In Proc. of The Joint Pacific Asian Conference on Expert Systems, Singapore, 1997

Information Brokering over the Information Highway: An Internet-Based Database Navigation System

Syed Sibte Raza ABIDI
School of Computer Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia

Abstract. *The profuse use of the information highway has promoted it as a meritorious medium for global information sharing; accessing and sharing information stored in different databases that are inter-linked by the internet. We present an internet-based database application, a 'Re-configurable Internet Information Broker' that provides (a) the functionality to dynamically connect and interact in real-time with remote databases via the internet; and (b) an 'intelligent' database navigation engine based on the notion of database virtual hierarchies -- a database navigation mechanism that allows users to navigate the database by envisaging it as a user-defined hierarchical structure.*

1. Introduction

The explosive growth of the "Information Highway" has radically transformed the norms of information processing. Growing number of business organisations, government institutions, academics, students and a variety of users are now embracing the information highway as an apt medium for electronic, infomation-centered communication. Advances in internet based technologies have generated tremendous opportunities for information-sharing by making available not only a plethora of applications, software and document archives, but also providing access to numerous informed people from various domains. In the background of such technological advances, now, it remains of interest to explore possibilities for transpiring 'intelligence' to typical internet-oriented operations. One possible area that demands research attention is the extraction of information from remote databases, via the internet, by utilising 'intelligent' internet-based database navigation mechanisms.

Indeed, today the need to have access to information that is both correct and complete is very real. For strategic reasons, such demands disregard geographical and time constraints -- information/data should be available from any database site in the world and furthermore the required data should be found and made available with the shortest possible time-delay, as and when required [1]. To address such demands the research proposal, then, entails dealing with two key technologies -- (a) the internet and (b) database technology.

In this paper, we present an internet-based database application, a ***'Re-configurable Internet Information Broker'*** -- a confluence of internet protocols, database technology and artificial intelligence techniques to realise an application that provides (a) the functionality to dynamically connect and interact in real-time with any remote database via the internet; and (b) an 'intelligent' database access engine that incorporates artificial intelligence (AI) based database navigation mechanisms to extract pertinent information from a variety of databases [2].

The proposed *Re-configurable Internet Information Broker (RIIB)* is a generic database manipulation tool which most attractively provides an open-ended interface, exploiting ODBC technology, that permits it to dynamically connect and retrieve information from a variety of databases situated at different sites: the databases may vary both in structure and implementation platform. On the database navigation front, RIIB is an intuitive, easy-to-use, visual solution to complex data extraction mechanisms, hereby providing end-users a completely customisable information access environment, without requiring the knowledge of any query language and the need for programming. RIIB queries databases in a way that end-users understand: there is no need to remember table or field names, no typing of queries and no submission of erroneous field values. By way of an user-friendly interactive session, users can build complex compound queries by simply *clicking and choosing* the various data fields and specifying whatever constraints deemed relevant. More so, RIIB introduces the notion of a database 'virtual hierarchy' and the automatic creation of 'child' databases.

2. A Novel Database Navigation Mechanism Based on a *Database 'Virtual Hierarchy'*

From a database navigation perspective, we argue that a database can be envisaged at two levels (i) the *physical level* -- the flat structure of the database where all records are stored at the same level, and (ii) the *virtual hierarchical level*, whereby the flat structure of the database is transformed to a tree-like organisation - a *virtual taxonomy* or a *virtual hierarchy*, where various fields are hierarchically organised, in a tree-like structure. The term 'virtual taxonomy' has been defined by Woods in the context of description of concepts in knowledge representation systems such that whenever a system "constructs an explicit collection of concept nodes ... the result is a subgraph of the virtual taxonomy". Woods' motivation for viewing a collection of 'descriptions' this way is that "although its structure is important, one never wants to make it explicit in the memory of a computer"[3].

For database navigational purposes, we introduce a shift towards a dynamic *database virtual hierarchy* ; providing users the provision to 'virtually' organise the database as a hierarchical structure based on user-specific priorities [4]. Each level of the database virtual hierarchy corresponds to a field of the records stored in the database. Database navigation can then be carried out by following appropriate branches through the virtual hierarchy, without needing to consider the database as a whole. Put simply, the virtual hierarchy can be envisaged as the (attribute-oriented) criteria for selecting pertinent records from a database. The word *virtual* is indicative of the fact that the physical structure of a database is not altered, rather the database is just to be viewed as a (virtual) hierarchical structure for navigation purposes. Since the hierarchy is virtual for all purposes, users can dynamically define the hierarchical structure of the same database in a variety of ways, i.e., a number of profiles of the same database can be realised by specifying different database virtual hierarchies.

Specification of a user-specific database virtual hierarchy is carried out during a user-friendly dialogue session. The user is provided a list of available fields, following which the user can specify the database's virtual hierarchy by simply selecting the fields of interest, and additionally suggesting the relevant values of the selected fields. The database's virtual hierarchy is simultaneously built according to the order in which the fields are selected: the top level of the hierarchy is the first selected field and so on. Indeed, different users would like to 'intelligently' view the same database for varying purposes and this can be achieved by specifying user-specific hierarchies. For instance, if a database containing records about text attributes is to be used by a translator then the top level of the virtual hierarchy should be the attribute *Language* (a typical hierarchy for this purpose may look like figure 1a), whereas if the same database is to be used for studying the nature of specialist texts then the top level is desired to be the attribute *Domain* (a typical hierarchy is shown in figure 1b).

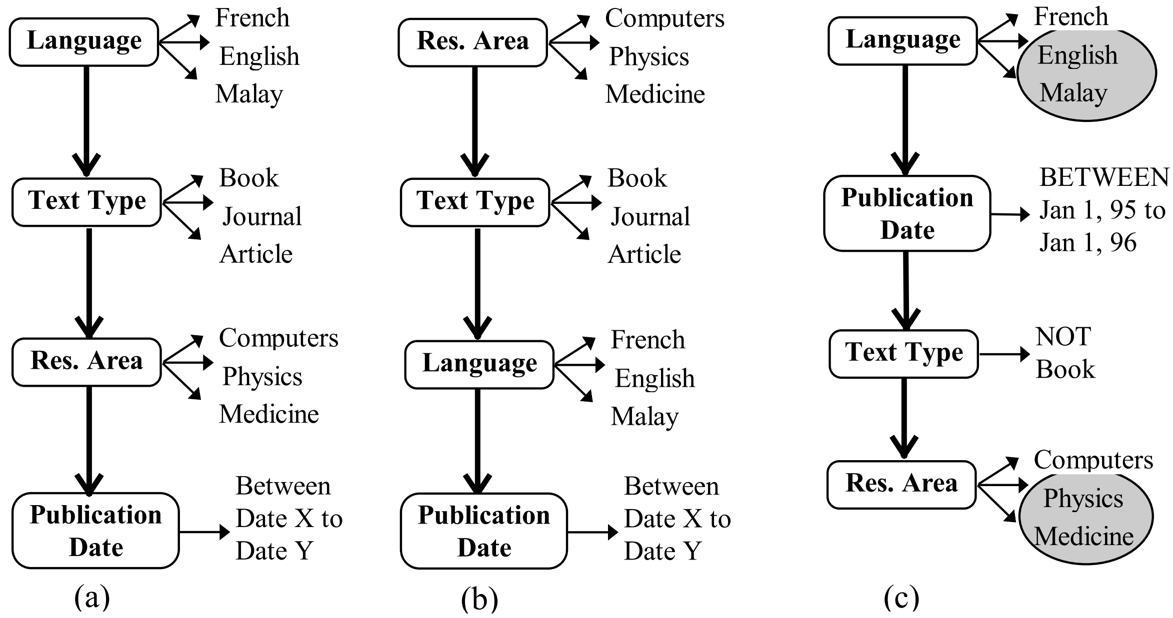

Figure 1: Virtual hierarchies for a database

Database navigation initiates from the top level of the database's virtual hierarchy database and subsequently proceeds to the next lower level and so on. At the top level (say L1) of the database hierarchy the scope of possible records that can be selected is the entire database. However, as the search proceeds to the next lower level (say L2), the scope of selectable records is reduced to the set of records found at level L1. This strategy can be envisaged as the filtering of records (based on the constraints imposed by the hierarchy) as we move down the database's virtual hierarchy, thus resulting in a fine-grained retrieval of highly specific records at the lowest level of the hierarchy. To elucidate the nature of the navigation mechanism and the scope of record selection consider the virtual hierarchy shown in figure 1c: the top level is the attribute *Language* with two selected values - English and German. The navigation mechanism works as follows: Initially, all records from the entire database that have *Language = English or German* are selected: say, the retrieved set of records is called Record Set 1. At the next lower level of the hierarchy, i.e. *Origination Date*, the scope of records to be searched is Record Set 1 and the search mechanism would retrieve from Record Set-1 all records that have an *Origination Date = BETWEEN Jan 1 1995 to 1 Jan 1996*, yielding say Record Set-2. Therefore, Record-Set 2 is a subset of Record-Set 1. Next, all records that have a *Text Type = Not Book* are selected from Record-Set 2, forming say Record Set 3. Finally, at the lowest level of the database's virtual hierarchy, from Record-Set 3 all records that have Domain = physics or medicine are collected to realise the final collection of records retrieved based on the database's virtual hierarchy.

More attractively, database navigation is interactive thus allowing the following features: (i) users can view and select (all or some) the records collected at each level of the hierarchy; (ii) users can dynamically customise the already specified database's virtual hierarchy. For instance, the hierarchy in figure 1c has *Language = English and German*. Here, the user can constrain the search by selecting only *Language = English* and not consider the other specified values of the *Language* attribute; and (iii) from each level the user can proceed both in a forward (next level down in the hierarchy) and backward (previous higher level in the hierarchy) direction, much like web-browsers. We therefore believe that, database navigation based on a database virtual hierarchy facilitates a more systematic visualisation and understanding of data stored in a database.

3. Architecture Of The Internet Information Broker

In architectural terms, RIIB can be envisaged to comprise three components: (i) *Internet Connector*; (ii) *Metadata Editor*; and (iii) *Database Navigator*. RIIB also implements a Database Manager, providing a suite of database management facilities. We briefly discuss below each component. Figure 2 illustrates the architecture of RIIB.

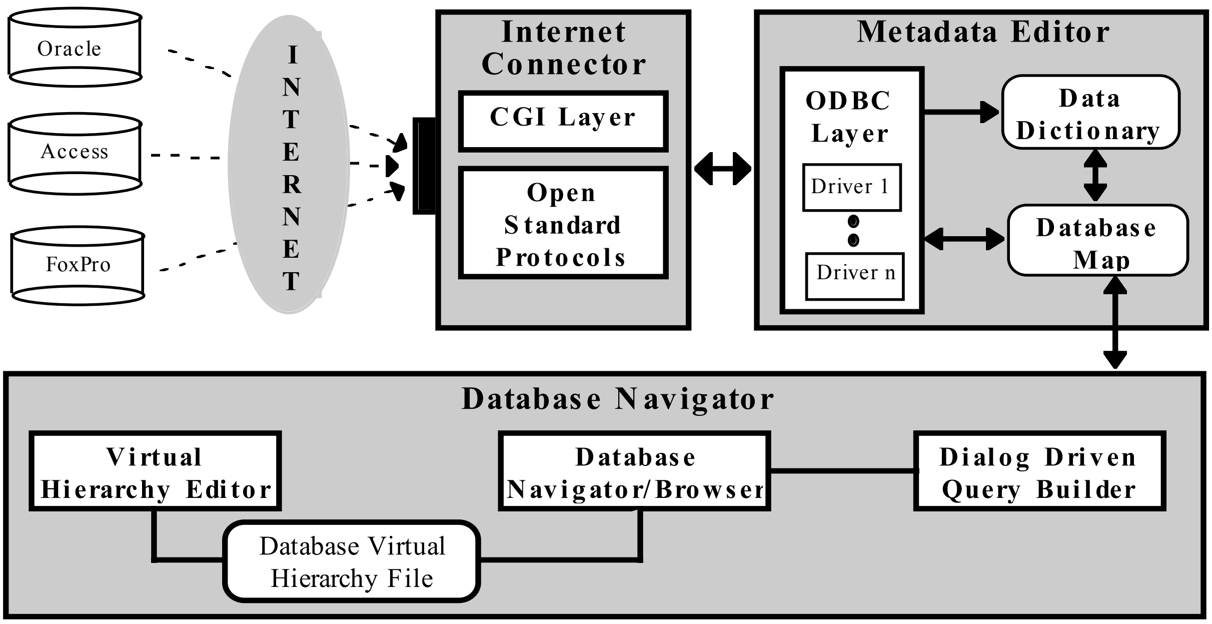

Figure 2: The architecture of the Information Broker

Internet Connector provides RIIB the necessary interface to connect with the internet and interact with various 'client' databases. The implementation of *open standard protocols* facilitates internet-based transactions and the *CGI (Common Graphical Interface) layer* working closely with the Metadata Editor, in particular the ODBC layer, provides an 'intelligent', dynamically re-configurable GUI front end to interact with the connected database.

Metadata Editor facilitates an internal, functional interface between the physical database and the database navigator. An *ODBC layer* allows the RIIB to connect with a variety of different databases. This is achieved by the dynamic creation of two files: (i) *Data Dictionary* file which contains the specification (tables, fields and types) of the connected database; and (ii) *Database Map* file that contains Prolog based facts responsible for the overall database navigation operations. The Database Map, which is also used for the creation of a database specific GUI, is based on the Data Dictionary's specifications. The use of the Data Dictionary and the Database Map thus allow RIIB to be reconfigured (on the fly) according to the specification of any connected database system.

Database Navigator provides an intelligent database navigation engine based on either a virtual hierarchy or on user-defined queries. It comprises three main modules: (i) *Virtual Hierarchy Editor* - a user-friendly dialogue box that allows the specification of user-specific database virtual hierarchies (manifested as the *Database Virtual Hierarchy File)*; (ii)

Dialogue Driven Query Builder which provides a dialogue based interface for building compound queries; and (iii) *Database Navigator/Browser* which provides a visual illustration of the navigation through the database. When navigating based on a virtual hierarchy, the Database Navigator/ shows the database hierarchy, the current level, the values of the attribute specified at the level in the hierarchy (with the provision to choose the various to dynamically customise the search), records retrieved at each level, and other navigational information and options

Database Manager implements various database management routines. A 'Electronic Form' based approach, based on CGI, is used for adding, modifying and deleting records.

The Internet Information Broker, currently under development, is implemented on the PC platform under the Microsoft Windows environment. The software is developed using the LPA Win-Prolog (version 3.1.) programming package.

4. Features Of The Re-Configurable Internet Information Broker

We discuss below the important database navigation and management features supported by RIIB

4.1. Dialogue Driven Query Building - click and choose:

Database navigation is guided by the constraints imposed by the user's query, however database users vary greatly in their level of database manipulating skills. RIIB relieves the user from the need to know any database query language. Rather, in our RIIB both a database virtual hierarchy and a compound query is built in an user-friendly interactive session where dialogue boxes showing the data fields are presented to the user. The user is just required to choose the data fields of interest and specify the constraints, if any. Put simply, to assist an efficient and informed transversal of the database our RIIB makes available all necessary information right at the fingertips of the user. Then, in an intuitive manner, the user can build a navigation scheme -- a virtual hierarchy or a compound query, by basically *clicking and choosing* the various data fields in a systematic manner.

4.2. Open-Ended Interface

RIIB provides an open-ended interface; by exploiting ODBC technology, RIIB can be configured (on the fly) to the specification of any database format and database platform (for instance Oracle, MS Access, FoxPro, etc.). This allows end-users to design their databases using whatever database products and suitable database specification but still be able to exploit the functionality of RIIB.

4.3. Creation of 'Child' Databases

Information brokering requires the efficient transportation of information from the 'parent' source to a secondary source - a 'child' database. RIIB allows the interactive creation of a *'child' database* that can be easily exported across various database and word processing platforms.

5. Conclusion

In conclusion, the problem of extracting information from databases is quite real [5], more so if the database in question is situated off-site. We believe that the highest level of functionality one can associate with our *Re-configurable Internet Information Broker* is **telepresence:** a capability that permits internet users to access information from a variety of databases resident on computers that may be either on- or off-site via a communication channel of their choice -- from a normal phone to a high-speed connection. More attractively, the RIIB provides a graphical gateway to databases (that may vary both in structure and platforms), making the use of databases more easy, efficient, informative and 'intelligent'.

We propose the notion of *database virtual hierarchies*: by incorporating virtual hierarchies users can impose their own structure on an existing database to retrieve pertinent information. The aim of the virtual hierarchy approach is to provide the flexibility of the traditional attribute-based information retrieval approach, but with the intuitive functionality of an explicit taxonomy approach. For that matter, we have demonstrated how database navigation based on a database's virtual hierarchy can be more informed, user-specific and interactive. We believe that, applications of RIIB could be numerous, such as an *'Electronic' Computer Shopping Mall* (a relational database of databases connected via a computer network), a front-end to a data mining system and other similar applications.

References

[1] Upton, D.M., & McAfee, A. "The Real Virtual Factory", in Harvard Business Review, July-August, 1996.

[2] Abidi, S.S.R. "Towards Information Brokering: An Intelligent Database Navigation and Management System", in Research & Development in Computer Science and Applications Conference, Kuala Lumpur, 1996.

[3] Woods, W.A. (1991). "Understanding Subsumption and Taxonomy: A Framework for Progress". in John. W. Sowa (eds) Principles of Semantic Networks. Morgan Kaufmann Publishers, California, 1991.

[4] Holmes-Higgin, P., Abidi, S.S.R., & Ahmad, K. "Virtual Text Corpora and Their Management", in Sixth EURALEX Int. Congress '94, Univ. of Amsterdam, 1994.

[5] Geppert, A, Dittrich, K.R. "Constructing the Next 100 Database Management Systems: Like the Handyman or Like the Engineer?", in ACM SIGMOD Record 23:1, (1994).