THE ROLE OF ADVANCED VR INTERFACES IN KNOWLEDGE MANAGEMENT AND THEIR RELEVANCE TO CAD

AHMAD RAFI, M.E.
Multimedia University
Malaysia
ahmadrafi.eshaq@mmu.edu.my

AND

PETER KARBOULONIS
Firesoft Technology Ltd.
United Kingdom
karboulonis@clara.co.uk

Abstract. This paper introduces knowledge management and computer aided visualisation as a key in establishing both valuation and value creation capabilities in the enterprise where dissemination of knowledge and effective sharing of information through collaboration spur creativity and stimulate business practices. The paper draws an original approach for the design and development of a universal information/knowledge visualisation tool and outlines the mechanics that enable the working prototype that focuses on CAD.

1. Introduction

Rapid progress in computer networks and a growing access to computers enable the emergence of new forms of highly dynamic organisations where dissemination of knowledge and effective sharing of information through collaboration spur creativity and stimulate business practices. More traditional horizontal enterprise infrastructures are strongly constrained and not well suited to support collaboration and the dissemination of knowledge unlike recent vertical business models.

2. Knowledge Management

In the new economy knowledge**,** assets and intellectual properties are expected to increasingly play a dominant role in establishing both valuation and value creation capabilities in such enterprises. Companies such as British Aerospace, Microsoft, Oracle and Autonomy are testament to that. Czerniawska and Potter (2001) examined the law of diminishing returns that observes that when new inputs are added to a process, outputs will not be increased by a similar amount. However in particular situations the law of increasing returns is present where the more goods are produced, the lower the costs or time to market. This discovery is credited to Theodore P. Wright, an engineer, who in 1922 calculated that the labour required in assembling a plane declined by 20 per cent with each doubling of production experience. The introduction of the learning curve and its widespread acceptance in business in the 1960s when applied by the Boston Consulting Group on a number of varied businesses showed the same pattern, costs fell as experience rose. Czerniawska and Potter further argue that understanding why this concept works is through the exploitation of information, i.e. the knowledge of the production process.

2.1 CASE STUDIES

In May 1997 British Aerospace (BAE) created the Virtual University (Heisig and Vorbeck, 2001) when it was decided that the company's growth depended on knowledge and know-how. When considering a high technology company such as BAE is easy to see that the company's knowledge and innovation is embedded within their products and customer services. Intelligent research and retrieval techniques based on Autonomy technologies (www.autonomy.com) were developed to provide users with relevant and timely information especially when considering the company's geographical spread. It is worth noting that different organisations and businesses have different needs when considering knowledge management solutions. Also of interest is IBM's effort in acquiring, validating and distributing elicited knowledge to its so-called Community of Practice, a cross-organisational network with over 4,000 members. IBM's interest is in the effective distribution of information, and tacit and explicit knowledge, and experiences and ideas that are structured in such a manner as to encourage exchange and re-use.

IBM's ICM knowledge management software tool supports the transformation of individuals associated with certain phrases, key words, skills, etc. This process is manual and costly, as it requires the time of a human expert in terms of reading contributions, assisting the contributors as well as servicing the database. Non-the-less of great value as the knowledge collected is of high quality. Alternative solutions exist that allow knowledge management systems to automatically identify valuable and relevant knowledge. Such a system is Casmir's Learning Engine (CLE) a system developed round proprietary neural network technologies (www.casmir.net).

2.2 HANDLING AND VISUALISING KNOWLEDGE

Knowledge management is concerned with the identification, encapsulation, and redeployment of knowledge at the workshop, operational, and designer level. The process requires for the design and development of knowledge models, knowledge representations and best practices. Support tools need to be developed and deployed to assist in the search for knowledge, its visualisation and effective navigation in an effort to reduce the number of redundant searches. The overall goal is to make users of such systems more effective in their information management and communication tasks where learning times are reduced, knowledge is retained, and personal satisfaction and performance are increased thus benefiting the enterprise.

According to Kluge et al. (2001), knowledge is the understanding of relations and casualties, and is therefore essential in making operations effective, building business processes, or predicting the outcomes of business models. They also differentiate between explicit knowledge (knowledge that can be structured and documented) and tacit knowledge (knowledge linked to human senses and experience). They do however acknowledge that the two categories are heavily interlinked and near impossible to separate in practice as for example the comprehension of a written document (explicit knowledge) may require a degree of experience (tacit knowledge). They finally perceive knowledge in terms of relationships rather than databases. Thus the accumulation and storing of information coupled with an ability to create, effectively visualise and identify relationships between these items of information is considered a vital part in the success of a knowledge management software system regardless of the of the industry that it serves.

3. Information Visualisation

According to Tufte (1994), the world is complex, dynamic, multidimensional; the paper is static and flat. He also poses the question "How are we to represent the rich and visual world of experience and measurement on mere flatland?" Representations of data beyond the textual can be employed to make information more accessible and to detect patterns and relationships as in C. J. Minard’s (1781-1870) graphical narrative of Napoleon’s army and its terrible fate in Russia. Different variables are plotted on a two-dimensional surface, army size, direction of movement, temperature, location, date and battle events. Researchers at Carnegie Mellon University employed SAGE (Chuah et al., 1995) an information visualisation system, using Minard’s original data. Their efforts revealed details and relationships beyond Minard’s original graphic showing that information visualisation is an effective tool for conveying complex ideas.

The rapid advancement and adoption of information visualisation in different areas is currently challenging system designers to find out strategies and tools to effectively discover, visualise and share particular types of information. Most common approaches include cone trees and tree structures as well new alternatives such as networked structures, hyperbolic views, trajectory and feature maps. All these concepts show marked diversity and tend to be used depending on the type of data and user requirements (Chen, 1999). Undirected graphs were chosen to represent the stored knowledge and its relationships despite the challenges they pose. The most popular algorithm for drawing undirected graphs was introduced by Eades (1984). It is referred to as the spring-embedder model and follows two aesthetic criteria: uniform edge lengths and symmetry. Subsequent research includes works from Kamada and Kawai (1989) Frutcherman and Reingold (1991) and Davidson and Harel (1996).

4. A Universal Tool for the Dissemination of Knowledge

The most common type of digitally stored information and knowledge exists in textual form where graphical and other representations are usually associated with sections of text. Current search engines are considered to be analytical, methodical, serial technologies suited to the activity of the left hemisphere of the brain where shapes and patterns that would tap in the right hemisphere have generally been neglected. It is suggested that by investigating the semantics of data links**,** classifications can be created regardless of physical boundaries. In turn these semantic connections can be given graphical form presented through a computer-based interactive graphic that improves cognition, reveals hidden relationships and enhances navigation.

4.1 A KNOWLEDGE MANAGEMENT TOOL FOR CAD

The project aimed to create a software tool that would assist users of CAD systems and students during the design process by taking advantage of accumulated knowledge. Useful Information stored within the databases would consist of text, graphics, video-clips, pointers to executable procedures and valuable external resources including references and sources on the Internet. It was important that a general structure and sets of relationships in terms of links were defined between the various imported elements stored in the databases, this was essential for the delivery of a useful and coherent system that was easy to use and navigate. For example, if a CAD user initiated a query about constructing a cog-wheel, the system was expected to automatically produce the required information consisting of a series of operations that would lead to the construction of a cog-wheel as well as other relevant material. This would point experienced and non-experienced users to information that would satisfy their query with an option to investigate further depending on a user's further needs.

Thus the need of being able to update the system and service the links effectively and within a reasonable amount of time would add to the usefulness and long life of the system. As new information was added and old information was deleted the prospect of redundant or broken links combined with the considerable effort required to link new information to old records became a main concern. The obvious problem was related to swiftly and effectively navigating the extracted knowledge and updating and servicing the proposed databases. Thus a system was designed to provide for the automatic generation of relationships between elements of information stored within the databases.

4.2 THE PROTOTYPE SYSTEM

A prototype system was designed consisting of a database, a knowledge-retrieving mechanism and a dynamic relationship generator. A simple text-based interface was also provided to test the system. To test the system**,** the glossary pages 122-125 from D. J. Bowman's book, The CAD/CAM Primer (8,000 words) was used as a pilot text for applying the knowledge retrieval algorithms. One hundred and thirty glossary terms and the text accompanying them were used to create eighty-eight database records containing this information. To increase efficiency and assist the knowledge retrieving mechanisms**,** a mirror database was created where words within records were first tokenised. Subsequently their inverse frequencies (1/f and $1/f^2$) were calculated and stored in the mirror database. Tokens with high frequencies were not included in the new database as their presence adversely affects knowledge extraction operations. Suppose that k (where k=14) filtered tokens were found within a knowledge base of n records (where n=8). Vectors ($SVector_n$) showing tokens present in each record are first generated for all records and stored in a database, i.e.

SV_1 = (0,0,0,1,0,0,1,0,0,0,0,0,1,1) … SV_9 = (1,0,0,0,1,0,0,1,1,0,0,0,0,1) → where 1 means a token is present, 0 when not

The SVector identifies content and requires minimum computation. Subsequently it can be used for *screening* purposes as it facilitates fast movement between relevant concepts stored within the knowledge base. A different vector (IdVector) that introduces context uses frequency to achieve more accurate results, i.e.

$IdVector_n$ = ((ts_1 * $idvt_1$), (ts_1 * $idvt_2$), … , (ts_{12} * $idvt_{12}$)), where: ts = attribute state (1 present, 0 not present),

idvt = identification value = (1/frequency) or (1/$frequency^2$)

The IdVector is used to calculate the knowledge weighing value of a record depending on filtered tokens being present, i.e.

$WRecord_n$ = (ts_1 * $idvt_1$) + (ts_1 * $idvt_2$) + …+ (ts_{12} * $idvt_{12}$)

Suppose the user made a query that when tokenised and filtered produced the following vector:

SV_0 = (1,1,1,0,0,1,0,1,1,1,0,1,0,0) → query vector

$Svectror_1$, the user query vector, is logically ANDed with SVector 1 – 9 for possible matches, i.e.

$SVector_0$ AND $SVector_1$ → not applicable

$SVector_0$ AND $SVector_2$ = (10100100100100)→ matches = 5, etc.

Records that have shown matches are ranked in terms of relevance dependent on the number of matches, in this case the *stack* looks like: 2, 4, 7, 5, 8, 9, 3, 6. However this simplistic mechanism does not work well and produces poor results, thus weighing factors need to be calculated for each of the records in the database in relation to the user query vector $SVector_0$. Subjective tests on users have shown that records with higher weighing values are most relevant to the user query and generally satisfy them. Calculating weights for all Sector quantities excluding $SVector_1$ that scored 0 matches are as follows:

$WRecord_2$ = 752.24, $WRecord_3$ = 427.70, $WRecord_4$ = 838.36, etc.

The new stack now looks as follows: 9, 4, 7, 2, 5, 3, 8, and 6. The new records will satisfy user queries a great degree better. When considering systems containing great numbers of records a further requirement arises. Assuming the system returned a finite number of records satisfying the user query how can the system ensure that these records do not to a great degree duplicate each other and that records with lower weighing scores are not excluded as they may contain useful information not present in the records high on the stack? Using the previous example let us assume that the system always presented the user with the top three records after a query was made, i.e. 9, 4, 7. The statement below refers to aspects of knowledge that the three stacked records, 9, 4 and 7 would cover if they were all read by the user, i.e. $SVector_9$ OR $SVector_4$ OR $SVector_7$ = (1,1,0,0,1,1,1,1,1,1,0,1,0,1)

The next statement refers to the common aspects covered by the three records on top of the stack compared to those found in the user query.

$SVector_9$ OR $SVector_4$ OR $SVector_7$ AND SVector0= (1,1,0,0,0,1,0,1,1,1,0,1,0,0)

Seven matches exist out of possible eight. A high score means that in terms of content records put forward by the system cover most topics within the query. A low score means that different records should be considered by the system lower on the stack. The above mechanisms also allow for the automatic creation of relationships between records and enable quick navigation between comparable concepts stored within the databases. To create such links dynamically all or subsets of selected records need to be considered against each other using the previous methods. The second phase of this research involves the development of a system that imports a database of over 12,000 records containing text (100-1,000 words per record), graphics, video clips, animations and other data items. A graphical interface based on unstructured graph technology will be introduced to enable fast and effective navigation.

5. Conclusion

This paper has outlined an original approach for the design and development of a universal information/knowledge visualisation tool and has outlined the mechanics that enable the working prototype. Of greater value is the ability of the system to automatically perform effective searches, identify and elicit knowledge, and automatically draw associations between concepts stored within unstructured databases. Its dynamic nature allows for the editing of the underlying databases with minimum effort without any concern of broken or redundant links as they are dynamically allocated by the system.

References

Chen, C.: 1999, *Information Visualisation and Virtual Environments*, Springer.

Chuah, M. C., Roth, S. F., Kolojejchick, J., Mattis, J. and Juarez, O.: 1995, SageBook: Searching data-graphics by content. *Proceedings of CHI* , ACM Press.

Czerniawska, F. and Potter, G.: 2001, *Business in a Virtual World*, Ichor Business Books (Purdue University Press).

Davidson, R. and Harel D.: 1996, Drawing Graphs nicely Using Simulated Annealing. *ACM Transactions on Graphics*, 15(4):301-331

Eades. P.: 1984, A Heuristic for Graph Drawing. *Congressus Numeratium*, 42:149-160.

Frutcherman, T. M. J. and Reingold, E. M.: 1991, Graph Drawing by Force Directed Placement. *Software – Practice and Experience*, 21:1129-1164.

Heisig, P. and Vorbeck, J.: 2001, *Knowledge Management: Best Practices in Europe*. Mertins, K., Heisig, P., Vorbeck, J. (eds), Springer.

Kamada, T. and Kawai, S.: 1989, An Algorithm for Drawing General Undirected Graphs. *Information Processing Letters*, 31(1):7-15.

Kluge, J., Stein, W. and Licht, T.: 2001, *Knowledge Unplugged*, Palgrave.

Tufte, E. R.: 1994, *Envisioning Information*, Graphics Press.

Tufte, E. R.: 1995, *The Visual Display of Quantitative Information*, Graphics Press.