The Ontology Viewer: Facilitating Image Annotation with Ontology Terms in the CSIDx Imaging Database

Amalia Kallergi
akallerg@liacs.nl

Yun Bei
ybei@liacs.nl

Fons J. Verbeek
fverbeek@liacs.nl

section Imaging & BioInformatics - Imagery & Media group
Leiden Institute of Advanced Computer Science (LIACS), Leiden University.
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

ABSTRACT

In the life sciences data must be described unambiguously. We apply this principle in our multi-modal bio-imaging database in which images are stored together with comprehensive metadata. We use ontology terms to describe the semantic content of images. Ontologies are obtained from dedicated ontology repositories in the life sciences. For our users, the process of image annotation with ontology terms was proven to be a challenging task. Therefore, we have made improvements on both usability and speed of annotation. We developed search facilities across our ontology collection and implemented a new graphical ontology viewer. This tool allows for both querying and visualizing ontology terms by means of a 2D-graph representation. Our viewer provides a means to collect ontology terms and at the same time familiarizes users with ontologies and their structure. In making these tools available we succeeded in our goals to reduce time and effort for accurate image annotation.

Author Keywords

Ontology, life sciences, annotation, graph, images

ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Web-based interaction.

INTRODUCTION

The Cyttron Scientific Image Database for Exchange (CSIDx) is a multi-modal imaging database for images produced in the life sciences [2,7]. In CSIDx, image annotation is a fundamental aspect of image submission and ontology terms, as extracted from life-sciences ontologies, are used to define the semantic content of an image. These ontologies with their intrinsic curation and relations between all terms help to obtain unambiguous annotations and to assess synonyms. Moreover, CSIDx aims to explore the added value of the ontological relations towards integration of images in representing biological concepts. In this section, we briefly introduce the scope and aim of the database and provide a short overview of the image annotation procedure.

CSIDx is built to support a wide range of imaging modalities and techniques and it is the backbone database of the Cyttron project [3], a consortium towards an integrated infrastructure for bio-imaging and modelling cells down to atomic detail. CSIDx is also a web-based community in which researchers from various institutes can share their image resources. The database is accessible via a web interface and the design is based on rich Internet application practices that allow for dynamic and responsive web applications. The system is developed, maintained and physically hosted by the Imaging and Bio-Informatics group at Leiden University.

In CSIDx, we propose that metadata as the information that describes an image is essential to support exchange and linking as well as analysis of images [1,2]. A key feature of CSIDx is linking of imaging modalities via concepts towards integration of functional concepts and to that end, an unambiguous annotation is required. Therefore, CSIDx stores both raw pixel data and user generated annotations. A major part of the CSIDx development is dedicated to tools that facilitate the process of an extended annotation by the image owner. The development process and design of new features is accomplished in close collaboration with users; i.e., biologists, structural biologists and others, whose feedback is registered via observation and informal interviews.

In order to assure a comprehensive annotation that also represents the actual image acquisition conditions, CSIDx maintains metadata about the 'who', the 'what' and the 'how' of an imaging experiment [1,2,7]. In this paper, we particularly address the metadata on what an image is about i.e. information about the biological phenomenon depicted or the phenomenon the image relates to. This annotation corresponds to the semantic content of the image and captures the interpretation of an image as given by the domain expert or the researcher responsible for the image

Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW2009), IUI2009, Feb 8 2009, Sanibel Island, Florida, USA.

acquisition. To assure accurate metadata [1,2] and to explore possible relations between the images, the 'what'-part of the annotation is expressed in ontology terms as extracted from life-science ontologies. In comparison with free text or user generated keywords, ontology terms guarantee consistency across the system and prevent ambiguities and spelling mistakes. Moreover, ontologies provide well-defined relations across the concepts which can be further explored in structuring or mining the image data. The use of domain specific ontologies also corresponds with the emerging practices in the field of life-science data repositories towards well-maintained and reusable resources by means of a common semantic representation.

In this paper, we describe our efforts and tools to support and facilitate image annotation with ontology terms. In particular, we describe our ontology viewer, a graphical tool developed to assist image annotation based on ontologies. CSIDx currently incorporates 37 life-science related ontologies, the majority of which are retrieved from the Open Biomedical Ontologies (OBO) Foundry [9]. The OBO Foundry is a platform to share biological ontologies in a common syntax and the maintained ontologies are available in a variety of formats such as OBO [9], OWL [10] and RDF [12]. The biological ontologies available are developed and maintained by researchers in the biomedical field and provide a fair overview of the domain specific knowledge and vocabulary in the field of life-sciences. In this manner, about half a million unique terms are available for annotation.

CHALLENGES OF USING ONTOLOGIES IN IMAGE ANNOTATION

In an earlier prototype of CSIDx, users could annotate images by selecting ontology terms from our ontology tree browser. This application depicts the hierarchical relation in the ontologies as a tree view; this view only displays the *subClassOf* relation. With an interactive tree view (cf. Figure 1), users can navigate by collapsing and expanding terms in the hierarchy and can select terms to be assigned to the image under annotation. The ontology tree browser parses ontologies in the OWL-format by means of the Jena framework [5]. This tool provides some control and structure in dealing with the available ontology terms as well as some means to navigate the ontologies. However, this approach was found to be insufficient for the extended annotation and usability requirements of CSIDx.

In fact, the introduction of ontology terms for image annotation was in itself a significant challenge for our users. Firstly, the majority of our users were not familiar with the exact concept of the ontology. Although all the ontologies used in our system are maintained by the bioscience community, hardly any of our users had prior extensive experience with ontologies. In particular, they had no mental image of the structure of an ontology and they demonstrated difficulties in comprehending that relations other than the child-parent relation of a hierarchy

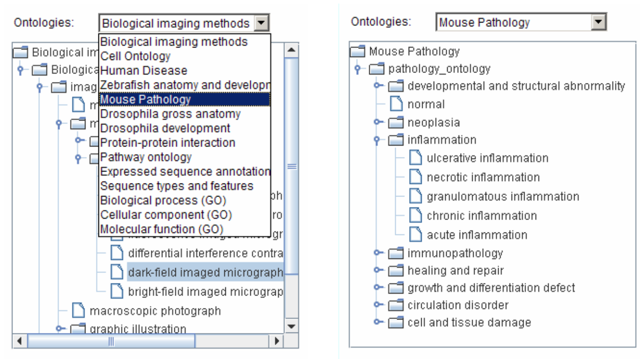

Figure 1. The ontology tree viewer.
Left: Selecting an ontology from a list.
Right: Selected ontology with expanding tree view.

may occur and that terms can be interconnected. During our sessions with the users, we often found ourselves sketching out ontology graphs on paper to explain an ontology. Secondly, most of our users were insufficiently familiar with the content of the ontologies. Simply, they did not have sufficient knowledge on what terms are to be found in each separate ontology. We expected that their biological knowledge would help them locate the terms of interest in the hierarchy but the ontology structure as given in the hierarchy view was not always matching the user's expectations. Additionally, the vast amount of terms available was difficult to manage. Even when a term was known or previously identified, clicking through the several levels of an ontology hierarchy to locate the term was time consuming and - from the user's point of view - unpractical and unacceptable.

FORM OF A SOLUTION

To address the challenges of using ontology terms for annotation, we examined the difficulties faced by our users and the limitations of the hierarchical visualization of the ontologies. Testing with our prototype provided useful knowledge on how users interact with ontologies. Through participatory evaluations, we learned that users need to learn the ontology content, to build a mental model of ontology structure and to extract information from ontologies. The complexity of the annotation task increases due to the lack of search facilities, the overwhelming amount of terms and the lack of experience with and understanding of ontologies. Hence, we implement a solution that aims to improve the annotation process both in terms of usability and in terms of ontology comprehension. In particular, we provide search facilities on the ontology terms corpus and implement the ontology viewer, a graphical tool used for both querying and visualizing ontology terms. In combination with these facilities and in order to reduce the annotation effort, the concept of MyTerms was also introduced in the workflow of the annotation process.

Construct	OWL-DL	Simplified Relation
SubClassOf	$A \subseteq B$	$\forall x\ [B(x) \rightarrow A(x)]$
Restriction	$A \subseteq \exists\ P.B$	$\exists x \exists y\ [A(x) \wedge B(y) \wedge P(x,y)]$
EquivalenceClass &IntersectionOf	$A \subseteq B \cap C$	$\forall x\ [A(x) \rightarrow B(x) \wedge C(x)]$
EquivalenceClass &UnionOf	$A \subseteq B \cup C$	$B \subseteq A,\ C \subseteq A$

Table 2. Indirect relations in OWL-DL are transformed in straightforward relations to be stored in a database schema

Querying Ontologies Using a Database Back End

From the user perspective, quick concept (keyword-like) searches across the ontologies are essential in order to complete an extensive image annotation with ontology terms. Keywords and textual descriptions of images as conceived by the image owner need to be mapped to existing ontology terms. Such a procedure is not easily accomplished without any search facilities especially when the user is not familiar with the content of the ontologies.

Ontology querying mechanisms can involve the use of dedicated RDF query languages such as SPARQL [13]. More elaborate forms of querying, like reasoning can be accomplished with reasoners such as Pellet [11] and KAON2 [8]. Although powerful, these mechanisms are heavily challenged when large or complex ontologies are involved and do not demonstrate fast performance in terms of speed of a query [4, Bei internal technical report].

CSIDx is focused on the domain of the life science where ontologies and controlled vocabularies tend to be enormous in size and/or are constantly updated and expanding. In addition, the ontology structure tends to include elaborate relations which result in increased complexity when querying or reasoning with the ontology. However, the web based character of CSIDx gives a high priority on speed and reactivity of the system. Being confronted with such a practical limitation, we adopted a solution with a Relational Database Management System (RDBMS) to support fast ontology queries. Specifically, our ontology resources were transformed from their original OWL format to a simplified schema that can be easily stored and queried by means of a RDBMS. Namely, the indirect relationships in OWL-DL are transformed into concise, direct relationships and the complete ontology structure is expressed as a directed graph of concepts and their relations that can be easily stored in a database schema. Examples of the transformations applied are given in Table 1. Such a representation definitely lacks the completeness, complexity and expressive power of the OWL- DL language but allows us to perform queries with high performance. For the purposes of image annotation, we believe that such a representation, although incomplete, is still able to provide a sufficient view on the domain knowledge.

By designing and populating the ontology database, which currently consists of 565,600 terms and 825,724 relations, we are able to support querying facilities across the ontologies. Users of CSIDx can search for a corresponding term by keywords using either the ontology viewer (cf. next section) or a simplified web search form.

MyTerms: A User Specific Collection of Terms for Annotation

To reduce the effort required for identifying annotation terms in the ontology collection, we have introduced the concept of MyTerms in the workflow of the annotation process. MyTerms is a collection of user specific ontology terms that are saved under a user's profile and can be reused across annotations. Prior to an actual image annotation, users can browse the ontology collection with the querying tools available looking for terms that are relevant to their study or field of research. During an image annotation, users assign terms to images by selecting terms from their own relevant subset (MyTerms) instead of the complete corpus of terms available. This process is an attempt to minimize the effort of searching for terms (search once, use in all subsequent annotations) and to reduce the overwhelming amount of ontology terms to a subset that is both meaningful to the user and easier to browse and use.

The MyTerm concept can be further elaborated to match the structure of our system. In CSIDx, users are organized in groups that correspond to their actual research institute or group and this organization is often used throughout the interface as a mechanism for exchanging shared resources, such as images or microscopes. Therefore, we also provide the possibility of sharing identified terms among group members, who are likely to work on a similar topic. In the case of group shared ontology terms, the time and effort spent by a group member to locate and identify useful terms across the ontologies profits all members of the research group. On the whole, MyTerms assure that the admiringly time-consuming process of mapping metadata to existing ontology terms does not need to be unnecessarily repeated.

THE ONTOLOGY VIEWER

The ontology viewer provides a graphical interface for querying ontology terms and a means to visualize the ontology structure. We believe that a graphical representation can assist our users in building a mental model next to building a collection of terms. In practice, it is a tool to assist building a MyTerms list and an attempt to demystify ontologies to our users by making the relations among ontology terms obvious. The application is developed in Java and deployed as a WebStart application. It can be accessed via the CSIDx web interface or used as a standalone application for registered users.

The ontology viewer (cf. Figure 2) consists of two major panels: a query form and a 2D viewer. In the query form, users search for ontology terms within an ontology by providing one or more keywords and by specifying the level of detail for the search. Queries can be performed on the label, synonym or definition of terms and keywords can be combined in an 'AND' or 'OR' query. Users can choose from the list of results to either visualize particular terms or directly add terms to their MyTerms list.

In the 2D viewer, the ontology structure is represented as a graph in which terms are graph nodes and relations are graph edges. Selected ontology terms, as collected from a query, are used to produce a sub-graph of the ontology graph. This sub-graph provides the local context for the selected nodes which are highlighted green to distinguish from their connected terms.

A short description with information on any given term can be obtained by mouse over the corresponding graph node. Regular graphical manipulations are supported on the ontology graph which can be zoomed, paned, rotated and sheared. In this manner, user can adjust the view to better understand the displayed relations. The ontology viewer also provides different graph layouts to support a more suitable or preferred arrangement in space, especially in the case of complex sub-graphs. The supported layouts are given in Table 2. To improve clarity of the presentation, both the text labels of either nodes or relations and nodes other than the selected nodes can be toggled on or off. The graph drawing and manipulation is implemented by means of the Java Universal Network/Graph Framework (JUNG) [6].

DISCUSSION AND CONCLUSIONS

Overall, the CSIDx ontology viewer provides an informative graphical interface to the collection of ontologies in CSIDx. As most ontologies are derived from the OBO matrix, this viewer is also an alternative graphical entry point to exploring the most popular and acknowledged ontologies in the domain of the life sciences. Importantly, for most CSIDx users, this interface is their first impression on biological ontologies and a first step towards familiarizing themselves with the concept and content of ontologies. Compared also to the web based querying facility that lacks the graph display, users have reported that the connected terms often help clarify ambiguities: when the label of a term can be explained differently depending on the context or when the description of a term is insufficient, the connected terms are often conclusive on the exact meaning of the term. As a result, users feel more confident that they have selected the proper term for precise annotation. The graph representation also seems to assist users in rethinking the way they translate their desired annotation to ontology terms. By exploring the ontology, users often conclude on more terms than they initially queried for and they often express the desire to automatically add to their user term list (MyTerms) the whole graph structure as displayed in the 2D viewer. Overall, the ontology viewer contributes towards a more complete and accurate annotation based on ontology terms.

Layout	Algorithm
KKLayout	The Kamada-Kawai algorithm
FRLayout	The Fruchterman-Rheingold algorithm
CircleLayout	A simple layout which places vertices randomly on a circle
SpringLayout	A simple force-directed spring-embedder
SpringLayout2	Another simple force-directed spring-embedder
ISOMLayout	Meyer's "Self-Organizing Map" layout

Table 2. Graph layouts available in the ontology viewer

As they familiarize themselves with the ontology structure, users demonstrate the wish to further interact with the ontology. Often, they request to expand the displayed nodes, a requirement that equals with interactively traversing the complete ontology. While the ontology viewer was basically aimed to provide some context for the queried terms rather than a complete overview of an ontology, we are interested to explore if the graph representation can be useful as a querying tool in itself.

While the contributions of a graphical interface for ontology exploration are encouraging, the overall performance remains an issue. Querying an ontology is satisfactory fast but displaying the graph structure has significant memory requirements and may halt for large ontologies. Also, while mapping familiar keywords to ontology terms, many users reported a difficulty in specifying which ontology to query in. In the current prototype, querying for a keyword in the whole collection of ontologies is not supported and needs further attention.

Our results can be represented by the following conclusions:

1. The Ontology viewer provides an intuitive interface for (novice) users; the options are self explanatory and the user is assisted in understanding ontology concepts while at the same time ontologies are queried and terms selected. The mapping to the graphs is very helpful to that respect.
2. The MyTerms list provides a good simplification to the otherwise “oversized” ontologies. Users can now use ontology concepts with ease in their image annotations.
3. To assure visibility in the interface of the ontology viewer a fast response to queries is required which can be provided through a transformation of the ontology structure to an RDBMS.

FUTURE DIRECTIONS

The work presented in this paper is the result of a participatory design trajectory; in the design phase we aimed to learn how we could bring the concept of annotation of images with ontologies across the users of the CSIDx database. The design process also included requirement generation by users. We accomplished this design phase with an artifact that is a fully working prototype, rather than proposing a final application. Now that we have gained sufficient information on how novice users can work with ontologies, we can make next steps towards observatory evaluations in which different annotation strategies can be tested. Further user evaluations by surveys will render sufficient data for statistical analysis on ontology interaction.

Annotations are the basic components of the semantic structure in CSIDx. Furthermore, the relations included in the ontologies provide additional material to be explored. Initially, we wish to investigate the direct relations as maintained in the RDBMS. Still, mechanisms to profit from the OWL-DL expressiveness can be expanded based on the existing annotation with ontology concepts.

REFERENCES

1. Bei Y, Belmamoune M and Verbeek FJ. (2006) Ontology and image semantics in multimodal imaging: submission and retrieval. Proc. SPIE Internet Imaging VII, Vol. 6061, 60610C1-C12, 2006.
2. Bei, Y., Dmitrieva, J., Belmamoune, M., Verbeek, F.J. (2007) Ontology Driven Image Search Engine. Proc. SPIE Vol. 6506, MultiMedia Content Access: Algorithms & Systems (Eds Hanjalic, A., Schettini, R., Sebe, N.), 65060G-1,65060G-10
3. Cyttron Project, http://cyttron.org
4. Gardiner, T., Horrocks, I., and Tsarkov, D. (2006) Automated benchmarking of description logic reasoners. In Proc. of the 2006 Description Logic Workshop. Volume 189
5. Jena A Semantic Web Framework for Java, http://jena.sourceforge.net/
6. JUNG Java Universal Network/Graph Framework, http://jung.sourceforge.net/
7. Kallergi, A., Bei, Y., Kok., P., Dijkstra, J., Abrahams, J.P., Verbeek, F.J. (2008) Cyttron: A Virtualized Microscope supporting image integration and knowledge discovery. In: Cell Death and Disease Series, ResearchSignPost Eds.Backendorf,Noteborn,Tavassoli):Proteins Killing Tumour Cells (In Press)
8. KAON2, http://kaon2.semanticweb.org/
9. OBO Download Matrix, http://www.berkeleybop.org/ontologies/
10. OWL Web Ontology Language Overview, http://www.w3.org/TR/owl-features/
11. Pellet Reasoner, http://pellet.owldl.com/
12. RDF Resource Description Framework, http://www.w3.org/TR/rdf-syntax-grammar
13. SPARQL Query Language for RDF, http://www.w3.org/TR/rdf-sparql-query/

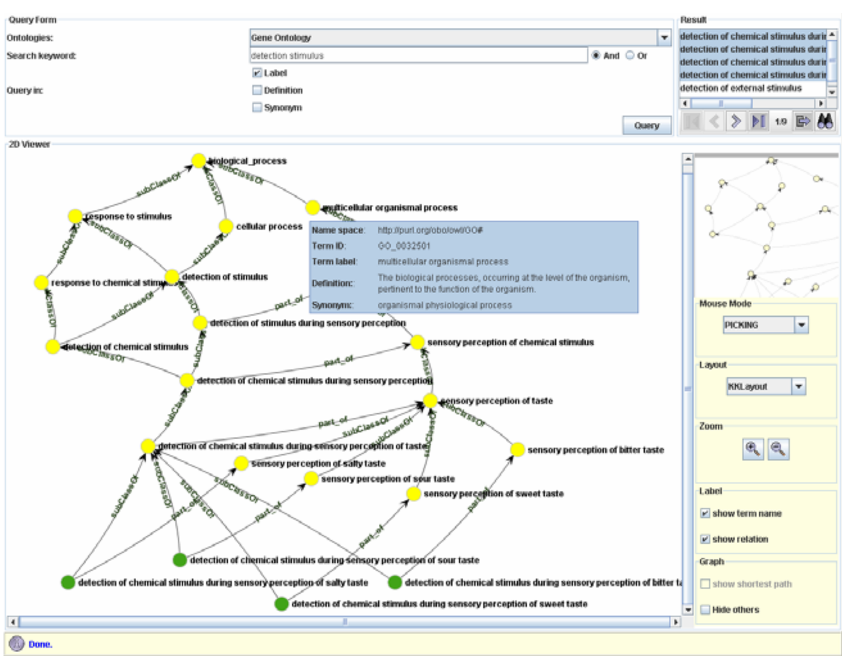

Figure 2. The ontology viewer with a KKLayout of the graph and highlighted the selected results