A Navigation System for Personalized Databases “StarMap”

Steven H. McCown  
Storage Technology Corporation  
steven_mccown@storagetek.com

Charles Milligan  
Storage Technology Corporation  
charles_milligan@storagetek.com

Abstract

In light of the proliferation of the World Wide Web (web), the future of large database management lies in addressing the problems of sifting through structured data sources (e.g., data warehouses) with the same tools that are used for sifting through unstructured sources (e.g., web). Doing so will enable faster and more specific decision making as well as facilitate reusable decision support components (e.g., a transferable set of decision support rules). The purpose of this paper is to describe a method called ‘StarMap’ by which users can intelligently and automatically navigate and sift through acquired data and identify information content. StarMap provides a highly customizable, graphical, user-view into acquired data that is integral to the concept of “Personalization of Databases.” StarMap provides individual users with the ability to navigate both large organized databases and other unorganized data sources using a tool that is as familiar as a favorite web browser.

1. Introduction

The purpose of data warehouses, databases, websites, and even files stored on a personal hard disk is to enable an individual to make a decision by providing that specific Decision Maker with immediate access to data and information. Due to the ever-increasing amount of such resources and the time required for humans to process them, automated decision support systems have been created.

Decision Maker’s are supported by a variety of decision support systems, expert systems, and private processes, which have generally been developed around special languages (e.g., SQL) for sifting and manipulating data and constructing queries. Generally, these processes focus on pulling together a subset of data and recommending a decision rather than allowing the Decision Maker to personally navigate through the data or recommendations and thereby determine a course of action. Therefore, the usability of such varied collections of data has been limited by the sophistication of the decision support systems. Two processes have significantly affected decision support. The first is that large collections of structured data are constantly being distilled into subsets (e.g., data warehouses into smaller more easily managed units known as ‘data-marts’). The second is that the availability of uncontrolled and unstructured data is growing at a much faster rate than structured and managed data.

In light of the proliferation of the World Wide Web (web), the future of large database management lies in addressing the problems of sifting through structured data sources (e.g., data warehouses) with the same tools that are used for sifting through unstructured sources (e.g., web). Doing so will enable faster and more specific decision making as well as facilitate reusable decision support components (e.g., a transferable set of decision support rules). As individual Decision Maker’s peruse their broadening fields of available data, they inevitably build personal caches of the results of inquiry and query searches.

The purpose of this paper is to describe a method called ‘StarMap’ by which users can intelligently and automatically collect metadata about the acquired data and navigate and sift through such metadata and identify information content. StarMap provides a highly customizable, graphical, user-view into acquired data that is integral to the concept of “Personalization of Databases.” StarMap provides individual users with the ability to navigate both large organized databases and other unorganized data sources using a tool that is as familiar as a favorite web browser.

The underlying work for StarMap is rooted in a research project conducted as part of an effort to empower large database systems users and is intended to be communicated as a matter of form and function rather than describing a completed commercial product. The StarMap approach allows users to manage and navigate (e.g., browse) their own databases without learning SQL, overtly generating tables, or doing little more than point-and-click to investigate new items. With server-side components, content creators may make their own StarMaps available for browsing by other users. Our vision for the future of data management is to empower individual decision-makers by enabling them to:
• Augment subsets of data with highly personalized metadata that allows extremely efficient navigation and information display;
• Invoke navigation aids using demonstrated web-based interaction tools; and,
• Conveniently share instances of these personalized StarMaps.

2. The Web – An analog for the Database

The Web has emerged as the world’s largest, but completely unmanageable database. It embodies an enormous number of data sources that can provide information to facilitate online research, publishing, e-commerce, artistic expression, recreation, etc. Although the Web is a tremendous step forward in the area of data availability, it has also highlighted the next set of database/user interface problems that must be solved. Some of these relate to current database and data warehouse usability, but some are new problems introduced by the Web. These are problems with managing the data as well as with interfacing to the data. This paper will treat the subject in the context of interfacing to the data (to review these problems in the context of managing the data please contact the authors for a copy of their paper ‘Personalization of Databases: “Pocket Kiosk”’). The problems facing StarMap can be summarized in the following classes:

• Ease of Inspection – Since contemporary Decision Maker’s are trying to deal with information overload, they demand that their support tools personalize the presentation to fit individual requirements. They need the ability to organize data in a form that allows convenient review. They also want a convenient set of tools, similar to what they now use to ‘surf’ the Web. This must allow them the ability to literally see the relationships in the data via graphical aids and to create and view new relationships on the fly.

• Amalgamation – Data is required from a wide spectrum of sources far beyond the domain of the enterprise control. Although different processing mechanisms are required for detailed analysis of each set of data, the navigation tools must be independent of these variations and present a single image to the user.

• Data Analysis – Decision Maker’s need the ability to modify the processes that are used for analyzing the data, and to do so in real time including the further refining of metadata at the moment of critical decision making. This must allow them to personally query the metadata using newly defined queries without relying on database experts. The information content of the data must be made navigable without the need to re-read the actual data.

• Personal Portability – Most decision makers have become quite mobile to the point of being nomadic and demand the ability to carry their decision support tools along with them (e.g., via a laptop or a PDA). They also require the ability to share those tools, supporting criteria, and the resultant decisions with others or to secure them from prying eyes and to verify that they have not been changed.

Most company’s now appreciate the existence of the Web as an extension to their network of data databases and sources. The databases available on the corporate network are often homogenous (e.g., Oracle) and searched via a common language (e.g., SQL). The Web provides an extremely large extension of disjoint sources that include large data warehouses, data marts, individual web pages, e-mail from other users, and newsgroup postings. While potentially useful information is available from each of these sources, the electronic formats may be vastly different. HTML files, text files, word-processing documents, postscript files, graphics files, audio and video clips, and Adobe .pdf files are examples of common electronic formats. Several commercial products attempt to provide solutions, but these generally relate only to web browsing or to a very specific application. They all have difficulty handling a large collection of files, or files of varying types (e.g., HTML, text, executables, and data from queries). The primary concern is to allow the user the flexibility of managing all these different types of data in a unified system that still retains the ability to be personalized by the individual user.

The most popular method of finding data content on the web is via query engines. According to Stox Broadcast Corp. (http://www.howestreet.com/isdn/990112/990112.113926.WCTI.html), “searching is the most popular activity on the Internet”. This is evident by the constant arrival of new web-based search engines. Another common search technique is to follow links from data to data. As with database queries, when desired web content is found the normal data accumulation process includes saving it to a folder on a hard disk, saving a reference (e.g., hyperlink) by using the browser’s ‘Favorites’ menu, or even printing the information. StarMap represents a new approach to saving the indexing metadata about a data source. StarMap also represents a new approach to displaying metadata and the results of additional user queries and provides a user-personalized view of this information. In the end, the only view of the world of available information that really matters is that of the individual user.

3. Characteristics of a StarMap

3.1. Ease of Inspection

Since contemporary Decision Maker’s are trying to deal with information overload, they demand that their
support tools personalize the presentation to fit individual requirements. The semantics of a StarMap allow a user to customize the different data types available for data collection and the different defined relationships that could exist between nodes of data. A node of data is that set of data collected at one time such as the contents of a specific web page or the capture of a document, file, or executable. The StarMap is not concerned with the actual capture of the data, that is a function that a companion process such as a database or data warehousing installation would provide. StarMap is concerned with access to the various types of metadata that are generated by the capture process and the presentation of this metadata to the user.

The user has full flexibility to assign specific presentation characteristics for the entire range of data and information acquired during the data capture process. The StarMap software identifies the relationship(s) between any two individual nodes of data by drawing a line between them. The properties of a line (e.g., width, style, color, arrowheads, and other special characteristics) have default values provided by the StarMap, but can be easily customized according to individual user preferences. The type of data, the context in which the data is captured, or the contents of the data node will generally be of special interest to the user.

**Figure 1. Personalized StarMap Semantics**

Each data node is assigned an icon whose properties (e.g., style, color, size, graphic, and any internal notations) may also be customized. Nodes may represent the sets of data captured from or identified in a database via a series of database queries. They may also denote a set of hyperlinked web pages, a set of captured documents or files sharing a specific context, or a set of multimedia clips on related topics. The links represent automatically identified relationships among the nodes of data. Figure 1 shows one example of user personalization of the StarMap presentation. When node and relationship information are combined in a particular StarMap rendering, they might appear as depicted in the Figure 2.

**Figure 2. Use of Semantics in a StarMap**

Decision Makers insist on a convenient set of tools, similar to what they now use to ‘surf’ the Web. One of the strengths of StarMap is that it can be integrated into the operation of any data acquisition process and combined with the browsing capabilities of the user’s system. The logic flow shown below in Figure 3 could typify how one might begin a data acquisition session with a StarMap and an office document system. Figure 4 shows an example of StarMap working within the user’s browser or document display application. The StarMap display can be subordinate to the application display and called out only by specific request. However, it can also share the window with the application and will maintain the current data node being accessed by the application in the window it is allocated. This will generally cause the bulk of a complex StarMap to be scrolled out of the window. When requested, the StarMap will use the entire application window as shown in Figure 5.
Users need the ability to organize the metadata in a presentation format that allows convenient review. They want the ability to literally see the relationships in the data via graphical aids and to create and view new relationships on the fly. There are a number of organization and presentation functions inherent in a StarMap. One is the ability to browse the metadata in an off-line manner (i.e., when there is no data capture function in process).

Another is the ability to display sketches of the data and metadata in a small window when the cue to do so is given. In the example shown in Figure 5 above, a ToolTip window displays a thumbnail sketch of the data or other user selected metadata because the StarMap software detects a mouse over event for the StarMap ‘current’ display object.

When browsing or sifting through the available data and information, the graphical display and ToolTips make it unnecessary to expand each node. Often the title or the icon will be sufficiently descriptive to allow the user to understand what to do next. When the user has a specific question, additional information can be displayed via a sequence of metadata messages as noted in the example above. The same functionality applies to the information about relationships as shown in Figure 6.

With this approach, a large number of data items can be manually browsed without opening the attendant application in order to look at individual data descriptions. Using tree structures to display the data nodes and relationships in a StarMap allows only one identified relationship to be displayed at a time. If the user chooses to maintain more than one relationship between any two data nodes (see Figure 1), there is provision that will also
allow the user to prioritize in which order the relationships will be displayed. Alternative sets of relationships can be identified on a sidebar. Displaying these relationships is done by entering the order preference in a StarMap pull down menu like the one shown in Figure 7. In order to display a single relationship for the whole StarMap, the user must identify that relationship as priority number 1 and leave all other possible selections blank. Modifying the relationship selections on the pull down menu (see Figure 7) will cause the display to be redrawn using the same node set while only displaying the selected relationship(s). If all relationships are noted at the same priority (including leaving them all blank), then all relationships possible will be available for display. Normally, only one relationship will be displayed at a time between any two nodes. The selection of the one that will be displayed by default will be the order identified in the basic StarMap tool. However there are user options and preferences that can modify this order. For example, one could select the temporal order in which the relationships were established as the criteria with a preference for newer relationships.

but were previously not displayed because they were lower priority to the Hyperlink relationship. If a temporal priority scheme was established earlier, then these would have been defined before the previously displayed relationships. Now that the priority order has been specifically reversed, the Content Similarity information will be displayed first. As noted earlier, normally when more than one relationship exists between two data nodes, only the first priority relationship is displayed. However, there are provisions for the multiplicity of relationships to be displayed. This can be done generally (which will usually create a StarMap that is not legible) or specifically by a mouse over event. A mouse over will display an information box for each relationship that exists describing the nature of the relationship including the current prioritization (see Figure 8).

Figure 6. StarMap Relationship Mouse Over

The results of the operation requested in Figure 7 are shown in Figure 8. The only displayed relationships are Content Similarities (first) and Hyperlinks that are permanently resident in a data node (second). Notice that several additional Content relationship lines have appeared. These were relationships that existed all along.

Figure 7. Selecting Relationships to Display
The discussion earlier relating to relationships can be repeated here in the matter of icons that are assigned to individual data nodes. Similarly, more than one can be assigned, and various rules for displaying them can be invoked. The example in Figure 9 shows the reordering of icon priorities in such a manner that the two data nodes identified by numbers 88 and 107 are changed. For example, the current information of importance about node 107 is that the source of the data is to be found in another StarMap, whereas before, the important information was that it was a video clip.

The Decision Maker’s also are anxious for any new functions that provide enhanced searching capabilities. As one can see in the previous figures, the richness of the presentation of the data nodes and their inter-relationships raises searching to a new level. Never again will the simple act of following a new link to a new site cause the user to lose a valuable set of context information. By double clicking on any data node in a StarMap, the application for that data node is invoked and the data is presented. However, the context of getting to that data node is still preserved in the StarMap and can be recalled with a bump of the mouse on the left of the screen (if the user selects this option) or a simple point-and-click.

One way to implement an initial approximation of the ideal StarMap is through innovative uses of tools such as the Hyperbolic Tree® from InXight Software, Inc.
Hyperbolic Tree provides an interactive three-dimensional view of a relationship tree as if it were mapped onto a hemispherical surface. To navigate the tree, users must simply click-and-drag any node identified within the view space. The use of multiple such trees integrated together will provide much of the StarMap functionality described above.

3.2. Amalgamation

By running a data capture engine, the user selects which items are added to the StarMap. Individual sets of data gleaned via a database query, copied files (documents), or downloaded web pages will be identified in the StarMap, and the appropriate data display application identified. Some capture engines allow downloading of entire or selected portions of web sites as if they were single entities. These can be identified in the StarMap either as a single entity or as a collection of entities with common capture context. As items are added, relationships must be immediately and automatically identified by the capture engine and the appropriate relationships supplied to the StarMap. Such relationships can be of the simple administrative type such as the order in which they were added, membership in a site (e.g., web server URL), or associated with a named entity (e.g., domain name for a company). Other relationships can be more data oriented such as identifying the existence of keywords in ‘text’ type additions. In addition, there are more advanced concepts such as the ‘subject’ of the associated data, type of content (e.g., those with audio or video), context of captured data, or other relationships. While not required for standard use of a StarMap, the inclusion of user-defined metadata from even more complex data capture processes is possible. Advanced users may want to index the nodes by type- or content-specific items (e.g., web pages by natural language so that future re-browsing could be done on English-only or French-only pages).

3.3. Data Analysis

During the creation of a StarMap, or during a later editing session, it is valuable to identify specific domains. This can be done by attaching notes to individual data nodes and by grouping data nodes together into sets. Most StarMap users create domain definitions specific to sets of problems being solved or queries being asked. For example, financial researchers who monitor company press releases may want the information grouped by product line. Military analysts who track news articles about a particular country may want the articles grouped into civilian and military subsets. Linkage to previous iterations (e.g., snapshots) of a data node may be desirable for certain change analysis queries. Layers of electronic ‘sticky’ notes may be associated with stored data without modifying the original content. This allows different versions of the data to be shown as desired (e.g., first-pass uses yellow notes, second-pass uses pink notes) and provides StarMap users with a mechanism to display a change history. Such a feature is useful for teams of researchers (e.g., professors and students working on research papers) and helps them avoid the age old “change / change back” cycle. Figure 10 shows a mechanism by which these node sets and their associated notes are displayed within a StarMap. Figure 11 shows where the StarMap user has identified both notes and sets. Highlighting each icon further identifies one set (in this case, the user had selected a bold red outline for highlighting). This set of nodes was selected for highlighting because of a mouse over of the set marker, which also displayed the user’s set description.
3.4. Personal Portability

A StarMap is a self-contained, self-organizing package of metadata. An individual StarMap may be transferred to other users as a fully functional package or it may be transferred with some of the functionality restricted (e.g., view-only – with disabled hyperlinks to actual data nodes). More important than sharing a StarMap with several individuals is the ability for the owner to republish it back to the data mart or data warehouse and automatically share it with a community of users. Once back in the data warehouse, this album of consistent and organized metadata becomes available to other interested users. If used judiciously, this would serve to minimize data transfers to other users. By first browsing through the library of pre-defined StarMaps, other users can find the most pertinent collections of data prior to actual downloading.

StarMaps may contain proprietary information (or ‘meta-information’) and will therefore have special security requirements. A StarMap must provide some form of application independent security. Data encryption and digital signatures are obvious choices for protecting the contents of a StarMap either in its entirety or by securing selected data or metadata contents. Access to StarMaps may also be controlled using encryption techniques that allow either individual users or specific groups to access the protected information (e.g., individual public keys or group keys). While many existing encryption programs acting independent of the StarMap could provide very strong encryption, they would also provide decrypted access to the information content as a whole. By putting the encryption routines within the StarMap access modules themselves, users may be granted read-only access to some or all of the contents. Such features are mandatory for ‘rented’ information sources such as an electronic library of copyrighted StarMaps. Additionally, some StarMaps, or the information sources they represent, may require multi-level security (e.g., a single electronic textbook containing both a teachers’ edition and a student’s edition).

Since a StarMap may be transmitted to any number of users on various types of computers and operating systems, the access software provides an optimal location for converting the contents to native formats. If a StarMap were transferred to a Personal Digital Assistant (PDA; as a travel agent might do as a service for major clients involved in tours), the icons would need to be consistent with the graphics of the PDA. In addition, any notes would have to be converted to a PDA-compatible format (e.g., ASCII text) and color images may be converted to gray scale in order to reduce the image size.

4. StarMap Applications

Since the StarMap is self-contained and may be transferred to other users or embedded within other objects (e.g., documents or even other StarMaps) and since it is self-organizing (i.e., the data is automatically filtered), it becomes the basis for innumerable user-level applications. Such applications include:

- Technical Research - Indexing engineering notebooks so critical to emerging technologies that rely on patent protection;
- Medical Records – Indexing a collection of a variety of data types and sources into a controlled record that can follow the patient anywhere even when the data itself is not digital;
- Online Product Announcements - Self-contained groups of text, audio, video and contact information could be organized and the StarMap used to guide potential customers through it all; and,
- Genealogy - Hierarchical pedigree information with links to additional historical content, family histories, scanned photographs, etc.

The security aspects of a StarMap allow it to act as a vehicle for delivering end-user published works without the need for an expensive publisher. Library applications become straightforward with no need for an actual librarian to perform the checkout and the checkin functions. The security embedded in the StarMap allows for automated processes to handle distribution and retiring
of copies according to specific rules (e.g., in two weeks the StarMap will no longer decrypt and becomes a heap of unreadable bytes - no need to return it to the library).

5. Conclusion

The principles embodied in the StarMap offer a fundamental technology base that is applicable to almost all applications that access very large databases. The development of new vertical applications that rely on some form of indexing is facilitated by the re-usable components contained within a StarMap. Applications only need ready access to simple and cost effective organizing tools to arise almost spontaneously. Users will now have the ability to perform continued specialization and analysis of the data currently contained within Data Warehouses and Data Marts. By enabling the users with these types of analysis tools, the ever-expanding burden on database managers will be eased. This will allow them to concentrate more effectively on mass data organization. The StarMap also allows users to tailor teaching and query sessions to very specific needs and to effectively combine data from a myriad of sources. Additionally, StarMaps, as smaller collections of organized information, can be re-integrated and used to augment existing massive data centers by providing new sources of information (i.e., refined data and metadata).