Submission s Assignment II	heet
	ontains questions to help you reflect about the operations applied to his assignment. You have to fill it in and submit it on Studium.
TASK 1: warm	ир
Number of recor	ds 9816 (after removing duplicates)
Number of attrib	utes (excluding User and Query) 500
TASK 2: freque	nt itemsets
What is the most	frequent item (i.e., single keyword)? "of"
	eries does it occur: 955
-	alue corresponds to a support count of 100 records? 100 / 98 6 = 0.0101874491
	ent itemsets have you found with this min-support? 28
• •	mum size of your frequent itemsets? 2
-	t of the support parameter
	ent itemsets for some selected values of support:
Mininum suppor	t Number of Frequent Itemsets 855
0.004	140
0.007	55
0.01	27
Which value of r	nin-support leads to the discovery of about 150 itemsets? 0.004 -> leads to 140
TASK 4: rule ge	eneration
How many rules	have you identified with min-confidence .8? 4
Indicate a high-c	onfidence rule $X \rightarrow Y$ where $Y \rightarrow X$ has lower confidence.
	York → new

Why is the confidence of $Y \rightarrow X$ lower than the confidence of $X \rightarrow Y$?

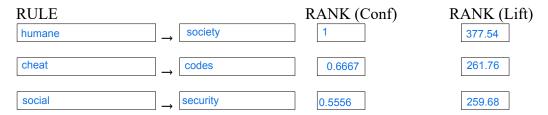
TASK 5: impact of confidence

Number of rules for some selected values of minimum confidence:

Mininum confidence	Number of Rules
0.1	18
0.3	14
0.5	7
0.7	5
0.9	1

TASK 6: rule interpretation

What are the top-3 interesting rules, and what is their rank w.r.t. confidence and lift?



What are the top-3 unexpected rules, and what is their rank w.r.t. confidence and lift?

