Chapter 5
A Semantic Approach for News Recommendation

Flavius Frasincar
Erasmus University Rotterdam, The Netherlands

Wouter IJntema
Erasmus University Rotterdam, The Netherlands

Frank Goossen
Erasmus University Rotterdam, The Netherlands

Frederik Hogenboom
Erasmus University Rotterdam, The Netherlands

ABSTRACT

News items play an increasingly important role in the current business decision processes. Due to the large amount of news published every day it is difficult to find the new items of one's interest. One solution to this problem is based on employing recommender systems. Traditionally, these recommenders use term extraction methods like TF-IDF combined with the cosine similarity measure. In this chapter, we explore semantic approaches for recommending news items by employing several semantic similarity measures. We have used existing semantic similarities as well as proposed new solutions for computing semantic similarities. Both traditional and semantic recommender approaches, some new, have been implemented in Athena, an extension of the Hermes news personalization framework. Based on the performed evaluation, we conclude that semantic recommender systems in general outperform traditional recommenders systems with respect to accuracy, precision, and recall, and that the new semantic recommenders have a better F-measure than existing semantic recommenders.

DOI: 10.4018/978-1-61350-038-5.ch005

INTRODUCTION

Finding the news items of interest is a critical task in many business processes. One such process is business intelligence which aims to gather, analyse, and use company-related data in order to support decision making (Luhn, 1958). While a lot of this information is represented by company internal data (e.g., product sales, costs, incomes, etc.), in the recent years, we observed a growing focus of attention for company external data whose processing is aimed to answer questions as how is the company perceived by the public? (business marketing), how are competitors reported in the media? (competitive intelligence), what are possible collaborators in other countries? (business internationalization), etc., (Saggion, Funk, Maynard, & Bontcheva, 2007) (Pang & Lee, 2008) (Castellanos, Gupta, Wang, & Dayal, 2010). News items, as rich sources of external company-related information, are increasingly exploited in business intelligence tasks.

The Web is one of the most popular platforms for distributing and consuming news items. There are several factors that contributed to this success story as for example the reduced cost for distributing and accessing news items, Web availability on a multitude of browsing platforms, world-wide information delivery and consumption, short amount of time required for news publication, etc. Unfortunately, the Web's success is also the cause of one of its most serious liabilities: the large number of daily published news items makes the process of finding the ones relevant to particular interests difficult. For business intelligence, companies are only interested in news items deemed relevant for their analytical processes, which for competitive reasons should be made available with minimal delay times.

One possible solution to deal with the news items overload problem is the use of recommender systems, which aim to propose previously unseen items, in our case news items, that are of interest to a certain user. Typically such recommenders employ a user profile and aim to recommend news items that best match this user profile. Currently, there are four types of recommender systems: content-based, collaborative filtering, semantics-based, and hybrid (Adomavicius & Tuzhilin, 2005). While the user profile is usually represented by the user's previously browsed items, the recommendation methods differ per employed recommendation method. The content-based recommenders propose items based on the lexical content of the previously viewed items, semantic recommenders use the semantic information of the earlier browsed items, collaborative filtering recommenders exploit profile similarities between different users, and hybrid recommenders are combinations of the previous recommenders.

In this chapter we focus on recommenders that use the information content in news items, be it lexical (as in content-based approaches) or semantic (as in semantics-based approaches). While content-based recommenders have previously been thoroughly investigated, it is only in the last years that researchers started to focus on semantics-based approaches for recommender systems. Also, a comprehensive study that compares the content-based recommenders with semantics-based recommenders is currently missing. Therefore one of the aims of this chapter is to produce such an investigation in the context of recommending news items. In addition, we would like to investigate multiple semantics-based approaches and compare their performances. The collaborative filtering and hybrid recommenders are considered outside the scope of this chapter.

In previous work (IJntema, Goossen, Frasincar, & Hogenboom, 2010) we have proposed a semantic recommender for news called Ranked-based Semantic Recommender (RSR). In this chapter we extend our previous work by considering not only the concepts directly related to the concepts from the user profile but also the concepts directly related to the concepts present in unread news items, which can help recommend more relevant news items than before. Our research is