Working the Crowd: Design Principles and Early Lessons from the Social-Semantic Web

Mathias Niepert
Indiana University
Department of Computer Science
mniepert@indiana.edu

Cameron Buckner
Indiana University
Department of Philosophy
cbuckner@indiana.edu

Colin Allen
Indiana University
Department of History and Philosophy of Science & Program in Cognitive Science
colallen@indiana.edu

ABSTRACT

The Indiana Philosophy Ontology (InPhO) project is presented as one of the first social-semantic web endeavors which aims to bootstrap feedback from users unskilled in ontology design into a precise representation of a specific domain. Our approach combines statistical text processing methods with expert feedback and logic programming approaches to create a dynamic semantic representation of the discipline of philosophy. We describe the basic principles and initial experimental results of our system.

General Terms

Social Semantic Web, Ontologies, Folksonomies, Provenance

1. INTRODUCTION

Until recently, research on the social web (Web 2.0) and semantic web has been largely segregated. This may not be surprising, as the two approaches seem to offer competing visions for the future of the Internet. Social web researchers devise ways to harness the "wisdom of the crowds" to structure web data around information obtained from collaborative social interactions between large numbers of amateur users. Semantic web researchers, on the other hand, emphasize the need for a technically precise backbone of formal ontologies developed by small groups of experts highly-trained in the best practices of ontology design. Cultural differences have further fueled misconceptions and misunderstandings between these two research communities, often leading them to regard one another with mutual skepticism.

Both approaches have had some striking successes. Web 2.0 applications like Wikipedia, Facebook, Del.icio.us, and Flickr have reshaped the way average users interact with the Web. A key strength of such approaches lies in their ability to obtain large amounts of information from unskilled volunteers and to combine information obtained from many different kinds of sources creatively. Such applications, however, face severe problems of data organization, validation, and integration, especially as they aspire to make data accessible and interoperable by organizing it according to semantic taxonomies. Some have proposed learning taxonomies from social tagging systems as a solution to this problem[2]. However, given that social tags are simply words applied to resources like documents and images, folksonomists have found themselves facing many of the same difficult problems that face researchers who try to induce taxonomies by processing natural language corpora. These problems include term ambiguity and the induced representation's lack of structural depth, precision, and reasoning capabilities.

While semantic web projects which impact the way the public is using the Web have largely failed to materialize, ontology-based approaches to data organization and integration have produced significant successes in certain domains, especially in bio- and medical informatics projects (such as the Gene Ontology) and in business applications. A factor severely hindering such approaches from being successfully applied to the Web at large, however, is that once elaborate and precise ontologies have been created, expertise in both ontology design and the relevant domain are required to populate and maintain them. Thus, semantic web projects have faced the dilemma of either hiring expensive "double experts" highly-skilled in both ontology design and the relevant domain or face inevitable data and user sparseness[3].

Fortunately, researchers are beginning to realize that not only is there no inherent opposition between these two approaches, but that their strengths and weaknesses are complementary[1, 5]. Thus, some have begun to call for the development of the "social-semantic" web, which would combine social web's facility for obtaining data from volunteer users with the semantic web's elegant and precise data representations. The combination of these two approaches faces its own unique set of problems, and large-scale social-semantic web projects which produce precise, high-quality data representation without presuming ontology design expertise of their users are still gleams in their future developers' eyes[4]. In this paper, however, we describe the Indiana Philosophy Ontology (InPhO) project as one of the first social-semantic web endeavors which aims to bootstrap feedback from users unskilled in ontology design into a precise representation of the domain. We will describe our ongoing solutions to some of the challenges facing this nascent area of research. At the InPhO project, we are developing a dynamic ontology for the domain of philosophy. This knowledge base is being deployed primarily to serve the metadata needs of the Stanford Encyclopedia of Philosophy (SEP) (although it has a wide array of other uses). Our approach combines statistical text processing with expert feedback to create a dynamic semantic representation of the entities described in the SEP's articles. While tagging approaches rely on users to spontaneously provide the needed feedback, our approach is based on the principle that if automated methods are used to guide users towards providing data which is most needed and for which they are most qualified, high-quality information can be obtained without placing undue demands on volunteer contributors.

2. INPHO: BASIC PRINCIPLES FOR A SOCIAL SEMANTIC WEB PROJECT

We believe that heavy user participation is key for social semantic web projects for keeping both the formal representation and its content up-to-date and of highest quality. In most cases, users experience top-down and static ontologies as too restrictive. Motivated by this consideration, we propose some basic principles for social semantic web projects which we strive to realize in the context of the InPhO project.

***Pragmatic* Ontology Design**
For many projects, especially those that rely on user participation, it is often unfeasible to design a *static* top-down ontology that models the targeted domain exhaustively. We believe that the social semantic web is better served by various specialized and dynamic ontologies that utilize semi-automated tools for information integration. Formal ontologies should be kept simple in the initial design phase and they should be iteratively and dynamically extended and populated through a combination of automated data processing methods, user feedback, and logical reasoning[9].

Ontology Extension as Iterative Relation Addition and Refinement
Many complex ontologies leave users bewildered by complications and thereby languish with huge sections almost entirely unpopulated. To ensure that data representations remain both relevant and well-populated, we believe that ontology design should be incremental and driven by user participation. For example, InPhO's *influenced-by* relation between philosophers can easily be populated by validating and integrating semi-structured data from Wikipedia[8]. However, the relation does not carry any specific information about what *kind* of influence and in which *area* of philosophy the influence took place. Hence, at later stages, one might decide to *refine* the relation by introducing a relation *influenced-in-area*, which relates an instance of the *influenced-by* relation to an instance of a philosophical area. Note that this is a form of tagging of pairs of entities. This is also supported by current W3C standards: OWL (and RDF in general) natively supports binary relations only, but allows several methods for modeling higher-order relations[1]. For example, the RDF standard allows relation instances to be treated as first-class citizens (reification). We believe that the pieces of information users are asked to provide should be kept as simple as possible and that the process should resemble the process of tagging. Projects that initially define intricate higher-order relations will have a hard time providing sufficient incentives for participation and will ultimately suffer from a lack of user contribution. Furthermore, we believe that formal ontologies (the set of relations and axioms) should grow with the practical needs of the individual semantic web projects and not vice versa.

Ontology Population as Iterative Data Addition, Validation, and Integration
Statistical text processing and other automated methods should be used to provide candidates for relation instances that can be verified and integrated using human feedback. The verification and addition of relation instances should resemble tagging as closely as possible. However, instead of tagging single web entities like documents, pictures, and videos, here *pairs* of entities are "tagged" with relationships

[1] http://www.w3.org/TR/swbp-n-aryRelations/

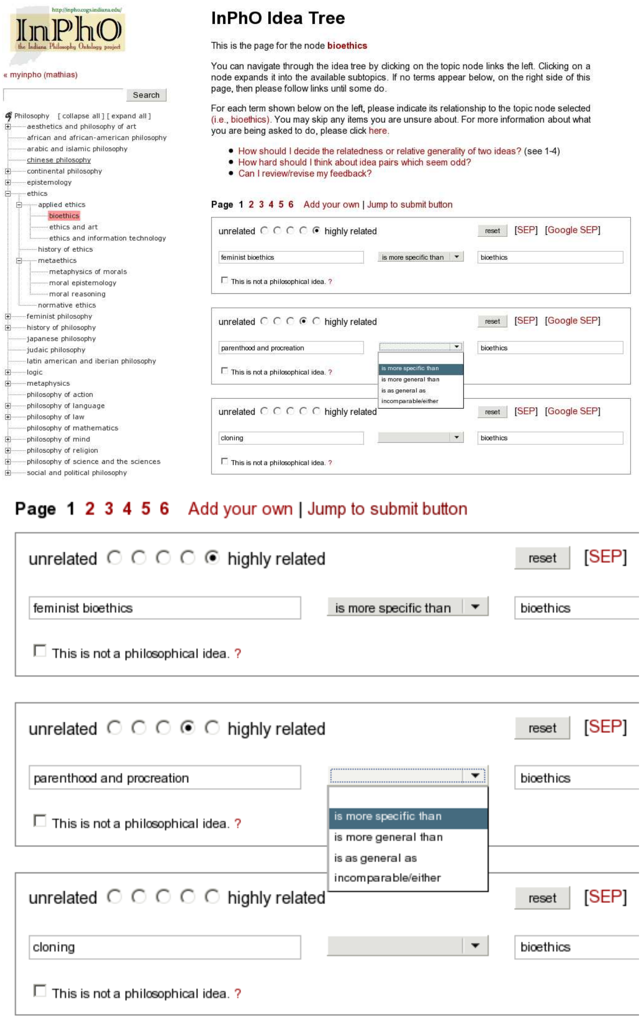

Figure 1: InPhO's "Idea Tree" interface which lets users quickly label relationships between pairs of philosophical ideas, ranked by statistical text processing algorithms.

holding between them, choosing from a predefined set of labels. For example, Figure 1 depicts one of InPhO's interfaces which provides users with pairs of philosophical ideas in their area of expertise for which they can evaluate the relatedness and relative generality. In addition, users should be able to add data in *batches* and have access to an API for data entry.

Stratified Participation; Provenance and Trust
Most Web 2.0 projects are powered by the "wisdom of the crowd," that is, many different users participating and collaborating to create large amounts of valuable (meta-)data. While we believe that large-scale semantic web projects will not succeed without leveraging the "wisdom of the crowd," we are also proponents of the position that the input of some users should be considered more trustworthy and reliable than others. InPhO allows users to provide areas of expertise in their personal profile and leverages this information to guide users to contribute in meaningful ways. Through InPhO's interfaces, all users are able to contribute to and populate the *uncertain* part of the ontology, and every piece of data is marked with detailed provenance information. When

logical reasoners are deployed to infer the taxonomic relationships, the provenance information is harnessed to resolve inconsistencies appropriately. For example, evaluations from users who are experts in this subfield of philosophy are valued higher than feedback from novice users. In addition, provenance information should be provided together with the instance data at all stages. For example, while birth and death date information is gathered by parsing external datasets and through contributions of InPhO's users, only the data verified by experts (i.e., authors and editors of the SEP) will be used as metadata for SEP entries.

Open Data Access and Open Community
Users should be able to download the populated ontology together with the provenance information and use it in external applications. An API should give direct access to write and read operations. The project's online community should be open to everyone and contributions should be visible and attributable to individual users.

3. INPHO: FIRST EXPERIENCES AND INITIAL RESULTS

As of now, the Indiana Philosophy Ontology[8] contains four main categories: person (subclass of FOAF::person[2]), document (from AKT[3]), organization (from SUMO[4]), and philosophical idea, as well as an initial set of non-taxonomic relations. The idea category contains a taxonomic decomposition of the space of philosophical ideas according to the disciplinary relatedness of their *contents* rather than according to their structural roles. For example, instead of dividing *idea about philosophy* into *concept*, *distinction*, *argument*, *counterexample*, and so on, the InPhO decomposes it into subareas of philosophy–e.g. *idea about metaphysics*, *idea about epistemology*, *idea about logic*, *idea about ethics*, *idea about philosophy of mind*. Each subarea is in turn decomposed into a series of issues considered fundamental to work in that subarea; for example, *idea about philosophy of mind* is decomposed into *idea about consciousness*, *idea about intentionality*, *idea about mental content*, *idea about philosophy of artificial intelligence*, *idea about philosophy of psychology*, and *idea about metaphysics of mind*. InPhO combines corpus-based measures of semantic similarity between words (for examples, see[7]) and a novel relative generality measure[8], to provide, for any given philosophical idea, a ranking of possible hyponyms and hypernyms, respectively (the interface is depicted in Figure 1). Using these carefully designed interfaces, InPhO's users can validate or falsify the estimates of semantic relatedness and relative generality of pairs of philosophical ideas, using a predefined set of possible labels. The relatedness is scored on a five-point scale from highly related to unrelated, and the generality can be evaluated using four different options: same level of generality, *idea*1 is more general than *idea*2, *idea*1 is more specific than *idea*2, and the two are incomparable. The generality of two ideas is deemed incomparable if they are entirely unrelated or if one idea can be both more *and* less general than the other, depending on the context. Of course, users may skip idea pairs or provide only partial information. The feedback is stored as first-order facts in our knowledge base, together with provenance data. For example, when a user with id 45 provides the information that an idea about neural networks is more specific than an idea about connectionism, and that they are highly related, the facts *msp(neural network, connectionism, 45)* and *s4p(neural network, connectionism, 45)* are added to the knowledge base. For each user, automatically computed trust scores and levels of expertise are stored to evaluate her reliability. A non-monotonic answer set program with stable model semantics is used daily on the set of first-order facts to construct the global populated ontology[9]. The taxonomy can be browsed online[5].

[2]http://xmlns.com/foaf/spec/
[3]http://www.aktors.org/publications/ontology/
[4]http://www.ontologyportal.org/

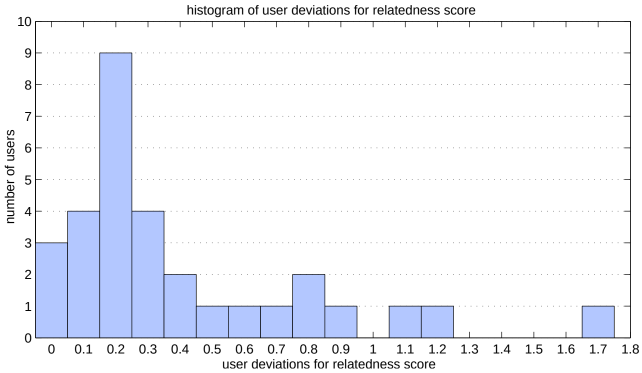

Figure 2: Histogram of deviations of relatedness scores among InPhO users with overlap $\geq$ 10.

	0 (%)	1 (%)	2 (%)	3 (%)	4 (%)
0	54 (3.8)	62 (4.4)	38 (2.7)	25 (1.8)	9 (0.6)
1	62 (4.4)	33 (2.4)	73 (5.2)	61 (4.3)	35 (2.5)
2	38 (2.7)	73 (5.2)	62 (4.4)	116(8.3)	84 (6.0)
3	25 (1.8)	61 (4.3)	116(8.3)	91 (6.5)	253(18.0)
4	9 (0.6)	35 (2.5)	84 (6.0)	253(18.0)	409(29.1)

Figure 3: Table depicting user agreement and disagreement on relatedness scores. Scores range from 0 (unrelated) to 4 (highly related). The entry in the i-th row and j-th column is the number of idea pairs that have been scored as i by one user and as j by a different user. The values in parentheses are the percentages with respect to all 1405 evaluations with overlap.

4. A FRAMEWORK FOR DATA-DRIVEN TRUST MEASURES

We introduce a general framework for the assignment of trust scores to individual users based on their deviation from other users' evaluations. A method to compute degrees of trustworthiness of users in a social network using semantic and social web data sources was recently proposed[6]. Here, we focus on trust scores that are computed using the users' evaluations of pairs of entities and their application to resolving feedback inconsistencies. Let U be the set of users, let A and B be two sets of individuals in the ontology, and let L be the set of possible *labels* that can be assigned to elements in $A \times B$. Let the *label distance* $dist : L \times L \rightarrow \mathbb{R}^+$ be a function that assigns to each pair of labels a non-negative real number. Let

[5]http://inpho.cogs.indiana.edu/taxonomy/

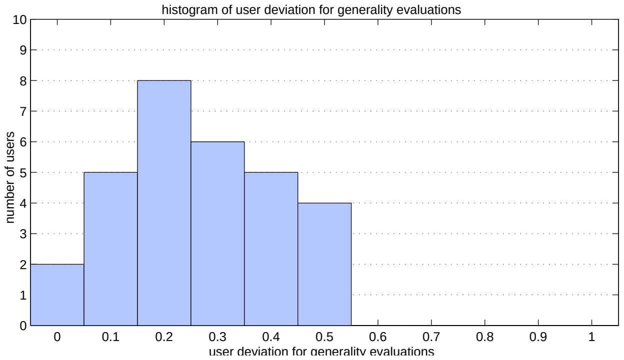

Figure 4: Histogram of users' deviation on relative generality labels with evaluation overlap $\geq$ 10.

	m.s. (%)	inc./e. (%)	same (%)	m.g. (%)
m.s.	489 (34.8)	127 (13.8)	79 (8.6)	33 (3.6)
inc./e.	127 (13.8)	19 (2.1)	37 (4.0)	32 (3.5)
same	79 (8.6)	37 (4.0)	35 (3.8)	49 (5.3)
m.g.	33 (3.6)	32 (3.5)	49 (5.3)	17 (1.9)

Figure 5: Table depicting user agreement and disagreement on generality evaluations. m.s.=more specific, m.g.=more general, same=same generality, inc./e.=incomparable/either more or less general, depending on the context. The values in parentheses are the percentages with respect to all 917 generality evaluations with overlap.

$E = \{(a, b, \ell, u) \mid a \in A, b \in B, \ell \in L, u \in U\}$ be the set of 4-tuples representing the user evaluations, that is, the assignments of labels in L to elements in $A \times B$ by the users in U. We define the *evaluation deviation* measure $D : U \to \mathbb{R}^+$ as

$$D(u) = \frac{1}{|N(u)|} \sum_{(a,b,\ell,u)\in E} \sum_{(a,b,\ell',u')\in E \text{ with } u\neq u'} dist(\ell, \ell'),$$

with $N(u) = \{(a, b, \ell', u') \in E \mid \exists (a, b, \ell, u) \in E \text{ with } u' \neq u\}$. Of course, the smaller the evaluation deviation, the higher the trust one can have in a particular user. The trust scores (some of which might be specialized to specific areas in philosophy) can then be used together with the users levels of expertise to enhance provenance information and settle feedback inconsistencies with increasing sophistication.

Initial Experimental Results

As of March 25th 2009, InPhO (currently in beta testing) has 92 registered users, 36 of which provided one or more of the 4,653 evaluations of 2,969 distinct pairs of ideas. The set of users consists of volunteers who registered after the InPhO system had been announced on several e-mail newsletters and blogs. They will soon be joined by the authors and editors of the Stanford Encyclopedia of Philosophy. 39 out of the 92 users have the highest level of expertise (published in the area) and 37 finished a graduate class in the area. From the 47 subareas of philosophy that are currently specified in the InPhO, 31 were covered by at least one expert. The contribution incentives are twofold: (1) users have their own personal account that displays type and number of contributions and several agreement statistics and (2) the more feedback a SEP author provides the better is her entry embedded in browse and search applications. However, we consider the objective of providing sufficient incentives for user participation an ongoing research and interface design challenge.

We are specifically interested in the extent of user agreement on evaluations of idea pairs with semantic relatedness and relative generality labels. Thus, in the remainder of the paper, A and B are the instances of the class *philosophical idea* in the ontology. Users can score the semantic relatedness of two philosophical ideas on a scale from 0 (unrelated) to 4 (highly related). Hence, for the relatedness score we have $L = \{0, 1, 2, 3, 4\}$ and $dist(\ell, \ell') = |\ell - \ell'|$. Figure 2 depicts the histogram of the *evaluation deviation* values for the 31 users who labeled the relatedness of one or more idea pairs that have also been evaluated by at least 10 other users (evaluation overlap $\geq$ 10). Except for some outliers, the majority of the users has a deviation of less than 0.5 where 4.0 is the possible maximum. Figure 3 shows the overall user agreement and disagreement. For example, only 9 out of 1405 overlapping evaluations (0.6%) have a label distance of 4, and 1153 out of 1405 overlapping evaluations (82%) have label distance of 1 or 0.

For the relative generality evaluations, $L = \{0, 1, 2, 3\}$ with 0="more specific", 1="more general", 2="same generality," and 3="incomparable/either more or less general." Here, we can define *dist* as $dist(\ell, \ell') = 1$ if $\ell \neq \ell'$ and $dist(\ell, \ell') = 0$ otherwise. Figure 4 depicts the histogram of the *evaluation deviation* values for the 30 users who labeled the relative generality of one or more idea pairs that have also been evaluated by at least 10 other users. All users have a deviation of less or equal than 0.5 where 1.0 is the possible maximum. Figure 5 shows the overall user (dis-)agreement on generality labels. For example, 489 out of 917 overlapping evaluations (52%) agree on the label "more specific", and there are only 33 overlapping evaluations (3.6%) with disagreeing labels "more specific" and "more general."

5. REFERENCES

[1] A. Ankolekar, M. Krötzsch, D. T. Tran, and D. Vrandecic. The two cultures: Mashing up web 2.0 and the semantic web. *Journal of Web Semantics*, 6(1):70–75, 2008.

[2] D. Benz and A. Hotho. Position paper: Ontology learning from folksonomies. In *LWA'07: Lernen, Wissen, Adaption, Workshop Proceedings*, pages 109–112, 2007.

[3] C. Buckner, M. Niepert, and C. Allen. From encyclopedia to ontology: Toward dynamic representation of the discipline of philosophy. *Synthese*. forthcoming.

[4] G. Correndo and H. Alani. Survey of tools for collaborative knowledge construction and sharing. In *Workshop on Collective Intelligence on Semantic Web*, 2007.

[5] T. Gruber. Collective knowledge systems: Where the social web meets the semantic web. *Journal of Web Semantics*, 6(1):4–13, 2008.

[6] T. Heath, E. Motta, and M. Petre. Computing word-of-mouth trust relationships in social networks from semantic web and web2.0 data sources. In *Proceedings of the Workshop on Bridging the Gap between Semantic Web and Web 2.0*, 2007.

[7] C. D. Manning and H. Schuetze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

[8] M. Niepert, C. Buckner, and C. Allen. A dynamic ontology for a dynamic reference work. In *Proceedings of JCDL*, pages 288–297. ACM Press, 2007.

[9] M. Niepert, C. Buckner, and C. Allen. Answer set programming on expert feedback to populate and extend dynamic ontologies. In *Proceedings of FLAIRS*, pages 500–505. AAAI Press, 2008.