



IT Systems Engineering | Universität Potsdam

Eclat - Algorithms for Pattern Mining

Uwe Hartmann, Peter Retzlaff

Agenda

2

1. Introduction
2. Eclat
3. Implementation
4. Evaluation
5. Extensions

Introduction

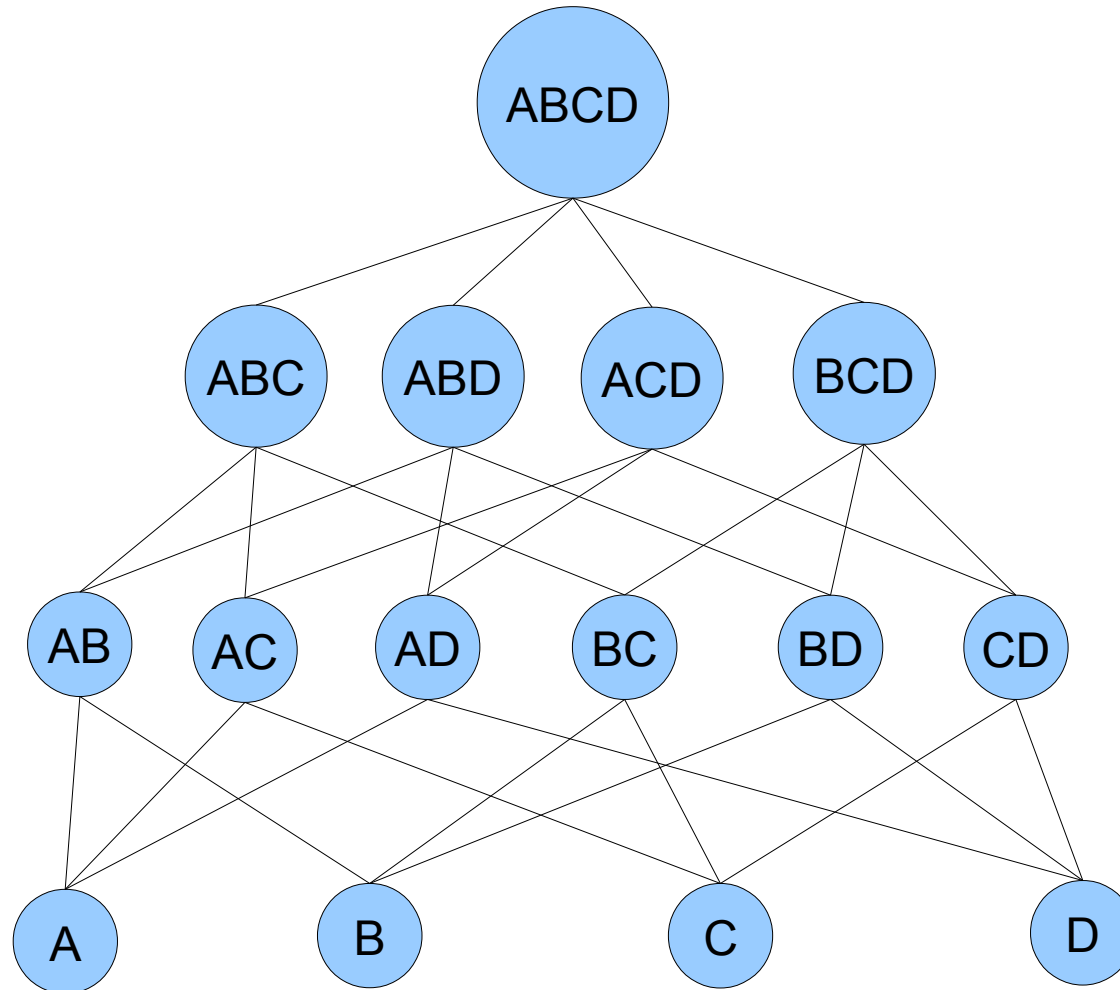
3

- M.J. Zaki, „Scalable Algorithms for Association Mining“, 2000

- challenges:
 - minimize I/O
 - limited main memory

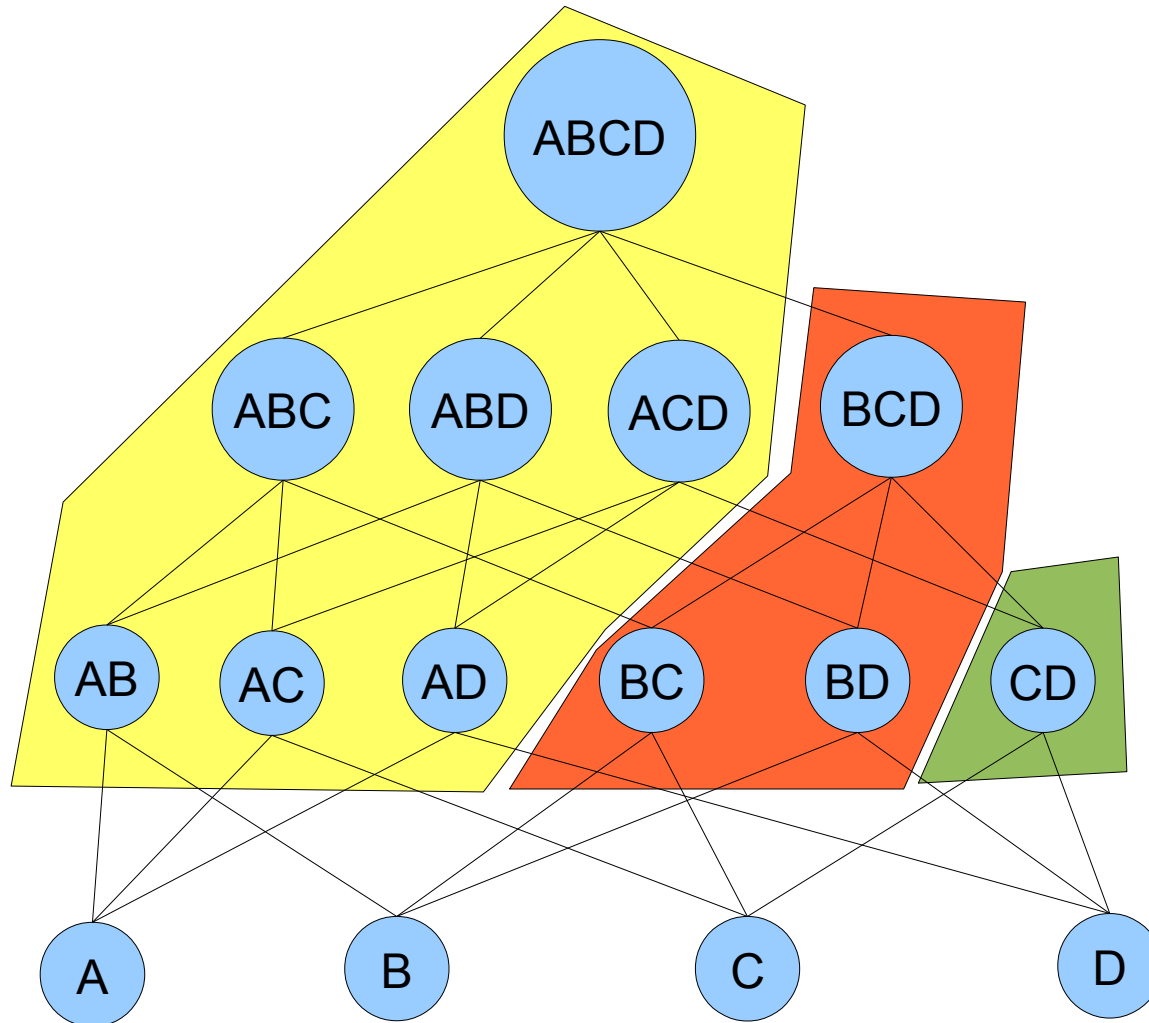
Eclat - Algorithm

4



Eclat - Partitioning

5



Eclat - Index

6

- use bottom-up search to find frequent itemsets in equivalence classes
- calculate support with inverted index

Item	Transactions
A	1, 2, 3, 4, 7, 8, 10
B	3, 5, 6, 7, 8
C	3, 4, 5, 8, 9
D	2, 10

Eclat - Example

7

Item	Transactions
beer	2,4,5
diapers	2,5
potato chips	1,2,4,5
soda	1,2,3,4,5



Item	Transactions
beer \cap diapers	2,5
beer \cap potato chips	2,4,5
beer \cap soda	2,4,5
diapers \cap potato chips	2,5
diapers \cap soda	2,5
potato chips \cap soda	1,2,4,5

Eclat - Example

8

Item	Transactions
beer \cap diapers	2,5
beer \cap potato chips	2,4,5
beer \cap soda	2,4,5
diapers \cap potato chips	2,5
diapers \cap soda	2,5
potato chips \cap soda	1,2,4,5



Item	Transactions
beer \cap diapers \cap potato chips	2,5
beer \cap diapers \cap soda	2,5
beer \cap potato chips \cap soda	2,4,5
diapers \cap potato chips \cap soda	2,5

Implementation

9

- F# (functional, .NET-based programming language)
- Data: Facebook-Likes →
public user profiles + list of liked pages
 - Likes: 44 000 000
 - Pages: 7 200 000
 - Profiles: 500 000
 - Database Size: 43GB

Implementation

10

CODE

Evaluation

11

■ Profiling

- 97.3% of time spend in intersect (not including time to load from db), Eclat only

■ Benchmarks:

- Eclat only, Msc = 10000; 9:00; peak memory 783MB
- Eclat only, Msc = 2000; Out of Memory
- Eclat with partition, Msc = 10000; 17:42; peak memory 996MB
- Eclat with partition, Msc = 2000; 306:17; peak memory 3744MB
- Eclat with partition, parallel, Msc = 2000; 118:54; peak memory 15054MB

Interesting

12

- Longest maximal frequent itemset with highest support:
 - Cooking, Jack Johnson, Kings Of Leon, MTV, Two and a Half Men, The Hangover, Harry Potter

Extensions

13

- Use Case
 - „people who liked this page, also liked...“

- Extension
 - LogEclat
 - MaxClique