

User-Induced Hyperlinks in Collaborative Tagging Systems

Ching-man Au Yeung^{*}, Nicholas Gibbins, Nigel Shadbolt
Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science
University of Southampton, Southampton, SO17 1BJ, United Kingdom
{cmay06r,nmg,nrs}@ecs.soton.ac.uk

ABSTRACT

This poster describes our study of user-induced hyperlinks found in folksonomies using association rule mining algorithms. We compare these induced hyperlinks with existing hyperlinks found among the same set of documents. Our experiments show that user-induced hyperlinks are very different in nature from existing hyperlinks. They tend to connect documents from different domains and documents that are highly related to each other as judged by the similarity between the tags assigned to them. Our study suggests that user-induced hyperlinks in collaborative tagging systems are very valuable in complementing the existing link structure on the Web to help users navigate between highly related Web documents.

Categories and Subject Descriptors

H.5.3 [Group and Organization Interfaces]: Web-based interaction; H.5.4 [Hypertext/Hypermedia]: User issues

General Terms

Experimentation, Human Factors

Keywords

Collaborative tagging, Folksonomies, Hyperlinks

1. DESCRIPTION OF THE POSTER

Hyperlinks are probably the most important elements on the Web, as their existence is the reason why the Web is a web: they allow users to jump from one hypertext document to another one, making navigation through the Web possible. Hyperlinks are in general embedded in documents. This

^{*}Ching-man Au Yeung is a PhD candidate at the University of Southampton under the supervision of Nigel Shadbolt and Nicholas Gibbins. Tel. +44 (0)23 8059 7684. Email. cmay06r@ecs.soton.ac.uk. Work described in this poster is an individual project.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

means that very often only the author of a hypertext document can decide to which other documents this one can link to. While the authors' decision may be necessary when hyperlinks are created for navigation within a particular Web site, it is conceivable that such author-created hyperlinks may not be adequate from the perspective of the readers. For example, the author may not be aware of some Web documents that are highly relevant, and thus fail to create hyperlinks to them. It may also be because that some highly relevant documents are created by rivals of the author and they may be competing to attract more readers. In this case a hyperlink between these documents are not likely to exist.

The recent years saw a surge of user-contributed content on the Web. In particular, descriptive keywords, commonly known as tags, are extensively used by users to organise and share online resources [3]. In social bookmarking sites such as Delicious, Web users maintain a collection of documents (Web pages identified by their URLs) that are categorised by their chosen tags. In general, two documents that are in the collection of the same group of users or that are assigned a similar set of tags can be considered as related to each other. From the perspective of hyperlinking, two such documents should be linked to each other such that Web users accessing the first one can be directed to the second one, and vice versa. This lead to some interesting research questions regarding collaborative tagging: (1) Are documents interested by the same group of users or assigned similar tags linked to one and other? (2) How different are the relations between documents induced by the collective user behaviour (user-induced hyperlinks) and the existing hyperlinks between the same set of documents?

This poster describes our study of using association rule mining [1] to discover user-induced hyperlinks in collaborative tagging systems and of comparing these user-induced hyperlinks with existing hyperlinks. We conduct our experiments using a large data set collected from Delicious, comprising of over 1 million users who have assigned over 800,000 tags to over 100,000 documents. The experiments reveal several interesting facts. Firstly, the majority of user-induced hyperlinks are not existing hyperlinks (an overlap of about only 10%). Secondly, very few of user-induced hyperlinks connect documents from the same domain, while a much larger proportion of existing hyperlinks do. This suggests that user-induced hyperlinks are more likely to direct users to new information. This is further confirmed by the finding that documents connected by user-induced hyperlinks are much more similar in the topics they address than those connected by existing hyperlinks. This maybe

due to the fact that many existing hyperlinks are for navigation purposes. We find that even many user-induced hyperlinks that connect documents from the same domain are not existing hyperlinks. This shows that clearly there is a gap between the perspective of the authors and that of the readers of the Web.

Implicit links having a nature similar to those mentioned in this poster are discussed in the literature. For example, Xue et al. [5] propose to mine query logs to discover implicit links between pages within a Web site. Shen et al. [4] propose to generate implicit links from search engine query logs for the purpose of classifying Web documents. However, no prior studies focus on discovering implicit links in folksonomies, which contain rich information about user preferences, user interests and document similarity. There are also very few comparative studies that look into the differences between implicit or induced hyperlinks and existing hyperlinks. Our work thus makes some important contribution in understanding the gap between authors and readers on the Web.

It would be worthwhile to consider something like an open hypermedia system [2] backed by collaborative tagging systems. Links between different Web documents are induced from the collective behaviour of the users, and are maintained externally with respect to the documents involved. These links represent the perspective of the users on how documents on the Web should be linked to each other. Given the tags assigned on these documents, it is also possible to give semantics to these induced links. For example, when we apply the technique on URLs all assigned the tag `cooking`, induced links between these URLs can be described by the tag, giving the users an idea of why these documents should be linked. In the future, we plan to compare user-induced links in collaborative tagging systems with implicit links discovered in Web logs or query logs.

2. REFERENCES

- [1] R. Agrawal, T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. In *SIGMOD '93: Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, pages 207–216, New York, NY, USA, 1993. ACM.
- [2] A. M. Fountain, W. Hall, I. Heath, and H. C. Davis. *MICROCOSM: an open model for hypermedia with dynamic linking*, pages 298–311. Cambridge University Press, New York, NY, USA, 1992.
- [3] S. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, April 2006.
- [4] D. Shen, J.-T. Sun, Q. Yang, and Z. Chen. A comparison of implicit and explicit links for web page classification. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 643–650, New York, NY, USA, 2006. ACM.
- [5] G.-R. Xue, H.-J. Zeng, Z. Chen, W.-Y. Ma, H.-J. Zhang, and C.-J. Lu. Implicit link analysis for small web search. In *SIGIR '03: Proceedings of the 26th Annual International ACM SIGIR conference*, pages 56–63, New York, NY, USA, 2003. ACM.