Building Concept Maps from Unified Medical Language System (UMLS) Dataset

Anmandla Sindhura Reddy
Computers Communications and Networking
The University of Alabama at Birmingham
USA
{sindhurar@gmail.com}

ABSTRACT: *The process of carrying out medical research involves time consuming and labor intensive tasks. There is a need to enhance the present scenario of medical research by using upcoming technologies. Most current computerized vocabularies use completely incompatible ways of classifying diseases. One vocabulary might organize certain diseases as acute or chronic, another might organize some disease based on the location. The Unified Medical Language System (UMLS) is an attempt to unify all medical vocabularies into a single hierarchy, like a kind of super-thesaurus, however the UMLS forgoes any attempt to organize information into a logically rigorous fashion. The aim of the project is to design a tool which would integrate UMLS and concept map technology. This tool could be used to cope up with the incredible chaos .The researchers can represent an idea or a concept in a graphical format, linking all the concepts together, resembling a growing neural network, effectively using the C-Map tools. The tool allows the conceptual information to be automatically converted into OWL and RDF [1], as needed and then into a concept map.*

Keywords: Unified Medical Language System, RDF, Medical vocabulary, Concept Maps

Received: 12 December 2010, Revised 1 March 2011, Accepted 12 March 2011

1.Introduction

The number of biomedical resources available to researchers has been increasing with high volumes of heterogeneous as well as complex information. The proliferation in biomedical sciences perhaps is higher than other fields. To solve the information paradox problem and aid the researchers in optimal access, many tools are emerged. One such significant tool is based on semantic models using concept maps as the platform and Unified Medical Language System (UMLS). The purpose of the UMLS is to enhance access to this literature by facilitating the development of computer systems that understand biomedical language. This is achieved by overcoming two significant barriers: “the variety of ways the same concepts are expressed in different machine-readable sources & by different people” and “the distribution of useful information among many disparate databases & systems”. However UMLS forgoes any attempt to organize information into a logically rigorous fashion and hence it would be easy for a researcher to have the data in the form of a concept map.

The labor intensive sequence of processes involved in medical research is suitably replaced by using this tool built by a suitable integration of concept map and UMLS. Normally a researcher would spend a good deal of time reading various articles and construing necessary information from the article. A lot of time could be saved if the ideas can be represented in a graphical format, linking all the concepts together, resembling a growing neural network, effectively using the C-Map tools. The following steps (table 1) are included in the process. First the electronic documents that pertain to the topic of medical research and its effects are identified. Secondly the information contained in the document is Converted into an XML-based format. Finally the information available in the XML-based format is converted into a concept.

2. Typical Research Process

Method	Overall Purpose	Advantages	Challenges
Documentation Review	When want impression of how program operates without interrupting the program is from review of applications, finances, memos, minutes etc	-Get comprehensive and historical Information -Few biases about information -Information already exists	-Often takes much time -Info may be incomplete -Need to be quite clear about what looking for
Observation	To gather accurate information about how a program actually operates, particularly about processes	-View operations of a program as they are actually occurring -Can adapt to events as they occur	-Can be difficult to interpret seen behaviors -Can be complex to categorize observations -Can influence behaviors of program participants
Focus Groups	Explore a topic in depth through group discussion e.g., about reactions to an experience or suggestion , understanding common complaints, etc.; useful in 1evaluation and marketing	-Quickly and reliably get common impressions -Can be efficient way to get much range and depth of information in short time	-Can be hard to analyze responses -Need good facilitator for safety and closure
Case Study	fully understand or depict client's experiences in a program, and conduct comprehensive examination through cross comparison of cases	-Fully depicts client's experience in program input, process and results. -Powerful means to portray program to outsiders	-Usually quite time consuming to collect, organize and describe -Represents depth of information , rather than breadth
Interviews	When want to fully understand someone's impressions or experiences, or learn more about their answers to questionnaires	-Get full range and depth of information -Develops relationship -Can be flexible with client	-Can take much time -Can be hard to analyze and compare - Can be costly - Interviewer can bias client's responses

Table 1. Comparision Of Various Research Process Along With Their Advantages And Challenges

The first step in a research is plan and research plans depend on information that is needed and available resources. More in depth research requires more time and cost increases as well. The table above shows some traditional methods of research and their advantages and challenges. These traditional methods are time consuming and they must be combined with new technologies to make them more efficient.

3. Challenges Faced By Researchers

It is usually observed that the medical researcher searches for an article on the Internet using an appropriate search engine and a keyword [2].The abundance of information available is overwhelming to researchers and using this technique may not

be a speedy way of integrating available information on a particular disease or disorder. What is needed is a search strategy based on the semantics of the information required. In other words, if a researcher searches for a particular disease or disorder, he must also be able to find the information related to it.

Normally, a researcher would spend a good deal of time reading various articles and collecting necessary information from them. The researcher will have to perform the following set of tasks to carry out the research activities:

- The researcher will have to identify a set of useful books and articles.
- The researcher will spend a good amount of time reading an article and construing necessary information from the article.
- The researcher will have to perform this task on many related and interesting articles to come up with a relation between the information presented in these articles [5].

The researcher will then build an idea or concept based on his understanding of the information construed by performing the above tasks.

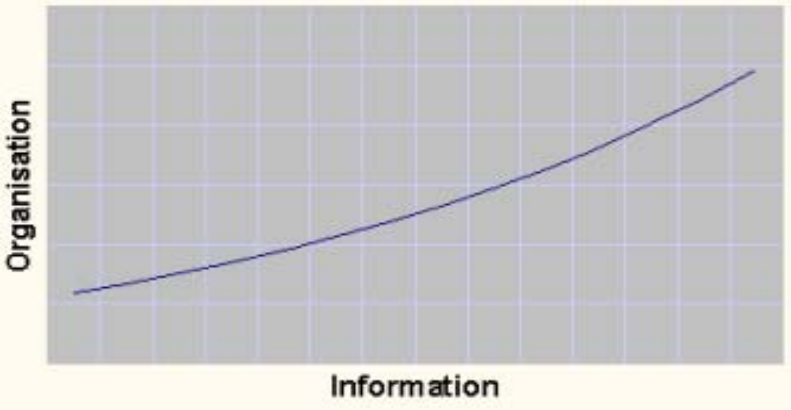

Figure 1. Graph Of Quality Of Information And NAeed Organisation[2]

The above graph shows that need to organize information increases as the quantity of information increases and hence the time required to do this increases as well.

4. Various Technologies For Medical Research

There are various sophisticated tools available for medical researcher. For example a simple excel sheet. The software behind the tool automatically searches for the data requested by the researcher and populates the Excel spreadsheet with the required data. In this manner, a researcher does not waste time in learning unnecessary interfaces but continues to use the tool on which he has obtained familiarity. The problem of preparing comprehensive information which can be used to bring research assistants or practicing physicians up to speed can be solved by concept maps[10].

5. Methods

5.1 Technologies Used

a) *Unified Medical Language System*

One of the big long-term goals of medical informatics is to develop a *Universal Medical Record* for all patients- Basically; a computer file a person can take with them wherever they go [5]. In a perfect world, any software a particular hospital uses would be able to understand this record because it is written in some kind of universally understood medical language- A difficult challenge in knowledge representation. The UMLS is an early step in making such a system possible. The UMLS is an attempt to unify all medical vocabularies into a single hierarchy, like a kind of super-thesaurus. Surprisingly, most current computerized vocabularies use completely incompatible ways of classifying diseases. One vocabulary might organize certain diseases as acute or chronic. Another might organize some disease based on the location, or maybe it may organize it underneath another disease which it is associated with- there is no uniform way (yet) of deciding how to classify such things. In order to cope with the incredible chaos these competing classification schemes bring into play, the UMLS forgoes any attempt to organize information into a logically rigorous fashion, focusing instead on being comprehensive and practical [6].

However, it is still an incredible accomplishment, since it can tie together over a hundred distinct source vocabularies and contains many millions of terms, which it is able to map between the disparate sources.

b) Java

Java has got many properties and hence it is used to write the program to convert the UMLS semantic network into concept map. Java is simple: Java was designed to be easy to use and is therefore easy to write, compile, debug, and learn than other programming languages. The reason that why Java is much simpler than C++ is because Java uses automatic memory allocation and garbage collection where else C++ requires the programmer to allocate memory and to collect garbage. Java is object-oriented: Java is object-oriented because programming in Java is centered on creating objects, manipulating objects, and making objects work together. This allows you to create modular programs and reusable code. Java is platform-independent: One of the most significant advantages of Java is its ability to move easily from one computer system to another [7]. The ability to run the same program on many different systems is crucial to World Wide Web software, and Java succeeds at this by being platform-independent at both the source and binary levels. Java is distributed: Distributed computing involves several computers on a network working together. Java is designed to make distributed computing easy with the networking capability that is inherently integrated into it. Writing network programs in Java is like sending and receiving data to and from a file. For example, the diagram below shows three programs running on three different systems, communicating with each other to perform a joint task[3].

c) C-Map tools

A concept map is a diagram showing the relationships among concepts. They are graphical tools for organizing and representing knowledge. They include concepts, usually enclosed in circles or boxes of some type, and relationships between concepts indicated by a connecting line linking two concepts.

The technique for visualizing these relationships among different concepts is called *Concept mapping*. In a concept map, each word or phrase is connected to another and linked back to the original idea, word or phrase. Concept maps are a way to develop logical thinking and study skills, by revealing connections and helping students see how individual ideas form a larger whole. Concept maps were developed to enhance meaningful learning in the sciences. A well made concept map grows within a *context frame* defined by an explicit "focus question," while a mind map often has only branches radiating out from a central picture. There is research evidence that knowledge is stored in the brain in the form of productions that act on declarative memory content which is also referred to as chunks or propositions. Concept maps have Psychological as well as Epistemological. The various features are as follows:

- Knowledge in a concept map is organized semantically.
- Ideas are arranged in networks of interconnected and interrelated ideas.
- Promotes deep processing of knowledge, and better understanding.
- Promotes the ability to apply knowledge in new situations.
- Improves understanding.

Research suggests that concept mapping facilitates problem solving [4].

6. Web Ontology Language

The OWL Web Ontology Language is designed for use by applications that need to process the content of information instead of just presenting information to humans. OWL facilitates greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics [11].

7. Integrating Concept Maps and UMLS

The interesting aspect of using a concept map tool is that it allows the domain expert to represent the knowledge in a graphical form which can be easily processed by the human mind while its computability can be harnessed by easily converting the concept map into an XML-based format. The conceptual information thus obtained in the form of XML acquires the attribute of machine-actability and can be used to create semantic data in the form of Resource Description Framework (RDF) [12] and Web Ontology Language (OWL) [11].

8. Process of Converting Text to C-Maps

The greatest challenge for storing domain knowledge of a particular medical research area is to store this information in a

human perceivable as well as machine-actable format. Concept maps can fulfill both of these requirements. While Concept maps can delineate a concept in the form of links and circles, it allows these concepts to be converted into XML-based ontology representation such as OWL and vice versa [10]. To achieve accumulation of domain knowledge in the form of concepts, the following steps will have to be performed:

1. Identify the electronic documents that pertain to the topic of medical research and its effects.
2. Convert the information contained in the document into an XML-based format.
3. Convert the information available in the XML-based format into a concept.

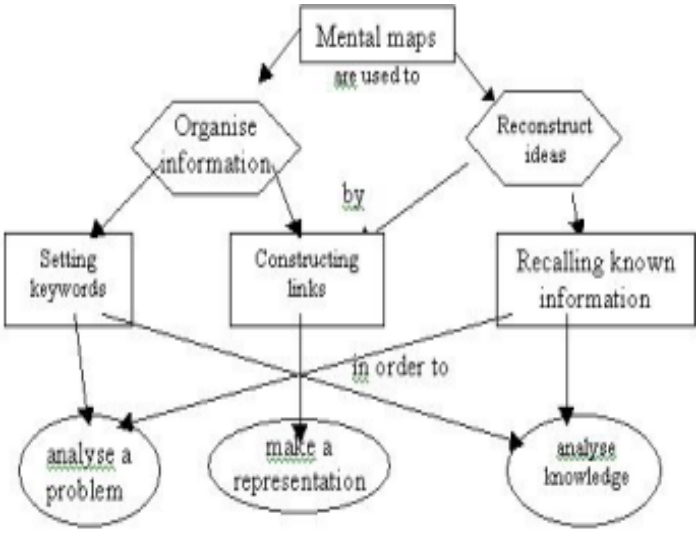

Figure 2. Example of a Concept Maps [2]

9. Alternatives

A possible alternative for the tool can be the following process

• Identify the electronic documents related to the research.

• Convert the information contained in the document into an OWL document using a tool such as TopBraid Composer developed by Franz Inc.

• Build a concept map using the OWL document thus obtained using CMap Tools

The concept map obtained by performing the above tasks can be easily used to depict the relationship between terminologies. Sometimes the research data can be populated on an Excel spreadsheet by the researcher, and it might be useful to convert this into a semantic model. The above mentioned steps can also be used to convert the data available on a spreadsheet. The CMap Tools software from IHMC has the ability to convert any document in OWL (XTM, XCM) into a concept map; however TopBraid Composer was used to convert textual data into OWL.

Although this process provides a path to build concept maps from textual data, it is required to refactor the data in OWL to build a sensible concept map.

10. Conclusion

The process of medical research is time consuming and involves labor intensive techniques. The present scenario can be greatly improved and the research process can be accelerated by using upcoming new technologies like Concept maps and semantic web. The tool has been used to successfully demonstrate the process of converting textual data into concept maps. In this way a new methodology has been provided for applying new visual tools to improve the area of medical research. The amount of time and labor involved in carrying out complicated tasks has been reduced thus making the research more efficient.

11. Acknowledgment

I would like to thank my advisor Dr. Murat M. Tanik for his constant guidance and encouragement throughout the entire coursework. I would also like to thank Varadraj P. Gurupur for his constant help on various concepts involved in the project.

References

[1] Overview of Basic Methods to Collect Information. Retrieved January 22, 2010, [online].Available: http://208.42.83.77/research/overview.htm

[2] Concept mapping as a Collaborative Tool for Enhanced Online Learning January 21, 2010, [online]. Available: www.ifets.info/journals/3_3/b01.html

[3] Java.Retrieved January 19, 2010.[online]. Available: http://www.webdotdev.com/nvd/content/view/1042/204/

[4] Gurupur,V., Tanju, M, N. (2009). Building Semantic Models Using Concept Maps for Medical Research. Retrieved January 05, 2010. [online]. Available: http://sdps.omnibooksonline.com/2009/.

[5] Unified Medical Language System. Retrieved January 18, 2010. [online]. Available: http://www.lisperati.com/tellstuff/umls.html.

[6] UMLS semantic network. Retrieved January 18, 2010.[online]. Available: http://www.lisperati.com/tellstuff/umls.html.

[7] Java.Retrieved January 19, 2010.[online]. Available: http://www.webdotdev.com/nvd/content/view/1042/204/

[8] CMAP tools for concept mapping and OWL authoring. Retrieved January 18, 2010[online]. Available: . http://www.mikepadgett.com/technology/technical/cmaptools-for-concept-mapping-and-owl-authoring

[9] Representing the UMLS Semantic Network Using OWL. Retrieved January 21, 2010,[online]. Available: http://springerlink.metapress.com

[10] Gurupur Varadraj, P., Kamdi Amit, S., Tolga Tuncer, Tanik Murat, M., Tanju Murat, N. (2009). Enhancing Medical Research Efficiency by Using Concept Maps.

[11] W3C Owl Web Ontology Language (2004). Retrieved January 21, 2010,[online].Available: http://www.w3.org/TR/owl-features/

[12] Allemang, D., Hendler, J. (2008). Semantic web for the working ontologist: effective modeling in RDFS and OWL. Elsevier Inc, Burlington.

Author Biography

Sindhura Reddy had completed her Bachelor of Engineering in Electronics and Communication Engineering from RTM Nagpur University, INDIA 2004-2008. She had completed her Master of Science Computers communications and Networking from University of Alabama, Birmingham USA 2010. Presently She is working with the same Department.