Application of Information Technology ■

Processes and Problems in the Formative Evaluation of an Interface to the Foundational Model of Anatomy Knowledge Base

LINDA G. SHAPIRO, PHD, EMILY CHUNG, BSC, LANDON T. DETWILER, BSC, JOSÉ L. V. MEJINO, JR., MD, AUGUSTO V. AGONCILLO, MD, JAMES F. BRINKLEY, MD, PHD, CORNELIUS ROSSE, MD, DSC

Abstract The Digital Anatomist Foundational Model of Anatomy (FMA) is a large semantic network of more than 100,000 terms that refer to the anatomical entities, which together with 1.6 million structural relationships symbolically represent the physical organization of the human body. Evaluation of such a large knowledge base by domain experts is challenging because of the sheer size of the resource and the need to evaluate not just classes but also relationships. To meet this challenge, the authors have developed a relation-centric query interface, called *Emily*, that is able to query the entire range of classes and relationships in the FMA, yet is simple to use by a domain expert. Formative evaluation of this interface considered the ability of *Emily* to formulate queries based on standard anatomy examination questions, as well as the processing speed of the query engine. Results show that *Emily* is able to express 90% of the examination questions submitted to it and that processing time is generally 1 second or less, but can be much longer for complex queries. These results suggest that *Emily* will be a very useful tool, not only for evaluating the FMA, but also for querying and evaluating other large semantic networks.

■ **J Am Med Inform Assoc.** 2005;12:35–46. DOI 10.1197/jamia.M1401.

The University of Washington Digital Anatomist (UWDA) vocabulary[1] was initially established to facilitate the correlation of anatomical concepts within the National Library of Medicine's Unified Medical Language System (UMLS).[2] UWDA's domain encompasses macroscopic and microscopic anatomy for all parts of the body and also includes, in a consistent and continuous semantic structure, extensive representations of subcellular and macromolecular anatomical entities. The latest UWDA version contains nearly 70,000 classes of anatomical entities associated with nearly twice as many terms. The authors have defined a high-level scheme for the UWDA knowledge base, enhanced it with 150 new kinds of relationships, and transformed it into a disciplined, expressive ontology. This enhanced, computable, anatomical knowledge source is known as the *Digital Anatomist Foundational Model of Anatomy* (FMA for short) (available at http://fma.biostr.washington.edu).[3–5]

Affiliation of the authors: Structural Informatics Group (LGS, EC, LTD, JLVM, AVA, JFB, CR), Department of Biological Structure (LTD, JLVM, AVA, JFB, CR), Department of Medical Education and Biomedical Informatics (LGS, JFB), Department of Computer Science and Engineering (LGS, EC, JFB) and the Department of Electrical Engineering (LGS), University of Washington, Seattle, WA.

Supported by National Library of Medicine grants LM007714 and LM06822 and Human Brain Project grant DC02310.

The name *Emily* is in honor of the second author, Emily Chung, who developed the initial prototype of this program as a summer undergraduate project.

Correspondence and reprints: Linda G. Shapiro, PhD, Box 352350, Computer Science and Engineering, 634 Paul Allen Center, University of Washington, Seattle, WA 98195; e-mail: <shapiro@cs.washington.edu>.

Received for publication: 05/27/03; accepted for publication: 09/07/04.

The FMA is a resource for many different groups of users. Anatomists may want to compare it with their own terminologies or other published compendia. Scientists studying other species can use it as a basis for comparison. Students of anatomy and the general public can use it to help them learn about the human body at their desired level of complexity. However, to be maximally useful, the FMA must be evaluated for accuracy and comprehensiveness.

Evaluation of the FMA presents challenges distinct from those described in published evaluations of UMLS and other controlled medical terminologies, which have largely focused on the comprehensiveness of concepts.[6–9] By contrast, evaluation of the FMA must also include assessment of the comprehensiveness and validity of the relationships that the FMA explicitly models. To ensure optimal results, evaluation must include participation of domain experts, in this case, anatomists. With a knowledge base the size of the FMA, it is not feasible to simply ask the experts to examine the entire model, paying particular attention to its relationships.

The FMA has been implemented in the Protégé-2000 frame-based system,[10,11] which has proved advantageous for authoring and curating the knowledge base. However, given the size of the knowledge base, it would be time-consuming and problematic for FMA-naive users to gain proficiency in the navigation of Protégé-2000. Furthermore, Protégé-2000 provides only a browsing interface, not a full query interface. Evaluators must therefore have access to an intuitive, easy-to-use query interface that allows them to ask systematically designed questions of the knowledge base. To meet this need, the authors have developed a simple and intuitive graphical user interface, called *Emily*, which allows the submission of queries composed of any combination of entities and relationships represented in the knowledge base. Formative evaluation of *Emily* shows that it is capable of expressing most queries given to it, it is easy to use by domain experts, and its response time is adequate for use in an evaluation study.

In the remainder of this paper, we describe the interface and its formative evaluation. We conclude that, with only small additions, the interface will become a useful evaluation tool, not only for the FMA, but also for other large semantic networks.

Background

The FMA, constructed using Protégé-2000,[10,11] explicitly defines classes of anatomical entities and relationships necessary for consistently representing the structure of the idealized human body. Protégé-2000 represents this knowledge, which is a semantic network, in a frame-based system.

Population of the FMA with terms that refer to anatomical entities is guided by a high-level abstraction composed of knowledge elements that the authors consider necessary for comprehensively modeling the structural organization of the human body.[4,5] These knowledge elements include the following:

- *Anatomy Taxonomy* (AT), an inheritance type hierarchy of anatomical entities;
- *Anatomical Structural Abstraction* (ASA), which specifies the structural relationships of the entities represented in AT;
- *Anatomical Transformation Abstraction* (ATA), which describes the morphological transformations of the entities represented in AT during the human life cycle (including prenatal development, postnatal growth, and aging);
- *Metaknowledge* (Mk), which comprises the principles, rules, and definitions according to which relationships are to be represented in the three other knowledge elements.

The classes of the AT represent entities at all levels of the biological organization of the body from the macroscopic (e.g., brain) to the microscopic, submicroscopic, and molecular (e.g., neuron, mitochondrial crista, alpha-tubulin). The ASA is an aggregate of structural relationships that exist between anatomical entities.[12] The ASA includes a Dimensional Taxonomy, which defines geometric entities of 0–3 dimensions, provides a classification of three-dimensional (3-D) shapes, and describes topological relationships such as parts, containment, adjacency and qualitative coordinates, branches, connectivity, continuity, and attachment. Figure 1 illustrates a portion of the taxonomy of anatomical relationships that are currently being instantiated, and Figure 2 shows how subclasses of these relationships (italicized in the following text) are used in the frame for a single anatomical entity, the esophagus.

As shown in the frame for esophagus (right half of Figure 2), the esophagus is a 3-D entity whose *inherent 3-D shape* is a hollow cylinder. It is *bounded-by* the external surface of esophagus. Its immediate superclass is organ with organ cavity (subclass hierarchy, or AT, shown on the left); its instances (e.g., John Doe's esophagus) are by design excluded from the AT. It is *part-of* foregut and upper gastrointestinal tract and *has parts* wall of esophagus, lumen of esophagus, cervical part of esophagus, thoracic part of esophagus, and abdominal part of esophagus. It has no branches or tributaries. It is *contained-in* superior mediastinal space, posterior mediastinal space, left posterior subphrenic space, and space of anterior compartment of neck. It *contains* no other anatomical entities, since only anatomical spaces can have contents, and therefore its lumen *contains* esophageal secretion and bolus of food (which would be shown in the frame for lumen of esophagus). It is *continuous-with* pharynx and stomach, and it is *adjacent-to* trachea, thoracic vertebral column, and thoracic aorta.

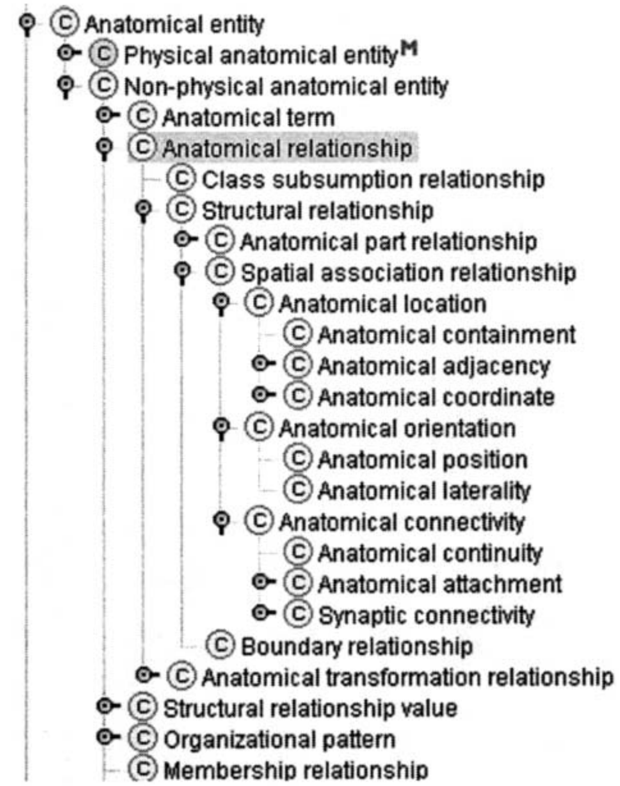

Figure 1. Some of the subclasses of anatomical relationship in the anatomy taxonomy (*is-a* hierarchy) of the Foundational Model of Anatomy. The class hierarchy (AT) is shown as it is viewed through the Protégé-2000 user interface.

This single example shows only a very small number of the 1.6 million relationships that are present in the FMA. As shown in Figure 2, it is possible to browse the FMA by clicking the values of specific relationships (e.g., clicking lumen of esophagus in Figure 2 will navigate to the frame for lumen of esophagus). It is also possible to browse the FMA over the Web using the online Foundational Model Explorer,[13] which presents a browsing interface based on Protégé. However, the sheer size of the FMA precludes browsing as a meaningful way to evaluate its accuracy and completeness.

The authors have developed a query interface to the FMA, called OQAFMA,[14] which accepts queries in the StruQL database query language.[15] Although end-user applications have been built on top of OQAFMA, the StruQL query language alone is too difficult for nonprogrammers.

Other graphical query languages have been developed for Protégé knowledge bases. For example, the ShriMP (Simple Hierarchical Multi-Perspective) visualization technique[16] has been made available to Protégé-2000 users through the Jambalaya interface.[17] ShriMP uses a nested graph view to present the semantic Web to users. Exploring the FMA

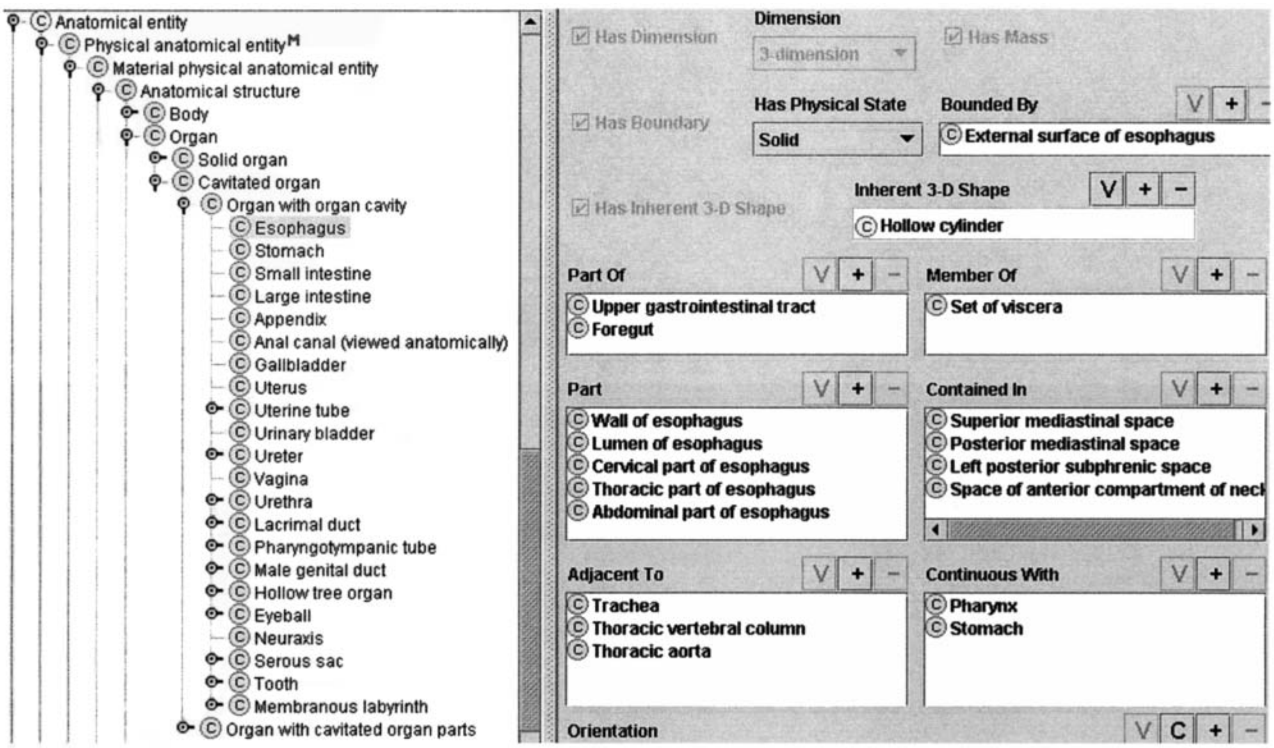

Figure 2. Structural relationships implemented within the frame of Esophagus in Protégé-2000.

through ShriMP/Jambalaya would only provide browsing, not querying capabilities.

Query interfaces have been developed for other large biomedical knowledge bases such as GALEN[18] through the GRAIL concept modeling language.[19] However, since this and many other knowledge bases are expressed in description logic rather than frames, the methods are not directly applicable to the FMA.

Design Objectives

The main functional objective of the *Emily* query interface is to allow users to submit queries concerning the many structural relationships of the ASA. Design objectives include the following: (1) the power to query both direct and closure relationships among anatomical entities, (2) the ability to ask for unknown relationships among given entities, (3) the ability to combine basic queries to ask more complex questions, and (4) a simple method for entering these queries so that nonprogrammers would require very little training to do so.

System Description

The *Emily* query interface to the FMA is a Java application, which accesses the Protégé-2000 API (a Java programming interface) to communicate with the FMA knowledge base (which itself is contained in a MySQL relational database). The *Emily* graphical user interface allows the user to pose *basic* queries that involve a single structural relationship between two anatomical entities and *compound* queries that involve more than one relationship. The names of anatomical entities and relationships available in *Emily* have all been based directly on FMA terms and relationships and include intuitive relationship name synonyms. *Emily* translates user-formulated queries into the appropriate method calls to the underlying Protégé-2000 library, which in turn retrieves the appropriate FMA data from the MySQL relational database. *Emily* reformats the values returned for display within its graphical interface.

Types of Queries

Emily can process two kinds of queries: basic and compound. A *basic query* has the form: < Subject Relation Object > where Subject and Object can be any anatomical entity or can be Unknown, and Relation specifies one of the structural relationships of the ASA or can be Unknown. The *part-of* relationship allows the user to submit queries for *is part of (directly)*, which would return the terms that refer only to those entities of which a given entity is a direct part, and for *is part of*, which would return those terms that refer to entities of which the given entity is a part in the closure sense: the entities of which the given entity is a direct part, the entities of which those entities are direct parts, and so on. For example, wall of esophagus *is part of (directly)* esophagus, but esophagus *is part of (directly)* foregut, *which is part of (directly)* gut, which *is part of (directly)* abdomen, which *is part of (directly)* trunk, which *is part of (directly)* body. Thus, in the closure sense, wall of esophagus *is part of* esophagus, foregut, gut, trunk, and body (among other entities).

Examples of basic queries include the following:

1. Esophagus *is contained in (directly)* unknown
2. Esophagus *is contained in* unknown
3. Unknown *is contained in* lumen of esophagus

4. Unknown *is part of* esophagus
5. Wall of esophagus *is part of* esophagus
6. Esophagus *unknown* gut
7. Wall of esophagus *unknown* stomach

Queries 1–4 each have one unknown and should return a set of zero or more anatomical entities. Query 5 is a yes/no question since it contains no unknowns; the response should be yes. Query 6 contains an unknown relationship between esophagus and gut. The response should be that Esophagus is *part-of* gut. Note from Figure 2 that Esophagus is *part of* Foregut, which is *part of* gut (not shown in Figure 2). Thus, query 5 illustrates the need to traverse relationship paths of lengths greater than 1. Similarly, query 7 asks for the unknown relationship between wall of esophagus and stomach. This query requires *Emily* to search through the database to produce the response that wall of esophagus *is continuous with* wall of stomach, which is *part of* stomach.

Compound queries allow the user to ask questions involving more than one relationship. They can be formulated in two different ways: (1) sets of linked queries, and (2) Boolean combinations of several queries. Sets of linked queries may generate or use variables whose values are sets of anatomical entities. Two queries can be linked by a common variable. An example is the query illustrated in Figure 3:

Unknown *is part of (directly)* Esophagus

generates the set {*part of esophagus, Lumen of esophagus, Wall of esophagus, Cervical part of esophagus, Thoracic, Abdominal part of esophagus*}. This set is then assigned to a variable, such as U1, which is automatically created and labeled by *Emily.* A second query (illustrated in Figure 4)

U1 *is continuous with (directly)* Unknown

looks for the *is continuous with (directly)* relationship between each separate element of U1 and other anatomical entities. The result is a tree structure that contains each element of U1 at the top level and the entities that satisfy the query as children of these top-level entities. For the above query, the tree structure contains the following information (where :: indicates children of an entity with respect to the continuous relationship).

Wall of esophagus::(Wall of stomach, Wall of pharynx)
Lumen of esophagus::(Cavity of pharynx, Cavity of stomach)
Cervical part of esophagus::(Pharynx, Thoracic part of esophagus)
Thoracic part of esophagus::(Cervical part of esophagus, Abdominal part of esophagus)
Abdominal part of esophagus::(Thoracic part of esophagus, Cardia of stomach)

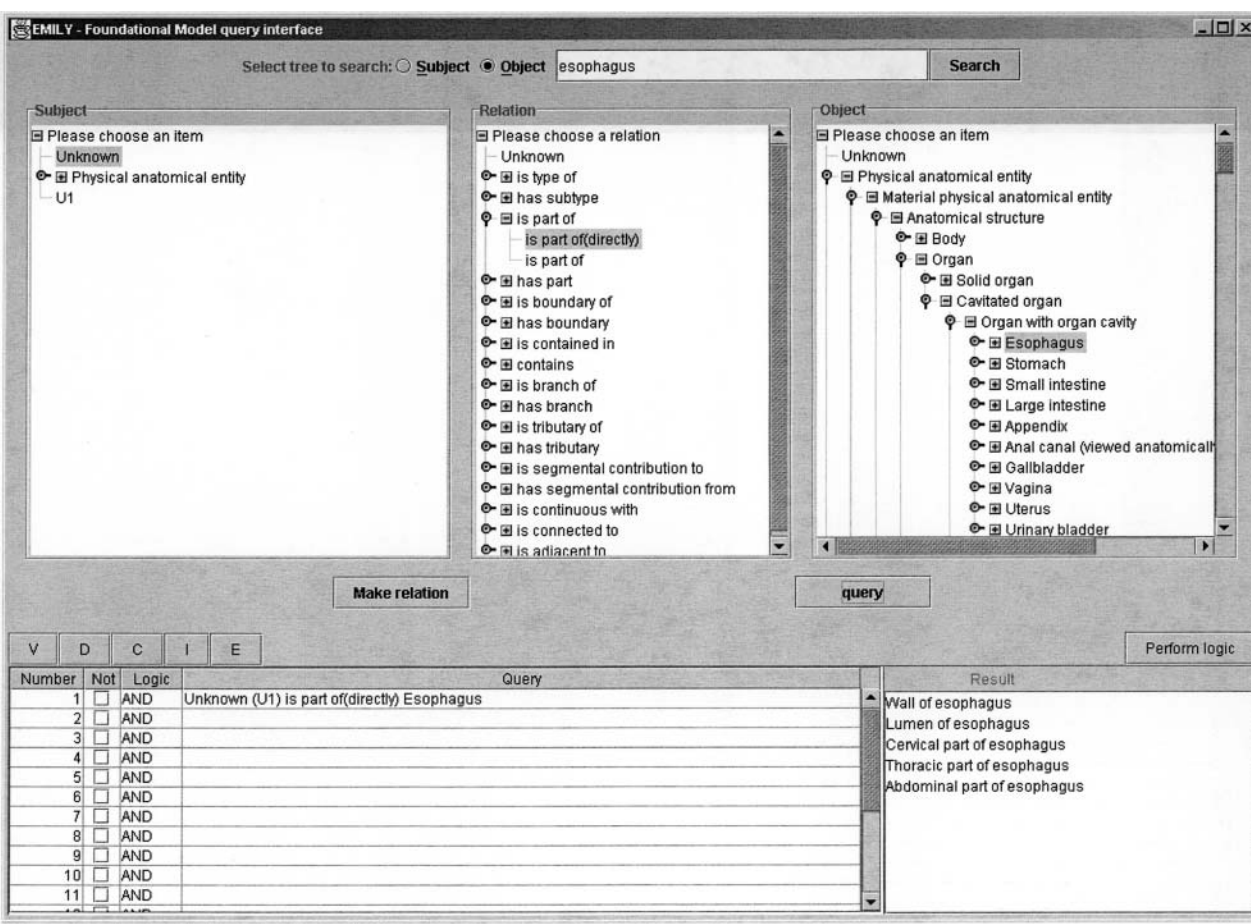

Figure 3. The *Emily* graphical user interface: processing of a basic query. The user has selected Unknown *is part of (directly)* Esophagus and clicked on the query button. The query is shown as a string in the Query column of the lower portion of the interface (for use in future compound queries), and the result set is displayed to its right in the Result column. The result set has been assigned to the system-generated variable U1. This set can now be used to formulate a new query U1 *is continuous with (directly)* Unknown, as discussed in the text. The interface for the compound query consisting of these two linked queries is shown in Figure 4.

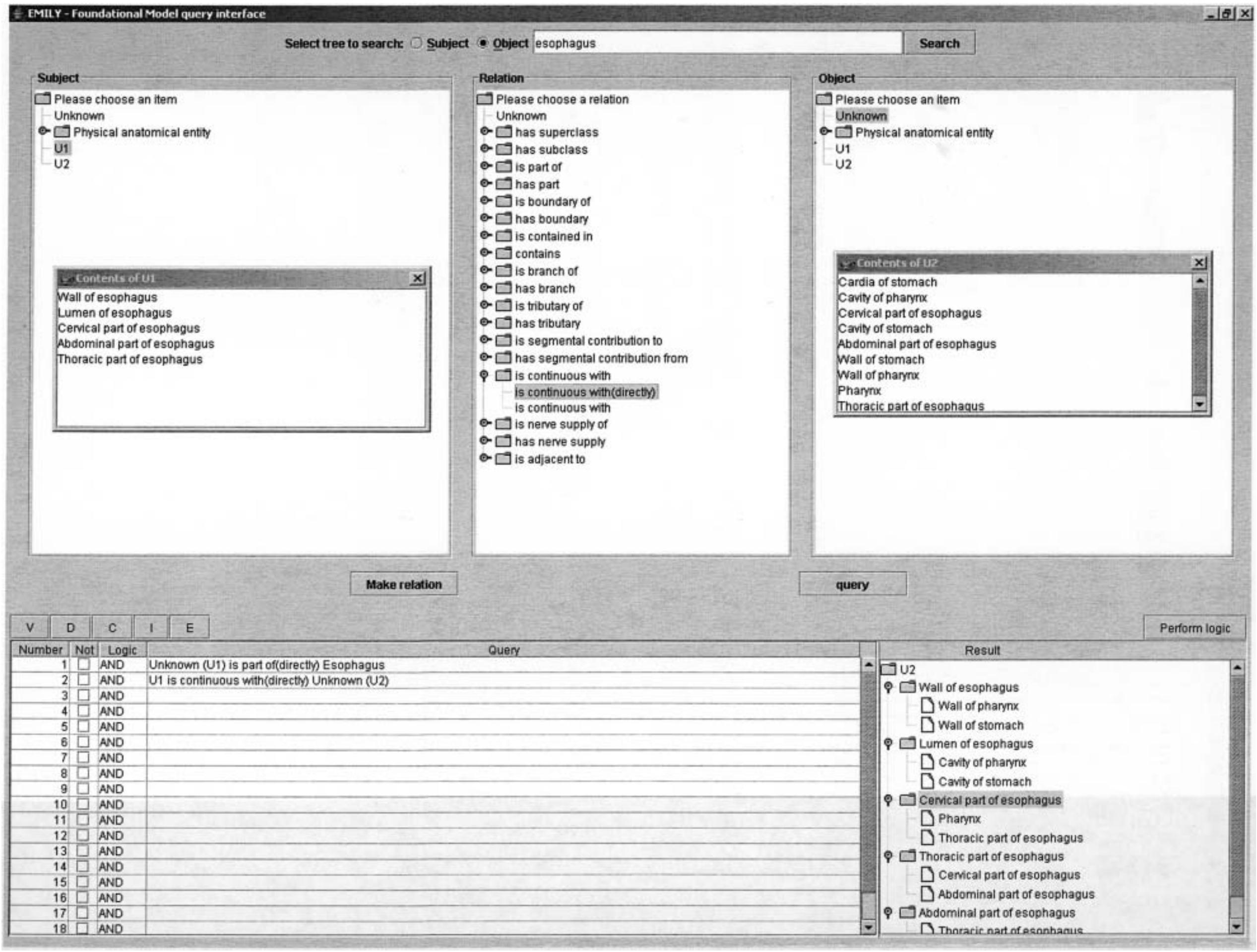

Figure 4. The *Emily* graphical user interface: processing of a compound query that is a set of linked queries derived from the Result shown in Figure 3. At the time of the snapshot, both of the linked queries have been executed, and the returned tree has been opened to show its full structure. Set U1 contains the result of the first query, and set U2 contains the results of the second query. Part of the returned tree structure is visible in the Result column.

In addition to the tree structure displayed as the query result, another variable, U2, is assigned to the set {Wall of stomach, Wall of pharynx, Cavity of pharynx, Cavity of stomach, Pharynx, Thoracic part of esophagus, Cervical part of esophagus, Abdominal part of esophagus, Cardia of stomach}, which is the set of leaves of the tree.

For Boolean combinations of several queries, results produced by each query are combined via Boolean operators. For example, if one asks, "What is continuous with the abdominal part of the esophagus that is not part of the esophagus?," the *Emily* query would be:

Unknown *is continuous with (directly)* Abdominal part of esophagus

and not

Unknown *is part of (directly)* Esophagus

The first query produces the set {Thoracic part of esophagus, Cardia of stomach}, the second produces {Wall of esophagus, Lumen of esophagus, Cervical part of esophagus, Thoracic part of esophagus, Abdominal part of esophagus}. The Boolean combination produces the set {Cardia of stomach}.

Answering "Unknown" Queries

Unknown relationship queries (such as Query 7 in the previous section) are interesting because there are many potential answers to a single query, but most of them are not very useful. *Emily's* unknown-relationship strategy is threefold:

1. The system will first search for direct and closure relationships between the two entities. The database can be indexed, so that finding either of these takes constant time.
2. Next, the system will search for specific, predefined, relational sequences that have been identified as important. For example, the composite relationship described by the regular expression *(has parts)*contains* specifies that any number of *has parts* relationships should be followed by one *contains* relationship. This composite relationship holds between heart and blood, since heart *has part* right atrium, which *has part* cavity of right atrium, which *contains* blood. Since regular expression components involve only direct relationships and closure relationships, indexing can be used to accelerate searches.
3. If no direct, closure, or predefined composite relationship is found between the two selected entities, the system will resort to a depth-limited, breadth-first search. Our

current system can search four levels of the FMA structure in an acceptable amount of time.

The Graphical User Interface

The *Emily* graphical user interface is shown in Figures 3 and 4. The upper part of the screen is for entry of basic queries. The user can select an anatomical entity or Unknown from each of the fields: Subject, Relation, and Object. When selecting an anatomical entity, the user can type in a term (or its synonym) directly into the top search box labeled "Select tree to search" or select a term by browsing through the entire hierarchy of terms from the AT (*is-a* hierarchy) of the FMA. Clicking on the Query button causes *Emily* to translate the query into appropriate calls to the Protégé API and to then display the result next to the representation of the query in the lower part of the screen. A basic query is processed in Figure 3. Figure 4 shows a compound query with its returned results. The search facility allows wild-card searches. For example, if a user types the string "cardiac*" into the search field, *Emily* returns the list of FMA terms starting with the word cardiac (Cardiac apex, Cardiac atrium, Cardiac border, etc.). Users can review previous results to select a prior term as a new query component.

Status Report

In this section, the authors describe a formative evaluation of the potential for *Emily* to be used as a tool for evaluation of the FMA.

To gain some independent measure of the reliability of *Emily*, the authors used two of several published compendia of anatomy examination questions[20,21] to establish whether *Emily* could process items selected from them. First, nine multiple-choice questions were selected which the authors considered representative of the different kinds of challenges that anatomists might encounter in presenting examination items to the FMA through *Emily*. The authors analyzed the results to better understand *Emily*'s capabilities for handling items of different levels of difficulty. Next, the authors performed a quantitative analysis on a different, larger set of question items to determine the proportion that *Emily* could answer correctly and to analyze the reasons for any failures. Finally, the authors performed a timing study on a third set of selected items to estimate *Emily*'s efficiency for query processing.

Selection of Representative Query Items for the Evaluation

Two basic difficulties with selecting examination questions arose: (1) due to the nature of the science of anatomy (which, like the FMA, is concerned strictly with describing the structure of the body), question items that required integration of knowledge of function, development, and even clinical correlation with pure structure-related information had to be eliminated (i.e., dropped from use in the evaluation) and (2) discrepancies in specificity between English-language expressions in the questions and the FMA, both in the naming of anatomical structures and of relationships, forced evaluators to translate "native" examination items into a format suitable for *Emily*. The translation from English examination questions to *Emily* queries was performed by a group of project anatomists. While several of the English expressions translated directly to relationships, others required some amount of human thought and anatomical knowledge to develop. The evaluation used the translations for the nine examination items, which are illustrated in detail in Appendix 1. For example, the first case in Appendix 1 involved an easy translation of "is continued as" from the examination item to "is continuous with (directly)" in *Emily*. The second example involved more complex translation of "supplies" to "is-nerve-supply-of" in *Emily*, as well as "arises from" to "is branch of (directly)" and "notches" to "is adjacent to." In the fourth example, a shortcoming of *Emily* was uncovered in that "lies anterior to" and "is lateral to" in examination items could only be translated as "is adjacent to" in *Emily*. The FMA representation allows attribute slots for "is adjacent to" that can hold values such as "lateral to" and "anterior to," but the current version of *Emily* cannot access those attribute slots. Several other examples illustrate the use of synonyms by *Emily*.

Emily's Ability to Return Correct Answers

This study attempted to measure the proportion of questions for which *Emily* could process and provide a correct answer. The evaluation secondarily attempted to gain insight into causes of query failures. The questions for this exercise were chosen from only one of the two reference texts.[20] The first 100 questions in each of the first seven chapters of this text were captured, considering each option in a multiple choice item as a "separate question" (i.e., a multiple choice question with five options counted as five separate questions for submission to *Emily*). In all, 700 questions were reviewed and those that required anything other than structural knowledge of human anatomy were removed, leaving 486 questions (see Table 1). The authors then eliminated those questions containing relationships not found in the FMA. Finally, 100 items were randomly selected from the remaining 412 candidate questions. Table 1 shows the results of this component of the study, broken down by chapters of the text.

As shown in Table 1, of the 100 questions that the authors attempted to submit to *Emily*, ten were answered correctly, ten could not be formulated due to *Emily's* inability to handle attributed relationships, and 80 were not answered because the required data had not been entered as yet into the FMA database. However, when the authors entered the data required by ten questions randomly selected out of the 80 unanswered questions, *Emily* answered all ten of them correctly. The following example illustrates the process:

Question. The greater sciatic foramen transmits the following structures, *EXCEPT*:

1. Superior gluteal vessels
2. Posterior cutaneous nerve of thigh
3. Piriformis muscle
4. Obturator internus
5. Inferior gluteal vessels

Answering this question through *Emily* first required the translation of the active term "transmits" into a structural relationship represented in the FMA. A foramen is classified in the FMA as an Anatomical conduit, which *contains*, rather than transmits diverse anatomical structures. Therefore, the authors translated this question into the following *Emily* query:

Greater sciatic foramen contains Unknown

Upon initially submitting this query, *Emily* returned an empty result set. Inspection of the knowledge base revealed that the values for the *contains* relationship in the frame of Greater

Table 1 ■ *Emily's* Performance on a Randomized Sample of Queries Derived from Published Anatomy Examination Questions

Chapter	Title	Purely Structural	Relation in FMA	Attempted by *Emily*	Answer Correct	Data Not in FMA	Problem Due to Inaccessible Attribute of Relation
1	Upper Limb	70	64	12	1	8	3
2	Lower Limb	69	63	16	3	11	2
3	Thorax	84	60	13	4	9	0
4	Abdomen	72	58	17	1	16	0
5	Pelvis and Perineum	87	78	20	0	17	3
6	Head and Neck	71	56	7	1	4	2
7	Nervous System	33	33	15	0	15	0
	Total	486	412	100	10	80	10

FMA = Foundational Model of Anatomy.

sciatic foramen had not been entered. Not surprisingly, when the query was resubmitted after entering the appropriate data, *Emily* returned the result set: {Superior gluteal artery, Inferior gluteal artery, Superior gluteal vein, Inferior gluteal vein, Superior gluteal nerve, Inferior gluteal nerve, Piriformis, Posterior femoral cutaneous nerve, Internal pudendal artery, Internal pudendal vein, Pudendal nerve, Sciatic nerve, Nerve to obturator internus, Nerve to quadratus femoris}.

Query Efficiency

To evaluate *Emily's* response time, the authors chose a set of ten representative queries, of varying degree of difficulty, and timed their processing. All efficiency tests were performed with both the *Emily* application and a local MySQL DBMS running on a PC with a 2.60 GHz Intel Pentium 4 processor and 1 GB of RAM. Each query was chosen because it was representative of a particular class of queries. The queries, along with a brief description, were as follows:

1. Heart *has part (directly)* Unknown
This query is a simple direct query but on a heavily populated slot (32 values).
2. Heart *has part* Unknown
This query is the transitive closure of query 1 and yields a highly populated result list (472 values).
3. Heart *has boundary (directly)* Unknown
Like 1, this is a simple direct query but on a lightly populated slot (1 value).
4. Heart *has boundary* Unknown
The transitive closure of 3, again with a lightly populated result set (1 value).
5. Heart *Unknown* Esophagus
An unknown relation query where the two entities are directly related by a single edge.
6. Heart *Unknown* Wall of right atrium
An unknown relation query that requires transitive closure (two edges).
7. Heart *Unknown* Pharynx
An unknown relation query that requires breadth first search (mixed relations three edges deep).
8. Right eye *Unknown* Heart
An unknown relation query for which there is no answer found. Queries with no answers typically take longest because *Emily* must search the entire tree from the subject node to a depth of 4 to determine that there is no answer.
9. Heart *Unknown* Right eye
The same unknown relation query as in 8 but with the subject and object transposed. While this appears to be the same query as 8, it is interesting because it illustrates the point that the query time is a function of the branching factor of the subject tree, not the object tree.
10. Heart *is adjacent to (directly)* unknown AND stomach *is continuous with* unknown
This query is actually a Boolean combination of two other queries: 1. Heart *is adjacent to (directly)* unknown, and 2. Stomach *is continuous with* unknown. The query time is the sum of the times to answer both subqueries (125 milliseconds for the first and 63 milliseconds for the second) plus the time it took to perform the Boolean AND operation (<1 millisecond). Processing times for these ten queries are illustrated in Table 2. Some queries were repeated to illustrate the effect of precaching data from the database (Sequential Run column).

Discussion

The formative evaluation indicates that the *Emily* relation-centric query interface allows anatomists who are nonprogrammers as well as programmers not trained in anatomy to enter both simple and relatively complex queries concerning the structural relationships among anatomical entities of the human body. About 90% of the structural queries selected from the published compendia of anatomy questions could be translated into the format required by *Emily*. Relatively

Table 2 ■ Processing Times of Ten Queries (Described in the Text) by *Emily*

Query No.	Sequential Run (Higher Runs Use Cached Data)	Time (msec)
1	1	328
1	2	16
2	1	2,672
2	2	531
3	1	47
3	2	15
4	1	16
4	2	<1
5	1	672
5	2	<1
6	1	2,703
6	2	94
7	1	9,594
7	2	828
8	1	42,608
9	1	198,591
10	1	188

simple extensions to the program should allow most, or all, of the query types to be answered if the content is present in the FMA. In addition, the response time for all but the most unlikely queries (queries 8 and 9 in Table 2) is acceptable for use in an interactive application.

However, the results also indicate a substantial need for human translation of English-language expressions into terms and relationships compatible with *Emily* and FMA representations. This need arises because anatomy questions in compendia of the type used in this evaluation[20,21] tend to use general terms and homonyms (the meaning of which is provided by the question's context), whereas the FMA terms are highly specific. Many of these translations require knowledge of anatomy. This requirement might not present a problem for well-trained anatomists who will be recruited to evaluate the FMA (although it might present a more substantial problem for novice students). The anatomists involved in the current formative evaluation were able to translate all options in the selected multiple choice type examination questions (see Appendix 1) into a format processable by *Emily*, except when these examination questions contained FMA relationships whose attributes were inaccessible to *Emily*. The answers to the nine translated questions in Appendix 1, derived from the results returned by *Emily*, were consistent in all instances with the published keys for the original questions. However, as noted, significant deficiencies in the comprehensiveness of the FMA prevented *Emily* from returning correct responses in 80 of 100 queries (Table 1).

These results suggest that, in the hands of domain experts, *Emily* may become a useful tool for evaluation of the FMA. Unless the evaluation is to be limited to low-level knowledge elements, such as comprehensiveness of content and equivalence of terms, the involvement of domain experts in the evaluations becomes inevitable. The involvement of domain experts has been advocated in the evaluation of medical informatics systems,[22] but cautions have also been sounded about such a strategy.[23] If domain experts, such as anatomists, look to their domain's time-honored sources as gold standards for evaluating a machine-based knowledge system, the information they provide will be of limited value to the system's developers. Therefore, these evaluators must be provided with insights into the different requirements for representing knowledge in traditional media versus formal systems, and they must be educated about the conceptual design of the latter systems. Meaningful input can be expected only from those participants in the analysis who have grasped the rationale accounting for the inherent differences between hardcopy and formal knowledge sources. Such an understanding will enable the evaluators to make the kinds of translations that the authors had to generate for the anatomy examination items.

The authors have learned a number of valuable lessons from this work. First, the relation-centric query format provided by *Emily* does allow most structural queries to be answered, but some cannot be formulated due to the inaccessibility of relationship attributes through the interface. There are many important structural questions that require attributed relationships, which the FMA provides but *Emily* does not yet answer. Second, the *Emily* interface was found to be easy to use by the anatomists who conducted the experiments. In particular, *Emily* allowed complex queries to be formulated that (1) could not be asked via the Protégé-2000 interface without a huge amount of search carried out interactively by a knowledgeable human and (2) could not be asked via a database query language without a large amount of knowledge about the structure of the database and some amount of programming capability. On the negative side, the authors have learned that some domain knowledge is required both for limiting queries to those that a domain-specific knowledge base can answer and for decomposing queries into the format that *Emily* can handle. A table that translates commonly used language into the more precise terminology of the FMA could be helpful here, but the construction of such a table is a large task and is not part of our current design plans. Finally, the biggest lesson learned was that without such a query interface, it would be very difficult to detect gaps in a knowledge base the size of the FMA.

The *Emily* interface is not specific to the FMA knowledge base; it could potentially be used with any large semantic network. The authors, however, are mainly interested in its use for both accessing and evaluating the FMA. To this end, *Emily* is being updated to handle attributed relationships and to add more power to the unknown relationship queries. A version of the program called *Emily Light* is being developed that can be executed from a Web site, without any downloading or installation required. This version will help the authors to identify gaps in the knowledge base and will make it easier for domain experts to evaluate the FMA, once it is more fully populated.

References ■

1. Rosse C, Ben Said M, Eno KR, Brinkley JF. Enhancements of anatomical information in UMLS knowledge sources. In: Gardner R, ed. Proceedings of the 19th Annual Symposium on Computer Applications in Medical Care, New Orleans, LA: Hanley & Belfus; 1995:873–7.
2. Lindberg DAB, Humphreys BL, McCray AT. The unified medical language system. Methods Inf Med. 1993;32:281–91.
3. Rosse C, Mejino JL, Modayur BR, Jakobovits RM, Hinshaw KP, Brinkley JF. Motivation and organizational principles for anatomical knowledge representation: the Digital Anatomist symbolic knowledge base. J Am Med Inform Assoc. 1998;5:17–40.
4. Rosse C, Shapiro LG, Brinkley JF. The Digital Anatomist foundational model: principles for defining and structuring its concept domain. Proc AMIA Fall Symp. 1998:820–4.
5. Rosse C, Mejino JLV. A reference ontology for bioinformatics: the Foundational Model of Anatomy. J Bioinform. 2003;36:478–500.
6. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. J Am Med Inform Assoc. 1996;3:224–33.
7. Humphreys BL, McCray AT, Cheh ML. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. J Am Med Inform Assoc. 1997;4:484–500.
8. McCray AT, Chech ML, Bangalore AK, et al. Conducting the NLM/AHCPR Large Scale Vocabulary Test: a distributed Internet-based experiment. Proc AMIA Fall Symp. 1997:560–4.
9. Bodenreider O, Burgun A, Botti G, Fieschi M, Le Beaux P, Kohler F. Evaluation of the Unified Medical Language System as a medical knowledge source. J Am Med Inform Assoc. 1998;5:76–87.
10. Grosso WE, Eriksson H, Fergerson RW, Gennari JH, Tu SW, Musen MA. Knowledge modeling at the millennium (the design and evolution of Protege-2000). Presented at the Twelfth Banff Workshop on Knowledge Acquisition, Modeling and Management, Banff, Alberta, 1999.

11. Noy NF, Musen MA, Mejino JLV, Rosse C. Pushing the envelope: challenges in a frame-based representation of human anatomy. Data Knowledge Engineering. 2004;48:335–59.
12. Mejino JLVJ, Rosse C. Symbolic modeling of structural relationships in the Foundational Model of Anatomy. Presented at the First International Workshop on Formal Biomedical Knowledge Representation (KR-MED 2004), Whistler Mountain, Canada, 2004.
13. Detwiler L, Mejino J, Rosse C, Brinkley J. Efficient Web-based navigation of the Foundational Model of Anatomy. Proc AMIA Symp. 2003:829.
14. 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;36:501–17.
15. Fernandez M, Florescu D, Levy H, Suciu D. A query language for a web site management system. SIGMOD Record. 1997;26:4–11.
16. Storey M, Wong K, Fracchia P, Müller HM. On integrating visualization techniques for effective software exploration. In: Proceedings, IEEE Symposium on Information Visualization, 1997, pp 38–45.
17. Storey M, Musen M, Silva J, et al. Jambalaya: interactive visualization to enhance ontology authoring and knowledge acquisition in Protege. Presented at: Workshop on Interactive Tools for Knowledge Capture, 2001; Victoria, BC.
18. Rogers J, Roberts A, Solomon D, et al. GALEN ten years on: tasks and supporting tools. In: Patel V (ed). Proceedings of the 10th World Congress on Health and Medical Informatics, MEDINFO 2001. Amsterdam: IOS Press, 2001, pp 256–60.
19. Rector A, Bechhofer S, Goble C, Horrocks I, Nowlan W, Solomon W. The GRAIL concept modelling language for medical terminology. Artif Intell Med. 1997;9:139–71.
20. Fitzgerald MJT, Golden JP. Anatomy 1600 Multiple Choice Questions. London: Butterworth, 1973.
21. Wilson JL. Anatomy 700 Questions: A USMLE Step I Review, 10th ed. East Norwalk, CT: Appleton and Lange, 1995.
22. Hripcsak G, Wilcox A. Reference standards, judges, and comparison subjects: roles for experts in evaluating system performance. J Am Med Inform Assoc. 2002;9:1–15.
23. Miller RA. Reference standards in evaluating system performance. J Am Med Inform Assoc. 2002;9:87–8.
24. Rosse C, Gaddum-Rosse P. Hollinshead's Textbook of Anatomy, 5th ed. Philadelphia: Lippincott, 1997.

Appendix 1 ■ Examples of Translations of Exam Questions for *Emily*

1. The spine of the scapula is continued as the
   a. Coracoid process
   b. Angle of the scapula
   c. Infraglenoid tubercle
   d. Supraglenoid tubercle
   e. Acromion process

Item 1 is a straightforward one for *Emily*; it exemplifies items requiring easy translation. The phrase "is continued as" translates directly to the ASA relationship *is continuous with (directly)*. Thus, the query to *Emily* can be phrased as

Spine of scapula *is continuous with (directly)* Unknown

for which *Emily* returns the result set {Acromion, Body of scapula}, which indicates that the correct answer is e. This question (and most of the others) can also be solved by executing several Boolean queries, such as

Spine of scapula *is continuous with (directly)* Coracoid process for which *Emily* will answer *no*.

2. The suprascapular nerve
   a. Supplies the infraspinatus muscle
   b. Arises from the lateral cord of the brachial plexus
   c. Notches the axillary border of the scapula
   d. All of the above
   e. A and B only

Item 2 illustrates the need for substantial translation, calling for knowledge of anatomy and of the FMA. Options a, b, and c require the translation of the phrases "supplies," "arises from," and "notches" since they are absent from the Foundational Model, as such. "Supplies" is a functional relationship and was translated into *is nerve supply of.* "Arises from" was translated to *is branch of (directly)*, and the term "notches" (which erroneously implies an active process) was translated as *is adjacent to*. Checking the validity of choices a, b, and c can be done with the following Boolean queries:

Suprascapular nerve *is nerve supply of* Infraspinatus.
Suprascapular nerve *is branch of (directly)* Lateral cord of brachial plexus.
Suprascapular nerve *is adjacent to* Axillary border of scapula.

The answer to the first query is *yes*, while the answer to the other two queries is *no*, which indicates that the correct answer to the examination question is choice a.

3. The thoracodorsal nerve
   a. Is a branch of the posterior cord of the brachial plexus
   b. Supplies the serratus anterior muscle
   c. Is cutaneous to dorsal surface thorax
   d. All of the above
   e. A and B only

Like item 2, item 3 presents some translation challenges to deriving the appropriate relationships for *Emily*, resulting in a need for inferring the correct answer from the returned results. To illustrate a different strategy, the authors translated options a and b as the following queries:

Thoracodorsal nerve *is branch of (directly)* Unknown
Thoracodorsal nerve *is nerve supply of* Unknown.

Anticipating that if the nerve had a cutaneous branch, *Emily* would return it, option c was submitted as

Thoracodorsal nerve *has branch (directly)* Unknown

For the first query, *Emily* returned Posterior cord of brachial plexus, indicating that option a is correct. For the second query, *Emily* returned Latissimus dorsi, indicating that option b is incorrect. Translation of option c remains problematic; however, *Emily* found no results in response to the third query, which suggests either that the thoracodorsal nerve has no branches or that no branches have been entered in the database. (The nerve, in fact, has no branches, but the FMA does not represent the absence of branches.) Regardless of the ambiguities of the translation and the results of the third query, both options d and e can be excluded on the basis of the information provided by *Emily*, leaving a as the correct answer, which tallies with the key. It may be of interest to note that the mechanics of many multiple-choice questions invite such reasoning from examination takers, which is independent of their knowledge of the domain that is being tested.

4. The axillary vein
   a. Is lateral to the axillary artery
   b. Is devoid of valves

c. Lies anterior to pectoralis minor
d. Is directly continuous with the brachiocephalic vein
e. None of the above

The authors selected this item to illustrate structural relationships that *Emily* is currently unable to handle. Options b and d are straightforward for *Emily*. Option d can be entered directly and the authors translate option b as

Axillary vein *has part (directly)* Valve of axillary vein.

Emily returns the answer *no* to each Boolean query. However, options a and c cannot be queried by the current version of *Emily*. The relationships "lateral" and "anterior" named in options a and c, respectively, are in fact attributes of the *adjacent to* slot of an anatomical entity. Although attributed relationships are represented in the FMA, the current version of *Emily* does not retrieve the attributes of relations. However, implementation of this capability is fairly straightforward and will be added to the next version.

5. The coronary sinus receives each of the following vessels **EXCEPT** the
   a. great cardiac vein
   b. middle cardiac vein
   c. anterior cardiac vein
   d. small cardiac vein
   e. posterior vein of the left ventricle

Item 5 illustrates the often-used examination item format that asks for an exception, which actually simplifies translation into *Emily's* format. Although it requires the translation of "receives" to the ASA relation *has tributary (directly)*, all four choices can be covered by a single query:

Coronary sinus *has tributary (directly)* Unknown

for which *Emily* returns the result set {Great cardiac vein, Posterior vein of left ventricle, Middle cardiac vein, Small cardiac vein, Oblique vein of left atrium}. This indicates that choice c is the required exception.

6. A tumor involving the fifth to twelfth thoracic vertebrae could affect each of the following structures in the posterior mediastinum **EXCEPT** the
   a. thoracic duct
   b. phrenic nerve
   c. azygos vein
   d. descending aorta
   e. esophagus

Item 6 is more typical of the type of questions included in anatomy examinations than the previous items. The challenge that it presents is that it does not explicitly state the relationship to be translated. The translated query is

Posterior mediastinum *contains (directly)* Unknown

and *Emily* returns the result set {Esophagus, Azygos vein, Descending aorta, Trunk of thoracic duct, Thoracic part of trunk of right vagus nerve, Thoracic part of trunk of left vagus nerve}. Since Phrenic nerve is not returned, choice b is the correct answer.

An alternative and more desirable (albeit more laborious) way to submit the question would be to query which of the structures listed in the options has an anterior adjacency in the posterior mediastinum to the fifth to twelfth thoracic vertebrae, which is the logical relationship to query. This approach, however, cannot be pursued until *Emily* can handle attributed relationships. Such an approach would reveal a flaw in the question, which calls for faulty reasoning. For the sake of expedience, our translation conforms to the faulty reasoning. The stem of the question restricts the options to structures located in the posterior mediastinum, yet the correct answer specified by the key can only be reasoned on the basis of a location other than the posterior mediastinum. The question provides an illustration of arriving at the right answer for the wrong reasons.

7. Each of the following is related to the lumen of the right ventricle **EXCEPT** the
   a. interventricular septum
   b. trabeculae carneae
   c. bicuspid valve
   d. anterior papillary muscle
   e. septomarginal band

This is a good question on which to use *Emily*'s unknown relationship query. Since the FMA constrains the term *lumen* to tubular structures, the authors need to use instead the term *cavity*.

The FMA does not allow plural terms, and so the authors use the singular term Trabecula carnea instead of trabeculae carneae in option b. Since there is an anterior papillary muscle in both right and left ventricles, in option d, the authors use the term specific for the muscle in the right ventricle. The authors therefore translate the question into the following queries:

Cavity of right ventricle *Unknown* Interventricular septum
Cavity of right ventricle *Unknown* Trabecula carnea
Cavity of right ventricle *Unknown* Bicuspid valve
Cavity of right ventricle *Unknown* Anterior papillary muscle of right ventricle
Cavity of right ventricle *Unknown* Septomarginal band

Emily returns nonempty result sets for all except choice c, which is the correct answer. Options c and e illustrate the need for enabling *Emily* to recognize synonyms. In the FMA, "bicuspid valve" and "septomarginal band" are synonyms of the preferred names Mitral valve and Septomarginal trabecula, respectively. *Emily* searches synonyms and foreign language equivalents of the entities included in a query, and it always returns preferred names in the results. Option d illustrates the difference between the specificity of terms in the FMA and general anatomical discourse, a topic better addressed in relation to the next examination question.

8. The left coronary artery bifurcates into the circumflex branch and the
   a. left marginal branch
   b. left ventricular branch
   c. anterior interventricular branch
   d. right marginal branch
   e. posterior interventricular branch

For Item 8, the authors translated *left coronary artery* as Trunk of left coronary artery and "bifurcates into the … branch" as the *has branch (directly)* relationship, yielding the query

Trunk of left coronary artery *has branch (directly)* Unknown

which produced the result set {Trunk of anterior interventricular branch of left coronary artery, Trunk of circumflex coronary artery, Trunk of variant atrial branch of left coronary artery}; so that option **c** is the correct answer.

Item 8 (and also Item 7) illustrates the difference between the specificity of terms in general anatomical discourse and the FMA. The term Left coronary artery is, in fact, a homonym for two distinct entities, which are readily distinguished by the context of the English sentence in which the homonym is embedded. The FMA does not allow homonyms and uses specific terms for each entity. The meaning of the term in the stem of Item 8 is suggested by the expression "bifurcates." The FMA's preferred name for this entity is Trunk of left coronary artery. The other meaning of the term Left coronary artery is implied by the expression (used just as commonly in anatomical discourse) "… supplies the left ventricle, the interventricular septum," etc. This meaning encompasses an entire arterial tree, which includes the trunk and all its branches. The preferred name of this entity in the FMA is Left coronary artery. The trunk and branches of this tree are represented in the FMA as parts of the tree, and specific branching relationships are modeled between the trunk, branches, and subbranches to symbolically represent the specific structure of the tree.

Our translation of the term "bifurcates" as "*has branch (directly)*" is not sufficiently specific, as indicated by the return of three rather than two branches, as a bifurcation is expected to yield. A bifurcation yields *terminal* branches, which are distinct from *lateral* branches given off along the trunk or a branch of the tree. Both *terminal* and *lateral* are attributes of the *branch* relationship, and as mentioned above, at the time of writing, are not processable by *Emily*.

9. The greater splanchnic nerve contains nerve fibers derived from each of the following spinal nerves **EXCEPT**
   a. T5
   b. T12
   c. T9
   d. T7
   e. T8

Item 9 illustrates the need for composing a compound query and transitive closure because inference is required to trace the complex and remote branching relationships through which nerve fibers are transmitted from a segment of the spinal cord through a set of spinal nerves and their branches to the Greater splanchnic nerve. These relationships are shown graphically in Figure 5[24] to illustrate the challenge for composing the queries and to demonstrate *Emily*'s capabilities for tracing complex relationships.

The trunk of the greater splanchnic nerve is formed by the union of its roots, which are branches of a set of sympathetic thoracic ganglia; each of these ganglia is connected by a sequence of branches to the trunk of a spinal nerve in a particular set. The stem of the item asks which of the options is *not* a member of this set. The intent of the item's author is to elicit from the examination taker (usually a student) a reasoning process that traces a nerve fiber through the structures that transmit such a fiber from a segment of the spinal cord to the trunk of the greater splanchnic nerve, as shown in Figure 5. Note, however, that to follow the fiber's path, the student need not necessarily know the names of the structures that transmit the fiber; however, the student will not be able to arrive at the correct answer without understanding the structural (or spatial) connections shown in Figure 5.

Emily can emulate the behavior of the student who understands these connections, provided the authors recognize in formulating the query that there must be transitive continuity between the greater splanchnic nerve and the spinal cord segments that contribute nerve fibers to this nerve. Therefore, the authors formulate the query

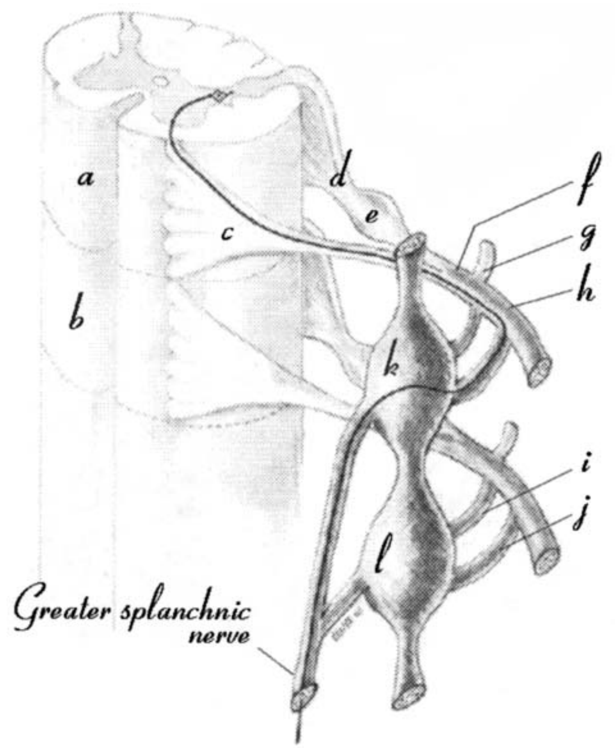

Figure 5. A schematic, graphic illustration of the continuities between neural structures through which *Emily* traces complex relationships in the FMA to answer examination question 9. Among others not illustrated, *Emily* navigates through the following structures and their relations in the FMA: *a*, T5 segment; *b*, T6 segment of the spinal cord; *c* and *d*, anterior and posterior roots of the fifth thoracic nerve; *e*, spinal ganglion of fifth thoracic nerve; *f*, trunk of fifth thoracic nerve; *g*, posterior ramus of fifth thoracic nerve; *h*, fifth intercostal nerve; *i*, gray communicating ramus; *j*, white communicating ramus; *k* and *l*, fifth and sixth thoracic ganglia. Contiguous structures corresponding to *c–i* are all parts of one spinal nerve, such as the fifth or sixth thoracic nerve. A black line represents a single nerve fiber that extends from T5 segment of the spinal cord to the greater splanchnic nerve. Not drawn to scale. Based on Rosse and Gaddum-Rosse.[24]

Greater splanchnic nerve *is continuous with* Unknown

Emily returns the result set {Fifth thoracic ganglion, Sixth thoracic ganglion, Seventh thoracic ganglion, Eighth thoracic ganglion, Ninth thoracic ganglion, Tenth thoracic ganglion, Fifth thoracic nerve, Sixth thoracic nerve, Seventh thoracic nerve, Eighth thoracic nerve, Ninth thoracic nerve, Tenth thoracic nerve, T5 segment, T6 segment, T7 segment, T8 segment, T9 segment, T10 segment}.

Interpretation of the result set requires us to recognize that in the Anatomical Taxonomy of the FMA Fifth thoracic nerve is a Thoracic nerve, which is a Spinal nerve, and that Fifth thoracic nerve corresponds to the shorthand expression T5 in option a of the item. Similar translations apply to the other options. Options a, c, d, and e are found in this set (as Fifth thoracic nerve, Ninth thoracic nerve, Seventh thoracic nerve, and Eighth thoracic nerve, respectively), but choice b is missing, which provides the correct answer.

The same conclusion is reached if the authors submit a series of Boolean queries for the continuities of each of the options. One such query would be

Greater splanchnic nerve *is continuous with* Fifth thoracic nerve

and so on. *Emily* returns *yes* for the fifth, ninth, seventh, and eighth thoracic nerves (i.e., T5, T9, T7, and T8). However, if the authors pose the query

Greater splanchnic nerve *is continuous with* Twelfth thoracic nerve,

the answer is *no*, which identifies the exception called for by Item 9.

Note that the answers to any of the above queries are not represented explicitly in the FMA. *Emily* deduced the query results by tracing the relations represented for each structure shown in Figure 5. Note also that *Emily* omits from the result sets a number of structures included in Figure 5 (e.g., roots and trunk of spinal nerve, intercostal nerve, communicating ramus). The explanation is that these structures are represented in the FMA as parts of a spinal nerve tree. As in the case of the coronary artery in Item 8, the FMA distinguishes between the trunk of the spinal nerve and the entire tree, which includes the roots, the trunk, and all branches (e.g., intercostal nerve, communicating ramus) of the tree. *Emily* takes advantage of such knowledge embedded in the FMA and returns only the names of the neural trees since this is the level at which the most general correct answer is first encountered.