ID2222 Assignment report 2

Discovery of frequent item-sets and association rules

Authors:

- Marco Dallagiacoma marcoda@kth.se
- Roberto Bampi bampi@kth.se

The purpose of this second assignment is to implement a system to discover frequent sets of items and their association rules.

The code is implemented as a pure-python file and it is compatible with python versions 3.5 and above. To run it, it is sufficient to execute the file using the python interpreter.

```Bash
~ $ python main.py dataset.dat 1000 0.5
# where dataset.dat is the input dataset,
# the second parameter is the support threshold (as the absolute number)
# and the third parameter is the confidence threshold
```

Solution

The solution is implemented as two python functions.

`apriori` takes a list of sets (the transactions) and the support threshold and returns a dictionary containing the frequent item-sets as keys and the support threshold as value.

`compute_association_rules` takes the frequent itemsets and the confidence threhsold and returns the associations as a set of pairs of itemsets in the form
`{((i1, i2, i3) -> (i4, i5, i6)), ...}`

The results are then printed on console.

Sample execution

Below is the execution of the program using the `simple.dat` dataset provided with the exercise.

Bash

```
~ $ python3 main.py simple.dat 2 0.5
FREQUENT ITEMSETS
-----------------
('B', 'E')
('E',)
('B', 'C', 'E')
('C',)
('B', 'C')
('B',)
('A',)
('A', 'C')
('C', 'E')
-----------------

ASSOCIATION RULES
-----------------
('E',) -> ('C',)
('B',) -> ('C', 'E')
('B',) -> ('E',)
('C',) -> ('E',)
('B', 'E') -> ('C',)
('B',) -> ('C',)
('E',) -> ('B',)
('C',) -> ('B', 'E')
('C',) -> ('B',)
('B', 'C') -> ('E',)
('C', 'E') -> ('B',)
('A',) -> ('C',)
('C',) -> ('A',)
('E',) -> ('B', 'C')```
```