

# Techniques and Applications of KDD

Willi Klösgen<sup>1</sup> and Jan Zytkow<sup>2</sup>

<sup>1</sup> German National Research Center for Information Technology  
D-53757 St. Augustin.

<sup>2</sup> Wichita State University

## Abstract

Increasing amounts of data are collected in application domains as diverse as finance, market data analysis, astronomy, manufacturing, and education. Large datasets can lead to knowledge essential in domain understanding and decision-making, but knowledge discovery requires intelligent and largely automatic methods and systems, supporting the analysts. The rapidly emerging field of Knowledge Discovery in Databases (KDD) provides these tools. It also assimilates methods of machine learning and discovery, statistics, databases, and visualization.

In this basic tutorial, we explain what is KDD and what are its basic objectives and methods. We introduce the knowledge discovery process that includes data pre-processing, selection, transformation, explorations for patterns, and qualification of pattern as knowledge. We present the most important data mining methods, the conditions for their successful application and their advantages in different application areas. Many methods and many forms of knowledge can be integrated in a modular, interactive and iterative KDD system. We discuss search efficiency, knowledge evaluation and refinement, the role of background knowledge, and languages designed to form different elements of knowledge. Finally we outline several key challenges.