OPEN PUBLICATION SYSTEM
Evaluating Users Qualification and Reputation

Gabriel Simões, Leandro Krug Wives and José Palazzo M. de Oliveira
II/UFRGS Av. Bento Gonçalves, 9500 Agronomia - Caixa Postal 15064 – CEP 91501-970 Porto Alegre, RS, Brazil

Keywords: Reputation, Collaboration, Trust, Quality assessment, Cooperation.

Abstract: In cooperative editing environments (e.g. Wikis), users can create and edit documents in a freely and cooperatively manner. However, sometimes it is interesting to identify if the contributions made by one user are really reliable, since users don't trust each other in an explicit way. This point is a central discussion about the open publishing truthfulness. While it is difficult to automatically identify the relevance of each user contribution, it is more plausible to evaluate their reputation as perceived by the community. In this paper we describe a model to evaluate the user's reputations in a Wiki community and the prototype developed for its evaluation. The basic assumption is that we are dealing with a homogeneous cooperative group on a limited knowledge context. This environment exists, for instance, in a cooperative group trying to consolidate organizational implicit knowledge into documents as a class-based report generation. This kind of environment is very useful to stimulate collaborative learning.

1 INTRODUCTION

The consolidation of Web 2.0 (Millard, 2006) brings more attention to open content edition environments. These environments work with spontaneous user's contributions to enlarge their contents. Wikipedia, the most successful Wiki application on the web, is equivalent to paper encyclopedias in terms of contents and, according to Giles (Giles, 2005), may be considered as trustful as closed revision environments. Apart from this evaluation, some criticisms arrive when we try to mix conflicting points of view maybe influenced by conflicts of interest. For instance, similar subjects may be interpreted with antagonistic perceptions or even ideological back-grounds.

The combination of the potentials of Wiki environments and the production of scientific knowledge developed in a collective manner allows productivity growth, researcher's integration and the development of a review process that is more transparent and interactive. Wiki environments, however, have problems related to the lack of trustworthy among users, since they generally don't know effectively each other. Trust is basic to any relationship in which the attitudes of the involved parties cannot be controlled (Jarvenpaa,) and it is usually addressed by reputation systems. These systems collect and distribute information regarding the behaviour of the individuals (Resnick, 2005).

In this sense, we developed a dynamic qualification mechanism based on reputation evaluation techniques. This mechanism can minimize the lack of trust problem in Wiki environments and qualify the users of a homogeneous community. This mechanism analyzes the user's reputation, and it is based in quantitative and qualitative data obtained from the Wiki environment and from other users' evaluations. With this information at hand, it is possible to create a rank to be employed as a relative index of users' reputation and to increase trust or confidence among users. An extension to the MediaWiki system was created to implement and evaluate the proposed qualification mechanism.

This research started from the experience of the last author with collaborative learning based in a research report generation by graduate students employing the Google Docs. All work was peer evaluated at the end of the course in a manual way. With the open cooperative editing environment the user reputation evaluation stimulates the individual work quality by a continuous ranking.

Simões G., Krug Wives L. and de Oliveira J. (2009).
OPEN PUBLICATION SYSTEM - Evaluating Users Qualification and Reputation.
In *Proceedings of the First International Conference on Computer Supported Education*, pages 199-204
DOI: 10.5220/0001968201990204

2 BACKGROUND

Reputation systems can be described as a computational implementation of the word-of-mouth information dissemination mechanism (Hu, 2006). These systems collect, distribute and aggregate feed-back about users' past behaviour. Their application can assist people in getting trust about other people, even if they don't previously know each other, or if they have a limited knowledge of the partners. According to Resnik et al. (Resnick, 2005), "Reputation systems seek to establish the shadow of the future to each transaction by creating an expectation that other people will look back on it".

Auction and e-commerce sites apply variations of reputation systems to provide some insurance to users. Collaborative environments can use variations of reputation systems to increase trust among its users. The community qualification assessment of individual researchers was recently considered as one of the central criteria for the evaluation process. Not only individual researchers are under social evaluation but also the conferences and journals are receiving evaluations based in the social perception of their importance (Butler, 2008).

We developed an alternative editing process that uses the approach developed in our group to support the open reviewing process (Oliveira, 2005). In this case, users can edit, comment and review documents created by other users. In this approach, the process is centered in the open edition and reviewing of scientific papers, which is an alternative approach to the blind or double blind review system, mostly adopted by the academia. Within this approach, the knowledge is collectively generated and reviewed in a transparent way.

We decided for existing Wiki environments to the production of scientific and technical documents, since they have the needed framework to manipulate texts, besides a good user management and version control. The main problem found in these environments is that most users have a limited knowledge of the members of the process (i.e., they may not know the other members). In an intercontinental research project that includes many participants, for example, this situation also happens. This is a consequence of the fact that interactions are mostly restricted to the exchange of data over the web, as physical meetings are very expensive, affecting how confidence or trust among authors/partners is established.

Reputation systems are employed to minimize this problem, and they create confidence between users of these systems. Recently, Google started the Knol service allowing users to write, evaluate, comment, review and contribute to other authors' works. Authors can accept or not the contributions made on their work by other authors, but the evaluations and reviews regard the whole document, not individual contributions or comments.

In our approach, every user knows what and who edited and contributed to each document, and they can evaluate other user's contributions and comments. One important point is that the process is still peer-reviewed, and reputation and confidence are yet important factors, but they are built on the bases of the social network. To evaluate this approach, we have extended the MediaWiki environment, incorporating some features, which are described in the next section, to implement the proposed reputation model.

It is important to state that this paper is based on the qualification mechanism conceptually developed in (Oliveira, 2005), which describes an open editing model in which there are three types of users: author, commenter and reviewer. When a person creates an account, he or she receives the 'commenter' status, which gives the ability to annotate documents, after a 'commenter' may be promoted to 'reviewer'. The basic idea to support this promotion is based in a comparison among the user rating and the paper rating, if the user has a rating that is equal or higher than the paper's rating, he/she will be allowed to review directly the text of the paper. This is an approach slightly different than the traditional Wiki process, in which every user can edit every page except for certain pages that are consolidated and blocked. Authors can comment and create new documents.

Reviewers are more qualified users that can also edit others documents. The role of a reviewer is also different from the role of the traditional reviewers involved in the academic reviewing process. In the traditional closed reviewing process, they may suggest changes to improve quality. Here, they directly contribute to the quality of the document by editing the text. Each person participating in the process is identified and all the actions are registered; the authors may accept or reject the received contributions.

We will validate the real-world operation of this approach in an on-going project for the publication of an experimental open edited version of a computer science journal, where the best papers, written by Ph.D. students, will be published using collective authoring, with the first author being the original writer of the document.

Next section describes our reputation model and how the user qualification is measured. It is important to state that, in this paper, we only address the roles of authors and reviewers.

3 REPUTATION MODEL

The qualification mechanism conceptually developed in (Oliveira, 2005) and implemented by our prototype give points to users according to their interaction with the system and also takes in account the evaluation their documents receive from the community. This is a continuous grading mechanism that allows a user to start from nil recognition and reach the better grade by a peer-to-peer assessment process. Considering previous evaluations, we expect to minimize Sybil attack problems. The Sybil attack is common in peer-to-peer systems when one entity responds by more than one identity, creating information bias (Doucer, 2002).

Interaction is a source of quantitative data. Each time a user access one page, we count one hit of this user in that page. User pages and documents created by the user are not taken in account. Consecutive accesses to the same page are computed if the interval between the accesses is greater than 24 hours; this is a heuristic to identify different accesses, perhaps composed by multiple pages reads during a specific time-period. The access rate is an indicative of the popularity of the document. It is clear that popularity is not an absolute quality indicator, but this happens also in the generally accepted impact index. In the extreme case, a paper may be referenced a lot of times as a counterexample but the reference counter is increasing. The discussion is related to a conceptual and philosophical debate about what quality and popularity are; then we decided to take the commonly accepted approach that a large amount of access indicates a good content.

On the other hand, qualitative data is based on user´s evaluation. All the documents available in the environment can be evaluated by any registered user. This approach is similar to the model found in reputation systems, in which evaluators indicate their grade of satisfaction in relation to the evaluated resource. To enable this evaluation, one effortless visual component containing five stars (Figure 1) was inserted on each page. Each star, from left to right, corresponds respectively to ‘very bad’, ‘bad’, ‘neutral’, ‘good’, and ‘very good’ in a five points Likert scale.

Figure 1: Visual evaluation component.

Qualitative evaluation measures the opinion of each user in relation to one specific document. When a document is evaluated by different users, there is a probability that it will receive different evaluations. However, as each user has a different qualification and reputation, the evaluation he or she gives must be related to this attribute; the most considered users, with great reputation, are more valorised in their opinions than the less considered ones. The underlying supposition is that we are working with a homogeneous cooperative group. For heterogeneous groups, with different and conflicting points of view, clustering mechanisms may be employed to identify diverse sub-communities.

We developed a method named EQ1 to deal with the different qualification of users. In this method, one positive evaluation of a more qualified user will count more than few negative qualifications of less qualified users. The purpose of this method is to generate confidence among users by an open and socially constructed reputation ranking. It is more plausible to have a more relevant and important evaluation from a well qualified user to the cooperative community, since this user has more social appreciation and reputation. We also worked with a method that does not take into account the user qualification for comparison purpose. It is named EQ2 method. Both methods are presented bellow.

3.1 EQ1 Method

Most qualification approaches are only quantitative-based, considering the quality as a side-effect of the quantitative data. The approach presented here is also quantitative, since it takes in ac-count the number of interactions performed by the users. However it is also explicitly qualitative, since it is based on the evaluations performed by the community about the level of approval of each document and on the evaluator’s reputation. EQ1 method was designed for the specific application described before, in which a Wiki system is employed to allow researchers edit and review documents in an open process, but it can be extended or adapted to other applications.

The qualification points produced by the EQ1 method generate a users ranking. This ranking is employed to generate a social confidence index, which is the central factor in this context. The

confidence points are also applied to suggest the quality of the documents assessed by the community. EQ1 aggregates characteristics from reputation systems, since it takes into account the evaluator competence or qualification in the evaluation. Then, better qualified users (or users with better reputation) give or take more points than lower qualified users (with low reputation).

The EQ1 algorithm adds to the document's author qualification the product of the normalized evaluator qualification by the given qualification value. This qualification is computed by the following equation (1).

$$PA = P'A + (F \,.\, N(PE)) \quad (1)$$

In this equation, given that A is the author of the document being evaluated and E is the evaluator, PA is the resulting qualification of the author, and P'A is his previous qualification (all authors start with a neutral qualification of 1). F is the multiplication factor, which can be -3, -2, 1, 2, and 3. These values correspond to the five criteria of evaluation, already stated: very bad (-3), bad (-2), neutral (1), good (2), and very good (3). N(x) is a normalization function employed to map the evaluator's qualification (x) to values between 0 and 1. Thus, N(PE) returns the normalized qualification of the evaluator, and consequently the final score is also ranged between -3 and 3.

Figure 2 shows an example of this process. The evaluator chooses his grade of satisfaction for the document he has just read, clicking on the corresponding star. This is translated to the corresponding numeric value and used in the equation.

Figure 2: Evolution of an author´s qualification.

3.2 EQ2 Method

This method does not take into consideration the evaluator's qualification, and was defined as the base-line of our system. Then we can analyze and compare the behaviour and the tradeoffs of our system against the basic method. EQ2 is computed by Equation 2.

$$PA = P'A + F \quad (2)$$

In this equation, as in the previous, PA is the resulting qualification of the author, P'A is his previous qualification and F is the qualification given by the evaluator.

4 QUALIFICATION FEEDBACK

The reputation of a user is created by the qualification mechanism. In the prototype, we have implemented two forms of user qualification feedback: the user qualification ranking and the user dashboard.

User Qualification Ranking. In this ranking, users with greatest qualification are located at the top of the list. The ranking consists on an ordered list, composed by the user identification and the associated qualification.

The ranking uses a decreasing order of qualification, and is dynamic generated. It is based on data available at the request time. To achieve better positions in the ranking, the user must access and write documents. These documents must be accessed and evaluated by other users to generate ratings. These ratings are added to the user's qualification to change the position in the ranking. The ranking is available to all users.

User Dashboard. Dashboards are graphic representations that allow quick visualization and comprehension of a data series (Butler, 2008). Dashboards are employed on business environments, keeping critical information available for decision takers.

In our case, dashboards are used to aggregate quantitative data about users and documents. They were implemented as a MediaWiki extension, and can be accessed from any user page, using a loupe icon. When someone clicks on the loupe, two graphs are shown: the bullet graph and the bar graph.

Bullet graphs (Figure 3) were created by Stephen Few (Few, 2006). A bullet graph can represent complex information. The graph is composed by a central bar that shows the results for the analyzed user, the vertical strong black line signs the mean achieved by the user and the small horizontal black line represents the average rating of the population. The graph also has a shadow area (the central region) that presents the standard deviation of the population. Figure 3 presents a bullet graph for the User A. Analyzing this graph, we can perceive that this user has good qualification, since the dark vertical central line (representing the individual User

A qualification) traverses the central horizontal line (representing de average qualification for all users) and also leaves behind the standard deviation, which is between -1 and 1, in this particular graph.

Figure 3: Bullet graph showing information about 'User A'.

The bar graph (Figure 4) is used to compare values, and in our case it is used to show five vertical bars representing the amount of evaluations the user receive on each qualification level ('very bad', 'bad', 'neutral', 'good', and 'very good'). Figure 4 shows the amount of evaluations that another user (User B) received on each category. Users with well evaluated documents will have the bars on the right higher than the on the left.

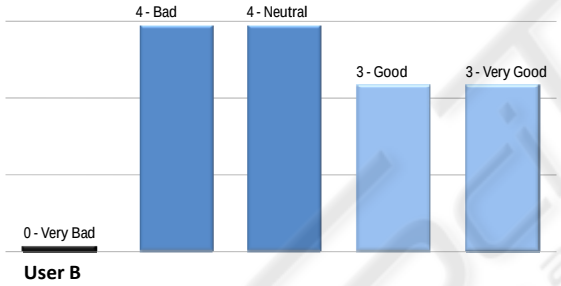

Figure 4: Bar graph used to show the amount of evaluations given for User B, on each category.

5 EVALUATION

To evaluate the proposed model, we have designed an experiment in which users were invited to evaluate the documents of other users. Data about their interactions and evaluations were collected and analyzed using EQ1 and EQ2. We were searching for variations in qualification rankings that confirm EQ1 effectiveness.

The experiment was composed by 10 pseudo-authors (users A to J) generating news about sports. To abbreviate the process, the texts were extracted from two main on-line Brazilian sports news services, O Globo and UOL , and their contents were related to nine soccer teams (teams T1 to T9) as if they were written by the ten writers. The central idea of the experiment was the evaluation of the ranking procedure not the writers' quality. After the document generation phase, each real user would focus on the evaluation of other user's documents. There were thirty documents in the total, some concerning local teams, from the same region of the users, and others involving teams from other regions of the country. Three documents were associated to each user, in the following manner: as there are three teams in the users' region (T1 to T3), users A, B and C received one team each; user D received one document from each team; moreover, the remaining users received documents from the other teams in a random fashion. We must state that T1 and T2 are from the same city and have large rivalry, T3 is neutral and the other teams are from different and distant regions. To have a homogeneous population, we have chosen the most part of the participating users to be supporters of T1. The distribution of documents concentrates documents from T1 in user A and from T2 in user B. If EQ1 is a method that overweight qualification of consensual users, the distribution that we use on the experiment will create a very qualified user (A) and a weak qualified user (B).

After the experiments, we confirm that most users have the team T1 as their favourite team and that is why User A received better evaluations, since his documents were from this team, and user B is negatively evaluated, since his documents are from team T2. The Figures 6 and 7 show the resulting normalized users´ evaluations, using, respectively, EQ1 and EQ2. The first three positions (bottom to up) are the same, but user B changes from the last position (using EQ1) to the seventh position (using EQ2). The graphs presented in these figures demonstrate that EQ1 privileges the consensus and the evaluations given by the more qualified users.

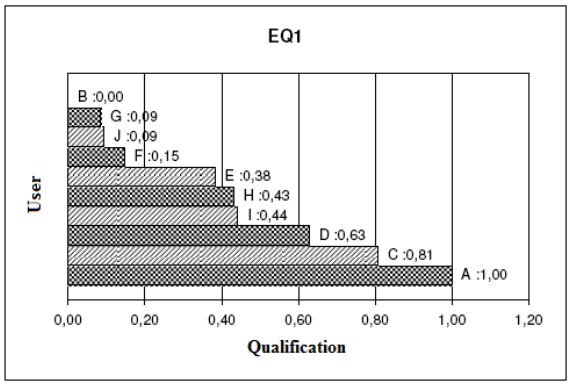

Figure 6: Users´ qualification using EQ1 (normalized).

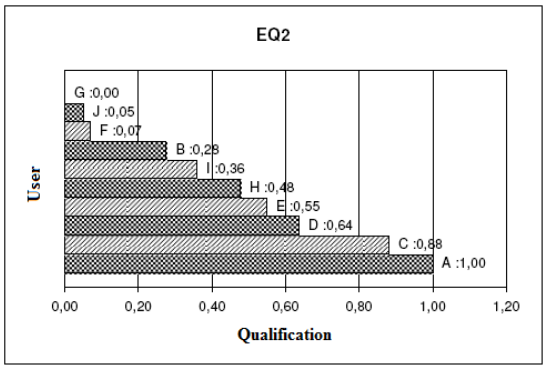

Figure 7: Users´ qualification using EQ2.

6 CONCLUSIONS

The open reviewing of documents is an open issue. A huge effort is being developed for the implementation of open access libraries but the quality assessment of this production needs to be assured. An interesting possibility is the open reviewing process. A simple alternative is the one proposed by Wikipedia, where all modifications are logged and a comparison among versions may be performed by user request. The main problem, in this case, is the absence of a clear acknowledgement of the reviewer's competence and trustfulness. A more recent proposal is the Google Knol service, in this case the authors and rewires must be identified and the revisions verified by the author. Our model and prototype offer a complete alternative to open publication and open reviewing of Web publications. With the social competence assessment of the participants, it is possible to develop a fair and independent papers quality evaluation.

The dynamic qualification mechanism present in this paper is an alternative to the generation of truth (confidence) among users of a Wiki system. It also addresses an interesting extension to the MediaWiki system, and users can edit, comment and review the documents created by other users, giving more transparency to the scientific knowledge production process.

The choice of the MediaWiki environment was appropriated, since it offers full Wiki functionality, including user and document management, version control, concurrence and consistency control, minimizing our development cycle. Besides that, it has interesting extension mechanisms that were used to carry out our qualification method. Finally, the MediaWiki environment is already known by many users, which minimizes the impact usually involved with the adoption of a new system.

The system has also other interesting applications, such as supporting collaborative work in graduation courses. In the case, students could use the environment to publish their works and to contribute in their colleagues documents. More qualified users should act as reviewers, giving more specific contributions and evaluations. Another interesting open possibility consists of employing the sys-tem as a submission and reviewing system for a scientific conference or journal, in order to analyze the differences between the traditional process and the proposed one.

ACKNOWLEDGEMENTS

This work was partially supported by research grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior (CAPES), Brazil.

REFERENCES

Butler, D., 2008. Free journal-ranking tool enters citation market. In *Nature*, 451, 3, 6 p.

Few, S., 2006. *Information Dashboard Design: the effective visual communications of data.* O'Reilly Media, Inc.

Giles, J., 2006 Internet encyclopaedias go head to head. In *Nature*, 438, 7070, 900-901.

Hu, N., Pavlou, P.A. and Zhang, J., 2006. Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of online word-of-mouth communication. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, ACM Press, 324-330.

Jarvenpaa, S.L., Tractinsky, N. and Vitale, M. Consumer trunst in an internet store. Inf. Technol. and Management, 1, 1-2, 45-71.

Millard, D.E. and Ross, M., 2006. Web 2.0: hypertext by any other name?. In *Proceedings of the Seventeenth Conference on Hyper-text and Hypermedia, HYPERTEXT'06*, ACM Press, 27-30.

Oliveira, J.P.M., 2005. Uma proposta para editoração, indexação e busca de documentos científicos em um processo de avaliação aberta (in portuguese). In *Workshop em bibliotecas digitais, WDL 2005; in conjunction with SBBD/SBES 2005*, 30-39.

Resnick, P., Kuwabara, K., Zeckhauser, R. and Friedman, E., 2005. Reputation systems. In *Commun. ACM*, 43, 12, 45-48.

Doucer, John R., 2002. The Sybil Atack. Peer-to-Peer Systems. In *LNCC 2429*. 251 - 260