



Contents lists available at [ScienceDirect](#)

Web Semantics: Science, Services and Agents on the World Wide Web

journal homepage: www.elsevier.com/locate/websem



Invited Paper

Semantic Web and Social Web heading towards Living Documents in the Life Sciences

Alexander Garcia-Castro^{a,*}, Alberto Labarga^c, Leyla Garcia^b, Olga Giraldo^d, Cesar Montaña^e, John A. Bateman^a

^a University of Bremen, Bremen, Germany

^b University of Bundeswehr, Neubiberg, Germany

^c University of Granada, Spain

^d International Center for Tropical Agriculture, Cali, Colombia

^e Los Andes University, Bogota, Colombia

ARTICLE INFO

Article history:

Received 3 June 2009

Received in revised form 15 October 2009

Accepted 26 March 2010

Available online xxx

Keywords:

Semantic Web, Social Web, Knowledge integration, Knowledge representation, Folksonomy

ABSTRACT

Rather than a document that is constantly being written as in the wiki approach, the Living Document (LD) is a document that also acts as a document router, operating by means of structured and organized social tagging and using existing ontologies. It offers an environment where users can manage papers and related information, share their knowledge with their peers and discover hidden associations amongst the shared knowledge. The LD builds upon both the Semantic Web, which values the integration of well-structured data, and the Social Web, which aims to facilitate interaction amongst people by means of user-generated content. In this vein, the LD is similar to a social networking system, with users as central nodes in the network, with the difference that interaction is focused on papers rather than people. Papers, with their ability to represent research interests, expertise, affiliations, and links to web based tools and databanks, are the central axis for interaction amongst users. To support this, we have also implemented a novel web prototype that enables researchers to accomplish three activities central to the Semantic Web vision: *organizing*, *sharing* and *discovering*. Availability: <http://www.scientifik.info/livingdocument>.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Although a significant portion of our knowledge about Life Sciences is stored in papers and online databases (DBs), the relationship amongst the information contained within papers, existing DBs and online resources is negligible. Digital libraries within the biomedical domain store information related to methods, biomaterials, topics, statements of problems being addressed, hypotheses, results, etc. But retrieving papers addressing the same topic and for which similar biomaterial has been used is not a trivial task. In order to improve search and retrieval, and also to enrich the available metadata, digital libraries should provide the facilities by means of which links can be established between atomic components of papers (domain terminologies, concepts, words, pieces of images or segments of video) and resources over the Web that are capable of processing and/or adding meaning to them. Although

tagging papers as a whole is now possible, this is still insufficient; marking and linking atomic components of papers is also necessary in order to enrich the metadata structure and to facilitate concept-based social interaction. For instance, for papers containing data types such as proteins, genes and metabolic pathways, digital libraries should link them to corresponding data entries in DB's, Knowledge Organization Systems (KOS), and/or existing online resources. Biomedical ontologies—such as those from the Open Biomedical Ontologies (OBO, <http://www.obofoundry.org>) initiative—could be used as an anchor point over which links are established and further expanded by collaborative tagging. Combining social tagging and ontology-based marking improves information retrieval and facilitates enriching metadata [1–3], thus encouraging methods that shift the creation of metadata from the individual to a collective [4].

The research reported here investigates how to support the annotation of atomic components of research papers in the Life Sciences by combining ontology-based and user-generated tags within a social network built upon the tagged concepts. To this end we propose the Living Document (LD); a document that lives on the web by interacting with other papers and resources related to the data types that it hosts. In this manner researchers can tag individual components of their papers drawing on categories from

* Corresponding author.

E-mail addresses: alexgarcia@gmail.com, cagarcia@uni-bremen.de (A. Garcia-Castro), alberto.labarga@scientifik.info (A. Labarga), leyla.garcia@ebusiness-unibw.org (L. Garcia), oxgiraldo@cgiar.org (O. Giraldo), cesar.montana@gmail.com (C. Montaña), bateman@uni-bremen.de (J.A. Bateman).

ontologies such as the Gene Ontology [5] (GO) via automatic tagging systems but without being exposed to the complexity of those ontologies. Researchers are also able to generate their own tags, or extend existing ones. The LD is therefore not a document that is constantly being written, as in the wiki approach, but rather one that acts as a document router drawing on both structured and organized social tagging and existing ontologies. Under this view, we see papers as containers of knowledge and argue that, as such, the hosted knowledge should be easily networkable with related resources. For instance, if we consider a sentence such as “IGFBP-2 expression is negatively regulated by PTEN and positively regulated by phosphatidylinositol 3-kinase (PI3K) and Akt activation [6]”, we can see that this contains valid ontology terms that could be linked to existing databases and relevant online resources. By the same token, domain experts reading the paper containing this sentence could enhance the annotation by providing, for example, more external resources relevant to the terms identified.

The framework presented here is based centrally on tags and a Semantic Web (SW) technology layer making use of such tags. From the sole perspective of software functionality, there are two main types of tags: (i) predefined tags, i.e., those coming from existing ontologies and (ii) user-generated tags, i.e., those created by the community, which collectively mediate interaction and provide the semantic tissue between papers. The combination of these two types of tags also makes it possible to build upon the knowledge contained in existing ontologies by associating their concepts with new user-defined tags. For instance, a paper in which drought tolerance is compared between *Oryza sativa* (rice) and barley in saline soils will have genes, metabolic pathways, geographical locations, atmospheric conditions attached to locations and specific periods of time, soil conditions, breeding conditions, etc. This paper could be enriched with both manual and automatic tags. Due to the interoperability infrastructure already built by bioinformaticians, biological DBs are highly interrelated. This makes it possible to execute crossed queries that retrieve integrated views: thus relating, for instance, molecular markers associated to those genes that have been studied, or geographic locations to online resources such as Google Maps. The underlying interest of the authors is then how to move from the *collected* intelligence closer to the *collective* intelligence within Life Sciences. As Berners-Lee highlights in his definition of the SW, “by better enabling computers and people to work in cooperation in an environment where information is given a well-defined meaning [7]”, we assume that structured and targeted collaboration is key to the realization of the SW vision. Life Sciences pose an ideal scenario for making this vision real as there are (i) communities of practice actively engaged in the development of their resources, (ii) ontologies being used by databases for annotation purposes and developed by communities of practice, (iii) highly interlinked and interrelated databases and (iv) analysis tools related to these databases. These features make it feasible to generate a concept-centric social network where papers accurately define relevant interaction paths and over which papers can be easily linked to external resources capable of consuming the data types they contain.

This paper is organized as follows. Related work is presented in Section 2. In Section 3 the LD approach, its corresponding architecture and functional features are presented. Finally, in Sections 4 and 5, we discuss the work done, sketching some consequences drawn from our first trials, and outline some relevant paths for the future.

2. Related work

There are several systems supporting various aspects of the problem we address here but no single system provides all

that is necessary. Resources such as Delicious (<http://delicious.com>), Bibsonomy (<http://www.bibsonomy.org>) and Connotea (<http://www.connotea.org>) facilitate the tagging of online resources and bibliographic references, thereby harnessing the collective knowledge that is modelled by collective tagging. Collaboration is thus based on similarities in tags and tagged objects. Within the biomedical community this notion of community annotation has recently started to be adopted more broadly. For instance, WikiProteins [8] delivers an environment in which it is possible to address a biological problem: the annotation of proteins. This allows the wider research community to directly benefit from the generation and peer-review of knowledge at minimal cost. The more annotations the system gets, the better the chances are for users to interact with other researchers who share similar interests/problems, for example working on the same motif, crystallographic method, etc. [8]. WikiProteins allows the annotation of proteins as a whole; however, annotating valid biological parts of a protein—e.g. positional features—is not possible. Within the context of a paper these valid biological parts could more usefully be assimilated to words/specialized-terminology/ontology terms. Another example that illustrates the usefulness of harnessing the collective intelligence is BIOWiki [9]; this collaborative ontology annotation and curation framework facilitates the engagement of the community with the sole purpose of improving an ontology. Similar to both WikiProteins and BIOWiki is the system myExperiment [10], which provides an environment where the community interacts on the basis of a common problem; in this case, however, sharing and reusing workflows is the main goal.

Within the publishing industry there has also been a series of efforts in promoting the use of Social Networks in combination with documents. BioMedExperts (BME, <http://www.biomedexperts.com>), for example, is a professional network in which literature references are used to support interaction. Although this system does not directly support tagging by users, it does support automatic tagging based on a reference terminology—thereby allowing the identification of researchers with similar interests. The system “Nature Network” (<http://network.nature.com/>) works in a similar way; it does not, however, facilitate any controlled vocabulary for annotating the literature references. Interestingly, the widely known PubMed system (<http://www.ncbi.nlm.nih.gov/pubmed>) also does not offer any kind of tagging system; nor does it make use of existing ontologies to classify documents—in this case, MESH (<http://www.nlm.nih.gov/mesh/meshhome.html>) is used to index documents. Moreover, although highly interrelated to DBs and analysis resources provided by the National Center for Biotechnology Information (NCBI), PubMed does not provide a direct relationship between the data types available in the abstracts and those NCBI resources. How semantic descriptors of resources, user profiles and enriched metadata may improve the usability of digital libraries has been investigated within a further system, that of the JeromeDL (<http://www.jeromedl.org>) project. This system offers a richer retrieval system and applies Semantic Web principles to the management of digital libraries. Finally, collaboration is enhanced still further in the Annotea project (<http://www.w3.org/2001/Annotea>) by supporting shared metadata based Web annotations, bookmarks, and their combinations; our own LD reuses concepts from Annotea within a more focused usage scenario.

Most of the investigated sites offer tagging systems for bio-related documents. However none of their functionalities addresses the problem of tagging atomic components of a paper within a social concept-centric network. Indeed, within the publishing industry in Life Sciences, publishers offer limited integration between the information contained in published papers and the Web [11]. For instance, no automatic or semi-automatic tagging systems are available, nor are there knowledge management facili-

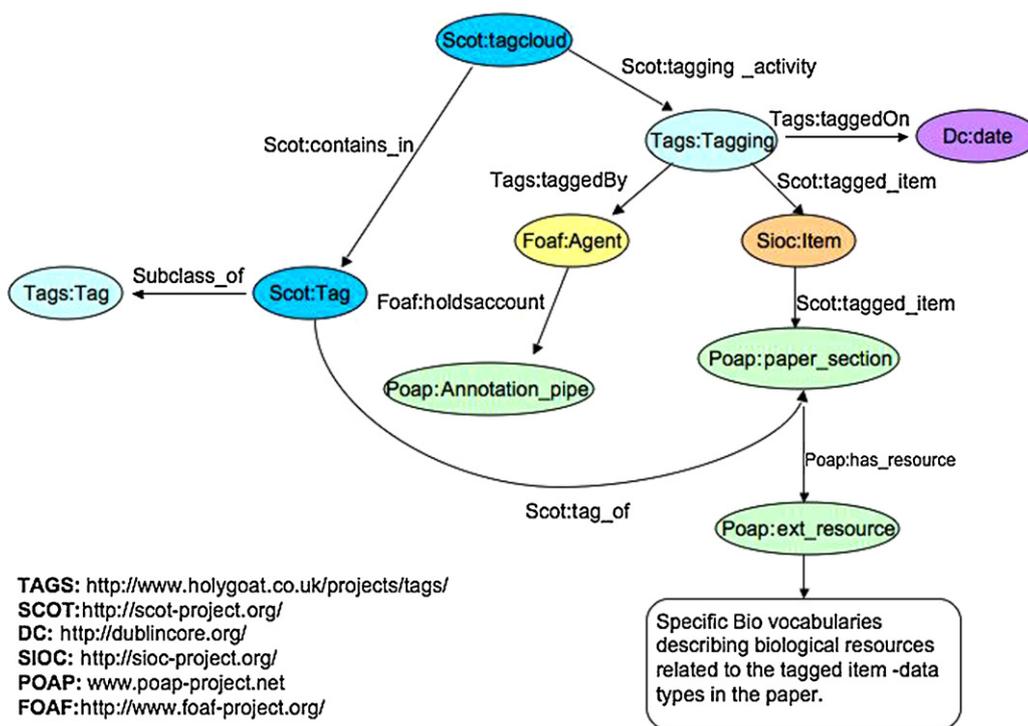


Fig. 1. Overview of POAP.

ties built over the papers and consistent with the data types sitting on them. Moreover, available application-programming interfaces (API) deliver functionalities that are not related to the contents of the stored paper.

3. The Paper-Of-A-Paper Ontology and the Living Document

One disappointing aspect of digital libraries is that papers, although in digital format, neither are interconnected nor interoperate with other valid resources capable of consuming the data types available in papers within Life Sciences. Most systems limit their functionalities to simple query facilities such as: “which other papers were published by an author”; given this, users are able to “jump” to a list of papers. Also, users can “click&jump” from bibliographic references, and then again “jump” to another paper. These are very limited functionalities for digital documents that live on the Web. The LD addresses these limitations not only by allowing researchers to define the network environment of a paper within its research context but also by predefining a network based on existing ontologies and resources. Papers are conceptually related to each other and to external resources. Papers have structural components such as authors, words, images, tables, etc.; the overall structure reflects the interests and lines of work of readers and writers. Seen in this way, therefore, a paper can be considered as a concept-based interconnection bus that provides interest-driven networking between the paper’s writers, external resources and readers.

3.1. Representing conceptual networks across papers

In order to facilitate the representation of the paper’s document structure and the conceptual networks that papers support, we have developed the Paper-Of-A-Paper Ontology (POAP, <http://www.mywikipaper.org>). POAP represents the network of concepts and associated external resources derived from the tagging activity. POAP interoperates with the Meaning Of a Tag (MOAT) ontology [12] and the Social Semantic Cloud of Tags (SCOT) ontol-

ogy [13]; both of these offer representations for the tag as well as for the social tagging activity. POAP was carefully designed to play a similar role to the Friend-Of-A-Friend (FOAF) ontology in human-centric social networks. The intersection between POAP and FOAF facilitates accurate interaction based on documents, in our case, research papers.

POAP extends these models in order for tags to represent networks of concepts across papers and external resources. A simple view that illustrates some sections of POAP and how it is interoperable with other models is presented in Fig. 1. Ultimately POAP aims to support, in coordination with models such as BioMoby [14], the discovery of services capable of consuming those valid biological data types available in papers.

3.2. Architecture and implementation

A general overview of the LD is presented in Fig. 2. Documents have sections (images, tables, words, phrases) and these can be tagged by users or by automatic pipelines. Other workflows can be easily added via the LD API.

For the purpose of our prototype we have developed the necessary infrastructure for the storage, indexing, annotation and retrieval of articles. Initially we have worked with papers from the Elsevier digital collection, but adding new data sources, such as PubMed, DBLP (<http://dblp.uni-trier.de>) or any other XML-based digital library, is also straightforward. The search and retrieval module has been developed on top of the Lucene (<http://lucene.apache.org/>) project framework; this is an open-source, high-performance, full-featured text search engine library written entirely in Java. Our system allows the user to search globally across Elsevier journals or individually in selected resources; it also facilitates filtering by authors or dates by using an advanced search. An overview of the architecture we have implemented is illustrated in Fig. 3.

Our prototype already makes use of annotation pipelines such as WhatIzIt (<http://www.ebi.ac.uk/whatizit>) [15]. Reflect (<http://reflect.ws>) will also soon be incorporated, while services

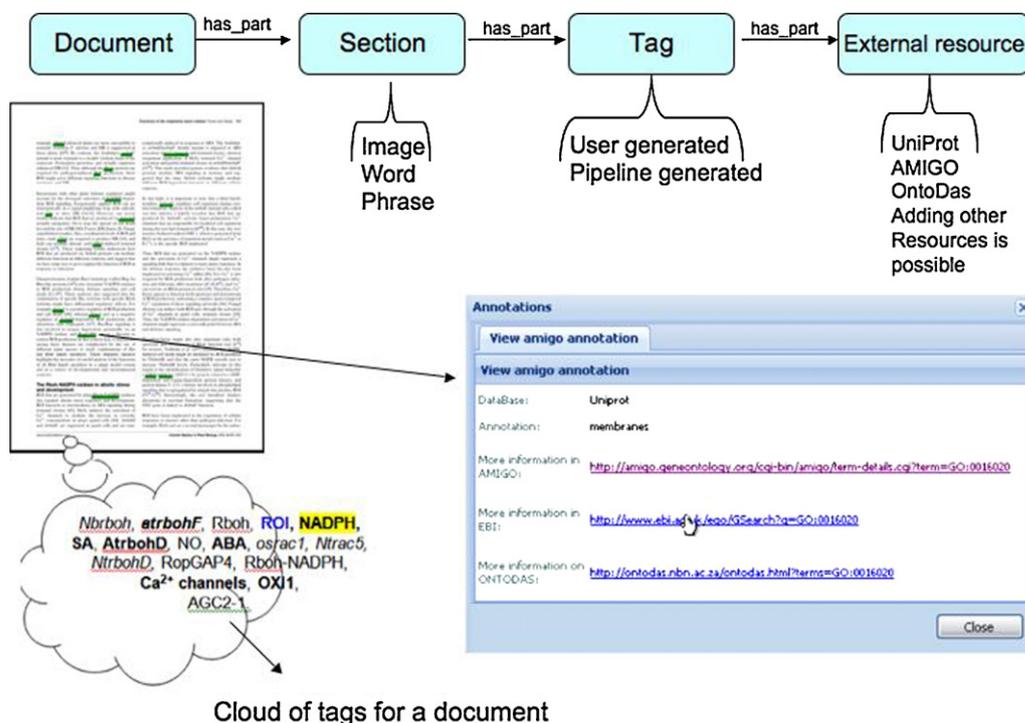


Fig. 2. General overview of the Living Document.

that have recently become available, such as the Open Biomedical Annotator [16], BioNotate (<http://bionotate.sourceforge.net/>) and some of the Web services offered by Bioportal [17], can also be plugged into our architecture as shown in Fig. 3. Annotation pipelines in general make it possible for users to generate their tags over specific sections (e.g. words, sentences) of papers. There are currently over 500 papers in our system and new documents

can be easily loaded. Users can share and reuse tags for improved definition of queries; it is also possible for users to specify external resources associated with tags and sections of the tagged papers.

For instance, by means of WhatIzIt users can tag words such as “NADPH” (nicotinamide adenine dinucleotide phosphate); this tag has a default external resource that provides more information about it, UniProt (www.uniprot.org). Other users could have

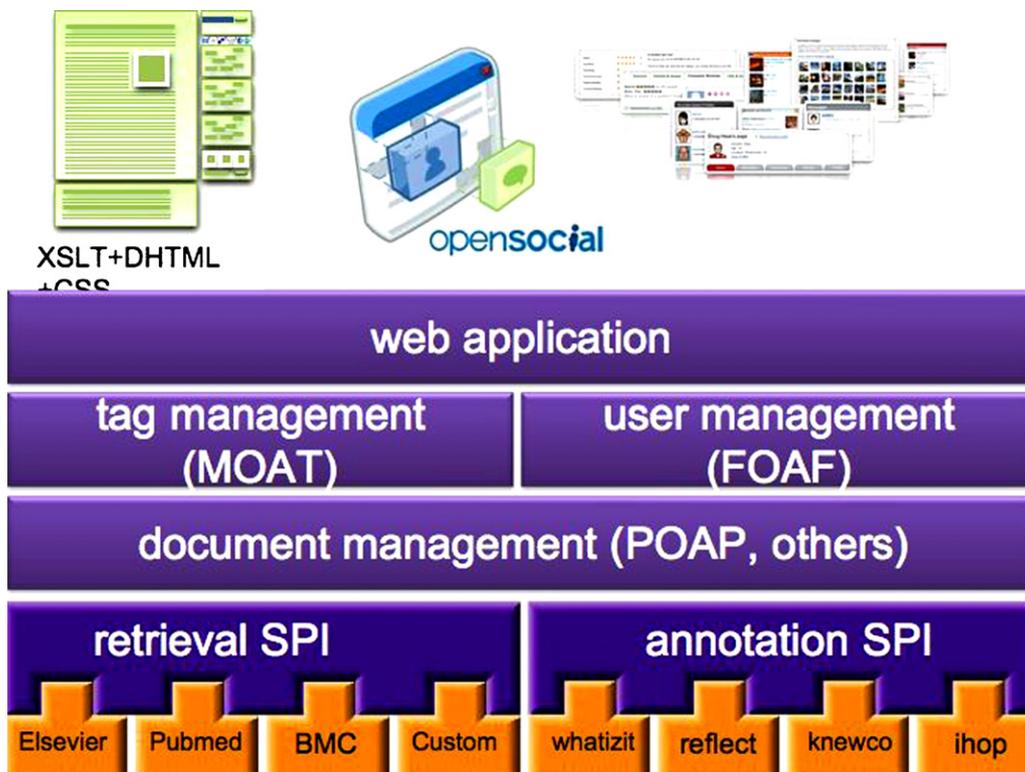


Fig. 3. General system architecture.



Fig. 4. Users are able to generate sets of automatically generated tags by using existing ontologies and/or annotation pipelines.

found it important to manually tag the term “Ca Channels” and link it to a different external resource. In order to support other digital libraries and different annotation pipelines our architecture supports the Service Provider Interface (SPI) paradigm for retrieval and annotation—see Fig. 3. The SPI is a software mechanism that supports replaceable components via a set of hooks. On top of these SPIs, we have built a semantic layer supporting MOAT, FOAF, POAP and other controlled vocabularies (also shown in Fig. 3). This makes it possible for the new metadata to be managed so that more expressive queries can be supported.

3.3. The LD prototype

The LD prototype itself allows the generation of two types of tags, those generated by an automatic workflow, such as WhatzIt, and those generated by human users. The automatic annotation infrastructure is built around the Monq software package (www.ebi.ac.uk/~kirsch/monq-doc/monq); this Java library enables the processing of text input streams based on regular expressions. The library binds regular expressions to actions that are automatically executed whenever a match occurs in the text stream being processed. A filter server is a computer program, also Java technology, which accepts TCP connections on a particular port from clients across the network. Each filter server specializes in recognizing the vocabulary of a particular terminology. Clients connect to a filter server and send a stream of text. The server then runs a collection of embedded discrete finite automata (DFA) on the incoming text to recognize and tag the terminology with XML tags.

Multiple filter servers can be cascaded to form processing pipelines, whereby the output of one filter becomes the input of the next. In this case the XML tags added by each server carry the data needed to accomplish the tasks of complex distributed text mining algorithms. Currently, we support the annotation of Swissprot protein/gene names, drug names, organism names, disease names, chemical entities and gene ontology terms. The pipeline also considers some disambiguation based on acronym resolution and term frequency. Protein/gene names resembling acronyms, for instance NPY (neuropeptide Y), are analyzed in order to disambiguate whether the target name is really a Protein/Gene name. If it is unclear, then the pipeline will assume that names with a high frequency in the British National Corpus (<http://www.natcorp.ox.ac.uk>) are common enough as to be considered relevant in the biomedical field. The result of this process is an extended XML document containing the annotated information.

Users can, for instance, tag the paper using GO as illustrated in Fig. 4; once the set of tags has been generated users are free to add

or modify the set of predefined tags. The functionality embedded within the set of predefined tags also includes a set of predefined links built over the tag. For example, as illustrated in the figure, the GO term “membranes” has a direct reference to its corresponding DB entry in AMIGO (<http://amigo.geneontology.org>). By default these tags are automatically linked to a further set of external resources such as ENTREZ (<http://www.ncbi.nlm.nih.gov/Entrez>), AMIGO, SRS (<http://srs.ebi.ac.uk>) and others. Both, the ontologies being used to support the automatic tagging and the corresponding external resources being used by default, can readily be extended.

The addition of new external resources, such as URIs capable of either consuming tags as input data types, or adding extra information, is currently achieved by typing the information required into a text field. By adding new tags and external resources, the clouds of tags for each paper are constantly being updated. Furthermore, users can use the clouds of tags in order to find more information and build more expressive and accurate queries, as suggested in Fig. 5. In addition, users can identify tags that are associated with collections of documents. This tag association is a measure indicating how closely related the documents are. More importantly, users can browse through the surrounding area of the tags so that they can contextually discover how valid the coincidence in tags may be. For instance, two documents may share the user-generated tag “<X>” but this coincidence alone may not be a sufficient indicator that the documents are actually related because one tag may have multiple meanings or be applied in different contexts (Fig. 6).

An interesting side effect of social tagging of this kind is the semi-automatic generation of social trust rankings. As users tag documents, these tags are easily identified by ownership. Thus any user can ask for some particular person's tags without including in the cloud of tags those generated by anyone else. A user can also select only those tags generated for a given paper by user X, Y, and W; this usage reflects the fact that people tend to trust tags generated by particular members of the community more than those produced by others.

4. Discussion

Folksonomies have recently gained attention from the research community [18], partly because of their rapid and spontaneous growth and partly because of the need for structuring and classifying information. Although social tagging is widely used and, as has been demonstrated by numerous applications, clearly beneficial, clouds of tags *per se* are not formal classification systems;



Fig. 5. Tags have references to external resources.

rather, they are a complementary form of organization system. For the task of finding information, taxonomies tend to be rigid and purely text-based search is not optimal. Tags introduce distributed human intelligence into the system [19]; also, tags are axes over which collaboration is supported. When researchers collaborate, the structure of the collaboration is based upon similarities in their work; an aspect of this similarity may be defined by the literature that researchers are reading. It has been observed that if two researchers use similar literature they are working on similar or conceptually related problems. Bibliographic references, keywords and abstracts are a valuable starting point for supporting information retrieval across large digital libraries and interaction based on similar interests. However; they do not offer the possibility for establishing networks of associated concepts across papers (NACAP); nor do they offer the facility for linking concepts from the paper to external resources (P2ext).

Both these features, NACAP and P2ext, are central to the structure of real collaboration amongst researchers. Extending support for this interaction by using social tagging over entire documents augments the support for the collaboration based on similar research interests. Moreover, the metadata thus generated can help researchers to retrieve an article when descriptive elements are not known. The main function of these tags is not to support a fixed

taxonomical classification, but a dynamic one [7], allowing the relevance and content of articles to keep pace with the evolution of scientific discovery. This dynamic annotation is the serendipity that may allow researchers to find other researchers working on similar areas as they find papers via generated tags. Providing a platform where both NACAP and P2ext can be exploited, both at the time of authoring the paper and post-publication, will greatly enhance social networking and information retrieval in Life Sciences. This enhancement allows knowledge to be discovered more expediently and facilitates the formation of sub-networks of collaboration over specific knowledge units, defined by tags. Social tagging provides direct insight into the knowledge conveyed within the body of a scientific paper.

Ideally social networks and ontologies should help offer an environment in which researchers can take advantage of collective knowledge. In principle, efforts such as Delicious and Connotea facilitate both, social interaction and harvesting the collective intelligence. Delicious offers a collective annotation facility for bookmarks in which the community interacts via the annotations they generate over their own bookmarks; the community as a whole benefits from everybody's knowledge as it is always possible to access everybody's annotations. Connotea is a more targeted environment in which users share their bibliographic references as

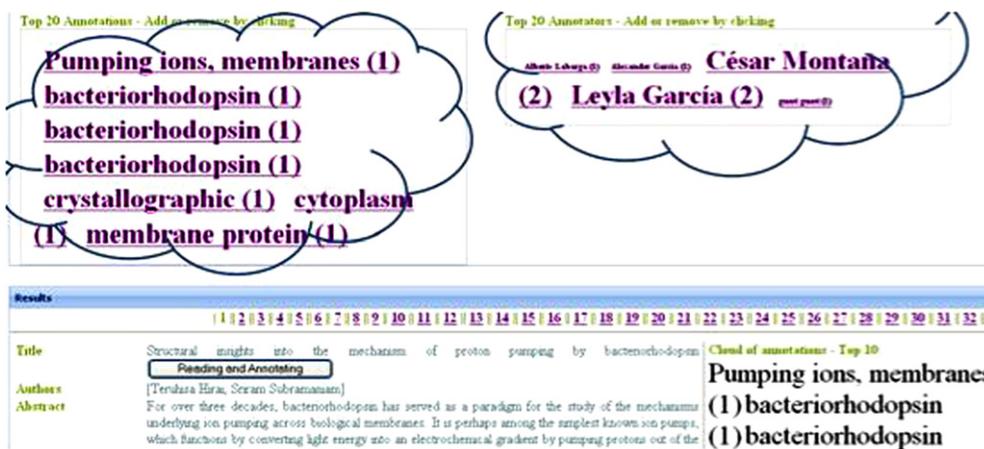


Fig. 6. Using tags to refine queries.

well as those annotations that describe the shared set of references. The Connotea approach relies on the assumption that users with similar interests should have bibliographic references in common. The limitation of Connotea is that it assumes that an annotation, independently from the nature of the annotated object, always has a structure similar to that of a bibliographic reference. In order to offer insights into papers, systems similar to Connotea facilitate the manipulation of abstracts. However, abstracts *per se* do not provide a full summary of the work described in some document, nor do they offer any way to integrate the document into existing knowledge. Connotea, as well as similar systems, does not support the entire structure of collaboration that is usually found within and across communities of researchers.

During the testing phase of our system, we had two trial researchers perform a collection of typical knowledge management and research tasks. During this work, both trial researchers moved from simple tagging into building classifications as well as using tags for refining queries. Interestingly, for those papers we used in our tests, the recommendations given by the original digital library were not always consistent with the paper discovered by the generated tags. The tag-based recommendation was more accurate. The axes, tag-tagged_object and tagger, over which the tagging was being performed were consistent with those described by MOAT and SCOT; such simple models proved to be easily extendible so that semantic and syntactic components of a scientific paper could be represented by POAP and coherently orchestrated with MOAT, SCOT and other existing tag-related models.

Concept Web Linker (CWL) is a similar approach to the one presented in this paper – see <http://conceptweblinker.wikiprofessional.org>. Both, this approach and ours, assume a paper deeply interconnected with other papers and with the Web. However, unlike CWL, the LD approach embraces a generative technology so users can actively generate the tools they need for the kind of information they want to manage. Furthermore, CWL does not allow users to improve the definition of queries based on available tags, thus limiting the usability of the cloud of tags. The generative capacity of a system should be understood as “its capacity to produce unanticipated change through unfiltered contributions from broad and varied audiences” [20]. Currently, the generative capacities in CWL are restricted; for instance, it is not possible to add annotation pipelines. More importantly, CWL does not make use of social networking technology. Generative technology of the kind we argue for enables an open market for which specialized plug-ins can be developed; in this way it will be possible to build better and more specialized mining tools over digital libraries.

5. Conclusions and future work

We have developed a Web prototype that allows the ontology-based or user-generated tagging of atomic components within the structure of scientific papers in Life Sciences. It also makes it possible for users to deploy tags in order to improve the definition of their queries and to better filter search results; users can always refine their searches adding terms to the queries from the cloud of tags. Finding related papers is then supported not only by those available tags, but also by the prototype’s use of eTblast (<http://invention.swmed.edu/etblast>); this allows the user to input an entire paragraph and returns MEDLINE abstracts that are similar to it. The LD enables authors to easily add scientific hyperlinks to their documents and research papers as semantic annotations, drawn from ontologies or provided by their peers; this links their papers to the Web in a meaningful way. The LD is complementary to the recently released MS Word Add On (<http://ucsdbioliot.codeplex.com>, May 11/09); however, it goes one

step further as it involves massive collaboration in the process of generating semantic annotations; it is also appropriate for existing digital libraries.

Our architecture is flexible so that loading XSLT files from other digital libraries is possible; we are currently starting to work with PubMed and DBLP (<http://dblp.uni-trier.de/>) for further testing. Another digital library of interest to us is Google Scholar; we are using Google’s latest API in combination with the Google Web Development Toolkit (GWT) and its corresponding extensions in order to exploit user and ontology generated tags. We are also carrying out more detailed evaluations not only concerning the tagging and its related operations but also regarding the added value required for communities to adopt and actively participate in scientific-oriented folksonomies. Our interest is the same: the intersection between the SW and the Social Web in Life Sciences. One interesting aspect that has arisen from our work with biologists is the need for intelligent interfaces—i.e., interfaces consistent with the information being delivered. We are consequently redesigning our interfaces to cater to this requirement. Support for browsers other than Internet Explorer (IE) as well as the release of the project to the SourceForge community is also being pursued. And finally, we are conducting further research into the relationship between ontologies and the new breed of ‘tagsonomies’ that naturally arises out of approaches such as ours.

Acknowledgements

Funding: Alexander Garcia acknowledges the financial support of the EU: Large-Scale Integrating Project OASIS, Grant Agreement # 215754.

References

- [1] S. Bindelli, et al., Improving search and navigation by combining ontologies and social tags, in: Proceedings of the OTM Confederated International Workshops and Posters on On the Move to Meaningful Internet Systems, Springer-Verlag, Monterrey, Mexico, 2008.
- [2] M. Chen, M.X. Liu, J. Qin, Semantic relation extraction from socially-generated tags: a methodology for metadata generation, in: Proceedings of International Conference on Dublin Core and Metadata Applications, 2008.
- [3] J. Hunter, I. Khan, A. Gerber, HarVANA—harvesting community tags to enrich collection metadata, in: ACM IEEE Joint Conference on Digital Libraries, JCDL, Pittsburgh, PA, USA, 2008.
- [4] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, Scientific American, 2001.
- [5] M. Ashburner, et al., Gene ontology: tool for the unification of biology. The gene ontology consortium, Nature Genetics 25 (1) (2000) 25–29.
- [6] R. Mehrian-Shai, et al., Insulin growth factor-binding protein 2 is a candidate biomarker for PTEN status and PI3K/Akt pathway activation in glioblastoma and prostate cancer, PNAS 104 (13) (2007) 5563–5568.
- [7] D. Gendarmi, F. Abbattista, F. Lanubile, Fostering knowledge evolution through community-based participation, in: 16th International World Wide Web Conference, Banff, Alberta, Canada, 2007.
- [8] B. Mons, et al., Calling on a million minds for community annotation. WikiProteins, Genome Biology (2008).
- [9] M. Backhaus, J. Kelso, BIOWiki—a collaborative annotation and ontology curation framework, in: 16th International World Wide Web Conference, Banff, Alberta, Canada, 2007.
- [10] D. De Roure, et al., The myExperiment Open Repository for Scientific Workflows, in: Open Repositories, Atlanta, Georgia, US, May 2009.
- [11] D. Shotton, et al., Adventures in semantic publishing: exemplar semantic enhancements of a research article, PLOS Computational Biology 5 (4) (2009).
- [12] A. Passant, P. Laublet, Meaning of a tag: a collaborative approach to bridge the gap between tagging and linked data, in: International World Wide Web Conference-Linked Data on the Web Workshop, China, 2008.
- [13] L.-H. Kim, J.G. Breslin, S.-K. Yang, H.-G. Kim, Social semantic cloud of tag: semantic model for social tagging, in: Agent and Multi-Agent Systems: Technologies and Applications, Springer, Berlin/Heidelberg, 2008, pp. 83–92.
- [14] M. Wilkinson, BioMOBY: an open-source biological web services proposal, Briefings in Bioinformatics 3 (4) (2002) 331–341.
- [15] D. Reibold-Schuhmann, et al., Text processing through Web Services: Calling Whatizit, Bioinformatics 24 (2) (2007) 296–298.
- [16] J. Clement, S. Nigam, M.A. Musen, The open biomedical annotator, in: AMIA Summit on Translational Bioinformatics, San Francisco, 2009.

- [17] N. Noy, Bioportal: a community-based repository for biomedical ontologies and data resources, in: AAAI 2009 Spring Symposium on Social Semantic Web: Where Web 2.0 Meets Web 3.0, 2009.
- [18] L.-H. Kim, et al., The state of the art in tag ontologies: a semantic model for tagging and folksonomies, in: International Conference on Dublin Core and Metadata Applications, Berlin, Germany, 2008.
- [19] T. Gruber, Ontology of folksonomy: a mash-up of apples and oranges, *International Journal of Semantic Web and Information Systems* 3 (2) (2005).
- [20] J. Zittrain, *The Future of the Internet and How to Stop It*, Yale University Press, New Heaven & London, 2008.