

Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource

Andrew Lih

*Journalism and Media Studies Centre
University of Hong Kong
alih@hku.hk*

Paper for the
5th International Symposium on Online Journalism (April 16-17, 2004)
University of Texas at Austin

Abstract

Wikipedia is an Internet-based, user contributed encyclopedia that is collaboratively edited, and utilizes the wiki concept – the idea that any user on the Internet can change any page within the Web site, even anonymously. Paradoxically, this seemingly chaotic process has created a highly regarded reference on the Internet.

Wikipedia has emerged as the largest example of participatory journalism to date – facilitating many-to-many communication among users editing articles, all working towards maintaining a neutral point of view — Wikipedia’s mantra. This study examines the growth of Wikipedia and analyzes the crucial technologies and community policies that have enabled the project to prosper. It also analyzes Wikipedia’s articles that have been cited in the news media, and establishes a set of metrics based on established encyclopedia taxonomies and analyzes the trends in Wikipedia being used as a source.

Abstract	1
Introduction	2
Overview	2
Wikipedia as Open Content	3
Keeping it social and neutral	4
The Knowledge Gap	5
Wikipedia Growth	5
Research	6
Methodology	7
Press Citations	9
Tools	9
Comparison	9
Results	10
Publications Citing Wikipedia	12
Articles Cited	13
Initial Results	14
Visual Comparison and Analysis	15
Top changers	16
Articles Passing the Threshold	17
Conclusions	19
Future research	19
About the Author	21
Acknowledgements	21
Appendix 1: Benchmark encyclopedia topics	22
Appendix 2: References to Wikipedia articles in the press - Jan 2003 to Mar 2004	24

Introduction

Over the last decade, the Web has become an essential tool for journalists and researchers. While information can be found quickly and easily using search tools such as Google or Yahoo, the problem is often not a lack of content, but rather the large volumes of stale and questionable information. Determining the accuracy of search results is a challenge for any Internet user. However, the advent of *participatory journalism* has provided a unique solution – it engages the news audience to participate in the process of rationalizing Web content, crafting the news and contributing knowledge in the *media ecology*.

In their work on the subject, *We Media, How audiences are shaping the future of news and information*, Shayne Bowman and Chris Willis define participatory journalism as¹:

The act of a citizen, or group of citizens, playing an active role in the process of collecting, reporting, analyzing and disseminating news and information. The intent of this participation is to provide independent, reliable, accurate, wide-ranging and relevant information that a democracy requires.

Web logs and bulletin boards are examples of this, by calling on audience feedback and contributions to help put stories in context. However, quickly emerging are *wiki* Web sites, where any user can immediately edit any page with one click of the mouse. It is wiki technology that has produced the largest form of participatory journalism to date – Wikipedia.

Overview

This paper describes the history and growth of the Wikipedia project (<http://www.wikipedia.org>), including milestones that have affected its growth and the dynamics of the community. It introduces a method for evaluating Wikipedia's content, as its article quality has not been systematically analyzed in academic research. The Historyflow project is one of the few that has attempted visualize the dynamic nature of Wikipedia's articles².

This study will help establish a set of metrics to evaluate the quality of articles among many factors – number of authors, number of edits, internal linking, external linking, article size and other metadata from Wikipedia. Subject taxonomy from the print version of the DK e.encyclopedia is used as a basic, representative hierarchy of encyclopedic knowledge. With that as a basis, statistics about Wikipedia's articles contained within this set will be used to show strong and weak subjects through two-dimensional visualization. The paper will then explore:

- Growth in the number of articles and user community, over its three year span
- Average article size and average number of edits per article over time
- Nature of the articles cited by the press during January 2003 – March 2004
- Analysis of how their metrics compare to benchmark subject statistics
- Breakdown of which publications have cited Wikipedia

Wikipedia as Open Content

Wikipedia is an Internet-based, volunteer-contributed encyclopedia that has become a popular online reference in just three years of existence (Figure 1). It has thousands of international contributors and is the largest current example of an *open content* wiki³. (The Hawaiian word for “quick,” WikiWiki, is the basis for the wiki name.) The goal of Wikipedia is to create an encyclopedia that can be shared and copied freely while encouraging people to easily change and improve the content. Each and every article has an “Edit this page” button, allowing anyone, even anonymous passersby, to add or delete any content on any page. What would surely seem to create chaos has actually produced increasingly respected content which has been evaluated and revised by the thousands of visitors to the site over time.

The project was started by Jimmy Wales, head of Internet startup Bomis.com, after his original concept for a strictly controlled, PhD-edited free encyclopedia ran out of money and resources, resulting in only a few hundred articles. At the time, the project was called Nupedia⁴ and in March 2000 had one full-time employee, Larry Sanger, who was editor-in-chief and co-coordinator for the project. Volunteers were solicited on the Internet, but there was a fairly complex working structure, including formal positions for writers, editors, peer reviewers, copyeditors and translators. Nupedia was built on the traditional structure of peer-reviewed academic publications, and according to their policy, "We wish editors to be true experts in their fields and (with few exceptions) possess Ph.D.'s."⁵ After the project failed to take off, Sanger ceased being a paid staff member and the project came towards a close.

Wikipedia Growth (English language) 2001-2004

Source: http://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia

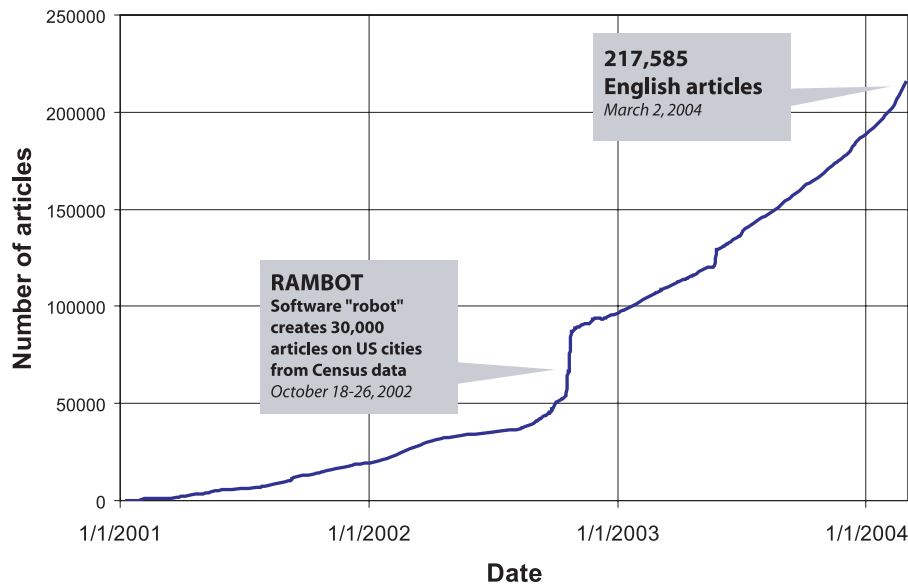


Figure 1 Wikipedia growth over its three year history

Not wanting the already-created content to stagnate, Wales and Sanger put the content onto their Web site in the form of a wiki in January 2001, and invited visitors to edit or add to the collection. It became a popular site shortly after, gaining mentions on the influential Slashdot technical community. In the first year it found a loyal following, and generated over 20,000 English language articles⁶ and spawned over a dozen language translations, despite the fact that it did not fully support internationalization. After two years, it reached a milestone of 100,000 English articles and in February 2004 at the three year mark, it exceeded 200,000 articles in English and 500,000 articles in 50 languages. In February 2004, it was adding articles at a rate of roughly 2,000 articles a day across all the various languages.

Keeping it social and neutral

The wiki concept is somewhat counterintuitive because the technical implementation itself provides no gate keeping function to ensure quality material is being contributed. Unlike typical creative efforts, no proof of identity or qualifications is needed to participate and a reputation tracking system is not used within the community. Given the description of how a wiki works, visitors to Wikipedia are often surprised the site works at all.

What allows this completely open editing system to function? Because wikis provide the ability to track the status of articles, review individual changes, and discuss issues, they function as *social software*, acting to foster communication and collaboration with other users. A wiki also tracks and stores every version of an article edited, so no operation is ever permanently destructive. As a foil to malicious contributors, it takes much more effort to vandalize a page than to revert an article back to an acceptable version⁷. While it may take five or ten seconds to deface one article, it can be quickly undone by others with just one click of a button. This crucial asymmetry tips the balance in favor of productive and cooperative members of the wiki community, allowing quality content to emerge.

However, while the technical means for managing the information and user interaction provides the mechanisms for the Wikipedia community to operate, they are not enough on their own. Founder Wales created an editorial policy of maintaining a neutral point of view (NPOV) as the guiding principle. “NPOV is an absolute non-negotiable requirement of everything that we do,” he says⁸, and according to Wikipedia’s guidelines⁹, “The neutral point of view attempts to present ideas and facts in such a fashion that both supporters and opponents can agree.” Guided by this policy, the grassroots project has implicitly adopted the same types of operational policies facing modern news operations – sticking to the facts, attributing sources and maintaining balance.

Some of the decisions are strikingly similar to those of other professional news organizations. For example, the Wikipedia community’s tendency to avoid the use of the word “terrorist¹⁰” is similar to a policy adopted by the Reuters news agency^{11,12}. Whether to reveal the name of the woman involved in the rape trial of basketball player Kobe Bryant has been a subject of intense debate within the community¹³. (The mainstream media in the United States has so far kept her identity private¹⁴.)

So far, the effort has created numerous reference-quality articles as wide ranging as the Hutton Inquiry, algorithms, social history of the piano, origins of the American Civil War and severe acute respiratory syndrome. As its quality has improved, many news publications have cited Wikipedia on subjects such as crony capitalism, folk metal, British “honours” system, Abdul Qadeer Khan and extinct animals. It has even been used in United States litigation, when in July 2003, a Wikipedia article on profanity was cited in a motion to dismiss a case in a Colorado court¹⁵.

The Knowledge Gap

Participatory journalism presents a major change in the media ecology because it uniquely addresses an historic “knowledge gap” – the general lack of content sources for the period between when the news is published and the history books are written (Table 1).

Traditional encyclopedias have typically served in this role, but their yearly publishing cycles and prohibitively high cost make them ill-suited for the task. Even conventional online encyclopedias, such as Britannica.com, work on six month to one year cycles for the creation of their articles¹⁶. Web logs and wikis have changed the landscape dramatically, as they provide not only timely distillation of information and generation of commentary, but do so in a hyperlinked context that allows users to explore, assess the merits of and contribute to the case before them. This is a fundamental shift in the relationship between reader and publisher, as communities have formed that have no strict boundaries between these roles.

Type	Sources	Type	Time Scope
Wisdom	Books, academic journals	Research and analysis	Years, decades, centuries, ad inf.
Knowledge	Magazines, encyclopedias	Secondary source	Weeks, months, years
Information	Newspaper, magazines, TV news, news wire	Primary source	Minutes, hours, days, weeks
Data	Stock quotes, sports scores, election results, economic statistics, interviews, press conferences	Live feed	Instantaneous, seconds

Table 1 Traditional journalistic sources and the Information Pyramid

Wikipedia Growth

Wikipedia has grown quickly over the last year because it uniquely fills this knowledge gap, as a constantly updated secondary source. For the month of March 2004, for example, the English Wikipedia had 2,743 active members and 521 “very active” members (those who contribute at least 100 edits per month).

Individual Wikipedias in other languages have also seen growth spurts from the bulk creation of articles. This is the result of software robots (or “bots”) which can add a large set of articles using databases of information. In the English language Wikipedia, 30,000 some articles were created by this method in October 2002, based on demographic information on all the individual towns in the United States. This meant that at the time, more than one in four articles in the English Wikipedia was about a U.S. town, which became a point of debate within the community on the merits of bulk addition of articles.

Another significant set of articles is the addition of “yearbook” articles, one for each year from the start of the common era (1 CE) to the present year and beyond. The purpose of these numbered articles is to have an index for significant anniversaries and point to events, births and deaths within the Wikipedia by year. Obviously, this straightforwardly adds 2,000 some articles to the database count. Depending on when a specific language Wikipedia decides to adds the “year” articles, this may provide a boost to the article count.

However for over a year, the growth numbers have stabilized to indicate clear trends in the articles, as shown in Figure 2.

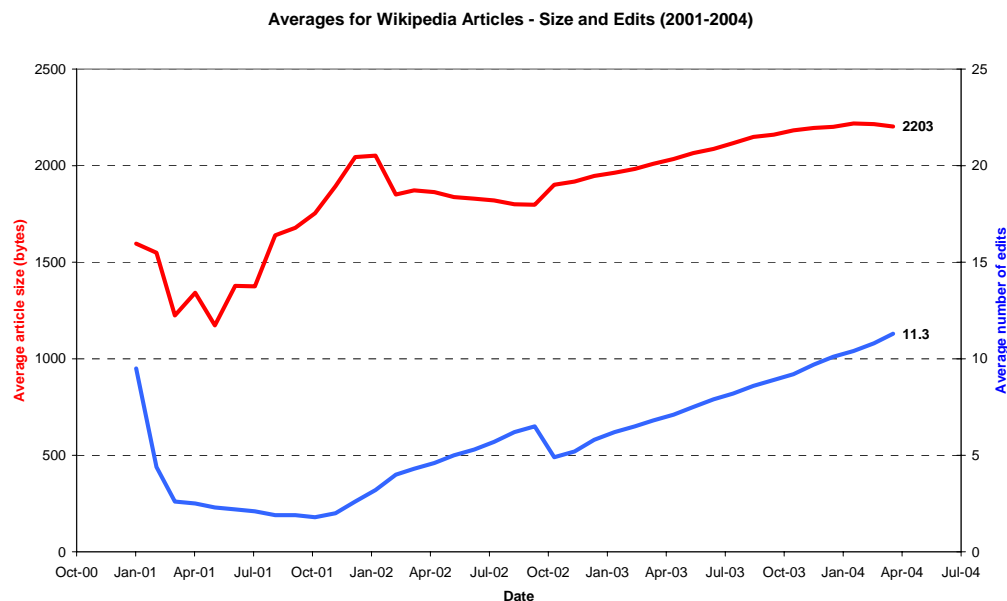


Figure 2 Average size of English Wikipedia articles and number of edits per article

The average number of edits per article for English Wikipedia has continued to climb steadily after the bulk addition of articles in 2002, and in March 2004, was roughly 11.3 edits per article. The average number of bytes per article has stabilized to just over 2 kilobytes, and is not likely to move far from this level in the near-term because of the sheer number of articles (over 200,000) for which this is an average.

Research

Wikis are just starting to receive recognition for generating credible collaborative content. Perhaps the toughest part of Wikipedia’s future is how to manage its own success. While Wikipedia has recorded impressive accomplishments in three years,

its articles have a mixed degree of quality because they are, by design, always in flux, and always editable. Even organizers of the Wikipedia project have commented on the uneven quality of the articles, and skeptics are even more wary of its content.

However, first time visitors are typically impressed with what the community has developed, considering the decentralized nature of the effort and the usefulness of its content. As more people use Wikipedia as a secondary source and as an adjunct to news content, they will desire more rigorous evaluation before putting trust in Wikipedia's articles.

Methodology

This study analyzes several quantitative metrics concerning the growth of Wikipedia articles, and its use as a source for news stories. The foundation of this study will be the analysis of Wikipedia articles which were cited in the English-language press from January 2003 to March 2004, including:

- Establishment of a benchmark using representative encyclopedia topics
- Breakdown of publications citing Wikipedia articles
- Analysis of articles cited
- Comparison of metrics against benchmark subjects

The first part of the study establishes a basic set of quantitative metrics to help compare the relative quality of the articles in the form of "reputation." While the content of the articles is important, the methodology instead focuses on the metadata concerning each article, namely the *edit history* information kept for every page in Wikipedia. The edit history is a log of each and every change made to the article in its entire lifetime, including time, date, username information and whether the change was marked as "minor" edit or a normal "major" edit. An example of the edit history for the article on "News" is shown in Figure 3.

News

Revision history

View (previous 100) (next 100) ([20](#) | [50](#) | [100](#) | [250](#) | [500](#)).

Legend: (cur) = difference with current version, (last) = difference with preceding version, m = minor edit

- (cur) (last) [04:31, 4 Apr 2004](#) [12.73.223.107](#)
- (cur) (last) [23:11, 6 Mar 2004](#) [Washington irving](#)
- (cur) (last) [23:10, 6 Mar 2004](#) [Washington irving](#) (*the claim that most news organizations aim for neutrality is not true in many culture*)
- (cur) (last) [01:55, 6 Mar 2004](#) [Washington irving](#) (*lnk to news management*)
- (cur) (last) [12:12, 5 Mar 2004](#) [Anthony DiPierro](#) (*rm ads*)
- (cur) (last) [06:05, 5 Mar 2004](#) [Washington irving](#) (*lnk to news values (could be incorporated into this article)*)
- (cur) (last) [03:45, 16 Feb 2004](#) [Meelar](#) (*added link to "editorial"*)
- (cur) (last) [m 15:52, 2 Feb 2004](#) [Snoyez](#) (*Reverted edit of Anshumangaur, changed back to last version by Patrick*)
- (cur) (last) [15:44, 2 Feb 2004](#) [Anshumangaur](#)
- (cur) (last) [00:48, 28 Jan 2004](#) [Patrick](#) (**[[Google#Google News]]*)
- (cur) (last) [00:39, 28 Jan 2004](#) [Patrick](#) (*scoop*)
- (cur) (last) [m 14:57, 8 Dec 2003](#) [Wikibot](#) (*robot interwiki standardization*)
- (cur) (last) [17:55, 24 Nov 2003](#) [Minesweeper](#) (*rm some comments; link to "news style"*)

Figure 3 Example of edit history for the article on "News"

For the purposes of the study, two metrics are used as a simple measure for the reputation of the article within the Wikipedia:

- **Rigor** (total number of edits for an article) – The assumption is that more editing cycles on an article provides for a deeper treatment of the subject or more scrutiny of the content. In Wikipedia, edits can be marked as major or minor, with the latter used for indicating something that can largely be ignored by others and inconsequential to the overall editorial position, such as fixing a typo or reformatting the page. Since this is a voluntary flag, and the use of the minor edit flag is inconsistent, at this time the study considers all edits, major or minor, as equal. In the future, a more intelligent decision could be used with minor edits in combination with the edit comments.
- **Diversity** (total number of unique users) – With more editors, there are more voices and different points of view for a given subject. Users come in the form of registered users (ie. User:Bob) or anonymous users, who do not register but show up as Internet addresses (ie. 192.168.0.10). The study tracks the number of unique users who have edited the article in question, regardless of whether they are registered or anonymous.

To establish a benchmark metric against which to compare the press-cited articles, a list of representative and widely accepted encyclopedia topics was chosen to provide a set of subjects which would be expected to be exemplars for accuracy and completeness. This also helps assess Wikipedia’s coverage of general topics of human knowledge.

The assumption is that a well-accepted set of general interest subjects in Wikipedia should be in good standing and reputation, because they have been heavily visited and edited by many different users on the Internet. Therefore, the number of edits and number of unique editors for this set of benchmark articles provides a good indicator of a “high level of quality” within the Wikipedia project.

This benchmark list was derived from the topics used for the Dorling Kindersley e.encyclopedia print edition. It was chosen primarily for DK’s reputation in producing modern, visual, up to date encyclopedias. This particular print edition was considered ideal because it was developed in conjunction with Google to produce a Web-based portal site (<http://www.dke-encyc.com>) which provides links to related Web sites for each article. Because it has a certain similarity to the Wikipedia approach to content and Internet linking, it was considered a respectable benchmark. The DK e.encyclopedia is a concise encyclopedia – a 448 page single volume work with 333 individual subjects divided into sections on space, earth, nature, human body, science and technology, people and places, society and beliefs, arts and entertainment and history. The complete list of benchmark subjects can be found in *Appendix 1: Benchmark encyclopedia topics*.

Topics from the DK encyclopedia did not always correspond directly with Wikipedia topics. In such cases, the closest topic was chosen from Wikipedia as a representation of its coverage of that area of knowledge. For example, DK’s more elaborate section title “Age of Empire” was instead substituted with Wikipedia’s article on Imperialism. Sometimes the substitution amounted to more than one entry. For DK’s entry “Canada, Alaska, and Greenland,” the three individual entries in Wikipedia were used instead. This departure from a direct one-to-one correspondence with the DK topics is not cause for concern since the endeavor is meant to gather accurate metrics

concerning the topics of most interest to general human knowledge and not simply to mirror the exact topic titles of one particular encyclopedia.

Once the diversity and rigor metrics are calculated for each entry from the benchmark subjects, the median of this set can be taken as a benchmark indicator of what constitutes “high quality.”

Press Citations

As Wikipedia has gained popularity, it has been cited increasingly more often in the press as a secondary source for additional reading, but it has also been cited directly as a source on historical facts and figures. Therefore, methods to determine whether Wikipedia articles are considered credible or reliable merits further investigation.

As an instantaneously changeable wiki web site, Wikipedia also has the unusual feature of being malleable and affected by wider exposure. As it is cited more often, it will bring more unique visitors to the site, which invites participation. Therefore, one additional goal of this research is to analyze how wider exposure for an article affects its quality and editing activity. Wikipedia articles that have been cited in the press from January 2003 to March 2004 can be compared against the median values found for the benchmark subjects, both before the citation and afterwards.

The Wikipedia.org site has an archive of which publications have referenced Wikipedia, which is maintained just like other wiki pages – ordinary users add and edit the collection of links. From these archives, a master list can be created and the metadata for these articles can be analyzed. The complete list used for this study can be found in *Appendix 2: References to Wikipedia articles in the press - Jan 2003 to Mar 2004*.

Tools

Because of the large number of articles requiring analysis, a custom set of software tools was written specifically for this project to extract the historical data for the hundreds of articles from the Web site <http://en.wikipedia.org>.

Scripts for automatically processing a list of topics and downloading the edit history were written using a combination of open source tools:

- **Perl** – a scripting language that can execute tasks, parse text strings and handle date-oriented functions.
- **Lynx** – a text mode browser that can retrieve and store the raw HTML code for each article
- **Cygwin** – the open source GNU UNIX utilities for Microsoft Windows, used for text processing and statistical analysis
- **Microsoft Excel** – for computing statistics and graphing functions.

Comparison

The data set to be analyzed is fairly large, with more than 300 benchmark subjects and over 100 target subjects. To make sense of trends and changes visualization, two dimensional plotting can be used to show general trends. To measure the change in an article since its citation in the press, the study examines the distance vector for each subject and attempts to find patterns and insight into the evolution of articles.

Results

The first task is to derive the diversity and rigor statistics for each of the 333 benchmark subjects derived from the DK e.encyclopedia after they are mapped onto the equivalent Wikipedia subjects. Perl scripts were used to download the history data for each of the articles, and to extract the unique number of users and total number of edits. The results plotted in Figure 4 show a nonlinear distribution among the subjects.

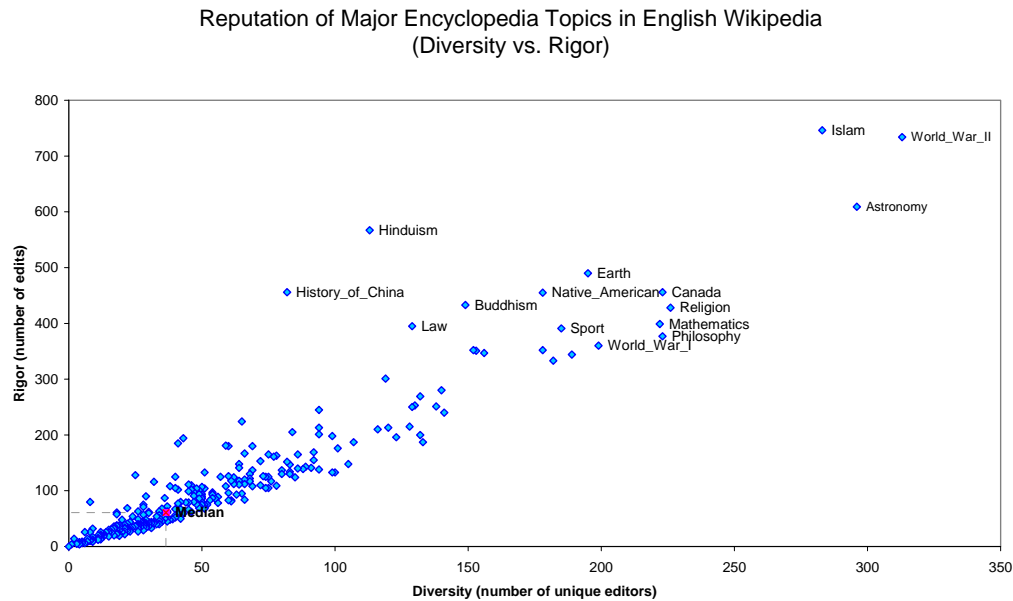


Figure 4 Benchmark subjects plotted, showing diversity vs. rigor

The most widely edited topics were World_War_II, Islam and Astronomy, residing in the far upper right hand corner of the graph. Other articles that stand out are Hinduism and History_of_China showing that while they have many edits, they were performed by relatively fewer individual unique editors than other subjects.

In the analysis of these benchmark topics, the median values for rigor and diversity were 61 and 36.5 respectively. The median is used because it is a more balanced look at the properties of the typical major article in Wikipedia, and neutralizes the large skew towards certain topics, or the mismatch between DK e.encyclopedia topics and Wikipedia's classification of topics.

Table 2 Statistics for Benchmark topics (n=333)

Metric	Max	Min	Average	Median
Edits (E)	746	4	98	61
Unique editors (UE)	313	1	51	36.5

The median indicates that a hypothetical article with 61 edits and 36.5 unique authors would be in the exact middle of the distribution of benchmark subjects. (As an example, subjects from the benchmark topics falling in this area include Cold War, continent, imperialism in Asia, mass media and samurai). For Wikipedia, these

benchmark are quite high, as the normal *average* number of edits per article in the entire English Wikipedia is fairly low, at roughly 11.3 edits per article¹⁷ (March 2004¹⁸).

If we examine the same data plotted logarithmically in Figure 5, which tries to compensate for the nonlinear distribution by scaling to powers of ten, it is clearer to see where the median value is situated.

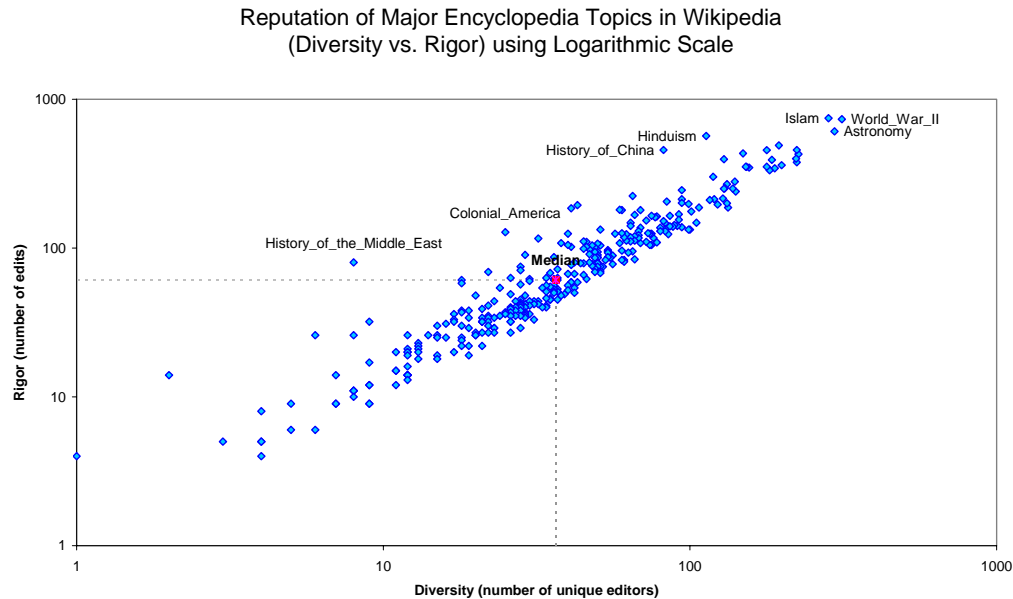


Figure 5 Benchmark subjects, diversity vs. rigor, logarithmic

The results provide some insight into the nature of the content and user community in Wikipedia. The historical topics `History_of_the_Middle_East`, `Colonial_America` and `History_of_China` are particularly interesting because their points lie significantly more towards the northwest portion of the graph, indicating that relatively few individuals are responsible for the edits. In the case of `History_of_the_Middle_East`, there were 80 edits done by only 8 unique users.

Publications Citing Wikipedia

With the benchmark numbers in hand, the study investigates how articles cited by the press compare to this benchmark. The following 72 news outlets cited, quoted or referred to Wikipedia in the English language “press” (many were Internet and television sites) from January 2003 to March 2004. Most were one-time occurrences, but there were some notable publications that made it a habit to refer to Wikipedia articles.

ADTmag.com	JapanToday.com	Star-Telegram
Aljazeera.info	Kansas City Star	Swiss Revue: official
Arizona Republic	Knoxville News Sentinel	publication for Swiss
Arutz Sheva	(Tennessee)	abroad
AskMen.com	Liberal Slant	Sydney Morning Herald
Associated Press	LinuxWorld.com	TechTV
Betterhumans	Loudon Times-Mirror	Techcentralstation.com
CBC Newsworld	Lowell Sun Online	The Age
Carlisle Sentinel	Metropolis	The Dallas Morning News
Channel Seven television	Mobile Register	The Globe and Mail
ChessBase News	New Straits Times	The National Business
Clovis News Journal	NewScientist.com	Review
Coeur d'Alene Press	NewsMax.com	The New Zealand Herald
Counterpunch.org	Norwich Bulletin	The Philadelphia Inquirer
Daily Telegraph online	OC Weekly	The Pocono Record
Embedded Systems	OS News	The South End Online
Programming	Occupational Hazards	The Straight Dope
Expatica	(magazine)	The Straits Times
Extremetech	Opinion Editorials	The Sydney Morning Herald
Greenville News	p2pnet.net	Time.com
Groklaw	Sacramento Bee	Townsend Letter for Doctors
Guardian	Salt Lake Tribune	and Patients
Hi-Desert Star	San Francisco Chronicle	Venezuelan News and
Houston Chronicle	San Jose Mercury News	Analysis
InPharm	SearchEnterpriseLinux.com	WOAI-TV San Antonio
Infoshop	Shargh	Wausau Daily Herald
International Railway Journal	Slashdot	

The most significant of these publications is the Daily Telegraph online, the Internet home of the United Kingdom-based newspaper, with 41 references to Wikipedia articles over the time period. It accounted for roughly a third of all mentions, as the publication consistently referred to Wikipedia articles in sidebars for further reading and background information.

Another publication that refers to Wikipedia on a frequent basis is Slashdot, a Web site for the technology community. (The phenomenon of being sent large amounts of visitors because of being mentioned by this popular Web site even has its own term – being Slashdotted.)

Table 3 Publications citing Wikipedia articles multiple times

Publication	Citations
Daily Telegraph online	41
Guardian	5
Sydney Morning Herald	5
Slashdot.org	4
NewScientist.com	3
ChessBase News	2

Articles Cited

During the 14 month period, 113 articles were referred to by various news outlets. In the raw data, there are actually 125 citations of Wikipedia but since some of the same subjects are referred to several times they are counted only once. Also, several publications did not specify exactly which Wikipedia article was used in their research or content, therefore ambiguous references were not counted in the final list.

These are the canonical names used in Wikipedia after any redirects or aliases are evaluated. For example, the title “System of government” in Wikipedia is an alias for one titled “Form of government.” Many news sources link to these aliases that are then redirected to the actual article title. (Either the articles were moved after the press citation, or the author may have followed a different link to the article.)

Therefore, the actual link referred to by the original article may not be exactly as listed below. The full set of data can be found in *Appendix 2: References to Wikipedia articles in the press - Jan 2003 to Mar 2004*.

11 March 2004 Madrid attacks	Graphics tablet	Mod
ALOHAnet	Great Depression	Moscow Metro
Abdul Qadeer Khan	Great Eastern Islamic Raiders' Front	Moscow Theatre Siege
Afghan National Army	Greater Tokyo Area	NLS
Ahmed Yassin	Grey goo	National Museum of Iraq
Akhmed Zakayev	Guy Green	Oggy Oggy Oggy
Al-Aqsa Martyrs' Brigades	Halabja poison gas attack	Otto von Habsburg
Alexander Graham Bell	Harry Gordon Selfridge	Palestinian Authority
Aramaic language	Heliopause	Pledge of Allegiance
Arcsecond	History of Iraq	Precautionary principle
Australian dollar	Holocaust denial	Prion
Begging the question	Howard Staunton	Provisional Irish Republican Army
Bloomsbury group	Human rights violations in Iraq	Roberto Calvi
Bluestein's FFT algorithm	Iain Duncan Smith	Russell Group of Universities
Breadboard	Immigration to Israel from Arab lands	Saddam Hussein
Breakdancing	Incompatible Timesharing System	Scintillation
British honours system	Indian National Congress	Semtex
Bubble fusion	Iraq Survey Group	Sigmund Jähn
Carpal tunnel syndrome	Jaffa (insult)	Stesichorus
Cassowary	Johnny Adair	Suez Crisis
Chen Shui-bian	Juliana of the Netherlands	Taha Yassin Ramadan
Chinese zodiac	Kings Cross fire	Thermobaric weapon
Civet	Laser cooling	Tsar Bomba
Commonwealth	Lavrenty Beria	USS Jimmy Carter (SSN-23)
Communist Party of China	Leap year	Valéry Giscard d'Estaing
Crisis of the Third Century	Lethal injection	Vendor lock-in
Crony capitalism	Lie	Vichy France
Damascus steel	Lien Chan	Vietnam War
Developing nation	List of extinct animals	Weapons of mass destruction
DrinkOrDie	List of Kings of Jordan	Who Wants to Be a Millionaire?
Earth's atmosphere	List of recipes	William Henry Cavendish-Bentinck, 3rd Duke of Portland
Establishment	Llanfairpwllgwyngyllgogerychwyrndrobwllllantysiliogogoch	Woody Guthrie
Eurofighter Typhoon	Loyalty Day	X Window System
Exciton	Lyndon LaRouche	
Execution by firing squad	Mafia	
Explosive material	Measles	
Faisal II of Iraq	Medici	
Folk metal	Microsoft Bob	
Form of government	Minix	
Fuck		
Fundamentalist Church of Jesus Christ of Latter Day Saints		

Initial Results

Because Wikipedia is an Internet-based project, the set of technology, science and computer related articles are well represented in the list of topics. However, other themes that seem to recur are those related to current events, colloquial usage (fuck, Jaffa, Oggy Oggy Oggy) and unusual terminology (folk metal, Llan...goch). It is difficult to imagine conventional encyclopedias having any substantial content for these latter subjects.

For each of the articles that were cited, two sets of statistics were created – the rigor and diversity metrics for all edits up until the date of the citation, and the same numbers for all edits up to April 6, 2004. These “before” and “after” snapshots were generated through Perl scripts that downloaded the edit history and filtered it based on the date of the article’s citation in the press.

The rigor and diversity measures for these articles vary quite a bit, which is similar to the nonlinear distribution of the reference encyclopedia topics. Figure 6 shows the “before citation” data on a two dimensional plot, where significant articles are labeled on the graph.

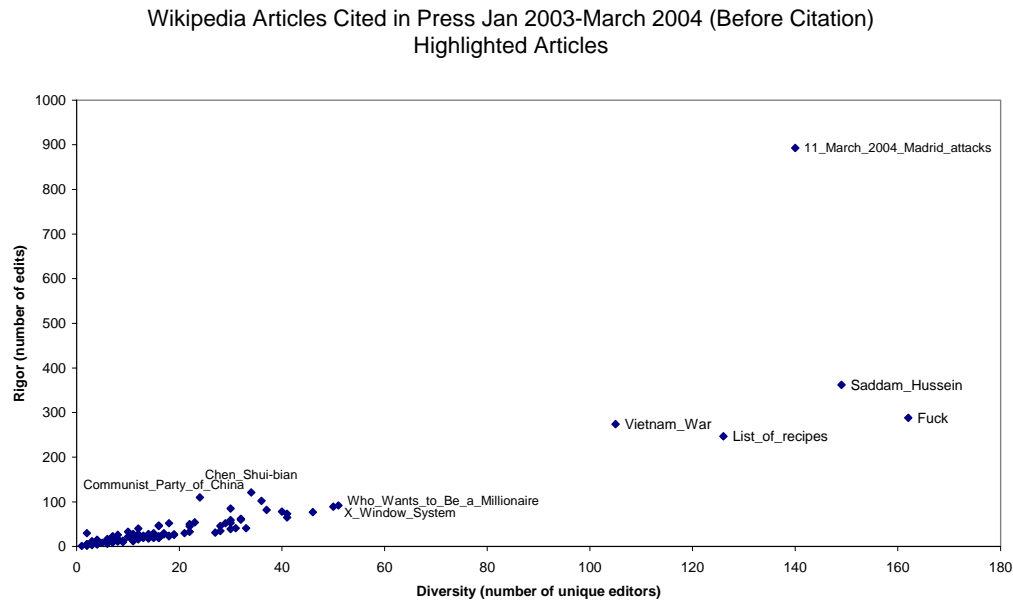


Figure 6 Wikipedia articles cited by the press, highlights

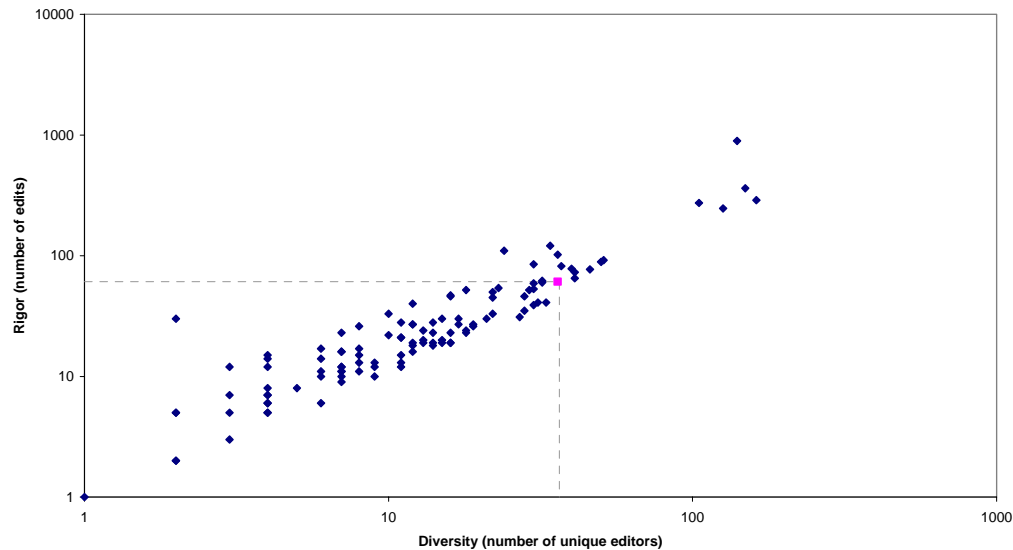
The 11_March_2004_Madrid_attacks article immediately stands out with dramatically more edits than the other articles (893) while having relatively fewer unique editors (140). There was a significant “news peg” for this single day terrorist attack on ten different locations in Spain. It was frequently updated throughout the crisis, by a set of dedicated Wikipedia users.

It is also interesting that the articles for Communist_Party_of_China and Chen_Shui-bian (recently re-elected president of Taiwan) exhibit the same characteristics as Asian-oriented subjects in the benchmark subjects – low diversity and high rigor. Again, this points to a smaller set of users making contributions to these articles.

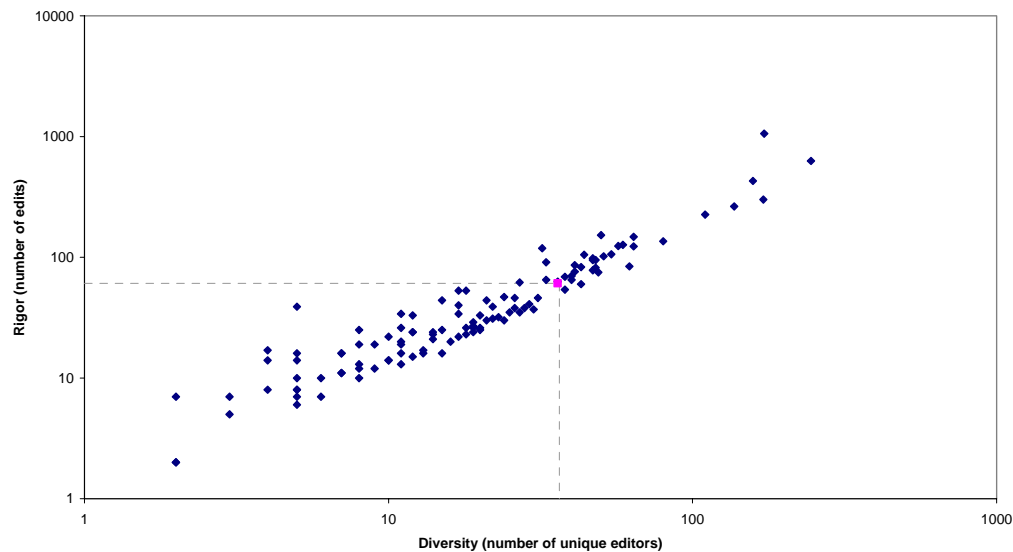
Visual Comparison and Analysis

With the metrics for both before and after press citations, a pair of two dimensional plots can be derived to provide insight on the evolution of these articles. For these graphs, a logarithmic plot is used to reveal more detail in the critical middle area. We can see that there is significant movement of articles from being below the benchmark point to being above it.

Wikipedia Articles Cited in Press Jan 2003-March 2004 (Before Citation)



Wikipedia Articles Cited in Press Jan 2003-March 2004 (After Citation)



Article quality *before* press citation. Before being cited, only 15.0% (17 of 113) of these articles were better than the benchmark standard. An article is considered better than the median if, at the time, either its rigor or diversity numbers exceed either of the median values: 36.5 editors or 61 edits.

Article quality *since* press citation. After the citation, the number of articles better than the benchmark standard doubled to 31.0% (35 of 113). Some of this is attributable to the fact that the article was newsworthy in its own right and had been updated because of other types of regular Wikipedia traffic, such as linking to Wikipedia from web logs or other sources. Further work needs to be done on tracking the activity immediately following the press mention to see if there is a more direct correlation. The best alternative is to use “referrer logs” which actually record the URL of the link that sent the traffic to Wikipedia. However, these logs are not currently available on Wikipedia.

Top changers

The articles that underwent the most change from before the citation in the press and after the citation are shown in Table 4. (A simple distance formula was used to determine the most change, with equal weighting to the number of unique editors (UE) as x , and the the number of edits (E) as y .)

$$d = \sqrt{(\Delta x)^2 + (\Delta y)^2}$$

In the future, it will be desirable to take other factors into account, such as major/minor edits, or to put greater weight the number of unique users, since an additional unique editor is a rarer occurrence than an additional edit.)

The article with the largest absolute change during the period was Saddam_Hussein. As can be expected, topics such as this, with relevancy to current events, will be high on the list as users are likely to seek out Wikipedia on their own for information about the topic. Because of this, the large amount of activity for topics not particularly timely, such as Vietnam War and Lyndon LaRouche, are interesting because they must be attributed to other phenomena – either the press citation or something else.

**Table 4. Top changing articles, since its citation in the press.
UE is unique editors, and E is number of edits**

	Article name	Before		After		Change		Distance
		UE	E	UE	E	d(UE)	d(E)	
1	Saddam_Hussein	149	362	245	628	96	266	282.79
2	Ahmed_Yassin	13	24	110	226	97	202	224.08
3	11_March_2004_Madrid_attacks	140	893	172	1060	32	167	170.04
4	Vietnam_War	105	274	158	429	53	155	163.81
5	Provisional_Irish_Republican_Army	15	20	48	95	33	75	81.94
6	Lyndon_LaRouche	17	27	47	98	30	71	77.08
7	Weapons_of_mass_destruction	22	50	51	102	29	52	59.54
8	Juliana_of_the_Netherlands	29	52	54	106	25	54	59.51
9	History_of_Iraq	16	47	47	95	31	48	57.14
10	Iain_Duncan_Smith	12	27	41	76	29	49	56.94

Of particular note are the articles for Ahmed Yassin, 11_March_2004_Madrid_attacks and Juliana_of_the_Netherlands because of their high rate of edits. Each of these articles had a flurry of editing between the time of mention in the press and the end of the study, April 6, 2004 (see Table 5). The Ahmed Yassin article was in fact the target of vandalism and subsequent intervention by a number of regular Wikipedia

users. This explains some of the activities, but the bulk of the edits were constructive ones from dozens of additional users over the course of two weeks.

Table 5. Highest rates of edits and unique users since citation in the press

Article name	Date mentioned	Days since April 6, 2004	Unique editors per day	Edits per day
Ahmed_Yassin	23-Mar-04	14	6.929	14.429
11_March_2004_Madrid_attacks	18-Mar-04	19	1.684	8.789
Juliana_of_the_Netherlands	22-Mar-04	15	1.667	3.600

As an example of the result of the high-traffic, and resulting quality of these articles, the lead paragraphs for each of these articles are listed below:

Sheikh Ahmed Ismail Yassin (*circa* [1937](#) - [March 22, 2004](#)) was the leader of [Hamas](#) until he was killed by an Israeli [helicopter gunship](#). Yassin co-founded Hamas with [Abdel Aziz al-Rantissi](#) in [1987](#), originally calling it the Palestinian Wing of the [Muslim Brotherhood](#). In addition to being nearly [blind](#), he was a [paraplegic](#) and was confined to a [wheelchair](#) after a sports accident in his youth.
(http://en.wikipedia.org/wiki/Ahmed_Yassin, Revision of 16:20, 6 Apr 2004)

The **11 March 2004 Madrid attacks** (also known as **11/3**, **3/11** and **11-M**) were a series of coordinated [terrorist](#) bombings against the [commuter train](#) system of [Madrid, Spain](#) on the morning of [11 March 2004](#). The [attacks](#) killed 191 people and wounded more than 1,800, making them the deadliest terror strikes in Europe since the [Lockerbie bombing](#) in [1988](#) and the worst terrorist assault in Spanish history.
(http://en.wikipedia.org/wiki/11_March_2004_Madrid_attacks, Revision of 05:41, 6 Apr 2004)

Juliana Louise Emma Marie Wilhelmina, Duchess of Mecklenburg-Schwerin, ([April 30, 1909](#) - [March 20, 2004](#)) of the [House of Orange-Nassau](#) was Queen of the [Netherlands](#) from her mother's abdication in [1948](#) to her own abdication in [1980](#).
(http://en.wikipedia.org/wiki/Juliana_of_the_Netherlands, Revision of 23:19, 2 Apr 2004)

The revisions on these are notable because each subject was in the news headlines during its day of citation. The first two items were related to ongoing news events with the situation changing quickly in Israel and Spain. The articles not only changed quickly, but they resulted in a large diversity of edits by many individuals. The entry for Juliana of the Netherlands was a subject of activity, but on a much smaller scale since her death itself was of relatively small global significance. The article's edits were largely to fill in historical highlights and particulars of her final years (she passed away in her sleep).

Articles Passing the Threshold

Of the 113 articles that were mentioned in the press, 18 of them moved from below the benchmark point to being above it since being cited in the press. These are listed in Table 6 and include a mix of different types of articles.

Table 6 Articles cited by the press, better than the benchmark point (April 6, 2004)

Article name	Before		After		Change	
	UE	E	UE	E	UE	E
Ahmed_Yassin	13	24	110	226	97	202
Aramaic_language	33	41	62	84	29	43
Juliana_of_the_Netherlands	29	52	54	106	25	54
Weapons_of_mass_destruction	22	50	51	102	29	52
Provisional_Irish_Republican_Army	15	20	48	95	33	75
Lyndon_LaRouche	17	27	47	98	30	71
History_of_Iraq	16	47	47	95	31	48
Australian_dollar	32	60	44	105	12	45
Form_of_government	23	54	43	83	20	29
Breakdancing	31	41	43	60	12	19
Alexander_Graham_Bell	30	39	41	86	11	47
Iain_Duncan_Smith	12	27	41	76	29	49
Execution_by_firing_squad	30	53	40	70	10	17
Palestinian_Authority	28	46	40	65	12	19
Explosive_material	30	59	38	69	8	10
Leap_year	28	35	38	54	10	19
Suez_Crisis	11	28	36	63	25	35
British_honours_system	22	45	27	62	5	17

Using the statistics, and with a basic knowledge of current affairs, much of the growth in the above topics can be linked to ongoing news events. However, some topics are likely to have benefited specifically from the mention in the press, because on their own, they do not have much of a “news peg” to contemporary news events:

- Alexander_Graham_Bell
- Australian_dollar
- Breakdancing
- Execution_by_firing_squad
- Explosive_material
- Form_of_government
- Lyndon_LaRouche
- Suez_Crisis

For example, a quick investigation of the edit history for Australian dollar (Figure 7) and Lyndon LaRouche (Figure 8) shows that indeed, a flurry of edits started from the day the following articles were written: "In the money"¹⁹ (The Sydney Morning Herald, Australia, January 24, 2004) and "LaRouche for president: The campaign that keeps on going"²⁰ (Loudon Times-Mirror, United States, January 27, 2004).

```
02:56, 28 Jan 2004 . . Paul A (sp.)
m 12:42, 26 Jan 2004 . . Gaz (Make lyrebird a link)
00:15, 26 Jan 2004 . . Arno (20c and 50c nonsense)
00:13, 26 Jan 2004 . . Arno (update)
00:13, 26 Jan 2004 . . Arno (update)
m 23:23, 25 Jan 2004 . . Silsor (format)
m 23:18, 25 Jan 2004 . . Silsor (unbold cent, link cent)
m 23:17, 25 Jan 2004 . . Silsor ([[2002]], [[2003]])
m 23:16, 25 Jan 2004 . . Silsor ([[United States dollar|US dollar]])
m 23:14, 25 Jan 2004 . . Silsor (header decapitalization)
m 23:13, 25 Jan 2004 . . Silsor (capitalization, format)
```

Figure 7 Edit history for Australian_dollar, after press citation

```

06:13, 31 Jan 2004 . . 68.164.188.118 (Fixed GM link)
m 23:38, 30 Jan 2004 . . DanKeshet (more copyedits, rm lots of extraneous
descriptions of hitler, mussoline, et al)
m 23:04, 30 Jan 2004 . . DanKeshet (oops! fix link)
23:01, 30 Jan 2004 . . DanKeshet (some cleaning up (meet me on talk:))
05:16, 28 Jan 2004 . . 67.10.50.234
08:40, 26 Jan 2004 . . J. Capone
08:32, 26 Jan 2004 . . J. Capone
07:42, 26 Jan 2004 . . 64.175.113.38
07:35, 26 Jan 2004 . . 64.175.113.38
01:41, 26 Jan 2004 . . 64.175.113.38 (By J.Capone - I am a researcher and
historian and have spend considerable time studying left and right social
movements. My additions are bracketed wtih "##". Hope this is helpfull to
Wikipedia!)

```

Figure 8 Edit history for Lyndon_LaRouche, after press citation

Suez_Crisis, Explosive_material and Form_of_government saw increased activity in the weeks following the publication date of articles but it is inconclusive whether there is a direct relationship between the press citation and the editing.

Conclusions

This study has provided a method for benchmarking the reputation of articles based solely on metadata, without requiring interpretation of article content. Though simple, this method of using information on edits and authors is immediately applicable to any of the 50 odd active language Wikipedia editions. The results indicate that there is a linkage between Wikipedia as a “working draft of history” and current news events. A great many Internet users will visit Wikipedia and contribute to it on their own volition, but the study points to clear cases where the citation in the press has driven traffic directly to articles and has improved them as a result.

Open content projects such as Wikipedia received their inspiration from the earlier open source software community that emerged from online collaboration for developing software. Linus Torvalds, leader of the Linux open source movement once said, “Given enough eyeballs, all bugs are shallow.” He was referring to software development, but it is equally relevant to Wikipedia.

This use of more “eyeballs” is a rather unique feature of participatory journalism, as it benefits directly from more traffic and more users making their impression by scrutinizing or contributing to content. This tight feedback loop between reading and editing provides for very quick evolution of encyclopedic knowledge, providing a function that has been missing in the traditional media ecology.

Future research

As Wikipedia grows more popular and widely recognized, it will be important to apply quantitative metrics to what people currently informally perceive as the overall “quality” of Wikipedia’s articles. This study is among the first to evaluate them in any systematic manner. Because of the relative lack of ground work on this subject, a simple model was chosen based solely on metadata – the edit history information.

The following areas for improvement provide opportunities for future work:

- **Content analysis of articles.** Currently, the methodology only tracks metadata since content analysis is a much more complicated task. Because of the nature of wikis, different subjects will have different organization based on

the evolution of articles within the project. For example, the `History_of_the_United_States` page is relatively short, because it links to twelve other section pages that go into much more detail. Other articles may keep all the information on one large page, such as for `Native_American` which has 11 sections and 19 subsections. Another complication is that the use of “standard elements” such as templates is not consistent across articles even in the same genre, such as history or art. However, some metrics worth pursuing include analyzing internal links (to other Wikipedia pages), external links and use of images. This information is recorded and stored, but not used in this simple model.

- **Detection of vandalism.** Vandalism, and subsequent repair of vandalism, currently does affect the analysis of the edit history. Users who deface articles will usually be “reverted” by a regular user on Wikipedia, and shows up as two additional edits – the vandal’s activity and the reversion activity. This adds two edits and possibly two unique users to the count. Repeat vandalism to the article will likely not increase the unique editors but will add to the edit count. Detecting vandalism can be done heuristically, since most Wikipedia regulars will mark vandalism or non-useful edits with a comment such as “Reverting...” or “rv” to indicate the elimination of unwanted content. Future work might be able to employ a good heuristic for detecting this and eliminating it from the data.
- **Time window for cause-effect evaluation.** Currently, the study measures the time from press mention until a specific cutoff time, April 6, 2004, which means article activity is measured for different lengths of time. Therefore articles from January 2003 have had a longer period of evolution than one mentioned in March 2004. In future work, a small time window (say 30 days) after the press mention can be used instead to assess the impact. Exactly what value to use is not immediately obvious, and requires some analysis to determine an ideal window of time. This is complicated by the different types of publication, described below.
- **Time accuracy of press citation.** The time of an article’s mention in various media sources is a challenge to model because there is different latency or “lag time” in terms of publication time and the resulting traffic to Wikipedia. For example, a citation in a Web site or television program can be pinpointed down to the exact day, if not exact hour, and the effects can be seen immediately. Newspapers can be accurate down to one day, but monthly magazines pose some problems because the activity directed from them will trickle in slowly over time. Currently, the methodology considers edits older than 24 hours before the date of the press mention as older edits, in order to ensure that the activity before any press mention-induced activity is recorded as the “before” behavior.
- **Differentiating among the nature of citations.** Different types of citations in the press are currently recorded in the data set, but not taken into account. Whether a citation is a link in the text, a quotation, or a link for further reading can make a difference in how much traffic is generated to the site.

- **Western civilization topics.** The topics from the DK e.encyclopedia are a good starting point for investigation, but are oriented more towards topics closer to Western civilization and culture. For example, there is a skewed focus on Britain, Ireland and the Americas, which is indicative of DK's roots as a UK publisher. As a result, some of Wikipedia's strongest sections are not scrutinized, such as mathematics, physics, technology, popular culture and Asian history. Selecting the representative subjects for a benchmark is not a simple task, since most encyclopedias are oriented towards a certain readership, which indirectly biases the set of categories used. So it is a challenge to find an optimal set of categories that can provide the structure for a common set of knowledge across the multiple languages, and therefore cultures, of Wikipedia.

About the Author

Andrew Lih is currently assistant professor at the Journalism and Media Studies Centre at the University of Hong Kong. He also serves as the center's director of technology. His research interests include online communities, mobile computing, virtual collaboration and participatory journalism. Email: alih@hku.hk

Acknowledgements

I wish to thank Dan Gillmor, Jerry Michalski and Sreenath Sreenivasan for their insight on online and participatory journalism; Wikipedia developers Magnus Manske, Tim Starling and Brion Vibber for their help in tracing the early technical history of Wikipedia.

Appendix 1: Benchmark encyclopedia topics

The following list of topics was used as a reference because it representative cross-section of general human knowledge. Topics below are based on those from the Dorling Kindersley e.encyclopedia, 2003 edition, with titles adapted to be compatible with entries in the English Wikipedia (<http://en.wikipedia.org/>).

Africa	Circulatory system	Fish	Lens (optics)
Agriculture	Civil society	Fishing	Lichen
Alaska	Climate	Flowering plant	Life
Alga	Cnidaria	Football World Cup	Light
Alloy	Coast	Force	Literature
American Civil War	Cold War	Forestry	Liver
American Revolution	Colonial America	Fossil	Lung
Amphibia	Color	French Revolution	Machine
Ancient Egypt	Comet	Fungus	Magnetism
Ancient Greece	Communication	Galaxy	Mammal
Animal	Computer	Game	Manufacturing
Animation	Computer network	Genetic engineering	Mars
Antarctica	Confucianism	Genetics	Mass media
Anthropology of religion	Conquistador	Geography	Materials science
Arachnid	Conservation	Great Depression	Mathematics
Architecture	Constellation	Greenland	Matter
Arctic	Construction	Gupta Empire	Maya civilization
Arthropod	Continent	Habitat	Measurement
Artist	Country	Heart	Megalith
Asia	Crusade	Heat	Mercury
Asteroid	Crustacean	Herbivore	Mesopotamia
Astronaut	Culture	Hinduism	Meteor
Astronomy	Dance	History	Mexico
Atmosphere	Decolonization	History of Africa	Microorganism
Atom	Design	History of Britain	Middle Ages
Australasia	Digestion	History of China	Middle East
Australia	Dinosaur	History of Ireland	Middle kingdoms of India
Aztec	Disease	History of Islam	Minoan civilization
Biology	Drawing	History of the Middle East	Mogul Empire
Biotechnology	Ear	Holy Roman Empire	Mollusk
Bird	Earth	Home cinema	Monarchy
Body	Earth science	Hormone	Mongols
Brain	Earthquake	Human rights	Moon
British Agricultural Revolution	East Africa	Ice	Motion
British Isles	East Asia	Ice age	Mountain
Bronze Age	Eastern Europe	Imperialism	Mouth
Buddhism	Echinoderm	Imperialism in Asia	Muscle
Byzantine Empire	Ecology	Inca Empire	Music
Canada	Economics	Indian Wars	Musical composition
Caribbean	Electricity	Indus Valley Civilization	Musical instrument
Carnivore	Electricity distribution	Industrial Revolution	Musical theater
Cartography	Electromagnetism	Industry	Nanotechnology
Cave	Electronics	Insect	Napoleonic Wars
Cell	Endangered species	Integrated circuit	Nation
Celt	Energy	International organization	Nationalism
Central Africa	Energy development	Internet	Native American
Central America	Engine	Interplanetary travel	Neptune
Central Asia	English Civil War	Invertebrate	Nervous system
Central Europe	Equality	Islam	New World Order
Chemical element	Erosion	Island	New Zealand
Chemical industry	Europe	Judaism	Nordic countries
Chemical reaction	Evolution	Jupiter	Normans
Chemistry	Evolution of Homo sapiens	Kidney	North Africa
Chinese Civil War	Exploration	Köppen climate classification	North America
Christianity	Extraterrestrial life	Law	Nose
Cinema	Eye		Nuclear reactor
Circuit	Family		Nutrition
	Fascism		

Observatory	Renaissance	South America	U.S. Eastern states
Ocean	Reproduction	South Asia	U.S. Western states
Oceania	Reptile	Southeast Asia	Universe
Olympic Games	Respiration	Southeast Europe	Uranus
Opera	River	Southern Africa	Venus
Orchestra	Robot	Space	Vertebrate
Ottoman Empire	Rock	Space Shuttle	Video game console
Painting	Rocket	program	Viking
Persian Empire	Roman Empire	Space observatory	Volcano
Philosophy	Russia	Space station	Völkerwanderung
Phoenicia	Russian Revolution	Sponge	War
Photography	Samurai	Sport	Weather
Physics	Satellite	Star	West Africa
Plant	Saturn	State	Western Europe
Pluto	Science	Storm	Wind
Poetry	Scientific Revolution	Sun	World War I
Politics	Sculpture	Super Bowl	World War II
Polynesia	Sense	Taoism	Worms
Popular music	Separation of mixture	Technology	Writer
Population	Shinto	Telecommunication	Writing
Pottery	Sikhism	The Enlightenment	Writing system
Printing	Skeleton	Theater	Zoroastrianism
Protest	Skin	Thirty Years' War	
Protestant	Slavery	Time	
Reformation	Society	Tooth	
Puberty	Soil	Toy	
Quantum mechanics	Solar system	Transport	
Rain	Sound	Tree	
Religion	South America	Tudor dynasty	

Appendix 2: References to Wikipedia articles in the press - Jan 2003 to Mar 2004

The following contains a list of all references to press citations of Wikipedia articles from January 2003 to March 2004. Some of the citations do not have a Wikipedia article listed next to them. This is because the article has ambiguously referred to Wikipedia as the source or as a general site for further reading. In these cases, the publication and date is noted below, but no corresponding article is provided. For a list which contains URLs, please refer to the original lists located at Wikipedia pages:

- http://en.wikipedia.org/wiki/Wikipedia:Press_coverage_2003#Wikipedia_as_a_source
- http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_as_a_press_source

Citation Title	Citation Publication	Date	Citation Type	Wikipedia article
The Straight Dope's column on leetspeak	The Straight Dope	10-Jan-03	Internet	Leet
We'll fight our own way, says UK general	Daily Telegraph online	11-Mar-03	UK	Suez Crisis
Liberal Slant's column on anti-French feelings in the US after the attack on Iraq	Liberal Slant	22-Mar-03	Internet	
Back to Babylon	The Age	29-Mar-03	Australia	History of Iraq
Woman may have founded ancient Rome	Daily Telegraph online	22-Apr-03	UK	Stesichorus
Minister in fees row snubs 'new' universities	Daily Telegraph online	22-May-03	UK	
Media blamed for exaggerating loss of antiquities	Daily Telegraph online	22-May-03	UK	National Museum of Iraq
Some words can make a war cry foul	Sydney Morning Herald	28-May-03	Australia	Weapons of mass destruction
Ford Motor Company profiles	Associated Press	1-Jun-03	United States	
Prince asks scientists to look into 'grey goo'	Daily Telegraph online	5-Jun-03	UK	Grey goo
Actress's friend faces extradition over terrorism	Daily Telegraph online	10-Jun-03	UK	Akhmed Zakayev
IRA 'provoked troops on Bloody Sunday'	Daily Telegraph online	16-Jun-03	UK	Provisional Irish Republican Army
Blair to phase out Lord Chancellor	The Globe and Mail	20-Jun-03	Canada	
Slivers of silver solve the problem of smelly socks	Daily Telegraph online	25-Jun-03	UK	Grey goo
Wharton: God doesn't bless only Americans	Salt Lake Tribune	26-Jun-03	United States	Woody Guthrie
GM slayer at the Staunton Memorial	ChessBase News	8-Jul-03	Germany	Howard Staunton

Citation Title	Citation Publication	Date	Citation Type	Wikipedia article
Vendor lock-in, part 1	SearchEnterpriseLinux.com	8-Jul-03	Internet	Vendor lock-in
Could nanobots destroy us?	Daily Telegraph online	9-Jul-03	UK	Grey goo
Royalists and communists return to the streets of Baghdad	Daily Telegraph online	15-Jul-03	UK	Faisal II of Iraq
Asteroid Armageddon Less Likely	Betterhumans	17-Jul-03	Canada	Earth's atmosphere
'Semtex' puff pastry shuts airport	Daily Telegraph online	25-Jul-03	UK	Semtex
Top universities want £5,000 fees	Daily Telegraph online	26-Jul-03	UK	Russell Group of Universities
Quantum logic gate created using excitons	Slashdot	11-Aug-03	Internet	Exciton
A sampling of early Christian anti-Semitism	Kansas City Star	16-Aug-03	United States	
Royal post for UN inspector	Daily Telegraph online	19-Aug-03	UK	Guy Green
'Saddam's Knuckles' captured in Mosul	Daily Telegraph online	20-Aug-03	UK	Taha Yassin Ramadan
Channel Seven Australia, the Gameshow 'Deal Or No Deal'	Channel Seven television	24-Aug-03	Australia	
Film brings East German spaceman out of the cold	Daily Telegraph online	28-Aug-03	UK	Sigmund Jähn
The Joys of Breadboarding	TechTV	9-Sep-03	Internet	Breadboard
Powell sees town where Saddam's gas killed 5,000	Daily Telegraph online	16-Sep-03	UK	Halabja poison gas attack
The Bloody Consequences of American and Iraqi Civil Wars	Aljazeera.info	23-Sep-03	Internet	Human rights violations in Iraq
'God's banker' case reopened	Daily Telegraph online	1-Oct-03	UK	Roberto Calvi
Murder squad revisit Roberto Calvi	Daily Telegraph online	4-Oct-03	UK	Roberto Calvi
'Mad Dog' Adair exposed as bungling liar in book	Daily Telegraph online	6-Oct-03	UK	Johnny Adair
Tory stalwarts challenge leader's policies"	Daily Telegraph online	7-Oct-03	UK	Iain Duncan Smith
The Great Displacement	Techcentralstation.com	13-Oct-03	Internet	Great Depression
Police seize Calvi scaffolding	Daily Telegraph online	15-Oct-03	UK	Roberto Calvi
Vichy mentally ill patients 'were not murdered'	Daily Telegraph online	17-Oct-03	UK	Vichy France
Israel's Raid on Syria, Stage Four in the Terror War	Counterpunch.org	18-Oct-03	Internet	Palestinian Authority
Russia 'ignoring' plight of siege poison victims"	Daily Telegraph online	23-Oct-03	UK	Moscow Theatre Siege
The mod mood	Knoxville News Sentinel (Tennessee)	23-Oct-03	United States	Mod
Books: Fight of the living dead	Sacramento Bee	30-Oct-03	United States	

Citation Title	Citation Publication	Date	Citation Type	Wikipedia article
What system prevails in Venezuela?	Venezuelan News and Analysis	1-Nov-03	Venezuela	System of government
Voyager 1 Reaches Interstellar Space	Slashdot	5-Nov-03	Internet	Heliopause
SCO Threatens to Sue Hollywood. Yeah, Right	Groklaw	7-Nov-03	Internet	Minix
Arafat 'diverted \$300m of public money to Swiss bank account'	Daily Telegraph online	9-Nov-03	UK	Al-Aqsa Martyrs' Brigades
US frigate takes Stars and Stripes back to Vietnam	Daily Telegraph online	20-Nov-03	UK	Vietnam War
Istanbul suicide bombers 'escaped capture after tip-off'	Daily Telegraph online	28-Nov-03	UK	Great Eastern Islamic Raiders' Front
Mafia lair stormed by 1,000-man hit team	Daily Telegraph online	30-Nov-03	UK	Mafia
'Explosive socks' seized by police in anti-terror raids	Daily Telegraph online	7-Dec-03	UK	Explosive material
Online resources on telephones	New Straits Times	8-Dec-03	Malaysia	Alexander Graham Bell
Computers: Cutting the Cord	Occupational Hazards (magazine)	11-Dec-03	Internet	ALOHAnet
Giscard the author joins Acad�mie that ignored Flaubert and Zola	Daily Telegraph online	12-Dec-03	UK	Val�ry Giscard d'Estaing
Can dynasty detectives unearth the Medici secrets?	Daily Telegraph online	13-Dec-03	UK	Medici
So close to achieving his evil dream but delusion and ineptitude led to his fall	Daily Telegraph online	15-Dec-03	UK	Saddam Hussein
Stalin's depraved executioner still has grip on Moscow	Daily Telegraph online	23-Dec-03	UK	Lavrenty Beria
The Eight Biggest Tech Flops Ever	Extremetech	23-Dec-03	Internet	Microsoft Bob
Bremer rejects Blair's 'secret labs' claim	Daily Telegraph online	29-Dec-03	UK	Iraq Survey Group
How it Sounds to the German Ear	Swiss Revue: official publication for Swiss abroad	Jan-04	Switzerland	
Web page potpourri	Townsend Letter for Doctors and Patients	Jan-04	Internet	Precautionary principle
Injections to replace Utah firing squads	Daily Telegraph online	5-Jan-04	UK	Execution by firing squad
Major Fraud, or just Playing The Game?	ChessBase News	6-Jan-04	Germany	Who Wants to Be a Millionaire
Palm-civets: The cute new face of SARS	The National Business Review	6-Jan-04	New Zealand	Civet
New Afghan army hit by 3,000 deserters	Daily Telegraph online	12-Jan-04	UK	Afghan National Army
Will Israel Become an Arab State?	NewsMax.com	12-Jan-04	Internet	Immigration to Israel from Arab lands

Citation Title	Citation Publication	Date	Citation Type	Wikipedia article
A DSP Algorithm for Frequency Analysis	Embedded Systems Programming	15-Jan-04	United States	Bluestein's FFT algorithm
Tune in, tune out, or starve	The South End Online	15-Jan-04	United States	Establishment
LWM Speaks with Richard Stallman	LinuxWorld.com	19-Jan-04	Internet	Incompatible Timesharing System
Happy Chinese New Year	Coeur d'Alene Press	22-Jan-04	United States	Chinese zodiac
Last King's Cross fire victim named	Daily Telegraph online	22-Jan-04	UK	Kings Cross fire
In the money	The Sydney Morning Herald	24-Jan-04	Australia	Australian dollar
LaRouche for president: The campaign that keeps on going	Loudon Times-Mirror	27-Jan-04	United States	Lyndon LaRouche
Navy removes submarine commander	Norwich Bulletin	27-Jan-04	United States	USS Jimmy Carter (SSN-23)
Jaffa insult quoted	The New Zealand Herald	27-Jan-04	New Zealand	Jaffa (insult)
Hip-hop dance history	San Jose Mercury News	28-Jan-04	United States	Breakdancing
All the Queen's men and the tales of knights	The Philadelphia Inquirer	29-Jan-04	United States	British honours system
Counting some numerically able animals	The Pocono Record	30-Jan-04	United States	Tsar Bomba
Moscow goes for homegrown TVMs	International Railway Journal	1-Feb-04	UK	Moscow metro
The Wonderful Proteinaceous Particle	Shargh	1-Feb-04	Iran	Prion
Prized literature: Annual used-book sale often yields overlooked gems	Arizona Republic	5-Feb-04	United States	William Henry Cavendish-Bentinck, 3rd Duke of Portland
Last of Bloomsbury Set dies aged 103	Daily Telegraph online	7-Feb-04	UK	Bloomsbury group
Between the grooves	The Sydney Morning Herald	7-Feb-04	Australia	Folk metal
Glamorous Gandhi relaunches 'first family' of India	Daily Telegraph online	8-Feb-04	UK	Indian National Congress
Why an atom spy is also a national hero	Guardian Unlimited	9-Feb-04	UK	Abdul Qadeer Khan
People lie more on the phone than by email	NewScientist.com	12-Feb-04	Internet	Lying
Animal attacks: 25 thoughts...	OC Weekly	13-Feb-04	United States	List of extinct animals
Gibson's dad stirs furor with anti-Jewish talk	The Dallas Morning News	20-Feb-04	United States	Holocaust denial
Habsburgs demand return of estates seized by Nazis in 1938	Daily Telegraph online	22-Feb-04	UK	Otto von Habsburg
Largest Lens Ever Discovered	Slashdot	22-Feb-04	Internet	Scintillation; Arcsecond
Corruption and the Political Process	CBC Newsworld	25-Feb-04	Canada	Crony capitalism
Roman coin confirms emperor's existence	Daily Telegraph online	25-Feb-04	UK	Crisis of the Third Century

Citation Title	Citation Publication	Date	Citation Type	Wikipedia article
Hotlist: Repetitive stress injury	Houston Chronicle	25-Feb-04	United States	Carpal tunnel syndrome
X Marks the Spot: Looking back at X11 Developments of Past Year	OS News	25-Feb-04	Internet	X Window System
What's up with Aramaic?	Star-Telegram	26-Feb-04	United States	Aramaic
How old are you now? Leap babies celebrate their special day	Lowell Sun Online	29-Feb-04	United States	Leap year
Top 10 Mobsters	AskMen.com	Mar-04	Internet	Mafia
Reference Page	Metropolis	Mar-04	United States	Harry Gordon Selfridge
An Allegiance to Truth	Opinion Editorials	1-Mar-04	United States	Pledge of Allegiance
St. David's Day	Guardian	1-Mar-04	UK	Llanfairpwllgwyngyllgogerychwyrndrob wlllantysiliogogoch
China begins to cut back on its 3.5m fleet of 'perk' cars	Daily Telegraph online	3-Mar-04	UK	Communist Party of China
New technique could give super-cool molecules	NewScientist.com	3-Mar-04	Internet	Laser cooling
Online: Father knows best	Guardian	4-Mar-04	UK	NLS
Oi, let's see wizard of 'was'	Sydney Morning Herald	5-Mar-04	Australia	Oggy Oggy Oggy
Cooking with the Internet?	Slashdot	6-Mar-04	Internet	List of recipes
Science runs into trouble with bubbles	Guardian	11-Mar-04	UK	Bubble fusion
Lawyers: Inmates died in 'agony' from injections	Greenville News	15-Mar-04	United States	Lethal injection
Leveraging Arab Fear	Arutz Sheva	16-Mar-04	Israel	King of Jordan
Budget pays attention to the poor, but...	The Straits Times	16-Mar-04	Singapore	Eurofighter Typhoon
Talk time: Todd Rundgren	Guardian	18-Mar-04	UK	Graphics tablet
Pharmalicensing.com: Articles: A Tour around Measles	InPharm	18-Mar-04	Internet	Measles
Media should stop begging the question	Mobile Register	18-Mar-04	United States	Begging the question
Spain aftershock: your say	The Sydney Morning Herald	18-Mar-04	Australia	March 11, 2004 Madrid attacks
Slogan diminishes our commonwealth	Carlisle Sentinel	19-Mar-04	United States	Commonwealth
Emus make themselves at home with Hi-Desert woman	Hi-Desert Star	19-Mar-04	United States	Cassowary
Taiwanese return home to vote: 1,500 from Bay Area expected to help decide island's future	San Francisco Chronicle	19-Mar-04	United States	Lien Chan; Chen Shui-Bian
Experts fear terrorists are seeking fuel-air bombs	NewScientist.com	21-Mar-04	Internet	Thermobaric weapons

Citation Title	Citation Publication	Date	Citation Type	Wikipedia article
Bridging the Divide might be good for the bottom line	ADTmag.com	22-Mar-04	Internet	Developing nations
Princess Juliana – an end of an era	Expatica	22-Mar-04	Internet	Princess Juliana
Fokken HAIL	p2pnet.net	22-Mar-04	Internet	Fuck
How will Sheik Ahmed Yassin's assassination affect the Middle East?	Time.com	23-Mar-04	United States	Sheikh Ahmed Yassin
Loyalty Day celebration set for return to Medford	Wausau Daily Herald	24-Mar-04	United States	Loyalty Day
On the cutting edge	Clovis News Journal	25-Mar-04	United States	Damascus steel
New Concerns Raised About Eldorado Compound	WOAI-TV San Antonio	26-Mar-04	United States	Fundamentalist Church of Jesus Christ of Latter Day Saints
US DrinkOrDie extradition fails	Infoshop	29-Mar-04	Internet	DrinkOrDie
You've got to be in Tokyo	JapanToday.com	30-Mar-04	Internet	Greater Tokyo Area

References

- ¹ Shayne Bowman and Chris Willis, *We Media, How audiences are shaping the future of news and information*, July 2003, <http://www.hypergene.net/wemedia/>
- ² Martin Wattenberg and Fernanda B. Viégas, “Historyflow: visualizing dynamic, evolving documents and the interactions of multiple collaborating authors, A preliminary report,” IBM Research, Collaborative User Experience research group.
- ³ “What is Wiki,” <http://wiki.org/wiki.cgi?WhatIsWiki>.
- ⁴ Nupedia archive, http://web.archive.org/web/*/www.nupedia.com/main.shtml
- ⁵ “Nupedia,” *Wikipedia: The Free Encyclopedia*, <http://en.wikipedia.org/wiki/Nupedia>
- ⁶ “Wikipedia:Modelling Wikipedia's growth,” *Wikipedia: The Free Encyclopedia*, http://en.wikipedia.org/wiki/Wikipedia:Modelling_Wikipedia's_growth
- ⁷ Andrea Ciffolilli, “Phantom authority, self-selective recruitment and retention of members in virtual communities: The case of Wikipedia”, *First Monday*, volume 8, number 12 (December 2003), http://firstmonday.org/issues/issue8_12/ciffolilli/index.html
- ⁸ Jimmy Wales, post to Textbook-1 mailing list, Mon Oct 20 07:50:15 UTC 2003, <http://mail.wikipedia.org/pipermail/textbook-1/2003-October/000496.html>
- ⁹ Wikipedia:Neutral_point_of_view, *Wikipedia: The Free Encyclopedia*, http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view
- ¹⁰ “Wikipedia:Words_to_avoid,” *Wikipedia: The Free Encyclopedia*, http://en.wikipedia.org/wiki/Wikipedia:Words_to_avoid
- ¹¹ “Reuters: Editorial Policy,” <http://about.reuters.com/aboutus/editorial/>
- ¹² Tom Glocer and Geert Linnebank, Letter to the editors of U.S. newspapers, October 2, 2001, <http://about.reuters.com/statement3.asp> (link no longer active)
- ¹³ “Talk:Kobe_Bryant’s_accuser,” *Wikipedia: The Free Encyclopedia*, http://en.wikipedia.org/wiki/Talk:Kobe_Bryant%27s_accuser
- ¹⁴ David Zeman, “THE KOBE BRYANT CASE: Spotlight shatters accuser's privacy,” *Detroit Free Press*, July 26, 2003.
- ¹⁵ The Smoking Gun, July 28, 2003, <http://www.thesmokinggun.com/archive/fword1.html>
- ¹⁶ Danny Bradbury, “A turn up for the books,” *The Independent (UK)*, March 24, 2004, p. 11.

-
- ¹⁷ Wikipedia Statistics:English, *Wikipedia: The Free Encyclopedia*,
<http://www.wikipedia.org/wikistats/EN/TablesWikipediaEN.htm>
- ¹⁸ Ibid
- ¹⁹ <http://www.smh.com.au/articles/2004/01/26/1074965468993.html>
- ²⁰ [http://www.timescommunity.com/site/tab1.cfm?](http://www.timescommunity.com/site/tab1.cfm?newsid=10876575&BRD=2553&PAG=461&dept_id=506040&rfi=6)
[newsid=10876575&BRD=2553&PAG=461&dept_id=506040&rfi=6](http://www.timescommunity.com/site/tab1.cfm?newsid=10876575&BRD=2553&PAG=461&dept_id=506040&rfi=6)