

MEASURING AUTHOR CONTRIBUTIONS TO THE MEDIAWIKI

Gökhan Akçapınar, Petek Aşkar

Hacettepe University

Dept. of Computer Education and Instructional Technologies, Ankara, Turkey

ABSTRACT

Wiki is a one of the most effective collaborative authoring tool which comes with Web 2.0. They bring about new vision to group work and help to take place in online learning environments rapidly. Although there is a great potential for learning environments, assessment process is time consuming and difficult if conventional methods are used. In this study, the purpose is to develop a tool to measure the author contribution in collaborative work within a wiki environment. By means of the developed tool it is possible to lessen time and energy to assess these contributions.

KEYWORDS

MediaWiki, Wiki, Collaboration, Author Contribution, Measuring, Tool.

1. INTRODUCTION

Assessment of the collaborative work in wiki environment is very important particularly when it is used in the field of education. Forte and Bruckman (2007), mention that if teachers cannot assess collaborative wiki work, then we cannot expect wiki to be adopted for formal education, despite the potential learning gains for students. When we analyze the structure of wiki we can easily see that they have some assessment difficulties because of their collaborative nature. Products may be assessed by using conventional methods, but in collaborative work not only product but also measuring individual contribution and collaborative process must be taken into account. For this reason alternative methods must be used in addition to qualitative analysis. One of these methods is quantitatively analysis of wiki database. It was also used in this study. There are more than a hundred wiki engines to create wiki based web sites and most of them have history feature that keeps every change conducted by users in database or log files. Although this is a great opportunity for us to measure individual contribution and collaborative process, without using any additional tool to automate this process it will be very complex and time consuming. In this paper, we first discuss existing approaches to measure users' contribution to wikis. Then we determine our metrics in the light of related research's findings and a tool was developed as WikLog to obtain data related to our metrics from database automatically.

1.1 Wiki

Wiki is one of the Web 2.0 technologies that promote the collaborative creation of information content. Leuf and Cunningham (2001) define a wiki as "a freely expandable collection of interlinked Web pages, a hypertext system for storing and modifying information-a database where each page is easily editable by any user with a forms-capable Web browser client". Wikis are increasingly being accepted as a new breed of collaborative technology. Wiki technology can impact knowledge management, and can support knowledge creation and sharing (Leuf and Cunningham 2001; Wagner 2004; Lamb 2004). The largest available example of wiki is Wikipedia. It is an online encyclopedia whose content was created collaboratively by volunteers all over the world. It's open to everyone's edit. Today English version has 2,653,979 articles and growing rapidly day by day (Wikipedia, 2008).

1.2 Measuring Author Contribution on Wikis

Although there is still very little research presenting quantitative analyses of Wikis, we take some previous works into consideration to mention a general outlook of this interesting field. Most of these studies are conducted on Wikipedia. The common purposes of these researches are using quantitative data to measure content quality or author contribution which leads to the creation of this content. In these studies a great deal of methods for automatic assessment of author contribution or article quality has been proposed. Initial studies used only number of edits as a metric to measure contents' quality. Anthony et al. (2005), show that there is a strong correlation between the quality of contributed contents and the level of contributions made by individual authors. That is to say the quality of contributions increases as the number of edits increase. Ortega and Gonzalez-Barahona (2007), extended this methodology, on the basis of quantifying per period parameters. They also analyzed users according to their activity level during specific periods of time. This is important to reveal users behavioral pattern and get a more accurate picture of the evolution in time. Some of them suggested edit counts and number of distinct author as a metric (Lih, 2004; Wilkinson & Huberman, 2007). Others use more complex metrics to measure author contribution to the wikipedia (Stvilia et al. 2005; Adler et al. 2008). There are few researches using quantitative data to measure student's contribution in educational wiki environments (Trentin 2008; Forte & Bruckman 2006; Forte & Bruckman 2007).

2. DEVELOPMENT PROCESS OF THE TOOL

2.1 Measures

We determine four metrics in the light of related research's findings to measure author contribution in wiki environment namely: number of pages, number of edits, number of links and number of words.

1. Number of pages: The number of pages created by the user within the total number of pages created by the members of users group.
2. Number of edits: The number of edits conducted by the user within the total number of edits conducted by the members of users group.
3. Number of links: The number of links created by the user within the total number of links created by the members of users group.
4. Number of words: The number of words created by the user within the total number of words created by the members of users group.

2.2 MediaWiki

The developed tool works on MediaWiki based wikis. There are more than a hundred wiki platforms for creating wiki based web site. One of them is MediaWiki. We choose it because it's a widely used. It's open source and free of charge. Wikipedia also use it as a wiki engine. In educational aspect MediaWiki logs some important data in database about student activity in wiki environment. But it's difficult to obtain some of the important data from database like students' words and links count since it's not especially produced for educational environments.

2.3 MediaWiki Database Structure

Before starting to develop tool, we analyze MediaWiki database to determine which tables can be used related to our metrics. MediaWiki version 1.16.0 which is the latest version when the tool was developed has 33 tables. We select seven tables concerning our metrics namely page, pagelinks, recentchanges, revision, site_stats, text and user. "User table" stores information about users such as *user_id*, *user_name*, *user_password*, *user_touched* storing last time a user made a change on the site etc. "Page table" can be considered as the "core of the wiki". Each page in a MediaWiki installation has an entry here which identifies it by title and contains some essential metadata. "Revision table" keeps metadata for every edit done to a page within the wiki. Every edit of a page creates a revision row. "Site stats" table contains a single row with

some aggregate info on the state of the site. These values are *total_views*, *total_edits*, *good_articles*, *total_pages* and *users*. “Text” table holds the wikitext of individual page revisions. “Pagelinks” table tracks all internal links in the Wiki.

2.4 WikLog

WikLog is a MediaWiki database analysis tool developed by the researchers based on C#.Net. It's directly connected to any MediaWiki database and receives data about user activities by sending SQL queries. Figure 1 shows the main window of the software. WikLog has eight windows and all of these are explained in detail below.

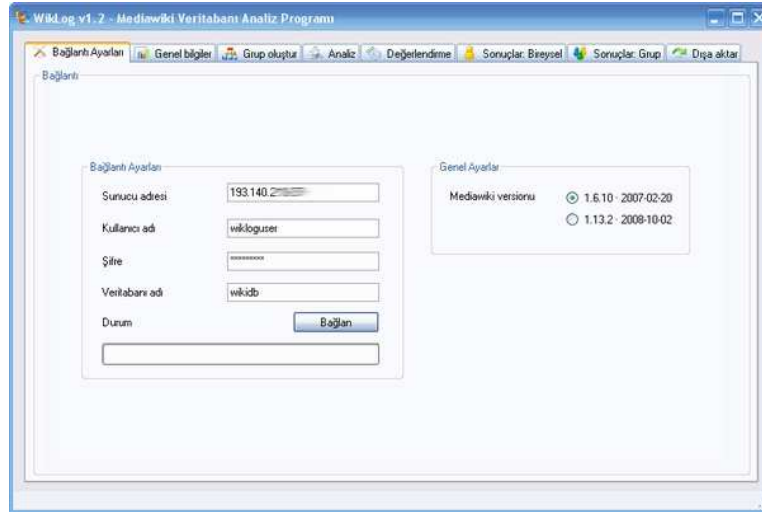


Figure 1. Main window

Connection: connection window is designed for entering information about database which is desired to connect. The following information must be supplied for connection. Server address: IP address for MYSQL database. Username: username for Wiki database. Password: wiki database user's password. Database name: Wiki database name. At this window evaluator can select MediaWiki software version. Latest MediaWiki version was 1.6.10 when the tool was developed but further versions are also supported by the tool.

General Statistics: at this window general statistics are shown about wiki such as total visits, total edits, content pages, total pages, registered users count and average edits per page.

Create Groups: at this stage evaluator assigns students to groups. These groups must be same to that of wiki environment. First evaluator enters groups count and click create button to create listboxes as much as groups count. Then list students. When the evaluator presses right click on selected students, names of all groups are listed in menu created by him and they are assigned by him to related group by selecting groups name on the menu. Because the process works dynamically there is no restriction in the number of groups. Since this process is time consuming evaluator has option to save group lists. Group list is saved as XML and imported whenever needed.

Analysis: This part is the core of the tool. All calculations belonging to individual and groups are conducted at this window by using relational queries among three different tables. Figure 2 shows these tables and relations. When analyses are conducted groups created in previous section are taken into consideration. While individual and group contributions are calculated at first stage nested loops are formed. The outer one is responsible for operations among groups and the inner one is responsible for operations within group. By using this method it is possible to analyze all users and all groups created dynamically in previous section. Dynamically generated queries are submitted to the database and gathering data related to our metrics like “number of new page” and “edits count”. Since it is not possible to obtain “number of links” and “number of words” directly from database, the calculation of these metrics is fairly difficult. Each edits conducted by users on wikis store in revision table and each line corresponds to an edit. Data stores in this table are explained in the previous section about database structure. Contrary to expectations the last line in

this table always stores current versions of the page, but doesn't store contribution conducted by the user. For this reason we developed an algorithm to make it possible to get user contribution and also our metrics "number of links" and "number of words". With the help of this algorithm each edit is compared with the previous one and the last edit which is the current versions of the page. Later differences are stored temporarily in a table as a quantitative value. This operation is repeated for every page created by each group.

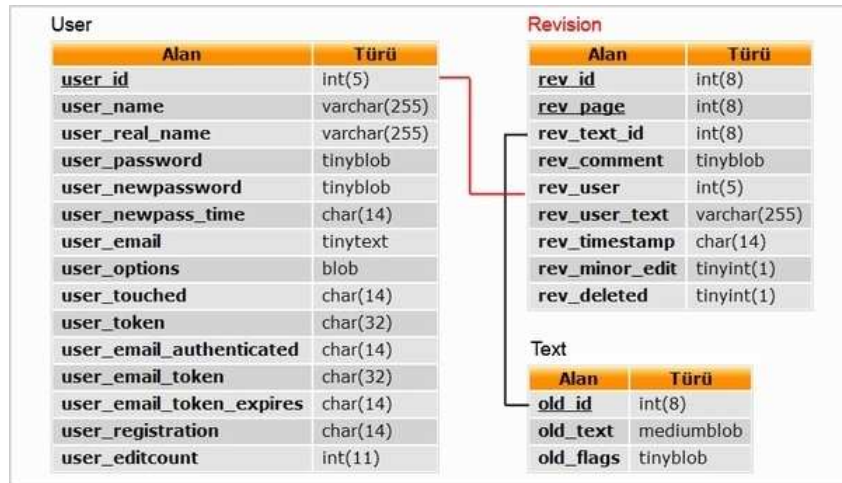


Figure 2. Tables and relations used for analysis

Assessment: when analysis process is completed each user's contribution is obtained quantitatively. At this window evaluator determines metrics' weight over total score.

Individual Scores: at this window presented all users' quantitative data about related metrics and their final scores which are calculated with respect to weights determined by the evaluator.

Group Scores: at this window presented all groups' quantitative data about related metrics and their final scores which are calculated with respect to weights determined by the evaluator.

Export: with the help of button on this window a data produced by the tool export to Microsoft Excel to use further assessment process.

2.5 Software and Hardware Specifications

Windows based computer and internet connection is required for using tool. In addition following applications must be installed on the system: Microsoft .NET Framework 2.0 or later, and MySQL.Net Connector. Although there is no need for specific hardware, analysis process will be faster depending on hardware and internet connection's speed.

3. CONCLUSION

Although writing process is very important teachers generally grade only the final product because of the difficulties in the assessment of writing process (Forte & Bruckma, 2006). Due to the structure of wiki environment, assessment is difficult if conventional methods are used. Most of wikis have page history feature that is important to track and assess students' involvement and contribution in the collaborative writing process. But without using any additional tool analysis of these data too much effort and time are necessary.

In this study a tool was developed for making the assessment automatically. Through the tool the assessment process automate and results are obtained in a short time. Contribution of users who rearrange or polish the content is at least as important as users' who create content (Adler et al. 2008). Tool can be improved in a way that it can equally score these two types of contributions. In further research new metrics can be added to increase accuracy of quantitative analysis.

REFERENCES

- Adler, B. T. et al, 2008. Measuring Author Contributions to the Wikipedia. *WikiSym '08: Proceedings of the 2008 international symposium on Wikis*. Porto, Portugal: ACM.
- Anthony, D. et al, 2005. *Explaining Quality in Internet Collective Goods: Zealots and Good Samaritans in the Case of Wikipedia*. Retrieved January 10, 2009, from Fall 2005 Innovation & Entrepreneurship Seminar at MIT: <http://web.mit.edu/iandeseminar/Papers/Fall2005/anthony.pdf>
- Forte, A., and Bruckma, A., 2006. From Wikipedia to the Classroom: Exploring Online Publication and Learning. *International Conference of Learning Sciences ICLS* (pp. 182-188). Bloomington USA: International Society of the Learning Sciences.
- Forte, A., and Bruckman, A., 2007. Constructing text: Wiki as a toolkit for (collaborative?) learning. *WikiSym '07: Proceedings of the 2007 international symposium on Wikis* (pp. 31-42). New York: ACM.
- Lamb, B., 2004. Wide Open Spaces: Wikis, Ready or Not. *EDUCAUSE Review* , 30 (5), 36-48.
- Leuf, B., and Cunningham, W., 2001. *The WIKI Way Quick Collaboration of the Web*. Addison-Wesley.
- Lih, A., 2004. Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource. *Proceedings of the Fifth International Symposium on Online Journalism*. Austin.
- Ortega, F., and Gonzalez-Barahona, J. M., 2007. Quantitative Analysis of the Wikipedia Community of Users. *WikiSym '07: Proceedings of the 2007 international symposium on Wikis* (pp. 75-86). Montreal, Quebec, Canada: ACM.
- Raman, M. et al, 2005. Designing knowledge management systems for teaching and learning with wiki technology. *Journal of Information Systems Education* , 16 (3), 311-321.
- Stvilia, B. et al, 2005. *Information quality discussions in Wikipedia*. Florida: Technical Report, Florida State University.
- Trentin, G., 2008. Using a wiki to evaluate individual contribution to a collaborative learning project. *Journal of Computer Assisted Learning* , 1-13.
- Wagner, C., 2004. Wiki: A Technology for Conversational Knowledge Management and Group Collaboration. *Communications of the Association for Information Systems* , 13, 256-289.
- Wikipedia., 2008. *Wikipedia:About*. Retrieved 12 20, 2008, from Wikipedia, the free encyclopedia: <http://en.wikipedia.org/wiki/Wikipedia:About>
- Wilkinson, D. M., and Huberman, B. A., 2007. Cooperation and quality in Wikipedia. *WikiSym '07: Proceedings of the 2007 international symposium on Wikis* (pp. 157-164). New York, NY, USA: ACM.