Name:Sharvari Yashwant Patil
PRN NO:2020BTECS00077

TY B.Tech. (CSE) – II [2022-23]

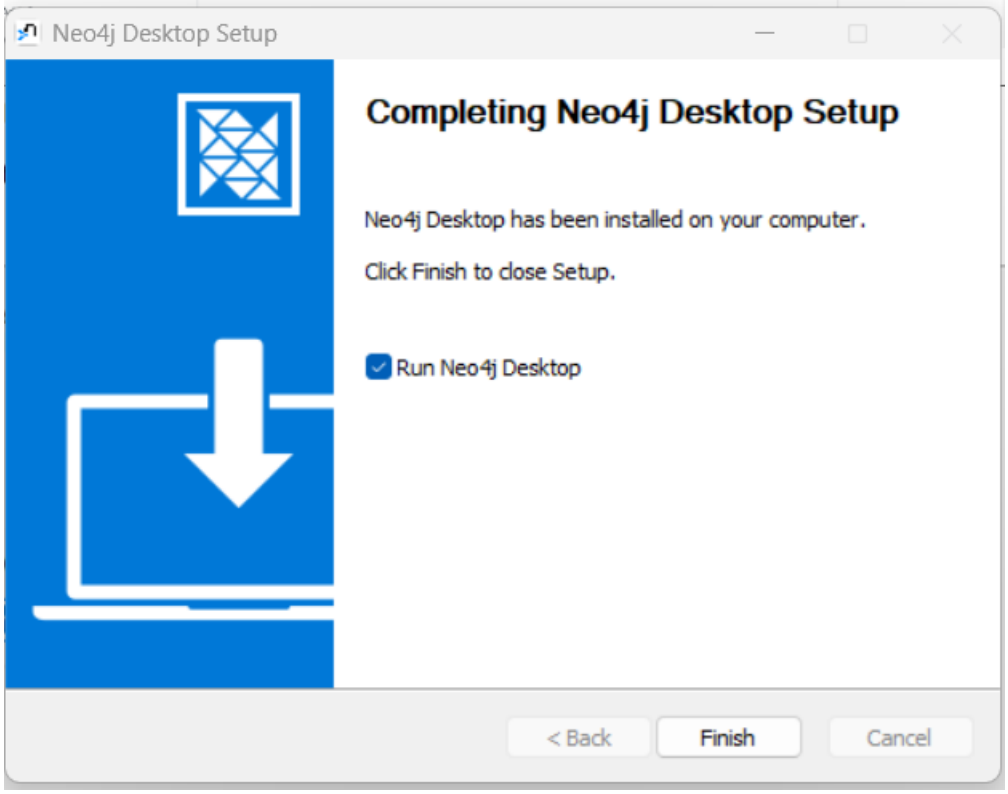
5CS372 : Advanced Database System Lab.

Assignment No. 11

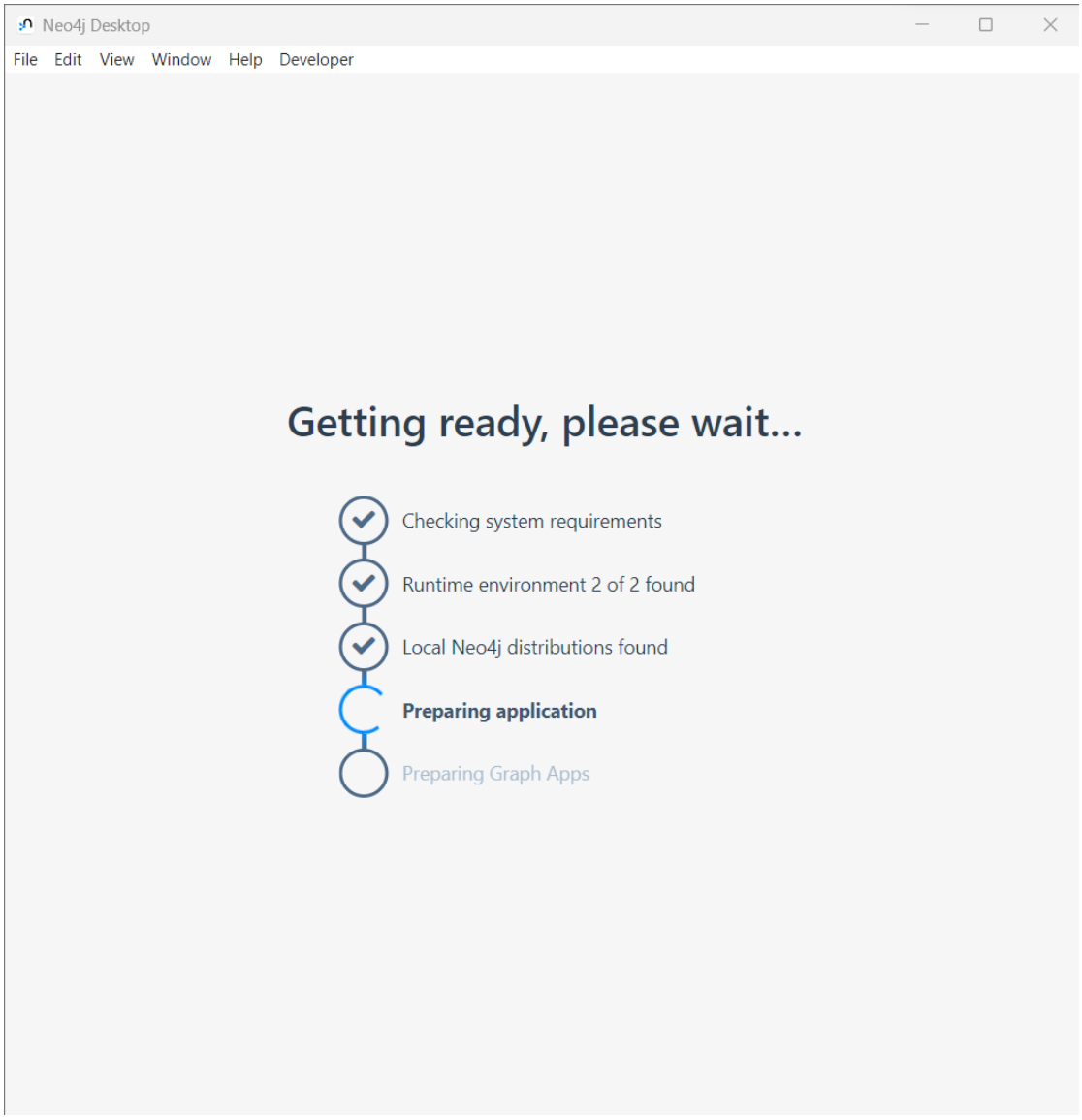
Neo4j Graph Database

Consider the "**Research Papers Database"** scenario as follows :

The research papers have authors (often more than one). Most papers have a classification (what the paper is about). The classifications form a hierarchy in several levels (for example, the classification "Databases" has the subclassifications "Relational" and "Object-Oriented"). A paper usually has a list of references, which are other papers. These are called citations.



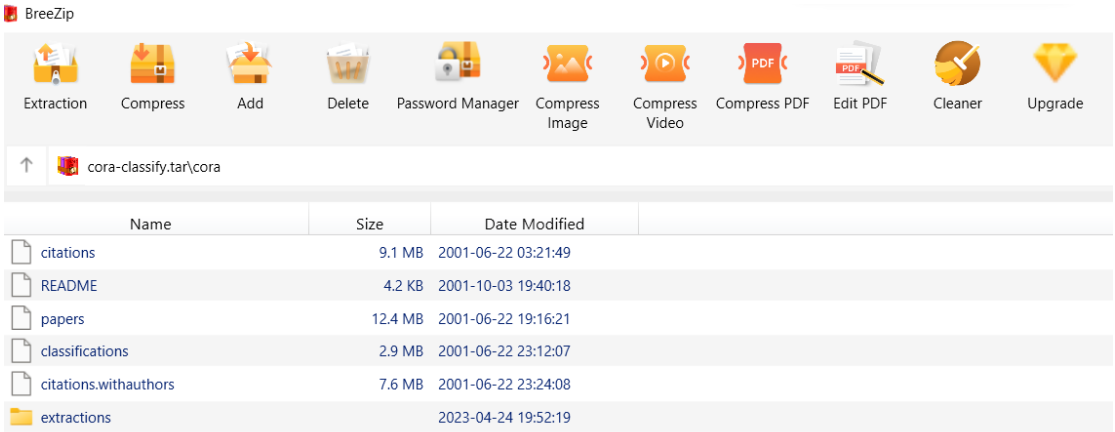
Name:Sharvari Yashwant Patil
PRN NO:2020BTECS00077



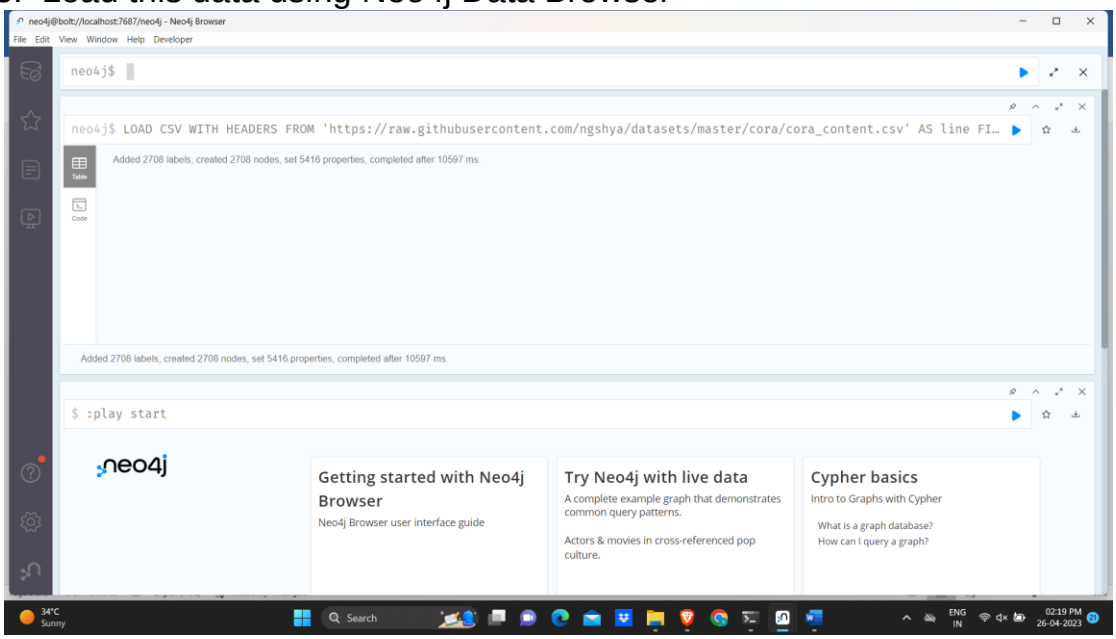
1. Design/model the graph database using Neo4j for above scenario.

2. Download the raw data from **Cora Research Paper Classification Project** : http://people.cs.umass.edu/~mccallum/data.html The database contains approximately 25,000 authors, 37,000 papers and 220,000 relationships.

Name:Sharvari Yashwant Patil
PRN NO:2020BTECS00077



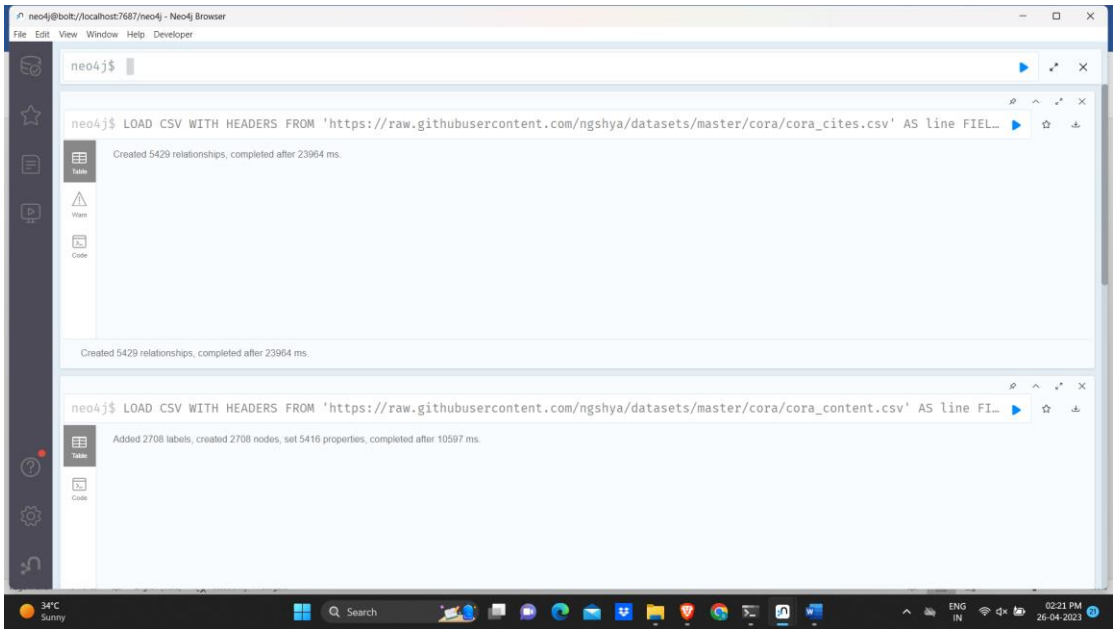
3. Load this data using Neo4j Data Browser



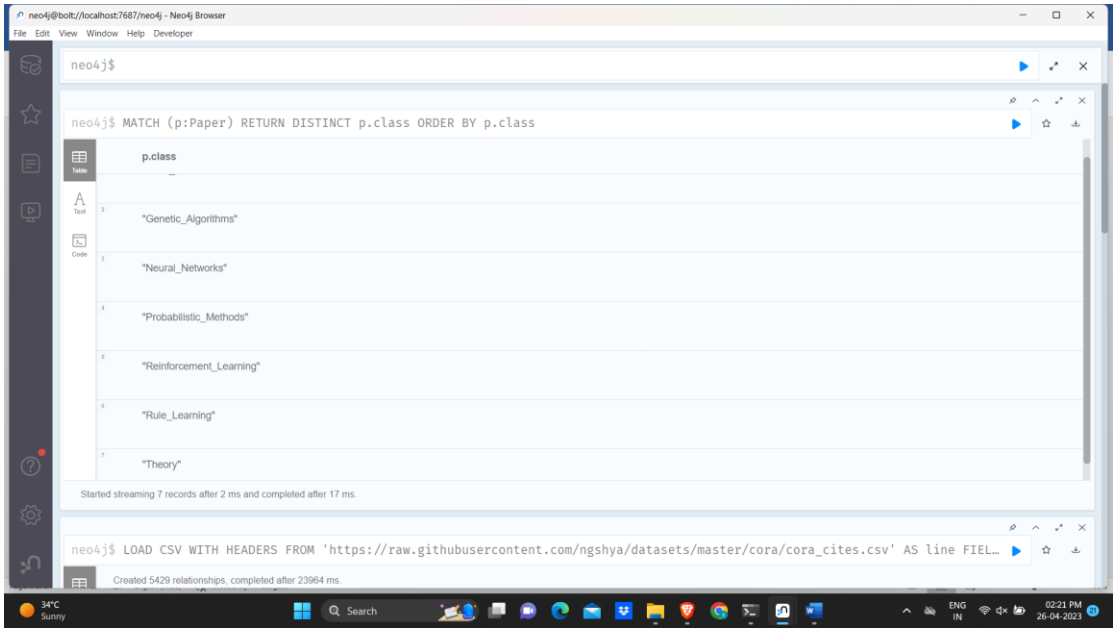
LOAD CSV WITH HEADERS FROM
'https://raw.githubusercontent.com/ngshya/datasets/master/cora/cora_content.csv'

AS line FIELDTERMINATOR ','
CREATE (:Paper {id: line.paper_id, class: line.label})

Name:Sharvari Yashwant Patil
PRN NO:2020BTECS00077



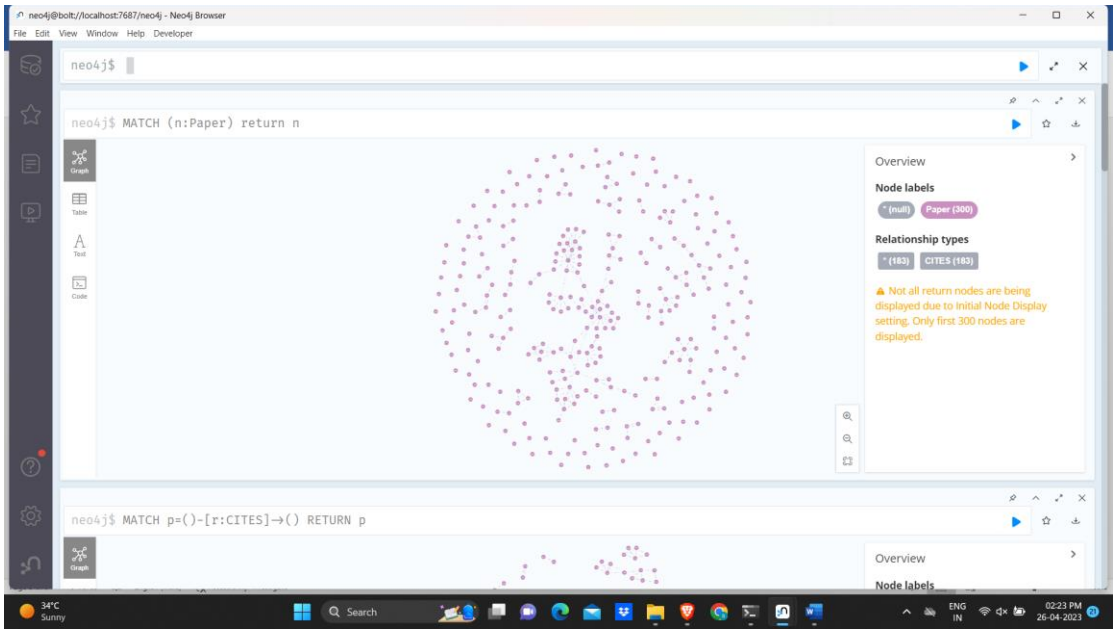
LOAD CSV WITH HEADERS FROM
'https://raw.githubusercontent.com/ngshya/datasets/master/cora/cora_cites.csv'
AS line FIELDTERMINATOR ','
MATCH (citing_paper:Paper {id: line.citing_paper_id}),(cited_paper:Paper {id: line.cited_paper_id})
CREATE (citing_paper)-[:CITES]->(cited_paper)



MATCH (p:Paper) RETURN DISTINCT p.class ORDER BY p.class

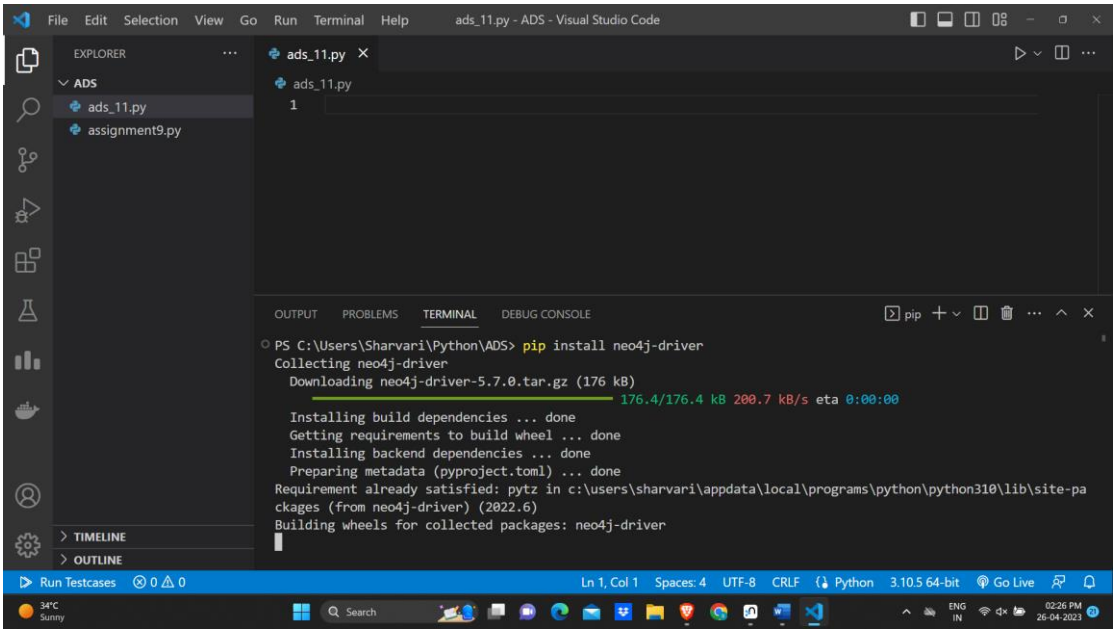
Name:Sharvari Yashwant Patil
PRN NO:2020BTECS00077

MATCH p=()-[r:CITES]->() RETURN p



4. Design the python based desktop application for any kind of search on above database. The application should able to answer queries like

Name:Sharvari Yashwant Patil
PRN NO:2020BTECS00077



a) Does paper A cite paper B? If not directly, does paper A cite a paper which in its turn cites paper B? And so on, in several levels.

```
import sys
import os
import tkinter as tk
from tkinter import *
import tkinter.messagebox

# For Neo4j Connection
from neo4j import GraphDatabase

class Neo4jConnection:

    def __init__(self, uri, user, pwd):
        self.__uri = uri
        self.__user = user
        self.__pwd = pwd
        self.__driver = None
        try:
            self.__driver = GraphDatabase.driver(
                self.__uri, auth=(self.__user, self.__pwd))
        except Exception as e:
            print("Failed to create the driver:", e)

    def close(self):
        if self.__driver is not None:
            self.__driver.close()

    def query(self, query, db=None):
```

Name:Sharvari Yashwant Patil
PRN NO:2020BTECS00077

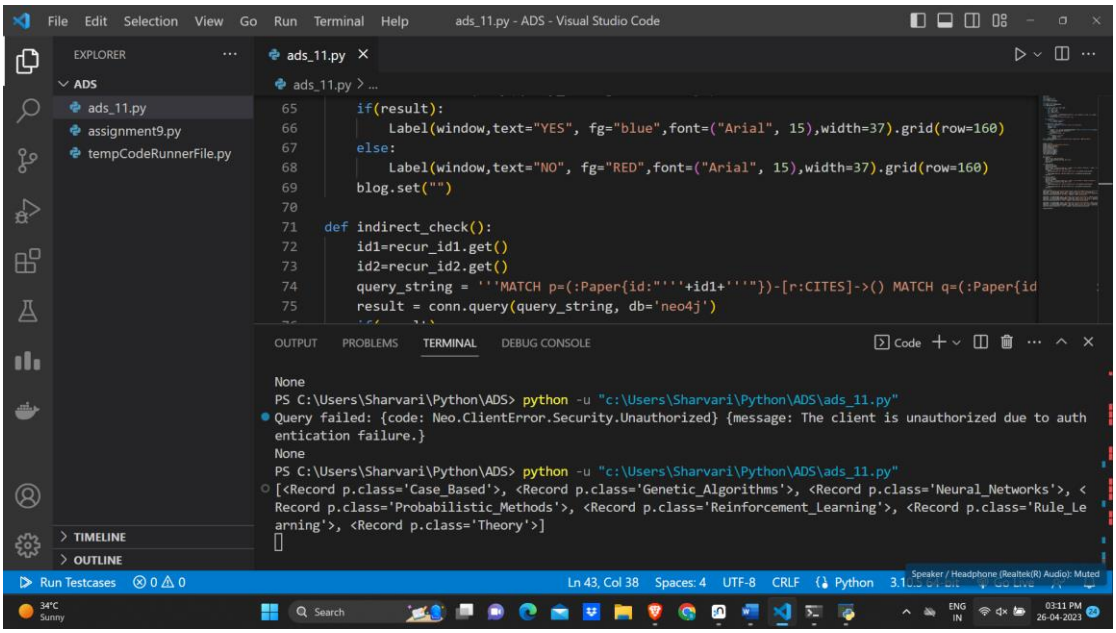
```
        assert self.__driver is not None, "Driver not initialized!"
        session = None
        response = None
        try:
            session = self.__driver.session(
                database=db) if db is not None else self.__driver.session()
            response = list(session.run(query))
        except Exception as e:
            print("Query failed:", e)
        finally:
            if session is not None:
                session.close()
        return response

conn = Neo4jConnection(uri="bolt://localhost:7687", user="neo4j", pwd="neo4j")
# ^ Neo4j Connected

window = tk.Tk()
window.title("Desktop App by Sagar")
window.geometry("700x500")
window.configure(bg="grey")
blog = tk.StringVar()
blog_title = tk.StringVar()
direct_id1 = tk.StringVar()
direct_id2 = tk.StringVar()
recur_id1 = tk.StringVar()
recur_id2 = tk.StringVar()

# submitting query

def submit():
    query_string = blog_title.get()
    result = conn.query(query_string, db='neo4j')
    print(result)
    blog.set("")

def direct_check():
    id1 = direct_id1.get()
    id2 = direct_id2.get()
    query_string = '''MATCH p=(:Paper{id:"'''+id1 + \
        '''"})-[r:CITES]->(:Paper{id:"'''+id2+'''"}) RETURN p'''
    result = conn.query(query_string, db='neo4j')
    if (result):
```

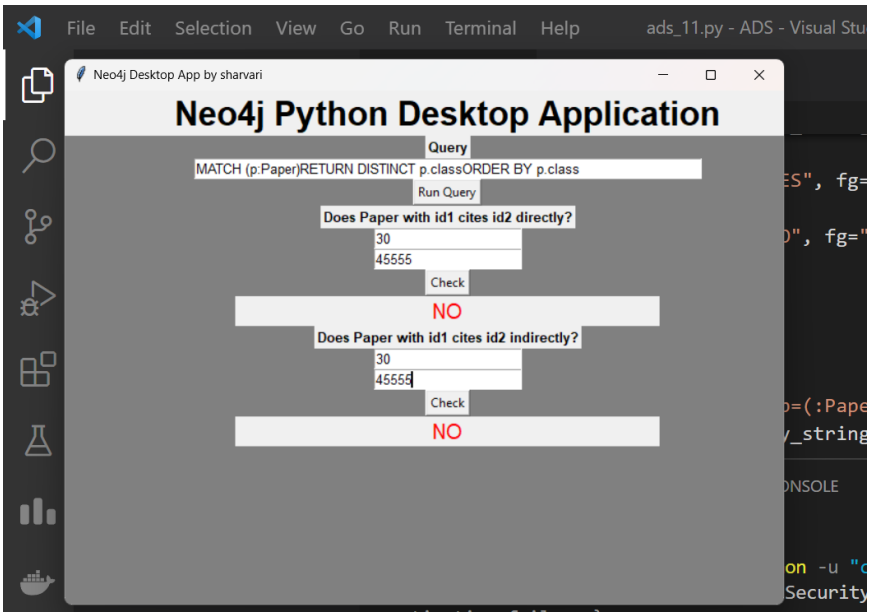
Name:Sharvari Yashwant Patil
PRN NO:2020BTECS00077

```
        Label(window, text="YES", fg="blue", font=(
            "Arial", 15), width=37).grid(row=160)
    else:
        Label(window, text="NO", fg="RED", font=(
            "Arial", 15), width=37).grid(row=160)
    blog.set("")

def indirect_check():
    id1 = recur_id1.get()
    id2 = recur_id2.get()
    query_string = '''MATCH p=(:Paper{id:"'''+id1 + \
        '''"})-[r:CITES]->() MATCH q=(:Paper{id:"'''+id2+'''"})
RETURN q'''
    result = conn.query(query_string, db='neo4j')
    if (result):
        Label(window, text="YES", fg="blue", font=(
            "Arial", 15), width=37).grid(row=220)
    else:
        Label(window, text="NO", fg="RED", font=(
            "Arial", 15), width=37).grid(row=220)
    blog.set("")

# tkinter window
Label(window, text="Neo4j Application", fg="black",
      font=("Arial", 25, 'bold'), width=37).grid(row=0, column=0)
name_label = tk.Label(window, text='Query', font=(
         'calibre', 10, 'bold')).grid(row=70)
     name_entry = tk.Entry(window, textvariable=blog_title,
font=(
         'calibre', 10, 'normal'), width=70).grid(row=80)
     sub_btn = tk.Button(window, text='Run Query',
command=submit).grid(row=110)

     name_label = tk.Label(window, text='Does Paper with id1
cites id2 directly?', font=(
         'calibre', 10, 'bold')).grid(row=120)
     name_entry = tk.Entry(window, textvariable=direct_id1,
                           font=('calibre', 10,
'normal')).grid(row=130)
     name_entry = tk.Entry(window, textvariable=direct_id2,
                           font=('calibre', 10,
'normal')).grid(row=140)
     sub_btn = tk.Button(window, text='Check',
command=direct_check).grid(row=150)

     name_label = tk.Label(window, text='Does Paper with id1
cites id2 indirectly?', font=(
```

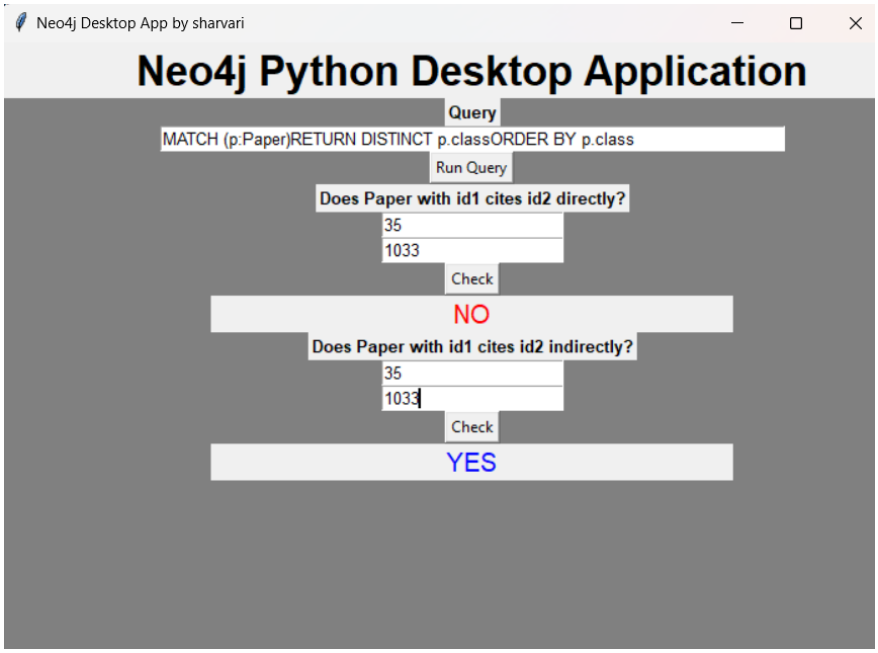
Name:Sharvari Yashwant Patil
PRN NO:2020BTECS00077

```
            'calibre', 10, 'bold')).grid(row=180)
        name_entry = tk.Entry(window, textvariable=recur_id1,
                              font=('calibre', 10,
'normal')).grid(row=190)
        name_entry = tk.Entry(window, textvariable=recur_id2,
                              font=('calibre', 10,
'normal')).grid(row=200)
        sub_btn = tk.Button(window, text='Check',
command=indirect_check).grid(row=210)

        window.mainloop()
```





Name:Sharvari Yashwant Patil
PRN NO:2020BTECS00077



a) Show the full classification of a paper (for example, Databases / Relational)

Note : Follow the submission guidelines.

Deadline : 16/04/2023

Dr. B. F. Momin
Course Coordinator