Using a Semantic Wiki as a Knowledge Source for Rich Modeling and Question Answering

Vinay K. Chaudhri[1], Mark Greaves[2], Daniel Hansch[3], Anthony Jameson[4], Frederik Pfisterer[3], Aaron Spaulding[1], and Moritz Weiten[3]

1: SRI international, 2: Vulcan Inc., 3: ontoprise GmbH, 4: DFKI

Abstract

The traditional focus of knowledge engineering has been to acquire knowledge that can be used in deductive inference. Acquiring such knowledge invariably requires users to undergo extensive training. There are, however, aspects of the knowledge necessary for deductive inference that require much less training and sophistication. For example, capturing knowledge involving ground facts and concept taxonomies requires much less training and sophistication than acquiring knowledge about deductive rules. On the basis of this observation and the recent successes of knowledge capture on the web, we are exploring the following hypothesis at the intersection between knowledge engineering and the semantic web: Can we use a semantic wiki to acquire symbolic knowledge on the web that can then be used in the context of a deductive question answering system? We are conducting this work in the context of Project Halo and the AURA system with the goal of answering Advanced Placement (AP) questions in physics, chemistry, and biology.

Introduction

As part of Project Halo, we have been developing a system called AURA (Chaudhri, John et al. 2007), (Clark, Chaw et al. 2007) that enables graduate students in science to encode knowledge from a science textbook in such a way that high school graduates can receive correct answers to Advanced Placement Exam questions concerning that knowledge. We have been focusing on three particular science textbooks for the domains of physics, chemistry, and biology (Campbell and Reece 2001), (Giancoli 2004), (Brown, LeMay et al. 2003). In a recent evaluation of the AURA system, we showed that the system was able to support knowledge capture and question answering such that the test scores on unseen questions in the three domains were in the range of 20-40%. We are in the process of addressing the limitations of AURA to improve its correctness scores. Even after those improvements, though, the knowledge capture model in AURA will still have the following limitations:

The graduate students who formulate knowledge − *knowledge formulation subject matter experts* or *KF SMEs* − require around 40 hours of training. Not every form of knowledge requires that level of training and sophistication. Examples of such forms of knowledge are

a. The atomic numbers of chemical elements
b. Solubility constants for various chemicals
c. The fact that cells can be eukaryotic or prokaryotic

AURA is currently a classical single-user desktop application: Each knowledge base must be constructed by a single KF SME; collaboration with other KF SMEs is not yet supported.

These limitations of the current AURA architecture present an ideal opportunity to leverage semantic web techniques and investigate their symbiosis with knowledge engineering. We are pursuing this idea by using a lightweight form of semantic web technology – the *semantic wiki* − as a source of ground facts and taxonomic knowledge. We anticipate that this new architecture will have the following benefits:

We will be able to make use of contributions from online contributors who have had much less training than the 40 hours that are required for the AURA system. As a result, the knowledge creation process will become more distributed, faster, and cheaper, and we will be able to leverage the efforts of a larger number of contributors.

We will be able to manage the evolution of concepts and link types in the semantic wiki so that the semantic wiki is both human readable and useful for deductive question answering.

Illustration of the Approach

Our approach is illustrated in Figures 1 and 2. Figure 1 shows how a user of the semantic wiki can annotate a scientific page that has been copied from the normal English-language Wikipedia into a separate environment based on the Semantic MediaWiki platform (Krötzsch et al., 2006). This extension of the MediaWiki platform

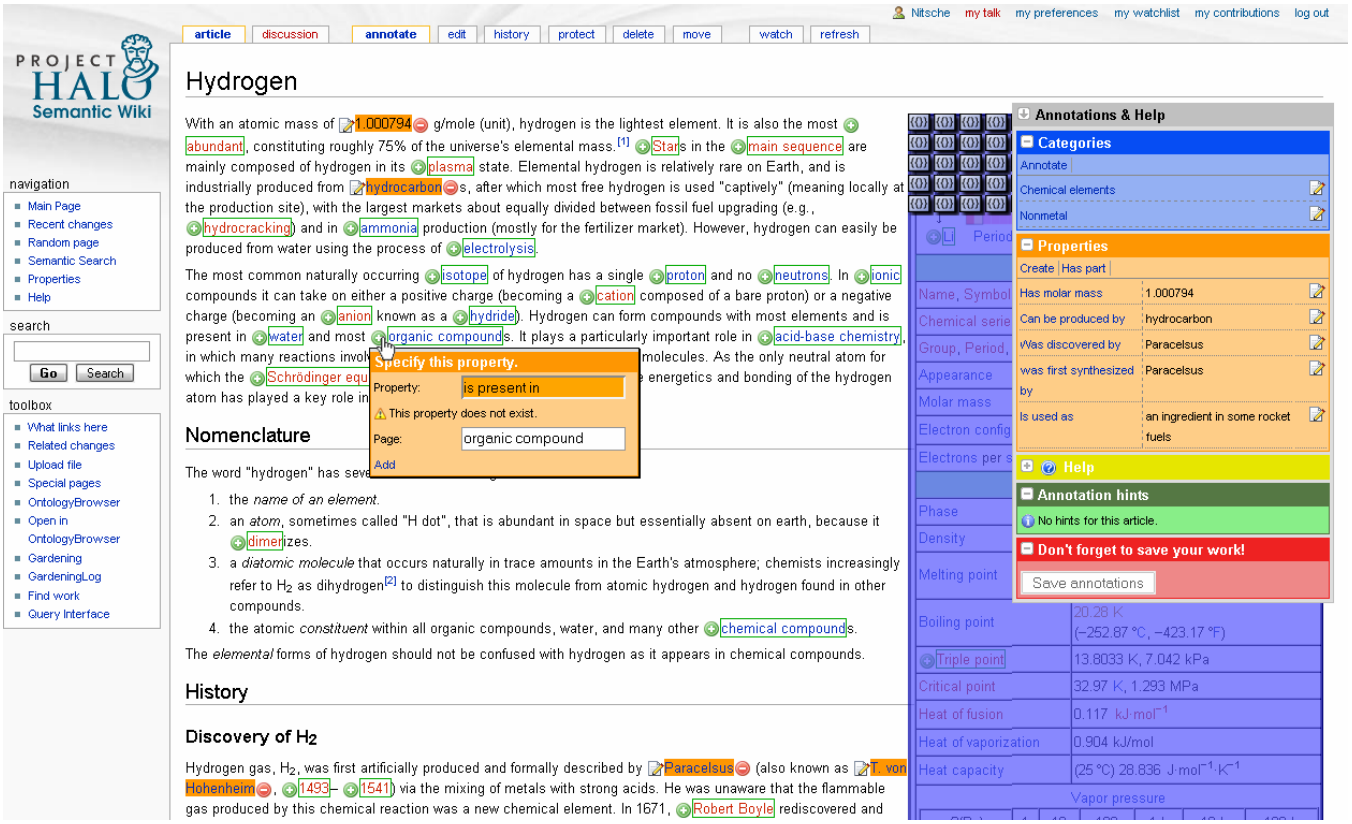

Figure 1. The Advanced Annotation Mode enables the creation and editing of typed links in a rendered view of the article without a need to edit the wiki source text.

(which is used by Wikipedia) supports *typed links*, each of which encodes a property (relation or attribute) of the entity described by the page in question. Figure 1 shows the *Advanced Annotation Mode* developed in the Halo project, which enables users to add and edit annotations without having to deal directly with the wiki source text. In this mode, each expression highlighted in a solid orange color corresponds to an underlying typed link. For example, the highlighted number in the first line indicates that the system has stored the fact that the atomic mass of hydrogen is 1.000794.

Terms highlighted with unfilled boxes are simply normal Wikipedia (untyped) hyperlinks. They are highlighted because they are likely candidates for annotation with typed links; but expressions that are not already hyperlinks can also be used for annotation.

The dialog box labeled "Specify this property" shows how a user can add an annotation for the fact that hydrogen is present in organic compounds.

The interfaces developed in the Halo project also include a number of other elements that are designed to make it easy for authors with virtually no training to annotate Wikipedia pages semantically (cf. Pfisterer et al., 2008). And in fact, science students working within the project have so far annotated hundreds of pages in this way, encoding thousands of facts.

Assuming that a semantic wiki exists that encodes knowledge that is useful for rich modeling and question answering, the question remains of how this knowledge can be imported into a rich modeling system such as AURA. The solution that is currently being integrated into AURA can be summarized as follows: Whenever the AURA KF SME searches for information in AURA, any matching pages in the semantic wiki are returned along with the normal search results. For each such matching page, the facts available for import into AURA are explicitly identified. (It is straightforward to identify such facts given the RDF associated with the page.) After examining the available facts, the AURA user makes a decision about whether to import them. If the decision is positive, the user invokes a mapping dialog in which to map the relevant concepts and properties onto the corresponding properties of AURA (see Figure 2). Once the KF SME has defined the mapping, a translation utility loads the desired facts into AURA in a representation that is compatible with the representation of knowledge that is already present in AURA. From then on, these facts are available to the AURA for answering questions.

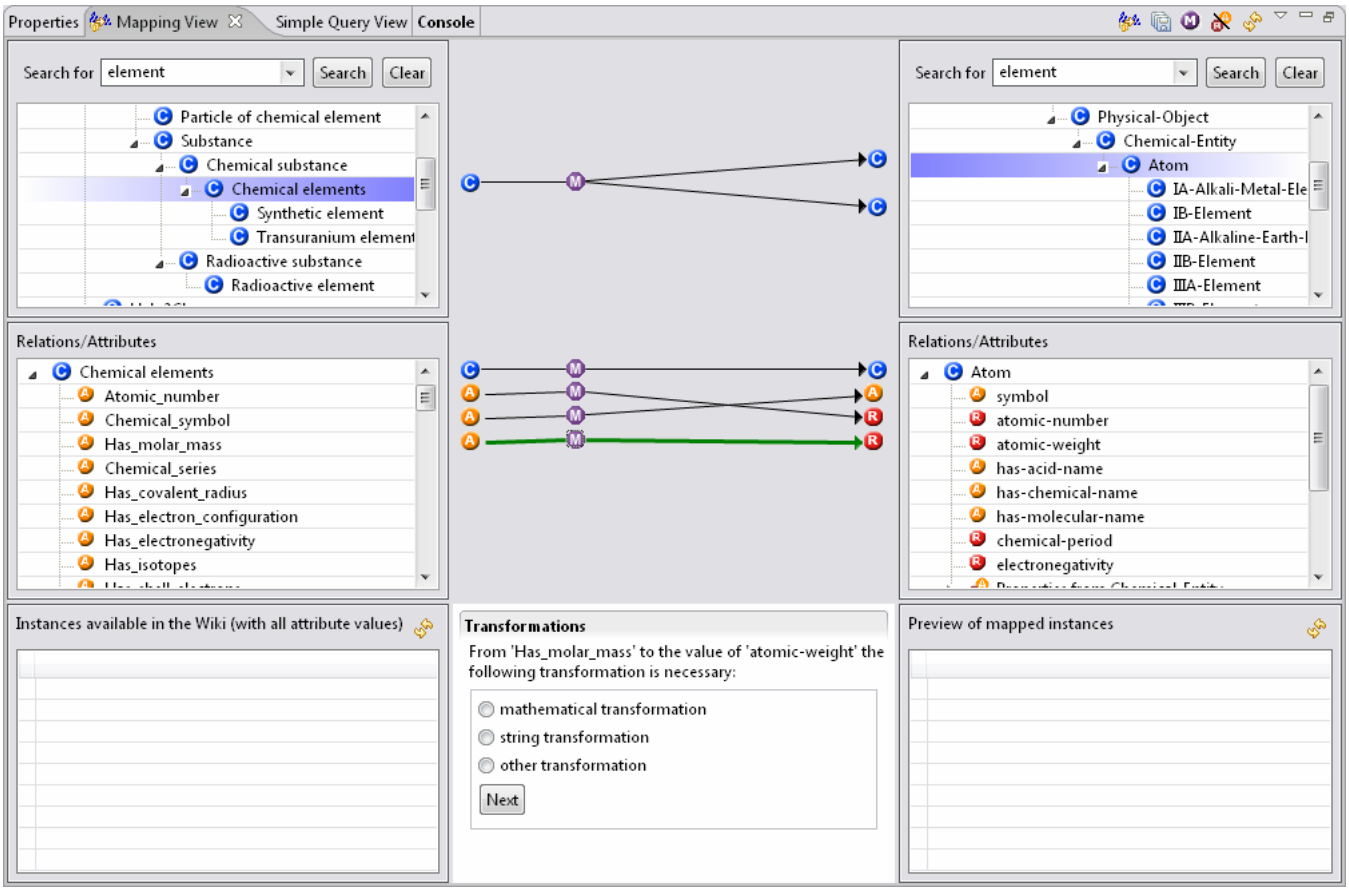

Figure 2. The mapping tool allows the KF SME to mediate between different ontology schemas in a declarative fashion using mapping patterns, filters, and transformations.

Status and Challenges

At present, we are focusing on creating a large knowledge base of semantically marked-up content in the semantic wiki. The tools for the importing of parts of this content into AURA are currently being integrated into AURA and tested. We have faced the following challenges so far.

Identifying annotations that can be useful for answering Advanced Placement questions: Since our ultimate goal is to answer AP questions, the semantic annotations introduced into the semantic wiki should be such that they can be useful for answering questions. In chemistry, there are a large number of ground facts, concerning properties such as atomic number, atomic weight, and dissociation constants, that can be annotated in a semantic wiki and that can also be useful for the answering of AP questions. In physics, much of the question answering involves the use of equations, and the ground facts are relatively few in number (one example would be the acceleration due to gravity at the Earth's surface). Because of this, we have been able to get only limited leverage by using the semantic wiki for physics knowledge. The biology domain also has a limited number of ground facts, but it is rich in taxonomic and partonomic information. Therefore, to be of significant usefulness, many of the annotations in biology pages in the semantic wiki have to encode partonomic and taxonomic facts.

Giving guidance on what to annotate: An author who is annotating pages in a scientific semantic wiki may need some guidance about what to annotate. In perhaps the simplest case, the guidance comes when an author uses the querying mechanisms of the Semantic MediaWiki (which have been enhanced in the Halo project) to retrieve facts from the wiki, only to discover that the facts returned by the system are incomplete or partly incorrect: It is then natural for the user to add or edit annotations until the result is satisfactory.

The user can also get an idea of what needs to be annotated by browsing the concepts, instances, and stored

facts in the *ontology browser*. This interface often quickly reveals gaps and inaccuracies, and it gives the user an opportunity to correct them immediately.

For partonomic relationships, we are investigating the use of AURA-like concept maps that could be generated from the data in the semantic wiki to provide a ready display of the partonomic information that has been added so far.

Finally, we have implemented a recommender system for annotators (inspired by the *SuggestBot* of Cosley et al., 2007) that lists possible annotation tasks on demand, taking into account not only the annotation gaps that the system has identified but also the interests of the current user, as reflected in that user's recent browsing behavior.

Representation mismatches between AURA and MediaWiki: An entity such as hydrogen is represented as an instance in the semantic wiki and as a class in the AURA system. This type of mismatch creates interesting challenges for the importing of knowledge from the semantic wiki into AURA. First, instances such as hydrogen must be mapped to classes in AURA. Second, since the user may want to import a large number of such entities at one time, we need to provide suitable interaction gestures, so that the mapping can be done in bulk as opposed to with a large number of specific actions on the part of the user.

Concluding Remark

The goal of combining lightweight, collaborative semantic annotation in a wiki with heavyweight knowledge formulation in a single-user system raises design and implementation challenges on a number of levels. Our experience so far indicates that these challenges can be met individually if sufficient attention is devoted to both technical and interaction design issues. But the question of just how effective this particular combination of approaches from the semantic web and the knowledge engineering areas really can be will begin to be answered only once the integration of AURA with the enhanced Semantic MediaWiki has been completed and thoroughly evaluated.

Acknowledgment

The research discussed in this paper is being funded by Vulcan Inc. in the context of the multistage project Halo.

References

Brown, T. L., H. E. LeMay, et al. (2003). Chemistry: The Central Science. New Jersey, Prentice Hall.

Campbell, N. A. and J. Reece (2001). Biology, Sixth Edition, Benjamin Cummings.

Chaudhri, V. K., B. John, et al. (2007). Enabling Experts to Build Knowledge Bases from Science Textbooks. International Conference on Knowledge Capture Systems (KCAP). Whistler, Canada.

Clark, P., J. Chaw, et al. (2007). Capturing and Answering Questions Posed to A Knowledge-Based System. International Conference on Knowledge Capture Systems (KCAP), Whistler, Canada.

Cosley, D., D. Frankowski, L. Terveen, and J. Riedl, (2007). SuggestBot: Using Intelligent Task Routing to Find Work in Wikipedia. In T. Lau and A. R. Puerta (Eds.), IUI 2007: International Conference on Intelligent User Interfaces (pp. 32-41). New York: ACM.

Giancoli, D. C. (2004). Physics Principles with Applications, Benjamin Cummings.

Krötzsch, M., D. Vrandecic, M. Völkel, H. Haller, and R. Studer (2006). Semantic Wikipedia. Proceedings of the 15th International Conference on the World Wide Web, Edinburgh, Scotland, pp. 585-594.

Pfisterer, F., M. Nitsche, A. Jameson, and C. Barbu (2008). User-Centered Design and Evaluation of Interface Enhancements to the Semantic MediaWiki. Proceedings of the CHI 2008 workshop on Semantic Web User Interaction, Florence, Italy.