

**Zsuzsanna Müller**

Department of Library and Information Science, Faculty of Informatics,

University of Debrecen, Hungary

[mullerzs85@gmail.com](mailto:mullerzs85@gmail.com)

INFORUM 2011: 17th Conference on Professional Information Resources  
Prague, May 24-26, 2011

**Abstract**

---

Nowadays - because of the appearance of internet - the access to information has become easier, the sheer number of information sources has been increasing and therefore there are very few topics that we cannot find information about on the web. As a result of this users often look for advice to the internet, instead of conventional knowledge centres, like libraries, however the internet is full of dangers. As there is no control mechanism in the www there can be found more and more anonym texts which contain incomplete, incorrect or inexact information misleading users. To sum it up the reliability of online documents is often disputed.

Librarians and information professionals have recognized the significance of these phenomena and were obliged to accept the new roles, trying to ensure the ways and means of evaluation in the field of information and information sources, which has meant, among others, compiling aids in order to make the evaluation process easier.

Taking these changes into consideration the most significant aims of current study was to develop a general criteria system to make the judgement of reliability of online information sources easier and to create a specific system for Wikipedia evaluation. The choice is Wikipedia because of its popularity, its encyclopedic nature and the fact, that this information resource is being applied more and more in the field of education. This research is expected to help not only Wikipedia-users, but will also be a useful aid for both librarians and information professionals

## Evaluation of the reliability of Wikipedia

The appearance of internet has meant radical change in the processes of information acquisition. On the one hand, the access to information has become easier, the sheer number of information sources has been increasing. On the other hand, there can be found more and more anonymous texts in the web containing incomplete, inexact and incorrect information. As a result of this the evaluation of reliability has become one of the most important issues of information acquisition.

The last few years more and more studies - e.g. by Rieh and Belkin [5]; Frieder, Kantor, Cool and Belkin [2]; Amento [1] - have been published to demonstrate that the certain criteria can be applied efficiently to prop up information seeking and evaluation. Libraries in some western countries like e.g. UC Berkeley Library or Cornell University Library have also created so called „How to evaluate web pages” aids to help their users filter unreliable information.

In spite of these efforts there are some shortcomings in the field of evaluation of online information sources. First, earlier studies focus on the examination of quality and not specifically of reliability. Second, previous researches on the topic are too specific: they show how information seeking behavior of members of certain social strata including e.g. students or professionals takes shape. At the same time, the factors to be applied have been determined in a way as to be able to evaluate any type of online documents with the result that these library aids are too general.

Taking these into consideration the aim of the current study is to develop criteria for evaluation in the case of specific information sources. The choice has been Wikipedia due to its popularity, its encyclopedic nature and the fact, that it is used more and more in the field of education.

### **Wikipedia**

Wikipedia is a free online, multilingual encyclopedia established in 2001. Now it contains 3,626,934 articles, and has 14,465,628 registered users [8]. As Wikipedia is based on a very popular web 2.0 tool, titled wiki technology - which means that Wikipedia can be edited collaboratively by anyone who has internet access – its content is often disputed. Wikipedia says ”what is contributed is more important than the expertise or qualification of the

contributor”, but this does not mean that the contributions are totally uncontrolled. There are more and more tools inside Wikipedia which have been created to control the editing process. First of all it is important to mention the five pillars 8 which are in fact the principles through which Wikipedia operates:

- Wikipedia is an online encyclopedia: it refers to the encyclopedic nature of this database, its structure and mechanism that should be taken into consideration by each user.
- Wikipedia has a neutral point of view: this means that the content of Wikipedia articles should be built objectively, i.e. without personal opinions. Besides, information in articles should be referenced to verify that the data in it are exact and correct.
- Wikipedia is free content: meaning that copyrights should be respected by each user.
- Wikipedians should interact in a respectful and civil manner: users should find consensus in the disputed parts of articles to preserve relative stability.
- Wikipedia does not have firm rules: this pillar encourages users to contribute, but not in a reckless way.

Wikipedia policies – which demonstrate standards that all users should follow - and guidelines - contain practices for following those policies - have also great significance [8]. As an example the verifiability policy refers to the appropriate citing. According to the reliability the banning, blocking, bot and page protection policies are of high significance along with the role of administrators.

Banning policy: users making wrong edits can be banned from Wikipedia.

Blocking policy: users worsening the content of Wikipedia can be blocked.

Bot policy: users can develop programs that automatically make corrections in the content.

Page protection policy: articles of high significance can be protected from vandalism by blocking them.

And finally, it is important to mention administrators who have more rights than an average editor to control the editing process in Wikipedia.

As it can be seen, there are numerous tools trying to ensure and improve the quality and reliability of content. The success of this quality control mechanism has been proved by several studies published in this field.

Stvilia, Smith, Gasser and Twidale [6] researched the quality work organization of Wikipedia. They proved that most of contributors take the issue of quality very seriously which results in

continuing development in the collaborative work, therefore also in the quality improvement process. They also claimed that authority – as well as cohesiveness and accuracy - has great importance in the quality assurance process.

Denise, Smith and Williamson [3] have examined edits and editors, with special reference to the motivation of the contributors. The result of the examination showed that the registered user edit more often than the unregistered, and the more often a user contribute, the higher reliability he/she has. This means that the quantity of contributions is in connection with the quality and reliability of content, therefore the authority of a contributor. Results also proved that anonym unregistered volunteers contribute on a higher level, than those registered.

Besides, Wilkinson and Huberman [9] demonstrated a strong correlation between the number of edits, number of distinct contributors and the article quality, while Voss [7] proved that the more people read an article the more errors are amended. This last one also means that the popularity is important factor in the quality of content.

As it can be seen, earlier studies, as well as the editing control processes inside Wikipedia provide useful information about the reliability of Wikipedia making evaluation much easier.

### **Criteria system for reliability evaluation of Wikipedia**

As a system of evaluation criteria of printed documents had been created earlier, they were a good starting point to determine criteria for reliability evaluation of Wikipedia. The frame of the criteria system was created on the basis of Patrick Wilson's cognitive authority theory [10] and Robert S. Taylor's quality model [4].

Patrick Wilson, an exponent of library and information philosophy researched the processes and factors in information acquisition, with special reference to cognitive authority. Wilson's basic concept was that people can construct knowledge in two ways, by personal experiences and in second-hand ways. As information acquired from others may represent uneven credibility, we have to examine it before being actually applied. He claimed that cognitive authority is a kind of influence on one's thoughts that one would consciously recognize as proper. To sum it up, having cognitive authority means reliability or credibility. Therefore investigation in this area should be the most important part in information evaluation. He determined the factors of cognitive authority like authorship, publishing data, document type and intrinsic plausability which are still valid, and can be perfectly used as the criteria of reliability for printed documents. The criteria and the participating elements of this process, delineated by Wilson are the following:

Authorship: a document is considered reliable if the author can be trusted or in other words has a reputation as trustworthy and knowledgeable in his respective field of study.

Publishing data: as Wilson has determined various data concerning the publisher, the scholarly or scientific journal as well as a review publication can equally be interesting in terms of interpretative power. This can mean the following: the interpretation of the reputation and scope of an individual publisher or journal in view of the subject material to be issued.

The significance of a type of a document can also be emphasized. E. g. the fact that the document in question is a dictionary or an encyclopedia can also have a special meaning.

Wilson's tenet about intrinsic plausability means that a document has only one chance to capture the attention of a reader and if it is successful we will continue reading because the writing appears reliable and interesting.

Taylor's model was created in the early 1980s for evaluating information systems like libraries or information retrieval processes. In spite of its age it is still relevant and it can also be used - after some changes - for framing the quality and therefore the reliability criteria of online information sources [4].

As it can be seen in Table 1., the first column of Taylor's model includes the categories of criteria which are important for the users when evaluating an information system. Criteria in the second column are more specific, they are actually the factors of the evaluation system. It is important to mention that each category contains more criteria, but according to the objective topic of the current study only the quality category was detailed in Table 1.:

**Table 1.** Taylor's Quality in Value-added Model

<b>Categories</b>	<b>Criteria</b>
ease of use	
noise reduction	
quality	accuracy
	comprehensiveness
	currency
	reliability
	validity
adaptability	
time-saving	
cost-saving	

After summarizing the results of earlier researches the criteria system was created according to the specific features of Wikipedia as it can be seen in Table 2.

**Table 2.** Reliability criteria for evaluation of Wikipedia

<b>Primary Criteria</b>	<b>Secondary Criteria</b>	<b>Tertiary Criteria</b>
<i>quality</i>	<i>accuracy</i>	
	<i>comprehensiveness</i>	
	<i>currency</i>	
	<i>reliability</i>	author
		content
		<b>entry qualifications</b>
		<b>verifiability</b>
	<i>validity</i>	

Terms in italics are from Taylor's model, the others from Patrick Wilson. The new, Wikipedia - specific reliability criteria are in bold.

As the four criteria – author, content, entry qualification and verifiability - are of uneven significance, they were weighing with the method of direct estimation, which can be seen in Table 3.

**Table 3.** Point system

<b>Main criteria</b>	<b>Secondary criteria</b>	<b>Scores to be given</b>
Author (2)	qualification, accomplishment	0-negative 1-neutral 2-administrator
Verifiability (2)	references, other bibliographical data	0-not satisfactory 1-satisfactory
Entry qualification (1)		0-negative 1-neutral 2-featured
Content (1)		0-not satisfactory 1-satisfactory

As an example this means that if verifiability in an article is unsatisfactory, the points to be given are two (2x1) or if the entry qualification is neutral, its point index is one (1x1).

According to the point system the maximum that can be given is eight, the point limit of reliability is five. However, people often do not know much about the topic they are searching; therefore they cannot judge the content. As a result of this, points of content criteria can be counted only as plus points.

### **Evaluation and its method**

Creating criteria of reliability was followed by an evaluation consisting of three stages. First, Wikipedia articles chosen from a special field of history, Greek stepmother mythology were evaluated in a conventional way, i.e. by comparing them to the similar articles from a reliable source, Encyclopedia Britannica (Examination I.). As my thesis was written about this historical topic, information in Wikipedia articles was also judged according to my own knowledge.

After that Wikipedia articles were evaluated by using the special criteria as well (Examination II.)

Finally, it is also important to make some comparison based on the results of earlier researches (Examination III.). A study by Wilkinson and Huberman [9] mentioned before demonstrated the following two useful rules:

- the older is an article, the more edits are made on it
- the more edits there are, the higher quality of its content is probable

Taking the above into consideration it would be useful to compare some data like e.g. the number of edits, number of contributors about the chosen articles and some random featured articles.

### **Results**

#### *Examination I.*

The results of Examination I. show that basic information can be found in Wikipedia articles as well as in the Encyclopedia Britannica. Wikipedia articles usually contain more complementary information, like supplementary applications, pictures or genealogy tables.

Wikipedia is more readily available and very easy to use. Though the reliability of Encyclopedia Britannica cannot be disputed, information in Wikipedia articles always need to be scrutinized more thoroughly, which makes its use more difficult. To sum it up, information

to be found in a sample of Wikipedia articles was judged as reliable. It is also important to mention that there cannot be found distinct articles about three mythological persons either in the printed or in online Britannica.

*Examination II.*

The results of Examination II. performed by using the specific criteria system are entirely different. Analysis of articles done with the criteria of authorship, content, verifiability and overall qualification shows the following results:

Investigating authorship was a very difficult and hard task, because to determine authority of authors, numerous edits and profile pages should have been examined. Results show that the number of authors for the articles are between 10 and 79, but really significant edits were made by 2 - 15 authors. Their reputation, therefore the acknowledgement and qualification can usually be determined as neutral.

Though verifiability is often unsatisfactory – as it can be seen in Table 4., the qualifications of articles are often neutral. As the point limit is 5 points, as was mentioned before, only three articles out of the ten can be qualified as reliable in Wikipedia.

**Table 3.** Verifiability and qualifications of articles in Examination II.

<b>Wikipedia Articles</b>	<b>Verifiability</b>	<b>Qualifications of articles</b>
1.Athamas	not satisfactory	neutral
2.Creusa	not satisfactory	negative
3.Hermione	not satisfactory	neutral
4.Hippolytus	not satisfactory	negative
5.Ino	satisfactory	neutral
6.Lamia	satisfactory	neutral
7.Nephele	not satisfactory	neutral
8.Phaidra	not satisfactory	neutral
9.Semele	satisfactory	neutral
10.Theano	not satisfactory	negative

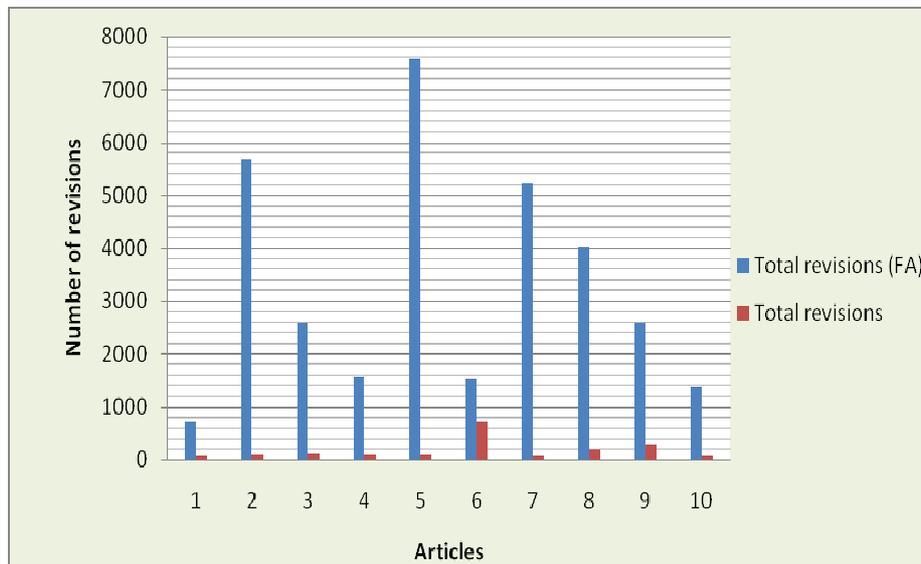
Though the difference between the results can be explained with more factors, the most significant one is the issue of verifiability.

### Examination III.

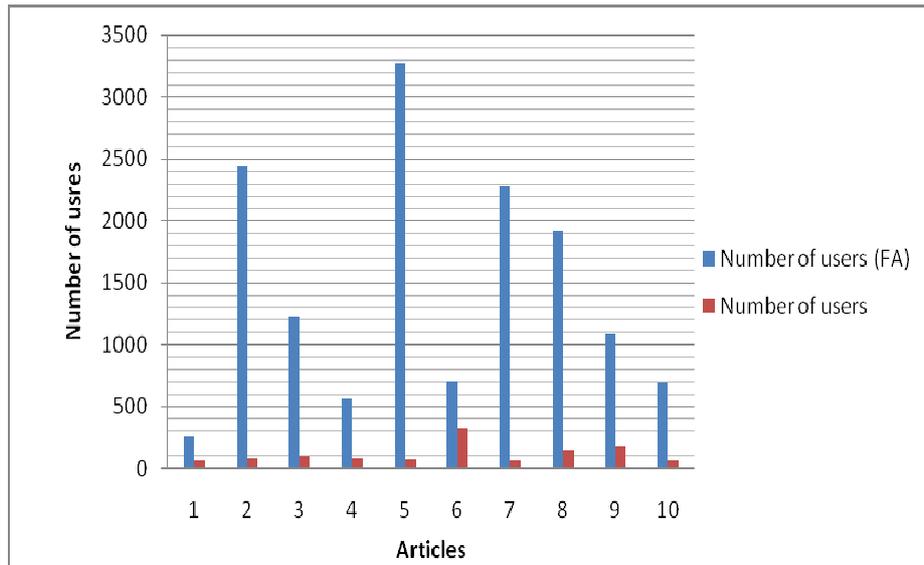
As featured articles are considered as the best in Wikipedia, ten out of them were randomly chosen (1. Attalus I., 2. Augustus, 3. Charles II. of England, 4. Demosthenes, 5. Gerald Ford, 6. Georg I. of Great Britain, 7. El Greco, 8. James I. of England, 9. Johannes Kepler, 10. Edward Teller) to be compared with the earlier sample. According to the researches dealing with content quality and reliability in Wikipedia, the following data were examined in the two groups of articles: number of revisions, number of contributors, and age.

As it can be seen below – Figure 1.,2.,3. -, number of edits and number of total revisions are higher in the featured articles than in the first sample, while the age of articles are nearly equal.

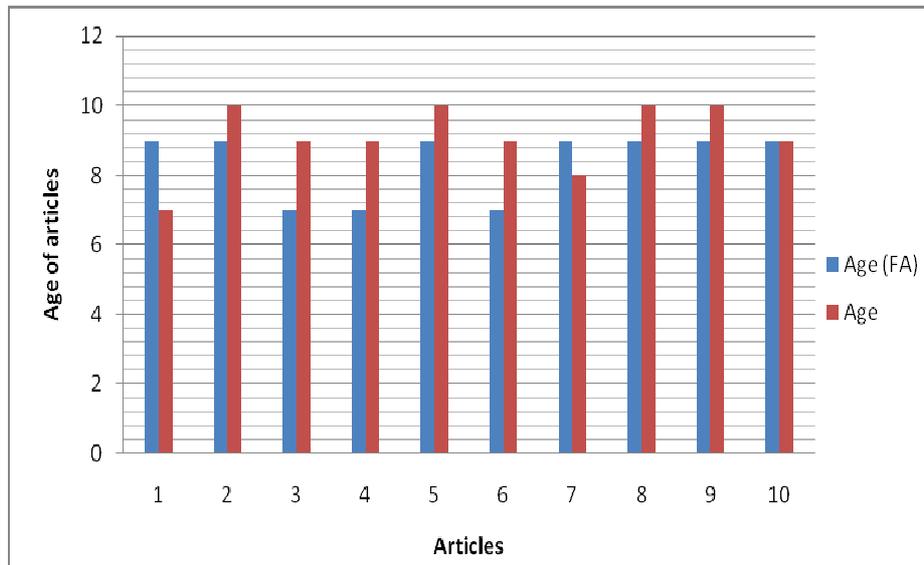
**Figure 1.** Number of total revisions of examined articles



**Figure 2.** Number of users of examined articles



**Figure 3.** Age of examined articles



This means that according to the quality policy of Wikipedia the articles of first sample have less authority, i.e. they are really less reliable.

It is also an interesting result, that Examination II. and III. also proved that articles titled Lamia and Semele are of the highest degree of authority.

### Conclusion

As the investigation performed with the criteria system shows we have different results than we have according to the conventional, content-based evaluation which simply confirms that it would be important to refine this system, maybe with enriching it with other Wikipedia-specific features like e.g. the open-source online reputation system, WikiTrust.

## References

- [1] Amento, Brian; Terveen, Loren; Hill, Will: Does Authority mean quality? predicting expert quality ratings of Web documents. *Proceedings of the 23rd annual international ACM SIGIR conference*, 2000. p. 296-303.
- [2] Cool C., Belkin, N. J., Frieder, O., Kantor, P. (1993): Characteristics of texts affecting relevance judgments. *Proceedings of the 14th National Online Meeting*, p. 77-84.
- [3] Denise, Anthony, Smith, Sean W., Williamson, Tim: The Quality of Open Source Production: Zealots and Good Samaritans in the case of Wikipedia. In.: *Technical report*, Department of Computer Science, Dartmouth College, 2007.
- [4] Eisenberg, M., Dirks, Lee: Taylor's Value-Added Model: Still Relevant After All These Years. *3<sup>rd</sup> Annual iConference*, 2008.
- [5] Rieh, S. Y., Belkin, N. J.: Judgement of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology*, 2002. 53. Vol. 2. Issue, p. 145-161.
- [6] Stvilia, B., Twidale, M. B., Gasser, L., & Smith, L. C. (2008). Information quality work organization in wikipedia. *Journal of the American Society for Information Science and Technology*, 2008. 59. Vol. 6. Issue, p. 983-1001.
- [7] Voss, Jacob: Measuring Wikipedia. *International Conference of the International Society for Scientometrics and Informetrics: 10th*, Stockholm, 24-28 July 2005.
- [8] en.wikipedia.org
- [9] Wilkinson, Dennis M.; Huberman, Bernardo A.: Cooperation and quality in wikipedia. *Proceedings of the 2007 international symposium on Wikis*. 2007. p. 157-164.
- [10] Wilson, P.. *Second-hand knowledge: an inquiry into cognitive authority*. Westport, Connecticut: Greenwood Press, 1983.