


Reconciling Knowledge in Social Tagging Web Services

View metadata, citation and similar papers at core.ac.uk

brought to you by  CORE

provided by idUS. Depósito de Investigación Universidad de Sevilla

¹ Universidad de Huelva. Department of Information Technology,
Crta. Palos de La Frontera s/n. 21819 Palos de La Frontera

² Universidad de Sevilla. Department of Computer Science and Artificial Intelligence,
Avda. Reina Mercedes s/n. 41012 Sevilla. Spain
gonzalo.aranda@dti.uhu.es, jborrego@us.es

Abstract. Sometimes we want to search for new information about topics but we can not find relevant results using our own knowledge (for example, our personal bookmarks). A potential solution could be the use of knowledge from other users to find what we are searching for. This solution implies that we can achieve some agreement on implicit semantics used by the other users. We call it *Reconciliation of Knowledge*. The aim of this paper is to show an agent-based method which lets us reconcile two different knowledge basis (associated with tagging systems) into a common language, obtaining a new one that allows the reconciliation of (part of) this knowledge. The agents use Formal Concept Analysis concepts and tools and it has been implemented on the JADE multiagent platform.

1 Introduction

The amazing growth of Web 2.0 provides powerful technologies for sharing information among users (members of social networks) as, for example, the social indexing of the digital objects of the Web. Collaborative tagging represents a very useful process for users that aim to add metadata to documents, objects, resources, urls, etc. Among other applications, the tagging enable users to achieve personal knowledge organization according to their own interests. Additionally, the Web 2.0 systems can extract (by means of Collective Intelligence methods) some global organization of the information (from a user's personal point of view). This way the collaborative tagging offers a pragmatic alternative to the semantic web ontologies. However, the gap between the personal organization of information and the global one (as well as between that of different users) makes the use of automated methods to reconcile them difficult.

These different ways are combined in tagging tools that the tag-based platform facilitates. This situations leads to a crowd of tagging systems. Moreover, inside of the platform and due to the preferences of the users, different tagging

* Partially supported by TIN2009-09492 project of Spanish Ministry of Science and Innovation, cofinanced with FEDER funds.

behaviours exist that actually obstruct the automated interoperability among tag sets. Despite the fact that the systems offer solutions to aid the understanding of the folksonomy that the users collectively build (tag clouds, tools based on related tag ideas, collective intelligence methods, data mining, etc.) Although tagging shows potential benefits, personal organization of information leads to implicit logical conditions that often differ from the global one. Tagging provides a sort of weak organisation of the information, very useful, but mediated by the user's behaviour. Therefore, it is also possible that user's tags associated with an object do not agree with the other users tags.

Formal Concept Analysis (FCA) is a mathematical tool that, applied to tagging systems, makes explicit the set of concepts that the user manages in tagging, as well as the structure of the relationship between them [7]. The concept lattice (a mathematical structure extracted by FCA methods) represents an intermediate structure between tagging (nonhierarchical and inclusive) and classical taxonomies (hierarchical and exclusive). Thus, FCA is useful to bridge the semantic gap providing a solid mathematical theory to tagging [7].

1.1 Motivation

Since that user's tagging reflect their own set of concepts about the documents, two of the main tools of FCA, namely the concept lattice and Stem basis, shows distinct results for different users (semantic heterogeneity). From the point of view of navigation by means of tags, the semantic heterogeneity makes the activity insecure. Thus, to ensure an efficient use of another user's tags, some reasoning on tags must be performed, in order to achieve some consensus (also represented by FCA tools) that allows navigation between different concept structures. In this scenario, it could be very important to attempt to delegate these tasks to intelligent agents (see fig. 1). Our aim is to show how the authors have solved this problem. The solution presented in this paper was designed in the framework of Mobile Web 2.0 project (Mowento), although the solution proposed is valid for any tagging system (in fact, in this paper the method is applied to a well known social bookmarking platform, *Delicious*¹).

The aim of Mowento is that anyone can publish content (on the WWW) both videos and photos from anywhere at anytime, without needing a next-generation mobile device[2]. Mowento allows users to annotate basic information semantically. This annotation is in principle very limited due to the usability of non-advanced mobile devices, which do not allow the use of complex applications for tagging. We address the challenge of creating a simple and effective labeling method for the content, which should be able to be properly labeled with a few clicks. The method consists of a series of hierarchically arranged menus whose construction algorithm is based on the Formal Concepts Analysis [1]. From the point of view of the Mowento server, the information is received and is automatically entered into a database, pending processing. From here, the multiagent system (programmed on JADE²) takes control of the process and performs its

¹ <http://delicious.com>

² <http://jade.tilab.com/>

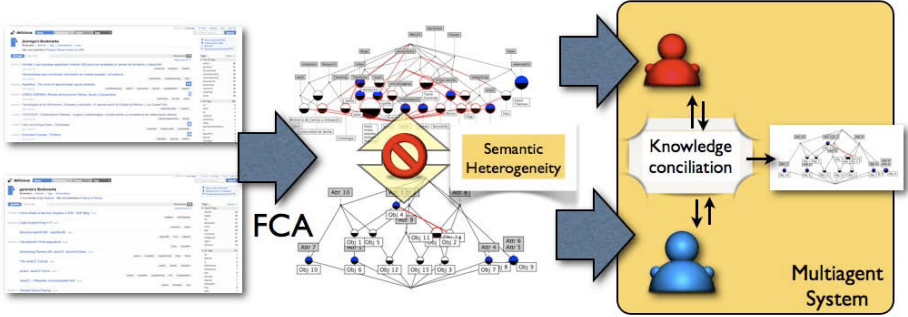


Fig. 1. Knowledge conciliation in social bookmarking represented by concept lattices

tasks. In this context several tagging problems have been solved by means of agents. Among others, this is the main aim of this paper, namely the agent-based reconciliation method. The solution method presented in this paper is also applicable to platforms with tag-based organization of information such as Delicious, which will be used as an example. Mowento is in experimental phase, and user generated content allocated in the project does not provide representative examples while personal bookmarks of authors in Delicious represent a good sample for showing results.

2 Tagging and Heterogeneity

In the case of bookmarking systems as *Delicious*, different features and users' behaviours represent a similar problem to one faced in the Mowento project: how to organise folksonomies by means formal ontologies (or better, ontologies on user's tags). Although tagging is useful to navigate among pages on the WWW, it can not be considered as a robust knowledge organization method. Some methods exist to integrate this kind of knowledge organization into SW realm [10]. These methods can be classified according to the semantics associated with tag sets (or folksonomies). For example, there are methods based on ontological definition of tags which use ad hoc ontologies, in order to formally describe the properties of tags (see [8]). Other methods are based on transforming folksonomies into ontologies (see, e.g., [12]), including ontologies designed for dealing with folksonomies [6] or more concrete proposals, as in [9].

2.1 Heterogeneity

As is argued in [5], tagging is fundamentally about sensemaking, a process in which information is categorized, labeled and, critically, through which meaning emerges [13]. Even in a personal tagging structure, boundaries of concepts and categories are vague, so some items are doubtfully labeled. Lastly, users also use the tagging for their own benefit, but nevertheless constitute a useful public good ([5]).

There exist several limitations to collaborative tagging in sites such as Delicious. The first one is that a tag can be used to refer to different concepts; that is, there is a context dependent feature of the tag associated with the user. This dependence limits both the effectiveness and adequacy of collaborative tagging. The limitation is called "Context Dependent Knowledge Heterogeneity" (CDKH). A second is the Classical Ambiguity (CA) of terms, inherited from natural language and/or the consideration of different "basic levels" among users [11][5]. CA would not be critical when users work with urls (content of url induces, in fact, a disambiguation of terms because of its specific topic). In this case, the contextualization of tags in a graph structure (by means of clustering analysis) distinguishes the different terms associated with the same tag [3]. However, CDKH is associated with concept structures that users do not represent in the system, but that FCA can extract. It is also possible CDKH is associated with the potential future use of the tagging (it can be used for classifying documents, for facilitating navigation among visited urls, to collect specific and temporal urls, etc.). Thus, navigation among concept structures of different users faced up with CDKH.

In the case of platforms such as Mowento, CDKH is a less important problem than with collaborative tagging such as Delicious. This is due to both the specific scope of the activities (reporting testimonials about events) and the common language represented by the tags offered by Mowento's mobile tagging widget. In Mowento, CDKH can occur only in specific concepts of the personal concept lattice. Thus, reconciliation is easier than collaborative tagging systems. However, in sites such as Delicious CDKH represents the main problem, because tags perform several functions as bookmarks (see [5]).

3 Agent-Based Reconciliation Knowledge Algorithm

To implement the algorithm, a solution has been chosen based on a multiagent system, which make the extension and distribution of our algorithm no big effort. The multiagent system has to satisfy some requirements, such as to be FIPA compliant, in order to facilitate communication and integration with other multiagent systems. We also thought that it should be, as far as possible, open source. Jade was selected since it is composed of a set of tools for developing agents and an execution platform where the agents can live. Another major point in this decision was that it is developed in Java, a multiplatform language. The reconciliation algorithm consists of the following sequence of steps (see fig. 2):

1. **Agent creation step:** It starts by creating two Jade agents, passing the agent names and Delicious data as parameters. They know the existence of each other within the platform, so it is not necessary to search -at Service Directory level, managed by the Directory Facilitator agent- another agent that offers the *reconciliation service*. White Pages registration is transparent to developers because it is already implemented in the Jade toolkit.
2. **Building formal contexts and Stem basis:** In this step, the agents work in parallel mode, with no interaction, by loading and setting their own knowledge base (KB). They work with the formal context which is built from the

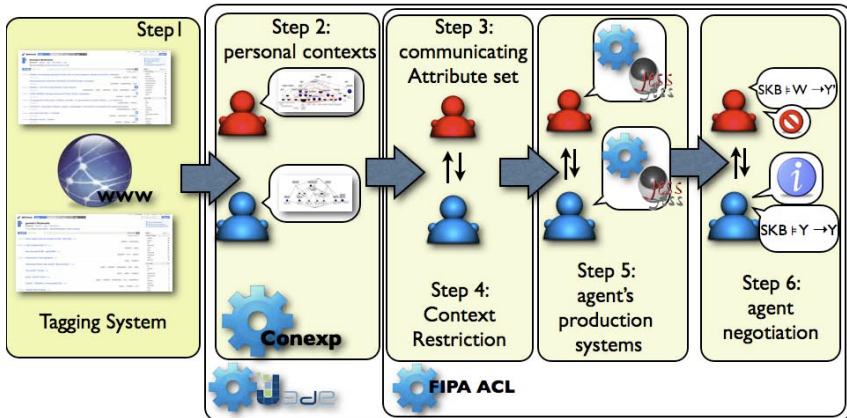


Fig. 2. Reconciliation algorithm

Delicious downloaded information, where the objects are the urls and the attributes are the associated tags. With data, the context is built, and concepts and Stem Basis (SB) are extracted. To obtain such elements it has been integrated with the *Concept Explorer* tool, ConExp³, which provides all the FCA algorithms that we need. It is developed in Java, it allows us a fast deployment of FCA algorithms. ConExp comes as a compressed file ".jar" to be included in the classpath application, and from there, we can instantiate the necessary objects for the computations.

3. **Initializing agent dialogue step:** Once the agent is initialized, he has to execute a double task related with communication. On one side, the agent sends its own language (attribute set) to the other agent. On the other side, the agent prepares itself for the reception of the same kind message from the other agent. For agent communication, we try to adjust the intention to FIPA performatives and its meaning, so that each message is associated with the best one, according to the content. Specifically, the sending of one language of an agent to another one is done through the **INFORM** performative.
4. **Restrictions of own formal contexts:** After this brief communication, agents reduce their languages (the attribute set) to the common language, restricting their formal contexts to that language. This restriction also implies that many objects are now outside of the restriction of the context, because it has discarded those that are not labelled with any common tags, and these contribute nothing to our knowledge base. With the restricted contexts, agents compute the new concept lattices, as well as their concepts and the Stem basis.
5. **Synthesizing the production system from Stem Basis:** From the stem basis, calculated in the previous step, agents consider the rules that have a support greater than zero. In this paper we call this set of implications *Stem Kernel Basis* (SKB). Based on the SKB, a production system is synthesized,

³ <http://sourceforge.net/projects/conexp/>

that it will serve later to suggest to the other agent the changes to objects so they can be accepted by the common ground. This production system (used for the new tags' suggestions) has been completely implemented, because the inference engine requirements were few and not worth the effort to integrate with any other engine, such as Jess⁴ or Drools⁵.

6. **Knowledge negotiation between agents:** To execute this step, a phase of implementation is necessary, which is clearly multiagent in character, in which a deep agent communication/negotiation is produced. Though a turn-based communication or alternating shifts could have implemented, a more asynchronous one that respects even more, the multiagent philosophy is preferred. The reason is that usual scenario consists of agents' KB of different sizes, so the communication needs of each agent will be different.
- The negotiation begins with the creation, by each agent of a new context where the common knowledge will be stored and will produce the results of the reconciliation. Then, a massive sending of all objects (associated tags included) to the other agent is performed and it waits for the objects and responses from it. All of these sent messages are described by the PROPOSE performative.
 - When an agent receives an object from the other one, we check whether the object satisfies all the implications of the agent's SKB, and if so, it includes it within the common context and it also sends an acceptance message to another agent (ACCEPT_PROPOSAL performative) so it can also include it in its common context.
 - If the object does not meet SKB, it introduces it into the production system, created from the SKB, and checks if any of the attributes obtained can be added to the object in order to be accepted by the SKB. This object is then sent back to the other agent as a "new object", restarting the negotiation about this object. If any suggestions are returned by the production system, we will send a message of rejection (REJECT_PROPOSAL performative) to the other agent to proceed to remove the object, as we did.
 - Once made the whole process of message exchanging and negotiation has finished, the agents will get a common context. So it can extract new concepts and suggestions from the stem basis. These represent a shared conceptualization

3.1 Example

As we explain above, Delicious has been chosen as a test environment to illustrate the method. For reasons of paper length, it is not possible to show the trace of the method. For the experiment, authors' accounts has been selected in Delicious (<http://delicious.com/garanda> and <http://delicious.com/jborrego>), which share common interests. This us find a significant common language, and

⁴ <http://www.jessrules.com/>

⁵ <http://www.jboss.org/drools/>

User	jborrego	garanda
Language	381	137
Bookmarks	358	536

User	jborrego	garanda
Common Tags	19	
Bookmarks	131	114
Implications	11	11

Fig. 3. User data statistics before(left) and after(right) reduce to common language

User A		User B
(tutorial) (robotics) \longrightarrow (ai)		(twitter) (blog) \longrightarrow (social) (web2.0)
(twitter) \longrightarrow (social) (web2.0)		(tutorial) (twitter) \longrightarrow (web2.0)
	(socialnetworking)	(facebook) \longrightarrow (twitter)
(facebook) \longrightarrow (haskell) (tutorial)		

Fig. 4. Rules before conciliating

(tutorial) (robotics) \longrightarrow (programming) (ai) (haskell) (blog)
(tutorial) (programming) (haskell) (blog) \longrightarrow (ai)
(twitter) \longrightarrow (social) (web2.0) (blog) (socialnetworking)
(facebook) \longrightarrow (social) (twitter) (tutorial) (haskell)
(web2.0) (blog) (socialnetworking)

Fig. 5. Some rules produced by reconciliation process

Conciliation	
Language	19
Bookmarks	245
Implications	21

Fig. 6. Size of conciliated knowledge

the conciliated knowledge could be more interesting. The size of the data referred to users' accounts (see fig 3, left) with attributes (language) and objects (bookmarks).

According to the multiagent protocol, it has to set the common language and reduce the context, leaving the common attributes and removing the objects with no tags in the common language. Results are in fig. 3, right (step 4). In fig. 4, part of the rule sets corresponding to both agents is depicted. Fig. 5 shows some of the rules after reconciliation.

Finally, we obtain a common context with a small number of objects and a greatly reduced number of implications with a support greater than zero (last step) (in fig. 6 is presented some information on this context).

4 Conclusions and Future Work

In this paper, a method to reconcile knowledge basis associated to tagging systems is presented. The method is based on FCA, and designed on a multiagent system, where agents collaborate in order to establish a common knowledge represented both a new tagging and a concept lattice. It is based on dialogs, so it is

interesting to compare them with some standard protocols, such as contract-net or similar, and in the near future to adopt one of them.

Reconciliation knowledge method can be applied to any tagging-based system. Experiments on Delicious show that after a small number of taggings on the same item, a nascent consensus seems to form and this consensus is not affected by the addition of new tags [5]. This stabilisation implies, for the conciliation method presented, that intentions of objects tend to be similar among users. Future work will be focused on extending the algorithm to find consensus ontologies (with a crowd of users) and, if possible, in a semi-automatic way.

References

1. Alonso-Jiménez, J.A., Aranda-Corral, G.A., Borrego-Díaz, J., Fernández-Lebrón, M.M., Hidalgo-Doblado, M.J.: Extending Attribute Exploration by Means of Boolean Derivatives. In: Proc. 6th Int. Conf. on Concept Lattices and Their Applications. CEUR Workshops Proc., p. 433 (2008)
2. Aranda-Corral, G.A., Borrego-Díaz, J., Gómez-Marín, F.: Toward Semantic Mobile Web 2.0 through Multiagent Systems. In: Håkansson, A., Nguyen, N.T., Hartung, R.L., Howlett, R.J., Jain, L.C. (eds.) KES-AMSTA 2009. LNCS, vol. 5559, pp. 400–409. Springer, Heidelberg (2009)
3. Yeung, C.M.A., Gibbins, N., Shadbolt, N.: Contextualising Tags in Collaborative Tagging Systems. In: Proceedings of the 20th ACM Conference on Hypertext and Hypermedia (2009)
4. Ganter, B., Wille, R.: Formal Concept Analysis - Mathematical Foundations. Springer, Heidelberg (1999)
5. Golder, S., Huberman, B.A.: The structure of collaborative tagging systems. *Journal of Information Science* 32(2), 98–208 (2006)
6. Gruber, T.: Ontology of Folksonomy: A Mash-up of Apples and Oranges. *Int'l. Journal on Semantic Web & Information Systems* 3(2) (2007)
7. Jäschke, R., Hotho, A., Schmitz, C., Ganter, B., Stumme, G.: Discovering shared conceptualizations in folksonomies. *Journal of Web Semantics* 6(1), 38–53 (2008)
8. Kim, H.-L., Scerri, S., Breslin, J., Decker, S., Kim, H.-G.: The state of the art in tag ontologies: A semantic model for tagging and folksonomies. In: International Conference on Dublin Core and Metadata Applications, Berlin, Germany (2008)
9. Knerr, T.: Tagging ontology- towards a common ontology for folksonomies (2006), <http://tagont.googlecode.com/files/TagOntPaper.pdf> (June 14, 2008)
10. Smith, G.: 2007 Tagging: People-Powered Metadata for the Social Web. First. New Riders Publishing, Indianapolis (2007)
11. Tanaka, J.W., Taylor, M.: Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology* 23(3), 457–482 (1991)
12. Van Damme, C., Hepp, M., Siorpaes, K.: FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies. In: ESWC 2007 workshop Bridging the Gap between Semantic Web and Web 2.0, May 2007, pp. 57–70 (2007)
13. Weick, K.E., Sutcliffe, K.M., Obstfeld, D.: Organizing and the Process of Sense-making. *Organization Science* 16(4), 409–421 (2005)