A Hypertext Development Methodology

Paul J.A. van Vliet
Oklahoma State University

Rick L. Wilson
Oklahoma State University

Hypertext has been alternatively defined as non-sequential reading and writing and as an associative database in which nodes containing information are interconnected by links. Users are able to move from node to node, using the links, in order to build a trail of interest through the available materials. Hypertext applications have already appeared in the areas of end-user support and multimedia, despite a lack of hypertext implementation standards.

So far, few database researchers have systematically assessed the process of hypertext development. Yet a development methodology is needed if hypertext is to take its place among other database architectures as an effective means of information storage and retrieval. This paper proposes a systematic hypertext development methodology based on insights gained in a number of hypertext development projects. This proposed methodology addresses the desired objectives of a hypertext system development approach.

Hypertext is an important and powerful information representation technology. As a unique way of storing and retrieving information, it has become a popular research topic in the area of database management. There have been a number of articles and books which have attempted to present a concise, somewhat universal view of hypertext systems. However, there have been few efforts to specify concrete system development procedures for building hypertext database systems. Since hypertext systems represent a significant departure from traditional database systems, it is apparent that established techniques may not be applicable. Thus, this paper is an initial step toward filling the need for determining a proper and sound development methodology for hypertext databases.

This paper will first provide a brief overview of hypertext concepts, followed by the identification of the necessary factors and design objectives of a unique hypertext development procedure. Next, the proposed methodology for developing hypertext databases will be presented. While the current state of hypertext research and technology is not yet at the stage in which a complete and prescriptive contingency model for hypertext development can be formulated, a number of hypertext design projects have resulted in insights and guidelines which a developer can use to build new hypertext databases, either using newly collected information or through the conversion of existing materials into a hypertext format. Finally, the paper will summarize and present potential productive research areas related to the development methodology.

The Hypertext Model

The Origins of Hypertext

In 1945 Vannevar Bush wrote the article “As We May Think,” in which he argued that the cumulative mass of knowledge - particularly scientific knowledge - had grown so large that existing methods of information storage and retrieval were no longer adequate to effectively access needed information. Bush proposed a device he called the “Memex,” in which this information would be indexed and cross-referenced so that when consulting a particular item, the user would be able to find related items, irrespective of the information’s
original location. In moving between related items, the user could then build a “trail of his interest through the maze of materials available.” (Bush, 1945, p. 107) While Bush’s idea of non-sequential and associative information access was positively received, the Memex proved to be technologically infeasible at the time. It took the development of digital computers and graphical user interfaces to realize some of the concepts of the Memex. (Conklin, 1987; Nyce & Kahn, 1989)

Ted Nelson, in 1965, conceived of the word “hypertext” to describe an all-encompassing information-collection consisting of interconnected documents. (Nelson, 1982) While his visionary project was never fully realized, the hypertext idea has been influential in the development of subsequent databases based on the idea of interconnected pieces of information, and the term hypertext has been generally accepted to denote such databases. (Conklin, 1987)

Hypertext databases remained scarce until the 1987 release of the HyperCard program by Apple Computer Inc. (Parsaye et al., 1989, pp. 232) This program provided hypertext capabilities to many microcomputer users and greatly increased the popularity of the hypertext concept. In subsequent years, other hypertext applications for microcomputers have become available, each with its own particular strengths, which has further increased the use of hypertext. (Begoray, 1990; Byers, 1987; Conklin, 1987)

Hypertext Overview

When Nelson originally coined the term hypertext, he defined it as non-sequential reading and writing. (Nelson, 1982) While most paper documents are still sequential, non-sequentiality is common in many types of writing. Common examples of non-sequential books are dictionaries and reference manuals. (Smith & Weiss, 1988) These books are usually not read sequentially, as readers will locate entries or topics of interest and read only these sections sequentially.

Since its inception, hypertext has developed into a database architecture which allows users to interconnect pieces of information of many types, not just plain text. (Parsaye et al., 1989, pp. 224; Smith, 1988) These pieces of information are linked together in either a hierarchical or network fashion, or in some hybrid combination. (Begoray, 1990) Because such hypertext structures can quickly become large and quite complex, a computer is needed to access and navigate through a hypertext database.

A hypertext database represents a radical departure from traditional database architectures. The hierarchical, network, and relational database architectures are essentially all record-oriented architectures, with data partitioned into files, records, and fixed-size fields. (Date, 1981, pp. 63) The hypertext approach is more free-form. Rather than splitting up the attributes of an entity into fields, the information is subdivided into semantic units, also known as chunks or nodes.

Usually, a node represents a single concept or idea. This requires that the distinction between nodes be explicit and well-defined. (Raymond & Tompa, 1988) The amount of information contained within each semantic unit, and/or its structure, may vary from node to node. In hypertext, the data shapes the database, and not vice versa. (Brockmann et al., 1989) The information contained within a node can typically be displayed on a single computer screen. (Conklin, 1987)

When more space is needed, some programs allow for the creation of longer nodes that scroll up from the bottom of the screen. (Fiderio, 1988)

Hypertext nodes are not connected by foreign keys or in fixed hierarchical or network structures, but by connections known as links. In hypertext applications, links allow a reader to move from one node to another. (Horn, 1989, pp. 8) There are essentially two types of links: organizational and referential. (Conklin, 1987; Frisse, 1988) Organizational links structure the information contained in such a manner so it can be systematically traversed. Most organizational links are defined by the developers of the original database. Users rarely add or alter organizational links.

On the other hand, referential links connect nodes which are semantically related in a way other than purely structural, such as a text and its explanatory footnotes. Some referential links are usually placed by developers. Later, more links may be added by the database users. (Conklin, 1987; Horn, 1989, pp. 27)

In a hypertext database, the collection of all nodes and the links that connect them is referred to as a hypertext graph. A graph can be visually represented as a network. (Begoray, 1990) Quite often, a hypertext graph will contain more information than a particular user wants or needs to know. The user will then limit the information search to a subset of the graph, relating to a sub-topic of the overall information collection. This subset is called a web (or context). Within a graph, webs identify subtopics and connections between webs indicate how the subtopics interrelate within the scope of the overall topic. Individual nodes may belong to several webs. (Begoray, 1990) As a user moves through a hypertext database from node to node, the path followed forms a linear sequence of nodes which make up the user’s “trail of interest.” (Bush, 1945) This trail is commonly referred to as a path. When the user moves through a hypertext database several times, he or she may follow different paths and, in the process, obtain different information.

A hypertext database can be consulted in three ways. (Conklin, 1987) The first method of data retrieval is to follow links from node to node, navigating through the database, and building an information trail of associated materials. Related to this method, a hypertext database can be consulted by navigating around the database using a graphical representation of the database structure called a “browser,” which permits the user to take the structure of the database into account when traversing it. The third consultation method is to search the database for a particular keyword, and examining only those nodes which contain the keyword.
Related Content

Semantic Integrity Constraint Checking for Multiple XML Databases
www.igi-global.com/article/semantic-integrity-constraint-checking-multiple/3360?camid=4v1a

INDUSTRY AND PRACTICE: What’s New? The Challenges of Emerging Information Technologies
www.igi-global.com/article/industry-practice-new-challenges-emerging/51191?camid=4v1a

Learning Classifiers from Distributed Data Sources
www.igi-global.com/chapter/learning-classifiers-distributed-data-sources/20744?camid=4v1a

The Algos Center: Information Systems in a Small Non-Profit Organization
www.igi-global.com/chapter/algos-center-information-systems-small/6204?camid=4v1a