An efficient and scalable ranking technique for mashups involving RSS data sources

Osama Al-Haj Hassan*, Thamer Al-Rousan1, Anas Abu Taleb1, Adi Maaita1

Ibros University, 11622 Amman, Jordan

ARTICLE INFO

Article history:
Received 2 October 2012
Received in revised form 5 April 2013
Accepted 15 June 2013
Available online 21 July 2013

Keywords:
Mashup Ranking Personalization Vector space model Web service Recommendation

ABSTRACT

Mashups are key category of Web 2.0 personalized applications. Due to personalization property of Web 2.0 applications, number of mashups hosted by a mashup platform is increasing. End-users are overwhelmed by the increasing number of mashups. Therefore, they cannot easily find mashups of their interest. In this paper, we propose a novel mashup ranking technique based on the popular Vector Space Model (VSM) for mashups that use RSS feeds as data sources. Mashups that are ranked higher would be more interesting to end-users. In order to evaluate our mashup ranking technique, we implement it in a prototype where end-users select mashups that they consider interesting. We implicitly collect the end-user mashup selections and record the outcome of our ranking technique, and then we analyze them. Recorded R-Precision value in our technique is on an average 30% higher than R-Precision value in binary ranking technique which shows an improvement in capturing mashups that resemble end-user interest. In our design, we make sure our mashup ranking technique scales well to increasing number of mashups.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

The main interest of Internet users is shifting towards Web 2.0 (Murugesan, 2007) applications. One of the main features of Web 2.0 that attracts Internet users is personalization. Web 2.0 personalized applications provide tools customized per end-user specific needs.

One of the icons of Web 2.0 personalized applications is mashups. Mashups enable end-users to remix feeds and further process them forming their own Web services (Beemer and Greg, 2009). This is why mashups in concept are Web services created by end-users. Mashups allow end-users to query the Web in a more personalized, flexible, and effective way. They enable end-users to extract data from multiple data sources and then combine data and apply further processing and refinement to it. Each end-user can create his/her own set of mashups; this is why mashups aid towards enhancing the world of personalization. A mashup example is found in Fig. 1. We state here that our work only considers RSS feeds as data sources for mashups.

Mashup platforms are those applications that offer a graphical user interface through which the end-user can perform the following tasks. First, the end-user can design his/her own mashups. Second, the end-user can browse other end-user mashups. Third, the end-user can execute mashups of his/her own or other end-users publicly available mashups. Examples of mashup platforms on the Web are Yahoo Pipes (Yahoo Inc., 2007) and Intel MashMaker (Intel Corp., 2007).

Another application that uses mashups in its core is mashup reader. Mashup readers are applications that allow end-users to design their own mashups and at the same time search and subscribe for mashups of their interest. Accordingly, end-users are able to follow updates in the result of execution of their favorite mashups.

One drawback of mashup platforms and mashup readers is that they do not provide a ranking mechanism for mashups. In mashup platforms, suppose the end-user is interested in browsing available mashups related to sport news of "LA Lakers" and "Kobe Bryant". The result of such browse request might involve large number of mashups and that makes it difficult for the end-user to select best mashups that satisfy his/her request. Therefore, it is important for end-users to get the result of their search request ordered based on mashup ranking which is unfortunately not provided by mashup platforms. In mashup readers, the end-user may care about following periodical updates about papers published by 'Springer' and 'Elsevier' only if they fulfill certain criterion such as having 'Web Services' topic and concerning 'Quality of Service'. Similarly, large number of mashups might satisfy this query. Consequently, mashup ranking is a necessity for end-users so that they can subscribe to mashups that conform to their demands.

* Corresponding author. Tel.: +96 2799834001; fax: +96 264711505.
E-mail addresses: osamahajhassan@hotmail.com, osama.haj@ipu.edu.jo (O. H. Hassan), thamer.rousan@ipu.edu.jo (T. Al-Rousan), anas.taleb@ipu.edu.jo (A. A. Taleb), adimaaita@ipu.edu.jo (A. Maaita).
1 Fax: +96 264711505.

1084-8045/S - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jnca.2013.06.004