Knowledge Mining

Mahesh S. Raisinghani
Texas Woman's University, USA

INTRODUCTION

Numerous conferences and several articles in scholarly and business journals have tried to get a handle on knowledge. The growth of knowledge consulting organizations signals a growing conviction that knowing about knowledge is critical to business success. Multiple factors have led to the current knowledge boom. The perception and the reality of a new global competitiveness are one driving force. Rapid change and increasing competition for the dollars, marks, and yen of increasingly sophisticated consumers have led firms to seek a sustainable advantage that distinguishes them in their business environments.

Knowledge is neither data nor information, though it is related to both. Most people have an intuitive sense that knowledge is broader, deeper, and richer than data or information. More and more, business leaders and consultants talk about knowledge as the key to a sustainable competitive advantage. Knowledge workers, knowledge-creating company, knowledge capital, and leveraging knowledge have become familiar phrases (Davenport & Prusak, 1998; Turban, McLean, & Wetherbe, 2003).

During his keynote speech at the Information Resources’ annual meeting in Boston, Massachusetts, Venkatraman (1998) discussed how companies manage their knowledge assets and how organizations have moved from the industrial economy to the knowledge economy (see Figure 1).

The purpose of this article is to discuss the knowledge concept of knowledge mining and address the following questions:

- How are the concepts of data life cycle and knowledge discovery related?
- What is the taxonomy of knowledge mining and its benefits?
- What is the role of knowledge in software development?

BACKGROUND: DATA LIFE CYCLE AND KNOWLEDGE DISCOVERY

To better understand how to manage data and knowledge, it is necessary to trace how and where data flow in organizations. Businesses do not run on data, they run on information and their knowledge of how to put that information to use successfully. Everything from innovative product designs to brilliant competitive moves relies on knowledge. However, knowledge is not readily available. In many cases, it is continuously derived from data. However, because of the difficulties, a derivation may not be simple.

The transformation of data into knowledge may be accomplished in several ways. The process starts with data collection from various sources. These data are stored in a database followed by storage in a data warehouse. To discover knowledge, the processed data may go through a transformation that makes them ready for analysis. The analysis is done with data mining tools, which look for patterns, to support data interpretation. The result of all these activities is generated knowledge. Both the data, at various times during the process, and the knowledge, derived at the end of the process, may need to be presented to users by using different presentation tools. As illustrated in Figure 2, the created knowledge is stored in a knowledge base (Turban et al., 2003).

Figure 1. Transition from the industrial economy to the knowledge economy

<table>
<thead>
<tr>
<th>Production</th>
<th>Capital</th>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economy of Scale</td>
<td>Economy of Scope</td>
<td>Economy of Expertise</td>
</tr>
<tr>
<td>Design</td>
<td>Expertise</td>
<td>Leverage</td>
</tr>
<tr>
<td>Data</td>
<td>Information</td>
<td>Knowledge</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Action</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Result</td>
</tr>
</tbody>
</table>
KNOWLEDGE MINING: TAXONOMY AND BENEFITS

All decision support systems use data, information, or knowledge. These three terms are sometimes used interchangeably. Data items refer to an elementary description about things, events, activities, and transactions that are recorded, classified, and stored but are not organized to convey any specific meaning. Data items can be numeric, alphanumeric, figures, sounds, or images. Knowledge consists of data items that are organized and processed to convey understanding, experience, accumulated learning, and expertise as they apply to current problem or activity (Grinstein, 2003; Grinstein, Kobsa, Plaisant, Shneiderman, & Stasko, 2003; Last, Friedman, & Kandel, 2003; Tan, Kumar, & Srivastava, 2004; Turban & Aronson, 1998).

The mental processing and representation of knowledge are complex activities, and our understanding is still rudimentary and subject to debate. A general concept for describing knowledge is an elusive as ever, though various key concepts have been developed from specific viewpoints in the cognitive sciences. Another way to define knowledge is to consider the way it is stored in human memory. Here, knowledge refers to a permanent structure of information stored in memory.

Han Koperski, Melli, Wang, and Zaane (1995) used the term knowledge mining as a practical synonym of knowledge discovery, not as an extension of it. At present, the use of the term is strongly associated as a synonym of knowledge discovery and data mining. In contrast, the term software mining is a special kind of knowledge discovery wherein the source data is already in the form of rules or program code.

Knowledge mining consists of the following four integrated components, designed to seamlessly guide the extraction process and contribute to providing corporations with a concise understanding of their business rules (Aiken, Muntz, & Richards, 1994; Chiang, 1995; Weiss, Buckley, Kapoor, & Damgaard, 2003; Yang, Hongji, & Chu, 2001; Yang, Hongji, Chu, Cheng, & Zhan, 2001):

- **System-wide knowledge recovery** provides an overall view of the business processes supported by an application. The system-wide knowledge recovery facility enables analysts to identify the programs in which particular business rules exist and then extract those rules from applications.
- **Program-level analysis** enables the structure and interrelationships within programs to be revealed. The program focuses on variable usage, paragraph calls diagrams, GO TO diagrams, execution paths, and complex queries.
- **Business rule extraction** enables concise business rules to be extracted from within legacy programs and across entire legacy systems. Support is provided by
  - variable-based techniques, which enable a business rule to be extracted based upon a specific variable within a program.
  - value-based specialization techniques, which contain embedded data that can be greatly simplified and specific rules that can be uncovered.