Semantic-Annotating of Research Lab Activities for Data Sharing (revised June 2015)

Erick Garcia[1,2], Supervised by: Dr. Natalia Villanueva Rosales[1,2]

[1]Department of Computer Science, [2]Cyber-ShARE Center of Excellence

egarcia87@miners.utep.edu, nvillanuevarosales@utep.edu

The University of Texas at El Paso

Abstract— The dissemination of products resulting from research efforts for validation, reproducibility, and reuse is an important task of a research group. Using web-based technologies facilitates this task. In this project, we will describe the process of generating the front-end of a system that accesses the knowledge base of the iLink research group. This front end uses frameworks compliant with web standards for publishing and annotating data. The frameworks that were used are Bootstrap and Code Igniter. While Bootstrap is mainly used for the front-end design of the website, Code Igniter's main purpose is the back-end data integration, which means the website is populated dynamically and integrates knowledge from various sources. The current front-end includes a website that makes uses of metadata from different sources to display the data for humans. Current work involves the creation of an interface (i.e. web-services) to allow ingestion of our annotated data by machines (i.e. software agents). Website: http://ilink.cybershare.utep.edu/.

Index Terms— **Metadata** is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. It is really important because it facilitates discovery of relevant information [2]. The **Semantic Web** is a standard promoted by the World Wide Web Consortium which aims at converting the World Wide Web from the current model largely unstructured documents and data into a common machine understandable framework that allows data to be shared and reused [2]. **Web-Services** are client and server applications that enable the communication between two electronic devices over a network [1].

1 Introduction

The main motivation for this project was that maintaining information about research group efforts is a tedious manual task. Sharing resources using natural languages, like English, are good for humans but hard for computers to process. Sharing resources using machine languages, like JSON, are great for machines but hard for humans to understand. Having a one stop place where resources of research groups can be dynamically populated from other sources and share information for both, humans and machines, would save time and also increase reusability of their resources.

2 Frameworks For Web Design

2.1 Bootstrap

Bootstrap is a front end framework that is full of features, javascripts, CSS for buttons, tables, fonts, etc. What is really useful of this framework is the grid system it features. It works assuming that you only need 12 columns for your web design and within those 12 rows and some containers you insert all the information you want the webpage to display. The grid system features adaptive and responsive columns that will change depending on the dimensions of your device. Using that and even specifying different behaviors for large, medium, small and extra small devices accomplished the HTML5 support for a mobile friendly website.

2.2 Code Igniter

Codeigniter is a powerful php back end framework that is elegant and simple with simple URI's. With this framework that uses the MVC (Model-view-controller). A model that directly interacts with the database, the view is that the user will see and what the developers will render, finally the controller will interact with the model and also if it is permitted the user will also be able to interact with the controller via JavaScript or get and post methods. The controller will send the parameters to the model. That is a very efficient method of getting a really neat, compact, and

organized code. Since the coding of the webpage was done by me and colleague Lorna Bustillos we needed a clear code and documentation which is another reason why we chose this framework.

2.3 Source Tree

As said previously the webpage was not done by just one person, so we used this software to collaborate and exchange code. This is a great user interface that interacts with the git repositories of iLink @Cybershare.

3 Semantic Annotations

Schema is a vocabulary that makes it easier for search engines to parse and interpret the information on your web pages more effectively so they can serve relevant results to users based on search queries. It can be used with many encodings, including RDFa, Microdata and JSON-LD. (Fig. 1)

3.1 Schema.org

Schema is a vocabulary that makes it easier for search engines to parse and interpret the information on your web pages more effectively so they can serve relevant results to users based on search queries. It can be used with many encodings, including RDFa, Microdata and JSON-LD. [5]

3.2 Microdata

Microdata is a new HTML5 specification that allows machine-readable data to be embedded in HTML documents. In other words, microdata helps computers understand what the contents of your web pages are saying. [5]

3.3 RDFA

RDFa is an extension to HTML5 that helps you markup things like People, Places, Events, Recipes and Reviews. Search Engines and Web Services use this markup to generate better search listings and give you better visibility on the Web, so that people can find your website more easily.

4 Data Integration and Ontologies

Data integration is the process of combining data from different sources. Data exchange refers to the process of transforming data that is structured under a source schema to data structured under a target schema that represents the source data accurately. There are various data exchange languages like RDF, XML, and JSON. Through ontologies describing a domain and the relationships of objects. Working together with research colleague Ana Villegas who is working on extracting the information of iLink from bigger ontologies so we can then integrate it to our source which is the website we are working on. Several presentations were done throughout the semester by our graduate colleagues which improved our knowledge on these areas.

5 Objectives

Design a web-based application for the iLink research group that can be:

1. Dynamically populated from heterogeneous sources,
2. Mobile-friendly for human consumption,
3. Use web standards for interoperability of other systems, and
4. Share resources using semantic annotations (meta-data) for machine processing.

6 Methodology

1. Design of a high-level data model (E/R diagram) that covers information retrieved from other sources and metadata (Fig. 2).

2. Evaluation for frameworks for web development that supports the creation of responsive mobile-, human-friendly website and the use of standards for interoperability

3. Develop the front- and back- end of a research group website through the use latest web-based technologies.

4. Comparing metadata standards and vocabulary to share research group's resources with semantic annotations.

5. Annotate semantically the information shared through the website.

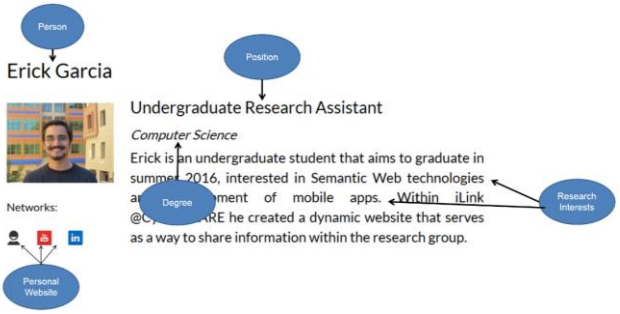

Figure 1. Differences between data that can be understood by humans vs. data that can be understood by machines.

6. Evaluate the web-based system.

7 Conclusion

The iLink research group's dynamic website was created. This website is: i) dynamically populated from other sources, ii) based on Model-View-Controller architecture, iii) based on standard web-languages and technologies including Bootstrap, Code Igniter, HTML5 and JSON.

The information shared on the iLink's website is prepared for semantic annotations to be fully understandable for both humans and machines. We are currently identifying controlled vocabularies to describe the information on the website, including Friend of a Friend [3] and schema.org [5]

Acknowledgment

I would wish to thank Doctor Natalia Villanueva for her supervision and her role as a mentor to this project. This work was supported by the National Science Foundation under Crest Grants HRD-0734825, HRD-1242122 AND CI-Team Grant OCI-1135525.

References

[1] Understanding Metadata (pp. 4-10). (2004). Bethesda, NISO Press National Information Standards Organization

[2] Semantic Web, Retrieved April 15, 2015 from http://www.w3.org/standards/semanticweb/data

[3] FOAF, Retrieved April 15, 2015 from http://www.foaf-project.org/.

[4] iLink Website: http://iLink.cybershare.utep.edu/.

[5] Schema.org, Retrieved April 15, 2015 from http://www.schema.org/.