A New Design for Knowledge Visualization based on User's Understanding of Wiki data

Mohsen Khoshgoftar †, Behrouz Minaei ††, and Ahmad Faraahi †††

† MSc Student, Tehran Payame Noor University
†† Assistant Proffessor, Computer Engineering Department, Iran University of Scinece & Technology
††† Assistant Proffessor, Computer Engineering Department, Tehran Payame Noor University

Summary
This paper represents a method for the knowledge visualization for the technical words in Newton's Telecom Dictionary in order to make users understand the intended topic and improve the results of search for resources in wikis by Google search engine. The WordNet dictionary and the Encyclopedia Wikipedia are used as the sources of knowledge acquiring. The new design of visualization named CloudViz Knowledge, in which there is a hierarchy of knowledge tree for the intended word along with its various levels of meaning, is also introduced here. Understanding the meaning of words and gaining mastery of the related knowledge assist users to achieve a desired research with more related results in a short period of time.

Key words:
Visualization, Wiki, Knowlage Discovery, Search Engine

1. Introduction

With the increasing growth of the internet, a vast amount of database from a variety of different areas of knowledge is available on the web. The main problem of data finding process is finding too much data, in which users face thousands or millions of web sites. Inaccessibility to data-finding experts and lack of knowledge about the rudiments of searching and information-finding make the process of information finding on the internet time-consuming and inefficient. On one hand, the search engines search the web based on words, not based on meaning. On the other hand, a considerable part of the web is hidden or inaccessible to the search engine. Institutes and many other scholars have worked on knowledge management and developing awareness. The purpose of the study is to provide a user adaptor for the visualization of the knowledge related to the word being searched in the form a new design in which user's understanding and mastery is increased. Furthermore, a new term is suggested, from among the words that are related to the topic being searched, which help users in finding useful resources in a short period of time. The outline of the paper is as follow:

Section 2 deals with the related literature. Section 3 indicates the aims of the visualization software. Section 4 represents the details of the design. Section 5 evaluates the design. And finally, chapter 6 proposes some suggestions for further research.

2. Review of the Related Literature

WordNet have established a lexical database with the aim of providing a combination of dictionary and thesauruses in order to be more usable and to automatically support artificial intelligent and text analysis applications in English language. The different meanings of a word differ from one synset to another. Most of the synsets are connected through semantic relations [1]. A visualized design is introduced for the relations of this web in the form of graph [Kidd E, 2009] which depicts the hierarchical classification of words, the process of finding different parts of a figure and answering the other questions [2].
Wikipedia is a free, open source encyclopedia provided in wiki in several languages, in which every one, with whatever knowledge, can write or copy or edit it. The main characteristic of these articles is the classification of their structures into different and similar parts. There is a definition of the word in the first paragraph of each article. Due to its open source nature, however, the articles are not reliable without references [3]. Several visualized designs concerning the different topics of Wikipedia are introduced. For example, Holloway, Todd, et.al. (2007) represented a method for analyzing and visualizing the semantic coverage of Wikipedia and its writers, which is shown in the form a mass of colorful spots [4].
Concerning the intelligent visualization of semantic wiki, the ENWiC software is capable of visualizing the content of a certain educational wiki [5]. The other study is the generalization of the method of radiant space-filling trees for the visualization of semantic wikis. There is a circle in which the related heading is observable through moving the mouse on its different parts [6]. PatViz is a development of JSP Wiki system which is used for the

Manuscript received December 5, 2010
Manuscript revised December 20, 2010

visualized representation of web notes of recorded data by the use of semantic wiki [7].

3. The Semantic Levels of Knowledge

Having worked on the terms, we learnt about the main components of user's understanding and its close relationship with data visualization. The mental image formed in human mind by physical objects is among the first components of visualization and understanding. Hearing the word *stork*, you imagine a bird with a long beak and long legs, i.e. the same bird, in our minds. The second component is visualization with the help of other images which are related to the intended object or concept. You imagine water with the help of a pot. Because it doesn't have color or a certain shape, you see it in the bed of a river or sea, inside a glass, or in the form of drops of rain through light rays and environment color. The third component of visualization is making use of the applications of the words. You don't have an image of air; however, you can see a flying hot air balloon filled with air. The forth component of user's visualization and understanding is paying attention to the attributes and properties of the word. You can't see depression, however, you call a person with a winced and unhappy face bending his/her shoulder forward while looking at the ground as depressed. In fact, in order to understand words and terms, after visualizing meaning and concept, we visualize their attributes, properties, characteristics and applications in our minds. Understanding of words is impossible without visualization. Therefore, for every technical term, we define, beside its lexical root based on which the technical points have a different definition, the three levels of meaning.

The first level of meaning is the words that are semantically and conceptually synonymous withs the intended word. For example, wiki, which is a technical word, has a literal meaning of "quick" in Hawaiian language and its root is the word *area* or *region*. The words *quick, spry, flying, fast, speedy, warm, straightaway, prompt, immediate, ready, agile*, and *nimble* are at the first level of its literal meaning, which are synonymous with the word "quick".

The second level of meaning is, in fact, the words that define the intended word technically. For example, the terms *Inter Linked Web Page, audit, create, Web Browser, Website, XYZYWYG Text Editor,* and *Markup language* are at the second level of meaning for the word "wiki".

The third level of meaning, which is called semantic vicinity radius, includes the other words that are somehow related to the word. We call this level semantic vicinity radius because we evaluate the degree of the relation of the related words to the intended word through the amount of its distance from the concept. For example, the word *Enterprise 2.0* is the closest word to wiki, because it is one of its technologies and is the best bed for that. *Web 2.0* stands in a farther radius because it is one of its components. *Web 1.0* is, in fact, the first generation of the web; so it stands on a much farther radius. This process continues for the word *web* and *internet*. Finally, with the help of root and the fact that root "is a kind of what", we follow the hierarchy of knowledge until we reach the primary content in order to reach the first word. Knowing the semantic relations of the intended term with the other words, the user will definitely have a better understanding and mastery of the word, and will do better in his/her search for the intended resources and data in wikis.

4. Visualized Algorithm and Software

We ran this design in the form of visualized software CloudViz.

4.1 Algorithm

From among different dictionaries of technical words we choose Newton telecom dictionary. Every word is explained in details in the dictionary. The key words of terms, which have a close relationship with the intended technical word, is extracted (as a second semantic level of meaning of the words), and they are saved in a table. Other than the explanation of terms, some references such as "see also" is provided about the other technical words which are used as vicinity radius. On the other hand, concerning the Wikipedia dictionary and the fact that it explains the meaning and defines the intended term in the first paragraph, some other words can be extracted as a first semantic level and vicinity radius. Each technical word may have a record in WordNet dictionary with the same spelling; and considering the fact that if the word is a noun or adjective (most of the words are nouns), its synonymous word is obtained from this dictionary and is saved in a separate table. If the technical word of Newton telecom dictionary is a compound word, we choose the main word; and if its spelling is different, e.g. *enterprise2.0* is spelled *enterprise* in WordNet. Then we find its synonymous words from WordNet dictionary and record it as a first semantic level on a separate table, and the option "is a kind of" is also used for depicting the knowledge tree.

4.2 Algorithm

Being able to make a conceptual database, we use visualized software called Cloudviz in order to have a better understanding of word relation and synonymous words. In this software, the image of a tree is used for showing knowledge tree and the image of knowledge cloud is used for indicating synonymous words and

related words. First, the intended word is entered in user interface by the user. Then, by pressing the "enter" key, the word is written in the center of the page. After that, the root and knowledge tree is drawn in the form of a tree at the bottom of the page, and the words from the three levels of meaning appear in three separate clouds (figure 1). Among the words, those that have a record in Newton telecom dictionary are in a different color. Tooltips of the word appears from Newton telecom dictionary by moving the mouse on them (figure 2- right); and the second level of meaning appears by clicking on the knowledge cloud (figure 2- left). The relation of the intended word with the other technical words can be seen in the form of a graph. From among these words, those that exist in Newton telecom dictionary can be seen in a different color.

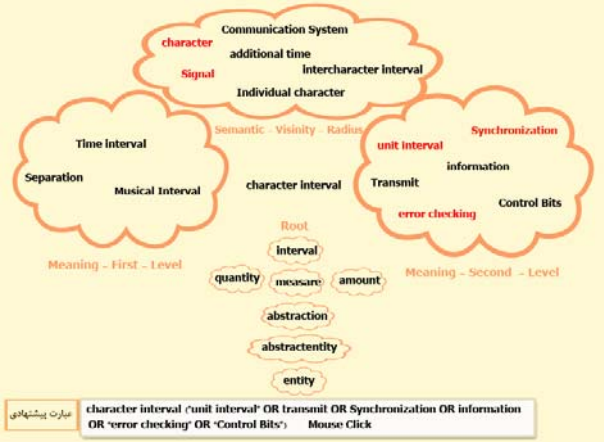

Figure 1. the words from the three levels of meaning

4.3 Suggested Term for Searching

If a user wants to search about a word, it is enough for him/her to press the search box so that the suggested terms (in this design, it is tried to use an appropriate combination of operators and the facilities of Google search engine along with the second level of meaning) represent some results to users. This way, the number of pages found by the search engine is fewer, and the found resources are more related and closer to the word being searched because of the existence of assisted related words. When searching, we use those words in second level of meaning that have already been used in the definition of the main word. User is also free, based on his/her realization, to use another combination of words from among the illustrated words so that he/she can find the related resources. By doing so, the spent time for searching the source and the required data in wiki data on the internet will be much shorter.

5. Evaluation of the Results

In order to evaluate the extent of efficiency of the design in user's understanding and mastery and also its effect on searching for resources through the internet, some questioners and the suggested software were provided. Then, they are given, along with some words, to the students majoring in computer engineering. The results of the survey were collected as follow:

About 80% of the students believed that the knowledge tree plays an important role in understanding the place of the word throughout the whole knowledge, and they also found it useful. 20% of the students, on the other hand, didn't consider it as a useful tool for obtaining information and new awareness.

About 85% of the participants considered the visualized design efficient for their understanding and mastery of the intended word. 100% of the participants believed that the results of the search for the suggested terms made them get the desired results in a shorter period of time while the number of unrelated pages decreased to zero or near zero, and the number of useful pages searched the same way as the user was few compared with the whole pages reviewed by the user.

In the meanwhile, the number of results found by Google search engine was reduced dramatically and much fewer pages were found by the search engine. It means that the pages that were unrelated or less related to the intended topic were filtered. 80% of the users evaluated their mastery of the internet, search engine and English language as intermediate.

At the end, the users suggested some words that they considered as useful for the results of the search. After investigation, some of these words were added to the related database. They also had some suggestions about applying a dictionary along with the design.

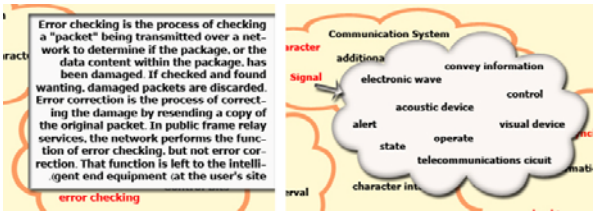

Figure 2. Tooltips of the word and second level of meaning appears by clicking

6. Suggestions for Further Research

One part of the research for the future work should focus on developing a database in various other subject areas. In our design, some part of Newton telecom dictionary is used and the technical words of computer, web, IT, etc were covered. There is a need for entering words from

other subject areas so that the design becomes comprehensive. Further, we used Google search engine in this study. This method can be evaluated using some other search engines so that the capabilities and efficiency of the best search engine can be found in this field. In addition, the developers of Google search engine can use this method in order to increase the efficiency of users' search results, and make a more powerful database using this idea.

References

[1] Miller A. "About WordNet".WordNet.available at URL : http: //wordnet. princeton.edu/.Accessed sep 16, 2010 .

[2] Kidd E."Visualizing WordNet relationships as graphs". Random Hacks. December 30, 2009.available at URL: http: //www.randomhacks.net/articles /2009/12/29/visualizing -wordnet-relationships-as-graphs. Accessed sep 16, 2010.

[3] Wales J,Sanger L."Welcome to Wikipedia". Wikipedia. Available at URL: http://en.wikipedia.org/wiki/Main_Page.Accessed sep 16, 2010.

[4] Holloway, Todd, Bozicevic, Miran and Borner, Katy. (2007) Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors. Complexity, Special issue on Understanding Complex Systems. Vol. 12(3), pp. 30-40. Also available as cs.IR/0512085.

[5] Espiritu C,Stroulia E,Tirapat T. "Visualizing Wiki Semantics As Topic Maps". Department of Computing Science, University of Alberta, Edmonton, Aberta, Canada. 2006. 8 pages

[6] Gedigian M. Semantic Wiki Visualizaton. U.C. Berkeley School of Information. 2008. 4 pages

[7] Giereth M, Ertl T. Visualization Enhanced Semantic Wikis for Patent Information,Visualization and Interactive Systems Institute. University of Stuttgart. 2008. 6 pages