Pattern Warehouse:
A Dedicated Pattern Management System

P

Vivek Tiwari
Maulana Azad National Institute of Technology, India

Ramjeevan Singh Thakur
Maulana Azad National Institute of Technology, India

INTRODUCTION

Today's world produces an enormous amount of data in a regular basis from various sources. Data in such huge volumes do not constitute knowledge i.e., they cannot be directly exploited by human beings and no useful information can be deduced simply by their observation. Thus, more elaborate techniques are required in order to extract the hidden knowledge and make these data valuable to the end-users (Terrovitis & Vassiliadis, 2003). Data mining (Romero, Oscar, & Abelló, 2010) was developed to help extract Knowledge from the raw data, using algorithms that could discover several statistic properties in the original data. Data mining produces results like association rules, clusters, decision trees and other structures that describe properties of the raw data. The common characteristic of all these techniques is that big portions of the available data are abstracted and represented by a small number of knowledge-carrying representatives, which we call patterns (Tiwari & Thakur, 2012). Patterns represent the huge quantity of heterogeneous data in compact and rich semantics way.

Nowadays, knowledge and pattern extracting methods are efficient, dynamic and apply on different kinds of data. As a result a lot of different and complex patterns are extracted. In order to someone to be able to exploit these patterns, an efficient and global (general) Pattern Warehouse Management System (PWMS) is required for handling (storing / processing / retrieving) patterns. A pattern- warehouse is a collection of persistently stored patterns. Pattern is interesting because it describes a recurrent behavior (e.g., in market segmentation, stock exchange analysis, etc) (Schneider, S. & Dirk, F., 2012). The pattern despite being already the result of some elaboration on some raw data, are not, usually, in a form that can lead us directly to real life results (Terrovitis & Vassiliadis, 2003). We need tools that will permit us to compare, query and store the pattern in order to retrieve the information when we want. This chapter describes the architecture of a system named 'Pattern Warehouse Management system' that provides such tools for pattern manipulation.

BACKGROUND

Nowadays, databases are huge, dynamic, come from different application domains and a lot of different and complex patterns can be extracted. Ordinary database management system (DBMS) for handling patterns can be used, but they are insufficient to handle a variety of patterns since patterns are very specific in their domain and structure. DBMS are designed for handling data which are semantically very poor. Data is rough, pattern is refined. Because of data size, many organizations only store 1 to 2 years worth of historical data. However, because pattern is so much more compact than data, the Pattern Warehouse is only a fraction of the size of the data warehouse, allowing the patterns of many years to be stored with ease. So pattern warehouse is a way to make pattern non volatile and pattern

DOI: 10.4018/978-1-4666-5202-6.ch161

management system allow to user to store, retrieve update patterns efficiently (Tiwari & Thakur, 2012). The patterns are not persistent by nature (Bartolini, Bertino, Catania, Ciaccia, Golfarelli, Patella, & Rizzi, 2003). Pattern gets lost when it goes out of memory. There is long and complex process behind the pattern generation. It is time and resource consuming also.

RELATED WORKS

There are various methods and techniques proposed by researchers specially for handling patterns. Following section represented techniques, frameworks, architectures of related works.

1. Patterns for Next-Generation Database Systems

They claim that the concept of pattern is a good candidate for generic representation of these novel information types (Bartolini et al., 2003). They discuss the main issue related to pattern handling. The work presented preliminary discussion on pattern representation. Authors introduced and argue on necessity of a separate and dedicated a pattern management system. The work also outlined the architecture of pattern base management system (PBMS). Authors insist the use of dedicated pattern storage system by discussing variety of patterns is available in huge amount now days. They introduced new idea of non volatile and persistent pattern. The given system is characterized by the following properties:

1. **Abstraction:** Within a PBMS, patterns are made first-class citizens thus providing the user with a meaningful abstraction of raw data to be directly analyzed and manipulated.
2. **Efficiency:** Introducing an architectural separation between the PBMS and the DBMS improves the efficiency of both traditional transactions on the DBMS and advanced processing on patterns.
3. **Flexible Querying:** The PBMS provides an expressive language for querying the pattern-base in order to retrieve and compare patterns.

Within the PBMS, there are three different layers. The pattern layer is populated with patterns. Patterns with similar structural characteristics have the same type (either built-in or user-defined) that is described in the type layer. The class layer holds definitions of pattern classes, i.e., collections of semantically related patterns. Classes play the role of collections in the object-oriented context and are the key concept in the definition of a pattern query language.

2. Modeling and Language Support for the Management of Pattern-Bases

Pattern management system is very new concept but originally it is based on data warehouse and database theme. By this idea, database like setting are useful for modeling, storing and querying patterns (Terrovitis, Vassiliadis, & Skiadopoulos, 2007). It givens logical foundations for the global setting of pattern management through a models like E-R Model, Star schema, Galaxy schema, Snowflake etc that covers data, patterns. Pattern retrieval is also a big issue because ordinary structured query language is not suitable. Predicates based system is used for comparing patterns and defining query operators. Two aspects of pattern managements are important. First, the formal foundations of a well specified logical model. Second, need necessary mechanisms for the querying of patterns and data. First issue is related to conceptual design of pattern-warehouse. Authors consider the larger problem of modeling, storing, and querying patterns, in a database-like setting and use a Pattern-Base Management System (PBMS) for this purpose. Specifically, (a) they formally define the logical foundations for the global setting of pattern management through a model that covers data, patterns, and their intermediate mappings; (b) They present formalism for pattern