OntoBlog: Informal Knowledge Management by Semantic Blogging

Aman Shakya, Vilas Wuwongse, Hideaki Takeda and Ikki Ohmukai

Abstract—**Blogs collect abundant information by providing easy and dynamic publishing interface. Semantic blogging attempts to provide structure to blog entries by enriching them with metadata. However, currently semantic blogging does not link up related items. So it is still difficult to navigate, retrieve and organize related blog entries. OntoBlog is a prototype semantic blogging system which attempts to facilitate informal knowledge management in blogs by interlinking related blog entries. It combines blogging and knowledge base technologies. This is done by semi-automatic semantic annotation of blog entries with ontology instances. Blog entries are automatically linked to related instances using language processing. The rich structure of ontology can enable useful semantic capabilities for knowledge management. Semantic navigation allows users to navigate through each blog entry to semantically related blog entries. Semantic search can be employed in blogs. Semantic aggregation collects blog entries relevant to the topic of interest and organizes them meaningfully. A prototype for computer department domain ontology has been implemented.**

Index Terms— **Blogging, informal knowledge management, ontology, semantic annotation**

I. Introduction

Blogs have made publishing information on the Web very easy. Blogging facilitates rapid publishing and results in abundant collection of information. Blogs harness the collective knowledge of large number of users. They serve as dynamic media showing the latest posted information at the top. However, apart from the reverse chronological order blog entries do not have much structure and organization. Hence, filtering, organizing and navigating blog entries are difficult in traditional blogging.

Semantic blogging [4], [5] provides well defined structure to blog entries in the form of metadata. It combines the desirable features of both blogging and the Semantic Web. Currently semantic blogging provides semantic structure to individual entries. However, these pieces of information are not interlinked with semantic relations. Thus, it is still difficult to navigate through related entries and organize contents meaningfully. Semantic blogging should not only be about publishing structured data but also linking related contents together based on semantics. To manage the abundant knowledge collected through blogging it is important to interlink related pieces of information and organize them.

In this paper we propose OntoBlog [1], a new semantic blogging prototype for informal knowledge management of blog contents. It links blog entries to an existing ontology and instances using semi-automatic semantic annotation. Semantic annotation helps in retrieving relevant resources; categorizing and organizing contents and navigating meaningfully [11], [13], [20]. OntoBlog attempts to provide an integrated platform to facilitate publication, semantic annotation and information utilization in blogs.

In Section II we discuss about knowledge management in blogs. A motivating example scenario is described in Section III. In Section IV we discuss about the implementation of the system using semantic annotation in blogs. We propose OntoBlog as an integrated environment. Some semantic services offered by the system are discussed in Section V. Section VI describes some related works. Finally, we conclude and point out some future works in Section VII.

II. Knowledge management in blogs

The Wikipedia definition says that "Knowledge Management (KM) comprises a range of practices used by organizations to identify, create, represent, and distribute knowledge for reuse, awareness and learning". In fact, KM is a vast area including many aspects. We only consider creating and sharing pieces of information informally and annotating them with semantic representations so that they can be organized and retrieved effectively. Though we only consider limited aspects of KM, the idea would be applicable and useful to many domains and organizations in general.

Blogging provides an easy platform for creating and sharing information snippets. Blogs can effectively capture informal knowledge from several users and cater to the entire community. Conventional database driven information systems are rigid and do not cover all types of information that people may want to share within an organization or community. Informal snippets in blogs can cover a wide variety of information. Cayzer [4], [5] discusses elaborately why blogs

Manuscript received November 15, 2007.

A. Shakya, H. Takeda and I. Ohmukai are with the National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan (e-mail: shakya_aman@nii.ac.jp, takeda@nii.ac.jp, i2k@nii.ac.jp).

V. Wuwongse is with the Asian Institute of Technology, Klong Luang, Pathumthani 12120, Thailand (e-mail: vw@cs.ait.ac.th).

[1] An online demo can be found at http://dutar.ex.nii.ac.jp/ontoblog/blog/default/

are suitable for managing information snippets.

However, blog entries are not semantically interlinked to each other making it difficult to navigate, organize and retrieve contents. We can try to address this issue by complementing the unstructured blog contents with well structured ontology and instances, i.e. a knowledge base. The semantic relations in the ontology can be used to interlink related blog entries.

Knowledge bases can be maintained well by organizations using available tools and methodologies. However, direct use of such knowledge bases is difficult for non-technical users. So a lot of high quality information remains locked in such knowledge bases. Coupling knowledge base with informal techniques like blogging can expose such valuable data for useful applications. Thus, these technologies can complement each other and transform blogs into informal knowledge management platforms. This can be done by semantic annotation of blog entries as detailed in Section IV.

III. EXAMPLE SCENARIO

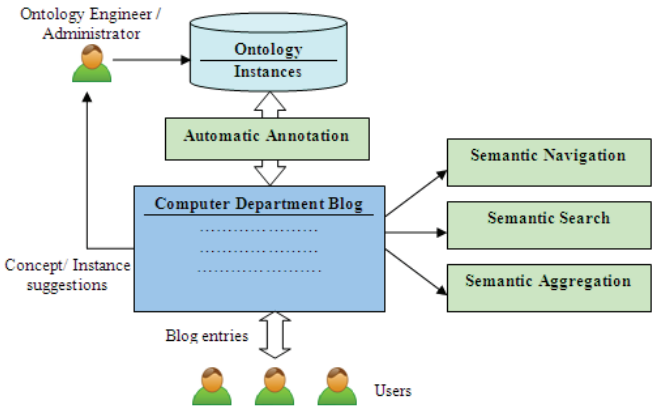

Fig. 1. Example scenario.

As an example domain, we consider the case of a computer department of a university. The department maintains an ontology with concepts like course, topic, teacher, research, etc. The ontology is maintained by the ontology engineer or administrator using available ontology management software. The knowledge base is populated with instances. The department also maintains a community blog as illustrated in Fig. 1. Registered users can easily publish entries on the blog. When publishing or updating a blog entry, the system automatically suggests instances related to the blog entry. The user may accept the suggestions or modify some choices as appropriate. If a related instance or concept is not shown by the system, the user may type in appropriate suggestion for a new instance and/or concept. The list of suggestions posted by several users can be viewed by the administrator and he/she can make appropriate additions or improvements to the ontology. The users can access the blog entries effectively with the help of semantic capabilities provided. When a blog entry is viewed, semantic navigation links are shown as related links. Search results are augmented by semantic search. Further, if the user is interested in some topic, semantic aggregation can gather relevant blog entries and organize them presenting a graphical table of contents.

IV. IMPLEMENTATION

To implement the scenario described above we propose linking blog entries to ontology and instances, as illustrated in Fig. 2. Blog entries are unstructured and scattered without explicit links among them. On the other hand, an ontology is very well structured with semantic links. Blog entries can be linked to ontology using Semantic Annotation, i.e., annotation that references an ontology. Semantic annotation can enhance information retrieval and improve interoperability [22]. The deep structure of the ontology can be leveraged to link up related blog entries and organize them for informal knowledge management.

A. Semantic Annotation in Blogs

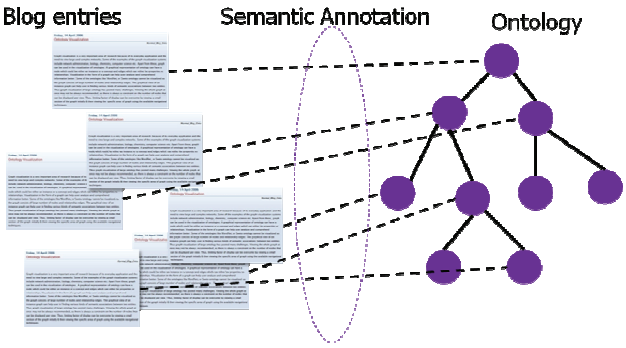

Fig. 2. Linking blogs and ontology by semantic annotation.

Blogs are somewhat different from other web information sources. Blog entries are self-contained snippets [4] of information or small contents [19]. We may consider a blog entry as a single discrete unit of information. Thus, annotations can be applied to the blog entry as a whole. Blogs facilitate easy and dynamic publishing. Therefore, there are some considerations for the application of semantic annotation to blogs.

Integrated authoring environment. Annotation would also become easy if we can integrate it with the dynamic publishing platform of blogs. Such an integrated environment has been pointed out as a requirement for semantic annotation systems by Uren et al. [22].

Automation. With manual annotation, the user has the burden of creating annotations. Providing useful annotation may also depend on the willingness of stakeholders [7]. Automatic or semi-automatic annotation with pre-existing information can help in solving these issues and makes the process of annotation fast and easy for the blogger.

Integrated services. As blog entries are scattered in the form of small discrete entries, it becomes essential to provide some services to relate these pieces of information together and present them to the user as an organized collection.

B. The OntoBlog Integrated Platform

We propose OntoBlog as an integrated authoring platform introducing the following features into blogging.

Semantic annotation. We use existing ontologies and instances to semantically annotate blog entries.

Integrated authoring. Semantic annotation is integrated with the authoring environment of the blog which helps the author to annotate entries easily at the time of blogging.

Semi-automatic annotation. The system automatically discovers related instances when blog entries are added and provides suggestions to the author.

Feedback for ontology maintenance. The users may suggest new concepts and instances if the system does not contain appropriate ones related to the blog entry.

Integrated services. The system demonstrates how semantic capabilities may be incorporated to utilize blog contents, by some example services.

A prototype implementation, similar to the example scenario, has been done considering the domain of the Computer Science and Information Management department of the Asian Institute of Technology, Thailand.

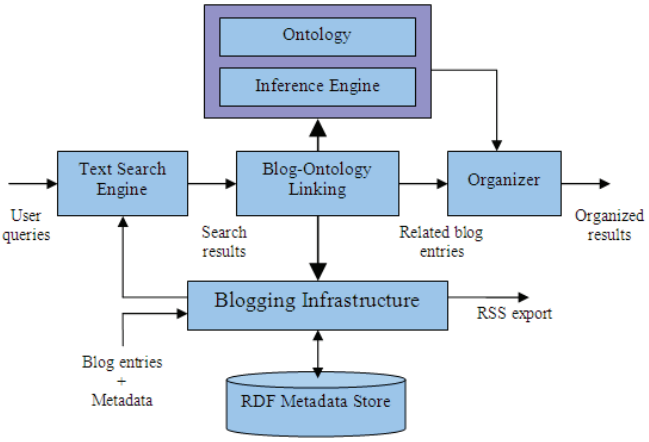

Fig. 3. System architecture.

The system architecture is shown in Fig. 3. The system is built upon a blogging infrastructure backed up by an RDF metadata store. The Blojsom blogging platform (http://blojsom.sourceforge.net/) has been used. Users post blog entries, optionally containing some metadata. The system has a text search engine to index blog entries. Blog entries are linked to related ontology instances by the blog-ontology linking component. The component automatically suggests related instances for blog entries and saves the selected annotations in the metadata store. The inference engine can deduce implicit relations between instances. Search results are augmented with related blog entries by finding linked related instances from the ontology. Related blog entries are finally organized into a navigable collection by the organizer based on the structure of the ontology. The system also exports metadata in extended RSS format.

C. Publishing Metadata

Besides normal text contents, blog entries may also contain metadata as shown in Fig. 4. In our test installation we used some publication types from the SWRC ontology (http://ontoware.org/projects/swrc/). Metadata is stored in the RDF metadata store as described in another work by us [21]. This provides a scalable storage unlike using a single RDF file for all metadata (as in [5], [4]). Jena has been used to manage operations on the RDF metadata store. Blog entries and metadata can be entered by several users thus harnessing the collective. Further, publishing metadata in RSS feeds makes way for aggregation of information from multiple blogs.

Fig. 4. OntoBlog interface.

D. Ontology and Inference

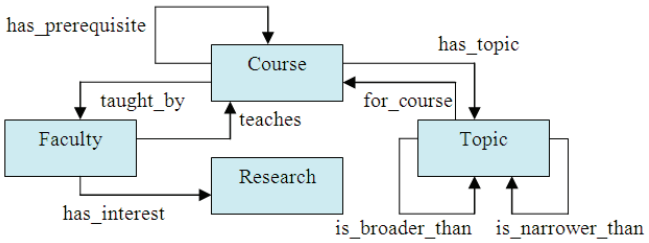

Fig. 5. A part of a computer department ontology.

The ontology may contain various concepts from the domain and a wide variety of relations, not just a topic hierarchy. A simple example ontology of a computer science department was constructed for testing the system. It was created using the SHOE Computer Department Ontology (http://www.cs.umd.edu/projects/plus/SHOE/onts/cs1.1.html). However, only few parts of the ontology have been used including concepts and relations depicted in Fig. 5.

Instances of the ontology are populated in the knowledge base. We can expect the availability of suitable knowledge bases [13], [20]. A knowledge base can easily be shared by multiple distributed blogs. Protégé has been used for managing the knowledge base.

The OWL Micro reasoner from Jena has been used for inference. We may have some axioms for inference like - "*for_course* and *has_topic* are inverse relations", "*is_broader_than* and *is_narrower_than* are inverse relations", "*teaches* and *taught_by* are inverse relations", "*has_prerequisite* and *is_broader_than* are transitive", etc.

E. Blog-Ontology Linking

The system links blogging to an existing ontology system by semi-automatic semantic annotation. Annotation can be automated by language processing of the blog entries. Language processing may not be as sophisticated as information extraction and wrapper mechanisms. However, simple lexical analysis can be very fast and quite effective [7]. The Porter stemming algorithm has been used (as in [5], [4]) which is a widely used technique in information retrieval. The system provides automatic suggestions to the user for annotation. The user can easily modify the options if some suggestions are not appropriate. Automatic annotation cannot be perfect even with other available sophisticated techniques. Moreover, relevance is a subjective matter and not possible for perfect automatic judgment. Providing suggestions to the user keeps the system flexible instead of making it totally automatic and rigid.

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:semblog="http://www.ait.ac.th/semanticblog#" >
  <rdf:Description rdf:about=
    "http://localhost:8080/semanticblog/blog/default/?
    permalink=RMI.html">
    <semblog:related-to rdf:nodeID="A0"/>
  </rdf:Description>
  <rdf:Description rdf:nodeID="A0">
    <rdf:type
      rdf:resource="http://www.w3.org/1999/02/22-rdf-
      syntax-ns#Bag"/>
    <rdf:_1>
      http://www.ait.ac.th/computerscience.owl#
      Java_Technologies
    </rdf:_1>
    <rdf:_2>
      http://www.ait.ac.th/computerscience.owl#
      Remote_Method_Invocation
    </rdf:_2>
  </rdf:Description>
</rdf:RDF>
```

Fig. 6. Example Blog-ontology linking.

An additional "keywords" element has been added to each concept in the ontology. For each instance, the "keywords" element contains a collection of words related to that instance. Whenever an entry is added or updated, the "keywords" for each instance are stemmed and matched against the stemmed blog entry. If any of the "keywords" is found in the stemmed text, the blog entry is considered as related to that instance. The discovered relations are encoded in the related-to property, as shown in Fig. 6. In the example, the blog entry with permalink "*RMI.html*" has been linked to "*Java_Technologies*" and "*Remote_Method_Invocation*".

F. Feedback for Ontology Maintenance

In case some related instances or concepts are not defined in the ontology, users may suggest a suitable instance or concept. The system provides a web-based form for new suggestions along with the automatic suggestions described in the previous section. The user may enter the name for the new instance, select the appropriate concept (or suggest a new concept) and post some remarks. The list of suggestions posted by various users can be accessed by the administrator on the blogging system itself. The feedback thus collected is useful for the administrator to maintain the ontology up-to-date by adding missing concepts and instances.

V. SEMANTIC SERVICES FOR KNOWLEDGE MANAGEMENT

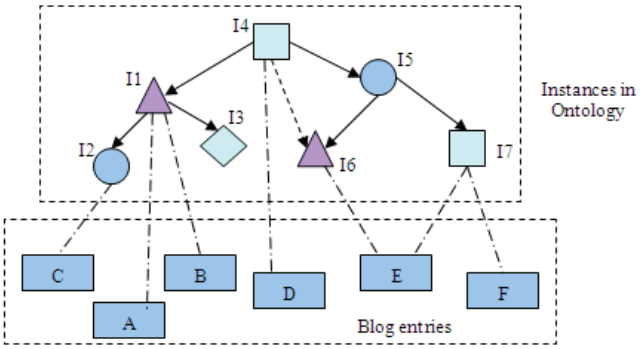

Fig. 7. Example Blog-ontology linking.

Semantic annotation of blog entries allows us to relate different blog entries using the structure of the ontology as illustrated in Fig. 7. In the figure, instances (I1-I7) in the ontology are represented by different shapes, each shape representing a concept. Instances are connected to each other by different relations (indicated by the solid arrows). Linking blog entries to ontology serves to link related blog entries implicitly. Blog entries (A to F) are annotated by the ontology by linking them to the instances, as shown by the dash-dotted lines. Blog entries 'A' and 'B' are related to each other because they are both mapped to the same instance 'I1'. Instance 'I1' is related to 'I2'. Hence, blog entry 'A' is indirectly related to 'C', which has been mapped to 'I2'. Instances may also be linked by implicit relations (shown by dashed arrow) that can be discovered by inference. Instance 'I4' is related to 'I6' by an inferred link. Thus, blog entry 'D' (mapped to 'I4') is related to 'E' (mapped to 'I6').

A. Semantic navigation

Semantic navigation helps the user to browse through related blog entries. For example, suppose we view a blog entry B about "Database Programming". The blog entry may be connected to {"computer programming", "databases", "software development", "Prof. Takeda"....}. "Computer programming" may be involved in the relations {"is taught by", "has prerequisite",..... }. Thus, there may be links like

[computer programming]
– is taught by – [Prof. Takeda]
– has prerequisite – [databases],etc.

Clicking on [databases] will lead to the blog entries related to databases. When a blog entry is opened, the semantic

navigation links are shown in a collapsible “Related to” block (shown in Fig. 8).

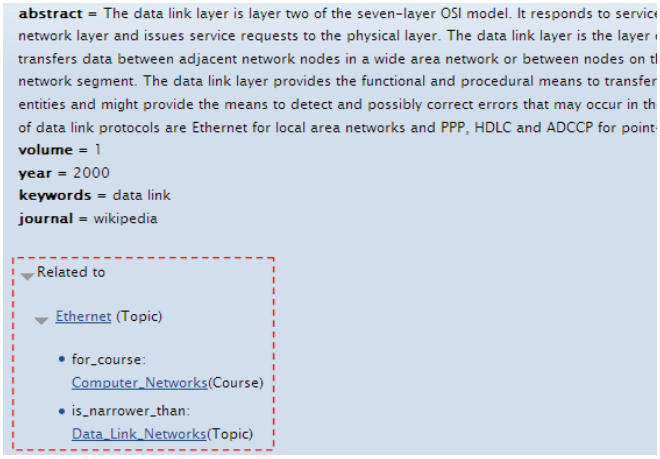

Fig. 8. Semantic navigation.

B. *Search*

The system provides indexed text search and metadata search. These are further augmented by semantic search. Guha et al. [9] have presented extensive research on semantic search along with sophisticated implementation. We do not intend to reproduce or extend the research on semantic search. Rather OntoBlog just provides a simple demonstration of its applicability. Semantic search may be particularly useful when text search alone does not produce enough results. Semantic search can be enabled or disabled in configuration settings. Further, the depth of semantic search can also be controlled.

C. *Semantic Aggregation*

Semantic aggregation can be introduced in the system to collect and organize search results relevant to a topic of interest. A simple algorithm is outlined below.

1. Get all relevant blog entries from search.
2. Find the set of instances S from the Ontology linked to each blog entry.
3. Find all relations between the instances in S
4. Visualize the related instances as directed graphs

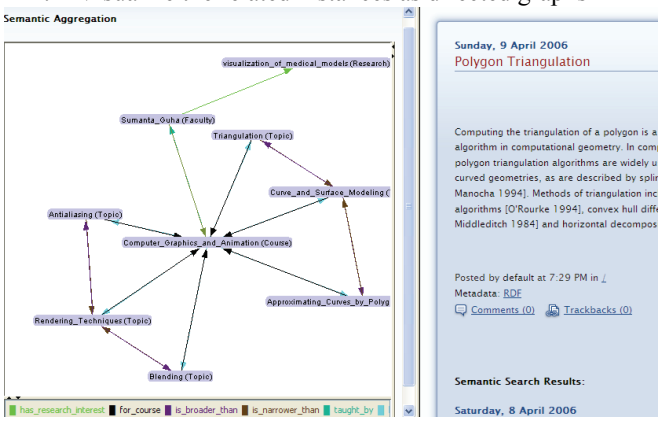

Fig. 9. Semantic aggregation.

GraphML has been used to represent the semantic aggregation graphs. Semantic aggregation is depicted in Fig. 9. The user runs semantic aggregation by searching on a topic of interest. The search results are listed on the right-hand side frame. Related instances from the ontology are aggregated and visualized on the left-hand side frame. The relation type between instances is identified by the color of the link and shown in an index. When a node is clicked, blog entries related to that node are displayed on the right-hand side.

VI. RELATED WORK

A number of works have been done in semantic blogging. The Semantic Blogging Demonstrator uses a category tree based on ‘broader than/narrower than’ relations [5], [4] to categorize blog entries. OntoBlog uses an ontology with wider variety of concepts and relations, rather than a simple taxonomy, and enables inferences. It also allows utilizing rich knowledge base of instances maintained using existing ontology management environments. The demonstrator emphasizes semantic view, navigation and query. However, tree browsing provided for semantic navigation is already a feature of many blogs [3]. Facet browsing is more like metadata search. OntoBlog offers a more intuitive way of semantic navigation by providing related links through each blog entry.

Karger and Quan [12] extended Haystack to enable users to view cross-blog reply graphs and track conversation in multiple blogs. Semblog [19] annotates content using FOAF metadata. SocioBiblog [21] facilitates sharing of bibliographic information in a social network based on extended RSS. Moller et al. [16] identify structural and content-related metadata in blogging - SIOC ontology [2] for structural metadata and FOAF, vCard, BibTex/SWRC, iCalendar, etc. for content metadata. In our case, the commenting mechanism generates some structural metadata. Though SWRC has been used as an example, any content metadata can be used. In addition, the semantic annotation metadata provides categorization metadata. semiBlog [16], [17], [18] utilizes data on the user’s desktop. But we still have to search metadata in desktop. More abundant metadata, of better quality, may be available in existing knowledge bases.

Uren et al. [22] present a detailed survey of semantic annotation tools and analyze them on the basis of a number of requirements. A large body of research on semi-automatic semantic annotation already exists including works like S-CREAM [10] and extraction ontologies [6]. Our attempt is to demonstrate the application of semantic annotation in blogs, not to build a sophisticated annotation system. However, OntoBlog satisfies some requirements [22] like automation, integrated environment, document-annotation consistency and separate annotation storage. Magpie [7], [8] automatically links instances identified in web documents to relevant ontological instance/class. It uses simple lexicon-based parsing and linguistic rules to identify instances. Magpie depends on external agents for semantic services. KIM [13], [20] uses an ontology a pre-populated with instances. It introduces

information retrieval based on IE techniques for the recognition of named entities.

VII. Conclusion and Future Work

In this paper attempted to show the possibility of informal knowledge management with blogs for utilizing the abundant information that can be gathered from the community. We proposed OntoBlog which demonstrates a new method for semantic blogging – integrating blogs with an existing knowledge base through semantic annotation. The semantic relations in the ontology can be exploited to turn blogs into a web of interrelated contents. We can effectively retrieve and organize contents when pieces of information are interlinked to each other through semantic relations.

We can explore mechanisms for the decentralized creation of the ontology in future. Collaborative techniques like folksonomy or semantic wiki could be utilized. The system is currently a community blog. Features like semantic navigation and search are yet to be employed across multiple blogs. The system can be made more powerful by introducing different types of inferences. Mature semantic search systems can be incorporated. Information retrieval mechanisms can be used to produce ranked search results. Language processing with stemming is quite basic. It can be made more accurate by handling lexical, semantical and syntactical variations [1]. Lexical variations can be dealt with technologies like Word-Net. Sophisticated automatic annotation using IE techniques can be incorporated to make the automation more robust.

References

[1] A. Arampatzis, T. van der Weide, C. Koster, and P. van Bommel, " Linguistically motivated information retrieval," in *Encyclopedia of Library and Information Science*, vol. 69, A. Kent, Ed. Marcel Dekker, Inc., New York, Basel, December 2000.

[2] J. G. Breslin, A. Harth, U. Bojars, and S. Decker, "Towards semantically-interlinked online communities," in *The Semantic Web: Research and Applications*, vol. 3532 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2005, pp. 500–514.

[3] S. Cayzer, "SWAD-Europe deliverable 12.1.4: Semantic blogging - lessons learnt". http://www.w3.org/2001/sw/Europe/reports/demo 1 report/

[4] S. Cayzer, "Semantic blogging and decentralized knowledge management," *Communications of the ACM*, vol. 47(12), December 2004, pp. 48–52.

[5] S. Cayzer, "Semantic blogging: Spreading the Semantic Web meme," in *Proceedings of XML Europe 2004*, Amsterdam, Netherlands, April 2004, pp. 18–21.

[6] Y. Ding, D. W. Embley, and S. W. Liddle, "Automatic creation and simplified querying of Semantic Web content: An approach based on information-extraction ontologies," in *Proceedings of the First Asian Semantic Web Conference (ASWC 2006)*, vol. 4185 of Lecture Notes in Computer Science, Beijing, China, 2006. Springer, pp. 400–414.

[7] M. Dzbor, J. Domingue, and E. Motta, "Magpie - towards a semantic web browser. In *The Semantic Web - ISWC 2003*, vol. 2870 of Lecture Notes in Computer Science, Florida, USA, 2003. Springer-Verlag Berlin Heidelberg, pp. 690–705.

[8] M. Dzbor, E. Motta, and J. Domingue, "Opening up magpie via semantic services," in *The Semantic Web - ISWC 2004*, vol. 3298 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2004, pp. 635–649.

[9] R. Guha, R. McCool, and E. Miller, "Semantic search," in *Proceedings of the twelfth international conference on World Wide Web*, Budapest, Hungary, 2003. ACM Press New York, USA, pp. 700–709.

[10] S. Handschuh, S. Staab, and F. Ciravegna, "S-CREAM - semi-automatic creation of metadata," in *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, Madrid, Spain, October 2002.

[11] J. Kahan, M. R. Koivunen, E. Prud'Hommeaux, and R. R. Swick, "Annotea: an open RDF infrastructure for shared web annotations," in *Proceedings of the 10th International World Wide Web Conference*, Hong Kong, May 2001, pp. 623–632.

[12] D. R. Karger and D. Quan, "What would it mean to blog on the semantic web?", *Journal of Web Semantics*, vol. 3(2), 2005, pp. 147–157.

[13] A. Kiryakov, B. Popov, D. Ognyanoff, D. Manov, A. Kirilov, and M. Goranov, "Semantic annotation, indexing, and retrieval," in *2nd International Semantic Web Conference (ISWC 2003)*, vol. 2870 of Lecture Notes in Computer Science, Florida, USA, 2003. Springer-Verlag Berlin Heidelberg, pp. 484–499.

[14] M. R. Koivunen, "Annotea and Semantic Web supported collaboration," in *ESWC 2005, UserSWeb workshop*, 2005.

[15] L. McDowell, O. Etzioni, S. D. Gribble, A. Halevy, H. Levy, W. Pentney, D. Verma, and S. Vlasseva, "Mangrove: Enticing ordinary people onto the Semantic Web via instant gratification," in *The Semantic Web - ISWC 2003*, vol. 2870, Florida, USA, 2003. Springer Berlin / Heidelberg, pp. 754–770.

[16] K. Moller, U. Bojars, and J. G. Breslin, "Using semantics to enhance the blogging experience," in *The Semantic Web: Research and Applications*, vol. 4011 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2006, pp. 679–696.

[17] K. Moller, J. G. Breslin, and S. Decker, "semiBlog - semantic publishing of desktop data," in *14th Conference on Information Systems Development (ISD2005)*, Karlstad, Sweden, August 2005.

[18] K. Moller and S. Decker, " Harvesting desktop data for semantic blogging," in *Proceedings of the 1st Workshop on The Semantic Desktop - Next Generation Personal Information Management and Collaboration Infrastructure* at ISWC2005, Galway, Ireland, 2005.

[19] I. Ohmukai and H. Takeda, "Semblog: Personal knowledge publishing suite," in *Proceedings of WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, New York, USA, 2004.

[20] B. Popov, A. Kiryakov, A. Kirilov, D. Manov, D. Ognyanoff, and M. Goranov, "KIM-Semantic Annotation Platform", in *2nd International Semantic Web Conference (ISWC2003)*, vol. 2870 of Lecture Notes in Computer Science. Springer, 2003, pp. 834–849.

[21] A. Shakya, H. Takeda, I. Ohmukai, and V. Wuwongse, "SocioBiblog: A decentralized platform for sharing bibliographic information," in *Proceedings of the IADIS International Conference WWW/Internet 2007*, vol. 1, P. Isaías, M. B. Nunes, J. Barroso, Eds. Vila Real, Portugal, October 2007. IADIS Press, pp. 371-380.

[22] V. Uren, P. Cimiano, J. Iria, S. Handschuh, M. Vargas-Vera, E. Motta, and F. Ciravegna, "Semantic annotation for knowledge management: Requirements and a survey of the state of the art," *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 4(1), 2006, pp. 14–28.