	Course	***Databases and Information Systems 2020***		
	Exercise Sheet	***7***		
	Points	–		
	Release Date	***July 7th 2020***	Due Date	***July 15th 2020***

1 Data Mining

Implement the Apriori algorithm for detecting frequent itemsets in shopping carts. The derivation of association rules is not required. For more information please refer to the lecture as well as the additional slides for this task.

Your results should contain lists of frequent itemsets containing {one, two, three, [...]} items each.

For the evaluation of your algorithm, please use the transactional data provided by us (see file `Transactions.zip` in Moodle) with a minimal support (minsup) of 1%: Every line is a whitespace-separated list of article identifiers and thus represents the articles acquired during a single transaction.

For Your Report

- Briefly describe how the algorithm and your software work.
- List how many frequent itemsets you obtained containing {one, two, three, ...} items.
  - Example: itemsets with 1 item: 120, itemsets with 2 items: 30, ... , itemsets with 5 items: 1
- Provide the found frequent itemsets containing two or more items together with their computed support value.
  - Example: [((A, B), 2.3%), ((A, C), 4%), ... , (A, B, C, D), 1.01%)]

Notes

- The transactional data can be read directly from the file and do not have to be imported into a database.
- You can use a programming language of your choice (as long we are able to understand the source code).
- You can use supporting libraries but no library(functions) which provide the full algorithm.
- Submit the report as pdf and a zip-archive with your source code (and a description how to run it) in Moodle.