# Hey! Ho! Let's Go! Explanatory Music Recommendations with dbrec

Alexandre Passant and Stefan Decker*

Digital Enterprise Research Institute,
National University of Ireland, Galway
firstname.lastname@deri.org

**Abstract.** In this demo paper, we present dbrec (http://dbrec.net), a music recommendation system using Linked Data, where recommendation are computed from DBpedia using an algorithm for *Linked Data Semantic Distance* (*LDSD*). We describe how the system can be used to get recommendations for approximately 40,000 artists and bands, and in particular how it provides explanatory recommendations to the end-user. In addition, we discuss the research background of dbrec, including the *LDSD* algorithm and its related ontology.

## 1 Research Background

### 1.1 Measuring Semantic Distance on Linked Data

The underlying research question behind dbrec is to understand how to define semantic distance [4] measures for resources that follow the Linked Data principles [1]. In order to define such measures, we first defined a theoretical model for Linked Data datasets, as follows.

**Definition 1.** *A dataset following the Linked Data principles is a graph $G$ such as $G = (R, L, I)$ in which $R = \{r_1, r_2, ..., r_n\}$ is a set of resources — identified by their URI —, $L = \{l_1, l_2, ..., l_n\}$ is a set of typed links — identified by their URI — and $I = \{i_1, i_2, ..., i_n\}$ is a set of instances of these links between resources, such as $i_i = \langle l_j, r_a, r_b \rangle$.*

In addition, we defined different functions that identify if and how two resources (represented by their URI, and following the Linked Data principles) are connected in such graphs.

**Definition 2.** *$C_d$ is a function that computes the number of* direct and distinct *links between resources in a graph $G$. $C_d(l_i, r_a, r_b)$ equals 1 if there is an instance of $l_i$ from resource $r_a$ to resource $r_b$, 0 if not. By extension $C_d$ can be used to compute (1) the total number of direct and distinct links from $r_a$ to $r_b$ ($C_d(n, r_a, r_b)$) as well as (2) the total number of distinct instances of the link $l_i$ from $r_a$ to any node ($C_d(l_i, r_a, n)$).*

* The work presented in this paper has been funded in part by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Líon-2).

L. Aroyo et al. (Eds.): ESWC 2010, Part II, LNCS 6089, pp. 411–415, 2010.

**Definition 3.** *$C_{io}$ and $C_{ii}$ are functions that compute the number of* indirect and distinct *links, both* outcoming and incoming*, between resources in a graph G. $C_{io}(l_i, r_a, r_b)$ equals 1 if there is a resource n that satisfy both $\langle l_i, r_a, n\rangle$ and $\langle l_i, r_b, n\rangle$, 0 if not. $C_{ii}(l_i, r_a, r_b)$ equals 1 if there is a resource n that satisfy both $\langle l_i, n, r_a\rangle$ and $\langle l_i, n, r_b\rangle$, 0 if not. By extension $C_{io}$ and $C_{ii}$ can be used to compute (1) the total number of indirect and distinct links between $r_a$ and $r_b$ ($C_{io}(n, r_a, r_b)$ and $C_{ii}(n, r_a, r_b)$, respectively outcoming and incoming) as well as (2) the total number of resources n linked indirectly to $r_a$ via $l_i$ ($C_{io}(l_i, r_a, n)$ and $C_{ii}(l_i, r_a, n)$, respectively outcoming and incoming)*

Based on these definitions, we defined different formula for computing *Linked Data Semantic Distance*[1] [3], *i.e.* the distance that exists between two resources within a Linked Data dataset[2]. These formula were defined using different criteria, using both direct and indirect relationships ($C_d$, $C_{io}$, $C_{ii}$), as well as using weights to give more importance to the links that are used only a few times in the dataset. Based on these different measures and user interviews, we decided to concentrate on a particular one, that we will simply name $LDSD$, and that uses both direct and indirect links as well as the aforementioned weights. This measure is defined as follows (Fig. 1).

$$LDSD(r_a, r_b) = \frac{1}{1 + \sum_i \frac{C_d(l_i, r_a, r_b)}{1+log(C_d(l_i, r_a, n))} + \sum_i \frac{C_d(l_i, r_b, r_a)}{1+log(C_d(l_i, r_b, n))} + \sum_i \frac{C_{ii}(l_i, r_a, r_b)}{1+log(C_{ii}(l_i, r_a, n))} + \sum_i \frac{C_{io}(l_i, r_a, r_b)}{1+log(C_{io}(l_i, r_a, n))}}$$

**Fig. 1.** The *LDSD* measure

### 1.2 The LDSD Ontology

Moreover, in order to represent the distances computed using the aforementioned *LDSD* measure in a machine-readable way (so that they can be queried and reused by third-party applications), we designed a related ontology, available at `http://dbrec.net/ldsd/ns`. This ontology has two main goals. On the one hand, it aims at modelling the distance between resources, so that they can be queried using SPARQL (and in some cases with its `FILTER` clause to limit resources to the ones at a particular distance from the seed one). On the other hand, its goal is to store some information about the way the distance has been computed, so that these distances can be explained, as we shall see in the upcoming section.

Using this ontology, the fact that Elvis Presley is at a distance of 0.09 from Johnny Cash, and that such distance can be explained because (among other) both share the value `http://dbpedia.org/class/yago/SunRecordsArtists` for their `rdf:type` property, which is shared by only 19 artists in the original dataset, can be represented as follows (Table 1).

[1] Note that we use the term *distance* while some of these measures are not symmetric.

[2] We shall note that theoretically, nothing prevents this dataset to be distributed as *e.g.* the Linking Open Data cloud.

**Table 1.** Excerpt of the distance explanation between Johnny Cash and Elvis Presley, using the *LDSD* ontology

```
@prefix ldsd: <http://dbrec.net/ldsd/ns#> .
<http://dbrec.net/distance/774a32aa-dede-11de-84a3
    -0011251e3563> a ldsd:Distance ;
   ldsd:from <http://dbpedia.org/resource/Johnny_Cash> ;
   ldsd:to <http://dbpedia.org/resource/Elvis_Presley> ;
   ldsd:value "0.0977874534544" .
<http://dbrec.net/distance/774a32aa-dede-11de-84a3
    -0011251e3563> ldsd:explain [
  a ldsd:IndirectOut ;
  ldsd:property <http://www.w3.org/1999/02/22-rdf-syntax-
       ns#type> ;
  ldsd:node <http://dbpedia.org/class/yago/
      SunRecordsArtists> ;
  ldsd:total "19"
] .
```

In addition, the *LDSD* ontology features mappings with the MuSim ontology[3] [2]. Additional mappings with SCOVO[4] — the Statistical COre VOcabulary — might be provided in the future (since the number of instance sharing a particular link can be considered as statistical data).

## 2 dbrec: Music Recommendations Using *LDSD*

Based on the previous findings, we implemented dbrec — `http://dbrec.net` —, a system that demonstrates how *LDSD* can be used in the realm of recommender systems. In particular, dbrec has been build by computing *LDSD* for all artists and bands referenced in DBpedia. While it does not involve cross-datasets

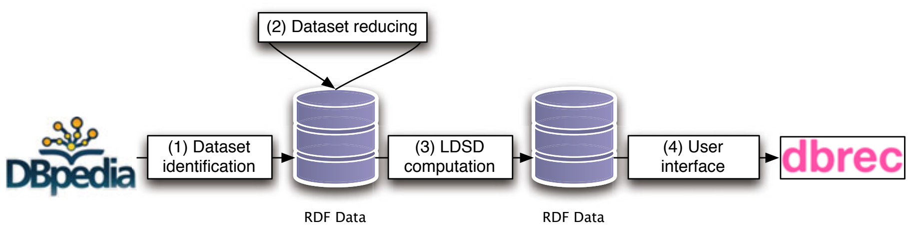


**Fig. 2.** Architecture of the dbrec framework

[3] `http://grasstunes.net/ontology/musim/musim.html`
[4] `http://sw.joanneum.at/scovo/schema.html`

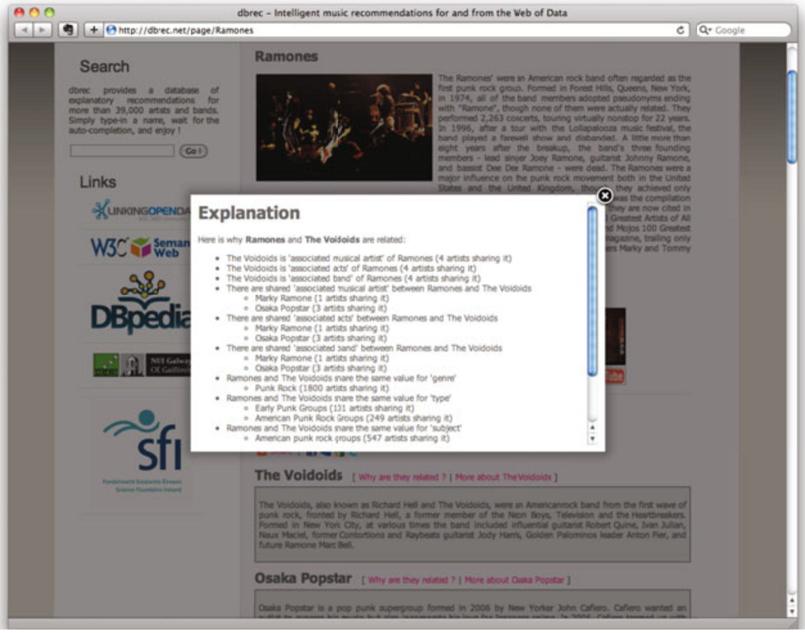


**Fig. 3.** Recommendations and explanations in dbrec

recommendations, it offers two major advantages: (i) there are more than 39,000 artists available in DBpedia for which recommendations can be build, and (ii) DBpedia also provides pictures and descriptions for most of them, that can be used when building the related user interface, as we will see in the next section. The system was implemented using the following steps (Fig. 2): (1) identify the relevant dataset from DBpedia; (2) reduce the dataset for query optimisation; (3) compute the distances using the *LDSD* algorithm and represent the results using the *LDSD* ontology; (4) build a user interface for browsing recommendations.

## 3 Purpose of the Demonstration

The purpose of the demonstration is to give a comprehensive overview of the dbrec system in use.

First, attendees will be able to check recommendations for the artists and bands of their choice. The search interface of dbrec provides auto-completion capabilities so that users simply start typing the name of a band, and the system suggests names for which recommendations are available in the system. Once validated, users have access to a live view of the recommendations, built using SPARQL queries applied to the RDF data resulting from the *LDSD* algorithm (this data being modelled with the aforementioned *LDSD* ontology). As we previously mentioned, dbrec relies on DBpedia to provide picture and description of each artist and band, hence enhancing the user experience when browsing recommendations. It also displays related YouTube videos about the main artist and provide links to share the recommendation on various Web 2.0 services such as Twitter and Facebook.

In addition, attendees will experience the explanation feature provided by dbrec, so that they can learn why the recommendations have been made. Indeed, when browsing recommendations, a pop-up providing such explanations can be displayed for each recommended artist or band. The explanations are provided using the information recorded about each measure, using the *LDSD* ontology as we previously detailed. By using this feature, attendees will discover why dbrec recommends $X$ if they are looking for information about $Y$.

The following figure (Fig. 3) displays the recommendation page for the Ramones, as well as a pop-up explaining why the Voidoids are recommended in that case[5].

In addition, we will show how these recommendations, by being available as Linked Data (*via* RDFa markup within the pages) can be browsed independently, and how they could be combined with other data for navigation and querying purposes.

## References

1. Berners-Lee, T.: Linked Data. In: Design Issues for the World Wide Web, World Wide Web Consortium (2006), `http://www.w3.org/DesignIssues/LinkedData.html`
2. Jacobson, K., Raimond, Y., Sandler, M.: An Ecosystem for Transparent Music Similarity in an Open World. In: International Symposium on Music Information Retrieval (2009)
3. Passant, A.: Measuring Semantic Distance on Linking Data and Using it for Resources Recommendations. In: Linked AI: AAAI Spring Symposium Linked Data Meets Artificial Intelligence. AIII (2010)
4. Rada, R., Mili, H., Bicknell, E., Blettner, M.: Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man and Cybernetics 19, 17–30 (1989)

[5] Live view at `http://dbrec.net/page/Ramones`