TAGGING SOCIETY THROUGH THE WEB AND CRAWLING BIG DATA FOR SOCIAL USES: METHOD AND IMPLICATIONS

Juan D. Borrero
Universidad de Huelva
Estrella Gualda
Universidad de Huelva

Tags, keywords freely chosen by users for annotating resources, offer a new way for organizing and retrieving web resources that closely reflect the users' interests and also automatically generate folksonomies. Social tagging systems have gained increasing popularity as a method of annotating and categorizing a wide range of different web resources. They also attract researchers in social sciences because they offer a huge quantity of user-generated annotations that reveal the interests of millions of people. The growing number of scientific publications concerning this issue on one hand, and the development of real social tagging systems on the other, such as social networks (Twitter), social bookmarking applications (Delicious), sharing systems (Flickr), and also in the e-commerce field (Amazon), confirm this tendence. The social aspects of these systems derive from the fact that the resources are tagged by the community, a feature known as collaborative tagging, which provides important metadata for investigators and others practitioners.

To date, the study using digital trace data methods continues to lack a theoretical framework, particularly in social science research. This paper presents a methodology to use big data from Social media in social research. At the same time, it applies this method to extract data from a particular social bookmarking site (*Delicious*) and shows the type of results that this type of analysis can offer to social scientists. Using data crawled from a large social tagging system, can have an important impact in the discovering of latent patterns, which are basics in order to provide effective recommendations to different actors.

Thus, identifying the mechanisms that underlie social tagging is important for several reasons: the socially tagged network obscures the local mechanisms that generate the network's structure; determine the importance of the users' themes of interest; a community's interest in a certain category enables us to discover other user interests, and this information is useful for making recommendations. Introducing folksonomies as basis for recommendations means that the usual binary relation between users and resources changes into a ternary relation between users, resources, and tags, more complex to manage.

This study also follows the structural analysis approach. Specifically, it has retrieved conceptual and methodological tools from the work of Marres and Rogers (2008). These authors understand the Web as a forum for relevant organizational interaction and for issue publicization. The analysis of such types of participation involves identifying agents that make up an "issue network" as well as examining the 'heat' of the affair that they have in common (Marres and Rogers 2008). In this article, we have taken a similar approach. More specifically, we have used Social Network Analysis for reviewing the issue network related to agriculture's globalization.

In our study, we traced a concrete key expression ('agriculture's globalization') and we considered that hyperlink structure we have constructed could be considered an 'issue network'. The results discussed in the paper come from the analysis of a sample of 851 users, 526 URLs and 1,700 tags from the *Delicious* classification system on the subject of globalization. We retrieved and analysed the composition of this issue network structure, its hyperlinking, key nodes and folksonomyes. More precisely, we trace the «agriculture's globalisation» network on Delicious community, analyse its structure and consider the collective representation of the connections between related tags. When aggregating all tags from a community, a collective representation of the connections between related tags and their strengths of association will emerge. These associations are typically visualized by tag clouds, in which different font sizes represent the strength of association of tags to a related tag or a resource. Tag clouds externalize the community's associations between tags and the strengths of associations. In this way, social tags are able to provide visual representations of the conceptual structure of a domain, which is built upon the knowledge of individuals who belong to a large Web community. The results obtained show main fields of very general interests and reveal the frames (economics, environment) associated with the agriculture's globalisation issue.

Finally, we also observed the phenomenon by which a small number of sites on the web attract a disproportionately high number of ties. These sites become web centres of gravity since the more ties the sites have attracted in the past, the more ties they are likely to capture in the future, and, because users will inevitably end up visiting the most prominent sites. Our social network analysis has identified that online communication companies have a competitive advantage in attracting links and user attention.

Hence, through the analysis main users and websites around globalization issue in Delicious emerged, and also the most important tags that were used by user to describe globalization were discovered. A central problem closely related to the opinion network is how to extract the hidden information. The exponential growth of the Internet confronts people with an information overload: They face too much data and sources able to find out those most relevant for him. Thus, our analysis provides an understanding of the social bookmarking systems that was not previously available. This research has attempted to convert several social indicators into socio-semantic analogs. It is also one of the first to apply information on the frame components in an attempt to establish the inherent ties within these networks. The implications of these methodology and findings for further research are discussed.