

# Extending QMBE Language with Clustering

Ana Azevedo, Polytechnic Institute of Porto/ISCAP, Portugal & Algoritmi R&D Center,  
University of Minho, Guimarães, Portugal

Manuel Filipe Santos, Algoritmi Center, Department of Information Systems, University of  
Minho, Guimarães, Portugal

---

## ABSTRACT

*Business Intelligence (BI) is an important area of the Decision Support Systems (DSS) discipline. Over the past years, the evolution in this area has been considerable. Similarly, in the last years, there has been a huge growth and consolidation of the Data Mining (DM) field. DM is being used with success in BI systems, but a truly DM integration with BI is lacking. This creates a gap between DM and BI systems. With the purpose of closing this gap a new DM language for BI, named as Query-Models-By-Example (QMBE), was envisaged and implemented with success, but addressing only classification rules. This paper presents an extension of QMBE language to include clustering. This represents one more step towards the integration of DM with BI, which constitutes an important issue.*

**Keywords:** *Business Intelligence (BI), Clustering, Data Mining (DM), Data Mining Language, Inductive Database, Query-By-Example (QBE), Query-Models-By-Example (QMBE), Relational Model*

---

## INTRODUCTION

Business Intelligence (BI) can be presented as an architecture, tool, technology or system that gathers and stores data, analyzes it using analytical tools, and delivers information and/or knowledge, facilitating reporting, querying, and, ultimately, allows organizations to improve decision making (Clark, Jones, & Armstrong, 2007; Kudyba & Hoptroff, 2001; Michalewicz, Schmidt, Michalewicz, & Chiriac, 2007; Moss & Shaku, 2003; Negash, 2004; Raisinghani,

2004; Thierauf, 2001; Turban, Sharda, & Delen, 2011). To put it shortly, Business Intelligence can be defined as the process that transforms data into information and then into knowledge (Golfarelli, Rizzi, & Cella, 2004). Being rooted in the Decision Support Systems (DSS) discipline, BI has suffered a considerable evolution over the last years and is, nowadays, an area of DSS that attracts a great deal of interest from both the industry and researchers (Arnott & Pervan, 2008; Clark, Jones, & Armstrong, 2007; Hannula & Pirttimäki, 2003; Hoffman, 2009;

DOI: 10.4018/ijdsst.2013100104

Negash, 2004; Richardson, Schlegel, & Hostmann, 2009; Richardson, Schlegel, Hostmann, & McMurchy, 2008).

Data Mining (DM) is being applied with success in BI and several examples of applications can be found (Linoff, 2008; Turban, Sharda, & Delen, 2011; Vercellis, 2009). Despite that, DM has not yet reached to non-specialized users. The authors consider that the real issue is related with the fact that the Knowledge Discovery in Databases (KDD) process, as presented by Fayyad, Piatetski-Shapiro, and Smyth (1996), is not fully integrated in BI. Consequently, its full potential could be not completely explored by decision makers using the systems. Currently, DM systems are functioning as separate isles.

The authors have point out three main reasons for DM to be not completely integrated with BI. Firstly, the models/patterns obtained from DM are complex and there is the need of an analysis from a DM specialist. This fact can lead to a non-effective adoption of DM in BI, being that DM is not really integrated on most of the implemented BI systems, nowadays. Secondly, the problem with DM is that there is not a user-friendly tool that can be used by decision makers to analyze DM models. Usually, BI systems have user-friendly analytical tools that help decision makers in order to obtain insights on the available data and allow them to take better decisions. Examples of such tools are On-Line Analytical Processing (OLAP) tools, which are widely used (Negash, 2004; Turban, Sharda, & Delen, 2011). Powerful analytical tools, such as DM, remain too complex and sophisticated for the average consumer. Finally, but extremely important, it has not been given sufficient emphasis to the development of solutions that allow the specification of DM problems through business oriented languages, and that are also oriented for BI activities. With the expansion that has occurred in the application of DM solutions in BI, this is, currently, of increasing importance.

Most of the BI systems are built on top of relational databases. As a consequence, DM integration with relational databases is an important issue to consider when studying DM

integration with BI. Codd's relational model for database systems is long ago adopted in organizations. One of the reasons for the great success of relational databases is related with the existence of a standard language – SQL (Structured Query Language). SQL allows business users to obtain quick answers to ad-hoc questions, through queries on the data stored in databases. SQL is nowadays included in all the Relational Database Management Systems (RDBMS). SQL serves as the core above which are constructed the various Graphical User Interfaces (GUI's) and user friendly languages, such as Query-By-Example (QBE's), included in RDBMS. It is also necessary to define a standard language for data mining, which can operate likewise for data mining. Some efforts are being made in order to overcome this issue. Efforts involve the definition of standards for DM that arises both by academics and by people in the industry. It is the authors' belief that the effective integration of DM with BI systems must involve final business users' access to DM models. This access is crucial in order to business users to develop an understanding of the models, to help them in decision making. With this in mind, the authors conceived and implementes a DM language, named as Query-Models-By-Example (QMBE), which is iterative and interactive in nature, thus allowing final business users to access and manipulate DM models. (Azevedo & Santos, 2012a; Azevedo & Santos, 2012b). Despite the importance of this language, it only addresses classification rules. This paper presents an extension to QMBE in order to also include clustering. The conception and implementation of QMBE language, despite including only classification rules, contributed to verifying the viability of allowing business users of BI systems to manipulate directly DM models, thus providing the possibility to explore the potential value of applying DM in the context of BI. Likewise classification rules, clustering is an important DM task in the realm of BI systems. As a result, extending QMBE to include clustering constitutes an additional contribution to allowing business users of BI systems to directly manipulated models,

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/article/extending-qmbe-language-with-clustering/105931](http://www.igi-global.com/article/extending-qmbe-language-with-clustering/105931)

## Related Content

---

### A Conceptual Model for Measuring Reverse Logistics Performance in Automobile Industry

Kaveh Khalili-Damghani and Maryam Najmodin (2014). *International Journal of Strategic Decision Sciences* (pp. 21-29).

[www.irma-international.org/article/a-conceptual-model-for-measuring-reverse-logistics-performance-in-automobile-industry/114626](http://www.irma-international.org/article/a-conceptual-model-for-measuring-reverse-logistics-performance-in-automobile-industry/114626)

### Computing Skills in Forecasting for Liquidity Risk Management in the Indian Banking Industry

Rituparna Das (2014). *Emerging Methods in Predictive Analytics: Risk Management and Decision-Making* (pp. 172-185).

[www.irma-international.org/chapter/computing-skills-in-forecasting-for-liquidity-risk-management-in-the-indian-banking-industry/107905](http://www.irma-international.org/chapter/computing-skills-in-forecasting-for-liquidity-risk-management-in-the-indian-banking-industry/107905)

### An Extended Model of Decision Making: A Devil's Advocate Workshop

David Sammon (2008). *Encyclopedia of Decision Making and Decision Support Technologies* (pp. 348-355).

[www.irma-international.org/chapter/extended-model-decision-making/11273](http://www.irma-international.org/chapter/extended-model-decision-making/11273)

### Multi-Criteria Analysis and Consequential Lifecycle Assessment of Solid Waste Management: An Integrated Approach

Anna Bernstad Saraiva, João Namorado Clímaco, Rogerio de Aragão Bastos do Valle and Claudio Mahler (2021). *International Journal of Decision Support System Technology* (pp. 1-17).

[www.irma-international.org/article/multi-criteria-analysis-and-consequential-lifecycle-assessment-of-solid-waste-management/282128](http://www.irma-international.org/article/multi-criteria-analysis-and-consequential-lifecycle-assessment-of-solid-waste-management/282128)

### A Decision Support System for Evaluation of Investments in a Computer-Integrated Production System

Hannu Kivijärvi and Markku Tuominen (2008). *Encyclopedia of Decision Making and Decision Support Technologies* (pp. 200-210).

[www.irma-international.org/chapter/decision-support-system-evaluation-investments/11256](http://www.irma-international.org/chapter/decision-support-system-evaluation-investments/11256)