Visualizer for concept relations in an automatic meaning extraction system

Nurul Amelina Nasharuddin, Jamaliah Abdul Hamid, Hamidah Ibrahim, Mohd. Hasan Selamat, Rusli Abdullah and Wan Malini Wan Isa
University Putra Malaysia, Serdang, Malaysia

Abstract

Purpose – The purpose of this paper is to discuss the visualizer interface that has been developed for the first phase of an automatic meaning extraction (AME) system.

Design/methodology/approach – AME system was developed to automatically extract concepts and their relations across texts from all domains of knowledge. One challenge for the developer is to create interface tools that help the users use the system. This paper describes a visualizer interface that can map the concepts and relations in the form of two-dimensional graph or network.

Findings – Using this visualizer, users can maximize the use of AME system by allowing the visualization of the concepts' networks results. Users can search for a concept and view the relationships of the concept to other concepts. Those relationships can be traced back to the source sentences in the original documents through the "Show Text" function.

Originality/value – This visualizer is useful in solving the problem of visualizing the relationships between concepts across varied domains of knowledge. The extraction of relationships in the AME system is based upon a unique connector-based relation extraction. It is particularly appropriate for target users such as the researcher, educators and learners. The visualizer implements the Java Universal Network/Graph Framework to provide a few functions that enable users to manipulate the concepts graph.

Keywords Java, Knowledge management, Computer software, Interface management, High level languages

Paper type Research paper

Introduction

Automatic meaning extraction (AME) system is an on-going project on the development of a system that creates ontologies automatically and interconnects knowledge ontologies across domains of knowledge. AME allows the creation and integration of ontologies of various domains to enable comprehensive mapping of knowledge. In the first phase of the system development, AME 1 automatically extracts the concepts and relations in multiple documents and creates the mapping of the concepts and relations in a semantic schema. Semantic or meaning schema is mapping between the nodes of the concepts and relations that correspond semantically to each other (Giunchiglia *et al.*, 2005). The mapping is discerned by computing the semantic relations (e.g. equivalence, more generality, and disjointness). According to Ţăndăreanu, a semantic schema is an abstract structure, which can represent knowledge by means of an appropriate interpretation. In AME system, we have incorporated the use of competition model between concepts and their relations. However, the competition model of the AME

VINE: The journal of information and knowledge management systems
Vol. 38 No. 2, 2008
pp. 232-240

0305-5728
DOI 10.1108/03055720810889860

system is not discussed here. This present paper only discusses the visualization interface of the AME system.

Apart from the concepts and relations mapping, AME 1 ensures the integrity of the knowledge patterns by tracing the relationship between concepts in a knowledge domain back to the original documents. However, an efficient and interactive interface is needed to aid the researchers, educators and learners who use this system so that they can view the network of the concepts easily, their relationships, and to source back the logic of the relationship to the original documents in the database.

To implement concepts visualization, we have built an interface or visualizer specifically developed for the AME 1 system. Visualization is the process of transforming data, information, and knowledge into visual form (Gershon *et al.*, 1997). The use of knowledge visualizations in tools that support the development of ontologies is becoming increasingly important. These tools focus on the structure of the ontology primarily its concepts and their relationships by using different schema visualization methods.

Through this interface for AME 1, users can begin a search for any concept. This triggers a search by AME for the concepts in all the documents in the database, and other secondary concepts related to the search concept. By secondary, we refer to other concepts related to the main search concept, and not the degree of importance of the relationships. The relationship of the searched concept to other concepts across multiple documents will be visualized in the interface. Users can next further select any one concept from the list of concepts and relationships' table to zoom in on that concept's immediate relationship to other concepts. The relationship amongst concepts displayed in the network can be traced back to the source sentences in the documents through the "Show Text" function. The visualizer is written in Java and will be embedded in the project's web site. The visualizer also includes the use of Java Universal Network/Graph (JUNG) library to visualize the concepts mapping in a clear and interesting approach where the whole mapping is visualized in a graph and the concepts and relationships in vertices or nodes.

Related work

Visualization of information using visual forms has been in research for many years. Recently, the interest of the community on knowledge management has brought attention to the effective representation and sharing of knowledge through the issue of knowledge visualization (Becerra-Fernandez *et al.*, 2004). Knowledge visualization which is a new field in today's research is defined as the use of visual representations to transfer knowledge between at least two individuals (Burkhard and Meier, 2004). Cañas *et al.* (2005) discussed and distinguished between information and knowledge visualization whilst Novak and Gowin (1984) regard the concept map as a visualization of knowledge and other media (images, videos, text, sound, etc.) as the visualization of information.

An efficient visual interface for both information and knowledge visualization's system enable users to interact with large volumes of data rapidly and users can reveal hidden characteristics and patterns of the data. Although AME 1 system still has not yet achieved full extraction of knowledge ontologies from the documents, we need to design its interface flexible enough so that additional functions transforming semantic schema to knowledge ontology in future is easy to be added. In order to represent meaningful relationships between concepts that have been extracted by the AME 1 system, we decided to implement the idea of concept map; a schematic diagram that consists of nodes and directed edges. Rather than a map, in this AME I interface the network of concepts and edges will be referred to as a graph. In conventional concept maps, two-dimensional graphical displays show concepts (usually presented within nodes or boxes), that are connected by directed edges encoding brief relationships between pairs of concepts. Usually, these relationships (referred to as linking words or linking phrases) comprise of verbs, prepositions or phrases for each pair of concepts. Figure 1 shows an example of the conventional concept map about "why we have seasons" by Novak and Cañas (2006).

Concept maps are traditionally given in a hierarchical view and the links annotated with arrowheads usually run top-down with a root node. This shows that inclusive concepts are at the highest levels while the more specific but less inclusive concepts are arranged below them. The linking phases in concept maps are usually words on the edges without having their own nodes. Maps that have similar concepts differ from one to another based on its context, usually constructed by different persons who are experts in their fields. Therefore, it is not possible to get the correct concept map about a particular topic as there can be many different representations of the topic, depending on the experts' viewpoints.

For AME 1 interface, the graphs are also formed by using concepts and relationships. But rather than having the linking phases attach on the edges, we decided to put them in

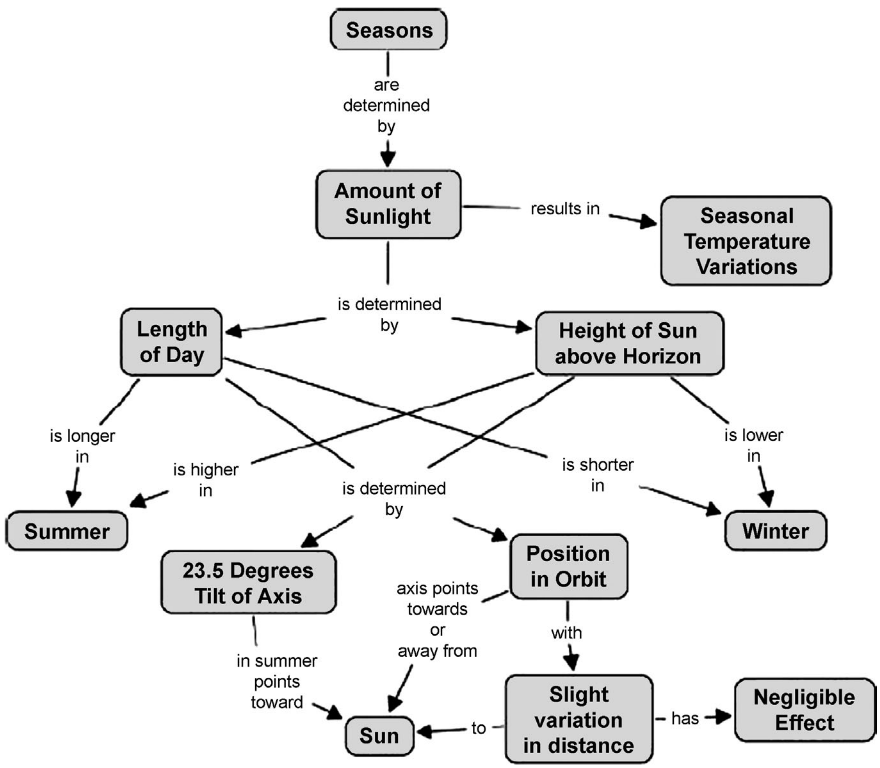

Figure 1.
Example of concept maps about seasons

their own nodes with different colours so that the user will have a clear difference between the concepts and their relationships. In this interface, the graph is not necessarily top-down and the concept that user selected to be the main concept will be drawn at the centre of the graph. Although the representations of the graph will be different each time depending on the user selection of which concept is selected as the main concept, the context of interpretation will be the same. This is so because AME automatically extracts the meaning schema from the texts and store it in the database, without any intervention from the experts. To facilitate users to centralize the selected concepts on the interface, the nodes are designed so that they can be manually rearranged and manipulated by the users. Users can also converge in on any subnetwork of concept relations displayed on the interface, and trace the logic of the relationship between concepts in the subnetwork back to the original documents in the database.

There are a quite number of tools and packages that can be used to built the visualization of ontology and many of them provide good features to aid the users in using it. Some of them are UCINET (Borgatti *et al.*, 2008), Pajek (Batagelj and Mrvar, 2004a, b), JGraph (Alder, 2007) and JUNG. UCINET and Pajek are examples of stand alone applications whose each provides a number of tools for visualizing and analyzing networks. These software can successfully visualize the network and provide the user with some powerful visualization features such as nodes manipulations and rich collection of graphical primitives and output formats. Still, they do not conveniently address programmatically by other applications, and so they are not well-suited to process large numbers of graphs. Furthermore, they are applications rather than libraries. Users can employ these applications but cannot write their own routines, nor are they able to take advantage of the capabilities of existing code, or to use methods that are not provided as part of the application.

We found that although JGraph (Alder, 2007) is an open-source library, oriented towards design layout, it barely met our needs in terms of graph operations. JGraph comes in three separate packages where only the first one is a free version. Moreover, the first package of JGraph comprises only the basic JGraph swing components.

To overcome the problem of getting access to and to use open source codes, we decided to implement a third-party Java library that can easily be adapted to the AME 1 environment. The library also enabled us to extend the functionalities of the existing codes. The JUNG (O'Madadhain *et al.*, 2003) is an open-source Java software library that provides a common and extendible language for the modeling, analysis, and visualization of data to enable it to be presented as a graph or network. It can be implemented in any Java-based applications and allows use of extensive capabilities of the Java API. One of the functionalities in JUNG that is suitable for visualizing ontology networks is that it provides a mechanism for annotating metadata to the graphs, entities and relations. This facilitates the creation of analytic tools for complex data sets that can examine the relations between entities as well as the metadata attached to each entity and relation.

The visualizer

To summarize, the visualizer helps researchers, educators and learners to use the AME 1 system and visualize the network of the concepts across all domains in a new approach. The visualizer uses graph to enable users to easily understand the structure of the network. The visualizer interface is created in Java Applet using Java language. The visualization of the ontology was created using JUNG framework.

The graph of the model in the visualizer is built from the inputs from the AME 1 database which are entity id, entity name and entity type. The entity id is a unique identifier that will be assigned for each of the nodes in the graph. Entity name is the noun or verb content that will be represented by each node. Entity type is used to differentiate between a concept and relationship when visualizing the concept network.

Documents uploaded into the database will be analyzed by the AME 1 system, and the system will automatically extricate the entity id, entity name, entity type, schema id and flow and store it into the MySQL database. The information needs to be captured since the visualizer needs to redraw the graph every time users enter a new query. The visualizer draws a graph based on the query input by the user and the graph is redrawn with every new input. The context that frames the relationship of all concepts remains the same, except that the concepts can be manipulated to become the centre of any graph network.

How to use the visualizer

The nodes of the graph are the concepts or relations in a text and the edges shows the connections between the nodes. Nodes representing a concept are red in colour, while nodes representing a relationship are green. A screen will pop up when the cursor hover a node, which shows the type and content of the node. Visualizer also includes a set of toolbar and attributes panes that allow access to various operations. Figure 2 is a screenshot of the visualizer that displays a visualization of a concept and its related concepts.

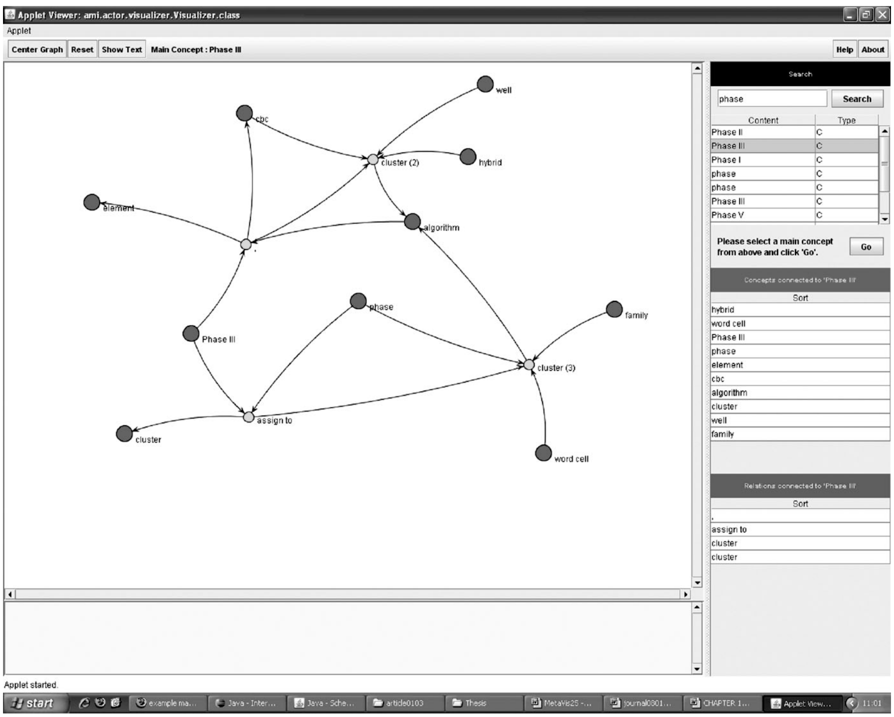

Figure 2.
The visualizer

The visualizer contains a graph viewer, attribute pane at the right side, toolbar and text pane at the lower side. The first operation when using the visualizer is "Search." A user needs to enter a word in the "Search" field. For example, in Figure 2 a concept "phase" is being searched. By clicking the "Search" button, a list of contents and its type which are all related to the word "phase" is displayed in a table. Users can select any content by clicking on it and click button "Go." The AME system will search in the database the entire concept network related to the searched concept and the network is visualized as a graph in the graph viewer with the selected concept at the centre. The concept table will display all the concepts that are related to the graph drawn. The same goes with the relationship table. When a user clicks on one of the concept or relationship in the list, the related vertex in the graph will be highlighted. The nodes in the graph can be manually rearranged by the users if they find the graph is cluttered.

The graph is not the same every time it is drawn because the positioning is automatically calculated by the interface. Furthermore, as we discussed earlier, we did not employ the idea of the concept map where the map has to run from top to down because each map may have different representations. In this interface, the way the graph is drawn is not as important as in the conventional concept maps. More important is whether the concepts are related accordingly to each other in an arrangement which we called a meaning or semantic schema. Every schema produced in the search will be stored in the database and can be recalled when drawing graphs.

AME 1 system also has included a function of text retrieval which is a process of searching for the portion of text from the documents that have been uploaded in the database that supports a particular semantic schema. Users select all the vertices related to each other in a meaning schema and click button "Search for Text." AME 1 system will analyze the input schemas and the most frequent sentence with the most co-occurrences of the schemas is assumed to be the most relevant proof to the schemas. The targeted sentence will be displayed in the text pane below the graph viewer. This will justify that each of schema in the graph is built truly based on the documents that have been processed (Figure 3). Semantic retrieval function has the potential to be an important tool in knowledge tracing especially in the next phases of AME system.

By clicking the "Center Graph" button in the toolbar area users can center a graph with the selected vertex as its center. Button "Reset" can move the graph back to its original coordinate. The last function provided to users is the "About" button where a pop up screen will be displayed which will briefly describe this interface and the functions it have.

Conclusion

Currently the visualizer can display a subnetwork of concept mapping and conceptual meaning schema of sentences in multiple documents. In the next development phase of the AME system, we will enhance the interface to portray a more complex knowledge schema with enhanced relationship type nodes. The important function provided by the present AME 1 system is to trace the logic path between concepts and their relationships to the original text document. The visualizer strengthens the case for the extraction of the conceptual relationships as well as the integrity of the concept schema presented. This interface also provides a powerful mind mapping tool for researchers. This interface is still new in the field of information visualization and the development and refinement is an on-going project.

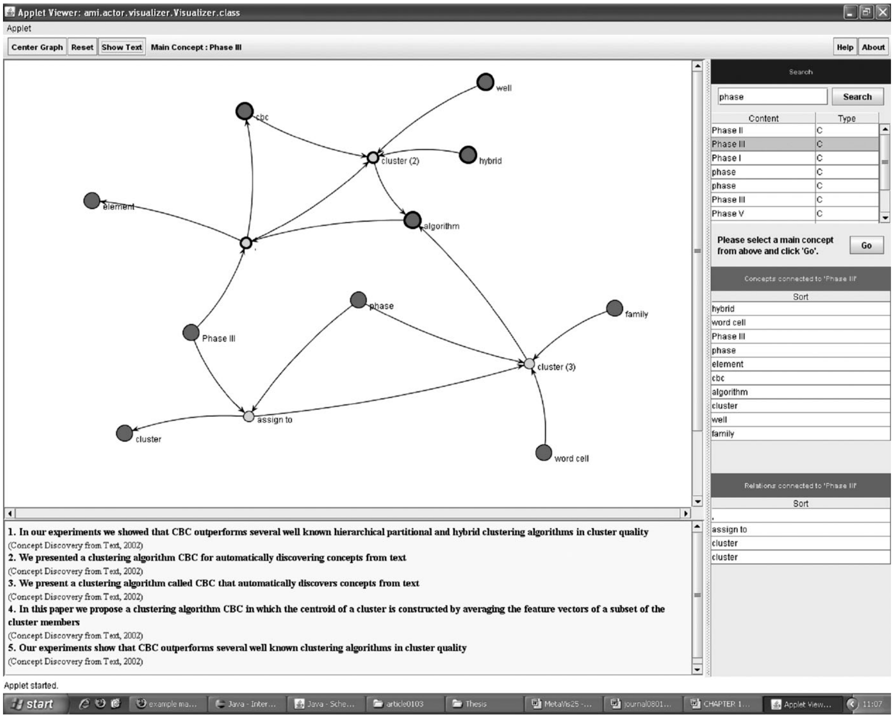

Figure 3.
The function of text retrieval in visualizer

Some of the unresolved issues that still need attention are the centrality and undo function. We need to develop a "central concept" function, possibly based on the hierarchical order, in order to filter the nodes to form a simpler graph. However, unlike the conventional concept maps, the hierarchy here is calculated after the extraction of concept networks. The undo and redo function are also a must since to allow for a lot of functions that can manipulate the graph and change its state.

As we continue to receive more feedbacks, we will possibly add in new functions and conduct experiments that will enhance the effectiveness of the visualizer. Once ready, this interface will be embedded in the project web site so that all researchers can access the system and give us feedback. If extended further, this interface can be incorporated in an artificial intelligence system for decision making to trace the logic of decisions documented in various texts. In a nutshell, since all texts contain related concepts, this automatic visualizer enables rapid online search of the mapping of concepts and their relationships.

References

Alder, G. (2007), *JGraph: The Java Graph Visualization Library*, JGraph Ltd, Northampton, Version 5.10, available at: www.jgraph.com

Batagelj, V. and Mrvar, A. (2004a), "Pajek: program for large network analysis", *Reference Manual: List of Commands with Short Explanation*, available at: http://vlado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.pdf

Batagelj, V. and Mrvar, A. (2004b), *Pajek: Package for Large Networks*, Version 0.98, University of Ljubljana, Ljubljana, available at: http://pajek.imfm.si/doku.php

Becerra-Fernandez, I., Gonzalez, A. and Sabherwal, R. (2004), *Knowledge Management: Challenges, Solutions, and Technologies*, Prentice-Hall, Upper Saddle River, NJ.

Borgatti, S., Everett, M. and Freeman, L. (2008), *UCINET 6 for Windows: Software for Social Network Analysis*, Version 6.182, Analytic Technologies Inc., Lexington, KY, available at: www.analytictech.com/ucinet/ucinet.htm

Burkhard, R. and Meier, M. (2004), "Tube map: evaluation of a visual metaphor for interfunctional communication of complex projects", *Proceedings of I-Know '04, Graz, Austria*, pp. 449-56, available at: www.know-center.tugraz.at/previous/i-know04/papers/burkhard.pdf

Cañas, A.J., Carff, R., Hill, G., Carvalho, M., Arguedas, M., Eskridge, T.C., Lott, J. and Carvajal, R. (2005), "Concepts maps: integrating knowledge and information visualization", in Tergan, S-O. and Keller, T. (Eds), *Knowledge and Information Visualization, LNCS*, Vol. 3426, Springer, Berlin, pp. 205-19.

Gershon, N., Eick, S.G. and Card, S. (1997), "Information visualization", *IEEE Journal of Computer Graphics and Applications*, Vol. 17 No. 4, pp. 29-31.

Giunchiglia, F., Yatskevich, M. and Giunchiglia, E. (2005), "Efficient semantic matching", *ESWC 2005*, pp. 272-89.

Novak, J.D. and Cañas, A.J. (2006), "The theory underlying concept maps and how to construct them", Technical Report IHMC CmapTools 2006-01, Florida Institute for Human and Machine Cognition, Pensacola City, FL, available at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf

Novak, J.D. and Gowin, D.B. (1984), *Learning How to Learn*, Cambridge University Press, New York, NY.

O'Madadhain, J., Fisher, D., White, S. and Boey, Y.B. (2003), "The JUNG (Java Universal Network/Graph) framework", Technical Report UCI-ICS 03-17, University of California, San Francisco, CA, available at: www.datalab.uci.edu/papers/JUNG_tech_report.html

Further reading

Cañas, A.J., Hill, G., Carff, R., Suri, N., Lott, J., Eskridge, T., Goméz, G., Arroyo, M. and Carvajal, R. (2004), "CmapTools: a knowledge modeling and sharing environment", in Cañas, A.J., Novak, J.D. and González, F.M. (Eds), *Concept Maps: Theory, Methodology, Technology, Proceedings of the First International Conference on Concept Mapping*, Universidad Pública de Navarra, Pamplona, pp. 125-33.

Fluit, C., Sabou, M. and Harmelen, F. (2000), "Ontology-based information visualization: towards semantic web applications", in Geromeinko, V. and Chen, C. (Eds), *Visualizing the Semantic Web: XML-Based Internet and Information Visualization*, 2nd ed., Springer, Berlin, available at: www.cs.vu.nl/ ~ frankh/postscript/VSW05.pdf

Rhodes, P., Kraemer, E. and Reed, B. (2006), "VisIOn: an interactive visualization ontology", *Proceedings of the 44th Annual Southeast Regional Conference ACM-SE 44*, ACM Press, New York, NY, pp. 405-10.

About the authors

Nurul Amelina Nasharuddin is at the Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Malaysia.

Jamaliah Abdul Hamid is an Associate Professor at the Department of Foundations of Education, Faculty of Educational Studies, Universiti Putra Malaysia, Serdang, Malaysia. Jamaliah Abdul Hamid is the corresponding author and can be contacted at: aliah@putra.upm.edu.my

Hamidah Ibrahim is an Associate Professor at the Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Malaysia.

Mohd. Hasan Selamat is an Associate Professor at the Department of Information System, Faculty of Computer Science and InformationTechnology, Universiti Putra Malaysia, Serdang, Malaysia.

Rusli Abdullah is at the Department of Information System, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Malaysia.

Wan Malini Wan Isa is at the Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Malaysia.