



**МІНІСТЕРСТВО ОСВІТИ, НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ**

**Лабораторна робота №9
Аналіз даних**

Підготував:

студент 4 курсу

групи ФІ-84

Коломієць Андрій Юрійович

E-mail: andrew.kolomiets.work@gmail.com

Київ – 2021

Лабораторна робота №9

Аналіз даних

Завдання на самостійну роботу

1. Детально ознайомитись із засобами перетворення типів даних в мові **Python**.
2. Самостійно встановити на комп'ютері систему **Gephi**.
3. Ознайомитись з основними режимами і можливостями системи **Gephi**.

Виконання завдання

Формування матриці суміжності концептів

Code

```
#!/bin/python3

import re
import string
import numpy as np

f = open("./Apple.json","r")
t = f.read()
f.close()

json =t.split('\n')
t=""

for i in range(len(json)):
    t=t+" "+json[i]

title = re.findall('"title" : "(.+)"source"', t)
t=""

for i in range(len(title)):
    t=t+" "+title[i]

t=re.sub('"textBody" :','.',t)
t=re.sub('[--%--||[\]\0-9",()$+»«-:;_...]', ' ',t)
t=t.upper()
```

```

sent =t.split('.')
print(sent)
f = open("./word.txt","r")
t = f.read()
f.close()
t=t.upper()
w =t.split('\n')

for i in range(len(w)-1):
    s = w [i].split()
    w[i]=s[1]

mtr = np.eye(len(w))

for i in range(len(w)):
    mtr[i][i]=0

stroka=""

for i in range(len(w)):
    stroka=stroka+";"+w[i]

print(stroka)

for i in range(len(sent)):
    for j in range(len(w)):
        for k in range(len(w)):
            if(j!=k):
                if (re.search(w[j],sent[i])):
                    if (re.search(w[k],sent[i])):
                        mtr[k][j]=mtr[k][j]+1

for i in range(len(w)):
    stroka=w[i]+";"
    for j in range(len(w)):
        a=int(mtr[i][j])
        b=a.__str__()
        if (j<len(w)-1):
            stroka=stroka+b+";"
        else:
            stroka=stroka+b
    print(stroka)

```

Using for regular expression

<https://regex101.com/>

Результати виконання програми занадто довгі тому варто їх було виокремити в файл **matrix.xls**. Але приблизний вигляд виводу наступний.

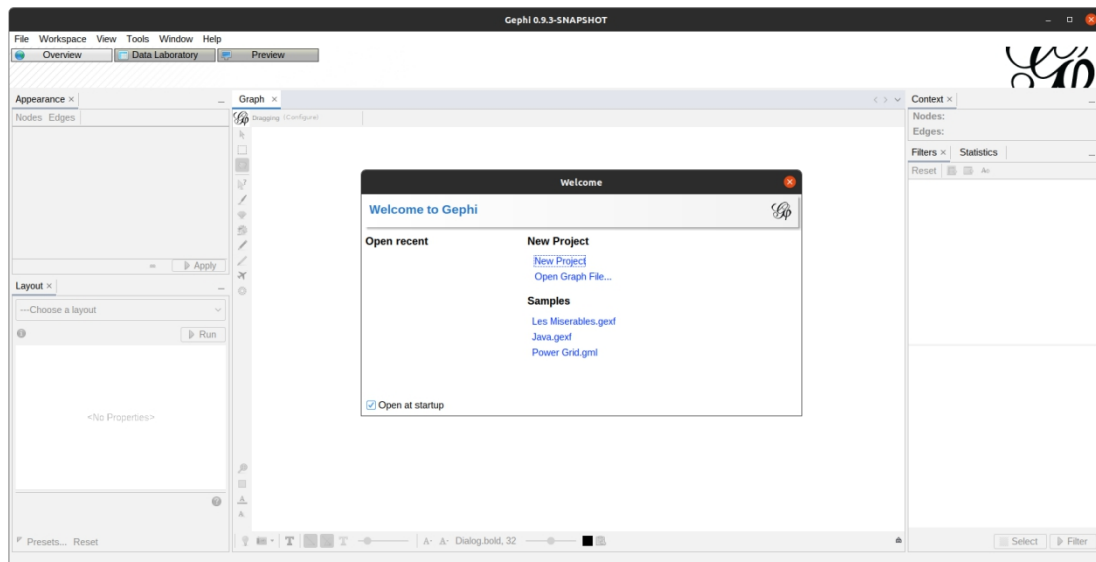
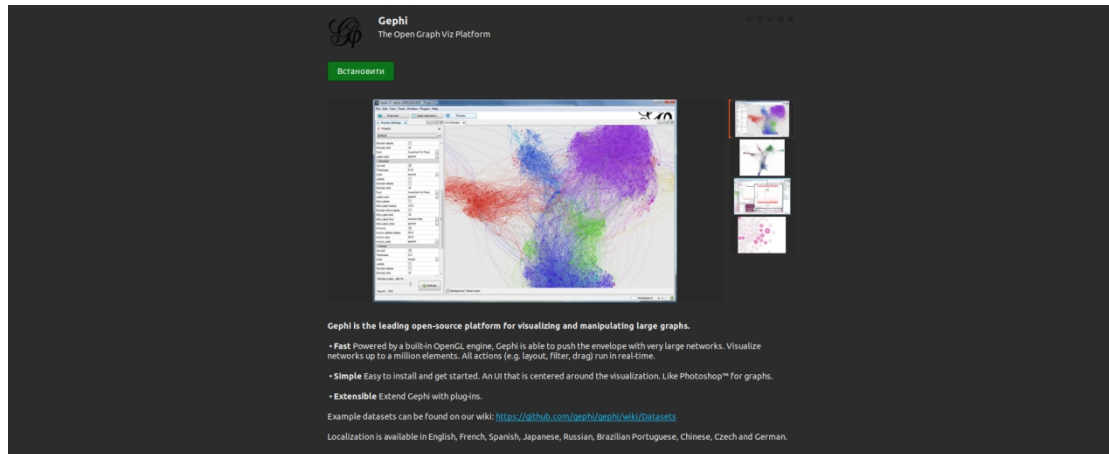
```
;APPLE;AND;IPHONE;NEW;EVENT;PRO;AIR;MAC;MACBOOK;HAS;IPAD;HAVE;ITS;SAID;ALSO;MACOS;STUDIO;
APPLE;0;71;41;35;35;37;27;39;16;22;20;22;18;15;10;3;13;19;18;7;17;15;9;19;11;12;7;14;10;1
AND;71;0;41;31;16;25;20;29;3;6;21;20;13;10;11;9;10;6;20;3;8;15;4;16;6;9;4;12;8;4;8;6;9;2;
IPHONE;41;41;0;32;17;16;20;13;2;5;22;9;8;10;7;0;11;9;4;11;12;2;7;2;6;10;2;0;1;3;8;5;0;2;8
NEW;35;31;32;0;19;20;23;20;9;3;16;5;6;6;4;0;9;10;2;3;9;2;7;0;3;9;2;1;2;1;6;4;0;0;4;1;0;
EVENT;35;16;17;19;0;8;20;18;8;2;15;3;0;2;3;0;9;18;0;0;14;0;9;0;2;7;0;0;0;4;2;5;0;0;5;2;12
PRO;37;25;16;20;8;0;11;23;13;3;2;7;6;12;2;1;1;5;8;3;10;6;9;4;3;3;0;4;1;4;9;4;4;0;2;2;1;0;
AIR;27;20;20;23;20;11;0;22;13;5;23;4;2;4;3;0;9;9;0;0;12;0;6;0;2;8;0;0;3;0;0;0;0;0;2;2;2;0
MAC;39;29;13;20;18;23;22;0;19;6;13;5;4;8;2;19;13;10;0;3;13;0;11;0;4;9;2;0;3;4;3;1;0;0;2;4
MACBOOK;16;3;2;9;8;13;13;19;0;3;2;2;2;7;1;1;0;6;0;1;8;0;6;0;1;0;0;0;0;1;0;1;0;0;0;2;0;0;2
HAS;22;6;5;3;2;3;5;6;3;0;2;1;5;3;5;2;0;1;0;4;5;0;2;8;6;0;2;0;1;5;0;0;2;0;1;0;0;4;2;0;5;7;
IPAD;20;21;22;16;15;2;23;13;2;2;0;5;0;0;3;0;9;7;0;2;6;0;2;0;3;8;2;0;5;0;0;0;12;2;2;2;0;
HAVE;22;20;9;5;3;7;4;5;2;1;5;0;7;4;5;1;0;1;6;1;1;6;1;4;8;2;2;6;5;1;1;3;4;0;1;0;1;0;0;0;4;
ITS;18;13;8;6;0;6;2;4;2;5;0;7;0;6;0;0;2;0;4;2;1;5;0;1;2;2;0;3;0;1;0;2;0;0;0;2;0;0;0;0;2;2
SAID;15;10;10;6;2;12;4;8;7;3;0;4;6;0;0;0;0;1;2;2;2;0;3;3;0;0;0;0;3;0;2;1;0;0;2;0;0;0;0;0;
ALSO;10;11;7;4;3;2;3;2;1;5;3;5;0;0;0;1;0;0;4;5;3;0;1;8;0;0;0;0;2;5;2;0;0;0;2;2;0;4;0;0;1;
MACOS;3;9;0;0;0;1;0;19;1;2;0;1;0;0;1;0;0;0;0;1;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;11;0
STUDIO;13;10;11;9;9;1;9;13;0;0;9;0;2;0;0;0;0;5;0;2;4;0;1;0;0;9;0;0;1;1;0;0;0;0;2;2;2;0;0;
```

Приведемо до оформленого вигляду в **Exel**. Вивід повний таблиці занадто громіздкий, наводиться лише частина.

	APPLE	AND	IPHONE	NEW	EVENT	PRO	AIR
APPLE	0	71	41	35	35	37	27
AND	71	0	41	31	16	25	20
IPHONE	41	41	0	32	17	16	20
NEW	35	31	32	0	19	20	23
EVENT	35	16	17	19	0	8	20
PRO	37	25	16	20	8	0	11
AIR	27	20	20	23	20	11	0

Інсталяція Gephi

У системах **Linux** дистриб'ютиву **Ubuntu** можна інстальювати простим чином скориставшись пакетним менеджером:



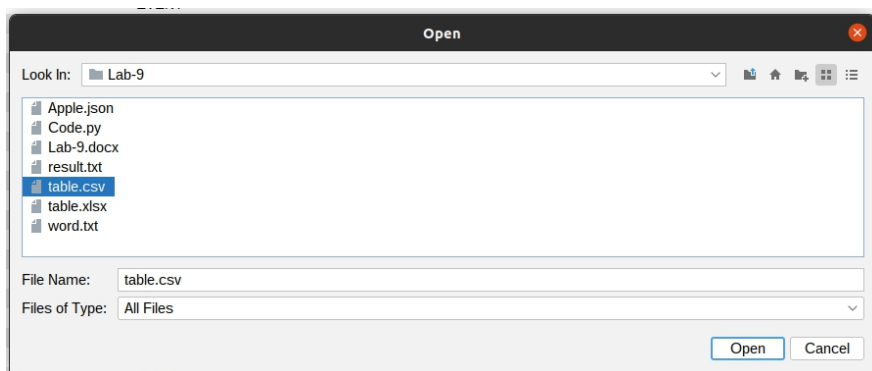
Ознайомлення з системою Gephi

Інтерфейс користувача системи **Gephi** включає три основні розділи (вікна):

- «**Лабораторія даних**»: тут зберігаються всі вихідні дані про мережі, а також додаткові розрахункові значення;
- «**Обробка даних**»: тут відбувається велика частина операцій користувача, зокрема, ручне редагування мереж, тестування макетів, встановлення фільтрів;
- «**Попередній перегляд**»: тут уточнюється форма виводу графу, як правило, за допомогою набору інструментів граф допрацьовується, в тому числі, і з естетичної точки зору. У цьому ж вікні реалізований виклик експорту графа в формати **PDF, PNG і SVG**.

Ці три основні розділи охоплюють множину вкладок, які дозволяють користувачу реалізовувати окремі функції.

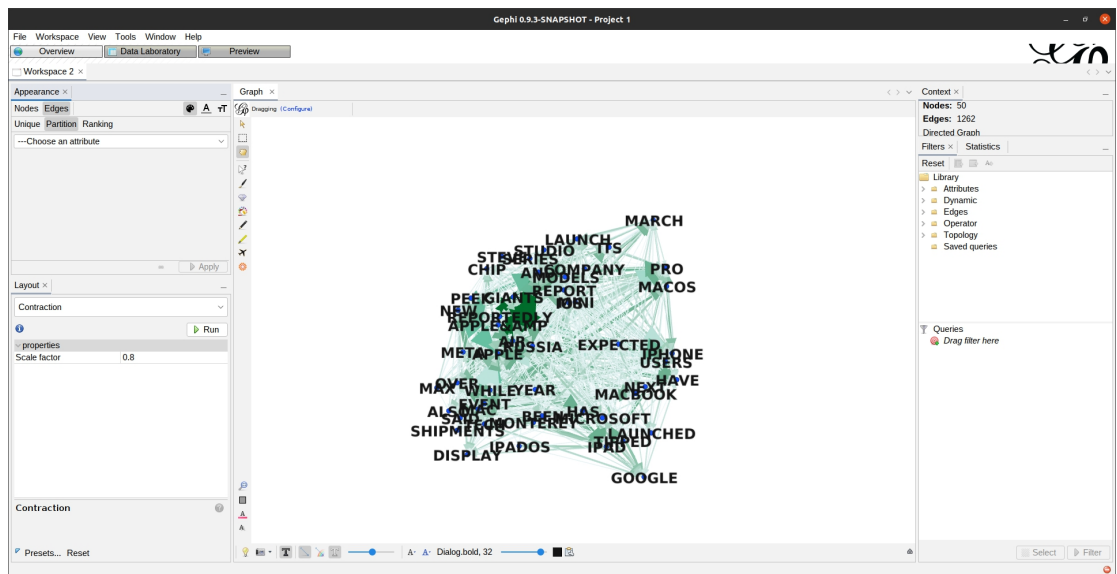
Завантажимо отриманий програмою файл в **Gephi**.



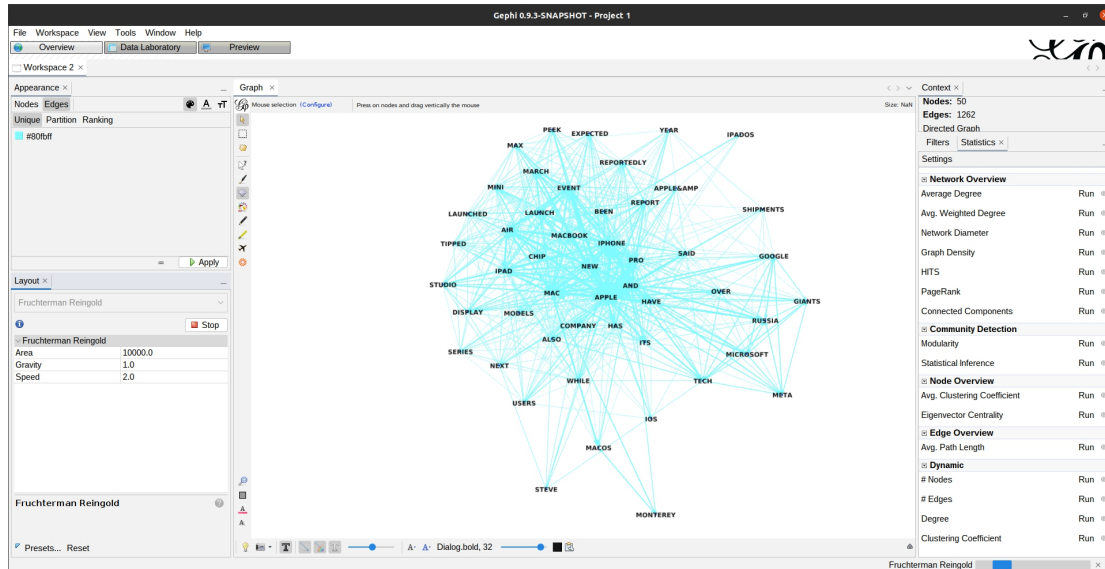
Перше, що ми бачимо перед собою, так це нашу таблицю.

Id	Label	Interval
APPLE	APPLE	
AND	AND	
IPHONE	IPHONE	
NEW	NEW	
EVENT	EVENT	
PRO	PRO	
AIR	AIR	
MAC	MAC	
MACBOOK	MACBOOK	
HAS	HAS	
IPAD	IPAD	
HAVE	HAVE	
ITS	ITS	
SMD	SMD	
ALSO	ALSO	
MACOS	MACOS	
STUDIO	STUDIO	
MARCH	MARCH	
RUSSIA	RUSSIA	
WHILE	WHILE	
LAUNCH	LAUNCH	
GOOGLE	GOOGLE	
MINI	MINI	
TECH	TECH	
BEEN	BEEN	
DISPLAY	DISPLAY	
IOS	IOS	
MICROSOFT	MICROSOFT	
USERS	USERS	
COMPANY	COMPANY	
MAX	MAX	
APPLE&	APPLE&	

Переходимо в сусіднє вікно, де можна здійснювати візуалізацію.



Практикуємо свої навички користування програмою.



Питання до практичної роботи

1. Що таке матриця суміжності концептів?

Матриця суміжності — один із способів представлення графу у вигляді матриці.

2. Які ви знаєте засоби перетворення типів даних в мові **Python**?

Засоби перетворення типів даних:

- ❖ Функція **int()**. Перетворення в ціле число
- ❖ Функція **str()**. Перетворення в рядок
- ❖ Функція **repr()**. Перетворення в рядок як літерал
- ❖ Функція **float()**. Перетворення в число з плаваючою комою
- ❖ Функції **ord()** та **chr()**. Перетворення кодів одиночних символів
- ❖ Тип даних - це абстрактне поняття, як в мові програмування довільний, *class* дозволяє визначати довільний тип даних. Тому можна створювати програми, що будуть за алгоритмом приводити дані до певного вигляду. Також існують автоматизовані вже програмні проекти, що дозволяють будувати довільні типи даних, в певній конкретній сфері їх застосування.

3. У яких дослідницьких проектах найбільш широко використовується пакет **Gephi**?

Gephi активно використовується в цілому ряді академічних дослідницьких проектів, зокрема соціологічних; також швидко здобув популярність серед журналістів. Зараз його користувацьке середовище значно розширилося — за допомогою цього пакета можна займатися будь-якою темою мережевого аналізу. **Gephi** використовувався, серед іншого, для візуалізації глобальної зв'язності контенту **New York Times** та вивчення мережевого трафіку **Twitter** під час соціальних заворушень; **Gephi** надихав створення **LinkedIn InMaps** і був використаний для візуалізації цілої мережі **Truthy**.

Загалом **Gephi** широко використовується в так званих «цифрових гуманітарних науках» (**Digital humanities**): в історії, літературі, політології тощо.

Спеціалісти з такого середовища активно беруть участь у подальшій розробці та популяризації продукту.

4. Назвіть основні режими роботи системи **Gephi**.

Інтерфейс користувача системи **Gephi** включає три основні розділи (вікна):

- > «Лабораторія даних»
- > «Обробка даних»
- > «Попередній перегляд»