Spatial-symbolic Query Engine in Anatomy*

A. Puget; J. L. V. Mejino Jr.; L. T. Detwiler; J. D. Franklin; J. F. Brinkley
Structural Informatics Group, Department of Biological Structure, University of Washington, Seattle, WA, USA

Keywords
Spatial relations, anatomy, query processing, ontologies, semantic web

Summary

Objectives: Currently, the primary means for answering anatomical questions such as 'what vital organs would potentially be impacted by a bullet wound to the abdomen?' is to look them up in textbooks or to browse online sources. In this work we describe a semantic web service and spatial query processor that permits a user to graphically pose such questions as joined queries over separately defined spatial and symbolic knowledge sources.

Methods: Spatial relations (e.g. anterior) were defined by two anatomy experts, and based on a 3-D volume of labeled images of the thorax, all the labeled anatomical structures were queried to retrieve the target structures for every query structure and every spatial relation. A web user interface and a web service were designed to relate existing symbolic information from the Foundational Model of Anatomy ontology (FMA) with spatial information provided by the spatial query processor, and to permit users to select anatomical structures and define queries.

Results: We evaluated the accuracy of results returned by the queries, and since there is no independent gold standard, we used two anatomy experts' opinions as the gold standard for comparison. We asked the same experts to define the gold standard and to define the spatial relations. The F-measure for the overall evaluation is 0.90 for rater 1 and 0.56 for rater 2. The percentage of observed agreement is 99% and Cohen's kappa coefficient reaches 0.51. The main source of disagreement relates to issues with the labels used in the dataset, and not with the tool itself.

Conclusions: In its current state the system can be used as an end-user application but it is likely to be of most use as a framework for building end-user applications such as displaying the results as a 3-D anatomical scene. The system promises potential practical utility for obtaining and navigating spatial and symbolic data.

Correspondence to:
Antoine Puget c/o James F. Brinkley
Structural Informatics Group
Box 357420
Department of Biological Structure
University of Washington
Seattle, WA 98195
USA
E-mail: antoine.puget@gmail.com

Methods Inf Med 2012; 51: 463–478
doi: 10.3414/ME11-01-0047
received: May 31, 2011
accepted: December 18, 2011
prepublished: May 22, 2012

1. Introduction

Anatomy is fundamental to understanding in medicine, which is why it is one of the first courses taught to health sciences students. In medical school anatomy is taught as a single corpus that gives students a holistic understanding of the entire field. However, in practice only very specific anatomical knowledge is needed in any given situation. Examples include 1) which lobes of the lung are related to chest landmarks, 2) what vital organs would potentially be impacted by a bullet wound to the abdomen, 3) which lymph nodes are likely to be affected by a nearby metastatic tumor, or 4) which vital organs near a tumor can be affected by a planned radiation treatment.

Although the answers to many of these questions are known to experienced practitioners, they are not readily available or accessible to beginning practitioners, and even experienced practitioners require help at some point to refresh their memory. Currently, the primary means for answering these kinds of questions is to look them up in anatomy textbooks, atlases or online sources. However, this process not only can be time-consuming but also inefficient given the number of possible sources, sometimes conflicting or inconsistent, and the difficulty of finding specific answers in any given source.

The increasing availability of online digital anatomy sources raises the potential of acquiring more efficiently the answers to specific questions. Example sources include the Digital Anatomist atlases [1], the Harvard Whole Brain Atlas [2], the Visible Human project and its many derivatives [3], and many commercial products such as Voxelman [4], ADAM [5], and the Google Body Browser [6]. However, at the present time these online sources largely mirror the print media: they are not linked together, and each source generally presents information in the same manner as the print books and atlases, which makes them good for learning but not computationally efficient for generating queries and answers to specific questions.

What we envision is a web application that permits a user such as a clinician to pose queries in natural language format or with the aid of some kind of a graphical interface. The application would determine the most useful sources for answering the query, would pose the query to each source, and then assemble the results into a single natural language or visual answer. Such a system should also be accessible to other programs such as those

* Supplementary material published on our website www.methods-online.com

used by radiation treatment planners, so the knowledge captured can be used for reasoning.

To build such a system several issues must be addressed: 1) how to specify the queries, 2) how to distribute them to multiple sources, 3) how to query the sources, 4) how to combine the results into a single answer, and 5) how to discover relevant sources.

To date we are not aware of any system that addresses all these issues, and most existing systems address only #3: how to query anatomical resources. For example, in our own work we have developed several query interfaces to our Foundational Model of Anatomy (FMA) ontology [6]. These interfaces include a natural language interface [7], a graphical interface [8], a StruQL [9] query language interface [10] and a SparQL [11] (semantic web query language) interface [12].

However, anatomical information is available both from images, such as atlas images and X-ray images, and from symbolic representations such as those captured in textbooks or anatomy ontologies. To date the FMA deals only with symbolic descriptions, whereas in this paper we add the ability to deal with spatial relations between anatomical entities. Although spatial relations can be represented in symbolic form (e.g. "anterior" can be represented in the same way as "part_of") there is nevertheless an intuitive distinction between a relation like "anterior" and one like "part_of", in that the former is associated with some geometric frame of reference. Thus, for the purposes of this report we define a "spatial" relation as a particular type of relation that exists in a geometric frame of reference.

It is possible to enumerate and represent symbolically all spatial relations in an ontology but such a task would be extremely tedious and daunting. We therefore propose a system that automatically generates spatial relations from images and translates them into human readable and machine-parseable symbolic knowledge. We then combine this calculated information with the existing symbolic knowledge that already exists in the FMA. This allows us to answer questions such as "What organs are anterior to the heart?"

For this study we use geometric methods to derive spatial relations. In particular we take a set of labeled 2-D images of the thorax and convert them into 3-D volumes, project spatial query objects such as lines and planes into these volumes and then determine which structures are intersected by those projected objects. Another example of such an approach is the TraumaSCAN system [13], which is designed to help ER physicians predict the likely structures impacted by a bullet wound. However, we not aware of any such system generalized for widespread use.

A caveat with this geometric approach is that the spatial relations are computed based on images from a single human individual (instance), whereas presumably if the relations were manually entered into an ontology by an anatomy expert the spatial relations represent class level anatomical knowledge based on experience observing the relations in a population of many individuals.

In particular, we deal with three types of entities: 1) anatomical labels on 2-D images derived from one particular individual, 2) image data about this particular individual, and 3) the FMA ontology which provides annotation labels for the set of images. Therefore our work aims to describe the spatial relations in a dataset obtained from one individual and these spatial relations may in fact mirror real-world relations in some other individuals but we cannot generalize the results as canonical for the entire population because of expected structural variations among its members. In this study, when we ask the question "What structures are anterior to the esophagus?" we only refer to instance level spatial relations based on the image data from one single subject. In future work, if labeled images from multiple subjects are swapped in using our framework, it should be possible to compute statistics over multiple instances and then infer class level relations for automatic incorporation into ontology.

2. Objectives

In this paper we describe an approach that combines spatial queries to a 3-D labeled volume with symbolic queries to the FMA, within the context of the semantic web.

The framework we describe addresses all but the last issue noted in Section 1 (discovering anatomical sources). Although we describe a specific approach to spatial queries over a specific 3-D labeled image volume the system allows other spatial query methods and other 3-D labeled volumes to be swapped in. In addition, the combination of the FMA, labels from the FMA associated with the image volumes, and the semantic web, together demonstrate the potential for linking multiple resources in a distributed anatomy query system.

We begin by describing our working definition of spatial relations, recognizing that this is a complex topic, and that competing definitions are abundant and some of them may not totally agree with our definitions. The system we have designed is modular, which allows different definitions to be used. We then describe the use of this approach for a specific 3-D labeled image dataset, the use of semantic web technologies to integrate these spatial queries with symbolic queries to the FMA, a graphical user interface for generating queries, and an evaluation of its accuracy and response time.

3. Methods

3.1 Spatial Relations

In traditional anatomy, spatial relations are defined in terms of qualitative anatomical coordinates (e.g. anterior, posterior), which are based on the standard anatomical position:

"The anatomical position is defined as the erect position with the arms at the sides and palms of the hands facing forward. The median plane bisects the body into right and left halves. Any plane parallel with the median plane is a sagittal plane. Medial describes a position nearer to the median plane, and lateral describes a position further from it. A coronal plane is at right angles to the sagittal plane and bisects the body into anterior and posterior portions. The position of a structure nearer the front of the body is described as anterior and that nearer the back of the body as posterior…

A transverse plane is horizontal and bisects the body into upper and lower parts. Superior means lying above, and inferior, lying below" ([14], page 8).

In order to compute these relations from a labeled 3-D image dataset, the above qualitative definitions must be related to the quantitative coordinate systems defined by imaging systems. Fortunately, imaging coordinate systems generally align with the anatomical planes. That is, individual image slices are acquired along one of the three anatomical planes (sagittal, coronal or transverse). The most common slice orientation is transverse, in which the Z-axis is along the superior-inferior direction, the X-axis is along the medial-lateral direction, and the Y-axis is along the anterior-posterior direction.

Given these coordinate directions, as well as an image volume in which each voxel is labeled with an anatomical structure name, the question then becomes, for any given relation and any query anatomical structure represented in the image volume as a series of contiguous labeled voxels, what target anatomical structures stand in a given relation with the query structure? The answer is not generally straightforward because anatomical structures vary widely in size and shape and therefore it is likely that only portions and not the entirety of both the query structure and the target structures are spatially related. The more appropriate question to ask then is, "What target structure or structures predominantly stand in a given spatial relation to the query structure?" Note that the qualifier "predominant" in many (but not all) cases denotes a matter of opinion or preference depending on the purpose of the query and the system being used to compute spatial relations that a particular user wants to express in the application or program.

Our basic approach is the following: for each query structure and for each type of spatial relation, we define a 3-D spatial query volume projected from the query structure, compute the intersection of each possible target structure with the spatial query volume, and compute the ratio between the intersection volume and the overall target structure volume as a percentage. A target structure is then said to stand in a given spatial relation to the query structure if its parts or its entirety lies within the projected spatial query volume of the query structure. How much of its parts (percentage) should be included to qualify as a valid result depends on the threshold percentage set by the user. In general, a threshold nearing 100% means that a target structure must be almost completely within the spatial query volume, whereas a threshold approaching 0% means only a few voxels of the target structure lies within the spatial query volume. The user can arbitrarily select any number between 0% and 100%. In this study, we arbitrarily set the threshold at 40% to accommodate a compromise between the anatomists in our group. The said threshold is not intended to address any specific application but rather to test the system, its accuracy and its response time.

In the following sections we describe this approach in more detail.

3.1.1 Transverse Direct Relations: Anterior, Posterior, Right-lateral, Left-lateral

▶Figure 1 illustrates the spatial query volumes for these relations. The query structure is depicted in a dark tone, query volumes are depicted as parallelepipeds, and possible target structures in the specified relations are depicted in a lighter tone. For the purposes of this paper we take a very strict definition of the meaning of these relations that we believe can reasonably meet expectations of anatomy experts. Thus, in ▶Figure 1, a target object is only "anterior" to a query object if it falls entirely within the projection ("shadow") of the query object in the anterior direction or a portion of it falls within the projection at or above the threshold set for this relation. If the target object is not within the projection of the query object in the anterior direction then it stands in some other relation to the query object, as for example the antero-right-lateral relation in ▶Figure 1. If the target object falls in both of these regions (e.g.

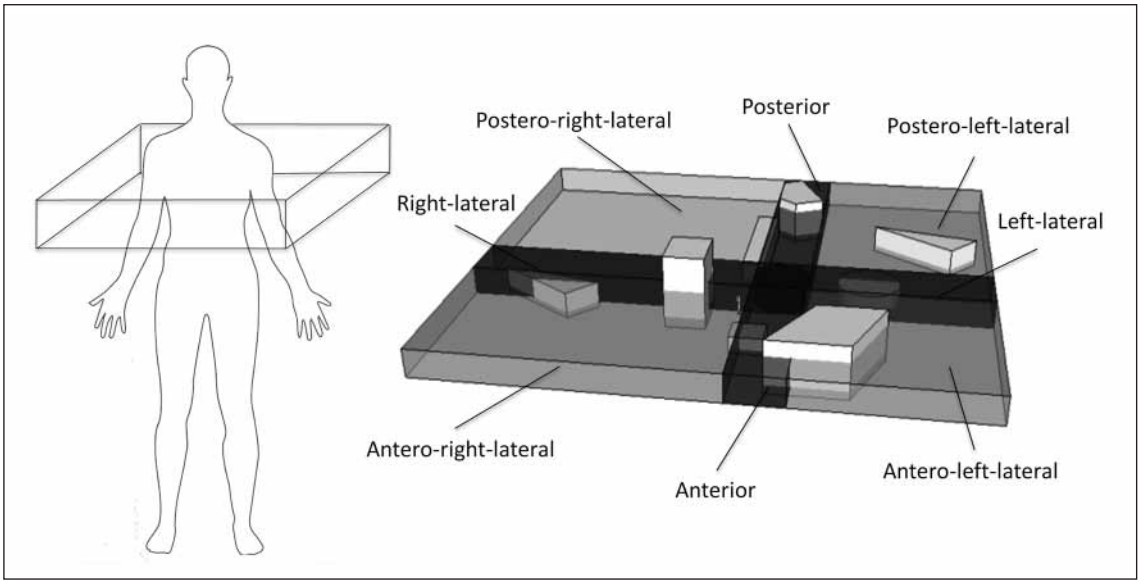

Fig. 1 Transverse direct and intermediate relations

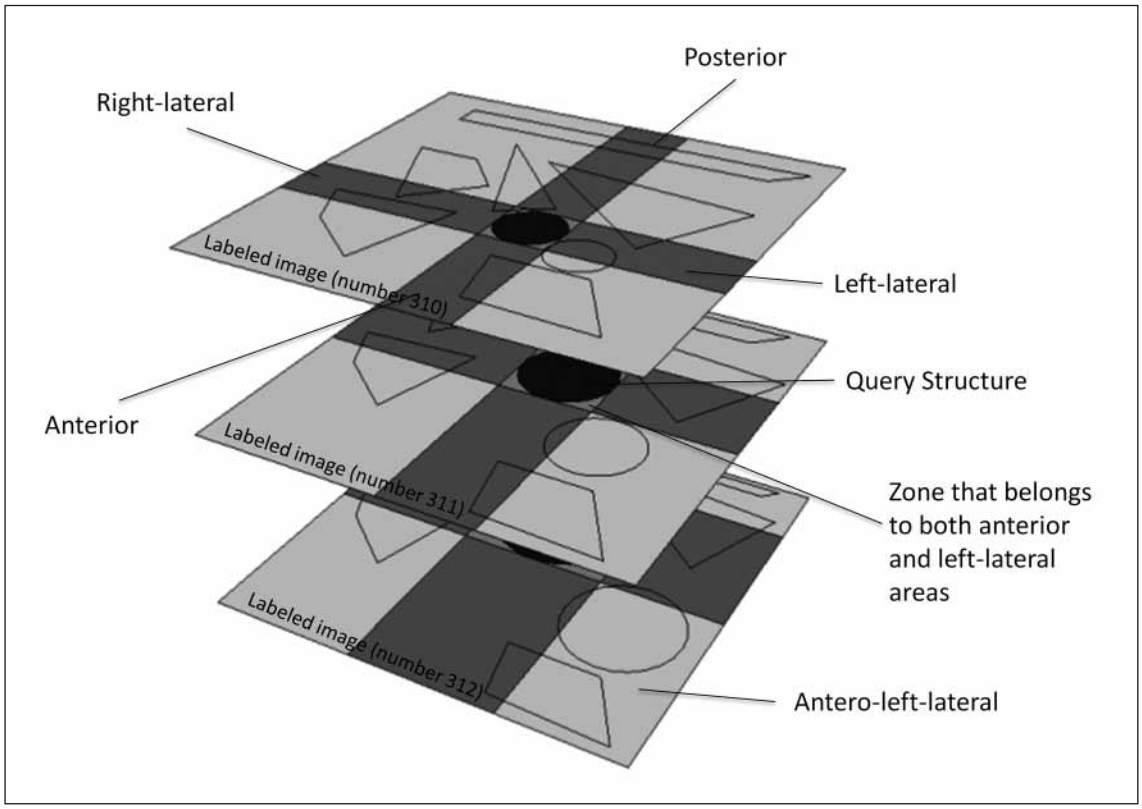

Fig. 2 Implementation of the transverse and direct spatial relation definitions on a stack of 2-D images

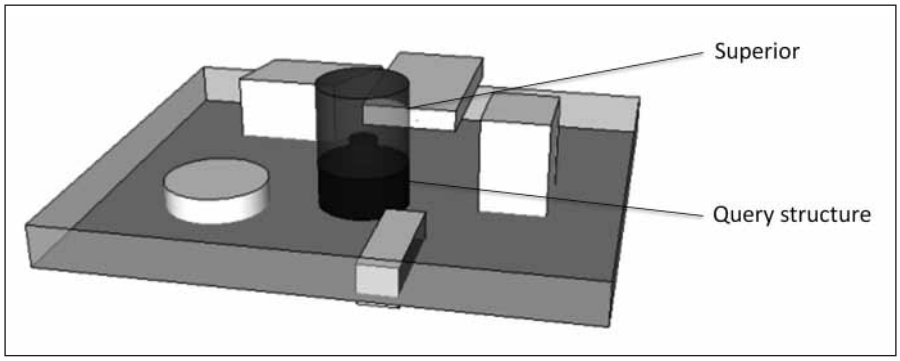

Fig. 3 Superior relation

anterior and antero-right-lateral) then it is considered to stand in relation to the query object in both spatial regions, as long as a percentage of the target object in each region is at or above the set threshold.

Although other experts may decide to be more liberal in their definitions we believe that all definitions must include at least the structures that satisfy our strict definitions. These definitions were in fact proposed by an anatomy expert as a sort of "least common denominator". However, our framework is modular so other definitions can be swapped in.

To compute the query volume in its full 3-D extent, consider the anterior relation (the same principle applies for all other transverse direct relations). We compute the spatial query volume as a reconstruction of spatial query slabs, where each slab corresponds to a single transverse image slice one voxel thick. As illustrated in ▶ Figures 1 and 2, for each such slice the spatial query slab is the portion of the image defined by the projection of the query structure in the same direction as the positive Y-axis in the image slice, which is the anterior direction. The boundaries of the slab emanate therefore from the boundaries of the query structure (as specified by an anatomy expert when labeling the image volume or when confirming the results of an automatic labeling scheme). Then, for each target structure whose cross section on a given slice intersects the query slab the number of voxels in that intersection is recorded. This procedure is repeated for each transverse image slice that falls within the minimum and maximum Z values of the query structure. The total percent involvement of the target structure is then the sum of the intersecting voxels on each slab divided by the total number of voxels in the target structure. For each query structure and for each spatial relation the result is therefore a list of all target structures that intersect the reconstructed query volume, together with their percent intersection.

3.1.2 Transverse Intermediate Relations: Antero-right-lateral, Antero-left-lateral, Postero-right-lateral, Postero-left-lateral

Determining query volumes for transverse intermediate relations relies on the same

mechanism as for transverse direct relations. There is only one difference when it comes to the definition of spatial query slabs: they are rectangles whose boundaries emanate from the minimum/maximum Y and minimum/maximum X values of the query structure on each slice. In ▶Figure 2, query slabs for the antero-left-lateral relation for example are defined by two points: the intersection of two lines tangential to the query structure on the slice where x is equal to the maximum X value of the query structure and y its maximum Y value, and the point in the bottom right-hand corner.

3.1.3 Sagittal Direct Relations: Inferior and Superior

Similarly, a target object is "superior" to the query object if it falls within the projection volume of the query object in the superior direction as described in ▶Figure 3 (or enough of it falls within the projection to exceed the threshold).

In implementing the superior relation we compute the spatial query volume (projection volume) as a reconstruction of spatial query slabs much like the other spatial relations (▶Fig. 4). For each slice starting from the first image of the dataset (highest slice in the body) the spatial query slab is an area of the image defined by the projection of the superior surface of the query structure in the same direction as the negative Z-axis (superior direction). That implies that a slab of any image includes an area with the same coordinates as the slab of the previous image (next image on the Z-axis in the positive direction) as we iterate through the image stack.

Then, for each target structure whose cross section on the given slice intersects the query slab the number of voxels in that intersection is recorded. This procedure is repeated for each transverse image slice that includes a query slab. Finally, the calculation of the total percent involvement of the target structures relies on the same procedure used for the transverse direct relations. A target structure is considered as being superior depending on the computed ratio and the threshold value set by the user. The same approach was used for the definition of the inferior query volume as well.

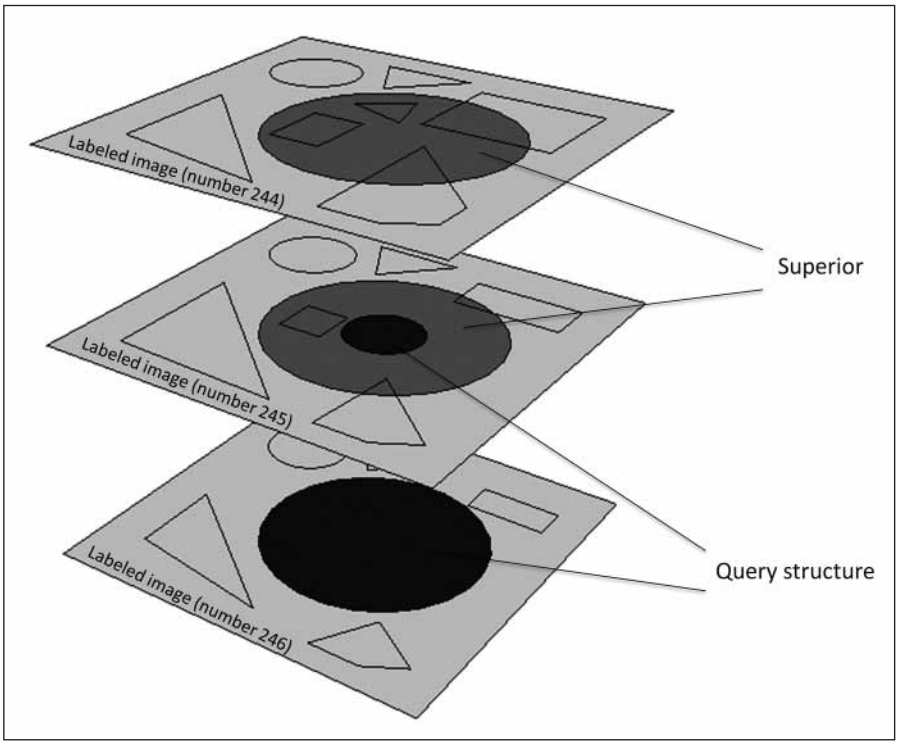

Fig. 4 Implementation of the superior relation definition on a stack of 2-D images

3.1.4 Filtering the Returned Lists

The above procedures return lists of target structures, together with their percentage of involvement in the query volume. Before returning the lists to the user they are pruned and rank ordered:

- Only the structures whose percentage involvement is greater than the user-specified threshold are retained.
- The lists of structures are sorted by the percentage of involvement in defined volumes.

Anatomical structures annotated in the images span different levels of granularity (tissues, organ parts, and organs) and it is up to the users to select the kinds of structures they need to identify in their queries. In this study we only allowed users to either retrieve all structures or just those that are organs, where we define "organ" according to the FMA: "Anatomical structure which has as its direct parts portions of two or more types of tissue or two or more types of cardinal organ part which constitute a maximally connected anatomical structure demarcated predominantly by a bona fide anatomical surface. Examples: femur, biceps, liver, heart, skin, tracheobronchial tree, ovary" [15].

3.2 Spatial Processor

3.2.1 Dataset: 2-D Images

The processor takes as input a set of 2-D labeled transverse images that stack to make a 3-D volume where the labels are terms from the FMA. For this study we used the Virtual Soldier [16] labeled dataset of the Visible Human, which consists of 411 axial label images with a total of 437 segmented and labeled structures and where each grey level is a label. As noted earlier, the Virtual Soldier dataset represents only a single individual (the convict whose body was used to create the Visible Human), and its images were semi-automatically labeled by members of the Virtual Soldier Project (VSP) team. Of the 437 segmented structures 121 qualified as organs by FMA standards. ▶Figure 5 shows a sample slice.

3.2.2 Spatial Query Processing

The 2-D labeled images are read into memory to form a 3-D volume defined by a list

of voxels that give the relative positioning of the anatomical structures. Instead of storing a list of 3-D points representing the voxels, 2-D upper and lower bounds in combination with maps are used for efficiency in query processing, making it possible to store the outline and positioning of the anatomical structures on each slice. ▶Figure 6 describes how the outlines are detected and stored for a slice. Each pixel is read and compared with the surrounding pixels along the medial-lateral direction (X-axis) to detect whether it composes the outline of the shape it represents. Each structure is related to a map that has the following structure: key: "z" (related to the image number); value: another map. For each map related to z, the key is "x" and the value is a point as described in ▶Figure 7: upper bound, lower bound, isolated point. Efficiency is not the primary goal at this stage and therefore, subsequent work should be done in order to conceive a more efficient spatial query processor.

After the entire dataset is read into a 3-D volume in memory, the processor is ready to handle queries and answer questions like "What is anterior to the lower lobe of the left lung?" given the anatomical definitions presented in Section 3.1 and the threshold value decided by the user. After the input query structure and the query spatial relation are selected, the spatial query slab is defined for each slice as in ▶Figure 2. The outline of the query structure was previously stored and therefore the query slab is easily determined for any slice in which the query structure is involved.

▶Figure 8 describes the mechanism of how the percentages of involvement are

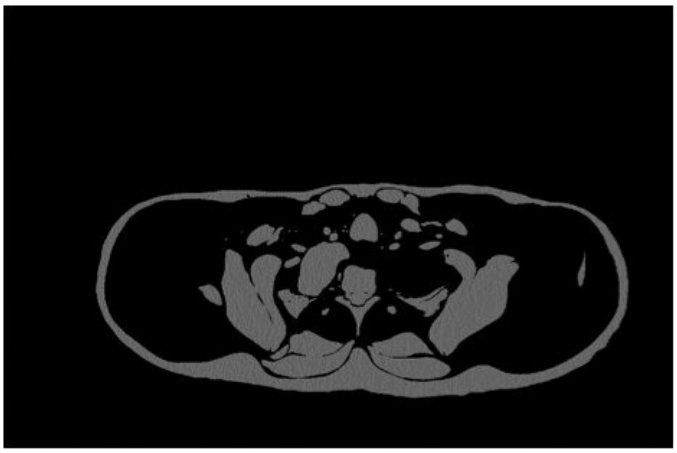

Fig. 5 Enhanced sample labeled image from the Virtual Soldier dataset

```
For reading purposes, consider that point(x,y) belongs to a structure named A

For y from 0 to (height-1), x from 0 to (width-1):
If (point(x,y) is not on the left of the image (x≠0) AND point(x,y) is not on the right of the image (x≠width-1))
        //3 possibilities
        If (pixelLeft and pixelRight do not belong to A)
                // The point is isolated, we add it to the list of points of A
                save point (x,y) for A;
        Else If (pixelLeft belongs to A but not pixelRight)
                // Upper bound
                save point(x,y) as an upper bound for A;
        Else If (pixelRight belongs to A but not pixelLeft)
                //Lower Bound
                save point(x,y) as a lower bound for A;
Else If (point(x,y) is on the right of the image (x=width-1))
        If (pixelLeft belongs to the shape)
                save point(x,y) as an upper bound for A;
        Else
                save point (x,y) for A;
Else If (point(x,y) is on the left of the image (x=0))
        If (pixelRight belongs to the shape)
                save point(x,y) as a lower bound for A;
        Else
                save point (x,y) for A;
```

Fig. 6 Pseudocode used on each slice for storing the 2-D outlines of the anatomical structures

computed for the target structures: each pixel of the query slabs is checked and the number of pixels involved for each target structure is recorded. Eventually the percentages of involvement are computed based on the number of pixels involved in the query slabs and the total number of pixels in the target structures, and a final check against the threshold value determines which target structures are considered posterior. Crossed pixels and pixels in a dark tone in ▶Figure 9 represent the

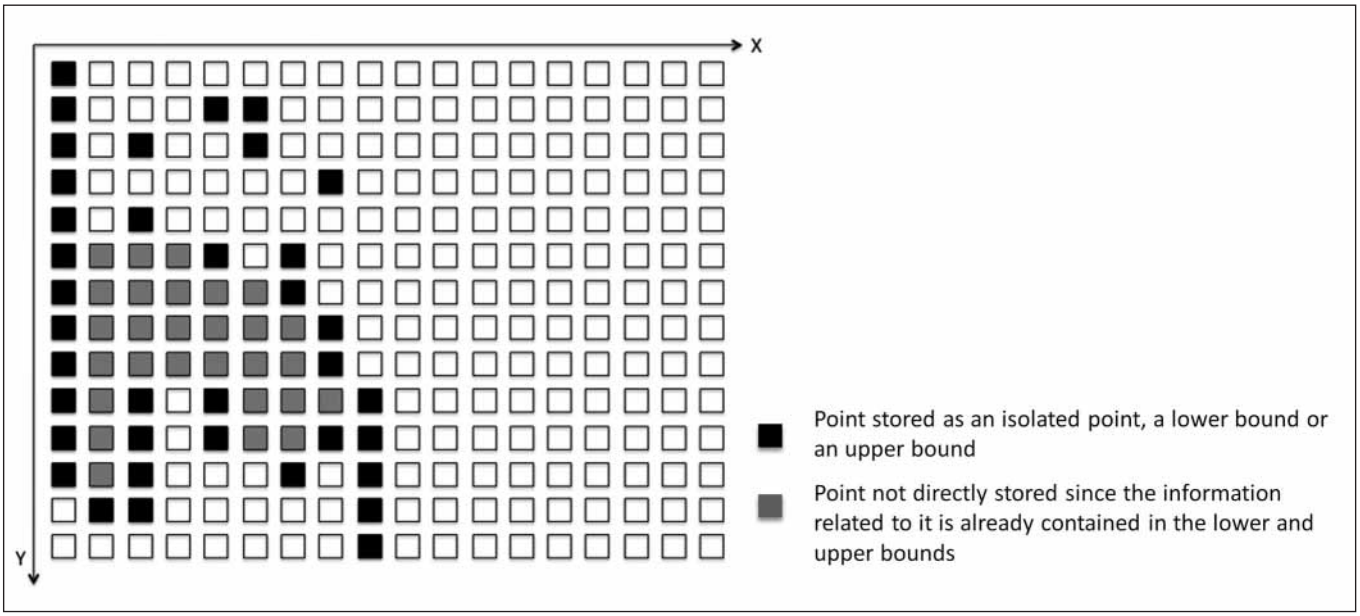

Fig. 7 Sample query slab

```
Set thresholdValue to 40%
Set queryStructure to the structure representing the lower lobe of the left lung //for example

For z from queryStructure.minZ to queryStructure.maxZ
        Set mapOutline to the outline of the query slab of queryStructure based on z and the spatial relation:
        posterior (in the negative Y direction) //See Fig. 9
        For each x in mapOutline
        For y from mapOutline[x] to 0
                Set structureInvolved to the structure located at (x,y,z)
                If (mapStructuresPixels contains structureInvolved)
                        //structureInvolved has already some pixels involved
                        Add one to mapStructuresPixels[structureInvolved]
                Else
                        //structureInvolved has its first pixel involved
                        Set one to mapStructuresPixels[structureInvolved]

//The system is now ready to compute the ratios and determine the posterior structures
For each targetStructure in mapStructuresPixels
//We compute the percent involvement based on the number of pixels contained in the 3-D query slab and the
total number of pixels in each structure
percentInvolvment=100 * MapStructuresPixels[targetStructure] / Total number of pixels of targetStructure;
        If (percentInvolvment > thresholdValue)
        Print targetStructure.name "is considered as posterior with" percentInvolvment "%"
```

Fig. 8 Pseudocode used for computing the percent involvement for the posterior relation in any dataset

query slab of a query structure on one slice for the posterior relation.

3.2.3 Migrating the Results to a Database for Further Use

At the current time the spatial query processor is not fast enough for real-time use. Therefore, we use the spatial query processor to precompute all possible spatial relations on our relatively small dataset, and then cache these results in a database. The database is composed of ten tables (one for each spatial relation). Each table contains three columns: the identifiers of the query structures in the dataset, the identifiers of the target structures for each query structure based on the spatial relation that the table deals with, and the percentages of involvement. ▶Table 1 shows an example. At this point the structure identifiers are only related to the dataset and not to the FMA structure identifiers, although that would be possible based on a mapping between FMA terms and terms used in the VSP dataset.

Note that the spatial database includes all structures in the dataset, both Organs and Organ parts. Only target structures identified as subclasses of FMA class "Organ" are returned in response to the query, including their corresponding percentages of involvement.

Table 1 Example data from the database for the "Anterior" table based on the Virtual Soldier dataset

Anterior		
Query Structure Identifier	**Target Structure Identifier**	**Percentage of Involvement**
181 (Lower lobe of the left lung)	42 (Left inferior lobar bronchus)	27%
181	50 (Central tendon)	41%
181	62 (Coronary sinus)	35%
181	67 (Descending thoracic aorta)	32%
181	92 (Fifth internal intercostal muscle)	26%
181	96 (Fifth anterior intercostal vein)	42%

3.3 Relating Spatial and Symbolic Knowledge via a Semantic Web Service

Spatial information is retrieved via the cached results from the spatial query processor and a web service. The web service enables users to select a single or multiple anatomical structures and define queries about spatial relations by setting up a threshold value related to each spatial relation (e.g. anterior). This information is combined with symbolic information from the FMA via a semantic web service.

3.3.1 Architecture

The semantic web service is composed of four main elements as described in ▶Figure 10: a graphical user interface, a query engine and two data sources. The web query interface (GUI) enables users to define spatial-symbolic queries, store sets of structures as intermediate results, and display the results of spatial queries. The GUI was implemented with Openlazslo [17], a platform for rich Internet applications that uses an XML language, which can compile to DHTML. With the GUI, it is possible to choose anatomical structures that will be involved in a query as well as the type of relation involved (symbolic and/or spatial). Finally, there is a possibility to create unions and intersections of

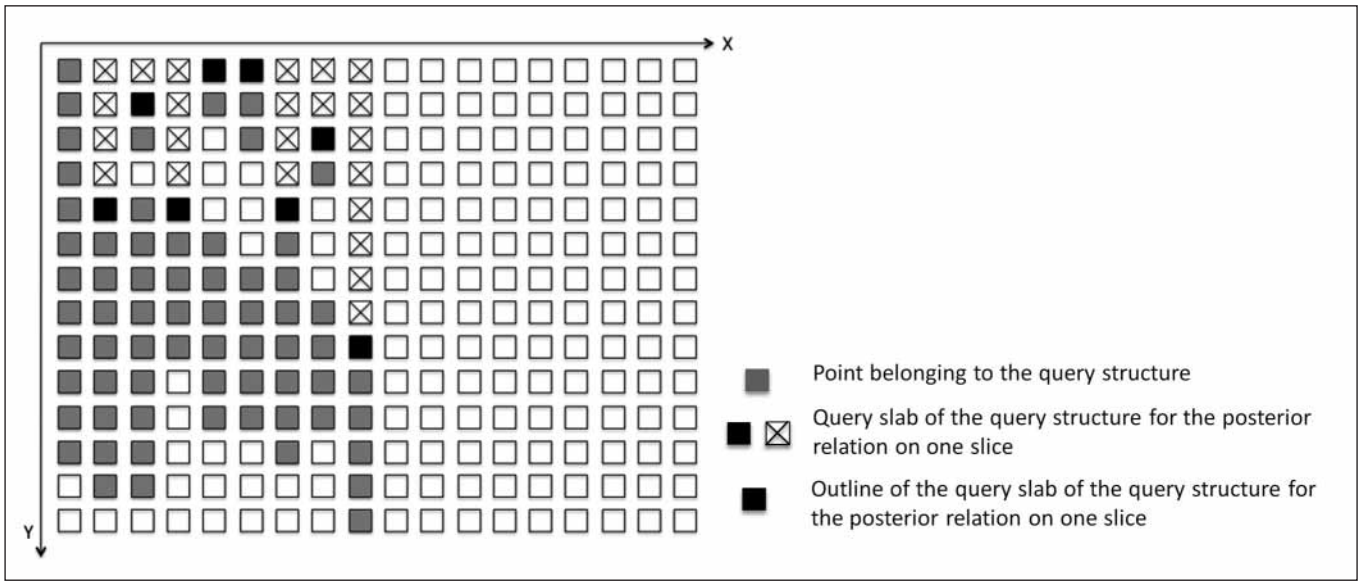

Fig. 9 Query slab for the posterior relation

sets to manage intermediate results when needed.

After the parameters are defined, they are sent to the query engine that inserts them into a template vSparQL query, where vSparQL [12] is a set of extensions we have created for the W3C-recommended SparQL [11] query language for RDF (Resource Description Framework), the foundation for the semantic web [18]. In particular we utilize the subquery feature of vSparQL that allows us to easily query over more than one RDF information source.

In our case the query engine queries two data sources: the spatial relation database and the FMA database, both of which are accessible as RDF triple stores. The spatial relation store is computed by the methods described in the previous section, whereas the version of the FMA we use is stored in OWL (web ontology language [19], which is built on RDF). The FMA represents entities and relations necessary for the symbolic modeling of the phenotypic structure of the human body, as for example, the types of structures (organs, cells) and spatio-structural relations such as branch-of and part-of.

The process of transferring spatial knowledge from a relational SQL database to a relational database as an RDF triple store is performed with the Jena framework, a semantic web framework for Java that provides persistent storage of RDF data in relational databases [20]. ▶Figure 11 describes the spatial RDF graph. Nodes represent anatomical structures and are identified by URIs, also used in the FMA ontology. An expert in anatomy performed the mapping between FMA terms and VSP terms in order to build the RDF graph with common anatomical structure identifiers (URIs). Each structure has associated properties: a name and the list of structures (only the list of anterior structures appears in ▶Figure 11 but the same principle applies for the other spatial relations). The list of anterior structures is a bag of blank nodes having two properties: a target structure's node (URI) and the percent of involvement.

3.3.2 Example Query

The query "Which organs are anterior to the lower lobe of the left lung (40%)?" requires handling spatial in symbolic form. Through this example, it is possible to show the mechanics of how the query engine works.

The first step is the definition of the question in the GUI (▶Fig. 12). The structure "Subject-Relation-Object" was chosen for the graphical user interface because it provides a means for posing numerous atomic questions over the FMA and spatial relations. Therefore any query can be divided into subqueries before getting the result of the initial query. In the question "Which organs are anterior to the lower lobe of the left lung? (40%)", the subject is unknown whereas the relation "anterior to-40%" and the object "the lower lobe of the left lung" are known. The GUI allows users to select a structure among a list of known structures. In ▶Figure 12, the spatial relation (anterior-40%) is selected and so is "organ" in the "Spatial and Symbolic" tab. That configuration will give the answer to the question directly with no intermediate results.

In order to determine the anterior organs, the appropriate parameters (name of the query structure, spatial relation and threshold value) are sent to the query engine that inserts them in a vSparQL query template and runs the query over the two RDF triplestores: the spatial database and the FMA (▶Fig. 13). vSparQL is a key element of the web service, making it possible to run only one query defined through the GUI and return information about both symbolic and spatial knowledge as a result.

Note that the use of Gleen [21] enables the web service to apply the transitivity property with SparQL. This property is essential to determine anatomical structures that are subclasses of organs (▶Fig. 13).

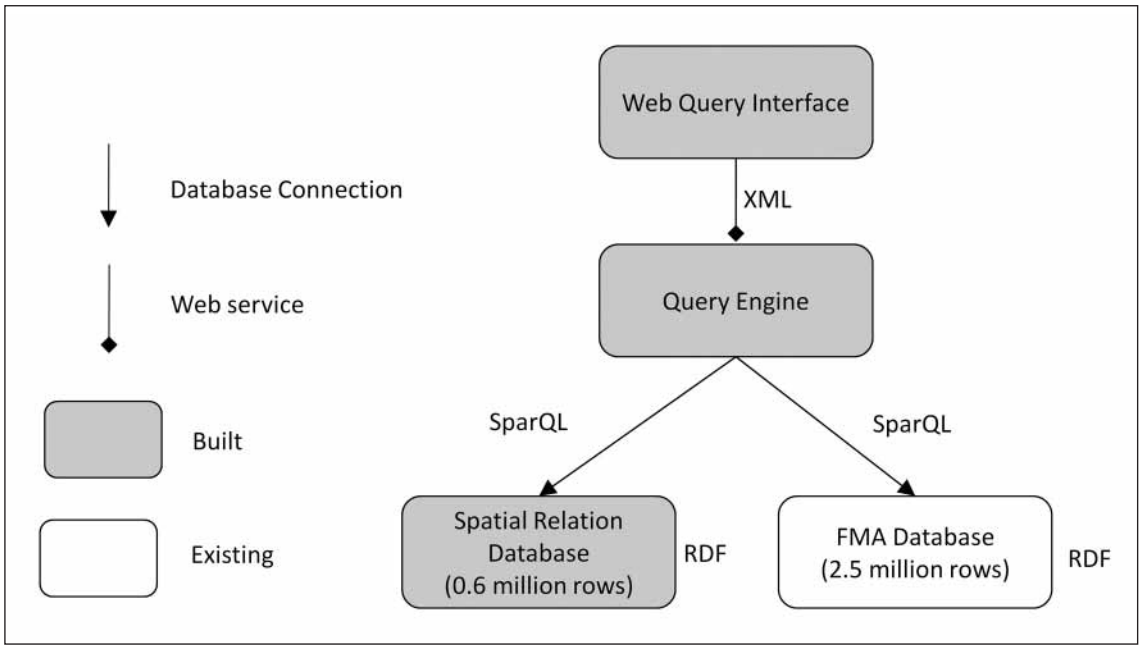

Fig. 10 Architecture of the symbolic-spatial web service

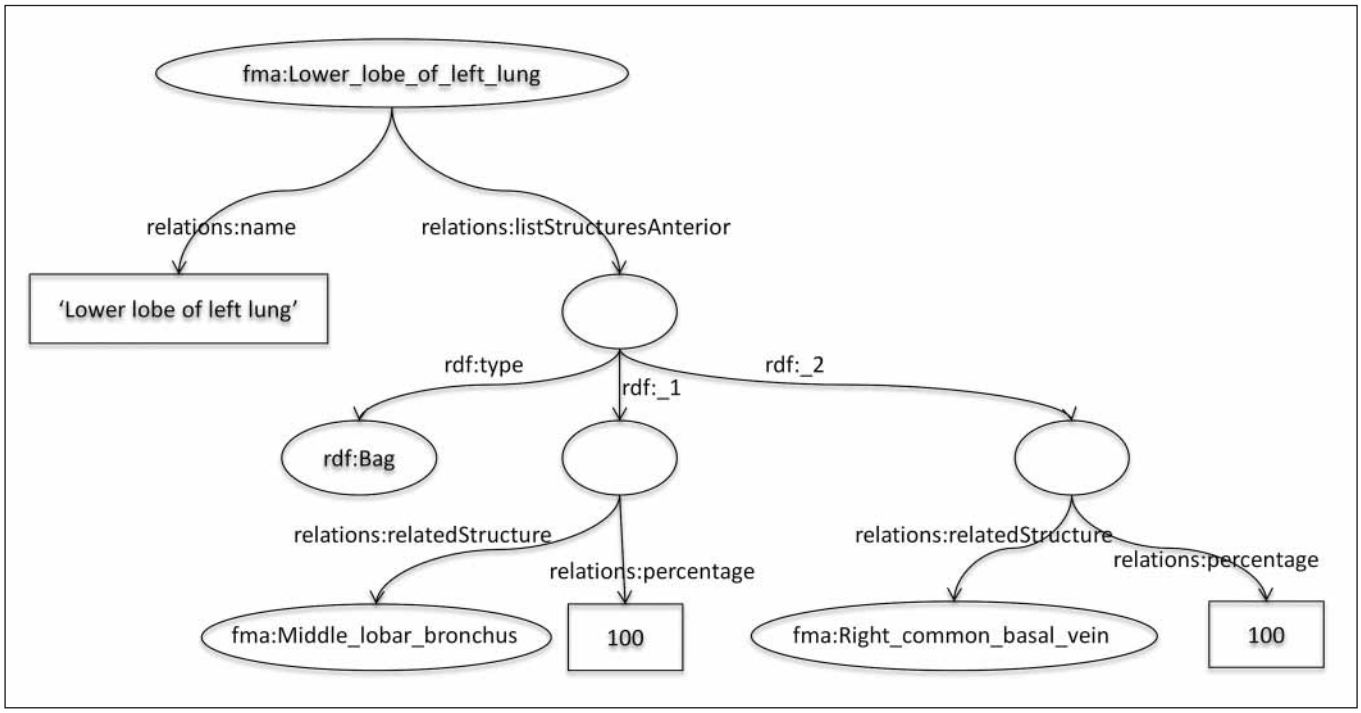

Fig. 11 Graph structure for a single anatomical structure

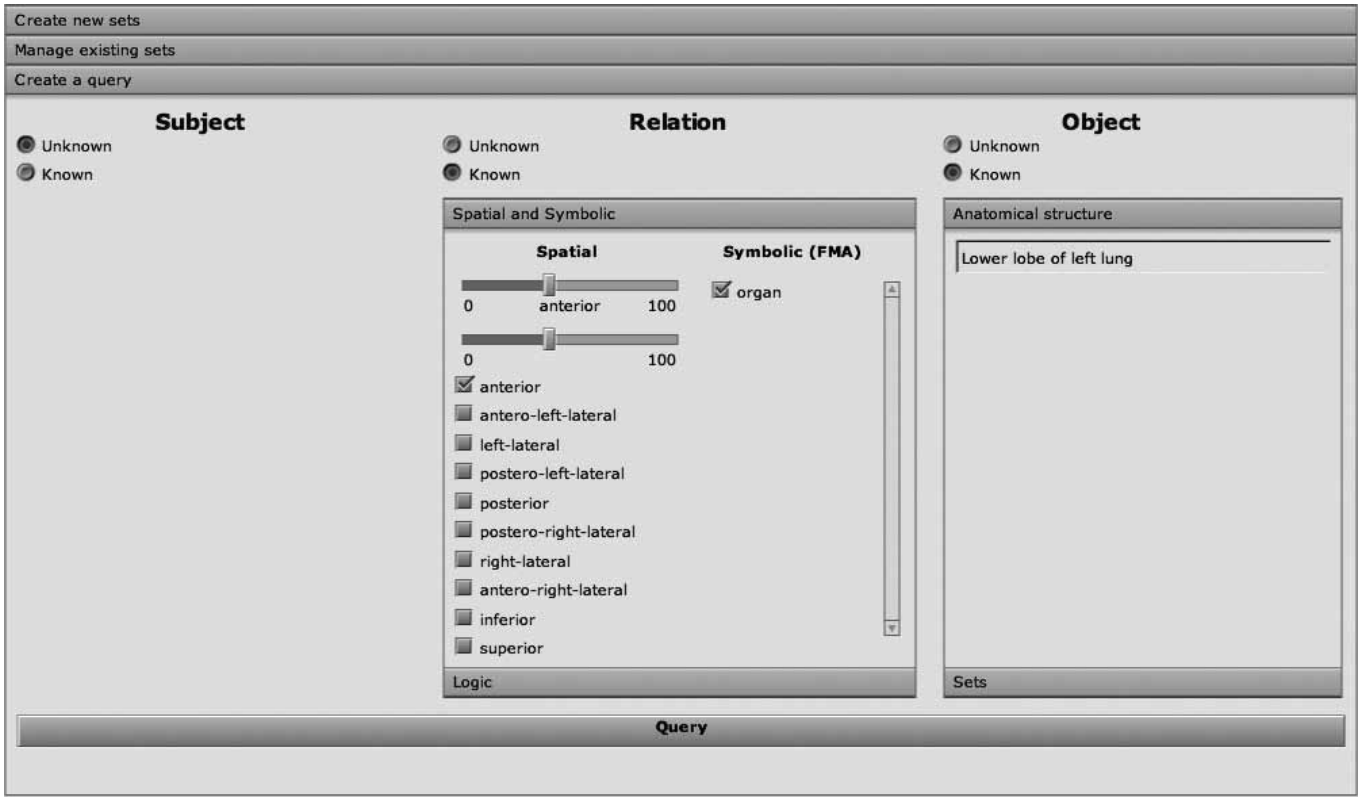

Fig. 12 Definition of a query through the web query interface

```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX gleen:<java:edu.washington.sig.gleen.>
PREFIX relations:<http://sig.biostr.washington.edu/spatialrelations#>
PREFIX fma:<http://bioontology.org/projects/ontologies/fma/fmaOwlDlComponent_2_0#>

SELECT ?organ ?percent
FROM NAMED <http://bioontology.org/projects/ontologies/fma/fmaOwlFullComponent_2_0>
FROM <anterior>
[
        CONSTRUCT { ?anterior relations:percentage ?percentage.
                fma:Lower_lobe_of_left_lung relations:anterior ?anterior. }
        FROM <http://sig.biostr.washington.edu/spatial>
        WHERE
        {
                fma:Lower_lobe_of_left_lung relations:listStructuresAnterior ?bag.
                ?bag rdfs:member ?node.
                ?node relations:percentage ?percentage.
                ?node relations:relatedStructure ?anterior.
                FILTER (?percentage>=40)
        }
]
WHERE
{       GRAPH<http://bioontology.org/projects/ontologies/fma/fmaOwlFullComponent_2_0>
        {
                ?organ gleen:OnPath ('[rdfs:subClassOf]*' fma:Organ) .
        }
      ?structure relations:anterior ?organ .
      ?organ relations:percentage ?percent .
}
ORDER BY DESC(?percent)
```

Fig. 13
vSparQL query: Which organs are anterior to the lower lobe of the left lung (40%)?

After clicking "Query" which is the last step, the list of organs appears in a new window (▶Fig. 14) and gives the answer to the question "Which organs are anterior to the lower lobe of the left lung? (40%)".

Note that if "organ" is not selected in the GUI, the query in ▶Figure 13 that uses Gleen to determine whether an anterior structure is an organ or not would not be run. Instead the system would run another query over the spatial database to return all the structures, including organ parts, that are anterior to the lower lobe of the left lung, as illustrated in ▶Figure 15.

4. Results

The data sources are composed of two relational databases representing RDF graphs. The spatial database consists of 0.6 million rows whereas the FMA database consists of 2.5 millions rows. In addition to having the ability to return all structures for each spatial relation, the web service also has the ability to filter spatial results based on a specified property of the FMA RDF graph. In this case, we limited the results to include only organs. For these two types of queries a template vSparQL query is hard-coded in the system. The query in ▶Figure 13 is used to only check whether a candidate structure is an organ. The ability to filter further the results using additional properties from the FMA, such as different class types other than organs, or relations such as parts or branches, could be readily added by the creation of new template queries like that in ▶Figure 13, and new GUI components that would allow users to select their particular constraints. Given this state of the system, a preliminary evaluation was performed to test accuracy and response time.

4.1 Accuracy Evaluation by Two Anatomy Experts

In this section we describe a preliminary evaluation of the accuracy of results returned by the queries. Since there is no independent gold standard yet available, we used two anatomy experts' opinion as the gold standard for comparison. Anatomy experts will likely have their own intuitions on the definition of spatial relations and on the percentage of the target structure volume that needs to be in the specified relation to the query structure. In this study the participating anatomy experts agreed on these two aspects and defined them as the test standard. They also specified the threshold percentage volume to the query engine.

To fully generate a gold standard against which the accuracy of the spatial query engine can be evaluated would require that

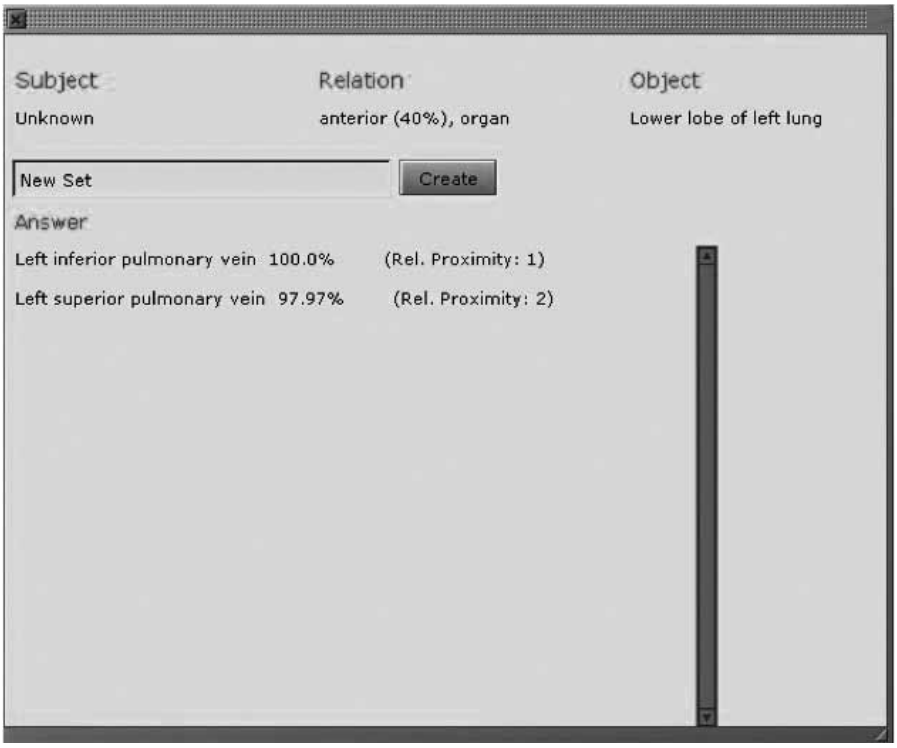

Fig. 14 List of organs anterior to the lower lobe of the left lung (40%)

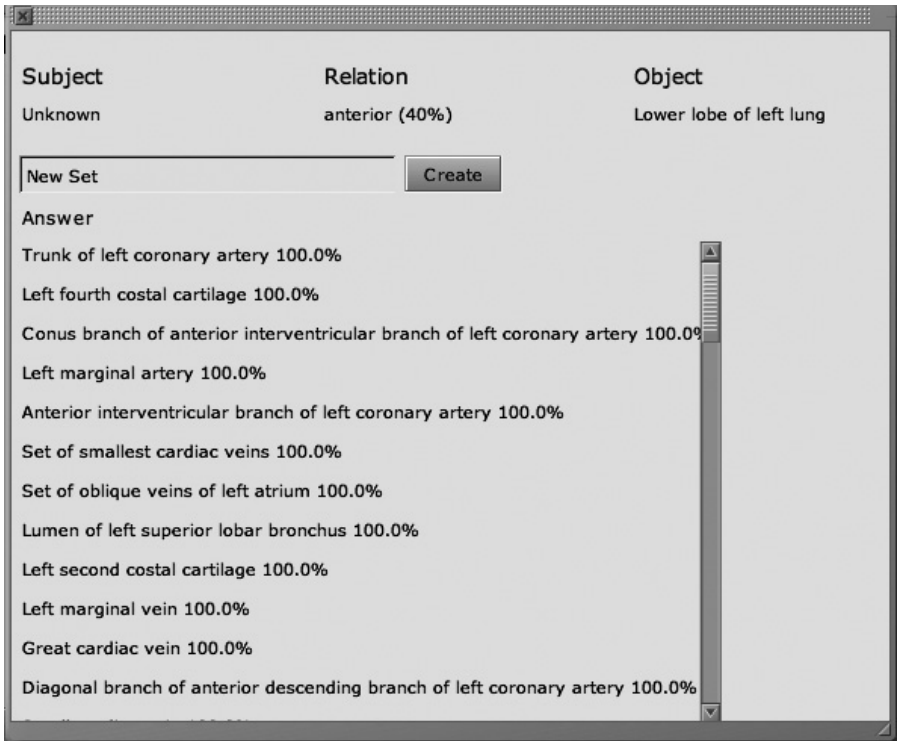

Fig. 15 List of structures anterior to the lower lobe of the left lung (40%)

each expert generates 437 × 10 = 4,370 lists at a given threshold, where 437 is the number of labeled structures in the VSP dataset, 10 is the number of relations, and each list is one expert's opinion of those structures that are in the specified relation at the specified threshold with the query structure. Given this list we could then evaluate precision and recall of the system. However, we deemed this task to be too daunting for the experts in this preliminary evaluation, so we limited our query structures to a sample of 43 randomly selected structures (i.e. 10% of the dataset) and the sagittal and transverse direct relations (superior, inferior, anterior, posterior, right-lateral and left-lateral). For each of these structures we ran the query engine at a fixed threshold (40%) and with the "organ" checkbox selected, and computed precision and recall only for the random sample.

To do this each expert examined the returned list for each query structure and each relation (43 × 6 = 258 lists) and indicated 'Y' for each returned organ if he thought it was indeed in the specified relation to the query structure and 'N' if he thought it was not. Thus, for each query structure the number of Ys represented the true positives and the number of 'N's represented the false positives.

To determine the false negatives (those organs the experts felt should be in the specified relation, but which the program missed) the experts provided a count of the missing organs, as well as the specific names of the missing organs as a validity check. Based on the FMA the number of organs in the dataset is 121 out of the total 437 structures. Thus, the experts dealt only with those organs in the list of 121, and in many cases it was immediately determined that none of these organs were involved in any of the specified relations. Using this procedure the number of false negatives for each query structure was just the count of the missing organs as supplied by the experts, and the number of true negatives (those that were not in the specified relation and which the computer did not flag) was the total number of possible organs less the true positives, the false negatives and the false positives.

These numbers were summed for the entire 43 query structures, leading to an

Table 2 Results of the evaluation by two anatomy experts for 43 randomly selected query structures for transverse and sagittal direct relations

	Anterior		Posterior		Right-lateral		Left-lateral	
	Expert 1	Expert 2	Expert 1	Expert 2	Expert 1	Expert 2	Expert 1	Expert 2
Number of true positives	57	52	10	7	13	4	17	3
Number of false positives	8	13	1	4	6	15	5	19
Number of false negatives	9	0	4	0	0	0	0	0
Number of true negatives	5,129	5,138	5,188	5,192	5,184	5,184	5,181	5,181
Precision	0.88	0.80	0.91	0.64	0.68	0.21	0.77	0.14
Recall	0.86	1.00	0.71	1.00	1.00	1.00	1.00	1.00
F-measure	0.87	0.89	0.80	0.78	0.81	0.35	0.87	0.24
	Inferior		**Superior**		**Overall**			
	Expert 1	Expert 2	Expert 1	Expert 2	Expert 1	Expert 2		
Number of true positives	113	15	32	23	242	104		
Number of false positives	3	101	4	13	27	165		
Number of false negatives	5	1	6	0	24	1		
Number of true negatives	5,082	5,086	5,161	5,167	30,925	30,948		
Precision	0.97	0.13	0.89	0.64	0.90	0.39		
Recall	0.96	0.94	0.84	1.00	0.91	0.99		
F-measure	0.97	0.23	0.86	0.78	0.90	0.56		

overall calculation of true positives, true negatives, false positives and false negatives, from which overall precision, recall, and F-value could be calculated, as in ▶Table 2. These overall values are also broken down according to relation type and expert in ▶Table 2. Details of these calculations are provided in the ▶Appendix.

To describe the quality of the gold standard an inter-rater reliability evaluation was performed with the calculation of Cohen's kappa. For each returned list of target structures for each spatial relation and each query structure the experts provided a count of common Ys and common Ns as well as the number of Ys and Ns expressed by one expert while the other expert specified respectively Ns and Ys. These numbers were summed for the query structures, leading to the calculation of the percentage of observed agreement, the percentage of expected agreement and Cohen's kappa (▶Table 3). Like precision, recall and F-value, these overall values were also broken down according to relation type in ▶Table 3. Details of these calculations are provided in the ▶Appendix.

4.2 Evaluation of Response Time

In addition to accuracy the system must return results in a reasonable amount of time in response to queries. As noted in Section 3.2.3 the response time was too slow when directly querying the spatial query processor as it is currently implemented. In this section we document this assertion, and in addition, show that the response time is greatly improved when the cached spatial triple store is accessed rather than directly querying the spatial query processor.

Analysis of the server logs from $437 \times 4 = 1{,}748$ queries (437 structures, 4 spatial relations) sent to the system when building the spatial database showed that the spatial query processor without the cached triple store has a mean response time equal to 29,078 ms (29 seconds) to respond to a query (▶Table 4). In contrast, given the database cache, the mean response time for

Table 3 Measure of inter-rater agreement for the evaluation with the calculation of Cohen's kappa

		Anterior	Posterior	Right-lateral	Left-lateral	Inferior	Superior	Overall
Number of occurrences	Expert 1: Y, Expert 2: Y (Agreement)	47	7	2	1	15	22	94
	Expert 1: N, Expert 2: N (Agreement)	5,132	5,189	5,188	5,184	5,086	5,164	30,943
	Expert 1: Y, Expert 2: N (Disagreement)	19	7	11	16	101	16	170
	Expert 1: N, Expert 2: Y (Disagreement)	5	0	2	2	1	1	11
Percentage of observed agreement		1.00	1.00	1.00	1.00	0.98	1.00	0.99
Percentage of expected agreement		0.98	1.00	1.00	1.00	0.97	0.99	0.99
Cohen's kappa		0.79	0.67	0.23	0.10	0.22	0.72	0.51

Table 4 Mean time response needed by the spatial processor and the web application to return a result

	Query Processor (437 structures)	Web Service (428 structures)	Web Service (121 organs)
Type of information requested	Spatial	Spatial	Spatial and symbolic
Mean: anterior, posterior, right-lateral and left-lateral	29,078 ms	265 ms	5,465 ms

Table 5 Time response needed by the spatial processor and the web application to return a result to a specific query

Anatomical Structure	Query Processor: What is anterior? 0% (all target structures)	Web Service: What is anterior? 0% (all target structures)	Web Service: Which organs are anterior? 0% (all target structures)
Right First Rib	38,641 ms	234 ms	5,625 ms
Left Sixth Costal Cartilage	2,500 ms	78 ms	5,687 ms
Lower lobe of the left lung	316,719 ms	641 ms	6,047 ms

a spatial query alone with the web application is equal to 265 ms ($428 \times 4 = 1{,}712$ queries: 428 structures, 4 spatial relations), and for the combined spatial-symbolic query (in which the system had to issue a separate subquery to the FMA) the mean response time was about five seconds. Note that some terms related to the VSP dataset were not mapped to FMA terms.

In ▶Table 5, there are three examples that show the response time when querying for anterior structures. The larger the anterior spatial query volume, the longer it takes for the processor to process a query and the response time is therefore very variable depending on target structures. Based on those values, using the spatial processor directly as a web service is not feasible for users' convenience. However, the web service delivers fast responses for any spatial query thanks to the cache.

The web service requires more time to run a spatial-symbolic SparQL query over the FMA database and the spatial database than it does to run a spatial query alone over the spatial database. The time needed is indeed nearly the same for any structure, which shows that the combination of spatial and symbolic knowledge dictates the mean response time.

5. Discussion

5.1 Evaluation

The preliminary evaluation on a random sample of 10% of the possible query structures shows that according to our strict criteria, and using two experts to define the threshold and query volume, and to evaluate the results, the system produced an overall F-measure of 0.90 for rater 1 and 0.56 for rater 2. This can be explained by the fact that expert 2 discarded results for query structures whose labels did not specify any laterality assignment, such as the eighth external intercostal muscle: is the query structure the right or the left muscle? On the other hand, the first evaluator inferred the sidedness of the query structures from the results obtained. If we only take into account the lateralized query structures (28 structures out of 43), the F-measure values would be 0.92

and 0.80 for expert 1 and expert 2 respectively.

The evaluation is based on two experts' opinion with a percentage of observed agreement of 99% for all returned target structures and Cohen's kappa coefficient reaching 0.51, which indicates a moderate agreement between expert 1 and expert 2. This is largely due to the approach taken by each evaluator in dealing with non-lateralized query structures and not on the performance of the tool itself. But if we take out the results for non-lateralized labels, Cohen's kappa coefficient goes up to 0.75, which indicates a much better agreement between experts 1 and 2.

Thus, these values are high enough to validate internally the gold standard. Keep in mind that a separate observer or expert may judge the results differently. We have yet to determine how the system would do on different datasets. In this study the evaluation mostly tests the computations we used in the system, as well as the labeling scheme for the original data.

We have identified a few false positive and false negative results and we believe that improper labeling of the original images likely accounts for the errors. We surmise that either the entire structure has not been accurately segmented or delineated thereby resulting in the erroneous percentage (reaching over 40%) of the structure falling within the shadow (the segmentation was extended beyond the true boundary of the structure) or the structure is not properly labeled, such as in the case of 'Fifth external intercostal muscle' where this label does not specify whether it refers to the right or the left version of the muscle. If the structure labeled is the left muscle then the 'right-lateral' relation with the right fifth rib is correct. However, if it is the right muscle, the relation is no longer valid and hence the system will register an error. We also detected some inconsistencies in the dataset where a structure has the 'organ part' label at certain cross section levels and then the 'organ' label in the other remaining sections, e.g. some pixels are labeled as 'left atrium' and 'left ventricle' in some cross sections but then as 'heart' in the succeeding sections. As a result the program identifies the heart as a smaller structure, which then easily falls within the 40% threshold, thereby leading to a false positive result. Another possible source of errors or inconsistencies is anatomical variation. The images we used for this work were taken from an individual whose structures may have some spatial properties that deviate from the canonical types and therefore the results may have been interpreted as false positives or false negatives by expert evaluators using canonical standards. Therefore to obtain more accurate results, we believe that more precise labeling of the datasets in strict conformance to FMA rules and conventions as well as using the same standards for evaluation are needed. This will likely lead to an F-measure closer to 1. Note also that the GUI of the system prevents users from defining queries with labels that are not included in the dataset so that there is no possibility of using a synonym of an anatomical structure, which eliminates a possible source of errors.

To further validate the system we built we would require a more complete evaluation by additional anatomy experts, examination of more structures and relations captured in additional datasets, enhanced or extended definitions of spatial relations, and involvement of more experts. However given the large number of possible variations in the spatial properties of anatomical structures as well as the different opinions of experts in interpreting the spatial relations, it would be a difficult challenge to set the gold standards for spatial relations that could be used to evaluate systems such as the one we proposed. However, the fact that our approach allows different definitions and labeled datasets to be swapped in, and can be widely accessed and used over the web means that our system could potentially be used to test different hypotheses by different experts, which may eventually lead to a consensus on the definitions of canonical spatial relations that most experts may agree upon.

From the technical standpoint, the system efficiently returned query results in a reasonable time as long as a cache is used. Taken together these results suggest that the program accurately calculates the spatial relations in a timely manner given accurate input data and strict adherence to our definitions of spatial relations.

5.2 Further Work

The web service's purpose is to provide the user or another application access to human readable and machine-processable spatial knowledge through the web service and web interface. In its current state the implementation can enable the use of any of the spatial relations in queries. However in this work we have only demonstrated queries on spatial relations dealing with general qualitative coordinates such as anterior, posterior, lateral, superior and inferior. We have yet to take advantage of other spatio-structural relations defined and implemented in the FMA, such as part-of, branch-of, continuous-with, etc. As noted in section 4 such relations could be enabled by writing new template vSparQL queries and new GUI components to populate those template queries. An interface modeled after our EMILY interface to the FMA [8] could prove useful in this case, and in fact some elements of EMILY have been incorporated in the current system.

Another area requiring further work relates to response time and the speed of the spatial query processor engine. The relatively small number of structures in the VSP dataset (437) permits caching of all possible results of spatial queries and storage in the spatial triple store. However, given that the actual number of structures is much larger than this (there are about 85,000 anatomical entities in the FMA), this approach may not scale up. Thus, methods from fields like computer graphics or geographic information systems should be adapted in order to enable real-time response from the spatial query processor without the need for pre-caching the results. Such a real-time processor would, in addition to the traditional spatial relations described in this paper (anterior, posterior, etc.) allow other types of relations, such as those anatomical structures impacted by a bullet, or those impacted by a needle during a biopsy or central venous line insertion.

Additional features that could enhance the utility of the application include the ability to:

- determine the proximity of target structures to query structures by measuring the distance between the center of mass

of the query structure and that of each target structure for any spatial relation;
- determine the precise ordered position of target structures to the query structure. For example, in some applications it may be important to know whether the sternum is in immediate anterior relation to the heart or whether there are other intervening structures between them;
- detect adjacency relations based on user specified conditions, such as immediate adjacency or adjacency based on some predetermined distance;
- detect connectivity relations and specify whether the relation is of type 'continuity' (no bona fide boundary between two structures) or 'attachment' (bona fide boundary between two structures);
- select the level of granularity required for a particular task or application, therefore providing the user the choice of selecting a given class of anatomical structures among the results (only organs can be selected with the current version of the application);
- display the results as a 3-D anatomical scene rather than as a simple list of structures. This would allow better validation of the correctness of the results.

6. Conclusions

In this report we have described a system that translates spatial information from an FMA labeled 3-D dataset into symbolic spatial knowledge, which can be queried to determine the precise spatial relations between anatomical structures that are of interest to users. Because the annotations are based on the FMA, the derived relations from this method can be used to enhance and extend the spatio-structural properties of anatomical entities that already exist in the FMA. Because we are dealing with a large number of structures and relations, our approach utilizes a pre-calculated triple-store cache to permit reasonable response time. Our system is implemented within the context of the semantic web to promote interoperability with other semantic web applications. Preliminary results and the evaluation suggest that the system is accurate enough within the constraints of the definitions that we established for the set of spatial coordinate relations we have used. However, additional extensive evaluations are needed to verify and confirm the validity of our results.

In its current state the system is likely to be of utility as a framework for building end-user applications rather than as an end-user application in itself. Although the web application could be used by end-users, or another end-user application could be built to access the web service, the system we developed still needs to be expanded and tested using other labeled datasets. Further evaluation by additional anatomy experts is needed before the tool is ready to be released for real-world application. However, we have laid the foundation where the system's modularity and web-based capability can provide the facility for different users to evaluate it over the web. Thus, once these verifications and expansions on this system and its extensions have been carried out and completed, the potential to provide a fast and efficient system to provide answers to anatomical queries over the web can be realized.

Acknowledgements

This work was partially funded by NIH grant HL087706. We thank Dr. Cornelius Rosse for his constructive comments and his highly valuable feedback on the article, and Marianne Shaw for letting us use her work on SparQL queries. The project page is available at http://sig.biostr.washington.edu/projects/spatial-symbolic.

References

1. Brinkley JF, Rosse C. The Digital Anatomist distributed framework and its applications to knowledge based medical imaging. JAMIA 1997; 4 (3): 165–183.
2. Johnson KA, Becker JA. 2001. The Whole Brain Atlas. http://www.med.harvard.edu/AANLIB/home.html.
3. Ackerman MJ. The Visible Human Project: a resource for education. Academic Medicine 1999; 74 (6): 667–670.
4. Hohne K, Pflesser B, Pommert A, Riemer M, Schubert R, Schiemann T, Tiede U, Schumacher U. A realistic model of human structure from the Visible Human data. Methods Inf Med 2001; 40 (2): 83–89. http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a-z/methods/contents/archive/issue/704/manuscript/5205/download.html
5. ADAM Software. ADAM Scholar Series; 1995.
6. Google. 2011. Body Browser. http://bodybrowser.googlelabs.com/ (requires Chrome).
7. Distelhorst G, Srivastava V, Rosse C, Brinkley JF. A prototype natural language interface to a large complex knowledhge base, the Foundational Model of Anatomy. In: AMIA Fall Symposium; 2003. pp 200–204. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1480167/
8. Shapiro LG, Chung E, Detwiler LT, Mejino JLV, Agoncillo AV, Brinkley JF, Rosse C. Processes and problems in the formative evaluation of an interface to the foundational model of anatomy. JAMIA 2005; 12: 35–46. http://jamia.bmjjournals.com/content/12/1/35.short
9. Fernandez M, Florescu D, Levy H, Suciu D. A query language for a web site management system. SIGMOD Record 1997; 26 (3): 4–11.
10. Mork P, Brinkley JF, Rosse C. OQAFMA Querying Agent for the Foundational Model of Anatomy: a prototype for providing flexible and efficient access to large semantic networks. J Biomed Inform 2003; 36 (6): 501–517. http://www.sciencedirect.com/science/article/pii/S1532046403001187
11. World Wide Web Consortium. 2010. SparQL query language for RDF. http://www.w3.org/TR/rdf-sparql-query/.
12. Shaw MS, Detwiler LT, Noy NF, Brinkley JF, Suciu D. vSPARQL: A view definition language for the semantic web. J Biomed Inform 2010; 44 (1): 102–117. doi:10.1016/j.jbi.2010.08.008
13. Ogunyemi OI, Clarke JR, Ash N, Webber BL. Combining geometric and probabilistic reasoning for computer-based penetrating-trauma assessment. JAMIA 2002; 9 (3): 273–282. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11971888 11971888.
14. Rosse C, Gaddum-Rosse P. Hollinshead's Textbook of Anatomy. 5th ed. Philadelphia: J.P. Lippincott; 1997.
15. Rosse C, Mejino JLV. A reference ontology for bioinformatics: the Foundational Model of Anatomy. J Biomed Inform 2003; 36 (6): 478–500. http://www.sciencedirect.com/science/article/pii/S1532046403001278
16. DARPA. The Virtual Soldier Project. 2011. http://www.virtualsoldier.us/
17. Laszlo Systems. 2011. OpenLaszlo. http://www.openlaszlo.org/.
18. World Wide Web Consortium. 2010. Resource Description Framework (RDF). http://www.w3.org/RDF.
19. World Wide Web Consortium. 2010. The OWL Web Ontology Language. http://www.w3.org/TR/owl-features/.
20. HP Labs. 2011. Jena – A semantic web framework for Java. http://www.openjena.org.
21. Detwiler LT, Suciu D, Brinkley JF. Regular paths in SparQL: Querying the NCI thesaurus. In: Proceedings, Fall Symposium of the American Medical Informatics Association; 2008. p. 161–165. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2656016/ PMCID: PMC2656016.

