

OntoGame: Games with a Purpose for the Semantic Web

Extended PhD Thesis Abstract

Katharina Siorpaes
STI, University of Innsbruck, Austria
katharina.siorpaes@sti2.at

1. Research Problem

A pre-requisite for the Semantic Web to become a reality is the availability of ontologies [1] and meta-data. In many cases, it might be necessary to align between different ontologies in order to ensure interoperability. The research in the area of semantic content authoring has brought up an inventory of mature techniques and tools for semantic content creation. However, there is a severe lack of semantic data available on the Web: one can only find few well-maintained ontologies, respective alignments and very little semantic annotation. For instance, a search on Watson¹ or Swoogle² for a tourism ontology does not deliver a proper tourism ontology even though travel and tourism ontologies have been created in many academic projects in the last couple years. Furthermore, one can observe very little involvement of Web users in the process of semantic content creation. However, this involvement is urgently needed: there are tasks that are trivial for a human user but still difficult for a computer [2, 3]. Conceptual modeling and semantic annotation are tasks that depend on human intelligence: even though approaches for automating these activities exist, the problem has not been solved completely yet and human input is required at some stage. Therefore, we are now confronted with the situation that even though the technology is available, there is very little semantic content which can be traced back to only little user involvement. We believe that this is caused by missing incentive structures: the effort of building ontologies currently outweighs the benefit.

2. Motivation and Contribution

This is in sharp contrast to the Web 2.0 movement, which has proper incentive structures in place [4-6]. In my thesis, I investigate intrinsic motivations of users for contributing to Web 2.0 applications and propose to define possible incentive models for the Semantic Web. More precisely, I propose to masquerade core tasks of weaving the Semantic Web behind on-line, multi-player game scenarios, in order to create proper incentives for humans to contribute. Doing so, I adopt the findings from the already famous “games with a purpose” by von Ahn [2], who has shown that presenting a useful task, which requires human intelligence, in the form of an on-line game can motivate a large amount of people to work heavily on this task, and this for free.

¹ <http://watson.kmi.open.ac.uk/WatsonWUI/>

² <http://swoogle.umbc.edu/>

The **contribution of my thesis** is (1) an overview of incentives for users to contribute to Web 2.0 applications, (2) a survey on serious games and games with a purpose, (3) a conceptual framework that aims at (a) defining incentives (more precisely, intrinsic motivations) for the Semantic Web and (b) describing how to hide semantic content creation and maintenance tasks behind online games. Furthermore, I will provide (3) a proof-of-concept implementation with four cool games scenarios that will be available to the general public. Finally, I will (4) evaluate the fun factor of the games and (5) analyze the output of the games checking the correctness and the usefulness of the resulting data. OntoGame is an approach to the massive generation of lightweight knowledge structures that can serve as a starting point for further axiomatization, as training sets for semi-automatic approaches, and that can be useful for machine learning techniques.

3. Related Work

Several “games with a purpose” have been described by **Luis von Ahn** and colleagues; they also coined the term “*human computation*”: The ESP game [7] aims at labeling images on the Web - two players, who do not know each other, have to come up with identical tags describing an image. Peekaboom [8] works similar and has the objective of locating objects within images. Verbosity [9] is a game for collecting common sense facts. Phetch [10] is a computer game that collects explanatory descriptions of images in order to improve accessibility of the Web for the visually impaired. Law, von Ahn, and colleagues [11] came up with a game called Tagatune for music and sound annotation based on tags. However, their current prototypes remain mostly at the **level of lexical resources only**, i.e. terms and tags and are not directly connected with Semantic Web research.

Liebermann and colleagues describe the game Common Consensus [12], which aims at collecting human goals in order to recognize goals from user actions and conclude a sequence of actions from these goals.

Another approach to collecting common sense knowledge is the FACTory Game³ published by **Cycorp**⁴: FACTory is a single-player online game that randomly chooses facts from the Cyc knowledge base [13] and presents them to the players. The player has to say whether the statement is true, false, doesn't make sense, or whether the user does not know. The answers are scored depending on accordance with the majority of answers.

A different type of games are so called passively multiplayer online games⁵, a term coined by **Justin Hall**. The idea of the PMOGs⁶ is to create avatars and game moves in multiplayer online games from user behavior on the Web. In other words, PMOGs translate e-mail content, chat logs, pictures, etc. into hunting parties, teams, puzzles, and so on.

4. Approach

In my PhD thesis, I propose to hide relevant tasks of semantic content authoring behind online games. In this section, I outline the challenges of my approach and the design

³ <http://game.cyc.com>

⁴ <http://www.cyc.com>

⁵ <http://passivelymultiplayer.com/PMOGPaper.html>

⁶ <http://www.passivelymultiplayer.com>

principles. Finally, I describe four cool OntoGame scenarios and explain how they address tasks in the Semantic Web lifecycle.

Challenges

The design of games for building the Semantic Web involves several challenges:

- (1) Conceptual model of the games: it is crucial to make sure that the games are interesting and deliver useful output at the same time. This involves not only nice user interface design but also methods to keep up interest. One example for this would be revealing information about the partner (gender, age, nationality, etc.).
- (2) Input data: For most game scenarios, a large corpus of knowledge, such as Wikipedia or YouTube, is required.
- (3) Deriving formal semantics: from the games, formal semantics must be extracted, i.e. exports in common languages such as OWL.
- (4) Cheating: ways to avoid cheating must be described and implemented.
- (5) Re-use and analysis of generated data: in order to increase the amount of diverging data gathered as well as further deepening the degree of detail of the data, one must find algorithms and mechanisms to re-use gained data.
- (6) Typical Mistakes: from first experiments, it is obvious that there are some cases where users tend to make mistakes, i.e. classifying something as a sub-class of a concept that is not correct. One has to find ways to avoid the impact of these “false friends”.

Design Principles

I build my work on OntoGame on the following design principles:

I. Fun and Intellectual Challenge

Fun and the game challenge are the predominant user experience. The actual tasks are very well hidden such that their serious and useful nature does not decrease the “fun factor”. Additionally, the games should comprise an intellectual challenge being fun and interesting at the same time.

II. Consensus

In our games, I adopt the “Wisdom of crowds” [14] paradigm. Groups only perform well under certain conditions: the group must be diverse, geographically dispersed, and members must be unable to influence each other. The settings of our games fulfill these requirements in order to tap the “wisdom of crowds”.

III. Massive Content Generation

The assumptions about the intelligence of groups are only true given mass participation. Our games aim at the massive generation of semantic content, and thus mass user participation.

Four Cool Scenarios

In order to evaluate the set of abstract game scenarios, four games were implemented⁷ that address the whole Semantic Web lifecycle (Fig. 1): certain tasks involved in ontology construction, alignment, and semantic annotation can be hidden behind online games. In a

⁷ OntoPronto is released, OntoTube and SpotTheLink are close to release, OntoBay implementation is starting now. All four scenarios can be expected by summer 2008 latest.

4 Katharina Siorpaes

nutshell, the OntoGame⁸ series includes the following games: **OntoPronto** is a game for annotating Wikipedia and for creating a huge general interest ontology (the English Wikipedia currently contains more than 2 Million articles). **SpotTheLink** aims at aligning the product and service classifications eCI@ss and UNSPSC, respectively their OWL counterparts eClassOWL and unspscOWL. **OntoTube** (Fig. 2) produces annotations for YouTube videos. A fourth upcoming scenario is called **OntoBay** and is a game for annotating eBay auctions by expressing the type of goods being offered using eClassOWL.⁹

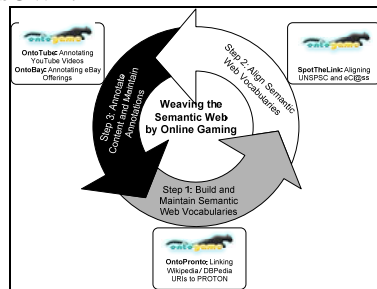


Fig. 1. Games for the Semantic Web Lifecycle



Fig. 2. Annotating YouTube

5. Evaluation and Preliminary Evidence

The objective of the evaluation is twofold: (1) to evaluate whether our first prototype creates an entertaining **gaming experience** and (2) whether the consensual conceptual choices of players in the **games are correct**. My hypothesis is that the majority of the players' decisions are ontologically correct and players will enjoy the games. We plan to release the four prototypes to the general public on several game platforms and make many users play the games. I will then analyze the resulting data: absolute number of games, absolute number of resources (Wikipedia articles, YouTube videos, etc.), ratio of single-player games, time invested by users, figures about the degree of consensus, and most importantly, the quality of conceptual salutation, i.e. mistakes that were made. For this purpose I will take representative samples and will ask experts to judge the correctness of the data. Furthermore, I will conduct surveys among players evaluating the fun factor, similar to the survey described in [15].

Preliminary evidence [15] indicates that this hypothesis is correct: OntoPronto, the first game of the OntoGame series, was released to the general public in Dec. 2007. Within the first two days, more than 200 players registered and played the game. The results of the analysis are promising: players make few mistakes and manage to find consensus in the majority of cases.

6. Expected Impact and Roadmap

Designers of semantic applications should start to think about incentives for users to invest time in those applications: in my thesis, I will provide helpful guidelines for adopting those

⁸ <http://www.ontogame.org>

⁹ A more detailed description of the games can be found in [16].

from Web 2.0 to Semantic Web. More precisely, the thesis will focus on games, implementing the motivation fun and competition. I believe that the games described in my PhD thesis have the potential to generate a huge amount of lightweight knowledge structures that are useful in several aspects: (1) use of the resulting data with very little or no changes as lightweight ontologies and annotations, (2) use of the resulting knowledge structures as a basis for domain ontologies for further axiomatization, (3) use as training data for semi-automatic approaches, and (4) for machine learning.

So far, OntoPronto has been released to the general public. OntoTube and SpotTheLink are currently being tested. All four scenarios are expected to be online and broadly published by summer 2008. So far, my work was published in [15-17].

References

1. Gruber, T.R., *Toward principles for the design of ontologies used for knowledge sharing*. International Journal of Human-Computer Studies, 1995. **43**: p. 907-928.
2. Von Ahn, L., *Games with a Purpose*. IEEE Computer, 2006. **29**(6): p. 92-94.
3. Von Ahn, L., et al. *CAPTCHA: Using Hard AI Problems for Security*. Eurocrypt 2003.
4. Marlow, C., et al., *Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead*, *Proceedings of the World Wide Web Conference (WWW2006)*. 2006, ACM: Edinburgh, Scotland.
5. Hemetsberger, A., *When Consumers Produce on the Internet: The Relationship between Cognitive-affective, Socially-based, and Behavioral Involvement of Prosumers*. The Journal of Social Psychology, 2003.
6. Kuznetsov, S., *Motivations of Contributors to Wikipedia*. ACM SIGCAS Computers and Society, 2006. **36**(2).
7. Von Ahn, L. and L. Dabbish, *Labeling Images with a Computer Game*, CHI 2004. ACM.
8. Von Ahn, L., *Peekaboom: A Game for Locating Objects in Images*, *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 2006, ACM: Montréal, Québec, Canada.
9. Von Ahn, L., M. Kedia, and M. Blum, *Verbosity: a game for collecting common-sense facts*, *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 2006, ACM: Montréal, Québec, Canada.
10. Von Ahn, L., et al., *Improving Accessibility of the Web with a Computer Game*, *Proceedings of the SIGCHI conference on Human Factors in computing systems CHI '06*. 2006, ACM.
11. Law, E., et al. *Tagatune*. in *ISMIR 2007*. 2007. Vienna, Austria: OCG.
12. Lieberman, H., D. Smith, and A. Teeters, *Common Consensus: A Web-based Game for Collecting Commonsense Goals*, in *Workshop on Common Sense for Intelligent Interfaces, ACM International Conference on Intelligent User Interfaces (IUI-07)*. 2007: Honolulu.
13. Lenat, D.B. and R.V. Guha, *Building Large Knowledge-based Systems: Representation and Inference in the Cyc Project*. 1990, Boston, Massachusetts: Addison-Wesley.
14. Surowiecki, J., *The Wisdom of Crowds*. 2003, New York: Anchor Books Random House.
15. Siorpaes, K. and M. Hepp, *Games with a Purpose for the Semantic Web*. IEEE Intelligent Systems, Special Issue on Semantic Web, Summer 2008. (Forthcoming)
16. Siorpaes, K. and M. Hepp, *OntoGame: Weaving the Semantic Web by Online Games*, in *European Semantic Web Conference (ESWC 2008)*. 2008, Springer LNCS: Tenerife, Spain.
17. Siorpaes, K. and M. Hepp, *OntoGame: Towards Overcoming the Incentive Bottleneck in Ontology Building*, in *Proceedings of the 3rd International IFIP Workshop On Semantic Web & Web Semantics (SWWS '07) co-located with OTM Federated Conferences*. 2007, Springer LNCS: Vilamoura, Portugal.