A Survey of Semantic Metadata Management Models for the Social Web

Johann Stan, Pierre Maret

An overview and research vision of the use of Semantic Technologies for User Modeling and Collaborative Tagging

Johann Stan[1,2]

[1] Alcatel-Lucent Bell Labs France,
Centre de Villarceaux
Route de Villejust
91620 Nozay, France
johann.stan@alcatel-lucent.com, johann.daigremont@alcatel-lucent.com
[2] Universite de Lyon, F-42023, Saint-Etienne, France
CNRS, UMR 5516, Laboratoire Hubert Curien, F-42000, Saint-Etienne, France
Universit de Saint-Etienne, Jean-Monnet, F-42000, Saint-Etienne, France

Abstract. This work aims at providing a review of recent work in semantic modeling of users. A second focus area is the tagging. Tagging links a user to a resource by a tag and we consider that this approach is suitable to partially answer the question: How semantic technologies can be leveraged to model social networks. Also this is a good reference work to see how users and resources can be interlinked and thus how a social network model can go beyond the traditional people-to-people model.

1 Introduction

The current developments in the field of the semantic web enables a new realm of applications. Semantic web made it possible to have the necessary tools to handle computer-understandable semantics. These tools, generally evolving from XML are used to enrich the description of web-pages. Languages like OWL (Web Ontology Language), RDF (Resource Description Framework), DAML+OIL are some of the most widely used languages. The advantage of these languages is the fact that they are machine readable and strongly related to Description Logics. A state-of-the art on this subject can be found in [1]. The RDF [2] language makes statements about resources in the form of triples: (subject, predicate, object). The subject denotes the resource and the predicate denotes the relationship between the subject and the object, which can be another resource or a literal. The OWL [3] language is a subset of RDF which allows a wide range of inferences providing more semantics in the way of expressing meaning.

2 Towards Semantic User Representation

We examine in this section recent approaches and models that address the issue of user modeling. We intend to show how ontologies emerged from ad-hoc modeling techniques as a powerful way to describe the user. The reason whu we adress the representation of individuals is that this allows to better understand the collaborative aspects.

FOAF (Friend-of-a-Friend) [4] is an ontology-based RDF vocabulary to describe people profiles, friends, affiliations, creations etcThe aim of FOAF is that of a completely decentralized machine-readable social network that is based on personal profiles. The profile contains mostly static data, like personal information, work history, links to contacts and services. The use of FOAF is currently way beyond popular social network communities, like LinkedIn. , which mostly addresses professional interactions, or more "fashionable playgrounds", like Facebook and its rival MySpace [3]. FOAF is currently used only by research communities or semantic-web developers. A person begins by describing himself or herself using the foaf:Person class, listing key attributes such as name, gender and resources relating to them. They can also list their interests, and each person is uniquely identified by using the foaf:mbox propery containing their email address. An alternative identification propery is foaf:openid conforming to the OpenID perspective of using a unique single URI to establish teh identity. The person in question then moves on to describing their friends, each friend is an instance of the foaf:Person class. FOAF is both machine-readable, and human-readable, and was adopted by LiveJournal, the blogging site, to offer the facility for each user to export their personal informatio. FOAF has seen a slow adoption by WEb 2.0 sites and services, this could be due to the lack of interest in the exportation of social information from one rival site to another and thus going beyond the walled gardens. It is, however, the most widely used specification for expressing personal and realtionship information within the Semantic Web community.

In their work, [5] investigates how tagging allows to infer data about user preferences or interests, allowing to create a user model. Tagging is the process where users label or annotate different ressources (web-pages for example) with the objective to share, organize or diffuse them. This is a concept strongly related to the web 2.0 and social networking. The way users employ tags might give an insight on different issues like how interested he is in the given ressource and the type of tags used (many synonims for example) can infer subjective details like level of creativity. Tests and a deep analysis is needed to better unerstand relations between tagging and these high-level concepts (preference, interest). Research is currently in progress examining how ontologies like Wordnet allows to categorize and to infer automatically the type of tag and relationship with the user (matching).

[6] proposes an architecture where semantic user profiles are used in a peer-to-peer mobile environment. In a ubiquitous mobile environment, service providers like a cinema are equipped with bluetooth-enabled devices to broadcast the service (movie being played for exemple). Passing people have mobile devices which store the owner's profile (interests, preferences, disinterests). Easy matching between user interests and service is necessary, and descriptive logic is considered

[3] www.linkedin.com, www.facebook.com, www.myspace.com

a good approach. The user profile is relatively simple, consisting simply of the union of interests and disinterests: A common domain ontology for concepts in both services and profile description is used to make possible this semantic matching using a reasoner. The system is completely peer-to-peer, allowing independence, cost effectiveness, scalability and most importantly, privacy management, which is the result of the fact that profiles are stored on the mobile device.

[7] argues that techniques like RDF and OWL together with ontologies are the key elements in the developpement of the next generation of user profiles. The User Profile Ontology presented grew out from a quite simple model containing semantical contact information encoded in the RDF language.

In the proposed ontology 1, which is built around three imported ontologies (Person Ontology, containing classes relevant only to the user; Organisation Ontology, containing business oriented information and a Common Ontology, containing information relevant to both persons and organisations) personal information like e-mail, telephone, Instant Messaging identifier, physical addresses are uniquely identified by a GenericContactIdentifier class. Social interactions inherit properties from an Event class. These interactions are classified into voice, text, real-time, online communications. The address book of the user is stored in the class ContactGroup. The article presents well the fraemework, but the structure of the profile ontology is not clear in all aspects.

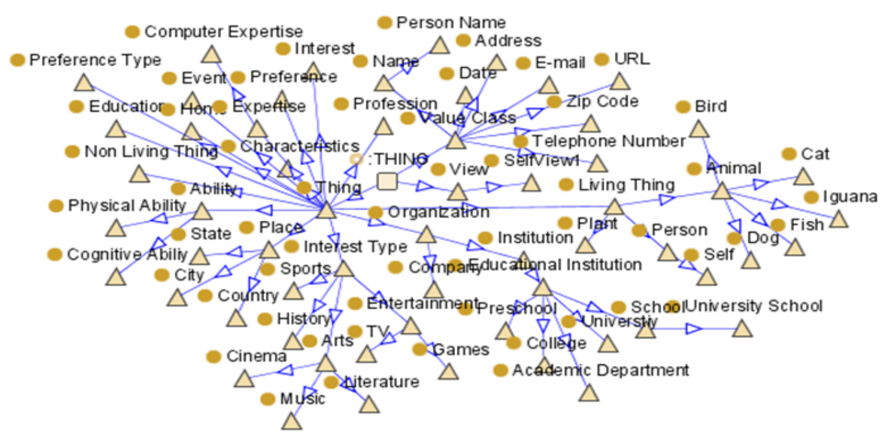

Fig. 1. The User Profile Ontology (some concepts were deleted for a better visibility) viewed with the Jambalaya Protege plugin

The applicability of the Profile Ontology is tested with a Dynamic Address Book application, which maintains dynamically the contact list of a user based on the frequency of communications. Rules are defined to decide when to remove or add a contact according to a given calling frequency threshold. Results show that this is a promising approach, but further evaluation of the system is needed. The framework is realised using among others the Java based Jena engine.

[8] presents an application-independent user profile ontology. The objective is to create a "general, comprehensive and extensible" user model taking into account existing literature, user- and context models. The use of ontologies in user profiling is a known issue, however the problem with existing proposals is that they respond only to application specific needs, mainly in personalized information retrieval and web search. It is important to stress out that the proposed ontology deals only with the static profile of the user, not the dynamic or contextual one (like current position, occupation or terminal).

Before presenting the proposed ontology model, we collect all the necessary concepts for a user profile considered in the literature, based on the same article. This can give an insight of what is currently considered important in describing end-users. A first observation is that dyamic, context-aware concepts are not yet well coupled with the static ones.

Concepts considered important in user profiling	**References**
Static profile User identity, characteristics, capabilities, universal preferences, state of the user, application-specific preferences **Dynamic profile (context)** Current activity, location, terminal, motion state, orientation **Parameters** Personal information (name, address, birthday) **General characteristics** Physical factors, abilities, disabilities, education, occupation, hobbies	[9]
Interests	[10] [11] [12] [13]
Preferences	[14] [13]
User expertise	[13]

Table 1. Concepts considered important in the literature for user profiling

The proposed ontology (Figure 2), where the main class is Person uses these concepts and many others to create a profile applicable in any kind of domain or application. Therefore, no restrictions are present in the ontology, it is completely up to the developer to personalize it according to the specific needs of the project. However, some restrictions could have been added like the fact that a friend of a Person can only be a Friend, which is still domain independent and represents general knowledge.

[15] addresses the issue of modeling users in a context-aware "smart home" environment. Static and dynamic user profiles are distinguished. The Amigo project aims to develop services for a smart home environment, which offers proactivity to users according to current context or situation. A machine-understandable shared user-model is needed between different services to make a unified way of

querying preferences and interests possible. As mentioned before, the user model is separated into two components:

- The context-aware static user profile: Tree-based representation of individual user preferences and personal data, grouped in agreement with user ontology representation in the system The context-aware dynamic user profile

This profile learns user behaviour from history of activities, learning meaning the ability to recommend a given topic in a given situation (for example a movie when Bob is alone at home on Friday night). Interaction history is stored in the form "Context"-¿"User Action", where context is a set of environment descriptors and user action any kind of interaction with the "smart home". Two machine learning techniques, CBR (Case-based reasoning) and SVM (Support Vector Machines) were used to test how items can be classified into terms like "good" or "bad" for a given context. Test results indicate that CBR gives good performances in almost any situation, since the retrieval of items is based on the overall similarity of context descriptors.

UPOS (User-Profile Ontology with Situation-Dependent Preferences Support) [16] is part of the Spice project It is probably the newest user profile ontology, addressing both static and context-aware aspects. This ontology, defined in OWL, allows creating situation-dependent sub-profiles. A user has a profile and a context (location or activity) associated. The notion of condition is defined, which includes a user, an operator and a context-value. For example, a condition can be: "if the context of user Bob equals the MyOffice location...". According to this condition, a corresponding sub profile can be applied that contains all personalization indications for services (e.g. not to use sms). This approach was inspired by the ETSI human factors group, which defined guidelines for designing context-aware user profiles. The most important guideline is to structure the profile into sub-profiles, each containing user preferences that correspond to a specific situation, as seen in the previous example.

3 Linking peoples and objects with the use of annotations/tags

The representation of tags and tagging activity has been going through the same process as user profiles: semantic technologies proved to be an important component in linking concepts together in the web 2.0 era. Many studies have been performed in a lot of disciplines via innovative approaches. However, without consistent structures and semantics, contributions of these studies are unable to analyze the social phenomenon relating the folksonomies. In order to operate social ecosystems on the Web, we need various technical and social analyses for folksonomies as well as formal representation for adopting the results. The semantics of tagging data is primarily about an agreement on the meaning among people or a community in the social space. A common semantics provides a way to share tag representation among services. We now provide an overview of a number of existing efforts that had the common aim of representing the concepts and operations of tags and tagging.

Gruber [17] describes tagging as a relation between:

- An object, i.e. the resource to be tagged. For example, a bookmark, a picture, a blog post etc.
- A tag, i.e. the tag associated with the resource
- A tagger, i.e. the agent -more generally a person -that created the link between the tag and the object
- A source, i.e. the space where the tagging action has been performed, e.g.: Flickr, del.icio.us.

Notably, Gruber defines the source as the scope of namespaces or universe of quantification for objects. This allows one to differentiate between tagging data from different systems and is the basis for collaborative tagging across multiple applications. P. Mika [24] already represented the tagging action from a theoretical point of view, but did not use this notion of source that Gruber introduces. Yet, while this model is widely commented, there is no currently available implementation to our best knowledge. It was also Gruber who defined the term ontology, as an explicit specification of a conceptualization of a domain [25].

Newman et al. [18] defined an ontology of tags and tagging, simply called the Tag Ontology 2, that describes the relationship between an agent, an arbitrary resource, and one or more tags. Thus, in his ontology, the three core concepts Taggers, Tagging, and Tags are used to represent the tagging activity. Contrary to Gruber [17], it does not represent the source of the tagging action. Yet, this ontology has been implemented (in OWL), is available on the Web, and is currently used in some projects such as Revyu.com, a review website combining Web 2.0 and Semantic Web technologies. Notably, in this ontology tags are represented as instances of the tags:Tag class which is assigned custom labels, i.e. the string representing the tag as seen by the user. Being instances of a class means that they are assigned a URI. URIs are a key feature of the Semantic Web, since, contrary to simple literals, they can be used as subject of triples, while literals can be only used as objects. This way, tags-identified by URIs can be linked together and people can semantically represent connection and similarities between tags. For this purpose the ontology introduces a tags:related property.

Yet, this relation does not have much semantics, since it does not define the nature of the relation, e.g. if this a linguistic variation or because it identifies a similar topic. Another limitation is that the ontology does not define any cardinality constraint on the number of labels a Tag can have. This can raise problems since it allows a Tag instance to have two completely disjoint labels (i.e. a Tag instance with labels "RDF" and "Paris"), which makes no sense from a tagging point of view. Still, this ontology reuses pre-defined Semantic

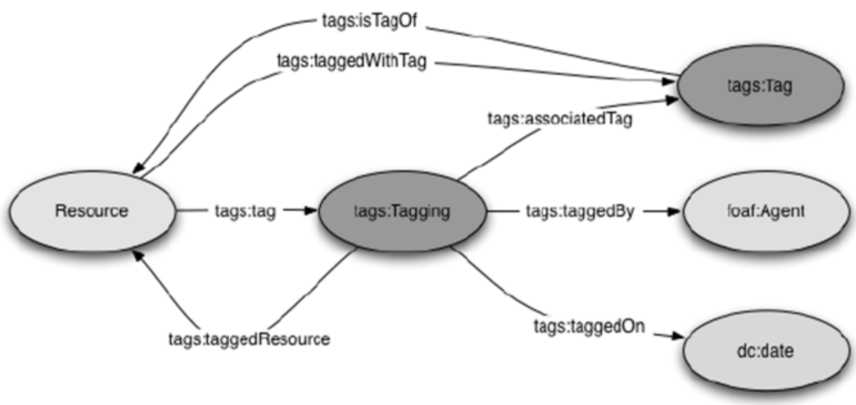

Fig. 2. Tagging Ontology of Newman

Web vocabularies, making it compliant with existing standards. SKOS properties are used to model relations between tags and the Tag class itself inherits from skos:Concept. DublinCore is used to represent the date of a tagging action, with subproperties of dc:date. Finally, the ontology relies on FOAF to identify the tagger of a tagging action thanks to foaf:Person.

SIOC (Semantically-Interlinked Online Communities [19], http://sioc-project.org) represented on 3 is an ontology-based framework aimed at interconnecting online community sites and internet-based discussions. The idea is to enable cross-platform interoperability so that conversation spanning over multiple online media (e.g. blogs, forums, mailing lists...) can be unified into one open format. The interchange format expresses the information contained both explicitly and implicitly in internet discussion methods, in a machine-readable manner. A similar approach is proposed by the OPSN (Open Portable Social Network, http://www.opsn.net/) initiative which also covers notification and synchronization of contacts across platforms. However there is no existing implementation,and privacy control for personal published information seems not to have been addressed yet. Online communities allow web users to express their thoughts, gain feedback and critique and interact with individuals who share a similar interest. Modern web users all have some kind of participation in this realm: forums, chat rooms, newsgroups and Social Networking Sites (SNS). Each community can be considered a walled garden, without link to others. The SIOC project focuses on ways to integrate and merge these gardens, providing bridges between the knoweldge that exists there.

The SIOC ontology uses existing specifications such as FOAF and RSS to define classes and their properties. The ontology represents knowledge existing in online communities with six main classes: Site, forum, post, role, usergroup and user. Thus, a flexible range is offered for capturing knowledge in online communities. The concept sioc: User can be thought of as the central point for the ontology, sioc: User is a subclass of the foaf: OnlineAccount class. The SIOC ontology also defines properties that relate each of the classes to the user in a similar manner to how a user interacts within an online community. SIOC allows the capture of interactions between individuals, and is capable of expressing the role an individual plants in an online social space.

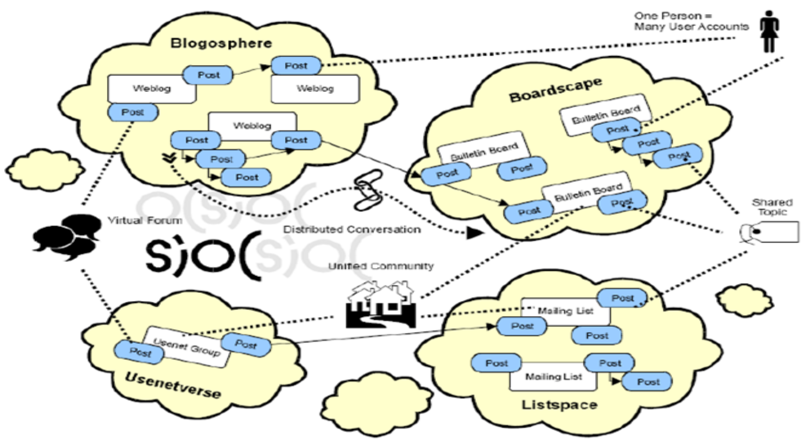

Fig. 3. Semantically Interlinked Online Communities

The Social Semantic Cloud of Tags (SCOT) (Figure 4) ontology aims to describe the structure and the semantics of tagging data and to offer social interoperability of the data among heterogeneous sources [20]. Both Tagcloud and Tag class in SCOT play a role to be able to represent social and semantic context of tagging, since both classes include users, tags, and resources and additional information to clarify tags' semantics. scot:TagCloud has properties that describe a certain user, tag spaces, number of tags, posts and co-occurrences and their frequencies, as well as updated information. The property scot:contains links scot:TagCloud to a set of scot:Tag instances. scot:Tag, as a subclass of tags:Tag from the Tag Ontology, describes a tag that is aggregated from individual tagging activities. SCOT allows the exchange of semantic tag metadata for reuse in social applications and enables interoperation amongst data sources, services, or agents in a tag space.

MOAT's -Meaning Of A Tag [21] (Figure 5)-goal is to provide a Semantic Web model to define the meaning of tags in a machine-readable way [9]]. To achieve it, MOAT defines:

– the global meanings of a tag, i.e. the list of all meanings

– the local meaning of a tag, i.e. the meaning of a tag in a particular tagging action.

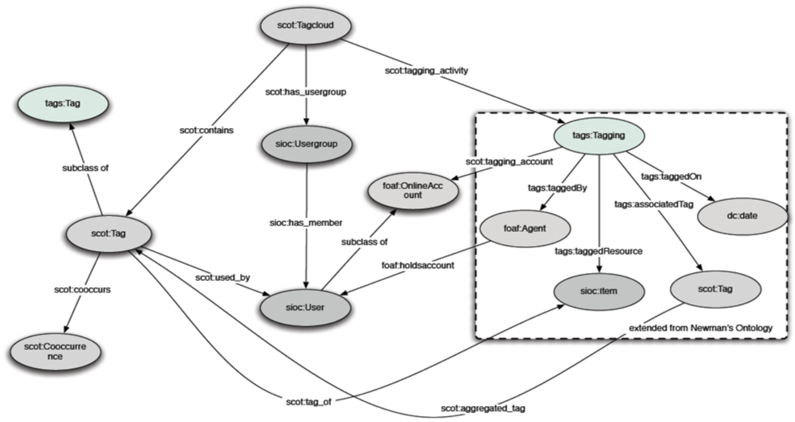

Fig. 4. Semantically Interlinked Online Communities

Indeed, for instance, the tag "Paris" can mean -depending on the user, context and other factors -a city in France, a city in the USA, or even a person. Yet when someone uses it in a tagging action, it has a particular meaning, for example the French capital. Thus, MOAT extends the usual tripartite model of tagging action to the following quadripartite model Tagging (User, Resource, Tag, Meaning). Moreover, MOAT introduces a social aspect that lets people share their tags -and their meanings -within a community by subscribing to a MOAT server, as they could do with the Annotea annotation server. They subscribe to a tag server in which they can share and update tag meanings, and use it when tagging content. When a user tag content, the client queries the server to retrieve tag meanings and let the user choose which one is the most relevant one, regarding the context.

4 Application Examples of the presented tag representation schemas

LODR (A Linking Open Data Tagging System) [22] is a prototype system based on the previously described MOAT vocabulary that allows to semantically enhance tags. Tagging is widely deployed, however the most important issue, that is the meaning of the tags, still needs to be resolved. The LODR system proposes to solve this issue by letting people give meaning to their tags using URIs of Semantic Web resourves from the Open Data initiative [23]. This means modeling facts as: "When I tag this picture 'apple', I mean http://dbpedia.org/resource/Apple Records, i.e. the record label, not the fruit". The system is based on a set of wrappers (currently available for 5 different services including Flickr and Slideshare), that parse the RSS feeds of user's data, extract items and related tags and translate it to RDF using SIOC and the Tagging Ontology. The data is then stored into a local triple-store and for each tagged item, the user can browse it and give meaning to its tags. To ease the process of chosing the right meaning, human-readable labels can be displayed instead of URIs. When no URI have been previously defined or when existing ones do not correspond to the meaning of the tag in the current context, a new URI can be added.

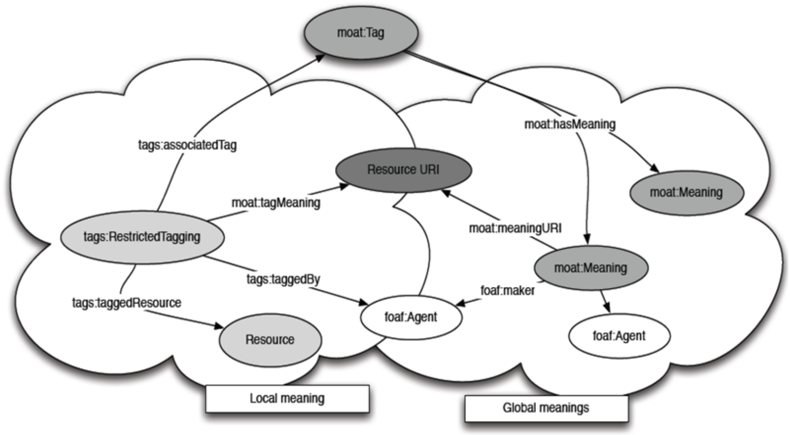

Fig. 5. Part of the MOAT ontology

Another reference work concerning modeling social tagging activity is proposed by Peter Mika in [Ontologies are us]. This model is important to study in that Mika demonstrates how tagging models and social networks come together. However the proposed model does not contain any novelties in the representation of tagging activity.

5 Comparison of the reviewed models

Our comparison will take into account several criteria. It is important to compare the modeling of the tagger in the different propositions. Several questions are to be considered: what user profile model is used to model the tagger? Does this model allow collaborative tagging, e.g. is the belonging of the user to a community considered? Another key question is the context of the tagger? What is the relationship of the tagger towards the resource? Also, we take into account the formal representation of the tags. What language is used to model association between tags, taggers and tags, tags and ressources? We also list existing implementations of the compared models and make a review of the applications. A key question that we will try to answer after this comparison is: which model could be efficient to model a linked community of individuals, objects and relationships, e.g. a social network that goes beyond the existing models, where relationships are declared explicitely. Indeed, objects can also link people: a photo, a video, a shared file etc.

Model	Namespace	Tagger Profile
Gruber	-	Tagger
Tag Ontology of Newman	tags	foaf: Agent
TagOnt	-	User
Echarte et al.	-	User
SCOT	scot	sioc: User
MOAT	moat	foat: Agent
NAO	nao	Party

Table 2. Comparison of tagging models

Model	When	Where	Format
Gruber	TagAssertTime	TagSource	OWL
Tag Ontology of Newman	taggedOn	taggedResource	OWL
TagOnt	isTaggedOn	hasTagged	OWL
Echarte et al.	hasDateTime	Source	OWL
SCOT	updated	tagspace	OWL
MOAT	taggedOn	taggedResource	OWL
NAO	created/modified	Resource	NRL

Table 3. Comparison of tagging models

Tagging is a ternary relation between a user (section 1), a resource and a tag. Several models were reviewed that intend to give a vocabulary and formmal representation of the tagging itseld with all the implications: the users, resources and tags.

In the following, we will review the AOA (Agent Oriented Abstraction) [**?**] model, which proposes a categorization of annotations/tags. AOA consideres four categories for annotations (semantic interpretation about a resource): 4 classes of knowledge annotations

- Ontology (structured semantic knowledge bases/classes)
- Communication (protocols of data exchange)
- Safety/Integrity (reachability of an annotation, resssource)
- Cognition

These annotation classes refer in a more general way to capabilities of an agent to associate semantics to a knowledge according to these annotation classes.

References

1. F. Baader et al., The Description Logic Handbook, Cambridge University Press, 2003.
2. "Resource Description Framework (RDF) / W3C Semantic Web Activity"; http://www.w3.org/RDF/.
3. "OWL Web Ontology Language Overview"; http://www.w3.org/TR/owl-features/.
4. D. Brickley and L. Miller, "FOAF Vocabulary Specification 0.91," Nov. 2007; http://xmlns.com/foaf/spec/.
5. F. Carmagnola et al., "Towards a Tag-Based User Model: How Can User Model Benefit from Tags?," 11th International Conference on User Modeling, vol. 4511/2007, 2007.
6. A. von Hessling, T. Kleemann, and A. Sinner, "Semantic User Profiles and their Applications in a Mobile Environment," Artificial Intelligence in Mobile Systems 2004.
7. V. Mendis, "RDF User Profiles - Bringing Semantic Web Capabilities to Next Generation Networks and Services," ICIN , 2007.
8. M. Golemati et al., "Creating an Ontology for the User Profile: Method and Applications," Proceedings of the First IEEE International Conference on Research Challenges in Information Science , 2007.
9. M. Tazari, M. Grimm, and M. Finke, "Modeling User Context," Proceedings of the 10th International Conference on Human-Computer Interaction, 2003.
10. J. Trajkova and S. Gauch, "Improving Ontology-based User Profiles," Proceedings of RIAO, 2004, pp. 380-389.
11. "Ontology-Based User Profiles for Search and Browsing," User Modeling and User-Adapted Interaction: The Journal of Personalization Research, Special Issue on User Modeling for Web and Hypermedia Information Retrieval, vol. 2003.
12. J. Teevan, S.T. Dumais, and E. Horvitz, "Personalizing Search via Automated Analysis of Interests and Activities," Proceedings of SIGIR 2005, 2005.
13. M. Golemati et al., "User Profile Ontology version 1"; http://oceanis.mm.di.uoa.gr/pened/?category=publications.
14. B. Kules, "User Modeling for Adaptive and Adaptable Software Systems," Adaptive User Interfaces: Principles and Practices, 2000; http://www.otal.umd.edu/UUGuide/wmk/.
15. E. Vildjiounaite et al., "Context-Dependent User Modelling for Smart Homes," 11th International Conference on User Modeling, vol. 4511/2007, 2007;
16. M. Sutterer , O. Droegehorn , et K. David, "UPOS: User Profile Ontology with Situation-Dependent Preferences Support ," Proceedings of the First International Conference on Advances in Computer-Human Interaction , pp. 230-235.
17. G. Thomas, "Ontology of folksonomy: A mashup of apples and oranges," Intl Journal on Semanic Web and Information Systems, 2007.
18. N. Richard , "Tag Ontology," http://holygoat.co.uk/owl/redwood/0.1/tags, 2005.
19. J.G. Breslin and U. Bojars: "Semantically-Interlinked Online Communities." (http://scos-project.org)
20. Hak Lae Kim, J. Breslin, S.K. Yang, H.G. Kim: Social Semantic Cloud of Tag: Semantic Model for Social Tagging. KES-AMSTA: 83-92
21. Alexandre Passant, Philippe Laublet: Combining Structure and Semantics for Ontology-Based Corporate Wikis. BIS 2008: 58-69

22. LODr – A Linking Open Data Tagging System. Alexandre Passant. In: Proceedings of the First Social Data on the Web Workshop, Karlsruhe, Germany, October 27, 2008, CEUR Workshop Proceedings, ISSN 1613-0073, online CEUR-WS/Vol-405//paper2.pdf
23. Chris Bizer, Tom Heath, Danny Ayers, and Yves Raimond. Interlinking open data on the web. In Poster, 4th Annual European Semantic Web Conference (ESWC2007), Innsbruck,Austria, 2007.
24. Peter Mika. Ontologies Are Us: A Unified Model of Social Networks and Semantics. In Proc. The Semantic Web ISWC 2005. 522–536.
25. T.T. Gruber. Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human Computer Studies. Volume 43, 2005, pages 907–928.