**Міністерство освіти і науки України**

**Національний університет «Запорізька політехніка»**

Кафедра програмних засобів

**ЗВІТ**

Дисципліна «Розробка прикладних програм»

Робота №4

Тема «Оброблення природньої мови»

**Виконав варіант 19**

Студент КНТ-122 Онищенко О. А.

**Прийняли**

Викладач Дейнега Л. Ю.

2024

Мета роботи

Ознайомитись з основними інструментами оброблення природньої мови, які входять у склад бібліотеки NLTK мови програмування Python.

Навчитися розв’язувати актуальні практичні завдання у галузі оброблення природної мови за допомогою бібліотеки NLTK.

Індивідуальне завдання

Реалізувати створення програмних відповідей Українською мовою на основі виконання SQL-запитів до бази даних, розробленої в лабораторній роботі №3.

Текст файлу

NAME='data'

USER='root'

PASSWORD='1313'

HOST='localhost'

PORT='3306'

import mysql.connector

db=mysql.connector.connect(database=NAME,user=USER,password=PASSWORD,host=HOST,port=PORT)

c=db.cursor()

CLIENTS\_TABLE='finance\_client'

PAYMENTS\_TABLE='finance\_payment'

PERIODIC\_PAYMENTS\_TABLE='finance\_periodicpayment'

import nltk

while 1:

HELP\_MESSAGE='''Загальні команди:

- поможи АБО допомога: виведення цього повідомлення

- вийти АБО вихід: закриття програми

- користувачі: виведення списку користувачів

- періодичні платежі: виведення усіх запланованих платежів

- платежі: виведення історії усіх платежів

Користувацькі команди (потребують імені користувача у запиті):

- баланс: виведення балансу користувача

- ліміт АБО кредит: виведення кредитного ліміту користувача

- менеджер АБО адміністратор: виведення статусу користувача

- користувач: виведення усіх даних про користувача

'''.strip()

user\_query=input('> ')

words=nltk.word\_tokenize(user\_query.lower())

def check\_any(

terms:list[str],

words:list[str]=words,

):

found=False

for term in terms:

for word in words:

if term in word:

found=True

break

return found

def check\_all(

terms:list[str],

words:list[str]=words,

):

all\_good=True

for term in terms:

found=False

for word in words:

if term in word: found=True

if not found:

all\_good=False

break

return all\_good

def clean\_query(

query:list[str],

):

return [

word for word in query

if 'буд' not in word

and 'ласк' not in word

and 'про' not in word

and 'дан' not in word

and 'інф' not in word

and 'чи' not in word

]

def execute\_query(

query:str,

):

c.execute(query)

rows=c.fetchall()

return rows

def get\_client\_by\_id(

client\_id:int,

):

q=f'SELECT id,name FROM {CLIENTS\_TABLE} WHERE id={client\_id}'

return execute\_query(q)[0]

def get\_client\_by\_name(

user\_name:str,

):

q=f'SELECT name,balance,credit,manager FROM {CLIENTS\_TABLE} WHERE name="{user\_name}"'

return execute\_query(q)[0]

def get\_balance(

user\_name:str,

):

q=f'SELECT name,balance FROM {CLIENTS\_TABLE} WHERE name="{user\_name}"'

return execute\_query(q)[0]

def get\_credit(

user\_name:str,

):

q=f'SELECT name,credit FROM {CLIENTS\_TABLE} WHERE name="{user\_name}"'

return execute\_query(q)[0]

def get\_manager(

user\_name:str,

):

client\_query=f'SELECT name,manager FROM {CLIENTS\_TABLE} WHERE name="{user\_name}"'

return execute\_query(q)[0]

if check\_any(['вих','вий']): break

elif check\_any(['пом','доп']): print(HELP\_MESSAGE)

elif check\_any(['пр','віт']): print(user\_query)

elif check\_any(['користувачі']) or check\_all(['всі','кор']) or check\_all(['усі','кор']) or (check\_any(['кор']) and len(words)==1):

q=f'SELECT name,balance,credit,manager FROM {CLIENTS\_TABLE}'

rows=execute\_query(q)

print(f'Користувачі ({len(rows)}):')

for name,balance,credit,manager in rows:

print(f'- {name} має {balance} на рахунку, {credit} кредитного ліміту, та {"Є" if manager else "НЕ є"} менеджером')

elif check\_any(['пер']):

q=f'SELECT amount,purpose,period,next\_date,client\_id FROM {PERIODIC\_PAYMENTS\_TABLE}'

rows=execute\_query(q)

print(f'Періодичні платежі ({len(rows)}):')

for amount,purpose,period,next\_date,client\_id in rows:

client\_id,client\_name=get\_client\_by\_id(client\_id)

print(f'- {purpose} кожен {"день" if period=="Day" else "місяць" if period=="Month" else "рік"}, наступний платіж {next\_date.strftime("%d.%m.%Y")} для {client\_name}')

elif check\_any(['пл']):

q=f'SELECT timestamp,purpose,amount,client\_id,kind,operation FROM {PAYMENTS\_TABLE}'

rows=execute\_query(q)

print(f'Платежі ({len(rows)}):')

for timestamp,purpose,amount,client\_id,kind,operation in rows:

client\_id,client\_name=get\_client\_by\_id(client\_id)

print(f'- {purpose} за {timestamp.strftime("%d.%m.%Y o %H:%M:%S")} на {amount} від {client\_name}, {"одноразове" if kind=="Single" else "періодичне"} {"зняття" if operation=="Withdrawal" else "внесення"}')

elif check\_any(['бал','гро']):

stripped\_words=clean\_query([w for w in words if 'бал' not in w and 'гро' not in w])

found=False

for word in stripped\_words:

try:

name,balance=get\_balance(word)

print(f'{name} має {balance} на рахунку')

found=True

break

except: continue

if not found: print('користувача не знайдено')

elif check\_any(['кр','лім']):

stripped\_words=clean\_query([w for w in words if 'кр' not in w and 'лім' not in w])

found=False

for word in stripped\_words:

try:

name,balance=get\_credit(word)

print(f'{name} має {balance} кредитного ліміту')

found=True

break

except: continue

if not found: print('користувача не знайдено')

elif check\_any(['мен','адм']):

stripped\_words=clean\_query([w for w in words if 'мен' not in w and 'адм' not in w])

found=False

for word in stripped\_words:

try:

name,manager=get\_manager(word)

print(f'{name} {"Є" if manager else "НЕ є"} менеджером')

found=True

break

except: continue

if not found: print('користувача не знайдено')

else:

stripped\_words=clean\_query([w for w in words if 'кор' not in w])

found=False

for word in stripped\_words:

try:

name,balance,credit,is\_manager=get\_client\_by\_name(word)

print(f'{name} має {balance} на рахунку, {credit} кредитного ліміту, та {"Є" if is\_manager else "НЕ є"} менеджером')

found=True

break

except: continue

if not found: print('користувача не знайдено')

Результати виконання

Процес роботи з програмою наведено нижче:

Зображення, що містить текст, знімок екрана, Шрифт, число

Автоматично згенерований опис

Зображення, що містить текст, Шрифт, знімок екрана, інформація

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, Шрифт, число

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Зображення, що містить текст, Шрифт, знімок екрана, ряд

Автоматично згенерований опис

Контрольні питання

Яким чином виконується оброблення природної мови в бібліотеці NLTK?

Бібліотека NLTK для мови програмування Python впроваджує багато способів обробки природньої мови. Основні засоби це **токенизація**, **чанкування** та **виділення** **частин мови**.

Для **токенізації** можна застосувати розділ бібліотеки nltk.tokenize та вбудовані у неї функції sent\_tokenize() для розділення тексту на речення або word\_tokenize() для розділення заданого тексту на список слів.

Для **виділення частин мови** можна застосувати функцію бібліотеки nltk під назвою pos\_tag() яка приймає список токенізованих слів тексту як параметр.

Для **чанкування** можна застосувати регулярні вирази для створення граматики як програмі потрібно виділяти чанки з тексту. Прикладом граматики чанку може бути "NP: {<DT>?<JJ>\*<NN>}". Після створення змінної граматики чанкування можна створити оброблювач чанків застосуванням функції бібліотеки nltk під назвою RegexpParser(), яка приймає граматику чанкування як аргумент. Для створення дерева чанків тепер необхідно використати метод prase об’єкту оброблювача чанків вигляду chunk\_parser.parse(text\_with\_pos\_tags). При виведенні такого об’єкту через методом draw() на екрані має з’явитися дерево чанків.

[Джерело](https://realpython.com/nltk-nlp-python/)

Які основні модулі входять до складу бібліотеки NLTK?

До складу бібліотеки оброблення природньої мови NLTK для мови програмування Python входять такі основні модулі:

* nltk.book
* nltk.chunk
* nltk.grammar
* nltk.metrics
* nltk.misc
* nltk.parse
* nltk

[Джерело](https://www.nltk.org/py-modindex.html)

Яким чином можна використати готові тексти для роботи програми на Python?

Для використання готових текстів з бібліотеки оброблення природньої мови NLTK для мови програмування Python потрібно спочатку завантажити їх викликом методу download() на об’єкт nltk. Після завантаження необхідних ресурсів можна імпортувати та використати їх включаючи відповідні назви до файлу:

from nltk.corpus import gutenberg

HOLY\_BIBLE=gutenberg.words('Bible-KJV.txt')

print(HOLY\_BIBLE)

[Джерело](https://www.geeksforgeeks.org/accessing-text-corpora-and-lexical-resources-using-nltk/)

Яким чином розбити текст на лексеми?

Яким чином побудувати граматику в NLTK?

Що таке біграми та яким чином їх визначити у тексті?

Що таке чанкінг та яким чином він виконується?

Яким чином та за допомогою яких засобів виконується семантична інтерпретація?

Що таке контекстно-вільна граматика? Наведіть приклади.

Яким чином визначити частину мови слів тексту?