Using Ontologies for Medical Image Retrieval - An Experiment

Jasmin Opitz, Bijan Parsia, Ulrike Sattler

The University of Manchester
{opitzj|bparsia|sattler}@cs.manchester.ac.uk

1 Introduction

Medical research and clinical workflows often involve collaboration between various institutions. Therefore, ontologies such as SNOMED CT[1] or the Foundational Model of Anatomy (FMA)[2] have gained acceptance as an important tool for a common standard of communication. Medical images are often stored in large databases or file systems along with information including radiologists' reports, findings and visual features, some of which are expressed in natural language. The latter tends to be ambiguous and difficult to interpret by machines [6]. The disadvantages of this practice are restricted querying capabilities and poor image retrieval. For example, if we want to find images that show a certain phenomenon, e.g. a neoplasm, we might not find all relevant images because the search term is defined ambiguously in the report (e.g. the report contains the term "cancer" or "tumor" instead). It is also possible that we find images which are completely unrelated to the search term because the report mentions the term in a different context (e.g. "This image shows a pleural effusion in a patient that suffered from a lung tumor five years ago."). The first example describes the situation where the search results have a low recall (we do not find all relevant images), the latter describes a low precision (we find irrelevant images).

We believe that the use of ontologies can help to improve querying and retrieval of medical images and therefore make them more easily accessible. Previous work shows that there has already been a significant research effort in the field of ontology-based image annotation and retrieval [3,5]. However, many of these approaches use ontologies merely to formalise the terminology in the image descriptions or to support users in annotating and querying image collections and build tools to assist them with these tasks. We describe an experiment in which we translate the information in the natural language reports to class and property assertions of the SNOMED CT ontology, so that they reflect the natural language description. Our goal is to investigate whether and how it is possible to use an existing, sufficiently expressive medical ontology for expressing image related information in such a way that the expressivity of the ontology-based annotations and the ontology's reasoning services allow the user to pose more expressive queries and lead to a higher recall and precision compared to keyword-based querying or text-mining-based approaches.

[1] http://www.ihtsdo.org/snomed-ct/
[2] http://sig.biostr.washington.edu/projects/fm/

Proceedings of OWL: Experiences and Directions 2009 (OWLED 2009),
Rinke Hoekstra and Peter F. Patel-Schneider, editors. http://www.webont.org/owled/2009

2 Experiment

2.1 Sample images

We collect a sample of 50 medical images and their natural language descriptions from EURORAD,[3] a web-based radiological case database provided by the European Society of Radiology. These images are available for the public and are accompanied by radiologist reports. In the experiment we focus on chest radiology images because they involve a complex anatomical structure and various phenomena. The sample contains images of different image types and projections, such as X-ray (frontal, lateral), Computerised Tomography (axial and multi-planar reconstructions), bronchoscopy and angiographs. The images show various groups of findings, such as cancer, infectious diseases, embolisms and effusions. For each of the 50 images we use the image's natural language description from the image caption and the heading of the clinical case. We store this information for all the images in an XML file, which was later used for comparing the performance of natural language and ontology-based image retrieval.

2.2 Building the annotation ontology

We built our annotation ontology from SNOMED CT. On the one hand, the ontology is suitable because it contains all the relevant terms that appear in the images' natural language descriptions. Furthermore, it uses a language that is closely related to the description logic $\mathcal{EL}$ [1,2] and thus to the OWL 2 profile OWL EL.[4] As a consequence, we can use existing DL reasoners for query answering. The queries of our experiment have been answered by the reasoner FaCT++ [7].

The full SNOMED CT ontology is very large as it consists of approximately 370.000 classes and it is difficult to manipulate such large ontologies with tools such as ontology editors. For the experiment a small subset is sufficient, i.e. the parts which are relevant for chest anatomy, chest pathology and medical imaging. We therefore extract a module [4] from SNOMED CT. Furthermore, we add an additional part to the annotation ontology which is specific for the task of image annotation and contains object properties such as *showsImageType*, *showsProjection* and *showsFinding*. We also create object properties that can be used to describe features of the image. Images can show an image type and a projection, e.g. "X-ray" and "lateral projection". An image also shows findings and body structures. SNOMED CT models clinical findings by using the *roleGroup* construct. *RoleGroup* is actually an object property that is used to describe disorders in SNOMED CT. In order to represent findings and their locations, a disorder can have a property *roleGroup* that splits into a conjunction of the two object properties *AssociatedMorphology* and *FindingSite*'. It is crucial to

[3] http://www.eurorad.org/

[4] http://www.w3.org/TR/owl2-profiles/

use the *roleGroup* construct in the annotations in order to model disorders and morphologic abnormalities in a consistent way. Furthermore, findings can have qualifier values, such as *large*, *solid*, *lobular* and *circumferential*. There are cases in which findings can be related to other findings. On the one hand, some images show metastases that spread from a primary tumor. For these cases we use the relation *derivingFrom*, e.g. in "chest wall metastasis *derivingFrom* carcinoma of lung". On the other hand, there are images that show diagnostic errors, e.g. the finding "Bronchocentric Granulomatosis" which appears to be a "carcinoma in situ of bronchus". We express such cases with the property *presentingAs*. The SNOMED CT module together with this set of object properties builds the TBox of the annotation ontology.

2.3 Image annotation

The actual image annotations form the ABox of the annotation ontology. For each image, class and object property assertions are added to the ontology. A natural language annotation usually contains several concepts, such as image type, findings, qualifier values and body structures, which stand in relation with each other. A typical example would be:

```
<image no="4005_01">
    <case no="4005" />
    <headline>
        Chest wall metastasis deriving from a carcinoma of the lung
    </headline>
    <type>Conventional X-ray of the thorax</type>
    <caption>
        At the level of the lower lobe (right lung) a well circumscribed
        tumour is visible.
    </caption>
</image>
```

For all medical concepts that appear in such a description we add instances of the respective SNOMED CT classes to the ontology. The individuals belonging to one image annotation are then linked with the appropriate object properties so that they build a structure that reflects the natural language annotation. Using class assertions as well as property assertions is the crucial part to capture the semantics of the textual descriptions. Figure 1 shows the ontology-based annotation for the image described above.

At the end of the annotation process we assembled an ontology with a TBox that represents general knowledge we have about features shown in the sample images and an ABox which consists of class and object property assertions that represent the image annotations.

2.4 Querying

The annotation ontology we constructed allows us to formulate an information request as a DL query. Hence, we can pose complex and precise queries in accordance with the image annotations and achieve higher recall and precision compared to keyword-based or text-mining-based querying. We think that we

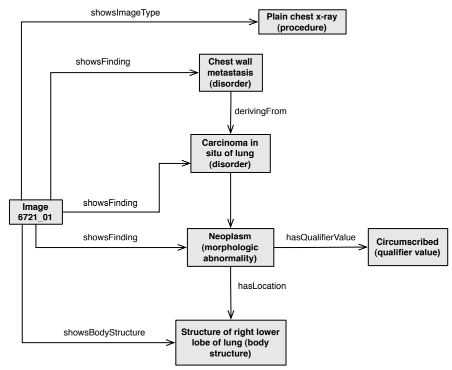

Fig. 1. An Example of an Ontology-based Image Annotation.

can achieve improvements in the following categories.

Semantic-based Query Answering
The way the ontology-based annotations have been constructed enables us to overcome differences in formulating concepts with the same semantics. In the original image annotations there have been various ways to express

- a concept (e.g. "left lower lobe of lung")
- the location of a finding (e.g. "calcific micronodules scattered throughout the lungs", "endobronchial lesion obstructing the right inferior bronchus")
- the behaviour of a finding (*derivingFrom*, *presentingAs*)

We can use the ontology to model the semantics of the image annotations in a consistent way. We expect that this will lead to higher recall and precision compared to retrieval methods that do not take synonyms or differences in phrasing terms into account, such as keyword-based retrieval.

Furthermore, findings can often be summarised under a more general finding or body structures are part of other body structures. It is possible to ask for all images that show a finding "neoplasm". For natural language annotations and keyword-based retrieval we would expect only to find such images which annotation contains the word "neoplasm". Ontology-based retrieval also involves the capabilities of thesaurus-based retrieval and takes the ontology's class hierarchy into account. The query would therefore return all images with annotations that contain the term *neoplasm* as well as other terms that are subclasses of *neoplasm*, such as *sarcoma* or *carcinoma*.

Nested Queries
In natural language effective query answering is very difficult because the expressions are phrased in various different ways. The ontology-based image annotations reflect a high level of detail using properties and conjunctions. We can generate a query in the same detailed way we would create an annotation, e.g.

by specifying image type, findings and body structures and how these terms relate to each other and therefore specify the search very precisely. This is also interesting for finding similar images because we can use an image's annotation or part of it as a query. We expect that complex and nested queries help us to increase the precision of search results.

In the experiment we compare ontology-based image retrieval to keyword-based retrieval. It has to be considered that it is not possible to pose a query in exactly the same way for both retrieval scenarios. With natural language descriptions we only do a simple keyword-matching search, whereas we can construct complex DL-based queries for the ontology-based annotations. In order to compare the search and retrieval capabilities for both kinds of image annotations, we create a set of representative queries that reflect the groups mentioned above. For each of these search scenarios we created a DL query for the image set with ontology-based annotations and a keyword query for the image set with natural language descriptions, which is in fact a conjunction of keywords. Since the experiment is based on a relatively small number of sample images, we can determine the optimal result of relevant images manually for each query beforehand. Then we pose the queries and note the images that were returned for each image set. These results are then compared with the optimal result in order to measure recall and precision.

3 Evaluation

3.1 Results

Table 1 shows a representative set of queries we generated for both kinds of image descriptions as well as their results. These queries are illustrative and should show conceptional types of search requests. Two different versions of the query are shown: a DL query in Manchester syntax for the ontology-based annotations and a keyword query for the natural language descriptions. The reasoner we used to answer the DL queries is FaCT++. Moreover, the table shows the optimal number of relevant images we would expect to be retrieved for each query as well as the actual search results. For both DL queries and keyword queries we differentiate how many of the retrieved images were relevant and irrelevant. The optimal number of images is determined in a way that we expect to find all members of a class, including members of subclasses.

3.2 Recall

As expected, the results confirmed that the images with ontology-based annotations reach an optimal level of recall. On the one hand this is due to the fact that the keyword queries cannot detect synonyms, such as *neoplasm* and *tumor*. The main reason, however, is that the image collection can be queried with respect to the class hierarchy and can therefore also find images which are subclasses of the classes specified in the query. This becomes very clear in the first example query, where we ask for all images that contain a finding *neoplasm*. Whereas the keyword query only find images with annotations that contain the word "neoplasm", the DL query also finds images with annotations that contain subclasses of *neoplasm*, such as *solitary fibrous tumor*, *leiomyosarcoma* or *renal cell carcinoma*. The same applies to query 2, where the natural language annotations contain various versions of expressing a *lesion of bronchus*, e.g. "endobronchial lesion", and cannot use the class hierarchy information to detect *metastasis* and *neoplasm* as subclasses of *lesion*.

In query 3, 5 and 6 the keyword-based queries have a low recall because of the variations in expressing the lung structure in natural language. Often these descriptions do not even contain the word "lung", but speak of "the right upper lobe" or other specific parts of the lung. Furthermore, the description of "upper" and "lower" vary. These terms tend to be equalised with "superior" and "inferior" or instead of speaking of a "lower lobe", radiologists often use phrases like "lung bases". We have encountered many more of such inconsistencies in expressing body structures in natural language, e.g. when referring to the bronchial structure, the pleura or the cardiovascular structure.

3.3 Precision

The example queries in Table 1 confirm our assumptions about the precision of the search results. Query 3 gives an example of the benefits we can achieve by defining explicit semantics for the annotations. We queried both image collections for images that show a "carcinoma of lung". The keyword query returned two images that we did not find with the DL query. These images have initially been diagnosed incorrectly with "carcinoma of lung", although they actually show the finding "bronchocentric granulomatosis". In the ontology-based annotations we avoided this by expressing it with the *presentingAs* property, so that the annotation does not state "carcinoma of lung" as a finding. Another example for an increase in precision is query 4. The keyword search for "X-ray images with lateral projection" returned an image that has the image type X-ray, but not a lateral projection. It was returned because the report contained the phrase "lateral wall of the aorta". This is another typical example where explicit semantics of the annotations can prevent us from finding images in which the search terms are used in a completely different context.

3.4 Complex queries

All of the example DL queries use properties to define the search objects. Whereas the first three example queries are relatively simple and just ask for images that show particular findings, the last three queries have a deeper nesting. The more properties we use and the deeper the nesting is, the more precise is the description of the desired results and it is therefore easier to isolate the relevant images from the image collection. This is mainly reflected in the increased precision, as mentioned above. Furthermore, using properties enables us to go beyond searching for the occurrence of keywords and define the semantics of our search. This is an important precondition if we want to search for similar images or make a broader or narrower request.

Table 1. Example Queries

Query				Results			
	DL	Keyword	Optimal	DL		Keyword	
				Re.	Irrel.	Rel.	Irrel.
1	'Disease (disorder)' and roleGroup some ('Associated morphology (attribute)' some 'Neoplasm (morphologic abnormality)' and 'Finding site (attribute)' some 'Body structure (body structure)')	neoplasm	25	25	0	5	0
2	'Disease (disorder)' and roleGroup some ('Associated morphology (attribute)' some 'Lesion of bronchus (finding)' and 'Finding site (attribute)' some 'Body structure (body structure)')	lesion, bronchus	7	7	0	1	0
3	'Disease (disorder)' and roleGroup some ('Associated morphology (attribute)' some 'Carcinoma in situ of lung (disorder)' and 'Finding site (attribute)' some 'Body structure (body structure)')	carcinoma, lung	3	3	0	3	2
4	Image and showsImageType some 'Plain chest X-ray (procedure)' and showsProjection some 'Lateral projection (qualifier value)'	X-ray, lateral	1	1	0	1	1
5	'Disease (disorder)' and roleGroup some ('Associated morphology (attribute)' some 'Mass (morphologic abnormality)' and 'Finding site (attribute)' some 'Structure of right lower lobe of lung (body structure)')	mass, right, lower, lobe, lung	6	6	0	1	0
6	Image and showsImageType some 'Computerized axial tomography (procedure)' and ('Disease (disorder)' and roleGroup some ('Associated morphology (attribute)' some 'Mass (morphologic abnormality)' and 'Finding site (attribute)' some 'Lung structure (body structure)'))	CT, computerized tomography, axial, mass, lung	10	10	0	2	0

4 Conclusion

We carried out an experiment in order to show that we can improve the accessibility of medical images by using ontology-based annotations to define image related information in addition to natural language descriptions. We took a sample of chest radiology images and their natural language reports and translated these reports to assertions of an ontology based on a subset of SNOMED CT. In order to evaluate the querying and retrieval performance of ontology-based image annotations and compare it with the natural language reports, we queried both kinds of image descriptions and measured recall and precision. We found out that we can achieve optimal recall results with ontology-based annotations because we can define the features of the images unambiguously and therefore overcome difficulties with synonyms and phrasing. Moreover, we can use the underlying class hierarchy and definitions in order to find relevant images whose annotations do not directly match the query, but can be inferred to do so by a reasoner. Ontology-based retrieval also showed improved precision compared to keyword-based retrieval. This is due to the fact that we can construct complex and detailed queries which can describe the desired images much more precisely. Furthermore, we can make use of the semantics of the annotations in order to ignore images that show the specified classes in a different context. The experiment has shown that it is possible to pose more expressive queries and achieve very good results on a relatively small collection of images. These benefits would have an even greater affect on a large collection of images, where natural-language-based querying typically cannot retrieve relevant images effectively. By using SNOMED CT we could show that we can re-use an existing ontology that is established in the medical domain.

The experiment also gave an insight in the feasibility of ontology-based image annotation. The mapping of natural language annotations to ontology-based annotations has been done manually for the sample images because we wanted to study the characteristics of the descriptions and the possible pitfalls in the mapping process. In fact, this helped us to understand how the ontology-based annotations have to be modelled in order to achieve full reasoning support.

References

1. F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence IJCAI-05*, Edinburgh, UK, 2005. Morgan-Kaufmann Publishers.
2. F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope further. In K. Clark and P. F. Patel-Schneider, editors, *In Proceedings of the OWLED 2008 DC Workshop on OWL: Experiences and Directions*, 2008.
3. B. Hu, S. Dasmahapatra, P. Lewis, and N. Shadbolt. Ontology-based medical image annotation with description logics. *IEEE International Conference on Tools with Artificial Intelligence*, 0:77, 2003.
4. E. Jiménez-Ruiz, B. C. Grau, U. Sattler, T. Schneider, and R. B. Llavori. Safe and Economic Re-Use of Ontologies: A Logic-Based Methodology and Tool Support. In S. Bechhofer, M. Hauswirth, J. Hoffmann, and M. Koubarakis, editors, *ESWC*, volume 5021 of *Lecture Notes in Computer Science*, pages 185–199. Springer, 2008.
5. D. L. Rubin, P. Mongkolwat, V. Kleper, K. Supekar, and D. S. Channin. Annotation and image markup: Accessing and interoperating with the semantic content in medical imaging. *IEEE Intelligent Systems*, 24(1):57–65, 2009.
6. I. Spasic, S. Ananiadou, J. McNaught, and A. Kumar. Text mining and ontologies in biomedicine: Making sense of raw text. *Briefings in Bioinformatics*, 6(3):239–251, September 2005.
7. D. Tsarkov and I. Horrocks. FaCT++ Description Logic Reasoner: System Description. In U. Furbach and N. Shankar, editors, *IJCAR*, volume 4130 of *Lecture Notes in Computer Science*, pages 292–297. Springer, 2006.