THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Usage of Online Analytics Programming And Information Retrieval: Business Intelligence (BI)

Samson Oluwaseun Fadiya
Girne American University, Mersin 10 Via Turkey, T.R.N.C

Abstract:
The information system has always been an important factor to a success of a business around it such as economic, political and social issues. Every organization must realize the importance of MIS, technology, mixing with the corporation's business strategies and techniques. To fully understand the concept of BI, it is often viewed in connection with data warehouse (DWH's) and online-analytical-programming (OLAP) technology and usage of Information Retrieval techniques of BI. Hence, this research activity is geared towards the investigation of the case of organizations; Evaluation of the effectiveness Business Intelligence (BI) is based on a database (DB) technology.

Keywords: *Business Intelligence, Data Warehouse, Information Retrieval (IR), Multidimensional IR engine (MIRE), Multidimensional Document Expression (MDDX), Symbiotic Context-Oriented Information (SCORE), On-line Analytical Processing (OLAP)*

1. Introduction

In order to receive complete business intelligence, integrating and examining text data are necessary. Analyzing the second, such information can be worked out. Sullivan [7] aimed to link relational data warehouses with a text document storage warehouse for that function. Online analytical processing (OLAP) technology is a multidimensional analysis of the immense volume of data from many views [2]. In this paper, we use OLAP technology as Text OLAP, as well as call the multidimensional analysis of text documents. Because Text OLAP, we engage one major text handling technologies: Information Retrieval (IR). Of a given document set, an IR system recovers the ones holding the keywords given as a user query.

For this role, we required the total BI program platform that incorporates IR technology as well as relational OLAP technology. We depict the example Text OLAP approaches based IR technology, and talk about the increase issues for getting total business intelligence BI. Operational report is also a paramount dimension of operational system; it generally involves significant reporting potentialities, which has a high value in doing business analysis and are justly component of BI strategy. In this paper, Section 2 centers on the Text OLAP demonstrated on IR technology respectively. Section 3 centers on the architecture of total business intelligence platform and research issues. Finally, we conclude in Section 4.

2. Analyzing Text File

2.1. Text OLAP demonstrated on Information Retrieval

Using IR technology, we can extract keywords from and compute their weights for text file. In an IR system, every document is assumed to consist of keywords, and a keyword-based query is used to retrieve the documents that are related with the keywords given in the query. For a particular retrieval model, e.g. the vector space model, the retrieved documents ranks in the query result. For more information on IR technology, Baeza-Yates et al. [1] Is a good reference.

2.2. Constructing Document Cube

As dimensional data, they used the keywords extracted from document contents, the categories of documents, and the metadata such as title, creator, date, and rights. Tseng et al. [8, 9, and 10] proposed a document warehouse for multidimensional analysis of text documents. However, they did not mention how the metadata and the keywords can be organized in a hierarchical form. As measure data, they used the identifiers and the number of corresponding documents. We can perform a multidimensional analysis for the document warehouse, for example, how many documents were created regarding 'business intelligence' between 2010 and 2013?

Each cell of the cube has the identifiers of corresponding documents stored in the document set. For the document cube, we can submit a multidimensional query. Tseng et. al.(2006) proposed a new form of query language called MDDX (Multidimensional Document Expression) specially designed for the document cube.

2.3. Constructing Text Cube

A modern proportion for semantic navigation of text data, which necessitates two new OLAP operations such as pull-up and push-down, is proposed by Lin et al.[5] it is a new data cube called text cube. It also has two unique IR measures, term frequency and inverted index so that other IR techniques and applications can be efficiently built. A term hierarchy, is freshly introduced in text-cube, to specify the semantic levels of and relationships among text terms. A user can pull-up or push down the term level along the hierarchy. The measures, term frequency and inverted index, are essential for IR tasks and are redefined for aggregated text data.

2.4. Computing Multidimensional IR

Lee and McCabe et al. [3, 4, and 6]. Suggested a multidimensional IR engine (MIRE), which is based on a multidimensional data model to utilize OLAP technology. The approach is only to build a multi-dimensional data model for text and allows users to search with the ease of OLAP techniques. The multidimensional IR engine system can provide for a multidimensional IR query, and new functionalities such as browsing through a document collection and quickly identifying patterns surrounding where, when and on what topic documents are scripted.

3. Integrate Business Intelligence Design

In order to employ total business intelligence, Text OLAP should be integrated with relational OLAP to enable Integral OLAP over both structured and unstructured data. In this section, we propose an architecture of the total business intelligence platform that integrates Text OLAP and relational, OLAP, and discuss research topics related with it.

3.1. Architecture of Integral Business Intelligence Design

In order to obtain Integral business intelligence, the total business intelligence platform needs another OLAP module that integrates Text OLAP and relational OLAP. The architecture of the total BI platform can be exemplified as shown in Figure 1. There exist two data sources: a relational data warehouse storing structured data and text document warehouse storing various kinds of documents including web pages from the Internet. The Relational OLAP module handles the former to provide some valuable idea, and the Text OLAP module based on IR handles the latter to provide some context surrounding the idea found. The Integral OLAP module handles Integrating relational OLAP and Text OLAP modules together.

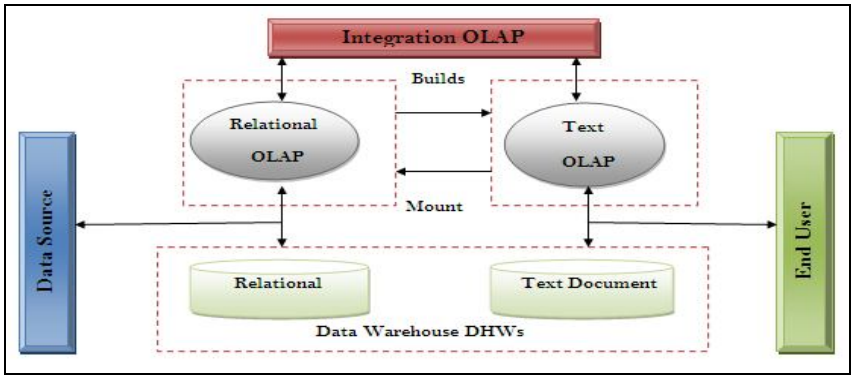

Figure 1: Architecture of Integral BI Design

3.2. Integration OLAP

Through the connecting mechanism like that of R-Cube or EROCS, the relational data in a relational data warehouse and the text documents in the text document warehouse can be actively integrated as shown in Figure 2. In the fused data warehouse, we have two builds tables connected with one another via the general attribute tables. For instance, in a relational data warehouse and the text document warehouse in Figure 2, we have two sets of typical dimension tables Attribute 1 and Attribute 2, and two builds tables, Builds 1 and Builds 2. The idea in the tables Builds 1 and Builds 2 are linked together when they have apparently the same attribute values in common. Hence, as illustrated in Figure 3, we can deduce fused data warehouse having one set of attribute tables and one complex build table with two kinds of values linked together according to the same.

We can take out a novel form of OLAP over the fused data warehouse, which we can name Integral OLAP. The platform architecture in Figure 1 has the ability, Integral OLAP. Integration OLAP is a principal one and has two OLAP's as physical components.

If we bring out an OLAP question via Integration OLAP module in Figure 1. It is disseminated to a relational OLAP module or Text OLAP module giving to the steps specified in the question, it is first exploited by Text OLAP and then by relational OLAP on the only idea that are linked to the Text OLAP answers. Hence, Integration OLAP has the benefit that individuals don't have to take out relational OLAP and Text OLAP independently; they can only take out just one.

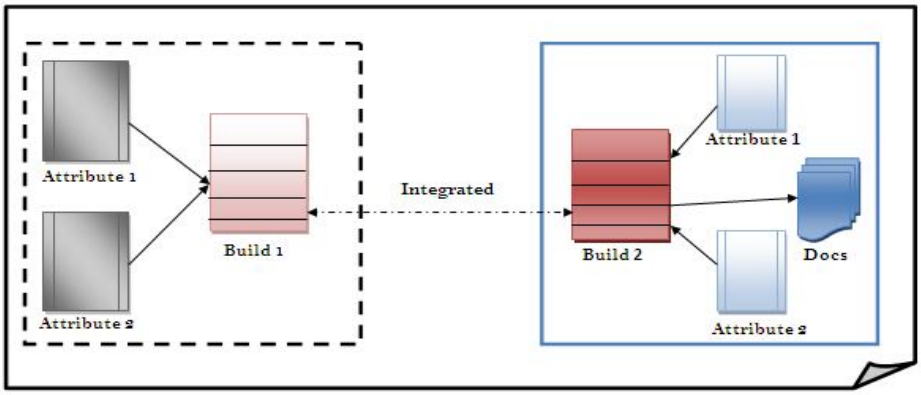

Figure 2: Integrated Data Warehouse

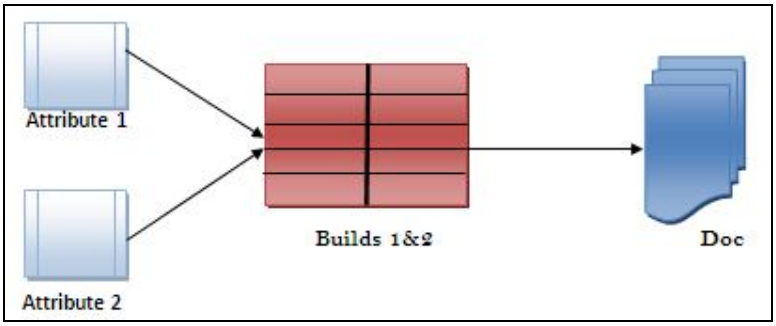

Figure 3: Aspects for Integrated Data Warehouse (DWH)

3.3. Outcome of Research

We call for more investigation for carrying out the architecture of the total BI platform that comprises of two heterogeneous elements: relational OLAP and Text OLAP. In details, the architecture allows the new ideas of integration OLAP. To totally take help of the combine BI platform architecture in merged situation, we encourage study of research topics like query processing and language: Integration OLAP is a principal one; the query computing is more complicated than relational OLAP or Text OLAP. In other word, we need an effective integration mechanism among relational OLAP and Text OLAP and more so, because individual interface the integration OLAP module, we need a new query language that incorporates both relational OLAP and Text OLAP

Finally, Meta precaution: that is for integration OLAP, we need mechanical sustain of metadata between the relational data warehouse and the text document storage warehouse. Keeping consistency between two severally developing data warehouses is difficult.

4. Conclusion

Now, creating and analyzing text data is necessary for complete business intelligence. Hence, we computed the representative works chosen for showing how the technologies of IR, can be used for Text OLAP. We also computed the representative works chosen for showing how we can, using IR technology for multidimensional analysis data.

Therefore, we recommended a combine BI platform architecture that can offer improved service toward total business intelligence. The architecture is established on the modern concept of OLAP. Which incorporates two heterogeneous OLAP offers users to search structured and unstructured data at a time for obtaining complete business intelligence, and to explore the information that otherwise could be omitted by relational or text OLAP when worked differently.

In conclusion, next research work should validate the idea of integration OLAP for searching its features more, and build up an effective system of integrating a system that of not the same OLAP sources and integration of OLAP for total BI platform comparable how to strengthen text documents and the performance areas.

5. Acknowledgements

This work is based on the projects “Emerging Markets Telecommunication Services (EMTS), trading as Etisalat", supported by Nigeria in partnership with Mubadala Development Company.

6. References

1. Baeza-Yates, R. and RibeiroNeto, B. (2010). Mordern Information Retrieval, Addison-Wesley.
2. Chaudhuri, S., & Dayal, U. (2008). An overview of data warehousing and OLAP technology. SIGMOD Rec., 26(1), 65-74.
3. Lee, J., Grossman, D., & Orlandic, R. (2011). MIRE: A multidimensional information retrieval engine for structured data and text. Proceedings of the International Conference on Information Technology: Coding and Computing, Las Vegas, NV, USA. 224-229.
4. Lee, J., Grossman, D., Frieder, O., & McCabe, M. C. (2009). Integrating structured data and text: A multi-dimensional approach. Proceedings of the International Conference on Information Technology: Coding and Computing, Las Vegas, NV, USA. 264-269.
5. Lin, C. X., Ding, B., Han, J., Zhu, F., & Zhao, B. (2008). Text Cube: Computing IR measures for multidimensional text database analysis. Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, 905-910.
6. McCabe, M. C., Lee, J., Chowdhury, A., Grossman, D., & Frieder, O. (2007). On the design and evaluation of a multi-dimensional approach to information retrieval (poster session). SIGIR '00: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece. 363-365.
7. Sullivan, D. (2009). Document Warehousing and Text Mining, John Wiley & Sons, Volume 63, 2009, pp 29-36.
8. Tseng, F. S. C. (2007). Design of a multi-dimensional query expression for document warehouses. Information Sciences, 174(1-2), 55-79.
9. Tseng, F. S. C., & Chou, A. Y. H. (2009). The concept of document warehousing for multi-dimensional modeling of textual-based business intelligence. Decision Support Systems, 42(2), 727-744.
10. Tseng, F. S. C., & Lin, W. P. (2006). D-tree: A multi-dimensional indexing structure for constructing document warehouses. Journal of Information Science and Engineering, 22(4), 819-841