EASAIER Semantic Music Retrieval Portal

Michael Luger*, Ying Ding*, Francois Scharffe*, Rubing Duan* and Zhixian Yan*
*University of Innsbruck
Technikerstrasse 21a, A-6020 Innsbruck, Austria
{firstname.lastname}@deri.at

***Abstract*—The vision of the Semantic Web is to lift the current Web into semantic repositories where heterogeneous data can be queried and different services can be mashed up. The Web becomes a platform for integrating data and services. Ontology or agreed consensus is the key issue to achieve that. In this paper, we present the EASAIER (*Enabling Access to Sound Archives through Integration, Enrichment and Retrieval*) Music Retrieval Portal, a prototype application that allows the integration of audio-related metadata according to the MusicOntology. This builds the foundation for a semantic search over natively heterogeneous sources and services.**

I. Introduction

With the advent of the Web, massive amounts of digitized multimedia content and metadata has been published online. This ranges from cultural heritage archives, instrument or speech recordings, to large music databases that provide information on popular releases and artists. When searching for information on a piece of music, using a search engine such as Google still requires manual work in filtering out undesired information. It is possible to use specific search portals that cover a certain branch or genre, but the content it contains may be missing information that is available from a different archive. Furthermore, any information that lies out of the scope of the topic of the database (like an event, or a related artist), can usually not be examined any further.

In this paper, we present the EASAIER Semantic Music Retrieval Portal that interlinks individual sources by lifting them to a Semantic Web context. We focus on two use cases, namely the HOTBED Traditional Scottish Music Archive, and the MusicBrainz database. This is accompanied by a Web 2.0 component, that takes into account user-annotated content from external sources.

In this section we have briefly shown the motivation of our work. Section II outlines the effort of mapping the data sources to the MusicOntology, which builds the foundation for semantic search in the EASAIER Music Retrieval Portal. In section III, we present the enriched value achieved by ontology based search and the significance of integrating the Web 2.0 component. Section IV concludes the paper with an outlook on future work.

II. Mapping Component

Heterogeneity in data representation prevents interlinking multiple sources for cross-archive search. The Semantic Web intends to improve the current Web as a shared and distributed knowledge environment that is able to overcome such heterogeneity issues. The foundation for this is agreed consensus in the form of an ontology, that provides a formal specification of the domain at hand.

A. The MusicOntology

The MusicOntology[1] provides concepts to describe music resources, such as artists, releases, tracks, performances, arrangements, etc. It is built on top of existing ontologies, such as FOAF[2], the Timeline Ontology[3], the Event Ontology[4], and the Functional Requirements for Bibliographic Records (FRBR) ontology[5].

B. HOTBED Mapping

The HOTBED database contains information on Scottish Traditional Music. It has been provided as a use case for the evaluation of ongoing research within EASAIER. The dataset has been delivered in the form of a relational database dump, and it contains information on artists, the instruments played by the artists and recordings along with the corresponding audio files.

Using D2R Mapping[6], those fields were translated into the corresponding representation in the MusicOntology. The modularity of the ontology allows the usage of a custom ontology for any instrument related fields. Using D2R server, the data has been published along with the defined mapping via a SPARQL query endpoint, that translates incoming queries into SQL and returns RDF triples.

C. MusicBrainz Mapping

Using the Openlink Virtuoso RDFView functionality, the MusicBrainz[1] dataset that is publicly available in the form of a relational database dump has been mapped to the MusicOntology[2]. A SPARQL query endpoint allows to perform queries on top of the resulting dataset. Each MusicBrainz resource, be it an artist, record or track is specified by a unique MusicBrainz ID which is part of the URI of each resource's RDF representation.

This mapping effort is related to several approaches to publish and interlink data sources on the Web, as part of the *Linking Open Data on the Semantic Web* W3C SWEO community project[7]. For instance, each MusicBrainz artist is related to the corresponding resource on DBPedia[8], an RDF representation of Wikipedia, via *owl:sameAs*.

[1] http://www.musicbrainz.org/

[2] Credit goes to Zitgist LLC, who have defined this mapping document. http://www.zitgist.com

III. The Music Retrieval Portal

The Semantic Search component that takes advantage of the afformentioned *ontologized* data sources is delivered in the form of a web application with a graphical search interface. The user can select multiple data sources to be queried and will be presented with browsable results. Currently it is possible to perform queries on top of the HOTBED and MusicBrainz data sources.

A. HOTBED Search

The application allows searching for artists and recordings by instruments, and a subsumption reasoning component makes it possible to get more advanced results. For example, searching by *Wind Instrument* also yields results about artists that play instruments that fall under the input's subcategories, such as *Woodwind Instruments*. A graphical representation of the instruments is provided in the form of a tree for convenience.

B. MusicBrainz Search

When searching for popular artists from the MusicBrainz source, the query string is sent to the official MusicBrainz Web Service to retrieve the MusicBrainz ID. We have chosen this step over a SPARQL query over our dataset to take advantage of the sophisticated MusicBrainz native search engine. The MusicBrainz ID is then used in the context of a SPARQL query that retrieves information about the artist, such as pictures, related websites, discography information, etc. In addition, we retrieve further artists that are related to the original artist over one ore multiple degrees, such as via band memberships, collaborations, or resolved aliases. Those artists can then be browsed inside the web application, and any such subsequent underlying queries directly rely on the URI rather than syntactical input.
The MusicBrainz Search also takes into account results from DBPedia, which is interlinked with the MusicBrainz RDF representation, in order to also provide information that is not available from MusicBrainz, such as genre, the Wikipedia description, years active, etc.

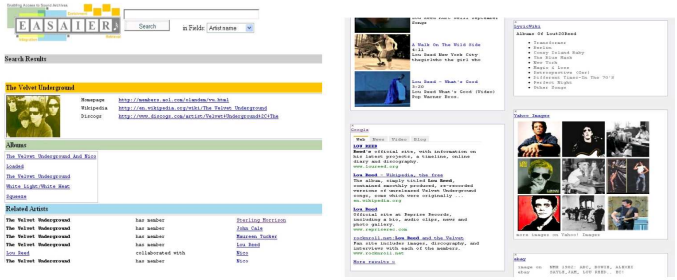

Fig. 1. Search results for artist Lou Reed.

C. Web 2.0 component

Furthermore, search results are complemented by web widgets that provide additional results from external sources. A query is posed to external APIs or RSS feeds, and partly those results are based on tagged keywords originating from a community effort. Currently, the EASAIER Music Retrieval Portal incorporates content from Google, YouTube, LyricWiki, Yahoo Images & Music, Amazon, eBay, and last.fm. An example of Music Retrieval Portal results is depicted in Fig. 1. In contrast to similar approaches that *mash up* individual data sources, this approach strongly relies on the structure of the MusicOntology. In other words, in the case of the MusicBrainz dataset, it is possible to browse artists in a structured manner, e.g. according to their unique MusicBrainz ID, and alongside provide unstructured or multimedia-based content that is aggregated by keywords.

IV. Conclusion and Future Work

In this paper, we have presented the EASAIER Music Retrieval Portal prototype, and we have shown how we can benefit from lifting data sources to an ontology and benefit from interlinking it with additional sources that are described according to the ontology. This application will be integrated as part of the EASAIER System that aims to build a framework to integrate digital sound archives in a Semantic Web context along with sophisticated feature extraction methods.

Acknowledgments

This work is funded by the European Comission under project EU-FP6-IST-033902. http://www.easaier.org/

References

[1] Frederick Giasson and Yves Raimond. Music ontology specification. Working draft, February 2007.
[2] Dan Brickley and Libby Miller. Foaf vocabulary specification. [Online] http://xmlns.com/foaf/0.1/. Namespace Document 27 July 2005 - ('Pages about Things' Edition).
[3] Yves Raimond and Samer A. Abdallah. The timeline ontology. OWL-DL ontology, [Online] http://purl.org/NET/c4dm/timeline.owl, 2006.
[4] Yves Raimond and Samer A. Abdallah. The event ontology. OWL-DL ontology, [Online] http://purl.org/NET/c4dm/event.owl, 2006.
[5] I. Davis and R. Newman. Expression of core frbr concepts in rdf, working draft. [Online] http://vocab.org/frbr/core, 2005.
[6] Christian Bizer. D2r map - a database to rdf mapping language. In *WWW (Posters)*, 2003.
[7] C. Bizer, T. Heath, D. Ayers, and Y. Raimond. Interlinking open data on the web. In *Demonstrations Track, 4th European Semantic Web Conference*, Innsbruck, Austria, 2007.
[8] Soren Auer and Jens Lehmann. What have innsbruck and leipzig in common? extracting semantics from wiki content. ESWC, 2007.