipedia's school for scandal has plenty more secrets to reveal. The Guardian, March 2008.
- [225] Andrew George. Avoiding tragedy in the wikicommons. Virginia Journal of Law and Technology, 12(8):1–42, 2007.
- [226] Reijo Kupiainen, Juha Suoranta, and Tere Vaden. Fire next time: or revisioning higher education in the context of digital social creativity. E-Learning and Digital Media, 4(2):128–137, 2007.

Annotation: Discuss some aspects of digial media and higher eductions.

- [227] Pattarawan Prasarnphanich and Christian Wagner. Explaining the sustainability of digital ecosystems based on the wiki model through critical mass theory. *IEEE Transactions on Industrial Electronics*, 58(6):2065–2072, June 2011.
- [228] Christian Wagner. Breaking the knowledge acquisition bottleneck through conversational knowledge management. Information Resources Management Journal, 19(1):70–83, January 2005.
- [229] Paolo Magrassi. Free and open-source software is not an emerging property but rather the result of studied design. arXiv, November 2010.

Annotation: Discusses free and open-source software and argues for a role of top-down coordination.

- [230] Eric S. Raymond. The cathedral and the bazaar. First Monday, 3(3), March 1998.
- [231] Kevin Crowston and James Howison. The social structure of free and open source software development. First Monday, 10(2), February 2005.

Annotation: An analysis of 140 Free and Open Source Software project hosted on SourceForge with a study of contributor interaction network.

[232] Peter Miller. Smart swarm: Using animal behaviour to change our world. Collins, London, 2010.

Annotation: Book about mass collaboration among animals and humans

[233] Joseph Reagle. Four short stories about the reference work, 2005.

Annotation: Discussion of history of reference work production.

[234] Mark Elliott. Stigmergic collaboration: the evolution of group work. M/C Journal, 9(2), May 2006.

Annotation: Discuss the idea of stigmergy in relation to Wikipedia.

- [235] Jeff Loveland and Joseph Reagle. Wikipedia and encyclopedic production. New Media & Society, January 2013.
- [236] Dan Wielsch. Governance of massive multiauthor collaboration Linux, Wikipedia, and other networks: governed by bilateral contracts, partnerships, or something in between. *JIPITEC*, 1(2):96–108, July 2010.
- [237] Martin Pekárek and Stefanie Pötzsch. A comparison of privacy issues in collaborative workspaces and social networks. *Identity in the Information Society*, 2(1):81–93, December 2009.
- [238] comScore, Reston. Media Alert: comScore Releases Worldwide Ranking of Top Web Properties, October 2006.

Annotation: Statistics on the worlds most visited Web sites during September 2006.

- [239] comScore, Reston. Ghosts and Goblins Drive Traffic to Greetings, Gift and Toys Sites in October, November 2006.
- [240] Jon W. Huss, III, Pierre Lindenbaum, Michael Martone, Donabel Roberts, Angel Pizarro, Faramarz Valafar, John B. Hogenesch, and Andrew I. Su. The Gene Wiki: community intelligence applied to human gene annotation. Nucleic Acids Research, 38:D633–D639, 2010.
- [241] Bruce Etling, John Kelly, Robert Faris, and John Palfrey. Mapping the Arabic blogosphere: Politics, culture and dissent. Research Publication 2009-06, Berkman Center, June 2009.

Annotation: A study of 35'000 active Arabic language blogs.

[242] Michaël R. Laurent and Tim J. Vickers. Seeking health information online: does Wikipedia matter? Journal of the American Medical Informatics Association, 16(4):471–479, July-August 2009.

Annotation: Describes an investigation of search engine ranking for health topics in Wikipedia with search engine optimization methods. With queries selected from MedlinePlus, NHS and NORD they show that the English Wikipedia is more often on the first place for the selected queries on Google as compared to .gov domain, MedilinePlus, Medscape, NHS Direct Online and a number of other domains. The also investigated health-related topics with seasonal effects.

[243] Marcia W. DiStaso and Marcus Messner. Forced transparency: corporate image on Wikipedia and what it means for public relations. *Public Rela*tions Journal, 4(2), Spring 2010.

Annotation: Content analysis of Wikipedia articles on companies as well as examination of Internet search engine ranking. 10 Fortune 500 companies were selected in the years 2006, 2008 and 2010 and examined for topic, tonality and length, number of edits, etc.

[244] Nathan Safran. Wikipedia is the SERPs: Appears on page 1 for 60% of information, 34% trransactional queries. Conductor Blog, March 2012.

Annotation: Reports a non-peerreviewed study on the the position of Wikipedia in Internet search engine results. Using 2'000 keywords from representative search engine queries it was found that the appearance of Wikipedia on page one depended on whether the query was information or transactional and on how many words were in the query.

[245] Sam Silverwood-Cope. Wikipedia: Page one of Google UK for 99% of searches. Intelligent Positioning, February 2012.

Annotation: Non-peer-reviewed study on the search engine ranking of Wikipedia. Wikipedia appear on Google's page one based on 1'000 random nouns queried.

- [246] American Customer Satisfaction Index. July 2011 and historical ACSI scores, July 2011.
- [247] American Customer Satisfaction Index. Benchmarks by industry: Internet social media, July 2013.

Annotation: ASCI benchmarks from 2010 to 2013 for websites such as Wikipedia, Pinterest, YouTube and Myspace.

[248] Klaus Larsen. Læger er glade for Wikipedia. Uqeskrift for Læger, May 2012.

Annotation: News report on a survey conducted by the Mannov company on the use of social media in professional contexts of Danish medical doctors.

[249] Mannov. Danske læger er vilde med sociale medier. Internet, May 2012.

Annotation: Report of Danish medical doctors use of social media for professional activities.

[250] Andrew G. West and Milowent. Examining the popularity of Wikipedia articles: catalysts, trends, and applications. The Signpost, February 2013.

Annotation: Analysis of Wikipedia page view statistics.

- [251] Andrew G. West, Jian Chang, Krishna Venkatasubramanian, Oleg Sokolsky, and Insup Lee. Link spamming Wikipedia for profit. In Proc. of the 8th Annual Collaboration, Electronic Messaging, Anti-Abuse, and Spam Conference, pages 152– 161, New York, NY, USA, 2011. Association for Computing Machinery.
- [252] Jonathan Band and Jonathan Gerafi. Wikipedia's economic value. SSRN, October 2013.

Annotation: Review of studies on the economic value of Wikipedia and some further estimates on replacement cost and consumer value.

- [253] Anselm Spoerri. What is popular on Wikipedia and why?. First Monday, 12(4), April 2007.
- [254] Sook Lim and Nahyun Kwon. Gender differences in information behavior concerning Wikipedia, an unorthodox information source. Library & Information Science Research, 32(3):212–220, July 2010.

Annotation: Studies gender difference in the use of Wikipedia.

- [255] R. Stuart Geiger and Aaron Halfaker. Using edit sessions to measure participation in Wikipedia. In Proceedings of the 2013 conference on Computer supported cooperative work, New York, NY, USA, 2013. Associations for Computing Machinery.
- [256] The ed17 and Tony1. WMF employee forced out over "paid advocacy editing". Wikipedia Signpost, January 2014.

Annotation: Blog post on paid editing on Wikipedia by a Wikimedia Foundation employee.

[257] Lynn Wasnak. How much should I charge?. In 2006 Writer's Market, Writer's Market, pages 68– 78. Writers Digest Books, 2006.

Annotation: Overview of freelance writer's word, page and project rates for different kinds of work.

[258] Susan L. Bryant, Andrea Forte, and Amy Bruckman. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work, pages 1–10, New York, NY, USA, 2005. ACM.

Annotation: Describes interviews with 9 Wikipedia contributors and some of their characteristics: Most of the contributors tell that their initial edit was for correcting a problem or extending a weak article. As novices they were not aware of the Wikipedia community. As more experienced contributors they get a sense of community and decrease article writing and increase administration.

[259] Timme Bisgaard Munk. Why Wikipedia: self-efficacy and self-esteem in a knowledge-political battle for an egalitarian epistemology. *Observatorio*, 3(4), 2009.

Annotation: Semi-structured qualitative interview with six Danish Wikipedia volunteers for investigating the movitation for contribution. The interviews were transcribed 'categorized and analyzed thematically'.

- [260] Xiaoquan Zhang and Feng Zhu. Intrinsic motivation of open content contributors: the case of Wikipedia. In Workshop on Information Systems and Economics, 2006.
- [261] Heng-Li Yang and Cheng-Yu Lai. Motivations of Wikipedia content contributors. Computers in Human Behavior, 26(6):1377–1383, November 2010.

Annotation: Reports on a survey among Wikipedia contributors about their motivation for sharing knowledge.

[262] Michael Restivo and Arnout van de Rijt. Experimental study of informal rewards in peer production. *PLoS ONE*, 7(3):e34358, 2012.

Annotation: An experiment where Wikipedia editors were given informal awards to see how it affected their productivity.

[263] Haiyi Zhu, Amy Zhang, Jiping He, Robert E. Kraut, and Aniket Kittur. Effects of peer feedback on contribution: a field experiment in Wikipedia. In Wendy E. Mackay, Stephen Brewster, and Susanne Bødker, editors, *CHI 2013*, pages 2253–2262, New York, NY, USA, 2013. Association for Computing Machinery.

- [264] Andrea Ciffolilli. Phantom authority, self-selective recruitment and retention of members in virtual communities: The case of Wikipedia. First Monday, 8(12), December 2003.
- [265] Guillaume Paumier. VisualEditor. MediaWiki wiki, July 2011.

Annotation: Early notes about the VisualEditor project for a WYSIWYG editor in the MediaWiki software.

[266] Martin Robbins. Is the PR industry buying influence over Wikipedia?. VICE, October 2013.

Annotation: News report on the controversial Wikipedia editing of the company Wiki-PR.

[267] Jonathan Corbet, Greg Kroah-Hartman, and Amanda McPherson. Linux kernel development: How fast it is going, who is doing it, what they are doing, and who is sponsoring it. Linux Foundation, September 2013.

Annotation: Statistics on Linux Kernel development.

[268] Eric Goldman. Wikipedia's labor squeeze and its consequences. Journal on Telecommunications and High Technology Law, 8(1):157–183, Winter 2010.

Annotation: Discussion of (possible future) sustainability problems for Wikipedia and what can be done to counter them.

[269] Magnus Cedergren. Open content and value creation. First Monday, 8(8), August 2003.

Annotation: Study on three open content projects: Wikipedia, Open Directory Project and the Prelinger Archives with a discussion of the motivation of contributors.

[270] Oded Nov. What motivates Wikipedians?. Communications of the ACM, 50(11):60-64, November 2007.

Annotation: Reports on an analysis of a survey among Wikipedia contributors for investigation of their motivation as contributors.

[271] Bernhard Hoisl, Wolfgang Aigner, and Silvia Miksch. Social rewarding in wiki systems — motivating the community. In Online Communities and Social Computing, volume 4564 of Lecture Notes in Computer Science, pages 362–371, Berlin/Heidelberg, August 2007. Springer.

Annotation: A suggestion for a computation of reputation of the individual authors in a wiki-system. Formulas are shown but it is not clear if the suggestion is tested on a real wiki.

[272] Arto Lanamäki, Mikko Rajanen, Anssi Öörni, and Netta Iivari. Once you step over the first line, you become sensitized to the next: towards a gateway theory of online participation. In Proceedings of the International Conference on Information Systems - Exploring the Information Frontier, ICIS 2015, Fort Worth, Texas, USA, December 13-16, 2015, 2015.

Annotation: Propose to view online participation under the framework of gateway theory as in contrast to antecedents-based explanations.

[273] Andrea Forte and Amy Bruckman. Why do people write for Wikipedia? incentives to contribute to open-content publishing. In Group 2005 workshop: Sustaining community: The role and design of incentive mechanisms in online systems. Sanibel Island, FL., November 2005.

Annotation: From interviews of 22 Wikipedia contributors to understand their motivation the study argues that the incentive system in Wikipedia resembles that of science.

[274] Seth Finkelstein. What the New Yorker article fraud tells us about Wikipedia. Infothought, March 2007.

Annotation: Blogpost which reflects on the Essjay controversy in Wikipedia.

- [275] Travis Kriplean, Ivan Beschastnikh, and David W. McDonald. Articulations of wikiwork: uncovering valued work in Wikipedia through barnstars. In Proceedings of the 2008 ACM conference on Computer supported cooperative work, pages 47–56, New York, NY, USA, 2008. ACM.
- [276] Aaron Shaw and Benjamin Mako Hill. Can social awards create better wikis?. Wikimania 2012, March 2012.

Annotation: Analysis of the effect of social awards on productivity in Amazon Mechanical Turk, the programming language Scratch and in Wikipedia.

[277] Richard Fisher. Just can't get e-nough. New Scientist, 192(2583/2584):34–37, 2006.

Annotation: Description of various instances of Internet addictions.

[278] Mike Cardosa, Katie Panciera, and Anna Rouben. Comparative study of Wikipedia users (2002 vs. 2005), 2007? Apparently only published on Rouben's website. Year of publication is unknown.

Annotation: A quantitative longitudinal study of Wikipedia users.

- [279] Jared Diamond, editor. The world until yesterday: What can we learn from traditional societies? Penguin Books, London, England, 2013.
- [280] Wikimedia Foundation staff and volunteers. Former contributors survey results. Wikimedia Strategic Planning, February 2010.

Annotation: Results about a survey on Wikipedia contributors that have stopped editing.

[281] Bongwon Suh, Gregorio Convertino, Ed H. Chi, and Peter Pirolli. The singularity is not near: slowing growth of Wikipedia. In Proceedings of the 5th International Symposium on Wikis and Open Collaboration. Association for Computing Machinery, 2009.

Annotation: Data analysis of the evolution of Wikipedia.

- [282] Camille Roth. Viable wikis: struggle for life in the wikisphere. In *Proceedings of the 2007 In*ternational Symposium on Wikis, pages 119–124. Association for Computing Machinery, 2007.
- [283] Dennis M. Wilkinson and Bernardo A. Huberman. Cooperation and quality in Wikipedia. In Proceedings of the 2007 International Symposium on Wikis, pages 157–164, New York, NY, USA, 2007. ACM.

Annotation: An investigation of what signals quality articles on the English Wikipedia. A growth model is suggested for the number of edits on an article, and this lognormal model shows good agreement with the observed number of edits. It is also shown that older articles ypically have more edits than newer articles and that quality articles have comparably more edits, many distinct editors, many edits in its talk page, a quick 'turnaround' and a high number of edits per editor when correcting for article age and Google PageRank. They write that featured articles with high Google PageRank have a lower number of talk page edits than articles with low PageRank. it is not clear if that effect is due to archiving.

[284] Dennis M. Wilkinson and Bernardo A. Huberman. Assessing the value of cooperation in Wikipedia. First Monday, 12(4), April 2007.

- [285] Rodrigo Almeida, Barzan Mozafari, and Junghoo Cho. On the evolution of Wikipedia. In International Conference on Weblogs and Social Media, 2007.
- [286] Sascha Segan. R.i.p usenet: 1980-2008. PC-Mag.com, July 2008.

Annotation: Discusses the decline of USENET

[287] Megan Garber. The contribution conundrum: Why did Wikipedia succeed while other encyclopedias failed?. Nieman Journalism Lab, 2011.

Annotation: A description of research on why Wikipedia was successful and other crowd-source online encyclopedia were not. The encyclopedia as a familiar product, focus on substantive content rather than technology as well as low transaction cost in participation were reported.

[288] Philippe Aigrain. The individual and the collective in open information communities. Extended abstract to 16th BLED Electronic Commerce Conference, June 2003.

Annotation: Discussion of reasons for success in Wikipedia, Slashdot and SPIP focussing on low transaction costs, clear vision and mechanisms and software tools to counter hostile contribution.

[289] Hsin-Yi Wu and Kuang-Hua Chen. A quantitative comparison on online encyclopedias — a case study of Wikipedia and Knol. In Proceedings of 2010 International Symposium on the Transformation and Innovation of Library and Information Science, November 2010.

Annotation: Related to the paper A quantitative study for online encyclopedias: comparison of Wikipedia and Knol. The present paper has information, e.g., about which articles the researchers studied.

[290] Kuang-Hua Chen and Hsin-Yi Wu. A quantitative study for online encyclopedias: comparison of Wikipedia and Knol. Journal of Library and Information Science Research, 7(1):32–38, December 2012.

Annotation: Comparison of 20 articles from Wikipedia and Google Knol with respect to page views, number of words, readability, number of citations, number of references and types of references.

[291] Jason Kincaid. Google announces plans to shutter Knol, Friend Connect, Wave, and more. TechCrunch, November 2011. Annotation: News articles about Google announcing the closing of Knol

[292] David M. Pennock, Steve Lawrence, C. Lee Giles, and Finn Årup Nielsen. The real power of artificial markets. Science, 291(5506):987–988, February 2001.

Annotation: A brief note reporting that artificial markets, such as Hollywood Stock Exchange and Foresight Exchange, are good predictors of future outcomes.

- [293] David M. Pennock, Steve Lawrence, Finn Årup Nielsen, and C. Lee Giles. Extracting collective probabilistic forecasts from web games. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, pages 174–183, New York, NY, USA, August 2001. SIGKDD, Association for Computing Machinery, ACM.
- [294] Shane Greenstein. The range of Linus' Law. *IEEE Micro*, 32(1):72–73, January/February 2012.

Annotation: Discusses Linus' Law with respect to Wikipedia.

- [295] Alexandre Bouchard, Percy Liang, Thomas Grifths, and Dan Klein. A probabilistic approach to diachronic phonology. In *Proceedings* of EMNLP-CoNLL, pages 887–896, 2007.
- [296] Paula Chesley, Bruce Vincent, Li Xu, and Rohini Srihari. Using verbs and adjectives to automatically classify blog sentiment. In Nicolas Nicolov, Franco Salvetti, Mark Liberman, and James H. Martin, editors, Proceedings of AAAI-CAAW-06, the Spring Symposia on Computational Approaches to Analyzing Weblogs, AAAI's Spring Symposium. The AAAI Press, 2006.

Annotation: Text sentiment analysis of blogs by using Wiktionary, verb classes and support vector machines classifier.

- [297] T. Zesch, C. Mueller, and I. Gurevych. Extracting lexical semantic knowledge from Wikipedia and Wiktionary. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, 2008.
- [298] T. Zesch, C. Mueller, and I. Gurevych. Using Wiktionary for computing semantic relatedness. In *Proceedings of AAAI*. Association for the Advancement of Artificial Intelligence, 2008.
- [299] Akiyo Nadamoto, Eiji Aramaki, Takeshi Abekawa, and Yohei Murakami. Content hole search in community-type content using Wikipedia. In Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services, pages 25–32, New York, NY, USA, 2009. Association for Computing Machinery.

Annotation: Conference article of 'Extracting content holes by comparing community-type content with Wikipedia'.

[300] Akiyo Nadamoto, Eiji Aramaki, Takeshi Abekawa, and Yohei Murakami. Extracting content holes by comparing community-type content with Wikipedia. *International Journal of Web Information Systems*, 6(3):248–260, August 2010.

Annotation: Describes a system for discovery of missing subtopics in a text by comparison against text in Wikipedia.

[301] Akiyo Nadamoto, Eiji Aramaki, Takeshi Abekawa, and Yohei Murakami. Content hole search in community-type content. In *Proceedings* of the 18th international conference on World wide web, pages 1223–1224, New York, NY, USA, 2009. Association for Computing Machinery.

Annotation: Short paper on the same topic as 'Extracting content holes by comparing community-type content with Wikipedia'.

- [302] Marijn Koolen, Gabriella Kazai, and Nick Craswell. Wikipedia pages as entry points for book search. In WSDM '09: Proceedings of the Second ACM International Conference on Web Search and Data Mining, pages 44–53, New York, NY, USA, 2009. ACM.
- [303] David N. Milne, Ian H. Witten, and David M. Nichols. A knowledge-based search engine powered by Wikipedia. In CIKM '07 Proceedings of the sixteenth ACM conference on Conference on Information and Knowledge Management, pages 445–454. Association for Computing Machinery, November 2007.

Annotation: Describes the Koru document retrieval system, that is based on a thesaurus constructed from Wikipedia data and features automatic query expansion. A user study with a comparison system shows that Koru performs better and more preferred.

[304] Olga Vechtomova. Facet-based opinion retrieval from blogs. *Information Processing and Manage*ment, 46(1):71–88, 2010.

Annotation: Describes an information retrieval systems with sentiment analysis and query expansion using Wikipedia.

[305] Julien Ah-Pine, Marco Bressan, Stephane Clinchant, Gabriela Csurka, Yves Hoppenot, and Jean-Michel Renders. Crossing textual and visual content in different application scenarios. *Multimedia Tools and Applications*, 42(1):31–56, 2009.

Annotation: Description of a system for combined text and image similarity computation evaluated on the Image-CLEF photo 2008 data set and demonstrated on Wikipedia data.

[306] Daniel Kinzler. WikiWord: Multilingual image search and more. In Wikimania, 2009.

Annotation: Describes the multilingual image retrieval system WikiWord based on data from Wikipedia.

- [307] Robert L. Chapman, editor. Roget's international thesaurus. HarperCollins Publishers, New York, NY, fifth edition edition, 1992.
- [308] Andrew Krizhanovsky. Synonym search in Wikipedia: Synarcher. In 11-th International Conference "Speech and Computer" SPECOM'2006. Russia, St. Petersburg, June 25–29, pages 474–477, 2006.

Annotation: Description of the Synarcher program that analyze Wikipedia with the Kleinberg HITS algorithm and present related terms to a query term in a graph visualization.

- [309] Michael Strube and Simone Paolo Ponzetto. WikiRelate! computing semantic relatedness using Wikipedia. In Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence, pages 1419–1424, Menlo Park, Califonia, 2006. AAAI Press.
- [310] Decong Li, Sujian Li, Wenjie Li, Congyun Gu, and Yun Li. Keyphrase extraction based on topic relevance and term association. *Journal of Information and Computational Science*, 7(1):293–299, 2010.
- [311] Razvan Bunescu and Marius Pasca. Using encyclopedic knowledge for named entity disambiguation. In Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-06), pages 9–16. Association for Computational Linguistics, 2006.
- [312] Silviu Cucerzan. Large-scale named entity disambiguation based on Wikipedia data. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pages 708–716. Association for Computational Linguistics, 2007.
- [313] Jun'ichi Kazama and Kentaro Torisawa. Exploiting Wikipedia as external knowledge for named entity recognition. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pages 698–707, Stroudsburg, PA, USA, June 2007. Association for Computational Linguistics.

- [314] Dominic Balasuriya, Nicky Ringland, Joel Nothman, Tara Murphy, and James R. Curran. Named entity recognition in Wikipedia. In Proceedings of the 2009 Workshop on the People's Web Meets NLP, pages 10–18. Association for Computational Linguistics, 2009.
- [315] Mohamed Amir Yosef, Johannes Hoffart, Ilaria Bordino, Marc Spaniol, and Gerhard Weikum. AIDA: an online tool for accurate disambiguation of named entities in text and tables. In *Proceedings of the VLDB Endowment*, volume 4, pages 1450–1453, 2011.

Annotation: Describes an online named entity extraction system with disambiguation using YAGO2, Stanford NER Tagger and PostgreSQL.

[316] Michael Cook, Simon Colton, and Alison Pease. Aesthetic considerations for automated platformer design. Association for the Advancement of Artificial Intelligence, 2012.

Annotation: Describes AN-GELINA3 that does notable person detection by looking up a person on Wikipedia, and the sentiment about the person is gauged by looking the person up on Twitter and using the AFINN word list for text sentiment analysis.

- [317] Daniel Kinzler. Automatischer aufbau eines multilingualen thesaurus durch extraktion semantischer und lexikalischer relationen aus der Wikipedia. Diplomarbeit an der abteilung für automatische sprachverarbeitung, Institut für Informatik, Universität Leipzig, 2008.
- [318] Eckhard Bick. WikiTrans: the English Wikipedia in Esperanto. In Constraint Grammar Applications, Workshop Proceedings at Nodalida 2011, volume 14, pages 8–16, 2011.
- [319] Eckhard Bick. Translating the Swedish Wikipedia into Danish. In Swedish Language Technology Conference 2014, 2014.

Annotation: Short description of a machine translation system for translating Swedish Wikipedia to Danish.

[320] Daniel H. Pink. Folksonomy. The New York Times, December 2005.

Annotation: News article explaining the at that time recently introduced term 'folksonomy'.

[321] Jakob Voß. Collaborative thesaurus tagging the Wikipedia way. arXiv, April 2006.

Annotation: Analysis of the Wikipedia category system.

[322] Tao Guo, David G. Schwartz, Frada Burstein, and Henry Linger. Codifying collaborative knowledge: using Wikipedia as a basis for automated ontology learning. *Knowledge Management Research & Practice*, 7(3):206–217, September 2009.

Annotation: Describes a semiautomated ontology learning system based on Wikipedia data.

- [323] Magnús Sigurdsson and Søren Christian Halling. Zeeker: A topic-based search engine. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Kongens Lyngby, Denmark, 2007.
- [324] Fabian Suchanek, Gjergji Kasneci, and Gerhard Weikum. YAGO: A core of semantic knowledge unifying WordNet and Wikipedia. In Carey L. Williamson, Mary Ellen Zurko, Peter F. Patel-Schneider, and Prashant J. Shenoy, editors, Proceedings of the Sixteenth International World Wide Web Conference (WWW2007), May 8-12, 2007, Banff, Alberta, CANADA, pages 697–706. ACM, 2007.
- [325] Georgi Kobilarov, Tom Scott, Yves Raimond, Silver Oliver, Chris Sizemore, Michael Smethurst, Christian Bizer, and Robert Lee. Media meets semantic web how the BBC uses DBpedia and Linked Data to make connections. In Lora Aroyo, Eyal Oren, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Marta Sabou, and Elena Simperl, editors, The Semantic Web: Research and Applications, volume 5554 of Lecture Notes in Computer Science, pages 723–737, Berlin/Heidelberg, 2009. Springer.
- [326] Brian Krebs. Wikipedia edits forecast vice presidential picks. washingtonpost.com, August 2008.

Annotation: News articles noting large activity in US vicepresidental candidate Sarah Palin Wikipedia article just before her candidacy became public knowledge.

[327] Cord Jefferson. Did an anonymous Wikipedia editor try to out the Petraeus affair back in January?. Gawker, November 2012.

Annotation: News story about an edit on Wikipedia that seemingly reveal information before it became public knowledge.

[328] Thomas Steiner, Seth van Hooland, and Ed Summers. MJ no more: using concurrent Wikipedia edit spikes with social network plausibility checks for breaking news detection. In Proceedings of the 22nd international conference on World Wide Web companion, pages 791–794, 2013.

Annotation: Describes a realtime breaking news monitoring service using Wikipedia edits.

- [329] Thomas Steiner. Comprehensive Wikipedia monitoring for global and realtime natural disaster detection. In Ruben Verborgh and Erik Mannens, editors, Proceedings of the ISWC Developers Workshop, volume 1268 of CEUR Workshop Proceedings, pages 86–95, 2014.
- [330] Murray Aitken, Thomas Altmann, and Daniel Rosen. Engaging patients through social media: Is healthcare ready for empowered and digitally demanding patients?. IMS Institute for Healthcare Informatics, January 2014.

Annotation: Analysis of social media, (YouTube, Facebook, Twitter and Wikipedia), in relation to social media engagement in health care.

- [331] Márton Mestyán, Taha Yasseri, and János Kertész. Early prediction of movie box office success based on Wikipedia activity big data. ArXiv 1211.0970, November 2012.
- [332] Brain Keegan. Does Wikipedia editing activity forecast Oscar wins?. www.brianckeegan.com, March 2014.

Annotation: Blog post with prediction of 2014 winners for 5 Academy Awards categories.

[333] Taha Yasseri and Jonathan Bright. Can electoral popularity be predicted using socially generated big data. it - Information Technology, 56(5):246–253, September 2014.

Annotation: Analysis of how good Google search volume and Wikipedia page views are for the prediction of elections in Iran, Germany and United Kingdom.

- [334] Kevin Makice. Twitter API: Up and running. O'Reilly, Sebastopol, California, 2009.
- [335] Stefanie Haustein, Isabella Peters, Judit Bar-Ilan, Jason Priem, Hadas Shema, and Jens Terliesner. Coverage and adoption of altmetrics sources in the bibliometric community. ArXiv.
- [336] Perry Evans and Michael Krauthammer. Exploring the use of social media to measure journal article impact. In AMIA Annual Symposium Proceedings, pages 374–381, January 2011.

Annotation: Altmetrics study of Wikipedia extracting scientific journal article citations based on PMID and DOI with comparison to Faculty of 1000 data. With longitudinal analysis they found articles to the cited increasingly faster.

[337] Jörg Waitelonis, Nadine Ludwig, Magnus Knuth, and Harald Sack. WhoKnows? evaluating linked data heuristics with a quiz that cleans up DBpedia. Interactive Technology and Smart Education, 8(4):236–248, 2011.

[338] Mihai Letia, Nuno Preguiça, and Marc Shapiro. Consistency without concurrency control in large, dynamic systems. ACM SIGOPS Operating Systems Review, 44:29–34, 2010.

Annotation: Describes a system from sharing mutable data.

- [339] Richard Bentley, Thilo Horstmann, Klaas Sikkel, and Jonathan Trevor. Supporting collaborative information sharing with the world wide web: The BSCW shared workspace system. In Fourth International World Wide Web Conference. World Wide Web Consortium, December 1995.
- [340] Y. Goland, E. Whitehead, A. Faizi, and D. Jensen. HTTP extensions for distributed authoring – WEBDAV, February 1999.
- [341] Honglei Zeng, Maher Alhossaini, Li Ding, Richard Fikes, and Deborah L. McGuinness. Computing trust from revision history. In Proceedings of the 2006 International Conference on Privacy, Security and Trust, volume 380 of ACM International Conference Proceeding Series, New York, NY, USA, October 2006. Association for Computing Machinery.

Annotation: Sets up a statistical model (a dynamic Bayesian network) for modeling of Wikipedia article trust by looking at the revision history. The set Bayesian priors based on whether the user is administrator, registered user, anonymous user or blocked user. Further the trust is based on the amount of insertion and deletions.

[342] Joshua E. Blumenstock. Size matters: Word count as a measure of quality on Wikipedia. In Proceedings of the 17th International World Wide Web Conference (WWW2008). April 21-25, 2008. Beijing, China, April 2008.

Annotation: Suggests word count as a simple measure for quality and shows that it perform well when classifying featured and random articles, — seemingly better than the methods of Zeng and Stvilia.

[343] Gerald C. Kane. A multimethod study of information quality in wiki collaboration. ACM Transactions on Management Information Systems, 2(1):4, March 2011.

Annotation: A logistic regression analysis of which features were important for quality article on Wikipedia.

[344] Daniel Nasaw. Meet the 'bots' that edi Wikipedia. BBC New Magazine, July 2012.

Annotation: News articles about bots operating on Wikipedia.

- [345] Aaron Halfaker, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. The rise and decline of an open collaboration system: how Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, 57(5):664–688, December 2013.
- [346] Pierpaolo Dondio, Stephen Barrett, Stefan Weber, and Jean Marc Seigneur. Extracting trust from domain analysis: A case study on the Wikipedia project. In Autonomic and Trusted Computing, volume 4158 of Lecture Notes in Computer Science, pages 362–373, Berlin/Heidelberg, October 2006. Springer.
- [347] B. Thomas Adler and Luca de Alfaro. A content-driven reputation system for the Wikipedia. In Proceedings of the 16th international conference on World Wide Web, pages 261–270, New York, NY, USA, 2007. Association for Computing Machinery.
- [348] B. Thomas Adler, J. Benterou, K. Chatterjee, Luca de Alfaro, I. Pye, and V. Raman. Assigning trust to Wikipedia content. Technical Report UCSC-CRL-07-09, School of Engineering, University of California, Santa Cruz, CA, USA, November 2007.
- [349] B. Thomas Adler, J. Benterou, K. Chatterjee, Luca de Alfaro, I. Pye, and V. Raman. Assigning trust to Wikipedia content. In *Proceedings of the* 4th International Symposium on Wikis, page 26, New York, NY, USA, 2008. Association for Computing Machinery.

Annotation: Describes a reputation system for Wikipedia which enable individual words to be colored according to computed 'trust'.

- [350] B. Thomas Adler, Luca de Alfaro, and Ian Pye. Detecting Wikipedia vandalism using wikitrust. In *PAN 2010 proceedings*, 2010.
- [351] Santiago M. Mola Velasco. Wikipedia vandalism detection through machine learning: Feature review and new proposals. In Lab Report for PAN at CLEF 2010, 2010.
- [352] Meiqun Hu, Ee-Peng Lim, Aixin Sun, Hady W. Lauw, and Ba-Quy Vuong. Measuring article quality in Wikipedia: models and evaluation. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, pages 243–252, New York, NY, USA, 2007. Association for Computing Machinery.
- [353] Christopher Thomas and Amit P. Sheth. Semantic convergence of Wikipedia articles. In IEEE/WIC/ACM International Conference on Web Intelligence, pages 600–606. IEEE, 2007.
- [354] Martin Potthast, Benno Stein, and Teresa Holfeld. Overview of the 1st international competition on Wikipedia vandalism detection. In *PAN* 2010, 2010.

- [355] Martin Potthast. Crowdsourcing a Wikipedia vandalism corpus. In Hsin-Hsi Chen, Efthimis N. Efthimiadis, Jaques Savoy, Fabio Crestani, and Stéphane Marchand-Maillet, editors, 33rd International ACM Conference on Research and Development in Information Retrieval (SIGIR 10), pages 789–790, New York, NY, USA, July 2010. ACM.
- [356] Martin Potthast and Teresa Holfeld. Overview of the 2nd international competition on Wikipedia vandalism detection. In Vivien Petras and Paul Clough, editors, Notebook Papers of CLEF 2011 Labs and Workshops, 2011.

Annotation: Report from a prediction competition on Wikipedia vandalism detection. The corpus was based on both English, German and Spanish Wikipedias. Three systems participated.

- [357] Sara Javanmardi, David W. McDonald, and Cristina V. Lopes. Vandalism detection in Wikipedia: A high-performing, feature—rich model and its reduction through Lasso. In Wiki-Sym'11, 2011.
- [358] Kermit K. Murray. Mass spectrometry on Wikipedia: Open source and peer review. Presented at the 55th ASMS Conference on Mass Spectrometry, June 3–7, 2007, Indianapolis, Indiana, June 2007.
- [359] Jon W. Huss, III, Camilo Orozco, James Goodale, Chunlei, Serge Batalov, Tim J. Vickers, Faramarz Valafar, and Andrew I. Su. The Gene Wiki for community annotation of gene function. PLoS Biology, 6(7):e175, July 2008.

Annotation: Description of the creation and addition of over 8000 gene articles in Wikipedia with an automated bot. Information is aggregated from Entrez Gene and a gene atlas for the mouse and human protein-encoding transcriptomes.

- [360] Andrew I. Su, Tim Wiltshire, Serge Batalov, Hilmar Lapp, Keith A. Ching, David Block, Jie Zhang, Richard Soden, Mimi Hayakawa, Gabriel Kreiman, Michael P. Cooke, John R. Walker, and John B. Hogenesch. A gene atlas of the mouse and human protein-encoding transcriptomes. Proceedings of the National Academy of Sciences of the United States of America, 101(16):6062–6067, April 2004.
- [361] Jennifer Daub, Paul P. Gardner, John Tate, Daniel Ramsköld, Magnus Manske, William G. Scott, Zasha Weinberg, Sam Griffiths-Jones, and Alex Bateman. The RNA WikiProject: Community annotation of RNA families. RNA, October 2008.

Annotation: Describes the WikiProject RNA where data and annotation about RNA are shared between Wikipedia and the Rfam database.

[362] Sylvain Brohée, Roland Barriot, and Yves Moreau. Biological knowledge bases using wikis: combining the flexibility of wikis with the structure of databases. *Bioinformatics*, 2010.

Annotation: Shortly describes shortly the WikiOpener MediaWiki extension, that can be used to include data from web-based database in a wiki. They state that the extension is used in two biological databases (wikis): CHDwiki and YTPdb.

[363] Magnus Manske. The game is on. The Whelming, May 2014.

Annotation: Blog post presenting the a gamification input for Wikidata: Wikidata — The Game.

[364] Thomas Pellissier Tanon, Denny Vrandecic, Sebastian Schaffert, Thomas Steiner, and Lydia Pintscher. From Freebase to Wikidata: the great migration. 2016.

Annotation: Description of content migration from Google Freebase to Wikidata with the Primary Sources Tool

- [365] Sisay Fissaha Adafre and Maarten de Rijke. Discovering missing links in Wikipedia. In Jafar Adibi, Marko Grobelnik, Dunja Mladenic, and Patrick Pantel, editors, Proceedings of the 3rd international workshop on Link discovery, pages 90–97, New York, NY, USA, 2005. ACM.
- [366] Wei Che Huang, Andrew Trotman, and Shlomo Geva. Collaborative knowledge management: evaluation of automated link discovery in the Wikipedia. In Andrew Trotman, Shlomo Geva, and Jaap Kamps, editors, *Proceedings of the SI-GIR 2007 Workshop on Focused Retrieval*. Department of Computer Science, University of Otago, 2007.
- [367] Michael Granitzer, Mario Zechner, Christin Seifert, Josef Kolbitsch, Peter Kemper, and Ronald In't Velt. Evaluation of automatic linking strategies for Wikipedia pages. In Proceedings of the IADIS WWWInternet Conference 2008. IADIS, 2008.
- [368] Rianne Kaptein, Pavel Sedyukov, and Jaap Kamps. Linking Wikipedia to the Web. In Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, 2010. ACM.

Annotation: Built and evaluate a system for prediction of external links within a Wikipedia article.

[369] Bahar Sateli, Marie-Jean Meurs, Greg Butler, Justin Powlowski, Adrian Tsang, and René Witte. IntelliGenWiki: an intelligent semantic wiki for life sciences. EMBnet.journal, 18 supplement B:50-52, 2012.

Annotation: Describes IntelliGen-Wiki a MediaWiki-based wiki with natural language processing (NLP) The data is represented via a Semantic MediaWiki extension. The NLP is handled with a General Architecture for Text Engineering (GATE) pipeline and may, e.g., identify organisms and enzymes.

[370] Bahar Sateli and René Witte. Natural language processing for MediaWiki: the semantic assistants approach. In Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration, New York, NY, USA, 2012. Association for Computing Machinery.

Annotation: Describes a system that integrates natural language processing with a MediaWiki-based wiki. NLP service is provided by a W3C standard web services running GATE NLP software. The system read data from a wiki and embed the result in the wiki. The system is demonstrated on several data/wiki: DurmWiki with cultural heritage data where the system extract terms for an index, ReqWiki for quality assurance in software requirements specifications on a wiki and GeneWiki with extraction of biological terms.

- [371] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. Scientific American, 284(5):28–37, May 2001.
- [372] Roberto Tazzoli, Paolo Castagna, and Stefano Emilio Campanini. Towards a semantic wiki wiki web. Poster at the International Semantic Web Conference, November 2004.

Annotation: Describes the semantic wiki called PlatypusWiki that is implemented in Java.

[373] Eyal Oren. Semperwiki: a semantic personal wiki. In Stefan Decker, Jack Park, Dennis Quan, and Leo Sauermann, editors, Proceedings of the ISWC 2005 Workshop on The Semantic Desktop - Next Generation Information Management & Collaboration Infrastructure. Galway, Ireland, November 6, 2005, volume 175 of CEUR Workshop Proceedings, November 2005.

Annotation: Describes the Semper-Wiki semantic personal wiki for the GNOME desktop environment.

[374] David Aumüller. Semantic authoring and retrieval within a wiki. In 2nd European Semantic Web Conference (ESWC 2005), 2005. Demos and Posters.

Annotation: describes the SHAWN Wiki, a semantic wiki implemented in Perl with semantic input in the wikimarkup and a query feature.

[375] Sebastian Schaffert. IkeWiki: a semantic wiki for collaborative knowledge management. In 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, 2006. WETICE '06, pages 388–396. IEEE, 2006.

Annotation: Describes the IkeWiki semantic wiki engine. It is implemented as a Java web application and has support for OWL-RDFS reasoning.

[376] Michel Buffa and Fabien Gandon. Sweetwiki: semantic web enabled technologies in wiki. In Proceedings of the 2006 International Symposium on Wikis, pages 69–78. Association for Computing Machinery, 2006.

Annotation: Description of the SweetWiki semantic wiki engine. It does not use a wiki markup language but rather XHTML and a WYSIWYG interface based on the Kupu editor. It relies on the CORESE semantic search engine.

[377] Hendry Muljadi, Hideaki Takeda, Jiro Araki, Shoko Kawamoto, Satoshi Kobayashi, Yoko Mizuta, Sven Minory Demiya, Satoshi Suzuki, Asanoby Kitamoto, Yasuyuki Shirai, Nobuyuki Ichiyoshi, Takehiko Ito, Takashi Abe an Takashi Gojobori, Hideaki Sugawara, Satoru Miyazaki, and Asao Fujiyama. Samantic mediawiki: a user-oriented system for integrated content and metadata management system. In IADIS International Conference on WWW/Internet, pages 261–264, 2005.

Annotation: Suggest a semantic wiki based on the MediaWiki software.

- [378] Max Völkel, Markus Krötzsch, Denny Vrandecic, Heiko Haller, and Rudi Studer. Semantic wikipedia. In Proceedings of the 15th international conference on World Wide Web, pages 585–594, New York, NY, USA, 2006. ACM.
- [379] Markus Krötzsch, Denny Vrandečić, and Max Völkel. Semantic MediaWiki. In Isabel Cruz, Stefan Decker, Dean Allemang, Chris Preist, Daniel Schwabe, Peter Mika, Mike Uschold, and Lora Aroyo, editors, The Semantic Web ISWC 2006, volume 4273 of Lecture Notes in Computer Science, pages 935–942, Berlin/Heidelberg, 2006. Springer.

- [380] David E. Millard, Christopher P. Bailey, Philip Boulain, Swapna Chennupati, Hugh C. Davis, Yvonne Howard, and Gary Wills. Semantics on demand: Can a semantic wiki replace a knowledge base? New Review of Hypermedia and Multimedia, 14(1):95–120, January 2008.
- [381] Jie Bao, Li Ding, and Deborah L. McGuinness. Semantic history: towards modeling and publishing changes of online semantic data. In John Breslin, Uldis Bojars, Alexandre Passant, and Sergio Fernández, editors, *Proceedings of the 2nd Workshop on Social Data on the Web (SDoW2009*, volume 520 of CEUR Workshop Proceedings, 2009.

Annotation: Describes an extension to Semantic MediaWiki so the revision history of a MediaWiki is available to the semantic query system.

- [382] Natasha Noy, Alan Rector, Pat Hayes, and Chris Welty. Defining n-ary relations on the semantic web. Technical report, W3C, April 2006.
- [383] Michael Backhaus and Janet Kelso. BOWiki a collaborative annotation and ontology curation framework. In WWW2007, 2007.
- [384] Denny Vrandečić. Wikidata: a new platform for collaborative data collection. In Proceedings of the 21st international conference companion on World Wide Web, pages 1063–1064, New York, NY, USA, 2012. Association for Computing Machinery.
- [385] Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledge base. Communications of the ACM, 2014.

Annotation: A general introduction to Wikidata.

- [386] Kingsley Idenhen. Preliminary SPARQL endpoint for Wikidata. Wikidata-l mailing list, March 2015.
- [387] Daniel Hernández, Aidan Hogan, and Markus Krötzsch. Reifying RDF: what works well with wikidata?. In Thorsten Liebig and Achille Fokoue, editors, Proceedings of the 11th International Workshop on Scalable Semantic Web Knowledge Base Systems, volume 1457 of CEUR Workshop Proceedings, pages 32–47. CEUR-WS.org, 2015.

Annotation: Describes four different methods to convert Wikidata data to a Semantic Web triple data structure and benchmark them with 5 different SPARQL engines.

[388] Marc Chevalier, Raphaël Charrondière, Quentin Cormier, Yassine Hamoudi, Valentin Lorentz, and Thomas Pellissier Tanon. Project pensées profondes. Master's thesis, ENS de Lyon, December 2014. Annotation: Describes a student project that built a question answering system, http://askplatyp.us, with the use of natural language processing and Wikidata data.

[389] Carrie Arnold, Todd Fleming, David Largent, and Chris Lüer. DynaTable: a wiki extension for structured data. In Proceedings of the 5th International Symposium on Wikis and Open Collaboration, New York, NY, USA, 2009. ACM.

Annotation: Short description of the DynaTable MediaWiki extension for structured data in tables.

- [390] Finn Arup Nielsen. Brede Wiki: Neuroscience data structured in a wiki. In Christoph Lange, Sebastian Schaffert, Hala Skaf-Molli, and Max Völkel, editors, Proceedings of the Fourth Workshop on Semantic Wikis — The Semantic Wiki Web, volume 464 of CEUR Workshop Proceedings, pages 129–133, Aachen, Germany, June 2009. RWTH Aachen University.
- [391] Finn Årup Nielsen, Matthew J. Kempton, and Steven C. R. Williams. Online open neuroimaging mass meta-analysis. In Alexander García Castro, Christoph Lange, Frank van Harmelen, and Benjamin Good, editors, Proceedings of the 2nd Workshop on Semantic Publishing, volume 903 of CEUR Workshop Proceedings, pages 35— 39, Aachen, Germany, 2012.
- [392] Sandro Pinna, Simone Mauri, Paolo Lorrai, Michele Marchesi, and Nicola Serra. XPSwiki: An agile tool supporting the planning game. In Extreme Programming and Agile Processes in Software Engineering, volume 2675 of Lecture Notes in Computer Science, page 1014, Berlin, 2003. Springer.

Annotation: Describes a wiki for supporting extreme programming. The wiki can associate forms with wiki pages.

[393] Anja Haake, Stephan Lukosch, and Till Schümmer. Wiki-templates: Adding structure support to wikis on demand. In *Proceedings* of the 2005 international symposium on Wikis, pages 41–51, New York, NY, USA, 2005. ACM.

Annotation: Describes a wiki with form-based input applied for annotation of scientific papers. The user may modify how the form is presented for display and for editing. XML and wiki markup combine in a language to define the form.

[394] Christoph Sauer, Chuck Smith, and Tomas Benz. Wikicreole: a common wiki markup. In Proceedings of the 2007 international symposium on Wikis, pages 131–142, New York, NY, USA, 2007. ACM.

Annotation: Describes the WikiCreole language that is a standardized wiki markup language.

- [395] Martin Junghaus, Dirk Riehle, Rama Gurram, Matthias Kaier, Mario Lopes, and Umit Yalcinalp. An EBNF grammar for Wiki Creole 1.0. ACM SIGWEB Newsletter, Winter 2007.
- [396] Martin Junghans, Dirk Riehle, Rama Gurram, Matthias Kaiser, Mário Lopes, and Umit Yalcinalp. A grammar for standardized wiki markup. In Ademar Aguiar and Mark Bernstein, editors, Proceedings of the 2008 International Symposium on Wikis, 2008, Porto, Portugal, September 8– 10, 2008. ACM, 2008.
- [397] Gabriel Wicke and Subramanya Sastry. Parsoid: How Wikipedia catches up with the web. Wikimedia blog, March 2013.

Annotation: A blog post on the Wikimedia Foundation Parsoid project that creates a HTML5 representation of MediaWiki wikitext.

[398] Andrew Orlowski. Revolting peasants force Wikipedia to cut'n'paste Visual Editor into the bin. The Register, September 2013.

Annotation: News article on the controversy surrounding the VisualEditor on Wikipedia.

- [399] Michael Hart. The history and philosophy of Project Gutenberg, 1992.
- [400] Lars Aronsson. Wikisource. Runeberg mailing list, January 2008.

Annotation: Swedish introduction to Wikisource and its inspiration from Runeberg.

[401] Imene Bensalem, Salim Chikhi, and Paolo Rosso. Building Arabic corpora from Wikisource. In 2013 ACS International Conference on Computer Systems and Applications (AICCSA). IEEE, 2013.

Annotation: Describes a system for collecting texts from the Arabic Wikisource to form a corpus that can be used in a text mining application for plagiarism detection.

[402] Michael F. Goodchild. Citizens as sensors: Web 2.0 and the volunteering of geographic information. *GeoFocus*, (7):8–10, July 2007.

Annotation: Editorial on volunteered geographic information systems

[403] Mordechai Haklay and Patrick Weber. Open-StreetMap: User-generated street maps. *Pervasive computing*, pages 12–18, October-December 2008.

Annotation: Introduces Open-StreetMap with discussion of among other issues the historic development, input and output methods, technical infrastructure, distributed maptile rendering, social collaboration, mapping parties, motivations, data accuracy and participation inequality.

- [404] Ruben Sipos, Abhijit Bhole, Blaz Fortuna, Marko Grobelnik, and Dunja Mladenic. Demo: HistoryViz visualizing events and relations extracted from Wikipedia. In Lora Aroyo, Eyal Oren, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Marta Sabou, and Elena Simperl, editors, The Semantic Web: Research and Applications: 6th European Semantic Web Conference, ESWC 2009 Heraklion, Crete, Greece, May 31-June 4, 2009 Proceedings, volume 5554 of Lecture Notes in Computer Science, pages 903–907, Berlin/Heidelberg, May 2009. Springer.
- [405] Abhijit Bhole, Blaz Fortuna, Marko Grobelnik, and Dunja Mladenic. Extracting named entities and relating them over time based on Wikipedia. Informatica (Slovenia), 31(4):463–468, 2007.
- [406] Navino Evans. Database system and method. United States Patent Application, February 2014. US20140046948 A1.

Annotation: Patent on some aspects of histropedia.

- [407] Juhani Eronen and Juha Röning. Graphingwiki a semantic wiki extension for visualising and infering protocol dependency. In Max Völkel and Sebastian Schaffert, editors, Proceedings of the First Workshop on Semantic Wikis From Wiki to Semantics co-located with the ESWC2006, Workshop on Semantic Wikis. ESWC2006, June 2006.
- [408] Emden R. Gansner and Stephen C. North. An open graph visualization system and its applications to software engineering. Software Practice and Experience, 30(11):1203–1234, 2000.
- [409] Frank Dengler, Steffen Lamparter, Mark Hefke, and Andreas Abecker. Collaborative process development using Semantic MediaWiki. In Knut Hinkelmann and Holger Wache, editors, WM2009: 5th Conference of Professional Knowledge Management, volume P-145 of Lecture Notes in Informatics, pages 97–107, Solothurn, Switzerland, March 2009. Bonner Köllen Verlag.
- [410] Jacek Jankowski and Sebastian Ryszard Kruk. 2LIP: The step towards the Web3D. In Proceedings of the 17th International World Wide Web Conference (WWW2008). April 21-25, 2008. Beijing, China, pages 1137–1138, April 2008.
- [411] Jacek Jankowski. Copernicus: 3D Wikipedia. In International Conference on Computer Graphics

and Interactive Techniques. ACM, 2008. Session: Design. Article No. 30.

Annotation: Description of 'Copernicus': A encyclopedia/browser that combines Wikipedia article reading with a 3-dimensional viewer in a two-layer interface paradigm. The article is presented as a transparent foreground, the 3D model in the background.

- [412] Jacek Jankowski. Copernicus adding the third dimension to Wikipedia. In Wikimania. Wikimedia, July 2008.
- [413] Joseph Corneli. GravPad. In Proceedings of the 6th International Symposium on Wikis and Open Collaboration, pages 30–31. ACM, 2010.

Annotation: Fairly briefly describes EtherPad with wiki and other extensions.

- [414] Priya Ganapati. \$20 Wikipedia reader uses 8-bit computing power. Wired Gadget Lab, July 2010.
- [415] Jon Brodkin. All of Wikipedia can be installed to your desktop in just 30 hours. Ars Technica, November 2013.

Annotation: News article on the XOWA application for automating downloading and offline displaying of Wikipedias and its sister project.

[416] Nina Hansen, Namkje Koudenburg, Rena Hiersemann, Peter J. Tellegen, Márton Kocsev, and Tom Postmes. Laptop usage affects abstract reasoning of children in the developing world. Computer & Education, 59(3):989–1000, November 2012.

Annotation: Field experiment on the effect of One Laptop Per Child deployment and children school engagement and performance. Note the word grade is quite confusingly used in two senses.

[417] Klint Finley. Wiki inventor sticks a fork in his baby. Wired, April 2012.

Annotation: News article about Ward Cunningham's Smallest Federated Wiki.

- [418] Ward Cunningham. Folk memory: A minimalist architecture for adaptive federation of object servers, June 1997.
- [419] Gérôme Canals, Pascal Molli, Julien Maire, Stéphane Laurière, Esther Pacitti, and Mounir Tlili. XWiki Concerto: A P2P wiki system supporting disconnected work. In Yuhua Luo, editor, Cooperative Design, Visualization, and Engineering, Lecture Notes in Computer Science, pages 98–106. Springer, 2008.

- [420] Clemens H. Cap. Towards content neutrality in wiki systems. Future Internet, 4(4):1086–1104, 2012
- [421] Victor Grishchenko. Deep hypertext with embedded revision control implemented in regular expressions. In Proceedings of the 6th International Symposium on Wikis and Open Collaboration, page 3, New York, NY, USA, 2010. ACM.
- [422] Craig Anslow and Dirk Riehle. Lightweight enduser programming with wikis. In Wikis for Software Engineering Workshop, 2007.
- [423] Alexandros Kanterakis. PyPedia: A Python development environment on a wiki. In EuroPython, 2012.
- [424] Darren W. Logan, Massimo Sandal, Paul P. Gardner, Magnus Manske, and Alex Bateman. Ten simple rules for editing Wikipedia. PLoS Computational Biology, 6(9):e1000941, 2010.

Annotation: Guidelines for scientists on how to write on Wikipedia.

[425] Erik Moeller. Wikipedia scholarly survey results. Technical report, Wikimedia Foundation, March 2009.

Annotation: Report on a survey on science and Wikipedia among 1743 self-selected respondents.

[426] Patricia L. Dooley. Wikipedia and the two-faced professoriate. In *Proceedings of the 6th Interna*tional Symposium on Wikis and Open Collaboration, New York, NY, USA, 2010. ACM.

Annotation: Survey among university factulty for use of Wikipedia and attitude towards Wikipedia's credibility, as well as a content analysis of the use of Wikipedia in research papers

[427] Alex Bateman and Darren W. Logan. Time to underpin Wikipedia wisdom. *Nature*, 468:765, December 2010.

Annotation: Short comment that encourages scientists to contribute to Wikipedia.

[428] Alexander L. Bond. Why ornithologists should embrace and contribute to Wikipedia. IBIS, 2011.

Annotation: Short article encouraging ornithologists to contributed to Wikipedia. Professors could also ask students to create or improve Wikipedia taxonomic articles.

[429] James Heilman. Why we should all edit Wikipedia. University of British Columbia Medical Journal, 3(1):32–33, September 2011.

Annotation: Short article encouraging healthcare professionals to get involved in Wikipedia contribution.

[430] Manu Mathew, Anna Joseph, James Heilman, and Prathap Tharyan. Cochrane and Wikipedia: the collaborative potential for a quantum leap in the dissemination and uptake of trusted evidence. The Cochrane database of systematic reviews, 10:ED000069, 2013.

Annotation: Short article on the collaboration between Wikipedia editors and the Cochrane Collaboration.

[431] Sook Lim. How and why do college students use Wikipedia. Journal of the American Society for Information Science and Technology, 50(11):2189–2202, November 2009.

Annotation: Examines how college students use (i.e., read) Wikipedia.

[432] Kristina Fiore. APA: Med students cram for exams with Wikipedia. MedPage Today, May 2011.

Annotation: News article reporting on a study by M. Namdari et al., 'Is Wikipedia taking over textbooks in medical student education?' that investigated which information sources medical students uses when taking an psychiatry exam. Question books, Upto-Date and Wikipedia were used the most.

[433] Jeff Maehre. What it means to ban Wikipedia: an exploration of the pedagogical principles at stake. *College Teaching*, 57(4):229–236, 2009.

Annotation: Discusses how students could use Wikipedia.

[434] Andrea Forte and Amy Bruckman. Writing, citing, and participatory media: wikis as learning environments in the high school classroom. *International Journal of Learning and Media*, 1:23–44, Fall 2009.

Annotation: Details a project were High School students wrote science wiki articles.

[435] Neil L. Waters. Why you can't cite Wikipedia in my class. Communications of the ACM, 50(9):15– 17, September 2007.

Annotation: Historian in Japanese history from Middlebury College describes the students of his using erroneous information on Wikipedia, the Wikipedia policy subsequently adopted by his department and the media coverage. He also makes general remarks on authority, accountability and validity.

[436] Noam Cohen. A history department bans citing Wikipedia as a research source. The New York Times, February 2007.

Annotation: Middlebury College history department has banned students from using Wikipedia as a source

- [437] Cullen J. Chandler and Alison S. Gregory. Sleeping with the enemy: Wikipedia in the college classroom. The History Teacher, 43(2):247–257, February 2010.
- [438] Alistair Coleman. Students 'should use Wikipedia'. BBC NEWS, December 2007.
- [439] Lisa Spiro. Is Wikipedia becoming a respectable academic source?. Digital Scholarship in the Humanties, September 2008.

Annotation: Blog article reporting an analysis of citations from academic journals in the humanities domain area to Wikipedia.

[440] Alireza Noruzi. Wikipedia popularity from a citation analysis point of view. Webology, 6(2), June 2009.

Annotation: Analysis of citation from scientific journals to Wikipedia using ISI Web of Science.

[441] Bradley Brazzeal. Citations to Wikipedia in chemistry journals: A preliminary study. Issues in Science and Technology Librarianship, Fall 2011.

Annotation: Analysis of citations in chemistry journals that goes to Wikipedia. 370 articles were found and the citations were grouped into categories. 33 of the citations cites numerical values of physical and chemical properties corresponding to 9 percent.

[442] Declan Butler. Publish in Wikipedia or perish. Nature News, December 2008.

Annotation: Article describing a new approach taken by scientific journal RNA Biology where author are required to also submit a Wikipedia article.

[443] Shoshana J. Wodak, Daniel Mietchen, Andrew M. Collings, Robert B. Russell, and Philip E. Bourne. Topic pages: PLoS Computational Biology meets Wikipedia. PLoS Computational Biology, 8(3):e1002446, 2012.

Annotation: Presents the concept of "Topic Pages" in the scientific journal PLoS Computational Biology. It is educational academic articles published in the journal that are meant to also be publised on Wikipedia.

[444] Spencer Bliven and Andreas Prlić. Circular permutation in proteins. *PLoS Computational Biology*, 8(3):e1002445, March 2012.

Annotation: The first so-called topic page of PLoS Computational Biology.

- [445] K. Wang. Gene-function wiki would let biologists pool worldwide resources. *Nature*, 439(7076):534, February 2006.
- [446] Jim Giles. Key biology databases go wiki. Nature, 445(7129):691, February 2007.

Annotation: News report on the "Wiki for Professionals", — a biology wiki.

- [447] S. L. Salzberg. Genome re-annotation: a wiki solution?. Genome Biology, 8(1):102, February 2007.
- [448] James C. Hu, Rodolfo Aramayo, Dan Bolser, Tyrrell Conway, Christine G. Elsik, Michael Gribskov, Thomas Kelder, Daisuke Kihara, Thomas F. Knight Jr., Alexander R. Pico, Deborah A. Siegele, Barry L. Wanner, and Roy D. Welch. The emerging world of wikis. *Science*, 320(5881):1289–1290, June 2008.

Annotation: A short letter noting the existence of biological wikis, such as EcoliWiki, GONUTS, PDBwiki and WikiPathways.

[449] Finn Årup Nielsen. Lost in localization: A solution with neuroinformatics 2.0. NeuroImage, 48(1):11–13, October 2009.

Annotation: Comment on a paper by Derrfuss and Mar suggesting the use of Web 2.0 techniques for neuroinformatics databasing and presentation of the Brede Wiki.

- [450] Robert Hoehndorf, Kay Prüfer, Michael Backhaus, Heinrich Herre, Janet Kelso, Frank Loebe, and Johann Visagie. A proposal for a gene functions wiki. In On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, volume 4277 of Lecture Notes in Computer Science, pages 669–678, Berlin/Heidelberg, November 2006. Springer.
- [451] Helen Pearson. Online methods share insider tricks. Science, 441:678, June 2006.
- [452] Brandon Keim. WikiMedia. Nature Medicine, 13(3):231–233, February 2007.

Annotation: An editorial describing wikis used in medicine and biology. Wikipedia, Digital Universe, Citizendium, Ganfyd, AskDrWiki and OpenWetWare are descriped.

[453] Michael Cariaso and Greg Lennon. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. *Nucleic acids research*, 40(Database issue):D1308–D1302, January 2012.

Annotation: Description of SNPedia, which is a wiki based on Semantic MediaWiki for recording of data related to single nucleotide polymorphisms (SNPs). In August 2011 the wiki had information about approximately 25.000 SNPs.

[454] Alexander R. Pico, Thomas Kelder, Martijn P. van Iersel, Kristina Hanspers ad Bruce R. Conklin, and Chris Evelo. WikiPathways: Pathway editing for the people. *PLoS Biology*, 6(7):e184, July 2008.

Annotation: Description of the WikiPathways database, http://www.wikipathways.org/ that is related to GenMAPP pathways database and the GenMAPP Pathway Markup Language (GPML) XML format. The wiki uses extended MediaWiki software.

- [455] Yoshinobu Igarashi, Alexey Eroshkin, Svetlana Gramatikova, Kosi Gramatikoff, Ying Zhang, Jeffrey W. Smith, Andrei L. Osterman, and Adam Godzik. CutDB: a proteolytic event database. Nucleic Acids Research, 35:D546–D549, January 2007.
- [456] Todd H. Stokes, J. T. Torrance, Henry Li, and May D. Wang. ArrayWiki: an enabling technology for sharing public microarray data repositories and meta-analyses. *BMC Bioinformatics*, 9, supplement 6:S18, 2008.
- [457] Henning Stehr, Jose M. Duarte, Michael Lappe, Jong Bhak, and Dan M. Bolser. PDBWiki: added value through community annotation of the Protein Data Bank. *Database*, 2010:baq009, 2010.
- [458] Eran Hodis, Jaime Prilusky, Eric Martz, Israel Silman, John Moult, and Joel L Sussman. Proteopedia - a scientific 'wiki' bridging the rift between three-dimensional structure and function of biomacromolecules. *Genome Biology*, 9:R121, 2008.
- [459] Robert Hoffman. A wiki for the life sciences where authorship matters. Nature Genetics, 40:1047– 1051, 2008.
- [460] Barend Mons, Michael Ashburner, Christine Chichester, Erik van Mulligan, Marc Weeber, Johan den Dunnen, Gert-Jan van Ommen, Mark Musen, Matthew Cockerill, Henning Hermjakob, Albert Mons, Able Packer, Roberto Pacheco, Suzanna Lewis, Alfred Berkeley, William Melton, Micholas Barris, Jimmy Wales, Gerard Meijssen, Erik Moeller, Peter Jan Roes, Katy Borner, and Amos Bairoch. Calling on a million minds for community annotation in WikiProteins. Genome Biology, 9:R89, May 2008.

Annotation: Describes an online system that integrates a number of bioinformatics databases. The system is

based on the OmegaWiki MediaWiki extension.

[461] Euan Adie. WikiProteins — a more critical look. Nascent, June 2008.

Annotation: A blog post with a critical examination of the then just launched WikiProteins biology wiki.

[462] Emilio J. Rodríguez Posada. Wikipapers, una recopilación colaborativa de literatura sobre wikis usando MediaWiki y su extensión semántica. In IV Jornadas Predoctorales de la ESI, Spain, December 2012. Escuela Superior de Ingeniería de Cádiz.

Annotation: Spanish description of the WikiPapers wiki for structured annotation of academic papers about wiki research.

- [463] Charles Wilson. Up and down states. Scholarpedia J., 3(6):1410, January 2008.
- [464]Melissa L. Rethlefsen. Medpedia. J. Med. Libr. Assoc., 97(4):325–326, October 2009.
- [465] Noam Cohen. A Rorschach cheat sheet on Wikipedia?. The New York Times, June 2009.
- [466] Michelle Paulson. Two German courts rule in favor of free knowledge movement. Wikimedia blog, December 2012.

Annotation: Blog post on German court cases involving Wikipedia information and right of personality.

- [467] Jenny Kleeman. Wikipedia ban for disruptive professor. *The Observer*, December 2007.
- [468] Eugenie Samuel Reich. Online reputations: best face forward. *Nature*, 473:138–139, May 2011.

Annotation: Describes reputation management among scientists

[469] Nature. Survey results: Best face forward. Nature Web site, May 2011.

Annotation: Statistics from poll on online reputation management by researchers

[470] Kelly Chandler-Olcott. A tale of two tasks: editing in the era of digital literacies. *Journal of Adolescent & Adult Literacy*, 53(1):71–74, September 2009.

Annotation: Short article arguing for Wikipedia as an opportunity for teaching students about collaborative writing in digital environments.

[471] Piotr Konieczny. Wikis and Wikipedia as a teaching tool. International Journal of Instructional Technology & Distance Learning, 4(1):15–34, January 2007.

- [472] Donald MacLeod. Students marked on writing in Wikipedia. Guardian Unlimited, March 2007.
- [473] Elizabeth M. Nix. Wikipedia: how it works and how it can work for you. The History Teacher, 43(2):259–264, February 2010.

Annotation: Reports on the experience with given Wikipedia writing assignment in a history course.

- [474] Associated Press. University class has students updating Wikipedia. FOXNews.com, November 2007.
- [475] Elizabeth Ann Pollard. Raising the stakes: writing about witchcraft on Wikipedia. The History Teacher, 42(1):9–24, November 2008.
- [476] Normann Witzleb. Engaging with the world: students of comparative law write for Wikipedia. Legal Education Review, 19(1 & 2):83–97, 2009.

Annotation: Reports on a Comparative Law course where students were to write and review Wikipedia articles.

[477] Kristine L. Callis, Lindsey R. Christ, Julian Resasco, David W. Armitage, Jeremy D. Ash, Timothy T. Caughlin, Sharon F. Clemmensen, Stella M. Copeland, Timothy J. Fullman, Ryan L. Lynch, Charley Olson, Raya A. Pruner, Ernane H. M. Vieira-Neto, Raneve West-Singh, and Emilio M. Bruna. Improving Wikipedia: educational opportunity and professional responsibility. Trends in Ecology & Evolution, 24(4):177–179, April 2009.

Annotation: Short report on how participants in a graduate seminar wrote Wikipedia articles

[478] Klaus Wannemacher. Experiences and perspectives of Wikipedia use in higher education. *International Journal of Management in Education*, 5(1):79–92, 2011.

Annotation: A review of students assignment in Wikipedia contribution as they were listed on Wikipedia.

[479] Joseph Bergin. Teaching on the wiki web. In Proceedings of the 7th annual conference on Innovation and technology in computer science education, page 195. ACM, 2002.

Annotation: A short description of the use of a wiki for course communication.

[480] Norm Friesen and Janet Hopkins. Wikiversity; or education meets the free culture movement: An etnographic investigation. First Monday, 13(10), October 2008.

Annotation: Describes the experiences from giving a course through Wikiversity.

- [481] John L. Hilton III and David A. Wiley. A sustainable future for open textbooks? the flat world knowledge story. *First Monday*, 15(8), August 2010
- [482] Suthiporn Sajjapanroj, Curtis J. Bonk, Mimi Miyoung Lee, and Meng-Fen Grace Lin. A window to Wikibookians: Surveying their statuses, successes, satisfactions, and sociocultural experiences. *Journal of Interactive Online Learning*, 7(1):36–58, Spring 2008.

Annotation: Survey among 80 contributors to the Wikibooks Wikimedia project as well as email interview with 15 individuals selected among the 80. Among the several results were that contributors/respondents tended to be young males.

[483] Gilad Ravid, Yoram M. Kalman, and Sheizaf Rafaeli. Wikibooks in higher education: Empowerment through online distributed collaboration. Computers in Human Behavior, 24(5):1913–1928, 2008.

Annotation: Reports on the extension and update of a textbook with wiki technology where students participate. The researchers found that for some classes the student participation was associated with higher grades on average.

[484] William Shockley. On the statistics of individual variations of productivity in research laboratories. Proceedings of the IRE, 45(3):279–289, March 1957.

Annotation: Study on the productivity of scientists by counting number of publications, finding a heavy-tailed log-normal distribution. Explanations for the phenomenon are put forward.

- [485] H. Sackman, W. J. Erikson, and E. E. Grant. Exploratory experimental studies comparing online and offline programming performance. *Communications of the ACM*, 11(1):3–11, January 1968.
- [486] Bill Curtis. Substantiating programmer variability. *Proceedings for the IEEE*, 69(7):846, July 1981.

Annotation: Small study on programmer productivity

[487] Alissa Skelton. Wikipedia volunteer editor reaches 1 million edits. Mashable, April 2012.

Annotation: Report on Justin Knapp that was the first to reach one million edits on Wikipedia.

[488] Dell Zhang. Wikipedia edit number prediction based on temporal dynamics only. ArXiv, October 2011. Annotation: Describes a machine learning-based system for predicting future number of edits of Wikipedia contributors

[489] Yutaku Yoshida and Hayato Ohwada. Wikipedia edit number prediction from the past edit record based on auto-supervised learning. In 2012 International Conference on Systems and Informatics (ICSAI), pages 2415–2419. IEEE, May 2012.

Annotation: Describes a machine learning-based system for predicting number of edits a contributor make on Wikipedia.

[490] Piotr Konieczny. We are drowning in promotional artspam. The Signpost, April 2015.

Annotation: Op-ed on the problem of promotional articles on Wikipedia.

- [491] Gamaliel. Sony emails reveal corporate practices and undisclosed advocacy editing. The Signpost, April 2015.
- [492] Ted Greenwald. How Jimmy Wales' Wikipedia harnessed the web as a force for good. Wired, March 2013.

Annotation: Interview with Jimmy Wales about Wikipedia.

[493] Mikael Häggström. Average weight of a conventional teaspoon made of metal. Wikiversity, November 2012.

Annotation: Example on a peerreviewed work published on a Wikimedia Foundation wiki. The small study weighed 19 different teaspoons and found an average weight on 25 grams.

- [494] Fiona Fidler and Ascelin Gordon. Science is in a reproducibility crisis: How do we resolve it?. Phys.org, September 2013.
- [495] John P. A. Ioannidis. Why most published research findings are false. PLoS Medicine, 2(8):e124, August 2005.

Annotation: Theoretical considerations and simulations

[496] Steven Goodman and Sander Greenland. Why most published research findings are false: Problems in the analysis. *PLoS Medicine*, 4(4):e168, April 2007.

Annotation: Critique of Ioannidis papers, e.g., the critique notes that Ioannidis is wrong when he only uses 0.05 for his simulation for his claim and not the actual p-value observed in the studies.

[497] Steven Goodman and Sander Greendland. Assessing the unreliability of the medical literature: A response to "why most published research findings are false". Working Paper 135, Johns Hopkins University, Dept. of Biostatistics, 2007.

Annotation: This is the indepth critique of Ioannidis paper summarized on an article in PLoS Medicine.

- [498] John P. A. Ioannidis. Why most published research findings are false: Author's reply to goodman and greenland. *PLoS Medicine*, 4(6):e215, June 2007.
- [499] John P. A. Ioannidis. Contradicted and initially stronger effects in highly cited clinical research. JAMA, 294(2):218–228, July 2005.

Annotation: Study on whether previous highly-cited clinical research papers are replicated.

[500] Xiaoquan Michael Zhang and Feng Zhu. Group size and incentives to contribute: a natural experiment at Chinese Wikipedia. American Economic Review, 101(4):1601–1615, June 2011.

Annotation: A study of what effect the censoring of Wikipedia in China had on the development of the Chinese Wikipedia.

[501] Nima Nazeri and Collin Anderson. Citation filtered: Iran's censorship of Wikipedia. Annenberg School for Communication at University of Pennsylvania, 2013.

Annotation: Report of an analysis of the Persian Wikipedia articles that Iranian authorities block.

- [502] Noam Cohen. Microsoft Encarta dies after long battle with Wikipedia. New York Times, March 2009.
- [503] Shane Greenstein. The reference wars: Encyclopædia Britannica's decline and encarta's emergence. Strategic Management Journal, 2016.
- [504] Mark Landler. Slow-to-adapt Encyclopaedia Britannica is for sale. New York Times, May 1995.

Annotation: News article on the troubling economy of Encyclopaedia Britannica.

[505] Dale Sheppard. Reading with iPads — the difference makes a difference. Education Today, pages 12–15, Term 3 2011.

Annotation: Study on reading comprehension with 43 Year 6 pupils where half was reading on texts on iPads.

[506] Maria Becher Trier. Pisa: Adgang til tablets gør ikke nødvendigvis elever dårligere. Folkeskolen.dk, December 2012. Annotation: Danish news article about the result of the 2012 Programme for International Student Assessment (PISA) in Denmark and the score of pupils in regard to whether they have access to tablet computers or not.

[507] Mark Graham. The problem with Wikidata. The Atlantic, April 2012.

Annotation: Article discussing Wikidata and its possible influence on the diversity of knowledge represented in Wikipedia.

- [508] Bernard Jacquemin, Aurelien Lauf, Céline Poudat, Martine Hurault-Plantet, and Nicolas Auray. Managing conflicts between users in Wikipedia. In Dominik Flejter, Slawomir Grzonkowski, Tomasz Kaczmarek, Marek Kowalkiewicz, Tadhg Nagle, and Jonny Parkes, editors, BIS 2008 Workshop Proceedings, Innsbruck, Austria, 6-7 May 2008, pages 81–93. Department of Information Systems, Poznań University of Economics, 2008.
- [509] Andreas Brändle. Too many cooks don't spoil the broth. In *Proceedings of Wikimania 2005*, 2005.
- [510] Sander Spek, Eric Postma, and H. Jaap van den Herik. Wikipedia: organisation from a bottom-up approach. From workshop Research in Wikipedia, on the Wikisym 2006, December 2006.
- [511] Andreas Neus. Managing information quality in virtual communities of practice. In E. Pierce and R. Katz-Haas, editors, Proceedings of the 6th International Conference on Information Quality at MIT, Boston, MA, 2001. Sloan School of Management.
- [512] Jenneke Fokker, Johan Pouwelse, and Wray Buntine. Tag-based navigation for peer-to-peer Wikipedia. Collaborative Web Tagging Workshop, WWW2006, 2006.
- [513] Nikolaos Th. Korfiatis, Marios Poulos, and George Bokos. Evaluating authoritative sources using social networks: an insight from Wikipedia. Online Information Review, 30(3):252–262, 2006.
- [514] Ulrike Pfeil, Panayiotis Zaphiris, and Chee Siang Ang. Cultural differences in collaborative authoring of Wikipedia. Journal of Computer-Mediated Communication, 12(1):88–113, October 2006.
- [515] Klaus Stein and Claudia Hess. Does it matter who contributes: a study on featured articles in the German Wikipedia. In Proceedings of the 18th Conference on Hypertext and Hypermedia, pages 171–174, New York, NY, USA, 2007. ACM.
- [516] Aniket Kittur, Ed Chi, and Todd Mytkowicz Bryan A. Pendleton, Bongwon Suh. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In Computer/Human Interaction 2007. ACM SIGCHI, 2007. This is not found in the CHI 2007 proceedings.

- [517] Fernanda B. Viégas, Martin Wattenberg, and Matthew M. McKeon. The hidden order of Wikipedia. In *Lecture Notes in Computer Sci*ence, volume 4564, pages 445–454, Berlin/Heidelberg, 2007. Springer.
- [518] Niels Mølbjerg Lund Pedersen and Anders Due. Wikipedia — viden som social handlen. Master's thesis, Department of Media, Cognition and Communication, University of Copenhagen, Copenhagen, Denmark, August 2006.
- [519] Anders J. Jørgensen. Den epistemiske hensigtsmæssighed af åbne kontra lukkede videnskabende samarbejdsformer med speciel hensynstagen til peer-reviews og internetfænomenet Wikipedia, January 2007.
- [520] Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells. Automatic extraction of semantic relationships for WordNet by means of pattern learning from Wikipedia. In Natural Language Processing and Information Systems, volume 3513 of Lecture Notes in Computer Science, pages 67–79, Berlin/Heidelberg, May 2005. Springer.
- [521] Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells. Automatic assignment of Wikipedia encyclopedic entries to WordNet synsets. In Advances in Web Intelligence, volume 3528 of Lecture Notes in Computer Science, pages 380–386, Berlin/Heidelberg, 2005. Springer.
- [522] Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells. Automatising the learning of lexical patterns: An application to the enrichment of WordNet by extracting semantic relationships from Wikipedia. Data & Knowledge Engineering, 61(3):484–499. June 2007.
- [523] Evgeniy Gabrilovich and Shaul Markovitch. Overcoming the brittleness bottleneck using Wikipedia: Enhancing text categorization with encyclopedic knowledge. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pages 1301–1306, Menlo Park, California, 2006. AAAI Press.
- [524] Evgeniy Gabrilovich and Shaul Markovitch. Computing semantic relatedness using Wikipedia-based explicit sematic analysis. In Proceedings of The Twentieth International Joint Conference for Artificial Intelligence, pages 1606–1611, 2007.
- [525] Denise Anthony, Sean W. Smith, and Tim Williamson. The quality of open source production: Zealots and good samaritans in the case of Wikipedia. Technical Report TR2007-606, Department of Computer Science, Dartmouth College, Hanover, New Hampshire, April 2007.
 - Annotation: An investigation of 7058 contributors on the French and Dutch Wikipedias, to answer whether contributions from registered users

- have higher quality than anonymous users. Each user is measured wrt. retention of their added text, while not considering, e.g., edit wars. They find that the retention is higher for anonymous users.
- [526] Deborah L. McGuinness, Honglei Zeng, Paulo Pinheiro da Silva, Li Ding, Dhyanesh Narayanan, and Mayukh Bhaowal. Investigations into trust for collaborative information repositories: A Wikipedia case study. In Proceedings of the Workshop on Models of Trust for the Web, Edinburgh, United Kingdom, May 2006.
- [527] Brian Bergstein. New tool measures Wikipedia entries. USATODAY.com, September 2007.
- [528] Thomas Rune Korsgaard. Improving trust in the Wikipedia. Master's thesis, Technical University of Denmark, Kongens Lyngby, Denmark, 2007.
- [529] Tor Nørretranders. Det generøse menneske: En naturhistorie om at umage giver mage. People'sPress, 2006.
 - Annotation: A book dubbing 'Homo generosus' and discussing issues such as altruism, gift economy, generosity, reciprocity, cooperation, moral, costly signals, the handicap principle, sexual selection and economic games.
- [530] Josef Urban, Jesse Alama, Piotr Rudnicki, and Herman Geuvers. A wiki for Mizar: Motivation, considerations, and initial prototype. In Serge Autexier, Jacques Calmet, David Delahaye, Patrick Ion, Laurence Rideau, Renaud Rioboo, and Alan Sexton, editors, Intelligent Computer Mathematics, volume 6167 of Lecture Notes in Computer Science, pages 455-469, Springer Berlin / Heidelberg, 2010. Springer.
- [531] Philip Ball. The more, the wikier. news@nature.com, February 2007.
- [532] Jo Brand. Verden i følge Wikipedia forandrer sig fra dag til dag. *Bibliotekspressen*, 3:10–14, 2005.
- [533] Gregory Druck, Gerome Miklau, and Andrew Mc-Callum. Learning to predict the quality of contributions to wikipedia. In Proceedings of the AAAI Workshop on Wikipedia and Artificial Intelligence (WIKIAI 08, pages 7–12. Association for the Advancement of Artificial Intelligence, 2008.
- [534] T. Hassine. The dynamics of NPOV disputes. In Proceedings of Wikimania 2005—The First International Wikimedia Conference, 2005.
- [535] Dennis Hoppe. Automatische erkennung von bearbeitungskonflikten in Wikipedia. Abschlussarbeit, Baus-Universitä Weimar, Weimar, February 2008.
- [536] Jonathan Isbell and Mark H. Butler. Extracting and re-using structured data from wikis. Technical Report HPL-2007-182, Digital Media Systems Laboratory, Bristol, November 2007.

- [537] Marek Meyer, Christoph Rensing, and Ralf Steinmetz. Categorizing learning objects based on Wikipedia as substitute corpus. In First International Workshop on Learning Object Discovery & Exchange (LODE'07). FIRE/LRE, 2007.
- [538] A. J. Reinoso, Felipe Ortega, J. M. Gonzalez-Barahona, and G. Robles. A quantitative approach to the use of the Wikipedia. In *IEEE Symposium on Computers and Communications*, 2009. ISCC 2009, pages 56–61, 2009.
- [539] Mattias Beerman. Collaborative web content management — wiki and beyond. Master's thesis, Department of Computer and Information Science, Linköpings University, December 2005.
- [540] Ian J. Deary. Intelligence: a very short introduction. Oxford University Press, Oxford, UK, 2001.

A Notes

There are numerous other studies on wikis, Wikipedia and related areas not summarized here. $^{508-530}$

The following are some notes and discovered links not yet incorporated into the body text:

A.1 Publications

- Sisay Adafre et al. (2007). Fact Discovery in Wikipedia.
- Maik Anderka, Benno Stein (2012). A Breakdown of Quality Flaws in Wikipedia, WebQuality.
- Denise L. Anthony et al. "Reputation and reliability in collective goods: the case of the online encyclopedia Wikipedia."
- 4. David Aumueller (2005). Semantic authoring and retrieval within a Wiki
- P. Ayers, "Researching Wikipedia current approaches and new directions," Proceedings of the American Society for Information Science and Technology, vol. 43, 2006, pp. 1–14.
- 6. Ball (2007). The more, the wikier. 531
- Hoda Baytiyeh , Jay Pfaffman (2009). Why be a Wikipedian, Proceedings of the 9th international conference on Computer supported collaborative learning, p.434-443, June 08-13, 2009, Rhodes, Greece
- 8. Robert P. Biuk-Aghai et al. (2009). Wikis as Digital Ecosystems: An Analysis Based on Authorship
- 9. Robert P. Biuk-Aghai, Cheong-Iao Pang and Yain-Whar Si (2014). Visualizing large-scale human collaboration in Wikipedia, Future Generation Computer Systems, Volume 31, pp. 120-133
- Alexandre Bouchard, Percy Liang, Thomas Griffiths, and Dan Klein. 2007. A probabilistic approach to diachronic phonology. In Proceedings of the 2007. In Proceedings of EMNLP-CoNLL, pages 887â @ S896.
- Jo Brand, Verden i f
 ølge Wikipedia forandrer sig fra dag til dag.⁵³² Drawing the attention of librarians
- Brändle, Andreas (2005): Zu wenige Köche verderben den Brei.
- 13. Michael Buffa et al. (2008). SweetWiki: A semantic wiki. Web Semantics: Science, Services and Agents on the World Wide Web Volume 6, Issue 1, February 2008, Pages 84-97. Semantic Web and Web 2.0. This is related to a previous conference contribution.³⁷⁶
- 14. Moira Burke, Robert Kraut (2008). Mopping up: modeling wikipedia promotion decisions
- 15. Jon Chamberlain, Udo Kruschwitz and Massimo Poesio. Constructing an Anaphorically Annotated Corpus with Non-Experts: Assessing the Quality of Collaborative Annotations

- Zhining Chen et al. (2010). Web Video Categorization based on Wikipedia Categories and Content-Duplicated Open Resources.
- 17. Paula Chesley, Bruce Vincent, Li Xu, and Rohini Srihari. 2006. Using verbs and adjectives to automatically classify blog sentiment. In Proceedings of AAAI-CAAW-06, the Spring Symposia on Computational Approaches to Analyzing Weblogs.
- Empirical Analysis of User Participation in Online Communities: the Case of Wikipedia. Giovanni Luca Ciampaglia, Alberto Vancheri
- 19. (About gender on Wikipedia). Collier et at.
- 20. Kino Coursey, Rada Mihalcea (2009). Topic identification using Wikipedia graph centrality. NAACL-Short '09 Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics.
- 21. Cragun, R. (2007). The future of textbooks? Electronic Journal of Sociology.
- Catalina Danis , David Singer (2008). A wiki instance in the enterprise: opportunities, concerns and reality, Proceedings of the ACM 2008 conference on Computer supported cooperative work, November 08-12, 2008, San Diego, CA, USA.
- DeHoust, C., Mangalath, P., Mingus., B. (2008).
 Improving search in Wikipedia through quality and concept discovery. Technical Report.
- Sylvain Dejean, Nicolas Jullien (2014). Big From the Beginning. Assessing Online Contributors' Behavior by Their First Contribution. Analysis of newcomers.
- Gregory Druck and Gerome Miklau and Andrew McCallum (2008). Learning to Predict the Quality of Contributions to Wikipedia⁵³³
- 26. Donghui Feng, Sveva Besana and Remi Zajac. Acquiring High Quality Non-Expert Knowledge from On-Demand Workforce
- 27. Bowei Du, Eric A. Brewer (2008). DTWiki: A Disconnection and Intermittency Tolerant Wiki
- 28. Förstner et al. Collaborative platforms for streamlining workflows in Open Science.
- Andrea Forte , Amy Bruckman (2007). Constructing text:: Wiki as a toolkit for (collaborative?) learning, Proceedings of the 2007 international symposium on Wikis, p.31-42, October 21-25, 2007, Montreal, Quebec, Canada
- Andrea Forte, Cliff Lampe (2013), Defining, Understanding and Supporting Open Collaboration: Lessons from the Literature
- E. Gabrilovich and S. Markovich (Israel Institute of Technology): "Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis", 2007

- Zeno Gantner and Lars Schmidt-Thieme. Automatic Content-Based Categorization of Wikipedia Articles
- 33. David Garcia, Pavlin Mavrodiev, Frank Schweitzer (2013). Social Resilience in Online Communities: The Autopsy of Friendster. http://arxiv.org/abs/1302.6109
- 34. E Graells-Garrido, M Lalmas, F Menczer (2015). First Women, Second Sex: Gender Bias in Wikipedia
- Andrew Gregorowicz and Mark A. Kramer (2006).
 Mining a Large-Scale Term-Concept Network from Wikipedia
- Guth, S. (2007). Wikis in education: Is public better? Proceedings of the 2007 international symposium on Wikis, 61–68.
- Ben Hachey et al. (2013). Evaluating Entity Linking with Wikipedia.
- 38. Aaron Halfaker, Aniket Kittur, John Riedl (2011). Don't bite the newbies: how reverts affect the quantity and quality of Wikipedia work.
- 39. Haigh. Wikipedia as an evidence source for nursing and healthcare students
- 40. Laura Hale (2013). User:LauraHale/Blocks on English Wikinews
- 41. Rainer Hammwöhner (Uni Regensburg): "Qualitätsaspekte der Wikipedia", 2007
- S. Hansen, S., Berente, N., and K. Lyytinen (2007). Wikipedia as rational discourse: An illustration of the emancipatory potential of information systems. 40th Annual Hawaii International Conference on System Sciences, 2007. HICSS 2007.
- 43. Rudolf den Hartogh (2015). The future of the past.
- 44. T. Hassine (2005). The Dynamics of NPOV disputes 534
- A. J. Head and M. B. Eisenberg, 2010, How today's college students use Wikipedia for courserelated research
- 46. Heilmann et al.: Wikipedia: a key tool for global public health promotion
- James M. Heilmann, Andrew G. West (2015).
 Wikipedia and Medicine: Quantifying Readership,
 Editors, and the Significance of Natural Language.
- 48. Benjamin Mako Hill. Cooperation in parallel: a tool for supporting collaborative writing in diverged documents.
- Matthew Honnibal, Joel Nothman and James R. Curran. Evaluating a Statistical CCG Parser on Wikipedia
- 50. Dennis Hoppe (2008). Automatische Erkennung von Bearbeitungskonflikten in Wikipedia 535
- Gerardo Iñiguez, Taha Yasseri, Kimmo Kaski, János (2014). Kertész. Modeling Social Dynamics in a Collaborative Environment.

- 52. Ironholds (2013). Why are users blocked on Wikipedia? http://blog.ironholds.org/?p=31
- Jonathan Isbell, Mark H. Butler (2007). Extracting and Re-using Structured Data from Wikis⁵³⁶
- Jordan, Watters (2009). Retrieval of Single Wikipedia Articles While Reading Abstracts.
- 55. Richard Jensen, "Military history on the electronic frontier: Wikipedia fights the was of 1812". Link
- 56. Elisabeth Joyce et al. (2011). Handling flammable materials: Wikipedia biographies of living persons as contentious objects.
- 57. Elisabeth Joyce et al. (2013). Rules and Roles vs. Consensus: Self-governed Mass Collaborative Deliberative Bureaucracies
- 58. Fariba Karimi, Ludvig Bohlin, Ann Samoilenko, Martin Rosvall, Andrea Lancichinetti (2015). Quantifying national information interests using the activity of Wikipedia editors.
- Brian Keegan et al. (2013). Hot off the Wiki: Structures and Dynamics of Wikipedia's Coverage of Breaking News Events
- 60. Maximilian Klein, Piotr Konieczny. Gender gap through time and space: a journey through Wikipedia biographics and the "WIGI" index.
- 61. Joachim Kimmerle , Johannes Moskaliuk , Ulrike Cress, Understanding learning: the Wiki way, Proceedings of the 5th International Symposium on Wikis and Open Collaboration, October 25-27, 2009, Orlando, Florida
- 62. Frank Kleiner, Andreas Abecker, Sven F. Brinkmann. WiSyMon: managing systems monitoring information in semantic Wikis, Proceedings of the 5th International Symposium on Wikis and Open Collaboration, October 25-27, 2009, Orlando, Florida
- 63. René König, "Wikipedia. Between lay participation and elite knowledge representation".
- 64. Krahn et al. (2009). Lively Wiki. A development environment for creating and sharing active web content.
- Markus Krötzsch, Denny Vrandecic, Max Völkel (2005). Wikipedia and the Semantic Web-The Missing Links. Semantic wiki.
- 66. S. Kuznetsov (2006). Motivations of contributors to Wikipedia.
- 67. Pierre-Carl Langlais (2013). Is Wikipedia a relevant model for e-learning?
- Kangpyo Lee et al. (2008). FolksoViz: A Subsumption-based Folksonomy Visualization Using Wikipedia Texts.
- Pentti Launonen, K. C. Kern, Sanna Tiilikainen (2015). Measuring Creativity of Wikipedia Editors.
- Yinghao Li et al. (2007). Improving weak ad-hoc queries using Wikipedia as external corpus

- David Lindsey. Evalutating quality control of Wikipedia's feature articles.
- Meyer et al. (2007). Categorizing Learning Objects Based On Wikipedia as Substitute Corpus.
 To help categorize unlabeled learning objects
- 73. Mietchen et al. 2011, Wikis in scholarly publishing
- Rada Mihalcea (University of North Texas): "Using Wikipedia for Automatic Word Sense Disambiguation", 2007
- 75. Rada Mihalcea and Andras Csomai. Wikify!: Linking Documents to Encyclopedic Knowledge. Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, CIKM 2007.
- N. Miller, "Wikipedia revisited," ETC: A Review of General Semantics, vol. 64, Apr. 2007, pp. 1 47–150.
- 77. Milne, Witten. Learning to link with Wikipedia
- 78. Minol K, Spelsberg G, Schulte E, Morris N. Portals, blogs and co.: the role of the Internet as a medium of science communication.
- Joseph Morris, Chris Lüer, DistriWiki: A Distributed Peer-to-Peer Wiki, Proceedings of the 2007 International Symposium on Wikis (WikiSym), Montréal, 2007, pages 69-74. PDF
- Muchnik L, Itzhack R, Solomon S, Louzoun Y. Self-emergence of knowledge trees: extraction of the Wikipedia hierarchies. Phys Rev E Stat Nonlin Soft Matter Phys. 2007 Jul;76(1 Pt 2):016106. Epub 2007 Jul 13.
- 81. Wyl McCully, Cliff Lampe, Chandan Sarkar, Alcides Velasquez, and Akshaya Sreevinasan (2011). Online and offline interactions in online communities. In Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11). ACM, New York, NY, USA, 39-48.
 Link
- 82. Timme Bisgaard Munk (2009), Why Wikipedia: self-efficacy and self-esteem in a knowledgepolitical battle for an egalitarian epistemology
- Kataro Nakayama, Takahiro Hara, and Shojiro Nishio (University of Osaka): "Wikipedia Mining for an Association Web Thesaurus Construction", 2007
- 84. Emmanuel Navarro, Franck Sajous, Bruno Gaume, Laurent Prévot, ShuKai Hsieh, Ivy Kuo, Pierre Magistry and Chu-Ren Huang. Wiktionary for Natural Language Processing: Methodology and Limitations
- 85. Oded Nov (2007). What motivates Wikipedians?²⁷⁰
- 86. Mathieu O'Neil, Cyberchiefs: Autonomy and Authority in Online Tribes.
- 87. Enrique Orduña-Malea, José-Antonio Ontalba-Ruipérez (2013). Selective linking from social platforms to university websites: a case study of the

- Spanish academic system. Webometrics including Wikipedia
- 88. Felipe Ortega et al. (2008). On the Inequality of Contributions to Wikipedia.
- Ayelet Oz, Legitimacy and efficacy: The blackout of Wikipedia.
- 90. Manuel Palomo-Duarte, Juan Manuel Dodero, Inmaculada Medina-Bulo, Emilio J. Rodríguez-Posada, Iván Ruiz-Rube (2012). Assessment of collaborative learning experiences by graphical analysis of wiki contributions
- 91. Rafaeli and Arial (2008), Online Motivational Factors: Incentives for Participation and Contribution in Wikipedia
- Rassbach, L., Mingus., B, Blackford, T. (2007).
 Exploring the feasibility of automatically rating online article quality. Technical Report. http://grey.colorado.edu/mingus
- R. D. M. Page, 2010, Wikipedia as an encyclopedia of life.
- 94. K.R. Parker and J.T. Chao, Wiki as a teaching tool, Interdisciplinary Journal of Knowledge and Learning Objects 3 (2007), pp. 57–72.
- 95. Ulrike Pfeil et al. (2006). Cultural Differences in Collaborative Authoring of Wikipedia 514
- 96. Giacomo Poderi (2009). Comparing featured article groups and revision patterns correlations in Wikipedia
- 97. Rafaeli, S., Hayat, T., and Ariel, Y. (2005). Wikipedia community: Users' motivations and knowledge building. Paper presented at the cyberculture 3rd global conference, Prague, Czech Republic, August.
- 98. Jessica Ramírez, Masayuki Asahara, Yuji Matsumoto (2013). Japanese-Spanish Thesaurus Construction Using English as a Pivot
- 99. Gilad Ravid, Yoram M. Kalman, Sheizaf Rafaeli (2008). Wikibooks in higher education: Empowerment through online distributed collaboration Computers in Human Behavior, Volume 24, Issue 5, September 2008, Pages 1913-1928
- Colleen A. Reilly (2011), Teaching Wikipedia as a mirrored technology
- 101. Reinoso et al., A quantitative approach to the use of the Wikipedia. 538
- 102. Antonio J. Reinoso, Juan Ortega-Valiente, Rocío Munñoz-Mansilla, Gabriel Pastor (2013). Visitors and contributors in Wikipedia
- 103. Christina Sauper and Regina Barzilay (2009). Automatically Generating Wikipedia Articles: A Structure-Aware Approach. ACL 2009.
- 104. Scheider et al (2010). A qualitative and quantitative analysis of how Wikipedia talk pages are used (same as "A content analysis: how Wikipedia talk pages are used")

- 105. Ralph Schoeder, Linnet Taylor (2015). Big data and Wikipedia research: social science knowledge across disciplinary divides.
- 106. Philipp Singer, Thomas Niebler, Markus Strohmaier, Andreas Hotho (2014). Computing Semantic Relatedness from Human Navigational Paths: A Case Study on Wikipedia.
- 107. Schroer, J.; Hertel, G. (2009). Voluntary engagement in an open web-based encyclopedia: Wikipedians and why they do it. Media Psychol. 12: 96–120.
- 108. Thamar Solorio, Ragib Hasan, Mainul Mizan (2013) Sockpuppet Detection in Wikipedia: A Corpus of Real-World Deceptive Writing for Linking Identities
- 109. Aline Soules (2015). Faculty Perception of Wikipedia in the California State University System
- 110. A. Souzis (2005). Building a semantic wiki. IEEE Intelligent Systems, pp. 87–91, Sep. 2005.
- 111. Valentin I. Spitkovsky, Angel X. Chang (2012). A Cross-Lingual Dictionary for English Wikipedia Concepts. Eighth International Conference on Language Resources and Evaluation (LREC 2012). http://research.google.com/pubs/archive/38098.pdf
- Michael Strube and Simone Ponzetto (EML Research): "Creating a Knowledge Base from a Collaboratively Generated Encyclopedia", 2007
- 113. Sangweon Suh , Harry Halpin , Ewan Klein (2006). Extracting common sense knowledge from Wikipedia
- 114. Klaus Stein, Claudia Hess (2007). Does It Matter Who Contributes? A Study on Featured Articles in the German Wikipedia. HT-07.
- 115. Klaus Stein, Claudia Hess (2008). Viele Autoren, gute Autoren? Eine Untersuchung ausgezeichneter Artikel in der deutschen Wikipedia. Web 2.0 Eine empirische Bestandsaufnahme 2008, pp 107-129
- 116. Olof Sundin. Debating Information Control in Web 2.0: The Case of Wikipedia vs. Citizendium.
- 117. Edna Tal-Elhasid and Meishar-Tal, 2007 Tal-Elhasid, E., and Meishar-Tal, H. (2007). Models for activities, collaboration and assessment in wiki in academic courses. Paper presented at the EDEN 2007 annual conference, Naples, Italy.
- Robert Tolksdorf and Elena Simperl (2006). Towards Wikis as semantic hypermedia. Semantic wiki.
- 119. Noriko Tomuro and Andriy Shepitsen. Construction of Disambiguated Folksonomy Ontologies Using Wikipedia
- 120. B. Ulrik, J. Jürgen (2007). Revision and corevision in Wikipedia. Pro Learn International Workshop Bridging the Gap Between Semantic Web and Web 2.0, 4th European Semantic Web Conference.

- 121. Benjamin Mako Hill (2013). Essays on volunteer mobilization in peer production
- 122. Timothy Weale, Chris Brew and Eric Fosler-Lussier. Using the Wiktionary Graph Structure for Synonym Detection
- 123. C. Wartena, R. Brussee (2008). Topic Detection by Clustering Keywords. Database and Expert Systems Application, 2008. DEXA '08. 19th International Workshop on
- 124. Daniel S. Weld et al. Intelligence in Wikipedia. Twenty-Third Conference on Artificial Intelligence (AAAI), 2008.
- 125. A. G. West, A. J. Aviv, J. Chang, I. Lee, Mitigating spam using spatiotemporal reputation, Technical report UPENN 2010.
- West and Williamson (2009). Wikipedia friend or foe.
- J. Willinsky, 2007. "What open access research can do for Wikipedia," First Monday, volume 12, number 3.
- 128. Thomas Wozniak (2012). Zehn Jahre Berührungsängste: Die Geschichtswissenschaften und die Wikipedia. Eine Be-standsaufnahme. : Zeitschrift für Geschichtswissenschaften 60/3, S. 247-264.
- 129. Li Xu (2007) Project the wiki way: using wiki for computer science course project management, Journal of Computing Sciences in Colleges, v.22 n.6, p.109-116, June 2007
- Mark Yatskar. For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia.
- 131. Ziqi Zhang and Jose Iria. A Novel Approach to Automatic Gazetteer Generation using Wikipedia
- 132. Torsten Zesch, Christof Mueller and Iryna Gurevych Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary Proceedings of the Conference on Language Resources and Evaluation (LREC), 2008.²⁹⁷ http://www.ukp.tu-darmstadt.de/software/jwpl/ http://www.ukp.tu-darmstadt.de/software/jwktl/
- 133. Zesch, T.; Mueller, C. and Gurevych, I. Using Wiktionary for Computing Semantic Relatedness. In Proceedings of AAAI, 2008
- 134. Zhang and Zhu, Intrinsic Motivation of Open Content Contributors: the Case of Wikipedia
- 135. Haifeng Zhao, William Kallander, Tometi Gbedema, Henric Johnson, Felix Wu. Read what you trust: An open wiki model enhanced by social context.

A.2 Further notes

• Aaron Halfaker (2013). Research: Wikipedia article creation

- Freebase, http://www.freebase.com/ http://www.wikicfp.com/ http://dbpedia.org/sparql
- Google Squared and Google Fusion Tables.
- International Music Score Library Project
- http://debategraph.org/
- Searching Wikipedia WikiWax "Semantic Wiki Search".
- Wikipedia as a teaching tool workshop.odp
- WikiXMLDB http://wikixmldb.dyndns.org is an XML database using Sedna
- Visualization researchers have used the many kinds of statistics extracted from Wikipedia in large-scale renderings of network indegrees. http://scimaps.org/maps/wikipedia/20080103/
 "An Emergent Mosaic of Wikipedian Activity" http://scimaps.org/maps/wikipedia/ New Scientist, May 19, 2007.
- The collaborative debugging tool http://pastebin.com/ with syntax highlightning.
- http://www.mkbergman.com/?p=417
- WikiMedia Foundation has maintained a computer system with fast response time. BSCW was slow
- User authentication "real name" legislation in South Korea. Link
- Research to the people
- Web service for generation of SIOC files for MediaWiki page
- http://fmtyewtk.blogspot.com/2009/10/mediawiki-git-word-level-blaming-one.html
- Survey of tools for collaborative knowledge construction and sharing.
- ... and wikis and Wikipedia have been the subject of many theses.⁵³⁹
- Game theory: contributing to the common good increase the value more than the contribution. Other apparent irrational behavior: voting.
 Performance on a work sample test is among the best predictor for job performance.⁵⁴⁰
 "people like to see their words in print".
- Xanadu
- Poor Man's Checkuser
- Visualizations: Link
- ZooKeys, Species ID wiki
- Encyclopedia of Earth, Environmental Health Perspectives
- Useful Chemistry
- Redaktionen editorial team

B Further links

https://meta.wikimedia.org/wiki/Research:CSCW_2015

https://blog.wikimedia.org/2015/02/03/who-links-to-wikipedia/ http://www.floatingsheep.org/2009/11/mapping-wikipedia.html http://www.floatingsheep.org/2010/11/geographies-of-wikipedia-in-uk.html http://en.wikipedia.org/wiki/Wikipedia:EDITORS#Demographics http://libresoft.es/Members/jfelipe/thesis-wkp-quantanalysis http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Vandalism_studies NOAM COHEN, Don't Like Palin's Wikipedia Story? Change It, http://www.nytimes.com/2008/09/01/technology/01link.html Wikilink manipulation. http://www.bluehatseo.com/how-to-overthrow-a-wikipedia-result/ http://grey.colorado.edu/mingus/index.php/Wikipedia_Quality_Pageviews_Correlation_Coefficient http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2229842 http://www.sitepoint.com/blogs/2008/09/04/just-how-powerful-is-wikipedia/ http://www.economist.com/search/PrinterFriendly.cfm?story_id=11484062 http://wikimania2010.wikimedia.org/wiki/Submissions/Personality_traits_of_Polish_Wikipedia_members Amichai {Hamburger, Lamdan, Madiel and Hayat in 2008 http://www.searchcrystal.com/versions/wikipedia/home.html 'In case anyone is interested, we tried to statistically map freebase types to dbpedia types, the key ideas are described here: http://domino.watson.ibm.com/library/cyberdig.nsf/papers/4D84639C32795569852574FD005EA539/\$File/rc24684.pdf A variant of this paper was just accepted to the International Semantic Web Conference (ISWC).' http://meta.wikimedia.org/wiki/Wikimedia_Foundation_Research_Goals $\verb|http://sunsite.rediris.es/mirror/WKP_research/DE_logging_200906.sql.gz|$ Kevin Clauson receiving hate mail. http://en.wikipedia.org/w/index.php?title=Wikipedia_talk:WikiProject_Medicine&curid=3209315&diff=255360019&oldid=255355224 http://thecaucus.blogs.nytimes.com/2008/07/19/the-sam-adams-project/ http://www.wikisym.org/ws2009/tiki-index.php?page=Corporate+Wiki+Metrics http://wikimania2007.wikimedia.org/wiki/Proceedings:AD1 http://meta.wikimedia.org/wiki/Edits_by_project_and_country_of_origin http://www.brooklynrail.org/2008/06/express/nobodys-safe-in-cyber-space http://smartech.gatech.edu/bitstream/1853/29767/1/forte_andrea_200908_phd.pdf http://strategy.wikimedia.org/wiki/Former_Contributors_Survey_Results http://wikimania2010.wikimedia.org/wiki/Submissions/The_State_of_Wikimedia_Scholarship:_2009-2010/Suggestions http://mako.cc/copyrighteous/20100627-00 http://meta.wikimedia.org/wiki/Contribution_Taxonomy_Project http://www.cs.washington.edu/homes/travis/reflect/ http://www.zotero.org/groups/wikipedia_research/items http://meta.wikimedia.org/wiki/Mind_the_Gap SuggestBot User:SuggestBot SuggestBot: using intelligent task routing to help people find work in wikipedia wikiFeed http://wikistudy.mathcs.carleton.edu

97

\mathbf{Index}

Æblerød, 17	Freebase, 41
addiction, 30	gamification, 41
altmetrics, 36, 37	Gardner, Sue, 15
altruism, 29	gender, 10, 12, 16, 20, 22, 23
	Gene Wiki, 25, 40
American Customer Satisfaction Index, 26	General Inquirer, 20
article trajectory, 17	genre, 14
associativity, 17	geotagging, 24
augmented reality, 37	GermaNet, 9
authorship, 21	Gini coefficient, 21
1 10 00 00 70	Gompertz model, 31
barnstar, 18, 29, 30, 53	Google, 34
bazaar, 24	9 ,
benevolent dictator, 24	Google Knol, 32
Bicholim conflict, 17	Google Scholar, 2, 25
biclique, 16	Google Trends, 25
bot, 15, 39, 40	Googlepedia, 44
Brede Wiki, 43, 44, 49	health, 36
Broadwell, Paula, 35	Heilman, James, 49
	HistoryViz, 44
certification, 39	,
CheiRank, 37	histropedia, 44
citation, 3, 47, 48	hoax, 9, 17
ClueBot, 19, 38, 39	HTML5, 43
ClueBot NG, 19, 38	hypertext, 32
clusterball, 16	image retrieval, 33
Cochrane Collaboration, 47	index generation, 41
collaborative writing application, 6, 7	infobox, 16, 23, 35, 40, 43
comScore, 25	IntelliGenWiki, 41
constitutional monarch, 24	Internet penetration, 52
court case, 14, 15	IRC, 4
coverage, 10–13	1100, 1
CoWeb, 50	Javascript, 34
critical mass, 24	• '
cross-cultural study, 16	Kaggle, 53
CrossRef, 13	Kane, Gerald C., 39
DBpedia, 1, 37	11 1 25
distributed wiki, 45	labor hours, 27
donation, 28	LaTeX, 52
donation, 28	Layar, 37
economy, 27, 28	Legolas2186, 17
edit-a-thon, 27	license, 21
egoboo, 17, 30	Linked Data, 1
election, 36	Linked Open Data, 42
Encarta, 7, 8, 32, 55	Linus' Law, 32
Encyclopædia Britannica, 7, 8, 11–14, 48, 55	Linux, 4, 12, 28
Everything 2, 14	List to Wikipedia, 35
Livery timing 2, 14	LOINC, 10
Facebook, 3, 20, 26	machine learning, 5, 19, 33, 38, 39
findvej.dk, 37	Macropædia, 13
Finkelstein, Seth, 24, 29, 53	Manske, Magnus, 12, 40–42, 56
Firefox, 44	markup, 43
flagged revision, 32	MD5, 39
folksonomy, 34	MediaWiki, 37, 40
formality, 14	Metaphacts, 40
free and open source software, 23, 24	Meyer, Bertrand, 39
free culture, 6	Microsoft, 19
*	,

Mix'n'match, 12	semantic wiki, 5 , 41 , 42
Monkey selfie, 15	sentiment analysis, 17, 20
motivation, 28–30	Slashdot, 26, 32
	SNPedia, 49
named entity recognition, 33	social movement, 6
natural language processing, 41	sockpuppet, 20, 53
network, 17	Spanish Fork, 27, 31
network analysis, 15–17	SPARQL, 42
network visualization, 16	Special:Nearby, 41
,	stigmergy, 25
oh shit graph, 31	STiki, 19, 38, 39
OmegaWiki, 49	Summers, Ed, 4, 13
One Laptop Per Child, 45, 55	Swartz, Aaron, 21, 32
online community, 6	, , ,
Open Access Reader, 54	Toolserver, 3–5, 18, 23, 35
OpenThesaurus, 9	transaction cost, 32
OpenWetWare, 49	translation, 34
opportunity cost, 28	trust, 38
	Twitter, 6, 13, 36
page view, 36	1 11111111, 0, 10, 00
page views, 3	up-to-dateness, 7, 8, 13
PageRank, 15, 16, 37	Usenet, 31
Palin, Sarah, 35	0301100, 01
Parsoid, 43	vandalism, 9, 17–19, 37–40, 51, 56
PDBWiki, 49, 54	Veropedia, 39
persistent word view, 21, 22	VisualEditor, 28, 43
personality right, 49	Vibratization, 20, 10
PISA, 55	Wales, Jimmy, 12, 21, 24, 47, 54
policy, 53	West, Andrew, 35
popularity, 25, 27	WhoKnows, 37
power law, 15	Wiki ShootMe, 41
prediction, 35, 36	Wiki Trip, 24
Prefuse, 16	Wiki-PR, 28
Primary Sources Tool, 41	WikiBlame, 21
productivity, 53	Wikibooks, 47, 52
PyPedia, 46	Wikibrowser, 45
	Wikibu, 39
Python, 46	WikiChanges, 35
qLabel, 34, 56	WikiChecker, 23, 35
- · · · · · · · · · · · · · · · · · · ·	WikiCite, 3
quality, 6, 7	·
query expansion, 33	WikiCreole, 43
question-answering, 42	WikiDashboard, 21
Damagar Davids 20	Wikidata, 3, 6, 12, 13, 34, 35, 37, 38, 40–43, 55, 56
Ramsey, Derek, 39	Wikidata Query editor, 42
Raymond, Eric, 24, 30, 32	Wikidata Query Service, 42
readability, 7–9	Wikidata Toolkit, 4
Reasonator, 40, 41, 44	Wikiganda, 20
recommender system, 16	WikiGenes, 49
reification, 42	wikiGraph, 45
replacement cost, 27	wikilink, 15, 16
reputation, 29	Wikimedia Commons, 34
review, 2, 33	Wikimedia Tool Labs, 3, 4, 35
revision history, 38	WikiOpener, 40
Rfam database, 18, 40, 49	WikiPathways, 49
RNA Biology, 48	Wikipedia, 19
0.1.1. 1: 00.40	Wikipedia Beautifier, 44
Scholarpedia, 32, 49	Wikipedia Live Monitor, 35
Semantic MediaWiki, 42, 43, 49	Wikipedia Natural Disaster Monitor, 35
semantic relatedness, 33	Wikipedia Recent Changes Map, 24
semantic stability, 40	WikipediaVision, 24
Semantic Web, 42	WikiPic, 34

```
WikiProtein, 54
WikiProteins, 49
Wikirage, 35
Wikiscanner, 19, 36
Wikisource, 43, 44, 47
WikiStory, 44
Wikistream, 4
Wikitravel, 31
Wikitrends, 35
WikiTrivia, 37
WikiTrust,\, \textcolor{red}{20},\, \textcolor{red}{38},\, \textcolor{red}{39}
Wikitude, 37
wikitweets, 13
Wikiversity, 51, 52, 54
WikiViz,\, {\color{red} 16}
WikiWord, 34
Wiktionary, 9, 33
WiViVi,\, {\color{red} {\bf 24}}
\mathrm{WordNet},\, \textcolor{red}{\mathbf{33}},\, \textcolor{red}{\mathbf{35}}
{\rm WYSIWYG},\, {\color{red} 43}
XOWA, 45
XWiki, 45
\rm YAGO,\, \color{red} 33,\, \color{red} 35
{\rm Zachte},\ {\rm Erik},\ {\color{red}5},\ {\color{red}24},\ {\color{red}31}
```