

GiveALink: A Social Recommendation System

Ben Markines
bmarkine@cs.indiana.edu

Lubomira Stoilova
lstoilov@cs.indiana.edu

Filippo Menczer
fil@indiana.edu

Department of Computer Science
School of Informatics
Indiana University
Bloomington, IN 47406, USA

ABSTRACT

GiveALink is a social bookmarking site where users may donate and manage their personal bookmark files online securely. The bookmarks are analyzed to build a new generation of intelligent information retrieval techniques to recommend, search, and personalize the Web. GiveALink does not use tags, content, or links in the submitted Web pages. Instead we present a semantic similarity measure for URLs that takes advantage both of the hierarchical structure in the bookmark files of individual users, and of collaborative filtering across users. In addition, we build a recommendation and search engine from ranking algorithms based on generality and novelty measures extracted from the similarity-induced network. Search results can be personalized using the bookmarks submitted by a user. We evaluate a subset of the proposed ranking measures by conducting a study with human subjects.

1. INTRODUCTION

Major search engines crawl the Web to populate their databases. When a user submits a query, results are generated and ranked using text similarity measures, the hyper-link structure of the Web, and click-through data from the company's servers. Social bookmarking tools on the other hand build upon the gregarious nature of individuals who establish semantic relationships by sharing URLs. This has led to an explosion of the "folksonomy" phenomenon, as witnessed by the multiplication and popularity of sites such as del.icio.us and citeulike.org.

Here we describe GiveALink, a system that goes beyond the tagging functionality of current bookmarking sites by actively exploiting collaborative filtering and the hierarchical structure of bookmark files, where present. Hierarchies display a finer, more structured representation of data in comparison to flat tagging systems. The collaborative or social aspect of GiveALink relies on aggregating information from donated bookmark files. Each bookmark file represents a person's notion of semantic similarity. Fig. 1 compares this approach with other recommendation systems.

GiveALink distributes the process of collecting data and determining similarity relations among all of its users. We use bookmark files as a convenient existing source of knowledge about what Web pages are important to people, and about the semantic structure in which they are organized.

Copyright is held by the author/owner(s).
NetSci2006, May 22–25, 2006, Bloomington, IN, USA.

Hierarchical Structure	DMOZ	GiveALink
Flat Structure	Classic Content/Link Based	Social Bookmarking
	Single or Shared Representation	Social or Collaborative Aggregation of Users

Figure 1: Two dimensions of recommendation systems: the structure of the knowledge representation space and the social aspect where each individual may contribute a shared or personal representation.

The URLs in our database originate from bookmark files donated and managed by users. We further determine similarity relationships and relevance to queries by mining the structure and attribute information contained in these files. Thus we propose a notion of similarity that is very different from the ones used by Google, Yahoo, and MSN. Our measure of similarity is not based on the content of the pages and not even on the Web link graph. Instead, it is an aggregate of the independent notions of semantic similarity contributed by different bookmark file owners.

Contributions of this work include: (1) A novel semantic similarity measure for URLs that takes advantage of both the hierarchical structure of bookmark files and collaborative filtering techniques. (2) Two ranking measures called generality and novelty, which are based on our similarity measure. (3) An algorithm for personalizing search results based on user bookmarks. (4) Data, such as our (anonymized) URL-to-URL similarity matrix, that is made freely available to the Web community in the hope that it will foster the development of novel and useful Web mining techniques.