Final Document

Xinformatics2011\_Porject\_Orange

May 3, 2011

Team Orange’s Cross-Topix

Bridging Music Silos using Semantic Technologies

X-Informatics Final Project

Spring 2011

Professor Peter Fox

Tetherless World Constellation

Rensselaer Polytechnic Institute

Team Members

Amanda Olyha- Senior, Physics and CS

Samuel Johnson - 4th semester (part time) MS IT, HCI Concentration

Tim Lebo - 2nd Year Ph.D. Cognitive Science, Tetherless World Constellation

Colin Anderson- Second semester MS in IT, MIS concentration

Yu - 1st Year PhD ECSE, Tetherless World Constellation 518-522-7669

Introduction

Choosing public domain music scores that are appropriate for instruction and performance can be challenging. Quality of score, related genres, compositions by composers are just a few issues that can arise. Therefore, our group chose to focus on developing a system dealing with library informatics. Our group wanted to use a similarity heuristic combined with crowd sourcing to combine webpages on two separate sites. Through the development of a “Social Machine” and a “See Also” box, musicians are able to find similar pages on different sites based on multiple parameters. It adds value to the user allowing them to easily find the same music

1. Use Case Development

Appendix

[**A. Project Pre-Definition** 2](#_Toc291191343)

[**ii. Evaluate system ideas** 2](#_Toc291191344)

[**PROPOSAL E.1** 2](#_Toc291191345)

[But perhaps "evaluating an existing system" would be easier b/c it is concrete and we can poke it. Perhaps we could evaluate the VIVO system. http://vivoweb.org/ It is a professional research "social networking" infrastructure that uses RDF as its representation. 2](#_Toc291191346)

[**Use case ideas** 2](#_Toc291191347)

[**PROPOSAL U.1** 2](#_Toc291191348)

[**Meetings** 2](#_Toc291191349)

[**Resources** 2](#_Toc291191350)

[**B. Meeting Notes** 3](#_Toc291191351)

[**C. Proposals** 4](#_Toc291191352)

[Tim defined objectives for proposals Done: everyone write 3 paragraphs proposing what we do. (Due Sunday evening) - for review by team on Monday. 5](#_Toc291191353)

**A. Project Pre-Definition**

**Initial Thoughts**

* This is our group project for <http://tw.rpi.edu/web/Courses/Xinformatics/2011>
* Slide 6 of <http://tw.rpi.edu/media/latest/Xinformatics2011_week6.ppt>
* Analysis of existing information system content and architecture, critique, redesign and prototype redeployment
* Pursuit of a detailed use case around a particular area of informatics includes developing a prototype IS, architecture, design, etc.
* Due May 3 (write up) and May 10 (presentation)
* That’s 7 (8) weeks
* Check in on progress in 3 weeks
* Did Peter outline what we would need to "evaluate", or would we come up with that criteria?

**Evaluate system ideas**

**Proposal 1**

But perhaps "evaluating an existing system" would be easier b/c it is concrete and we can poke it.  
 Perhaps we could evaluate the VIVO system. http://vivoweb.org/  
 It is a professional research "social networking" infrastructure that uses RDF as its representation.

**Use case ideas**

Proposal  
 Tim ran into some poorly managed MRI handling that we could develop a use case for.

**Meetings**

Interest in music.  
 Amanda is the orchestra librarian.  
 Safari books - hard to navigate. Went flash and it is terrible.

**Resources**

Sheet music - physical stuff. Copyright.  
 Libraries  
 <http://www3.cpdl.org/wiki/>  
 <http://imslp.org/wiki/>  
  
 LOD cloud of music. Music Genome Project (Pandora)  
  
<http://richard.cyganiak.de/2007/10/lod/lod-datasets_2010-09-22_colored.html>  
  
  
[<http://tw.rpi.edu/instances/TimLebo>](about:blank) a foaf:Person .  
  
<http://tw.rpi.edu/web/person/TimLebo>  
  
<http://validator.linkeddata.org/vapour>  
  
  
<http://en.wikipedia.org/wiki/Composer>  
  
<http://dbpedia.org/resource/Composer>  ←----  
  
validating dbpedia’s URI for composer:  
  
http://validator.linkeddata.org/vapour?vocabUri=http%3A%2F%2Fdbpedia.org%2Fresource%2FComposer&classUri=http%3A%2F%2F&propertyUri=http%3A%2F%2F&instanceUri=http%3A%2F%2F&defaultResponse=dontmind&userAgent=vapour.sourceforge.net  
  
<http://www3.cpdl.org/wiki/index.php/Emanuele_d%27Astorga>  
  
2 questions to ask:   
1) fitting requirements of project  
2) can we do it in a month.

**B. Meeting Notes**

**A history of meetings Orange Group - Spring 2011**

Meeting notes: Tuesday 2011 April 05

Discussed the Possibilities of the Project  
Instructor: time period, “no half notes”  
independent musician: I want sheet music. arrangements.  
create an ontology? keep it small  
emotional component - music at fourth of july, at a wedding.

Meeting notes: Thursday 2011 April 07

Tonight’s Goals   
1. Subset of Wiki  
2. Develop example content  
3. How search that?  
4. Small model  
  
 1. We want to Facilitate Browsing  
 2. Be able to address number of instruments  
  
Browsing, NOT search.  
Don’t know composer, style, piece.  
Upbeat music for marching band.  
Given a set of instruments, what works?

Meeting Notes, Thursday 2011 April 07

Define Project- much debate and Tim draws nice

Meeting Notes, Thursday 2011 April 14

Discussed progress of work  
Set up goals for the night  
 Yu - Make various diagrams for the project  
 Colin and Amanda - Develop use cases  
 Tim - Scraping of two Wiki’s

Samuel – Started the joining of the Wiki – complained at size of file. Almost a throw down over file size

Meeting Notes, Thursday 2011 April 21

Discussed progress of work

Yu – created all diagrams for the project, created Sparkle endpoint

Amanda – Created Use cases, ran two cases for similarities

Colin - developed use cases

Tim – Organized the data set management design, reviewed and modified document, triple store installation with Sir Patrick West, coordinated Sparkle with Sir Patrick West and future Doctor Chen (Tim is our master)

Samuel – Finished implementing the joiner, and developed a read me (very lengthy assignment)

Tonight’s work

Yu – Testing triples on Sparkle

Amanda – Coming up with histogram of similarity measure

Colin – Taking notes, documenting work, beginning paper

Tim –

Samuel -

**C. Proposals**

Tim defined objectives for proposals  
Done: everyone write 3 paragraphs proposing what we do. (Due Sunday evening) - for review by team on Monday.

* What resources we will use
* What benefits we will provide.
* How we conform to each request in the assignment.
* What technologies we will use
* The kind of user would benefit
* How what we’d do fits into the topics discussed in class
* what deliverable we will provide

**Proposal by Yu:**  
We are trying to redesign the digital music library, after analyzing a set of online music libraries, leveraging the social, cognitive and domain concerns that could make a music information system much easier accessible for the music fans. The music libraries that we reference are those such as IMSP-Pertrucci Music Library, Free Choral Sheet Music etc.  
  
The reason why we redesign the online music library is based on the observation that the pages are information-overloaded and the navigation is neither intuitive nor effective.  These are something we are trying to modify and optimize. In a word, we will redesign the music information system by providing more user-friendly interface that better navigate the user to the piece of information that required. We will also make better arrangement of the presentation of the pages such that the user could immediately get what the pages tells them in a short time.  
  
As we see the requirements in the assignment, we would definitely draw the diagram of the conceptual model again that better illustrate the infrastructure of all the information system is found in components. What’s more, we will implement a simple prototype of the system within web browser to illustrate our solutions.  
  
Technologies that we might use are Flash and Action script, PHP etc. All that related to UI design might help. (I independently designed and implemented several Flash and Action scripts apps before, hope it could help in designing our new UI )  
  
The users that could benefit from our design could be more than the professional musicians. Music lovers without so much domain expertise could also find the scores or recordings.  
  
To fit into the topics of the materials covered in class, we need definitely consider cognitive, semiotics and social concerns towards a good information system. Therefore, we each might need to be responsible for a particular section of the concerns mentioned in class. In implementing the system, we should collect the suggestions from each of us and realize the functionalities accordingly.  
  
A deliverables that I could think of is a demo Flash application or a set of linked webpages, which shows how user could better get the information according to our re-newed schemes.  
  
**Proposal by Sam:**  
I actually found what Yu wrote to be in line with what my thoughts were, so I’m going to attempt to extend and flesh-out what he wrote, rather than attempt to come up with something completely new (my changes/additions in *italic*):  
  
**Proposal by Yu (*with extensions/additions by Sam*):**  
We are trying to redesign the digital music library, after analyzing a set of online music libraries, leveraging the social, cognitive and domain concerns that could make a music information system much easier accessible for the music fans. The music libraries that we reference are those such as IMSP-Pertrucci Music Library, Free Choral Sheet Music etc.  
*I’d propose focusing on professional or amateur musicians, rather than music fans. The typical music fan isn’t interested in the sheet music itself so much as the product of that sheet music. Because the libraries we discussed on Friday are libraries of sheet music rather than recordings of musical performances, I believe that we should assume any users of our redesigned system would be musicians intending to use the sheet music, rather than fans looking for music to “consume.”*  
  
The reason why we redesign the online music library is based on the observation that the pages are information-overloaded and the navigation is neither intuitive nor effective.  These are something we are trying to modify and optimize. In a word, we will redesign the music information system by providing more user-friendly interface that better navigate the user to the piece of information that required. We will also make better arrangement of the presentation of the pages such that the user could immediately get what the pages tells them in a short time.  
*I’m less concerned with the interface itself than the way the information is tagged and organized. We’ll need to come up with a few use cases specifying the reasons a musician comes to one of these sites looking for sheet music, and focus on helping that user achieve his or her goal.*  
  
*As a semi-professional musician myself, I can both add to our domain knowledge in creating these use cases, as well as provide contacts to professionals in the area who rely on the databases as they currently exist. That will help us when we create the use-cases the project requires.*  
  
As we see the requirements in the assignment, we would definitely draw the diagram of the conceptual model again that better illustrate the infrastructure of all the information system is found in components. What’s more, we will implement a simple prototype of the system within web browser to illustrate our solutions.  
*Again, my (and Amanda’s) experience within the domain will be useful in creating a set of diagrams modeling the information. This will be particularly interesting when contrasted with people who have little experience with music--the combination could potentially be very effective.*  
  
*There are any number of different potential attributes for any given music, and I think that will give a rich complexity to the problem that will make it challenging. While that’s what makes it worthwhile, we should also be aware that it won’t be straightforward. I’m interested to talk more about what Tim started talking about, because I’m not totally sure I understood it. For any given score, there’s an immense amount of available meta-data that could be associated with it, and finding ways to make that meta-data accessible by the users, whether through better searching, or better organization, would be a good direction to go.*  
  
Technologies that we might use are Flash and Action script, PHP etc. All that related to UI design might help. (I independently designed and implemented several Flash and Action scripts apps before, hope it could help in designing our new UI )  
*I don’t have any knowledge of Flash or ActionScript (I’m too cheap to pay for Adobe’s development tools); my UI language of choice is JavaScript, but I’m sure we can find a way to sort out that difference.*  
  
The users that could benefit from our design could be more than the professional musicians. Music lovers without so much domain expertise could also find the scores or recordings.  
*As I said above, I don’t know that I agree with this, but I’d be interested in hearing more of how a non-musician would make use of musical scores. We need to be clear about whether we’re including recordings within our domain. I believe the only actual sound files that are available on these sites are MIDI versions of the musical scores.*  
  
To fit into the topics of the materials covered in class, we need definitely consider cognitive, semiotics and social concerns towards a good information system. Therefore, we each might need to be responsible for a particular section of the concerns mentioned in class. In implementing the system, we should collect the suggestions from each of us and realize the functionalities accordingly.  
  
A deliverables that I could think of is a demo Flash application or a set of linked webpages, which shows how user could better get the information according to our re-newed schemes.  
*We might also want to consider what Tim talked about i.e., finding ways to integrate these databases into the existing semantic web, which would enable much broader use by the academic community, as well as the application of existing informatics tools and techniques in the searching/browsing of the musical scores. I’d be interested to hear from Tim more about how that could work.*  
  
**Proposal by Amanda:**  
  
I agree with most of what has been said above, but there are a few things I would like to add:  
  
Like Sam had said above, I think the users of this redesigned system will probably be for professional or amateur musicians, although we may not want to only limit ourselves to this.   We can determine the specifics once we create our specific use cases.  
  
With this in mind, we should have a strong, well-defined search use case planned out since it will be one of the main points of entry to the system.  Along with a search, I think the browse options should also be well-designed.  Because of all the metadata that we could potentially use, browsing can be a very strong feature.  
  
**Colin’s addition:**  
  
I agree with much that has been said above so I will add my two cents for what its worth.  
  
I agree with Sam and Amanda that the functionality of the system should be our main concern. However, I would not disregard the UI as extraneous. Yu’s expertise here could make the system more popular and easier to use for all involved. Before we accept or disregard technologies to use as well, we need to get together and decide what is best for the project.  
  
I think we could look at this as an educational tool as well. We are focusing on amateur and professional musicians, but we could also include music instructors (choral, band, orchestra) to the mix as well.  
  
I am concerned with the “Use Case” itself. Reading the instructions for the final assignment, the use case needs to be be well defined and the scope of this project needs to very tight. Based on what I have read, and what I thought the use case should be, we really need to nail that down at Monday’s meeting.  
  
  
**Proposal by Tim:**

* Steps
  + Install a local Mediawiki
    - Install the Semantic Mediawiki (SMW) extension
    - copy several sample values to local wiki
    - annotate using SMW markup
    - associate annotated pages to LOD cloud and ontologies
    - demonstrate local queries
    - import appropriate ontologies (including FRBR)
    - access the data as RDF dumps.
  + Contact existing wiki owners and convince them to install the sem media wiki extension
    - demonstrate markup
    - train community
    - demonstrate queries
    - ask them what problems they have and document them.
  + explore LOD music
    - establish linking among existing cloud data via the sem media wiki
    - accumulate some of the LOD into a sparql endpoint
    - accumualte some of the wiki RDf into same sparql endpoint
  + crawl existing mediawikis and try to grab some structure
    - encode as RDF establishing URIs as if they had a SMW installation.
  + index physical sheet music
    - don’t publish, just provide metadata and pointers to where it is phsyciall and whom to contact.
  + Additional UI
    - javascript?
    - actioncript?
* Resources:
  + a server machine with admin access to host web stuff
  + SMW software
  + RDF tools/crawlers
  + Music ML?
* Benefits:
  + answering new questions
  + person finding music would benefit
  + linking to Library of Congress?
  + I like Sam’s idea of focusing on musicians proper.
* Deliverables
  + use case documentation
  + prototype system
  + people saying we are awesome and helped them.
  + music ontology