USE of ONTOLOGY-BASED MULTI-AGENT SYSTEMS in the BIOMEDICAL DOMAIN

Maja Hadzic, Elizabeth Chang
Curtin University of Technology
School of Information Systems
Perth, Western Australia, 6845, Australia
E-mail: hadzicm@cbs.curtin.edu.au, change@cbs.curtin.edu.au

Abstract

Coordination, cooperation and exchange of information is important to the medical community. We design a new ontology, called Generic Human Disease Ontology (GHDO), by merging and aligning existing medical ontologies. The concepts of the GHDO are organized into the following four dimensions: Types, Symptoms, Causes and Treatments of human diseases.

We also design a multi-agent system framework over different information resources. The multi-agent system uses the common GHDO ontology for query formulation, information retrieval and information integration.

We conclude that this intelligent dynamic system provides opportunities to collect information from multiple information resources, to share data efficiently and to integrate and manage scientific results in a timely manner.

1 INTRODUCTION

In order to make all available data really useful, one needs tools that will access and retrieve exactly the information one needs. The available online information needs to be intelligently queried. We have chosen the application domain of human disease research and control to investigate. A characteristic of the domain is that trusted databases exist but their schemas are often poorly documented, if at all. The management of resources and services is important in the biomedical community and associated smaller communities of people committed to a common goal. The network of biomedical databases forms a loose federation of autonomous, distributed, heterogeneous data repositories ready for information integration.

Our research is centered on the task of formalizing and combining the knowledge regarding human diseases into a single coherent unifying framework. We aim to develop a methodology to access, extract and manipulate information from various information resources. Ontologies may be seen as shared formal conceptualization of domain knowledge. Therefore, ontologies constitute an essential resource for enabling interoperation in an open environment such as the internet.

The multi-agent system makes use of this ontology for the purpose of intelligent and dynamical information retrieval. Within the multi-agent system, ontology is used at the different levels:
Firstly, ontology is used to **locate and retrieve** requested information. Information content within an information resource can be described using an ontology. Only then, an agent committed to this ontology is able to "understand" the information contained within these resources and is able to exactly locate and retrieve the requested information.
Secondly, ontology is used to enable cooperatively working **agents to communicate** with each other during the process of the information retrieval. Use of ontology permits coherent communication and facilitates sharing of the information among different agents.
Thirdly, ontology is used to **analyze and manipulate** the retrieved information. In this way, the redundant and/or inconsistent information is removed. Only relevant information is selected, assembled and presented to the user.
Fourthly, ontology is used to **present** the retrieved information to the user in a meaningful way. The retrieved information is presented to the user in a way that makes it easier for the researcher, physician or patient to have an overview of the requested knowledge regarding human disease of interest. Moreover, the inherited organisation of ontologies adds a taxonomical context to search results, making it easier for the researcher to spot conceptual relationships in data.

In this paper, we introduce an ontology-based multi-agent model for the information retrieval and representation of biomedical knowledge related to human diseases. The ontology is realized in multi-agent system designed to aid medical researchers, physicians and patients in retrieving relevant information regarding human diseases. But we believe that the way we approach our problem is applicable to other knowledge domains as well.

This paper is structured as following. In Section 2, we discuss related work in the biomedical domain. We describe how we designed Generic Human Disease

Ontology in Section 3. In Section 4, we describe four different types of agents and four different phases in the process of problem solving within the multi-agent system. Finally, in Section 5, we conclude and provide our final remarks.

2 BASELINE AND RELATED WORK

Agent Cities [1] is a multi-agent system (MAS) composed of agents that provide medical services. The MAS contains agents that allow the user to search for medical centers 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. Each agent platform supports agents that offer services similar to those that can be found in a real city (facilities, amenities, information and commercial services).

AADCare agent architecture [2] comprises multiple layers of knowledge, a working memory, a communications manager and a human-computer interface. The three layers of knowledge which form the key part of the AADCare architecture are domain knowledge (a knowledge base covering specific medical domains), inference knowledge (generic, declarative inference rules which specify inference relations between domain knowledge, existing patient information and possible new data) and control knowledge (applies the inference knowledge to the domain knowledge in order to generate new inferences whenever new data is added to the working memory). The agents themselves are implemented using the layered architecture which combines a number of AI and agent techniques.

BioAgent [3] is a mobile agent system suitable to support bioscientists during the process of genome analysis and annotation. An agent is associated to a given task and it travels among multiple locations (called places) 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 (HMDS) is a medical diagnostic system [4. This system combines the advantages of the holonic paradigm, multi-agent system technology and swarm intelligence in order to realize Internet-based diagnostic system for diseases. Each agent has a certain responsibility. Some agents may represent experts on a broader field of diseases while others may be experts on (occurrences of) one specific disease. Like ants in an ant colony, the proposed agents collaborate in order to provide a reliable medical diagnosis.

Agent Cities and AADCare are designed as multi-agent architecture. BioAgent and HMDS are systems that make use of mobile agents. The use of agents within the medical community is quite encouraging. Yet, none of the above mentioned agent architectures make use of the ontologies. As rich form of domain knowledge representation, ontologies were in the first place brought into the computer and information society for the purpose of being used by the agents. We propose an ontology-based multi-agent system in which the ontologies are used for the purpose of intelligent information retrieval.

3 HUMAN DISEASE ONTOLOGY

We identified two potentially user categories of the system:

1. medical researches that are mainly interested in *causes* of a disease, and
2. physicians and patients that are faced with a situation of a disease and are mainly interested in *symptoms* and *treatments* of a disease.

Having this two users categories in mind, we construct Generic Human Disease Ontology (GHDO) [5]. The GHDO has four main branches:

1. disease types, describing different types of a disease;
2. phenotype, describing symptoms of a disease;
3. causes responsible for a specific disease which can be genetic and/or environmental;
4. treatments, giving an overview of all treatments possible for a particular disease;

Top-level hierarchy of the GHDO is illustrated in the Figure 1.

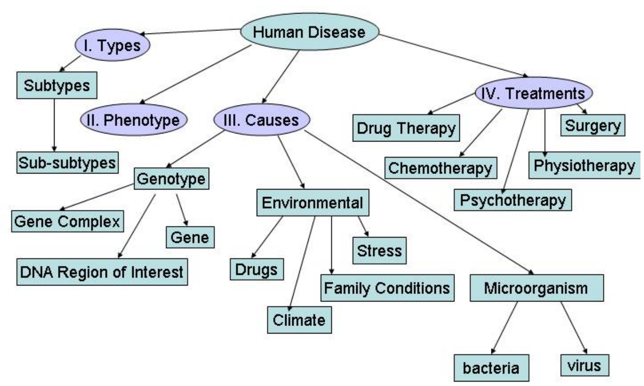

Figure 1. Top-level hierarchy of the GHDO

The information presented in this figure state that a disease may have different types that also further may be divided into subtypes and sub-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 or family conditions. Microorganisms that may cause a disease may be virus or bacteria. Possible treatments for a disease can be drug therapy, chemotherapy, surgery, psychotherapy or physiotherapy.

The four different branches (sub-ontologies) of the GHDO ontology can serve as a reference point against which the concepts from the existing medical ontologies can be reorganized, aligned and merged. Researchers in the medical ontology-design field have developed different terminologies and ontologies in many different areas of medical domain. In order to obtain some uniformity across different ontologies, definitions from other published and consensual ontologies can be reused [6]. Lots of applications already use the existing terminologies like UMLS [7] and LinkBase [8]. Rather then creating a new terminology, we decided to use the concepts from the existing medical ontologies. The way that these concepts are organized within the existing ontologies is not suitable to be used by our system. So, we use terminology from the existing ontologies but organize the concepts in a way that can be used for our purpose and by our system.

4 MULTI-AGENT SYSTEM AND PROBLEM SOLVING PROCESS

In Figure 2, we show different types of agents used in our multi-agent system:

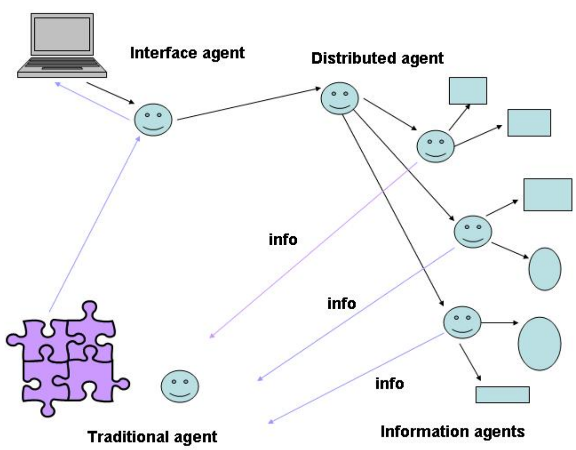

Figure 2. Interface, distributed, information and traditional agents

- *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 distributed agents.

- *distributed* 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 traditional agents.

- *traditional* 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.

The four different phases in the process of problem solving within our multi-agent system are query specification, problem decomposition, atomic problems solution, result sharing and analysis, and solution synthesis.

Query specification. A user may only be interested in a part of information presented by GHDO. Accordingly to user's request, the overall problem to be solved by the multi-agent system is constructed as Specific Human Disease (SHDO) template from Generic Human Disaese Ontology (GHDO) by *interface* agents. For example, a user is interested in symptoms (phenotype) and causes of some disease. The query is structured as SHDO template by interface agent. This template is composed of two sub-ontologies: Symptoms and Causes subontologies.

Problem decomposition. The SHDO template is decomposed into smaller subproblems by *distributed* agents. This kind of decomposition is hierarchical so that subproblems are further decomposed into smaller sub-subproblems and so on. The SHDO template is first decomposed into its four subontologies (disease types, symptoms, causes and treatments). These subontologies are further decomposed into smaller sub-subontologies. The goal of the problem decomposition is to reach a stage where the subproblems are of an appropriate granularity so that they may be solved by individual information agents. A task assigned to an information agent can be composed of more atomic actions. The grain size of subproblems is important, and decomposition can continue by information agents until the subproblems represent atomic actions that cannot be decomposed any further. The different levels of decompositions will often represent different levels of problem abstraction. Each of these different levels in the problem solving hierarchy represents the problem at the progressively lower level of abstraction.

The process of problem decomposition and task assignment assumes that the agents must have the appropriate expertise to do this. They must have knowledge of the task structure and must know how the task is put together. For example, distributed agents needs to know which information agents are suitable to perform a particular action so it knows in what way different tasks need to be assigned to different information agents. This is the reason why ontology is used to represent domain knowledge as well as the task structure in our system.

Atomic problems solution. In this stage the subproblems identified during the problem decomposition phase are individually solved by *information* agents. Usually, a task assigned to individual information agents is composed of more atomic actions. The information agents perform atomic actions and migrate from one to another database in order to accomplish their overall task.

Result sharing and analysis. Different information agents share information relevant to their subproblems and are cooperatively exchanging information covering different areas of the originally defined SHDO template. The solution is developed progressively. The final result progresses from the solution to small problems that are gradually refined into larger more abstract solutions.

Traditional agent compares and analyzes information coming from different information resources. As we see in the Figure 3, information regarding "DNA region of interest" is coming from three different information agents. The relevant information needs to be selected by the distributed agent and in the next stage, incorporated into SHDO template. In this case, "DNA region of interest" contains information about regions in human DNA which may potentially contain a gene that may be responsible for the development of requested disease if mutation (abnormal change of gene structure) of this gene occurs. The information agents may provide for example, following information for the case of manic-depression [9]. A part of this information is presented in the Table 1. The numbers represent chromosomes in human DNA that may contain the gene of interest (2, 10, 12, 17 and X chromosome) followed by the precise region of this chromosome where this gene is positioned (p13-16, q21-24, q23-24, q11-12, q24-25 etc.).

A1	2, p13-19	**10,q21-24**	17,q11-14	**17,q11-12**	X, q24-27
A2	10,q21-26	10,q21-25	**12,q23-24**	17,q11-13	**X, q24-25**
A3	2, p13-17	**2, p13-16**	**12,q23-24**	12,q23-26	17,q11-13

Table 1. Information retrieved by different information agents regarding DNA region of interest.

The traditional agent compares this information on two levels. Firstly, it assembles all information together such as information regarding chromosomes 2, 10, 12, 17 and X respectively. For each of the chromosomes, it compares information regarding the chromosome regions. In the Table 1, for example, for the chromosome 17 we have regions: q 11-13 (information provided by Agent 2) and q11-12 and q11-14 (information provided by Agent 1). In this context, a smaller DNA region of chromosome means being closer to the novel gene that, if mutated, causes specific disease. A researcher looking to exactly locate this gene is thus closer to his/her goal. That is reason why smaller regions of chromosomes are selected by traditional agent to be incorporated into the SHDO template. In the example of chromosome 17, region q11-12 would be selected.

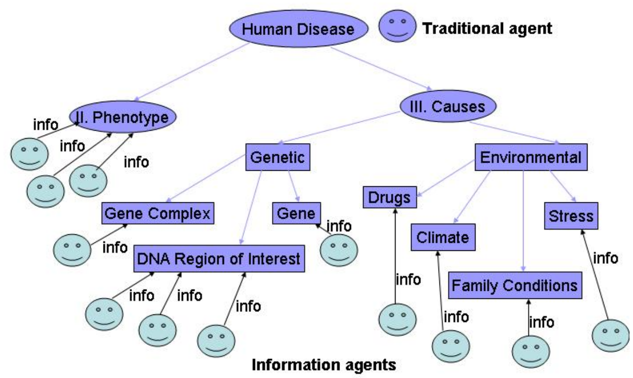

Figure 3. Information retrieved by different information agents is assembled together by traditional agent into SHDO

Solution synthesis. In this stage, the information is assembled together into SHDO template and results in Specific Human Disease Ontology (SHDO) [10]. In the example from Table 1, the following information would be selected and incorporated into SHDO template: chromosome 2, region p13-16; chromosome 10, region q21- 24; chromosome 12, region q23-24; chromosome 17, region q11-12 and chromosome X, region q24-25. In this stage solutions to atomic problems are integrated into an overall solution by *traditional* agents.

As in problem decomposition, this stage may be hierarchical with partial solutions assembled at different levels of abstraction. The use of ontology for representing the domain of knowledge in an organized way is equally important in the solution synthesis stage. The retrieved information is compared, analyzed, assembled together and added on the SHDO template that was constructed at the beginning by the interface agent. After the final reorganization of information within the SHDO, the result is presented to the user as answer to his/her query.

5 CONCLUSION

The system is composed of the two main components: ontology and multi-agent system. Ontologies are high expressive knowledge models and as such increase the system expressiveness and intelligence. We show how the ontologies can be used by multi-agent systems in intelligent information retrieval processes. The ontologies can be used to support some important processes involved in the information retrieval such as posing queries by the user, problem decomposition and task sharing among different agents, result sharing and analysis, information selection and integration, and structured presentation of the assembled information to the user.

The ontology based multi-agent system described in this paper has a number of obvious but quite important advantages:

- it supports the work of scientists in gathering information on highly specific research topics of human disorders, and allows users on a world-wide basis to *intelligently* access new scientific information much more quickly;
- shared knowledge improves research *efficiency* and effectiveness, as it helps (a) to avoid unnecessary redundancy in doing the same experiments, such as the examination of the same region of a DNA sequence, and (b) the determination of for example, which part of DNA sequence needs to be further examined in order to find the gene responsible for a disease;
- it forms a basis of *interoperation*, by allowing distributed but autonomous and heterogeneous resources to function in a world-wide cooperative environment: this makes it possible to split effectively a big task between different research teams;

In our work, we introduce an ontology-based multi-agent model for the information retrieval and representation of biomedical knowledge related to human diseases. The ontology is realized in multi-agent system designed to aid medical researchers, physicians and patients in retrieving relevant information regarding human diseases. But we believe that the way we approach our problem is applicable to other knowledge domains as well.

We have already stared implementing the system in our research centre. However, lots of work still remains, such as security concerns, upload the testbed system on-line for testing and validation, and development of user view interfaces.

REFERENCES

[1] Moreno, A., Isern, D. 2002, '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)*, pp. 589-590.

[2] Huang, J., Jennings, N. R., Fox, J. 1995, 'An Agent-based Approach to Health Care Management', *International Journal of Applied Artificial Intelligence*, vol. 9, no. 4, pp. 401-420.

[3] Merelli, E., Culmone, R., Mariani, L. 2002, 'BioAgent: A Mobile Agent System for Bioscientists', *Proceedings of the Network Tools and Applications in Biology Workshop Agents in Bioinformatics (NETTAB02).*

[4] Ulieru M. 2003, "Internet-Enabled Soft Computing Holarchies for e-Health Applications", *New Directions in Enhancing the Power of the Internet*, Springer-Verlag Heidelberg, pp. 131-166.

[5] Hadzic M., Chang E. 2005, 'Ontology-based Support for Human Disease Study', *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS-38)*.

[6] Noy, N.F., Musen, M.A. 2000, 'PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment', *Proceedings of the seventeenth national conference on Artificial Intelligence (AAAI-2000)*, pp. 450-455.

[7] Bodenreider, O. 2004, 'The Unified Medical language System (UMLS): integrating biomedical terminology', *Nucleic Acids Res*, vol. 32, no. 1, pp. 267-270.

[8] Montyne, F. 2001, 'The importance of formal ontologies: a case study in occupational health', *Proceedings of the international workshop on Open Enterprise Solutions: Systems, Experiences, and Organizations (OES-SEO2001).*

[9] Liu, J., Juo, S.H., Dewan, A., Grunn, A., Tong, X., Brito, M., Park, N., Loth, J.E., Kanyas, K., Lerer, B., Endicott, J., Penchaszadeh, G., Knowles, J.A., Ott, J., Gilliam, T.C., Baron, M. 2003, 'Evidence for a putative bipolar disorder locus on 2p13-16 and other potential loci on 4q31, 7q34, 8q13, 9q31, 10q21-24, 13q32, 14q21 and 17q11-12', *Mol Psychiatry*, vol. 8, no. 3, pp. 333-342.

[10] Hadzic M., Ulieru M., Chang E., "Ontology-Based Holonic Diagnostic System for the Research and Control of New and Unknown Diseases", *Proceedings of the IASTED International Conference on Biomedical Engineering*.