Using the Annotation Ontology in Semantic Digital Libraries

L. Jael García Castro[1], Olga X. Giraldo[2], Alexander García Castro[3]

[1] Universität der Bundeswehr München, Werner-Heinsenberg-Weg 39,
85779 Neubiberg, Germany
w31blega@unibw.de
[2] National University of Colombia
Palmira, Valle, Colombia
oxgiraldo@unal.edu.co
[3] University of Bremen, Bibliothekstrasse 1,
28359 Bremen, Germany
cagarcia@uni-bremen.de

Abstract. The Living Document Project aims to harness the collective knowledge within communities in digital libraries, making it possible to enhance knowledge discovery and dissemination as well as to facilitate interdisciplinary collaborations amongst readers. Here we present a prototype that allows users to annotate content within digital libraries; the annotation schema is built upon the Annotation Ontology; data is available as RDF, making it possible to publish it as linked data and use SPARQL and SWRL for querying, reasoning, and processing. Our demo illustrates how a social tagging system could be used within the context of digital libraries in life sciences so that users are able to better organize, share, and discover knowledge embedded in research articles. Availability: http://www.biotea.ws/videos/ld_ao/ld_ao.html

Keywords: Social semantic web, digital libraries, Web 3.0

1 Introduction

Semantic Digital Libraries (SDL) make extensive use of meta-data in order to support information retrieval and classification tasks. Within the context of SDLs, ontologies can be used to: (*i*) organize bibliographic descriptions, (*ii*) represent and expose document contents, and (*iii*) share knowledge amongst users [1]. There have been some efforts aiming to make use of ontologies and Semantic Web technology in digital libraries; for instance, JeromeDL (http://www.jeromedl.org) allows users to semantically annotate books, papers, and resources [2]. The Bricks project (http://www.brickscommunity.org/) aims to integrate existing digital resources into a shared digital memory; it relies on OWL-DL in order to support, organize and manage meta-data [1]. Digital libraries within the biomedical domain store information related to methods, biomaterial, research statements, hypotheses, results, etc. Although the information is in the digital library, retrieving papers addressing the same topic and for which similar biomaterial has been used is not a trivial task [3]. Ontologies have shown to be useful for supporting the semantic annotation of scientific papers [4] –and thereby facilitating information retrieval tasks. However, as ontologies are often incomplete users should be able to provide additional metadata [3, 5]. Collaborative social tagging and annotation systems have recently gained attention in the research community [6, 7]; partly because of their rapid and spontaneous growth and partly because of the need for structuring and classifying information. Collaborative social tagging is considered exemplary of the WEB2.0 phenomena because such sites use the Internet to "harness" the collective intelligence. It has been observed that several users can tag a resource; tags used for individual resources tend to stabilize overtime [8]. Our implementation uses the Annotation Ontology (AO) [9] for supporting the automatic and manual annotation of research articles. Annotations may be rooted in existing ontologies or provided by users; we are supporting the tagging of atomic components within papers –*e.g.* words, tables, figures. The content of the paper and the corresponding tags are being presented as linked data, this facilitates the interoperability between the paper and external resources –*e.g.* databases, repositories for experimental data, etc. Our approach aims to facilitate sharing, linking, and integrating knowledge across digital libraries and online resources. It also aims to support concept-based collaboration.

2 Enhancing Digital Libraries with the Annotation Ontology

The AO is built upon the Annotea Project (http://www.w3.org/2001/Annotea/); it is also compatible with Newman's (http://www.holygoat.co.uk/projects/tags/), MOAT [7] , and SKOS (http://www.w3.org/2004/02/skos/) ontologies. The AO supports free and semantic annotation over the paper; it facilitates tagging the paper as a whole as well as portions of it, *i.e.* atomic annotation. It also provides facilities for curation, provenance, authoring and versioning. Annotations are not limit to tags but also include notes, comments, erratum, etc.

Our prototype, the LD, makes it possible for users to annotate papers as well as specific sections of them, *e.g.* words, sentences, images, tables, etc. It also interoperates with automatic annotation tools such as Whatizit (http://www.ebi.ac.uk/webservices/whatizit). Annotations are used to improve search and retrieval of papers; it also makes possible to find related papers and researchers. Within the LD, the AO is used to represent the network of concepts and related resources derived from the annotations; in this sense, the AO applied to papers plays a similar role to that played by FOAF in human-centric social networks. The LD facilitates discovering links and improving interaction across papers and researchers.

An atomic annotation is shown in Fig. 1. The document is internally represented by an XML as it is the format used by the publisher; however RDF is also possible. The annotated elements are identified by using XPointer technology (http://www.w3.org/TR/WD-xptr). The provenance is based on FOAF ontology while tagging reuses Newman's and MOAT ontology. The annotation states a related meaning for the term “partial sequence on psy promoter” to the GeneBank (http://www.ncbi.nlm.nih.gov/genbank/) term AB005238, since the meaning is linked to a well established ontology, the type of the annotation is Qualifier.

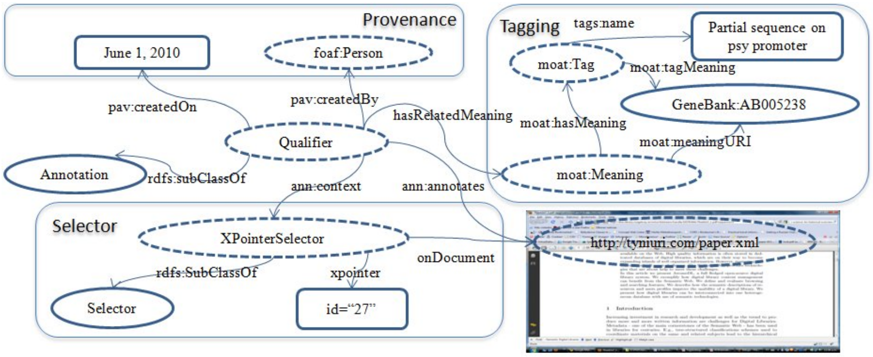
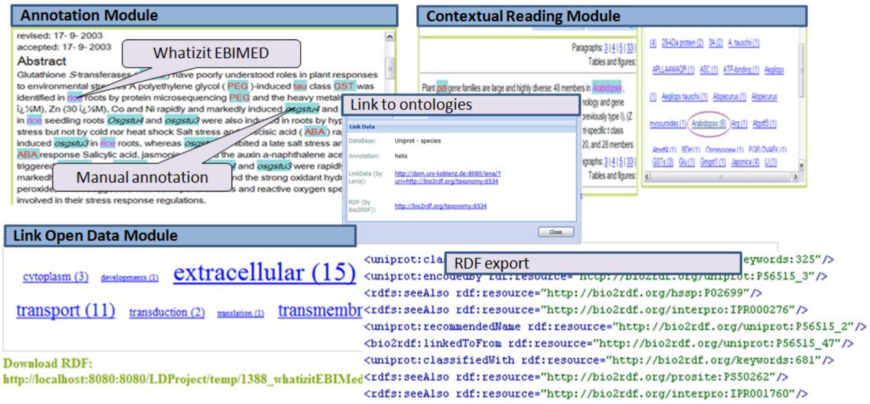


Fig. 1. LD and AO in action

The ***search & retrieval module*** is based on that one usually provided by digital libraries; it uses clouds of annotations and annotators to facilitate navigation and filtering. Once a paper is selected, the ***annotation module*** allows users to identify annotations on the paper, using different colors for different types of annotations, *i.e.* manual and automatic annotations, and also to distinguish amongst categories, *i.e.* species, proteins, genes, etc. It also allows users to manage their annotations and to link them to external resources. Additional information on automatic annotations is provided: links to specialized sources such as UniProt (http://www.uniprot.org). The ***contextual reading module*** allows easily navigating across the paper by jumping from one annotation to other. The ***linked open data module*** allows exporting annotations as RDF, making it possible to use query and reasoning languages such as SPARQL and SWRL. An overview of the LD modules is showed in Fig. 2.

3 Final Remarks

“Less is more” illustrates the collaboration dynamic that embodies the Long Tail (http://en.wikipedia.org/wiki/Long_Tail) principle within the Social Web; a huge number of people providing relatively small contributions that collectively are substantial and significant. Current available metadata in digital libraries is not enough as to support quires such as “*retrieve papers for which microarrays have been used in liver mice*”. By making it possible for ontologies and free-provided terms to live together within the scaffold granted by the AO executing such complex queries is possible. It also facilitates the enrichment of the available metadata. In addition, presenting the paper as RDF allows going beyond the PDF without compromising the business model most publishers have –selling access to the full content of the document. The LD approach offers an environment in which researchers harness the collective intelligence as they are building networks based on similar reading practices. Our future work includes: *i*) enhancing meta-data on authors and co-authors, *ii*) allowing users to organize networks, use social consensus mechanisms, and create relationships between annotations, and *iii*) better orchestrating the LD with existing biomedical ontologies, *e.g.* improving the user interface for large ontologies.



revised: 17- 9- 2003
accepted: 17- 9- 2003

Abstract

Glutathione *S*-transferases have poorly understood roles in plant responses to environmental st A polyethylene glycol (PEG)-induced tau class GST was identified in rice roots by protein microsequencing PEG and the heavy metal ï¿½M), Zn (30 ï¿½M). Co and Ni rapidly and markedly induced osgstu4 and in rice seedling roots *Osgstu4* and *osgstu3* were also ind in roots by hyp stress but not by cold nor heat shock Salt stress an cisic acid (ABA) ra induced *osgstu3* in rice roots, whereas osgs ibited a late salt stress ar ABA response Salicylic acid, jasmoni the auxin a-naphthalene ace triggere and *osgstu3* were rapidl marked nd the strong oxidant hydr peroxide s and reactive oxygen spe involved in their stress response regulations.

Link Data

DataBase: Uniprot - species

Annotation: helix

LinkData (by Lena): http://dom.uni-koblenz.de:8080/lena/?uri=http://bio2rdf.org/taxonomy:6534

RDF (by Bio2RDF): http://bio2rdf.org/taxonomy:6534

Download RDF:
http://localhost:8080:8080/LDProject/temp/1388_whatizitEBIMed

```
<uniprot:cla                                  eywords:325"/>
<uniprot:encodedBy rdf:resource="http://bio2rdf.org/uniprot:P56515_3"/>
<rdfs:seeAlso rdf:resource="http://bio2rdf.org/hssp:P02699"/>
<rdfs:seeAlso rdf:resource="http://bio2rdf.org/interpro:IPR000276"/>
<uniprot:recommendedName rdf:resource="http://bio2rdf.org/uniprot:P56515_2"/>
<bio2rdf:linkedToFrom rdf:resource="http://bio2rdf.org/uniprot:P56515_47"/>
<uniprot:classifiedWith rdf:resource="http://bio2rdf.org/keywords:681"/>
<rdfs:seeAlso rdf:resource="http://bio2rdf.org/prosite:PS50262"/>
<rdfs:seeAlso rdf:resource="http://bio2rdf.org/interpro:IPR001760"/>
```

Fig. 2. LD: Modules and Characteristics

References

1. Kruk, S., Haslhofer, B., Piotr, P., Westerski, A., Woroniecki, T.: The Role of Ontologies in Semantic Digital Libraries. European Networked Knowledge Organization Systems (NKOS) Workshop, Spain (2006)
2. Kruk, S., Woroniecki, T., Gzella, A., Dabrowski, M.: JeromeDL -a Semantic Digital Library. International Semantic Web Conference -Semantic Web Challenge, Korea (2007)
3. Garcia-Castro, A., Labarga, A., Garcia, L., Giraldo, O., Montaña, C., Bateman, J.A.: Semantic Web and Social Web heading towards Living Documents in the Life Sciences. Web Semantics: Science, Services and Agents on the World Wide Web **8** (2010) 155-162
4. Shotton, D., Portwin, K., Klyne, G., Miles, A.: Adventures in semantic publishing: exemplar semantic enhancement of a research article. PLoS Computational Biology **5** (2009)
5. Pafilis, E., O'Donoghue, S.I., Jensen, L.J., Horn, H., Kuhn, M., Brown, N.P., Schneider, R.: Reflect: augmented browsing for the life scientist. Nat Biotech **27** (2009) 508-510
6. Kim, H.-L., Scerri, S., Breslin, J., Decker, S., Kim, H.-G.: The State of the Art in Tag Ontologies: A Semantic Model for Tagging and Folksonomies. International Conference on Dublin Core and Metadata Applications, Germany (2008)
7. Passant, A., Laublet, P.: Meaning Of A Tag: A Collaborative Approach to Bridge the Gap Between Tagging and Linked Data. International World Wide Web Conference - Linked Data on the Web Workshop, China (2008)
8. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. Journal of Information Science **32** (2006) 198-208
9. Ciccarese, P., Ocaña, M., Das, S., Clark, T.P.a.B.-o.: AO: An Open Annotation Ontology for Science on the Web. Bio-ontologies, USA (2010)