

WRS: The Wikipedia Recommender System

Thomas Lefèvre, Christian Damsgaard Jensen, and Thomas Rune Korsgaard

Informatics and Mathematical Modelling
Technical University of Denmark
Christian.Jensen@imm.dtu.dk

Abstract. In 2005, the Wikipedia became the most popular reference website on the Internet and it has continued to grow in size and popularity ever since. With the increasing reliance on the Wikipedia comes issues of the credibility and provenance of content. In order to address these issues, we have developed a Recommender System for the Wikipedia, which allows the users of the Wikipedia to rate articles in order to guide other users about the quality of articles. This rating system provides both an incentive for authors to improve articles and a quantifiable measure of the perceived quality of articles.

1 Introduction

The Wikipedia is an online encyclopedia that is collaboratively edited by users on the Internet. The Wikipedia's philosophy is that anyone who wishes to share their knowledge about a subject can edit the article on that subject. The process through which Wikipedia content is added and modified is largely unregulated, which has raised concerns about the credibility of the Wikipedia [1,2] and there are plenty of examples of erroneous information that has propagated through the Wikipedia [3,4]. It is, however, apparent that the Wikipedia has gained the trust of the Internet population, despite the fact that there is nothing inherently trustworthy about the Wikipedia. It is therefore important to provide Wikipedia users with a simple and intuitive way to assess the trustworthiness of the content they are reading. We have therefore developed a recommender system, which provides users with an assessment of the quality of Wikipedia articles based on the feedback from other users who read the same article.

In order to preserve backward compatibility with the existing architecture, the Wikipedia Recommender System (WRS) is implemented as a proxy between the Wikipedia site and the user's browser. We have based the WRS on an extensible web-proxy technology called Scone [8]. The Scone proxy intercepts HTML documents, among other things, tokenizes them and allows user developed plugins to manipulate these tokenised documents before they are handed over to the user's browser. The WRS is implemented as a plugin to Scone, which makes it possible to intercept Wikipedia requests and collect recommendations from other users about the article and present an aggregated value that represents these recommendations to the user through the WRS user interface that is injected into the browser. The user can then use this interface to read estimated ratings as well as rate articles himself. The general architecture of the WRS is shown in Figure 1.

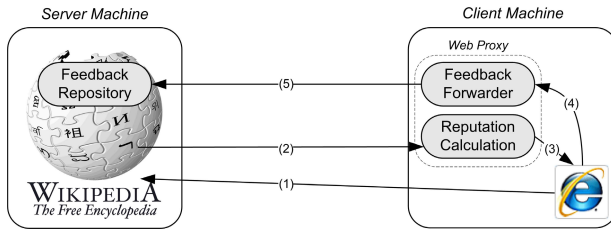


Fig. 1. Architecture of the Wikipedia Recommender System

When a user requests a Wikipedia article (1), the Reputation Calculation module extracts the rating data from Wikipedia (2). It then uses this data to calculate an aggregated rating of the article which is passed to the user interface embedded into the web page (3). By using the interface (4), the user can rate the article and the Feedback Forwarder will compile the necessary data, and upload it to the appropriate location on Wikipedia (5). The Feedback Repository, which stores the recommendations from WRS users and the WRS client proxy are described in greater details in the following.

2 Feedback Repository

A Recommender System needs to store all the recommendations in a place that is accessible at all times, by all users. The storage should also require the least amount of work to setup, maintain and operate compared to its target environment. Since the Wikipedia allows user editing, it is possible to use this mechanism to store meta-data, such as recommendations, inside the Wikipedia articles, e.g., as HTML comments that will not be rendered by the browser. This makes the WRS compatible with all Wikis based on the MediaWiki software, not just the Wikipedia.

Storing the recommendations inside the existing Wikipedia means that implementation of the WRS required no modifications to the MediaWiki software and it requires no additional maintenance or operation procedures on the Wikipedia servers. Moreover, recommendations are available to all users who have access to the Wikipedia, because they are stored on the same servers. The first part of the recommendation contains the rating of the article which is readable by humans, but the second part is a BASE64 encoded digital signature by the user who submitted the rating.

The meta-data associated with the WRS can be stored 'invisibly' in articles via the wikitext markup language, which makes it easy to store the meta-data anywhere on the Wikipedia. However, storing the meta-data inside Wikipedia articles causes problems, because it results in longer load times for all users, not only the users of the WRS. This is not considered good behavior by ordinary Wikipedia users, so another solution is needed. Fortunately, registered users are given their own user-page, which is part of the Wikipedia and can be edited by everybody. We have therefore created the user `recommendations` which stores all the recommendations for all users of the WRS. We have created sub-pages in this user's user-page, which mirrors the rated articles from

the main Wikipedia, but these sub-pages only contain the recommendations. As an example this naming convention, consider the article http://en.wikipedia.org/wiki/Bass_Strait, which becomes http://en.wikipedia.org/wiki/User:Recommendations/Bass_Strait

3 WRS Client Proxy

The client proxy consists of two components, the Feedback Forwarder and the Reputation Calculator. The Feedback Forwarder prepares the user recommendation and signs it, then prepares the meta-data and uploads it in the correct location in the Wikipedia. The Reputation Calculator extracts and calculates on this meta-data, which will be explained in further detail in the following.

When the user requests a Wikipedia article, the Reputation Calculator accesses the Feedback Repository and extracts the recommendations. It then verifies the recommendations, making sure that they are valid and properly signed. The Reputation Calculator then aggregates the recommendations for the article, based on the ratings in the recommendations and the requesting user’s trust in the recommenders. The aggregated rating is then presented to the user through the WRS user interface shown in Figure 2. The WRS user interface shows the aggregated quality indication at the top (the value “3.0” in Figure 2. Below this line, the user is asked whether he agrees with this value (“Was This Information Useful to You?”) and the user is asked to provide his own rating at the bottom of the window. The usefulness indicator and the user’s own rating are used to calculate the user’s trust in the recommenders.

The WRS Client Proxy maintains a database of all users who have rated articles accessed by the user. When the user then rates an article on the Wikipedia, the Reputation Calculator compares this rating with the recommendations received from other users. Recommendations that contain ratings close to the user’s own rating are considered positive and result in a higher trust in these recommenders. Similarly, recommendations with ratings that are very different from the user’s own rating will be considered negative and the user’s trust in those recommenders will decrease. The user’s trust in



Fig. 2. The WRS user interface

the recommender is used together with the rating in the recommendation to calculate the aggregated article rating. If the user's trust in a recommender is high, the rating will contribute more to the aggregate rating. This means that the value of the WRS to an individual user depends on the user's rating of articles, which build trust in the other users as recommenders, but at the same time contribute recommendations to the WRS community.

4 Discussion

The WRS addresses the important problem of providing Wikipedia users with essential information about the credibility of Wikipedia articles. The developed prototype of the WRS integrates seamlessly with the existing Wikipedia infrastructure in a way the benefits users of the WRS without penalising Wikipedia users who do not use the WRS.

References

1. McHenry, R.: The Faith-Based Encyclopedia. TCSDaily.com (November 15, 2004) (visited March 14, 2009)
2. Denning, P., Horning, J., Parnas, D., Weinstein, L.: Wikipedia Risks, in *Inside Risks* 186. Communications of the ACM 48(12) (2005)
3. Seigenthaler, J.: A false Wikipedia 'biography'. USA TODAY (November 29, 2005)
4. Orłowski, A.: Avoid Wikipedia, warns Wikipedia chief, It can seriously damage your grades. The Register (June 15, 2006)
5. Korsgaard, T.R.: Improving Trust in the Wikipedia. M.Sc. Thesis, Department of Informatics & Mathematical Modelling, Technical University of Denmark (2007)
6. Korsgaard, T.R., Jensen, C.D.: Reengineering the Wikipedia for Reputation. In: Proceedings of the 4th International Workshop on Security and Trust Management (STM 2008), Trondheim, Norway, pp. 71–84 (2008)
7. Jensen, C.D., Korsgaard, T.R.: Dynamics of Trust Evolution: Auto-configuration of dispositional trust dynamics. In: Proceedings of the International Conference on Security and Cryptography (SECRYPT 2008), Porto, Portugal (2008)
8. Scone website, <http://www.scone.de> (visited March 15, 2009)