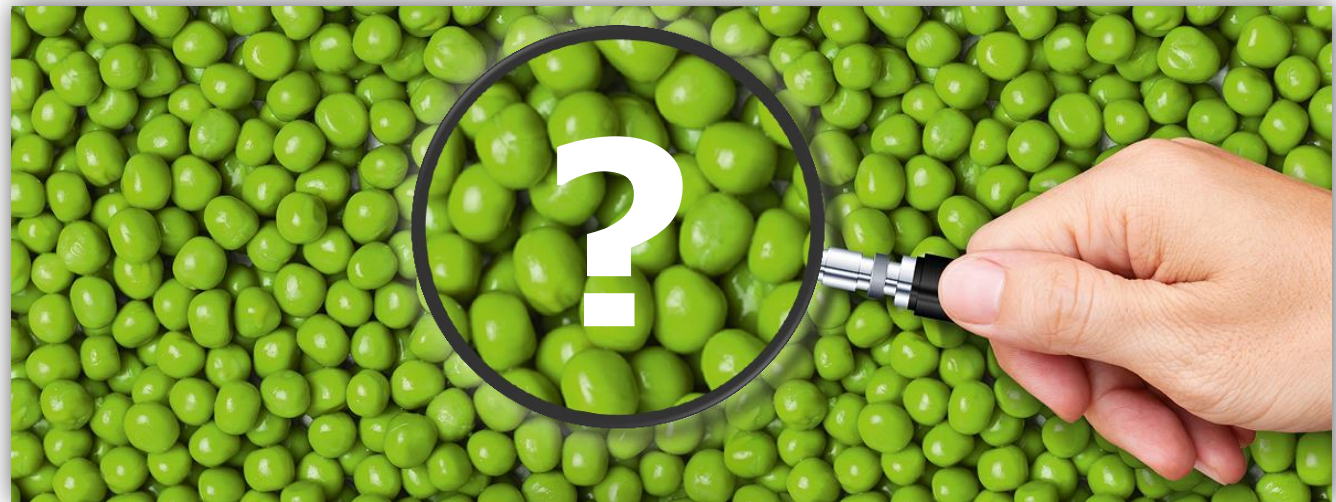


Advanced Data Profiling

Kick-off



Team Building

2

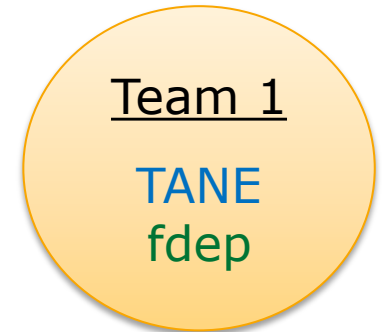


Topics

3

Group 1: TANE, FUN, FD-Mine

- Candidate generate-and-test approaches
- Pruning based



Group 2: Dep-Miner, FastFDs

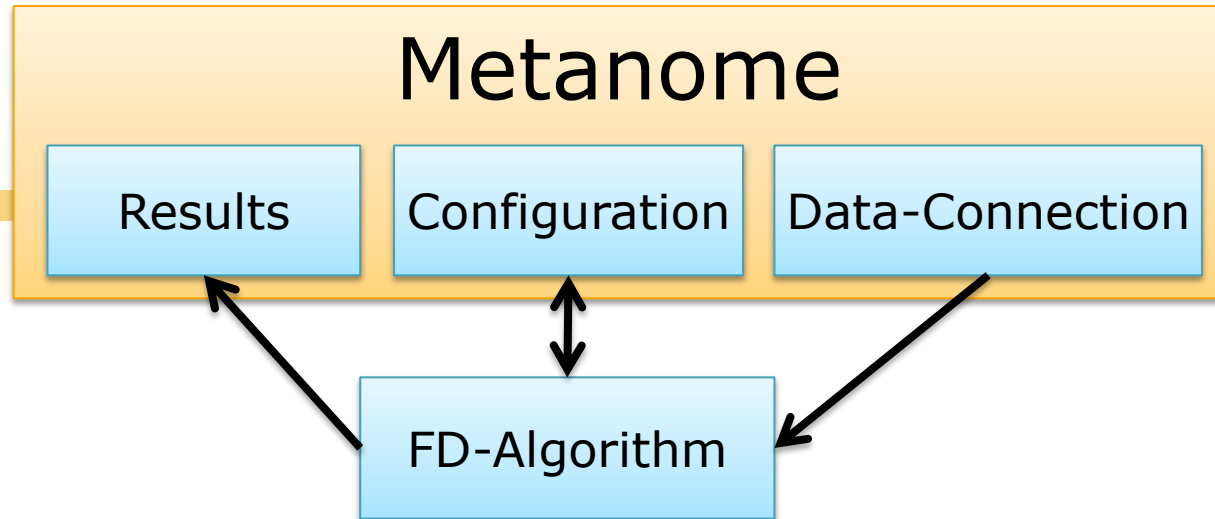
- Formal concept analysis approaches



Group 3: fdep

- Minimal cover approach
- Machine learning concepts





- Prerequisites:
 - Java 1.6
 - Maven
- Getting started:
 - Access Metanome via Maven repository (skeleton project)
 - Read the interface documentation and how-to-use guide
- Ongoing development:
 - Read the email-notifications for interface changes
 - Use "maven package **-U**" to get the new version
- Guidelines:
 - Use Maven for your dependencies
 - Use the fastUtils library for data structures with primitive types

Data

5

- Source format:
 - MySQL-Database
 - (Bonus: DB2 and CSV-Files)
- Dataset:
 - Plista Statistic Data
 - Your dataset!
- Be generic with your input source, e.g.
 - `Java.sql.DatabaseMetaData` not supported in all DB2 versions
 - `INTERSECT` not supported by MySQL

Organization

6

- Read the paper until next week
- Discuss your algorithms and decide for the most promising approach
- We then meet again in your teams (A-2.2):
 - 15:15 team 1
 - 15:45 team 2
 - 16:15 team 3
- Goal of this meeting:
 - Discussing your algorithms
 - Choosing the algorithm to be implemented (first)
 - Discussing first ideas for possible improvements