BROADENING THE PERSPECTIVE ON CLASSIFICATION SYSTEMS IN THE WEB: ANALYZING WEB CLASSIFICATION AS A SITUATED ACTIVITY WITHIN COMMUNITIES OF PRACTICE

Fernando Figueira Filho, João P. de Albuquerque* and Paulo Lício de Geus
Institute of Computing, University of Campinas, Brazil
**University of São Paulo, Brazil*

ABSTRACT

The terms "social web" and "web 2.0" have been used to refer to cost-effective platforms for collaborative knowledge production in the World Wide Web. In this context, recent web knowledge-sharing tools incite a growing change in the role of users from mere consumers to active knowledge producers. In this paper, we broaden the perspective on the approaches to information classification currently used in the web 2.0 by drawing on social theories that consider the production and consumption of categories as a situated activity that takes place within Communities of Practice. We analyze current approaches to information classification in the web (ontology-based *versus* social tagging approaches), arguing that approaches that impose standardization are not well suited to large-scale information systems in the web and its complex information ecology, where the existence of heterogeneous groups of users demands multiple perspectives of a given knowledge domain. We finally suggest that additional contextual information can be gathered in order to improve current classification approaches and the web systems in which these approaches are embedded.

KEYWORDS

Classification, communities of practice, web 2.0, semantic web.

1. INTRODUCTION

In recent years, we noticed a paradigm change in the World Wide Web. New web application functionalities incite a growing change in the role of a user from a mere information consumer to an active knowledge producer. The phenomena leveraged scientific interest to what has been called "the social web" or "Web 2.0" (O'Reilly 2005). Web logs (blogs), problem-solving discussion forums, wiki-based tools are a few examples of a new class of knowledge-sharing applications which have been considered as a cost-effective platform for collaborative knowledge production.

Although we observe an increase in the amount of information that is produced collaboratively, a closer look shows that the scope of the collaborative activity has never been so fragmented. In such scenario, users are part of wider social structures that have different implicit assumptions and classification systematics. This background is particularly important if we consider that information consumption and production are two situated activities that could occur in different contexts. As such, Bannon and Bødker define the *web information space* as "one of the most open—in the sense of accessible—electronic space that exists, while at the same time […] one of the most closed—in the sense that due to the heterogeneity of users and possible use situations, the possible interpretations of the information that is presented is impossible to know" (Bannon and Bødker 1997).

Considering this background, in this paper we draw attention to the wider implications of classification in the web. Several approaches have been recently employed—from more formalized initiatives such as those based on ontologies, to less structured approaches such as those based on folksonomies. We draw upon the

* This author would like to gratefully acknowledge the support of the Alexander von Humboldt Foundation and of FAPESP (The State of Sao Paulo Research Foundation).

concept of communities of practice (Wenger 1999) to characterize and contrastively analyze both approaches and finally derive implications for design of web information spaces, considering heterogeneous groups of collaborators contributing to content classification. We conclude that co-existing classification perspectives are only manageable when some level of contextual information works as a boundary object to convey meaning to those that collaborate producing and consuming information within the same web information space.

The remainder of this paper is organized as follows: in Section 2 we characterize two opposite trends currently employed in the web as regards to information classification (ontology-based versus social tagging approaches). In Section 3 we provide a revision of our theoretical framework, which serves as a basis for an analysis of web classification systems in Section 4. In Section 5, we draw some implications for the design of web information spaces. We finish the paper with a conclusion in Section 6.

2. CLASSIFICATION IN THE WEB

Presently, there are two different approaches in the World Wide Web to the classification of available information. On one hand, the *semantic web* (Berners-Lee et al. 2001) is based on a set of representation standards (e.g. RDF (Lassila and Swick 1999) and ontology description languages (e.g. OWL (Smith et al. 2004)). The central tenet of the semantic web is the use of ontologies, as opposed to the *social tagging* approach, currently popular among web applications, which could be characterized by its distributed and eminently collaborative classification strategy. The next sections discuss both approaches in turn.

2.1 Ontologies

According to (Gruber 1995), ontology is "an explicit specification of a domain conceptualization". In other words, ontologies are standardized classification models based in a controlled vocabulary and, in this sense, are similar to other standardization efforts developed along human history, such as the living organisms taxonomy in the biology and the international classification of diseases (ICD) prepared by the World Health Organization. Ontology description languages support the composition of a grammar for the usage of its terms in the controlled vocabulary to express something meaningful within a specified domain of interest. As such, the grammar contains formal constraints to the way in which the terms of the controlled vocabulary can be put together (i.e. it specifies what is the meaning of a well-formed statement, assertion, query etc.).

The adepts of the semantic web claim that formal specifications could bring considerable benefits to the web. The major problems faced today by information retrieval mechanisms are related to the fact that, except for formatting rules, hypertext is solely based on natural language. As a consequence, ambiguity problems arise such as polysemy—i.e. a word, term or phrase that have multiple senses (e.g. mouse could refer to the computer device, to an animal or to the surname of a famous Walt Disney character)—and synonymy—when multiple terms have the same sense (e.g. car and automobile). Once applied to the web, ontologies could therefore provide disambiguation, among others benefits such as the employment of inference rules to discover related terms in a web search. In brief, ontologies promise to bring more intelligence to the web, but the benefits do not come cheap, as discussed bellow.

Most ontology projects strive for structural and temporal stability, that is, ontology engineers assume that the structure, vocabulary and conceptual relationships of the ontology are to be defined once and then be kept relatively stable along time. This way the initial requirements remain valid, which avoid restructuring efforts that could be expensive and time-demanding. However, in real-world settings, ontologies suffer from what it is called the "knowledge acquisition bottle-neck" (Zhou 2007)—i.e. the limitations of the approach to automatically evolve its classification schema in order to reflect changes in a given real-world domain.

Hepp (Hepp 2007) points out three fundamental limitations of an ontology-based approach. First, in several knowledge domains there is a significant delay between the updates performed by ontology engineers and the conceptual dynamics of the domain, i.e. new concepts become relevant, while others fall rapidly into obsolescence. This can be considered one of the major drawbacks of the ontology approach as regards to large-scale information systems. The second limitation of ontologies comes from an economic perspective: in various real-world situations, the savings obtained through the automation that is enabled by the ontology cannot cover the resources spent to maintain the ontology itself. The reason is that some level of human

intervention is required even in the state-of-the-art approaches on ontology management—e.g. techniques for automated ontology learning are still semi-automatic, since at least the outcomes generated by the technique have to be manually validated (Zhou 2007). Third, considering that an ontology-based approach generally depends on human intervention, there is a need in some cases to cope with the gap between the natural language known by humans, and the formal language used to specify the ontology. Even if we consider a group of specialists in a given domain, few of them would have knowledge in description logics and languages such as RDF and OWL. Although this problem could be mitigated with tools that attempt to bridge the gap between formal and natural languages, in most cases it is hard to guarantee that the final product will meet its initial requirements.

2.2 Social Tagging

Social tagging has been growing in popularity among web services. *Tags* are terms that are collaboratively associated with web content (e.g. a video, a web page, a photo) as annotations. The outcome of this collaborative classification activity has been called folksonomy (from *folk* and *taxonomy*). We observe the application of this approach in several services available in the web, such as content-sharing portals (e.g. Flickr), social bookmarking tools (e.g. del.icio.us), among others.

The operation of a web system which supports a folksonomy (e.g. del.icio.us) is basically the same in every application: once authenticated, users have the option to associate tags to contents of their interest. Tags can be associated with contents shared by the users themselves or assigned to contents shared by other users. There is no control over the terms which can be applied (i.e. in principle, every character sequence is a valid tag, with minor exceptions).

As a product of a collaborative classification strategy, folksonomies are advantageous to the extent that classification efforts are distributed among a large set of users, usually composed by thousands of collaborators. As a consequence, folksonomies are less susceptible to the knowledge-acquisition problem mentioned before. Indeed, this collaborative character allows for new emergent informational niches and concepts to be readily incorporated in folksonomies. A second advantage of folksonomies in comparison to more traditional approaches such as directories (or folders) is related to the paradigm for organizing contents into categories. In a folksonomy, if we consider tags as categories, a given content could be put into more than one category at the same time and thus it does not belong to any category exclusively. This is certainly advantageous if we consider the multiple facets of the classification of any given web content. A third benefit of social tagging systems is to offer a low entry barrier, which works as an incentive to user participation (Bannon and Bødker 1997). In particular, social tagging systems are easy to use and can be integrated as free plug-ins into web browsers (e.g. del.icio.us). Additionally, users are not obliged to use a controlled vocabulary to classify content, which exempts users from memorizing pre-defined terms or concepts. In this sense, folksonomies develop a community-driven vocabulary, which reflects the terms used by people to label things under their own classification systematics.

However, folksonomies also suffer from ambiguity problems, just as common hypertext currently employed in the web. Particularly, terms in a folksonomy do not have explicit relations among each other, conversely to a directory-based approach, in which terms hold parent-child relationships. Furthermore, a given term does not have relation to formally specified concepts and, consequently, does not allow the inference of semantic relations (e.g., there is no rule that relates the terms *car* and *automobile*). These characteristics lead to problems in information retrieval as regards to the precision of search results (i.e. retrieval of irrelevant information) and recall (i.e. the omission of relevant results).

Some hybrid approaches try to overcome these problems by exploring implicit semantic relationships in folksonomies (Begelman et al. 2006)(Wu et al. 2006), which are automatically inferred based on patterns of co-occurrence among terms, i.e. two terms are often used to categorize a given content, so there is a high probability that they are related. However, the type of semantic relationship (e.g., part-of, is-a) is a way more difficult to be derived solely by using automated methods, which are inherently less accurate than human supervised methods for classification.

In summary, choosing between an ontology-based or a social tagging approach demands the designer to take several trade-offs into account, many of them including aspects that cut-across technical boundaries. Earlier works have pointed out technical characteristics in every discussion about classification approaches under the assumption that such ontologies or folksonomies are easily usable. This leads to our main research

question: is a classification approach meaningful to all actors and in its multiple situated instances? Our question considers both classification approaches as black boxes, concealing their technical details and turning the focus of analysis to the social use context of their classification systems.

3. THEORETICAL FRAMEWORK

In the aim of developing an alternative perspective for the analysis of classification systems, this session introduces our theoretical framework based on the concepts of Communities of Practice (Lave and Wenger 1991)(Wenger 1999), Boundary Objects (Star and Griesemer 1989), and on the distinct perspectives on context proposed by Dourish (Dourish 2004). Those concepts are part of a series of studies on science and technology (Star 1999)(Star 1991)(Star and Strauss 1999)(Star 1992) and on human cognition (Lave 1988)(Suchman 1988)(Hutchins 1995).

3.1 Communities of Practice and Semantics

The concept of communities of practice (CoP) has its origins in social learning theories developed by anthropologists such as Lave (Lave 1988) in late the 80's. According to Wenger (Wenger 1999), "communities of practice (CoPs) are groups of people who share a concern or a passion for something they do, and learn how to do it better as they interact regularly". These communities are not necessarily affiliated to an organization and, consequently, are not defined based on traditional organization principles such as pre-defined hierarchies, inflexible division of labor and formal norms of conduct. Moreover, CoPs are not necessarily geographically co-located and its members could interact solely by using a virtual medium (e.g. web systems). The concept was primarily designed based on the observation of informal learning processes. Recently it was applied to the study of the impacts of classification systems in a broad range of information systems (Bowker and Star 1999).

Central to the definition of CoPs is the concept of practice. For Wenger (Wenger 1999), "practice is first and foremost a process by which we can experience the world and our engagement with it as meaningful". The centrality of practices for understanding social processes corresponds to a point of view employed in different branches of contemporary social theory, as pointed out by Reckwitz (Reckwitz 2002). Against this background, we are interested not only in what people do, but rather in how people's actions are rendered meaningful to them. The question on 'how a regular practice affects the sense of meaning' is a major concern to our analysis as long as we are investigating how to deal with different contexts—content consumption and production—when considering a heterogeneous group of people that might have different classification practices for a given content.

The reasons for such differences are diverse. For instance, domain knowledge is a major cause of misunderstanding when people with different backgrounds interact producing to and consuming information from a particular system. People that participate regularly are more likely to know specific vocabulary and jargons which are crucial to the proper understanding of certain practices. Linux specialists, for example, know several words that are required when using keywords to filter information in the web. In many cases such as in web forums, novices depends on expert help to solve technical problems because they are unable to find relevant information solely by using vocabulary that they know. After interacting in the community of practice, in turn, novices become experts through a process called legitimate peripheral participation (Lave and Wenger 1991). As a central pivot of the communication between experts and novices, categories play a significant role as boundary objects, as we analyze below.

3.2 Categories as Boundary Objects

The concept of boundary object (Star and Griesemer 1989) is especially useful in the context of information classification systems. As Star and Griesemer put it: "boundary object is analytic concept […] which both inhabit several intersecting social worlds […] and satisfy the informational requirements of each of them. Boundary objects are objects which are both plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites"

As such, categories are boundary objects that have an important role in the juggling of meanings created by people's interaction. In the real world, boundary infrastructures serve multiple CoPs simultaneously by bringing into play stable regimes of boundary objects such that any given CoP can interface with the information system and pull out the kinds of information objects it needs. However, as Bowker and Star (Bowker and Star 1999) put out, these boundary infrastructures are not perfect constructions. For instance, contingencies caused by the practical use of boundary objects as standardized categories are solved in the real-world by employing work-arounds such as the use of ad hoc nonstandard categories, which is a suitable work-around in direct, face-to-face human interaction. Web information systems, in contrast, pose a limitation as regards to the software systems in which a given classification perspective is embedded. As such, user access to available information is inevitably mediated by the categories employed in web content classification.

3.3 Context as Representation versus Context as Interaction

As boundary objects, categories cannot be considered individually: when put together in a particular situation, they contribute to build a social context. Different fields of computer science and information systems have considered the problem of context. However, understanding what is context and how it may be used to improve the user experience when using information systems is still an open issue in several real-world applications. In particular, there is a need to devise principles which can help designers to cope with the several contexts involved in the heterogeneity of large-scale information systems.

As Dourish (Dourish 2004) points out, a widespread form to take context into account is to consider it a representational problem. According to this perspective, context and semantic relations are a form of information, and hence could be encoded and represented much as other information is encoded and represented in software systems. Furthermore, semantic information is considered relatively stable. Although the precise elements of a semantic representation might vary from application to application, they do not vary from instance to instance of an activity. The determination of relevance of any potential contextual element can be made once and for all.

In contrast to this perspective, Dourish proposes an alternative view of context, inspired by sociological investigations of real-world practice. According to Dourish (Dourish 2004), "contextuality comes about only when it is mutually recognized by the parties to some interaction, drawing on their everyday, cultural, common-sense understandings of the nature of the social world". Taking into consideration the social contexts of information production and consumption, the aim of any information system is to make the production activities and their outcomes meaningful in the context of consumption. As such, as opposed to the aforementioned representational perspective, we understand semantics as a *relational property*, in a sense that it emerges as a consequence of the articulation of meanings among users within a particular information system.

4. THE WITHIN AND THE BETWEEN IN LARGE-SCALE CLASSIFICATION SYSTEMS

Based on the theoretical framework of the previous section, we look at the classification approaches of web information systems in a different fashion. With the advent of information technology and the creation of virtual environments for social interaction, new possibilities for knowledge management are emerging. However, in large-scale, heterogeneous environments such as the web, enabling the creation of Communities of Practice (CoPs) with the purpose of knowledge exchange is a challenging task for several reasons. Firstly, web information spaces have to satisfy information requirements of distinct CoPs, since the wide dissemination of web systems entails the usage of a system by several, heterogeneous groups of users. Secondly, knowledge articulation amongst members of the same CoP occurs in a different manner than articulation that cut across community boundaries. Furthermore, the practices of knowledge sharing also vary amongst the members of a CoP with different levels of domain expertise, i.e. knowledge sharing between experts happens in a different way than that with novices. Therefore, from this perspective a web classification system has to be flexible enough in order to provide meaning *within* a given community and *between* different communities.

Brown and Duguid (Brown and Duguid 1999) state that "knowledge moves differently *within* communities than it does *between* them". As regards to knowledge classification and to how knowledge should be categorized, a democratic common information space will always have to cope with the tension between the ambiguous (outsider, *naïve*, strange) and the naturalized (at home, taken-for-granted). Indeed, the more at home you are in a community of practice, the more you forget the strange and contingent nature of its categories seen from the outside (Bowker and Star 1999). In this sense, naturalization is an outcome of the membership process in a CoP. Web information systems and its information classification approaches have a fundamental role in the management of this tension, thus enabling the information flow *within* and *between* CoPs.

Therefore, the process of mediation imposed by the classification approach which is embedded in a web system could either explore the whole potential of the common information space and satisfy the informational requirements of its different participants or, in contrast, under-utilize the available infrastructure and limit the virtual experience. Under this perspective, categories, people and information assets form a complex scenario to a system designer. The needs for a classification system which can cope with this complexity push our investigation to a higher level of abstraction, thus permitting us to revisit the approaches of ontology and social tagging addressed in Section 2, so as to derive some implications for the design of web information spaces.

5. IMPLICATIONS FOR DESIGN

The first question a system designer should consider is not whether using an ontology or a folksonomy-based approach, but if the classification system is suited to the intended social context of use, which means that it democratically provides meaning to all users in the information system. As an example, consider a technical web forum such as the Ubuntu Forums (cite). In such environments, more experienced users help newcomers to solve technical problems. Although it is easy to produce information by posting new messages in a particular forum, reusing the available knowledge is still an open issue, basically because certain categories attributed to technical content do not properly convey meaning to a novice user, making it difficult for newcomers to filter, recognize and analyze available content. In this sense, regardless of whether a category is produced automatically or manually, is part of a controlled or non-controlled vocabulary, or other dualities presented along this paper, the primary goal of deciding which approach to use considering only technical constraints is misleading and should be revised to take into account the social classification practices of the intended users in the first place.

In a certain way, the constant criticism deferred towards ontology-based approaches is not related to the ontologies themselves, but to how knowledge is acquired, updated and used. The ideology behind the semantic web is to conceive semantics as a representational problem as seen in Section 4. In this perspective, categories are semantically related based on information pre-conceived in data structures, which could be automatically produced—as is the case of hybrid and some semi-automatic ontology-based approaches—or manually constructed by specialists in a given domain. In all these cases, semantic information is conceived as *a priori.* Categories and their instances, i.e. contents, are related based on structural information that is independent of the situation in which those categories were in fact used. Following this strategy, semantics (i.e. the pre-conceived structural relations between terms) and activity (i.e. acts of search and classification) are separable.

This separation between semantics and activity proves very problematic when we consider current large-scale web information systems, for in these systems the activities of production and consumption of categories cannot be assumed to happen within an homogeneous group of users. For instance, although the categories used to classify the information may be familiar to information producers, they may not be familiar to the information consumer—i.e. they are not part of the information consumer's vocabulary or are not readily available—in a cognitive sense, during the consumption activities. In those cases, the designer should come up with alternatives to contextualize such a category by making references to pieces of domain knowledge that are familiar to the user.

In the aforementioned example of a Linux web forum, consider that a user does not have experience with *grep*, a command-line Linux tool, but is trying to accomplish a task that demands its use. As such, the word *grep* detached from a context that is meaningful to the user does not have any utility. However, when it is

presented to the user along with sense-making references, the user has an opportunity to learn more about the tool and associate its function to their current task. This association can be easily done, for example, by presenting the word *grep* in a brief explanation about what the tool actually does. This explanation could be a web forum post or a summary of a threaded discussion. This position is thus intrinsically related to Dourish's vision on context as an interactional problem posed in (Dourish 2004). Under this perspective, rather than defining semantics in advance, the alternative view argues that semantics is defined dynamically. In a classification system, this means that meaning is not defined a priori. Instead, it depends on occasioned properties, instances of action and particular parties to that action.

6. CONCLUSION

Current investigation on methods to web content classification follows two main streams: a) the *semantic web*, which employs formalization in order to explicitly describe a domain conceptualization; and b) *social tagging*, which has given rise to community-driven classification schemes so-called folksonomies. Although this two-fold differentiation is particularly relevant when considering technical matters, it reveals itself incomplete to take into account the social practices of classification in the context of large-scale, heterogeneous information systems.

In this paper, we argue for a broadened perspective that considers the classification of contents in web information systems as a social practice, which takes place within the shared context of a Community of Practice (CoP). In this way, we depart from a technical view of semantics that understands context as a representational problem that could be fully expressed by term structures independent from their situations of use. Instead, we consider that the production and consumption of categories in web information systems are fully-fledged situated activities that build upon and engender a set of classification practices shared within a CoP. As such, web classification systems must be seen as complex information ecologies in which people have different levels of domain knowledge and distinct classification systematics, depending on which CoPs they belong to, and on their level of domain knowledge in each community. Of crucial importance here is that some categories may function as boundary objects, i.e. acting as a bridge that allow communication between members of different CoPs or between members of a CoP with different level of domain expertise.

In this manner, following this new conceptualization of classification in web information systems, finding, interpreting and reusing available knowledge could be hard activities depending on what extent the employed category system proves familiar to a given user. To address this issue, we propose that designers should pay greater attention to the social context and its particular classification practices, in order to make sure that the categories embedded into a classification system can be readily articulated in the practices of users. One possible strategy, for instance, is to make sure that contextual information about previous classification practices involving a given category is provided as a boundary object. This contextual information produced by other users within the system would be able to bridge the gap between members of distinct communities of practice or people holding different levels of domain knowledge about a particular subject. Instead of pressing for homogeneity and universal standardization, this approach could thus improve the support to diversity and heterogeneity, contributing to better exploit and foster the richness of large-scale information spaces as the web.

REFERENCES

Bannon, L. and Bødker, S. 1997. Constructing common information spaces in *ECSCW'97: Proceedings of the fifth European Conference on Computer-Supported Cooperative Work*. Norwell, MA, USA: Kluwer Academic Publishers, pp. 81-96.

Begelman, G. et al. 2006. Automated Tag Clustering: Improving search and exploration in the tag space. *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland*.

Berners-Lee, T. et al. 2001. The semantic Web. *Scientific American*, 284(5), pp. 28-37.

Bowker, G. and Star, S.L. 1999. *Sorting Things Out: Classification and Its Consequences*. MIT Press, Cambridge, MA.

Brown, J.S. and Duguid, P. 1999. Organizing Knowledge. *Reflections*, 1(2), pp. 28-44.

Dourish, P. 2004. What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8(1), pp. 19-30.

Gruber, T.R. 1995. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5/6), pp. 907-928.

Hepp, M. 2007. Possible Ontologies-How Reality Constrains the Development of Relevant Ontologies. *IEEE INTERNET COMPUTING*, pp. 90-96.

Hutchins, E. 1995. *Cognition in the Wild*. MIT Press, Cambridge, MA.

Lassila, O. et al. 1999. Resource Description Framework (RDF) Model and Syntax Specification. *W3C Recommendation*, 22, pp. 2003-2004.

Lave, J. and Wenger, E. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

Lave, J. 1988. *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*. Cambridge University Press.

O'Reilly, T. 2005. *What is Web 2.0 | O'Reilly Media*.

Reckwitz, A. 2002. Toward a Theory of Social Practices: A Development in Culturalist Theorizing. *European Journal of Social Theory*, 5(2), pp. 243-263.

Smith, M.K. et al. 2004. OWL Web Ontology Language Guide. *W3C Recommendation*, 10.

Star, S.L. and Griesemer, J.R. 1989. Institutional Ecology, Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19(3), p. 387.

Star, S.L. and Strauss, A. 1999. Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW)*, 8(1), pp. 9-30.

Star, S.L. 1991. Power, Technologies and the Phenomenology of Standards: On Being Allergic to Onions. *J, Law, ed. A Sociology of Monsters*, pp. 27-57.

Star, S.L. 1999. The ethnography of infrastructure. *American Behavioral Scientist*.

Star, S.L. 1992. The trojan door: Organizations, work, and the ``open black box". *Systemic Practice and Action Research*, 5(4), pp. 395-410.

Suchman, L.A. 1988. Representing practice in cognitive science. *Human Studies*, 11(2), pp. 305-325.

Wenger, E. 1999. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press.

Wu, X. et al. 2006. Exploring social annotations for the semantic web. *Proceedings of the 15th international conference on World Wide Web*, pp. 417-426.

Zhou, L. 2007. Ontology learning: state of the art and open issues. *Information Technology and Management*, 8(3), pp. 241-252.