OntoBlog: Linking Ontology and Blogs

Aman Shakya
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku
Tokyo 101-8430, Japan
shakya_aman@nii.ac.jp

Vilas Wuwongse
Asian Institute of Technology
Klong Luang, Pathumthani 12120
Thailand
vw@cs.ait.ac.th

Hideaki Takeda
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku
Tokyo 101-8430, Japan
takeda@nii.ac.jp

ABSTRACT

Semantic blogging attempts to enhance traditional blogging by using Semantic Web technologies. Blog entries are semantically enriched by metadata. However, authoring metadata is not easy for normal users. Currently semantic blogging only offers limited semantic capabilities. It is still difficult to navigate through semantically related entries, search and organize relevant blog entries. OntoBlog attempts to solve these issues by linking blogs to existing ontology maintained using available ontology management environment. OntoBlog is a prototype semantic blogging system which employs semi-automatic semantic annotation of blog entries using ontology instances. Blog entries are automatically mapped to related instances using language processing techniques. The rich structure of ontology with different semantic relations, enhanced by inference, can enable useful semantic capabilities. 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.

Categories and Subject Descriptors

H.4.3 [**Information Systems Applications**]: Communications applications—*information browsers*

General Terms

Design

Keywords

Blogging, ontology, Semantic Web, semantic annotation

1. INTRODUCTION

Blogging has become very popular on the Web and is growing rapidly [21]. Blogs make publishing information on the Web very easy for any user. But filtering, organizing and navigating blog entries is difficult in traditional blogging. Semantic blogging attempts to address these issues by providing a well defined structure to blog entries in the form of metadata. It combines desirable features of both blogging and the Semantic Web and extends blogging to informal knowledge management [5, 4]. Semantic blogging may also be viewed as annotation to blog entries. We may annotate blog entries with metadata based on some standard vocabulary. However, manual entry of metadata is cumbersome to the users and error-prone. Moreover, semantic blogging should not only be about publishing metadata on the Web but also linking related blog entries together based on semantics.

In this paper we propose OntoBlog, a new semantic blogging prototype which links blog entries to an existing ontology and instances. It allows semi-automatic annotation of blog entries with instances of the ontology. Any domain ontology structured with various semantic relations may be used and further enhanced by inference. Semantic annotation can help us to retrieve relevant resources; categorize and organize contents and navigate meaningfully [11, 13, 20]. OntoBlog attempts to provide an integrated platform to facilitate publication, semantic annotation and information utilization. It demonstrates possibilities for semantic capabilities in blogs by the use of an existing ontology. On the other hand, blogging can be used to obtain feedback from the users for the improvement and maintenance of the ontology.

In Section 2 we discuss about annotation and semantic annotation in blogs. In Section 3 we propose OntoBlog[1] as an integrated environment. A motivating scenario is described in Section 4. Section 5 describes implementation of the system. Some services offered by the system are discussed in Section 6. In Section 7, we discuss some initial experiences testing the system. Section 8 describes some related works. Finally, we conclude and point out future works in Section 9.

2. ANNOTATION

Annotations[2] are comments, notes, explanations, or remarks attached to any document or a selected part of the document. Annotation metadata can be used not only for describing content, but also to organize and classify it [11].

[1]An online demo can be found at http://dutar.ex.nii.ac.jp/ontoblog/blog/default/
[2]http://www.w3.org/2001/Annotea/

Annotation can also help in information retrieval process [13, 20] and organizing search results [14].

2.1 Automatic Semantic Annotation

Annotation that references an ontology has been termed *semantic annotation*. Semantic annotation can enhance information retrieval and improve interoperability [23]. With manual annotation, the user has the burden of creating annotations. Providing useful annotation may also depend on the willingness of stakeholders [7]. Issues may also arise about the authenticity and quality of manual annotations. Automatic or semi-automatic annotation with pre-existing information can help in solving these issues. Automatic annotation systems may provide suggestions, but still require intervention by user; or acquire annotations automatically on a large scale. Uren et al. [23] describe different ways to provide automatic support for annotation:

- Wrappers can exploit the structure of web pages to identify pieces of information for annotation.
- IE (Information Extraction) systems using supervised / unsupervised learning
- Natural language processing

2.2 Semantic Annotation in Blogs

OntoBlog is an application of semantic annotation to blogs. 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 unit of information. A characteristic of blogs is dynamic publishing. Blogs are user-oriented and provide an easy mechanism for frequent publication. Therefore, there are some considerations for the application of semantic annotation to blogs.

Integrated authoring environment. Blogging provides an easy platform for publishing. Annotation would become easy if we can integrate it with the publishing platform. An integrated environment providing a single point of entry interface for publication and annotation has been pointed out as a requirement for semantic annotation systems by Uren et al. [23].

Automation. Automation makes the process of annotation fast and easy for the blogger. Utilizing existing ontology instances for annotation can be considered as automatic metadata creation for blog entries.

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.

3. THE ONTOBLOG INTEGRATED PLATFORM

We propose OntoBlog as an integrated authoring platform which links existing ontology to blogs. It incorporates the following features into blogging.

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

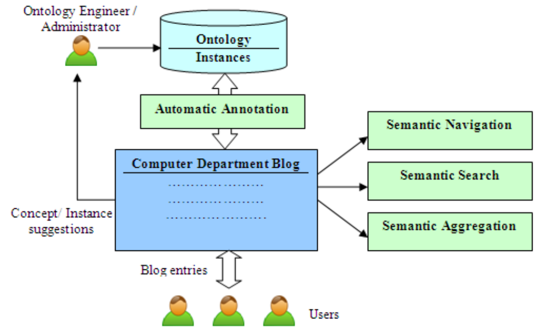

Figure 1: Example scenario

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. It can be used as a feedback channel which would be helpful to the administrator to improve and maintain the knowledge base up-to-date.

Integrated services. The system demonstrates how semantic capabilities may be incorporated to utilize blog contents properly, by some example services. Semantic navigation and search services are provided to retrieve and browse related blog entries. Organization of related blog entries is provided by exploiting semantic relations in the ontology.

4. 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 has been populated with instances of courses, topics, teachers, etc. 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 instance name as suggestion and select the proper concept, or suggest a new concept. The list of suggestions posted by users can be viewed by the administrator. He/she can evaluate the suggestions of various users and 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. Suppose our blog entry is related to databases. Then we will

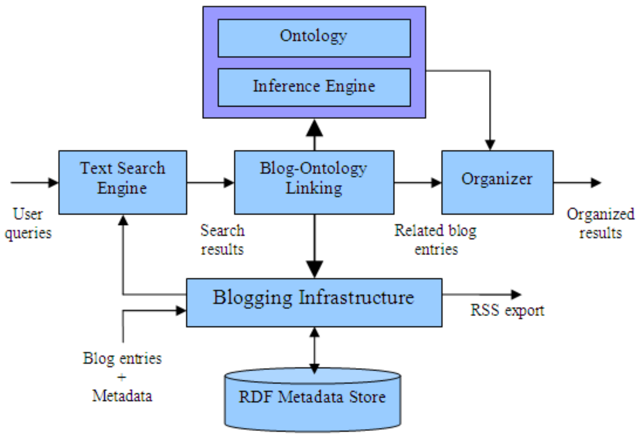

Figure 2: System architecture

Figure 3: OntoBlog interface

have a link for the 'Database' course and further links like - *taught by* 'Prof. Vilas Wuwongse' and *has prerequisite* 'Data structures'. When the user clicks on 'Data structures', all blog entries related to that course are listed. Search results are augmented by *semantic search*. Further, if the user is interested in some topic, he/she may use *semantic aggregation* to gather blog entries relevant to the topic and organize the collection. Instances are shown inter-related to each other as a directed graph. Implicit relations may also be revealed by enabling inference. The graph serves as a table of contents for the topic of interest.

5. IMPLEMENTATION

The system architecture is shown in Fig. 2. The system is built upon a blogging infrastructure backed up by an RDF metadata store. The Blojsom blogging platform has been used in the system. Blojsom[3] offers extensibility with plug-in architecture. Users post blog entries to the system. The blog entries may also contain some metadata. The system has a text search engine which indexes the blog entries and performs keyword based search for user queries. 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. All the related blog entries obtained are finally organized meaningfully into a navigable collection by the organizer based on the structure of the ontology. The system also exports blog entries in RSS format with embedded metadata.

5.1 Publishing Metadata

Besides normal text contents, semantic blog entries may also contain metadata as shown in Fig. 3. In our test installation we used some publication types from the SWRC ontology[4] for the metadata schema. We can use any other metadata provided that an appropriate schema is available. Metadata is stored in the RDF metadata store as described in our other work [22]. The RDF metadata store uses a MySQL relational database to store RDF metadata. This provides a scalable storage unlike using a single RDF file for all metadata (as in [5, 4]). The Jena framework[5] 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 intelligence in the community. Further, publishing metadata in RSS feeds makes way for aggregation of information from multiple blogs.

5.2 Ontology and Inference

The ontology contains 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. The ontology was created referring to the SHOE Computer Department Ontology[6]. However, only few parts of the ontology have been used including concepts and relations depicted in Fig. 4.

Instances of the ontology are populated in the knowledge base. We can also expect the availability of suitable knowledge bases. There are many knowledge bases in various domains about real world entities being maintained by several groups and enterprises [13, 20]. The knowledge base maintained in a server can easily be shared by multiple distributed blogs.

Various mature ontology management software packages are available which can be used to create and maintain the knowledge base. Protégé[7] has been used to create and maintain the ontology and instances.

The OWL Micro reasoner from the reasoning subsystem in Jena has been used for inference. It can be replaced easily by other reasoners if more powerful inference is desired. The example ontology uses some axioms for inference.

[3]http://blojsom.sourceforge.net/
[4]http://ontoware.org/projects/swrc/
[5]http://jena.sourceforge.net/inference/index.html
[6]http://www.cs.umd.edu/projects/plus/SHOE/onts/cs1.1.html
[7]http://protege.stanford.edu/

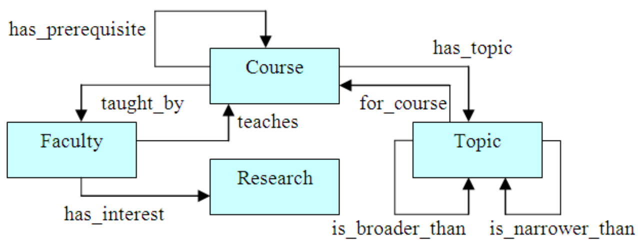

Figure 4: A part of a computer department ontology

For e.g. - "*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.

5.3 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 techniques like information extraction and wrapper mechanisms. However, simple lexical analysis can be very fast and quite effective for annotating by existing instances [7]. The Porter stemming algorithm has been used (as in [5, 4]). Stemming is a widely used technique in information retrieval and though it is a simple method, produces quite good results. 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.

A "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. Instances related to the blog entry are automatically discovered and suggested to the user. The discovered relations, after user's approval, are encoded in the *related-to* property, as shown in Fig. 5. In the example, the blog entry with permalink "RMI.html" has been linked to "Java_Technologies" and "Remote_Method_Invocation" in the ontology. Section 6 further illustrates how blog entries are linked to ontology instances to enable semantic capabilities.

5.4 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 for blog-ontology linking 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 in the ontology or refining them. The administrator uses an existing ontology management software to make updates based on the suggestions. The suggestion also shows the permalink of the concerned blog entry so that it can easily be annotated with the newly defined instance.

```
<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>
```

Figure 5: Example Blog-ontology linking

6. SEMANTIC SERVICES

Semantic annotation of blog entries allows us to relate different blog entries using the structure of the ontology as illustrated in Fig. 6. 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'. A blog entry may be related to multiple instances (like 'E' related to 'I6' and 'I7'). 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').

6.1 Semantic Navigation

Semantic navigation helps the user to browse through related blog entries. The following algorithm is used for the purpose.

1. While viewing a blog entry B, get all related instances from the Blog-Ontology relations. Let I be the set of these instances.

2. For each instance i in I

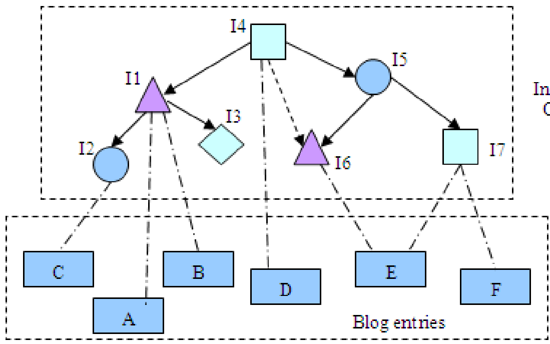

Figure 6: Example Blog-ontology linking

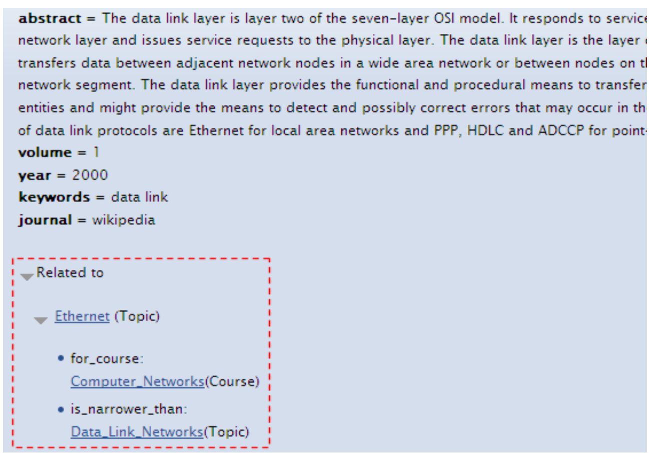

Figure 7: Semantic navigation

(a) Find all relations R involving i

(b) For each relation r in R, find instances related to i by r.

3. Render all instances with hyperlinks.

For example, suppose we view a blog entry B about "Database Programming". The blog entry may be connected to, I = "computer programming", "databases", "software development", "Prof. Takeda" .... "Computer programming" may be involved in R = "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. 7).

6.2 Search

The system provides indexed text search and metadata search. These are further augmented by semantic search. Semantic search attempts to augment and improve traditional search results using data from the Semantic Web. 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 implementation of semantic search for demonstration of its applicability. The algorithm is outlined below.

1. Find all instances in the ontology with keywords matching the search text. Let the set of instances obtained be I.

2. Set final list of instances L = I.

3. For each instance i in I

   (a) Let J be the set of instances directly related to i.

   (b) L = L union J

4. For each instance in L, get all related blog entries from Blog-Ontology relations.

5. Return the total blog entries as result.

Semantic search may be particularly useful when text search alone does not produce enough results. It may be better to have more results than just few accurate results. Semantic search can be enabled or disabled using configuration settings. Further, the depth of semantic search can also be controlled. If semantic search produces excess results in some cases, the search depth can be decreased or semantic search can be disabled altogether.

6.3 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 for semantic aggregation 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.

For example, let the search results for some search contain following blog entries.

entry A - related to - Computer programming, data structure
entry B - related to - XML, database, Prof. Takeda
entry C - related to - Java, OOPS, etc

Then, S = Computer programming, data structure, XML .... The instances in S may be related with each other by relations like prerequisite, broader/narrower, taught by, is related to, etc.

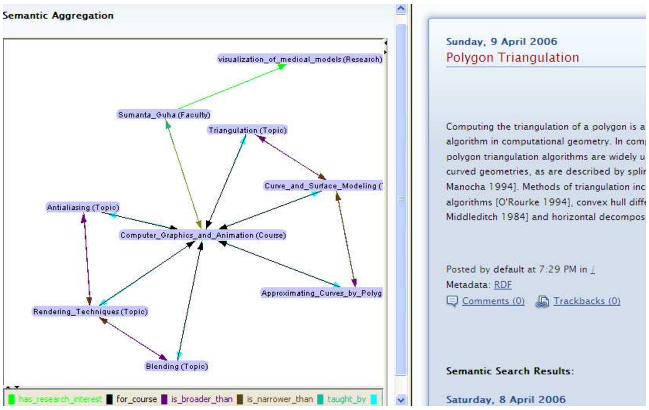

Figure 8: Semantic aggregation

Prefuse[8] has been used to visualize the graphs. GraphML[9] has been used to represent the semantic aggregation graphs. Semantic aggregation is depicted in Fig. 8. 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 frame.

6.4 Other Features

The system generates RSS feeds with embedded metadata. As demonstrated in our other work [22], RSS feeds from multiple blogs in the community can be aggregated. However, the new ontology-based features are yet to be applied on such aggregate from multiple blogs. The system also provides a bi-directional commenting mechanism, employs bookmarklet for easy publishing and generates FOAF profile for interconnecting the community as described in [22].

7. SOME EXPERIENCES

We attempted to test some features of the system by conducting simple experiments. The data for populating instances in the ontology was mostly adapted from the website of the Computer Science and Information Management department of the Asian Institute of Technology[10]. Data about 15 courses, 93 topics, 8 faculty members, 10 researches and 15 universities have been entered. About 100 dummy blog entries related to different courses, topics, researches, etc were populated manually. 10 students from the department used the system to help in experiments. Statistical treatment was not possible at this stage due to the limited number of users.

Automatic blog-ontology linking. While populating blog entries, right and wrong suggestions were noted. The relevance of the suggestions to the blog entries was judged by human subjects. The result showed about 84% right suggestions and 16% wrong suggestions. We cannot expect perfect results from simple language processing like stemming alone. "Keywords" defined to identify an instance, also affects the process. It is difficult to exhaustively list keywords related to any instance.

Table 1: Semantic search results

Search text	Text Search		Semantic Search	
	Returned	Relevant	Returned	Relevant
computer graphics	2	2	10	9
payment	1	1	7	4
java	10	8	14	9
relational database	1	1	10	8
ontology	2	2	8	4
recursion	1	1	10	6
firewall	2	2	9	7
computer networks	3	1	17	14
system analysis	3	3	10	7
compiler	0	0	13	6
informatio n system	2	2	20	16

Semantic search. Semantic search was tested with a number of search texts related to some course or topic as shown in Table 1. The table verifies that semantic search can produce additional relevant results not produced by text search alone hence increasing the recall of the system.

Semantic aggregation. Semantic aggregation was tested only to see if relevant instances are retrieved during the process. The precision was roughly found to be about 77.83%. But semantic aggregation provides much more than simply retrieving relevant instances.

User feedback. Many features of the system are subjective in nature and difficult to evaluate by experiments. Traditional information retrieval metrics like recall and precision are not suitable enough for evaluation of a system like this [13, 20]. So user feedback about the system was also collected. The questionnaires had questions about the effectiveness of semantic navigation, semantic search, inference, semantic aggregation, metadata search and RSS aggregation. The users had good response for most features. However, it was not so easy for the users to judge relevance of semantic search results because the results were not directly related to the query.

8. RELATED WORK

A number of works have been done in semantic blogging. The Semantic Blogging Demonstrator[11] uses a category tree based on 'broader than/narrower than' relations [5, 4] to cat-

[8]http://prefuse.sourceforge.net/
[9]http://graphml.graphdrawing.org/
[10]http://www.cs.ait.ac.th/ (retrieved April, 2006)
[11]http://www.semanticblogging.org/semblog/blog/default/

egorize blog entries. OntoBlog uses an ontology with wider variety of concepts and relations, rather than a simple taxonomy, and enables powerful inferences. It also allows utilizing rich knowledge base of instances maintained using existing ontology management environments. It uses OWL which is a more powerful ontology language than SKOS used by the demonstrator. OWL allows us to define several types of relations and offers powerful inferences which may be needed in future. The demonstrator mainly 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 different way of semantic navigation by providing related links through each blog entry which is more intuitive way of browsing. Search has been enhanced to explore semantic relations. Furthermore, OntoBlog organizes the search results by semantic aggregation. Works like [12], [19] and [16, 18, 17] are generic. We used computer science department domain (also used by Mangrove [15]) just as a simple example. OntoBlog can be used in any domain provided that a knowledge base is maintained. The demonstrator provides a category chooser functionality which works based on simple language processing. We also used a similar technique for automatic annotation by ontology instances.

Karger and Quan [12] extended Haystack[12] to enable users to view cross-blog reply graphs and track conversation in multiple blogs. Semblog [19] annotates content using FOAF metadata of users and exports using extended RSS. Our other work [22] facilitates sharing of bibliographic information in a social network based on extended RSS. Möller et al. [16] identify structural and content-related metadata in blogging. The SIOC ontology [2] has been used for structural metadata. FOAF, vCard, BibTex/SWRC, iCalendar, etc. have been used for content metadata. In our case, existing commenting mechanism generates some structural metadata. Though SWRC has been used as an example, any content metadata can be used. In addition, the metadata produced by automatic semantic annotation can be considered as categorization metadata. semiBlog [16, 18, 17] emphasizes generating metadata by utilizing data on the user's desktop. But the user still has to search metadata assuming that it exists in his/her desktop. More abundant metadata, of better quality, may be available in existing knowledge bases than one's desktop.

Uren et al. [23] present a detailed survey of annotation frameworks and 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 significant 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 satisfying all the requirements mentioned in [23]. OntoBlog satisfies some requirements like automation, integrated environment, document-annotation consistency and separate annotation storage. Magpie [7, 8] automatically creates a semantic layer over web documents and links instances identified in the document to relevant ontological instance/class. It uses simple lexicon-based parsing and linguistic rules to identify instances. Magpie depends on external service providers for providing semantic services. KIM [13, 20] uses an ontology with a pre-populated knowledge base of instances. IE techniques have been employed for the recognition of named entities in documents. It also introduces indexing and retrieval based on named entities.

[12] http://haystack.lcs.mit.edu

9. CONCLUSIONS AND FUTURE WORK

In this paper we proposed OntoBlog, a prototype semantic blogging system which semi-automatically annotates blog entries with instances of an ontology. It links together the well established technologies of blogging and ontology management. The system is helpful in organizing the contents in blogs through semantic capabilities and also receiving feedback from the users for ontology maintenance to some extent. The deep semantic structure of ontology provides a rich way of classifying and organizing relevant entries and can enable better navigation and search capabilities in blogs. OntoBlog provides an integrated platform for publishing and information utilization in blogs.

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 for the concerned domain. 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 inference and inference engines. Mature semantic search systems can be incorporated. Information retrieval mechanisms can be used to produce ranked search results. The system may be extended to handle complex hierarchical metadata schema. Language processing with stemming used for the demonstration prototype 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 WordNet. Sophisticated automatic annotation mechanisms using IE techniques can be incorporated to make the automation more robust. Supervised or unsupervised learning can further augment such IE techniques. It would also be good to support semantic blogging clients like semiBlog [16, 18, 17] to utilize data from the user's desktop.

10. ADDITIONAL AUTHORS

Additional authors: Ikki Ohmukai (National Institute of Informatics, email: i2k@nii.ac.jp)

11. REFERENCES

[1] A. Arampatzis, T. van der Weide, C. Koster, and P. van Bommel. Linguistically motivated information retrieval. In A. Kent, editor, *Encyclopedia of Library and Information Science*, volume 69. 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*, volume 3532 of *Lecture Notes in Computer Science*, pages 500–514. Springer Berlin / Heidelberg, 2005.

[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*, 47(12):48–52, December 2004.

[5] S. Cayzer. Semantic blogging: Spreading the semantic web meme. In *Proceedings of XML Europe 2004*, pages 18–21, Amsterdam, Netherlands, April 2004.

[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)*, volume 4185 of *Lecture Notes in Computer Science*, pages 400–414, Beijing, China, 2006. Springer.

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

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

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

[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*, pages 623–632, Hong Kong, May 2001.

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

[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)*, volume 2870 of *Lecture Notes in Computer Science*, pages 484–499, Florida, USA, 2003. Springer-Verlag Berlin Heidelberg.

[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*, volume 2870, pages 754–770, Florida, USA, 2003. Springer Berlin / Heidelberg.

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

[17] K. Möller, 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. Möller 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, November 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)*, volume 2870 of *Lecture Notes in Computer Science*, pages 834–849. Springer, 2003.

[21] L. Rainie. The state of blogging. http://www.pewinternet.org/PPF/r/144/report_display.asp, 2005.

[22] A. Shakya, H. Takeda, I. Ohmukai, and V. Wuwongse. A publication aggregation system using semantic blogging. In G. Li, Y. Liang, and M. Ronchetti, editors, *The Semantic Web - ASWC 2006 Workshops Proceedings*, pages 55–62, Beijing, China, September 2006. Jilin University Press.

[23] 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*, 4(1):14–28, 2006.