GRAPHICAL WEB BASED TOOL FOR GENERATING QUERY FROM STAR SCHEMA

MOHAMMED F R ANBAR

UNIVERSITI UTARA MALAYSIA
2009

GRAPHICAL WEB BASED TOOL FOR GENERATING QUERY FROM STAR SCHEMA

A thesis submitted to the Graduate School, College of Arts and Sciences in partial fulfilment of the requirements for the degree Master of Science (IT)

Universiti Utara Malaysia

By

MOHAMMED F.R. ANBAR

(Matric No: 800310)

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Master of Science in IT degree from University Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor or, in their absence by the Academic Dean College of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to University Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to

Dean (Academic) College of Art and Sciences
University Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman.

ABSTRACT

Novice users have difficulty to generate structured query language from the star schemas because they are not familiar with formulating SQL queries and SQL syntax. This study proposed graphical web based tool to generate queries from star schema and represent the data in tabular or graphical forms which help novice user to formulate SQL query. A prototype for a web based tool to generate the query has been developed using Java Server Pages programming language. The developed tool can facilitate complex query construction which is faced by non-technical and/or novice users. The output of SQL query is presented in tabular and graphical forms which can help users especially top management in better understanding and interpreting query results.

.

ACKNOWLEDGEMENTS

First, I would like to express my appreciation to Allah, the most merciful and, the most compassionate, who has granted me the ability and willing to start and complete this study. I do pray to His Greatness to inspire and enable me to continue the work for the benefits of humanity.

After that, my most profound thankfulness goes to my supervisor Prof Ku Ruhana Ku-Mahamud for her scientifically proven and creativity encouraging guidance and great support in this study.

Last, I wish to thank my Father, Mother who were always there for me by giving everything they have, my brothers and sisters for their love and support.

Thank you UUM.

Mohammed F.R Anbar

April 12, 2009

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SQL	Structured query language
OLAP	On-line analytical processing
OLTP	Online transaction processing
HTML	HyperText Markup Language
IDE	Integrated Development Environment
JSP	Java Server Pages
DW	Data warehouse
NL	Natural language
SQUARE	Specifying Queries as Relational Expressions
TRC	Tuple relational calculus
DDL	Data definition language
DML	Manipulation language
DBMS	Database management system
RDBMS	Relational Database Management system
RUP	Rational Unified Process
UML	Unified Modeling Language
WWW	WORLD WIDE WEB
HTTP	Hyper Text Transfer Protocol

NO	TITLE OF THE FIGURE	PAGE

CHAPTER ONE

INTRODUCTION

Database applications which are deployed in many corporations for the purpose of storing and retrieving data in a structured way has become ubiquitous and indispensable. Many transactions are being performed by users daily. These transactions represent the questions asked by users in the natural language which are formulated as structured query language (SQL) to deal with database. The ways of formulating queries depends on user knowledge and experience. Ultimately, the output of queries represents the answers to user questions.

A query language is a specialized language in which a user requests information from a database. These are typically of a higher-level than programming languages. They are classified in to two types, procedural and non procedural. Procedural is where the user instructs the system to perform a sequence of operations on the database. That will compute the desired information. Nonprocedural is where the user specifies the information desired without giving a procedure in obtaining the information. A complete query language also contains facilities to insert and delete topples as well as to modify parts of the existing topples.

Most of data set which represents the outputs of SQL query are presented in the form that users take a lot of time in browsing the data. This makes the users uneasy

The contents of the thesis is for internal user only

References:

Agarwal, D., Barman, D., Gunopulos, D., Young, N. E., Korn, F., & Srivastava, D. (2007). Efficient and effective explanation of change in hierarchical summaries. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining.*

Baresi, L., & Morasca, S. (2007). Three empirical studies on estimating the design effort of Web applications. *ACM Transactions on Software Engineering and Methodology*, *16* (4), 40-45

Becker, K., Cardoso, M. O., Nichele, C. M., & Frighetto, M. (2000). Mail-by-example: a visual query interface for email management. *Proceedings of the working conference on Advanced visual interfaces*. 280-281.

Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta, A., & Paraboschi, S. (2001). Designing data marts for data warehouses. *ACM Transactions on Software Engineering and Methodology, 10*(4), 452-483.

Borgman, C. L. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science, 47*(7), 493-503.

Brambilla, M., Ceri, S., Fraternali, P., & Manolescu, I. (2006). Process modeling in Web applications. *ACM Transactions on Software Engineering and Methodology (TOSEM), 15*(4), 360-409.

Cao, B., & Badia, A. (2007). SQL query optimization through nested relational algebra. *ACM Transactions on Database Systems (TODS), 32*(3)

Chapple, M. (2000). Data Definition Language. Retrieved February 11th, 2009 From http://databases.about.com/od/sql/a/sqlfundamentals_2.htm

Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record, 26*(1), 65-74.

Chen, Z., & Li, T. (2007). Addressing diverse user preferences in SQL-query-result navigation. *Proceedings of the 2007 ACM SIGMOD international conference on management of data.* 641-652.

Cl, B. (1999). The users mental model of an information retrieval system: an experiment on a prototype online catalog. *International Journal of Human-Computer Studies, 51*(2), 435-452.

Conallen, J. (2000). Modeling Web Applications with UML. *White paper, Conallen Inc., http://www. conallen. com/whitepapers/webapps/ModelingWebApplications. htm, March 1999*.

Conallen, J. (2002). *Modeling Web Applications with UML* (2nd ed.). USA: Addison-Wesley.

Connolly, T. M., & Begg, C. E. (2005). *Database Systems: A Practical Approach to Design, Implementation, and Management*: USA: Addison Wesley Publishing Company.

Deshmukh, H., Malavia, J., & Carter, w. J. (2003). *SCWCD exam study kit : Java web component developer certification.* Greenwich: Manning Publications.

France, R., & Kobryn, C. (2001). UML for software engineers. *Proceedings of the 23rd International Conference on Software Engineering.* 705-706

Ginige, A., & Murugesan, S. (2001). Web engineering: an introduction. *IEEE Multimedia, 8*(1), 14-18.

Hao, J., & Mendes, E. (2006). Usage-based statistical testing of web applications. *Proceedings of the 6th international conference on Web engineering.* 17 - 24

Huo, J. (2008). *KMVQL: a visual query interface based on Karnaugh map*. Proceedings of the working conference on Advanced visual interfaces.

Inmon, B., & Kelly, C. (1994). The Twelve Rule of Data warehouse for a Client/Server World. *Data Managment Review 4*(5), 6-16.

Jarke, M. (2003). *Fundamentals of Data Warehouses.* New York: Springer-Verlag.

Jones, S. (1998). Graphical query specification and dynamic result previews for a digital library. *Proceedings of the 11th annual ACM symposium on User interface software and technology.*

Jung, H., & Lee, G. G. (2002). Multilingual question answering with high portability on relational databases. *COLING-02: proceedings of the 2002 conference on multilingual summarization and question answering - Volume 19*

Kaner, C. (2003). *What is a Good Test Case?*

Kappel, G., Pröll, B., Reich, S., & Retschitzegger, W. (2006). *Web Engineering-The Discipline of Systematic Development of Web Applications.* USA: John Wiley & Sons, Inc.

Kern, J., & Garrett, C. (2003). Effective Sequence Diagram Generation. *Borland (www. borland. com/together/white_papers)*.

Kim, W. (1982). On optimizing an SQL-like nested query. *ACM Transactions on Database Systems (TODS), 7*(3), 443-469.

Krippendorf, M., & Song, I. Y. (1997). The translation of star schema into entity-relationship diagrams. *Proceedings of the 8th International Workshop on Database and Expert Systems Applications.*

Law, A. S., Freer, Y., Hunter, J., Logie, R. H., McIntosh, N., & Quinn, J. (2005). A Comparison of Graphical and Textual Presentations of Time Series Data to Support Medical Decision Making in the Neonatal Intensive Care Unit. *Journal of Clinical Monitoring and Computing, 19*(3), 183-194.

Malinowski, E., & Zimányi, E. (2006). Hierarchies in a multidimensional model: From conceptual modeling to logical representation. *Data & Knowledge Engineering, 59*(2), 348-377.

Martin, R. C. (1997). UML Tutorial: Part 1--Class Diagrams. *Engineering Notebook Column, C++ Report*.

Nourie, D. (2006). Java Technologies for Web Applications. Retrieved January 13th, 2009 From http://java.sun.com/developer/technicalArticles/tools/webapps_1/.

Pokorny, J. (1998). Data Warehouses: a Modelling Perspective. In: Evolution and Challenges in System Development. *Proceeding of 7th International Conference on Information Systems.*

Ramakrishnan, R., & Gehrke, J. (2003). *Database Management Systems* (3 ed.). New York: McGraw-Hill Professional.

Rob, P., & Coronel, C. (2008). *Database Systems: Design, Implementation, and Management*. USA: Thomson Course Technology.

Rowen, W., Song, I. Y., Medsker, C., Arynth, I., & Ewen, E. (2001). An Analysis of Many-to-Many Relationships Between Fact and Dimension Tables in Dimensional Modeling. *Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2001) Interlaken, Switzerland, June 4, 2001.*

Sibley, E. H., & Taylor, R. W. (1973). A data definition and mapping language. *Communications of the ACM, 16*(12), 750-759.

Stoffel, K., Davis, J. D., Rottman, G., Saltz, J., Dick, J., Merz, W., et al. (1998). A graphical tool for ad hoc query generation. *Proceedings of the AMIA(American Medical Informatics Association) Symposium*. 503-7.

Nishith. (2005). Database vs. Data Warehouse. Retrieved Febreuary 2nd, 2009, From http://opensourceanalytics.com/2005/11/02/database-vs-data-warehouse/

Weininger, A. (2002). Efficient execution of joins in a star schema. *Proceedings of the ACM SIGMOD international conference on Management of data.*