Personalised, Serendipitous and Diverse Linked Data Resource Recommendations

Milan Dojchinovski(✉) and Tomas Vitvar

Web Intelligence Research Group, Faculty of Information Technology,
Czech Technical University in Prague, Prague, Czech Republic
{Milan.Dojchinovski,Tomas.Vitvar}@fit.cvut.cz

Abstract. Due to the huge and diverse amount of information, the actual access to a piece of information in the Linked Open Data (LOD) cloud still demands significant amount of effort. To overcome this problem, number of Linked Data based recommender systems have been developed. However, they have been primarily developed for a particular domain, they require human intervention in the dataset pre-processing step, and they can be hardly adopted to new datasets. In this paper, we present our method for personalised access to Linked Data, in particular focusing on its applicability and its salient features.

Keywords: Personalisation · Recommendation · Linked data · Semantic distance · Similarity metric

1 Introduction

Due to the huge and diverse amount of information, the actual access to a piece of information in the Linked Open Data (LOD) cloud still demands significant amount of effort. To overcome this problem, number of Linked Data based recommender systems have been developed. However, they have been primarily developed for a particular domain (e.g., music or videos), they require human intervention in the dataset pre-processing step, and they can be hardly adopted to other datasets.

In our related EKAW 2014 research paper [3], we described in detail our method for "personalised access to Linked Data". In this paper, we focus on its usage in different domains and datasets, and its salient features. Section 2 describes the resource recommendation workflow. Section 3 presents a resource recommendation case study from the Web services domain. Finally, Section 4 highlights its salient features.

2 Linked Data Resource Recommendation

The prototype of our method implements the workflow depicted in Figure 1. It starts with (i) Linked Data datasets and user profiles import, followed by

P. Lambrix et al. (Eds.): EKAW 2014 Satellite Events, LNAI 8982, pp. 106–110, 2015.
DOI: 10.1007/978-3-319-17966-7_11

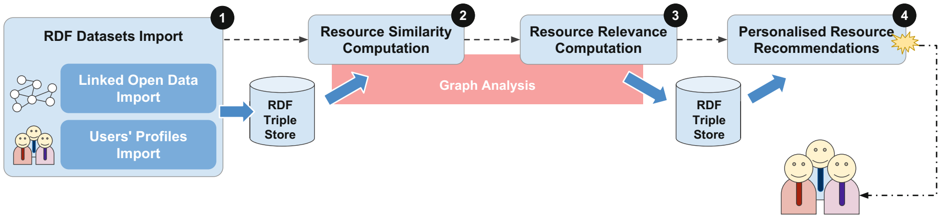

Fig. 1. Overview of the resource recommendation workflow

(ii) graph based data analysis, and ends with (iii) generation of personalised Linked Data resource recommendations.

RDF Datasets Import. First, it is necessary to identify one or more datasets which contain the resources of interest. The datasets can be single domain datasets such as GeoNames, or multi-domain datasets such as DBpedia. It is also necessary to provide users' profiles information which encodes users' activities and their relations to other people and resources. According to the statistics as of April 2014 [5], 48% of all datasets in the LOD cloud are datasets providing social information such as people profiles data and social relationships among them. Note that the RDF data is imported and analysed in its original form. In other words, we do not perform any domain-dependent selection or filtering of the information. For comparison, the movie recommender described in [1] uses only a subset of the information, defined as a set of RDF properties.

Graph Analysis. Our resource recommendation method is based on the collaborative filtering technique and it recommends resources from users with interests in similar resources. To measure the user similarity, we perform graph based analysis of the users' RDF context graphs. To this end, we introduce a novel domain–independent semantic resource similarity metric, which takes into account the *commonalities* and *informativeness* of the shared resources. Our assumption is that the more information two resources share, the more similar they are. And second, the more informative resources they have in common, the more similar they are. The resources informativeness is primarily incorporated to differentiate informative shared resources from non-informative (e.g., *owl:Thing* or *skos:Concept*).

Figure 2 illustrates computation of similarity between two users, *Alfredo* and *mlachwani*. The users have 6 resources in common, which convey different amount of information content. According to our assumption, resources with high graph degree value convey less information content, than resources with lower graph degree. Considering the whole Linked Web APIs dataset, the Facebook API with node degree of 418, is more informative and carries more similarity information than the Twitter API, with node degree of 799. Facebook API is characterised with a lower degree due to its lower usage in mashups, leading to a lower number of incident links.

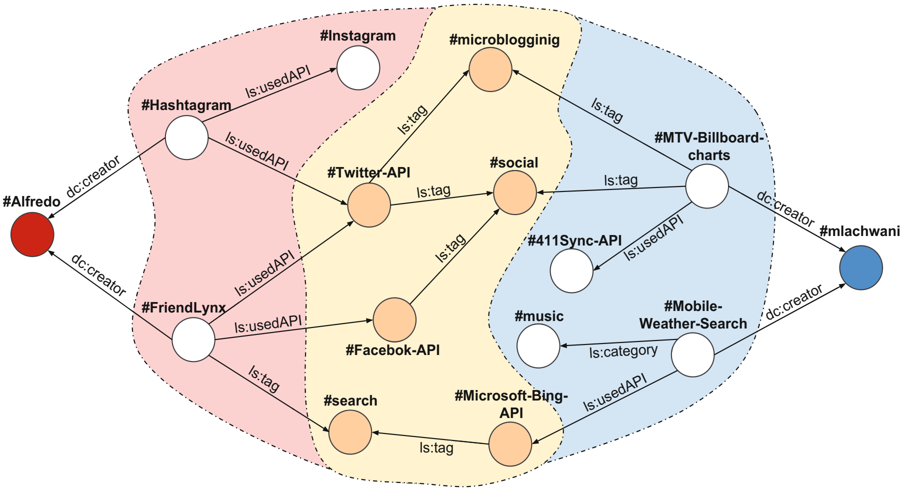

Fig. 2. Excerpt from the Linked Web APIs dataset with resource context graphs with context distance of 3

Personalised Resource Recommendations. The similarities between the resources representing users are then used to compute the relevance of each resource candidate. For a user requester, the method first computes the *connectivity* between each similar user (i.e., a user similar with the user requester) and each resource candidate. The connectivity is computed as the amount of gained information between the similar user and the resource candidate. The final relevance score aggregates the user similarity scores and the connectivity scores for the resource candidate. Finally, the method returns the top-n most relevant resources for the user requester.

3 Case Study: Recommendation of Linked Web APIs

In this section we show a resource recommendation case study on the Linked Web APIs dataset [2], which provides RDF representation of the ProgrammableWeb.com service repository as of April 24th 2014. The datasets provides descriptions for 11,339 Web APIs, 7,415 mashups and 5,907 users.

Step 1: Identification of Requesters and Items of Interest. We start with identification of resources representing requesters and resources representing items of interest. In our case, requesters are users represented with the *foaf:Person* class and items of interests are Web APIs represented with the *wl:Service* class. If we want to develop a music recommendation system for recommendation of artists and bands, it will be necessary to specify the class describing those resources of interest. In DBpedia those are resources of type *dbpedia:MusicalArtist* and *dbpedia:Band*.

Step 2: Data Analysis. Next, the method computes the similarities between resources representing users (i.e., *foaf:Person*) and computes the relevance of each resource candidate for each user. In this step, the resource information content is also computed and considered.

Step 3: Generate Resource Recommendations. Finally, the method can generates a list of top-n most relevant resources for the user requester.

4 Salient Features

Serendipitous and Diverse Recommendations. For any recommender system it is important to accommodate the difference between the individuals in order to produce accurate, while at the same time serendipitous and diverse recommendations. The results from our evaluation [3] show that our method satisfy these requirements and also outperforms the traditional personalised collaborative filtering methods and non-personalised methods.

Adaptability. One drawback of the existing Linked Data based recommendation systems [1,4] is their applicability on different datasets due to the metric used for computation of the resource similarity. The metrics used in [1,4] are not suitable for computation of similarities between resources in datasets (i.e., the Linked Web APIs dataset), where the graph distance between the resources is more than two. In comparison, our method can be easily adapted to any dataset by setting the size of the resource context. In datasets, where the users in the RDF graph are close to each other, will require setting lower distance, while in datasets where the users are far, will require higher distance. Choosing small context distance in datasets where the users are far from each other, can possibly lead to no overlap of the resource context graphs, and no similarity evidenced.

Cross-Dataset and Domain Recommendations. As shown in Section 2, without any restrictions our method can consume and benefit from one or more Linked Data datasets from different domains, and at the same time produce resource recommendations for these domains and datasets. In contrast, the methods presented in [1,4] use only subsets of Linked Data datasets, which need to be defined in advance for the particular domain.

Acknowledgments. This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/104/OHK3/1T/18. We also thank to Programmableweb.com for supporting this research.

References

1. Di Noia, T., et al.: Linked open data to support content-based recommender systems. In: Proceedings of the 8th International Conference on Semantic Systems, I-SEMANTICS 2012, pp. 1–8. ACM, New York (2012)
2. Zaremba, M., Dojchinovski, M., Kuchar, J., Vitvar, T.: Personalised graph-based selection of web APIs. In: Cudré-Mauroux, P., Heflin, J., Sirin, E., Tudorache, T., Euzenat, J., Hauswirth, M., Parreira, J.X., Hendler, J., Schreiber, G., Bernstein, A., Blomqvist, E. (eds.) ISWC 2012, Part I. LNCS, vol. 7649, pp. 34–48. Springer, Heidelberg (2012)

3. Dojchinovski, M., Vitvar, T.: Personalised access to linked data. In: Janowicz, K., Schlobach, S., Lambrix, P., Hyvönen, E. (eds.) EKAW 2014. LNCS, vol. 8876, pp. 121–136. Springer, Heidelberg (2014)
4. Passant, A.: DBrec — Music recommendations using DBpedia. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) ISWC 2010, Part II. LNCS, vol. 6497, pp. 209–224. Springer, Heidelberg (2010)
5. Schmachtenberg, M., Bizer, C., Paulheim, H.: Adoption of the linked data best practices in different topical domains. In: Mika, P., Tudorache, T., Bernstein, A., Welty, C., Knoblock, C., Vrandečić, D., Groth, P., Noy, N., Janowicz, K., Goble, C. (eds.) ISWC 2014, Part I. LNCS, vol. 8796, pp. 245–260. Springer, Heidelberg (2014)