IEEE Intelligent Systems
www.computer.org/intelligent

Games with a Purpose for the Semantic Web

Katharina Siorpaes and Martin Hepp

Vol. 23, No. 3
May/June 2008

Games with a Purpose for the Semantic Web

Katharina Siorpaes, *Semantic Technology Institute, University of Innsbruck*

Martin Hepp, *Bundeswehr University Munich and Semantic Technology Institute, University of Innsbruck*

Weaving the Semantic Web requires that humans contribute their labor and judgment for creating, extending, and updating formal knowledge structures. Hiding such tasks behind online multiplayer games presents the tasks as fun and intellectually challenging entertainment.

Making the Semantic Web a reality includes many tasks that even leading-edge computer technology can't perform fully automatically, but that humans with moderate training can master. Examples include creating or aligning ontologies and annotating data such as Web pages, e-commerce offerings, Flickr images, and YouTube videos. So, many humans must contribute substantial labor and intelligence, at least to generate training sets for semiautomatic approaches. Given the current slow progress, the job would still eventually be done if the ontologies we need and the data to be annotated were static. However, building the Semantic Web is a continuous challenge because domains and their respective representations change, requiring ceaseless maintenance.

Ontologies should reflect a view on the domain of interest that many people share. This means that ontology engineering, alignment, and annotation are by their very nature community efforts—humans must interact to yield useful results. While Web 2.0 applications enjoy great popularity, with people willing to spend time adding tags and extending tag sets, the Semantic Web lacks sufficient user involvement almost everywhere, as evidenced by the shortage of mature ontologies, industrial-strength ontology alignments, and substantial annotations. In a nutshell, we need to increase user involvement in building the Semantic Web by orders of magnitude.

You could argue that it's just a matter of time and that Semantic Web technology simply isn't yet mature enough. However, we assume that the Semantic Web's incentive structures are fundamentally flawed, whereas those of Web 2.0 applications are sufficient. Traditional ontology engineering, for example, detaches the effort from the benefits. Building an ontology alone doesn't improve your access to existing knowledge, while others can enjoy an ontology's added value without investing in its construction. So, those investing resources in creating or improving an ontology won't necessarily realize sufficient benefits from its use. The same holds for bringing about other key pieces of the larger Semantic Web puzzle, such as heavyweight annotations or ontology alignments.

We try to fix the broken incentive scheme for the Semantic Web by providing fun as a form of reciprocity and by letting users develop reputation—that is, establish and shape a positive identity. Our key approach is to have humans master those tasks and solve those problems that are too demanding to be automated as challenges in multiplayer online games. By doing so, the players generate semantic content and so unknowingly weave ontologies, alignments, and respective metadata for the Semantic Web. Our work adopts Luis von Ahn's "games with a purpose" paradigm (see the "Related Work in Games with a Purpose" sidebar) for the next generation of the Web. We focus on the challenges involved in building the

1541-1672/08/$25.00 © 2008 IEEE
Published by the IEEE Computer Society

Related Work in Games with a Purpose

Luis von Ahn and his colleagues have described several "games with a purpose."[1] They also coined the term "human computation." The ESP game aims to label images on the Web: two players, who don't know each other, must come up with identical tags describing an image.[2] The game has been extremely popular. Von Ahn observed that some people play the game more than 40 hours per week, and within a few months of the initial deployment on 25 October 2003, the game collected more than 10 million consensual image labels, and this without paying a cent to the contributors. Peekaboom works similarly and aims to locate objects in images.[3] Verbosity is a game for collecting commonsense facts.[4] Phetch is a computer game that collects explanatory descriptions of images to improve Web accessibility for the visually impaired.[5] Very recently, Edith Law and her colleagues introduced Tagatune, a game for tagging music and sound.[6] However, their prototypes remain mostly at the level of lexical resources—that is, terms and tags aren't directly connected with Semantic Web research.

Henry Lieberman and his colleagues describe Common Consensus, a game that aims to collect human goals to recognize goals from user actions and conclude a sequence of actions from these goals.[7] Another approach to collecting commonsense knowledge is Cycorp's Factory Game (http://game.cyc.com). Factory is a single-player online game that randomly chooses facts from the Cyc knowledge base[8] and presents them to the player. The player can say that the statement is true, is false, or doesn't make sense or that he or she doesn't know. Answers are scored on the basis of their accordance with most answers.

Justin Hall introduced *passively multiplayer online games*.[9] PMOGs aim to create avatars and game moves in multiplayer online games from user behavior on the Web (www.passivelymultiplayer.com). In other words, PMOGs translate email content, chat logs, pictures, and other content into hunting parties, teams, puzzles, and so on. In a sense, PMOGs take the opposite direction from our approach. They move from content and Web action to game scenarios, whereas we derive formal Web content from user input in game scenarios.

Other researchers have presented research on intrinsic motivations, especially in Web 2.0 settings.[10,11] The difference between OntoGame and serious games[12] is that serious games aim primarily to educate and train.

Apart from Verbosity, Common Consensus, and Factory, we don't know of any other research that uses computer game scenarios for collecting and codifying knowledge at the conceptual level, and none of those applications is currently linked to Semantic Web efforts or the current Semantic Web technology stack.

References

1. L. Von Ahn, "Games with a Purpose," *Computer*, vol. 29, no. 6, 2006, pp. 92–94.
2. L. Von Ahn and L. Dabbish, "Labeling Images with a Computer Game," *Proc. Conf. Human Factors in Computing Systems* (CHI 04), ACM Press, 2004, pp. 319–326.
3. L. Von Ahn, "Peekaboom: A Game for Locating Objects in Images," *Proc. Conf. Human Factors in Computing Systems* (CHI 06), ACM Press, 2006, pp. 55–64.
4. L. Von Ahn, M. Kedia, and M. Blum, "Verbosity: A Game for Collecting Commonsense Facts," *Proc. Conf. Human Factors in Computing Systems* (CHI 06), ACM Press, 2006, pp. 75–78.
5. L. Von Ahn et al., "Improving Accessibility of the Web with a Computer Game," *Proc. Conf. Human Factors in Computing Systems* (CHI 06), ACM Press, 2006, pp. 79–82.
6. E. Law et al., "Tagatune," *Proc. Int'l Conf. Music Information Retrieval* (Ismir 07), Austrian Computer Soc., 2007, pp. 361–364.
7. H. Lieberman, D. Smith, and A. Teeters, "Common Consensus: A Web-Based Game or Collecting Commonsense Goals," *Proc. Workshop Common Sense for Intelligent Interfaces, ACM Conf. Intelligent User Interfaces* (IUI 07), ACM Press, 2007.
8. D.B. Lenat and R.V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*, Addison-Wesley, 1990.
9. J. Hall, "Passively Multiplayer Online Games," Feb. 2007, http://passivelymultiplayer.com/20070214-PMOGPaper.html.
10. A. Hemetsberger, "Fostering Cooperation on the Internet: Social Exchange Processes in Innovative Virtual Consumer Communities," *Advances in Consumer Research*, vol. 29, 2002, pp. 354–356.
11. C. Marlow et al., "Tagging, Taxonomy, Flickr, Article, ToRead," *Proc. World Wide Web Conf.* (WWW 06), ACM Press, 2006, pp. 31–40.
12. D. Michael and S. Chen, *Serious Games—Games That Educate, Train, and Inform*, Thomson Course Technology, 2006.

Semantic Web that are doable for a human player but hard for a computer. From the data produced in the games, we derive representations in common Semantic Web standards, such as OWL, which are then made available via HTTP to Semantic Web search engines and other Semantic Web applications.

Motivating users

If we look at what could motivate people to dedicate their valuable time to building the Semantic Web, we see four options:

- We can hope that enough people will contribute simply because it's good for the world. Building solely on altruism is however likely not sufficient; most of humankind's major achievements offered their contributors stronger incentives (for example, fame, wealth, or both).
- We could pay contributors. But, who will provide the money? The Semantic Web can't be built just from taxpayers' money via public funding.
- We can set up the Semantic Web's authoring mechanisms to balance contribution with immediate benefits. For example, to use a Semantic Web search engine, a user would have to contribute five minutes a day producing semantic data. This approach might be promising, but it requires much additional research and implementation work.
- We can arrange for nonmonetary incentives other than an immediate gain in personal access to Web data.

Wikipedia created a setting that has continued to motivate a huge number of

people to contribute a large amount of human labor, intelligence, and knowledge. Elsewhere, we show that about 230,000 change operations occur each day in the English version of Wikipedia alone—almost 7 million per month.[1] For Wikipedia, Stacey Kuznetsov names the following nonmonetary, indirect incentives apart from pure altruism as sources of motivation for contributors:[2]

- *Reciprocity.* Altruistic contributors receive a benefit in return.
- *Community.* "Wikipedians […] feel needed;" there is a "sense of common purpose and belonging."
- *Reputation*. Contributors "develop identities in order to be respected, trusted, and appreciated by peers."
- *Autonomy.* Contributors enjoy the "freedom of independent decision."

Games for weaving the Web

Von Ahn has already shown that masquerading content-authoring tasks as games is promising, in particular when considering the sheer number of hours that online gamers play every day (Luis von Ahn: "Human Computation," Google TechTalk, 2006, http://video.google.com/videoplay?docid=-8246463980976635143). One principle of successful games for this purpose is to design them such that cooperation is the dominant strategy—that is, only consensual solutions are awarded with points. Because a shared representation—understandable by and agreed upon by multiple human actors—is at an ontology's heart, such games also show a promising fit for the Semantic Web.

In addition, the sheer number of potential players on the Web creates a human resource of unique potential. If we can make, on average, 50 individuals around the globe play every moment and run our games for half a year, they'll contribute 216,000 hours (50 * 24 * 180) of intellectual work. If we assume an average wage of US$10 for conceptual-modeling tasks, which is likely much less than actual wages, we get work done that would cost more than US$2 million on the labor market. In one of our games, we observed an average of about four conceptual choices per two-minute game round—that is, two players produced about two conceptual-modeling decisions per minute, or one per player per minute. So, on the basis of our numbers, we can gather more than 12 million conceptual-modeling choices in half a year (216,000 * 60 = 12,960,000).

For our games, we apply the following design principles for the games:

- *Fun and intellectual challenge.* Fun and intellectual challenge are the predominant user experience. The actual tasks, such as annotating a resource or specifying a label for a relationship between classes, are hidden so that their serious and useful nature doesn't decrease the gaming fun. Additionally, the games should provide an intellectual challenge so that they're both fun and interesting.
- *Consensus.* Our games adopt the "wisdom of crowds" paradigm:[3] If many say that A is an instance of B, A is likely an instance of B. It has been reported that groups perform well only under certain conditions: the group must be diverse and geographically dispersed, and its members must be unable to influence each other. Our games' setting fulfills these requirements to tap the wisdom of crowds.
- *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.

> Most useful game scenarios require a large set of input data to play with; otherwise, a player may be faced with the same challenge over and over again.

When hiding semantic-content creation behind online games, we face the following 10 key challenges:

1. *Identifying tasks in semantic-content creation.* We must identify the relevant tasks that *should* be done and that *can* be done. In other words, not all tasks are suitable for being masqueraded behind online games and thus feasible for a broad audience.
2. *Designing game scenarios.* On the basis of the collection of relevant semantic-content-creation tasks, we must make the games' conceptual design such that they achieve the targeted goal and are fun to play.
3. *Designing a usable, attractive interface.* It is much more difficult to produce a user interface that's suitable for online games for ontology building or semantic annotation than for the rather simple tagging challenges in existing games with a purpose. A major challenge is that the navigation must force users to move along existing knowledge structures in most of our games.
4. *Identifying reusable bodies of knowledge.* Most useful game scenarios require a large set of input data to play with; otherwise, a player may be faced with the same challenge over and over again. So we must identify suitable knowledge corpora that fit to our gaming scenarios and can be reused with limited effort. Examples include Wikipedia articles and YouTube videos.
5. *Preventing cheating.* If users can fool the system with false responses, this might deteriorate the data we derive from the games. Thus, we must both develop a setting that makes cheating difficult and minimize the impact of such cheating on the derived formal content. We basically rely on von Ahn's techniques: First, all games are anonymous (that is, players don't know each other's identity, so they can't communicate). Second, we can sporadically present known challenges—that is, challenges for which the correct results are stored in our system—to test players. Only if players succeed in those tests, we consider their further input in our data. Finally, we must design the games so that the most promising strategy is entering correct answers, which goes hand-in-hand with the wisdom-of-crowds paradigm.
6. *Avoiding typical pitfalls.* It's easier to counteract intentional cheating than unintentional bad input. When building ontologies or annotating content, players can unintentionally agree on wrong choices. An example of this is the incorrect usage of subClassOf relations—that is, establishing a subClassOf relation while another type of relation exists. To address this problem, we can identify a couple of

challenges with such pitfalls to test players, and ignore their input if they don't master the challenge. Alternatively, the game's user interface can guide players and warn them about possible pitfalls.

7. *Fostering user participation.* We can't assume that the games alone will generate sufficient user involvement. We aim to provide additional incentives to make users start and continue playing. This could include informing users by email that they're about to lose their top-10 ranking, or revealing information about their playing partner, such as gender or nationality.
8. *Deriving semantic data.* We must develop robust algorithms that let us derive formal representations from the data produced in the games played without additional human intervention—ideally, in standards such as OWL or RDF Schema. As a starting point, we'll apply simple threshold mechanisms—for example, assume the most popular choice among those choices that have been the consensual output of games to be the correct one. If we know the option space's size, we could even determine the confidence level for that choice using standard statistics.
9. *Efficient distribution of labor.* Good games won't simply pose the same set of challenges to players randomly, even when the set of challenges is large. Instead, the games should direct players' intellectual contributions to those parts of the underlying task that, at that moment, most urgently need attention and most immediately benefit from additional labor. For example, we might want to present each challenge only until at least one team has solved it consensually, and then proceed to the remaining challenges. When at least one team has mastered a single challenge, we'd direct user input to confirming previous choices. The underlying approaches for directing contributed play time to the most meaningful tasks can of course be much more complex and are an important part of efficient game designs.
10. *Scalability and performance.* The game infrastructure must deal properly with numerous parallel games, and all mechanisms for attracting players should try to balance the distribution of players over time.

Our goal isn't just to show that the "games with a purpose" idea is applicable to some Semantic Web tasks. Instead, we aim to provide a series of games that cover the full life cycle of weaving the Semantic Web, from building and aligning vocabularies to annotating data. Figure 1 illustrates this approach.

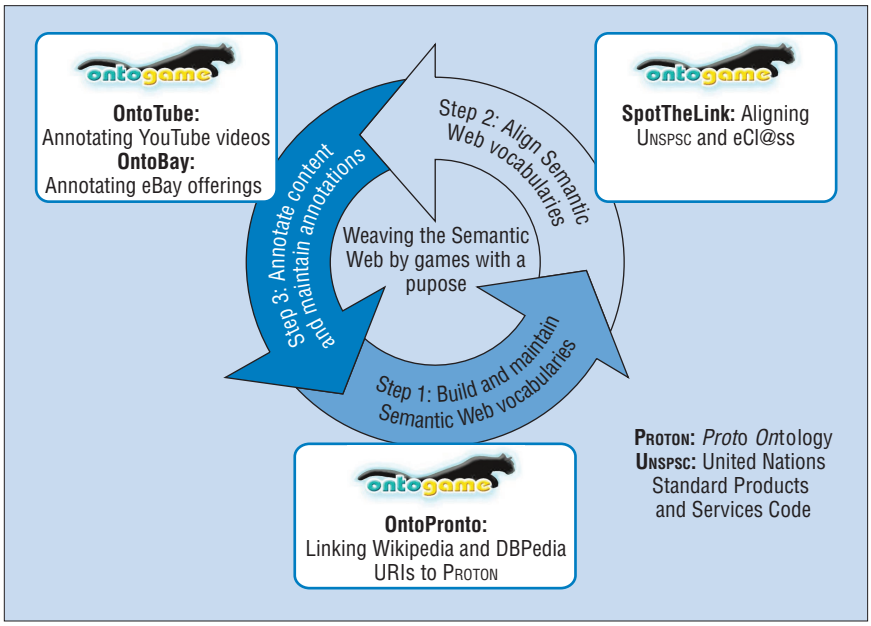

Figure 1. The OntoGame series of games aims to cover the complete Semantic Web life cycle. We work on games for the construction of ontologies (OntoPronto), the semantic annotation of data (OntoTube and OntoBay), and the alignment of ontologies (SpotTheLink).

We start with games that help create large, general-purpose domain ontologies by making Wikipedia or DBPedia entries[4] subclasses or ontologically significant individuals of PROTON (http://proton.semanticweb.org).[5] Then, we show how a game scenario can help specify semantic alignments between two large ontologies derived from classification schemas for products and services—namely, eClassOWL[6,7] and unspscOWL, an upcoming OWL variant of the popular UN Standard Products and Services Code (UNSPSC) classification. Ellen Schulten and her colleagues have described interoperability between those two standards (with more than 20,000 classes) as a major e-business interoperability challenge that an ontology infrastructure might solve.[8] The third set of games aims at annotating Web resources, such as eBay offerings or YouTube videos.

Games for ontology building and maintenance

At present, several tasks involved in constructing ontologies can't be completely automated. One or more of these tasks can be hidden behind an online game to make players unknowingly build ontologies. We describe several of these tasks here (see Mike Uschold and Martin King's classic article for an overview[9]).

Collecting named entities. We must identify and informally describe relevant conceptual elements of the respective domain and assign a unique key. These elements can be entity types and their attributes or relationship types. Respective tasks include producing a list of entity types or spotting names of missing attributes for a known concept.

Typing named entities. We must determine the type of conceptual element according to the distinctions of the applicable ontology metamodel for each named entity. For example, many popular ontology metamodels support classes, properties, and individuals as core types, and OWL DL and OWL Lite require the sets of individuals and classes to be disjoint. So, in OWL DL, we must decide whether each conceptual element is a class or an instance, even though we could argue that a conceptual element can serve as both—that is, these are just two possible roles for an element.

Adding taxonomic and nontaxonomic relations. We can enrich a flat collection of

ontological elements by adding taxonomic and nontaxonomic relations. The most prominent form of this task is arranging the concepts into a subsumption hierarchy by introducing subClassOf relations.

Axiomatization. Depending on the ontology's degree of expressivity, formally constraining the interpretation of ontology elements is often desirable. For example, we might want to include disjointness axioms.

Modularization. For large domains, defining subsets of concepts—either based on their ontological nature or by target application—can make the vocabulary more manageable.

Lexical enrichment. Ontology engineering methodologies tend to focus on formal means for specifying ontologies, but we also need informal means to describe the ontology elements' intended semantics, such as natural language labels or synonyms. However, relating a conceptual element to terms or synonym sets requires careful human judgment. Without it, inconsistencies between the ontology's informal and formal parts might result. An example of this task is augmenting an existing ontology with references to WordNet synsets. Translating natural language definitions into foreign languages is another interesting game scenario.

Games for ontology alignment

In an open environment such as the Web, multiple, partly overlapping ontologies will inevitably evolve and be used. To improve access to the related information, the elements of overlapping ontologies must be aligned. Because ontologies evolve with conceptual dynamics in domains and our understanding of the world, aligning these elements is a continuous rather than a one-time task. Jérôme Euzenat and Pavel Shvaiko distinguish four ontology-matching techniques:[10]

- *terminological* techniques that rely on lexical resources in the ontology,
- *structural* techniques that focus on the relations between entities (that is, ontology elements),
- *extensional* techniques that compare entity extensions, and
- *semantic* techniques that exploit formalized knowledge.

Despite significant advancement in automatic matching of ontologies, current systems can't perfectly match real-world ontologies automatically. The less formal the input ontologies, the less likely a machine will be able to reliably determine the proper semantic relationships between elements from two ontologies. Again, this recommends online games as a paradigm for soliciting the respective contributions from humans.

Online games could, for example, support the following three core tasks:

- *Spotting the most closely related conceptual element in a given second ontology.* Given a conceptual element of one ontology, the players must find the closest conceptual element in another ontology.

> The less formal the input ontologies, the less likely a machine will be able to reliably determine the proper semantic relationships between elements from two ontologies.

- *Selecting the most specific type of semantic relationship between two conceptual elements.* Given a pair of conceptual elements, the players must agree on the most specific semantic relationship between them. For example, players must spot whether two classes are truly equivalent or only partly overlapping, or players must spot the relationship between an attribute "age" and a class "adult."
- *Validating the implications of a given semantic relationship between two conceptual elements.* Given a pair of conceptual elements and a suggested semantic relationship between them, users must correctly spot this choice's semantic implications. We could, for example, translate the implications into phrases in simplified English and ask users to confirm whether these implications hold in all cases.

We can use these tasks in combination—for example, to ground a domain ontology in a standard top-level ontology such as Proton or to map two ontologies for the same domain.

An important issue in this type of games is the trade-off between expressivity and precision on one hand and the suitability for a large audience on the other. Most researchers with experience in conceptual modeling know the difficulties of teaching the subtleties of subClassOf relations versus parthood, or the differences between narrowerThan and subClassOf relations.

Broadly, we can use the most relevant set-theoretic relations, such as equivalence (=), more general ($\supseteq$), disjointness ($\perp$), and subsumption ($\subseteq$),[10] or instead limit the choice to less-specific relations used in the context of thesauri—for example, simple broaderThan/narrowerThan relations. Currently, we follow the latter approach and use a subset of the alignment relations defined by the W3C's Simple Knowledge Organization System (SKOS, www.w3.org/2004/02/skos) because they're more suitable for broad audiences. The subset comprises these relations:

- *equivalent* (=);
- *broaderThan*, a concept that's in some way more general in meaning;
- *narrowerThan*, a concept that's in some contexts more specific than another concept, without implying subsumption;
- *related*, some degree of semantic proximity (that is, an associative semantic relationship exists); and
- *partlyOverlappingWith*, in which an overlap exists in these concepts' meaning but neither concept is a proper subset of the other.

As an extension, we're investigating using strict subClassOf and disjointness ($\perp$) relations for large audiences. We hope to use paraphrasing in simplified English to make these relations comprehensible to lay audiences. As for the existing data sources, we can use respective games to align ontologies in the Swoogle (http://swoogle.umbc.edu) or Watson (http://watson.kmi.open.ac.uk) repositories or to establish a mapping between eClassOWL and unspscOWL.[6–8]

Games for annotating content

Content annotation involves describing existing resources' semantics using existing ontologies' vocabularies. This is true not only for manual approaches but also for creating the training sets that semiautomatic

annotation methods require. Resources can range from Web pages, images, videos, and sound files to Web services, and the description can refer to the resource's functional and nonfunctional properties.

Games that support content annotation would present a relevant resource or excerpt thereof, randomly selected from the set of resources, and ask players to describe the resource's respective aspect using choices in the game that map to elements in a suitable domain ontology.

Such games could involve

- presenting Google pages for a given keyword and having players describe it using a form based on a subset of the PROTON ontology;
- presenting eBay offerings and having players describe the type of goods being offered using references to eClassOWL; and
- presenting YouTube movies and having players describe genre, actors, and topics. To describe actors, we could use the subset of DBPedia[4] elements that are classified as actors, establishing a link between actors and their films.

The OntoGame series of games

Instead of hard-coding a small set of gaming scenarios, we developed a generic gaming platform for Semantic Web games. This lets us implement new scenarios or modify existing ones quickly. In particular, we hope to reuse a substantial amount of functionality. We also decided to store all game-related data in a native RDF store. This simplifies integrating gaming results from multiple games.

Each game is an online game for two players. A player is teamed with an anonymous partner, which prevents cheating because the players can't communicate directly. The system shows both players a modeling choice or other type of challenge. As much as possible, we hide the abstract flavor of conceptual modeling by using natural language, illustrations, or examples. Next, both players select a concept or an instance from a finite set of choices. If both players make the same choice, they earn points and move to the next challenge. Most challenges are multilevel—that is, after players reach consensus on the first level, they receive a question at a higher level of detail. Players can try to master this next, more difficult level, or they can skip it, while still keeping the points they've earned. To maximize the game output's quality, we encourage users to skip a challenge rather than make a wild guess. So, we don't penalize players for skipping a challenge by, for example, deducting the points earned so far.

The scheme for awarding points to challenges can vary by game scenario. Generally, we award more points for more difficult tasks. For example, when classifying an instance, finding consensus on the deepest PROTON class in the fifth level of the subsumption hierarchy earns more points than mastering only the first challenge. So, it pays to try to master the difficult challenges.

Each game round lasts a fixed amount of time; players try to solve as many challenges as possible during that time. In some scenarios, the system can offer an extra time credit to make especially difficult challenges more rewarding.

> The scheme for awarding points to challenges can vary by game scenario. Generally, we award more points for more difficult tasks.

Although we try to ensure that the number of human players is always even, it's possible that there won't be a partner for a lone player wanting to play. We therefore implemented a single-player mode in which the player plays against previous games' challenges and user input. However, people are more motivated when they know they're playing against a real human being in real time. So, a functionality lets members from our development team wait in an idle mode for single players wanting to play.

OntoPronto: Turning Wikipedia into a huge domain ontology

The URIs of the more than 1.8 million entries of the English Wikipedia are reliable identifiers for countless useful conceptual entities.[1] For example, Wikipedia contains more than 220,000 URIs for types of products and services, so it's eight times larger than eCl@ss or UNSPSC, the two largest categories for products and services. Parallel to this approach, Sören Auer and Jens Lehmann derived formal knowledge structures from Wikipedia, harvesting all named entities and part of the semantic content in RDF under the brand DBPedia.[4]

By grounding those 1.8 million conceptual elements properly in PROTON, we can create the largest general-interest ontology for annotating Web resources—1.8 million identifiers for everything from artists to high schools and from products to organizations—and are then able to use all PROTON generalizations in queries. We use PROTON because the specific Wikipedia concepts will benefit from a general, easy-to-understand upper ontology. DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering, www.loa-cnr.it/DOLCE.html) and SUMO (Suggested Upper Merged Ontology, www.ontologyportal.org) are available alternatives but seemed more difficult for lay audiences to understand. Because PROTON is expressed in OWL DL, we'll release our results in OWL DL. This creates the following challenges. First, OWL DL and OWL Lite require disjoint sets of individuals and classes. So, we must decide whether each Wikipedia entry element is more important as a class or an instance. Although we could argue that a conceptual element can serve as both, OWL DL requires us to make the choice. So, in practice, we must decide whether the entity's more relevant role is grouping similar objects (that is, serving as a class) or referring to a single conceptual entity (that is, serving as an instance).

Game scenario. In OntoPronto, players see an excerpt—the first paragraph of a randomly chosen Wikipedia article (http://en.wikipedia.org/wiki/Special:Random). They must judge whether the most relevant use of this entry is to refer to sets of objects (a class) or to a single object (an individual). For example, "Olympic Games" is more important as the set of Olympic games, whereas "John Lennon" is more important as a single entity. If the players reach consensus, they earn points and move to the next step, in which they must agree, iteratively, on the most specific PROTON class related to the current entry. Figure 2 shows the game's two phases.

Players receive 20 points for agreeing on whether the entry is more useful as a class or

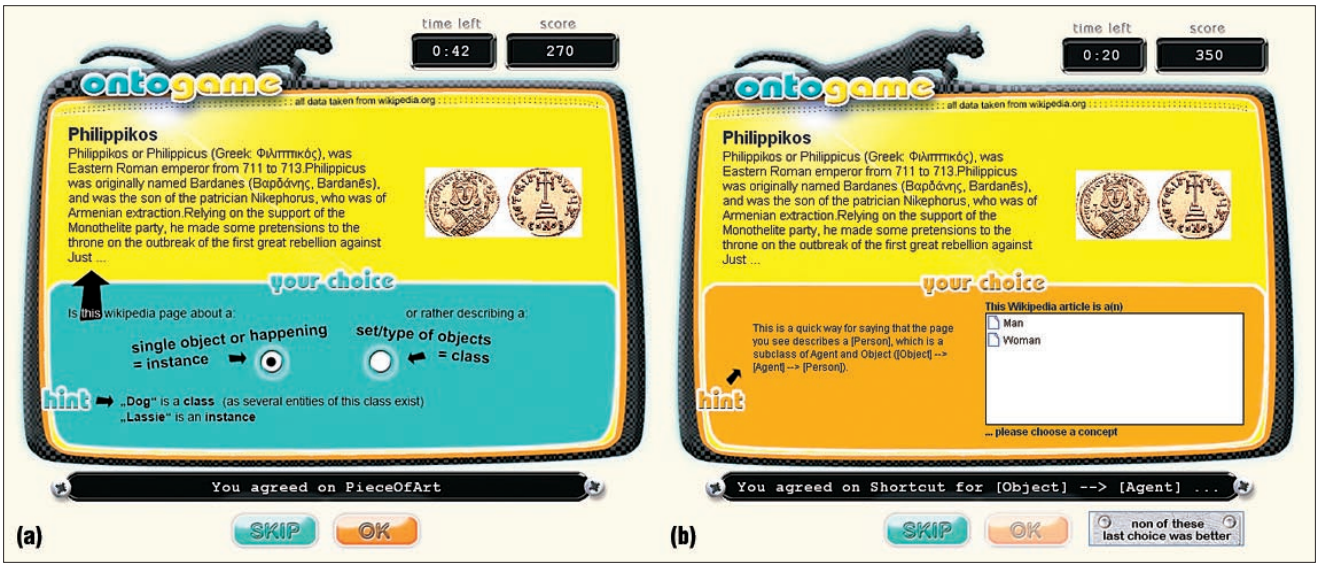

Figure 2. Phases in the OntoPronto game. (a) In phase 1, players must agree whether the most useful ontological role of a Wikipedia article is to serve as class or as individual. (b) In phase 2, they must consensually choose the most specific Proton branch related to that article.

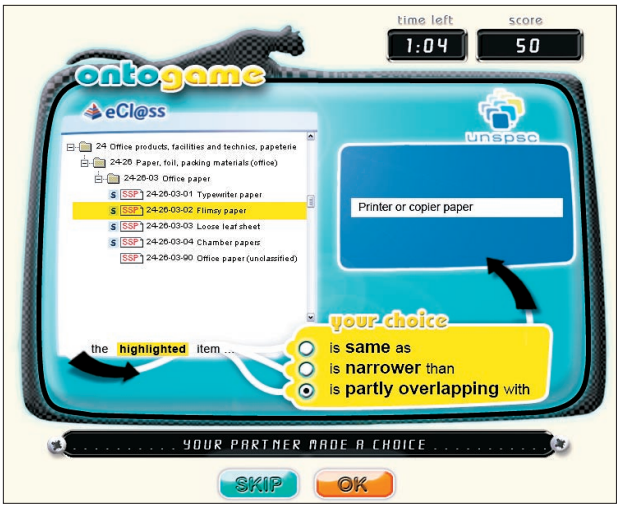

Figure 3. SpotTheLink: Mapping the UN Standard Products and Services Code (Unspsc) and eCl@ss. Players must select a related class in eCl@ss and the most suitable type of relation.

as an individual. Consensus on the first level of Proton buys them additional 10 points; on the second level, 20 points; on the third, 30; and so on. The more Wikipedia articles the team classifies consensually within two minutes, the more points they earn. A chosen class might have subclasses that are unsuitable for the Wikipedia article. For such cases, we give players the option of choosing "None of these—last choice was best." The points from all game rounds accumulate for each player, so players must continually return and continue to play to keep their ranking among the top players.

Intellectual challenge. The game's main challenges are abstracting from a concrete thing to a generalization and judging whether a page stands for multiple (tangible or intangible) objects or for a single object. Positive side effects are that players learn about different topics when looking at Wikipedia articles and that they become familiar with the Proton ontology.

Through a nightly batch run that implements filtering and postprocessing, we derive a large OWL DL ontology from the consensual choices. This ontology is available at www.ontogame.org/ontologies.

SpotTheLink: Mapping eCl@ss and the Unspsc

eCl@ss and Unspsc are the two most important product and service categorization standards, and establishing mappings between them for achieving data interoperability is one of semantic technology's oldest candidate applications.[8]

Game scenario. In this game (see figure 3), the players are presented a randomly chosen class from Unspsc, a set of possible relations, and the eCl@ss tree. The challenge is to reach consensus on a class from eCl@ss and the most specific kind of relation between the Unspsc and the eCl@ss classes. As we mentioned earlier, we derived the relations from traditional mapping relations[10] and the SKOS vocabulary: *sameAs*, *narrowerThan*, and *partlyOverlappingWith*. Before choosing an eCl@ss branch, players can view the branch's subclasses to better understand that branch. Players can choose multiple classes. SpotTheLink is currently in an early prototype stage only.

Intellectual challenge. The game's challenge is to navigate to the suitable branch in eCl@ss and choose the relation between two classes in accordance with your partner. We'll export the resulting alignments between eClassOWL and unspscOWL to the general public, using a threshold mechanism for filtering those alignments that are meaningful and most likely correct.

OntoTube: Annotating YouTube

According to the *Wall Street Journal*, YouTube hosted more than 6 million videos in

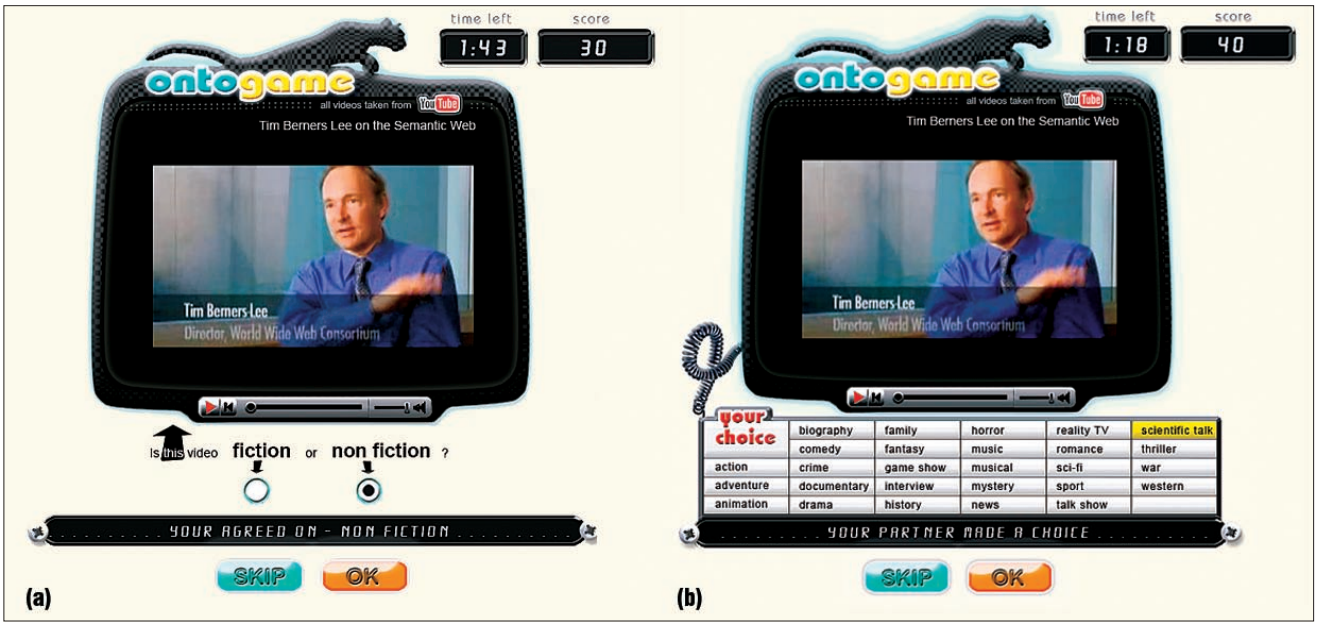

Figure 4. OntoTube: (a) Players must choose whether a video is fiction or nonfiction, a task that is usually trivial for a human user; and (b) players must agree on the topic of the YouTube video.

2006, and the total time people spent watching YouTube videos in its first year is equal to 9,305 years.[11] Obviously, YouTube has a lot of content, but the amount of available metadata is limited. A rich semantic annotation of YouTube content would therefore be useful. For example, we could establish links between people, topics, or locations related to a YouTube video and the respective Wikipedia or DBPedia entry. This would let a user find all videos showing New York or featuring John Lennon, thus combining Wikipedia and YouTube content.

Game scenario. In OntoTube, the players view a randomly chosen YouTube video, which starts playing immediately but can be stopped or fast-forwarded at any moment. For each video, the players must agree on answers for a set of questions derived from the video content ontology (see table 1). The more questions the players answer consensually, the more points they earn. The number of points depends on the question's difficulty—that is, players earn more points for achieving consensus on a video's general topic than on whether the video is black-and-white or color. Figure 4 shows two screenshots of this prototype.

For this scenario, we developed a simple domain ontology (available at www.ontogame.org/evaluation) that describes video content, derived from MPEG-7 (www.chiariglione.org/MPEG/standards/mpeg-7/mpeg-7.htm) and the Internet Movie Database (www.imdb.com). IMDB has a huge user base, and we're interested in what users are searching for when they search for videos. However, IMDB focuses on movies that were usually published by big production companies, whereas YouTube largely contains user-generated video content, which is a new type of video content. Therefore, we're cooperating with media scientists working on a classification of user-generated content to improve the underlying ontology of the OntoTube game.

Table 1. Questions regarding videos in OntoTube.

Question	Points
Is the video fiction or nonfiction?	10
Is the video black-and-white or color? (The system can of course determine this automatically by looking at the movie's color data; however, we can use the answers later to identify wrong or fraudulent user input.)	10
The video's genre can be best described as _____. (Players choose answers from a list of 27 genres, as used by the Internet Movie Database.)	30
Was the video produced by a private person or by a company? (The players can also answer that this question doesn't apply.)	10
The language of the video is _____. (Players choose from a list of languages including "no language.")	20
Generally, the video is about _____. (Players can select a Wikipedia article that represents what the video is about.)	40
The location of the video is _____. (Players select a Wikipedia article representing the location of the video.)	40
The time period the video plays in is _____. (Players choose from a list of time periods.)	40

Intellectual challenge. Players must quickly grasp the video's content. Because the games last only two minutes, players must extract the respective facts without wasting too much time. In addition, each player has individual controls (play and fast-forward). Some questions are relatively easy, guaranteeing that users experience success. Other questions

Survey Questionnaire

In addition to analyzing the data collected through the games, we posted an online survey, which 35 players completed:

1. Tell us about your background (several choices possible):
   - I have a background in the Semantic Web.
   - I have a background in computer science.
   - I do not have a technology-related background.
   - None of the above applies.
2. How often have you already played OntoGame?
   - More than 20 times.
   - Between 5 and 20 times.
   - Less than 5 times.
3. Assess the overall gaming fun of OntoGame:
   - Addictive
   - Cool
   - So-so
   - Boring
4. Were the rules of the game hard to understand? (Yes/No)
5. The intellectual challenge of the game is …
   - Too easy
   - Too hard
   - Demanding
   - Easy
6. Would you play the game again? (Yes/No/Maybe)
7. What do you like/dislike about OntoGame? (free text)

Table 2. Number of games per player (n = 271).

Number of rounds played	Number of players	Percentage
0	53	19.6
1	102	37.6
2	47	17.3
3	69	25.5

Table 3. Number of game rounds played—comparing log data and survey results.

Number of game rounds played	Log data		Survey
	Population	Percentage	Percentage
Between 0 and 4	231	85.2	56
Between 5 and 20	36	13.3	34
More than 21	4	1.5	9

are more difficult because of a wide range of possible choices. For describing a video's topic and location, the game lets users select appropriate Wikipedia URIs, linking the game more closely with OntoPronto.

You can play the game at www.ontogame.org/ontotube. We'll make the annotations available for access via HTTP shortly.

OntoBay: Annotating eBay offerings

Another game scenario that's currently under development is OntoBay (see Figure 1), which aims at annotating eBay offerings with the eCl@ss standard to allow more sophisticated search by categories or features.

Evaluation

We performed a preliminary analysis to evaluate

- whether the prototype creates a positive gaming experience and can motivate users to spend time playing, and
- whether players' consensual conceptual choices are of sufficient quality for the Semantic Web at large.

Overview and user data

We released OntoPronto to the general public on 16 December 2007. On 31 December 2007, we saved a snapshot of the data collected by the game (all raw data and additional material are available at www.ontogame.org/evaluation). In total, 271 players registered; of these, 240 said they were male (89 percent) and 31 said they were female (11 percent).

An online survey among registered players complemented the analysis of the data collected during the games. The survey consisted of seven questions for assessing the fun factor, rule comprehensibility, intellectual challenge, and other aspects (see the "Survey Questionnaire" sidebar for the original questions). The online survey was available for eight days, and 35 players completed it. Evidence suggests that the players in the sample aren't representative of the population of all players, but the data can still serve as a preliminary indication.

Gaming fun and motivation

To get a tentative understanding of whether OntoPronto creates sufficient gaming fun, we analyzed how much time each user spent playing the games. Table 2 summarizes the number of rounds per player. More than 80 percent (218) of those who registered tried the game at least once. Of these players, more than 50 percent (47 + 69 = 116) played at least twice, and 32 percent (69) of those who tried the game at least once played three or more rounds. Because this data is based on the full log data of all rounds, it suggests that even our early prototype can create a substantial amount of gaming fun.

A related question in the survey let us estimate the degree of bias among survey participants. According to their answers, 56 percent of them played the games fewer than five times, 34 percent between five and 20 times, and 9 percent more than 20 times. From the log files, we know that the population contains more people who tried the game never or fewer than five times.

As you might expect, happy players are a bit overrepresented in the survey (see Table 3), but the basic patterns seem similar.

Even with our small-scale example, we acquired more than 35 hours of human labor for conceptual-modeling tasks over a period

of two weeks and derived an extension to PROTON with more than 300 new elements.

As for the remaining survey questions, almost all the survey participants (81 percent) found the game's rules easy to understand. Forty-one percent found the game's fundamental intellectual challenge demanding, 44 percent found it easy, 9 percent found it too easy, and 6 percent found it too hard. Nine percent said that the overall gaming fun was "addictive," 41 percent rated the game "cool," 44 percent found it to be "so-so," and only 6 percent found it "boring." Forty-seven percent indicated that they would play the game again, 47 percent said that they might play again, and only 6 percent indicated that they wouldn't play again.

The last question gave users a chance to comment. Generally, the feedback was positive. One participant said he loved the game, many emphasized that they liked the idea very much, and some described the game as a lot of fun. Two participants especially liked reading the Wikipedia articles and learning about a wide range of topics. One participant described "learning PROTON" as a nice side effect. Several commented that a PROTON tutorial would help them get to know the hierarchy better, especially for articles that are more difficult to classify (for example, abstract things).

Some players mentioned that they enjoyed playing with a human partner but that single-player mode wasn't interesting and was easy to recognize. Some survey respondents indicated that they would like to know more about their partners. We've already addressed this request: from now on, the game will display partners' nationalities and age (if specified during registration). We're also working on a functionality that lets players communicate with each other after the game and reveal contact information if desired. This correlates to von Ahn's experience with the ESP game. It's also in line with a comment by one participant, who felt that increasing the social aspect of the game—that is, allowing for more human-to-human communication—would increase the game fun.

One participant described OntoGame as "constructive entertainment," another as "simple and straightforward." Some participants complained that some pages were repeated (especially in single-player mode), which was a flaw in the early version. Some also said that similar classifications (such as persons, football teams, and cities) become dull when repeated. This is partly because 23 percent of all Wikipedia entries relate to persons.[1] As a first solution, we introduced shortcuts to frequently needed PROTON concepts.

Table 4. Degree of consensus per challenge (n = 2,905).

Degree of consensus	Number of challenges	Percentage
Teams either skipped the first task or came to no consensus about it.	374	12.9
Teams completed only the first task consensually.	297	10.2
Teams completed both tasks consensually.	2,234	76.9
Teams completed both tasks consensually, with consensus on the leaf level.	1,291	44.4

Table 5. Conceptual quality of consensual solutions (n = 2,234).

Quality of the solutions	Number of challenges	Percentage
Teams made the correct consensual choice for tasks 1 and 2.	2,172	97.2
Teams found the correct PROTON concept but made an incorrect decision on class versus instance.	12	0.54
Teams made the correct decision on class versus instance but didn't find a valid PROTON concept.	50	2.24

Consensus

Given the subject's complexity, we evaluated how easily users reached consensus (see table 4). In 13 percent of the cases, players either skipped the first task (class versus instance) or didn't agree. Of approximately 10 percent of all challenges played, players finished only the first task consensually. In almost 77 percent of the challenges, the teams completed both tasks consensually. Even though not all teams achieved leaf-level consensus, the consensual solutions produced by multiple teams were always in the same branch—varying only by depth in the hierarchy.

Conceptual quality of user choices

Finally, we wanted to know the quality of the conceptual choices made in those challenges in which players completed both tasks consensually (that is, of the 2,234 challenges for which the players agreed both on class versus instance and on a PROTON class). For each of the 365 Wikipedia entries played at least once, we determined the set of all chosen PROTON classes and then judged manually whether those choices were feasible. Some choices allowed room for argument; if even an expert was in doubt, we assumed the decision was correct. Also, for us, a PROTON class that's correct but not the most specific counts as a correct choice (for example, some players classified John Lennon as a person instead of a male person). We also judged whether the decision as to the entry's most relevant role (class versus instance) was correct. Table 5 and figure 5 summarize the results. As they show, the share of clearly wrong choices was only 2.8 percent (62 challenges).

A significant share of wrong choices was due to the fact that the game originally didn't let players keep the current PROTON class if subclasses existed—they could only skip or select a subclass in case further specializations of the superclass were available. Before we introduced the "last was best" option, many players always played to the last level. Consequently, players classified cities as a capital or local capital in 35 challenges even though those cities weren't a capital, just because they couldn't select the more general choice "city" except by the unintuitive use of the skip button. Another popular mistake (made six times) was classifying a class of plants or animals as an instance. For example, some players classified the fish species *Indian whiting* as individual. To prevent this kind of mistake, we've improved the task description. Finally, we also counted as incorrect when players didn't really understand the meaning of an article or PROTON concept. For instance, some participants classified a bank or a radio station as

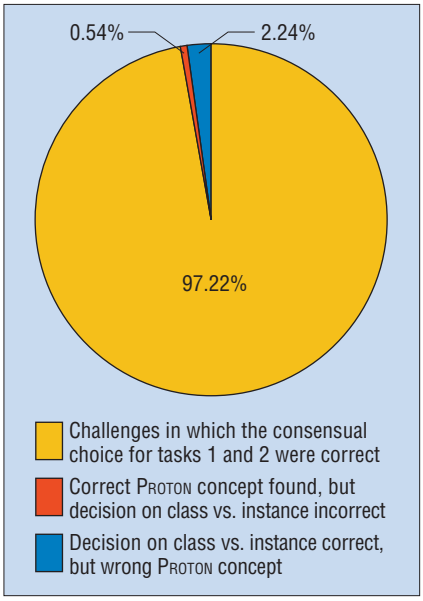

Figure 5. Conceptual quality of complete consensual solutions (*n* = 2,234). Most games were completed with a correct consensual choice, both judging the ontological nature and selecting a PROTON class.

services, when actually they're commercial institutions that provide services. Increasing the game's functionality should help players avoid these typical pitfalls.

To extend this evaluation, we'll check how many of these pitfalls our ontology export algorithm will filter out—that is, whether there are cases in which the most popular judgments for the same entry were wrong. We'll publish our full raw data on www.ontogame.org/ontologies in the near future.

The games we've presented are the first prototypes of the OntoGame series. We're extending and improving the scenarios in several directions.

We intend to further motivate players by emphasizing the games' social component. When players reach a certain number of points, for example, they can reveal information about themselves to their partners (for example, their age and location). To increase our pool of players, we'll give users points for inviting others to play. An adaptive points system—that is, players earn more points for difficult tasks—could make especially difficult challenges more attractive. Another extension is a more meaningful partner selection. Currently, the system pairs partners randomly. However, the games could take user preferences (for example, "I want to play only with players from Florida") into account to create a more enjoyable gaming experience.

For most players, the games are far more entertaining when played with a real person. To extend live mode and address the problem of always needing many players to be online, we're working on an offline mode in which a user can play some rounds of the game and then have his or her partner (who is notified by email, as in remote chess) complete those rounds later. We're also experimenting with allowing more than two players on a team.

In addition, we're working on autocompletion and suggestion functions to avoid friction (and frustration) due to lack of consensus between players because of lexical variants or spelling mistakes. We can even extend this to a combination of user input and machine learning, leading to a true combination of human and computer intelligence. In addition, especially for the OntoTube scenario, we must deal with inappropriate content. YouTube already allows marking such content, but we plan to integrate functionality for flagging it. Finally, scalability is important for increasing gaming fun.

You can always find our most recent work at www.ontogame.org.

The Authors

Katharina Siorpaes is a researcher and PhD student at the Semantic Technology Institute (STI) at the University of Innsbruck. As part of her PhD research, she's using games for semantic content creation and for providing an incentive for users to weave the Semantic Web. Furthermore, she's analyzing technology for the community-grounded evolution of ontologies. Siorpaes received her MSc in computer science from the University of Innsbruck. Contact her at katharina.siorpaes@sti2.at; www.katharinasiorpaes.net.

Martin Hepp is a professor of general management and e-business at the Bundeswehr University Munich, where he heads the E-Business and Web Science research group, and a professor of computer science at the University of Innsbruck, where he heads the Semantics in Business Information Systems research group. He created eClassOWL, the first industry-strength ontology for products and services, and is working on using Semantic Web technology for business applications—namely, e-commerce and enterprise-resource-planning systems management. Hepp received his PhD in business information systems from the University of Würzburg. He's a member of the IEEE Computer Society. Contact him at mhepp@computer.org; www.heppnetz.de.

Acknowledgments

We thank Werner Huber, Michael Waltl, and Roberta Hart-Hilber for their support in building our game prototypes, and Bruno Haid from System One for valuable feedback and support. The Austrian BMVIT/FFG (Bundesministerium für Verkehr, Innovatio, und Technologie/Forschungsförderungsgesellschaft) funded this work under the FIT-IT (Forschung, Innovation, Technology - Informationstechnologie) Semantic Systems project myOntology (grant number 812515/9284), www.myontology.org.

References

1. M. Hepp, K. Siorpaes, and D. Bachlechner, "Harvesting Wiki Consensus: Using Wikipedia Entries as Vocabulary for Knowledge Management," *IEEE Internet Computing*, vol. 11, no. 5, 2007, pp. 54–65.
2. S. Kuznetsov, "Motivations of Contributors to Wikipedia," *ACM SIGCAS Computers and Soc.*, vol. 36, no. 2, 2006, article 1.
3. J. Surowiecki, *The Wisdom of Crowds*, Anchor Books, 2003.
4. S. Auer and J. Lehmann, "What Have Innsbruck und Leipzig in Common? Extracting Semantics from Wiki Content," *Proc. European Semantic Web Conf.* (ESWC 07), LNCS 4519, Springer, 2007, pp. 503–517.
5. I. Terziev, A. Kiryakov, and D. Manov, *Base Upper-level Ontology (BULO) Guidance*, SEKT Deliverable D1.8.1, SEKT Consortium, 2005, http://proton.semantic-web.org/D1_8_1.pdf.
6. M. Hepp, "Products and Services Ontologies: A Methodology for Deriving OWL Ontologies from Industrial Categorization Standards," *Int'l J. Semantic Web and Information Systems*, vol. 2, no. 1, 2006, pp. 72–99.
7. M. Hepp, "eCl@ssOWL: The Products and Services Ontology," 2006, www.heppnetz.de/projects/eclassowl.
8. E. Schulten et al., "The E-Commerce Product Classification Challenge," *IEEE Intelligent Systems*, vol. 16, no. 4, 2001, pp. 86–89.
9. M. Uschold and M. King, "Towards a Methodology for Building Ontologies," *Proc. Workshop Basic Ontological Issues in Knowledge Sharing*, 1995, http://i-rescue.net/~oplan/documents/1995/95-ont-ijcai95-ont-method.pdf.
10. J. Euzenat and P. Shvaiko, *Ontology Matching*, Springer, 2007.
11. L. Gomes, "Will All of Us Get Our 15 Minutes on a YouTube Video?" *Wall Street J.*, 30 Aug. 2006.