Hands-on: Frequent pattern mining and association rules

The objective is to process a set of documents related to questions asked during an exam. You will find the documents attached to this exercise.

Data: see Arche (Données pour mini projet)

Questions:
For *cf* folder do:

1. Load the set of documents in the *cf* folder.
2. Count, sort and display the number of words in all documents.
3. Filter by removing the words using French stopwords. You will find a document containing the stop words in French-Stopwords.txt
4. Display the top-10 of words.

The goal of the next questions is to apply frequent itemset mining and association rules algorithms

5. Transform each document into a set of transactions: one transaction = one document
6. Filter by by removing the words using French stopwords in each transaction.
7. Apply frequent itemset mining algorithm by varying the ***minsup***=0.2, 0.3, 0.5, 0.8
8. Sort the results by displaying the top-k
9. Apply the association rules algorithm by varying the ***minconf***=0.2, 0.5, 0.8
10. Sort the results by displaying the top-k

For *cp* folder do:

1. the same question [1-10]

Send me the Spark source code and a Readme file.