

# Finding It on Google, Finding It on del.icio.us.

Jacek Gwizdka and Michael Cole

Department of Library and Information Science,  
School of Communication, Information and Library Studies, Rutgers University  
4 Huntington St, New Brunswick, NJ 08901, USA  
ecd12007@gwizdka.com, mcole@scils.rutgers.edu

**Abstract.** We consider search engines and collaborative tagging systems from the perspective of resource discovery and re-finding on the Web. We performed repeated searches over nine-months on Google and del.icio.us for web pages related to three topics selected to have different dynamic characteristics. The results show differences in the resources they provide to the searcher. The resources tagged on del.icio.us differ strongly from the top results returned by Google. The results also suggest the changes in the most recently tagged web pages may be associated with the level of activity in user communities and, indirectly, with external events.

**Keywords:** Folksonomy, Collaborative tagging, Resource discovery, Search.

## 1 Introduction and Motivation

Collaborative tagging of electronic resources has been described from the perspective of social navigation [4], distributed cognition [10], semiotic dynamics [1], and knowledge sharing and resource discovery [7]. In this project, we take the latter perspective and ask if delicious provides an additional dimension to information search resources on the Web. We pose our question in the context of communities as identified by users engaging in topical tagging activity. The users select Web resources and, by tagging them, provide an additional layer of information. We are interested in finding out if the selected Web resources are different from search engine results. We are also interested in examining changes in the kind and the level of tagging activity over long time periods.

## 2 Methodology

The Google and Delicious APIs were used make daily requests between 25.06.2006 and 1.01.2007, and then again in March 2007. The successful requests sent to Delicious and Google within a 6-hour window on the same day were paired. The Google results captured the top 19 documents. The Delicious results returned the tags related to the request and the most recent 27-30 URLs and the tags produced by users for those URLs. Five kinds of searches were conducted 1) `world cup`, 2) the phrase

“world cup” on Google and `worldcup` on Delicious, 3) “web design” on Google and `webdesign` on Delicious, 4) `social tagging folksonomy`, and 5) `socialtagging folksonomy` (“social tagging” folksonomy` on Google).

### 3 Results and Discussion

Overlap was calculated by counting exact matches of normalized URLs and for domains where any 'www.' prefix was stripped. Duplicates in a given day's data were removed for overlap and rank calculation. Table 1 compares the Delicious results with the top two pages (n=19) of Google results for submission of the same query over all days.

**Table 1.** Overlap of Delicious and Google search results

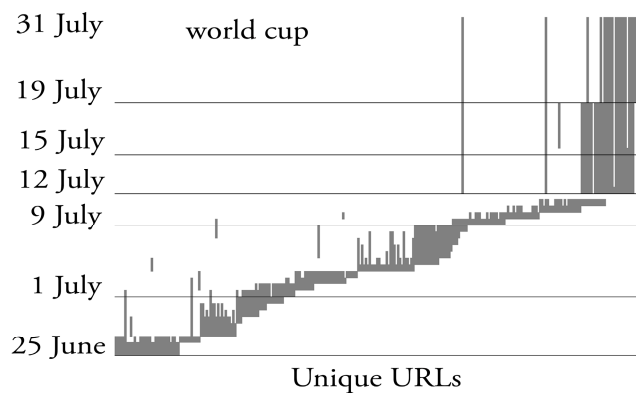
Query	worldcup	world cup	webdesign	social tagging folksonomy	socialtagging folksonomy
Total URLs	1402	3107	1129	1106	794
Total domains	1204	1762	1081	1081	723
URLs overlap %	1.5%	0.1%	0.6%	2.0%	0.0%
Domain overlap %	9.3%	8.3%	2.1%	8.2%	1.1%

Overlap between first page results from major search engines is low [2, 6]. Spink, et al. [9] found that 85% of the results were unique and the overlap between any pair of search engines was about 11%. They concluded that search engines appear to have different capabilities. We considered Delicious and Google as alternative search resources. Compared to Spink et al., we observed dramatically lower overlap even though overlap was calculated using only the top two pages of Google results. It seems unlikely Delicious users were tagging top ranked results retrieved from Google. This suggests users are tagging pages arrived at by other means, perhaps using very different queries to search engine, or by other Web search mechanisms altogether, for example recommendations or references in email or blogs. The mechanisms of user discovery of these pages may expose content that has not acquired (or was not designed to acquire) features that result in higher page ranks in the Google algorithm. It is also possible that taggers use search results but choose to tag only those items that are not in the top ranks, perhaps because they believe the top ranked results will remain easily found by the search engine. Future work includes considering Delicious overlap against multiple search engines.

Delicious users from the US tend to be better educated and wealthier than average [8]. Professionals or those with expertise in an area may be tagging information and resources relevant to their needs and specific interests. These documents are likely to have low ranks for search engine queries using general terms a non-expert might use in seeking relevant information on the Web.

We used ‘world cup’ to observe tagging activity on Delicious for a transient event, the 2006 FIFA World Cup. A rise in overlap was observed around the time of the conclusion of the event (July 9), when it might be expected that more of the new content produced by the event had been indexed by Google, and possibly that these new URLs had acquired sufficient link authority to be highly ranked. Figure 1 shows

the development of activity in new Web resources being tagged with “world” and “cup” while the 2006 FIFA event was taking place and shortly afterwards. It is interesting to see that many new resources were tagged daily until shortly after the conclusion of the event, when tagging activity declined and the bookmarks stabilized. Tagging activity for other communities (‘webdesign’ and ‘social tagging folksonomy’ not shown here) exhibit different dynamics, and we conclude that the level of activity may be used to characterize interaction within communities and how they are influenced by external events.



**Fig. 1.** Unique URLs (horizontal) being tagged on del.icio.us with “world” and “cup” over the period of five weeks (from June 25 to July 31,06). FIFA’2006 World Cup ended on July 9.



**Fig. 2.** Tags clouds with 64 most frequent words that co-occurred with “world” and “cup” in two month-long periods. July 2006 (FIFA World Cup in Germany) and March 2007 (Cricket world cup in Jamaica). “World” does not appear because it was on the stop-word list.

Figure 2 presents two tags clouds with 64 most frequent words that co-occurred with “world” and “cup”. The clouds are shown for two month-long periods, during which two big world cup events took place. The co-occurring tags are quite different in each case. We can certainly learn a bit about a world cup event, or events that were taking place during each period. For example, it is clear that the first is related to soccer and FIFA (Federation of International Football Associations), while the second to cricket and ICC (International Cricket Council). Since the latter was during winter, we also note “skiing” among the tags. We can see what world cup taggers found of particular interest. The French football player Zinedine Zidane’s infamous headbutt and the murder of cricket coach Bob Woolmer. We can identify countries that were involved in each event, but we cannot easily tell where each event was located. Clearly, this short analysis uses additional knowledge. Future work includes examining if tags could be used to provide machine support for information seeking tasks.

## 4 Conclusion

Considered as a search resource, Delicious results have very little overlap with the top pages of Google results, both in URLs and domains. The Delicious results appear to capture transient events and evolving developments in certain domains that are not reflected in Google’s top results. This provides evidence that Delicious offers a resource that may provide a new dimension for Web searching beyond the collaborative tagging activity. Furthermore, Delicious may provide source of information about the level and the kind of activity in different communities.

## References

1. Cattuto, C.: Semiotic dynamics in online social communities. *The European Physical Journal C - Particles and Fields* 46, 33–37 (2006)
2. Chignell, M.H., Gwizdka, J., Bodner, R.C.: Discriminating meta-search: A framework for evaluation. *Information Processing and Management* 35, 337–362 (1999)
3. Coenen, T., Kenis, D., Damme, C.V., Matthys, E.: *Knowledge Sharing over Social Networking Systems: Architecture, Usage Patterns and Their Application* (2006)
4. Dieberger, A., Dourish, P., H.K., Resnick, P., Wexelblat, A.: *Social navigation: techniques for building more usable systems*, vol. 7(6), pp. 36–45. ACM Press, New York (2000)
5. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. *Journal of Information Science* 32(2), 198–208 (2006)
6. Gordon, M., Pathak, P.: Finding information on the world wide web: The retrieval effectiveness of search engines. *Information Processing and Management* 35, 141–180 (1999)
7. Macgregor, G., McCulloch, E.: Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review* 55(5), 291–300 (2006)
8. Rainie, L.: *28% of Online Americans Have Used the Internet to Tag Content*. Pew Internet and American Life Project (2007)
9. Spink, A., Jansen, B., Blakley, C., Koshman, S.: A study of results overlap and uniqueness among major Web search engines. *IP&M* 42, 1379–1391 (2006)
10. Steels, L.: Collaborative tagging as distributed cognition. *Pragmatics & Cognition* 14(2), 287–292 (2006)