

Full citation: Gottgroy, P., Kasabov, N., & MacDonell, S. G. (2004) An ontology driven approach for knowledge discovery in Biomedicine, in Proceedings of the 8th Pacific Rim International Conference on Artificial Intelligence (PRICAI). Auckland, New Zealand, Springer-Verlag, Vol. 3157, pp.53-67.

An ontology driven approach for knowledge discovery in Biomedicine

Paulo Gottgroy, Nik Kasabov, Stephen MacDonell
Auckland University of Technology, School of Computer and Information Sciences,
Knowledge Engineering and Discovery Research Institute – KEDRI
AUT Tech Park 581-585 Great South Road
Ronald Trotter House Private Bag 92006
1020, Auckland, New Zealand
{paulo.gottgroy, nkasabov, stephen.macdonell}@aut.ac.nz
<http://www.kedri.info>

Abstract

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 Biomedicine, 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.

1. INTRODUCTION

The explosion of data and the growing number of disparate sources are exposing researchers to a new challenge -how to acquire, maintain and share knowledge from the large and distributed databases. In the biomedical domain, for instance, the problem of discover knowledge from biomedical data and making biomedical knowledge and concepts sharable over applications and reusable for several purposes is both complex and crucial. It is central to support the decision in the medical practice as well as 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 hypothesis.

Ontology is defined in the artificial intelligence literature as a specification of a conceptualization. Ontology specifies at a higher level the classes of concepts that are relevant to the 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.

Although ontology-engineering tools have matured over the last decade, manual ontology acquisition remains a tedious, cumbersome task that can easily result in a knowledge acquisition bottleneck. Therefore, in the context of evolving processes, ontologies should be created and refined automatically.

The knowledge discovery process is comprised of different phases, such as data preparation, cleaning and transformation, and each of these phases or steps in the life cycle might benefit from an ontology-driven approach which leverages the semantic power of ontologies in order to fully improve the knowledge discovery process and in the symmetric direction applies data analysis, data visualization and mining techniques to discovery semantic relationships, identify missing concepts, cluster concepts, thus refining and improving the ontology model.

Opportunities arise from the combination of ontology engineering and KDD process in order to improve both activities. KDD research can contribute to ontology learning through the application of data analysis, data visualization and mining techniques in order to identify new relationships among known concepts, to identify clusters and as a consequence define a better ontology hierarchy, and even help to discover new concepts among the data. Ontologies can help the knowledge discovery process introducing a new semantic layer to the process and moving it from a data driven approach to a knowledge driven approach.

This paper describes research in progress on a new methodology for leveraging the semantic content of ontologies to improve knowledge discovery in databases and to use data mining techniques to build, evolving, adapting and learning ontologies. In the first part of this paper we present and then argue 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, some current initiatives are presented, and, in particular, we present an initial biomedical ontology case study.

The second part presents our ontology driven framework and identifies the tasks required to combine ontology and machine learning. Finally we draw some conclusions and present the ongoing work towards the application of this framework in a data-mining environment.

2. “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 [1].

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.

2.1 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 [2].

In [3] Barnes 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 [4-6]. These initiatives have in common the fact of using ontologies to represent their ‘conceptual framework’.

In recent years ontology structures [7] have been increasingly used to provide a common framework across disparate systems, especially in bioinformatics [8],

medical decision support systems [9], and knowledge management [10].

The use of ontology is a key towards structuring biological data [3] 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) [4], 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 [11] and DiscoveryInsight [5].

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. 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. 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 [2]. Conceptual research can encompass many fields without limitation. So what is still needed is a way to manage 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.

2.2 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 [12] 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 [13], in providing a common framework for several systems and problem solving methods.

3. LINKING CONCEPTUAL AND THEORETICAL RESEARCH

In [2] 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. 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, [14] 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” [13]. 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.

4. INFOGENE MAP

Infogene Map is a case study that aims to build a multi-dimensional biomedical ontology (fig. 1), 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.

4.1 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.

4.1.1 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 permits the creation of new relationships acquired from the expert or through the data mining process.

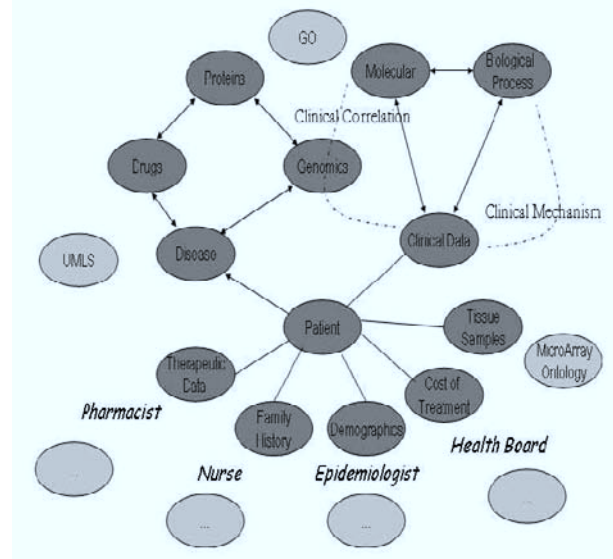


Figure 1: Ontology scope

4.1.2 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.

4.1.3 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.

4.1.4 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.

4.1.5 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.

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 ongoing 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.

4.1.6 Disease Gene Map

This ontology is the core of the Infogene Map. It is responsible for build the gene/disease relationship.

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?”

4.2 Infogene Map characteristics

Infogene Map is a frame-based ontology developed using Protégé 2000 [15]. 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.

4.3 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 fully the two first stages – Ontology Integration and Disease/Gene Map. The third stage is part of the framework that integrates ontology and data mining. This framework and current stage is covered in the next section.

4.3.1 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 ontology [16,17]. At this initial stage Infogene Map includes two of the most used ontologies, Gene Ontology and Unified Medical Language System (UMLS) terms, to represent respectively genes and biomedical knowledge as shown in figure 2.

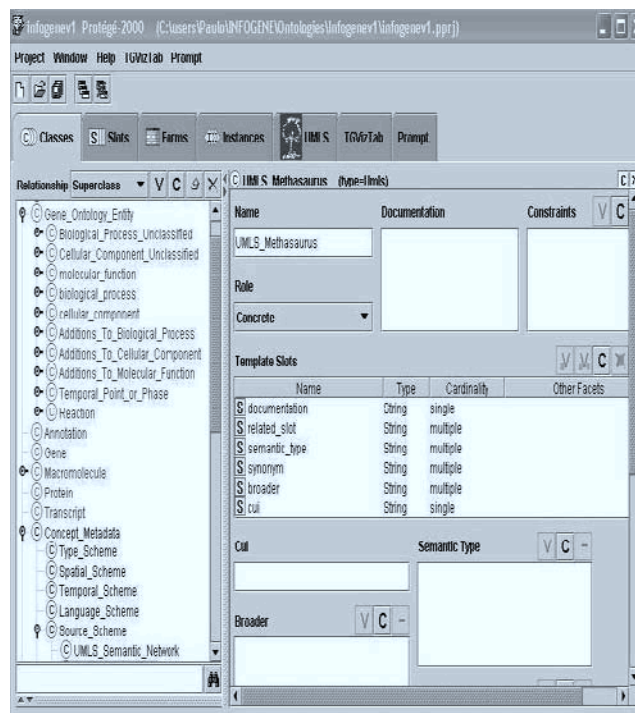


Figure 2: Ontology

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. Domain knowledge is acquired using knowledge acquisition forms built in Protégé 2000 based on interviews with experts.

4.3.2 Disease/Gene Map

Infogene Map is primarily focused on the gene-disease relationship. We are representing graphically (fig. 3), 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 [18]. 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 [19].

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.

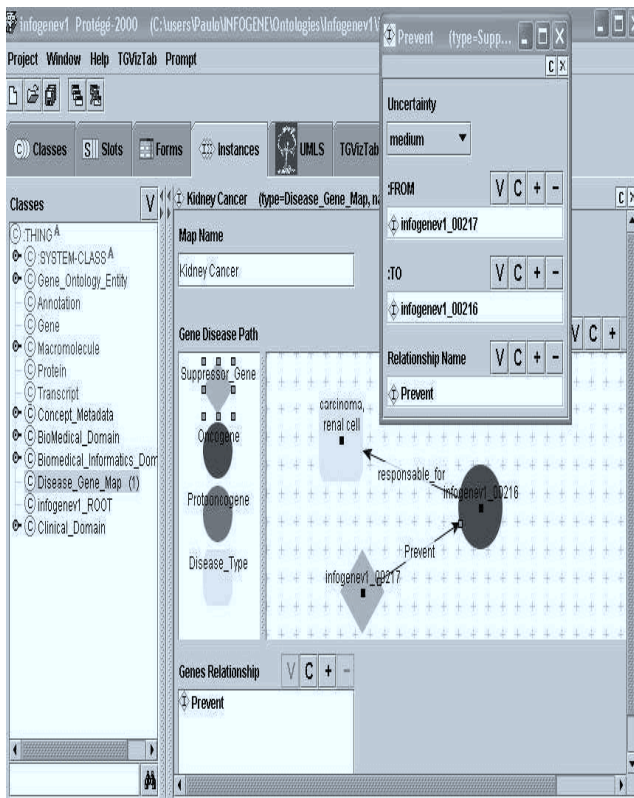


Figure 3: Gene/Disease map example

5. ONTOLOGY DRIVEN KNOWLEDGE DISCOVERY

Most of the current works in ontology learning, which combines ontology and machine learning, is focused on the application of a small number of mining techniques to extract some useful information to help the ontology engineer in the process of building ontologies. However, many other learning opportunities arise from the use of the entire knowledge discovery in databases life cycle and its techniques.

The knowledge discovery process is comprised of different phases, such as data selection, transformation, so each of these phases or steps in the life cycle might benefit itself from an ontology-driven approach.

Onto4KDD has been defined as an application of ontologies in order to improve the KDD process, for example, domain ontologies to support hypothesis driven approaches, such as in bioinformatics where many predictions and interpretations of biological data are made by comparing the data in hand against existing knowledge. The objective is to use ontologies to refine search tasks and even save computing time.

Another approach is KDD4Onto. This approach is focused on the application of mining techniques in order to learn and build ontologies. Currently, ontologies are mostly learnt from text [20,21] by an integration of lexicons, taxonomies, and ontologies, in the natural language processing and computational linguistics areas. Another incentive for this development comes from the Semantic Web research where ontology acquisition from text is a significant issue and has been well documented in recent years. Somewhat less traditional is the role of ontologies in the knowledge discovery in database field that includes learning from schemata and learning from

instances.

Although some researchers are addressing one of the above approaches, rare is the research that encompasses both perspectives. In the following section we present our ontology driven approach that integrates a data mining environment and ontologies in both directions Onto4KDD and KDD4Onto.

5.1 Onto4KDD4Onto

The current interest in ontologies is the latest instance of Artificial Intelligence's alternation of focus between content theories and mechanism theories [22]. Sometimes, some mechanism, such as rule systems, frame languages, neural nets, and fuzzy logic, excites the Artificial Intelligence community. The mechanisms are proposed as the key to making intelligent machines. At other times, we realize that, however wonderful the mechanism, 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.

This research is based on a hybrid approach that integrates content theory and mechanism theory. It defines an ontology driven knowledge discovery process framework (fig. 4) for leveraging the semantic content of ontologies to improve knowledge discovery in complex and dynamic domains and to use the KDD techniques to learn and refine ontologies.

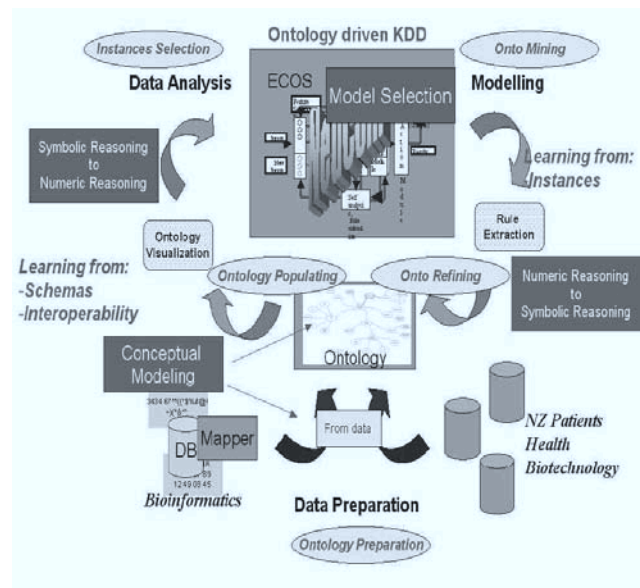


Figure 4: The Ontology Driven Knowledge Discovery process

In a nutshell ontologies can support the KD process by inserting a semantic layer to the process, that is, giving meaning to the results; and KD techniques can be useful to refine the ontology model by uncovering new concepts and relationships among them or even helping the ontologists in the modelling process.

There are five steps in the ontology driven knowledge discovery process:

- Ontology preparation;

- Ontology population;
- Instance selection;
- Ontology mining;
- Ontology refining.

5.1.1 Ontology Preparation / data preparation

Ontology learning is the set of techniques applied to reduce the effort expended during the knowledge acquisition to build ontologies. It aims to integrate different methods to assist a knowledge engineer in the ontology building process.

Ontology preparation is the process of analysing a data model and an ontology model in order to match conceptually both models. In this phase ontologies are refined, integrated or built through a learning process. In [23] we present a detailed explanation about this phase.

There are two different approaches to ontology learning in this phase:

- Ontology learning from schemata;
- Ontology learning for interoperability

Ontologies are used in this phase to integrate different databases and to merge with other available ontologies. Integrating databases enables a shared conceptualization about the domain that at the same time improves the quality of the “input data source”.

5.1.2 Ontology Population

Ontology population is the act of populating the knowledge base with the instances acquired from the databases. This is the next natural step after the conceptual model matching. There are different approaches to integrate and populate ontologies and databases [24,25]. In brief, after two concepts have been identified as similar the instances of the ontology and/or the records of the database are stored in the ontology to build the knowledge base.

5.1.3 Instance Selection / Feature Selection

Instance selection is the final step before the selection and application of a mining technique. In the KDD process it is one of the most important steps. In bioinformatics, for example, in a gene expression analysis, the number of samples may be small but the number of genes can be incredibly huge, thus any support for gene selection, should help to identify those that are significant in the problem analysis, reduce the amount of computing time needed and give more meaning for the results. Ontologies/Instance selection plays an important role in acquiring previous knowledge that can help the bioinformatician in the selection of features.

Instance selection from the ontology requires explicit knowledge about the domain. 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, can be reused in a quite different context, to form new concepts or hypotheses. Navigation through the hierarchy and through the relationships increases the intelligence density transforming the data into explicit knowledge represented in the ontology.

On the other hand, KDD techniques help the ontologists to identify new relationships among concepts when using data analysis techniques, such as PCA

(principal component analysis), ICA (independent component analysis), and when visualising the data through 2D, 3D, Surface Plotting.

Ontology visualization has been improved with previous work and techniques in the data visualization field. At the same time, data visualization is enhanced with the “knowledge visualization feature” represented within the ontology.

5.1.4 Onto mining / data mining

Onto-mining is mainly recognised as ontology learning from instances. Most of the research in this area is concentrated on learning from textual and semi-structured resources in the process of building an ontology. However less has been said about the use of ontologies as an inference mechanism. The reasoning within the links and relationships of an ontology can help the mining techniques to produce improved results. Ontologies can help mining techniques in the same sense as the application of mining techniques in a good data warehouse lead to better results than when applied to raw data.

Data mining techniques can help to identify new relationships among concepts within the ontology, in turn refining or updating the ontology model. However, there are many mining techniques and cross disciplines techniques unexplored, such as FCA, neural networks, and statistical methods.

Our understanding here is that both fields have much to contribute to each other and a broad vision is needed to integrate them. In the Future Work section we present a brief introduction about our ongoing research to integrate ontologies and Neucom (A Neurocomputing Environment for Data Mining, Knowledge Discovery and Intelligent Decision support systems).

5.1.5 Onto refining/building

The main task in this phase is to translate numeric results into symbols. For each technique selected as a learning approach the result format must be analysed and a mapping technique applied to translate it into symbols that can be represented in the formalism used by the ontology. Some of the research in this direction includes the same merging techniques utilized in ontology integration, such as FCA [26]. This step was included in the framework in order to encompass the field of research that is concerned with formalism interoperability.

6. CONCLUSIONS

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 ontologies, 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.

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. In this paper we present our first step towards a methodology to integrate both content and mechanism approaches – ontology driven knowledge discovery process. The framework identifies a sequence of necessary steps in a life cycle that integrates both ontology engineering and knowledge discovery. It guides the selection of techniques that are suitable to improve both Onto4KDD and KDD4Onto.

We believe that this work contributes to both ontology engineering /KDD and Biomedicine by integrating their processes in a common process that leverages the content power of ontologies and the learning power of data mining techniques in order to build biomedical ontologies.

In the next section we present a brief vision of the ongoing project integrating the data mining environment developed in our research group and Protégé.

7. FUTURE WORK

Our ongoing research is focused on the integration of Neucom and Protégé. Our aim is to develop and test Neucom as an ontology-learning tool, and, at the same time, integrate it with an ontology editor in order to add semantics to Neucom and to build an ontology editor environment.

Neucom is a self-programmable, learning and reasoning computer environment based on connectionist (Neurocomputing) modules. It is based on the theory of Evolving Connectionist Systems (ECOS) that enables the environment to adapt to news inputs and evolve its behaviour over time.

Protégé is an ontology editor and a knowledge-base editor developed at Stanford. It has plug-ins that extends the environment, such as inference and reasoning, and visualization. Protégé was selected considering its features and its already available plug-ins that support visualisation, merging, and semantic web integration.

Our approach takes advantage of ECOS evolving characteristic accumulating knowledge of the databases as new inputs are added to the model. It is appropriate for and applicable to knowledge discovery from data because KDD is an iterative process where any change in one of the source databases, or in the environment, might represent an input to a new knowledge discovery process. In the symmetric direction, the results acquired from ECOS are represented in an ontology adding meaning to it, and allowing reuse of the knowledge discovered.

This ongoing work aims to build an evolving environment that shares both state-of-the-art of knowledge discovery and ontology engineering. Currently we are applying some of Neucom features in order to learn from ontology instances. The integration with Protégé is achieved manually by some plug-ins developed by us and by the Protégé community.

Additional features will be developed in order to better merge ontologies and databases. Some of the already promised techniques are Formal Concept Analysis, Theory of information Flow and Conceptual clustering.

8. REFERENCES

- 1 Bodenreider, O., Mitchell, J., & McCray, A. (2003). Biomedical ontologies. In Pacific Symposium on Biocomputing 2003: World Scientific, 2003.
- 2 Blagosklonny, M. V. & Pardee, A. B. (2002). Unearthing the gems. *Nature*, (v. 416, pp. 373).
- 3 Barnes, J. C. (2002). Conceptual biology: a semantic issue and more. *Nature*, (v. 417, pp. 587-588).
- 4 GO, "Gene Ontology," Retrievable from the Internet 15/10/2003 at: <http://www.geneontology.org>, 2003.
- 5 BOWISDOM, "Biowisdom," Retrievable from the Internet 15/10/2003 at: www.biowisdom.com/. 2003.
- 6 PHARMGKB, "Pharmgkb," Retrievable from the Internet 15/10/2003 at: www.pharmgkb.org/. 2003.
- 7 Gruber, T. (2002). What is an Ontology? Retrievable from the internet 28/05/2004 at: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>.
- 8 Köhler, J. & Schulze-Kremer, S. (2002). The Semantic Metadatabase (SEMEDA): Ontology based integration of federated molecular biological data sources. In *Silico Biology*, vol. 2.
- 9 Burgun, A., et al (1999). Sharing knowledge in medicine: semantic and ontologic facets of medical concepts. Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference.
- 10 Andrade, H. & Saltz, (2003). Towards a Knowledge Base Management System (KBMS): An Ontology -Aware Database Management System (DBMS)."
- 11 TWO, R., "GO-KDS," Retrievable from the Internet 15/10/2003 at: www.go-kds.com/. 2003.
- 12 NCI, "Unified Medical Language System," Retrievable from the Internet 15/10/2003 at: www.nlm.nih.gov/research/umls/. 2004.
- 13 Noe, K. (2002). The Structure of Scientific Discovery: From a Philosophical Point of View, In *Progress in Discovery Science*, p. 31-39, LNAI 2281, S. Arikawa and A. Shinohara, Eds. Berlin Heidelberg: Springer-Verlag, (2002).
- 14 Bray, D. (2002). Reasoning for results. *Nature*, (v. 412, pp. 863).
- 15 Protégé 2000, <http://www.protege.stanford.edu/>. (2004).
- 16 Bodenreider, O., et al (2002). Evaluation of the UMLS as a terminology and knowledge resource for biomedical informatics. In *ACL'2002 Workshop "Natural Language Processing in the Biomedical Domain*.
- 17 Gómez-Pérez, A., Fernández-López, M., Corcho, O. (2004). *Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer-Verlag, Berlin Heidelberg New York (2004).
- 18 Kasabov, N. (2002). Evolving connectionist systems for adaptive learning and knowledge discovery: methods, tools, and applications. *IEEE Intelligent Systems*, 2002, 24-28.
- 19 KEDRI, Neucom (2004). www.theneucom.com.
- 20 Asunción G.P. & David M. (2003) Survey of ontology learning methods and techniques. Retrievable from the Internet 08/05/2004 at www.ontoweb.org.
- 21 Velardi, P., Fabriani, P., and Missikoff, M. (2001). Using text-processing techniques to automatically enrich a domain ontology, In *International conference on Formal Ontology in*

Information Systems. Ogunquit, Maine, USA: ACM Press, p.270-284 (2001).

22 Chandrasekaran, B.J., & Benjamins, V.R. (1999). What are ontologies, and why do we need them? IEEE Intelligent Systems, 1999, 14(1094-7167): 20-26 (1999).

23 Gottgroy, P., Kasabov, N. & MacDonell, S. (2003). An ontology engineering approach for Knowledge Discovery from data in evolving domains. Proceedings of the Data mining IV. Southampton, Boston: WIT press (2003).

24 Rubin D.L., et all. (2002). Automatic data acquisition into ontologies from pharmacogenetics relational data sources using declarative object definitions and XML. In: Proceedings of the Pacific Symposium on Biology, Lihue, HI, Eds. R.B. et all (2002).

25 Johannesson P. (1994). A Method for Transforming Relational Schemas into Conceptual Schemas. In 10th International Conference on Data Engineering, Ed. M. Rusinkiewicz, pp. 115 -122, Houston,IEEE Press.

26 Stumme, G. and Maedche, A. (2001). FCA-Merge: Bottom-up merging of ontologies. In 17th International Joint Conference on Artificial Intelligence (IJCAI'01), Seattle, Washington, USA, (2001).