Assignment 5

Amazon wants to study the purchasing behavior pattern of their online customers. The problem they are specifically interested in is recommendation, i.e., recommending to a customer the products he/she might be interested in. To achieve this goal, we first construct a table like the following

Customer	Purchased Products	Date
1	A,B,C,N	01/03/2013
1	B,C,D,A,F	02/05/2013
2	A,C,V,N,J	03/03/2013
3	A,C,J,D	03/04/2014

where the first column is customer ID, the second column represents the products they purchased at the date in the third column. Those capital letters (A,B,C,B,F,J,N,V) are products.

Then we will mine "frequent purchasing patterns" from those historical purchasing records. Here every purchasing pattern is a set of products that the customers are frequently buying. To describe how frequent that a set of products were purchased, we define a concept called support. The support of a set of products is just the number of purchasing records that contain those products, divided by the total number of purchasing records. For example, for the purchasing record table shown above, the support of product A is 100%, the support for A,B,C is 25%, the support for A,C is 100%, the support for B,C is 50%. We call a purchasing product set is frequent, if its support exceeds a pre-specified threshold, and the number of events contained in a purchasing products is usually referred to as the length of the product set.

Please write a java program, that takes a purchasing record table file (the file could be in cdv or txt file format), and a support threshold as the input parameters. Output all (i.e., of different lengths) frequent purchased product sets whose support exceed the threshold.

For example, if your input table is the above and your support threshold is 50%, then you should output
{A};{B};{C};{N};{J};{D};{A,B};{A,C};{B,C};{A,N};{A,J};{A,D};{C,N};{C,J};{C,D};{A,B,C};{A,C,N};
{A,C,D};

You must upload a *single* ".java" file containing your source code. Your public class must include a public static method called "main" from which the execution starts.