

# A Quantitative Comparison on Online Encyclopedias - A case Study of Wikipedia and Knol

Hsin-Yi Wu and Kuang-hua Chen

Department of Library and Information Science, National Taiwan University, Taipei, Taiwan.

Email: qoo.pikachu@gmail.com; khchen@ntu.edu.tw

## Abstract

The study attempts to use a quantitative approach to compare two kinds of online encyclopedias, that is, folk-oriented Wikipedia and expert-oriented Knol. Based on two main sampling rules, 20 topics were chosen from each encyclopedia. The comparison is based on six indices - page views per year, text words, readability, page strength, citation numbers, and citation types. All the data for calculating six indices are collected from online encyclopedias' websites, document analyses and tools of websites. Eight health related topics out of 20 topics are examined in this preliminary study and Wilcoxon's rank sum test is used to determine statistical significance. The results show that five out of six indices are of statistical significance at level of 0.05. The suggestions to Wikipedia and Knol are also presented. The initial findings may benefit the related researches in the future.

## Introduction

When Internet starts to flourish, a great deal of data and information is created by various persons from all over the world. Too much information results in an increasing necessity of well-organized information. In this circumstance, online encyclopedias are created accordingly, e.g., Wikipedia, Knol, Citizendium and Scholarpedia. Among these valuable resources, Wikipedia receives the most attention. The reason is that experts, laymen, and even the average Internet users can make their own edifications. Since Wikipedia has been regarded as one of "trustworthy" resources, any controversial content will have influences on the Internet uses. For example, an article of Wikipedia had incorrectly linked former journalist John Seigenthaler to the assassination of Robert Kennedy and John F. Kennedy. Another example is the former MTV VJ and pod casting pioneer Adam Curry was accused of omitting Wikipedia's references to Kevin Marks, another early pod casting luminary (Terdiman, 2005). Since these controversial events have happened, researchers and teachers began to question the quality, accuracy and credibility of Wikipedia. Some schools in America even ban students to cite Wikipedia.

In order to improve quality and avoid controversy, Knol, Citizendium and Scholarpedia emerged to emphasize the role of editors. For instance, Scholarpedia invites experts to write and review articles. Citizendium also welcomes experts and the public to write articles, but articles have to be reviewed by experts. Knol is a blog-type encyclopedia paying attention to the background of authors. After surveying online encyclopedias, this study attempts to use a quantitative approach to investigate present status of Wikipedia and Knol; namely, to explore differences between folk-oriented and expert-oriented encyclopedias. The main reason we choose Knol out of Citizendium and Scholarpedia as the research subject is that it has accumulated more than 100,000 articles and extensively covers various topics which are appropriately compared with Wikipedia.

According to the online encyclopedias' websites, document analyses, and tools of websites, this study develops six indices to compare 20 topics sampled from two main sampling rules. The Wilcoxon's rank sum test has been applied to analyze the significance of page views per year, text words, readability, page strength, citation numbers, and citation types. To summarize, this study aims to investigate differences between Wikipedia and Knol, and the results will deep our understanding of the folk-oriented and the expert-oriented encyclopedias.

## Research Design

The steps involved in this study are shown in Figure 1. The three crucial designs are choosing subjects of the study, setting up two main sampling rules, and developing six indices. The researcher surveys the famous online encyclopedias, Wikipedia, Knol, Citizendium, and Scholarpedia and chooses research subjects to compare different characteristics. Totally 21 different characteristics consist of the follow-up research. They are time of creation, years, creator, meaning of encyclopedias' name, language, classification, community, articles numbers, character of community, account system, information quality assurance agent manage, management, character of articles, edit mode, cooperation model, copyright, work coordination artifact, content quality evaluation and selection process, high quality articles review and removal processes, high quality articles status assignment processes, and revenue model.

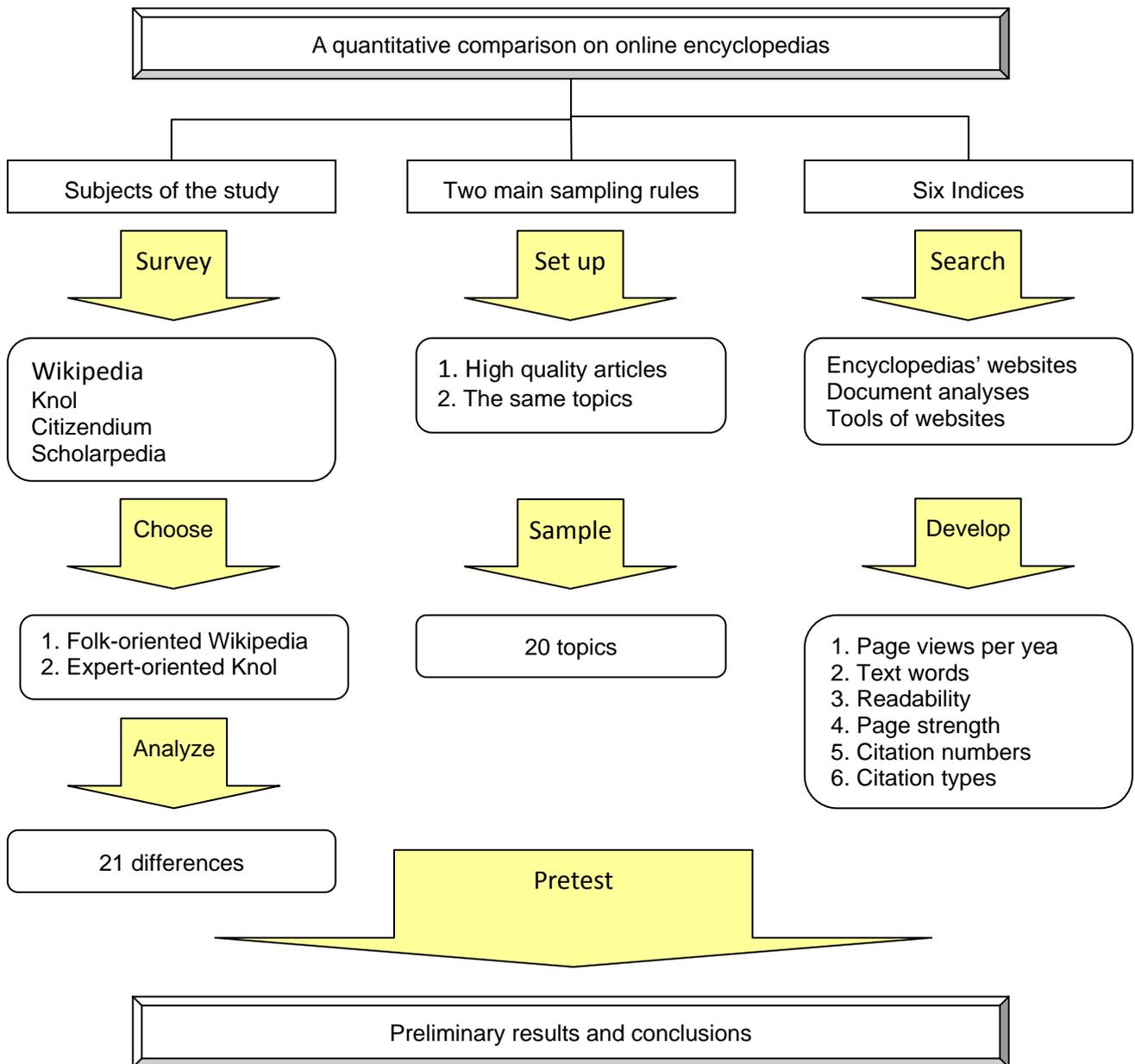


Figure 1: Research process

The study has set up two main sampling rules to select topics. Rule 1 is choosing the high quality articles in each encyclopedia. In Wikipedia, the high quality article is so-called “featured article”, which need to pass the featured article criteria, are finally decided by the community consensus. (Wikipedia, n.d.) In Knol, that is called “top picks article” selected by an algorithm, such as page views, article rating and reviews. (Knol, n.d.) Rule 2 is sampling the similar topics from those high quality articles for comparison’s purpose. Original sampling design is to sample at least 1 article based on each of 12 popular categories of Knol, which are society, business, regional content, Internet, music, shopping, health, entertainment, industries, science, reference and education. However, the researcher is unable to sample at least 1 article for each category due to few similar articles, so the sampling result is not the same as expected.

This research also develops six indices which are page views per year, text words, readability, page strength, citation numbers, and citation types as listed in Table 1. Page views per year collected from the online encyclopedias’ websites. It represents the popularity of topics and it is the total article page views divided by the age of article. The uses of text words, readability, citation numbers and citation types have been discussed in literature. (Nielsen, 2007; Rector, 2008; Song & Jeng, 2009) Text words calculate the total words except reference and further reading sections. Readability is used to examine whether textbooks are appropriate for students to read (Song & Jeng, 2009) and also is used to evaluate blogs, pages, etc. in many other researches. The readability test tool takes the text on web page and gives a score for the most used readability indicators including Gunning-Fog Score, Flesch Kincaid Reading Ease, Flesch Kincaid Grade Level, SMOG Index,

Coleman Liau Index, and Automated Readability Index as listed in Table 2. (Read-able.com, n.d.) Citation numbers calculates the total number of references, which usually appear at the end of the article. Citation types consist of three different references types, and the researcher will calculate the percentage of each type and then use JCR Web (n.d.), LibraryThing (n.d.) and Trifecta (n.d.) to collect data of all kinds of journals, books, and web pages respectively.

Table 1. The excerpts of 6 indices

Indices	Sources	Description
Page views per year	Online encyclopedias' websites	Total article page views divided by the age of article
Text words	Document analysis	Total words except reference and further reading sections
Readability	Document analysis	Gunning-Fog Score Flesch Kincaid Reading Ease Flesch Kincaid Grade Level SMOG Index Coleman Liau Index Automated Readability Index
Page strength	Tools of websites	Ranking at Google for 1st 4 words in title tag Page rank of homepage, page rank of page Google URL mentions Google blog search links to page Yahoo site explorer links to domain Yahoo site explorer links to page
Citation numbers	Document analysis	Total number of references
Citation types	Document analysis	Journals, Books, Web pages

Table 2. 6 Readability indicators

Readability	Formula
Gunning-Fog Score	Grade = $0.4 \times \{ASL + 100 \times (CW/words)\}$
Flesch Kincaid Reading Ease	0 to 100 = $206.835 - (1.015 \times ASL) - (84.6 \times ASW)$ The higher the number, the easier the text is to read.
Flesch Kincaid Grade Level	Grade = $(0.39 \times ASL) + (11.8 \times ASW) - 15.59$
SMOG Index	Grade = $3.1291 + 1.0430 \times (\sqrt{CW/sentences})$
Coleman Liau Index	Grade = $5.89 \times (characters/words) - 0.3 \times (sentences/words) - 15.8$
Automated Readability Index	Grade = $4.71 \times (characters/words) + 0.5 \times (words/sentences) - 21.43$

Note: ASL: Average sentence length (i.e., the number of words divided by the number of sentences); ASW: Average number of syllables per word (i.e., the number of syllables divided by the number of words); CW: Three or more syllables words.

Besides aforementioned indices, Trifecta can measure the popularity, influence and ranking ability of pages, blogs and entire domains in the World Wide Web through the results of ranking at Google for 1st 4 words in title tag, page rank of homepage, page rank of web page, Google URL mentions, Google blog search links to page, Yahoo site explorer links to domain, and Yahoo site explorer links to page (SEOMoz, n.d.). Each topic will have six grades, and the Wilcoxon's rank sum test is used to find whether the six indices are statistically significant at level of 0.05. The pretest and its preliminary results will be described in the next section.

### Preliminary Results

Based on the 2 main sampling rules, this study totally samples 20 topics, including Alzheimer's disease, autism, Helicobacter pylori, influenza, lung cancer, multiple sclerosis, subarachnoid hemorrhage, tuberculosis, Alfred Russel Wallace, Archaeopteryx, galaxy, global warming, Hubble Space Telescope, Australia, Canada, India, Israel, film noir, Islam and search engine optimization. According to the classification in Wikipedia and Knol, these topics are divided into 8 and 6 categories respectively. In Wikipedia, those are health and medicine, biology, physics and astronomy, geology, geophysics and meteorology, geography and places, media, religion, mysticism and mythology and computing. In Knol, those are health, science, regional content, entertainment, religion and internet. The pretest samples 8 health related topics, Alzheimer's disease, autism, Helicobacter pylori, influenza, lung cancer, multiple sclerosis, subarachnoid hemorrhage and tuberculosis.

The research can obtain the grades for page views per year, text words, readability, page strength, citation numbers and citation types using the resources discussed in previous section. SPSS is used to carry out the

Wilcoxon's rank sum test for these grades and finds that five indices are statistically significant at level of 0.05. The results for six indices and their pretest values are listed in Table 3. The asterisk denotes the values are statistically significant.

Table 3. The preliminary results

	Page views per year	Text words	Readability	Page strength	Citation numbers	Citation types(J)
Z	-2.521 <sup>a</sup>	-.980 <sup>a</sup>	-1.980 <sup>b</sup>	-2.527 <sup>a</sup>	-2.524 <sup>a</sup>	-2.032 <sup>b</sup>
Asymp. Sig. (2-tailed)	.012 <sup>*</sup>	.327	.048 <sup>*</sup>	.012 <sup>*</sup>	.012 <sup>*</sup>	.042 <sup>*</sup>
Average of Wikipedia	145264.120	7347.000	8.500	0.629	141.000	8.602
Average of Knol	10144.500	6130.625	9.875	0.243	17.000	12.372

Note: \*: denotes significance; **a**: based on positive rank; **b**: based on negative rank; **J**: all kinds of journals

There is no difference in text words but the other 5 indices show significantly different. In page views per year, page strength and citation numbers, the average values of Wikipedia are higher than Knol. In readability and citation types of journals, the average values of Knol are higher than those of Wikipedia. It is interesting to observe that each online encyclopedia possesses its own features. In other words, folk-oriented Wikipedia has the characteristics of popularity, influence, ranking ability, and also offers a large number of reference resources. Expert-oriented Knol provides more difficult articles and cites more authority resources. To improve Wikipedia, it needs to control citation numbers and cites more authority resources to increase accuracy and credibility. For Knol, it needs to do search engine optimization. Besides, according to the National Adult Literacy Survey (NALS), approximately 21 percent of the adult population in the United States have low literacy skills, defined as reading at the 6 grade level or below, while another 27 percent may have limited literacy ability, defined as lacking general reading and numeric proficiency to function adequately in society (Shalowitz & Wolf, 2004), so not only Wikipedia but also Knol need to increase its readability to fit much more general readers.

## Conclusions

This study reports preliminary results on the differences between folk-oriented and expert-oriented online encyclopedias. We used Wilcoxon's rank sum test to examine the pretest 8 health related topics. Five out of six indices show significant differences between Wikipedia and Knol. Those are page views per year, readability, page strength, citation numbers and journal in citation types. As there are differences in 5 indices, the initial findings may benefit the future related research. Furthermore, the result can help online encyclopedias improve their quality from different viewpoints. In addition, users could choose appropriate articles from different valuable resources based on our investigations. However, the study is unable to sample at least 1 article for each category of Knol and the comparisons among Wikipedia, Citizendium, and Scholarpedia are not carried out due to the same reason, either. The researchers have to conduct further studies with consideration of these shortcomings.

## References

- Nielsen, F. (2007). Scientific citations in Wikipedia. *First Monday*, 12. Retrieved May 27, 2010, from [http://131.193.153.231/www/issues/issue12\\_8/nielsen/](http://131.193.153.231/www/issues/issue12_8/nielsen/)
- Rector, L. (2008). Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review*, 36, 7-22.
- Shalowitz, D. I. and Wolf, M. S. (2004). Shared decision-making and the lower literate patient. *Journal of law, medicine & ethics*, 32, 759-64.
- Song, P.-J. and Jeng, C.-C. (2009). Research on and application of readability of approved elementary school English textbooks in Taiwan. *Journal of Textbook Research*, 2, 55-80. (in Chinese)
- Terdiman, D. (2005). *Wikipedia vs Encyclopaedia Britannica*. Retrieved August 16, 2010, from <http://www.silicon.com/technology/networks/2005/12/16/wikipedia-vs-encyclopaedia-britannica-39155109/>
- Knol (n.d.) Knol. Retrieved August 16, 2010, from <http://knol.google.com/>
- LibraryThing (n.d.) *LibraryThing*. Retrieved August 16, 2010, from <http://www.librarything.com/>
- Wikipedia (n.d.) *Wikipedia*. Retrieved August 16, 2010, from <http://www.wikipedia.org/>
- JCR Web (n.d.) *Journal Citation Report*. Retrieved August 16, 2010, from <http://isiknowledge.com/jcr>
- Read-able.com (n.d.) The Readability Test Tool. Retrieved Oct. 4, 2010, from <http://read-able.com/>.
- SEOMoz (n.d.) *Trifecta*. Retrieved Oct. 4, 2010, from <http://www.seomoz.org/trifecta>