



A Comparison of the Historical Entries in Wikipedia and Baidu Baike

Wenyi Shang^(✉) 

Department of Information Management, Peking University,
No. 5 Yiheyuan Road, Haidian District, Beijing, China
shang-wen-yi@pku.edu.cn

Abstract. This research chose two representative online encyclopedias, Wikipedia (English) and Baidu Baike, to compare their performance on historical entries. This research purposefully chose 6 entries and developed a framework to evaluate their performance in accuracy, breadth, depth, informativeness, conciseness and objectiveness. The result shows that: Wikipedia is superior in most cases while Baidu Baike is a little better in the entries on Chinese history. The operating mechanism is the main reason for it. The result implies that in the field of history, well-established online encyclopedias can be reliable for common users although improvement such as developing operating mechanisms is still needed.

Keywords: Wikipedia · Baidu Baike · Historical entries · Entry quality

1 Introduction

As a new way of creating and spreading knowledge, online encyclopedias are on the way of breaking the monopolization of the traditional academic world in knowledge dissemination. However, since they are edited by common users rather than scholars, their reliability is questioned. Among them Wikipedia receives the most attention, while Baidu Baike, a Chinese online encyclopedia is doubted by Chinese scholars.

Among the areas involved in online encyclopedias, humanity subjects, especially history, are attracting more and more attention since online resources are commonly used in popularization process of them. We thus focused on historical entries, evaluating the performance of Wikipedia and Baidu Baike. We try to develop a framework in order to evaluate their performance through multiple dimensions, therefore the research question is purposed as follows: how do Wikipedia and Baidu Baike perform in accuracy, breadth, depth, informativeness, conciseness and objectiveness?

2 Related Work

Rector (2008) did a research about the accuracy, breadth and depth of the historical article, directing at the reliability of historical entries in Wikipedia. The research provided a method of judging the quality of entries in online encyclopedias. The three factors were also emphasized by Spinellis and Louridas (2008).

More mature methods have been applied by researchers in order to judge the quality. Hu et al. (2007) have designed models based on author authority, review behavior and partial reviewership of contributors. Wöhner and Peters (2009) offered new metrics based on the lifecycle of articles, referring to the changes of contributions. Blumenstock (2008) proposed a simple metric, word count, to measure article quality.

Chinese scholars have noticed the difference of Wikipedia and Baidu Baike and tried to explain it. Liao (2014) used cultural factors to expound the difference. And in the following year, he focused on the mechanism of collaborative filtering and compared network gatekeeping of the encyclopedias (Liao 2015).

3 Methods

We used purposeful sampling method to choose 6 sample entries. We generated a list of randomly selected entries by continuously clicking the link of “random article”, which is a function of Wikipedia that provides a link to another entry randomly. From the list, we chose first 6 historical entries based on the following criteria:

- They should also be contained in Baidu Baike.
- They should be attributed to the three categories, historic place, historical event and historical figure, with each category contains two entries.
- They should be attributed to three eras, ancient (before 476 A.D.), medieval (476 A.D.–1648 A.D.) and pre-modern (1648 A.D.–1914 A.D.). Each era should contain two entries. (Modern history is ignored since it may be effected by political reasons.)
- They should be attributed to four domains of World history, European, Asian (not including Chinese history), African, and American. Each domain contains one entry.
- Two entries on Chinese history are separately chosen.

The first 6 qualified entries selected from the list include *Fushimi Castle*, *Ciudad Bolivar*, *Battle of Torgau*, *Battle of Tangdao*, *Ezana of Axum* and *Fu Jian*.

In order to compare the above 6 entries in Wikipedia (English) and Baidu Baike, we divided the articles in both encyclopedias into many information items. Each piece of words that conveys a certain fact or judgement independently is considered as an “information item”. Each item is given a binary value, where 1 means correct and 0 means incorrect, by referring to the authoritative historical works, traditional encyclopedias like Encyclopedia Britannica and news from official website. After judging all items, we got the precision rate of each encyclopedia.

Moreover, we also took breadth and depth into consideration. We picked out the information which appears in one encyclopedia but not in the other, and invited five graduates majored in history to give grades. Each information item received two grades: relevancy between the item and the article, and the depth of the item. The regulation of grading is based on a rubric of 0–5, where 0 means it is totally irrelative to the article or it is of no value at all, and 5 means it mightily matches the article or it is highly valuable in depth. After data collection, we summed up the grades to get the

final grade of breadth of each entry. The formula is as follows (n is the number of information, and $G_j r_i$ stands for the grade of relevancy of information i by graduate j):

$$\sum_{j=1}^5 \frac{\sum_{i=1}^n G_j r_i}{5} \quad (1)$$

And the grade of depth of each entry by each person is measured by the average score of every information. The formula is as follows (n is the number of information, and $G_j d_i$ stands for the grade of depth of information i by graduate j):

$$\sum_{j=1}^5 \frac{\sum_{i=1}^n G_j d_i}{5n} \quad (2)$$

Both grades are measured with Pearson correlation coefficient in order to judge the inter-coder trustworthiness. Besides, because the grades given by graduates may be biased to some extent, the results were verified by referring to authoritative works.

The difference in languages was also considered in order to compare the informativeness. We fitted a function with a scaling factor of 1.9003 based on 5 paragraphs chosen from 5 entries in Encyclopedia Britannica and its Chinese translation. Based on researches such as Yu (1989), who calculated a factor of 1.735 and Wang (2004), who calculated a factor of 1.76, we affirmed our scale is reasonable since there is no obvious difference. Thus, the length is measured after dividing the Chinese words by the scale.

The framework of evaluation is shown in Fig. 1.



Fig. 1. Framework of evaluation

4 Findings

Wikipedia is superior in accuracy, breadth, depth and informativeness over Baidu Baike (except for the entry on Chinese history), and is more concise and objective

Wikipedia has a higher precision rate in 5 out of 6 entries, and is 95.57% vs 88.03% in average. The difference between the encyclopedias in World history is clearer. Besides, the superiority in accuracy of Wikipedia appears in the usage of multimedia.

In the 4 entries on World history, Wikipedia has a higher grade in breadth. Moreover, in 4 entries out of 6, Wikipedia has a higher score of depth. The correlation is significant between most of the graders at the 0.05 level (except for grader D and E in relevancy, and grade A, C and B, E in depth), therefore, although still need further evaluation, results can be preliminarily explained as inter-coder reliable.

Table 1. Precision rate and average grades of breadth and depth (W stands for Wikipedia and B stands for Baidu Baike)

Entry	Precision rate of article		Precision rate of picture		Grade of breadth		Grade of depth	
	W	B	W	B	W	B	W	B
Fushimi Castle	95.65%	85.00%	100%	100%	41.4	23.8	3.44	3.44
Ciudad Bolivar	96.92%	94.12%	100%	N/A	79	12.2	3.93	3.13
Battle of Tangdao	92.00%	90.24%	100%	100%	23.4	40.2	3.50	3.04
Battle of Torgau	97.87%	85.71%	100%	0%	46.2	12.2	3.46	3.73
Ezana of Axum	95.45%	76.92%	100%	N/A	40	16.4	3.84	3.50
Fu Jian	95.52%	96.20%	N/A	50%	23.6	68	3.80	3.74
Average	95.57%	88.03%	100%	62.5%	42.3	28.8	3.66	3.43

Table 2. Information contained and normalized words per information items

Entry	Words contained		Number of information items		Number of pictures		Normalized words per information items	
	W	B (normalized)	W	B	W	B	W	B
Fushimi Castle	467	296	23	20	3	9	38.57	28.10
Ciudad Bolivar	1305	207	65	17	15	0	38.15	23.18
Battle of Tangdao	479	1361	25	41	1	2	36.40	63.10
Battle of Torgau	1049	631	47	28	3	1	42.40	42.86
Ezana of Axum	459	484	22	13	1	0	39.64	70.69
Fu Jian	1650	1849	67	79	0	2	67	44.48
Average	901.5	804.7	41.5	33	3.8	2.3	43.70	45.49

Wikipedia contains more words (normalized) in 3 out of 6 entries, however, since more words may result from redundancy, we calculated the number of information items in both encyclopedias to compare informativeness more accurately. Under this method, Wikipedia contains more information in all 4 entries on world history. Moreover, the number of pictures appearing in the 6 entries adds up to 23 in Wikipedia, while only 14 appear in Baidu Baike. Therefore, Wikipedia is more informative in general.

In order to compare the conciseness, we divided the number of normalized words by the number of information items in each entry, and found that the result is 45.40 in Baidu Baike and 40.32 in Wikipedia in average, implying Wikipedia is more concise.

We also judged their performance on objectiveness. In Wikipedia, there is no obvious subjective description in any of the entries, however, Baidu Baike showed non-neutrality especially in entries on Chinese history. For instance, in the entry Fu Jian, when Baidu Baike described the event of his coronation, it used the description “he arrogated the title of ‘heavenly prince’”, showing a strong emotional tendency.

Baidu Baike performs a little better in the entries on Chinese history

Above data shows that although Wikipedia is better under most circumstances, Baidu Baike is somewhat better in entries on Chinese history. Table 1 shows that the precision rate is relatively close in the entries on Chinese history, and that Baidu Baike performs better in breadth on those entries. Table 2 suggests that it contains more information.

5 Discussion

Referring to previous researches and studying the mechanisms of the encyclopedias, we try to briefly explain the findings in chapter 4.

The difference of accuracy may result from the fact that unlike Wikipedia, Baidu Baike lacks a mature consultation system to alleviate the problem caused by the editors' casualness. Because participation behavior of editors is mainly driven by interest, they lack enough motivation for precise verification. However, Wikipedia provides a way for consulting, the talk page. It is an auxiliary tool for the editors to discuss the content and correct the error, according to Wang (2004), which results in its high accuracy.

The difference of breadth results from the better organizing system and taxonomy of Wikipedia. Jia and Li (2013) found that Wikipedia has various taxonomies while Baidu Baike has only one, therefore entries in Wikipedia interact with each other more intimately. So more related entries will be considered when the editor is editing an entry. The result of depth is highly influenced by the talk page of Wikipedia, which provides editors a means to share information, making the editors with better academic attainment to participate more easily. Besides, the weakness in the academic field of World History in China also results in relative less valuable source for the editors.

The interest motivation influenced the result on informativeness. Editors tend not to include information beyond their interest, while Wikipedia editors can add information more easily with the help of talk page. The difference in conciseness can be attributed to the consultation system of Wikipedia, and that in objectiveness is a result of the neutrality principle. The first basic principle of Wikipedia is "articles should present an unbiased or neutral description of the entry," in contrast, there is no such restraint for editors of Baidu Baike. Personal emotional tendency has an impact on the words used in the entries. For instance, traditional view on legitimacy resulted in biased attitude.

Finally, resulted from a lack of the ancient books, study in the field of Chinese history in western world is relatively insufficient. In Baidu Baike, entry *Fu Jian* is completely based on *Jin Shu*, however, citations of ancient books in entries on Chinese history in Wikipedia are relatively insufficient. The lack of usage of ancient books resulted from language gap seriously influences Wikipedia's performance on Chinese history.

6 Conclusion

The result shows that Wikipedia is superior in general, while Baidu Baike is a little better in entries on Chinese history. Unlike previous studies that paid attention mainly on one encyclopedia, the research focused on the difference between historical entries of Wikipedia and Baidu Baike, which has never been studied previously.

Improvement is needed since the research is still preliminary. As a small-scale study, the size of sample needs increment, and the inter-coder trustworthiness need further analysis. Besides, more factors can be included in, such as relationship between entries and qualities of links provided. Future study may extend the research field to other subjects. Although difference between subjects results in different methods for collecting and analyzing data, the framework for evaluation is repeatable. It can be adopted in other studies on online encyclopedias, as well as specific methods such as dividing the article into information items and normalizing the words in different languages.

Based on findings of the study, we come to the conclusion that well-established online encyclopedias like Wikipedia can be reliable for common users in the field of history, proper evaluation on the online encyclopedias should not be totally negative. However, limitations still exist. Improving the reliability of online encyclopedias is not only of necessity, but is of great importance since they deeply influence the process of knowledge dissemination as well.

Online encyclopedias may improve their qualities by developing their operating mechanism. Building a better-organized knowledge community can be a good idea, which means constructing a communication platform for professionals to share ideas.

Moreover, since the development of Digital Humanities acts as an avoidable tendency, online encyclopedias should pay more attention to the quality of the historical entries and other entries on humanity subjects, trying to involve themselves in the tendency. In this way, they can fully exploit their advantages in the digitalization era and play a more important role in dissemination and development of humanity knowledge.

References

- Blumenstock, J.E.: Size matters: word count as a measure of quality on Wikipedia. In: International Conference on World Wide Web, Peking, China (2008)
- Hu, M., Lim, E.P., Sun, A., Lauw, H.W., Vuong, B.Q.: Conference on Information and Knowledge Management, Lisbon, Portugal (2007)
- Jia, J., Li, Y.: Comparative analysis of the category organization systems of the Chinese Wikipedia and Baidu Baike. *Inf. Stud.: Theory Appl.* (6), 114–118 (2013). (Chinese)
- Liao, H.: The cultural politics of user-generated encyclopedias: comparing Chinese Wikipedia and Baidu Baike. Unpublished doctor's thesis. University of Oxford, Oxford, UK (2014)
- Liao, H.: Harnessing the power of collaborative filtering. Comparing the network gatekeeping of Baidu Baike and Chinese Wikipedia. *China Perspectives*, (4) (2015). <http://chinaperspectives.revues.org/6854>
- Rector, L.H.: Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Ref. Serv. Rev.* **36**(1), 7–22 (2008)
- Spinellis, D., Louridas, P.: The collaborative organization of knowledge. *Commun. ACM* **51**(8), 68–73 (2008)
- Wang, F.: Enlightenments of synergic edition system on the networking of traditional encyclopedias. *Sci. Technol. Publ.* (11), 76–80 (2014). (Chinese)

- Wang, K.: *The Development and Application of Bilingual Corpus*. Foreign Language Teaching and Research Press, Beijing (2004). (Chinese)
- Wöhner, T., Peters, R.: Assessing the quality of Wikipedia articles with lifecycle based metrics. In: *International Symposium on Wikis*, Orlando, Florida, USA (2009)
- Yu, Y.: *The Selected Readings and Translation of English and American Masterpieces*. Xi'an Jiaotong University Press, Xi'an (1989). (Chinese)