

# INTEGRATING SOCIAL NETWORKS WITH EDUCATIONAL PLATFORMS

José Carpio Cañada

*Departamento de Tecnologías de la Información*

*Universidad de Huelva*

*Escuela Politécnica Superior*

*jose.carpio@dti.uhu.es*

## ABSTRACT

Social networks are widely used today. Millions of people are connected with web sites like Facebook, Tuenti or Twitter. These websites let their users to easily send messages, share pictures or write anything you want to do today. Students at any level are social networks users too. With this project we want to provide a way to integrate social network users with educational platforms like Moodle. In this paper we present a recommendation utility as example for teachers who want to create their own Facebook applications and then we present a way to connect this application with Moodle.

## KEYWORDS

educational learning, recommendations, blogs, social networks

## 1. INTRODUCTION

People use social networks every day. How we can use that to improve the relationship with our students? Trying to give an answer to this question we implemented a Facebook application to make recommendations as example for the teachers interested in use the power of social networks. We try to recommend websites to the social networks users. A Blog (a contraction of the term "Web log") is a web site, usually maintained by an individual with dated entries of commentary, descriptions of events, or other material such as graphics or video. There have been a number of studies (Technorati, 2008; P. Verna, 2008) aimed at understanding the size of the Blogosphere (the collective community of all blogs). The exponential rise in the number of blogs from thousands in the late 1990s to ten hundred in 2005 (G. Mishne & M. de Rijke, 2006) and hundred of millions today (Technorati, 2008) has created a need for effective access. Browsing among so much information is no easy task, even more when blogs are often updated several times a day, and generic search engines (Google, Yahoo, Altavista, etc) do not update their index so frequently. Another drawback of current search engines is that they perform only string searches; they not include any semantic information. For example, a searching "Granada" (which, in Spanish, is the city of Granada, in Spain, several other Granada cities all over the world, and a pomegranate, and a grenade) will give a list of web pages about tourist information, explosives or fruits. That means, is difficult to find which weblogs is talking about a particular topic one is interested in. Disadvantages like that, motivate study of new technologies to generate better results in web knowledge extraction (web mining), specifically from weblogs.

This paper is based on applying automatic association rules extraction algorithm. That data mining algorithm, help us to find weblogs not evident associations. We use 50 thousands links from *delicious.com* tagged as *blogs* to construct the association rules that let us to recommend to users similar blogs to read. By using weblog tagging, instead of hyperlinks, our algorithm extracts information from user themselves, who make associations by choosing a few weblogs out of the huge amount of them available. This also means that rules extracted should be expressed as "Readers who chose this (these) weblog(s) also chose this other". It is often the case that weblogs that appear in the same rule, but that does not happen always, for several possible reason: not all weblogs use hyperlinks, and hyperlinks are unidirectional, so hyperlinked need not be aware

of that; even more so, those who read a blog often do not know other blogs that link to that one, even as that fact implies that they are related.

This work is organized as follows: after a brief exposition of state of the art, a priori algorithm is briefly explained in section 3; further on, a formal problem description and data mining phases is given. Finally, results are presented along with conclusion details and future works.



## 2. STATE OF THE ART

Association rules are widely used, and can be considered a mainstream procedure for recommendations. Amazon.com, through extensive analysis of clickstreams (visits through their pages), buying habits, as well as other data available for them (wish lists, lists from listmania) has one of the most extensive databases of recommendations in the industry, so that readers recommendations are quite accurate. Many other e-commerce sites use same kind of algorithms, but, for the time being, website recommendation is much more elusive.

Main problem is that, even as Amazon's items database is huge, websites number, and even of weblogs, is almost as big. Besides, Amazon features static item with known characteristics, while websites evolve, offer different facets, and change completely, even more so in the case of weblogs, which can be said to be created collaboratively by the author and its readers through comments.

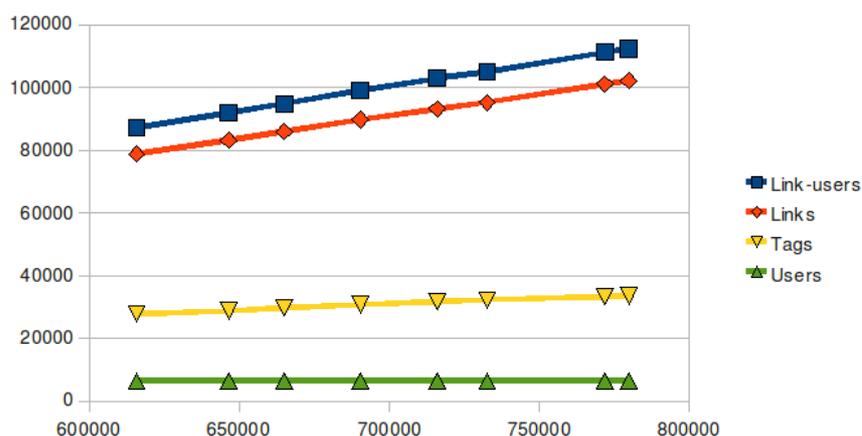
Weblogs have several mechanisms that allow readers to discover others: one of them is simply hyperlinks; if weblogs author finds something interesting elsewhere, he refers to it by adding comments with hyperlink in its own weblog. Another mechanism is called trackback: a weblog can inform another that he is linking to it, or talking about a similar topic, by calling a function in a particular way; trackbacks appear alongside commentaries in the post they refer to.

These mechanisms, however, are often not enough. Every weblog works in its own way: some of them do not link at all (think about literary weblogs, for instance, or pure personal diaries), others link more often to news sites than to other weblogs (mainly collective weblogs), and others, simply, are not aware of the huge amount of sites available out there.

So far, we have proposed using community detection mechanisms (J. J. Merelo-Guervós, et al., 2004). that help authors and readers to know which community do they belong, and, by linking, what kind of communities emerge; however, this only tells a part of the story. By using survey result, and applying rule extraction algorithm, we can find new associations that could not have been discovered by analyzing links.

### 3. RESULTS

We have one basket per user: A total amount of 6500 baskets to make the experiments with approximately 10000 URLS, 38. The above section explain how we obtained the information and how many tagged links we worked with.



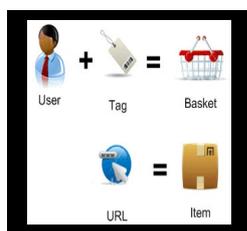
From our point of view, a blog site correspond with one URL tagged for somebody with one of the follows tags blog, blogs, weblog, weblogs, bitacora or bitacoras. Obviously, that is not always true. Somebody can tag a url using one of this tags to a web site that not correspond exactly with a blog. When we create the rules with a certain support, the rules with under a minimum support will not be taken into account. We will said that a Blog site is what a lot of people tag as blog.

The study has been done on a poll taken among weblog readers and authors, who answered, among others, the following question: ¿Which are the last 3 weblogs you have read?

Data was stored in XML file; after extraction, it had to be cleansed. Some cases, weblog titles or names were used instead of URLs; and even if addresses were used, they sometimes missed the final slash (/) or the initial http://www. At the end of this preprocessing phase, we obtained baskets with 3 URLs, with each weblog represented by a single URL. Apriori algorithm described above was then applied to this set of items, and results analyzed.

Rules were also analyzed to find which was the minimum confidence and support that guaranteed reasonable results. Since even the most popular site (MiniD) was voted only by a fraction of the pollsters, a

very small support should be expected; confidence went up with the minimum support, and finally, a level of confidence of around 50%, for a support of around 0.2%, was found reasonable, as is shown in figure 1.



**Figure 1: Association rules Users, tags and urls**

Obviously, the number of rules found changed with confidence level and support chosen, as is shown in figure 2; there were very few rules with a support over 20%, which is to be expected, due to the dispersion of the readership. The confidence of the rules which was finally used in this study was actually quite low, as is shown in figure 3, but due to the huge number of weblogs available, and the number of weblogs that were mentioned in this study (around a few thousands), the fact that two weblogs are associated even in a few polls is quite relevant.

After a detailed analysis of results using our knowledge about the blogosphere and of many of the weblogs used in this study, we have been able to find the relationship among weblogs appearing in the association rules. For example, the following rule:

`bandaancha.st ==> barrapunto.com (0.692308, 9)`

with a confidence factor of 0.69 and support of 9 rules represents two collective weblogs with a big audience (in fact, Barrapunto, at <http://barrapunto.com>, is possibly the most popular blog of the Spanish blogosphere, as was also found in survey results shown in [http://tintachina.com/archivo/los\\_weblogs\\_mas\\_leidos.php](http://tintachina.com/archivo/los_weblogs_mas_leidos.php)) and with a strongly technical, and more weakly political, background. In this case, there are several links between both sites, and the result is not altogether unexpected. However, in the following example:

For this kind of recommendations expert advice would be needed, since they are not so obvious as the previous one. In fact, paper's authors, who are heavy participants in Barrapunto, did not even know about the first one. Using this association rules we can recommend similar weblogs to readers of a given site, without human intervention.

Several groups of tests have been done, changing minimum support and minimum confidence. Minimum support has been reduced until 0,001 in order to obtain 96 rules with minimum confidence of 0,6. For a minimum support of 0,1 no rule is obtained. With a minimum support of 0.0006, 5800 rules were obtained.

From these results we would like to remark the values obtained for rules support. It is very low, but we think they are relevant enough. We think there are two main reasons. The first one is the wide variety of “products” in our “supermarket” which, following example proposed in (Kleinberg, 1999), is very big. Our articles are URLs pointing to weblogs. We have counted more than 2160 articles in this example and we are

working with a database of 1473 baskets. We think this is the main reason for a so low min support: great diversity of articles and low number of baskets.



## 2. CONCLUSION

This paper presents a blog recommendation system integrated with Facebook platform as example for teachers who want to use the social network power to improve the learning experience of their students. At this point we have not information of teachers and students feedback. Our next work will try to obtain user feedback to evaluate the complete experience and will know what kind of applications will give better benefits to the people involved in an educational process.

## ACKNOWLEDGEMENT

We are grateful to the Teaching Innovation Service of the University of Huelva that promoted this project.

## REFERENCES

- R. Kumar, et al. (2003). 'On the bursty evolution of blogspace'. In *WWW '03: Proceedings of the twelfth international conference on World Wide Web*, pp. 568-576. ACM Press.
- Technorati (2008). 'State of the Blogosphere / 2008'. [Online]. URL: <http://technorati.com/blogging/state-of-the-blogosphere/>
- Blogpulse (2009). 'Blogpulse stats'. URL: <http://www.blogpulse.com/index.html>.
- P. Verna (2008). 'The Blogosphere: A Mass Movement from Grass Roots'. URL : [http://www.emarketer.com/Report.aspx?code=emarketer\\_2000494](http://www.emarketer.com/Report.aspx?code=emarketer_2000494)
- G. Mishne & M. de Rijke (2006). 'A Study of Blog Search'. In *Advances in Information Retrieval*, vol. Volume 3936/2006, pp. 289-301. Springer Berlin / Heidelberg.
- J. J. Merelo-Guervós, et al. (2004). 'Clustering web-based communities using self-organizing maps'. *WBC04, Proceedings of the IADIS International Conference* pp. 158-165.