Containing the Semantic Explosion

Jaimie Murdock
Indiana University
Cognitive Science Program
810 Eigenmann Hall
Bloomington, IN 47405
jammurdo@indiana.edu

Cameron Buckner
University of Houston
Department of Philosophy
654 Agnes Arnold Hall
Houston, TX 77204
cjbuckner@uh.edu

Colin Allen
Indiana University
Cognitive Science Program
814 Eigenmann Hall
Bloomington, IN 47405
colallen@indiana.edu

ABSTRACT

The explosion of semantic data on the information web, and within digital philosophy, requires new techniques for organizing and linking these knowledge repositories. These must address concerns about consistency, completeness, maintenance, usability, and pragmatics, while reducing the cost of double experts trained both in ontology design and the target domain. Folksonomy approaches address concerns about usability and personnel at the expense of consistency, completeness, and maintenance. Upper-level formal ontologies address concerns about consistency and completeness, but require double experts for the initial construction and maintenance of the representation. At the Indiana Philosophy Ontology (InPhO) Project, we have developed a general methodology called *dynamic ontology*, which alleviates the need for double experts, while addressing concerns about consistency, completeness and change through machine learning over a domain corpus, and concerns about usability and pragmatics through human input and semantic web standards. This representation can then be used by other projects in digital philosophy, such as the Stanford Encyclopedia of Philosophy (SEP) and PhilPapers, along with resources outside of digital philosophy enabled by the LinkedHumanities project.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems
; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods—*representations, semantic networks*

1. INTRODUCTION

"Information explosion" and "semantic web" are metaphors that have become clichés. Like many popular phrases they capture some important aspects of the situation, while disguising others. There has been rapid growth in the availability of both new and old materials on the Internet. The result of this rapid expansion is not, however, the pile of shrapnel that "explosion" might suggest. Rather, there is a highly linked set of pieces captured by the phrase "semantic web", which represents the connectivity but suggests a rigid approach to meaning that has fueled skepticism, and obscures the extent to which the possible semantic relations themselves grow exponentially as the number of linked sources increases. Thus, we prefer to characterize the situation as a "semantic explosion in the information web". This semantic explosion constitutes perhaps the most challenging problem that automatic methods for dealing with the information web must face.

Consider the problem of linking concepts as they occur in one philosophical resource to the concepts in another. Take, for instance, the term "realism". It is not adequate to treat the term as a proxy for a concept and link every occurrence of "realism" to every other because the term's meaning is context-sensitive according to whether it occurs in ethics, metaphysics, or political philosophy (among others), not to mention even finer contexts right down to the level of individual passages by authors who may use terms idiosyncratically. Even if one has disambiguated the term within a given digital resource or project, there remains the problem of how to link the disambiguated senses to occurrences of the term in other digital philosophy projects. As the web of information grows, so too do the interactions among its parts. So, noticing that realism is contrasted with idealism in some contexts (or databases), but with anti-realism in others, leads one to the question of whether the relationship between idealism and anti-realism is synonymy, and whether idealism in political philosophy is at all related to idealism in metaphysics. Furthermore, in trying to connect "realism" as that term is disambiguated in the Indiana Philosophy Ontology, to that term as it appears in other sources of philosophical content, such as the refereed journal articles covered in the PhilPapers database or the crowd-sourced Wikipedia entries on philosophical topics, there is a huge challenge in determining which "realism" belongs with which.

An appealing idea, at this point, is to regiment all of this into one overarching computational ontology that precisely fixes all the possible meanings. We think, however, that this one-size-fits-all approach is unlikely to succeed for a number of reasons. There may be pragmatically useful alternative ways of representing relationships among the data.[3] While we agree that a wide variety of digital humanities projects can benefit from the development of taxonomic schemes that make use of certain techniques of the computational ontologist, there are important differences between ontology design for the humanities and the approaches favored in other areas. Ontology science has evolved in large part to suit the needs of large projects in medicine, business, and the natural sciences. These domains share a cluster of features: their underlying structures have a relatively stable consensus, projects are amply funded, and a primary goal is often to render interoperable large bodies of data. In these projects, the best practices often require hiring so-called "double experts" – knowledge modelers highly trained in both ontology design and the subject domains – to produce a representation in the early stages of a project which is optimally comprehensive and technically precise.

There is a cluster of digital humanities applications, however, for which these practices are not ideal. These involve projects with principles of open-access and domains without the ample funding of the natural sciences. Additionally, ontologies for domains in which structural understanding is controversial or constantly evolving and projects which utilize computational ontologies to enhance search or navigation through asynchronously-updated digital resources must account for the dynamic nature of their resources – whether it is in the underlying corpus or in the judgments of the experts providing feedback on domain structure. On the positive side, these areas often offer more opportunities to collect feedback from users who are domain experts (but who lack expertise in ontology design).

Digital philosophy has many of the features described in the previous paragraph, and different projects have pursued different approaches to taxonomizing the subject matter of the discipline. The InPhO project starts with a top-level structure — a basic taxonomy provided by the editorial structure of the SEP — and a list of key terms assembled from various sources. We then data-mine the text of the encyclopedia to derive statistical hypotheses about term-relatedness. These hypotheses are presented to experts in a simple question format. Their answers drive an automated reasoning system to populate the seed taxonomy. This approach contrasts with PhilPapers, which employs a folksonomy-type approach to classifying the articles in its database. Another contrast is provided by DBpedia, which uses automated tools to extract lists of concepts and philosophers from the markup language used in the crowd-sourced Wikipedia articles. Our goal is to align these various representations based on semantic information embedded in them and the texts on which they are based. Given a taxonomy of concepts, for example, nearest neighbors and other related terms can help determine whether "realism" at a given location in one structure should be mapped to the same term appearing in another. The methods need to be as automatic as possible so that they can continue to be deployed even as the structures provided by these different projects change as a result of new input.

2. ONTOLOGY DESIGN

Effective ontology design for the humanities faces a number of sometimes conflicting desiderata. Formal approaches emphasize consistency, and in scientific contexts this can be achieved by axiomatizing the meanings of the terms represented. However, in the humanities, term meanings are among the most contested aspects of the disciplines. Another desideratum is completeness of coverage, which is hampered, however, by inadequate techniques for automatic term discovery, and vagueness about whether certain terms even belong to a given discipline. Usability is yet another desideratum, but usability by whom or by what? Formal ontologies stress machine readability and reasoning. However, given the semantic complexity and context sensitivity of terms in the humanities, usability by humans may be a more appropriate goal. The context-sensitivity also raises pragmatic issues relative to the various audiences. In a scientific gene ontology, for instance, it is clear that the end users are experts in the field for whom a large degree of consensus exists. In the humanities, however, there is much less consensus among experts, and disagreement is even encouraged. Representations of humanities disciplines, including philosophy, need to allow for the range of interpretations that different users will provide. Finally, scholarship in the humanities consists of suggesting novel interpretations of existing texts, new arguments and criticism, and novel concepts, necessitating not just the addition of new materials to existing databases, but continuously contributing to the semantic explosion as these new approaches interact with the existing structures. It is essential to automate as far as possible the maintenance of any digital representation of philosophy, lest the existing structures become quickly obsolete and abandoned.

In the following sections, we review the main approaches to ontology design, folksonomy and formal ontology. We conclude by outlining the principles of our favored approach, which we call "dynamic ontology", which attempts to leverage the strengths of each approach in semi-automatically generating structured representations of target domains.

2.1 Folksonomy

Social web (Web 2.0) and semantic web research were, for a time, conducted largely independently. Indeed, initial explorations of social computing were driven by skeptics of the grand unifying vision of the semantic web (e.g. [37]), who explicitly proposed "folksonomy" as an alternative method. This mutual antipathy 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, emphasized 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 — though influential researchers have now recognized that the two approaches are not only not in conflict, but can even be complementary [9].

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.

Since Thomas Vander Wal coined the term *folksonomy* in 2004 [37], there has been a surge of research on the effectiveness of folksonomy (see review in [36]). The use of the term itself is not precise, but a *folksonomy* is usually regarded as particular kind of knowledge base, one resulting from or induced upon the vocabulary derived from the collective tagging of shared resources by users in an online community. Folksonomy as a method comes with many advantages — the collection and organization of tags is virtually free, and the population of the knowledge base with resources with community relevance is guaranteed. Reviewers of taxonomic approaches have been encouraged by research on the "Wisdom of the Crowds", believing that the precision and recall of emergent tag behavior, once stabilized, will be superior to alternative methods.

From the beginning, critics recognized that folksonomy would face a variety of serious challenges. Mathes (2004) noted that tagging-based approaches inherently faces the problems of ambiguity, inconsistent orthography, and the unnoticed synonomy "inherent to an uncontrolled vocabulary". Many have worried about the idiosyncratic nature of tagging (characterized as the "long tail" phenomena, which describes the tendency of tag distributions to have a large number of rarely-used terms), though some research has shown that a stabilization of terminology can be reached in a community after a small amount of initial tagging behavior[7, 11]. Other studies, however, have shown that individual tagging behavior can evolve over time[4] as users become more familiar with the resources, raising challenges of intra-user lexical stability — though such behavior can gravitate towards "netlingo" tags that are not suitable for many taxonomic purposes (e.g. "fail", "toread", and "yum" — see [16]).

Many tagging systems have components designed to facilitate the stabilization of vocabulary — del.icio.us suggests commonly used tags, for example. Experts have also worried about the shallow depth of the taxonomic schemes induced on tags — Quintarelli (2005) noted the their lack of hierarchy, together with the concomitant difficulty in scaling the method up to organize larger knowledge bases. If tags are freely submitted by users, one must also worry about simply invalid tags; a study by Stvilia and Jorgensen shows that 37% of Flickr tags used in the Flickr Commons Project were misspelled or otherwise invalid, though this could be reduced to 15.3% with some simple pre-processing rules. A further worry of Kroski (2005) is that folksonomies are subject to "gaming"; because folksonomy systems often treat each user as an equal peer, they are vulnerable to "unethical users" who might "propagate tags ... in order to corrupt a system" (as a result, such systems would be wise to exploit user provenance data — e.g. see the ExpertRank system of John and Seligmann 2006 , and see Koutrika et al. 2007). Some have suggested that training users in tagging might help mitigate some of these problems [10], though other research has shown that users often balk at such training [19] and if financial incentives were required this approach would begin to incur the costs associated with double experts. In addition, Stvilia et al. 2011 found that the relationship between user tagging experience and perception of tagging quality is complex, with age and tagging experience being inversely related to the perceived suitability of tag-supplied terms, but Flickr familiarity and indexing experience having a positive relationship with term rating.

As clusters of tagging behaviors emerge, a further challenge is presented when one tries to use folksonomies to support tasks traditionally ascribed to ontologies — such as supporting reasoning and data interoperability. A variety of systems have been devised to leverage tagging libraries into ontologies, either using automated information-extraction or by designing tools which help users arrange tags in taxonomic relationships; but since tags are merely words applied to resources, these approaches face many of the same challenges that are faced by systems which attempt to learn taxonomies directly from text, including synonymy, polysemy, slang, inconsistent lexical forms and misspellings, and varying levels of generality. Marchetti et al (2007) have proposed that providing semantic support to taggers from resources like Wordnet and Wikipedia can help mitigate some of these challenges. Several approaches have been proposed to learn taxonomies and ontologies by using statistical techniques on tag distributions as a solution to this problem [28, 12, 1, 35], though all struggle with challenges posed by unregulated vocabulary and none offer the same rich level of structure as manually-encoded ontologies.

2.2 Upper-level Ontology

The grand vision of the formal "ontological" approach to the semantic web is to take a multi-layered approach to modeling reality. The task is divided into two levels: Lower-level domain ontologies are constructed to describe the entities of interest in specific domains; the types in the lower-level ontologies are then linked into a so-called "upper-level ontology", intended to describe the most basic, enduring features of reality. While by the nature of the method change is much more challenging on the formal ontological approach than with folksonomy, some of the largest formal domain ontology projects aspire to dynamism; the Gene Ontology project, for example, claims to offer "a controlled vocabulary that can be used for dynamic maintenance and interoperability between genome databases" [17]. Such dynamism is possible in the context of large biomedical informatics projects because they involve the efforts of very many dedicated biomedical informatics specialists working with manually designed taxonomies and ontologies. New data come pre-annotated because of the sophisticated equipment used for sequencing and other experiments. These features are only possible for deep-pocketed projects in domains studying relatively stable structure (though conceptual structure even in biology may not be so stable as one would think [8]).

The most significant challenges facing formal ontology are economic. Once elaborate and precise ontologies have been created, semantic web projects have faced the dilemma of either hiring expensive "double experts" to populate and maintain them or face inevitable data and user sparseness [3]. A further economic challenge is posed by the fact that projects developing domain-level ontologies are never sure which upper-level ontology should be linked to. Upper-level ontologies have now been an active area of research for fifteen years, and the diversity of choices appears to be increasing rather than decreasing. Modelers are now faced with a bewildering choice between a variety of inconsistent upper-level ontologies — including SUMO, DOLCE (and DnS), BFO, GFO, IDEAS, Cyc (and UMBEL), PROTON, OCHRE, and Sowa's [29, 20]. Debates in this area are bitter and protracted, given that there is often a significant commercial gains to be won by emerging as the "one ontology to unite them all". Many formal ontologists have by now abandoned the goal of selecting a single upper-level ontology [31], and recently attempts have been made to map the diverging upper-level ontologies into each other, such as COSMO (constructed largely out of categories from Cyc and SUMO) The most serious effect of these "ontology wars" has been that the population of elaborately-designed ontologies by the large amount of data already available on the web has languished while the battles are fought. Frustration with this process has in turn driven interest in alternative approaches to interoperability, such as the Linked Data initiative [2] which tries to obviate the need for upper-level ontologies by directly linking data in shared repositories (such as DBPedia and Freebase) .

More broadly, the debate over formal ontologies is situated within a paradigm shift within artificial intelligence. The original vision of logic-based AI held that computers could display intelligence if only we could encode enough explicit expert knowledge into their systems. Though it quickly became apparent that this was a hopeless approach, the grand vision of "just getting enough knowledge" formally-specified continued to live on in the Cyc project for decades (as the largest remaining attempt in true artificial intelligence). Nowadays, even Cycorp has largely conceded this point — themselves turning away from grand visions of passing the Turing test with more specific practical goals, such as database translation. The push towards manual encoding and population of formal ontologies in the semantic web can further be seen as the last gasp of this knowledge-based approach to AI. Meanwhile, IBM's DeepQA system, showcased in Watson, starkly illustrates the lesson that, outside of a few specialized applications, it will simply never be practical to encode every scrap of knowledge in a clean, precise formal system [6]. Using a complex and heterogeneous system consisting of layers upon layers of diverse statistical methods, heuristics, partial ontologies, and ad-hoc tuning, the DeepQA methodology demonstrates that double experts are too expensive, and knowledge evolves too quickly for them to keep up with the problem at any hourly rate — especially in domains like the humanities. Practical intelligence, rather, requires a vastly more efficient tangle of statistical and ontological prowess, with both humans and computers contributing only what they do best.

3. DYNAMIC ONTOLOGY

At the InPhO project, we have developed a methodology for ontology population called *dynamic ontology*, which alleviates the need for double experts, while addressing concerns about consistency, completeness and change through machine learning over a domain corpus, and concerns about usability and pragmatics through human input and semantic web standards. Dynamic ontology follows a three-stage pipeline of data mining, feedback collection, and machine reasoning, summarized in Section 3.1. The core of dynamic ontology is the marriage of both human and computational resources in the design process. While human experts may be locally-blinded by their own familiarity with a subdomain, algorithmic processes can keep perspective over the entire corpus. Similarly, while data mining techniques may struggle with word sense disambiguation, human feedback can easily resolve such inconsistencies. This set of checks and balances helps maintain consistency in the resulting ontology.

As mentioned in the introduction, our pragmatic approach recognizes the likelihood that there is no single, correct view of the discipline. However, even if other projects do not agree with the InPhO's taxonomic projection, our statistical data and expert evaluations may still be useful for filling out alternative representations of the discipline. By exposing our data from each of the three steps in our procedure through an easy-to-use API, we enable the adoption of our system by other projects seeking alternative ways to construct meaningful and useful representations of the discipline. Additionally, by offering an open platform, we invite other projects to contribute relevant data and expert feedback to improve the quality of the service. By enabling linkages between the different representations, it becomes possible for end users to move among the different digital philosophy resources and make semantically interesting connections based upon their understanding of the concepts involved.

Thus, for example, the InPhO does not seek to replicate the cross-referencing structure of the SEP, but it provides data that the editors can use to select appropriate cross-references for the entries. Also, by providing links from each entry in the SEP to a dedicated InPhO page, readers can explore the concept network given by the InPhO representation, and use the InPhO portal as a way to discover other related resources outside the SEP, or to navigate back to related SEP articles via the InPhO taxonomy. Eventually, given fuller integration with PhilPapers, for example, it will be possible for end users and developers to navigate the conceptual space, the bibliographic network, and the linkages between specific thinkers using resources from all the various data providers.

3.1 The InPhO Workflow

Data Mining — Natural Language Processing (NLP) techniques are used over an external corpus (the SEP) to generate a lexicon of concepts and statistical hypotheses about semantic relevance and generality relations among various topics in the corpus. From this corpus we generate a co-occurrence graph in which each node represents a term in our set of keywords. An edge between two nodes indicates that the terms co-occur at least once. For each edge, the directed J-measure [32, 23] and conditional entropy [30] is calculated bidirectionally. The J-measure calculates the interestingness of inducing the rule "Whenever idea i is mentioned in a fragment of text, then idea j is mentioned as well" (for details see Niepert et al. 2007). This is used as a proxy for semantic relevance of term i to term j. While the J-measure can be used to estimate semantic distance, we are currently investigating alternative measures of semantic distance, as reviewed in Resnik (1999). We then apply to each node an informational metric of entropy. Entropy is used as a proxy for the generality of a each term, on the assumption that more general terms will have higher entropy. By combining these relevance and entropy measures, we obtain a directed estimate of hypernymy/hyponymy–the basic building blocks of taxonomies. Further details on data mining techniques can be found in [23].

Feedback Collection — The statistical hypotheses about hypernymy and hyponymy are presented to domain experts through online interfaces located both on the InPhO website and through the SEP editorial interface. Evaluations are presented as pairs of concepts, with a slider to indicate the relatedness of two terms, and a selection of whether the first term is more specific, more general, as general as, or incomparable with the second term. Users self-report levels of expertise when they sign up for our system, and each feedback fact is recorded with provenance information. This allows us to stratify feedback by self-reported education level and leverage expertise to resolve feedback inconsistencies. Experiments have been conducted on the effects of this stratification [22] and upon feedback collected from Amazon Mechanical Turk users[5]. Further details on feedback collection can be found in [25].

Machine Reasoning — User feedback is then combined with the statistical measures as the input for our machine reasoning program, which uses answer set programming to output a taxonomic view of the discipline. To reduce computational complexity, a *seed taxonomy* created by domain experts is added to the input. Variations in the answer set program, the subset of user feedback used, the data mining techniques, or the seed taxonomy can allow us to generate different representations of the discipline. These variations can then be evaluated against the external corpus to find the most suitable population method [22]. Further details on the answer set programming techniques can be found in [24].

4. FUTURE DIRECTIONS

We are currently engaged in two projects that will contribute to the management of the semantic explosion on the information web that we described in the introduction. These projects will expand the range of digital philosophy projects and enable connections to other digital databases not solely concerned with philosophy. This expansion has at least two different fronts. On the one hand, as different pieces of the semantic and social webs become connected, the appropriate linkages between entities in these pieces need to be established. In the LinkedHumanities project, in collaboration with the University of Mannheim and jointly funded by the DFG in Germany and the NEH in the United States, we are exploring ways of matching entities across the various databases that contain semantic information about the concepts and major figures already represented by the InPhO. In the Digging by Debating project with partners in the UK, and joint funding from the NEH and JISC in the UK, we will be attempting to map the interactions among philosophy and the sciences across various timescales, using data from Hathi Trust, PhilPapers, and InPhO. The resulting tools will enable users to discover and represent arguments appearing both in historical texts and current articles.

5. ACKNOWLEDGMENTS

The work in this paper was funded by the National Endowment for the Humanities Office of Digital Humanities, Grant No. PW-50401-09.

6. REFERENCES

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

[2] C. Bizer and T. Berners-Lee. Linked data âĂŞ the story so far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.

[3] C. Buckner, M. Niepert, and C. Allen. From encyclopedia to ontology: Toward a dynamic representation of the discipline of philosophy. *Synthese*, 182:205–233, 2011.

[4] T. Coates. Two cultures of fauxonomies collide, 2005. http://www.plasticbag.org/archives/2005/06/two

[5] K. Eckert, M. Niepert, C. Niemann, C. Buckner, C. Allen, and H. Stuckenschmidt. Crowdsourcing the Assembly of Concept Hierarchies. In *Proceedings of the 10th ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, Brisbane, Australia, 2010.

[6] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, and K. Aditya. Building watson: An overview of the deepqa project. *AI Magazine*, 31(3), 2010.

[7] S. A. Golder and B. A. Huberman. The structure of collaborative tagging systems. *Journal of Information Science*, 2005.

[8] P. Griffiths and K. Stotz. Experimental philosophy of science. *Philosophy Compass*, 3(3):507–521, 2008.

[9] T. Gruber. Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web & Information Systems*, 3(2), 2005.

[10] M. Guy and E. Tonkin. Folksonomies tidying up tags? *D-Lib Magazine*, 12(1), 2006.

[11] H. Halpin, V. Robu, and H. Shepard. The dynamics and semantics of collaborative tagging. In K. Moller, A. de Waard, S. Cayzer, M. Koivunen, M. Sintek, and S. Handschuh, editors, *Proceedings of the 1st Semantic Authoring and Annotation Workshop*. CEUR-WS.org, 2006.

[12] H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. In *WWW 2007*. ACM Press, 2007.

[13] A. John and D. Seligmann. Collaborative tagging and expertise in the enterprise. studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*, 10(1), 2006.

[14] G. Koutrika, F. A. Effendi, Z. GyÃ şngyi, P. Heymann, and H. Garcia-Molina. Combating spam in tagging systems. In C. Castillo, K. Chellapilla, and B. Davidson, editors, *AIRWeb '07: Proceedings of the 3rd International Workshop on Adversarial Information Retrieval on the Web*, pages 57–64. ACM Press, 2007.

[15] E. Kroski. The hive mind folksonomies and user-based tagging, 2005. http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/.

[16] C. Lu, J. Park, and X. Hu. User tags versus expert-assigned subject terms: A comparison of librarything tags and library of congress subject headings. *Journal of Information Studies*, 36(6):763–779, 2010.

[17] A. M, C. A. Ball, and J. A. Blake. Gene ontology: tool for the unification of biology. *Natural Genetics*, 25:25–29, 2000.

[18] A. Marchetti, M. Tesconi, F. Ronzano, M. Rosella, and S. Minutoli. Semkey a semantic collaborative tagging system. tagging and metadata for social information organization; a workshop at www2007, banff, alberta. In *Tagging and Metadata for Social Information Organization; a workshop at WWW2007*, 2007.

[19] K. Markey. Twenty-five years of end-user searching, part 2: Future research directions. *Journal of the American Society for Information Science and Technology*, 2(8):1123–1130, 2007.

[20] V. Mascardi, V. Cordi, and P. Rosso. A comparison of upper ontologies. Technical report, University of Genova Computer Science Department, 2006.

[21] A. Mathes. Folksonomies - cooperative classification and communication through shared metadata. Technical Report LIS590CMC, University of Illinois, 2004.

[22] J. Murdock, C. Buckner, and C. Allen. Evaluating Dynamic Ontologies. In A. Fred et al., editor, *IC3K*

2010, CCIS, pages 258–275. Springer-Verlag, Berlin Heidelberg, 2012.

[23] M. Niepert, C. Buckner, and C. Allen. A dynamic ontology for a dynamic reference work. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, page 297. ACM, 2007.

[24] M. Niepert, C. Buckner, and C. Allen. Answer set programming on expert feedback to populate and extend dynamic ontologies. In *Proceedings of 21st FLAIRS*, 2008.

[25] M. Niepert, C. Buckner, and C. Allen. Working the crowd: Design principles and early lessons from the social-semantic web. In *Proceedings of Workshop on Web 3.0: Merging Semantic Web and Social Web 2009 (SW)ˆ2*, Turin, Italy, 2009. CEUR Workshop Proceedings.

[26] E. Quintarelli. Folksonomies power to the people. ISKO Italy - UniMIB Meeting, 2005.

[27] P. Resnik. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of artificial intelligence research*, 11(4):95âĂŞ130, 1999.

[28] P. Schmitz. Inducing ontology from flickr tags. In *WWW2006: Collaborative Web Tagging Workshop*. ACM Press, 2006.

[29] S. K. Semy, M. K. Pulvermacher, and L. J. Obrst. Toward the use of an upper ontology for u.s. government and u.s. military domains: An evaluation. Technical Report MTR 04B0000063, The MITRE Corporation, 2004.

[30] C. E. Shannon. *A mathematical theory of communication*. University of Illinois Press, Urbana, Illinois, 1949.

[31] B. Smith. Ontology. In L. Floridi, editor, *Blackwell Guide to the Philosophy of Computing and Information*, pages 155–166. Blackwell, 2004.

[32] P. Smyth and R. Goodman. An information theoretic approach to rule induction from databases. *IEEE Transactions on Knowledge and Data Engineering*, 4(4):301–316, 1992.

[33] B. Stvilia and C. Jorgensen. Member activities and quality of tags in a collection of historical photographs in flickr. *Journal of the American Society for Information Science and Technology*, 61(12):2477–2489, 2010.

[34] B. Stvilia, C. JÃűrgensen, and S. Wu. Establishing the value of socially created metadata to image indexing. *Library & Information Science Research*, 2011.

[35] J. Tang, H. f. Leung, Q. Luo, D. Chen, and J. Gong. Towards ontology learning from folksonomies. In *IJCAI'09: Proceedings of the 21st international joint conference on Artificial intelligence*, pages 2089–2094. Morgan Kaufmann Publishers Inc., 2009.

[36] J. Trant. Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*, 10(1), 2009.

[37] T. V. Wal. Folksonomy definition and wikipedia, 2005.