Sudatta Chowdhury and G.G. Chowdhury
Department of Computer and Information Sciences
University of Strathclyde, Glasgow G1 1XH

Using DDC to create a visual knowledge map as an aid to online information retrieval

Abstract:
Selection of search terms in an online search environment can be facilitated by the visual display of a knowledge map showing the various concepts and their links. This paper reports on a preliminary research aimed at designing a prototype knowledge map using DDC and its visual display. The prototype knowledge map created using the Protégé and TGViz freeware has been demonstrated, and further areas of research in this field are discussed.

1. Introduction

Web search engines and digital libraries usually expect the users to use search terms that most accurately represent their information needs. Finding the most appropriate search terms to represent an information need is an age old problem in information retrieval. Keyword or phrase search may produce good search results as long as the search terms or phrase(s) match those used by the authors and have been chosen for indexing by the concerned information retrieval system. Since this does not always happen, a large number of false drops are produced by information retrieval systems. The retrieval results become worse in very large systems that deal with millions of records, such as the web search engines and digital libraries.

Vocabulary control tools are used to improve the performance of text retrieval systems. Thesauri, the most common type of vocabulary control tool used in information retrieval, appeared in the late fifties, designed for use with the emerging post-coordinate indexing systems of that time. They are used to exert terminology control in indexing, and to aid in searching by allowing the searcher to select appropriate search terms. A large volume of literature exists describing the design features, and experiments with the use, of thesauri in various types of information retrieval systems (see for example, Furnas et.al., 1987; Bates, 1986, 1998; Milstead,1997, and Shiri, Chowdhury and Revie, 2002).

2. Ontology and Bibliographic Classifications Systems

Interests in information organization and vocabulary control have been revived with a new area of study on ontologies. An ontology can be defined as a formal, explicit specification of a shared conceptualization. Gilchrist (2003) notes that Vickery was one of the first in the LIS field to draw attention to the emergence of the term ontology in knowledge engineering and information science. Two of the oldest and most widely-known ontologies are WordNet and CYC. WordNet, a lexical tool, was developed by the Cognitive Science Laboratory at Stanford University, and contains some 100,000 word meanings organized in a taxonomy (Fellbaum, 1998; www.wordnet.com). The CYC ontology (Cycorp, n.d.) provides a foundation for common sense reasoning by developing ontologies for a wide variety of domain-specific applications. A large number of domain-specific ontologies have been built for applications such as machine translation, enterprise modeling, knowledge re-use and information retrieval.

Overviews of some of these projects, and the tools used by them can be found in Ding (2001), Ding and Foo (2002), Fensel (2001), and DIOGENE (http://www.diogene.org).

While many researchers are concentrating on building ontologies in different disciplines, others have concentrated on using the existing knowledge organization and management tools, like classifications and thesauri, instead of trying to re-invent the wheels altogether. Mitchell and Vizine-Goetz (2002) review research on the applications of DDC as a knowledge organization tool. Cochrane and Johnson (1996) have proposed use of the thesaural relationships in Dewey as an aid to bibliographic searching, and Pollitt (1998) has experimented on the derivation and presentation of facets from Dewey as an aid to online searching. Olson and Ward (1998) have experimented with mapping a Women's Thesaurus to the DDC while Koch and Vizine-Goetz (1999) have experimented with the mapping of engineering terminology. Many researchers have used traditional library classification and indexing tools, such as DDC, Library of Congress (LC), Library of Congress Subject Headings (LCSH), etc., to index Internet and digital library resources with a view to producing better retrieval. Typical examples of these approaches include the BUBL Link (http://www.bubl.ac.uk), INFOMINE (http://infomine.ucr.edu/), and Knowledge Class (http://www.uky.edu/~xlin/kclass.html),CyberDewey (http://www.anthus.com/CyberDewey/CyberDewey.html), and CyberStacks (http://www.public.iastate.edu/~CYBERSTACKS/homepage.html).

3. Research Objectives

Visualization techniques have been used by some web search tools, such as Kartoo (http://www.kartoo.com), and Renardus (www.renardus.org).This paper reports on a research that aims to create a visual knowledge map using DDC. The basic hypothesis of this research is that, display concepts on a visual knowledge map will be useful for formulating better queries producing better retrieval. If users can see how their chosen search terms map on the overall knowledge map, they can make up their mind on several things, such as (1) whether the general area of knowledge that the search terms has mapped onto reflects their area of interest, (2) whether or not to probe further on to the mapped knowledge domain to select the specific area of interest, (3) whether or not to select an alternative search term to map onto the right area of interest onto the knowledge map, and so on. This in turn will allow the users to retrieve better results, by selecting the most appropriate search terms, and by selecting a specific domain (if the search term maps onto more than one domain on the knowledge map). Thus, such a system will allow the users to select the most appropriate search terms, and thereby retrieve better results.

4. The Knowledge Map

This paper reports on the first phase of a research that aims to build a simple prototype knowledge map with the nodes representing the major disciplines, and the major topics within the disciplines. A simple ontology development tool called *Protégé* with a plug-in program for visualization, called *TGViz*, has been used to build the prototype. This paper reports on an early stage of this research whereby a prototype knowledge map has been created using some DDC main and sub-classes. The knowledge map can be displayed on the screens showing the various nodes and links. The following sections of the paper provide a brief description of the simple knowledge map and its visual display. The paper then outlines some further areas of work that will eventually lead to the development of an interface to support users in query formulation and modification using a visual knowledge map based on the DDC as the backbone ontology.

5. Protégé

Protégé-2000 is a tool that allows to build an ontology that can be extended and visually displayed on the screen using simple plug-ins. The tool has been developed by Stanford Medical Informatics Group at the Stanford University School of Medicine. The tool and the

associated plug-ins can be downloaded for free from the Protégé website (http://protege.stanford.edu/). The tool comes with documentation and demonstration and a sample knowledge base that can be used for learning purposes. The TGViz plus-in has been developed at the university of Southampton, and can be downloaded from http://www.ecs.soton.ac.uk/~ha/TGVizTab/TGVizTab.htm.

6. The Prototype Knowledge base

A prototype ontology has been created using the main classes, and some selected subclasses of DDC. Although Protégé has a number of advanced features to help users build an ontology, for this prototype, only the simple features supporting the creation of classes and subclasses have been used. The *TGViz* plug-in has been used for the creation of the visual knowledge map. Figures 1 to 4 show the knowledge map – the left side shows the hierarchy while the right side of the screen provides a visual display of the various nodes and their links. It may be noted that the demonstration is not meant to show a complete knowledge map, and therefore should not be looked at to show the complete Dewey Class and subclasses. Although this will eventually happen once a full-blown knowledge map using all the required Dewey classes and subclasses are created. However, the objective of this paper, and the demonstration that follows, is to show how DDC can be used to build and visually display an ontology and a knowledge map which in turn can help the users formulate and expand queries in an online search session.

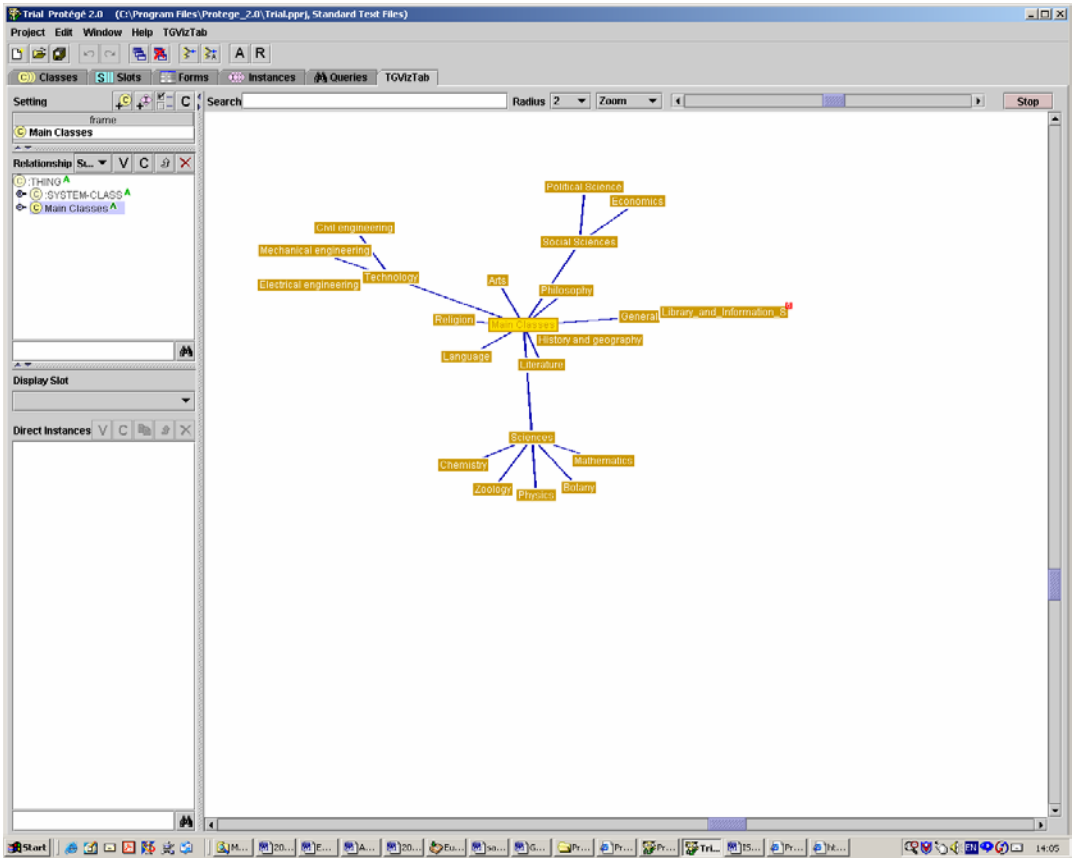

Figure 1: Visual map of the Main Classes

Figure 1 shows a demonstration of the main classes and their subclasses, etc. The user can use the hierarchy of classes and subclasses that appears at the left side of the screen to display more subclasses, or can double click on any node in the visual map whereby the map will change and to display the links from the chosen node. For example a double-click on the node 'Library and Information Science' will display a visual map shown in Figure 2 and a double-

click on Digital libraries from Figure 2 will display a map shown in Figure 3, and a double click on the node 'Digital library types' will display the map shown in Figure 4.

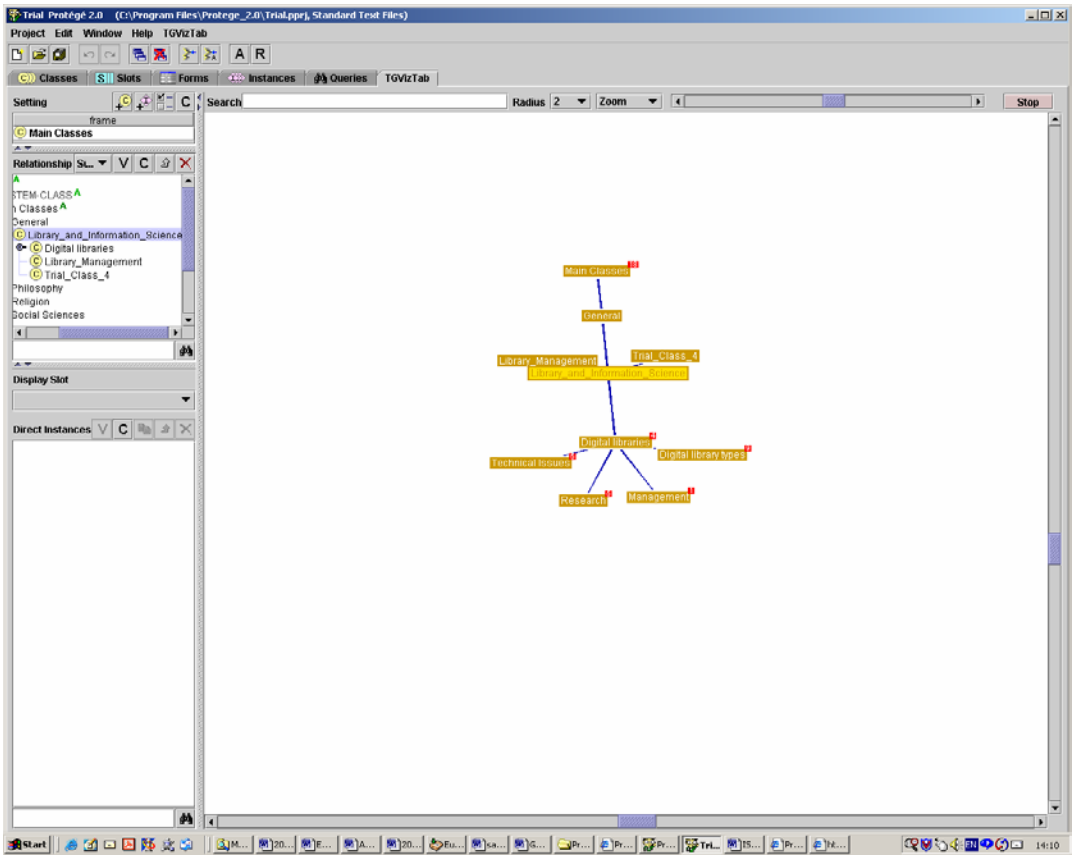

Figure 2: Subclasses of 'Library and Information Science' node

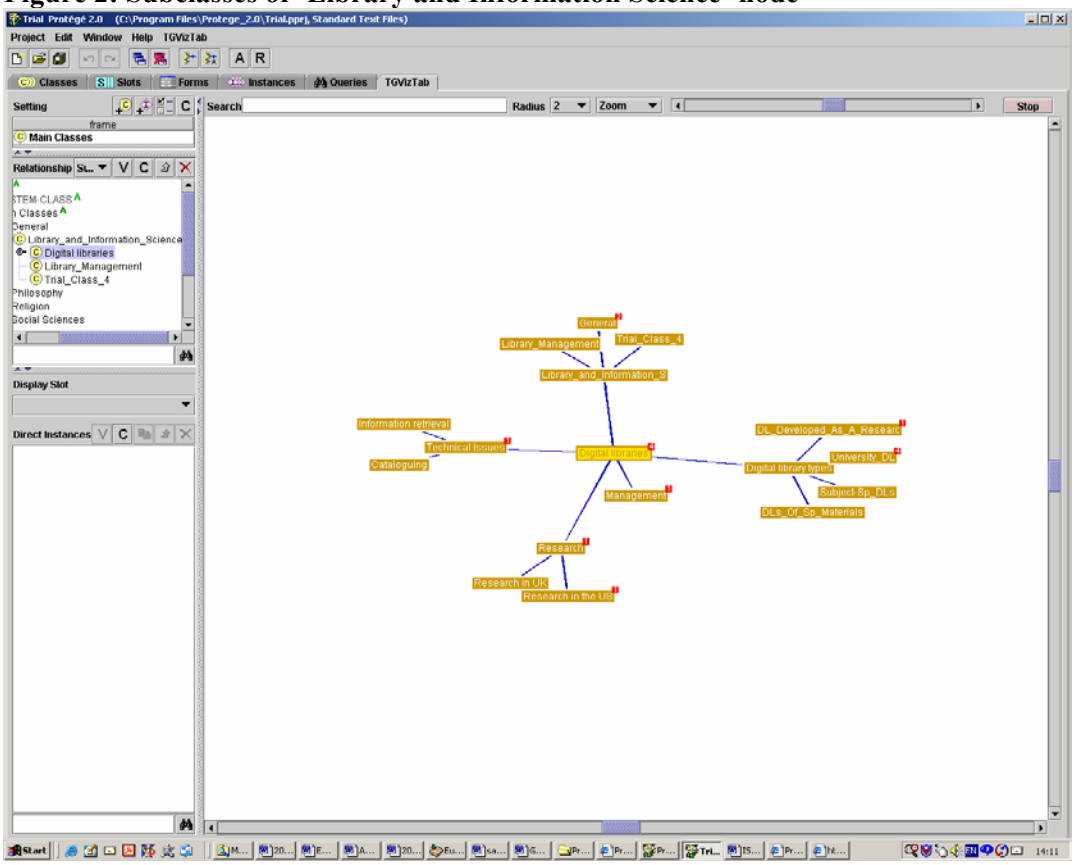

Figure 3: A visual map of the topic 'Digital Libraries'

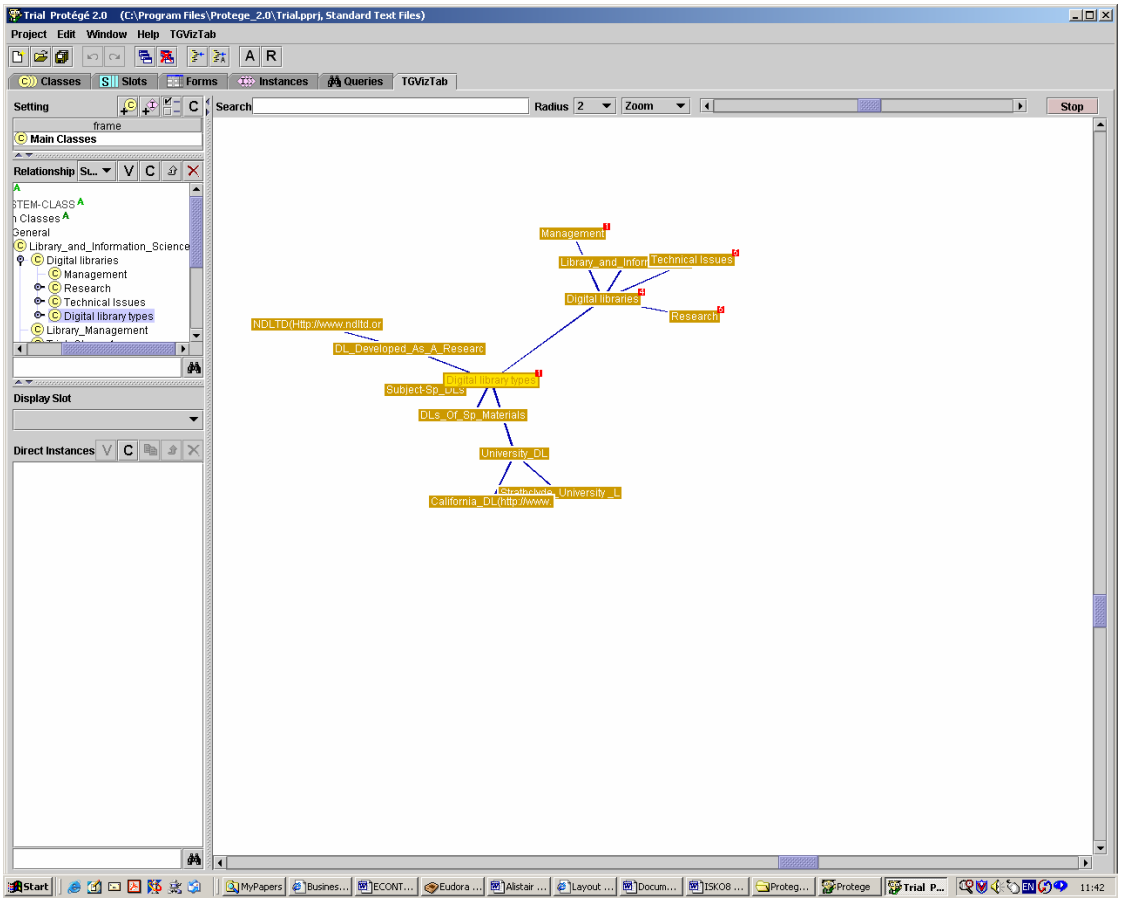

Figure 4: Sub-classes of Digital Library Types

7. Future Work

The preliminary work reported in this paper shows that it is feasible to build and display a knowledge map using DDC as the backbone. The next step will be to interface this with web search engines and digital libraries thereby allowing the users to use the visual knowledge map to formulate and refine queries and conduct a web search. Of course the knowledge map displayed here is very simple; more work is needed to build a complex ontology structure with appropriate links among the various concepts. Once the ontology and the interface is complete, the next phase should be the evaluation where end-users should be involved to conduct search for electronic information using the visual knowledge map, and to comment on its usefulness. Provided appropriate funding support is available, it is expected that DDC will have a new role to play in the web and digital library information access systems.

References

Bates, M.J. (1986) Subject access in online catalogs: a design model, Journal of the American Society for Information Science, 37 (6): 357-376.

Bates, M.J. (1998). Indexing and access for digital libraries and the Internet: human, database, and domain factors, Journal of the American Society for Information Science 49(13):1185-1205.

Cochrane, Pauline, and Johnson, Eric (1996). Visual Dewey: DDC in a Hypertextual Browser for the Library User. In Knowledge Organization and Change: Proceedings of the 4th International ISKO Conference, edited by R. Green. Frankfurt/Main: INDEKS Verlag: 95-106.

Cycorp (n.d.). [http:// www.cyc.com] Accessed 19/02/2004

Fellbaum, C.(ed.) (1998). WordNet an electronic lexical database. Cambridge, Mass: MIT Press.

Fensel, D. (2001). Ontologies: a silver bullet for knowledge management and electric commerce. Berlin: Springer.

Furnas, G.W., Landauer, T.K., Gomez, L.M. and Dumais, S.T. (1987). The vocabulary problem in human-system communication, Communications of ACM 30(11): 964-971.

Gilchrist, A. (2003). Thesauri, taxonomies and ontologies – an etymological note. Journal of Documentation, 59(1): 7-18.

Gruber, Tom. What is an ontology? [www-ksl.Stanford.edu/kst/what-is-an-ontology.html] Accessed 19/02/2004

Vizine-Goetz, Diane and Koch, T. (1999). Automatic Classification and Content Navigation Support for Web Services: DESIRE II Cooperates with OCLC. Annual Review of OCLC research, 1998. [http://www.oclc.org/research/publications/1990-1999.htm]. Accessed 19/02/2004

Milstead, J. L. (1997). Use of thesauri in the full-text environment. In: Pauline A. Cochrane and Eric H. Johnson (eds.) Visualizing Subject Access for 21st Century Information Resources: Proceedings of the 34th Annual Clinic on Library Applications of Data Processing, Graduate School of Library and Information Science, University of Illinois.

Mitchell, Joan, S. and Vizine-Goetz, D (2002). A Research Agenda for Classification, 2002. [http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000003953]. Accessed 19/02/2004

Olson, Hope. A. and Ward, Dennis. B. (1998) "Charting a Journey across Knowledge Domains: Feminism in the Dewey Decimal Classification." In Structure and Relations in Knowledge Organization: Proceedings of the 5th International ISKO Conference, edited by W. Mustafa el-Hadi, J. Maniez, & A.S. Pollitt, 238-244. Würzburg: Ergon Verlag.

Pollitt, A. Steven (1998) "The Application of Dewey Decimal Classification in a View-Based Searching OPAC." In Structure and Relations in Knowledge Organization: Proceedings of the 5th International ISKO Conference, edited by W. Mustafa el-Hadi, J. Maniez, & A. S. Pollitt, 176-183. Würzburg: Ergon Verlag.

Shiri, Ali.A., Revie, C. and Chowdhury, G. (2002). Thesaurus-assisted search term selection and query expansion: a review of user-centred studies. Knowledge organization. 29(1): 1-19.

Ying, Ding and Foo, S. (2002). Ontology research and development. Part 1 – a review of ontology generation. Journal of Information Science, 28(2): 123 – 136.