**Part B Report**

**Maya Tanne 315342253 | Gal Bodek 316047968**

**Project Goal:** Identifying distinguishing patterns (Gene Teams) between two classes of bacteria.

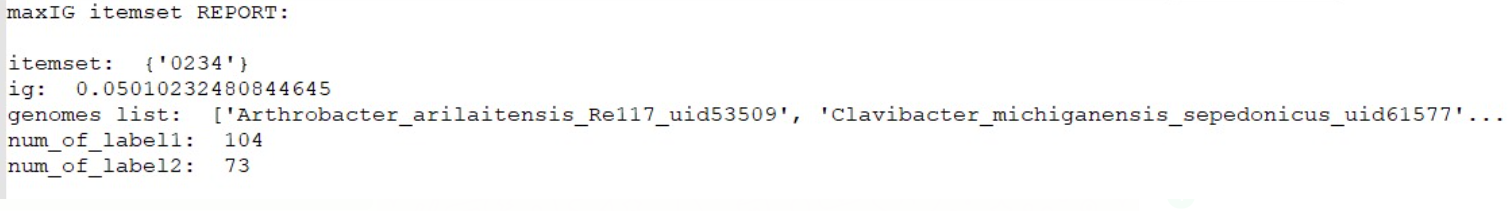
**Chosen Class labels**: Animal, Plant

**Number of Animal and Plant in *bactTaxa\_habitat***: 250 (see: taxa\_of\_classes.txt)

**Number of Transactions**: 190

(The reason for the difference between these two numbers is the fact that some uid's in *bactTaxa\_habitat* do not exist in *cog\_words\_bac* file)

**MIN\_SUP**: 172

**The Output Report of our program**:

**Conclusions:**

Our goal was to find itemsets which best distinguish between the two classes we chose: Animal and Plant. In practice, our searching space is very wide, hence our program could not handle it without choosing a high MIN\_SUP. As a result, on the first run on the original dataset, the itemset with the max IG was very common in both class labels. That caused our dataset to be significantly reduced when removing the transactions that contain that itemset. Therefore, on the next run, when updating the FPtree, no itemset met our MIN\_SUP, the tree was empty, and the program ended.

The only itemset we received was very common in both class labels and was not informative enough to draw conclusions from.

Upon dealing with part C, we hope to improve the efficiency of our program so we could run it with a lower MIN\_SUP, to get more informative results.