

Grid Services Complemented by Domain Ontology Supporting Biomedical Community

Maja Hadzic and Elizabeth Chang

Curtin University of Technology, School of Information Systems,
GPO Box U1987 Perth,
Western Australia 6845, Australia
{hadzicm, change}@cbs.curtin.edu.au

Abstract. This paper describes the increasing role of ontologies in the context of Grid computing for obtaining, comparing and analyzing distributed heterogeneous scientific data. In the communities of people committed to a common goal, the management of resources and services becomes very important. We chose the application domain of human disease research and control. A characteristic of the domain is that trusted databases exist but their schemas are often poorly or not documented. The network of biomedical databases forms a loose federation of autonomous, distributed, heterogeneous data repositories ripe for information integration. Grid services will provide a dynamic way to use resources in such a large distributed scientific environment while the use of ontology enables the system to carryout reasoning at 3 levels: a) available information in all Bio-Databases (Grid nodes) worldwide, b) reasoning about the retrievable information from each node, c) reasoning about the retrieved information and presenting it in a meaningful format for users. We adopted the ontology design methodology of DOGMA and developed Generic Human Disease Ontology (GenDO) that contains common general information regarding human diseases. The information is represented in 4 “dimensions”: (a) disease types, (b) causes (c) symptoms and (d) treatments. We illustrate how this GenDO helps to produce Specific Human Disease Ontologies (SpeDO) on request. We show how the combination of two different but complementary techniques, namely Grid computing and ontology, results in a dynamic and intelligent information system. The two approaches together, being complementary, enable the system as a whole.

1 Introduction

Recent developments in integrating parallel and distributed computing, combined with improvements in overall network bandwidth have made it possible to add a new dimension to distributed computing: the Grid. Grid offers data management facilities and access to distributed resources by providing cross-institutional integration of information and resources in an environment. Grid means resource integration and collaboration [13].

The biomedical community is a distributed one and involves the storage and analysis of experimental and observational data. A large body of knowledge has become available through the Internet. The information sources have complete autonomy and they are continually extending their content. Also, each area of biomedical research generates its own databases. In this community, sharing of information inside an area as well as between different research areas is essential and data from one source often must be combined with data from other sources to give users the information they desire. This network of biomedical databases forms a loose federation of autonomous, distributed, heterogeneous data repositories ripe for information integration.

The systematic growth of research efforts in biomedicine resulted in vast amounts of observational, experimental and theoretical data being scattered around the world. Two fundamental challenges in biomedical science are the management of the available information and the extraction of useful information from large data sets. There is also a need for cooperation of multi-disciplinary teams located at geographically dispersed sites on a single experimental level as well as on a higher level. Sometimes, on a higher level, information from one area in biomedicine must be linked with information from other areas (e.g. to link information about genetic causes with the information about environmental causes in order to get an overall picture of all causes responsible for a particular human disease) in order to form a network of evidence. We support a collaboratory effort in which biomedical scientists and researchers may utilize distributed computing resources to discover, access, select, and analyze data from information resources worldwide.

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.

Grid services will provide a dynamic way to use resources in such a large distributed scientific environment. It will constitute a distributed, collaborative, and high-volume computing environment that poses particular new challenges to the efficient and effective design of data and transactions. Another major advantage Grid offers is the freedom of information resources. In a Grid environment the resources may come and go, may belong to different institutions, have different usage policies and pose different requirements on acceptable requests. Grid applications, at the same time, may have different constraints that can only be satisfied by certain types of resources with specific capabilities.

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 such as the Web on the Grid. We illustrate how ontologies can be developed for the knowledge domain of biomedical and bioengineering research. We chose the application domain of human disease research and control since it necessarily 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.

In a Grid environment, information structured in ontologies may become crucial to many of the operations necessary to obtain and analyze desired data. For example, a user may want to make a collection of data files regarding only symptoms of a human

disease, but the user may not know the physical location, the name of each individual file etc. At a higher level of interoperability, shared ontologies between different systems, and mappings of a domain ontology onto a service, are important components of a service-based open architecture and re-use of tools on a semantic basis.

In the Section 2 (Related Work), we discuss related work. In the Section 3 (Ontology Data Repository on Grid for Human Disease Study) we describe extraction of the relevant information used to build the ontology. Section 4 (Principles of Building Generic Human Disease Ontology) describes the four main branches of Generic Human Disease Ontology made by using DOGMA Modeler. Section 5 (From Generic to Specific Human Disease Ontologies) illustrates on the examples the ontology as a tool for physicians (section 5.1.) and for researchers (section 5.2.). In the Section 6 (Comparisons, Discussion and Conclusions) we discuss the combination of the two complementary techniques and give final remarks.

2 Related Work

Ontology based bioinformatic work includes the Riboweb ontology[1], the Gene Ontology (GO) [6] and the TAMBIS Ontology while L&C's LinkBase® and UMLS are designed to support human disease studies.

The TAMBIS Ontology, Transparent Access to Multiple Bioinformatics Information Sources [15], uses ontology to enable biologists to ask questions over multiple external databases using a common query interface. The RiboWeb Ontology can be helpful for scientists studying ribosome related diseases, but it doesn't support study of other much more numerous diseases. Gene Ontology provides us with information about all genes within an organism and the TAMBIS Ontology represents all nucleic acids and proteins, but scientists studying a particular disease are only interested in genes and proteins responsible for that particular disease.

LinkBase® by L&C incorporates recent results involving a very large commercially available formal domain ontology. It is reported [12] 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®.

The Unified Medical Language System (UMLS) [3] project develops and distributes multi-purpose, electronic "Knowledge Sources" and associated lexical programs. System developers can use the UMLS products to enhance their applications in systems focused on patient data, digital libraries, Web and bibliographic retrieval, natural language processing, and decision support. Researchers will find the UMLS products useful in investigating knowledge representation and retrieval questions.

None of the above mentioned ontologies make use of Grid services to access and retrieve the significant information. In this paper, we show how the combination of the Grid computing and ontology can be very useful for the biomedical community.

The Grid is proposed as the new distributed computing. Originally conceived as a means of sharing resources on demand, the Grid's vision and reach has rapidly evolved to intelligent middleware for flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources. These kinds of

services enable our system to make Specific Human Disease Ontologies on request. Inter-community technology exchange and inter-disciplinary research can generate inspirational innovations. The inter-community and inter-disciplinary information exchange is important for us when constructing the Generic Human Disease Ontologies in 4 “dimensions”. For example, medical researchers examining causes of a specific human disease need to exchange the information with medical researchers working on drug design to prevent or cure that particular disease. The Grid and Semantic Web are now drawing closer together through Web Services and have a new off-spring “the Semantic Grid”, the application of knowledge technologies from the Semantic Web to both Grid applications and deep Grid infrastructure [13].

MyGrid is a project targeted at developing open source high-level middleware to support personalised *in silico* experiments in biology on a Grid. A number of BioGrid projects are underway, including the Asia Pacific BioGrid, the North Carolina BioGrid, the Canadian BioGrid, the EUROGRID project and the Biomedical Informatics Research Network. MyGrid is building services for integration such as resource discovery, workflow enactment and distributed query processing. The target users of myGrid are tool and service providers who build applications for a community of biologists. Early prototypes of myGrid services were developed and tested with use cases based on the functional analysis of clusters of proteins, identified in a microarray study of genes showing circadian rhythms in *Drosophila melanogaster* (fruit fly). Following this, a distributed system has been developed to meet the requirements of researchers studying the genetics of Graves’ disease [16]. On the contrary, our intelligent computer system is constructed that way so that it supports research, study and control of all human disease.

3 Ontology Data Repository on Grid for Human Disease Studies

Central to the Grid concept are communities of people committed to a common information-dependent goal. Medical researchers consist of teams with heterogeneous members with different capabilities. There does not exist a unique organization that has all the required resources or skills and team members to be distributed around the globe. Hence, the Grid should enable resources sharing and usage co-ordination in dynamic, virtual, multi-institutional organizations.

Grid computing is not only about accessing computing resources, but more about accessing remote data sources like stored medical and biological information in large quantities. But it would be very time consuming to figure out for each database one may need, what is in it, what is the value of the information, where it fits into the whole knowledge world and how one can access it. This is where ontologies are needed: a way to capture and present in the computer, knowledge all people in a certain community share. For instance, 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 Grid services to provide a dynamic way to use resources and services in such a large distributed scientific environment. Secondly, we need domain ontology to describe data and resources in a way that is understandable and usable by the target community.

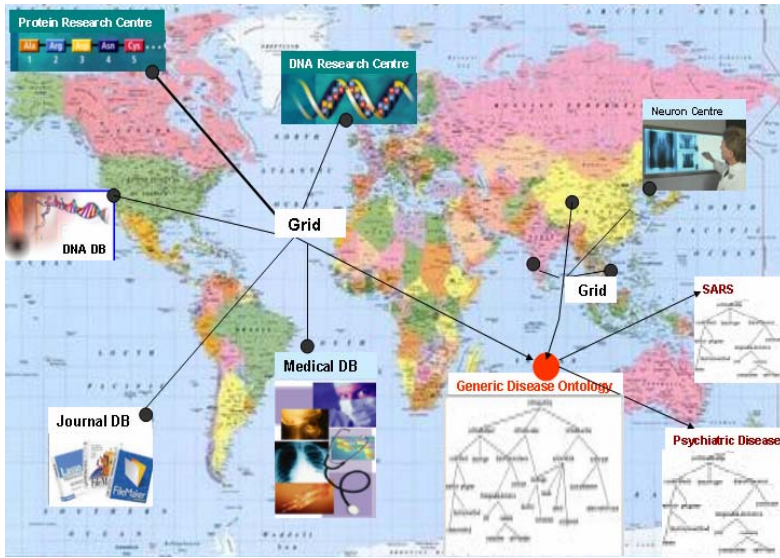


Fig. 1. Information from different databases worldwide used to create Generic Disease Ontology. Specific Ontologies such as Psychiatric Disease or SARS Ontology, derived from Generic Disease Ontology

Ontologies can effectively integrate distributed world wide research in the area of disease by aligning and merging relevant information from publication and medical databases, DNA and protein databases, research institutes, health departments, hospitals etc. [5]. Grid middleware can provide the required distributed collaborative platform as well as easy access to resources. Another major advantage of using the Grid is that it respects complete autonomy of the existing ontology nodes. Each of the existing nodes can withdraw or join the Grid whenever it is necessary. This is very important when generating on request Specific Human Disease Ontologies as we show in Section 5.

A grid-computing-based middleware system helps extracting relevant available information related to disease research from around the world . After analysis, combination and interpretation of the information according to an agreed structured representation of domain knowledge by using ontology, 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. Generic and Specific Human Disease Ontologies (see models in Fig. 1.) make it possible for researchers to carry out integrated studies involving in general multiple factors to be considered. The proposed solution provides a real-time information resource that assists researchers and physicians to analyze the different factors and the relationships between them as well as different types of diseases. Figure 1 shows a pictorial presentation of the Human Disease Ontologies deployed on a Computing Grid.

4 Principles of Building Generic Human Disease Ontology

A body of formally represented knowledge is based on conceptualisation. Conceptualisation is 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 conceptualisation, implicitly or explicitly. An explicit specification of this agreed conceptualisation is called an ontology [7]. In the sequel we shall adopt the DOGMA formalism [11], [14] for the description and terminology involving ontologies.

Ontological commitments are formal agreements (expressed in DOGMA as views, rules, and constraints) to use the shared vocabulary in 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 *facts* between *terms*. Terms are often organized hierarchically in taxonomy. Facts in DOGMA are always true only within a *context*. A lexicon L consists of a finite set of semantically meaningful concepts, denoted by C and a finite set of Relationships R ($L = C \cup R$). An ontology is a formal specification of a shared conceptualization, that is, the knowledge structure that describes the semantics of an information source by commitment to a lexicon L.

The conceptual framework of our GenDO methodology and prototype will be based on such a formal theory of ontology. Indeed, we will extract relevant information from publication and medical databases, DNA and protein databases, research institutes, health departments, hospitals etc. Upon the analysis and combination of the

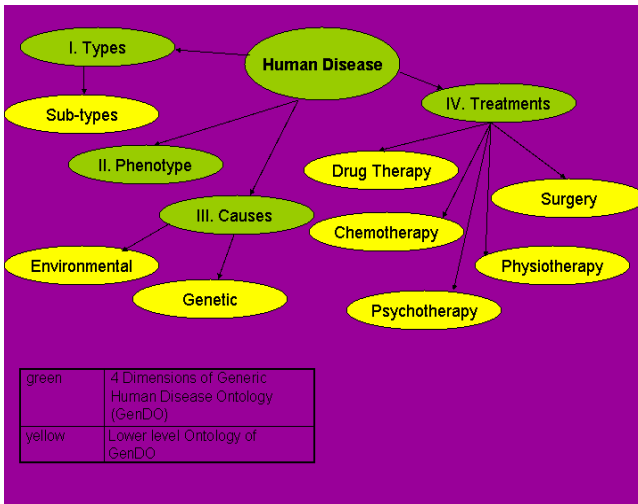


Fig. 2. Generic Human Disease Ontology and its four main subontologies: type, phenotype (sympptoms), cause and treatment

information, the result will be 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 inherited organisation of ontologies 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 GenDO has four main branches: (1) types, describing different types of a disorder; (2) causes responsible for that disorder which can be environmental and/or genetic; (3) phenotype, describing symptoms of a disease; (4) treatments, giving an overview of all treatments possible for that particular disease as well as treatments efficiency. This ontology helps to produce SpeDO as illustrated in Section 5. In the Fig. 2. we show four main branches of the GHDO. Terms within GHDO are much more numerous than shown and are validated for existence against concepts from a biomedical lexicon such as UMLS Metathesaurus [3].

Consider a vocabulary $\mathbf{V} = (\mathbf{T}, \mathbf{R})$ where \mathbf{T} is a set of terms denoting concepts, and \mathbf{R} is a set of relationship names. As a simple example, we develop a small generic ontology representing the main concepts, identified in a given (implicit) context. Let $\mathbf{T} = \{\text{disease, type, subtype, sub-subtype, phenotype, treatment, drug therapy, chemotherapy, physiotherapy, surgery, psychotherapy, cause, genotype, gene, gene complex, DNA region of interest, environment, stress, climate, family conditions, drugs, micro-organism, bacteria, virus}\}$ that represent the lexicon of user's world of diseases, and $\mathbf{R} = \{\text{has, isof, isa, is caused by, is responsible for, is cured by, cures, shows, characterizes}\}$ that represent relationships (roles) for this domain. The DOGMA Modeler uses ORM [8] notation 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 ontology explains that a disease may have (1) different types which also may be further divided into subtypes etc. Each disease is caused by (2) 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 stress, climate, drugs or family conditions. For each disease, there is (3) 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.

5 From Generic to Specific Human Disease Ontologies

By combining grid services with a prototype of Generic Human Disease Ontology (GenDO), we extract and align the relevant information from publication and medical databases, DNA and protein databases, research institutes, health departments, hospitals etc. The Specific Human Disease Ontologies (SpeDOs) are specified and generated when a user queries the system. The GenDO stands here central as a link be-

tween multiple heterogeneous information resources on one side and the users on the other side. With its four main branches (types, causes, phenotypes and treatments of a disorder) it serves as a template. Grid services then “feed” applications committed to this GenDO ontology with relevant data required by a user which results in SpeDOs.

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. This has been illustrated on the following examples. The examples are intended to show typical, common problems researchers and physicians encounter. 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. Researchers and physicians are strongly connected because they are working towards the same goal, but on different knowledge levels.

5.1 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.

Use case one: 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 three components involved in the search (symptoms (phenotype), causes and treatments). In this case, the derived SpeDOs have the “phenotype”, “cause” and “treatment” branches.

By entering the symptoms into the system, she/he may be able to retrieve the information regarding that disease. It is also possible that different diseases are showing the same or similar symptoms, so that the physician retrieves more than one SpeDO (in Fig. 3. we show two different SpeDOs). In that case, it may become useful to look for some significance in the causes of the disorders.

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 two: Physician can identify the disease and wants to consider possible treatments. It is common that there is more than one (drug) treatments possible for a particular disease (see Fig. 4.). A physician will wish to look at all the options possible before choosing one. Choosing medication is also a personal thing because not all the 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.

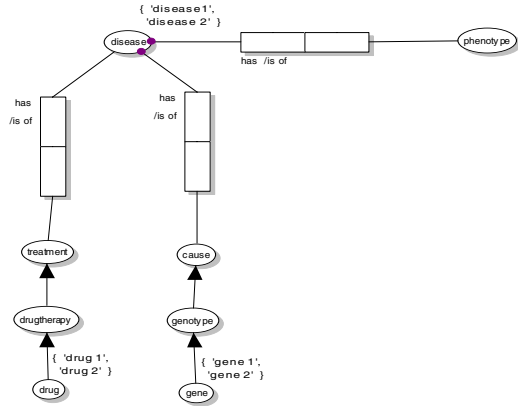


Fig. 3. Two different diseases caused by mutations of different genes and treated by different drugs showing same symptoms

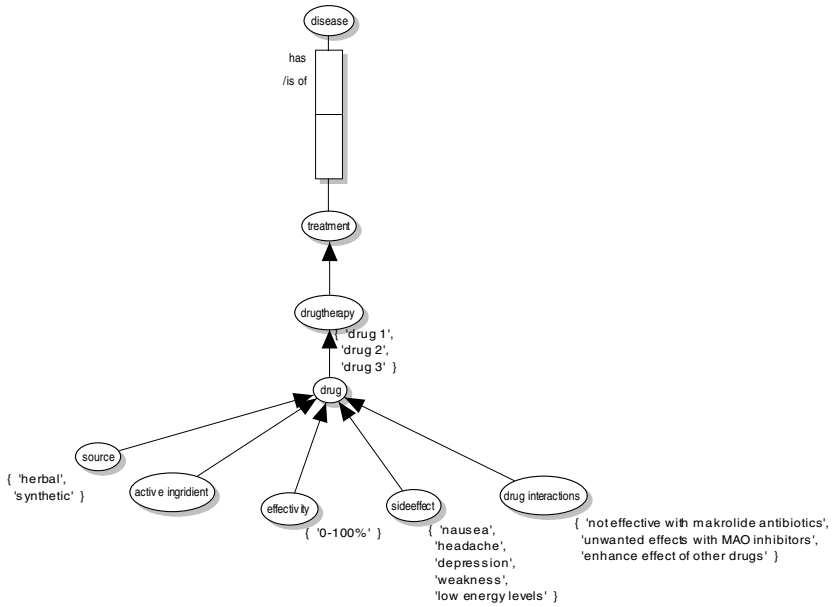


Fig. 4. Different drugs target same disease

5.2 Ontologies as Tools for Researchers

When a biomedical researcher uses our system, she/he will in general mainly be interested in one specific of the four possible components of our system, namely causes or treatments depending of her/his research area. Researcher working on drug discovery

would be more interested in the “treatment” branch. We show another example where the derived Specific Human Disease Ontology has only the “cause” branch.

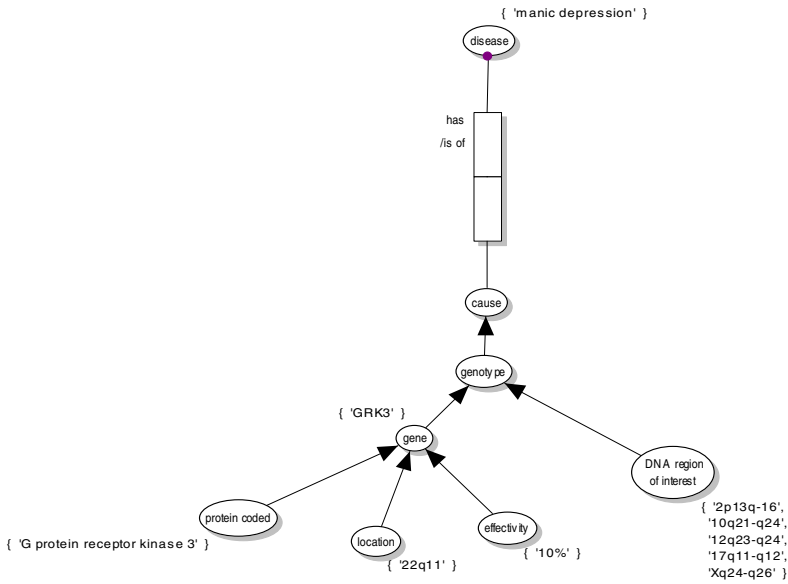


Fig. 5. Genetical causes of manic-depression, current research

Use Case Three: 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.

By querying our system and getting back significant information systematically represented (see Fig. 5.), 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], [9], [10]. 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 Fig. 6.). Because of the length of DNA sequence it obviously is 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 resources.

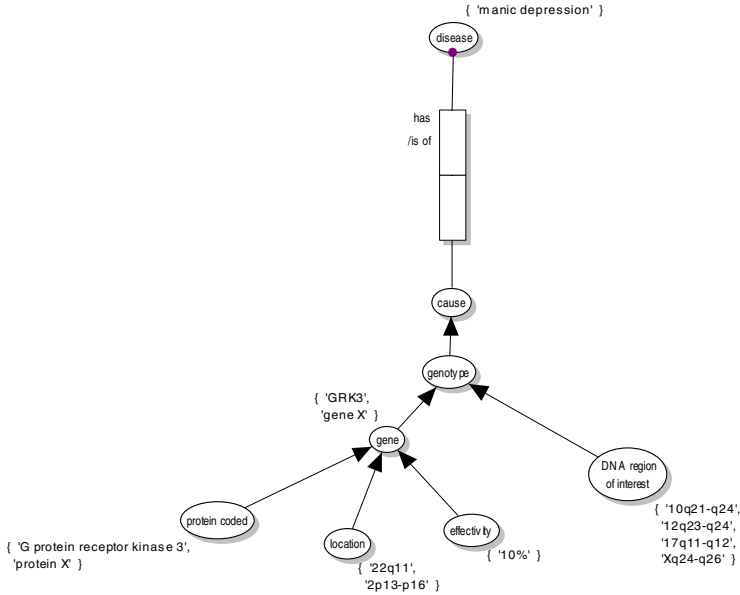


Fig. 6. Genetical causes of manic-depression, future research if gene of interest found on chromosome 2

6 Comparisons, Discussion and Conclusions

The development of an integrated Ontology deployed on Grid for the purpose of accessing, retrieving and representing the active knowledge about human disorders 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) to direct future work, such as the determination which part of DNA sequence needs to be further examined in order to find the gene responsible for a disease;
- it forms 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 combining 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 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 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 on a human disorder.

In this paper we show how the combination of two different but complementary techniques, namely Grid computing and ontology, results in a dynamic and intelligent information system. This is especially important in the communities of people committed to a common goal such as medical researchers and physicians. The Grid enables resources sharing and usage co-ordination in dynamic, virtual, multi-institutional organizations. The ontologies provide a way to describe data and resources in a way that is understandable and usable by the target community. The two approaches together, being complementary, enable the system as a whole.

Acknowledgments

This paper has been finalized during Maja Hadzic’s research visit at the VUB STAR-Lab in Brussels. The authors would like to thank Prof. Robert Meersman, Sven Van Acker, Andriy Lisovoy and other research team members of VUB STARLab for helpful discussions and useful suggestions.

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 L., “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 vol. 1609, Springer-Verlag Heidelberg, pp.30-45, 1999.
6. Gene Ontology Consortium, “Gene Ontology: tool for the unification of biology”, *Nat Genet*, vol. 25, pp.25-29, 2002.

7. Gruber, T. R., "A Translation Approach to Portable Ontology Specifications", *Knowledge Acquisition*, vol. 5, no. 2, pp. 199-220, 1993.
8. Halpin T.: Information modeling and relational databases, 3rd edition, Morgan-Kaufmann (2001).
9. 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.
10. 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.
11. Meersman R., "Semantic Ontology Tools in Information System Design", *Proceedings of the ISMIS 99 Conference*, Lecture Notes in Computer Science, vol. 1609, Springer-Verlag Heidelberg, pp.30-45, 1999.
12. Montyne Frank, "The importance of formal ontologies: A case study in occupational health", *Proceedings of the OES-SEO 2001 Rome Workshop*, Luiss Publications, 2001.
13. Carole Goble, "The Grid Needs you! Enlist Now", *Confederated International Conferences DOA, CoopIS and ODBASE 2002 Proceedings*. Lecture Notes in Computer Science, Springer-Verlag Heidelberg, ISBN 3-540-20498-9, pp.589-600, 2003.
14. Spyns P., Meersman R., Jarrar M., "Data modelling versus Ontology engineering", *SIGMOD Record*, vol. 31, no. 4, pp.12-17, 2002.
15. 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.
16. Stevens R. D., Robinson A. J. and Goble C. A., "MyGrid: personalised bioinformatics on the information grid", *Bioinformatics*, vol.19, pp.302-304, 2003.