

Duck-miner: A Tool for Discovering and Handling Knowledge from Large Databases

(abstract)

Anastasia Tzebeleka

Database Group, Information Systems Laboratory
Department of Informatics, University of Piraeus, Greece
<http://isl.cs.unipi.gr/db>

May 2005

The overall aim of this thesis was to extend WEKA, a freely available open-source data mining tool implemented in JAVA, in order to support new tasks such as storing/ retrieving extracted knowledge in/ from Databases. The elaboration of this project involved understanding the architecture of WEKA data mining platform as well as studying background theory and techniques of data mining, database management and Java programming.

More specifically, the thesis focused on the following areas:

- ✓ Extend WEKA to store in relational database tables the results (patterns) extracted by using machine learning and data mining techniques, so that stored knowledge can be easily reused further. We implemented this possibility for the results generated from four algorithms supported by WEKA: J48 (classification using decision trees), SimpleKMeans and EM (clustering), Apriori (association rule mining).
- ✓ Extend WEKA to enable the retrieval of patterns previously stored in relational database tables. Users can query the pattern bases, where results from past data mining algorithm executions are stored, through a friendly, interactive environment.
- ✓ Add new capabilities for the Apriori algorithm. Users can terminate execution at the exact moment when large itemsets are produced (and before association rules are formed by frequent itemsets) as well as control the content of the results being displayed.

The software developed as thesis result, called Duck-miner, is an extended version of WEKA containing all the current features of that tool as well as the new features mentioned above. Database management was handled by Microsoft SQL Server 2000, while source code was developed in Java. Like WEKA, Duck-miner is a user-friendly, powerful tool which can be used both for professional and educational purposes.