

©2008 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Ontology Support for Biomedical Information Resources

Tharam S. Dillon, Elizabeth Chang, Maja Hadzic

Digital Ecosystems and Business Intelligence Institute, Curtin University of Technology, Australia

E-mail: t.dillon@curtin.edu.au, e.chang@curtin.edu.au, m.hadzic@curtin.edu.au,

Abstract

The increasing body of distributed and heterogeneous information and the autonomous, heterogeneous and dynamic nature of information resources are important issues hindering effective and efficient data access, retrieval and knowledge sharing. The importance of ontologies has been recognised within the biomedical domain and work has begun on developing and sharing biomedical ontologies. In this paper, we define ontology and ontology commitments and explain the main characteristics and representations of ontology models. Ontologies are highly expressive knowledge models and as such increase expressiveness and intelligence of a system. We highlight the significance of ontologies in a variety of semi-automatic and automatic tasks, and provide an illustrative example of an ontology-based multi-agent system designed to intelligently retrieve information about human diseases from a number of heterogeneous and dispersed information resources.

1. Introduction

Biomedical information is distributed all over the world. The rapid increase in available information makes the search for needed information complicated. Extensive time is required to locate the appropriate information by browsing and searching. When researchers need to retrieve information about proteins, the following tasks need to be performed [1]:

- identify appropriate information resources
- understand the content of those resources
- target queries to relevant resources
- communicate with the information resources
- retrieve the relevant data
- transform data between different formats
- merge results from different information resources

The available information is distributed over various information resources. Lots of the available information is repeated within numerous databases. Portions of this information may be related to each other and/or may overlap. The information is found in different formats that may not be compatible with each other. Two different databases may contain information on the same topic but the way this information is structured and organised within the databases may be different. In these situations,

it is quite difficult to automatically compare and analyse available information.

Efficient storage, use and management of knowledge require standardised organisation using a common underlying knowledge base. Information and the basic concepts used to describe this information needs to be shared. A unifying framework that represents shared understanding of the domain knowledge would make the intelligent management of the available information possible.

In most cases, the information resources exist autonomously and are designed independently from others. The information resources are quite heterogeneous in their content, data structure, organisation, information management and the like. Heterogeneity of the information resources makes their integration difficult. One of the main issues is the dynamic nature of the internet environment. Content within the information resources is changeable as it is continuously updated and modified.

We identified the following underlying issues hindering the effective and efficient data access, retrieval and knowledge sharing [1]:

1. The increasing body of distributed and heterogeneous information;
2. lack of an underlying knowledge base to enable shared understanding of concepts to facilitate information retrieval, knowledge integration and sharing; and
3. most of databases and information resources are autonomous, heterogeneous and dynamic (content is continuously updated).

In this paper, we discuss the role of the ontology within biomedical domain. We introduce ontologies in Section 2. We clarify ontology concepts including ontology community and ontology commitment (Section 3) and generalization and specialization of ontologies (Section 4). In Section 5, we discuss properties of ontologies. In Section 6, we highlight the significance of ontologies in a number of semi-automatic and automatic tasks within the biomedical domain. Characteristics and representation of ontology models are discussed in Section 7 and 8 respectively. In Section 10, we discuss some significant biomedical systems and in Section 10, we give an illustrative example of an ontology-based multi-agents system designed for intelligent retrieval of information. The paper is concluded in Section 11.

2. Ontology

An ontology, particularly in medicine, grew out of a perceived need for a controlled vocabulary [2,3]. The importance of ontologies has been recognised within the biomedical domain and work has begun on developing and sharing biomedical ontologies [4,5].

An ontology is an enriched contextual form for representing domain knowledge. Ontologies are highly expressive knowledge models and as such increase expressiveness and intelligence of a system. Ontologies have been suggested as a mechanism to provide applications with domain knowledge and to facilitate information sharing, support intelligent information retrieval and enable cooperation.

An ontology provides a shared common understanding of a domain and a means to facilitate knowledge reuse by different applications, software systems and human resources [6,7]. Using the ontologies to formalize the 'common knowledge' will reduce the possibility of duplicating experiments by different research groups (such as examining the same region of a DNA sequence), thus saving time and resources. This will create a cooperative environment, allowing coherence between different research teams in big research tasks. Gap analysis in biomedical research will also be easier. Research projects (past and present) would be mapped onto the target domain as specified by the ontologies. During the planning stage all research projects could use these maps to see where overlaps, redundancies, complementary research and (more importantly) no research has occurred or is occurring.

3. Ontology Community and Commitment

One way of categorizing human diseases is to describe them as genetically simple (Mendelian) or complex (multifactorial). Simple diseases manifest the presence of a mutated gene and the disease. Mental illnesses do not follow Mendelian patterns but are caused by a number of genes usually interacting with various environmental factors. Some mental illnesses may be caused by a physical dysfunction of the brain but it is not known what triggers this exactly and the exact cause of many mental illnesses are unclear. For example, genetic analysis has identified candidate loci on human chromosomes 4, 7, 8, 9, 10, 13, 14 and 17. There is some evidence that environmental factors such as stress, life-cycle matters, social environment, climate etc. are important. The characterization of mental illnesses is made even more complicated through the existence of different illness types, such as depression, bipolar disorder and schizophrenia, and illness subtypes, such as bipolar I and II. Despite major medical advances, the identification of

genetic and environmental factors responsible for many mental illnesses still remains unsolved and is therefore a very active research focus today. Mental health community could greatly benefit from a Mental Health Ontology that will enable them to share the research results effectively and efficiently.

Usually, a team of researchers who need to share information in a specific domain propose an ontology. The researchers may be inspired to design a common ontology for various reasons [8] such as:

- common understanding of the domain knowledge (knowledge within a community of people committed to a common goal)
- sharing of the domain knowledge
- reusing of the domain knowledge
- analyzing domain knowledge
- making domain assumptions explicit

Ontologies represent how a specific community understands part of the world. It is important for all members of the community to reach an agreement in (1) their understanding and definition of domain knowledge, and (2) representation of this knowledge through an ontology. A definition may be added to an ontology only with the consensus of all community members. No member can alter or override the definitions in an ontology only according to his/her preferences.

Ontological commitment is described as the agreement about the ontology concepts and relationships between those concepts within the ontology [9]. When a user, application or agent commit to an ontology, there is an agreement with respect to the meaning of the concepts and relationships represented. Furthermore, there is agreement to use the shared vocabulary coherently and consistently. Ontological commitment to the semantics (meaning) of the ontology concepts and their relationships in the common ontology includes commitments to the axioms, rules, constraints etc. stated in this ontology.

Community members, when committed to their common ontology, are able to use this ontology for various purposes. Any community can commit to ontologies developed by other communities. Many ontologies are already available in electronic format via ontology libraries. These ontologies can be imported into a new ontological development environment and can be reused by another application. A group of communities can accept and commit to an ontology with more general terms and fewer constraints. These communities can use such general ontologies to develop their own specialized ontologies.

Even though the Agent Communication Language (ACL) is generally used for agent communication [10,11], different agents can communicate with each other using the same ontology. When different agents commit to an ontology, this common ontology defines the

vocabulary and enables the different agents to exchange the information efficiently [12]. The agents can meaningfully communicate about a domain even when designed to be used by different applications [13].

In some situations, an agent can commit to several ontologies at the same time where each ontology is used in communication with a particular agent. In other situations, ontologies of different agents may contain knowledge complementary to each other. Those ontologies can then be merged into a large ontology that will serve as a common ontology for the agents. Of course, here it is very important to have these different ontologies formatted in the same way so that the agents can easily accept and process the new ontologies.

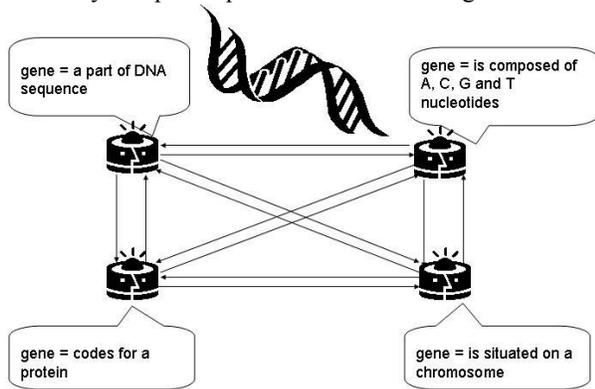


Figure 1 Different definitions of the same concept

Ontology commitments play an important role in the ontology design. It is possible for the same term to be defined differently within different ontologies from the same domain even though its meaning is the same i.e. they refer to the same concept. For example, within Genetics, the concept gene can be defined in different ways. All these statements in Figure 1 are correct and all of them need to be taken into account when defining the concept gene. Ontology design is an art in itself; it needs to be done by a number of domain experts and ontology engineers. Different perspectives need to be considered during the ontology design. Ontology commitments need to be carefully evaluated against the project objectives.

4. Generic and Specific Ontologies

Generic ontologies contain core concepts and relationships sharable across a wide domain. Generic ontologies are at a higher level of abstraction and contain less detailed information than specialized ontologies. Usually, more detailed specialized ontologies are derived from the generic ontologies (Figure 2). These new specialized ontologies are more detailed as they refine and extend the general descriptions present in generic ontologies.

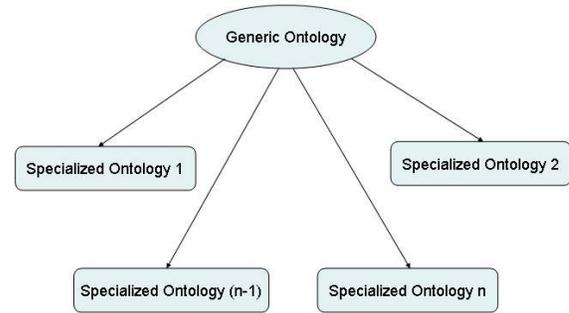


Figure 2 Ontology specialization

Generic ontologies are more likely to be shareable by a large number of communities. Usually, specialized ontologies are for a smaller community that makes part of a larger community. In our work [22], we use the principle of generic and specialized ontologies. We have designed a Generic Ontology of Human Diseases (GHDO) and, for example, Diabetes, Bipolar Disorder and Huntington Disease Ontologies have been derived as specialized versions of this GHDO.

Adding constraints can specialize an ontology to be used by a specific community. In the “Ontolingua”, library of ontologies, there are ontologies with general terms and minimum constraints that can be used by many communities to develop specialized ontologies [14]. We can add new terms along with their definitions to specialize an ontology for a subcommunity [15].

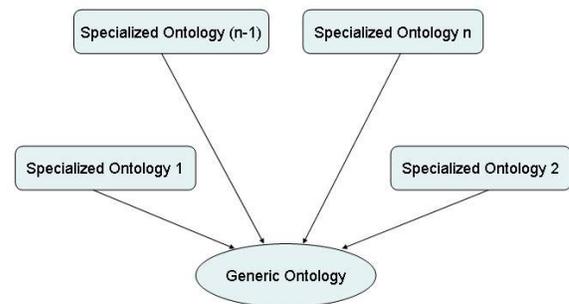


Figure 3 Ontology generalization

A generic ontology consists of a minimal number of axioms while a specialized ontology has a large number of axioms and needs a very expressive language. It is possible to progress gradually from generic to specialized ontologies through gradual increase of the number of axioms.

We can also build a generic ontology which is based on a number of specialized ontologies. Such a process is referred to as ontology generalization and is shown in Figure 3. This process aims to resolve the conflicts

between definitions of terms in the ontologies. The result of such a process is an ontology that can be accepted and shared by the participating communities.

5. Properties of Ontologies

Some essential properties of ontologies are:

- ontologies describe and are shared across a specific domain
- ontology concepts and relations are unambiguously defined in a formal language by axioms and definitions
- ontologies can be understood by computers
- relationships between ontology concepts determine the ontology structure
- ontology users agree to use the ontology terms consistently

Four important functions of ontologies are that they enable:

- integration of heterogeneous information resources
- design of intelligent search engines which look for the meaning of information over its appearance in the text
- access to the relevant information for automated agents
- communication between cooperatively working agents

Additionally, some ontologies such as UMLS [16] are useful resource of medical knowledge.

Ontologies are knowledge models that can be shared between different information resources even when they are operating independently from each other. A domain ontology can be designed to represent the knowledge shared between the different information resources. Each information resource needs to commit to this common ontology. This would enable integration of heterogeneous information resources and effective and efficient cooperation, collaboration and knowledge sharing between them.

Specific and targeted searches are very difficult with most search engines as they look for the specific string of letters within the text rather than its meaning. Use of ontologies significantly increase the expressive power and enable the search engines to perform semantic-based meaningful searches. The intelligent information retrieval system based on ontology technology will perform searches based on the context and meaning of the information, rather than simply its appearance in the text. One of the most significant breakthroughs in this area was the design of GOPubMed [17]. GOPubMed uses the Gene Ontology to perform searches over the information available in the PubMed database.

If the available information is to be more accessible to automated agents, it is essential that the meaning of this information can be understood by such agents. Typically, an ontology has a hierarchical structure, and describes domain concepts and their properties. As such, an ontology is a collection of shared and precisely defined terms that can be used in metadata to describe Web content in a machine-understandable manner. Adding information that describes Web content in a machine-understandable way enables Web information resources to become more accessible to automated processes [18]. People and software systems are able to access, extract and analyze information from different organizational sites when those sites operate on the basis of the same underlying ontology. Through this, ontologies play an important role in efficient and effective information access and retrieval.

Agents working cooperatively can communicate with each other by means of a common ontology. Ontology is machine readable, and support agent communication by defining and providing a shared vocabulary to be used in the course of communication. Ontologies do not only provide a definition of the terms that can be used in communication, but also the definition of the world in which an agent grounds its actions. Different agents of a system can reach a shared understanding by committing to the same ontology. This enables them to make statements, communicate knowledge and ask different queries. Use of ontology permits coherent communication and easier information sharing between different agents, enabling agents to cooperate and coordinate their actions.

6. Use of Ontologies

Various communities have recognized the importance of ontologies. Computer based ontologies constitute an essential resource for enabling interoperation and cooperation in an open environment such as the Internet. The cooperation factor is becoming more and more important every day; efficiency of the system is based on cooperative work between different people, agents, communities, applications etc.

Ontologies can be used to encode knowledge for an intelligent information system and, in this way, enable it to interrogate the reorganised and heterogenous information efficiently. The methodology proposed by Szirbik [21] provides an insight into how this engineering will change the way medical information is managed.

In order to access, manage and analyze the available information more effectively and efficiently and make the most out of this information, we need to aim for the implementation of ontologies in the following areas:

- (1) Data semantics. A biomedical ontology provides a model of biomedical concepts and relationships that can

be used to form a semantic framework for many data storage, retrieval and analysis tasks. Such a semantic framework could be used to support a variety of important tasks, such as the querying of heterogeneous information sources or for automatic data analysis such as data mining.

(2) Knowledge representation. An ontology is an expressive way to represent domain knowledge. Use of a shared ontology enables the users to fully understand the domain as they can look simultaneously at the different aspects of the problems and identify possible links. When all information on a specific topic is put together, it is also much easier to identify research areas where more examination is needed. Moreover, the inherited organization of ontologies adds a taxonomical context to search results and the user can identify conceptual relationships within the data. This can be especially important in studying complex diseases such as mental illness where many different factors, genetic and environmental, may cause these illnesses.

(3) Knowledge sharing. Ontologies can be designed to create shareable forms of knowledge. Different biomedical database systems may organize basic concepts differently for their own purposes, but the ontology, as the shared knowledge component, will make the integration of various data formats possible and enable sharing of the available knowledge. The common knowledge is going to reduce the necessity of undertaking same experiments, such as examining the same region of the DNA sequence, by different research groups, thus saving their time and their resources. Moreover, the shared ontologies can be used to help create a worldwide cooperative environment making it possible to take on a big task by dividing it among different research teams.

(4) Knowledge management. Ontologies can be used to provide expressive knowledge descriptions needed to structure, access, manage and maintain large sources of heterogeneous information within the target community. Use of ontologies for knowledge management facilitates effective organization of the existing knowledge and addition of new knowledge. It would be of the greatest importance to increase the value of the available information through the use of ontologies to help access and management of the biomedical data and information. It is important that existing knowledge becomes available to researchers. If aware of the research done by other medical research teams, scientists will be able to progress much faster, coming closer to the answer for their common problem. For example, extraction of data patterns from information regarding different genetic and environmental casual factors responsible for specific mental illness types will facilitate finding out the exact

combinations of the casual factors involved as well as their individual influences on a specific mental illness type.

(5) Intelligent information retrieval. Ontologies can be used for systematic annotation of information available through various information resources. The annotation makes it easier to understand the content of the information resource and retrieve relevant information. Most of the current search engines do not operate in this way and often retrieve voluminous results that are completely outside the domain user is interested in.

(6) Mediation. Ontologies can be used as mediator to bridge the heterogeneities between different data resources, information services and the user applications.

(7) NLP (Natural Language Processing) applications. Ontology provides factual and contextual information about the different knowledge sources and enables the machines to understand, analyze and extract the knowledge available through the sources.

(8) Communication between different agents. Different agents can communicate with each other using a common ontology. This ontology defines the vocabulary used in communication that enables the different agents to communicate efficiently.

7. Characteristics of Ontology Models

Analogous to the characteristics of the data models [19], we developed a similar scheme for the ontology models:

- Concept is an unit of thought. Concepts may have abstract meaning (e.g. human diseases) and physical meaning (e.g. gene).
- Term is a lexical representation of a concept (or name of a concept).
- Context is used to group related terms in an ontology. For example, in the context of human diseases, we group terms used to describe and represent the knowledge of human diseases (Figure 4).

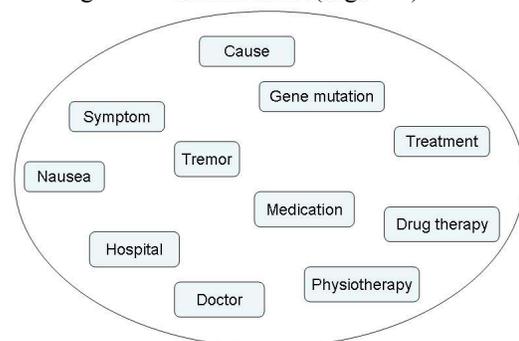


Figure 4 Few concepts in the context of human disease

- Terms exist in various relationships with each other. For instance, the term ‘gene’ is related to the term ‘DNA chromosome’ because a gene is situated on a DNA chromosome (Figure 5).

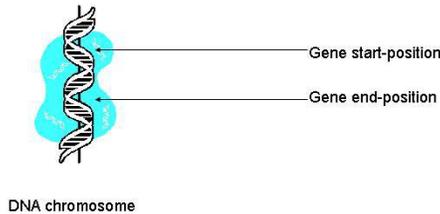


Figure 5 Gene is situated on a DNA chromosome

- The knowledge representation changes over time. The level of detail of the ontologies is related to the level of detail of the information. It is often necessary to filter and omit details of information in order to make the situation clearer by representing only information relevant to the concept within the context. At other times, this is inadvisable because the requested knowledge can be found only in the detailed information [20]. For example, when determining the exact position of a gene within human DNA, we firstly locate a chromosome (out of 23 different human chromosomes) that contains this gene. At this stage, the ontology associates only a chromosome with the term ‘gene position’. Secondly, we locate an arm of this chromosome (p or q arm) where the gene of interest is situated. At this stage, the ontology associates p or q chromosome arm with the term ‘gene position’. Finally, we locate the exact region within the chromosome arm, such as p11. At this final stage, the ontology associates the exact region of a chromosome with the term ‘gene position’. The knowledge model that describes the position of a gene within human DNA needs to be updated as new knowledge becomes available. Generally, the knowledge model needs to be updated when new knowledge emerges.
- In processes such as information retrieval, information needs to be integrated at different levels of detail. It may be necessary to navigate at different levels in the ontology structure, select the relevant information and integrate this information in a controlled way. Information may need to be combined at different levels of detail but merged on a specific level. A mechanism for integrating ontologies should provide these services.

- Incorporation of a new concept within an ontology may require presence of an additional concept. Similarly, deletion of an ontology concept may make presence of another ontology concept redundant. For example, gene mutation is a change of the gene structure which may cause a disease. If we introduce term ‘gene’ into an ontology, we may also need to introduce the term ‘DNA’. Because the ontological model is a semantic framework of concepts and their relationships, the model extension/reduction may change ontology structure significantly.

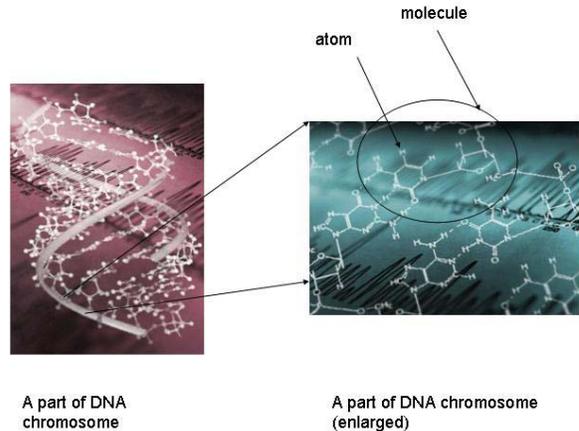


Figure 6 Double-helix DNA is composed of molecules, and molecules are composed of atoms

- Different terms may be incorporated into the ontology to enable definition of a new term. For example, as we show in Figure 6, human DNA is a double-stranded helix composed of four different nucleotides that consist of a base (Adenine (A), Cytosine (C), Guanine (G) and Thymine (T)), a phosphate molecule and a sugar molecule (deoxyribose). The nucleotides are grouped together in a helix forming human DNA. Different atoms are grouped together to form a molecule and different molecules are grouped together to form the DNA helix. The level of detail described by an ontology will depend on the purpose of the ontology design.

8. Representation of Ontology Models

Hierarchies have been widely used to represent ontologies. Two main reasons for this are that: (1) hierarchies are similar to the way we organize the mental models of the world in our minds, and (2) hierarchies allow for the generalization/specialization mechanism in information processing. However, in addition to these generalization/specialization relationships, one can have association relationships between concepts and also

composition relationships which can also have a hierarchical structure.

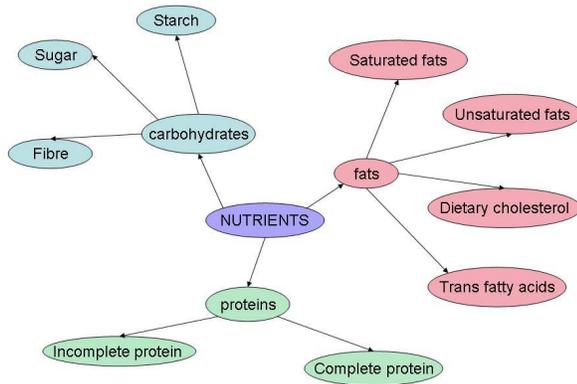


Figure 7 Top-level ontology of the Nutrients Ontology

The generalization/specialization relationship (i.e. “is-a” relationship) between two concepts is an example of a hierarchical relationship between concepts. In the example from Figure 7, we represent top-level of Nutrients Ontology that describes the basic classification of nutrients on carbohydrates, proteins and fats. At the top of hierarchy, we have the concept ‘nutrients’. On the following level down, we have three concepts: ‘carbohydrates’, ‘proteins’ and ‘fats’. The reason why particularly those three different groups are chosen to classify the concept ‘nutrients’ depends on many factors e.g. ontology use and purpose. The ontology has further branches as follows: ‘carbohydrates’ can be ‘starch’, ‘sugar’ or ‘fibre’, ‘proteins’ can be ‘incomplete protein’ or ‘complete protein’ of vegetable or animal origin respectively. ‘Fats’ can be ‘saturated fats’, ‘unsaturated fats’, ‘dietary cholesterol’ and ‘trans fatty acids’. The knowledge in the Nutrients Ontology is represented through hierarchical relationships between different concepts. Knowledge about specific foods, such as honey, can be represented by composition and association relationships. Honey is mainly composed of sugar (80%) and includes fructose, glucose, sucrose, maltose and melezitose.

Although ontologies often assume the form of a taxonomic class hierarchy, they are not restricted to hierarchies in any way. In fact, ontologies may take on much more general and complex structures. The ontology concepts may be grouped together to form layers of related hierarchy trees, networks, cubical structures or any other shapes. The way the ontology concepts are grouped together depends on the purpose of ontology design and on the complexity of knowledge it represents.

9. Some Significant Contributions to the Biomedical Domain

The importance of ontologies has been recognised within the medical community and work has begun on developing and sharing biomedical ontologies [22]. Gene Ontology (GO) (<http://www.geneontology.org/>) project works on establishing consistent descriptions of gene products in different databases by using the GO to annotate major repositories for plant, animal and microbial genomes. The Unified Medical Language System (UMLS) [23] is a collection of many biomedical vocabularies and it consists of Met thesaurus, Semantic Network, SPECIALIST Lexicon and a number of software tools. There are 1 million biomedical concepts in UMLS, and 135 semantic types and 54 relationships are used to classify these concepts.

Use of the ontologies in information retrieval, extraction and data mining has been reported. In the sequel, we introduce some of the existing tools. A large number of these tools have the annotation system as a part of the tool. Textpresso [24] perform semantic searches through *Caenorhabditis elegans* literature using ontology of 14,500 terms which is based on Gene Ontology. BioIE [25] is an extraction system of the information covering biological interactions. It annotates the results using the terms of Gene Ontology. GOPubMed [26] applies Gene Ontology to perform meaningful searches through PubMed database and classifies the results into categories and subcategories as determined by the GO structure. GenIE [27] extracts information about biochemical pathways, sequences, structures and functions of genomes and proteins. It uses both simple processing techniques as well as syntactic and semantic analysis based on a domain ontology. GENIA [28] corpus covers information about biological reactions concerning transcription factors in human blood cells. All the MEDLINE abstracts on this topic and their titles have been marked-up for biologically meaningful terms, and these terms have been semantically annotated using the GENIA ontology. GENIA ontology is a taxonomy of 47 biologically relevant nominal categories.

UMLS have been used to support information retrieval and analysis process. MedScan [29] utilizes a specially developed context-free grammar and lexicon designed from Gene Ontology and UMLS. It processes sentences from MEDLINE abstracts and produces a set of logical structures corresponding with the meaning of each sentence. Medical World Search Engine [30] uses the UMLS as its built-in knowledge of medical terminology and a selected set of medical resources to perform its searches. BioAnnotator [31] uses domain-based dictionaries (UMLS, LocusLink and GeneAlias) for recognizing known terms within the given text. Due to

the incomplete nature of the used dictionaries, a rule engine has been designed for discovering new terms.

Multi-agent systems are being increasingly used within the biomedical domain. One way of classifying agent-based systems is non-mobile multi-agent system (e.g. Agent Cities [32] and AACare [33]) and mobile multi-agent systems (e.g. BioAgent [34] and Holonic Medical Diagnostic System [35]). The multi-agent systems differ from each other in the purpose they were designed for. Some of them are designed for a specific hospital, its physicians and patients and the available medical services. The information available to these systems is limited to a specific institution and these multi-agent systems help the management of the already available information. They do not have the purpose of gaining new knowledge regarding the disease in question. For example, Agent Cities is a multi-agent system composed of agents that provide medical services. The multi-agent system contains agents that allow the user to search for medical centres satisfying a given set of requirements, to access his/her medical record or to make a booking to be visited by a particular kind of doctor. AACare is an agent architecture which provides a decision support system for physicians. It connects patient's records with the domain knowledge such as knowledge regarding a specific disease, a knowledge base of clinical management plans, a database of patient records etc. The other two multi-agent systems have specific tasks but the information that needs to be retrieved is not limited to a specific hospital or institution but is retrieved from the Internet. BioAgent is a mobile agent system where an agent is associated to the given task and it travels among multiple locations and at each location performs its mission. At the end of the trip, an information integration procedure takes place before the answer is deployed to the user. Holonic Medical Diagnostic System architecture is a medical diagnostic system that combines the advantages of the holonic paradigm, multi-agent system technology and swarm intelligence in order to realize Internet-based diagnostic systems for diseases. All necessary/available medical information about a patient is kept in exactly one comprehensive computer readable patient record called computer readable patient pattern (CRPP) and is processed by the agents of the holarchy.

10. Our Combined Multi-Agent/Ontology Approach

Ontologies have the potential to support management, searching, interpretation, categorisation and indexing of the information in disparate medical databases. We have designed Generic Human Disease Ontology (GHDO) [36] to have the following four branches or subontologies:

- *disease types*, describing different types of a disease;
- *phenotype*, describing disease symptoms;
- *causes* responsible for that disease that can be genetic, environmental and/or microorganism;
- *treatments*, providing an overview of all treatments possible for a particular disease;

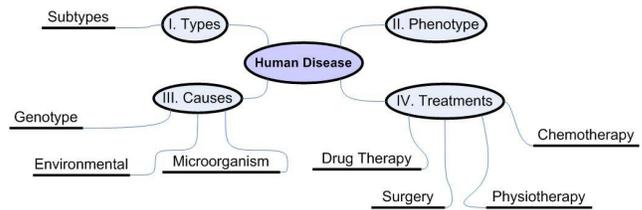


Figure 8 Top level hierarchy of Generic Human Disease Ontology

The information presented in the Figure 8 states that a disease may have different types that also further may be divided into subtypes. For each disease, there is a corresponding phenotype (or observable characteristics of an ill individual), namely symptoms of a disease. Each disease is caused by cause(s) which can be genetic (genotype), environmental or a disease may be caused by a microorganism. Genetic causes can be a mutated gene, a complex of genes or DNA region of interest. DNA region of interest is a region in the DNA sequence that potentially contains a gene responsible for the disease. This region needs to be further examined in order to correctly locate the mutated gene. Environmental causes of a disease can be stress, climate, drugs, family conditions etc. Microorganisms that may cause a disease may be virus or bacteria. Possible treatments for a disease can be drug therapy, chemotherapy, surgery, or physiotherapy.

A multi-agent system can be designed that uses the ontology for the purpose of retrieval and analysis of the relevant information. This multi-agent system can be designed to use ontologies to run most of its functions, such as information location and retrieval, agent's communication, information analysis and manipulation, information presentation etc [37].

The organization of the different types of agents within this multi-agent system is presented in Figure 9. This is a distributed multi-agent system with the agents as its processing nodes. All the agents within this information system are dependent on each other with the respect to the same goal. To be able to achieve this goal, they are performing different tasks and cooperatively working on different levels within this multi-agent system.

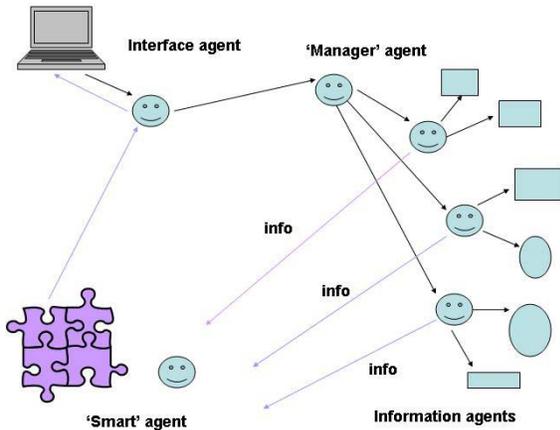


Figure 9 Interface, ‘Manager’, Information and ‘Smart’ agent

In our multi-agent system architecture, we distinguish following agent types [38]:

- *Interface* agents to assist the user in forming queries as well as to present the retrieved and assembled information back to the user. Interface agents communicate user’s request to the ‘Manager’ agents.
- ‘*Manager*’ agents decompose the overall task into smaller tasks and assign these subtasks to the various Information agents.
- *Information* agents retrieve the requested information from a wide range of biomedical databases. Each information agent may have a set of databases assigned to it. The information agents send the retrieved information to the ‘Smart’ agents.
- ‘*Smart*’ agents analyze this information, assemble it correctly and send to the Interface agent directing it back to the user as an answer to his/her query.

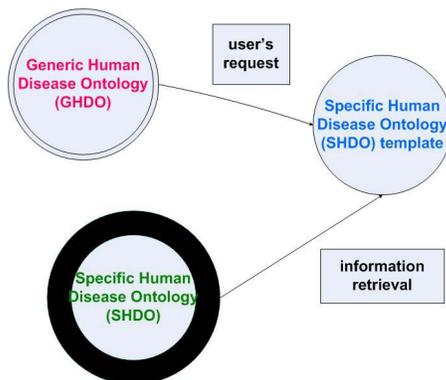


Figure 10 GHDO, SHDO template and SHDO

Interface agents use the GHDO to construct the Specific Human Disease Ontology (SHDO) template (see Figure 10). SHDO template is a formalized version of the overall problem to be solved by the multi-agent system as specified by the user. Manager agent uses the SHDO template and assigns corresponding tasks to Information agents. Information agents retrieve the relevant information and pass it to the Smart agent. The Smart agent analyses, selects and compiles all the retrieved information into the SHDO template. This step results in Specific Human Disease Ontology (SHDO) which is presented to the user as answer to her/his query.

Recently, we have considered using of data mining technologies in combination with this systems and this is explained in [39,40].

In the biomedical domain, we have developed:

- 1) Human Disease Ontology [36]
- 2) Mental Health Ontology [41]
- 3) Protein Ontology (<http://proteinontology.info/>) [42]
- 4) Privacy Ontology for Health Information [43]
- 5) Trust and Reputation Ontologies [44]

11. Conclusion and future work

This paper can bring a better understanding of the unique role, use and significance of ontologies within the biomedical domain. We have discussed ontology properties, community and commitments including generic and specialized ontologies, explained characteristics and representation of ontology models and highlighted the numerous roles of ontologies within biomedical systems. We have provided an example to illustrate the use of the ontologies within multi-agent systems in intelligent retrieval of human disease information. Implementation information will be given in further publications.

References

- [1] M. Hadzic, E. Chang, "Web Semantics for Intelligent and Dynamic Information Retrieval Illustrated Within the Mental Health Domain," *Advances in Web Semantics*, Springer, 2007.
- [2] J.J. Cimino, "In defense of the desiderata", *Journal of Biomedical Informatics* 2006, vol.39, pp.299-306.
- [3] Smith, "From concepts to clinical reality: An essay on the benchmarking of biomedical terminologies", *Journal of Biomedical Informatics* 2006, vol.39, pp.288-298.
- [4] W. Ceusters, P. Martens, C. Dhaen, B. Terzic, "LinkFactory: an advanced formal ontology management System", *Proceedings of Interactive Tools for Knowledge Capture (KCAP 2001)*.
- [5] Burgun, "Desiderata for domain reference ontologies in biomedicine", *Journal of Biomedical Informatics* 2006, vol.39, pp.307-313.
- [6] Gómez-Pérez, "Towards a framework to verify knowledge sharing technology", *Expert Systems with Applications* 1996, vol.11, no.4, pp.519-529.
- [7] Gómez-Pérez, "Knowledge sharing and reuse", *The Handbook on Applied Expert Systems*, CRC Press, 1998, pp. 1-36.

- [8] T. Gruber, "Towards Principles for the Design of Ontologies Used for Knowledge Sharing", *International Journal of Human and Computer Studies* 1995, vol.43, no.5-6, pp.907-928.
- [9] T. Gruber, "A Translation Approach to Portable Ontology Specifications", *Knowledge Acquisition* 1993, vol.5, no.2, pp.199-220.
- [10] Y. Labrou, T. Finin, Y. Peng, "Agent Communication Languages: The Current Landscape", *IEEE Intelligent Systems* 1999, vol.14, pp.45-52.
- [11] T. Finin, Y. Labrou, J. Mayfield, "KQML as an agent communication language", *Proceedings of the third international Conference on Information and Knowledge Management (CIKM 1994)*, pp.456-463.
- [12] Van Aart, R. Pels, G. Caire, F. Bergenti, "Creating and Using Ontologies in Agent Communication", *Proceedings of the Workshop on Ontologies in Agent Systems (OAS 2002) at the first International Joint Conference on Autonomous Agents and Multi-agent systems (AAMAS 2002)*.
- [13] Stuckenschmidt, "Exploiting Partially Shared Ontologies for Multi-Agent Communication", *Proceedings of the sixth international workshop on Cooperative Information Agents (CIA 2002)*, pp. 249-263.
- [14] A. Farquhar, R. Fikes, J. Rice, "The Ontolingua Server: A tool for collaborative ontology construction", *International Journal of Human-Computer Studies* 1997, vol.46, no.6, pp.707-727.
- [15] P.R.S. Visser, D.M. Jones, T.J.M. Bench-Capon, M.J.R. Shave, "Assessing heterogeneity by classifying ontology mismatches", *Proceedings of Formal Ontology in Information Systems (FOIS 1998)*, pp.148-162.
- [16] O. Bodenreider, "The Unified Medical Language System (UMLS): integrating Biomedical terminology", *Nucleic Acids Research* 2004, vol. 32, no. 1, pp. 267-270.
- [17] A. Doms, M. Schroeder, "GOPubMed: exploring PubMed with the Gene Ontology", *Nucleic Acid Research* 2005, vol. 33, pp. W783-786.
- [18] I. Horrocks, P.F. Patel-Schneider, F. Van Harmelen, "Reviewing the Design of DAML+OIL: an Ontology Language for the Semantic Web", *Proceedings of the Eighteenth National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence (AAAI/IAAI 2002)*, pp.792-797.
- [19] B. Chandrasekaran, J. Josephson, V. Benjamins, "What Are Ontologies, and Why Do We Need Them?", *IEEE Intelligent Systems* 1999, vol.14, no.1, pp.20-26.
- [20] R.W. Langacker, *Foundations of Cognitive Grammar: Theoretical Prerequisites*, Stanford University Press, 1987.
- [21] N.B. Szirbik, C. Pelletier, T. Chausalet, "Six methodological steps to build medical data warehouses for research", *International Journal of Medical Informatics* 2006, vol.75, pp.683-691.
- [22] W. Ceusters, P. Martens, C. Dhaen, B. Terzic, "LinkFactory: an advanced formal ontology management System", *Proceedings of interactive tools for Knowledge Capture (KCAP 2001)*.
- [23] O. Bodenreider, "The Unified Medical Language System (UMLS): integrating Biomedical terminology", *Nucleic Acids Research* 2004, vol. 32, no. 1, pp. 267-270.
- [24] H. Muller, E. Kenny, P. Sternberg, "Textpresso: An ontology-based information retrieval and extraction system for biological literature", *PLoS Biology* 2004, vol. 2, no. 11, p.e309.
- [25] J. Kim, J. Park, "BioIE: Retargetable information extraction and ontological annotation of biological interactions from the literature", *Journal of Bioinformatics and Computational Biology* 2004, vol. 2, no. 3, pp. 551-568.
- [26] A. Doms, M. Schroeder, "GOPubMed: exploring PubMed with the Gene Ontology", *Nucleic Acid Research* 2005, vol. 33, pp. W783-786.
- [27] P. Cimiano, U. Reyle, J. Saric, "Ontology-driven discourse analysis for information extraction", *Data and Knowledge Engineering* 2005, vol. 55, pp. 59-83.
- [28] J. Kim, T. Ohta, Y. Tateisi, J. Tsujii, "Genia corpus – semantically annotated corpus for bio-textmining", *Bioinformatics* 2003, vol. 19, pp. i180–182.
- [29] S. Novichkova, S. Egorov, N. Daraselia, "Medscan, a natural language processing engine for Medline abstracts", *Bioinformatics* 2003, vol. 19, no. 13, pp. 1699-1706.
- [30] H. Suarez, X. Hao, I. Chang, "Searching for information on the internet using the UMLS and medical world search", *Proceedings of the Annual AMIA Fall Symposium* 1997, America, pp. 824-828.
- [31] S. Mukherjea, L. Subramaniam, G. Chanda, S. Sankararaman, R. Kothari, V. Batra, D. Bhardwaj, B. Srivastava, "Enhancing a biomedical information extraction system with dictionary mining and context disambiguation", *IBM Journal of Research and Development* 2004, vol. 48, no. 5/6, pp. 693-701.
- [32] A. Moreno, D. Isern, "A first step towards providing health-care agent-based services to mobile users", *Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS'02)* 2002, pp. 589-590.
- [33] J. Huang, N. R. Jennings, J. Fox, "An Agent-based Approach to Health Care Management", *International Journal of Applied Artificial Intelligence* 1995, vol. 9, no. 4, pp. 401-420.
- [34] E. Merelli, R. Culmone, L. Mariani, "BioAgent: A Mobile Agent System for Bioscientists", *Proceedings of the Network Tools and Applications in Biology Workshop Agents in Bioinformatics (NETTAB02)* 2002.
- [35] M. Ulieru, "Internet-Enabled Soft Computing Holarchies for e-Health Applications", *New Directions in Enhancing the Power of the Internet* 2003, Springer-Verlag Heidelberg, pp. 131-166.
- [36] M. Hadzic, E. Chang, "Ontology-based Support for Human Disease Study", *Proceedings of the Hawaii International Conference on System Sciences (HICSS38 2005)*, pp.143a.
- [37] M. Hadzic, T. Dillon, P. Wongthongtham, E. Chang, *Ontology-based Multi-agent Systems* (in final draft), Springer, 2008.
- [38] M. Hadzic, E. Chang, "Ontology-based multi-agent systems support human disease study and control", *Frontiers in Artificial Intelligence and Applications* (special issues on Self-organization and Autonomic Informatics), IOS Press, 2005, pp.129-141.
- [39] M. Hadzic, R.A. Cowan, "Three Fold System (3FS) for Mental Health Domain", *Proceedings of the International IFIP Workshop On Semantic Web & Web Semantics (SWWS 2007)* in conjunction with On The Move Federated Conferences (OTM 2007).
- [40] M. Hadzic, F. Hadzic, T. Dillon, "Mining of Health Information from Ontologies", to appear in the *Proceedings of the International Conference on Health Informatics (HEALTHINF 2008)*.
- [41] M. Hadzic, R. D'Souza, F. Hadzic, T. Dillon, "Synergy of Ontology and Data Mining: Increasing Value of the Mental Health Information within PubMed database", *Proceedings of the Second IEEE International Digital Ecosystems and Technology Conference* 2008, Thailand, pp. 600-603.
- [42] A. S. Sidhu, T. S. Dillon, E. Chang, "Integration of Protein Data Sources through PO", *Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA 2006)* 2006, Poland, pp. 519-527.
- [43] M. Hecker, T.S. Dillon, "Ontological Privacy Support for the Medical Domain", *Proceedings of the National e-Health Privacy and Security and Symposium* 2006, Australia.
- [44] E. Chang, T. S. Dillon, F. Hussain, *Trust and Reputation for Service-Oriented Environments: Technologies For Building Business Intelligence And Consumer Confidence*, John Wiley & Sons, 2006.