



Försättsblad till skriftlig tentamen vid Linköpings universitet

(fylls i av ansvarig)

Datum för tentamen	2014-08-26
Sal	
Tid	8-12
Kurskod	TDDD41 + 732A31
Provkod	TEN1
Kursnamn/benämning	Data mining - clustering and association analysis
Institution	<i>IDA</i>
Antal uppgifter som ingår i tentamen	7
Antal sidor på tentamen (inkl. försättsbladet)	3 + cover page
Jour/Kursansvarig	Patrick Lambrix
Telefon under skrivtid	2605
Besöker salen ca kl.	9:30, 10:45
Kursadministratör (namn + tfnr + mailadress)	Carita Lilja, 1463, carita.lilja@liu.se
Tillåtna hjälpmedel	dictionary
Övrigt (exempel när resultat kan ses på webben, betygsgränser, visning, övriga salar tentan går i m.m.)	
Vilken typ av papper ska användas, rutigt eller linjerat	
Antal exemplar i påsen	7 + 1

Institutionen för datavetenskap
Linköpings universitet

EXAM
732A31 and TDDD41
Data Mining –
Clustering and Association Analysis
August 26, 2014, kl 8-12

Teachers: Patrick Lambrix, José M Pena

Instructions:

- Start each question at a new page.
- Write at one side of a page.
- Write clearly.
- If you make assumptions about a question, that are not explicitly stated, you need to write these down. (These assumptions cannot change the exercise or question.)

Help: dictionary

GOOD LUCK!

6. FP grow algorithm (2p+1p+1p+1p=5p)

- a. Run the FP grow algorithm on the following transactional database with minimum support equal to one transaction. Explain step by step the execution.

Transaction id	Items
1	C, B, A
2	D, C, A
3	A, B
4	A, B
5	A, D
6	A, D

- b. Repeat the exercise 6a with the following additional constraint: Find the frequent itemsets that do not contain the itemset CD. Explain step by step the execution. Make clear when and how the constraint is used. Incorporate the constraint into the algorithm, i.e. do not simply run the algorithm and afterwards consider the constraint.
- c. Let the items A, B, C and D have a price of respectively 1, 2, 3 and 4 units. Repeat the exercise 6a with the following additional constraint: Find the frequent itemsets whose total cost is greater than 2. Explain step by step the execution. Make clear when and how the constraint is used. Incorporate the constraint into the algorithm, i.e. do not simply run the algorithm and afterwards consider the constraint.
- d. What is the main advantage that the FP grow algorithm has over the Apriori algorithm ?

7. Constraints and lift (1p+1p+1p=3p)

- a. Give an example of a constraint that is both monotone and antimonotone. If you think it is not possible, explain why.
- b. Apply the Simple algorithm to the frequent itemset ABC on the database in exercise 6 in order to find association rules with confidence greater or equal than 50 %.
- c. Give an example of an association rule with lift greater than one and another example of a rule with lift smaller than one.