A Web-based User Interaction Framework for Collaboratively Building and Validating Ontologies

Arno Scharl
MODUL University Vienna
Department of New Media Technology
scharl@modul.ac.at

Albert Weichselbraun,
Gerhard Wohlgenannt
Vienna University of Economics & BA
Institute for Information Business
{aweichse, wohlg}@ai.wu-wien.ac.at

ABSTRACT
Collaborative ontology building requires innovative navigational instruments that improve content exploration and the creation of shared meaning. Building upon an existing architecture for automated ontology learning from unstructured textual resources developed by the authors, this paper presents a Web-based user interaction framework encompassing three major components: (i) real-time visualizations of ontology evolution with time interval and confidence sliders, (ii) traditional ontology editing environments, and (iii) multi-player online games leveraging social networking platforms in the tradition of games with a purpose. A prototype in the environmental domain will showcase the integration of powerful search capabilities with novel graph-based interfaces for guiding novice and expert users alike.

Author Keywords
Ontology engineering, tightly-coupled interfaces, graph visualization, collaborative editing, games with a purpose, social networking platforms.

ACM Classification Keywords
H5.2. User Interfaces; H.5.3 Group and Organization Interfaces; H.3.3 Information Search and Retrieval

INTRODUCTION

Visual representations of ontologies, delivered through Web-based interfaces, with simple and rapid user feedback mechanisms generate semantic knowledge while browsing the content repository, with minimal cognitive effort on behalf of the user. Online verification games can assist this process by validating ontology elements and creating an entertaining yet productive setting where large-scale ontology evaluation becomes feasible. Such feedback mechanism and online games are effective means to leverage collective intelligence, which is described as "combining of behavior, preferences, or ideas of a group of people to create novel insights" [6]. Gathering collective intelligence from communities in a structured way and comparing the results with the opinions of individual domain experts [8] can help understand the cognitive processes and structures underlying expert behavior [4].

IHC 2008 – VIII Simpósio Sobre Fatores Humanos em Sistemas Computacionais. October 21-24, 2008, Porto Alegre, RS, Brazil.
 ISBN 978-85-7669-203-4

ONTOLOGY LEARNING AND VISUALIZATION

The *Media Watch on Climate Change* [14], which is shown in Figure 1, uses a tightly-coupled visual interface to integrate semantic and geospatial technologies. The ontology view on the right side results from semi-automated ontology learning. The system applies natural language processing techniques on its content repository to identify related concepts and relations. Spreading activation selects the most relevant concepts for inclusion into the domain ontology and combines information from multiple evidence sources to decide on the new concept's position [2]. The ontology learning component serializes the ontology in the RDF format and forwards the results to the visualization component based on *Graphviz* [11], an open source software project that provides different graph layouts and options for optimizing the resulting visualizations.

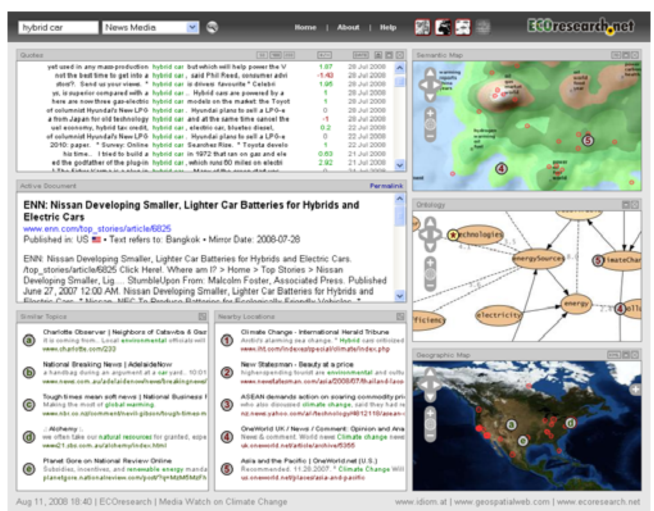

Figure 1. Media Watch on Climate Change [14]

To achieve a readable and easy understandable representation of the output ontology, the visualization component only translates a subset of the ontological constructs into the DOT language. The component visualizes hypernym and hyponym relations using arrows pointing to the more general concept and displays other types of relations by connecting the affected concepts using a dashed line. The numbers next to these lines indicate the relative strength of the relation between the connected concepts.

The *Media Watch on Climate Change* uses Web Map Server (WMS) technology to serve the finished graph – a flexible and generic solution that allows panning and zooming operations in multiple views. It seamlessly integrates the ontology view with a two-dimensional information landscape and a geographic map. We are in the process of extending this tightly-coupled view based on WMS technology with three-dimensional Java portlets. The current version of the portal already supports "*knowledge planets"* [5], the three-dimensional rendering of semantic maps using the Java SDK of NASA World Wind [15], and the projection of geographically referenced information onto virtual globes through customized KML code. The next release of the portal will provide the rendering of all supported visualizations through a common Java-based architecture.

COLLABORATIVE ONTOLOGY EDITING

For non-expert users, searching and navigating within ontologies is difficult. The visual interface described in this section facilitates intuitive high-level access to the content in the knowledge base (see Figure 2). Users have access to the full history of domain ontologies at the touch of a button. Concepts and relations are visualized according to their confidence values (indicating an element's level of support by empirical evidence), with the option to display or hide instance data. For any point in time, the user can adjust the confidence level and hence the ontology's appearance. Evidence sources are reflected by icons attached to the respective concept, relation or property. To depict the *evolution of ontological structures,* it is not only possible to display the state of the ontology at any given point in time, but also dynamic charts that show longitudinal patterns such as rising, declining or cyclic confidence values. Users can determine the level of granularity and the *minimum confidence value* of displayed objects. Graphical slider elements represent both confidence threshold and time interval.

Three levels of domain concepts and relations will be distinguished: *core, extended and peripheral*. Core domain terminology is always included in the domain ontology. Core domain relations are the essential relations between those concepts. Extended domain terminology is included (excluded) in the ontology depending on the increase (decrease) in coverage in the content base [10]. Graphical markup will allow visualizing the differentiation between permanent and temporary elements; e.g., by color coding the graph nodes. Peripheral terminology has not gained enough confidence to be included in the ontology.

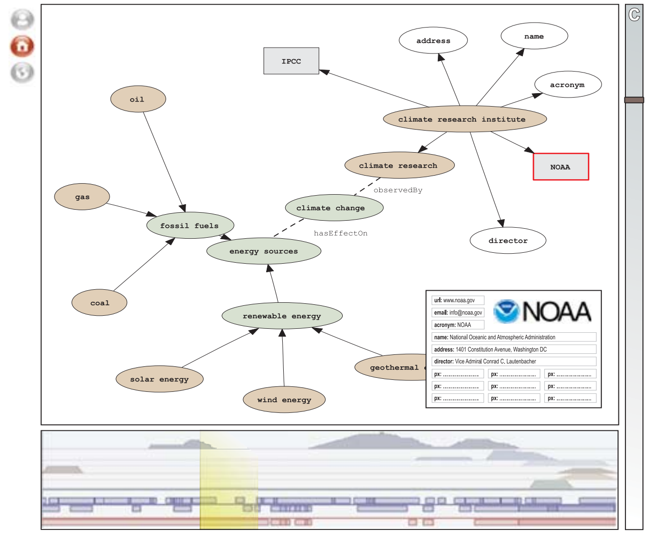

Figure 2. Mockup of collaborative editing view with time interval and confidence sliders

Feedback Mode

After publishing the domain ontology and entering the ontology graph feedback mode, users can take various actions: (i) *Concepts.* Mark a concept as correct and relevant, thereby increasing its confidence value. Alternatively, report a concept as having low or no relevance for the domain. If a certain number of users agree, the concept is removed from the ontology. If two concepts are marked as synonymous, the better denominator has to be chosen. Disagreements trigger so-called "games with a purpose" [7] to resolve the conflicting evidence (see below). The removal of synonyms requires procedures for handling orphaned relations, which are usually merged with the relations of the remaining concept. (ii) *Relations.* Report low relevance and perceived errors (e.g., no relevant relation between the concepts), or change the relation type (e.g., mark relations currently assumed to be taxonomic as non-taxonomic and vice versa). Situations of low confidence or conflicting evidence trigger games with a purpose. (iii) *Properties.* Provide feedback on property relevance and appropriateness of data types.

Editing Mode

The user interaction framework outlined in this paper does not intend to create a new ontology editor, but to develop new paradigms of interacting with highly dynamic, evolving and confidence-annotated ontologies. For expert users who are familiar with ontology editors, proven solutions such as OntoWiki [16] can be integrated as a separate view, disregarding desktop ontology editors like Protégé [17] that lack dynamic Web-based interfaces. With the ontology editor, users can enter new concepts including their properties, or specify the type of currently unlabeled relations. Users can also correct or replace existing labels. These actions will often create conflicts and the need for validation. Repositioning concepts in the ontology graph can be reduced to the task of evaluating relations, therefore games with a purpose can be used to process positioning suggestions.

GAMES WITH A PURPOSE

Multi-player online games are an innovative way to address conflicting evidence and confirm ontological entries with low confidence values. Games have been used successfully to solve problems that computers cannot yet solve, such as the tagging of images [1] and the alignment of ontologies and content annotation [7]. The main challenges in such games with a purpose are creating a user experience that motivates users to play the game while generating useful data, and ensuring that the process yields unbiased results. Games should be triggered on the fly to solve conflicting evidence on the concept, property, relation and instance levels. Compared to information extracted from text and structured sources automatically, user input is a scarce resource. The search-test-stop paradigm [3] can help optimize query strategies for multiple evidence sources [9].

We intend to apply this approach to managing queries to unstructured, structured and social evidence sources (including games) in a way that optimizes the *overall* confidence of the retrieved concepts. To allocate human resources most effectively, a scheduler will continuously update the list of pending verification tasks, assigning each task a priority value that takes into account the urgency of the problem (e.g., whether the information is required for a real-time application), and the expected information gain based on the pre-assigned confidence value.

Game Design

The search-test-stop module forwards selected entities to symmetric and asymmetric verification games. Users are not aware whether their input is matched against real-time input from other anonymous players, against stored records from previous users, or against system-generated values. If a certain number of players agree on a solution, it will be assumed correct and taken out of the game. There are two types of games:

- *Closed-ended games* provide a set of responses, either in binary format (Are A and B synonymous?), scaled (How relevant is this concept for the domain?) or multiple choice (Choose the correct relation type between A and B). Users receive points for matching answers.
- *Open-ended games.* Similar to image tagging games, users have a limited amount of time to make arbitrary suggestions. They get points for suggestions that match the other player's input. Popular suggestions can then be fed into closed-end games to refine the confidence values.

Social Networking Platforms

The large number of potential players and effective viral mechanisms built into popular social network platforms such as Facebook [12] help achieve a critical mass of social evidence. Users should be encouraged to invite their friends to install the game; e.g., by awarding a certain percentage of the points made by the players whom they have invited. The interface not only shows the players with the highest scores, but also those who are currently playing. Within the IDIOM Project [13], the social verification game for sentiment detection depicted in Figure 3 has been developed [18]. The findings in terms of usability and user motivation will be applied for designing the more complex case of ontology games. Merging several types of games in a meta-gaming platform (e.g., combining sentiment detection, image tagging and ontology learning) should increase the game's overall attractiveness, reduce the risk of cheating and allocate collective intelligence more efficiently by prioritizing tasks across game types.

Figure 3. US Election 2008 Sentiment Quiz [15]

Evaluation and Quality Control

The importance of accurate evidence for the quality of the resulting ontology calls for mechanisms to make sure that the games yield unbiased results. On social networking sites, users can identify other players and might collaborate to manipulate the game; e.g., by agreeing in advance on the answers on a limited number of questions.

A set of simple measures ensures high quality output: (i) hide the identity of the other player; (ii) analyze the temporal distribution of answers; (iii) assign trust values to each player, which in turn determine the impact of their answers on the confidence matrix – e.g. insert questions with known answers into the exercise queue and identify users who score low on these questions; (iv) avoid exploitable patterns in the sequence of answers, as users who identify the pattern could quickly gain credits without solving the puzzle.

APPLICATION TO THE ENVIRONMENTAL DOMAIN

Online communities emerge through cooperation and social exchange. They depend on and benefit from the dynamic maintenance of shared knowledge, non-hierarchical modes of cooperation, and distributed decision-making. The user interaction framework presented in this paper recognizes and supports the social construction of meaning in communicative processes to improve the quality of decisions, build trust and help resolve conflicts among competing interest. Collaborative ontology building ensures that discussions or recent findings are understood by all members of a virtual community. It facilitates the transition from conflict positions to shared meaning and replaces isolated, compartmentalized knowledge by the ability to deal effectively with complexity, uncertainty, and dynamic change.

The environmental domain has been chosen as a showcase for several reasons: (i) Due to its complex and dynamic nature, environmental knowledge represents a true challenge for ontology engineering; (ii) there is significant demand for environmental ontologies among various stakeholders; (iii) typically, members of environmental communities are intrinsically motivated and can be expected to actively participate in collaborative modeling and games with a purpose; (iv) the interface of the *Media Watch on Climate Change* shown in Figure 1 already includes an ontology view that can be extended accordingly.

OUTLOOK AND CONCLUSION

Successful games can motivate users to spend significant amounts of time on productive tasks [1]. Applying the game concept to more complex tasks such as ontology learning poses new challenges. With the exception of image labeling games, previous projects usually featured simple multiple-choice questions. Ontology building games deployed within an expert community (for example, environmental advocacy platforms and networking sites) should benefit from users with a high level of intrinsic motivation to help build a shared knowledge repository by playing the games. Unless group features are supported and accepted by the communities, this is not applicable to games deployed on social networking sites. In such a generic setting, the main motivation to play the games will be the entertainment value that they provide, or additional incentives such as awards and prizes that might be built into the application.

Besides refining the framework and advancing the interface technologies, future research will investigate what differentiates experts from novices in a given domain. Psychological research, for example, has shown that the level of expertise influences the fixation of domain-specific objects, the perception of relational elements (experts: parallel; novices: serial), and the visual span [4].

ACKNOWLEDGEMENT

The technologies underlying this paper were developed within the research projects IDIOM (www.idiom.at) and RAVEN (www.modul.ac.at/nmt/raven), funded by the Austrian Ministry of Transport, Innovation & Technology and the Austrian Research Promotion Agency within the strategic objective FIT-IT (www.fit-it.at). The Sentiment Quiz was developed by Walter Rafelsberger. The authors would like to thank Philipp Cimiano and Reinhard Fischer for their valuable feedback on earlier versions of this article.

REFERENCES

1. Ahn, L.v. (2006). “Games with a Purpose”, *Computer*, 39(6): 92-94.
2. Liu, W., Weichselbraun, A., Scharl, A. and Chang, E. (2005). “Semi-Automatic Ontology Extension Using Spreading Activation”, *Journal of Universal Knowledge Management*, 0(1): 50-58.
3. MacQueen, J. (1964). “Optimal Policies for a Class of Search and Evaluation Problems”, *Management Science*, 10(4): 746-759.
4. Reingold, E.M., Charness, N., Schultetus, R.S. and Stampe, D.M. (2001). “Perceptual Automaticity in Expert Chess Players: Parallel Encoding of Chess Relations”, *Psychonomic Bulletin & Review*, 8(3): 504-510.
5. Scharl, A. (2007). "Towards the Geospatial Web: Media Platforms for Managing Geotagged Knowledge Repositories", *The Geospatial Web - How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*. Eds. A. Scharl and K. Tochtermann. London: Springer. 3-14.
6. Segaran, T. (2007). *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. Beijing: O'Reilly.
7. Siorpaes, K. and Hepp, M. (2008). “Games with a Purpose for the Semantic Web”, *IEEE Intelligent Systems & their Applications*, 23: 50-60.
8. Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. London: Little, Brown.
9. Weichselbraun, A. (2008). Strategies for Optimizing Querying 3rd Party Resources in Semantic Web Applications. *3rd Int’l Conference on Software & Data Technologies*. Ed. J. Cordeiro et al. Porto, Portugal: 111-118.
10. Weichselbraun, A., Scharl, A. and Liu, W. (2008). Capturing and Classifying Ontology Evolution in News Media Archives. *19th Int’l Conference on Database and Expert Systems Applications (DEXA-2008), 7th International Workshop on Web Semantics*. Turin, Italy.

ONLINE RESOURCES

11. Graphviz. http://www.graphviz.org/.
12. Facebook. http://www.facebook.com/.
13. IDIOM Research Project. http://www.idiom.at/.
14. Media Watch on Climate Change. http://www.ecoresearch.net/climate/.
15. NASA World Wind. http://worldwind.arc.nasa.gov/.
16. OntoWiki. http://ontowiki.net/Projects/OntoWiki/.
17. *Protégé*. http://protege.stanford.edu/.
18. US08 Sentiment Quiz. http://www.ecoresearch.net/election2008/facebook/