

FRBRPedia: a Tool for FRBRizing Web products and Linking FRBR Entities to DBpedia

Fabien Duchateau*
Norwegian University of
Science and Technology
NO-7491 Trondheim, Norway
fabierend@idi.ntnu.no

Naimdjon Takhirov
Norwegian University of
Science and Technology
NO-7491 Trondheim, Norway
takhirov@idi.ntnu.no

Trond Aalberg
Norwegian University of
Science and Technology
NO-7491 Trondheim, Norway
trondaal@idi.ntnu.no

ABSTRACT

The FRBR model has received much attention due to its potential for greatly improving user interaction with digital libraries. However, the amount of information found on the Web is far larger than in digital libraries. In this demo, we present an approach to transform Web-based resources to a FRBR compatible form, a process known as FRBRization. The FRBRized collection is then linked to DBpedia, thus providing a basis for information sharing and verification.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

Keywords

FRBR, FRBRization, Linked Data, Entity Matching, Data Visualization

1. INTRODUCTION

Functional Requirements for Bibliographic Records (FRBR) is an entity relationship model proposed by the International Federation of Library Association (IFLA) in 1998 [3]. This semantically rich model provides a framework for the representation of the bibliographic universe, thus enabling advanced capabilities for end-users in terms of search or navigation. The challenge now deals with the conversion of the data stored in existing formats into FRBR. For example, the products stored on the Web are represented in different formats such as HTML, XML, various databases, etc.

In this paper, we demonstrate an approach to extract and convert products from Amazon bookstore into the FRBR model. We link the generated FRBR works to Linked Open Data cloud (LOD) by using entity matching techniques. One of the goals with linking to LOD is to verify the FRBRized information. We have used the DBpedia knowledge base since it contains potentially a great amount of information.

2. OVERVIEW OF OUR TOOL

We briefly describe how we transform a Web product from Amazon into the FRBR model and how we link the FRBR

*The author carried out this work during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme.

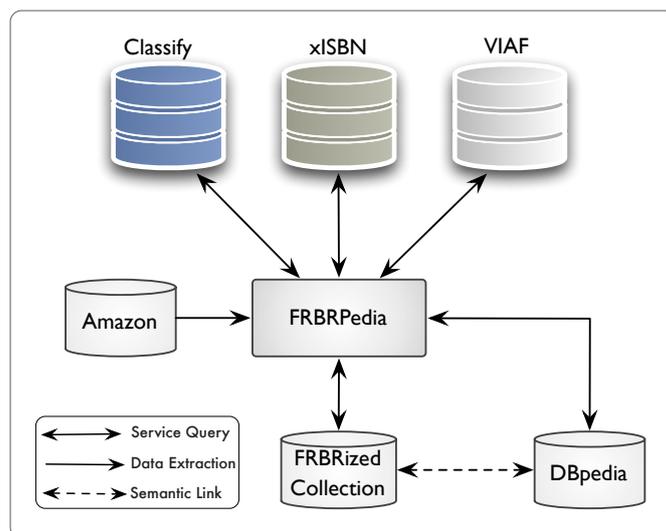


Figure 1: The Architecture of the System

work to DBpedia. Figure 1 depicts the architecture of our system, which is detailed below.

Based on the product information extracted from Amazon bookstore, we first need to create (at least) four FRBR entities: work, manifestation(s), expression(s), and actor(s). The first step consists of identifying the work. We use the external service *OCLC Classify API*¹ to fulfill this goal. We send a query (e.g., ISBN number) to retrieve the work corresponding to the product. Although OCLC includes a large collection of works, it occurs that the service does not return any result, for instance with products specific to a country. In the future, we plan to integrate other methods for identifying a work such as z39.50². The second step of our FRBRization process deals with the discovery of related manifestations for the previously identified work. Similarly, we chose an OCLC Service, *xISBN*, since a search in the same database increases our chances of finding related manifestations. The *xISBN* Web service returns ISBNs and other information associated with an individual intellectual work that is represented in the WorldCat catalog. From the set of related manifestations, we can automatically generate the third category of FRBR entities: expressions. By analyzing

attributes such as language and translator, we are able to create expressions, whose identifier is automatically generated. Next, the actors who contributed to this work are identified through the Virtual International Authority File (VIAF), a joint project between national libraries. This service aims at gathering authoritative names from many libraries into a global service freely available on the Web. Finally, the work, its expressions, manifestations and actors are linked during the last step. The final output of our FRBRization process is a set of RDF files for each entity type.

Linking the FRBR work to DBpedia is performed in two steps. First, we exploit information contained in the work (title, creator, categories, etc.) to query a DBpedia service named Lookup³. Indeed, the large amount of data stored in DBpedia cannot be fully searched without this kind of specific service, which returns a set of DBpedia entities related to the query. Our tool sends different queries to Lookup and it analyses all returned URIs with matching techniques. The goal of this second matching step is to rank the DBpedia entities and detect which one corresponds to the FRBR work. Therefore, we apply various similarity measures [2] between the attributes shared both by the work and the DBpedia entities (i.e., title, creator, date, categories and type). The computed similarity values are then aggregated into a global score using a weighted average function. This global score enables the ranking of all DBpedia entities for a given work, and the ones with the highest scores are presented to the user for validation. Once the user has validated the correct DBpedia entity, we finally display the FRBRized data enriched with DBpedia information.

3. DEMONSTRATION

Imagine we would like a FRBRized version of Agatha Christie's novel *And Then There Were None* sold on Amazon⁴. To achieve this goal with our tool, the user has two options. The former is a direct access at the webpage of FRBRpedia⁵. The latter is using a plugin which is activated when the user browses an Amazon product. This browser plugin⁶ is embedded using the GreaseMonkey extension for major browsers. The script is installed from the FRBRpedia webpage and the browser automatically detects if the user is in the product display page of amazon.com. In this case, a link appears on the product page. By clicking this button, the user is led to the FRBRpedia page⁵. The main difference between the direct access and the plugin usage is that the identifier (ISBN or Amazon ASIN number) for the product is automatically filled in with the plugin.

With our Agatha Christie example, the ISBN *0312330871* appears in the input field for the identifier. Once the form is submitted, the tool generates FRBR entities and it discovers the corresponding DBpedia entity (see Section 2). Depending on the number of manifestations for the work, the performance of the whole process varies from one second to one minute. In Figure 2, the FRBR entities, represented as RDF files, are shown in separate tabs with a syntax highlighting feature to promote readability. The corresponding DBpedia entity is also displayed in the fifth tab. Our tool has auto-

³<http://lookup.dbpedia.org>

⁴<http://www.amazon.com/dp/0312330871/>

⁵<http://november.idi.ntnu.no/frbrpedia/>

⁶<http://november.idi.ntnu.no/frbrpedia/amzn2frbr.user.js>

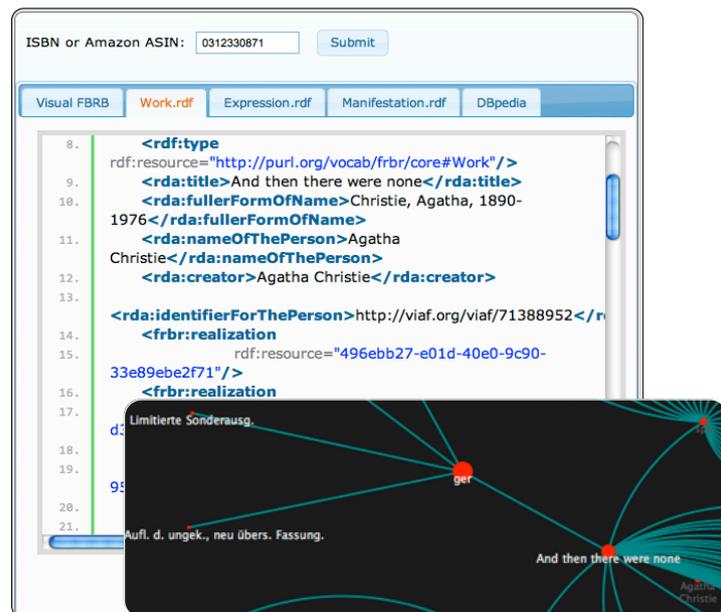


Figure 2: Screenshot of FRBRpedia

matically selected as corresponding DBpedia entity, the one with highest similarity value. In our example, the selected DBpedia entity is *dbpedia.org/resource/And_Then_There_Were_None*. Additionally, the system creates a JSON file which is used for visualization of the FRBR entities. This visualization is performed using the HyperTree included in the JavaScript Visualization Toolkit [1]. The bottom part of Figure 2 depicts a fragment of visualization of Agatha Christie's work. Namely, a german translation is displayed along with different manifestations.

4. CONCLUSION

In this demo, we have presented a tool to transform a Web-based product into the FRBR model. The resulting FRBR work has been linked to its corresponding DBpedia entity for verification and linking. We plan to extend our tool to include other services such as LibraryThing, Spotify, LastFM as well as other knowledge bases such as Freebase and OpenCyc.

5. REFERENCES

- [1] N. G. Belmonte. Creating interactive data visualizations for the web with the javascript infovis toolkit 2.0. In *YOW! Developer Conferences*, Brisbane, Australia, 2010.
- [2] J. Euzenat and P. Shvaiko. *Ontology matching*. Springer-Verlag, Heidelberg (DE), 2007.
- [3] IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional requirements for bibliographic records, final report. *UBCIM publications; new series*, 19(1), November 1998.