Introducing Docear's Research Paper Recommender System

Joeran Beel
Docear
Magdeburg, Germany
beel@docear.org

Stefan Langer
Docear
Magdeburg, Germany
langer@docear.org

Marcel Genzmehr
Docear
Magdeburg, Germany
genzmehr@docear.org

Andreas Nürnberg
OvGU, FIN, DKE Group
Magdeburg, Germany
andreas.nuernberger@ovgu.de

ABSTRACT
In this demo paper we present Docear’s research paper recommender system. Docear is an academic literature suite to search, organize, and create research articles. The users’ data (papers, references, annotations, etc.) is managed in mind maps and these mind maps are utilized for the recommendations. Using content based filtering methods, Docear’s recommender achieves click through rates around 6%, in some scenarios even over 10%.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – information filtering

General Terms
Management, Design

Keywords
recommender systems, user model, mind map, mind mapping, research paper recommender system, content based filtering

1. INTRODUCTION
Literature management, i.e. searching, organizing and creating literature, is important for researchers and students. Especially the search for relevant literature is challenging due to the millions of articles and books being published every year and the fact that most search services such as the ACM Digital Library focus only on publications of selected publishers (e.g. ACM). In addition, full-texts are often not freely accessible and need to be paid unless ones university or library has a subscription for the publisher.

Our open source tool Docear (www.docear.org) supports researchers with literature management by bundling several applications that help in searching, organizing, and creating academic literature [2]. Docear has the unique feature of utilizing mind maps for information management. That means users organize their data in a tree-like data structure and not in a table or with social tags. Figure 1 shows an example of a mind map created as draft for this paper. The mind map created outlines the skeleton of this paper and the nodes (i.e. the “entities” with the words) link to articles in which the information was originally found. More information on Docear can be found in [2].

In this paper we present the research paper recommender system which we developed for Docear.

Permission to make digital or hard copies of part or all 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. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

JCDL’13, July 22–26, 2013, Indianapolis, Indiana, USA.
ACM 978-1-4503-2077-1/13/07.

2. RELATED WORK
One of the earliest research paper recommender systems was TechLens [8]. TechLens used the two most popular recommendation approaches – content based filtering (CBF) and collaborative filtering (CF). In CBF, the words of a user’s documents are taken to build a user model, and documents that contain the same words as the user model are recommended. In CF, similar users are determined by comparing how they rated items (the more often two users rated items alike, the more similar they are assumed to be). Then, items in the collection of the similar user are recommended to the other user.

3. DOCEAR’S RECOMMENDER SYSTEM
Docear has access to quite diverse, and quite a lot, information about its users. With Docear, users search for literature, they organize their literature, and they draft their own literature. That means Docear knows what a researcher is currently looking for, which articles a researcher already knows, which ones he is interested in, and which articles he is currently working. In addition, Docear allows its users to make annotations in PDF files, i.e. adding comments or highlighting passages they consider
important. Consequently, Docear does not only know which papers a researcher currently is reading but also which parts of an article he considers particularly important. This allows highly personalized recommendations.

We are experimenting with different variations of content based filtering for Docear. All variations have in common that from a user’s mind map collection a user model is build and the user model is matched with Docear’s Digital Library containing around 1.8 million research articles from various disciplines (Figure 2). The Docear desktop software synchronizes all mind maps of a user every couple of minutes with Docear’s server. When a user requests recommendations, Docear sends the request to Docear’s Digital Library. This creates a user model and returns ten recommendations which are all accessible in full-text. Some users have reservations against having their mind maps transferred to Docear’s server and having them analyzed. Therefore, Docear allows users to turn recommendation off which about 2/3 of users do (when users turn recommendations off, they still can use Docear to manage their literature).

Currently, we are experimenting with different variations of content based filtering. For instance, we are varying the number of mind maps and nodes that are analyzed for creating a user model (e.g. all mind maps the user created vs. only the last edited mind map), the user model size and the type of TF-IDF being used. Based on an evaluation with 938 users and 32,790 delivered recommendations, the average click through rate (CTR) is 6.03%. However, for specific variations such as when stop-words are removed, TF-IDF is applied instead of term frequency only, and only user requested recommendations are analyzed, click through rates increases to 10.31% (Figure 4), and in some specific scenarios even more. In comparison, a typical click through rate in advertisement is 0.5% [6]. Research paper recommender systems typically achieve CTRs of around 8% [7]. Further details on the recommendation algorithms will be published in a forthcoming paper.

Finally, we plan on extending the recommender's performance by offering more potentially relevant recommendations. When more papers are added to the digital library, there will be more potentially relevant recommendations. For instance, journals and conferences, research grants, and for students we would like to recommend university programs.