

Of Mice and Men: Design of a Comparative Anatomy Information System

Ravensara S. Travillian MS, MA^{1,2}; John H. Gennari PhD^{1,2}; Linda G. Shapiro PhD^{1,2,3}

¹Structural Informatics Group, ²Dept. of Medical Education & Biomedical Informatics,

³Dept. of Computer Science & Engineering; University of Washington, Seattle, WA 98195

In previous work, we proposed an approach called the Structural Difference Method (SDM) to correlating the anatomy of Homo sapiens with selected species¹, using the Foundational Model of Anatomy (FMA)^{2,3} as a framework and graph matching as a method, for determining similarities and differences between species. In this paper, we present the design of a comparative anatomy information system that utilizes the SDM and allows users to issue queries to determine the similarities and differences between two species. Our system will serve as a pilot project for cross-species anatomical information collection, storage, and retrieval. The underlying data structure of a mapping, and the syntax and semantics of the system's query language, are presented.

INTRODUCTION

The goal of this work is to build a comparative anatomy information system to which users can pose queries about the similarities and differences between two species. In our prior work¹ we have developed an approach called the Structural Difference Method (SDM), in which each species is represented by an attributed graph, and graph matching is used to determine the similarities and differences. Using the Foundational Model of Anatomy (FMA) for humans^{2,3} as our framework, we have developed a partial mouse anatomy ontology (MAO) that can be used for comparisons. To compare two species, a mapping between them must be constructed and represented as a computer data structure. Since both the FMA and the MAO are implemented in the Protégé 2000 frame-based knowledge system, we have designed the mappings between mouse and human as Protégé classes that can link the two ontologies and provide a resource for a query system.

The information system proposed in this paper will accept queries posed by the user about similarities and differences in human and mouse anatomy. The implementation of the pilot version of the comparative anatomy system will be a single database of mappings, from which the query engine will access and return a result set.

The proposed graphical user interface for the application is shown in Figure 1. The system is currently under development, so this proposed interface is a mockup, but serves to illustrate the kinds of queries supported by the system. Because the user interface

is designed to conform to our syntax (described below), the user is not required to remember the syntax of the different queries—rather, she can form a query by selecting choices which serve to specify the form of the query. The Mapping direction, From structure, To structure, and Query fields are options selected by the user to specify the information requested from the database, and the Mapping results box is where the application displays the results of the processed query. For example, if the user wanted to ask “How does the human lung differ from the mouse lung?”, she would select Mapping direction: Human \mapsto Mouse, From structure: Lung, To structure: Lung, Query: differs-from, and then she would click the Submit button. The results would appear in the Mapping results box, and would include such information as the human left lung has 2 lobes while the mouse left lung has only 1, the human right lung has 3 lobes while the mouse right lung has 4, and so forth.

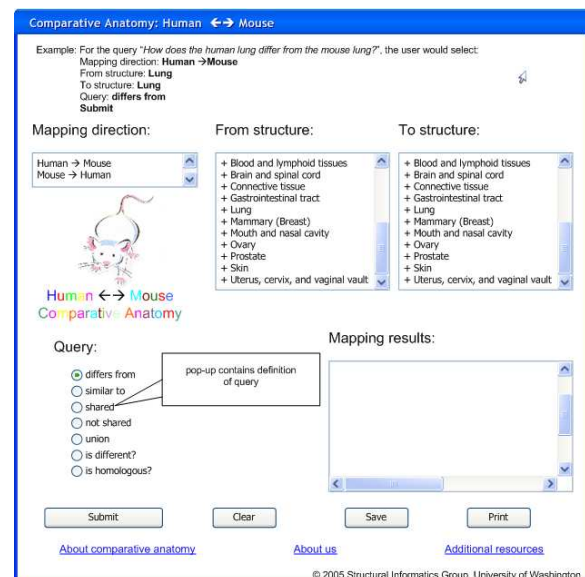


Figure 1: Proposed graphical user interface for comparative anatomy application.

The Mapping direction list box specifies which species is the source and which is the target of the mapping, thus permitting unidirectional as well as bidirectional queries. The + sign by the structure names indicate that a hierarchy can be expanded, and it is by selecting a class in the hierarchy that the granularity

of the comparison is specified. We expand the hierarchy initially along the *part-of* links, rather than *is-a*, as our experience shows that use of the parthood is the most intuitive way for bio-researchers to think about anatomy. Additionally, use of the parthood is consistent with the JAX mouse anatomy ontology⁴.

The Query radio buttons allow the user to select a query relationship, which will use the specified structures as arguments. Once a query relationship has been specified, the user can proceed by clicking the Submit button, or she can click the Clear button to reset the selections and start over again.

When a query has been submitted, the mapping result set—*i.e.*, a set of descriptions of similarities and differences for that structure across those species, as calculated by the SDM—is returned by the application in the Mapping results box. The results of the query can be saved by clicking the Save button, or printed by clicking the Print button.

The anatomical mapping data structure and the syntax and semantics of the system’s query language are particularly significant, and will be discussed in more detail below.

ANATOMICAL MAPPING

Mappings are the data structure at the heart of the proposed information system. In order to be able to create a mapping of anatomical structures across species, the structures must be formally represented in a way that supports the mapping. In earlier work, it has been demonstrated that the Foundational Model of Anatomy (FMA) describes multiple directed acyclic graphs (DAG)⁵, and so, in order to create the mappings, we develop the appropriate graphs, one for the human structure, and one for the mouse structure. In these graphs, the nodes represent anatomical entities, and the edges represent the structures among those entities. This representation is consistent with the one used by Hayamizu’s Adult Mouse Anatomical Dictionary⁶, as well, so we expect to be able to extend this system to other ontologies.

A Mapping, then, is a correspondence between a structure in the source species and a structure in the target species. As developed in Travillian 2003, there are two main kinds of mappings: Node mappings and Edge mappings, corresponding to the components of the directed graph described by the FMA. The structures which are mapped across species are selected on the basis of homology (evolutionary relatedness); homoplasy (similarity of appearance) and analogy (similarity of function) are not considered in creating mappings.

At a conceptual level, a Mapping across Species between Anatomical structures can be represented as in Figure 2, which shows Mappings

between the human and mouse Prostates at the Organ level. Note that the graph is a composite of the *is-a* and *has-member* graphs; these relationships were selected to emphasize differences, as the mouse and human prostates differ in some important and non-intuitive ways which are demonstrated efficiently in the *is-a* hierarchy. The fact that the mouse has five prostates, where the human has one, and that homologies between the mouse organs and the parts of the human organ have not been definitively established, represents important considerations in comparative medicine. Since different pathologies or resistance to pathology (prostatic carcinoma, benign prostatic hyperplasia, or no disease) arise in different parts of the human prostate, and since those parts and their pathologies correlate to the embryonic origin of the structures, the importance of establishing those homologies to draw the Mappings is clear.

The edges of the graph in green represent isomorphisms, or anatomical identity: one-to-one, onto, and structure-preserving. For example, the anatomical abstraction Lobular organ in the mouse is isomorphic to the Lobular organ in the human. The edges of the graph in blue represent non-isomorphic matches. For example, there is a 5:1 mapping between the five different mouse prostate organs and the single human prostate. The edges in red represent null mappings. For example, there is no corresponding Set of human prostates to map to the Set of mouse prostates, so that constitutes a null mapping. A single mapping can answer a query such as “What is the structure in the human that corresponds to the liver in the mouse?”.

A unidirectional comparison consists of a hierarchy of Mappings. A root for the mapping, and the depth to which the comparison is to be pursued, are chosen, and all the mappings for structures in the hierarchy beginning at the root and proceeding to the chosen depth, make up the unidirectional comparison. A unidirectional comparison can thus answer a query such as “What are the structures in the mouse mammary gland that are missing in the human mammary gland?”. A cross-species, or bidirectional, comparison consists of two complementary unidirectional comparisons.

To implement this functionality, the underlying Mapping data structure contains pointers in both directions between species: *i.e.*, the human can be either the source or the target species, as can the mouse. Both directions are necessary for a complete answer to queries on similarities and differences between species, as, from the user’s point of view, the answer returned to the query “what is the difference between the human and mouse prostates?” should be the same as the answer returned to the query “what is the difference between the mouse and human prostates?”. This

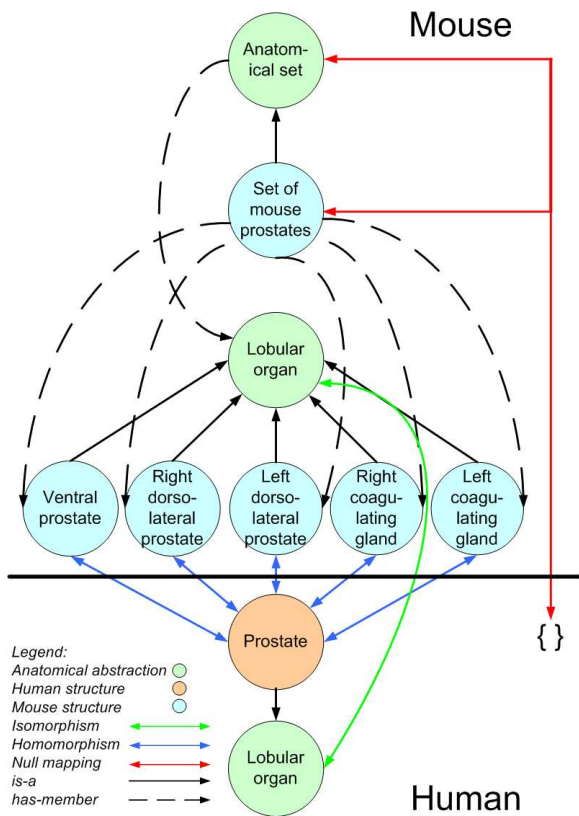


Figure 2: Conceptual mapping between the human and mouse prostates.

data structure provides that consistency of response, yet at the same time allows a more refined query to return a more granular answer, depending on the level of detail the user wishes to specify. Although the usual query will be bidirectional, there will be users who want information in one direction only. For example, a user may want to know what prostatic lobe in the human is homologous to the murine dorsal prostate. This structure will be able to accommodate those queries as well.

The FMA is implemented in Protégé 2000⁷. Following the suggestion of Bernstein and Pottinger⁸ that mappings should be first-class objects, we have implemented mappings as Protégé classes. Mappings are implemented in Protégé in the following manner: the Protégé template slots for Mapping are the two Species being compared (as in Figure 3), and the two corresponding Anatomical structures. Species names are required to always be single; Anatomical structures can be 1 or more in a particular Species. Cardinality specifies whether the correspondence is 1:null, null:1, 1:1, 1:many, many:1, many:many, many:null, or null:many.

The slots for Node mapping are inherited verbatim from the Mapping class; Edge mappings

Node mapping

Mapping Slots	
	Name
S	Species 1
S	Anatomical entity 1
S	Species 2
S	Anatomical entity 2
S	Cardinality

Edge mapping

Mapping Slots	
	Name
S	Relationship
S	Species 1
S	Anatomical entity 1
S	Species 2
S	Anatomical entity 2
S	Cardinality

Figure 3: Mapping templates. The Edge mapping template has the same slots as the Node mapping template, plus the additional Relationship slot.

have the additional slot Relationship to describe which relationship is being compared across Species for the given Anatomical structures. These examples demonstrate the definitions for the different kinds of Mappings. A Cross-species comparison is made up of all the Mappings of the Anatomical structure at the level under comparison. We propose to use these structures to return answers to anatomical queries about similarities and differences between these structures—the Mapping contains the information about similarity and differences of particular discrete structures, and the Cross-species comparison provides the context (hierarchy) for those structures in relation to other anatomical structures.

QUERIES

For the purpose of defining this comparative anatomy information system, it is useful to draw a distinction between different kinds of queries, based on how many models the system handles at a time. These classifications will specify what types of queries our system handles, and what is outside its scope. We define the classification of a query as follows:

Single-species queries hold for species models taken one at a time. For example, in the human, the Heart is *inside* the Thoracic cavity, so the query “what is the relationship between Heart and Thoracic cavity [*implied: in the human*]?” is a single-species query. Note that a single-species query can be simple or compound; the classification of the query refers to the number of species models participating in the query, NOT to the complexity of the query. Single-species queries currently can be the basis of queries in the FMA using the *Emily* graphical user interface⁹, and involve existence, location, connectivity, and similar features of anatomical structures.

Two-species queries hold for species models taken two at a time, and are the basis of what is unique

about our proposed system. They involve comparisons between anatomical structures across two different species, such as “how is the human prostate different from the mouse prostate?”. Two-species queries involve similarity, difference, homology, identity, and synonymy of anatomical structures in two different species, as described below. While the concepts of homology, identity, and synonymy overlap to some degree in natural language, the syntax below suffices to deal with them at the level of the users’ needs. Higher-degree queries represent future work, and will explicitly not be treated in this specification. We propose to develop the syntax for two-species queries, as follows.

Syntax

The following syntax represents a textual abstraction of our allowable cross-species queries.

```
<query> ::= <concept> <SDM relationship>
           <concept>
<concept> ::= <species> <anatomical
                entity> | <unknown> |
              <result set>
<SDM relationship> ::= differs-from |
                       similar-to | union | shared |
                       not-shared | is-different? |
                       is-homologous?
```

We propose to use this syntax as the basis for queries and responses about anatomical similarities and differences between the human and the mouse. This notation represents an abstraction of the basis for the queries and responses; there will be a low-level syntax that is used by the system for accessing and returning information, as well as a higher-level graphical user interface for the users of the system.

Semantics

While the details remain to be determined, some of the semantics of the query language are already emerging from the information gathered to date. Queries will be of two major types, set and Boolean. Set queries will return result sets, such as the set of shared mappings between two species for a structure at a given level of granularity. Boolean-type queries will, for example, return T or F when the user queries whether structures in two different species map to each other. The semantics of the proposed operators are as follows.

Set queries: The set query operators are *differs-from*, *similar-to*, *shared*, *not-shared*, and *union*.

- *species1.anatomical-entity1 differs-from species2.anatomical-entity2* returns the difference between anatomical-entity1 in species1 and anatomical-entity2 in species2. If anatomical-entity1 and anatomical-entity2 are isomorphic, it will return null.

- *species1.anatomical-entity1 similar-to species2.anatomical-entity2* returns the complement of the set returned by (*species1.anatomical-entity1 differs-from species2.anatomical-entity2*), which is all of the similarities between *species1.anatomical-entity1* and *species2.anatomical-entity2*.
- *species1 shared species2* returns the set of non-null mappings between anatomical entities of *species1* and those of *species2*.
- *species1 not-shared species2* returns the set of null mappings between anatomical entities of *species1* and those of *species2*. In other words, it is the inverse operation of *shared*.
- *species1 union species2* returns the set of all (null as well as non-null) mappings between anatomical entities of *species1* and those of *species2*.

Boolean queries: The Boolean query operators are *is-different?* and *is-homologous?*.

- *species1.anatomical-entity1 is-different? species2.anatomical-entity2* returns T if *species1.anatomical-entity1* does not map to *species2.anatomical-entity2*, and F if the two anatomical entities do map to each other.
- *species1.anatomical-entity1 is-homologous? species2.anatomical-entity2* returns F if *species1.anatomical-entity1* does not map to *species2.anatomical-entity2*, and T if the two anatomical entities do map to each other. In other words, it is the inverse operation of *is-different?*.

These Boolean and set query operators suffice to deal with the questions of similarity and difference that a user would ask the system about the comparisons between mouse and human anatomy, and this aim serves to provide the structure (syntactic and semantic) for those operators.

CONCLUSION

Many contemporary observers^{10,11} have remarked upon the increasing need for extrapolating information from one species to another, which has been highlighted by contemporary research in bioinformatics, genomics, proteomics, and animal models of human disease, as well as other fields. The amount of anatomical and associated medical information emerging from animal modeling in comparative medicine and comparative genomics is increasing at an exponential rate, calling for innovative techniques in evaluating, organizing, and managing that information for researchers and clinicians. In addition, the increasingly interdisciplinary nature of medical research has

greatly increased the base of users, and the corresponding need, for such an information system. Therefore, in addition to rigorous attention to the quality of the anatomical information involved, such a system must be flexible and extensible enough to accommodate different information views, depending on the needs of the user—whether a bench scientist, a clinician, or a student.

In order to successfully manage this information, a systematic, principled way of correlating anatomical information across species is needed. Because the FMA has the necessary qualities to serve as the basis for a sound and complete pan-vertebrate metamodel, we base our information system on FMA models of the human and of the mouse.

In this paper, we propose a pilot comparative anatomy information system to meet this need. Our proposed information system builds on our previous work in correlating the anatomy of *Homo sapiens* with selected species, using the Foundational Model of Anatomy (FMA) as a framework, and graph matching as a method. It will be able to answer queries regarding cross-species similarities and differences in structural phenotypes, and it addresses important scientific questions in both medical informatics and comparative anatomy.

In informatics, the problem of ontology alignment has been a promising research area for decades, yet the inherent complexity of comparing such different anatomical data at so many levels of resolution for so many species poses a challenge far greater than the domain of most ontology alignments, and carries the promise of developing techniques and tools that can be applied to genomics ontology alignment problems, taken as another level of anatomical complexity. As well, in comparative anatomy, the structure and organization of massive amounts of anatomical data in one resource will serve multiple purposes of making information accessible and visualizable in different views for different users with different information needs, as well as for identifying gaps and inconsistencies in the scientific literature to facilitate future research. We hypothesize that our system will prove to be an initial step toward meeting these needs.

ACKNOWLEDGEMENTS

This work was supported by the University of Washington National Library of Medicine Informatics Training Grant (1T15LM07441-01).

REFERENCES

- 1 Travillian RS, Rosse C, Shapiro LG. An approach to the anatomical correlation of species through the Foundational Model of Anatomy. *AMIA Annu Symp Proc.* 2003:669-73.
- 2 Rosse C, Mejino JL Jr. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform.* 2003 Dec;36(6):478-500.
- 3 Rosse C, Shapiro LG, and Brinkley JF. The Digital Anatomist Foundational Model: Principles for defining and structuring its concept domain. *Proceedings 1998 American Medical Informatics Association Annual Symposium*, November 1998.
- 4 The Jackson Laboratory (JAX) Mouse Genome Informatics. Available at <http://www.informatics.jax.org/>. Accessed December 14, 2004.
- 5 Mejino JL Jr, Rosse C. Conceptualization of anatomical spatial entities in the Digital Anatomist Foundational Model. *Proc AMIA Symp.* 1999:112-6.
- 6 Hayamizu TF, Mangan M, Corradi JP, Kadin JA, Ringwald M. The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data. *Genome Biology* 2005;6:R29.
- 7 Noy NF, Crubezy M, Fergerson RW, Knublauch H, Tu SW, Vendetti J, Musen MA. Protege-2000: an open-source ontology-development and knowledge-acquisition environment. *AMIA Annu Symp Proc.* 2003:953.
- 8 Bernstein, PA, Levy AY, Pottinger RA. A Vision for Management of Complex Models. Microsoft Research Technical Report *MSR-TR-2000-53*, June 2000.
- 9 Shapiro LG, Chung E, Detwiler LT, Mejino JL Jr, Agoncillo AV, Brinkley JF, Rosse C. Processes and problems in the formative evaluation of an interface to the foundational model of anatomy knowledge base. *J Am Med Inform Assoc.* 2005 Jan-Feb;12(1):35-46. Epub 2004 Oct 18.
- 10 Linazasoro G. Recent failures of new potential symptomatic treatments for Parkinson's disease: causes and solutions. *Mov Disord.* 2004 Jul;19(7):743-54.
- 11 Noble M, Dietrich J. Noble M, Dietrich J. The complex identity of brain tumors: emerging concerns regarding origin, diversity and plasticity. *Trends Neurosci.* 2004 Mar;27(3):148-54.