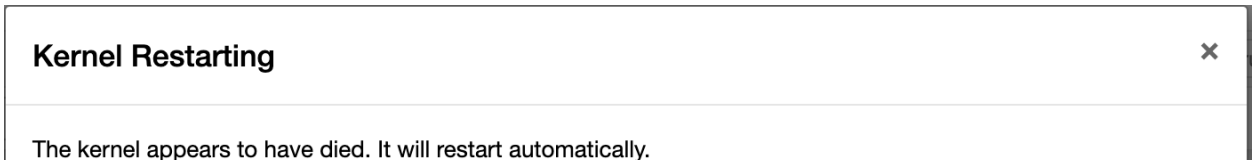
Problem A

In [2]:
```
import pandas as pd
import numpy as np
from collections import defaultdict

def process_data(file_path):
    # read the file
    with open(file_path, 'r') as file:
        lines = file.readlines()

    # create a dictionary to hold each item per transaction
    itemsets = defaultdict(list)
    all_items = set()

    # loop over each line and add items to the dictionary
    for i, line in enumerate(lines):
        items = line.strip().split(' ')
        all_items.update(items)
        for item in items:
            itemsets[i].append(item)

    # create a DataFrame with zeros for all items and all transactions
    df = pd.DataFrame(0, index=np.arange(len(itemsets)), columns=sorted(lis

    # update the DataFrame with the items in each transaction
    for trans, items in itemsets.items():
        df.loc[trans, items] = 1

    # print ARFF file header
    print("@RELATION news_items")
    for column in df.columns:
        print(f"@ATTRIBUTE item_{column} {{0, 1}}")
    print("@DATA")

    # print ARFF file data
    for i, row in df.iterrows():
        print(','.join(map(str, row.tolist())))
```

In [ ]:
```
process_data("./kosarak.dat.txt")
```

Question B

My Jupyter keeps crashing when I try to run this so I don't know the exact time.



In [1]:
```
import time

start_time = time.time()

# process_data("./kosarak.dat.txt")

end_time = time.time()

print(f"The script took {end_time - start_time} seconds to run.")
```

```
The script took 0.00014519691467285156 seconds to run.
```

Problem C

Same issue as problem 2, I don't have the actual file so I can't do it. If I did, here are the steps I would follow based on my research:

- Download the Weka application.
- In the Weka GUI Chooser window, click on "Explorer".
- In the "Preprocess" tab, click on the "Open file..." button, and open the .arff file.
- Then onte the summary of the data in the "Current relation" panel.

Problem D

Since I don't have the file, can't execute on this one either. But here are the steps I would have followed:

- In the "Preprocess" tab, click on "Open file..." to open the ARFF file.
- Select the FP-Growth algorithm: Go to the "Associate" tab, click on the "Choose" button, and select "FPGrowth" from the list.
- Configure the algorithm: Click on the box next to the "Choose" button to open the configuration dialog. Here, we can set the minimum support and confidence levels. The minimum support should be set to 49500/990002 (the fraction of the total number of instances), and the minimum confidence to 0.99. Close the dialog by clicking "OK".
- Run the algorithm: Click on the "Start" button to start the algorithm. The results will be displayed in the right-hand panel.

Problem E

Same issue as above