Expert Systems With Applications 000 (2016) 1–14

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A recommendation approach for consuming linked open data

Jonice Oliveira[1,*], Carla Delgado, Ana Carolina Assaife

Universidade Federal do Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Received 6 April 2016
Revised 16 October 2016
Accepted 16 October 2016
Available online xxx

Keywords:
Information retrieval
Linked open data
Recommender systems

ABSTRACT

Most of linked open data (LOD) applications focus on the search and visualization of information, not efficiently using the links among objects in different data sources and the semantics of their relations. This work aims to create a LOD-consuming approach that uses recommendation techniques based on items' description, their relations, users' interests and social network. The proposed approach was instantiated by an application that uses movie related LOD. The results obtained in our experiments were promising: accuracy of the recommendations generated was equal or better, compared to other recommender algorithms used in conventional (not LOD) scenario.

1. Introduction

The World Wide Web (WWW, or Web for short) changed the way people access and share knowledge (Berners-Lee, Cailliau, Groff, & Pollermann, 1992). Barriers in document publication and sharing have been considerably reduced, causing the global information space to grow. In spite of the efforts to improve data consumption, the data published in the Web still form knowledge islands. Also, document search is still based on keywords, heterogeneous data models and data types, and no pattern to resource identification has been established. As documents were primarily made to be understood by humans, document integration used to be a manual process. This makes the computational processing and information inference on published documents very difficult.

Because of the difficulties in automatically extracting meaning from documents, the Web evolved to the so called "Web of Data", incorporating mechanisms for a better semantic representation and integration of Web publications (Bizer, Heath, & Berners-Lee, 2009). The concept of Linked Open Data (LOD) establishes the principles for data publication and data relation, and motivates the creation of links with associated meaning among data from several sources (Bizer, Cyganiak, & Heath, 2007). In LOD, the information is represented in RDF (Resource Description Framework) triple graphs (Klyne & Carroll, 2004) where world elements are represented by unique identifiers called URIs (Uniform Resource Identifier) (Berners-Lee, Fielding, & Masinter, 1998). What makes LOD particularly interesting is the huge volume of structured data available, which can be consumed by humans or artificial agents (computer programs) and also, the links among items of different sorts, as for example, links among person data bases to movies' data bases.

With this mobilization from the community and with the growing volume of published data, research was conducted and applications developed to improve the means to consume and explore the Web of Data. According to Bizer et al. (2009), the main advantage that linked data brings to the user is the possibility to access related or combined data from distributed and even heterogeneous data sources. Also, according to the same authors, using the Web as a unique global database has a couple of challenges, as for example the construction of applications that ease the user interaction with the data. This way, new contributions to the fields of data navigation, data search and data consumption is welcome.

The support for users that consume information published as LOD is still very limited: such applications are in majority focused in the search and visualization of the retrieved information. They are able to show related vocabulary, exhibit the data in facets, but they do not suggest related information to catch the user interest (Franz, Koch, Dividino, & Staab, 2010). It is still a challenge to make the consumption of LOD attractive to the end user. Our goal is to tackle the problem of easing the consumption of LOD to the end user, as well as support the construction of applications that would motivate the user to consume LOD. We believe that this is an important step towards using the web as a unique global data repository to serve all sorts of end users' applications.

After studying the related works on the field of end users' LOD consumption and navigation, we noticed that recommenda-

* Corresponding author.
E-mail addresses: jonice@dcc.ufrj.br, jonice@gmail.com (J. Oliveira), carla@dcc.ufrj.br (C. Delgado), ana.assaife@ppgi.ufrj.br (A.C. Assaife).

[1] Authors would like to thank CNPq, CAPES and FAPERJ for partially supporting this research work.

http://dx.doi.org/10.1016/j.eswa.2016.10.037
0957-4174/

tion mechanisms are applied to ease and motivate the consumption of such data (Heitmann & Hayes, 2010; Di Noia, Cantador, & Ostuni, 2014; Musto, Lops, Basile, de Gemmis, & Semeraro, 2016, to cite some). In this scenario of recommendation systems using LOD, previous works used to combine LOD with other closed systems, as for instance Facebook, MySpace and LastFM (Mirizzi, Noia, Ragone, Ostuni, & Sciascio, 2012; Heitmann & Hayes, 2010; Passant, 2010), envisioning the improvement of the recommendations by the use of more structured and semantically richer data about the user. The improvement of recommendations using only LOD data – on a way to work only with open data on a self-sustainable LOD universe, and not requiring user information – was an open problem. The presented approach is based on this challenge: working with the recommendation of LOD using one or more datasets – and only LOD datasets – enabling the end users to navigate, by the recommendations, in LOD.

This paper describes a recommendation approach to LOD, using one or more datasets. This is the main contribution, considering that previous works are limited to the use of only one dataset. The approach here presented is based on three recommendation types:

(1) Recommendation of resources of the same type of the input resource, considering that the input resource is a resource that interests the user, or that the user likes;
(2) Recommendation of social networks, that recommends people from the same domain as the input resource;
(3) Recommendation of resources of any type that are related to the input resource, using the textual description of the resources to find related items. This empowers the identification of related resources from different types. For instance, when the user searches for a movie, it is possible to recommend people, books, soundtracks or any other resource that the algorithm considers similar to the movie given as start point for the search.

As we said previously, our goal is to tackle the problem of easing the consumption of LOD to the end user. Consequently, we would like to analyze how our approach could be feasible and efficient from the user's point-of-view. An experiment, divided in two stages, was conducted to validate this work. Firstly, we conduct a quasi-experiment, aimed at evaluating the end user perception of the recommendation technique. In this stage, a group used our system and we collected information about the execution and their perception. These results were used in the second phase of the evaluation, which was a quantitative off-line analysis. In this stage, we measured a set of recommendation metrics for the proposed algorithms (RICCI et al., 2011), such as: accuracy, prediction accuracy by ranking, novelty and utility. All the analysis based on the user's feedback and the four metrics were positive and, consequently, we could affirm that this approach brings contributions for Consuming Linked Open Data scenario.

This paper is organized as follows. Section 2 presents the works related to the use of LOD in recommendations. Our proposal is detailed in section 3. In Section 4 our proposal is applied to the movie scenario, and the validation experiments are described in Section 5, as well as the results obtained. Conclusions and future work come in Section 6.

2. Related work

Considerations on how to take advantage of LOD have been raised since real LOD bases became available (Heath, 2008). Due the absence of works in the area, Passant (2010) proposed a set of measures to compute the semantic distance in LOD in order to measure similarity between two resources, and discussed how such measures could be used in applications, such as resources' recommendations. The measures proposed were based on the abundance of links among the resources and do not consider the ontology hierarchy describing this data. Passant conducted an experiment to decide which of the distances (direct, indirect, balanced and combined distances) would be more appropriate for a recommender system. The combined balanced distance was the one performing best.

Later, the same author (Passant & Decker, 2010; Passant, 2010) applied the created methodology to develop a system to recommend songs called dbrec (Passant, 2010). The output of this system was validated by making a comparison to Last.fm, a music recommender system already established. As we will see later, the main difference to our work is the restriction of the items to be compared. In (Passant, 2010) just resources that have the properties dbpedia:Band and/or dbpedia:MusicalArtist are taken into account. In our approach, broader criteria are used (correlated properties, similarities in descriptions for a resource). Another aspect observed in this related work is that it just uses one database, thus not benefiting from the links found in several bases and also DBPedia.

Still in the music domain, the Seevl application is worth mentioning (Passant, 2011). It is a Chrome plugin that provides music recommendations, biographies and complementary information to youtube videos. Once extracted and converted to RDF, the data about artists is translated into a common vocabulary using the Music Ontology Raimond et al. (2007), forming a graph of musical entities like artists, bands, genres, labels, and others. Seevl also provides combined search of properties, making it possible to search "all artists and bands of rock genre that played in a particular festival". In spite of presenting the used architecture and the steps in the process, this work did not bring any information about the implementation of the recommendation algorithms used. But as it is from the same author of the previous works described in this section, and using data from the same domain, we believe that Seevl is a mature application developed after other more theoretical works from the same author.

Also in the field of media files, MORE (Mirizzi et al., 2012) is an application developed to complement Facebook by recommending movies. MORE uses the data from bases of triplets in the domain, and compute movie similarities. The special thing about this application is that besides looking at the content, it also considers the user's Facebook profile. By using as input information the movies that the user marked as interesting in his or her Facebook profile, MORE intends to overcome the "cold start" problem, i.e. having some information even about users that have not yet rated any movie. The main contribution of this work is that it provides a semantic recommendation algorithm, called Semantic Vector Space Model. The algorithm is an adaptation from the Vector Space Model (Salton, Wong, & Yang, 1975). In spite of exploring the same domain as our work (movies), this work is not strongly related to ours as it was not our goal to use user profile information, as this information is not always available. Our goal was to generate recommendations from RDF databases.

Other project worth mentioning is RKBExplorer.com (Glaser, Millard, & Jaffri, 2008). It is a semantic web application that aims to present unified versions from a considerable number of different data sources. Up to now, the application works in the domain of people and publications in Computer Science. The data about people, publications and institutions is in RDF, and have been provided by the project partners in ReSIST (Glaser & Millard, 2007). This application provides an interface where the user specifies two resources of his or her interest and the type of the relation between them, and gets as an output the properties that link these resources. This application requires that the user knows both the input resources and the type of each resource. This is a limitation in the sense that there is no way to surprise the user recommending novel items. In our approach the user does not need to know anything about the resources nor the datasets.

Please cite this article as: J. Oliveira et al., A recommendation approach for consuming linked open data, Expert Systems With Applications (2016), http://dx.doi.org/10.1016/j.eswa.2016.10.037