

# Lab 3 – Pattern & Association Mining #1

*Data Mining, Spring 2017*

*Mathias Löwe (malw@itu.dk) | Daniel Sørensen (dafr@itu.dk)*

## *Today's Lab: Patterns!*

# Pattern & Association Mining #1

- Today you will be searching for frequent patterns in some simple transactional data.
- You will implement the apriori algorithm to accomplish this.
  - Page 248-254 (chapter 6.2.1-6.2.2) in the book.
- A simple code structure is provided to help you get started.

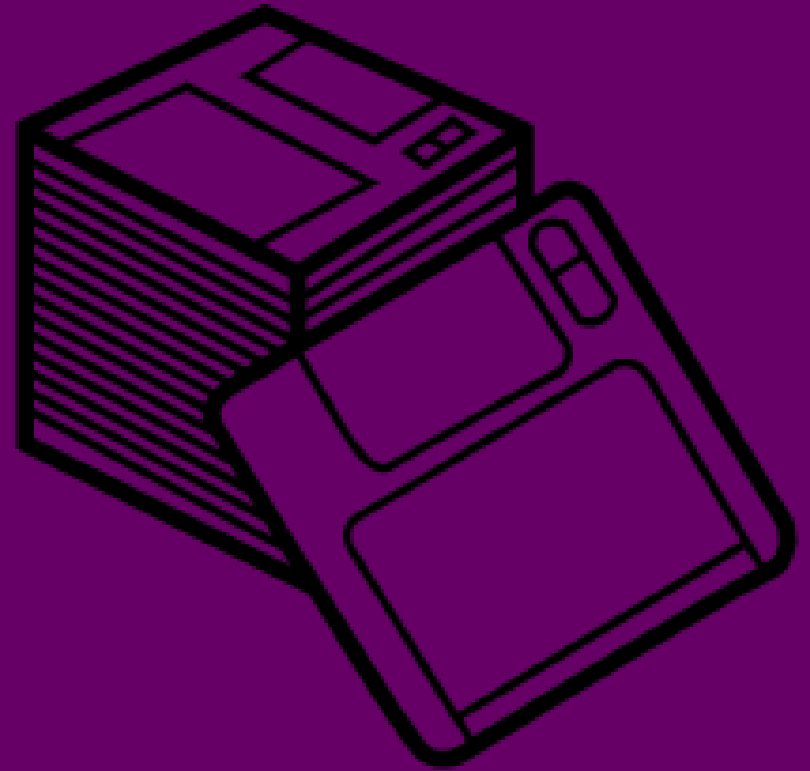
419385	05 374859 62	52534 637936	93190 461 94	98104 57418
4 724	896430 8939	37 978 13 95	04 986334408	26 038 6862
715830	12574 970 43	808 82012998	041498186 44	26441517385
64761	4349 90 7487	57 555864594	5053 086890	26375 974 86
458 79	349 75300499	5632 03 753	826378 29108	593036 8931
0 9722	41044 145 26	066789 8127	16303 08 430	88827 77 92
72812	3564976 409	2710339 8273	3394 0510 62	768031925644
29565	416634 1489	4616 914 68	46 2752 2312	0784989 072
605 72	102952413 8	210103 1888	47817 158 62	754933 58474
09883	81788993	4410 9469 36	47 21 13461	073 78315 27
029 56	8335 285 59	144010796329	1521811 572	0043527011
082347	95 44963 791	7306 29675	78185 0 7	113441 99090
2 541	4663 80 002	8 1962687	39386906 071	09 688 9122
317650	8 16 32 20	85 455 5109	328 17 58 66	8129 442 855
79 543	272634 0592	39 315028 89	34901 187 91	73237 71 32
27 55	43185317 391	497425 3938	0982035413	7395 0563588

# Code Provided

- Two classes
  - Apriori
  - ItemSet
- The ItemSet class is used to encapsulate information of sets of transaction items constructed during the algorithm.
- The Apriori class is where you should implement the algorithm.
  - Methods
    - Main
    - apriori
    - generateFrequentItemSets
    - joinSets
    - generateFrequentItemSetsLevel1
    - countSupport
- Code provided makes use of the HashTable java data structure, which is used to store <Key, Value> pairs. Values can then be retrieved based on their key. Is in this instance used to store <ItemSet, Integer> pairs, where the integer is used to store the support value for the item set.

# The Data

- The transactional data is simple and is only made up of integers
- See it as different records of sales, where each number is an item with id=1, id=2 and so on.
- The data set is provided in the code as the `TRANSACTIONS` two-dimensional integer array in the `Apriori` class.
- Also included is the two-dimensional integer array `BOOK_TRANSACTIONS` in the `Apriori` class
  - Has the data the book uses in its example
  - Use this to check your implementation



# Lab Overview

- First take a look at the code provided.
- Then start working on your apriori implementation
  - Suggested order of implementation of methods in the Apriori class:
    - countSupport
    - joinSets
    - generateFrequentItemSetsLevel1
    - generateFrequentItemSets
    - apriori
    - main

*Thanks for listening!*