Harnessing Manpower for Semantics Acquisition

Jakub ŠIMKO*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
jsimko@fiit.stuba.sk

Nowadays, the amount of information on the Web grows extremely fast. In order to be able to search the Web and utilize its content, we require scalable methods for acquiring information about individual resources. This information, opposite to the heterogeneity of web resources, must be homogeneous to be easily processed by machines. Nowadays, the role of such homogeneous meta-layer above the Web is played by the keyword search indexes within web search engines.

Semantic Web principles were created to provide a worldwide framework for creating richer web resource annotation than keyword indexes. The Semantic Web can be seen as a meta-layer on the common web: a collection of web resource annotations unified under universal and widely accepted domain models. With such structure, the web invisibility problem [2] is easy to solve – it can be used to create various forms of web abstractions to browse. The solving of the problem of sophisticated queries becomes also trivial – besides the structured information and knowledge, Semantic Web standards provide logical reasoning frameworks suitable for question answering. Although there is much work done in the field of automatic semantics acquisition, human effort on building the semantics is still a need. In response to that we study approaches for creating semantics on the Web with accent on games with a purpose (GWAP), which are computer games that transform human intelligence tasks into entertaining games [5, 3].

In our previous research, we introduced *Little Google Game* (LGG), a game with the purpose of acquiring a general term relationship network. Little Google Game is a web search game of query formulation, in which the player's task is to minimize number of search results returned for the given single-term query (e.g., query "sea" which yields approximately 2 billion results). The player can expand the query using only *negative* search terms that reduce the original result set of web resources decorated with negative terms (e.g., "sea –blue –fish –deep"). The game is single player and exercises a ranking system to motivate competition among players [4].

---

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2011 PeWe Workshop, April 1, 2011, pp. 67–68.

We have validated the semantic soundness of the relationships within the network and shown that the game can discover relationships that remain hidden to automated corpora mining methods. Now we look at the LGG network as a potential source of ontological facts. We conducted experiments to disclose how many of the relationships are present in the major knowledge base of Wikipedia using the Wikipedia Miner tool and what kind of relationships are mostly present in the network, by evaluating them manually and also comparing them to the facts in the knowledge base of ConceptNet. We propose two methods for naming the LGG network relationships: first, a modification of Little Google Game itself forcing players to disclose predicate-like terms related to existing bigrams and second, automated sentence mining with web search engines helping to gather relevant sentences.

We are also interested in the design of games with a purpose for the Semantic Web in general. GWAPs are a potent tools for harnessing human knowledge and hypothetically applicable to any problem. Creating a GWAP is however, not so straightforward and is very specific with each problem. Our goal is to reach a set of best practices and guidelines to unify the creation of GWAPs (for the Semantic Web) since many design aspects recur (e.g., reaching the entertainment, elimination of cheating, validation of user inputs). One possible way may be a modification of a classical game design methodology, such as MDA [1], or methodology for transformation of existing classical games to games with a purpose.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11.

References

[1] Hunicke, R., Leblanc, M., Zubek, R.: MDA: A formal approach to game design and game research. In *Proc. of the Challenges in Games AI Workshop, Nineteenth National Conf. of Artificial Intelligence,* pp. 1–5, 2004.

[2] Marchionini, G.: From finding to understanding. *Communications of the ACM* Vol. 49, No. 4,, 41-46, 2006.

[3] Siorpaes, K., Hepp, M.: Games with a purpose for the semantic web. *IEEE Intelligent Systems*, Vol. 23, 50–60, 2008.

[4] Šimko, J., Tvarožek, M., Bieliková, M. Little Google Game: Creation of Term Network via Search Game. In *Proc. of Datakon 2010,* 2010 (in Slovak).

[5] von Ahn, L., Dabbish, L.: Designing games with a purpose. *Communications of the ACM*, Vol. 51, No. 8, 58–67, 2008.