MuzkMesh: Interlinking Semantic Music Data

Mayank Singhi, Ying Ding, Yuyin Sun
School of Library and Information Science
Indiana University
Bloomington, USA
{msinghi, dingying, yuysun}@indiana.edu

***Abstract*—The vision of the Semantic Web is to lift 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. The paper discusses the MuzkMesh music portal which mashups existing semantic music data from the Linked Open Data (LOD) bubbles and other common APIs. It aims to demo the power of semantic integration and useful use scenarios on music retrieval and entertainment.**

Keywords: Semantic mashup, music data, data integration, Linked Open Data

I. Introduction

With the decade-round endeavor from the Semantic Web believers, researchers and practitioners, Semantic Web has made remarkable progress recently. It has raised significant attentions from US, UK governments, and European Commission who are willing to deploy Semantic Web technologies to enhance the transparency of eGovernment. The Linked Open Data (LOD) initiative is on their way to convert current document web into data web and to further enable various data and service mashups. The fast adoption of Semantic Web technologies in medical and life science has created impressive showcases to the world. All these efforts march a crucial step to enable the takeoff and the success of the Semantic Web.

The vision of the Semantic Web is to lift current Web into semantic repositories where heterogeneous data can be queried and different services can be mashed up [2]. The Web becomes a platform for integrating data and services. Ontology or agreed consensus is the key issue to achieve that. Especially in cultural heritage area, cross-media and cross-archival retrieval turn out to be the slogan in this area. The paper discusses the MuzkMesh music portal (http://www.muzkmesh.com) which mashups existing semantic music data from the Linked Open Data (LOD) bubbles and other common APIs. It aims to demo the power of semantic integration and useful use scenarios on music retrieval and entertainment. This paper is outlined as followings. Section 1 gives the general introduction. Section 2 shows the related work. Section 3 discusses the major functions of MuzkMesh. Section 4 illustrates its architecture and implementation. Section 5 concludes the paper and points out future work.

II. Related work

Web2.0 has revolutionized the ways of data acquisition through content syndication feeds (RSS or ATOM feeds) or Open APIs. RSS or ATOM feeds are one of the most popular public streaming data. They are the content syndication formats which are used for publishing new content regularly. They are widely used for newspaper articles (such as Nature News[1]), library updates[2], shared calendars[3], video updates[4], professional social network updates (such as LinkedIn provides RSS feeds for user's own social network updates[5]) and so on (see RSS hotlist[6]). Open APIs enable websites to interact with each other by using SOAP, Javascript and other web technologies. They are forming important trends for Web2.0 applications with the goal of integration and mashup of different data and web services. It aims to convert the current web into programmable web[7] so that silos of web data can be unlocked; different content can be remixed; productivity can be unleashed. OpenSocial [8] initiative launched by Google pioneers this area by defining a common API for social application across multiple websites including social network giants, such as MySpace, orkut, XING, LinkedIN, Plaxo and so on. Open APIs trigger massive mashup applications which seamlessly combines content from different sources into an integrated content. For example, 2Realestate Auctions website[9] mashups live eBay Real Estate listings with Google Maps so that it shows houses and lands for sale almost all over the USA with the Google Map locations for them. By mashing BBC news and Google Maps, BBC News maps [10] present their news in Google Maps. The Oscars Birthplaces Map [11] mashups IMDB data with Google Maps so that birth cities of Oscar winning best actors and actresses are mapped out on Google Maps. FOAFster[12] mashes up MyBlogLog, FOAF objects,

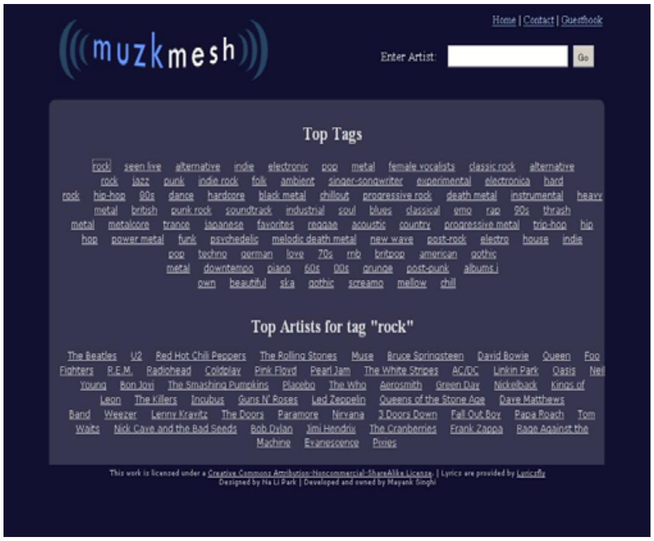
[1] http://www.nature.com/webfeeds/
[2] http://libraries.mit.edu/help/rss/barton/
[3] http://www.rsscalendar.com/
[4] http://www.youtube.com/rssls
[5] http://blog.linkedin.com/blog/2008/03/network-updates.html
[6] http://radio.xmlstoragesystem.com/rcsPublic/rssHotlist
[7] http://www.programmableweb.com/
[8] http://code.google.com/apis/opensocial/
[9] http://www.2realestateauctions.com/
[10] http://dev.benedictoneill.com/bbc/
[11] http://www.mibazaar.com/oscars/
[12] http://kentbrewster.com/foafster/

del.icio.us, Digg, Last.fm, LinkedIn, StumbleUpon, Upcoming, Tumbler, Twitter, and YouTube to provide the visualization of friend networks. EducationSearch.Net [13] enables you to search location, career, industry/salary by mashuping data from Flickr, Google Maps and YouTube. TrendyNewz aggregates and visualizes trends on Google, Yahoo and Amazon by mashuping Google, Yahoo, Amazon and YouTube data. eFanMusic[14] merges data about favorite artists, album information, mp3 songs, lyrics, biographies, podcasts and eBay by mashuping APIs from Amazon, Digital Podcast, eBay, Eventful, Google Ajax Feeds, OpenStrands, SeeqPod and YouTube.

Semantic Web glues together static data and streaming data which provides the unique contribution on linking static and streaming data. Semantic Web technologies aim to convert the current document web into data web where data are typed based on ontologies or metadata. Ontologies and the Semantic Web languages (XML, RDF and OWL) add semantics to data which enable computers to automate data exchange; federate different data sources and make some inferences. The current initiative of Linked Open Data [15] (LOD) led by W3C SWEO Community Project is one of the core initiatives in the Semantic Web area for interlinking data and knowledge [1]. LOD data grows from over 500M RDF triples in May 2007 to around 4B RDF triples in April 2009. LOD data sets vary from Wikipedia data (DBpedia, wikicompany), geo data (Geonames, World Factbook, Eurostat), news data (BBC News), government data (US Census Data, Gov-track), to music data (MusicBrainz, Jamendo, Magnatune) [3]. All these data are in RDF and data are federated via RDF Links (such as rdf:seeAlso, owl:sameAs, or foaf:knows). LOD data can be either dumped or accessed via SPARQL endpoints which are web services enabling users to query a knowledge base via the SPARQL language. Using LOD data to remix and re-connect different content and knowledge are already on the way. In 2008, Linked Open Data triplification challenge[16] has attracted many applications. For example, DBTune [17] utilizes the music data sets from LOD bubbles to create iTune-like application; LinkedMDB [18] uses DBPedia, Geonames, RDF Book Mashup datasets form LOD bubbles to create a linked movie database.

III. MUZKMESH

MuzkMesh is a mashup application that uses web services of Last.fm, MusicBrainz, Youtube, Lyricsfly and google maps and aims to help users find new artists or new songs. MuzkMesh provides two methods to find an artist and the artist's songs. The first method provides a search interface, using which any artist can be searched. The second method is presented on the homepage in the form of "Top Tags". The homepage presents top 70 tags that Last.fm users have used. When a tag string is clicked, it shows the "Top Artists" section that contains the top 50 artists that are tagged with the particular string as shown in Figure 1.



Top Tags

rock seen live alternative indie electronic pop metal female vocalists classic rock alternative rock jazz punk indie rock folk ambient singer-songwriter experimental electronica hard rock hip-hop 80s dance hardcore black metal chillout progressive rock death metal instrumental heavy metal british punk rock soundtrack industrial soul blues classical emo rap 90s thrash metal metalcore trance japanese favorites reggae acoustic country progressive metal trip-hop hip hop power metal funk psychedelic melodic death metal new wave post-rock electro house indie pop techno german love 70s rnb britpop american gothic metal downtempo piano 60s 00s grunge post-punk albums i own beautiful ska gothic screamo mellow chill

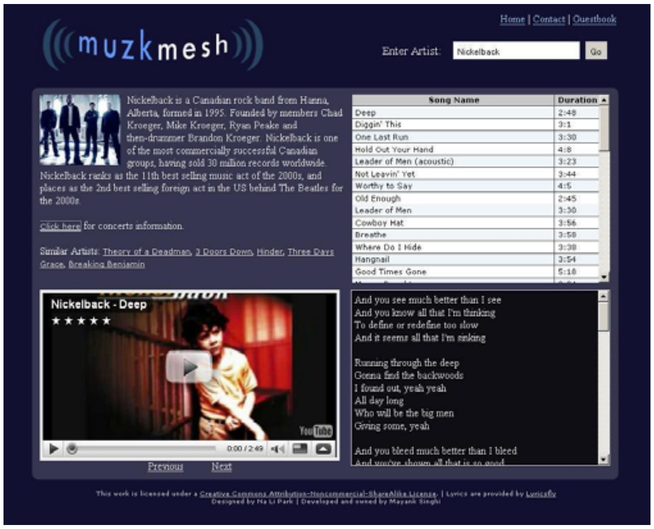
Top Artists for tag "rock"

The Beatles U2 Red Hot Chili Peppers The Rolling Stones Muse Bruce Springsteen David Bowie Queen Foo Fighters R.E.M. Radiohead Coldplay Pink Floyd Pearl Jam The White Stripes AC/DC Linkin Park Oasis Neil Young Bon Jovi The Smashing Pumpkins Placebo The Who Aerosmith Green Day Nickelback Kings of Leon The Killers Incubus Guns N' Roses Led Zeppelin Queens of the Stone Age Dave Matthews Band Weezer Lenny Kravitz The Doors Paramore Nirvana 3 Doors Down Fall Out Boy Papa Roach Tom Waits Nick Cave and the Bad Seeds Bob Dylan Jimi Hendrix The Cranberries Frank Zappa Rage Against the Machine Evanescence Pixies

This work is licensed under a Creative Commons Attribution-Noncommercial-ShareAlike License. | Lyrics are provided by Lyricsfly
Designed by Na Li Park | Developed and owned by Mayank Singhi

Figure 1. Top tags and top artists

After the artist is searched or clicked, MuzkMesh takes the artist name and finds the biography summary and random songs by the artist. MuzkMesh displays the summary and song list as shown in Figure 2. When a particular song is clicked, MuzkMesh finds the song video from Youtube and the lyrics from Lyricsfly which are displayed on the same page along with summary and songs. MuzkMesh also displays 5 similar artists and provides a link to display the upcoming concerts laid on google map (see Figure 3).



Nickelback is a Canadian rock band from Hanna, Alberta, formed in 1995. Founded by members Chad Kroeger, Mike Kroeger, Ryan Peake and then-drummer Brandon Kroeger. Nickelback is one of the most commercially successful Canadian groups, having sold 30 million records worldwide. Nickelback ranks as the 11th best selling music act of the 2000s, and places as the 2nd best selling foreign act in the US behind The Beatles for the 2000s.

Click here for concerts information.

Similar Artists: Theory of a Deadman, 3 Doors Down, Hinder, Three Days Grace, Breaking Benjamin

Song Name	Duration
Deep	2:48
Diggin' This	3:1
One Last Run	3:30
Hold Out Your Hand	4:8
Leader of Men (acoustic)	3:23
Not Leavin' Yet	3:44
Worthy to Say	4:5
Old Enough	2:45
Leader of Men	3:30
Cowboy Hat	3:56
Breathe	3:58
Where Do I Hide	3:38
Hangnail	3:54
Good Times Gone	5:18

And you see much better than I see
And you know all that I'm thinking
To define or redefine too slow
And it seems all that I'm sinking

Running through the deep
Gonna find the backwoods
I found out, yeah yeah
All day long
Who will be the big men
Giving some, yeah

And you bleed much better than I bleed
And you've shown all that is so good

This work is licensed under a Creative Commons Attribution-Noncommercial-ShareAlike License. | Lyrics are provided by Lyricsfly
Designed by Na Li Park | Developed and owned by Mayank Singhi

Figure 2. Artist's summary, songs, video and lyrics

[13] http://educationsearch.net/

[14] http://www.efanmusic.com/

[15] http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

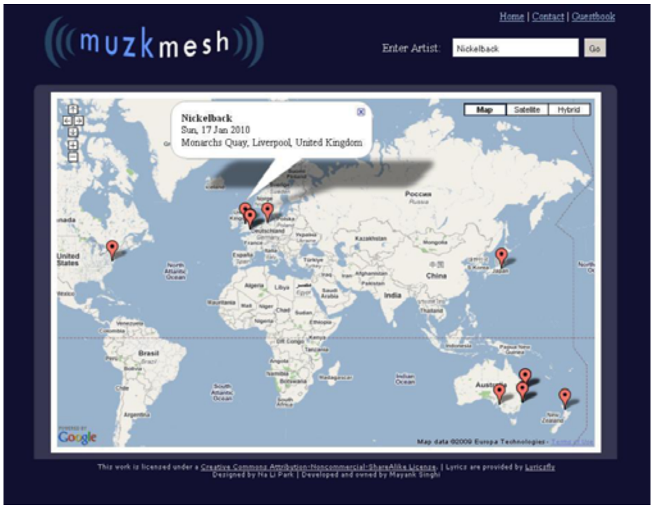
[16] http://blog.semantic-web.at/2008/07/31/lod-triplification-challenge-nominees/

[17] http://dbtune.org/

[18] http://www.linkedmdb.org/

IV. Architecture and Implementation

MuzkMesh communicates with Last.fm, MusicBrainz, Youtube and Lyricsfly through HTTP interaction. MuzkMesh sends HTTP request to the web services and receives the results in XML format. Google maps provide javascript API that can be used to map locations.



Nickelback
Sun, 17 Jan 2010
Monarchs Quay, Liverpool, United Kingdom

This work is licensed under a Creative Commons Attribution-Noncommercial-ShareAlike License. | Lyrics are provided by Lyricsfly
Designed by Na Li Park | Developed and owned by Mayank Singhi

Figure 3. Concerts displayed on a Google map

MusicBrainz provides web service which acts as an interface to the MusicBrainz database. The database contains a huge amount of music metadata, all maintained by the MusicBrainz community. The service's architecture follows the REST design principles.

The Last.fm API allows users to call methods that also respond using REST design principles. The methods, required parameters and returned result format are well documents on the Last.fm API's web page.

The Youtube's data API allows a program to perform many of the operations available on the YouTube website. It is possible to search for videos, retrieve standard feeds, and see related content.

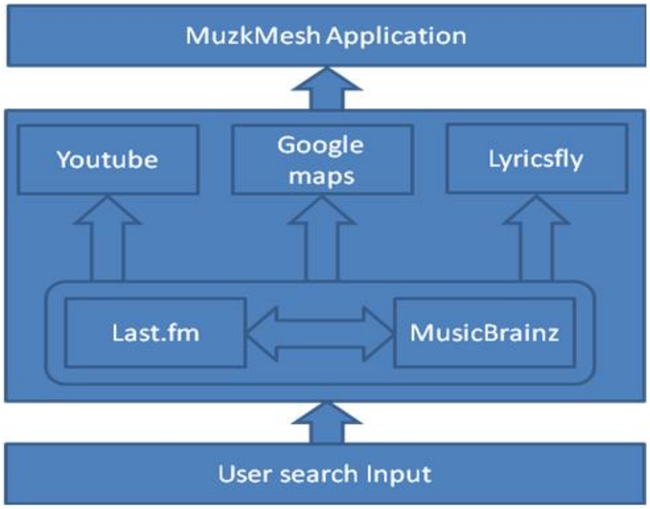


Figure 4. MuzkMesh architecture

Lyricsfly also provide an API, available for anyone's use. They allow temporary and limited access to their API by providing a temporary key. After the user conforms to their requirements mentioned on their site, Lyricsfly provides a permanent key which gives full access to the content.

When an artist is searched or clicked, MuzkMesh takes the artist name and sends an http request to Last.fm to search for the artist. Last.fm sends the result in XML format. From the result, only the first matched artist name and MusicBrainz id (mbid) is used and rest are discarded. The mbid is a unique number given to an artist by MusicBrainz. Using the mbid, a request is again sent to Last.fm to get artist's biography summary, similar artists and events information. The mbid from Last.fm is also used to send a request to MusicBrainz to receive the tracks list of the artist. Each track name is converted to a hyperlink. When a track is clicked, MuzkMesh constructs a query "<artist name> <song name>" and issues a request to Youtube and Lyricsfly to search for the video and lyrics respectively. Youtube serves the results in a RSS feed and only the first found result is displayed. Lyricsfly delivers the result in XML format and the first found lyrics are displayed.

V. Evaluation

Evaluation was based on the feedback from five users. The users mentioned that the song list, retrieved from MusicBrainz was not updated and did not include lot of new songs. The Youtube search did not perform very well as sometimes MuzkMesh failed to retrieve any video for some lesser known songs or artists. Overall, Youtube's search precision was high but the recall was poor. Lyricsfly also sometimes failed to retrieve the lyrics of the selected song. The reasons for Youtube and Lyricsfly's poor search could be that video and lyrics do not exist for a particular song or the query we construct is not proper.

VI. Conclusion

This paper discusses the demo of mashing up different music data or services. There remain several improvements that can further enhance the functionalities of MuzkMesh. The first improvement that we need is to retrieve a list of songs that includes new popular songs as well as old songs. This can be done with the help of Last.fm. Last.fm provides a method to get the popular songs by an artist. The popular songs from Last.fm combined with the list retrieved from MusicBrainz would make a comprehensive list of songs. The search performance also needs to be improved when looking for the video and lyrics. One prospect to improve the search could be to make a two stage search. In the first stage, we keep our original query of "<artist name> <song name>" as it is. If there is no video and/or lyrics found, then we go to stage 2. In stage 2, the query string could be changed to only artist's name or only song's name and search again. We need to do further study on how much the precision and recall would be affected by this method.

A future feature will be to enable users to add a song as their favorite and play the favorite videos at a later time in the form of a playlist. We will construct a RDF based database to store user's favorite songs, videos and lyrics.

REFERENCES

[1] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak and S. Hellmann, “DBpedia - A Crystallization Point for the Web of Data,” Journal of Web Semantics, Special Issue on the Web of Data, 2009, pp.164-165.

[2] C. Bizer, T. Heath, T. Berners-Lee, “Linked Data - The Story So Far,” Journal on Semantic Web and Information Systems, Special Issue on Linked Data, 2009.

[3] Luger, M., Ding, Y., Scharffe, F., Duan, R. and Yan, Z., “EASAIER Semantic Music Retrieval Portal,” Poster at the 2nd International Conference on Semantics And Digital Media Technologies (SAMT), Genova, Italy, Dec 5-7, 2007.