Using Semantic Technology to Address Gender Stereotyping in Videogame Recommender Systems

Rebecca Little, ASU Polytechnic

CRA-W CREU End of Year Report

1. INTRODUCTION

'Big Data' is the now-ubiquitous term used to describe the massive amount of data available in nearly every domain. Big Data has a myriad of applications that are literally revolutionizing the way many industries work, but due to the unstructured and even inaccessible nature of most datasets it can be impossible to create applications that leverage Big Data's power without the development of tools to find, transform, analyze, and visualize data [1]. Organizing data in a meaningful way is thus an open challenge in Big Data. One approach to Big Data integration is the use of semantic technologies, in particular data categorization using ontologies. The idea of enhancing knowledge acquisition through the structuring of data is not new; see [2] and early proposals for the world-wide web which included the vision of a Web of Data, or the Semantic Web. Applying this old idea to the new problem of Big Data is a viable approach to better prepare data for decision-making purposes.

One particular area in which an explosion of data has occurred is in entertainment (movies, books, television shows, videogames, etc.).The field of recommender systems was created as a response to the huge amount of content now available to consumers. One of the first and most popular methods of designing recommender systems was the ratings-based method, but in the past decade several systems have integrated the use of ontologies in order to utilize content metadata other than ratings. The use of semantic technologies has been instrumental in addressing some of the identified issues with recommender systems, such as limited content analysis, overspecialization, and the new user problem [3]. This project addresses a subset of overspecialization in videogame recommender systems: stereotyping of demographic segments leading to overspecialized and even inaccurate videogame recommendations. For example, a middle-aged woman will be recommended a puzzle/matching game like Bejeweled, while a teenage boy will be recommended a first-person shooter like Call of Duty, regardless of their respective personal preferences.

2. BACKGROUND AND RELATED WORK

There are currently three subsets of videogame recommender systems. The most primitive yet pervasive are online quizzes, usually generated by online quiz sites. These quizzes are trivialized by the goal of entertaining users and a limited amount of data concerning both users and content. The second subset is online forums like [4], wherein users can request as well as provide suggestions to other users in a thread-based format. While the suggestions are generally very accurate due to the heuristic nature of forums, this method lacks scalability, usability, and accessibility. The final subset is suggestion software; GamesLikeFinder.com is a hand-selected collection of game recommendations which sacrifices scalability by not leveraging Big Data, while TasteKid.com is a suggestion site built for all entertainment content which follows the ratings-based approach and thus suffers from limited content analysis and overspecialization. Videogame suggestion software which successfully integrates Big Data without stereotyping demographic segments does not currently exist.

There are several examples of the use of semantic technologies to improve recommender systems; they are in a wide variety of application domains, although none in the videogame domain. [5] uses content metadata pulled from a domain ontology to enhance the quality of discovered patterns. [6] overcomes overspecialization by applying reasoning techniques that makes their system flexible and ultimately allows it to offer accurate, enhanced suggestions. [7] produces recommendations using the semantic

information of items and user demographic data. While these examples successfully address some of the issues named in [3], there is not literature on recommender systems that address demographic stereotyping.

3. APPROACH AND UNIQUENESS

Instead of collecting demographic data about users, like gender and age, the author proposes to collect user metadata using the social network APIs of videogame platforms. [8] suggests using metadata such as badges and trophies earned and amount of time played to generate a better picture of a user than purely demographic metrics. This approach has not yet been applied to videogame recommender systems; the author would like to discover whether this approach can be expressed using an ontology in order to improve videogame suggestions by associating attributes pulled from playing metrics with users. As an example, user Mary has played the game Destiny for Playstation 4 for 35 hours; she is associated with the attribute <playtime = 35> for <videogame = Destiny> on <platform = Playstation 4. This user is thus strongly correlated with Destiny, which has the attribute <genre = RolePlayingGame FirstPersonShooter>. Mary will be recommended games that are also role playing games and first-person shooters.

This project was completed in three phases during the Spring of 2015. The first phase entailed creating a videogame ontology and the second entailed creating a user metadata ontology. The final phase involved integrating the two ontologies using semantic reasoning.

3.1 First Phase: Videogame Ontology

There exist several large videogame databases which expose their data through APIs. The author chose TheGamesDB.net, a database with more than 25,000 games, because of its accessibility and ease of use, although future work may include pulling from multiple databases. The author created a custom ontology, VideoGame, based on Schema.org's CreativeWork type. Even though Schema.org provides a VideoGame type, the author departed from it because (a) Schema does not use the OWL format and (b) there was a lot of disparity between properties defined by Schema.org's VideoGame type and the metadata provided by videogame databases. The ontology displayed in Figure 1 correlates exactly to much of the videogame data returned by the chosen videogame database.

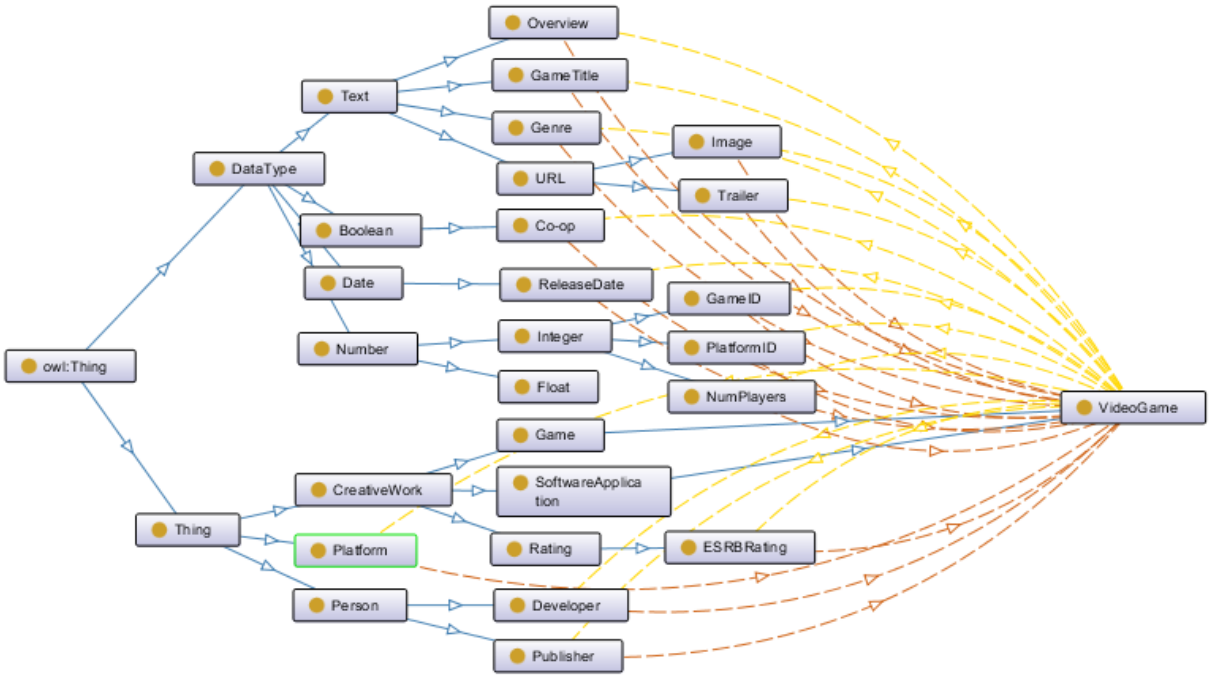

Figure 1: VideoGame Ontology

3.2 Second Phase: User Metadata Ontology

The author has identified Steam as the videogame platform social network to be used to obtain user metadata. Steam is a network for PC games, and it is the only platform for which an official API is provided. APIs for Playstation Network and Xbox Live exist, but are not official; therefore, they do not allow for username verification and are not as fully supported. Future work may include pulling from these APIs as well in order to cover a larger cross-section of users, namely users of the two major gaming consoles. As per the Steam API terms of use, the application is not allowed to directly ask the user for his or her Steam username. Instead, it implements the OpenID OAuth 2.0 decentralized authentication platform, which allows users to login to Steam without ever revealing their username to the hosting application. The usernames are verified as being that of the particular user and a 64-bit Steam ID is returned; this ID is then used to retrieve user metadata from Steam's ISteamUser API. Two API requests have been identified as useful for this application: GetPlayerAchievements, which returns a list of achievements for the passed user; and GetOwnedGames, which returns a list of games a player owns along with some playtime information. These two requests were used to construct the user metadata ontology displayed in Figure 2.

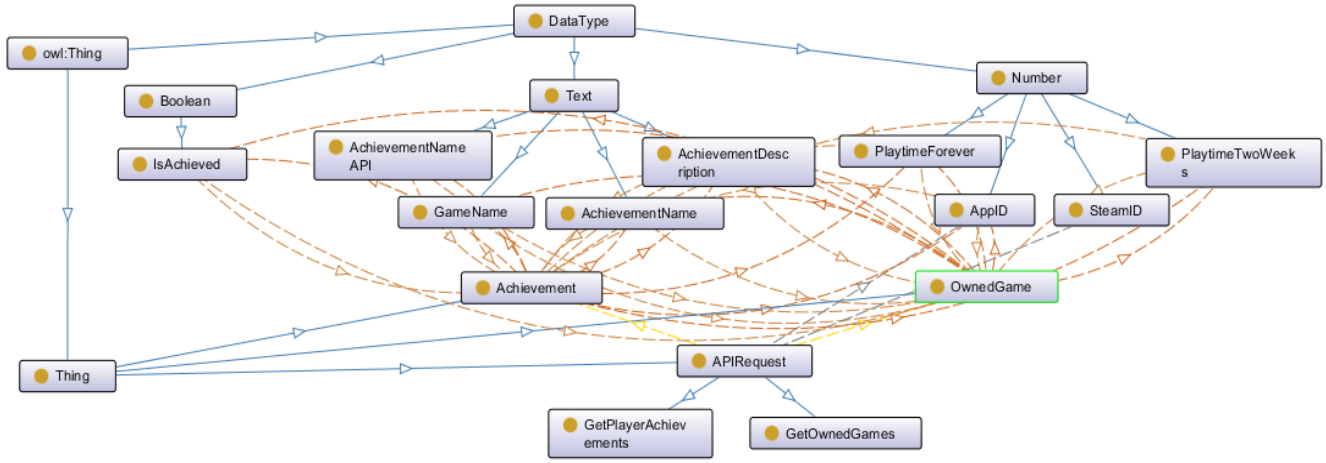

Figure 2: User Metadata Ontology

3.3 Third Phase: Integration

The data from both TheGamesDB.net and the Steam API is returned in XML format, which is converted to RDF form via a RESTful RDF translator service found at [9]. The returned data is stored in a SQL database. A web application was constructed using Angular, a framework for single page applications built off of jQuery. The web application uses the Java Persistence API (JPA) to communicate with MySQL. Users log in to Steam through the OpenID login authentication, and a request is then sent to the Steam API using GetOwnedGames. The games are sorted server-side by playtime; then, for the top five games, the application queries MySQL for games of similar and ESRB rating. The idea is that the user would probably like games that are similar in overall type and quality. The similar games are then displayed to the user, including their picture, description, and other information.

4. CONTRIBUTIONS AND FURTHER WORK

The result of this research was 1) the videogame ontology, 2) the user metadata ontology, and 3) the web application.

4.1 Contributions

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.

4.1 Further Work

There are several areas in which further work could improve this application. First, the system only uses videogame play time to identify games the user has liked in the past. While this appears to be a better indicator of good videogame recommendations than the use of demographic segments, integrating achievements as initially proposed would surely improve suggestions. The problem is that there is no

way to sort achievements into categories; that is, Steam's achievement data does not include some sort of classification. A possible solution would be to parse the achievement descriptions for key words pertaining to different playing types mentioned in [8]; the only alternative is heuristic hand-classification of achievements, which is outside of the scope of this research. In order to automatically classify achievements based on keywords found in their descriptions, however, a dictionary of keywords needs to be created. This is a possible area of further study.

Another area of future effort is the way in which games from the system's database are identified as being similar to games liked by the user. Currently, the system looks for games that are similar in regards to genre and ESRB rating. It would be useful to expand this search to take into account more data, especially considering the amount of data that the system already has concerning each videogame.

Finally, there is the user interface itself. There is not currently any options for users to customize the kind of results they obtain from the system. It would be a simple matter to abstract some of the SQL query parameters to the user interface. This method would introduce more complexity to the web application, but due to the choice of Angular as the framework used, there would not be an untenable amount of rework needed.

5. ACKNOWLEDGEMENTS

The author would like to acknowledge the CREU program, including the CDC and the CRA-W, for making this project possible through their generous research grant. The author would also like to thank Dr. Srividya Bansal for providing mentorship and guidance throughout the project. Finally, the author acknowledges the Ira A Fulton Schools of Engineering for supporting the Big Data and Semantic Computing Lab ASU Polytechnic.

6. REFERENCES

[1] Kandogan, E., Kieliszewski, C., Ozcan, F., Roth, M., Schloss, B., and Schmidt, M.-T. Data for All: A Systems Approach to Accelerate the Path from Data to Insight. In 2013 IEEE International Congress on Big Data (BigData Congress), June 2013, 427–428.

[2] Berners-Lee, Tim. 1989. History. Information Management: A Proposal. http://www.w3.org/History/1989/proposal.html.

[3] Adomavicius, G. and Tuzhilin, A. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions, 17, 6, 734-749.

[4] Reddit. 2015. GamingSuggestions. http://www.reddit.com/r/gamingsuggestions.

[5] Adda, M., Djeraba, C., Missaoui, R., and Valtchev, P.

2007. Toward Recommendation Based on Ontology-Powered Web-Usage Mining. Internet Computing, IEEE, 11, 4, 45-52.

[6] Blanco-Fernandez, Y., Gil-Solla, A., Lopez-Nores, M., Pazos-arias, J., and Ramos-Cabrer, M. 2008. Providing entertainment by content-based filtering and semantic reasoning in intelligent recommender systems.Consumer Electronics, IEEE Transactions, 17, 2, 727-735.

[7] Kong, Fan-Sheng and Wang, Rui-Qin. 2007. Semantic-Enhanced Personalized Recommender System. In 2007 International Conference on Machine Learning and Cybernetics, August 2007, 4069-4074.

[8] Lees, Jennie. 2014. Stop Recommending Bejeweled: Using Data Science to Broaden Game Audiences. In Grace Hopper Celebration of Women in Computing, October 2014.

[9] Stolz, Alex. RDF Translator. E-Business and Web Science Research Group, Universitat der Bundeswehr Munchen. 2014. http://rdf-translator.appspot.com/