Building and Querying RDF/OWL Database of Semantically Annotated Nuclear Medicine Images

Kyung Hoon Hwang[1] · Haejun Lee[1] · Geon Koh[1] · Debra Willrett[2] ·
Daniel L. Rubin[2,3]

Abstract As the use of positron emission tomography-computed tomography (PET-CT) has increased rapidly, there is a need to retrieve relevant medical images that can assist image interpretation. However, the images themselves lack the explicit information needed for query. We constructed a semantically structured database of nuclear medicine images using the Annotation and Image Markup (AIM) format and evaluated the ability the AIM annotations to improve image search. We created AIM annotation templates specific to the nuclear medicine domain and used them to annotate 100 nuclear medicine PET-CT studies in AIM format using controlled vocabulary. We evaluated image retrieval from 20 specific clinical queries. As the gold standard, two nuclear medicine physicians manually retrieved the relevant images from the image database using free text search of radiology reports for the same queries. We compared query results with the manually retrieved results obtained by the physicians. The query performance indicated a 98 % recall for simple queries and a 89 % recall for complex queries. In total, the queries provided 95 % (75 of 79 images) recall, 100 % precision, and an F1 score of 0.97 for the 20 clinical queries. Three of the four images missed by the queries required reasoning for successful retrieval. Nuclear medicine images augmented using

semantic annotations in AIM enabled high recall and precision for simple queries, helping physicians to retrieve the relevant images. Further study using a larger data set and the implementation of an inference engine may improve query results for more complex queries.

Keywords Image retrieval · Nuclear medicine · PET · Controlled vocabulary · Protégé · AIM · ePAD

Introduction

As the use of nuclear medicine imaging modalities such as positron emission tomography-computed tomography (PET-CT) has increased quickly, rapid and accurate retrieval of similar images may provide a component of decision support in the interpretation of PET-CT images. There are two approaches to searching for similar images: text-based and content-based methods [1]. In text-based searches, the images are retrieved by finding matching text strings in text descriptions associated with the images (e.g., in the DICOM header or in figure captions). The interpretations of medical images are usually recorded in free text; hence, text search could be an attractive option for finding similar images. However, to enable text search, natural language processing is needed to extract the relevant information from free text. In addition, if image abnormalities are not described in sufficient detail, the value of searching radiology reports could be limited.

The second approach to searching for similar images is content-based image retrieval (CBIR). In CBIR, the images are retrieved on the basis of features found in radiology reports or in the images themselves such as color, shape, and texture. Commercial CBIR systems have been developed, such as QBIC [2], Photobook [3], Virage [4], VisualSEEK [5], and

✉ Daniel L. Rubin
dlrubin@stanford.edu

[1] Department of Nuclear Medicine, Gachon University Gil Medical Center, Incheon, South Korea

[2] Department of Radiology, Stanford University, Stanford, CA, USA

[3] Department of Medicine (Biomedical Informatics Research), Stanford University, Stanford, CA, USA

Netra [6]. Eakins divided image features used in CBIR into three levels [7]:

- Level 1. Primitive features, such as color, texture, shape, or the spatial location of image elements. A typical query example is "find pictures like this."
- Level 2. Derived attributes or logical features involving some degree of inference about the identity of the objects depicted in the image. A typical query example is "find a picture of a flower."
- Level 3. Abstract attributes involving complex reasoning about the significance of the objects or scenes depicted. A typical query example is "find pictures of a beautiful lady."

The majority of CBIR systems offer mostly level 1 retrieval, while a few experimental systems provide level 2 retrieval. None provide level 3 retrieval so far.

Prior CBIR systems characterized images using global features such as a color histogram, texture values, and a shape parameter. However, for medical images, systems using global image features fail to capture the relevant information [8]. In medical images, the clinically useful information is usually highly localized in small areas of the images, i.e., the ratio of pathology-bearing pixels relative to the rest of the images is small. Thus, in the case of medical images, global features, such as color, texture, and shape, often cannot effectively characterize the content required for clinically relevant searches (e.g., "find images that contain a liver lesion that looks like this one").

Recently, semantically structured text-based databases for medical records that use eXtensible Markup Language (XML) [9], Resource Description Framework (RDF) [10], Web Ontology Language (OWL) [11], and controlled vocabulary to encode the data have been introduced. With the RDF, the knowledge pertaining to medical image interpretation can be stored as a set of triples, where each element of the triple can be referenced by an explicit uniform resource identifier (URI) to which resources can be linked. By linking to well-defined medical terminologies, such as RadLex [12–14], Foundational Model of Anatomy (FMA) [15], and SNOMED CT [16, 17], an RDF-based approach can explicitly refer to a formalized set of concepts. With query languages that make use of triple-based patterns, such as the Simple Protocol and RDF Query Language (SPARQL), it is possible to create detailed queries for specific information.

The Annotation and Image Markup (AIM) project has developed methods to capture the semantic meaning of medical images and calculations of pixel data with graphical drawings placed on the image, thus providing a semantic infrastructure for the images [18]. AIM not only describes the semantic content in images using ontologies but also provides interchangeable encoding in DICOM-SR, XML, HL7 CDA, and OWL [19, 20].

A recently developed and freely available software platform, the electronic Physician Annotation Device (ePAD) [21, 22] facilitates the recording of image measurements and annotations in the AIM format using controlled vocabularies such as RadLex and FMA. In this work, we built on the inherent information structure in AIM to construct a semantically structured database of nuclear medicine image reports. We queried this database to evaluate if AIM-encoded annotations improve the retrieval rate of nuclear medicine images. The level of improvement was assessed by comparing the retrieval efficiency of AIM-encoded annotations by ePAD with that of manual searches by physicians.

Materials and Methods

Image Data

This study was approved by the Institutional Review Board and written consent was waived. Retrospectively, 100 nuclear medicine PET-CT studies (sets of images) and associated radiology reports were selected. The metadata of the PET-CT images and their associated interpretations were anonymized.

Building an RDF/OWL Database of Nuclear Medicine Images

We created AIM annotation templates [23] specific to the nuclear medicine domain (especially for oncologic PET-CT) for annotation of PET-CT images on the ePAD platform. These templates were created to input information on lesions, such as anatomic location, number, size, kind of lesion, metabolism, description of interval change, and measured parameters, according to controlled vocabularies and aimed at easy recording of regional lesions (for example, thoracic lymph node lesions on PET-CT). We imported PET-CT images and PET-CT AIM annotation templates into ePAD. We used these templates to annotate 100 PET-CT studies in AIM format (Fig. 1) using controlled vocabulary (RadLex, Foundational Model of Anatomy, and SNOMED CT). PET-CT annotations recorded by the ePAD platform [24] were converted to RDF/OWL. The hierarchy and link to RDF/OWL data were viewed and managed through Protégé (Fig. 2). The PET-CT annotations were converted from XML to RDF/OWL format using an SAX parser programmed in JAVA, which collected the information items to be stored and converted them to an RDF/OWL file. An ontological knowledge hierarchy of oncological and nuclear medicine controlled vocabulary was also stored in the database.
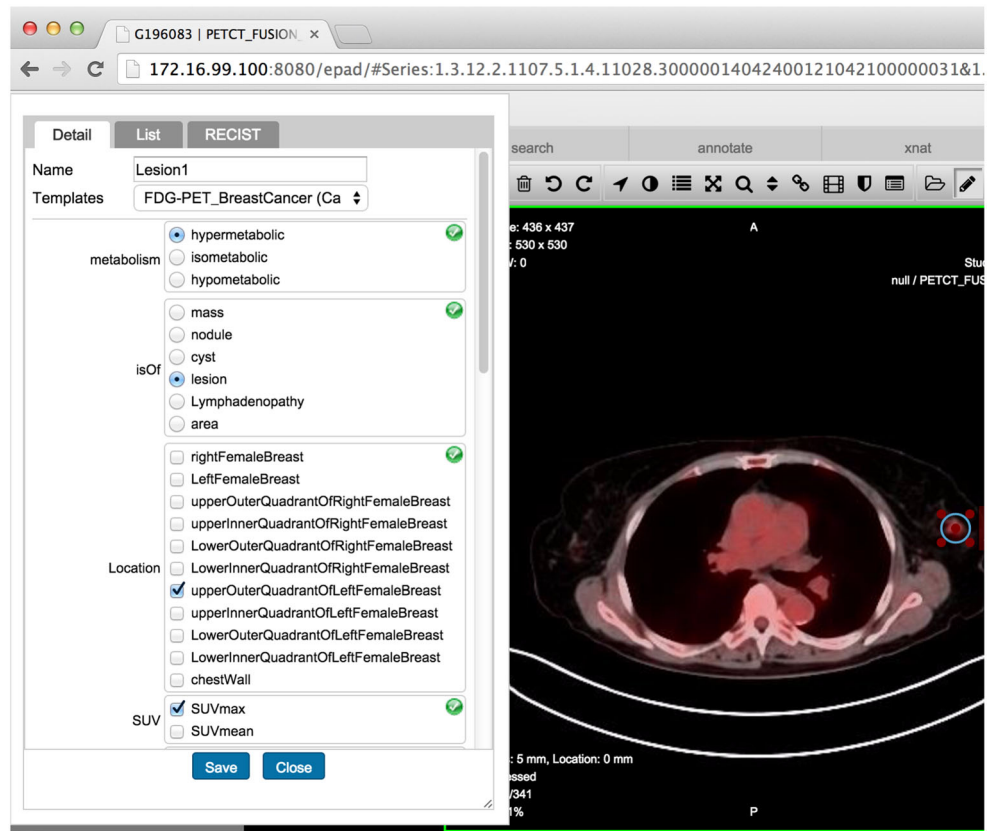
Querying, Gold Standard, and Evaluation

We created 20 clinically relevant image feature queries
(Appendix) that are useful to nuclear medicine physicians yet
difficult to perform with text-based queries. We used these

queries to evaluate image retrieval from the constructed RDF/
OWL database of 100 nuclear medicine PET-CT imaging stud-
ies. The 20 query items were selected from common queries
required by nuclear physicians but are difficult to perform on
image annotation database in free text format. The queries

Fig. 2 View of the hierarchy and
links of RDF/OWL PET-CT
annotation data on the Protégé
Ontology Tool. The *left column*
shows the RDF/OWL data
hierarchy and the *right bottom
row* demonstrates the linked class
or instance
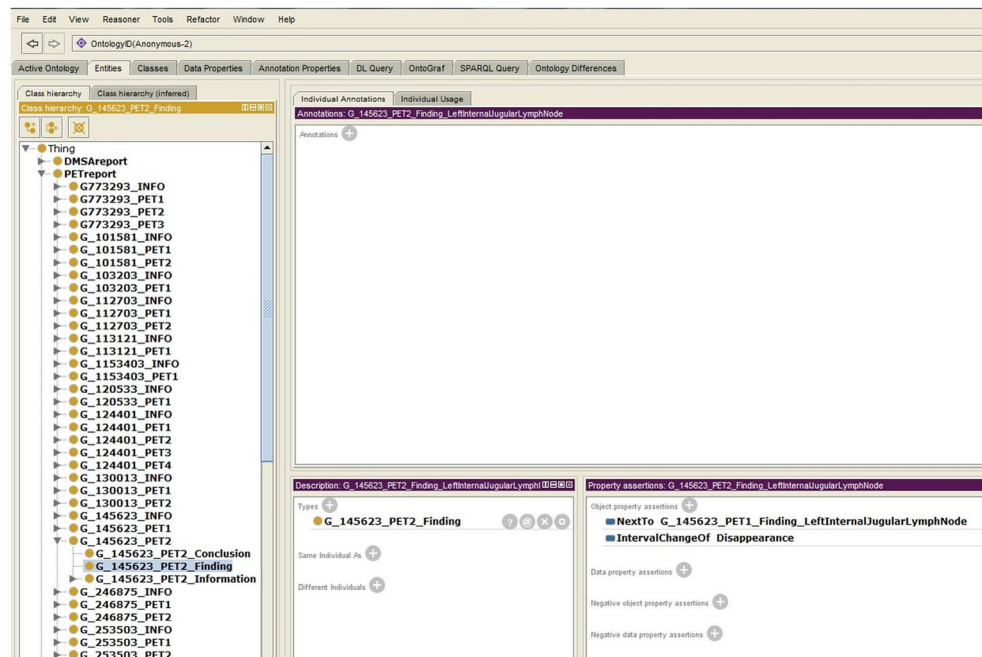
covered a variety of PET-CT imaging scenarios, such as the selection of PET-CT images under specific conditions, criteria, and even staging. The 20 queries were classified as simple or complex based on the complexity of the "where" clause. A query such as "retrieve PET-CT studies where lesions in the gallbladder have maximum SUV greater than or equal to 3.0" was classified as a simple double query because the "where" clause of the query consisted of only two filter functions (maximum SUV and location). A total of 11 simple queries were included, consisting of simple single queries (two queries), simple double queries (four queries), and simple temporal queries (five queries). The rest were complex queries, consisting of complex temporal queries (six queries; e.g., clinical guideline—for example, the Fleischner criteria), and complex reasoning queries (three queries) such as staging. One example of complex query is "Retrieve the PET-CT study containing the lymph node lesion, which showed no interval change for more than 2 years" because the where clause of the query included a relatively complex date logic as well as a simple filter function of location. The queries were performed using the SPARQL plug-in on the Protégé platform (Fig. 3). As the reference standard, two nuclear medicine physicians (HL and GK) reviewed all the images, for the same queries, and manually retrieved the relevant images by searching the free text format database of associated radiology reports. We compared the SPARQL query results with the manually retrieved results and calculated recall, precision, and F-measure.

Results

Compared to the manual query results (as the gold standard) by the two nuclear physicians, the SPARQL queries on the RDF/OWL database of the PET-CT images annotated by ePAD demonstrated a 98 % recall, 100 % precision, and an F1-score of 0.99 for simple queries (Table 1). For simple single queries, the SPARQL query results showed 96 % recall (22 of 23 images). The missed image was a nodule in the upper lobe of the right lung, recorded as the right lung in the radiology report in free text and in the RDF/OWL file. This image was retrieved by the two physicians using inference from the image slice number. For the simple double and simple temporal SPARQL queries, both query results were 100 % recall (11/11 and 19/19 images, respectively).

Fig. 3 An example of a SPARQL query (**a**) and the images returned by the query (**b**). The text of the query was "Retrieve the PET-CT study with the peritoneal space lesion showing a sequentially decreased FDG uptake (SUVmax) of more than 20 %." This change is apparent in the retrieved images (shown with *arrows*)
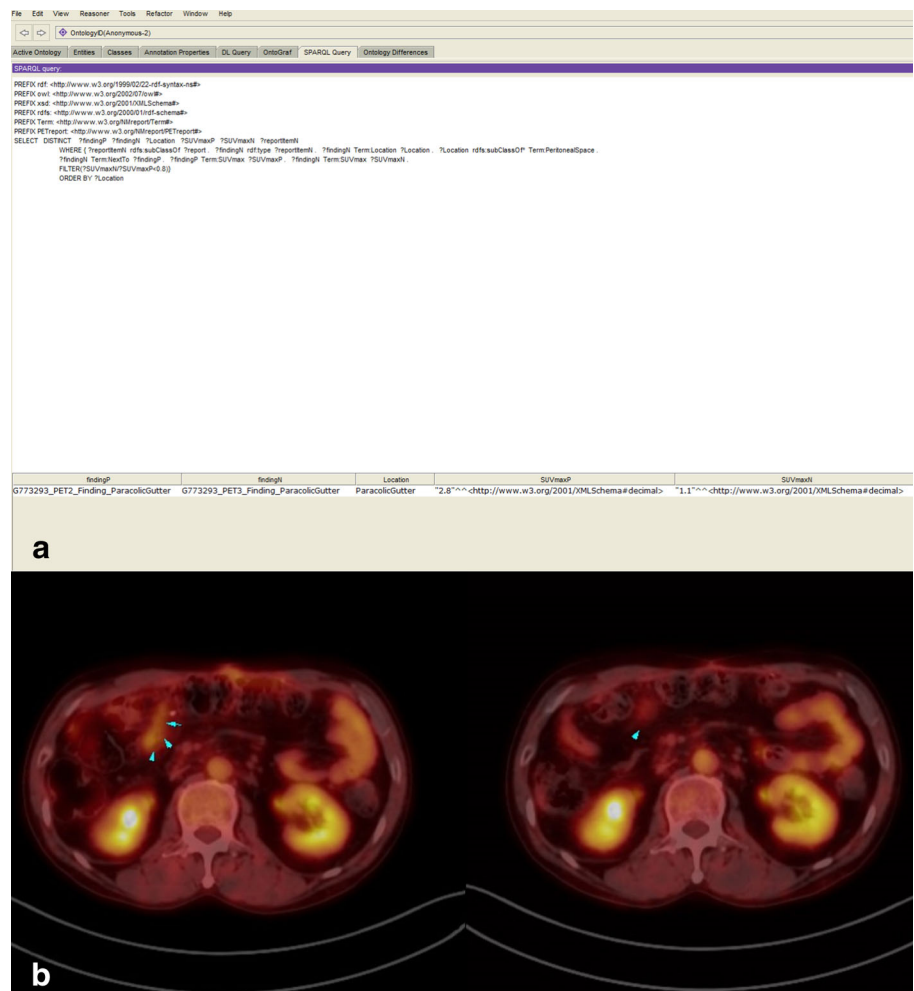
Table 1 SPARQL image retrieval results obtained by querying the RDF/OWL PET-CT database

	SPARQL-retrieved images	Physician-retrieved images	Recall (%)	F1
Simple queries	52	53	98.1	0.990
Simple single queries	22	23	95.6	0.978
Simple double queries	11	11	100	1.0
Simple temporal queries	19	19	100	1.0
Complex queries	23	26	88.5	0.942
Complex temporal queries	19	19	100	1.0
Complex reasoning queries	4	7	57.1	0.727
All SPARQL queries	75	79	94.9	0.974

For complex queries, the SPARQL queries demonstrated 89 % recall, 100 % precision, and an F1-score of 0.94 (Table 1). The SPARQL query results showed 100 % recall for the complex temporal queries (19/19 images; e.g., clinical guidelines). For the three complex reasoning queries; the SPARQL queries missed three PET-CT images, resulting in only 57 % recall (4/7 images).

In total, the SPARQL queries yielded 95 % (75 of 79 images) recall, 100 % precision, and an F1-score of 0.97 for 20 clinical query items. Three of the four images missed by the SPARQL queries required reasoning (Table 1).

Discussion

RDF provides a simple declarative data model of triples (subject, predicate, and object) to describe resources. OWL is an ontology language for the Semantic Web, extending the RDF and RDF Schema; it became a World Wide Web Consortium (W3C) recommendation in 2004. The amount of data encoded using RDF/OWL has increased in diverse areas of application, such as social networks, geographic locations, books, films, and bioinformatics. SPARQL is a query language for RDF/OWL data and the official W3C recommendation for the semantic web [25]. SPARQL provides an efficient way of accessing a variety of RDF/OWL data sources, and it is expected that an increasing number of content providers will make their data available for SPARQL query. To date, however, these semantic web technologies have proliferated largely outside of the medical domain. In addition, their use in radiology is novel, to our knowledge.

RDF/OWL databases with SPARQL query systems were applied in the medical domain relatively recently. They have shown favorable performance for mining drug-drug interactions [26] and in data integration between clinical and research data [27]. RDF/OWL and SPARQL query systems may have value in searching typically large, unstructured medical imaging database.

We developed methods to convert AIM annotations produced using the ePAD platform into RDF/OWL data. ePAD enables interpreters to record image annotations on a lesion-by-lesion basis. AIM annotations link the information about a lesion to the image on which it appears. This enables the creation of a structured medical imaging database, which can be queried for useful information related to the images, such as finding similar images or evaluating the temporal change in a lesion in sequential follow-up images. This can be useful when evaluating radiologic criteria such as the Fleischner criteria [28], which are composed of conditions based on temporal change in lesions.

Compared to manual retrieval by physicians, our RDF/OWL database with a SPARQL query system showed very high recall and precision in simple and complex queries except for queries requiring reasoning. In complex temporal queries such as radiologic criteria, our system demonstrated a high accuracy compared with manual retrieval. This, together with SPARQL's flexibility in combining query conditions, suggests a great potential in application to other medical data, such as electronic medical records. As our database contained only PET-CT imaging data from 100 studies, we could not test a wide variety of radiologic or clinical criteria. However, we expect much more useful querying if more clinical data such as electronic medical records are available.

Three out of the four images missed by our system were in response to queries requiring reasoning. This is not surprising since we did not design our system to resolve queries requiring high-level reasoning, except for anatomical reasoning. Future work could address this gap by adding a computerized inference engine to our system. The inclusion of inference would require an ontology with detailed concept terms and relationships to provide the semantic data structure needed to make inferences with query terms (e.g., to perform query expansion).

Our study has some limitations. First, we used only PET-CT image data from the oncology domain. We tested only a limited set of queries, although we expect that other queries based on the structured data captured in AIM will provide similarly good results. Secondly, the physicians used not only text-based image retrieval, but also context, and semantic understanding. We could not eliminate the impact of these other confounders from the

physicians enrolled in the study. Finally, our data set consists of only 100 studies. However, we believe this number of studies is reasonable to assess our preliminary results. Future additional studies on larger data sets would be helpful.

Conclusions

An RDF/OWL database of PET-CT images annotated using AIM showed very high recall and precision for simple image feature queries using SPARQL. These queries help physicians to retrieve relevant images for comparison and decision support. Further studies using larger data sets and including an implementation of inference may improve image query performance.

Compliance with Ethical Standards   This study was approved by the Institutional Review Board and written consent was waived.

Appendix

SPARQL Query

Simple Single Query

(1) Retrieve PET-CT studies containing lesions with maximum SUV greater than 10.0.
(2) Retrieve PET-CT studies containing lesions in the upper lobe of the right lung.

Simple Double Query

(3) Retrieve PET-CT studies where lesions in the laryngopharynx have maximum SUV greater than or equal to 3.0.
(4) Retrieve PET-CT studies where lesions in the gallbladder have maximum SUV greater than or equal to 3.0.
(5) Retrieve PET-CT studies where lesions in the lower lobe of the left lung have maximum SUV greater than or equal to 3.0.
(6) Retrieve PET-CT studies where lesions in the cavitated organ have maximum SUV greater than or equal to 3.0.

Simple Temporal Query

(7) Retrieve PET-CT studies where lesions are in the lymph nodes and maximum SUV decreased by more than 20 % in the next PET-CT studies.
(8) Retrieve PET-CT studies containing mild hypermetabolic lesions that disappeared in the next PET-CT studies.
(9) Retrieve PET-CT studies containing moderate hypermetabolic lesions that disappeared in the next PET-CT studies.
(10) Retrieve PET-CT studies containing severe hypermetabolic lesions that disappeared in the next PET-CT studies.

(11) Retrieve PET-CT studies containing lesions with maximum SUV greater than or equal to 3.0 and showing no interval change in the next PET-CT studies.

Complex Temporal Query (Including Radiologic Criteria)

(12) Retrieve PET-CT studies with lymph node lesions that showed no interval change for more than 2 years.
(13) Retrieve PET-CT studies with lung lesions that showed no interval change for more than 3 years.
(14) Retrieve PET-CT studies with thyroid lesions that showed no interval change for more than 1 year.
(15) Retrieve PET-CT studies with thyroid or lymph node lesions that showed no interval change for more than 2 years.
(16) Retrieve PET-CT studies with stomach lesions that showed no interval change for more than 1 year since the recommendation of an endoscopy.
(17) Retrieve PET-CT studies containing lung nodules sized 0.4 ~ 0.6 cm in diameter that showed no interval change for more than 6 months (the Fleischner criteria).

Complex Reasoning Query

(18) Retrieve PET-CT studies that demonstrated lymphoma in stage 3.
(19) Retrieve PET-CT studies that demonstrated gastric lymphoma in stage 2
(20) Retrieve PET-CT studies that show a potential of biliary obstruction.

References

1. Liu Y, Zhang D, Lu G, Ma W-Y: A survey of content-based image retrieval with high-level semantics. Pattern Recogn 40:262–282, 2007
2. Faloutsos C, Barber R, Flickner M, Hafner J, Niblack W, Petkovic D, Equitz W: Efficient and effective querying by image content. J Intell Inf Syst 3:231–262, 1994
3. Pentland A, Picard RW, Scaroff S: Photobook: content-based manipulation for image databases. Int J Comput Vis 18:233–254, 1996
4. Gupta A, Jain R: Visual information retrieval. Commun ACM 40:70–79, 1997
5. Smith JR, Chang SF: VisualSeek: a fully automatic content-based query system. Proceedings of the Fourth ACM International Conference on Multimedia, ACM Multimedia'96, Boston, MA, Nov 1996
6. Ma WY, Manjunath B: Netra: a toolbox for navigating large image databases. Proceedings of the IEEE International Conference on Image Processing 1:1997, pp 568–571
7. Eakins JP: Towards intelligent image retrieval. Pattern Recogn 35:3–14, 2002
8. Brodley C, Kak A, Shyu C, Dy J, Broderick L, Aisen AM: Content-based retrieval from medical image database: a synergy of human interaction, machine learning and computer vision. Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99), Orlando, FL, Jul 1999
9. Bray T, Paoli J, Sperberg-McQueen CM: Extensible Markup Language (XML) 1.0, W3C recommendation. Available at http://www.w3.org/TR/REC-xml. Accessed 15 September 2016

10. Lassila O, Swick RR: Resource Description Framework (RDF) Model and Syntax Specification, W3C recommendation. Available at http://www.w3.org/TR/PR-rdf-syntax. Accessed 15 September 2016

11. Bao J, Kendall EF, McGuinness DL, Patel-Schneider PF: OWL 2 Web Ontology Language Quick Reference Guide (Second Edition), W3C recommendation. Available at http://www.w3.org/TR/2012/REC-owl2-quick-reference-20121211/. Accessed 15 September 2016

12. Langlotz CP: RadLex: a new method for indexing online educational materials. Radiographics 26:1595–1597, 2006

13. Rubin DL: Creating and curating a terminology for radiology: ontology modeling and analysis. J Digit Imaging 21:355–362, 2008

14. Hong Y, Zhang J, Heilbrun ME, Kahn Jr, CE: Analysis of RadLex coverage and term co-occurrence in radiology reporting templates. J Digit Imaging 25:56–62, 2012

15. Rosse C, Mejino Jr, JL: A reference ontology for biomedical informatics: the foundational model of anatomy. J Biomed Inform 36:478–500, 2003

16. Sherter AL: Building a vocabulary. A new, improved version of SNOMED has the potential to ease the collection and analysis of clinical data. Health Data Manag 6:76–77, 1998

17. Nachimuthu SK, Lau LM: Practical issues in using SNOMED CT as a reference terminology. Stud Health Technol Inform 129:640–644, 2007

18. Mongkolwat P, Kleper V, Talbot S, Rubin DL: The National Cancer Informatics Program (NCIP) Annotation and Image Markup (AIM) foundation model. J Digit Imaging 27:692–701, 2014

19. Rubin DL, Mongkolwat P, Kleper V, Supekar K, Channin DS: Medical Imaging on the Semantic Web: Annotation and Image Markup, Association for the Advancement of Artificial Intelligence, 2008. Spring Symposium Series, Stanford, 2008

20. Channin DS, Mongkolwat P, Kleper V, Rubin DL: The annotation and image mark-up project. Radiology 253:590–592, 2009

21. Rubin DL, Rodriguez C, Shah P, Beaulieu C: iPad: Semantic annotation and markup of radiological images. AMIA Annu Symp Proc: 626–630, 2008

22. Channin DS, Mongkolwat P, Kleper V, Sepukar K, Rubin DL: The caBIG annotation and image Markup project. J Digit Imaging 23:217–225, 2010

23. Mongkolwat P, Channin DS, Kleper V, Rubin DL: Informatics in radiology: An open-source and open-access cancer biomedical informatics grid annotation and image markup template builder. Radiographics 32:1223–1232, 2012

24. Moreira DA, Hage C, Luque EF, Willrett D, Rubin DL: 3D markup of radiological images in ePAD, a web-based image annotation tool. IEEE 28th International Symposium on Computer-Based Medical Systems, Proc, 2015, pp 97–102

25. Prud'hommeaux E, Seaborne A: SPARQL Query Language for RDF. W3C Recommendation. Available at http://www.w3.org/TR/rdf-sparql-query. Accessed 25 September 2015

26. Pathak J, Kiefer RC, Chute CG: Using linked data for mining drug-drug interactions in electronic health records. Stud Health Technol Inform 192:682–686, 2013

27. Mate S, Kopcke F, Toddenroth D, Martin M, Prokosch H-U, Burkle T, Ganslandt T: Ontology-based data integration between clinical and research systems. PLoS ONE, 2015. doi:10.1371/journal.pone.0116656

28. MacMahon H, Austin JHM, Gamsu G, Herold CJ, Jett JR, Naidich DP, Patz EF, Swensen S: Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner society. Radiology 237:395–400, 2005