



A SURVEY ON SEMANTIC ANNOTATION TOOLS FOR KNOWLEDGE MANAGEMENT

Pooja Kherwa
Assistant Professor

Maharaja Surajmal Institute of Technology, Janakpuri, New Delhi - 110058

Abstract - Support for information and knowledge exchange is a key issue in the information society. To reduce the time wasted in searching and to reduce associated user frustration much more selective user access is needed. This is possible by semantic information processing of online documents. Knowledge management in an organisation are used for managing knowledge resources in order to facilitate access and reuse of knowledge. Semantic annotation is about assigning to the entities in the text, links to their semantic descriptions. This sort of metadata provides both class and instance information about the entities. Semantic annotation is applicable for any type of text- web pages, regular documents etc. For semantic annotation, there are various manual, semiautomatic and full automatic tools are developed by various organizations like mindswap.org, ontotext.org etc. In this paper, we are presenting analysis and review of some of these tools according to their applicability for an application domain in knowledge management. Our review and analysis may help the research community in selecting an appropriate tool for extracting the relevant and desired information from huge knowledge base of an organisation.

I. INTRODUCTION

The semantic web purposes annotating document content using semantic information. The result is web pages with machine interpretable markup to create annotation with well defined semantics.[1]

Manual Annotation: Manual annotation is more easily accomplished today, using authoring tools such as Semantic Word[2], which provide an integrated environment for simultaneously authoring and annotating text. However, the use of manual annotation is often leads with errors due to factors such as annotator familiarity with the domain, amount of training, personal motivation and complex schemas [3]. Manual annotation is also an expensive process. Another problem with manual annotation is the volume of existing documents on the Web that must be annotated to become a useful part of the Semantic Web.

Semiautomatic Annotation: Semiautomatic annotation of documents has been proposed. Semiautomatic means, as opposed to completely automatic, are required because it is

not yet possible to automatically identify and classify all entities in source documents with complete accuracy[4]. All existing semantic annotation systems rely on human intervention at some point in the annotation process [5].

Automated annotation: Automated annotation provides the scalability needed to annotate existing documents on the Web, and reduces the burden of annotating new documents. Other potential benefits are consistently applying ontologies, and using multiple ontologies to annotate a single document.

As a motivating example of what can be achieved once documents are given semantic markup. Consider the medical imaging and advanced knowledge technologies(MIAKT) project. MIAKT has developed problem environment for use in the medical domain. In MIAKT the annotation make the knowledge contained in unstructured sources(x-ray available in structured form, allowing both accurate and focused retrieval and knowledge sharing for a given patients) Moreover the annotation can be used to provide automated services ,for e.g ,they can be processed using natural language generation software to automatically draft textual report about the patient, the diagnostic information that is available and assessment made about the data by the medical team, a task which usually consumes doctors' valuable time[6].

The paper is organized as follows, In section 2, we are discussing about analysis criteria for semantic annotation tools, In section 3, we are presenting our analysis and review in tabular form according to the criteria discussed in section 2, In section 4, we have conclude the paper by adding feature that should be supported by the future generation semantic annotation tools.

Analysis criteria for semantic annotation tools

Analysis criteria that we have chosen for our review process of semantic annotation tools are overlap to some extent with the criteria set out by Handschuh et al.[7]. Analysis criteria for semantic annotation tools can be categorised as follows:

Standard format

For annotation systems in particular, standards can provide abridging mechanism that allows heterogeneous resources to be accessed simultaneously and collaborating users and



organizations to share annotations. It is the activity of the w3c in developing and promoting international standards for the semantic web.

Ontology support

Second criteria is concerned with whether our annotation tool is supporting appropriate ontology format or not. Ontology support is also concerned with the problem of ensuring consistency between ontologies and annotation with respect to ontology changes.

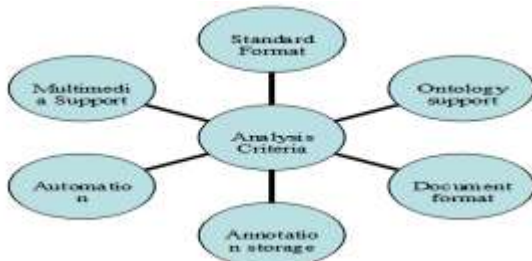


Fig.1:Analysis criteria for semantic annotation tools

Support for heterogeneous document format

Semantic web standards for annotation tend to assume that the document being annotated are in web native formats such as HTML and XML. In knowledge bases of organization we found documents in many different format including word processor files, spreadsheets, graphics files. So this is important for annotation tool, whether it support multiple document format or not.

Annotation storage

Another important criteria is where annotation will be stored, the semantic web model assumes that annotation will be stored separately from the original document whereas the “word processor model assumes that comments are stored as an integral part of the document.

Automation

Another important criteria is annotation of document or knowledge base is automatic or manual.

Multimedia support

Multimedia annotation is the next phase of development for annotation, expanding the range of files types like images video, audio. That can be annotated with the annotation tools.

Analysis and review of semantic annotation tools

Table 1: Analysis and review [17]

Annotation tool	Standard format	Ontology support	Document format	Annotation storage	Automation	Multimedia Support
Amaya [8]	RDF(S),Xlink,Xpointer	Annotation Server	HTML,XHTML and XML	Local or annotation server	No	No
Mangrove[9]	RDF	Annotation server	HTML,Email	RDF database	No	No
Vannotaa[10]	XML	Annotation Server	MPEG-2,JPEG2000,Direct3D	Annotation server	No	Yes
OntoMat[6]	DAML,OIL,OWL,S QL	OntoBroker annotation inference server	HTML,Deep Web	Annotation Server,embedded in webpage,separate file	Yes	No
M-Ontomat- Annotizer[11]	XML,RDF(S),DOLCE	OntoBroker annotation inference server	MPEG-7	Annotation server	Yes	Yes
SMORE	RDF(S)	Ontology server and ontology editing	HTML,text ,email and images	Embedded in webpage	Yes	Yes
Open ontology forge [12]	RDF(S),XML,Xlink,Xpointer	Local,editable ontologies	HTML,text,images	Local RDF or XML file	Yes	Yes
MnM[13]	RDF(S),DAML+OIL	Ontology server	HTML,Text	Embedded in webpages	Yes	No
PANKOW[14]	HTML	Pattern Based Annotation	HTML	RDF knowledge Base	Yes	No
KIM[15][16]	RDF(S),OWL	KIMO,	HTML	RDF knowledge Base	Yes	No



II. CONCLUSION

In this paper, we have present an analysis and review of semantic annotation and its various tools ,which may help the researcher in selection of appropriate tool for their application domain. Intelligent documents created by semantic annotation, would bring the advantages of semantic search and interoperability. our review of existing semantic annotation system indicates that although today we have lots of semantic annotation tools are available, each of them are fully enriched with annotation features. Still there is requirement for semantic annotation tools with more enriched features like: Linguistic annotation, Commentry annotation, Natural language processing support, to create a more intelligent knowledge management system

III. REFERENCES

- [1]. T. Berner-Lee, J. Hendler, O. Lassila, The Semantic Web, *Sci. Am.* (2001) 34-43.
- [2]. Tallis, M., Semantic Word Processing for Content Authors in Second International Conference on Knowledge Capture, (Sanibel, Florida, 2003).
- [3]. Bayerl, P.S., Lungen, H., Gut, U. and Paul, K.I., Methodology for reliable schema development and evaluation of manual annotations in Workshop on Knowledge Markup and Semantic Annotation at the Second International Conference on Knowledge Capture (KCAP), (2003).
- [4]. Yesilada, Y., Harper, S., Goble, C. and Stevens, R., Ontology Based Semantic Annotation for Enhancing Mobility Support for Visually Impaired Web Users in K-CAP 2003 Workshop on Knowledge Markup and Semantic Annotation, (2003).
- [5]. Maedche, A. and Staab, S. Ontology Learning for the Semantic Web. *IEEE Intelligent Systems*, 16 (2). 72-79.
- [6]. K. Bontcheva, Y. Wilks, Automatic report generation from ontologies: The MIAKT approach, in: *Proceeding of the 9th international Conference On Applications of Natural Language to information systems*.
- [7]. S. Handschuh, S. Staab, R. Studer, Leveraging metadata creation for the Semantic Web with CREAM, *KI '2003—advances in artificial intelligence*, in: *Proceedings of the Annual German Conference on AI*, September 2003, 2003.
- [8]. V. Quint, I. Vatton, An Introduction to Amaya, *W3C NOTE* 20-February-1997, 1997 (<http://www.w3.org/TR/NOTE-amaya-970220.html> accessed on 28 July 2004)
- [9]. L. McDowell, O. Etzioni, S. Gribble, A. Halevy, H. Levy, W. Pentney, D. Verma, S. Vlasheva, Enticing ordinary people onto the SemanticWeb via instant gratification, in: *Proceedings of the 2nd International Semantic Web Conference (ISWC 2003)*, October 2003, 2003.
- [10]. R. Schroeter, J. Hunter, D. Kosovic, Vannotea, A collaborative video indexing, annotation and discussion system for broadband networks, in: *Proceedings of the K-CAP 2003 Workshop on "Knowledge Markup and Semantic Annotation"*, October 2003, Florida, 2003.
- [11]. S. Bloehdorn, K. Petridis, C. Saathoff, N. Simou, V. Tzouaras, Y. Avrithis, S. Handschuh, Y. Kompatsiaris, S. Staab, M.G. Strintzis, Semantic annotation of images and videos for multimedia analysis, in: *Proceedings of the 2nd European Semantic Web Conference (ESWC 2005)*, 29 May–1 June 2005, Heraklion, Greece, 2005.
- [12]. N. Collier, A. Kawazoe, A.A. Kitamoto, T. Wattarujeekrit, T.Y. Mizuta, A. Mullen, Integrating deep and shallow semantic structures in open ontology forge, in: *Proceedings of the Special Interest Group on Semantic Web and Ontology*, vol. SIG-SWO-A402-05, 2004.
- [13]. Vargas-Vera M., E. Motta, J. Domingue, M. Lanzoni, A. Stutt, F. Ciravegna, MnM: A tool for automatic support on semantic markup, *KMi Technical Report*, September 2003, TR Number 133, 2003.
- [14]. P. Cimiano, S. Handschuh, S. Staab, Towards the self-annotating web, in: *Proceedings of the 13th International World Wide Web Conference (WWW 2004)*, May 17–22, 2004, New York, NY, 2004.
- [15]. B. Popov, A. Kiryakov, D. Ognyanoff, D. Manov, A. Kirilov, M. Goranov, Towards Semantic Web information extraction, in: *Proceedings of the Human Language Technologies Workshop at 2nd International Semantic Web Conference (ISWC2003)*, 20 October 2003, Florida, USA, 2003.
- [16]. B. Popov, A. Kiryakov, D. Ognyanoff, D. Manov, A. Kirilov, KIM—a semantic platform fo information extraction and retrieval, *Nat. Lang. Eng.* 10 (3/4) (2004) 375–392.
- [17]. Victoria Uren a, Philipp Cimiano b, Jos´e Iria c, Siegfried Handschuh d, Maria Vargas-Vera a, Enrico Motta a, Fabio Ciravegna c, Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semantics: Science, Services and Agents on the World Wide Web* (2005)