# Learning Contextualised Weblog Topics

Paolo Avesani, Marco Cova, Conor Hayes, Paolo Massa
ITC-irst
via Sommarive 18
38050 Povo, Italy
{avesani, cova, hayes, massa}@itc.it

## ABSTRACT

The blogosphere refers to the distributed network of user opinions published on the WWW. Whereas centralized review sites such Amazon.com previously allowed users to post opinions on goods such as books and CDs, blogging software allows users to publish opinions on any topic without constraints on predefined schema. However, centralized review sites such as Amazon.com have one significant advantage: reviews pertaining to a single topic are collected together in one place, allowing readers to peruse a diverse range of opinions quickly. In this paper we examine how such a topic-centric view of the Blogosphere can be created. We characterise the problems in aligning similar concepts created by a set of distributed, autonomous users and describe current initiatives to solve the problem. Finally, we introduce the Tagsocratic project, a novel initiative to solve the concept alignment problem using techniques derived from research in language acquisition among distributed, autonomous agents.

## 1. INTRODUCTION

Weblogging has increasingly become an important part of the information economy found on the Internet [13, 16]. Its great benefit is that it allows ordinary people to easily publish opinions based upon their experiences. This type of information, sometimes highly subjective, has great value for other Internet users who can make use of it to take decisions or simply to inform themselves. This could be viewed as a problem solving exercise where the goal of the user is to find as many points of view on a particular topic as possible. For example, having such knowledge would enable the user to make a decision on whether to purchase a particular item. The value of this type of user opinions was initially harnessed by retail sites like Amazon.com where contributors can build up extensive portfolios of reviews.

However, blogs offer the user many advantages over proprietary review sites like Amazon.com. For one thing, the independent blogger owns and has complete control over the content he/she posts. In contrast, the proprietary site requires the reviewer to give up ownership and has the right to delete or censor the reviewer's work. More importantly, the reviewer is limited to posting annotations that pertain *in a relevant way* to the stock items in the proprietary site where the concept of *relevance* is determined and enforced by the retailer. Consequently, a specific context in which a user may refer to an item may not be tolerated if it is not considered relevant to the sales objective of the retailer. Associated with this problem is the fact that a site like Amazon.com only supports concepts that refer to stock items and does not allow the user to define higher level or related concepts. For instance, a reviewer can review a specific novel by Raymond Chandler but cannot post an opinion on the class of books termed "detective fiction". In contrast, the emergence of blogging sites allows users full control to post comments, reviews and links on any topic and to define the local categories to contain their topics.

Centralized proprietary review sites like Amazon.com do have one significant advantage. Because they provide a limited number of topics and list them centrally, it is easy to find all the reviews that pertain to a single novel. Conversely, blog entries relating to the same novel may be distributed across multiple servers, each entry indexed by the local category selected by the user. The key observation is that while review sites are *topic-centric*, organising information around the schema describing the stock inventory, blog sites are *user-centric*, publishing the user's perspective on multiple topics, defined and categorised in a local way by the user (see Figure 1) [12]. A key issue is how to produce a topic-centric view of the blogosphere so that the blogger and his/her readers can quickly find related blog entries on a given topic. We could view this as a classification problem where the goal is to classify or label the blog entry in such a way that it can be quickly mapped to other blog entries which consider the same topic.

In this paper we describe the issues affecting this problem. In Section 2 we define the problem and present an idealized use case scenario involving a hypothetical *topic-centralizing* service. In Section 3 we outline the requirements to achieve such a service while in Section 4 we describe possible solutions and specific initiatives to address this problem. Section 5 introduces the Tagsocratic project, a novel initiative to solve the concept alignment problem using techniques derived from research in language acquisition among distributed, autonomous agents.

## 2. DESCRIBING THE PROBLEM

The success of the Internet and the WWW is based upon the design of an open set of protocols which allowed multiple heterogeneous networks to be easily linked together. Although such protocols have enabled an unprecedented flow of information, the original specifications of the WWW did not consider two important issues: allowing users to easily inject information into the network and standardising the semantics of the information being posted to the web. The first issue is being addressed by blogging software, whereby blog users can quickly post annotations to the Web on any

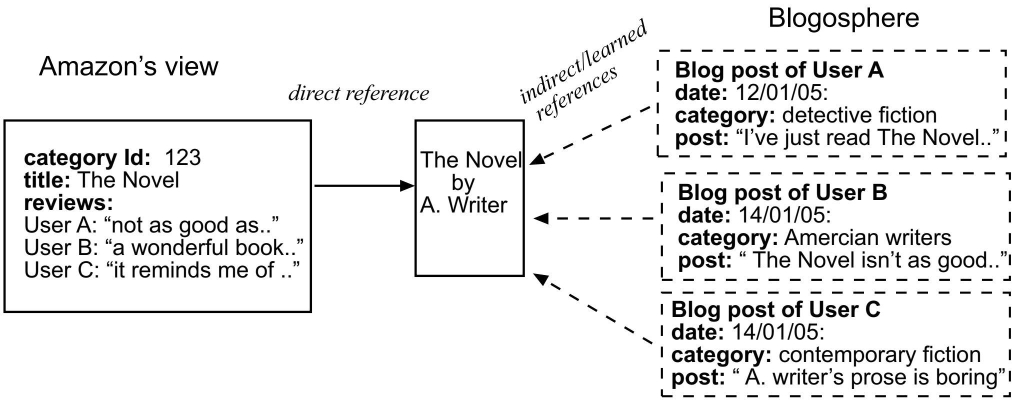
Amazon's view
category Id: 123
title: The Novel
reviews:
User A: "not as good as.."
User B: "a wonderful book.."
User C: "it reminds me of .."
direct reference
The Novel
by
A. Writer
indirect/learned
references
Blogosphere
Blog post of User A
date: 12/01/05:
category: detective fiction
post: "I've just read The Novel.."
Blog post of User B
date: 14/01/05:
category: Amercian writers
post: " The Novel isn't as good.."
Blog post of User C
date: 14/01/05:
category: contemporary fiction
post: " A. writer's prose is boring"

**Figure 1: Topic-centric vs. User-centric views.**

topic. However, the blogging phenomenon exacerbates the problems posed by the lack of semantic protocols for the Internet. Although there is no constraint on what information can be posted, blogs often take the form of a series of annotations on topics of shared interest [4]. As bloggers write independently and autonomously there is no standard way of organising the blogosphere so that the posts that relate to a particular topic can be automatically indexed together. Some blog software allows users to define tags with which to label their posts. However, the semantics of the tag is defined locally by the user rather than relating to a globally understood concept.

Clearly, there are benefits if these distributed information sources can be organised so that the reader (or blogger) can view related opinions on a single topic or concept. For example, the prominence given to user reviews on proprietary review sites like Amazon.com suggests their importance in providing sales advice to the potential customer. In Figure 2 we present a use case of the type of topic-centric service we require for the blogosphere. The objective is to provide an on-line mapping service for locally defined blog entry categories. Thus, given an entry posted by user Alice under category $C$, user Bob can retrieve a list of categories defined by different users whose topic matches Alice's $C$. In the use case scenario depicted in Figure 2, Bob is visiting Alice's blog. He finds posts about the activity of blogging and notices that Alice categorizes them under the category `blogs`. Bob would like to view other posts available in the blogosphere about the same topic. The problem, of course, is that other bloggers may use different categories to describe the *blogging* topic. Thus, Bob contacts our hypothetical service, requesting blog entries from categories mapped to Alice's `blogs` category; the mapping engine then returns a list of entries from categories aligned to the `blogs` category. The returned matches include entries from Carl under his `blogging` category and entries from Dave labeled `PhD` (Dave is doing a PhD on the effect of blogs on society). However, Bob does not receive any entries from Eve whose `blogs` category simply stores links to various blog engines web sites.

## 3. ANALYSING REQUIREMENTS

In this section we broadly define the issues in providing a service like that described in the previous section. As we have described up to the point, the problem is how to enable a set of distributed content providers (bloggers) to easily reference shared concepts. An analysis of the problem suggests two ways forward. The first involves agreement on an *ex ante* semantic framework which allows users to define and make reference to concepts unambiguously. This type of eager approach is being promoted by the Semantic Web project [5, 11]. The second is to adopt a less structured approach in which local semantic representations are negotiated *ex post*. An example of the second would be a technique that *learns* the relationships between local concepts [3, 2]. We argue that the following dimensions should define the solution that is adopted for this problem.

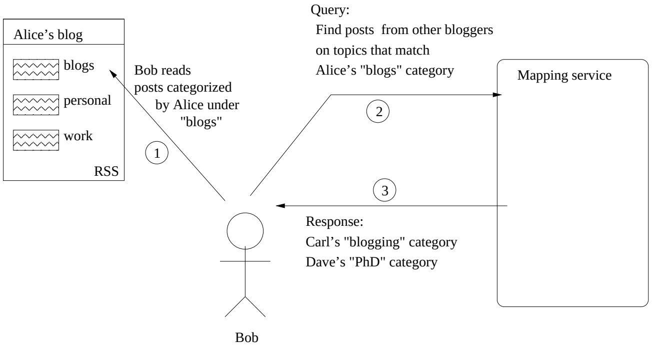

**Figure 2: Tagsocratic use case.**

**Architecture:** centralized vs. distributed. One of the characteristics of the blogosphere is that it is not centralized. Blogs are hosted throughout the Web using different types of blog software. The question is whether a solution requires a blog to be registered with a centralized service or whether a purely distributed solution can be deployed. A distributed approach may require participating peers to implement a communication and/or processing interface in their blog service. A centralized approach is less invasive but creates a single point of failure.

**Semantic Coordination:** ontology vs. lexicon. A key issue is how to share the same meanings for a collection of topics. Ontology-based approaches aim to coordinate with respect to a common interpretation of a formal representation of concepts. Lexical-based approaches focus the attention on the reference language, giving less emphasis on explicit representation of the semantics.

**Representation:** autonomous vs. standardized. Currently, blog users are unconstrained by the categories they use for their blog entries. Thus they may organise subjects meaningfully and easily according to their perspective or local context. The advantage of adopting an ex-ante semantic framework approach is that related concepts can be easily indexed together. The disadvantage lies in getting agreement on the semantic components to be used and in convincing bloggers to adopt the convention. The current approach allows users to organise and categorise their work as they please. The clear disadvantage is that the relationship between locally defined categories must be learned or hand encoded.

**Sustainability:** high work load vs. low work load. A secondary issue is how to provide a means whereby the user can categorise the subject matter correctly with relatively low cognitive load. For example, for a simple categorization system based on free form tags, the work required may be quite low. For a more complex system based on large taxonomy of categories, the user will have significantly more work to do to find a matching category from the elements available.

**Locality:** context-based vs. context-free. Related to the issue of autonomy is "locality" which refers to how a solution caters to the context in which a blogger or a group of bloggers create categories. At one end of the scale, a solution which imposes a global semantic perspective disregards the fact that meaning may be local. At the other end of the scale, a completely local perspective makes it difficult to link easily to local communities. A middle ground is to recognise that meaning is local, but not so fragmented that local clusters that have shared interests cannot agree on a common lexicon.

**Dynamics:** static vs. evolutionary. Given that the blogosphere is a particularly dynamic environment, any solution must be robust to the introduction of new categories, new bloggers and the issue of concept drift, whereby agreed categories may no longer be accurate to describe the concepts contained within.

## 4. SOLUTION DIRECTIONS

The previous section introduced the issues that underlie the design and the development of a solution for blog interoperability. Although this is ambitious project there are several research initiatives that attempt at least a partial solution. These can be characterized with respect to two principle dimensions: a methodological perspective and a technical perspective. The first dimension takes into account the strategy of agreement: *all-to-one*, *one-to-one* or *many-to-many*. The second dimension refers to the process of semantic assessment.

Let us start with a brief survey of the possible strategies for achieving an agreement on the meaning of a collection of tags.

**(i) All-to-one**. The simplest possible way to handle a shared lexicon is to have a centralized repository where we can store the official and up-to-date version of the common agreement. the tacit assumption is that all users will reference the centralized lexicon (all-to-one). On the one hand, this scenario is quite simple: the users are in charge of mapping their local lexica with respect to the reference lexicon. The reference lexicon, thus, acts as a bridge between the tags of any two users. On the other hand, it is not at all clear how meaningful agreement on a reference lexicon can be achieved. Furthermore, this kind of approach tends to neglect sociological and sustainability issues (see above) [6].

The SKOS project (Semantic Knowledge Organisation Systems) is an example of this type of approach with in Semantic Web project [17]. SKOS provides a basic framework for building concept schemes. SKOS Core supports the RDF description of language-oriented knowledge organisation systems (KOS) such as thesauri, glossaries, controlled vocabularies, taxonomies and classification schemes.

The Open Directory Project [7] is an other representative initiative that aims to build a comprehensive taxonomy as a centralized web directory. Unlike SKOS, there is less emphasis on the explicit formal definition of semantics. The meaning of each category is defined implicitly through a collaborative classification of web contents.

TopicExchange.com [14] is a similar initiative but specifically tailored for the blogosphere. It is a centralized site where anyone can create a channel which is represented by a label: for example, one user has created the channel "weblog_research". It is possible to notify TopicExchange.com that a blog post is related to a channel using a method called TrackBack whereby a small message is sent to the channel page. In this way, TopicExchange.com can show in the page for a specific channel all the blog posts that have sent a TrackBack to that page. In this system, there is no central control and anyone is free to create whatever channel he/she likes.

The solutions arranged along this all-to-one strategy suffer from the same drawback: vulnerability as a single point of failure. Moreover it is also worthwhile to note that comprehensive solutions, such as a centralized lexicon, usually fail to satisfy domain specific requirements.

**(ii) One-to-one**. The one-to-one strategy, known also as peer-to-peer, is based on the idea of pairwise mapping of lexica. The notion of reference and official lexicon is replaced by an independent pairwise negotiation between two autonomously defined lexica [1]. It is not necessary that mappings have to be negotiated between every pair of peers. However, meaningful coverage of the community implies a quadratic factor of growth.

Of course, the negotiation of agreement between two heterogeneous lexica is a knowledge intensive and time consuming task. For this purpose new solutions have been conceived to automate this process. The research effort is twofold: reasoning on mapping at the linguistic level or learning mapping by document analysis. In the first case, the mapping hypothesis is formulated by looking only at the linguistic knowledge encoded in the labels [8, 15], in the second case the interpretation of the labels is supported by a process of learning by examples [9]. While in principle, this is much more sustainable than the previous all-to-one strategy, the one-to-one strategy is computationally expensive.

**(iii) Many-to-many**. A more recent strategy is concerned with a fully distributed strategy. The basic intuition is to look at the problem of a shared lexicon as a language game [18]. The agreement on a common denotation system is conceived as a result of trial end error interactions. The meaning of a collection of labels emerges as a process of coordination among a distributed community of peers. A language game is designed as an iterative pairwise session.

Each session is defined by two players, a speaker and an hearer. The speaker sends the hearer a label and an example that is representative of the semantics of the label. The hearer looks at its own lexicon to assess the local semantic for such a label. In this strategy there is no notion of a centralized reference lexicon. Instead, a distributed lexicon emerges whereby each peer learns where topics that match its own interests can be found. One of the main advantages of a many-to-many strategy is that the mappings occur at run time and such a process is evolutionary; hence it can cope with a natural phenomenon on the blogosphere that is the emergence of new labels to express new concepts.

Like the methodological perspective, the technological perspective addresses three main approaches: syntactic (or keyword based), semantic (or linguistic based) and inductive (or example based).

**(i) Syntactic (or keyword based)**. The baseline technique to assess the relationship between two tags is represented by the keyword matching, i.e. syntactic equivalence. Using this technique it is easy to deliver useful services that can be used to aggregate in one place all the blog posts published by their authors on their blog and tagged by them under the same label. For example, subscribing to the feed of all the posts tagged as "blogging" is very interesting if one wants to monitor comments about the blogging phenomena and this can be done on Technorati[1]. We can produce even more results by taking into consideration, not only the category, but the entire content of the blog posts as Feedster does [2]. However, there is a huge problem. It is not possible to automatically monitor posts that speaks of concepts similar to "blogging" but use a different tag.

**(ii) Semantic (or linguistic based)**. Of course syntactic approach is not effective in practice. Proliferation of many different variations of the same tag ("blog", "blogging", "weblog", ...) prevents the recognition of synonymous labels. Simple linguistic heuristics, like stemming, can help to reduce noise. The synonyms can be detected even with the help of linguistic resources like WordNet [10]. In WordNet, english nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. However, while synonyms are reasonably managed, the phenomenon of polysemy is much harder. Polysemy occurs when the same label is used to refer to two different meanings, not necessarily related. The recognition of the correct interpretation for ambiguous labels is even harder because they are not available in a context of use like a sentence. Moreover the linguistic resources, like WordNet, tend to have a poor coverage of domain specific labels.

**(iii) Inductive (or example based)**. All the research in the semantic web is concerned with the explicit representation of meanings. Such effort is producing many formal languages to encode concepts, but usually a formal representation is not self-explanatory. The main drawback is that the adoption of a well-defined ontology requires preliminary training and significant maintenance costs.

An alternative approach based on a bottom-up strategy is to have an implicit encoding of meaning. For example, the right interpretation of a given tag can be derived by a collection of representative posts. Statistical machine learning provides powerful techniques to induce the correct classification of a piece of text given a set of examples. Of course, the inductive process is prone to error and automated text classifier has a margin of approximation, but learning by examples seems much more sustainable rather than a process of promoting standards.

An example of a recent and successful online service using blog's category labels is Technorati. Technically, Technorati is simply an aggregator: it fetches all blog entries that have user defined categories associated with them and offers web pages containing all posts tagged under the same category label[3]. The service provided is very useful as it allows anyone to monitor a chosen keyword. However it requires the user to know precisely what she is looking for and this cannot always be the case. However, the service does not support synonyms; for example, posts tagged under "blog", "blogging" and "blogs" are shown in 3 different pages. Nor does it support polysemy. For example, the posts tagged under "python" are all presented in one page, notwithstanding the fact that some of them may refer to the animal while other to the programming language. A researcher studying pythons would be bothered with non relevant posts about the programming language.

## 5. TAGSOCRATIC PROJECT

The objective of the Tagsocratic project is to provide an on-line matching service for blog entries categories. Our goal is to allow a user to find posts categorized by other users under labels that are semantically equivalent to a chosen one. We wish to overcome the limitations of previous approaches, retaining, however, the advantages of working with tags and category labels. In our approach, the semantics of other users' categories are automatically learnt by the system. The learning takes place using the language games technique. The usage patterns of the user (which we call local context) is taken into account. This allows us, for example, to handle situations where two bloggers use the same category label with totally different meanings. From the functional point of view, Tagsocratic tackles the situation presented in our initial use case (see Figure 2). We now discuss in more details our design and implementation choices for Tagsocratic.

With regard to the architecture, we are proposing a centralized solution: the service is available through a web site, that stores a repository of blogs, a matching engine, a mapping layer and the presentation layer (see Figure 3). The blogs repository stores a copy of users' blogs. The matching engine implements the language games technique used to match the semantics of categories defined by different users on the basis of sample posts stored in the repository. The mapping layer provides the relationships between users' categories by leveraging the matchings found by the matching engine. Finally, the presentation layer simply allows users to perform queries and visualize responses.

The motivations for choosing the centralized approach over the distributed one are of two types. The first has a social aspect. A distributed approach would require users to install and run the distributed version of Tagsocratic on top of their blogs. In many cases, blogs are stored on hosting sites that do not allow users to install and run customized

[1] http://www.technorati.com

[2] http://www.feedster.com

[3] For example, you can visit all the posts tagged under the category "blog" at http://technorati.com/tag/blog

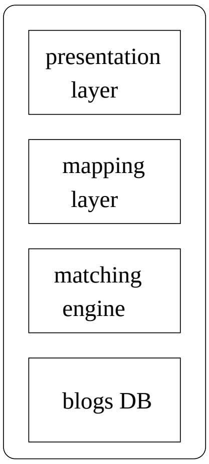

**Figure 3: Topic-centric vs. User-centric views.**

features. Even where possible, this would require a considerable amount of user effort. Secondly, there are technical reasons: the centralized approach simplifies the implementation effort and is independent of the specific blog engine used by bloggers.

In terms of the remaining points made in Section 2 our choices are dependent upon the characteristics of the language games paradigm. As we have seen previously, a language game consists of a series of pair-wise interactions between players in a community. During an interaction, peers try to negotiate a global, user-independent label to refer to shared topics. We use the term *global tag* for such a globally defined label and *local tag* for the locally used category label. Each player associates a local tag with the global tag and exchanges a sample of blog posts indexed under the local tag. Then each assesses whether the received posts are compatible with the local category. If so, the association between the global tag and the local tag is strengthen; otherwise, it is negatively reinforced. For a more detailed description of the language games mechanism, please refer to [3].

The result of these interactions is the convergence inside the community to a set of common global tags and each player maps local tags to the set of global tags. This agreement effectively corresponds to the negotiation of a shared, common lexicon for category names. The following table provides an example of such common lexicon.

| *User* | *Local tag* | *Global tag* |
|---|---|---|
| Alice | blog | blogging |
| Alice | personal | Hobbies |
| Carl | blogging | blogging |
| Dave | blog | blogEngines |

Using this common lexicon, it is possible to map Alice's blog category to Carl's blogging category and discriminate between Alice and Dave's blog categories.

In terms of Section 3, the semantic resolution is of the mapping type, whereby meaning is referred to rather than explicitly modeled. It is autonomous in the sense that bloggers continue to independently use the set of tags of their choice, without being forced to adopt an externally imposed schema.

Finally, the blog domain is too vast and disparate for global alignment to be achievable or meaningful. Thus, we introduce the concept of context to limit the set of bloggers that are involved in a lexicon negotiation process. Consequently, a number of implicit and explicit mechanisms are in place to define and allow the evolution a user's context.

The following table summarizes the main characteristics of the language games-based Tagsocratic project we are experimenting with.

| *Design dimension* | *Design choice* |
|---|---|
| Architecture | centralized |
| Semantic Coordination | lexical-based |
| Representation | autonomous |
| Locality | context-based |
| Dynamics | evolutionary |

Tagsocratic faces a number of challenges, both under the functional and the technological point of view. We have already discussed some of these issues in Sections 3 and 4. With regard to the matching technique, the most critical aspect is the need to face the evolving nature of the domain. At any time, new bloggers may enter the system and others may leave, requiring an adjustment in the negotiated lexicon of a community; the usage of tags is subject to the "concept drift" phenomenon, that is, over time the same label is used to refer to slightly different concepts; bloggers can mis-categorise posts, etc. The language games paradigm, with its negotiation-based, iterative approach deals effectively with changing scenarios [18]. Furthermore, the system offers a number of pro-active and passive techniques to elicit feedback on proposed matchings from users.

Under the technological point of view, we are especially assessing the availability and scalability issues (both in terms of CPU, disk and bandwidth) of the system. The current prototype of the system is based on the J2EE Java Servlet technology. Interaction between the service and users is offered through simple REST interfaces. Results of the service are presented via HTML pages and RSS feeds.

## 6. CONCLUSIONS

In this paper we consider the problem of providing a topic-centric view of the blogosphere. This would allow web users access to a range of views on particular topic from bloggers throughout the Web. At the same time we recognise that it is important not to compromise the autonomy of bloggers. We characterise the solution to this problem in terms of architecture, techniques for resolving semantic equivalence, blogging autonomy, cognitive sustainability, consideration for local context and ability to cope with a dynamic system such as the blogosphere. Although a top-down solution to the issue of semantic interoperability has been proposed, namely the Semantic Web, we suggest that a bottom-up approach is more sustainable. Hence we introduce the Tagsocratic project, whereby we learn the mapping between locally defined categories using a technique called language games.

## 7. ACKNOWLEDGMENTS

This work was partially funded by the DiCA project, thanks to a grant of INRM (Istituto Nazionale Ricerca Montagna) and PAT (Provincia Autonoma di Trento).

# 8. REFERENCES

[1] K. Aberer, P. Cudré-Mauroux, and M. Hauswirth. Start making sense: The Chatty Web approach for global Semantic agreements. *Semantic Web Journal*, 1(1), 2003.

[2] A. Anjewierden, R. Brussee, and L. Efimova. Shared Conceptualisations in Weblogs. In *Proceedings of BlogTalk 2.0*, 2004.

[3] P. Avesani and A. Agostini. A Peer-to-Peer Advertising Game. In M. Orlowksa, M. Papazoglou, S. Weerawarana, and J. Yang, editors, *First International Conference on Service Oriented Computing (ICSOC-03)*, pages 28–42, Berlin Heidelberg, 2003. Springer-Verlag LNCS 2910.

[4] J. Bar-Ilan. An outsider's view on "topic-oriented blogging". In *Proceedings of the 13th Internation World Wide Web Conference*, 2004.

[5] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, May 2001.

[6] S. Cayzer. Semantic Blogging and Decentralized Knowledge Management. *Commun. ACM*, 47(12):47–52, 2004.

[7] dmoz: Open Directory Project. http://dmoz.org/.

[8] H. H. Do and E. Rahm. COMA - A System for Flexible Combination of Schema Matching Approaches. In *Proceedings of the 28th International Conference on Very Large Databases*, pages 610–621, 2002.

[9] A. Doan, P. Domingos, and A. Y. Halevy. Reconciling Schemas of Disparate Data Sources: A Machine-Learning Approach. In *SIGMOD Conference*, 2001.

[10] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. The MIT Press, 1998.

[11] D. R. Karger and D. Quan. What would it mean to blog on the semantic web? In S. A. McIlraith, D. Plexousakis, and F. van Harmelen, editors, *Third International Semantic Web Conference (ISWC-04), Hiroshima, Japan*, volume 3298 of *Lecture Notes in Computer Science*. Springer, 2004.

[12] A. Mathes. Folksonomies: Cooperative classification and communication through shared metadata. Computer Mediated Communication, LIS590CMC (Doctoral Seminar), Graduate School of Library and Information Science, University of Illinois Urbana-Champaign, December 2004.

[13] B. A. Nardi, D. J. Schiano, and M. Gumbrecht. Blogging as social activity, or, would you let 900 million people read your diary? In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 222–231. ACM Press, 2004.

[14] S. Paquet and P. Pearson. A Topic Sharing Infrastructure for Weblog Networks. In *Second Annual Conference on Communication Networks and Services Research (CNSR'04)*, pages 301–304, Fredericton, N.B., Canada, May 2004.

[15] E. Rahm and P. A. Bernstein. A survey of approaches to automatic schema matching. *VLDB Journal: Very Large Data Bases*, 10(4):334–350, 2001.

[16] D. J. Schiano, B. A. Nardi, M. Gumbrecht, and L. Swartz. Blogging by the rest of us. In *Extended abstracts of the 2004 conference on Human factors and computing systems*, pages 1143–1146. ACM Press, 2004.

[17] SKOS: Simple Knowledge Organisation System. http://www.w3.org/2004/02/skos/.

[18] L. Steels and A. McIntyre. Spatially Distributed Naming Games. *Advances in Complex Systems*, 1(4):301–323, January 1999.