Exploring the Use of Semantic Technologies for Big Data Integration - Final Report

CREU Final Reports 2015

Student Researchers: Rebecca Little
Kristel Basra
Ashley Mannon

Advisor: Srividya K. Bansal

Institution: Arizona State University

Webpage: http://creu-asu.weebly.com

Goals of the Project

The goal of this project is to explore the use of semantic technologies to connect, link and load data into a data warehouse. The specific objectives included:

(i) creation of a semantic data model via ontologies to provide a basis for integration and understanding knowledge from multiple sources;
(ii) creation of integrated semantic data using Resource Description Framework (RDF) as the graph data model;
(iii) extracting useful knowledge and information from the combined web of data for use in innovative applications.

More specifically this project addressed 2 domains of data:

(i) data from video game platforms about users that includes badges, trophies earned, time spent on playing a game, etc.
(ii) real-estate and crime rate data.

Process

The project was conducted in the following phases:

(i) Literature Review: this phase involved reading and discussing research articles in the field of data integration, identifying public datasets available in various domains, identifying innovative apps that could improve quality of life, and learning semantic technologies such as OWL (Ontology language), Protege (OWL editor), RDF (Semantic data format), and SPARQL (Semantic Query language).

(ii) Requirement analysis: At the beginning of this phase two domains were picked for the research project - home buying advisor system and video game suggestion system. A survey monkey survey was conducted to understand some of the most important factors one would look for when
buying a house. A competitive analysis of various game recommender systems was done and categorized into online quizzes, forums, and suggestion software.

(iii) Ontology design: This phase involved reusing existing schemas and ontologies and extending them to create ontologies for the two domains: Videogames and user metadata; and Real estate and Crime data. The ontologies were designed and created in OWL language using Protege visual ontology editor.

(iv) Semantic Data generation: this phase involved converting public datasets found into semantic data in Resource Description Framework (RDF) format.

Conclusions and Results

Videogames and recommender systems represent a cross-section of areas in which women are underrepresented. While women make up nearly half of all gamers, they compose less than five percent of videogame programmers and about twenty-five percent of general programmers. By identifying areas in which gender disparity exists and discovering ways to view users as a personality rather than a demographic, this project helped to advance gender equality in the fields of both computer science and videogames. It also addressed the relatively overlooked issue of demographic stereotyping in recommender systems in general.

Moving to a new home is a difficult process especially getting information about the neighborhood and its safety. With the use of data integrated from multiple domains, a variety of information can be collected about a house, its surroundings, amenities, and the value of the area. Our survey results show that home buyers consider crime rates and home price range as the top factors influencing their home buying decision.

This project resulted in the design of 2 ontologies: the first combines information about video games from GamesDB.net and user metadata from Steam gaming platform; the second combines real estate data and crime data from a public data provider called Quandl.

Future work includes creation of a video game recommender system and home buying advisor application using the ontologies and semantic data generated from this project.

One of the students on this project has decided to pursue higher studies as a result of this project and might continue this work towards a Master’s thesis.

Presentations and Publications

- GHC 2015 Poster (Under review): Using Semantic Technology to Address Gender Stereotyping in Videogame Recommender Systems by Rebecca Little
- Tapia 2016 Paper (in preparation)