Hadzic, M. and Cowan, R.A. (2007) Three Fold System (3FS) for mental health domain. Lecture Notes in Computer Science , 4806 . pp. 1355-1364.

Three Fold System (3FS) for Mental Health Domain

Maja Hadzic, Roberta Ann Cowan

Curtin University of Technology, Digital Ecosystems and Business Intelligence Institute (DEBII)
GPO Box U1987 Perth,
Western Australia 6845, Australia
{m.hadzic, r.cowan}@curtin.edu.au

Abstract. Along with an increase in the number of mentally ill people, research into all aspects of mental health has increased in recent years. In all disciplines information is the key to success but major problems adversely affect the efficiency and effectiveness that available mental health information is used. These relate to the lack of existing infrastructure to support effective access and information retrieval, and lack of tools to analyze the available information and derive useful knowledge from it. In this paper we explain how the ontology, multi-agent system and data mining technologies can be implemented within the mental health domain to effectively address these issues. The synergy of these frontier technologies may result in an intelligent information infrastructure that provides effective and efficient use of all available information.

Keywords: mental health research, health information systems, ontology-based multi-agent systems, data mining, intelligent information retrieval.

1 Introduction

The World Health Organisation predicted that depression would be the world's leading cause of disability by 2020 [1]. The exact causes of many mental illnesses are unclear. Mental illness is a causal factor in many chronic conditions such as diabetes, hypertension, HIV/AIDS resulting in higher cost to the health system [2]. The recognition that mental illness is costly and many cases may not become chronic if treated early has lead to an increase in research in the last 20 years.

Information technologies must be effectively implemented within health domain. In their paper, Horvitz-Lennon *et al.* [2] state that we need to fully embrace information technology and its potential for improving service efficiency and develop a better information infrastructure for patient care.

In the health portfolios of government the need for information access is understood. The US Department of Health and Human Services (www.hhs.gov) has the major goals of (1) constant access to health information for the health care consumer, and (2) improved tracking of chronic disease management for public health [3]. This has led to the development of a portal www.health.gov for agency information such as Healthfinder.gov (www.healthfinder.gov) and MedlinePlus®

(www.nlm.nih.gov/medlineplus/) for information to the public, and ONCHIT (Office of the National Coordinator for Health Information Technology) to institute heath information transfer. The Australian Federal Government developed a public gateway to health information (www.healthinsite.gov.au) and Health Connect (www.health.gov.au/internet/hconnect/publishing.nsf/Content/home) for health information transfer. The UK Department of Health offers health information to citizens via NHS Direct (www.nhsdirect.nhs.uk/) for information to the public and NHS Connecting for Health (www.connectingforhealth.nhs.uk) for heath information transfer.

Specific and targeted searches are very difficult with current search engines. For example, a search for "genetic causes of bipolar disorder" using Google provides 960,000 hits consisting of a large assortment of well meaning general information sites with few interspersed evidence-based resources. The information provided by the government sites is not necessarily returned on the first page of a 'google' search. A similar search on Medline Plus retrieves all information about bipolar disorder plus information on other mental illnesses. The main problem of the current search engines is that they match specific strings of letters within the text rather than searching by meaning. This is where the ontologies come into their own as they enable searching by meaning of the information rather then its appearance in the text.

Moreover, we believe that the synergy of ontology with multi-agent and data mining systems can result in an intelligent information infrastructure which will meet and efficiently address the needs of the mental health community. Such a 'three-fold system (3FS)' will be an innovative breakthrough within the community and would fit within the goals and criteria set out by Coye and Kell [4]. This innovation is presented as follows. An overview of some of the related works is given in Section 2. The issues related to the volume and management of the existing information and the issues specific to the mental health domain are discussed in Section 3. In Section 4, we discuss the three major technologies that can bring a breakthrough in this complicated situation. We provide a conceptual framework for a possible implementation of such system in Section 5. The benefits of the resulting system are discussed in Section 6. This paper is concluded in Section 7.

2 State of play

The importance of ontologies has been recognised within the medical community and work has begun on developing and sharing medical ontologies [5]. The Unified Medical Language System (UMLS) [6] project develops and distributes multi-purpose, electronic 'knowledge sources' and associated lexical programs. LinKBase [5, 7] contains over 5 million medical concepts, relationships and terms. The UMLS and LinK-Base are to large ontologies for the mental health projects, although a subset of them could be used to design a mental health ontology. Ontology reuse techniques, such as ontology alignment and merging, can be used for this purpose.

Use of agent-based systems enables us to model, design and build complex medical information systems. Multi-agent systems are being used more and more in the medical domain. Agent Cities [8], BioAgent [9], Holonic Medical Diagnostic System

HMDS [10] are all agent-based systems. Some agents are designed for a specific hospital and their databases (e.g., Agent Cities). The information available to these systems is institution specific and the agents help with the management of that information within a closed system. The agents cannot gather new knowledge regarding the disease in question from outside. Other agents are designed for the searching the internet, such as BioAgent and HMDS, which search for biochemical and medical diagnostic information respectively. No agents have been designed specifically for mental health domain as yet.

Ontologies were initially designed for agents to perform intelligent actions. Only in the Service Agent Layer (i.e., Ontology Service) of the BioAgent [9], does an ontology exist. In this case, the ontology is only used to provide semantics on data that is locally stored, not to interrogate the wider information domain of the internet. An ontology-based multi-agent system for human disease study and control has been proposed by Hadzic and Chang [11], and is in early implementation stage.

Within the biomedical and health field, data mining techniques have been predominately used for tasks such as text mining, gene expression analysis, drug design, genomics and proteomics [12]. Text mining uses techniques from data mining [13]. The data analysis necessary for microarrays has necessitated data mining [15]. Recently, use of data mining methods has been proposed for the purpose of mapping and identification of complex disease loci [16]. However, the proposed methods are yet to be implemented. Use of data mining techniques within the biomedical field is encouraging and is useful as a point of reference. Some of the existing data mining techniques, with minor modifications, could be used for mental health information seeking.

3 An alarming combination: Petabytes of medical data and complexity of mental illnesses

New modern techniques in medicine, science and information technology are providing huge, rapidly accumulating amounts of information. To extract and analyze the data poses a much bigger challenge for researchers than to generate the data. Experienced scientists, clinicians, mental health professionals and doctors are overwhelmed with this situation. As the size of the existing corpus of knowledge is so large, the possibility of searching it effectively is very low. In most cases some important information is being neglected.

Information regarding mental illness is dispersed over various resources and it is difficult to link this information, to share it and find specific information when needed. The information covers different mental illnesses with a huge range of results regarding different disease types, symptoms, treatments, causal factors (genetic and environmental) as well as candidate genes that could be responsible for the onset of these diseases. We need to take a systematic approach to making use of enormous amount of available information that has no value unless analysed and linked with other available information from the same mental health domain.

Three major problems are slowing advancement in mental health research: (1) issues related to the lack of an existing infrastructure to support effective *access* to the relevant information (2) lack of an existing infrastructure to support effective information *retrieval* of distributed information with inconsistent structures; and (3) lack of tools to *analyse* this information and derive useful knowledge from it. Currently, there is no lack of information but its sheer volume and its structure is a factor hindering research.

Medical information alone totals several petabyte [17]. In 2007, a researcher setting a genetic human trial for "bipolar disorder" would have needed to sift through a many results, from a variety of sources (for example, Entrez Gene www.ncbi.nlm.nih.gov/entrez/) to have found that the human gene loci 2p13-16 are potential positive sites for this disorder (the information originating from the research reported by Liu *et al.* [18]).

As research continues, new papers or journals are frequently published and added to databases and more and more of this published information is available via the internet. Problematically, no collaborative framework currently exists to help inform mental health researchers and practitioners of the latest research and where and when it will become available.

Portions of the information or data on the internet may be related to each other, portions of the information may overlap, and portions of the information may be semi-complementary between one another. No knowledge based middleware is available to help identify these issues.

The complexity of mental illnesses adds further complications to research and makes control of the illness even more difficult. Mental illnesses do not follow Mendelian patterns but are caused by a number of genes usually interacting with various environmental factors [19]. There are many different types of severe mental illness [2], for example depression, bipolar disorder, schizophrenia. Genetic research has identified candidate loci on human chromosomes 4, 7, 8, 9, 10, 13, 14 and 17 [18]. There is some evidence that environmental factors such as stress, life-cycle events, social environment, economic conditions, climate etc. are important [2, 20]. Usually in any medical research, one research team concentrates on one aspect, be it genetic or environmental, for one type of disorder. But for mental illnesses, in order for research to be inclusive, we need to look at many causal factors simultaneously. We need a tool that combines and examines all the causal factors together and derives relationships between the illness-causing factors and specific illness type.

To overcome the currently complex and complicated situation, an intelligent and efficient information system needs to be designed that does not require researchers to sift through the same or similar results from different databases. This expert system needs to be able to link information from the heterogeneous and disparate databases, and then take the leap forward of providing newly mined data patterns.

4 Strength of the three-fold system (3FS)

The combination of multi-agent systems, ontology and data mining, the 'three-fold system (3FS)', has the power to take the data contained within the health information technology systems and release unbounded, unconfined knowledge.

Agents are intelligent software objects capable of autonomous actions and are sociable, reactive and proactive in an information environment [21]. Agents can answer queries, retrieve information from servers, make decisions and communicate with systems, other agents or with users. A multi-agent system provides a distributed collaborative platform and as such determines the system dynamics.

Ontologies were brought into the computer and information society to be used by various agents and for different applications addressing the problems of various knowledge domains. Ontology provides a shared common understanding of a domain and means to facilitate knowledge reuse by different applications, software systems and human resources [22, 23]. Ontologies are highly expressive knowledge models and as such increase expressiveness and intelligence of a system.

Ontologies represent the domain knowledge and can be used to support some important processes within a multi-agent system such as: problem decomposition and task sharing among different agents, result sharing and analysis, information retrieval, selection and integration etc [11]. Ontology and multi-agent technologies can be used together to enable efficient and effective access, extraction and manipulation of the information from various information resources.

An ontology-based multi-agent system would enable us to retrieve a list of all possible causes of mental illness (genetic and environmental). We could use such a system to retrieve the information about "genetic causes of bipolar disorder". The latter result would be a small but very precise subset of the former. This result would be much more informative compared to the results of current search engines which would retrieve ad hoc information about bipolar disorder. Additional application of data mining techniques on the results retrieved through an ontology-based multi-agent system would take us a step further by identifying precise combinations and patterns of illness-causing factors associated with each illness type.

Data mining is a set of processes that is based on automated searching for actionable knowledge buried within a huge body of data to extract information and find hidden patterns and behaviours for the purpose of making predictive models for decision making and new discoveries. Data mining helps find patterns and knowledge that are embedded in the data and requires exploration and analysis of large quantities of target data for the purpose of better understanding and deriving knowledge regarding the problem at hand. Data mining draws work from areas including database technology, machine learning, statistics, pattern recognition, information retrieval, neural networks, knowledge based systems, artificial intelligence, high-performance computing and data visualization [24].

The primary role of data mining technology in the improving the efficiency of the information retrieval process has been recognised by Chen [25]. Some advantages of data mining include:

- efficient processing of large and complex data (scalability)
- automatically analysing, detecting errors and inconsistencies, classifying, and summarizing the data with no human intervention (automation)
- extracting novel and useful patterns which leads to new knowledge and discoveries (knowledge extractions)
- combining the advantages of various disciplines (multi-disciplinary nature)

- reducing the costs and time associated with the data analysis as a result of its automation (cost and time efficiency).

An efficient, intelligent and dynamic information retrieval system can be designed to unify the advantages of the multi-agent system, ontology and data mining technologies. The synergy of the different but complementary techniques gives the resulting system the following important characteristics:

1. dynamic behaviour, through implementation of a multi-agent system
2. intelligent behaviour, through use of ontology
3. reasoning abilities of constructive nature, through implementation of data-mining technologies.

These characteristics enable the 3FS to use the available information effectively and efficiently and to derive new high quality information.

5 An implementation of the three-fold system (3FS)

We propose the 3FS to include Generic Mental Illness Ontology (GMIO), GMIO-based multi-agent system and Data mining agents.

Generic Mental Illness Ontology (GMIO) can be developed to contain general mental health information. Four sub-ontologies can be designed as a part of the GMIO to represent knowledge about illness sub-groups (e.g., clinical depression, postnatal depression), illness causes (such as environmental and genetic), phenotypes (which describe illness symptoms) and possible treatments. The ontology and sub-ontologies will serve as template to generate Specific Mental Illness Ontologies (SHIO), the information specific to an illness in question (e.g., bipolar, depression, schizophrenia).

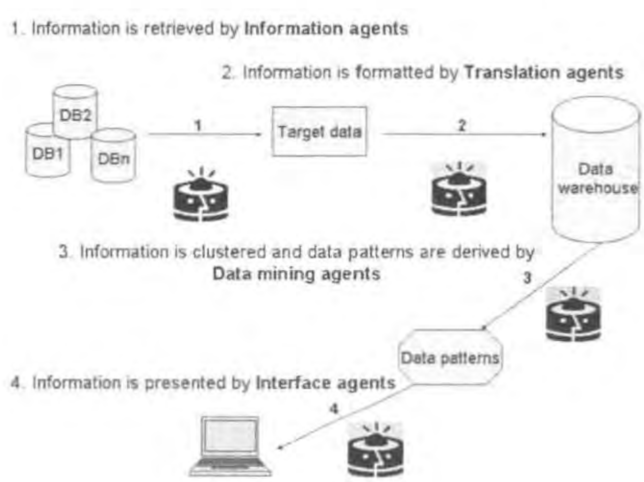

Figure 1 GMIO-based Multi-agent system

The GMIO needs to be effectively utilised within a multi-agent system. We need to define a multi-agent system architecture that will be based on GMIO and will enable the agents to collaborate effectively. A possible solution that includes different agent types (Information, Translation, Data mining and Interface agents) is represented in Figure 1. *Information agents* (Arrow 1) have the task of recognising and retrieving relevant information from various information resources. Each Information agent may have a set of assigned information resources that it needs to visit in order to gather requested information. GMIO can be used to annotate the information resources. This will enable the Information agents to understand available information and retrieve requested information effectively. The retrieved information can be represented as instances of GMIO defining concepts and their relations. Information agents hand over the retrieved information to *Translation agents* (Arrow 2). If the information retrieved by Information agents is raw and in various formats, it needs to be put into the same format. This is the task of Translation agents. When all the retrieved information is put into the same format it is locally stored in a data warehouse. *Data mining agents* (Arrow 3) compare and analyse the formatted information simultaneously in order to derive new knowledge and patterns from the information. The results are presented in a meaningful format to the user. *Interface agents* (Arrow 4) can be designed to assist with this.

Two different types of data mining agents can be designed to have different functions within the system. *Clustering agent* (1) may have the task of locating relatively homogenous subgroups in the given set of data. For example, a Clustering agent can cluster all data related to a specific disease type within a cluster. *Pattern discovery agent* (2) may be designed to discover data patterns typical for illness in question. For example, the Pattern discovery agent may provide information about precise combination of genetic and environmental illness causing factors associated with a specific illness type. This results in hidden patterns and new discoveries, results which have a potential to bring a breakthrough in the research, control and prevention of mental illnesses. Currently, no tools exist to provide this or a similar facility in mental health domain.

Protégé [27] developed by Stanford University can be utilized for the modelling of the ontologies. Java Agent DEvelopment framework (JADE) can be used to develop the multi-agent based system. JADE is a software framework that simplifies the implementation of agent applications in compliance with the FIPA (Foundation of Intelligent Physical Agents) specifications for interoperable intelligent multi-agent systems and it offers a general support for ontologies. Data mining application can also be implemented in Java. The ontology can be stored as a computer readable description of knowledge and the agents can use this knowledge for intelligent actions.

6 Benefits for the mental health domain

Mental illness is becoming one of the major problems of our society [28]. The number of mentally ill people is increasing globally each year. This introduces major costs in economic and human terms, to the individual communities and the nation in general, both in rural and urban areas.

Access to and efficient use of, available information through an intelligent information infrastructure will greatly assist mental health research. This research will assist in prevention of mental illness and assist in delivery of effective and efficient mental health services. Such a system will dramatically increase the value of the huge body of information which is currently heterogeneously stored and spatially dispersed. The vision is for this system to provide information for all levels of inquiry; for example, from the high level researcher wanting to know effects of environmental variables on the loci of chromosome 2, through the general practitioner needing to know the dosage of a drug combination for a clinical depressive alcoholic through to the carer of a schizophrenic child needing respite care. Thus the bridges between the mental health and general health sectors, that Horvitz-Lennon et al [2] aspire to will be constructed.

The resulting system will:

- support physicians in early diagnosis and treatment of mental illnesses as well as in provision of high quality health services through greatly increasing the effectiveness and accuracy of such services;
- support patients and their caregivers in dealing with, managing and treating illness;
- provide the general public accurate, reliable and up-to-date information that promotes understanding of mental illness. This will support management and prevention of mental illness.
- provide medical researchers support in advancing their research in identifying the disease causing factors and effective patient treatments as well as in reducing the possibility of redundant research (saving research time, effort and resources). Modern science and technology together with the associated medical research have the opportunity to ameliorate mental illnesses that are afflicting a substantial sector of the population.
- support the development of new technologies and facilitating development of technologies for maintaining good health;
- reduce the cost of the mental health budget and enhance the spending power of money spent on health and medical research by providing better information access [29];

The system described above goes some way to delivering what Patel et al [26] say is 'necessary to transform the quality of mental health care'. The system delivers a quality improvement in resources applicable to a diverse set of mental illnesses, improves the infrastructure for evidence-based interventions and brings in innovation for quality improvement in mental health care.

The 3FS principle could just as easily be applied to other medical and health domains as well as to the knowledge domains from other disciplines. The developed system can be used internally in a specific organization such as a pharmaceutical company or consortium such as Universities Australia, or externally in a government-supported public information service.

7 Conclusion

In this paper we have discussed numerous problems associated with use of the available information within mental health domain. We described the 3FS which unifies multi-agent systems, ontology and data mining technologies to effectively address information access, retrieval and analysis.

The implementation of such synergetic system will result in positive transformation of world-wide mental health research and management to a more effective and efficient regime. This system is highly significant and innovative, and represents a new generation of information-seeking tool. This is not another search engine. The totality of this innovation in information seeking comes from combining multi-agents, ontology and data mining. The beauty of this innovative tool is that the complexity is hidden from the client. Users will simply see targeted answers to their questions. The agents, ontology and data-mining working in concert remove complexity to where it should be, not on the human client's computer screen but within the tool.

References

1. Lopez, A. D., Murray, C.C.J.L.: The global burden of disease, 1990-2020. Nature Medicine vol. 4, pp. 1241-1243 (1998).
2. Horvitz-Lennon, M., Kilbourne, A.M., Pincus, H.A.: From silos to bridges: meeting the general health care needs of adults with severe mental illnesses. Health Affairs vol. 25, no. 3, pp. 659-669 (2006).
3. Anderson, G.F., Frogner, B.K., Johns, R.A., Reinhardt, U.E.: Health care spending and use of information technology in OECD countries. Health Affairs vol. 25, no. 3, pp. 819-831 (2006).
4. Coye, M.J., Kell, J.: How hospitals confront new technology. Health Affairs vol. 25, no. 1, pp. 163-173 (2006).
5. Ceusters, W., Martens, P., Dhaen, C., Terzic, B.: LinkFactory: an advanced formal ontology management System. Proceedings of Interactive Tools for Knowledge Capture (KCAP 2001) (2001).
6. Bodenreider, O.: The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Research, vol. 32, no. 1, pp. 267-270 (2004).
7. Montyne, F.: The importance of formal ontologies: A case study in occupational health. Proceedings of the Open Enterprise Solutions: Systems, Experiences, and Organizations (OES-SEO 2001), Luiss Publications (2001).
8. Moreno, A., Isern, D.: A first step towards providing health-care agent-based services to mobile users. Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'02), pp. 589-590 (2002).
9. Merelli, E., Culmone, R., Mariani, L.: BioAgent: A mobile agent system for bioscientists. Proceedings of The Network Tools and Applications in Biology Workshop Agents in Bioinformatics (NETTAB02) (2002).

10. 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).
11. Hadzic, M., Chang, E.: Ontology-based Multi-agent systems support human disease study and control. Edited by Czap, H., Unland, R., Branki, C., Tianfield, H., in Frontiers in Artificial Intelligence and Applications (special issues on Self-organization and Autonomic Informatics), vol. 135, IOS Press pp. 129-141 (2005).
12. Zaki, M., J., Wang, J., T., L., Toivonen, H., T., T.: Data mining in bioinformatics: report on BIOKDD'03. SIGKDD Explorations, vol. 5, no. 2, pp. 198-200 (2003).
13. Hearst, M.A.: Untangling text data mining. Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, pp. 3-10 (1999).
14. Shatkay, H., Feldman, R.: Mining the biomedical literature in the genomic era: An Overview', Journal of Computational Biology, vol. 10, no. 6, pp. 821-855 (2003).
15 Piatetsky-Shapiro, G., Tamayo, P.: Microarray data mining: Facing the challenges. SIGKDD Explorations, vol. 5, no. 2, pp. 1-6 (2003).
16. Onkamo, P., Toivonen, H.: A survey of data mining methods for linkage disequilibrium mapping. Human Genomics, vol. 2, no. 5, pp. 336-340 (2006).
17. Goble, C.: The grid needs you. Enlist now. Proceedings of the International Conference on Cooperative Object Oriented Information Systems, pp. 589-600 (2003).
18. 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. Molecular Psychiatry, vol. 8, no. 3, pp. 333-342 (2003).
19. Smith, D. G., Ebrahim, S., Lewis, S., Hansell, A. L., Palmer, L. J., Burton, P. R.: Genetic epidemiology and public health: hope, hype, and future prospects. The Lancet vol. 366, no. 9495, pp. 1484-1498 (2005).
20. Craddock, N., Jones, I.: Molecular genetics of bipolar disorder. The British Journal of Psychiatry vol. 178, no. 41, pp. 128-133 (2001).
21. Wooldridge, M.: An Introduction to Multiagent Systems. John Wiley and Sons (2002).
22. Gómez-Pérez, A.: Towards a framework to verify knowledge sharing technology. Expert Systems with Applications vol. 11, no. 4. pp. 519-529 (1996).
23. Gómez-Pérez, A.: Knowledge sharing and reuse. The Handbook on Applied Expert Systems, pp. 1-36 (1998).
24. Sestito, S., Dillon, T.S.: Automated Knowledge Acquisition. Prentice Hall of Australia, Sydney (1994).
25. Chen, W. L.: Chemoinformatics: Past, Present, and Future. Journal of Chemical Information and Modelling, vol. 46, pp. 2230-2255 (2006).
26. Patel, K.K., Butler, B., Wells, K.B.: What Is Necessary To Transform The Quality Of Mental Health Care. Health Affairs vol. 25, no. 3, pp. 681-693 (2006).
27. Protégé: an Ontology and knowledge-based editor 2006. Stanford Medical Informatics, Stanford University, School of Medicine: http://protege.stanford.edu/
28. Australian Government, "Department of Health and Ageing Factbook 2006," http://www.health.gov.au/internet/wcms/publishing.nsf/Content/Factbook2006-1 (accessed 20 February 2007)
29. Garber, A.M.: PERSPECTIVE: To Use Technology Better. Health Affairs, W51-W53 (2006).