

Database Keyword Search and Data Mining: Any Connections?

Zhengxin Chen
Department of Computer Science
University of Nebraska at Omaha
Omaha, NE 68182-0500, USA

Today's huge amount of data is inevitably stored in various types of databases, so any serious study involved in dataology has to be database-centered. The amount of data available today is usually referred to as an ocean. But the number of research fields or technologies used to explore the data is also a vast terrain. Two such fields are database keyword search (KWS) and data mining (DM).

Traditionally database search has been conducted in certain database languages, such as SQL. KWS is trying to change that. Because of the nature of the involved tasks, KWS and data mining are two very different research fields in database research: KWS is about DB content retrieval, is about the connection of individual tuples scattered in various tables, and is computational expensive by nature; while DM is about aggregation, summary, and finding hidden knowledge patterns behind the data. These two research fields are far apart from each other, requiring very different algorithms. But are there any interesting connections between them, since they are cultural siblings (both born at the end of last decade)? This is the topic we want to address in this presentation.

We start our discussion from a quick review on the nature of data mining, which is related to foundations of data mining. Aimed at finding hidden knowledge patterns, data mining is not only about finding intrinsic relationships among various pieces of data; as indicated in our earlier presentation [1], the next phase of data mining should be characterized by mining the *behavior* (such as physical processes or social events) which *produces* the complex form of data, rather than just focusing on the data itself (which appears on the surface). [Think about Brownian motion, for example.] Is there any role KWS can play in data mining or behavior mining? It could be the case, because by providing a unified treatment of structured, unstructured and semistructured data and offering a more "natural" way (since keywords are more closely related to natural language than to computer languages such as SQL) for integrated DB search, KWS can help better understanding the contents of the data. In this sense, KWS is potentially an effective way to assist data mining and behavior mining.

Although very different algorithms are needed in KWS and DM, technically they can be complementary to each other. For example, by conducting DB search in a more natural way, KWS may reveal unexpected associations among values of different attributes. Such kind of associations cannot replace standalone association rule mining, cluster analysis or other data mining tasks, but can offer significant insight (or direction) about what to be mined. In addition, KWS usually requires construction of intermediate data structures such as graphs or trees, based on which database search can be conducted. It is possible that some of these data structures also serve as the basis for certain data mining tasks, such as frequent subgraph mining.

KWS and data mining can also be related to each other through the use of ontologies. For example, ontologies can help identifying the right keywords for database search. Along with appropriate ontologies, KWS may also offer useful hints to data miners to determine what to mine.

Some recent research has already hinted about the connection between KWS and DM. For example, starting from exploiting a relational DB system for integrating and exploring complex interrelationships between a DB and a collection of potentially related documents, the approach described in [2] leads to the process of extracting and querying all the relationships. Querying the discovered relationships requires searching only the newly extracted relationships using keyword search.

This talk contains personal observations and speculations. The speaker is willing to share his thoughts and welcomes critics from colleagues with various research perspectives.

The organization of this talk is as follows:

- Brief review of KWS
- Brief review of data mining/text mining and behavior mining
- General discussion on differences and relationships between KWS and data mining
- A brief examination of some related work
- Issues to be explored and other remarks

References

[1] Z. Chen, From data mining to behavior mining, *Int. J. Inf. Tech. Dec. Making*, 5(4), 703-712, 2006.

[2] C. Garcia-Alvarado and C. Ordonez, Keyword search across databases and documents, *Proc. Keys'10*