Evaluation of Ontology Merging Tools in Bioinformatics

P. Lambrix, A. Edberg

Pacific Symposium on Biocomputing 8:589-600(2003)

EVALUATION OF ONTOLOGY MERGING TOOLS IN BIOINFORMATICS

P. LAMBRIX, A. EDBERG

Department of Computer and Information Science Linköpings universitet, 581 83 Linköping, Sweden

Ontologies are being used nowadays in many areas, including bioinformatics. One of the issues in ontology research is the aligning and merging of ontologies. Tools have been developed for ontology merging, but they have not been evaluated for their use in bioinformatics. In this paper we evaluate two of the most well-known ontology merging tools with a bioinformatics perspective. As test ontologies we have used Gene Ontology and Signal-Ontology.

1 Introduction

Ontologies define the basic terms and relations comprising the vocabulary of a topic area, as well as the rules for combining terms and relations to define extensions to the vocabulary. They are being used nowadays in many areas, including bioinformatics, for several reasons. They are used for communication between people and organizations by providing a common terminology over a domain. They provide the basis for interoperability between systems. They can be used for content explicitation for information sources and serve as an index to a repository of information. They can also be used as a basis for integration of information sources and as a query model for information sources. To assist users in developing and maintaining ontologies a number of tools have been developed and comparative studies of ontological engineering tools have been performed. In [DSWKB99] Ontolingua, WebOnto, ProtégéWin, OntoSaurus, ODE and KADS22 were compared. The authors used ontologies concerning academia and university studies for testing. In [HP02] we used Gene Ontology to compare the basic functionality of the ontology tools Protégé-2000, Chimaera, DAGEdit and OilEd.

Within the bioinformatics area there are a number of bio-ontologies, each with their own focus, that cover different aspects in molecular biology such as molecular function and cell signaling. Many of these ontologies contain overlapping information. For instance, a protein can be involved in both cell signaling and other biological processes. In applications using ontologies it is therefore of interest to be able to use multiple ontologies. However, to obtain the best results, we need to know how the ontologies overlap and align them or merge them into a new ontology. Another reason for merging ontologies is that it allows for the creation of ontologies that can later be composed into larger ontologies. Also, companies may want to use de facto standard ontologies and merge them with company-specific ontologies. There exist a number of tools for merging ontologies that have been used outside the bioinformatics area. In this paper we evaluate how well these tools work for merging bio-ontologies. We have chosen to evaluate Protégé-2000 with PROMPT and Chimaera, which are, currently, two of the main ontology merging tools. We describe these tools in section 3. We looked at a number of bio-ontologies and decided to use two ontologies for testing: Gene Ontology, which has become a de facto standard and Signal-Ontology. The overlap between these ontologies gives interesting test cases. In section 2 we briefly describe these ontologies. In section 4 we discuss the evaluation criteria, the evaluation method and the set-up of the evaluation. Section 5 discusses, because of space limitations, a high-level view of the results of the evaluation. More details can be found in [Edb02].

2 Ontologies

2.1 Gene Ontology

The Gene Ontology (GO) Consortium is a joint project with as goal to produce a structured, precisely defined, common and dynamic controlled vocabulary that describes the roles of genes and proteins in all organisms [GO]. Currently, there are three independent ontologies publicly available over the Internet: biological process, molecular function and cellular component. The biological process ontology deals with biological objectives to which the gene or gene product contribute. A process is accomplished via one or more ordered assemblies of molecular functions. The molecular function ontology deals with the biochemical activities of a gene product. It only describes what is done without specifying where or when the event takes place. The cellular component ontology describes the places where a gene product can be active. The GO ontologies are becoming a de facto standard and many different bio-databases are today annotated with GO terms. The ontologies grow continuously. The terms in GO are arranged as nodes in a directed acyclic graph, where multiple inheritance is allowed.

2.2 Signal-Ontology

The purpose of the Signal-Ontology (SO) [SO] project is to extract common features of cell signaling in the model organisms, try to understand what cell signaling is and how cell signaling systems can be modeled. SO is a publicly available controlled vocabulary of the cell signaling system. It is based on the knowledge of the Cell Signaling Networks databank [TNK98] and treats

complex knowledge of living cells such as pathways, networks and causal relationships among molecules [TT00]. SO consists of four different conceptual classes: signal motif, reaction, signal family and cellular function. In the signal motif class the units of the signaling network that are conserved through species are defined. In the reaction class the biochemical properties of individual molecular reactions in the signaling pathways are defined. The signal family class provides a functional classification of signal molecules and the cellular function class classifies biological functions controlled by the cell signaling system. SO is composed of signal modules that are minimal units of signal processing common to the model species. Each module models a concept extracted from biological information and constitutes a unit in the signal transduction. The ontology consists of a flow diagram of signal transduction and a conceptual hierarchy of biochemical attributes of signaling molecules.

3 Tools for merging ontologies

3.1 Protégé-2000 with PROMPT

Protégé-2000 is software for creating, editing and browsing ontologies developed by Stanford Medical Informatics. The design and development of Protégé-2000 has been driven by two goals: to be compatible with other systems for knowledge representation and to be an easy to use and configurable tool for knowledge extraction. Protégé-2000 is available as free software and should be installed locally. It also has a number of plug-ins, among others PROMPT, which is an algorithm for merging and aligning ontologies [PROMPT, NM00].

When merging two ontologies PROMPT creates a list of suggested operations. An operation can, for instance, be to merge two terms or to copy a term to the new ontology. The user can then perform an operation by choosing one of the suggestions or by specifying an operation directly. PROMPT performs the chosen operation and additional changes that follow from that operation. The list of suggestions is then updated and a list of conflicts and possible solutions to these conflicts is created. This is repeated until the new ontology is ready. PROMPT was previously called SMART and a high-level description of the algorithm is given in [NM99].

3.2 Chimaera

Chimaera is developed by the Knowledge Systems Laboratory at Stanford University to provide assistance to users for browsing, editing, merging and diagnosing of ontologies. It was built on top of the Ontolingua Distributed Collaborative Ontology Environment. The initial goal was to develop a tool that could give substantial assistance for the task of merging knowledge bases produced by different users for different purposes with different assumptions and different vocabulary. Later the goals of supporting testing and diagnosing ontologies arose as well. The user interacts with Chimaera through a browser such as Netscape or Microsoft Internet Explorer [MFRW00].

The two main tasks when merging two ontologies in Chimaera are to merge two semantically identical terms from different ontologies so that they are referred to by the same name in the resulting ontology, and to identify terms that should be related via the is-a, disjointness or instance relationships and provide support for introducing those relationships. Chimaera also supports the identification of the locations for editing and performing the edits. Today, Chimaera has support for merging taxonomies of terms and for merging attributes. The support for merging relations, functions, facets, individuals and axioms will be developed later. To assist the user Chimaera generates name resolution lists that suggest terms that are candidates to be merged or to have taxonomic relationships not yet included in the merged ontology. Chimaera also generates a taxonomy resolution list where it suggests taxonomy areas that are candidates for reorganization. On the basis of these lists the user decides what should be done.

4 Evaluation Methods

In this section we present the evaluation of the ontology merging tools Protégé-2000 with PROMPT and Chimaera. We only discuss the systems in terms of their support for merging ontologies in some detail. We have also studied the systems in a more general study of ontology tools and refer to [HP02] for detailed results concerning the other functionality of these systems.

4.1 Methods

The evaluation of the tools has been divided into two parts. In the first part we evaluated the systems based on a number of predefined criteria. The evaluation was done through a literature study and tests using GO and SO. The criteria that were evaluated are presented in section 4.3.

In the second part of the evaluation the user interface of the systems was evaluated. For this we used an empirical approach. In this technique end users test the systems. This approach is usually time-consuming and requires the use of test persons, but has the advantage that the problems that are discovered are real problems. The use of this method requires the development of good instructions for the test persons. In our evaluation eight test persons participated, four with a computer science background and four with a biology background. The biologists were used to work with computers. They had no experience of working with ontologies. As evaluation method we used the REAL method [Löw93]. This method was chosen for its ease of use, because it requires few test persons and it usually gives good results. Before the actual evaluation the test persons were given information on ontologies, what they are, how they can be represented, what they are used for etc. Then the test persons familiarized themselves with the tools using a tutorial and by doing another evaluation [HP02]. This took about 1.5 to 2.5 hours per person. The actual evaluation for each system was divided into two parts. During the first part the test persons were given a number of tasks to perform. They had a manual on paper and the system help (when available) for support. They were asked to think loud and an evaluator took notes during the process. Afterwards, the test persons were asked to fill in a questionnaire. The questionnaire consisted of questions with coded values. The test persons were also asked to write comments on their grading. In total this took about 1.5 to 2.5 hours per person.

4.2 Description of test ontologies

In our tests two chosen 'cases' from GO and SO were used. Each case consists of one part of SO and one part of GO. Each case was chosen in such a way that there was an overlap between the GO part and the SO part. The first case, *behavior*, contains approximately 60 terms from GO and approximately 10 terms from SO. The second case, *immune defense*, contains approximately 70 terms from GO and approximately 15 terms from SO. A complete listing of the terms that are included can be found in [Edb02]. We used more terms from GO than from SO because the granularity of GO is higher than the granularity of SO for these topics. The terms in SO and GO include concepts, attributes and instances. The concepts are arranged in a hierarchy. Concepts may have attributes. The attributes are not hierarchically organized. Each concept may also have instances.

4.3 Evaluation criteria

Our criteria for the first part of the evaluation are partly based on previous work on evaluating ontology tools [DSWKB99] and ontology exchange languages for bioinformatics [McE00]. We investigated the availability of the tool. We tested the systems during a period of time and checked their stability. Further, we looked at the representation language that is used by the tools. For the merging we evaluated the tools with respect to the functionality they provide. This deals with issues such as what kind of ontologies can be merged and whether consistency checking is performed. We also looked at the assistance that is given to the user while merging. Both tools generate suggestions for merging. We looked at how the suggestions are generated and evaluated their quality measuring precision and recall using our test ontologies. The precision measures how many of the suggestions are relevant, while recall measures how many of the total number of the relevant suggestions the systems actually proposed. Further, we also measured the time it took to merge the test ontologies using the tools.

The user interface was evaluated using the REAL approach. The aspects studied in this evaluation are Relevance, Efficiency, Attitude and Learnability. Relevance measures how well a user's needs are satisfied by the tool. Efficiency measures how fast users can perform their tasks using the tool. The subjective feelings towards the tools are measured by attitude. Finally, learnability measures how easy or difficult it is to learn the tool for initial use as well as how much a user can remember the workings of the tool.

5 Evaluation Results and Discussion

5.1 General

Protégé-2000 with PROMPT is available as free Java software and is installed locally. Chimaera is available over the Internet and requires a relatively fast connection to be able to work efficiently. With high use of the network even simple operations can take a long time to perform.

Protégé-2000 with PROMPT is not entirely stable. During about 35 hours (excluding user interface evaluation time) of work the Protégé-2000 program had to be terminated in an abnormal way five times. Together with PROMPT this happened about half of the times we worked with merging. During about 10 hours (excluding user interface evaluation time) of work with Chimaera, the system has been stable. On one occasion the server at Stanford was not accessible.

Both Protégé-2000 with PROMPT and Chimaera allow import and export of ontologies in different formats. Protégé-2000 also allows user-programmed plug-ins to import and export to other formats. The internal representations in both systems are based on the Open Knowledge Base Connectivity (OKBC), which is an application programming interface for frame-based knowledge representation systems [CFFKR98]. The internal language of Protégé-2000 is less expressive than the language of Chimaera. Both are frame based and are expressive enough to represent most of the current bio-ontologies, including our test ontologies.

5.2 Merging

Protégé-2000 with PROMPT allows for merging and copying concepts, attributes and instances from two ontologies to a new ontology. PROMPT generates suggestions for merge operations on concept and attribute levels together

Table 1: Quality of suggestions

Tool	Case	Suggestions	Correct	Missing	Recall	Precision
PROMPT	В	3	3	2	0.6	1
Chimaera	В	16	4	1	0.8	0.25
PROMPT	ID	4	4	5	0.44	1
Chimaera	ID	10	4	5	0.44	0.4

with an explanation of why these suggestions were made. When the user decides to follow the suggestion, also a number of other changes (logically derived) are made automatically and new suggestions may be derived too. PROMPT also identifies possible conflicts that could occur as a result of the merging and proposes possible solutions. Concepts in the original ontologies that are not merged need to be copied into the new ontology.

The generation of suggestions is based on similarities in the concept and attribute names, but in [NM99] the authors suggest that the similarity measure is not hard-coded and that users may define their own similarity measures. For the tests, however, we used the system as it is provided in the distribution. Extensions to the basic algorithm are being made, such as Anchor-PROMPT [NM01] where paths between merged concepts in the two ontologies are investigated to find new suggestions.

Chimaera provides about seventy commands in the user interface, thereby providing a taxonomy and slot editor. The applicable commands at each point in time are made available by the interface. Some of these commands are related to ontology merging such as 'merge classes' and 'move class x to become a subclass of class y'. There are also commands related to diagnosis that, among others, check for incompleteness, cycles and value-type mismatches.

As a help to the user, Chimaera generates a list of concepts and attributes that are candidates for merging. The generation of this list is based on similarities in names, definitions, acronyms, name extensions, names that appear as suffixes in other names etc. Also a list with suggestions for areas that may need restructuring is generated based on heuristics. For instance, one heuristic looks for concepts that have direct sub-concepts in both ontologies and suggest these as restructuring areas [MFRW00]. A difference between PROMPT and Chimaera is that Chimaera suggests concepts and places *where* actions may need to be taken, while PROMPT also suggests *what* actions should be taken.

In table 1 we show the results regarding precision and recall for the cases behavior (B) and immune defense (ID), respectively. The 'suggestions' column informs about the number of suggestions that were made. The 'correct'

Table 2: Time (in minutes) for merging

Tool	Case	Merging based on suggestions	Additional time	Total
PROMPT	В			15
Chimaera	В	8	0	8
PROMPT	ID			26
Chimaera	ID	10	5	15

column shows how many of the suggestions were relevant while the 'missing' column shows how many of the relevant suggestions were missing. In total there were 5 possible cases for merging in behavior and 9 possible cases in immune defense. PROMPT got perfect precision in both test cases, while Chimaera's precision was below 50%. However, PROMPT gave much fewer suggestions than Chimaera. Chimaera's recall was higher than PROMPT's recall, but both performed relatively poorly on immune defense. Examples of concepts to merge that were missing were: 'inflammatory response' and 'inflammation', 'antigen processing and presentation' and 'antigen presentation', and 'complement activation' and 'complement signaling'. The last pair would have been found if we compiled a synonym list based on the synomyms given in GO and SO. The two other pairs might have been found with less restrict string matching.

In table 2 we show the results regarding the time it took to merge the test ontologies for the cases behavior (B) and immune defense (ID), respectively. For Chimaera we calculated the time to merge the ontologies based on the suggestions and the additional time it took to merge the concepts and attributes that were not suggested by the program. For PROMPT we only calculated the complete time as much of the work was spent on copying the concepts and attributes that did not need to be merged. It is clear that merging was much faster with Chimaera than with PROMPT. This may be a key factor when we want to merge large ontologies.

5.3 User interface

The results from the questionnaire are found in tables 3 and 4. In this section we summarize these results as well as the observations reported from the comments the test persons gave during the evaluation.

Relevance The test persons felt it was better to use PROMPT than Chimaera for the merging of ontologies. This depended, among others, on the fact that it takes relatively long time to get a response for some operations in Chimaera.

Question	Mean	Min	Max
Relevance			
To use PROMPT for merging ontologies feels (bad=1, good=10)	8.25	3	10
Efficiency			
To merge ontologies with PROMPT feels (difficult=1, easy=10)	5.50	3	8
To choose wanted operation feels (difficult=1, $easy=10$)	6.63	3	10
The overview over the ontologies you wish to merge feels	8.50	5	10
(bad=1, good=10)			
The overview over the terms, attributes and instances	7.13	4	9
you merge feels $(bad=1, good=10)$			
Attitude			
To merge ontologies with PROMPT seems (boring=1, fun=10)	5.88	1	10
To decide whether two terms, attributes or instances should		6	9
be merged or not seems (difficult=1, $easy=10$)			
The naming of the operations is (unclear=1, clear=10)	8	4	10
The naming feels $(inconsistent=1, consistent=10)$	8.50	6	10
Learnability			
To learn how to merge ontologies seems (difficult=1, easy=10)	6.25	4	9
The available help feels (insufficient=1, sufficient=10)	5.63	2	8
To learn how to load the ontologies you wish to merge feels		5	10
(difficult=1, easy=10)			
To learn how to merge terms, attributes and instances feels		4	10
(difficult=1, easy=10)			
To learn how to copy the terms, attributes and instances	7	4	10
that you do not wish to merge, feels (difficult=1, easy=10)			

Efficiency The test persons found that it was better to use PROMPT than Chimaera for specific operations and the fact that some operations with Chimaera took a long time was one of the reasons for the lower score. However, they also thought that merging ontologies required too much work with PROMPT. In particular, the need to remember and copy all the concepts, attributes and instances that were not merged was considered awkward. The overview over the ontologies in the tool was considered to be better in PROMPT than in Chimaera. In particular, the fact that the original ontologies were kept and represented in different colors in PROMPT was considered good. In Chimaera it was difficult to see where in the hierarchy a concept was situated. The choice of operations to perform was easier in PROMPT than in Chimaera. This may be because Chimaera allows for many more operations than PROMPT,

Question	Mean	Min	Max
Relevance			
To use Chimaera for merging ontologies feels (bad=1, good=10)	6.25	2	10
Efficiency			
To merge ontologies with Chimaera feels (difficult=1, easy=10)	5.25	2	9
To choose wanted operation feels (difficult=1, $easy=10$)	6	1	10
The overview over the ontologies you wish to merge feels	5.25	2	9
(bad=1, good=10)			
The overview over the terms and attributes you merge feels	5.63	2	8
(bad=1, good=10)			
Attitude			
To merge ontologies with Chimaera seems (boring=1, fun=10)	3.88	1	8
To decide whether two terms or attributes should be		5	9
merged or not seems (difficult=1, $easy=10$)			
The naming of the operations is (unclear=1, clear=10)	5.75	1	10
The naming feels (inconsistent=1, consistent=10)	7.63	6	10
Learnability			
To learn how to merge ontologies seems (difficult=1, easy=10)	6.63	3	10
The available help feels (insufficient=1, sufficient=10) $($		5	10
To learn how to load the ontologies you wish to merge feels	6.63	2	10
(difficult=1, easy=10)			
To learn how to merge terms, attributes and instances feels	7	2	10
(difficult=1, easy=10)			

Table 4: Evaluation of graphical user interface for Chimaera

which may make it harder to find and choose the right operation.

Attitude The test persons found it was more fun to use PROMPT than Chimaera. Chimaera's graphical user interface was considered to be boring and unclear. Also, it was felt that the names of the operations were more selfexplaining in PROMPT than in Chimaera.

Learnability To learn how to merge ontologies felt equally hard for both systems. The available help for Chimaera was considered better than the help for PROMPT. The hardest to learn was how to copy the concepts, attributes and instances that are not merged in PROMPT.

5.4 Method critique

Although we have some experience with ontologies, we are not the common end users of ontology merging tools. This may have an influence on the results of the first part of the evaluation. However, we have tried to minimize this influence by studying other evaluations, learning to work with the tools and clearly defining the criteria before the actual evaluation.

The test persons in the user interface evaluation were novices in using ontologies and ontology merging. Therefore, there may have been an influence on the way the test persons performed the given tasks. In particular, there may have been a larger influence for the results of Chimaera than for PROMPT as the choice of operations in Chimaera is significantly larger.

We were also concerned about the fact that doing a task in one tool may make it easier to perform a similar task in the next tool. Therefore, to minimize the effects on the results of having done a similar task before we varied the order in which the different test persons evaluated the two tools.

As we wanted to investigate whether a test person's knowledge of biology would influence the results in the user interface evaluation, we took test persons that work daily with biology and persons that have only high school knowledge of biology. However, there was no significant difference between the results of these two groups.

6 Conclusion

In this paper we have evaluated Protégé-2000 with PROMPT and Chimaera for their use as ontology merging tools in bioinformatics. We used two cases based on Gene Ontology and Signal-Ontology as test cases. Both systems are expressive enough to model most of the current bio-ontologies. The user interface of Protégé-2000 with PROMPT was considered better than Chimaera. It gave a better overview over the ontologies and it was easier to work with. Chimaera, however, provides more functionality and better help. It is also much faster to merge ontologies with Chimaera. This may be an important factor when merging large ontologies. The quality of the suggestions could be improved for both systems and work is ongoing by the developers. We actually implemented our own algorithm for merging using synonym lists and less restricted string matching and got better results for precision and recall [Edb02]. Another issue that may be taken into consideration are the results of [HP02] where it is shown that Protégé-2000 with PROMPT has characteristics that make it appropriate to use in the early phases of ontology creation and Chimaera is a good choice for later phases such as analysis and maintenance. As the tools allow for some common import languages, using both tools may actually be an option.

References

[CFFKR98] Chaudhri, V., Farquhar, A., Fikes, R., Karp, P., Rice, J., 'OKBC: A Programmatic Foundation for Knowledge Base Interoperability', *Proceedings of* the Fifth National Conference on Artificial Intelligence, pp 600-607, Madison, WI, USA, 1998.

[Chimaera] http://www.ksl.stanford.edu/software/chimaera/

- [DSWKB99] Duineveld, A.J., Stoter, R., Weiden, M.R., Kenepa, B., Benjamins, V.R., 'Wondertools? A comparative study of ontological engineering tools', Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management, pp 4.6.1-4.6.20, Banff, Canada, 1999.
- [Edb02] Edberg, A., Ontology merging in bioinformatics, M.Sc. thesis LiTH-IDA-Ex-02/62, Department of computer and information science, Linköpings universitet, 2002. In Swedish.
- [GO] The Gene Ontology Consortium, 'Gene Ontology: tool for the unification of biology', Nature Genetics, 25:25-29, 2000. http://www.geneontology.org/
- [HP02] Habbouche, M., Pérez, M., Evaluation of ontology tools for bioinformatics, B.Sc. thesis LiTH-IDA-Ex-Ing-02/10, Department of computer and information science, Linköpings universitet, 2002. In Swedish.
- [Löw93] Löwgren, J., Human-Computer Interaction. What every system developer should know. Studentlitteratur, Lund, 1993.
- [NM99] Noy, N.F., Musen, M., 'SMART: Automated Support for Ontology Merging and Alignment', Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management, Banff, Canada, 1999.
- [NM00] Noy, N.F., Musen, M., 'PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment', Proceedings of Seventeenth National Conference on Artificial Intelligence, pp 450-455, Austin, TX, USA, 2000.
- [NM01] Noy, N.F., Musen, M., 'Anchor-PROMPT: Using Non-Local Context for Semantic Matching', Proceedings of the IJCAI01 Workshop on Ontologies and Information Sharing, 2001.
- [McE00] McEntire, R., Karp, P., Abernethy, N., Benton, D., Helt, G., DeJongh, M., Kent, R., Kosky, A., Lewis, S., Hodnett, D., Neumann, E., Olken, F., Pathak, D., Tarczy-Hornoch, P., Toldo, L., Topaloglou, T., 'An Evaluation of Ontology Exchange Languages for Bioinformatics', Proceedings of the Eight International Conference on Intelligent Systems for Molecular Biology, pp 239-250, San Diego, California, 2000.
- [MFRW00] McGuinness D., Fikes R., Rice J., Wilder S., 'An Environment for Merging and Testing Large Ontologies', Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning, pp 483-493, Breckenridge, CO, USA, 2000.
- [PROMPT] Protégé (http://protege.stanford.edu/index.html) with PROMPT (http://protege.stanford.edu/plugins/prompt/prompt.html)
- [SO] Signal-Ontology, http://ontology.ims.u-tokyo.ac.jp/signalontology/
- [TNK98] Takai-Igarashi, T., Nadaoka, Y., Kaminuma, T., 'A Database for Cell Signaling Networks', Journal of Computational Biology 5(4):747-754, 1998. CSNDB: http://geo.nihs.go.jp/csndb/
- [TT00] Takai-Igarashi, T., Takagi, T., 'Cell Signaling Ontology', Third Annual Bio-Ontologies Meeting, La Jolla, CA, USA, 2000.