Building Evolving Ontology Maps for Data Mining and Knowledge Discovery in Biomedical Informatics

Paulo Gottgtroy[1], Nik Kasabov[2], Stephen MacDonell[3]
[1] Knowledge Engineering and Discovery Institute, Auckland University of Technology
AUT Tech Park 581-585 Great South Rd, Auckland, New Zealand
{paulo.gottgtroy[1], nkasabov[2], smacdone[3]@aut.ac.nz}

Abstract¾

The explosion of biomedical data and the growing number of disparate data sources are exposing researchers to a new challenge - how to acquire, maintain and share knowledge from large and distributed databases in the context of rapidly evolving research.

This paper describes research in progress on a new methodology for leveraging the semantic content of ontologies to improve knowledge discovery in complex and dynamic domains. It aims to build a multi-dimensional ontology able to share knowledge from different experiments undertaken across aligned research communities in order to connect areas of science seemingly unrelated to the area of immediate interest. We analyze how ontologies and data mining may facilitate biomedical data analysis and present our efforts to bridge the two fields, knowledge discovery in databases, and ontology learning for successful data mining in large databases. In particular we present an initial biomedical ontology case study and how we are integrating that with a data mining environment.

Keywords¾ **Biomedical discovery, Ontology, Data mining**

I. INTRODUCTION

The explosion of biomedical data and the growing number of disparate data sources are exposing researchers to a new challenge - how to acquire, maintain and share knowledge from large and distributed databases in the context of rapidly evolving research. Blagoskolonny and Perdee's proposal presented the "Conceptual Biology" challenge, to build a knowledge repository capable of transforming the current data collection era into one of hypothesis–driven, experimental research. In doing so we must consider that in addition to research-informed literature biomedical data is tremendously diverse and can consist of information stored in genetic code, identified in genomics and proteomics research by discovering sequencing patterns, gene functions, and protein-protein interactions, along with experimental results from various sources, patient statistics and clinical data. This might include, for example, the information collected in clinical patient data for clinical trial design, geographical and demographic data, epidemiological data, drugs, and therapeutic data as well as from many other dimensions that may appear not to be relevant when looked at from one specific research perspective but can turn to be crucial from another perspective. The problem of making diverse biomedical knowledge and concepts sharable over applications and reusable for several purposes is both complex and crucial. It is, however, central to enabling comprehensive knowledge-acquisition by medical research communities and molecular biologists involved in biomedical discovery.

Biomedical discovery itself is an intrinsically complex and risky process. One of the aspects of the biomedical discovery process is its iterative nature in terms of analyzing existing facts or data, to validate current hypotheses or to generate new ones. Opportunities arise by the simple act of connecting different facts and points of view that have been created for one purpose, but in light of subsequent information, they can be reused in a quite different context, to form new concepts or hypotheses.

From a philosophical point of view, discovery can be defined as "the act of becoming aware of something previously existing but unknown". This broad definition includes both kinds of scientific discovery: factual and conceptual. The former typically happens during the investigation of current "known" facts or data. The latter emerges from different points of view of "unknown" facts or data, and frequently finishes with a paradigm shift. Thus it is necessary for scientific discoveries to use "imagination" as well as reasoning. Ontologies can be used to facilitate both forms of scientific discovery in providing a common framework for several systems and problem solving methods. Ontology specifies at a higher level the classes of concepts that comprise the application domain and the classes of relations that exist between these concepts classes. The ontology captures the intrinsic conceptual structure of the domain. Therefore given a

domain, its ontology forms the heart of any system of knowledge representation for that domain.

This paper describes research in progress on a new methodology for leveraging the semantic content of ontologies to improve knowledge discovery in complex and dynamic domains, and aims to build a multi-dimensional ontology able to share knowledge from different experiments undertaken across aligned research communities in order to connect areas of science seemingly unrelated to the area of immediate interest.

In the first part of this paper we present and then argument the "Conceptual Biology" statement adding medical knowledge acquired from data in addition of literature facts. Biomedical Ontology is discussed as a solution to integrate different knowledge dimensions, and some current initiatives are presented.

In the following section we analyze how ontologies and data mining may facilitate biomedical data analysis and present our efforts to bridge the two fields, knowledge discovery in databases, and ontology learning for successful data mining in large databases. In particular we present an initial biomedical ontology case study and how we are integrating that with a data mining environment we are developing - Neucom. Finally we make some conclusions and show future directions for research.

II. "CONCEPTUAL BIOLOGY" AND BIOMEDICAL ONTOLOGIES

Biological knowledge is evolving from structural genomics towards functional genomics. The tremendous amount of DNA sequence information that is now available provides the foundation for studying how the genome of an organism is functioning, and microarray technologies provide detailed information on the mRNA, protein, and metabolic components of organisms [BOD 03].

At the same time, millions of easily retrievable facts are being accumulated from a variety of sources in seemingly unrelated fields, and from thousands of journals.

Biological knowledge is evolving so rapidly that it is difficult for most scientists to assimilate and integrate the new information with their existing knowledge.

A. Beyond Conceptual Biology

Considering the facts above, Blagoskolonny and Perdee discuss the emergence of "Conceptual Biology" – the iterative process of analyzing existing facts and models available in published literature to generate new hypotheses. They state, "The conceptual review should take its place as an essential component of scientific research". In doing so, new knowledge can be generated by 'reviewing' these accumulated results in a concept-driven manner, linking them into testable chains and networks [BLA 02].

Barnes [BAR 02] has increased Blagoskolonny and Perdee's proposal complexity through the argument that "scientists have traditionally worked in discrete communities, creating discipline-specific language." The natural consequence is that today we are faced with an overwhelming array of nomenclature for genes, proteins, drugs and even diseases.

The problem for scientists trying to perform 'conceptual' searches precisely and in a comprehensible manner is evident and has been addressed by different groups [PHA 03, GO 03, BIO 03]. These initiatives have in common the fact of using ontologies to represent their 'conceptual framework'.

In recent years ontology structures [GRU 93, SOW 02] have been increasingly used to provide a common framework across disparate systems, especially in bioinformatics [KOH 02, GLA 02, SCH 02], medical decision support systems [BUR 99, CHE 01], and knowledge management [VEL 01, AND 02]. *Ontology* is defined in the artificial intelligence literature as a specification of a conceptualization [GRU 02]. Ontology specifies at a higher level the classes of concepts that are relevant to the application domain and the relations that exist between these classes. Ontology captures the intrinsic conceptual structure of a domain. For any given domain, its ontology forms the heart of the knowledge representation [CHAN 02].

The use of ontology is a key towards structuring biological data [BAR 02] in a way that helps scientists to understand the relationships that exist between terms in a specialized area of interest, as well as to help them understand the nomenclature in areas with which they are unfamiliar.

Gene Ontology (GO) [GO, 03], for example, has been used to "produce a controlled vocabulary that can be applied to all organisms even if knowledge of genes and proteins is changing". GO is the basis for systems that address the problem of linking biology knowledge and literature facts, such as GO-KDS [TWO 03] and DiscoveryInsight [BIO 03].

GO KDS classifies the full collection of 12 million Medline publications. It is based on a classification system that uses GO term identification (go-id) to match it with MESH terminology in order to find abstracts in the Medline publications. However, it uses GO for terminology matching (syntax matching) instead of semantic concept matching and cannot be used to share and reuse information from different domains.

Another approach is used by Biowisdom, which employs its own set of ontologies to guide conceptual research for information retrieval from PubMed in the drugs domain. Biowisdom uses DiscoveryInsight to facilitate the formation of lateral 'connections' between scientific documents and records available to the

pharmaceutical industry. The ontologies help the association of key concepts, perhaps known by different names at different times or in different fields, and aims to stimulate the generation of insight into the complex network of factors involved in a disease process and generates novel, patentable ideas for therapeutics.

However, in addition to research-based literature the amount of data produced daily by medical information systems and medical decision support systems is growing at a staggering rate [LEU 00]. We must consider that scientific biomedical information can include information stored in the genetic code, but also can include experimental results from various experiments and databases, including patient statistics and clinical data. Large amounts of information and knowledge are available in medicine [VEL 01]. Making medical knowledge and medical concepts shared over applications and reusable for different purposes is crucial.

In biological systems, everything is interconnected, and ostensibly unrelated fields are related — the separation of biology into different disciplines is artificial [BLA 02]. Conceptual research can encompass many fields without limitation. So what is still needed is a way to manag the context of the search, so that terms having different meaning in different contexts can be retrieved appropriately. We also need ways to enable scientists to cross disciplines and search in areas outside their expertise, so that they can extract information critical for new discoveries. Biomedical ontologies are the best opportunity in this regard.

B. Biomedical Ontologies

Biomedical ontologies is an organizational framework of the concepts involved in biological entities and processes as well as medical knowledge in a system of hierarchical and associative relations that allows reasoning about biomedical knowledge.

Biomedical ontologies should provide conceptual links between data from seemingly disparate fields. This might include, for example, the information collected in clinical patient data for clinical trial design, geographical and demographic data, epidemiological data, drugs, and therapeutic data, as well as from different perspectives as those collected by nurses, doctors, laboratory experts, research experiments and so on.

At the same time the framework should reuse and integrate as many as possible different ontologies. The ontologies should integrate terminologies, such as UMLS [NCI 03] as well as domain specific ontologies, such as disease ontologies and GO, in order to support the knowledge discovery process.

Furthermore, to leverage the maximum power of biomedical ontologies, it must be used for information retrieval as well as in the data preparation phase of knowledge discovery as the basis for a "semantic preparation phase" that will allow us to facilitate both forms of scientific discovery, factual and conceptual [NOE 02], in providing a common framework for several systems and problem solving methods.

III. LINKING CONCEPTUAL AND THEORETICAL RESEARCH

In [BLA 02] the authors define the term "conceptual research" using the following metaphor: *"Connecting separate facts into new concepts is analogous to combining the 26 letters of the alphabet into languages. One can generate enormous diversity without inventing new letters. These concepts (words), in turn, constitute pieces of more complex concepts (sentences, paragraphs, chapters, books)."*

They argue that by searching successive pairs of terms, a chain or network of connections can be generated, and they use this metaphor to distinguish it from automated data-mining and from conventional theoretical biology [BRA 02]. In their point of view it is not a distinct type of science, but rather it has a different source– literature facts.

In the same direction, moving from an era of data collection into one of hypothesis driven research, Dennis Brady [BRA 02] discussed the importance of artificial models as another source of information – computer models. His argument is based on the power of these models to guide new hypotheses in a biomedical discovery process.

From a philosophical point of view, these works are complementary rather than divergent. Discovery can be defined as "the act of becoming aware of something previously existing but unknown" [NOE 02]. This broad definition includes both kinds of scientific discovery: factual and conceptual. The former typically happens during the investigation of current "known" facts or models. The latter emerges from different points of view concerning "unknown" facts or data that appear not to be relevant when looked at from one specific research perspective, and frequently finishes with a paradigm shift. Thus it is necessary for scientific discoveries to use "imagination" as well as reasoning.

IV. ONTOLOGIES AND DATA MINING

The current interest in ontologies relates to the change of focus of the area of Artificial Intelligence (AI) from content theories to mechanism theories that are a significant part of rule systems, frame languages, neural nets, fuzzy logic, constraint propagation, or unification [CHA 99]. The mechanisms are considered the secret of making intelligent machines. At other times, we realize that, however wonderful the mechanism are, it cannot do much without a good content theory of the domain on which it is to work. Moreover, we often recognize that once good content theory is available, many different mechanisms might be

used equally well to implement effective systems all using essentially the same content.

Techniques are essential, as is the careful collection of quantitative data. But without ideas to give them shape and meaning, those endless successions of base sequences, expression profiles, electrical recordings and confocal images are featureless [BRA 02].

Our approach integrates both content theories and mechanism theories. The Evolving Connectionist Systems (ECOS) [KAS 02] paradigm that is aimed at building on-line, adaptive intelligent systems that have both their structure and functionality evolving over time is used as a mechanism to find new relationship and patterns from the data. The rules extracted update the content, represented as an ontology that is used as a knowledge visualization tool for another data mining process.

A. Semantic Preparation

At the initial phase of the knowledge discovery process we use the ontology to guide our data selection and preparation, as shown in figure 1. In this way we use the ontology to enhance semantically the data preparation phase.

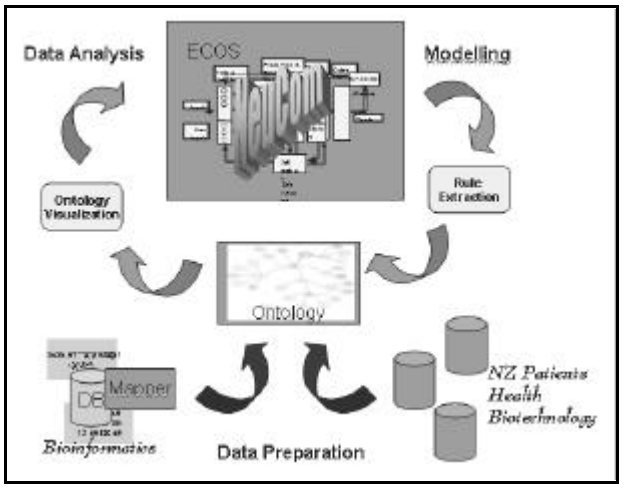

Fig.1 Proposed framework.

We do not presume that ontology is complete at the time a new data mining application is begun on the contrary; we believe that new domains will bring new types of variables and knowledge about them. However, we also believe that data mining is not simply the one-time application of a program to a new database. In our own work, data mining frequently starts with small pilot studies and manual bias space search, including feature construction. With preliminary confirmation that the programs can find some interesting relationships, more data and greater expectations are introduced.

V. INFOGENE MAP

Infogene Map is a case study that aims to build a multi-dimensional biomedical ontology, figure 2, able to share knowledge from different experiments undertaken across aligned research communities in order to connect areas of science seemingly unrelated to the area of immediate interest.

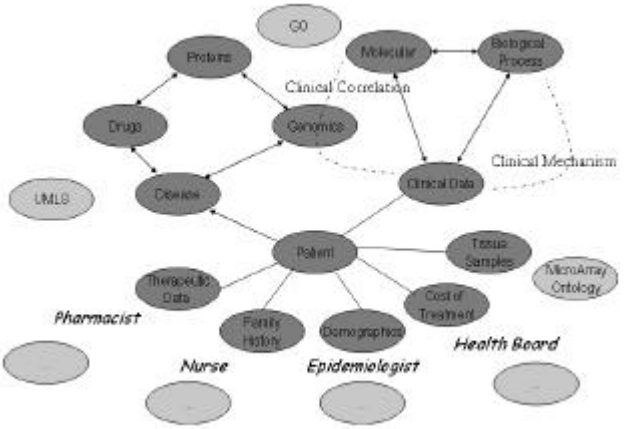

Fig.2 Ontology scope.

A. Infogene Map Ontologies

There are currently six ontologies included or developed in the Infogene Map. Each of them represents a specific domain in the Biomedical area.

A.1 Concept Metadata

Concept Metadata Entity is responsible for define a flexible knowledge representation for any concept present in the other specific ontologies. This entity contains schemes to represent the following knowledge:

Type Scheme – allow us to import and represent various type of information, such as image, text.

Spatial Scheme – represent the geographic knowledge.

Temporal Scheme – represents the time notion in the ontologies.

Language Scheme – allow us to acquire concepts in a language and maintain link with synonyms in other language.

Source Scheme – give the flexibility to acquire information from different sources of information, such as UMLS, clinical data, and maintain its independence of the original source.

Relationship Scheme – represents known relationships, such as, part_of, responsible_for, and permit the creation of new relationships acquired from the expert or through the data mining process.

A.2 Biomedical Domain

This entity represents the biomedical knowledge in the Infogene Map. It includes abstracts concepts, such as organism, and more concrete concepts, such as disease and its instances.

The biomedical concepts uses the concept metadata to define its source of information and any other abstract dimension needed to well represent its knowledge.

A.3 Biomedical Informatics Domain

Biomedical Informatics domain represents the common knowledge between biomedical domain and bioinformatics domain. Each subclass of this entity, such as oncogene, inherits characteristics from its domain and properties related with the biomedical informatics domain.

A.4 Clinical Domain

Clinical domain classes are responsible for represent the clinical knowledge contained in laboratories results, signs, drugs and so on.

The subclasses are mainly multi-inherited from biomedical domain and its instances are directly updated from databases.

A.5 Gene Ontology

Gene ontology represents the bioinformatics knowledge in the Infogene Map. This entity is directly imported to our ontology and its instances are included through annotations tools developed by the Protégé 2000 team.

In the current stage we maintain GO included in the main ontology and use it without alter its knowledge representation.

In order to keep our ontology aligned with the current on going research projects around the world, every gene represented in the Infogene Map is an instance of GO. At the same time, we are updating the included GO project monthly.

A.6 Disease Gene Map

This ontology is the core of the Infogene Map. It is responsible for build the gene/disease relationship, as shown in the section B.2.

Each instance of this ontology represents an experiment and is traceable through a query language that allow us, for example, to answer questions, such as "which genes are related with Leukemia?"

B. Infogene Map characteristics

Infogene Map is a frame-based ontology developed using Protégé 2000 [PRO 03]. It uses some of its tools, such as Prompt, to support the ontology engineering process.

It includes knowledge acquisition tools that allow domain expert and ontology engineers to built and refine the knowledge representation at the same time that populate instances in the knowledge base.

Infogene Map is being integrated with data mining tools in order to learn and acquire new knowledge from the knowledge discovery process.

C. Project development stages

We have defined three development stages for this project:

- Ontology Integration
- Disease/Gene Map
- Ontology Automatic Learning

The first stage, Ontology Integration, is directly related to ontology engineering issues. Specially, we are dealing with the best practices to build ontologies in the biomedical domain. The second and third stages are related with ontology/data mining integration and involve learning techniques in both research fields.

The current version covers the two first stages – Ontology Integration and Disease/Gene Map.

C.1 Ontology Integration

It is well accepted in the ontology engineering community that reuse is a key factor in projects that aim to integrate different domains or different sources of information under the umbrella of an ontology [BOD 02, GOM 99, TAB 01]. At this initial stage Infogene Map includes two of the most used ontologies, Gene Ontology and Unified Medical Language System (UMLS) terms [NCI 03], to represent respectively genes and biomedical knowledge as shown in figure 3.

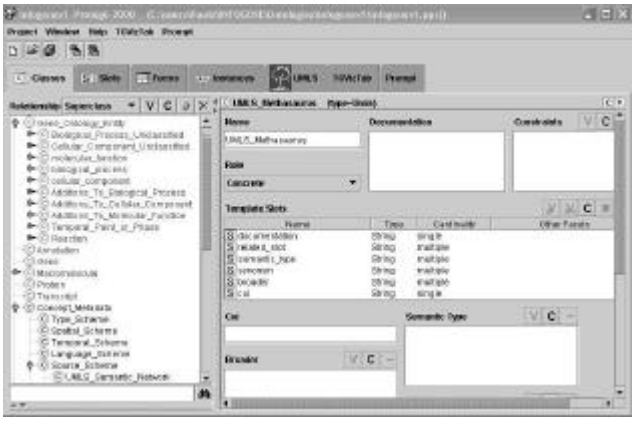

Fig.3 Ontology.

Our biomedical ontology was projected to be generic enough to integrate different sources of information and types of information. In the current development phase we are representing biomedical concepts based on the UMLS semantic network, UMLS metathesaurus, the knowledge

acquired from the domain expert as well as from knowledge acquired directly from clinical databases.

The first version is able to import knowledge directly from flat files and relational databases, and uses Protégé UMLS tab to import metathesaurus directly from the UMLS knowledge server. UMLS semantic network terms are included from scratch based on the UMLS semantic navigator [BOD 01]. Domain knowledge is acquired using knowledge acquisition forms built in Protégé 2000 based on interviews with experts.

C.2 Disease/Gene Map

Infogene Map is primarily focused at this stage on the gene-disease relationship. We are representing graphically, figure 4, these relationships in a way that enables visualisation and creation of new relationships. We are using additional properties to define and weight those items of knowledge acquired from ECOS. This approach enables us evolve the maps as new knowledge is discovered, by the use of the data mining techniques available in the Neucom environment [KED 03].

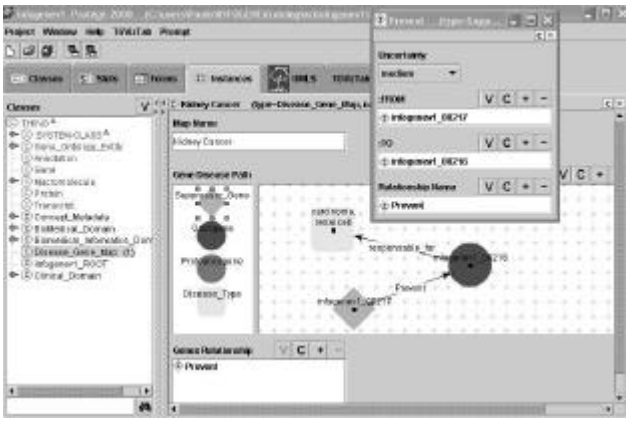

Fig.4 Gene/Disease map example.

Neucom is self-programmable learning and reasoning computer environment based on connectionist (Neurocomputing) modules. Neucom learns from data, thus evolving new connectionist components. The modules can adapt to new incoming data in an on-line incremental, life-long learning mode, and can extract meaningful rules that would help people discover new knowledge in their respective fields.

There are two disease/gene maps being developed as case studies: Leukemia and Kidney (renal) cancer. Both cases are using Gene Ontology to represent genes and UMLS definitions to represent the relationships among diseases and biomedical concepts. Each map is an instance of the experiment realized and these experiments can be further explored by queries in the knowledge base.

The maps are shown in different visualizations and can be used as a knowledge acquisition tool to support the domain expert during her or his analysis.

At this stage we are using Neucom separately from the Protégé environment. The next version will work in an integrated manner with Neucom using a knowledge management model. Some of these features are already developed but remain isolated from this current version.

VI. DISCUSSION

All really big discoveries are the result of thought, in biology as in any other discipline. Allostery, genes, DNA structure, chemi-osmosis, immunological memory, ion channels were all once just an idea [BRA 02].

A knowledge repository that is sharable and capable of moving the current data collection era into one of hypothesis–driven research is essential to support new biomedical discoveries. The conceptual biology and theoretical biology proposals are start to taken us in this direction. However, in order to be able to evolve the ontology map with the huge amount of information produced daily worldwide, any knowledge repository must be flexible enough to represent information from diverse sources of information and in different formats and be able to represent dynamic relationships.

Modeling these data interactions, learning about them, extracting knowledge, and building a reusable knowledge base applying the state of the art of AI and soft-computing will guide future research and practice and this is in the core of our research.

Although content theories and mechanism theories have been viewed as divergent approaches, we believe that a hybrid system that integrates and leverages the best of both theories is a sound approach to support a knowledge discovery process capable of evolving in environments where the process is developing, changing over time in a continuous manner.

Our effort is an attempt to integrate both paradigms aiming at leveraging the semantic content of ontologies to improve knowledge discovery. Neucom has a solid set of data analysis and modelling tools and its integration with the Infogene Map Ontology is enhancing semantically the results of the data mining.

Additionally, reusing models significantly reduced the time and costs of building a new application. Reusing knowledge components across different applications and domains helps to acquire expert knowledge and accurately describe the reasoning process. However, a methodology to guide this reuse process is still needed. Moreover, the integration of Gene Ontology and UMLS is a research question that is being followed by our group and many others around the world [CAN 03, YU 99].

VII. CONCLUSION

Although our approach is directed towards many of the current problems in the ontology integration area, the next stage is still an open question. Automatic ontology learning from data is a major challenge for the next development phase.

Transparent integration between Protégé and Neucom has to be completed in order to facilitate the knowledge acquisition by the domain expert him/herself.

The amount of information represented on the Web and the advance of semantic web will guide our future implementation. Infogene Map will be translated to different representation formalisms, such as OWL [OWL 03], to be able to acquire and represent web sources of information.

The biomedical ontology is currently small but will be extended to include life style and other patient related variables as well as including other diseases that are being investigated in our research group.

Our main contribution at this stage is in putting together the best practice in ontology engineering and the best techniques in data mining in order to facilitate the biomedical discovery process. We believe that Infogene Map contributes to biomedical discovery bringing together different dimensions acquired from diverse perspectives and useful for different groups of users such as researchers, doctors, nurses and pharmacists.

Furthermore, integrating this knowledge with Neucom will allow us to introduce a new dimension of knowledge produced by artificial models using the best mathematical modeling techniques.

REFERENCES

[BOD 03] BODENREIDER, O., MITCHELL, J., and MCCRAY, A., "Biomedical ontologies." presented at Pacific Symposium on Biocomputing 2003: World Scientific, 2003.

[BLA 02] BLAGOSKLONNY, M. V. and PARDEE, A. B., "Unearthing the gems." Nature, vol. 416, pp. 373, 2002.

[BAR 02] BARNES, J. C., "Conceptual biology: a semantic issue and more." Nature, vol. 417, pp. 587-588, 2002.

[GO 03] GO, "Gene Ontology," Retrievable from the Internet 15/10/2003 at: http://www.geneontology.org, 2003.

[BIO 03] BIOWISDOM, "Biowisdom," Retrievable from the Internet 15/10/2003 at: www.biowisdom.com/. 2003.

[PHA 03] PHARMGKB, "Pharmgkb," Retrievable from the Internet 15/10/2003 at: www.pharmgkb.org/. 2003.

[GRU 93] GRUBER, T. R., "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," presented at International Workshop on Formal Ontology, Italy, 1993.

[SOW 02] SOWA, J. F., "Ontology, Metadata, and Semiotics," vol. 2002, 2002.

[KÖH 02] KÖHLER, J. and SCHULZE-KREMER, S., "The Semantic Metadatabase (SEMEDA):Ontology based integration of federated molecular biological data sources," In Silico Biology, vol. 2, 2002.

[GLAS 02] GLASS, A. and KAROPKA, T., "Genomic Data Explosion - The Challenge for Bioinformatics," in Advances in Data Mining - Applications in E-commerce, Medicine, and Knowledge Management, Lectures Notes in Artificial Intelligence, P. Perner, Ed., 2394 ed. Berlin: Springer, 2002, pp. 80-98.

[SCH 02] SCHULZE-KREMER, S., "Ontologies for molecular biology and bioinformatics," In Silico Biology, vol. 2, 2002.

[BUR 99] BURGUN, A., BOTTI, G., FIESCHI, M., and LE BEUX, P., "Sharing knowledge in medicine: semantic and ontologic facets of medical concepts," presented at Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference, 1999.

[CHE 01] CHEAH, Y. N. and ABIDI, S. S. R., "Augmenting knowledge-based medical systems with tacit healthcare expertise: towards an intelligent tacit knowledge acquisition info-structure," presented at Computer-Based Medical Systems, 2001. CBMS 2001. Proceedings. 14th IEEE Symposium on, 2001.

[VEL 01] VELARDI, P., FABRIANI, P., and MISSIKOFF, M., "Using text processing techniques to automatically enrich a domain ontology," in Proceedings of the international conference on Formal Ontology in Information Systems. Ogunquit, Maine, USA: ACM Press, 2001, pp. 270--284.

[AND 01] ANDRADE, H. and JOELSALTZ, "Towards a Knowledge Base Management System (KBMS): An Ontology - Aware Database Management System (DBMS)."

[GRU 02] GRUBER, T., "What is an Ontology?" vol. 2002, 2002.

[CHA 99] CHANDRASEKARAN, B. J., J.R.; BENJAMINS, V.R., "What are ontologies, and why do we need them?" Intelligent Systems, IEEE [see also IEEE Expert], vol. 14, pp. 20-26, 1999.

[TWO 03] TWO, R., "GO-KDS," Retrievable from the Internet 15/10/2003 at: www.go-kds.com/. 2003.

[BIO 03] BIOWISDOM, "DiscoveryInsight," Retrievable from the Internet 15/10/2003 at: www.biowisdom.com/press_15.htm, 2003.

[LEU 00] LEUNG, K. S., CHAN, L. W., and MENG, H., "Intelligent data engineering and automated learning," presented at the IDEAL 2000: data mining, financial engineering, and intelligent agents: second international conference. New York, 2000.

[NCI 03] NCI, "Unified Medical Language System," Retrievable from the Internet 15/10/2003 at: www.nlm.nih.gov/research/umls/. 2003.

[NOE 02] NOE, K., "The Structure of Scientific Discovery: From a Philosophical Point of View," in Progress in Discovery Science, LNAI 2281, S. Arikawa and A. Shinohara, Eds. Berlin Heidelberg: Springer-Verlag, 2002, pp. 31-39.

[BRA 02] BRAY, D., "Reasoning for results," Nature, vol. 412, pp. 863, 2002.

[KAS 02] KASABOV, N., "Evolving connectionist systems for adaptive learning and knowledge discovery: methods, tools, applications VO - 1," presented at Intelligent

Systems, 2002. Proceedings. 2002 First International IEEE Symposium, 2002.

[PRO 03] PROTEGE, "Protégé 2000," Retrievable from the Internet 15/10/2003 at: http://www.protege.stanford.edu/. 2003.

[KED 03] KEDRI, "Neucom," Retrievable from the Internet 15/10/2003 at: www.theneucom.com, 2003.

[OWL 03] (OWL), W. O. W. W. G., "OWL," Retrievable from the Internet 15/10/2003 at: www.w3.org/2001/sw/WebOnt/. 2003.

[BOD 01] BODENREIDER, O., "Medical Ontology Research." National Center for Biomedical Communications, National Library of Medicine 2001.

[BOD 02] BODENREIDER, O., MITCHELL, J., and MCCRAY, A., "Evaluation of the UMLS as a terminology and knowledge resource for biomedical informatics." presented at ACL'2002 Workshop "Natural Language Processing in the Biomedical Domain", 2002.

[GOM 99] GOMEZ-PEREZ, A., "Ontological engineering: A state of the art." Expert Update, vol. 2, pp. 33-43, 1999.

[TAB 01] TABOADA, M., DES, J., JOSE, M., and MARIN, R., "Diagnosis systems in medicine with reusable knowledge components," Intelligent Systems, IEEE [see also IEEE Expert], vol. 16, pp. 68-73, 2001.

[CAN 03] CANTOR, M. N., N, S. I., HARTEL, F., BODENREIDER, O., and LUSSIER, Y. A., "An evaluation of hybrid methods for matching biomedical terminologies: Mapping the Gene Ontology to the UMLS." Stud Health Technol Inform, pp. 62-67, 2003.

[YU 99] YU, H., FRIEDMAN, C., RHZETSKY, A., and KRA, P., "Representing genomic knowledge in the UMLS semantic network." presented at Am Med Inf Assoc, 1999.