Soft Computing Agents for e-Health
Applied to the Research and Control of Unknown Diseases

Mihaela Ulieru
Canada Research Chair
Director, Emergent Information Systems Lab.
The University of Calgary
T2N 1N4 Alberta, Canada
ulieru@ucalgary.ca

Maja Hadzic, Elizabeth Chang
Curtin University of Technology
School of Information Systems
GPO Box U1987, Perth 6845
Western Australia, Australia
hadzicm, change@cbs.curtin.edu.au

Abstract- This paper presents an Ontology-based Holonic Diagnostic System (OHDS) that combines the advantages of the holonic paradigm with multi-agent system technology and ontology design, for the organization of unstructured biomedical research into structured disease information. We use ontologies as 'brain' for the holonic diagnostic system to enhance its ability to structure information in a meaningful way and share information fast. To integrate dispersed heterogeneous knowledge available on the web we use a fuzzy mechanism ruled by intelligent agents, which automatically structures the information in the adequate ontology template. Our vision of how this system implementation should be backed by a solid security shield that ensures the privacy and safety of medical information concludes the paper.

Keywords- Human disease ontology template, internet-enabled diagnosis, heterogeneous knowledge integration, soft computing agents, secure health information systems, control and treatment of unknown diseases.

1. INTRODUCTION

In today's global world fast and reliable medical diagnosis is of vital importance as can be seen, for example, from the recent problems with SARS or the bird flu. Such highly contagious and lethal diseases can threaten the world if they are not fought immediately and with high efficiency and reliability. However, to do so, it is, first of all, necessary to quickly and surely diagnose the disease regardless of where the case is encountered in the world. While, after a short while, the identification of the disease at its hot spots may become routine, its diagnosis at more remote/unlikely places will remain the challenge. As such, of major importance is the rapid creation of an appropriate knowledge structure easily accessible on the Web, encoding the most up-to-date information regarding the new disease, and capable of easy, continuous updates from the various medical communities working on the disease understanding and relief.

A Holonic Diagnosis System for e-Health applications was proposed by Ulieru [24]. It consists of a medical holarchy, Figure 1, that is a community of people and/or virtual entities (hospitals, clinics, databases, medical devices) committed to a common information-dependent goal (e.g. to contain and control a new epidemic, such as SARS). In virtue of its ability to self-organize [25] the holonic diagnosis system is capable to cluster all the resources to be involved in diagnosis, prediction and progression monitoring of the disease at stake and manages the flow of information and interactions throughout the holarchy according to the particular need to be dealt with [22],[23].

Figure 1: Medical Holarchy

Medical holarchies, Figure 2 can act as a primary response to the needs and requirements of today's healthcare system, especially to the need for unimpeded access to healthcare services and ease of workflow management throughout the medical system. Moreover, backed by a solid search mechanism and a consistent knowledge gathering and representation engine, the system can dynamically retrieve information and create new knowledge to support the continuous discovery of treatments for new diseases [9], or the immediate access to vital information in case of an emergency. During an e-Health rescue operation, novel e-Health technologies can be used, e.g. for patient care authentication by a wireless fingerprint sensor that accesses their profile from a remote database which can be accessed via the e-Health (support) holarchy [24].

Depending on indicators such as blood pressure and the health history of the patient, a first diagnosis is compiled using automated decision support systems [26]. Electronic logistics support provides information about the next available and suitable hospital, initiate staff assembly and emergency room preparation, and provides on-the-fly patient check-in.

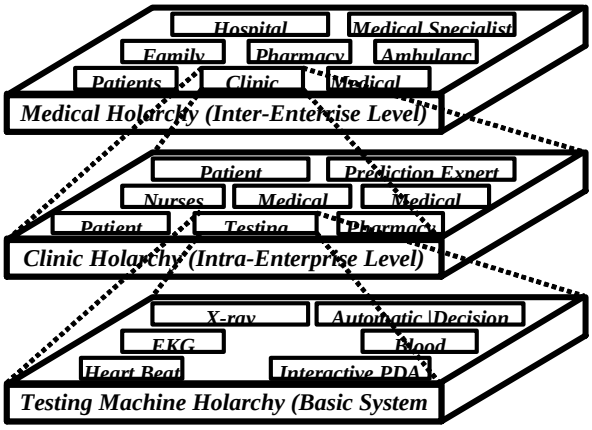

Figure 2: e-Health Holarchy

Planning and scheduling of resources on all levels of the e-Health holarchy enables reconfiguration and flexibility by selecting functional units, assigning their locations, and defining their interconnections (e.g. reallocating hospital beds to cope with the victims, finding the nearest hospital with the appropriate facilities, respectively medical specialists, etc.

With the advent of the Semantic Web [33] the WWW world is evolving from a simple a repository for information, towards a distributed, collaborative, and high-volume computing environment that poses particular new challenges to the efficient and effective design of data and transactions. To make the information more accessible using machine-readable meta-data there have been several research efforts of which *ontology engineering* is a key component. A body of formally represented knowledge is based on conceptualization, namely an abstract, simplified view of the world that we wish to represent for some purpose, usually involving computers. It consists of a set of objects, concepts and other entities about which knowledge is being expressed (often called the *universe of discourse*) and of relationships that hold among them. Every formal knowledge model is committed to some conceptualization, implicitly or explicitly. An explicit specification of this agreed conceptualisation is called ontology [8].

A shared ontology defines a common understanding of specific terms together with their relationships and rules of use, in order to allow communication between systems on a semantic level. Classical techniques and methodologies are largely inadequate because of the inherently autonomous and heterogeneous nature of the information resources, which forces applications to share data, respectively services, often without prior knowledge of their structure respectively functionality. Computer based ontologies may be seen as shared formal conceptualization of domain knowledge and therefore constitute an essential resource for enabling interoperation in an open environment supported by the OHDS on the Web.

The Ontology-based Holonic Diagnostic System (OHDS) [9] sets up on knowledge discovery from ontologies, such as medical issues, health matters, disease factors, DNA etc, and knows who is doing a particular research, what work has been done and which research group has the most up-to-date results, which database on the web is needed, what is in it, what is the value of the information in that database, where it fits into the specific disease knowledge and how to access it, who's work are related to each other or overlapping with each other or complementary to each other etc. It supports searches, translations, categorization, indexing (through ontology and agents), downloads, uploads and correlates disease information to dynamically create knowledge for the diagnosis, control and treatment of new, unknown diseases.

In this paper we build on these previous results in the development of medical holarchies by strengthening the knowledge organization using latest advances in ontology design and development.

2. STATE-OF-THE-ART IN e-HEALTH ONTOLOGY DEVELOPMENT

The development, dissemination and utilization of common communication standards, vocabularies and ontologies [18] for health care is a very hot research topic, given the proliferation of e-Health technologies. There are several consortia in which IT specialists join forces with medical experts to develop such standards. The EU's CEN/TC 251 aims is to achieve compatibility and interoperability between independent systems, to support clinical and administrative procedures, technical methods to support interoperable systems as well as requirements regarding safety, security and quality. The US standardization bodies, the American Society for Testing and Materials' Committee on Healthcare Informatics (ASTM E31) and Health Level Seven (HL7) are involved in similar work. ASTM E31 is developing standards related to the architecture, content, storage, security, confidentiality, functionality, and communication of information while HL7 is mainly concerned with protocol specifications for application level communications among health data acquisition, processing, and handling systems.

Bioinformatics and health care informatics are fields that already have active communities developing ontologies, yet the application of such ontologies as GALEN [30], Unified Medical Language System (UMLS) [3], Systematized Nomenclature of Human and Veterinary Medicine (SNOMED), has lagged behind their potential, despite the huge drive by health care professionals to bring bioinformatics and health care information into clinical workstations and onto the Internet. The main reason appears to be that these existing ontologies are being developed to meet different needs, each with its own representation of the world, suitable to the purpose it has been developed for. There is as yet no common ontology. Of those that are being developed, GALEN provides a common terminology that is currently of limited scope, while UMLS lacks a strong organizational structure, and SNOMED provides only diagnosis nomenclature and codification [32].

Other ontology based bioinformatics work includes the Riboweb ontology [1], the Gene Ontology (GO) [7], the TAMBIS Ontology (Transparent Access to Multiple Bioinformatics Information Sources) [21], and L&C's LinkBase®.

TAMBIS, uses ontology to enable biologists to ask questions over multiple external databases using a common query interface.

LinKBase® by L&C incorporates recent results involving a very large commercially available formal domain ontology. It is reported [17] to currently contain over 5.000.000 knowledge entities of various types: concepts, relationships, terms etc. These entities represent medicine in a way that can be understood by algorithms. Consistency is maintained through a description-logic based knowledge system called LinKFactory®.

While neither Riboweb, Gene or TAMBIS Ontology deal with human diseases and do not answer disease questions, the LinKBase project has been commercialized and is not available for everyone. The application domain of human disease research and control involves resources of medical, genetic, environmental and treatment data. A characteristic of the domain is that trusted databases exist but their schemas are often poorly or not documented for outsiders, and explicit agreement about their contents is therefore rare.

For this reason, we adopted the ontology design methodology of DOGMA (Developing Ontology-Guided Mediation for Agents) [16].In this approach database schema elements, as well as linguistic elements are represented as lexons combining the knowledge domain. Knowledge about their usage (such as constraints, rules, etc.) is kept rigorously separate and is specified as part of the formal commitment of an application to these lexons. This so-called double articulation permits a high degree of scalability, an essential requirement for agent-based computing. A second fundamental aspect of DOGMA is that it distinguishes data models (which are embedded in specific applications) from proper ontologies (this should be application-independent) [5], [20]. The mapping of a data model to an ontology (in DOGMA) precisely constitutes its formal semantics, in fact reified as part of a commitment.

3. INFORMATION RESOURCES FOR OHDS

Medical researcher teams are heterogeneous. No single institution has all the required resources or skills and team members capable to cover all the health related issues at the global health level (such e.g. new epidemics). Hence the OHDS should enable resources sharing and usage co-ordination in dynamic, virtual, multi-institutional organizations by accessing remote data sources like stored medical and biological information in large quantities. But it would be very time consuming to evaluate the information from each

database one may need, such as where it fits into the whole knowledge world and how one can access it. This is where ontologies are needed as a means to capture and represent in the computer knowledge shared by all people in a certain community. For example, one could want to combine a medical data source in Europe with a biological data source in China in order to perform an analysis. Firstly, we need the OHDS services to provide a dynamic way to use dispersed heterogeneous resources and services in such a large distributed scientific environment. For this we have developed a soft computing methodology that will be presented in subsection 3.1. Secondly, we need a way to describe data and resources in a way that is understandable and usable by the target community. For this we have developed generic templates [10], named generic human disease ontologies (GHDO) from which specific templates, named specific human disease ontologies (SHDO) can be derived by the OHDS framework. The GHDOs and SHDOs will be described in subsection 3.2, the OHDS framework is presented in section 4 while the OHDS mechanism will be detailed in section 5.

3.1. Soft Computing Framework for Heterogeneous, Dispersed Medical Knowledge Integration

As a post internet framework, the OHDS enjoys an unusually large number of high-quality, complex, but extremely heterogeneous information resources, which furthermore are often made available through site-specific services only. For the integration of the heterogeneous knowledge gathered from various sources we have developed a consensus analyzer designed using soft computing technology [26].

The contribution of several knowledge sources to the development of a knowledge base brings enormous value, but at the same time it presents a big challenge to the knowledge engineers. Communication between dispersedly located expert sources has to be supported by an adequate interface, various expert opinions have to be reconciled, eliminating contradictions and choosing the most encompassing solution in each case, privacy and security issues have to be dealt with adequately [14], etc.

To cope with this we propose a methodology capable to integrate disperse heterogeneous sources of knowledge into a unified ontology-based template. The methodology consists of the following steps:

Find the patterns for the disease at stake. This will determine the backbone on which the gathered knowledge will be structured.

Reconcile the differences between various knowledge sources. The differences are investigated using an automatic consensus analyzer capable to determine where these differences occur and reconcile them by embracing both views within a broader, more generic rule/template.

Determine and Test the Core Rule Set. The result of this reconciliation process will be a *core template for human disease ontology* encoding fundamental diagnostic knowledge regarding a particular disease.

Each time a discrepancy between the existing template and a new knowledge source is encountered both the template and the new source are analyzed by the Consensus Analyzer which evaluates the 'distance' between them and the point of minimum consensus (the point of maximum conflict) – where the template and new source clash most. To evaluate this distance we use soft competitive learning [6].

Suppose we are developing consensus in a universe $X = \{x_1, x_2, ...,x_n\}$; a fuzzy relation **R** of order n will have elements r_{ij} encoding the preferences given to x_i relative to x_j. $r_{ij} = 1$ implies that alternative *i* is definitely preferred to alternative *j*. At the other extreme we have maximal fuzziness, where $r_{ij} = r_{ji} = 0.5$. Two common measures of preference are defined here as average fuzziness (F) in R and average certainty (C) in R**:**

$$F(R) = \frac{tr(R^2)}{n(n-1)^{1/2}} \qquad (1)$$

$$C(R) = \frac{tr(R \cdot R^T)}{n(n-1)^{1/2}} \qquad (2)$$

where *tr* is the trace and T is the transposed of the matrix. The measure F(R) averages the joint preferences in **R** over all distinct pairs in the Cartesian space X x X. F(R) is proportional to the fuzziness or uncertainty about pair wise rankings. Conversely the measure C(R) averages the individual dominance of each distinct pair of rankings.

The two measures are related:

$$F(R) + c(R) = 1 \quad (3)$$

Measures of preference can be useful in determining consensus. We define three type of consensus as follows:

Type I consensus: There is a clear choice, say alternative *i* (the ith column is all zeros) and the remaining (n-1) alternatives all have equal secondary preference (i.e 1/2).

Type II consensus: There is one clear choice say alternative *i* but the remaining (n-1) alternatives all have definite secondary preference (i.e 1).

Type Fuzzy consensus: Occurs when there is a unanimous decision for the most preferred choice, say alternative *i* but the remaining (n-1) alternatives have infinitely many fuzzy secondary preferences.

From the degree of preferences measures given in previous equations we can construct a distance to consensus metric defined as

$$m(R) = 1 - (2 \cdot C(R) - 1)^{1/2} \quad (4)$$

Where:

$$m(R) = 1 - (2 / n)^{1/2}$$

for a Type I consensus relation (5)

$$m(R) = 0$$

for a Type II consensus relation (6)

When $n > 2$, the distance between Type I and Type II consensus increases with n as it becomes increasingly difficult to develop a consensus choice and simultaneously rank the remaining pairs of alternatives. The value of distance to consensus quantifies the dynamic evolution of a group as the group refines its preferences and moves closer to a Type I or Type II or Type Fuzzy consensus. The vast majority of group preference situations eventually develop into Type Fuzzy consensus, Types I and II being typically only useful as boundary conditions.

The consensus analyzer works with consensus rooms. Each room represents a topic on which the knowledge needs to be reconciled. As an example, in Figure 3 we present a consensus room designed for the development of a standardized knowledge base for glaucoma progression monitoring/follow-up. Consider that some experts access the system via the web on a secured internet ring. Once the expert has selected the consensus room, the system will provide her/him with the case data, that is, all the input variables used by the intelligent algorithm and its fuzzy rules.

The possible Follow-up alternatives are presented together with the input variables. Those alternatives are extracted from the expert system in order to present to the expert only with the feasible ones.

The lower part of the screen shows the ranking matrix (a fuzzy relation) that the expert should fill in with his/her preferences about the alternatives, as a pairwise comparison. All the opinions are aggregated using fuzzy weighted aggregation operators. The already mentioned two very simple aggregation methods are the max (optimistic combination) and the min (pessimistic combination). We use here the max aggregation operator, that is, the optimistic combination.

A Final Relation **R** is obtained from this step. This fuzzy relation **R** of order n will have elements rij encoding the preferences given to xi relative to xj. $rij = 1$ implies that alternative i is definitely preferred to alternative j.

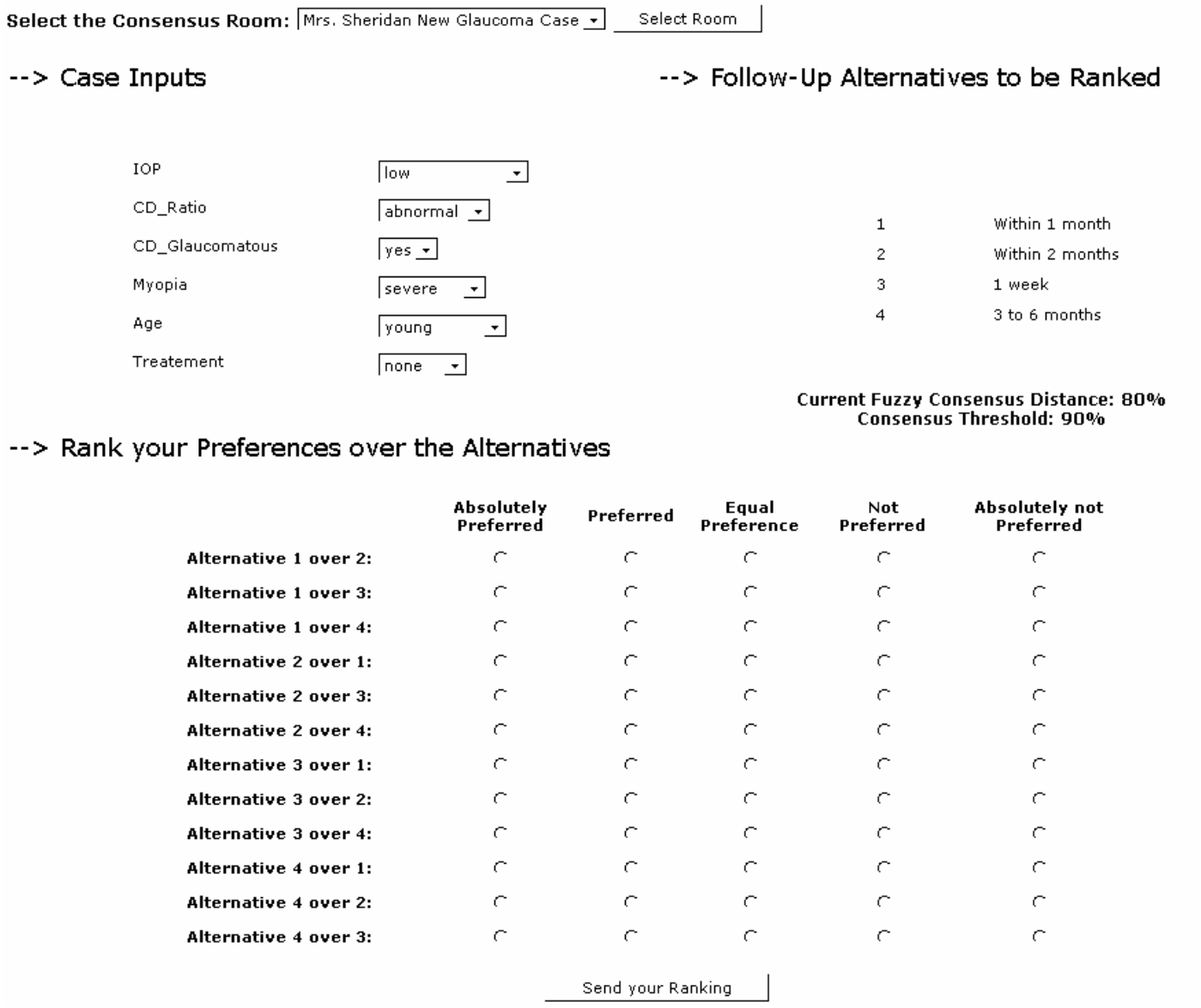

Figure 3: Consensus Room

The value of **distance to consensus** quantifies the dynamic evolution of the group of experts as the group refines its preferences and moves closer to a Fuzzy consensus.

The system has a predefined distance to consensus independent from the case being analyzed. We defined the distance to be 90%. Finally when the distance to consensus predefined is reached, the rule is integrated in the knowledge base. Otherwise, a new round of opinion occurs by providing feedback with the results to each specialist.

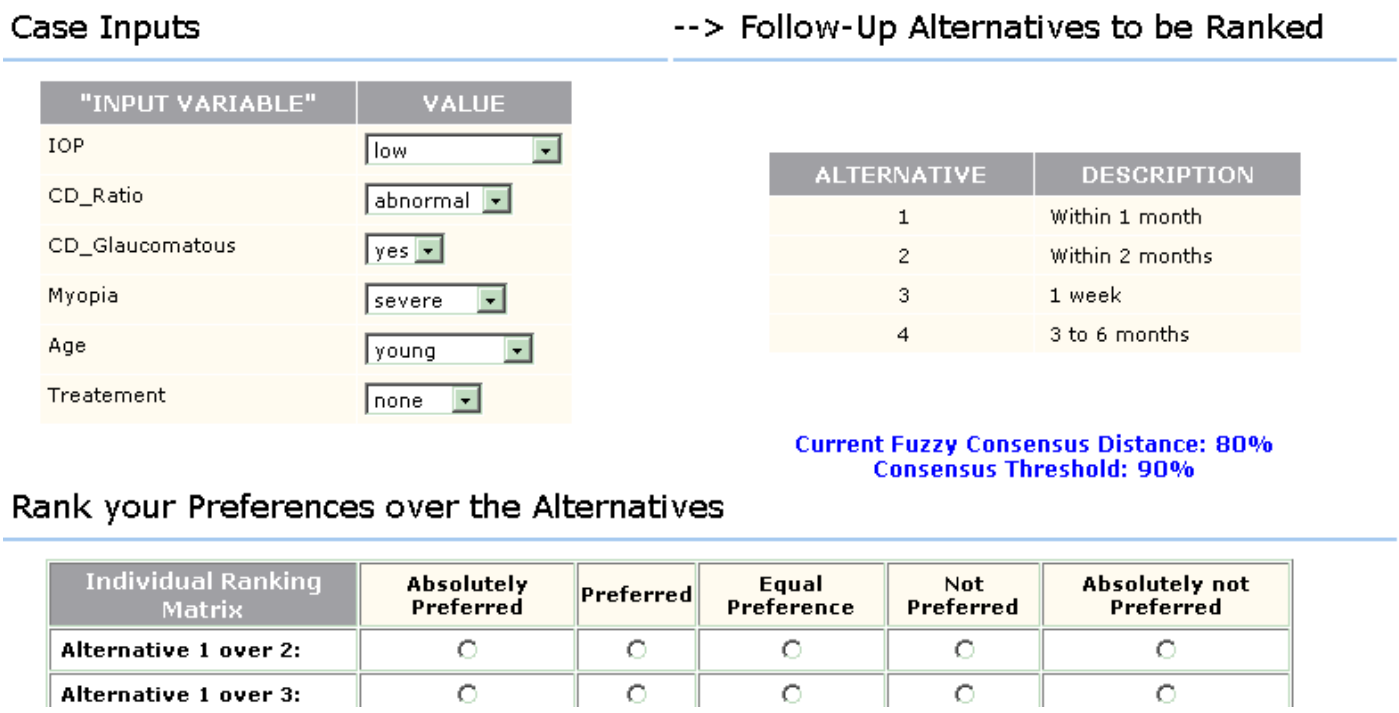

Figure 4: Current Consensus and Threshold Determination

For the Example in Figure 4:

Consensus Room: Mister X

Input Variables: Under _treatement: NO; IOP (Intra Ocular Pressure): low; CD_Ratio: abnormal; Glaucomatous_CD: Yes; Visual Field: Early Loss

Alternatives: A1: Within 1 month; A2: Within 2 months; A3: 1 Week; A4: 3 to 6 Months

Doctor 1 Ranking: A1-A2: Absolutely Preferred; A1-A3: Absolutely Preferred; A1-A4: Absolutely Preferred; A2-A1: Absolutely Not Preferred; A2-A3: Equal Preferred; A2-A4: Equal Preferred; A3-A1: Absolutely Not Preferred; A3-A2: Equal Preferred; A3-A4: Equal Preferred; A4-A1: Absolutely not Preferred; A4-A2: Equal Preferred; A4-A3: Equal Preferred;

Doctor 2 Ranking, The follow-up alternatives comparison results in: A1-A2: Absolutely Preferred; A1-A3: Absolutely Preferred; A1-A4: Absolutely Preferred; A2-A1: Absolutely Not Preferred; A2-A3: Equal Preferred; A2-A4: Equal Preferred; A3-A1: Absolutely Not Preferred; A3-A2: Equal Preferred; A3-A4: Equal Preferred; A4-A1: Absolutely not Preferred; A4-A2: Equal Preferred; A4-A3: Equal Preferred;

The 'Absolutely' preferred option having the max value, we obtain the R fuzzy relation applying the max operator to each opinion, so Alternative 1 is the choice of the two doctors. Then we get the F and C values to 0 and 1 respectively, so finally M will be equal to 1, that means 100% of consensus.

Based on the consensus metrics, the ontology template is tuned to embrace all opinions as much as possible. This means that the rules obtained will be positioned in the equidistant point to all expert opinions. Once the distance to consensus predefined is reached, the rule is integrated in the knowledge base. In case of strong discrepancies between the actual template and the new knowledge source, the generated rule may not make sense as its generality may render it useless. However, this will not affect the system overall, as the rule will be implicitly overlooked by the fuzzy reasoning process.

3.2. Ontologies as Patterns of Medical Knowledge

In our vision ontologies can effectively integrate distributed world wide research in the area of disease by aligning and merging relevant information from publications and medical databases, DNA and protein databases, research institutes, health departments, hospitals etc. As such, the OHDS can provide the required distributed collaborative platform as well as easy access to resources. We designed the Generic Human Disease Ontology (GHDO) as a template with four main branches [10], Figure 6: (1) *types*, describing different types of a disorder; (2) *phenotype*, describing symptoms of a disease; (3) *causes* responsible for that disorder which can be environmental and/or genetic; (4) *treatments*, giving an overview of all treatments possible for that particular disease as well as treatments efficiency. This template helps to produce Specific Human Disease Ontologies (SHDO) as it will be illustrated in section 6. The ontology explains (Figure 6) that a disease may have (1) different *types* which also may be further divided into subtypes etc. Each disease is caused by (3) *cause(s)* which can be genetic (genotype) or environmental. Genetic causes can be a mutated gene, a complex of genes or a region in the DNA sequence that potentially contains a gene responsible for the

disease and needs to be further examined. Environmental causes can be: viruses/bacteria, stress, climate, drugs or family conditions. For each disease, there is (2) corresponding phenotype namely, observable characteristics of an ill individual and (4) *treatments* possible for the disorder that can be drug therapy, chemotherapy, surgery, psychotherapy or physiotherapy.

Another major advantage of using the holonic structure is that it respects complete autonomy of the existing ontology nodes. Each of the existing nodes can withdraw or join the holarchy whenever it is necessary [11]. This is very important when generating on request Specific Human Disease Ontologies as we will show in Section 6.

Figure 5 shows a pictorial representation of the information integration from different sources world-wide. The retrieved information is organized within the Generic Human Disease Ontology and its four different dimensions. The proposed solution enables researchers to analyze the different factors, the relationships between them and different types of diseases simultaneously. After analysis and combination of the information, the result is presented in a way that makes it easier for the user to have an overview of the up-to-date knowledge about a specific disorder.

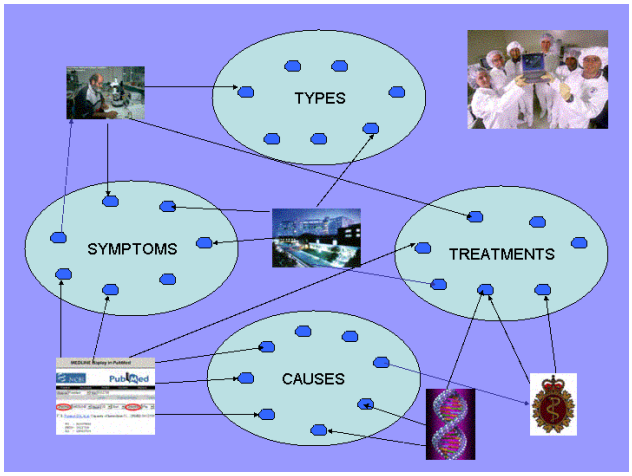

Figure 5: Combining the information from different databases worldwide into the four dimensions of Generic Human Disease Ontology (GHDO) Template Pattern

4. PRINCIPLES OF BUILDING GENERIC HUMAN DISEASE ONTOLOGY (GHDO)

Ontological commitments are formal agreements (expressed in DOGMA as views, rules, and constraints [16]) to use the shared vocabulary in a coherent and consistent manner. Shared vocabulary is different for different knowledge domains. Our knowledge domain is going to have its own vocabulary written in an ontological lexicon. An *ontology base* consists of *lexons*, expressing (usually linguistically derived) *facts* between *terms*. Terms are often organized hierarchically in taxonomy, by promoting the subsumption fact into an implicit, special, and axiomatically defined relationship. Facts in DOGMA are always true only within a *context*, defined for any lexon as carried by an identifiable source, usually a document.

In Figure 6, we show the four main branches of the GHDO. Of course, terms within the GHDO are much more numerous than shown and are validated for existence against concepts from a biomedical lexicon such as e.g. UMLS Metathesaurus [3].

We first illustrate the notions of commitment as a constrained interpretation and of (first order) well-formed formula (wff) through examples. Consider a vocabulary **V = (T, R)** where **T** is a set of terms denoting concepts, and **R** is a set of relationship names.

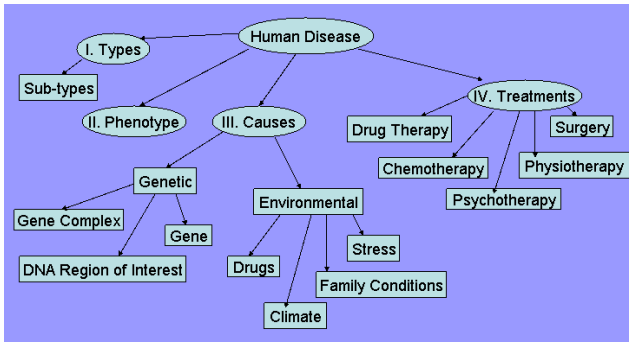

Fig. 6: Generic Human Disease Ontology and its four main subontologies: type, phenotype (symptoms), cause and treatment.

For illustration we will develop a small generic ontology representing the main concepts, identified in a given (implicit) context. Let **T** = {disease, type, subtype, sub-subtype, phenotype, treatment, drug therapy, chemotherapy, physiotherapy, surgery, psychotherapy, cause, genotype, gene, gene complex, DNA region of

interest, environment, stress, climate, family conditions, drugs, micro-organism, bacteria, virus} that represent the lexicon of user's world of diseases, and **R** = {has, isof, isa, is caused by, is responsible for, is cured by, cures, shows, characterizes} that represent relationships (roles) for this domain. Within DOGMA Modeler, the object-relation model (ORM) [12] notation is also used to represent relationships and commitments such as "each disease is caused by at least one cause" and "each disease shows at least one phenotype". The relationships can be represented through the following binary relations, called lexons or facts:

- *has (disease, type); isof (type, disease);*

This means that *"disease has a type"* and *"type is of a disease"*.

- *shows (disease, phenotype); characterizes (phenotype, disease);*

This means that *"disease shows a phenotype"* and *"phenotype characterizes a disease"*.

- *is caused by (disease, cause); is responsible for (cause, disease);*

This means that *"disease is caused by a cause"* and *"cause is responsible for a disease"*.

- *is cured by (disease, treatment); cures (treatment, disease);*

This means that "disease is cured by a treatment" and "treatment cures a disease".

5. HOLARCHIC STRUCTURE AND MECHANISM

In this section, we illustrate how ontologies can be dynamically developed for the knowledge domain of biomedical and bio-engineering research, using the OHDS framework.

In case of knowledge collection, manipulation, organization and discovery for human diseases the proposed OHDS framework can be very useful [9]. The holonic structure (Figure 7) is a nested hierarchy of four holarchies in which each of the four GHDO dimensions template is associated with one holarchy. By sending a request to the Mediator Agent of the OHDS the process is started. From there it infiltrates the hierarchy till it reaches the leaves. The record is interpreted and analyzed at the higher levels of the hierarchy while collection of the data happens at the lower level holarchy.

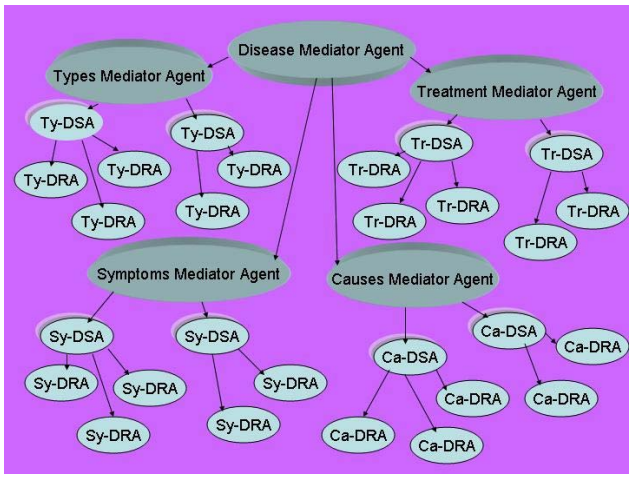

Figure 7: OHDS structure

HOLARCHY MEDIATOR AGENTS (HMA)

Each holarchy has a single entry point, named Mediator Agent. The OHDS has as main entry point the Disease Mediator Agent and in turn each branch has its own mediator agents, respectively Types, Symptoms, Causes and Treatments Mediator Agents. Their task is to decide what other subordinate Disease Specialist Agents - DSAs or Disease Representative Agents – DRAs need to be activated in order to retrieve the information requested by the user. Another task is to integrate the retrieved information coming from DSA via DRA in another direction.

DISEASE MEDIATOR AGENT (DMA)

The DMA interacts with the user and decides which of the four holarchies needs to be deployed in order to generate SHDO requested by user. For example, sometimes a user may be interested only in causes of a disease so that there is no need to deploy Types, Symptoms or Causes holarchy. Also, each of the four holarchies has significant databases assigned to it. Some databases contain information only regarding for example, symptoms of a disease so that, e.g. there is no need for agents from the Cause holarchy to visit those databases. Another task of the DMA is to combine the information coming in another direction from the four holarchies and present it to the user as a single unit.

DISEASE SPECIALISTS AGENTS (DSA)

Holarchy inner nodes represent Disease Specialist Agents (DSAs). They represent decision makers and are specialists on a specific dimension of GHDO. We differentiate Types, Symptoms, Causes and Treatments Diseases Specialists Agents (Ty-DSA, Sy-DSA, Ca-DSA and Tr-DSA). Each DSA will focus on a task which corresponds to its level of knowledge namely, after subordinate agents (DRAs) have returned their data it interprets, compares, and evaluates them in order to define a proper ranking among all the delivered data. The ranking is done by the HMA using two different types of matching as it will be described further. An important task of a DSA is to interpret the incoming data and come to a conclusion on whether there is sufficient evidence for the likelihood of a specific disease. If not, the DSA has to decide - on the basis of the delivered information - whether it makes sense to consult other DRAs or, if this seems to be unpromising, whether to advertise the request on the Internet. This is especially promising if there is suspicion that the disease is a so far unknown or imported one, thus one that is very rare in the living space of the patient/medical unit.

DISEASE REPRESEANTATIVE AGENTS (DRA)

The leaves are so-called Disease Representative Agents (DRAs). We differentiate Types, Symptoms, Causes and Treatments Diseases Representative Agents (Ty-DRA, Sy-DRA, Ca-DRA and Tr-DRA). Each DRA is an expert on a lower level concept within GHDO. Note that DRAs differ from DSAs in that they need to *recognize* the significant information inside the appropriate database and *retrieve* that information. This information is then passed over to the DSA which will do the *analysis* and *comparison* of the retrieved information so that only "new" information will be passed over to the respective mediator agents. For example, article_1 claims that a gene located somewhere on chromosome 6 is responsible for a disease in question, while article_2 gives more precise information regarding the gene of interest such as location 6p11-p17. Ca-DRA retrieves both articles while Ca-DSA passes over only information from article_2 to the CMA. CMA will do the matching and assign the value '6p11-p17' to the concept 'DNA region of interest', telling the user that the DNA sequence positioned on chromosome 6 between p11 and p17 potentially contains a gene which may be causing the specific disease. In this way we keep the presented information updated and

also do the selection of the information before presenting it to the user and present only the key-information. This is especially important when lots of information regarding a specific topic is available.

THE HOLARCHIC MECHANISM

For the information integration process, the Holarchy Mediator Agents perform two different types of matching. First one is matching of the template of GHDO with the incoming information and assigning values to the concepts from GHDO (for example, to assign the name 'GRK3' to the concept 'gene name' from GHDO.) In its decision process on what to do with all the input that may be provided by the lower level agents the Disease Mediator Agent (DMA) not only relies on its knowledge but also on the experiences it made in the past. For this reason, latest version of SHDOs regarding the same disease requested by some other user before, are saved in a pattern store, making it possible to do the second type of matching. If a difference is found, the new SHDO should be checked for its consistency. If the difference is consistent, the latest version should be saved and used next time for matching. The DMA needs to be enriched with sufficient knowledge/intelligence to be able to interpret the incoming information and also to relate it to its knowledge/experience. Moreover, it may be that relevant data/examinations are missing and that more information may be needed and thus lower level agents need to be activated until the process is completed.

The achieved results can have different levels of certainty. In the best case, the information that was provided to the DRA and combined together by Mediator Agents, provides all the data and information that is needed in order to conduct a comprehensive search on the SHDOs as requested by the user. In less fortunate cases the Electronic Health Record (HER) may only provide a part of the optimal set of information and data requested. In such a case where the already available information in the SHDO does not exclude a disease, the result of its analysis comes with a set of tasks, examinations, and tests that are suggested to be performed by the medical institution in order to further verify (or invalidate) the hypothesis.

6. HOW DO THE GHDO AND OHDS WORK TOGETHER

The conceptual framework of our OHDS methodology and prototype is based on the formal theory of ontology described in section 4. The system extracts relevant information from publications and medical

databases, DNA and protein databases, research institutes, health departments, hospitals etc. Upon the analysis and combination of the information, the result is presented in a way that makes it easier for the user to have an overview of the up-to-date knowledge about a specific disorder. Use of ontologies provides us with a more controlled and systematic way to perform information retrieval. Moreover, the holarchic/nested organization of ontologies enables implicit inheritance which adds taxonomical context to search results, making it easier for the researcher to spot conceptual relationships in data. The latter fact is important for instance in the case of complex human disorders where one looks for relationships between different factors that are simultaneously responsible for each of the many types of disorders.

The GHDO links the user to multiple heterogeneous information resources via its four main branches. Each mediator agent in the GHDO template (Figure 7) has a consensus analyzer (recall subsection 3.1.) that supports the OHDS to integrate the dispersed knowledge on each branch of the holarchy into the SHDO relative to the case/disease investigated. In this way, using the GHDO the OHDS can derive SHDOs on request. The SHDOs are specified and generated when a user queries the system.

The source information covers different areas of interest with respect to human diseases in order to allow different user categories (each having specific intentions), to query the system. Researchers are constantly searching for and adding more information to the already existing pool of knowledge regarding a particular disorder. Physicians are directly in contact with patients and are using all significant information to help and treat the patients. Especially when a new disease epidemic starts spreading, researchers and physicians are strongly connected because they are working towards the same goal, but on different knowledge levels.

6.1 Ontology as Support Tool for Physicians

If a medical professional queries the system, she/he will mainly be interested in two of the four components of our system, namely symptoms and possible treatments of a particular disorder. There are some exceptions to this rule, such as in the next use case, when a new disease is encountered by the physician.

Use case 1: Physician cannot identify the disease. A physician may have a patient showing some symptoms of a disease but he may not be able to say what kind of disease it is. At this stage, it is

recommended to keep as many as possible components involved in the search: symptoms (phenotype), causes and treatments. In this case, the derived Specific Ontologies have the “phenotype”, “cause” and “treatment” branches. By entering the symptoms into the system, the doctor may be able to retrieve the information regarding that disease. It is also possible that different diseases are showing the same or similar symptoms, such that the physician retrieves more than one SHDO as we show in Figure 8. In such a case, it may be useful to look for some significance in the causes of the disorders, as we explain in the sequel.

Use case 1_a: causes of the disease are not known. On the basis of the key symptoms the doctor will chose one (set of) disease(s). This disease becomes the doctor's working hypothesis, her/his most likely choice. The doctor then starts to gather evidence in support of the working hypothesis, always keeping in mind the set of alternative hypotheses. Such a process relies on all kinds of information, e.g., information that is gained by interrogating the patient or by conducting necessary (physical or instrument- or tool-based) examinations and tests. It will be assumed that all this data and information will be stored in so-called medical records for patients or *patient records* for short, which follow the GHDO template. It will be assumed that all necessary/available medical information about a patient is kept in exactly one comprehensive computer readable patient record that is a set of SHDOs for the specific conditions of the particular patient. This enables the patient record to be processed by agents because the ontology assigns the unequivocal semantics to the record and, thus, defines how the agent may understand, interpret and process it.

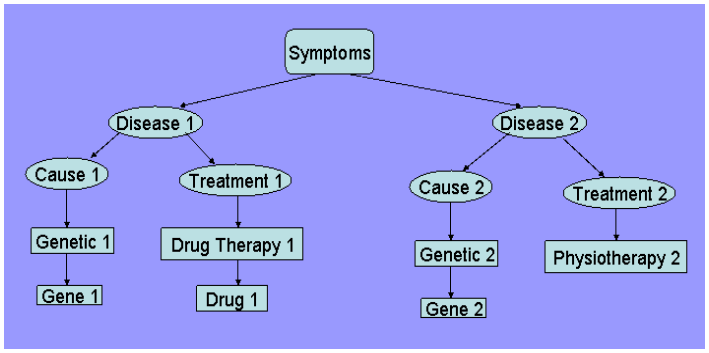

Figure 8: Two different diseases caused by mutations of different genes and treated by different methods showing same symptoms

Use case 1_b: cause of the disease is known, e.g. a gene mutation. For example, in case of disease_1, gene_1 is mutated and thus causes this disorder. And disease_2 is caused by mutation of gene_2. The physician can do the screening of the patients' DNA to check if gene_1 or gene_2 is mutated. If mutation found in gene_1, the patient has disease_1 and if gene_2 mutated the patient suffers from disease_2.

Only when the patient is correctly diagnosed, the physician may consider possible treatments for the patient. Our information system therefore also reduces the risk of misdiagnosis.

Use case 2: Physician can identify the disease and wants to consider possible treatments. It is common that there may be more than one (drug) treatments possible for a particular disease (see Figure 9). A physician will wish to look at all the options possible before choosing one. Choosing medication is also a personal thing because not all people respond in the same way to same medication. At this point a medical professional might for instance consult our ontology-based information system to do a one-component search (treatments). In this case, the derived Specific Ontology has only the "treatment" branch.

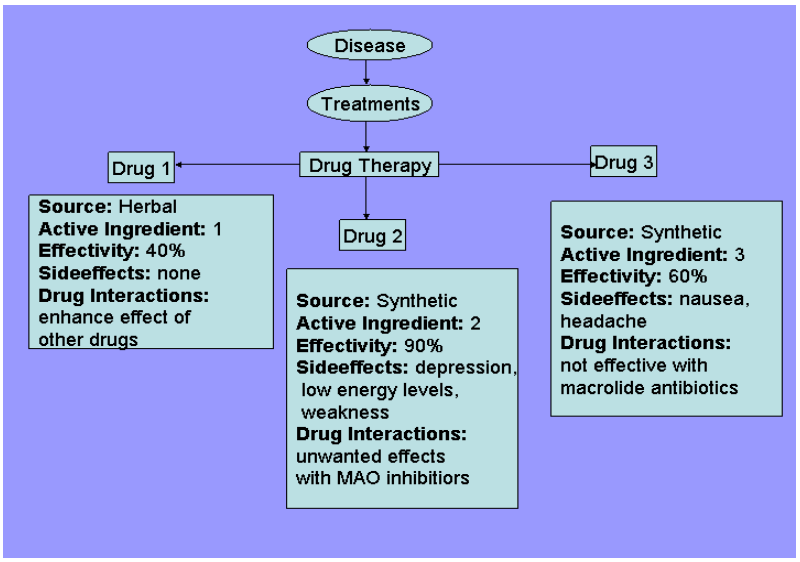

Figure 9: Different drugs target same disease

6.2. Ontology as Support Tool for Research

The biomedical researcher using our system may be interested in one specific of the four possible components of our system. E.g. a researcher working on drug discovery would be more interested in the "treatment" branch. We show another example where the derived SHDO has only the "cause" branch.

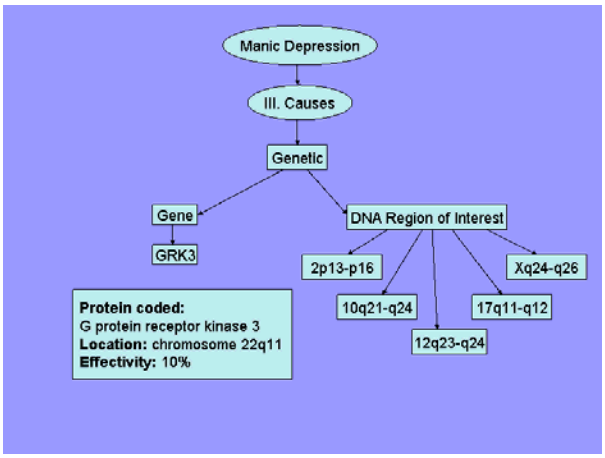

Figure 10: Genetic causes of manic-depression (current research)

Use case 3: Researcher examines possible causes of a disorder. Often not all the causes responsible for a particular disorder are known, e.g. in the case of manic-depression (Figure 10). By querying the OHDS and getting back significant information systematically represented, the researcher is able to identify some regions of interest in the DNA sequence such as regions 2p13-16, 10q21-24, 12q23-24, 17q11-12 and Xq24-26 on chromosomes 2, 10, 12, 17 and X respectively [2], [4], [13], and [15]. Those regions need to be further examined in order to find a gene and a mutation inside that gene.

If a new gene is found on one of the already identified DNA regions of interest, our model will now have four instead of five instances of the term "DNA region of interest" and one more instance of the term "gene" (see Figure 11). Given the length of the DNA sequence it is obviously much easier for a researcher to target a specific area of a chromosome such as 2p13-16 than the whole chromosome 2. Further research, may allow her/him to narrow down the region of interest to, for example 2p14-15. Because of the agreed semantics in a

shared ontology it will be easier for the next person to continue the research in the same direction and possibly to locate the gene of interest. This aspect of cooperation between different teams increases productivity by saving time and research.

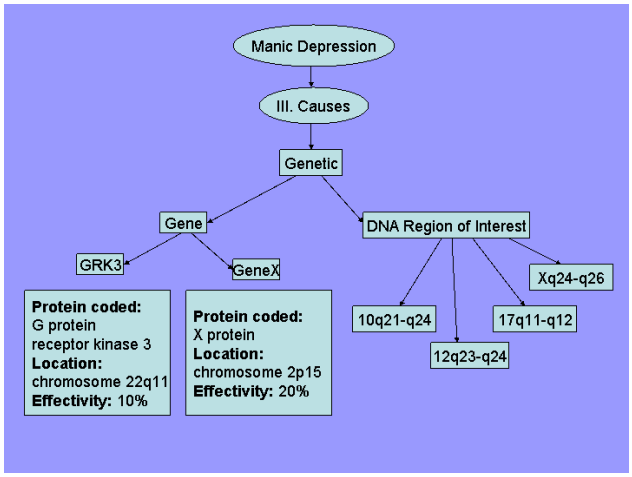

Figure 11: Genetic causes of manic-depression, future research if gene of interest found on chromosome 2

7. IMPLEMENTATION VISION

The proposed OHDS will support the doctors in the diagnostic, treatment and supervision processes of the evolution of new epidemics, based on the exploration of all data pertinent to each case and on the scientific data contained in various professional databases. From an architectural perspective, the OHDS (Figure 12) consists of an *Education&Consultation System* [28] to provide evidence-based guidelines of care to clinicians, and the *Consensus Analyzer* [26] to constantly update and refine these knowledge templates based on as new knowledge sources are parsed. Users will access the OHDS *directly* via a Web-based user interface, or indirectly by using their clinical system. The OHDS is also integrated with other electronic health information infrastructure services, such as patient and provider registries. In addition to structured data, the OHDS system uses high-resolution diagnostic imaging supported by various networking infrastructures. The medical specialist will interact in real-time with the various data collected, unified, and explored by the OHDS agents.

With respect to the reconciliation of standards of care we are working on the creation of an R-MIM for the OHDS based on the HL7 e-Health communication standard [29].

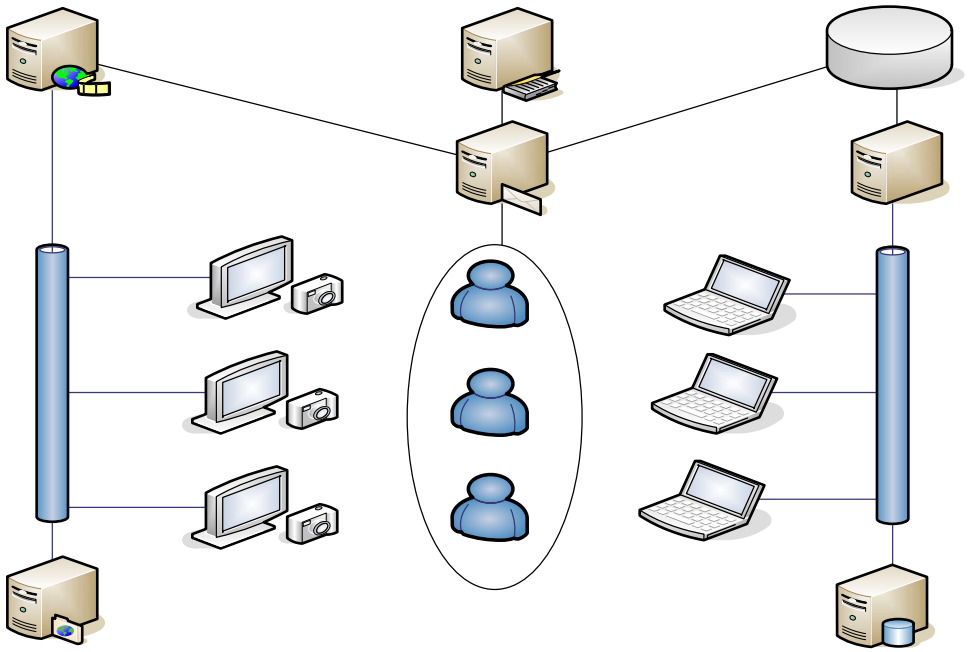

Figure 12: Solution Domain Architecture

In a medical holarchy such as our OHDS a major challenge is to develop scalable, secure web based services where the security and privacy framework is meant for the access to and the protection of sensitive information as it travels across the boundaries of individual organisations, in compliance with the Privacy of Information Act [31]. Therefore the major implementation concern is the secure web-data manipulation by medical specialists, while dealing with patients affected by diseases calling for a highly specialized knowledge and expertise. In this context a platform able to interact with a plethora of databases and other forms of information storage and retrieval methods is a must. The interaction of doctors with the information has to be secured through encryption and through a complex process of authentication and authorization. Some of the required security technologies have already been developed for other industries, e.g., in the area of electronic commerce. Other technologies such as patient-consent dependent role-based access control and

person-oriented audit trails are not readily available to date. We are currently working on the system implementation – for more details see [29].

8. COMPARISIONS, DISCUSSION AND CONCLUSIONS

In this paper an ontology-based holonic diagnostic system was presented that unifies the advantages of multi-agent system technology with those of an integrated ontology for the purpose of representing the active knowledge about human disorders. The self-organizing, emergent behavior of the resulting system supports the medical researcher/specialist, especially in cases in which the kind of disease the patient is suffering from is not certain or easily diagnosable. The ontology-based development supports the containment and control of new diseases by enabling dynamic knowledge discovery as follows:

- a *computer-based ontology* 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, e.g. which part of DNA sequence needs to be further examined in order to find the gene responsible for a disease;
- ontologies are the 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;
- constructing the data patterns which combine different genetic and environmental causes and different disease types, will facilitate the sorting out of the exact combinations of the genetic and environmental factors involved as well as their individual influences on a specific complex disease type such as e.g. depression, thereby assisting medical professionals to diagnose, treat and possibly prevent the disorder.

The four “dimensions” (phenotype, cause, treatment and type) are each built for a different purpose and are orthogonal to each other. The “Types” sub-ontology is more a classifying ontology and is strongly

hierarchically supported. It does not provide a user with much scientific information. This ontology is based on classification. The “Phenotype” sub-ontology is more descriptive than the others and is based on observation and diagnosing characteristics of the ill individual. The “Cause” sub-ontology is providing a user with scientifically proven facts and is strongly based on scientific research. The “Treatment” sub-ontology is a combination of classifying and research ontology. Modeling available treatments is research work but, for example all the discovered drugs can be further hierarchically classified. All four “dimensions” are different from each other and each “dimension” is unique. But jointly they give an overall picture and a good overview of knowledge about a human disorder.

The holarchic structure (Figure 7) provides the required distributed collaborative platform as well as easy access to resources. In the case of human diseases, we use the research publications and medical databases, DNA and protein databases, research institutes, health departments, hospitals etc as information resources. The specific information requested by a user is aligned and merged into the GHDO which results in SHDOs.

The innovation in our work lies in the combination of holonic architectures, multi-agent technology for managing and subtracting un-structured bio-medical research results into structured disease information for end users and development of Human Disease Ontologies which act as spinal cord for the diagnostic system. Involving soft computing agents/holons to integrate dispersed knowledge sources into the ontology template, and by this refining the generic template via feed-back from the specific ones, results in a powerful mechanism for dynamic building of new knowledge, on the spot, as new epidemics emerge. In addition a reference model for secure health information processing was developed as a means for the OHDS implementation. So far we have developed complete upper and lower ontologies. However, lots of work still remains, such as implementation of local agent interactions, security concerns, upload the testbed system on-line for testing and validation, test the Ontology and development of user view interfaces.

ACKNOWLEDGMENTS

Dr. Ulieru’s work on holonic diagnosis was partly sponsored by NSERC Canada under Collaborative Health Research Project grant.

REFERENCES

[1] Altman R., Bada M., Chai X.J., Whirl Carillo M., Chen R.O., Abernethy N.F., "Riboweb: An Ontology Based System for Collaborative Molecular Biology", *IEEE Intelligent Systems*, vol. 14, no. 5, pp.68-76, 1999.

[2] Barrett T.B., Hauger R.L., Kennedy J.L., Sadovnick A.D., Remick R.A., Keck P.E., McElroy S.L., Alexander M., Shaw S.H., Kelsoe J.R., "Evidence that a single nucleotide polymorphism in the promoter of the G protein receptor kinase 3 gene is associated with bipolar disorder", *Molecular Psychiatry*, vol. 8, pp.546-557, 2003.

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

[4] Craddock N, Jones I., "Molecular genetics of bipolar disorder", *The British Journal of Psychiatry*, vol. 178, no. 41, pp.128-133, 2001.

[5] Deray T., Verheyden P., "Towards a semantic integration of medical relational databases by using ontologies: a case study", *Proceedings of the ISMIS 99 Conference, Lecture Notes in Computer Science 1609*, Springer-Verlag Heidelberg, pp.30-45, 1999.

[6] Fries T. P., "Consensus development in fuzzy intelligent agents for decision making", *Proceedings of SSGRR 2001, International Conference on Advances in Infrastructure for Electronic Business, Science and Education on the Internet*, L'Aquila, Italy, 2001.

[7] Gene Ontology Consortium, "Gene Ontology: tool for the unification of biology", *Nat Genet*, vol. 25, pp.25-29, 2002.

[8]Gruber T. R., "A Translation Approach to Portable Ontology Specifications", *Knowledge Acquisition*, vol. 5, no. 2, pp. 199-220, 1993.

[9] 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*, Innsbruck, Austria, 2005.

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

[11] Hadzic M., Chang E., "Role of the Ontologies in the context of Grid computing and application for the human disease studies", *Proceedings of Interanational Confonference on Semantics of a Networked World-Semantics for Grid Databases*, Paris, France, 2004.

[12] Halpin T.: Information modeling and relational databases, 3rd edition, Morgan-Kaufmann, 2001.

[13] Hattori E., Liu C., Badner J.A., Bonner T.I., Christian S.L., Maheshwari M., Detera-Wadleigh S.D., Gibbs R.A., Gershon E.S., "Polymorphisms at the G72/G30 gene locus, on 13q33, are associated with bipolar disorder in two independent pedigree series", *American Journal of Human Genetics*, vol. 72, no. 5, pp.1131-1140, 2003.

[14] Jennett P.A., Affleck H. L., Hailey D., Ohinmaa A., Anderson C., Thomas, R., Young, B., Lorenzetti, D., Scott, R.E. "The Socio-economic Impact of Telehealth: A Systematic Review". *Journal of Telemedicine and Telecare*. In Press.

[15] 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., "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, 2003.

[16] Meersman R., "Semantic Ontology Tools in Information System Design", *Proceedings of the ISMIS 99 Conference*, *Lecture Notes in Computer Science 1609*, Springer-Verlag Heidelberg, pp.30-45, 1999.

[17] Montyne F., "The importance of formal ontologies: A case study in occupational health", *Proceedings of the OES-SEO 2001 Rome Workshop*, Luiss Publications, 2001.

[18] Nealon J. L., Moreno A., "The Application of Agent Technology to Healthcare", *Proceedings of the First Agentcities Workshop: Challenges in Implementing Open Environments, AAMAS 2002*, 2002.

[19] Principe J. C., Euliano N. R., Lefebvre W. K., "Neural and adaptive systems". *Fundamentals through simulations*, John Willey and Sons, 2000.

[20] Spyns P., Meersman R., Jarrar M., "Data modelling versus Ontology engineering", *SIGMOD Record*, vol. 31, no. 4, pp.12-17, 2002.

[21] Stevens R., Baker P., Bechhofer S., Ng G., Jacoby A., Paton N.W., Goble C.A., and Brass A., "TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources", *Bioinformatics*, vol. 16, no. 2, pp.184-186, 2002.

[22] Ulieru M., Unland R., "Enabling Technologies for the Creation and Restructuring Process of Emergent Enterprise Alliances", *International Journal of Information Technology and Decision Making*, vol. 3, no. 1, pp. 33-60, 2004.

[23] Ulieru M., "Emergence of Holonic Enterprises from Multi-Agent Systems: A Fuzzy-Evolutionary Approach", Invited Chapter in *Soft Computing Agents: A New Perspective on Dynamic Information Systems*, IOS Press, pp. 187-215, 2002.

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

[25] Ulieru M., "Adaptive Information Infrastructures for the e-Society", Proceedings of the International Workshop on Engineering Self-Organizing Applications AAMAS, pp. 167-185, 2004.

[26] Ulieru M., Rizzi M., “A Cooperative Approach to the Development of Expert Knowledge Bases Applied to Define Standard of Care in Glaucoma”, *Proceedings of CoopIS, Lecture Notes in Computer Science 2888*, Springer-Verlag Heidelberg, pp. 235-243, 2003.

[27] Ulieru M., Norrie D., “Fault Recovery in Distributed Manufacturing Systems by Emergent Holonic Re-Configuration: A Fuzzy Multi-Agent Modeling Approach”, *Information Science 7669*, pp. 101-125, 2000.

[28] Ulieru M., Crichton A. C. S., Rizzi M., Karanicolas C., “Using Soft Computing to Define Standards of Care in Glaucoma Monitoring”, *International Journal of Soft Computing: A Fusion of Foundations, Methodologies and Applications*, Springer Verlag, 2003

[29] Ulieru M., Ionescu D., “Privacy and Security Shield for Health Information Systems”, *Proceedings of IEEE-IECON 2004*, Busan, South Corea, 2004,

[30] Zhang S., Bodenreider O., “Comparing Associative Relationships among Equivalent Concepts across Ontologies”, *Presentation at Medinfo 2004,* 2004.

[31] http://www.epic.org/

[32] http://www.snomed.org/

[33] http://www.w3.org/2001/sw/