МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Київський національний університет імені Тараса Шевченка
Навчально-науковий інститут філології
Кафедра української мови та прикладної лінгвістики

Укладання корпусу українськомовних текстів соціальної мережі Інстаграм

Кваліфікаційна робота
освітнього ступеня «бакалавр»
за спеціальністю 035 «Філологія»,
спеціалізацією 035.10 «Прикладна лінгвістика»,
галузі знань 03 «гуманітарні науки»
ОПП «Прикладна (комп'ютерна) лінгвістика та англійська мова»
студентки IV курсу
Марини ХАРЧЕНКО

Науковий керівник:
к.філол.н., доц. Оксана ЗУБАНЬ

«Допущено до захисту»
Протокол № 11 засідання кафедри
української мови та прикладної лінгвістики
ННІФ від 01.06.2023
Завідувач кафедри_____**Сергій Різник**

КИЇВ – 2023

ЗМІСТ

ЗМІСТ..... 2
ВСТУП..... 3
РОЗДІЛ 1. СОЦІАЛЬНІ МЕРЕЖІ ЯК КОМУНІКАТИВНА ВЗАЄМОДІЯ. ІНСТАГРАМ..... 6
1.1. Вплив соціальних мереж на інтернет-спілкування..... 6
1.2. Функціональні особливості соціальної мережі Інстаграм.....9
Висновки до розділу 1..... 11
РОЗДІЛ 2. ТЕКСТОВИЙ КОНТЕНТ В ІНСТАГРАМ..... 12
2.1. Особливості мовлення соціальної мережі Інстаграм..... 12
2.2. Текст в Інстаграм..... 13
2.3. Стиль комунікації..... 14
2.4. Українськомовний контент в Інстаграм.....16
Висновки до розділу 2.....19
РОЗДІЛ 3. ФОРМУВАННЯ БАЗИ ДАНИХ ТЕКСТІВ ІНСТАГРАМ ВОЄННОГО ПЕРІОДУ.....20
3.1. Добір та систематизація текстового матеріалу..... 20
3.2. Робота з Instaloader..... 21
3.3. Структура бази даних текстів Інстаграм..... 23
3.4. Аналіз зібраних текстів..... 24
Висновки до розділу 3.....27
РОЗДІЛ 4. СТВОРЕННЯ КОРПУСУ ТЕКСТІВ ВОЄННОГО ЧАСУ СОЦІАЛЬНОЇ МЕРЕЖІ ІНСТАГРАМ..... 29
4.1. Корпус і його особливості.....29
4.2. Створення корпусу текстів (програмне забезпечення)..... 31
4.3. Параметри пошуку та користувацькі можливості корпусу текстів..... 39
Висновки до розділу 4.....41
ВИСНОВКИ.....43
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... 46
ДОДАТКИ.....50

ВСТУП

Щороку використання соціальних мереж невпинно збільшується та стрімко розвивається. Паралельно з цим набуває розвитку один з новоутворених напрямів лінгвістики – мережева комунікація, що розкриває в собі такі важливі питання як спілкування на просторах інтернету (інтернет-спілкування), україномовність, роль тексту в рекламних кампаніях і маркетингу, а також інші важливі питання сучасності. Саме ці питання дають лінгвістам велику перспективу для дослідження мережевої комунікації на просторах соціальних мереж і платформах.

Спілкування – невід’ємна частина життя сучасного суспільства, тому з кожним роком віртуальна комунікація стає більш важливим елементом у повсякденному житті кожної людини і дає доступність спілкуватися у дистанційному режимі. Це спрощує вирішення багатьох питань, які раніше вимагали більше зусиль і часу: передавання великого за обсягом матеріалу у текстових форматах, комунікація з людьми на великих відстанях у різних точках світу, обговорення важливих тем і обмін думками на велику аудиторію тощо. Наприклад, з початком російсько-української війни люди масово почали об’єднуватись і вести інформаційну війну на різних соціальних платформах, поширюючи на широку аудиторію важливу інформацію. Вони розповсюджують важливі новини, факти та докази про події, що відбуваються, з метою просвітити широку аудиторію та усвідомити ситуацію. Люди також використовують соціальні мережі для організації зборів коштів на гуманітарну допомогу, медичне обладнання та інші потреби, пов'язані з конфліктом. Вони створюють пости, рекламні кампанії та закликають спільноту до активної підтримки і благодійності.

Соціальні мережі сьогодні виконують набагато більше функцій, ніж просто забезпечення розваг, відпочинку і публікації фотографій. Вони стали невід'ємною складовою сучасного способу спілкування і мають велике значення для розвитку бізнесу.

Серед популярних соціальних мереж є Фейсбук, Ютуб, Тік-ток, Твіттер та Інстаграм, який став найпотужнішим інструментом комунікації, бізнесу, навчання.

Усі попередньо згадані платформи мають різну функціональність, мету створення, аудиторію, візуальну концепцію, проте дещо є спільним – текстовий блок, що використовується різним способом (пости, тексти до фото та відео, коментарі, повідомлення тощо).

Інтернет-комунікація об’єднує мільярди людей з усього світу, тож важливо звернути увагу, що для багатьох людей це стає основним каналом комунікації. Це має значний вплив на появу новоутворених слів, словосполучень, нових каналів комунікації (текст на картинках, на відео), стилі мовлення, знакові системи. Напркилад, за допомогою використання знаків і емодзі можна легко передати емоції, почуття і переживання.

Також інтернет-комунікація відіграє важливу роль у розвитку української мови. З початком російсько-української війни у соцмережах стрімко зросло використання української мови, а також почали з’являтись нові слова, їх тлумачення та фрази. Серед наукових праць і досліджень особливостей інтернет-спілкування є роботи Зайцевої С., Дзюбіної О. [5], Єльнікової Н. [6], Івахненко Т. [8], Малєнкової Д. [14], Маковецької-Гудзь Ю. [13], Сидорова М. [25] та інші.

Актуальність теми дослідження. Актуальність цієї теми обумовлена необхідністю дослідження особливостей та можливостей інтернет-спілкування у соціальній мережі Інстаграм. Це допоможе сучасному суспільству правильно використовувати і розуміти методи комунікації на просторах Інтернету. А також актуальність полягає в дослідженні текстів воєнного періоду та виявленні новоутвореної лексики.

Мета і завдання дослідження. Метою роботи є створення корпусу текстів воєнного часу за матеріалами соціальної мережі Інстаграм. Відповідно до мети були поставлені такі **завдання**:

- дослідити особливості текстового блоку на просторах Інстаграму, включаючи інтернет-комунікацію;
- знайти і дослідити тексти різних сторінок в Інстаграм з початку воєнного періоду;

- зібрати і завантажити тексти, що були знайдені;
- створити базу даних зібраних текстів;
- виконати очищення текстів для подальшої роботи;
- створення програми: робота з Google Drive, підключення до таблиці, обробка текстів та їх даних, створення веб-застосунку у вигляді корпусу текстів.

Об’єктом дослідження є українськомовні тексти соціальної мережі Інстаграм, **предметом** – програмний функціонал оброблення, класифікації та пошуку лінгвістичних об'єктів у текстових даних.

Практичне значення одержаних результатів полягає в тому, що створений корпус текстів може послужити зручним застосунком для лінгвістичного аналізу українськомовних текстів.

Методологічною базою дослідження є метод теоретичного та емпіричного аналізу, включаючи аналіз, синтез, індукцію, використання програмного забезпечення та бібліотек\ модулів – Instaloader, os, re, pandas, gspread, streamlit та інші.

Матеріалом дослідження є 220 українськомовних текстів, дібраних зі соціальної мережі Інстаграм за період 24 лютого 2022 року – 8 травня 2022 року.

Структура та обсяг дипломної роботи. Кваліфікаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаної літератури та додатків. Обсяг кваліфікаційної роботи роботи – 53 сторінки.

РОЗДІЛ 1. СОЦІАЛЬНІ МЕРЕЖІ ЯК КОМУНІКАТИВНА ВЗАЄМОДІЯ. ІНСТАГРАМ

1.1. Вплив соціальних мереж на інтернет-спілкування

Соціальна мережа – це структурна платформа, яка слугує засобом комунікації користувачів через обмін інформацією, повідомленнями, фото і відео. Спілкування в реальному житті значно відрізняється від комунікації у віртуальних умовах, але все ж таки другий спосіб все більше і більше заохочує людей саме до такого типу зв'язку з іншими. В наш час соціальні мережі набувають зовсім іншого значення та мети. Такі платформи як Фейсбук, Твіттер, Інстаграм та інші стають комунікативним простором. Це нові і малодосліджені канали комунікації, які є необмеженими засобами дослідження для лінгвістів, тому що інтернет-спілкування поєднує в собі писемний і усний способи комунікації.

Вплив соціальних мереж на інтернет-спілкування можна розглядати з різних аспектів. Головними перевагами соціальних мереж у засобах віртуальної комунікації є [14, 8]:

1. Доступність і знайомства. Інтернет-спілкування надає більше можливостей, ніж реальне. По-перше, соціальні мережі дозволяють підтримувати комунікацію з людьми, незважаючи на місцезнаходження, часові пояси. Також користувач може знайти нові знайомства та співрозмовників на просторах соціальних мереж, які дають можливість побачити профілі цих людей, знайти спільні інтереси за допомогою текстового блоку в Інстаграм (почитати пости, опис сторінки тощо).
2. Формати комунікації. Користувач за допомогою соціальних мереж має змогу обрати для себе зручний формат комунікації – текстовий формат (повідомлення, пости, коментарі), аудіо- та відеоформат, фотографії. У сучасному світі текстовий формат – найбільш популярний спосіб спілкування і обміну інформації, оскільки його використовують не лише у постах, повідомленнях, коментарях, а й у інфографіці, відео, сторіс для легкого сприйняття інформації.

3. Обговорення на велику аудиторію та висловлюваня своєї думки. Завдяки соціальним мережам люди можуть висловлювати свої думки з більшою легкістю, ніж у реальному житті. Як сказав М.Сидоров у своїй праці: «Текстовий обмін інформацією, з одного боку, приховує більшість емоційних реакцій співрозмовника, що дає йому можливість висловити те, що він не наважився б сказати у розмові віч-на-віч» [25, ст. 120]. Якщо взяти до уваги різносторонність цього питання, то інакше користувач не може передати свої емоції і почуття повністю (хоча емодзі допомагають у цьому питанні), а також побачити емоції співрозмовника.
4. Інформаційне джерело. Соціальні мережі послуговуються важливим інформаційним ресурсом – тут можна дізнатися про важливі новини, події, тренди. У сучасному світі часто ведуться інформаційні війни, які розповсюджуються уже швидко. Тож краще з обережністю ставитись до сприйняття тієї чи іншої інформації.
5. Просування та бізнес. Соціальні мережі стали одним з головних інструментів комунікації бізнесу і клієнтів. Вони допомагають бізнесу за допомогою активності на своїх сторінках залучати більшу аудиторії і потенційних клієнтів, використовуючи рекламні тексти, пости з різним типом контенту, манери спілкування у особистих повідомленнях тощо.

Інтернет-спілкування є вільним та легким для комунікації, оскільки у рамках соціальних мереж користувачі відчувають свободу і спілкуються тим стилем і мовою, які є для них зручними. Це і вплинуло на формування і виявлення комунікативних особливостей на просторах Інтернету. “Комунікативна поведінка користувачів соціальних мереж зумовлена певними правилами мережевого етикету, що розкривається у широкому застосуванні зниженої лексики, жаргонізмів, сленгізмів, вживанні великої кількості абревіатур та скорочень, використанні смайликів (від англ. smiley – піктограма, що виражає емоцію), що надають емоційно-експресивного забарвлення повідомленню” – зазначає у своїй праці Н. Єльнікова [6, ст. 44].

До особливостей мовлення і спілкування у соціальних мережах належать [14, ст.10], [13, ст.661]:

1. Використання скорочень і абревіатур. Ось як описує використання скорочень у своїй праці Сидорук Г.: “Інтернет простір переповнений скороченнями, які зумовлені особливостями спілкування в ньому, необхідністю зважати на свій і чужий час, передати максимум інформації мінімальними зусиллями, швидко і в доступних для огляду обсягах.” [26, 293].

   Для інтернет-мереж це було завжди актуально, оскільки пости, повідомлення, коментарі мають обмежену кількість символів, тож користувачі використовують скорочення і абревіатури, які стали частиною повсякденного спілкування. За приклад такої лексики можна взяти відомі скорочення “ЛОЛ” (Laughing Out Loud), “ОМГ” (Oh My God) тощо.
2. Емодзі, включаючи стікери та анімовані картинки, теж стали невід’ємною частиною віртуальної комунікації. Вони слугують для вираження емоцій і привертання уваги співрозмовника, оскільки текст не завжди може виразити емоції. Наприклад, користувачі часто використовують такі графічні символи, що мають своє тлумачення: :) – посмішка, =O – здивування, =D – сміх тощо.
3. Запозичення з інших мов. Завдяки інтернет-спілкуванню українська мова насичена багатьма новими словами іншомовного походження. У повсякденному мовленні – як і в реальному, так і на просторах Інтернету – часто використовуються такі слова як “сторіс”, “лайкнути”, “зафоловити”, “селфі” тощо.
4. Соціальні мережі значно вплинули на дотримання правил граматики і орфографії. Так, можна помітити, що користувачі часто користуються для експресивності слова чи тексту повторення літери чи знаку пунктуації: “!!!”, “Сиииииильно”.

Інтернет-спілкування у соціальних мережах внесло свої зміни у мовну систему, а також стали повсякденним ритуалом більшості населення. Стало очевидно, що комунікація у соціальних мережах потребує побудови повноцінної

системи інтернет-спілкування. Тож ця тема потребує детального вивчення та дослідження.

1.2. Функціональні особливості соціальної мережі Інстаграм

Інстаграм – одна з популярних соціальних мереж, яка базується на обміні інформацією (фото, відео, текстова комунікація) між двома або більше користувачами. Це платформа, яка використовується не лише у рамках особистісних цілей, а й для розвитку брендів і бізнесу. У сучасному світі жоден підприємець не починає свою справу без такого важливого інструмента як Інстаграм. Текстовий блок соціальних мереж є важливим етапом у розробці маркетингової стратегії, а також комунікації між користувачами і бізнес-сторінкою бренду.

Інстаграм став повноцінним способом показати себе всьому світу. Одним з основних переваг є почуття свободи і прояву творчості. Звісно, у віртуальному житті людина має змогу створити свій інтернет-образ і стиль, і стати ким завгодно. Не менш важливою є можливість вільно висловити свою думку і бачення стосовно будь-якої ситуації. Функціонал мережі постійно розширюється, але є усталені засоби комунікації з читачами, у яких використовується інтернет-спілкування у текстових форматах. Вони допомагають розповісти про себе і показати частинку свого життя.

Ці типи контенту безсумнівно впливають на залучення нової аудиторії і багато реакцій на будь-яку дію в Інстаграм, пошук знайомств і друзів. Статистика і комунікація з читачами залежить від правильного використання цього функціоналу, до якого входять шапка профілю (ім’я користувача, біографія, категорія профілю тощо), сторіс та функціонал для їх оформлення, пост, текст, рілс (reels), співавторство, відмітки, збережені сторіс, а також кнопок швидкого зв’язку.

Щороку ми спостерігаємо за оновленнями від розробників, які полегшують створення контенту, покращують якість наших сторінок та, найголовніше, піднімають активність і статистику. Щось лише оновлюється, щось з’являється нове або стає недоступним для використання.

Так, наприклад, у 2016 році у користувачів з’явилася можливість записувати відео тривалістю до 15 секунд, які зникають через добу після публікації, а вже з 2022 року такі відео можна записувати тривалістю до 1 хвилини. Така функція називається *stories* (сторіс, історії), яка має найбільший попит і найкраще піднімає статистику. Згодом можливим стало додавання стікерів/ фотографій/ місцезнаходження/ позначення інших акаунтів через @/ опитування та інших функцій. Також є можливість групувати сторіс у ‘Вибрані’, завдяки чому історії будуть доступні для перегляду не 24 години, а завжди. Тобто можна створювати тематичні рубрики, які можна подивитися в будь-який момент. Ця стрічка вибраних знаходиться під шапкою профілю.

Пост або ж *публікація* — це те, що користувачі бачать в своїй стрічці, тобто фото і відеоматеріали від різних акаунтів. Пости стали менш актуальними з появою сторіс, але про них не варто забувати. Пости мають відповідати тематиці сторінки і мати дописи на певні теми, які б зацікавили читачів. Існують різні типи контенту, тож варто балансувати і дотримуватись послідовності. Дописи складаються з фото/ відео і тексту до них. Також зараз стала актуальною *інфографіка* і часто використовується для постів. В Інстаграм найбільш популярним типом контенту був пост, що включає в себе фото (або декілька) і текст до нього. Це одна з найкращих комбінацій, щоб привернути увагу до своєї сторінки і особистості. Проте зараз популярності набуває відеоконтент: це може бути просто відео з текстом для нього, це може бути інтерв’ю тощо.

Також є легкий спосіб колаборації і залучення аудиторії – *співавторство*. Така функція дає можливість опублікувати пост двом профілям на своїх сторінках, при чому на обох сторінках буде видно двох авторів.

Оскільки відеоконтент стає найбільш актуальним типом контенту, користувачі отримали таку функцію в Інстаграм як *reels* (рілс) - функція, що дозволяє записувати відео (у форматі Тік Ток) зі звуком і різними ефектами тривалістю до хвилини. Якщо викласти рілс, то у обліковому записі автоматично з'являється спеціальна вкладка, де будуть публікуватися ролики подібного формату. Цей формат є швидким способом

зацікавити користувачів і змусити перейти на сторінку, щоб переглянути інший контент і в результаті почати слідкувати за акаунтом.

Висновки до розділу 1

Соціальні мережі продовжують впливати на розвиток інтернет-комунікації і мови загалом. Це можна обгрунтувати появою нової лексики, використання її не лише у віртуальній комунікації, а й у реальному і живому спілкуванні. З’являються нові фрази, скорочення, абревіатури, емодзі тощо. Варто відзначити, що спілкування в Інтернеті не сприяє поліпшенню граматичних і орфографічних навичок, але має вплив на мовну поведінку суспільства в цілому.

РОЗДІЛ 2. ТЕКСТОВИЙ КОНТЕНТ В ІНСТАГРАМ

2.1. Особливості мовлення соціальної мережі Інстаграм

Інтернет-комунікація на просторах Інстаграму має свої особливості. В такій соціальній платформі як Інстаграм комунікація може бути реалізована через різні види. Ці види максимально наближені до комунікації у реальному житті. До таких видів належить міжособистісна комунікація (відбувається між двома користувачами у приватних повідомленнях, коментарях один одному), міжгрупова комунікація (відбувається між різними групами людей, активне обговорення на певну тему в коментарях є прикладом цьому виду), комунікація між користувачем і суспільством (соціальні мережі також надають можливість виражати свою думку, погляди або ідеї публічно перед широкою аудиторією, також сюди можна долучити комунікацію з брендом і його аудиторією), а також монологічне мовлення, тут варто згадати слова Кромської А. «Монологічне мовлення в умовах спілкування в соціальних мережах набуває спрямованості на діалог». І це доводить, що будь-яка інформація, яку користувач публікує у соціальній мережі, стає видимою для всіх інших і є у загальному доступі. Інстаграм дає свободу, яка проявляється у тому, що люди можуть надавати особисту інформацію, якими вони бажають ділитися. Це може бути неправдива і вигадана інформація. У соціальній мережі Інстаграм практикуються такі явища, коли деякі користувачі представляють себе у своїх профілях такими, якими вони не є у реальному житті. Це може включати показ вигляду, стилю життя, досягнень або соціального статусу, які можуть бути прикрашеними або навіть вигаданими.

Соціальні мережі змінили традиційне спілкування. Можна виділити декілька груп мовних одиниць, що стали невід'ємною частиною мовлення у соціальній мережі Інстаграм: експресивно забарвлена лексика (*бомба-ракета, агонь, крутий, стрьомний, не комільфо тощо*), неологізми (*селфі, хештег, лайк, сторіс, пост, блогер, смм, хакер, айтішник тощо*), інтернет-меми (зазвичай їх поділяють на фразові , словосполучення і слова, які можуть виражатись різними частинами мови. Для прикладу можна взяти актуальні фрази для України *хто не скаче – той москаль,*

порохоботи, зеленоботи, орки тощо), емодзі і смайли, хештеги (їх використання допомагає позначити конкретну тему або ключове слово. Хештеги допомагають згрупувати повідомлення за тематикою і полегшують пошук відповідних контенту), скорочення і абревіатури (використовуються для економії місця і обсягу тексту) і інші лексичні параметри.

2.2. Текст в Інстаграм

Інстаграм – це вже не просто соціальна мережа, що заснована на візуальному змісті. Зараз ця платформа набирає великих обертів і дає великі можливості користувачам. Вона розвиває інтернет-комунікацію і лексичну базу використання на просторах Інтернету. Текстовий блок Інстаграму є основою для просування і комунікації з читачами і клієнтами.

Текст в Інстаграмі має сильний вплив на сприйняття та розуміння контенту. Він допомагає встановити зв'язок зі своєю аудиторією, передати свої ідеї та спілкуватися з іншими користувачами. Крім того, текст є інструментом для вираження особистості, творчого підходу та самовираження в Інстаграмі. Сьогодні текст використовується всюди: у постах, сторіс, рілс, коментарях, приватних повідомленнях, у шапці профілю, у рекламних креативах і співпраці з блогерами.

Основними завданнями текстового блоку в Інстаграмі є [21, п.2]:

1. Комунікація: текст дозволяє користувачам спілкуватися один з одним, обмінюватися думками, коментувати пости та взаємодіяти в коментарях. Це створює можливість для взаємодії, обговорення та побудови спільноти в Інстаграмі.
2. За допомогою тексту в Інстаграмі користувачі мають можливість виразити свої думки, почуття, ідеї та переконання, що забезпечує унікальну атмосферу та враження від профілю або посту.
3. Інформування – не менш важливе завдання. Текст може бути використаний для передачі інформації про події, акції, новини, продукти або послуги.

Користувачі можуть ділитися корисними порадами, розповідати про свої досягнення або надавати контекст до своїх фото та відео.

4. Текст в Інстаграмі є невід'ємною складовою брендування та маркетингу, використовується для зацікавленості, взаємодії з аудиторією, підтримки репутації та впливу на споживачів.
5. Текст в Інстаграмі може бути використаний для створення наративу, хронології подій або розповіді про особистий досвід. Він допомагає створити контекст та розширити історію, що доповнюється фото чи відео.
6. Вираження творчості: користувачі можуть використовувати текст для вираження своєї творчості, наприклад, шляхом написання віршів, цитат, мотиваційних фраз або коротких історій.
7. Свою експертність люди можуть передавати через Інстаграм. Користувачі можуть використовувати текст для ділової та професійної інформації, надавати поради, викладати свої знання або ділитися експертними думками зі своєю аудиторією. Це дозволяє побудувати авторитет та експертність в певній галузі.

Текст слугує створенню корисного контенту і комунікації з читачами. Це важливий елемент Інстаграму, без якого ця соціальна платформа не стала б такою популярною у використанні, якою вона є сьогодні.

2.3. Стиль комунікації

Відомо, що візуальний контент, візуальна комунікація на просторах Інтернету є однією з ефективних та затребуваних форм взаємодії. Проте текст є основою і початком для подальшої роботи і просування бренду. Мовні особливості тексту і певні характеристики бренду – ідеальна комбінація для комунікації з користувачами.

Американський маркетолог Лаура Райс [40] правильно підкреслює важливість візуального образу в успішному позиціонуванні бренду. Вона порівнює вербальну концепцію, яка включає словесні прийоми та рекламні тексти, з "цвяхом", який потребує "молотка" - візуального образу, щоб зміцнити та підсилити цю концепцію у свідомості споживачів. Вербальна концепція, така як унікальний підпис, слоган або

повідомлення, відображає сутність бренду та його цінності. Проте, візуальний образ додає емоційний та запам'ятовувальний ефект до цієї концепції. Він може передавати атмосферу, стиль, ідею чи повідомлення бренду за допомогою інфографіки, кольорів, фотографій або логотипу. Використання візуального образу дає змогу створити сильні асоціації з брендом і сприяє легкому запам'ятовуванню. Комбінація вербальної концепції та візуального образу створює потужний інструмент для ефективної комунікації зі споживачами та будує сильну брендову ідентичність. Тому текст, який просуває на просторах соціальних мереж не буде успішним без голосу бренду, це є його основою.

Tone of voice або *голос бренду* – це єдиний стиль взаємодії компанії з аудиторією для всіх каналів комунікації. Інакше кажучи, це те, як звучить бренд в усіх його вербальних комунікаціях – у розсилках, спілкуванні з клієнтами чи читачами у повідомленнях і коментарях, на сайті, в постах тощо.

Для створення tone of voice потрібно набагато більше, ніж знати назву бренду і його спрямування. Щоб створити голос бренду існує декілька етапів та рекомендацій [43]:

1. Створення бренду і брендової стратегії (для чого створений бренд, його цінності, місія, особливості, параметри, за якими можна відрізнити бренд від колег чи конкурентів).
2. Аналіз цільової аудиторії (розуміння хто може бути потенційними клієнтами чи читачами бренду, дослідження їх болей, потреб, цінностей, активностей. Голос бренду повинен бути спрямований на вивчену цільову аудиторію).
3. Вибір тону комунікації та мови. Важливо визначити тон, в якому бренд спілкується з аудиторією. Це може бути формальний, неформальний, дружній та інші. Мова теж може бути формальна або неформальна, використання професійної чи повсякденної лексики тощо.
4. Емоційний зв’язок. Голос бренду має вміти створювати емоційний зв'язок з аудиторією. Це може бути за допомогою використання певних

слів, тону комунікації або способу підходу до споживачів. Важливо, щоб голос бренду викликав позитивні емоції та спонукав до взаємодії.

5. Популяризація голосу бренду повинна відбуватись на всіх можливих платформах і каналах комунікації – веб-сайт, соціальні медіа, рекламні матеріали тощо. Це створює єдність і підсилює впізнаваність бренду.

Голос бренду використовується як і в звичайних текстах для постів і сторіс, так і для рекламних текстів. Рекламні тексти - це тексти, які спрямовані на просування або презентацію товарів, послуг, ідей з метою зростання їх продажів.

2.4. Українськомовний контент в Інстаграм

Соціальна мережа Інстаграм об’єднує мільярди користувачів щодня з різних точок світу. Це свідчить про використання багатьох мов на просторах Інстаграму. Ця платформа підтримує багатомовність (станом на 9 червня 2023 року Інстаграм налічує 28 мов для використання додатку – англійська, українська, польська, португальська, німецька, данська, хорватська, чеська, фінська, тайська, словацька та інші) і надає можливість користувачам комунікувати та публікувати контент різними мовами. Користувачі можуть вибирати мову, на якій вони хочуть створювати свої пости, сторіз та описи. Це дає змогу виражати себе та спілкуватися з аудиторією рідною мовою. Багатомовність в Інстаграмі сприяє різноманітності та культурному обміну, та дає користувачу можливість розповісти про цінності, культурну спадщину своєї країни.

Українська мова в Інстаграмі відіграє важливу роль. Вона використовується для створення контенту, спілкування з аудиторією та вираження ідентичності українських користувачів. Багато людей обирають українську мову для своїх постів, сторіс та описів, що сприяє поширенню використання української мови в цифровому середовищі. Українська мова допомагає відтворювати унікальність, культуру та індивідуальність української спільноти.

Мовне питання також варто розглядати на просторах соціальних мереж. Воно полягає у використанні української мови при написанні постів, створення рекламних текстів, спілкуванні з користувачами. Використання української мови в коментарях, запитаннях та відповідях сприяє зрозумілому комунікаційному процесу та побудові зв'язку між учасниками спільноти.

Але якщо розглянути використання української мови в Інстаграм з обох сторін – період до початку російсько-української війни і після – то можна зробити висновок, що ситуація різко покращилась і україномовного контенту стало в рази більше. До війни україномовний контент міг включати різноманітні теми, такі як культура, мистецтво, мода, спорт, їжа, подорожі та багато іншого. Він допомагав користувачам виразити свою особистість, зацікавленість та спілкуватися з аудиторією. З початком війни україномовний контент відразу змінився та набув нових акцентів. Він став платформою для висловлення підтримки один одного, поширення важливих новин, для зборів і допомоги.

Україномовний контент звісно зазнав змін після початку воєнних дій, але не менш важливим є ставлення і цінність української мови на просторах Інстаграму. Згадуючи слова Маковецької-Гудзь Ю.: «Україномовними найчастіше є спільноти українських компаній, ЗМІ, політичних чи громадських організацій, блоги та сайти міст (що мають свою сторінку у соцмережах), а саме тому і більшість інформації в них подана українською літературною мовою» [13, ст.658], то можна дійти висновку, що ще донедавна на просторах українського Інстаграму користувачі будь-якого типу сторінки і її тематики, окрім вищезгаданих напрямів, використовували суржик або російську мову для спілкування.

Наразі ситуація зазнала суттєвих змін. Більшість бізнес-сторінок ведуть свої соціальні мережі українською мовою, а також багато блогерів та інфлюенсерів перейшли на українську мову у своєму Інстаграмі. Блогерство грає важливу роль у популяризації української мови у соціальних мережах. Зазвичай блогери мають великі аудиторії і звісно це є для багатьох людей мотивацією і прикладом.

Блогерство також відіграє важливу роль у використанні української мови на просторах соціальних мереж. Інстаграм – це та платформа, де українських блогерів і

інфлюенсерів найбільше. Вплив блогерів на українську мову можна розглядати як двостороннє явище. З одного боку, вони значно впливають на формування мовного стилю та лексики серед своєї аудиторії. Часто вони використовують неформальну мову, скорочення, суржик та інші стилістичні особливості, які відображають їхню особистість та спосіб комунікації. Блогери також активно використовують англійську та інші мови, особливо в галузях, де це є поширеним явищем (наприклад, у моді, красі, фітнесі, технологіях тощо). Це може сприяти появі запозичень, кальків та нових сленгових виразів серед блогерів та їхньої аудиторії.

З іншого боку, блогери також сприяють збереженню та популяризації української мови. Багато з них активно використовують українську мову у своєму контенті, підтримують акції та кампанії на підтримку української мови, спонукають свою аудиторію до її використання. Таким чином, вплив блогерів на українську мову можна розглядати як процес взаємодії між новаторськими тенденціями та збереженням мовних традицій.

Україномовний контент відображає розповіді про виживання, духовність, згуртованість спільноти та підкреслює важливість мови як символу національної ідентичності. Користувачі українського Інстаграму вперше за всю історію існування так міцно об’єднались та створили інформаційну силу по всьому світу, коли розпочалась російсько-українська війна у лютому 2022 року. Користувачі активно використовували Інстаграм для поширення новин, фотографій, відео та історій, що стосуються війни. Вони ділилися своїми думками, емоціями та досвідом, створюючи потужну інформаційну силу. Хештеги, такі як #SaveUkraine, #PrayforUkraine, #StopRussianAggression та інші, стали поширеними в Інстаграмі, допомагаючи залучити увагу до ситуації в Україні. Завдяки активному використанню Інстаграму українськими користувачами під час російсько-української війни, вдалося привернути увагу світової спільноти до конфлікту та ситуації в Україні. Також цей медійний простір став важливим каналом для збирання гуманітарної допомоги, передачі важливих повідомлень та організації акцій підтримки.

Висновки до розділу 2

Текст в Інстаграм є важливим інструментом для передачі повідомлень, вираження думок та спілкування з аудиторією. Він дозволяє користувачам розповідати історії, надавати контекст до візуального змісту та створювати взаємодію зі своїми підписниками. Текст може впливати на сприйняття публікацій, викликати емоції та стимулювати обговорення.

Інтернет-комунікація в Інстаграм також має значний вплив на мову. Завдяки широкому поширенню соціальних мереж і месенджерів, люди активно спілкуються в письмовій формі, що призводить до змін у мовній практиці. Інтернет-комунікація впливає на скорочення слів, використання емотиконів та смайликів, абревіатур, хештегів, а також на розробку нових мовних форм – інтернет-сленгу та метамови.

Після початку воєнних дій у лютому 2022 року український Інстаграм став місцем, де люди могли висловити свою громадянську позицію, надати підтримку та об'єднатися для розв'язання спільних проблем. Він довів, що соціальні мережі можуть використовуватися не лише для розваг та особистих цілей, а й як потужний інструмент для змін у суспільстві.

РОЗДІЛ 3. ФОРМУВАННЯ БАЗИ ДАНИХ ТЕКСТІВ ІНСТАГРАМ ВОЄННОГО ПЕРІОДУ

3.1. Добір та систематизація текстового матеріалу

Пошук, збір текстового матеріалу та організація бази даних вимагали чіткого алгоритму для роботи. Першим етапом був пошук відповідних текстів і їх збереження на сторінці. Цей процес вимагав чітких етапів і критеріїв, за яким відбувався пошук – дата написання, тематика сторінки, розмір тексту тощо. Наступним етапом було завантаження збережених текстів з особистої сторінки в Інстаграм на комп'ютер для подальшої роботи. З метою збору та збереження текстів з початку воєнного періоду було прийнято рішення про використання швидкого та ефективного підходу, і для цього було використано програмування. Для розробки програми була обрана мова програмування Python. Головним елементом у написанні програми для автоматичного завантаження текстів для створення бази даних був Instaloader – це інструмент, що був використаний для завантаження зображень та відео разом із підписами та іншими метаданими з Instagram.

В результаті збору і завантаження текстових даних з Інстаграм нам вдалось створити базу даних зібраного текстового матеріалу з його метаданими:

1. База даних створена в Google Таблицях для зручності заповнення її потрібним матеріалом і легким доступом до нього.
2. В таблиці збережена та кількість текстів, яку вдалось зібрати за період початку війни. Тексти датуються від 24 лютого 2022 року до 8 травня 2022 року.
3. До кожного тексту в таблиці окремо на Google Диску представлені фото- та відеоматеріали, кожні знімки занесені в окрему папку, яка матиме назву відповідно до ID (номеру) тексту.
4. Тексти зібрані зі сторінок Інстаграм, охоплюючи профілі різного характеру – бізнес-сторінок, блогерів, політичних діячів тощо.
5. Також у корпусі надана інформація щодо дати опублікування поста на сторінці автора і дата занесення тексту в таблицю. При внесенні текстів було дотримано послідовність, що сприяла дослідженню експресивності і змін написання текстів того чи іншого профілю.

6. Слід зазначити, що з кожного профілю зібрано по декілька текстів на початку війни.
7. Для процесу збору було створено окрему сторінку в Інстаграм, куди зберігались вибрані тексти, а звідти за допомогою програмування було завантажено інформацію про пост (дата, фото) і сам текст. Кожен текст з усіма його додатками скачувався в окрему папку на комп'ютері. В сам корпус тексти і інформація переносилась вручну.

3.2. Робота з Instaloader

Для збору і збереження текстів було вирішено обрати легкий та пришвидшений спосіб виконання цього завдання, тож для реалізації всіх етапів для створення бази даних текстів з початку воєнного періоду було використано програмування. Для створення програми було обрано мову програмування Python. Для розробки коду було використане інтегроване середовище розробки для мови програмування Python – PyCharm.

Також для роботи була використана система керування пакунками pip і модуль os для успішної роботи з операційною системою та файловою системою ПК. Важливим елементом у роботі був Instaloader. Це популярна бібліотека для мови програмування Python, яка дозволяє виконувати завантаження даних з Інстаграм. Ця бібліотека надає широкі можливості для отримання різноманітної інформації з профілів, постів, сторіс, коментарів та інших елементів Інстаграм. До основних функцій бібліотеки належать – завантаження профілів (можливість отримати інформацію про публічні профілі користувачів, таку як ім'я, біографія, кількість підписників тощо), завантаження постів (можливість завантажити пости з обраних профілів, включаючи зображення, відео, текстові та інші метадані), завантаження коментарів (можливість отримати коментарі до постів або сторіс з обраних профілів) та інші можливості.

Створена програма спрямована на завантаження особистих збережених постів (допис) соціальної мережі Інстаграм. При запуску роботи програми першочергово

відбувається авторизація на вказану сторінку – користувач вводить свій пароль, щоб отримати доступ до сторінки та збережених дописів.

Далі автоматично створюється папка, куди зберігаються всі завантажені збережені пости. Кожен пост зі своїми метаданими (текст, фото, дата) зберігається окремо в кожній папці. Завантаження автоматично припиняється, якщо всі збережені пости з Інстаграм збереглись. Код програми та поетапний опис до нього [див. Додаток 12]:

```
#Імпортуємо Instaloader та модуль os
import instaloader
import os

#Отримуємо екземпляр Instaloader
loader = instaloader.Instaloader(compress_json=False)

#Авторизуємось на своїй сторінці Інстаграм, де збережені пости
def save_insta_collection():
    USERNAME_P = "cursach2022"
    PASSWORD_P = "пароль"

    loader.login(USERNAME_P, PASSWORD_P)
    generalFolder = 'MyPosts'

   #Підключаємось до нашого профілю
    profile = instaloader.Profile.from_username(loader.context, USERNAME_P)

   #Створюємо каталог
    try:
        os.mkdir(generalFolder)
        print("Directory ", generalFolder, " Created ")
    except FileExistsError:
        print("Directory ", generalFolder, " already exists")
    os.chdir(generalFolder)

    #Завантажуємо кожен пост та метадані до нього
    try:
        for post in profile.get_saved_posts():
            loader.download_post(post, '{})'.format(post.pcaption))
            print("Post ", post.pcaption, " saved")

    except IndexError:
        print("You have no saved posts yet.")

    print("All posts saved")


save_insta_collection()
```

3.3. Структура бази даних текстів Інстаграм

В результаті була створена таблиця з усіма зібраними текстами та іншою інформацією (див. Додаток 1) і папка з фото (див. Додаток2).

Створена база даних у вигляді таблиці має 11 колонок:

1. ID тексту.
2. Текст.
3. Джерело (пост в Інстаграм/ сторіс в Інстаграм).
4. Автор (назва сторінки в Інстаграм).
5. Тематика сторінки.
6. Дата створення тексту.
7. Дата внесення тексту до БД.
8. Тема (чітких тем не було від авторів, тому в таблиці подана коротка інформація про пост, тобто про що йдеться у тексті).
9. Мова, якою написаний текст.
10. Візуальні елементи (наявність фото та\ або відео).
11. Примітки.

База даних містить 220 зібраних текстів, що мають різну функціональну спрямованість, а саме:

1. 50 текстів бізнес-сторінок.
2. 50 текстів сторінок блогерів і інфлюенсерів.
3. 50 текстів персональних сторінок політичних діячів.
4. 50 текстів сторінок різних піднаправлень – сторінки воєнних, новин (тсн, 1+1, Україна) тощо.
5. 20 рекламних текстів, що зустрічались під час збору (пости, сторіс).

Обсяг постів різний – від одного речення до максимального обсягу символів для поста в Інстаграм. Через обмежений доступ символів у десятках постів

інформація подана за допомогою інфографіки. Також наявне співавторство (коли пост має двох авторіві пости публікуються на обох сторінках).

3.4. Аналіз зібраних текстів

Активність користувачів Інстаграм почала набувати стрімкого зросту, коли розпочались воєнні дії 2022 року. Конфлікти та інші ситуації в країні мають вплив на велику аудиторію, тож ціль та мета соціальних мереж зазнала змін з подіями останніх років. Люди використовують соціальні медіа для спілкування, обміну новинами\ фото- та відеоматеріалом, для вираження своїх поглядів чи емоцій. Мільйони користувачів найпопулярніших соціальних мереж об’єднались і активно продовжують працювати на інформаційному фронті до сьогодні. В наш час інформаційна війна є не менш важливою, адже її завдання полягає в тому, щоб вплинути на інформацію супротивника з метою підриву його боєздатності і навпаки – захистити свій інформаційний потік від ворожого впливу.

Проаналізувавши зібрані тексти можна сказати, що наша інформаційна армія має досить потужну силу і вплив. Сторінки будь-якого направлення намагались якомога більше розповісти про події, звернутись до когось і закликати читачів і не тільки до тих чи інших дій, попросити про допомогу тощо.

Війна вплинула і на написання самих текстів. Багато зібраних текстів написані двома, а то й більше мовами, серед яких є турецька, іспанська, англійська, російська. Звичайно, тексти, до прикладу, блогерів і політичних діячів значно відрізняються. Перші звертаються більше до народу, своїх колег російського ринку[див. Додаток 1, тексти 35, 53, 54, 63], другі – до всесвітньо відомих людей з офіційними заявами і проханнями [див. Додаток 1, тексти 71, 77, 84]. У перших спостерігається набагато більше ненормативної лексики (у однієї з блогерок навіть є пост про лайку українською) [див. Додаток 1, текст 85] і ‘легших’ постів для сприйняття, присутній більш розважальний контент.

Відомо, що на початку подій бізнеси і компанії тимчасово припинили свою роботу, але продовжували підтримувати комунікацію з клієнтами. Допомагали, кому це було потрібно. Проте згодом все почало відновлювати свою роботу і відповідно у таких (про відновлення роботи і тд) текстах почали використовувати зовсім незвичні слова і навіть неологізми. Декілька прикладів:

1. Використання рядків з пісні гурту Мері ‘Я з України’. Так мережа доставки суш і ролів повідомила про відновлення роботи у декількох містах:

   “ВИ СА ЛЬВОВА?
   HI, SUSHI ICONS з України — це навпроти Європи, це там, де Хрещатик, це там, де Дніпро.”

2. Використання популярних фраз:

   “WHAT’S YOUR SUPERPOWER?
   I AM UKRAINIAN”

   “Доброго дня!) Ми з України!
   Доброго вечора, ми з України!”

3. Також в деяких постах спостерігається використання таких слів як ‘джавелін’, ‘байрактар’, ‘укриття’, ‘наш Крим’, які привертають увагу читача.

Спостерігаються зміни у значеннях деяких слів, з’явились нові асоціації. Так, в одному пості путіна назвали ‘кремлівським параноїком’, орки – це вже не істоти з фантастичних творів, кремезні варвари зі звіриними рисами, а військові рф. Ще одне цікаве словосполучення – корумповані посіпаки – продажні та підкупні люди (“…ця інформація стане нищівним ударом для Путіна та його "корумпованих посіпак".”) Слова росія, путін написані з маленької літери, а от українці, мир, свобода, герої, армія – з великої. Навіть офіційні сторінки популярних телеканалів називають нашу армію ‘котиками’ (‘Наші військові за будь-яких обставин

залишаються справжніми котиками, правда ж?’). Також яскравим прикладом змін є слова ‘мрія’ – відтепер в українців є дві головні асоціації з цим словом (мрія про щось і незламну Мрію, тобто літак) і бандерівське смузі, так називається коктейль Молотова. Оскільки Інстаграм вимогливо перевіряє лексику в текстах і блокує пост при виявленні чогось забороненого, нецензурного, українські користувачі у свої постах в багатьох словах і словосполученнях, які могли б вплинути на блокування, використовують цифри і іншомовні букви. Наприклад, у таких словах ‘бандерівські смузі’, ‘протитанкові’, ‘в0ює’, ‘холєр@’ можна спостерігати такі зміни. У примітках до тексту вказані ці спостереження.

Щодо рекламних текстів, то в цьому направленні теж почали використовувати патріотичний сенс. В текстах якомога частіше зустрічаються синьо-жовті сердечка, емодзі у вигляді українського прапора, тепер вже відомі всьому світові фрази тощо. У такий спосіб, це вплинуло не лише на хорошу статистику і підняття активності, продажі послуг і товарів, а й на патріотичне виховання. Теперішня ситуація в Україні завдала тенденцію і спонукає бренди до використання у своїх текстах патріотичної ідеї. Так, серед зібраних текстів можна спостерігати пости про традиційні цінності, культурну спадщину тощо.

Слід згадати про використання *хештегів* в текстах. Хештег – це ключове слово або фраза, яким передує символ «#» (октоторп), використовувані веб-технологіями для структурування текстових повідомлень за темою або типом. Короткі повідомлення в соціальних мережах, таких як: Instagram, Twitter, Facebook — можуть бути помічені хештегом, включаючи в себе як одне слово, так і більше об'єднаних слів (але без пробілів). Хештеги дають можливість групувати подібні повідомлення, таким чином можна знайти хештег і отримати набір повідомлень, які його містять. У зібраних текстах використано велику кількість хештегів, найбільш популярні з них – #StopRussianAggression, #StandWithUkraine, #UAразом, #prayforukraine, #ДякуюЗСУ, #СлаваУкраіні!, #SaveKidsOfUkraine, #PeaceForUkraine, #nowarinukraine, #STOPWAR, #WORLDFORUKRAINE, #SaveMariupol, #AzovRegiment.

Особливу увагу хочеться звернути на хештег #BuchaMassarce, цей хештег мені вдалось знайти лише в одному пості. З масовим поширенням постів з цим хештегом і фото/ відео масових вбивств у місті Буча соціальні мережа Інстаграм і Фейсбук почали блокувати дописи #Bucha і #BuchaMassacre. Їх блокування почалось через автоматизовані системи, які сканують насильницькі зображення на вищезгаданих соціальних платформах. Проте компанія блокує не всі хештеги і фото/ відео і додає попереджувальні мітки до деяких дописів і приховує відверті зображення.

Висновки до розділу 3

За допомогою чіткого дотримання алгоритму було успішно зібрано тексти, опубліковані з початку російсько-української війни, та створено базу даних. Процес завантаження текстів з Інстаграму був виконаний за допомогою програмування (мова програмування Python, середовище – Pycharm) і відповідних для роботи бібліотек (Instaloader, pip, os та інші). Це допомогло нам зекономити час і пришвидшити процес створення бази даних. Кожен текст зберігався в окрему папку з іншими метаданими до нього. Наступним етапом було створення бази даних.

Було вирішено, що така інформаційна база буде створена в Google Таблиці через її зручність і функціональність. У базі даних міститься 220 текстів, які відносяться до різних функціональних спрямованостей. Конкретні категорії текстів та їх кількість такі: 50 текстів належать до бізнес-сторінок, 50 текстів стосуються сторінок блогерів і інфлюенсерів, 50 текстів пов'язані з персональними сторінками політичних діячів, 50 текстів представляють різні піднапрямки, включаючи сторінки воєнних, новинних ресурсів (таких як ТСН, 1+1, Україна) та інші, 20 текстів є рекламними, що зустрічалися під час збору даних (пости, сторіс). Таким чином, база даних містить текстовий матеріал з різних джерел, що охоплюють різноманітні сфери та функціональність. Створена база даних є цінним ресурсом для різних видів досліджень як аналіз соціальних мереж (дослідити і зрозуміти інтернет-комунікацію, спосіб спілкування блогерів, користувачів соціальних мереж тощо), лінгвістичні дослідження (тексти можуть бути використані для аналізу мовних особливостей, стилістичних засобів, тематичних та контекстуальних відмінностей між різними

категоріями текстів), маркетингові дослідження (рекламні тексти можуть бути використані для аналізу маркетингових стратегій, просування продуктів і послуг, а також оцінки ефективності рекламних кампаній) тощо.

Активність користувачів Інстаграму стрімко зросла після початку воєнних дій у 2022 році. Це сильно вплинуло на широку аудиторію, змінивши ціль та призначення соціальних мереж, особливо на просторах українського Інстаграму. Мільйони користувачів головних соціальних мереж активно працюють на інформаційному фронті й досі. Даний розділ розглядає особливості написання постів українських користувачів під впливом війни, які включають в себе появу нової лексики, іншомовних слів та фраз, використання хештегів, ненормативної лексики та емодзі. Війна на інформаційному фронті внесла свої корективи у інтернет-спілкування та текстові блоки Інстаграму.

РОЗДІЛ 4. СТВОРЕННЯ КОРПУСУ ТЕКСТІВ ВОЄННОГО ЧАСУ СОЦІАЛЬНОЇ МЕРЕЖІ ІНСТАГРАМ

4.1. Корпус і його особливості

Створення корпусу – це систематизація і організація текстових даних для подальшого використання у лінгвістичних дослідженнях або розробці мовних моделей. Корпус – це колекція текстів або мовних зразків, яка може включати писемний текст, усну мову, переклади та інші типи лінгвістичних даних.

Сьогодні корпусний метод активно застосовується у лінгвістичних дослідженнях різних мов. Мовні корпуси можуть мати різні розміри, типи, структури, зміст та призначення. Збір і аналіз писемних і усних мовних даних дозволяють вивчати окремі явища мови і розкривати закономірності функціонування мовних одиниць на різних рівнях. Розвиток корпусної лінгвістики в Україні є активним процесом, що отримує все більшу увагу і визнання. Останні роки в Україні активно розробляються та створюються україномовні корпуси, що містять текстові дані з різних джерел, включаючи літературні тексти, медійні матеріали, наукові дослідження тощо. Це сприяє розвитку української корпусної лінгвістики та забезпечує наявність ресурсів для досліджень української мови.

Одним з найбільших і успішних проектів корпусної лінгвістики в Україні є Корпус української мови. Корпус української мови (КУМ) – це один з найбільших лінгвістичних ресурсів, який містить широкий спектр текстів українською мовою. КУМ включає писемні тексти різних жанрів та стилів, такі як законодавчі, наукові, поетичні тексти, публіцистика, художня проза, фольклорні тексти тощо. Корпус української мови працює як інформаційно-довідкова система [22] та має тестову електронну версію, яка представлена на лінгвістичному порталі Mova.info; її обсяг – понад 100 млн слововживань [22].

Створення корпусу включає вибір джерел текстів, таких як книги, статті, газети, блоги, соціальні медіа, інтерв'ю, записи розмов тощо. Потім проводиться збір цих текстів і їх обробка, включаючи анотування, класифікацію та індексацію. Під час анотування додаються метадані до текстів, наприклад, інформація про автора, дату, жанр тощо. Класифікація допомагає організувати тексти за категоріями або

темами, а індексація дозволяє швидко знаходити та отримувати доступ до конкретних текстів або їх фрагментів. Цей процес дозволяє дослідникам та лінгвістам проводити різноманітні аналізи текстів, вивчати мовні явища, розвивати мовні моделі та здійснювати порівняльні дослідження між різними мовами чи варіантами мови.

До особливостей корпусу відносяться: різноманітність мовних даних (мовний корпус містить тексти з різних жанрів, стилів, часу та контекстів. Це надає можливість вивчати різноманітні мовні явища і варіації в мовленні); розмір та обсяг (мовні корпуси можуть мати різні розміри, від невеликих масштабів до великих – чим більший обсяг корпусу, тим більше даних доступно для аналізу і досліджень); представлення реального мовлення (корпуси мають на меті відображати реальне мовлення, яким користуються носії мови в різних контекстах. Це дозволяє вивчати природну мову та розуміти її використання в реальному житті); доступність та використання (корпуси можуть бути доступними для дослідників і лінгвістів з метою аналізу, перевірки гіпотез, розробки мовних моделей, вони використовуються для розвитку мовознавства, машинного перекладу, автоматичного розпізнавання мови та інших областей).

Залежно від цілей дослідження та використання, можна виділити кілька типів мовних корпусів. Наприклад, білінгвальний корпус містить тексти на двох різних мовах і використовується для перекладу та вивчення взаємодії між мовами. Монолінгвальний корпус складається з текстів лише однієї мови і дозволяє проводити аналіз мовних явищ та розробку мовних моделей. Корпус усного мовлення містить записи розмов, інтерв'ю та інші усні мовленнєві зразки і використовується для дослідження фонетики та соціолінгвістики. Корпус писемного мовлення містить письмові тексти, такі як книги, статті, листи, і дозволяє проводити аналіз граматики та стилістики писемного мовлення. Корпус спеціалізованої мови містить тексти з певних галузей, таких як медицина, юриспруденція, технології, і використовується для розвитку термінології та спеціалізованого аналізу. Корпус паралельних текстів містить тексти на одній мові, але у паралельних варіантах і використовується для машинного перекладу та порівняльного аналізу мови.

4.2. Створення корпусу текстів (програмне забезпечення)

Кінцевим етапом дослідження є створення корпусу текстів за укладеною базою даних. Підготовка корпусу включає кілька важливих етапів, які допомагають забезпечити якість і готовність корпусу для подальшого використання в лінгвістичних дослідженнях. Створення корпусу є досить часомістким і складним процесом, що включає в себе декілька важливих етапів:

1. Вибір джерел. Визначається, з яких джерел будуть зібрані текстові дані. Для створення корпусу було обрано одне інформаційне джерело – соціальну мережу Інстаграм. Інстаграм є популярною платформою, де користувачі діляться своїми фотографіями та публікують підписи, коментарі, хештеги тощо. Збір текстових даних з Інстаграму дозволяє отримати широкий спектр мовних зразків, які відображають різноманітні жанри, стилі, контексти та варіації мовлення. Такий вибір джерела дозволяє отримати цікаві дані для лінгвістичних досліджень та аналізу мовних явищ у соціальних медіа.
2. Збір текстів. Цей процес відбувався тільки з одного джерела – соціальної мережі Інстаграм. Використовуючи програмування і бібліотеку Instaloader вдалось уникнути ручного збору текстів та втілити все автоматично, окрім внесення матеріалу до бази даних. У вище написаному розділі розповідається детальніше про процес завантаження текстів та інших метаданих з Інстаграм, і оформлення всього матеріалу в базу даних. Створення пошуку відбувалось на основі вже попередньо створеній БД.
3. Очищення текстів. Зібрані тексти можуть містити небажані елементи тож виконується очищення текстів від небажаних елементів, таких як HTML-теги, емодзі, хештеги.
4. Класифікація та індексація. Тексти можуть бути класифіковані за різними параметрами, такими як жанр, автор, тематика тощо. Для зручного доступу до текстів була проведена індексація текстів та створення відповідних метаданих.

5. Написання програм для створення саме корпусу – робота з Google Drive, обробкою даних з таблиці, текстів, створення і керування веб-додатком тощо.

Ці етапи підготовки корпусу важливі для забезпечення якості даних, їхньої структурованості та використання корпусу у лінгвістичних дослідженнях.

Для створення корпусу текстів було обрано мову програмування Python. Для розробки коду було використане інтегроване середовище розробки для мови програмування Python – PyCharm.

Також для роботи був використаний інший програмний функціонал, включаючи бібліотеки, модулі та функції:

1. re (import re) – модуль для роботи з регулярними виразами, використовується для пошуку числа в імені папки.
2. logging (import logging) – модуль для логування помилок та подій у програмі.
3. pandas (import pandas as pd) – бібліотека для обробки та аналізу даних, використовується для створення DataFrame (це основна структура даних в pandas, що дозволяє представляти дані у вигляді табличної структури з рядками та стовпцями).
4. gspread (import gspread) – бібліотека для роботи з електронними таблицями Google, використовується для зчитування даних з електронної таблиці.
5. os (import os) – це вбудована бібліотека Python, яка надає функції для взаємодії з операційною системою. Вона дозволяє виконувати різноманітні операції, такі як робота з файловою системою, отримання і налаштування змінних середовища, взаємодія з процесами та багато іншого.
6. datetime (from datetime import datetime) – клас для роботи з датою і часом, дозволяє виконувати різні операції з датами та часом, такі як створення об'єктів дати та часу, отримання поточної дати та часу,

форматування та розбір рядків, виконання арифметичних операцій, порівняння дат та часу тощо.

7. streamlit (import streamlit as st) – це бібліотека на основі Python, яка дозволяє швидко створювати інтерактивні веб-додатки для візуалізації даних, машинного навчання та інших завдань обробки даних. Вона пропонує простий та інтуїтивно зрозумілий спосіб відтворювати код Python у веб-браузері, що дозволяє вам швидко створювати і змінювати додатки. Streamlit також має вбудовані функції для візуалізації даних, інтерактивного вибору параметрів та взаємодії з користувачем.
8. config (from config import IMG_URL) – модуль config імпортується з файлу `config.py`, із нього використовується змінна `IMG_URL`. Його основна мета - забезпечити зручний спосіб конфігурації параметрів програми з використанням зовнішнього конфігураційного файлу.

Дослідження проводилось на попередньо зібраних матеріалах, що були систематизовані як база даних в Google Таблиця. Ця таблиця бази даних містить інформацію про текстові дані, зібрані з постів та сторісів в Інстаграм. Кожен запис у таблиці включає унікальний ідентифікатор тексту, сам текст, вказівку на джерело (пост або сторіс), автора (назву сторінки в Інстаграм), тематику сторінки, дату створення, дату внесення тексту у корпус, коротку інформацію про тему тексту, мову, якою написаний текст, візуальні елементи (фото або відео) та примітки. Ця структура дозволяє зберігати та організовувати дані для подальшого використання у мовних дослідженнях та аналізі.

Створення корпусу – це досить велика за об’ємом робота. Процес написання коду відбувався у декілька етапів: робота з Google Drive, обробка даних з електронних таблиць, обробка текстових даних, створення сторінки Streamlit, обробка всіх даних і відображення результатів на сторінці Streamlit.

Проект має декілька програмних кодів, кожен з яких виконує свою частину завдання: drive_utils.py (див. Додаток 3), spreadsheet_utils.py (див. Додаток 4), text_utils.py (див. Додаток 5) і головний програмний код – main.py (див. Додаток 6).

Для подальшого розуміння роботи програми, ми вирішили описати кожен код окремо. Нижче ми подали опис кожного файлу, його завдання і опис головних функцій, що використовуються у програмі.

Файл **drive_utils.py** містить декілька функцій для роботи з Google Drive, пов'язаної з обробкою зображень і папок. Основні функції, що містяться у файлі:

1. “get_all_image_folders(folder_id: str, service) -> list” – ця функція отримує всі папки зображень з Google Drive. Вона приймає ідентифікатор “folder_id” папки, з якої потрібно отримати папки зображень, і об'єкт “service”, який використовується для взаємодії з Google Drive API. Функція повертає список файлів зображень (об'єктів з ідентифікатором та ім'ям файлу).
2. “get_list_images(folder_id, service)” – ця функція отримує список зображень із заданої папки `folder_id` в Google Drive. Вона також використовує об'єкт “service” для взаємодії з Google Drive API. Функція повертає URL першого зображення у форматі JPEG, якщо таке зображення знайдено в папці, або “None”, якщо зображень немає.
3. “preprocess_image_folders(folders: list) -> pd.DataFrame” – ця функція перетворює список папок зображень у вхідному форматі в об'єкт “pd.DataFrame” (DataFrame) з двома стовпцями: “link” (посилання на папку) і “name” (ім'я папки). Функція повертає отриманий DataFrame.

Файл **spreadsheet_utils.py** містить декілька функцій для обробки даних з електронних таблиць (spreadsheet) із використанням бібліотеки pandas та gspread. Основними функціями цього програмного коду є:

1. “preprocess_spreadsheet(sp_df: pd.DataFrame) -> pd.DataFrame” – ця функція виконує попередню обробку даних з електронної таблиці. Вона приймає DataFrame “sp_df” з даними з електронної таблиці і повертає оброблений DataFrame. В процесі обробки виконуються деякі операції, такі як заміна порожніх значень на “pd.NaT”, заповнення пропусків в стовпці “ТЕМАТИКА СТОРІНКИ” за попереднім значенням, перетворення стовпців “ДАТА

СТВОРЕННЯ” і “ДАТА ВНЕСЕННЯ” на об'єкти типу “datetime”, перейменування стовпця “№ ТЕКСТА” на “text_number” і видалення деяких непотрібних стовпців.

2. “read_spreadsheet(url: str, creds) -> pd.DataFrame” – ця функція зчитує електронну таблицю з вказаного URL. Вона приймає рядок “url”, що містить посилання на таблицю, і об'єкт “creds”, який використовується для авторизації. Функція зчитує дані з таблиці за допомогою бібліотеки gspread, створює DataFrame з отриманими даними і передає його до функції “preprocess_spreadsheet” для попередньої обробки. Після чого повертає оброблений DataFrame з даними.
3. “make_clickable(link)” – ця функція приймає посилання “link” і повертає рядок HTML зображення, якщо посилання не є порожнім, або порожній рядок, якщо посилання відсутнє.

Файл програми **text_utils.py** містить декілька утилітних функцій для обробки текстових даних. Опис функцій, які містяться у файлі:

1. “find_matches(text: str, lemma: str, tp) -> bool” – ця функція перевіряє, чи містить текст “text” лему “lemma”. Вона приймає текст, лему та інструмент обробки тексту “tp” (наприклад, SpaCy) і повертає кортеж, що складається з результату обробки тексту та прапорця, який показує, чи було знайдено співпадіння леми в тексті. Ця функція дозволяє попередньо обробити текст посту і отримати потрібні нам атрибути. Тобто на цьому етапі відбувається лематизація.

2. “find_by_phrase(text: str, word: str)” – ця функція знаходить фразу “word” в тексті “text” і позначає її тегом “”. Вона повертає новий текст, в якому знайдені фрази виділені тегом “”, і прапорець, який показує, чи було знайдено фразу в тексті.

3. “find_texts(word: str, text_df: pd.DataFrame, tp, mode: str = "lexemma") -> pd.DataFrame” – ця функція знаходить тексти зі співпадіннями слова або леми в залежності від заданого режиму (“mode”). Вона приймає слово, DataFrame з

текстовими даними “text_df”, інструмент обробки тексту “tp” і режим пошуку (“mode”). Режим може бути “lexemma” (шукає співпадіння за лемою) або “word form” (шукає співпадіння за словоформою). Функція повертає DataFrame зі знайденими текстами, які містять співпадіння.

4. “merge_and_save_df(left, right, save: bool, word: str) -> pd.DataFrame” – ця функція об'єднує два DataFrame (“left” і “right”) за спільним стовпцем “text_number” і зберігає результат у файлі Excel, якщо параметр “save” встановлено в “True”. Вона повертає об'єднаний DataFrame.

Також для роботи були використанні змінні: “SAVE_DIR” – шлях до теки, в якій зберігаються результати, “HEADER_ROW” – номер рядка, що містить заголовки стовпців у таблиці, “DATA_START_ROW” – номер рядка, з якого починаються дані у таблиці.

Програмний файл **main.py** – основний код у роботі. Цей файл містить головний код програми, який керує веб-додатком на основі бібліотеки Streamlit. Додаток призначений для обробки даних з електронних таблиць Google та відповідного зображення з Google Drive. Основні функції та їх опис:

1. “main()” – основна функція програми. Вона налаштовує сторінку Streamlit, отримує вхідні дані від користувача, виконує обробку даних та відображає результати на сторінці. Вона викликає функції з модулів “drive_utils.py”, “spreadsheet_utils.py” та “text_utils.py” для отримання та обробки даних.

2. “SPREADSHEET_URL”, “IMAGE_FOLDER_ID”, “SCOPES”, “CREDS_PATH” – змінні, що містять URL електронної таблиці, ідентифікатор папки зображень, області доступу та шлях до файлу з обліковими даними.

3. “st.set_page_config(layout="wide")” – налаштування конфігурації сторінки Streamlit для широкого макету.

4. “st.columns(2)” – створення двох колонок на сторінці.

5. “read_spreadsheet(url=SPREADSHEET_URL, creds=creds)” – виклик функції для зчитування електронної таблиці за вказаним URL та обліковими даними.

6. “find_texts(text_df=text_df, word=word, mode=mode, tp=text_preprocessor)” – виклик функції для пошуку текстів зі співпадіннями слова або леми в залежності від заданого режиму.

7. “get_all_image_folders(folder_id=IMAGE_FOLDER_ID, service=service)” – виклик функції для отримання списку папок зображень з Google Drive

8. “preprocess_image_folders(folders=folders)” – виклик функції для попередньої обробки списку папок зображень.

9. “merge_and_save_df(filtered_df, imagelink_df, save, word)” – виклик функції для об'єднання двох DataFrame та збереження результату.

10. “result_df['link'] = result_df['link'].apply(lambda link: get_list_images(link, service))” – отримання посилання на відповідне зображення для кожного рядка в DataFrame.

11. “result_df['link'] = result_df['link'].apply(make_clickable)” – застосування функції `make_clickable` до стовпця “link” для створення посилань на зображення.

12. “result_df.to_html(index=False, escape=False, col_space=300)” – конвертування DataFrame у формат HTML.

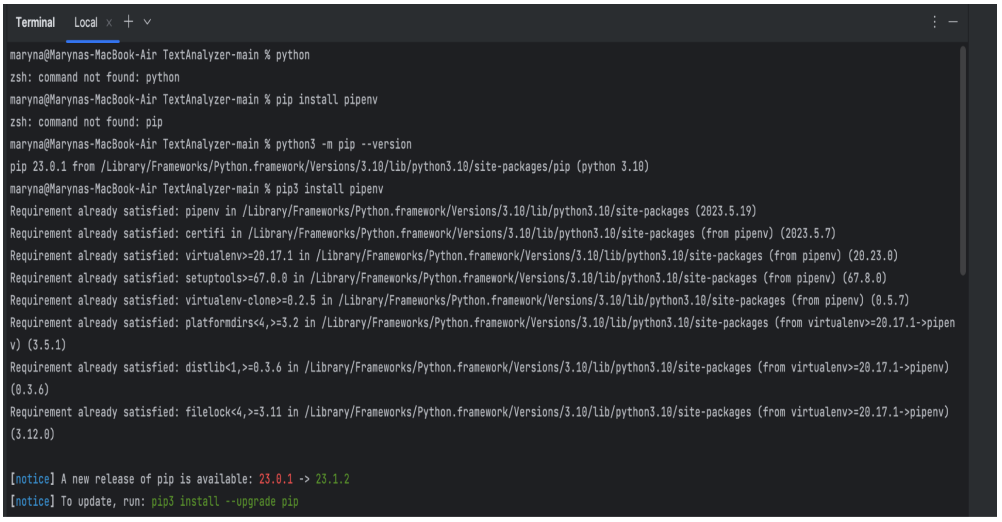
13. “st.write(html_df, unsafe_allow_html=True)” – виведення результатів на сторінку Streamlit у вигляді HTML.

14. “logging.basicConfig(level=logging.INFO)” – налаштування рівня логування на рівень INFO.

15. “if __name__ == '__main__':” – перевірка, чи файл “main.py” виконується безпосередньо, і виклик функції “main()”.

Цей файл об'єднує функціональність різних модулів та налаштовує веб-додаток для відображення та обробки даних з електронних таблиць та зображень з Google Drive.

Для того, щоб запустити комплексну роботу всієї програми наступним етапом було встановлення pipenv (див. Додаток 7) – це інструмент для керування залежностями у проектах Python, що дозволяє зручно встановлювати та оновлювати пакети. Він також створює віртуальне середовище, яке ізолює залежності проекту від системних, що полегшує управління проектом та його розгортання.

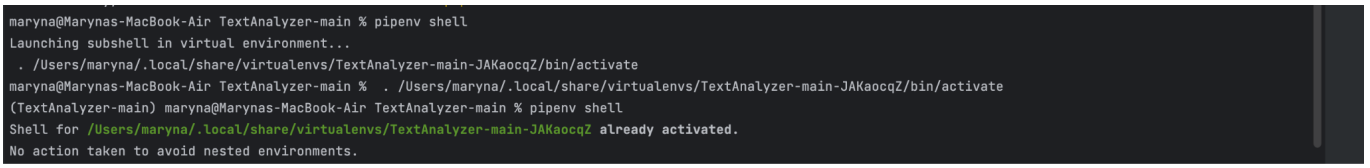


Мал. 4.1. Встановлення pipenv

Наступним етапом були: pipenv install – встановлення всіх залежностей для проекту (див. Додаток 8):

Мал.4.2 Встановлення всіх залежностей для проекту

Pipenv shell – активація середовища (див. Додаток 9):



Мал. 4.3. Активація середовища

Python -m spacy download uk_core_news_sm – встановлення мовної моделі для роботи з українською мовою (див. Додаток 10). І останній етап – запуск програми за допомогю streamlit run main.py (див. Додаток 11):

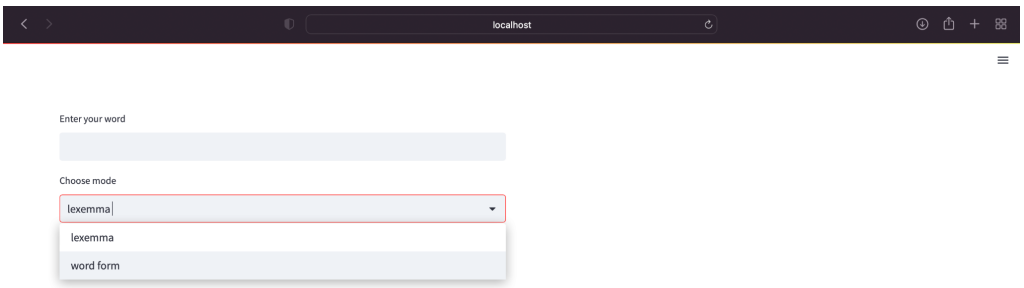
Мал. 4.4. Запуск веб-додатку з корпусом текстів

Були успішно виконані завдання на основні програмування і використання елементів програмної реалізації – використання функцій, бібліотек, модулів, змінних тощо. Труднощі виникали з імпортуванням pipenv, проте це питання було вирішено за допомогою модуля pip і роботи у консолі.

4.3. Параметри пошуку та користувацькі можливості корпусу текстів

Створений корпус містить зібрані тексти українського Інстаграму різних профілів з початку російсько-української війни.Програма може слугувати як об’єкт для подальших лінгвістичних досліджень, особливо вживання української мови на просторах соціальних мереж, а також зміни у використанні лексики з початку воєнних дій у лютому 2022 року.

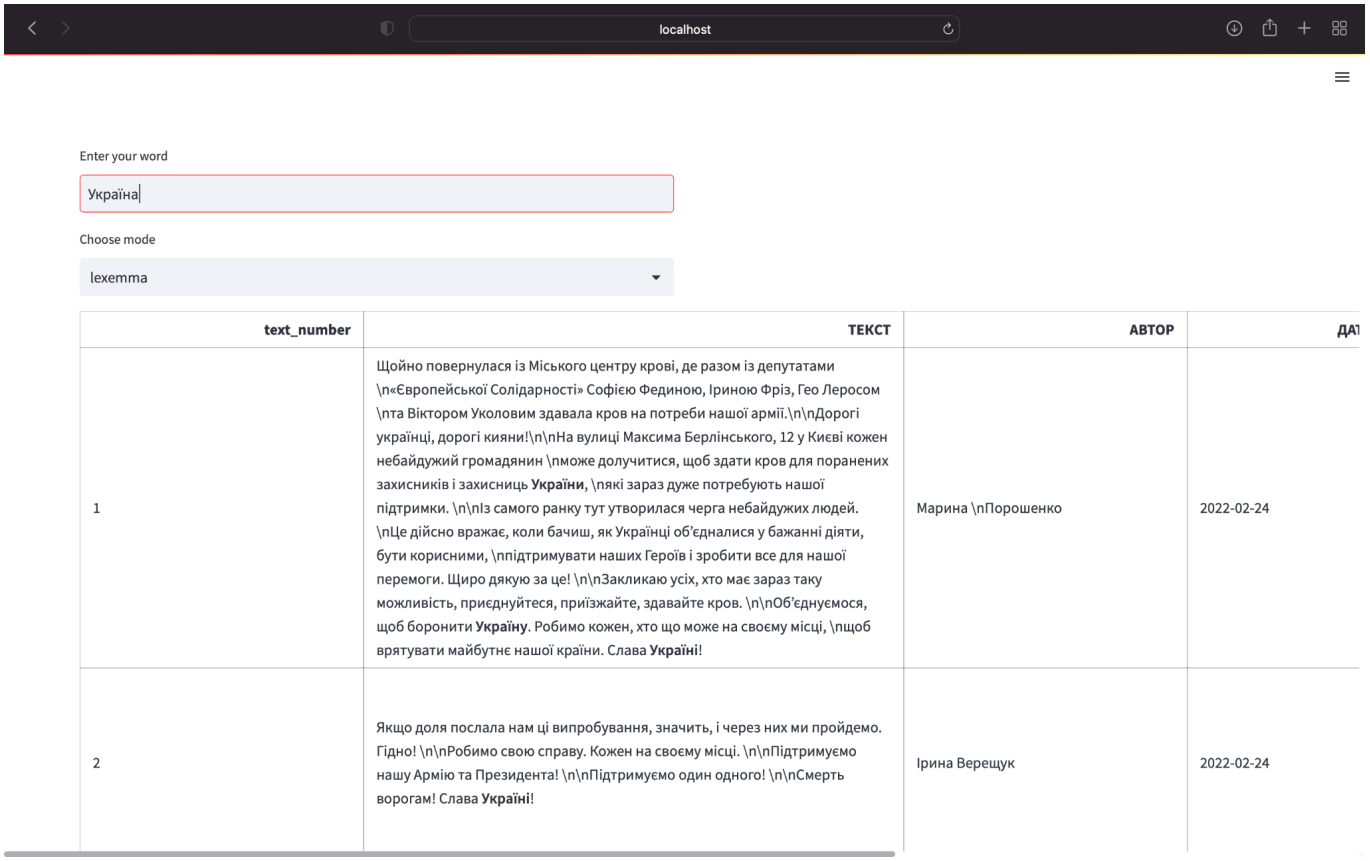
Програма успішно працює і створює корпус за допомогою бібліотеки Streamlit – працює для створення інтерактивних веб-додатків з візуалізацією даних. При запуску коду програма запускає вікно веб-версії, де користувач бачить порожнє вікно з двома функціями: enter word – користувач може ввести будь-яке слово для пошуку в корпусі та choose mode – за допомогою цього критерію користувач має змогу обрати параметр, за яким буде відбуватись пошук введеного слова: за лемою чи словоформою:



Мал. 4.5. Запуск коду та початкове вікно створеного корпусу

При вводі слова користувачем відбувається пошук введеного слова. У результаті програма виводить всі результати, знайдені у корпусі. Структура корпусу складається з номеру тексту у корпусі, тексту та інших метаданих до нього, а саме – автор тексту, мову тексту (інколи трапляються двомовні тексти), фото поста та дата публікації поста.

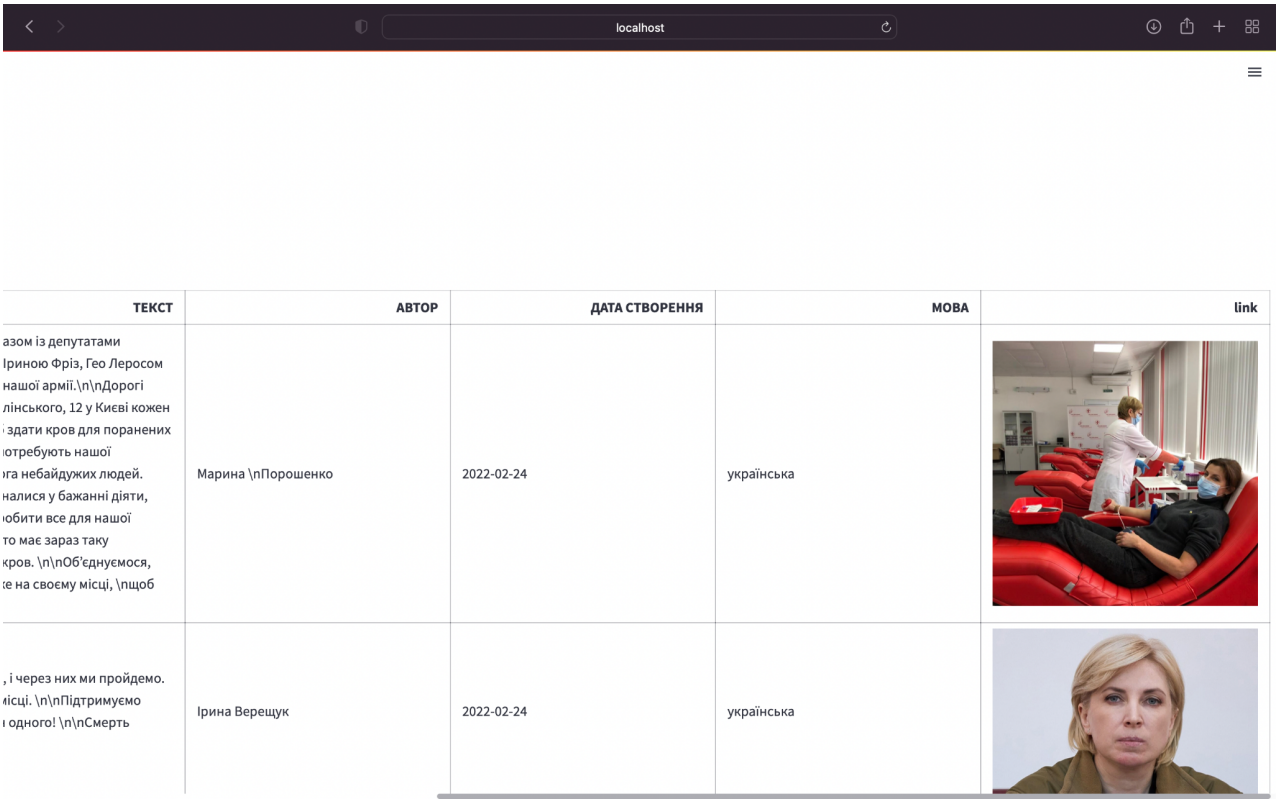
Для перевірки роботи програми ми вирішили дослідити процес роботи на основі слова “Україна”.



Мал. 4.6. Пошук введеного слова за лемою і результат пошуку

Пошук слів за лемою займає більше часу, ніж за словоформою. За результатами корпус виділяє у текстах слово, яке вводиться користувачем у пошукову систему. Також до кожного тексту підтягуються його метадані – автор, мова, дата публікації і фото (підтягується лише перше фото поста, оскільки траплялись дописи з кількома фотографіями). На жаль, не вдалось реалізувати

прикріпленні відео до корпусу, тому на місцях, де має бути відео, програма виводить пустий квадрат. Знайти відеоматеріал можна у базі даних (див. Додаток 1).



Мал. 4.7. Інші метадані, прикріплені до слова у корпусі

У корпусі відображено тексти воєнного періоду різних людей, блогерів і бізнес-сторінок. Всі вони мають різний тип написання, різну лексику і оформлення. Проте їх об’єднує одна важлива річ – угрупування і спільні зусилля. Майже всі тексти пов’язані з війною – хтось просто виражає свої емоції, хтось висловлює підтримку, інші пишуть двомовні пости для залучення більшої аудиторії з усього світу або відкривають збори для військових. Тому тексти у корпусі мають багато спільної лексики і фраз.

Висновки до розділу 4.

Створення корпусу є важливим етапом для збереження та аналізу текстових даних з різних джерел. Цей процес є складним і часомістким, проте дозволяє на практиці дослідити зібрані тексти і полегшити роботу пошуку слова.

Створення корпусу з текстами українських користувачів Інстаграму під час воєнних дій дозволяє зберегти і проаналізувати цінну інформацію, що відображає думки та переживання українського населення в цей нелегкий період. А також може слугувати для подальших лінгвістичних досліджень і інших завдань – документація подій, аналіз мови та інтернет-спілкування під час цього періоду тощо. Корпус, що зроблений для дослідження, показав не лише особливості у текстах, написаних під час російсько-української війни, а й згрупованість користувачів українського Інстаграму, їх спільної інформаційної боротьби у текстовому і візуальному форматі.

ВИСНОВКИ

Однією з популярних соціальних мереж є Інстаграм. Дана платформа вважається не просто інструментом для поширення своїх фотографій і відеоматеріалів. Наразі це сильний елемент у розвитку бізнесу, особистих профілів і, врешті-решт, має сильну інтернет-комунікацію. Інстаграм відіграє важливу роль у сучасній інтернет-комунікації. Ця соціальна мережа надає можливість з легкістю спілкуватися, обмінюватися думками і ідеями через коментарі, приватні повідомлення та інші функції. Вона створює зручну платформу для взаємодії між користувачами з усього світу, незалежно від їх географічного розташування і часового поясу. Звісно, невід'ємною частиною є текстовий блок, без нього не було б комунікації між користувачами. Він використовується всюди – пост, коментарі, приватні повідомлення, рекламні тексти і креативи, опис про себе тощо.

Варто зазначити, що український Інстаграм зазнав значних змін після початку російсько-української війни. Більшість українських блогерів, лідерів думок, бізнес-сторінок перейшли повністю на використання української мови у своєму контенті. На просторах інтернету з'явилось багато нової лексики, хештегів тощо. Це допомагає користувачам й досі поширювати свої пости, думки, збори на широку аудиторію. Завдяки хештегам увесь світ бачить наслідки воєнних дій на територіях України. Тож ми вирішили дослідити ці тексти детальніше і зібрати їх у корпус для подальшого використання у лінгвістичних дослідженнях.

Кінцевою метою дослідження було створення корпусу текстів воєнного періоду на просторах Інстаграм. Для досягнення мети були виконані такі завдання:

1. Дослідити особливості текстового блоку на просторах Інстаграму, включаючи інтернет-комунікацію. Соціальні мережі продовжують впливати на розвиток комунікації в Інтернеті та мови загалом. Це проявляється у появі нових слів і фраз, які використовуються як у віртуальному спілкуванні, так і в реальному житті. Також спостерігається використання скорочень, абревіатур, емодзі та інших елементів. Важливо відзначити, що спілкування в Інтернеті не завжди

сприяє поліпшенню граматичних та орфографічних навичок, але воно має вплив на мовну поведінку суспільства в цілому.

2. Процес пошуку текстів різних сторінок в Інстаграм з початку воєнного періоду успішно був завершений для подальшого дослідження. Тексти за цей період дійсно змінились – в постах, сторіс, рекламі почали використовувати синьо-жовті емодзі, використовувати нові фрази, більше постів спрямовано на допомогу ЗСУ, новини, контакти надійних волонтерів, культурну спадщину України. Тексти тісно пов’язані з патріотичною ідеєю і звичайно такі тексти впливають на патріотичне виховання. Так, до прикладу, на сторінці однієї зі студій краси стало більше постів про традиції українок ще з стародавніх часів, про українське вбрання, зачіски, чоловічі стрижки, весільні традиції, фільми.
3. Далі знайдені тексти були зібраними і завантаженими для створення бази даних. Зібрані тексти датуються з 24 лютого 2022 року. Ми обрали легший спосіб завантаження текстового контенту, щоб пришвидшити процес створення корпусу. Завантаження відбувалось за допомогою програмування. В результаті кожен текст з його метаданими був завантажений в окрему папку.
4. Важливим завданням на шляху до створення корпусу є база даних. Отже, було створено базу даних, яка включала в себе тексти і метадані до них, а саме – фото, дату публікації, автора, мову написання тексту, тематику сторінки тощо. Створена база даних зібраних текстів базується в Google Таблиці (див. Додаток 1).
5. Для легшої організації програми створення корпусу було здійснено очищення тексту від небажаних елементів, тож ми очистили тексти від таких елементів як емодзі, хештеги тощо.
6. Кінцевим етапом було написання програм для створення корпусу. Це потребувало виконання попередніх завдань, а також створення алгоритму і класифікації, за якими буде працювати програма. Для реалізації створення корпусу текстів було обрано мову програмування Python. Для написання коду було використане інтегроване середовище розробки PyCharm, спеціально призначене для роботи з мовою програмування Python. Крім того, в процесі

роботи було використано інші програмні компоненти, включаючи різноманітні бібліотеки, модулі та функції – Instaloader, os, re, pip, gspread, streamlit та інші. Всі завдання були успішно виконані – пошук у корпусі працює без помилок і видає результат з усіма метаданими (номер тексту, сам текст, дата публікації, мова написання тексту, фото поста). Єдине, що не вдалось реалізувати – прикріплення таких метаданих як відео до корпусу.

Отже, створений корпус текстів воєнного періоду може послужити зручним застосунком для лінгвістичного аналізу українськомовних текстів. А також може бути використаним для дослідження рекламних кампаній та соціального впливу, досліджуючи різні аспекти комунікації та ефективності повідомлень. Війна справді має значний вплив на написання текстів і їх важливість у соціальних мережах. У ситуації війни та конфліктів, коли доступ до візуальної інформації може бути обмеженим або небезпечним, текст стає основним засобом передачі інформації та спілкування з громадськістю.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Балог В. Передумови створення корпусу словників. *Лексикографічний бюлетень*. С. 174–176.
2. Барсукова О. Компанія Meta пояснила, чому Facebook та Instagram блокують хештеги про Бучу і фото міста. *Українська правда*. 2022. URL: https://life.pravda.com.ua/society/2022/04/5/248112/.
3. Бойко С. Соціальні мережі як інструмент інформаційної війни. *Українське суспільство в умовах війни: виклики сьогодення та перспективи миротворення* : матеріали II Всеукр. науково-практ. конф., м. Маріуполь, 15 черв. 2018 р. Маріуполь: ДонДУУ, 2018. 331 с., м. Маріуполь, 15 черв. 2018 р. С. 126–129.
4. Горлач П. З'явився словник неологізмів війни від проєкту АрміяInform. *Суспільне Культура*. 2022. URL: https://suspilne.media/264155-zavivsa-slovnik-neologizmiv-vijni-vid-proektu-armia inform/.
5. Дзюбіна О. КОМУНІКАТИВНИЙ АСПЕКТ СОЦІАЛЬНИХ МЕРЕж fACeBOOK I TwiTTeR. *Вісник дніпропетровського університету імені альфреда нобеля.*2016. С. 218–221.
6. Єльнікова Н. І. Мовний аспект комунікації в українських версіях соціальних мереж. *Сучасні проблеми правового, економічного та соціального розвитку держави.*, м. Харків, 12 січ. 2017 р. 2017. С. 43–46.
7. Жуковська В. Вступ до корпусної лінгвістики : Навч. посіб. Житомир : Вид-во ЖДУ ім. І.Франка, 2013. 142 с.
8. Івахненко Т. Культура мовлення в інтернет-комунікації. С. 97–99.
9. Калита О. Електронна комунікація: мовнокомунікативні особливості й жанрова диференціація. С. 102–106.
10. Коваленко М. Як зняти та викласти відео в Instagram Reels: повне керівництво. *Ramediaplus*. URL: https://ramediaplus.com.ua/iak-zniaty-ta-vyklasty-video-v-instagram-reels/.
11. Кульчицький А., Дігтярьова І. Інформаційні технології у якості зброї. 2017. С. 47–48.

12. Курбан О. Проблема критичності мислення при споживанні медіаконтенту в умовах інформаційної війни. *Синопсис: текст, контекст, медіа*. 2022. Т. 28 № 1. С. 21–26.
13. Маковцеька-Гудзь Ю. Українська мова в соціальних мережах. *Мовні і концептуальні картини світу*. 2014. Т. 1, № 47. С. 658–663.
14. Малєнкова Д. Мова українських блогерів у соціальній мережі «інстаграм». Херсон, 2020. 51 с.
15. Оборіна О. Класифікація рекламних текстів як лінгвістична проблема. URL: http://xn--e1aajfpcds8ay4h.com.ua/pages/view/601.
16. Свириденко А. Ефективність соціальної реклами: світовий досвід та українські реалії. Київ, 2015. 87 с.
17. Учасники проектів Вікімедіа. Інфографіка – вікіпедія. *Вікіпедія*. URL: https://uk.wikipedia.org/wiki/Інфографіка.
18. Учасники проектів Вікімедіа. Корпусна лінгвістика – Вікіпедія. *Вікіпедія*. 2008. URL: https://uk.wikipedia.org/wiki/Корпусна_лінгвістика.
19. Учасники проектів Вікімедіа. Корпус української мови – Вікіпедія. *Вікіпедія*. 2022. URL: https://uk.wikipedia.org/wiki/Корпус_української_мови.
20. Учасники проектів Вікімедіа. Орки – вікіпедія. *Вікіпедія*. 2011. URL: https://uk.wikipedia.org/wiki/Орки.
21. Картинка чи текст в Інстаграм: що важливіше?. *TextElit*. URL: https://textelit.com.ua/shho-vazhlivishe-v-instagram/#Q_2.
22. Лінгвістичний портал mova.info. *Лінгвістичний портал MOVA.info*. URL: http://www.mova.info.
23. Семененко К., Юрченко В., Скригун Н. Маркетингова діяльність у соціальних мережах “facebook” и “instagram”. *Вчені записки ТНУ імені В. І. Вернадського*.2019. Т. 30. С. 164–167.
24. Семен Н., Казімова Ю. Соціальна мережа instagram як сучасна платформа популяризації журналістського контенту. 2019. С. 95–97.
25. Сидоров М., Табаков Д. Інтернет як засіб соціальної комунікації та соціального впливу. *Політичний менеджент*. 2008. С. 119–125.

26. Сидорук Г. Інтернет-скорочення як засіб мовної економії. *Лінгвістика і поетика тексту*. 2016. С. 294–298.
27. Согорін А. Реклама як засіб патріотичного виховання (за результатами масового та експертного опитувань). *Українськиц соціум*. 2016. С. 134–137.
28. Українська мова у соцмережах: що змінилося після початку повномасштабної війни? *Український соціологічний портал*. URL: https://usp-ltd.org/ukrainska-mova-u-sotsmerezhakh-shcho-zminylosia-pislia-pochatku-povnomasshtabnoi-vijny/.
29. Учасники проектів Вікімедіа. Хештег – вікіпедія. *Вікіпедія*. 2013. URL: https://uk.wikipedia.org/wiki/Хештег.
30. Учасники проектів Вікімедіа. Цільова аудиторія – Вікіпедія. *Вікіпедія*. 2012. URL: https://uk.wikipedia.org/wiki/Цільова_аудиторія.
31. Шевчук І. Бізнес у соціальних мережах : Навч. посіб. Львів, 2021. 215 с.
32. Delin J. Brand Tone of Voice: a linguistic analysis of brand positions.
33. Gspread. *PyPI*. URL: https://pypi.org/project/gspread/.
34. Instagram і Facebook приховують дописи з хештегами звірств у Бучі. *Читомо | Культура читання і мистецтво книговидання | Усе про книги | Читомо - портал про культуру читання і мистецтво книговидання*. URL: https://chytomo.com/instagram-i-facebook-prykhovuiut-dopysy-z-kheshtehamy-zvirstv-u-buchi/.
35. Instagram. *Instagram*. URL: https://www.instagram.com.
36. Install instaloader – instaloader documentation. *Instaloader – Download Instagram Photos and Metadata*. URL: https://instaloader.github.io/installation.html.
37. Logging facility for Python. *Python documentation*. URL: https://docs.python.org/3/library/logging.html.
38. Pandas. *PyPI*. URL: https://pypi.org/project/pandas/.
39. Pip. *PyPI*. URL: https://pypi.org/project/pip/.
40. Ries L. Visual hammer. *Goodreads*. URL: https://www.goodreads.com/book/show/19254161-visual-hammer.

41. PyCharm: the python IDE for professional developers by jetbrains. *JetBrains*. URL: https://www.jetbrains.com/pycharm/.
42. Re Regular expression operations. *Python documentation*. URL: https://docs.python.org/3/library/re.html.
43. Skillsetter. Tone of Voice – як розробити щирий голос бренду. *Skillsetter*. URL: https://skillsetter.io/blog/how-to-tone-of-voice-ua.
44. Streamlit • A faster way to build and share data apps. *Streamlit • A faster way to build and share data apps*. URL: https://streamlit.io.
45. Welcome to python.org. *Python.org*. URL: https://www.python.org.

ДОДАТКИ

Додаток 1. База даних зібраних текстів. URL: https://docs.google.com/spreadsheets/d/1_fUOpWWZZp7Kjxtb9wB8lGZlLdU0D0ohb96LkxF6aLg/edit?usp=sharing

Додаток 2. Папка з метаданими до текстів (фото, дата створення, сам текст). URL: https://drive.google.com/drive/folders/1d7gMluoS-ImeufrGctOvevLfLWfjgR6N?usp=sharing

Додаток 3. Код програми і опис до нього. URL: https://drive.google.com/file/d/1z7317WOHpDZZ3AVAbJmthsCjfIPTnmny/view?usp=share_link

Код drive_utils.py містить декілька функцій для роботи з Google Drive, пов'язаної з обробкою зображень і папок.

Додаток 4. Код програми і опис до нього. URL: https://drive.google.com/file/d/1aRUIo7AxK71P3mh5yrGEOb1bBU0lb_sW/view?usp=share_link

Файл spreadsheet_utils.py містить декілька функцій для обробки даних з електронних таблиць (spreadsheet) із використанням бібліотеки pandas та gspread.

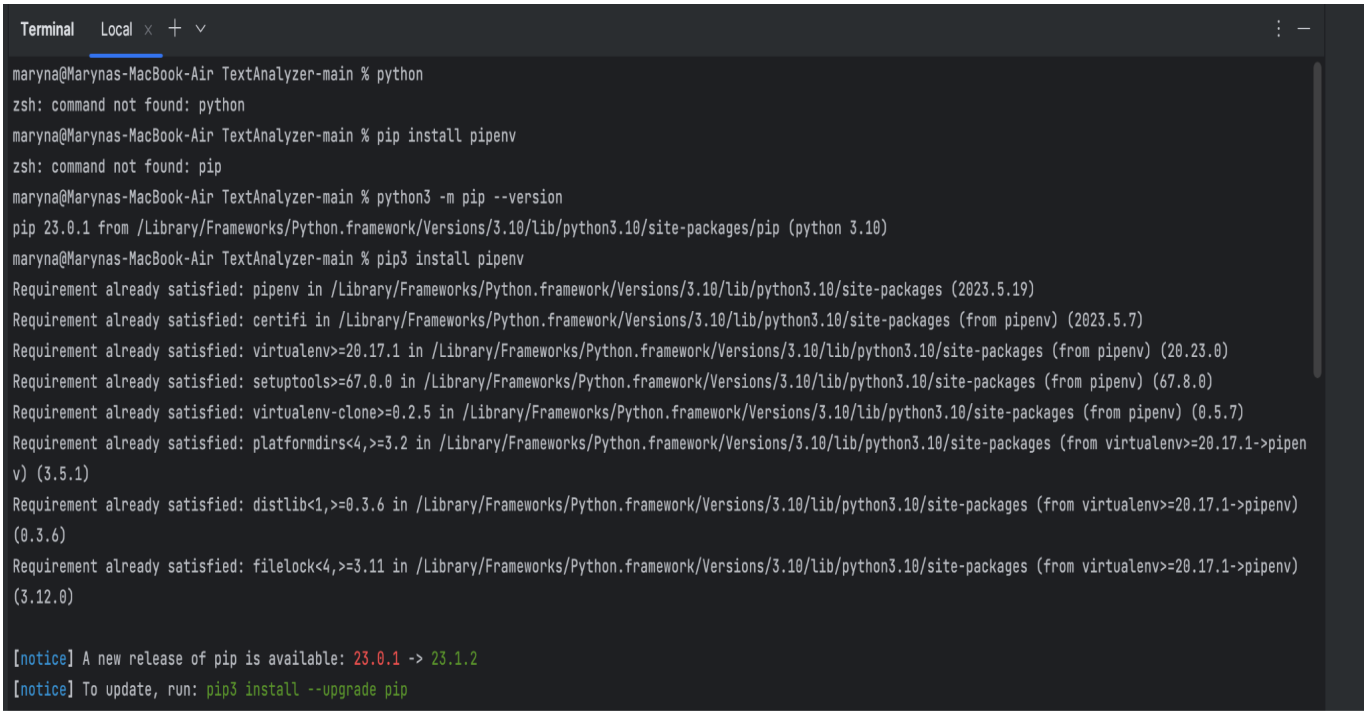
Додаток 5. Код програми і опис до нього. URL: https://drive.google.com/file/d/1zwpC0KmqiFaeUaL6lb1Fv4LPJwk3tW63/view?usp=share_link

Файл програми text_utils.py містить декілька утилітних функцій для обробки текстових даних.

Додаток 6. Код програми і опис до нього. URL: https://drive.google.com/file/d/17yUzLqdUBE72PoTaEtUOFr6zo56KuXpF/view?usp=share_link

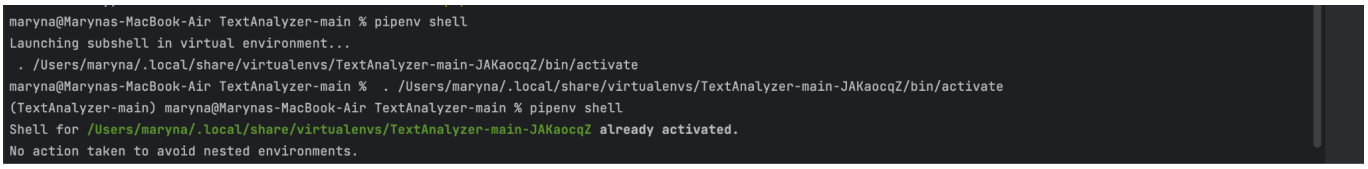
Програмний файл main.py – основний код у роботі. Цей файл містить головний код програми, який керує веб-додатком на основі бібліотеки Streamlit. Додаток призначений для обробки даних з електронних таблиць Google та відповідного зображення з Google Drive.

Додаток 7. Встановлення pipenv:



Додаток 8. Встановлення всіх залежностей для проекту за допомогою pipenv install:

Додаток 9. Активація середовища за допомогою pipenv shell:



Додаток 10. Встановлення мовної моделі для роботи з українською мовою за допомогою python -m spacy download uk_core_news_sm:

```
(TextAnalyzer-main) maryna@Marynas-MacBook-Air TextAnalyzer-main % python -m spacy download uk_core_news_sm
Collecting uk-core-news-sm==3.5.0
  Downloading https://github.com/explosion/spacy-models/releases/download/uk_core_news_sm-3.5.0/uk_core_news_sm-3.5.0-py3-none-any.whl (14.9 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 14.9/14.9 MB 11.1 MB/s eta 0:00:00
Requirement already satisfied: spacy<3.6.0,>=3.5.0 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from uk-core-news-sm==3.5.0) (3.5.3)
Requirement already satisfied: pymorphy3>=1.0.0 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from uk-core-news-sm==3.5.0) (1.2.0)
Requirement already satisfied: pymorphy3-dicts-uk in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from uk-core-news-sm==3.5.0) (2.4.1.1.1663094765)
Requirement already satisfied: dawg-python>=0.7.1 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from pymorphy3>=1.0.0->uk-core-news-sm==3.5.0) (0.7.2)
Requirement already satisfied: docopt>=0.6 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from pymorphy3>=1.0.0->uk-core-news-sm==3.5.0) (0.6.2)
Requirement already satisfied: pymorphy3-dicts-ru in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from pymorphy3>=1.0.0->uk-core-news-sm==3.5.0) (2.4.417150.4580142)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (1.0.4)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (1.0.9)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (2.0.7)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (3.0.8)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (8.1.10)
```

```
Terminal  Local
0->uk-core-news-sm==3.5.0) (1.24.3)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (2.31.0)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (1.10.7)
Requirement already satisfied: jinja2 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (3.1.2)
Requirement already satisfied: setuptools in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (67.8.0)
Requirement already satisfied: packaging>=20.0 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (23.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (3.3.0)
Requirement already satisfied: typing-extensions>=4.2.0 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4->spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (4.6.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (3.1.0)
Requirement already satisfied: idna<4,>=2.5 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (1.26.15)
Requirement already satisfied: certifi>=2017.4.17 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (2023.5.7)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from thinc<8.2.0,>=8.1.8->spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (0.7.9)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from thinc<8.2.0,>=8.1.8->spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (0.0.4)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from typer<0.8.0,>=0.3.0->spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (8.1.3)
Requirement already satisfied: MarkupSafe>=2.0 in /Users/maryna/.local/share/virtualenvs/TextAnalyzer-main-JAKaocqZ/lib/python3.10/site-packages (from jinja2->spacy<3.6.0,>=3.5.0->uk-core-news-sm==3.5.0) (2.1.2)
✔ Download and installation successful
You can now load the package via spacy.load('uk_core_news_sm')
(TextAnalyzer-main) maryna@Marynas-MacBook-Air TextAnalyzer-main %
```

Додаток 11. Запуск програми за допомогою streamlit run main.py:

```
(TextAnalyzer-main) maryna@Marynas-MacBook-Air TextAnalyzer-main % streamlit run main.py

  You can now view your Streamlit app in your browser.

  Local URL: http://localhost:8501
  Network URL: http://192.168.0.103:8501

  For better performance, install the Watchdog module:

  $ xcode-select --install
  $ pip install watchdog
```

Додаток 12. Код програми завантаження текстів і опис до нього. URL: https://drive.google.com/file/d/1ZHT9DdWQLqwp8dQN5eT5qQUw6vZyZpbW/view?usp=share_link