Semantically Linking Humanities Research Articles and Music Artists

Mariano Mora-McGinity[1]*, Gary Ogilvie*[2]*, György Fazekas*[1]

(1) Queen Mary University, London
(2) Northampton Universtiy, Northampton
m.mora-mcginity@qmul.ac.uk, gary.ogilvie@northampton.ac.uk

ABSTRACT

The paper describes research carried out to provide a tool designed to offer music researchers data and resources linked to other science areas and domains. The tool gathers data from the internet and stores it semantically. Most of this data consists of publications and articles about music related issues, such as artists, styles, music tags and keywords. The data is offered to researchers in a *faceted* manner, allowing the user to navigate the data through an interface, in the hope of allowing her to discover new resources which might be of value to her research.

KEYWORDS: Semantic metadata, Research, Music, Humanities, Social Sciences, Linked data.

1 Introduction

Research in the study of popular music crosses a variety of disciplines and topic areas, ranging from those in music practice to understanding music within the remit of popular culture, gender studies, medical science and so on. Its place within a broad and extensive range of subject areas sees a need for improved information recovery. Although online databases exist that offer analysis and chart data, such as the Academic Charts[1], few offer linked academic data to the music artist. Our research looks to build a database of semantic linking that provides a centralised location for music scholars to access information across many disciplines related to their research.

2 Related Work

There is much work binding humanities and music research. For example, Korsyn critiques the methodology in music research and analyses the role external factors, such as media conditions and politics play in disrupting this research (Korsyn, 2003). Other elements, such as the truthfulness of popular music history archives in the United Kingdom are challenged (Frith, 2013, p. 11), which can disrupt the accuracy of future research in this subject area. Efforts have been made to utilize semantic web technologies to find relationships between different and possibly hitherto unrelated areas of knowledge. (Heim et al., 2010), as well as using web services to increase knowledge retrieval in the web (Marchionini, 2006). The field of visualization of linked data is large and increasing quickly. A survey on different approaches can found in (Dadzie and Rowe, 2011). Our approach to visualizing the data has been based on *faceted browsing*, an account of which can be found in (Oren et al., 2006). However, to our knowledge, little or no literature has been written where the proposed outcome is the improvement of linked data concerning the technological data methods used in music research.

3 Requirements

3.1 Research

Over the course of this short-term research project there have been three main methods of research development in requirements analysis: local interviewing, dissertation analysis and data collation through an online survey. Phase one of the research process was to identify key research areas undertaken in the field. At this first phase, ten research topics were recorded from local students across eight subject areas including sociology, gender studies, and technology and production. Information from a further nine articles was also collection from Bath Spa University's online student journal known as 'PopThink'[2]. Comparative analysis of the two sets of information highlighted six common research areas including studies on subcultures, neurosciences, and technology and production. This highlights a point of focus by initially constricting sources based on common research areas.

3.2 Common Research Areas

To distinguish which research areas are most prominent in the study of popular music it is important to analyse the collected data for multiple occurrences of these research areas. This applies to data obtained across a variety of UK educational institutions and courses. The results of this analysis can be used to develop the data model to source the correct and relevant article

[1] http://www.academiccharts.com

[2] https://sites.google.com/a/bathspa.ac.uk/popthink/home

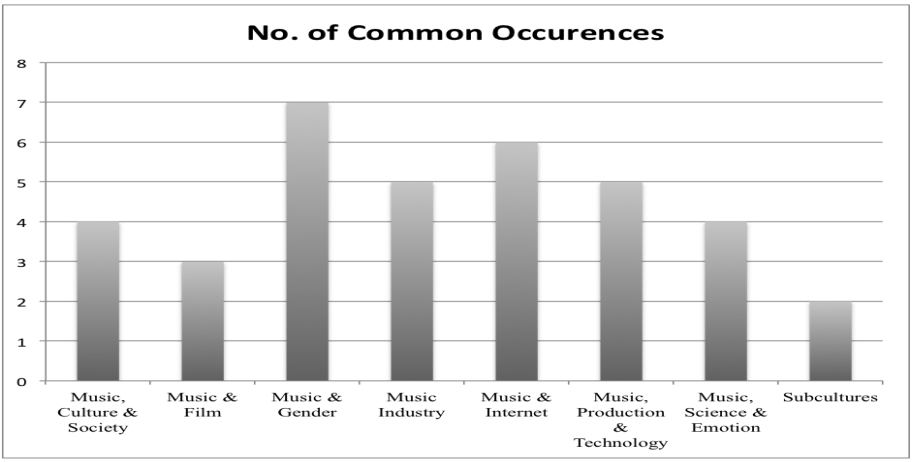

Figure 1: Common research areas

data through the means of natural language processing. Currently the project is in a proof of concept stage and article data is being acquired through web resources Mendeley and Core with the intention of including numerous other resource types in the future. Figure 1 is a chart demonstrating common research areas that have been identified so far. Although other areas of study have also been acknowledged the results below indicate those that appear in the results two or more times. From the chart it is clear that student interest in music and gender is very popular.

4 Methods

The main objective of the project is to offer researchers a hub of information resources. These resources are presented in the form of linked data, thus connecting information which might be valuable to the researcher and which might have been missed otherwise.
The following section outlines the methods employed in the research project. The outline follows the three main concerns of the project's realization: how to *acquire* data, how to *link* data with data and how to *visualize* data. We briefly discuss the first two aspects and explain with some more detail the visualization in the results section.

4.1 Data acquisition

Our main method of gathering data has been to query several services available on the web.Our main source of information about artists was musicbrainz[3], which includes an artist-id universally accepted by most other web resources, and echonest[4]. From these resources we have gathered information such as, for example, biographies, blogs that mention the artist, similar artists, tags related to the artist, etc. This data has been used to select and parse articles about each artist. The articles have been pulled also from web resources, mainly mendeley[5] and core[6]. Articles offer metadata such as authors, type of article, publication, keywords and tags, etc. all of which might be interest to researchers looking for related data.

[3] https://musicbrainz.org/
[4] http://the.echonest.com/
[5] http://apidocs.mendeley.com/home/public-resources
[6] http://core.kmi.open.ac.uk/search

4.2 Data Linking

Our semantic database links artists with research related to them. The two main ontologies used to bind these two domains together are the *Music Ontology*[7] and the bibliographic ontology *fabio*[8]. Artists are linked to similar artists and to metadata contained in articles related to them. Many of the classes and predicates can be found in the ontologies used. We created several classes and predicates to help bind relations. We have developed predicates such as *isAbout*, *isSimilarTo*, *isRelatedTo*, *isSimilarlyTaggedAs*, etc.

5 Results

Our project is aimed at researchers: our goal is to provide an interface for them which will allow them to research better, faster and in ways which they might not have thought of. The main motivation while designing the data visualization interface is to facilitate the researcher's navigation through the data. There are many conceivable ways to enter a connected graph. Our intention has been to offer the researcher different ways to begin the search and different possibilities to continue at each step of the navigation. For this reason we decided on a *faceted browsing* approach. We felt other alternatives such as network visualizations do not make explicit a categorization of relationships which we considered important for researchers. Hence the decision to present data in tree format. Figure 2 shows a tree as may be seen by a researcher. The tree's branches represent different *facets* or taxonomies for browsing new data. An instance of a branch is, for instance, a new sub-tree of journals that publish articles about that artist. Opening that node uncovers new branches of disciplines included in that particular journal, as well as all nodes to all articles published in that journal, as well as all authors who have published in that journal. The user reaches eventually a leaf with no further facets related to the original root. This leaf serves as the root of a new tree, and the user finds herself in a new domain from which to continue browsing. Such happens, for instance, when the user reaches a node containing a discipline, a keyword or a tag. At other times, a leaf will direct to an external resource, such as the website of an article or a journal. A right-hand side panel displays important article data including publication data, discipline, author, abstract and where possible a web link to the original text. Users have the appropriate amount of information to discover whether the article is appropriate to their research and to locate the article should a link be unavailable. Centralisation of relevant resources is the key element in this data model. Long names, such as article or journal titles are abbreviated in the tree. This can be difficult at first glance to identify an article that suits a particular research area. The user must click on each article to see the full title in the right panel. Also, it is currently not possible to search by keyword and therefore all articles relating to an artist, or discipline, are shown. Implementation of this is crucial to displaying not only articles relevant to the artist, but those also relevant to the research area.

[7] http://musicontology.com/

[8] http://www.essepuntato.it/lode/http://purl.org/spar/fabio

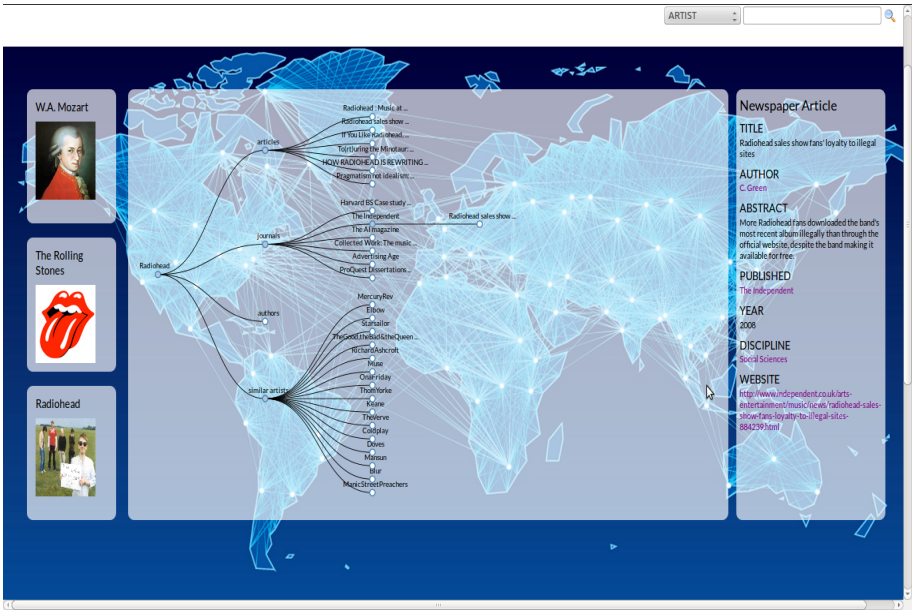

Figure 2: View of the data visualization interface

References

Dadzie, A. and Rowe, M. (2011). Approaches to visualizing linked data: A survey. *Semantic Web*, 1.

Frith, S. (2013). Writing the history of popular music. In Dauncey, H. and Guern, P. L., editors, *Stereo: Comparative Perspectives on the Sociological Study of Popular Music in France and Britain*, pages 11–22. Farnham: Ashgate Publishing Limited.

Heim, P., Lohmann, S., and Stegemann, T. (2010). Interactive relationship discovery via the semantic web. In *ESWC 2010: Proc., 7th Extended Semantic Web Conference*.

Korsyn, K. (2003). *Decentering Music: A Critique of Contemporary Musical Research*. Oxford University Press, Oxford.

Marchionini, G. (2006). Exploratory search: from finding to understanding. *COMMUNICATIONS OF THE ACM*, 49(9).

Oren, E., Delbru, R., and Decker, S. (2006). Extending faceted navigation for rdf data. In *ISWC*, pages 559–572.

Porter, S. and Whitcomb, M. (2003). The impact of contact type on web survey response rates. *Public Opinion Quarterly*, 67:579–88.

Reips, U. (2002). Standards for internet-based experimenting. *Experimental Psychology*, 49:243–56.

S.D. Gosling, S. Vazire, S. S. and John, O. (2004). Should we trust web-based studies? a comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59:105–17.