

Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка
Навчально-науковий інститут філології
кафедра української мови та прикладної лінгвістики

**ДІАЛОГОВА СИСТЕМА ДЛЯ ВИВЧЕННЯ ПАРЕМІЙ УКРАЇНСЬКОЇ МОВИ НА
ОСНОВІ TELEGRAM-БОТА**

Кваліфікаційна робота бакалавра
студента 4 курсу
освітньої програми
*«Прикладна (комп'ютерна) лінгвістика
та англійська мова»*,
спеціальності – 035.10 Філологія (прикладна
лінгвістика)
галузі знань – 03 гуманітарні науки
Назарія КУРПІТИ
Науковий керівник:
к.філол.н., доц. Оксана ЗУБАНЬ

«Допущено до захисту»
Протокол засідання
кафедри української мови та прикладної лінгвістики
протокол № 15 від «6» 06 2024 року
завідувач кафедри _____ (підпис)
к.філол.н., доц. Сергій РІЗНИК

КИЇВ – 2024

Анотація. Дослідження присвячене розробці Telegram-бота для вивчення українських паремій. Актуальність роботи зумовлена ризиком скорочення використання українцями прислів'їв та приказок у повсякденному мовленні, а також необхідністю створення сучасних інструментів для їх популяризації та глибшого розуміння. Об'єктом дослідження є українські прислів'я та приказки, предметом — комп'ютерна лексикографічна модель бази даних українських паремій та структура Telegram-бота. Метою роботи є створення Telegram-бота, що допомагатиме користувачам правильно добирати та вивчати українські паремії з поясненнями їх значень. Для досягнення мети було поставлено такі завдання: дослідження методологічних засад пареміології та пареміографії; проведення соціолінгвістичного експерименту для з'ясування використання паремій у повсякденному мовленні українців; проведення експертизи ефективних технологій для розроблення програмного коду проєкту; екстракція прислів'їв та приказок із вебсайту джерельної бази; редагування та класифікація отриманих паремій; створення та наповнення електронної бази даних; лінгвістичний аналіз бази даних; розробка Telegram-бота з функціоналами пошуку та навчання. Кваліфікаційна робота складається з трьох розділів. У першому розділі висвітлено теорію пареміології та пареміографії, методи класифікації паремій, досягнення вітчизняних науковців, проведення експерименту й аналіз його результатів. Другий розділ присвячений створенню бази даних українських паремій, включаючи експертну оцінку джерел, вебскрейпінг, редагування та класифікацію пареміологічних одиниць, розроблення структури бази даних та її лінгвістичний аналіз. У третьому розділі розглянуто процес вибору технологій, розробки програмного коду та демонстрацію його функціонала. Результати дослідження демонструють, що створений Telegram-бот є ефективним і зручним інструментом для вивчення українських паремій, адже об'єднує у собі інтерактивне навчання з автоматизованою перевіркою знань. Він забезпечує правильний добір і пояснення паремій як для носіїв української мови, так і іноземців.

Ключові слова: пареміологія, пареміографія, паремія, Telegram-бот, семантична класифікація, вебскрейпінг, програмування.

Abstract. The study is devoted to the development of a Telegram bot for learning Ukrainian paremias. The relevance of the work is due to the risk of a decrease in the use of proverbs and sayings by Ukrainians in everyday speech, as well as the need to create modern tools for their popularization and deeper understanding. The object of the study is Ukrainian proverbs and sayings, the subject is a computer lexicographic model of the Ukrainian paremia database and the structure of the Telegram bot. The aim of the study is to create a Telegram bot that will help users to make proper choices of Ukrainian paremias and learn them with explanations of their meanings. To achieve this goal, the following objectives were set: to study the methodological foundations of paremiology and paremiography; Conducting a sociolinguistic experiment to determine the use of paremias in everyday speech of Ukrainians; conducting an examination of effective technologies for the development of the project's program code; extraction of proverbs and sayings from the source database website; editing and classification of the received paremias; creation and filling of an electronic database; linguistic analysis of the database; development of a Telegram bot with search and learning functions. The paper consists of three chapters. The first chapter covers the theory of paremiology and paremiography, methods of paremias' classification, achievements of national scientists, conducting an experiment and analyzing its results. The second chapter is devoted to the creation of a database of Ukrainian paremias, including expert evaluation of sources, web scraping, editing and classification of paremic units, development of the database structure and its linguistic analysis. The third chapter discusses the process of choosing technologies, developing program code, and demonstrating its functionality. The results of the study demonstrate that the created Telegram bot is an effective and convenient tool for learning Ukrainian paremias, as it combines interactive learning with automated knowledge testing. It ensures the correct selection and explanation of paremias for both native Ukrainian speakers and foreigners.

Keywords: paremiology, paremiography, paremia, Telegram bot, semantic classification, web scraping, programming.

ЗМІСТ

ВСТУП.....	5
РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ УКРАЇНСЬКОЇ ПАРЕМІОЛОГІЇ ТА ПАРЕМІОГРАФІЇ.....	8
1.1. Лінгвістична інтерпретація та класифікація паремій у сучасному мовознавстві.....	8
1.2. Українська пареміографія: витоки й сучасність.....	15
1.3. Паремії у мовній практиці українців: соціолінгвістичний експеримент.....	19
Висновки до розділу 1.....	23
РОЗДІЛ 2. СТВОРЕННЯ БАЗИ ДАНИХ УКРАЇНСЬКИХ ПАРЕМІЙ.....	25
2.1. Експертна оцінка електронних лексикографічних джерел дослідження.....	25
2.2. Вебскрейпінг: аналіз ефективних технологій.....	28
2.3. Екстракція та редагування пареміологічних одиниць із вебсайту.....	30
2.4. Розроблення структури бази даних та імпорт даних.....	33
2.5 Лінгвістичний аналіз бази даних українських паремій.....	36
Висновки до розділу 2.....	41
РОЗДІЛ 3. РОЗРОБЛЕННЯ TELEGRAM-БОТА.....	42
3.1. Вибір Python-бібліотеки для розроблення бота.....	43
3.2. Реєстрація Telegram-бота та отримання токена.....	44
3.3. Поетапна реалізація програмного коду Telegram-бота.....	46
3.3.1. Розроблення функціонала пошуку.....	47
3.3.2. Розроблення функціонала навчання.....	49
3.4. Демонстрація роботи Telegram-бота.....	50
ВИСНОВКИ.....	58
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	61
ДОДАТКИ.....	68

ВСТУП

Паремія — прислів'я, приказка, видове позначення малих фольклорних жанрів [19]. Саме прислів'я і приказки за своєю природою є найбільш поширеними в усному мовленні. Ці пареміологічні одиниці, дослідники часто також називають «зерна народної творчості», «перлини народної творчості», «народна мудрість» тощо [4]. Вони виникали й використовувалися впродовж тисячоліть і передавалися як вираження знань, філософії та істини з покоління в покоління. Тож цілком логічним є те, що науковці зацікавилися цим мовним феноменом і почали досліджувати його з різних поглядів.

За власними спостереженнями, попри те, що прислів'я і приказки вважаються найпоширенішими видами паремій, у період сьогодення їх використання все більше і більше скорочується. Крім того, якщо запитати пересічну людину, які вона знає прислів'я і приказки, то навряд чи вона назве більше як три-чотири одиниці й навряд чи це будуть якісь особливі приклади, застосування яких є рідкісними.

Та навіть за умов, що українці знають й використовують прислів'я та приказки, чи розуміють вони їх значення? Це завдання не з простих. Діти ще у школах вивчають ці види паремій, їм навіть присвячені окремі теми на уроках, однак розуміння тлумачення потребує багато навичок: розуміння різноплановості прислів'я чи приказки, їх підтекст, сенс слів, що несуть за собою образність значення і т.ін.. Однак опанування значення паремій є лише початковим етапом їх засвоєння. Існує також проблема створення умов, за яких прислів'я та приказки переходять із пасивного словника людини до активного. Саме це питання ми й поставили собі на меті вирішити, визначаючи актуальність нашого дослідження.

Здійснивши попереднє дослідження ресурсів з пареміями на просторах інтернету, було виявлено декілька програм/сайтів, які містили в собі ці одиниці. Але майже всі вони є лишень електронною збіркою паремій, онлайн-виданням. Саме тому, **актуальність** нашої роботи зумовлена необхідністю вебзастосунків не тільки з реєстром пареміологічних одиниць, а і з їхніми тлумаченнями та зі

зручним користувацьким функціоналом для роботи і їх вивчення. Платформою, на якій ми розроблятимемо такий програмний проєкт, ми обрали Telegram.

Мета роботи: створити Telegram-бот для правильного добору та вивчення українських паремій з поясненнями їх значень. Щоб досягнути нашої мети, нам необхідно виконати такі **завдання**:

- 1) дослідити проблеми та завдання галузей пареміології та пареміографії;
- 2) дослідити види пареміологічних одиниць та пареміографічних підходів класифікації паремій;
- 3) провести соціолінгвістичний експеримент, щоб з'ясувати використання прислів'їв та приказок у повсякденному мовленні українців;
- 4) здійснити експертизу вебзастосунків для автоматичного формування реєстру паремій;
- 5) відредагувати та прокласифікувати отримані прислів'я та приказки;
- 6) створити й наповнити електронну реляційну базу даних українських паремій;
- 7) провести лінгвістичний аналіз бази даних;
- 8) провести аналіз ефективних технологій для розроблення бота;
- 9) розробити Telegram-бота з функціоналами пошуку та навчання;

Об'єктом дослідження є українські паремії.

Предметом дослідження є комп'ютерна лексикографічна модель бази даних українських паремій та структура електронної діалогової системи Telegram-бота для вивчення паремій.

Методи дослідження: метод анкетування; метод семантичної класифікації; метод комп'ютерного моделювання; метод автоматичного аналізу та синтезу; метод об'єктно-орієнтованого програмування.

Інформаційна база дослідження: мови програмування Python [63] та Java [56], їхні бібліотеки: asyncio [52], logging [59], random [65], re [66], aiogram [51], sqlite3 [71], json [57], pymorphy2 [61], tokenize_uk [56], jackson [54], commons-io [53], jsoup [58].

Практичне значення цієї роботи полягає в тому, що створена діалогова система «Приказки і прислів'я: бот-помічник» [73] є ефективним і зручним інструментом для вивчення, правильного добору і розуміння українських паремій у повсякденному мовленні як носіїв української мови, так і іноземців, які виявили інтерес до вивчення української мови. Крім того, такий онлайн-застосунок покликаний популяризувати українські паремії, які відображають ідентичність українського етносу, його мудрість і філософію.

Матеріал дослідження: прислів'я та приказки з тлумаченнями, дібрані автоматично із вебсторінки «Словник прислів'їв та приказок з поясненнями» [49]. Кількісна характеристика реєстру бази даних – 230 пареміологічних одиниць.

Структура кваліфікаційної роботи:

У першому розділі оглянуто проблеми лінгвістичної ідентифікації пареміологічних одиниць у мовознавстві, проаналізовано лінгвістичну інтерпретацію понять, що входять у пареміологію, розглянуто запропоновані науковцями ідеї, щодо організації лінгвістичної матерії, проведено дослідження вітчизняних робіт з української пареміографії та проведено соціолінгвістичний експеримент для встановлення думки народу, щодо використання прислів'їв та приказок у їх мовленнєвій практиці.

У другому розділі викладено поетапне формування джерельної бази нашого дослідження, детально описано обрані ефективні технології для виконання поставлених завдань, а також процес створення та наповнення електронної реляційної бази даних, проаналізовано отриману базу даних.

У третьому розділі проведено дослідження найпопулярніших бібліотек для створення Telegram-ботів, охарактеризований процес реєстрації Telegram-бота, детально описано етапи створення програмного коду нашого проєкту та розроблення усього функціонала Telegram-бота, проаналізовано та продемонстровано роботу нашої кінцевої розробки.

Під час роботи було опрацьовано понад 70 джерел, з яких 35 англійською мовою.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ УКРАЇНСЬКОЇ ПАРЕМІОЛОГІЇ ТА ПАРЕМІОГРАФІЇ

1.1. Лінгвістична інтерпретація та класифікація паремій у сучасному мовознавстві

Історично, паремії вивчалися з фольклористичної, лінгвістичної та лексикографічної точок зору з неоднаковими методами та цілями, що призвело до появи різноманітних термінів, які іноді перетинаються, іноді доповнюють один одного. Різні спроби дати визначення цьому явищу мали різний ступінь успіху, але існують досить загальні згоди щодо основних тлумачень.

Пареміологія – філологічна дисципліна про паремії – народні вислови, виражені реченнями або короткими ланцюжками речень, якими передаються елементарна сценка чи найпростіший діалог [5]. Пареміологи розглядають питання визначення, форми, структури, стилю, змісту, функцій, значення і цінності пареміологічних одиниць. Постає важливе питання, яке час від часу хвилює багатьох науковців, а саме: чи відіграють паремії помітну роль у сучасному світі, чи у високорозвинених культурах вони перебувають на шляху до вимирання? Сучасні результати пареміологічних досліджень показують, що паремії все ще живі та ілюструють комунікативні функції, які вони виконують [37].

Пареміографія – це практична та науково-методична діяльність щодо відбору, систематизації й фіксації у збірниках і словниках прислів'їв, приказок та інших паремій за алфавітним чи тематичним принципом [12]. Пареміографи займаються збиранням і класифікацією пареміологічних одиниць. Учені ще не дійшли до одностайної думки про місце та статус паремій в ієрархічній будові мови; досі немає чітких та універсальних критеріїв для їх відбору, аналізу, класифікації й опису, через що виникають розбіжності серед дослідників, і звідси уже випливає відсутність системності у методах дослідження, різні способи тлумачення, неточні або спірні судження та висновки. [6]. Донедавна вони досліджувалися здебільшого тільки в межах фразеології. Проте на сучасному

етапі розвитку мовознавчої науки можна впевнено стверджувати, що пареміологія набула виокремлених рис. Однією з таких є синтаксична завершеність паремій. Пареміологічні одиниці відзначаються семантичною неоднорідністю: одні з них зберігають пряме номінативне значення, що впливає із суми значень складових компонентів, в інших, розвивається ще й другий семантичний план – узагальнено-метафоричне, переносне значення, яке безпосередньо не формується значенням складників, тобто різні паремії по-різному транслюють національну культуру, національний світогляд, національні пріоритети [11]. Повертаючись до основної теми, далі ми більш детально приділимо увагу самим пареміологічним одиницям.

Паремії розпочали своє існування ще за доантичних часів. Їхнє виникнення оповите загадковістю і невідомістю, а перше видове тлумачення сприймалося як висловлення. Старіші дослідження цього різновиду словесності дотримувалися уявлення про те, що їх походження лежить у душах людей, що паремії – це свого роду накопичення мудрості, що вони сформувалися з досвіду віків без цілеспрямованих зусиль окремих осіб. Сьогодні сучасні дослідники вважають, що прислів'я та приказки є найбільш "філософським" за своєю формою жанром у найдавнішій усній творчості [30]. Паремії могли перебувати в "анонімному" вжитку ще до того, як з'явилася їхня перша історично зафіксована згадка.

Антрополог Реймонд Ферт висунув такі перші загальні висновки стосовно виникнення паремій: За якихось обставин, будь-яка людина повинна була висловити почуття спільноти з того чи іншого питання, словами, які припали до душі, іншим членам групи. Тоді ця фраза передавалася по колу і врешті була прийнята як зручний спосіб вираження [35]. З такого опису можна припустити три основні алгоритми створення паремії (в нашому випадку саме прислів'я/приказки) серед прадавніх цивілізацій:

1. Конкретне викладення думки однією особою у відповідь на певні обставини;

2. Сприйняття викладеної думки спільнотою в цілому як такого, що відповідає більш загальній ситуації, і здається, що завдяки своїй особливій формі, вона дає влучне вираження їхніх думок і почуттів;
3. Можлива модифікація переносності або ж його істинного значення з плином часу, під впливом несвідомого процесу з метою збереження відповідності думки зі суспільним тенденціям [41].

Не варто забувати, що паремії продовжують виникати й що у них, очевидно, є майбутнє в сучасності. Причиною цього твердження може слугувати загальночасовий характер відтворюваних явищ або декларованих понять в пареміях [42]. Загальночасовий це такий, який не підвладний “зношенню” з часом, постійно актуальний, з поколінь у покоління, з епохи в епоху. Найчастіше він виражений теперішнім часом дієслова, що передає формальний вияв загальночасовості паремій, наприклад: *«Газда умирає, а газдиня другого шукає»*, *«Тримайся за землю-матінку, вона одна не видасть»*; *«Хто друга в біді залишає, той сам у біді пропадає»*. Використання в одиницях пареміології форми теперішнього часу означає єдність дії з суб’єктом поза конкретним часовим виміром, абстрагування від будь-якої дії в певний період, стереотипність її повторень [12].

Морфологічним індикатором загальночасовості є прислівник *завжди*: *У «суперечці з курами тарган завжди винен»*; *«На печі завжди червоне літо»*; *«Тюлень у морі не журиється — риба завжди під боком»*.

Загальночасовість у пареміях виявляється також у формі майбутнього часу дієслова, яку кваліфікують як реалізацію теперішнього постійного. Наприклад: *«Бездонної бочки водою не наповниш»*.

Категорія загальночасовості в пареміях має також формальне вираження дієсловом минулого часу: *«Як їв, то аж упрів, а як працював, то аж змерз»*; *«Взявся за гуж — не кажи, що не дуж»*; *«На годину спізнився, за рік не доженеш»*.

Як уже зазначалося, лінгвістична ідентифікація пареміологічних одиниць у зарубіжному й вітчизняному мовознавстві проводилася в аспекті фразеології.

Декілька лінгвістів (Л. Авксентьєв, Н. Амосова, В. Жуков, Д. Мальцева, В. Мокієнко, О. Молотков, С. Ожегов, В. Телія та ін.) такий підхід заперечували [12]. Підґрунтям відмежування паремій від фразеології стала їх кваліфікація як афористичних фразеологізмів. З того часу розпочинається відокремлення прислів'їв, приказок, афоризмів, крилатих висловів від фразеологічного й лексичного масивів, звуження меж фразеології [31]. У нашому розумінні, диференційними ознаками паремій для їх виокремлення є також те, що власне фразеологізми виражають поняття, а не судження; вони є семантичними й функціонально-граматичними еквівалентами слів, тоді як пареміям не притаманна така функція.

А. Крикманн [17] виокремив низку ознак, за якими певний сегмент мовленнєвого потоку може сприйматися як паремія:

- 1) відрізок мовлення не відповідає загальному контексту;
- 2) його пряме значення видається досить тривіальним;
- 3) прямий сенс відрізка мовлення здається неправильним;
- 4) відрізок мовлення у прямому витлумаченні здається безглуздим;
- 5) відрізок мовлення має виразну художню форму.

До паремій зараховують прислів'я, приказки (приповідки) та їх жанрові різновиди – вітання, побажання, прокльони, порівняння, прикмети, каламбури, тости. Також включають інші найкоротші фольклорні жанри: загадки, прикмети, повір'я, віщі сні, ворожби, задачі, головоломки, скоромовки, замовляння, небилиці, нісенітниці, казкові формули та ін. [19]. Основними елементами, на чому сходиться більшість науковців, все ж потрібно вважати прислів'я та приказки.

Прислів'я та приказки – стислі, стійкі вислови, які в образній формі відображають різноманітні явища життя [4]. Вони є узагальненням життєвого досвіду й повсякденних спостережень, що дають право формулювати погляди на різні теми буття. У своїй сумі прислів'я та приказки становлять начебто звід правил, якими людина має керуватися у повсякденному житті. Вони рідко тільки констатують якийсь факт, скоріше рекомендують чи відкидають, схвалюють або

засуджують, остерігають чи повчають, бо за ними стоїть історія поколінь [41].

Ці дві одиниці зазвичай не розмежовують, проте це не означає, що вони однакові, вони відрізняються певними структурними особливостями. Прислів'я – довершений за змістом вислів, який становить граматично й інтонаційно оформлене судження, як правило, у формі складного речення, що має двочленну структуру [19]. Приказка – це образний вислів чи мовний зворот, що висловлює незавершену думку, є елементом ширшого судження, має форму кліше [19]. Прислів'ям властиве повне вираження думки, приказка висловлює думку неповно, часто є частиною прислів'я [21].

Прислів'я

***У річку вскочить і сухий вискочить:
на все тями вистачить.***

***Багатому і під гору вода тече, а
бідному і в долині треба криницю
копати.***

***Береженого Бог береже, а козака
шабля стереже.***

Приказка

У річку вскочить і сухий вискочить.

Багатому і під гору вода тече.

Береженого Бог береже.

Але межа між цими паремійними одиницями залишається все ж недостатньо вираженою, адже інколи, прислів'я може поєднувати ознаки приказки.

Коротко витлумачимо й деякі інші жанрові різновиди паремій, які так чи інакше, грають велику роль у становленні українського паремійного фонду. Вони є і залишаються невід'ємною частиною історії, народної творчості, народної мудрості, нашого фольклору.

Загадка — компактний поетичний твір, який має метафоричне запитання в основі [19]. Її розгадка вимагає здатності порівнювати життєві явища на основі їх спорідненості чи подібності. Часто ця подібність може обмежуватись лише кількома або навіть однією ознакою. Тому для розгадування загадки потрібна не лише певна кількість знань, але й кмітливість, вдумливість та здатність бачити як конкретне, так і щось абстрактне, завуальоване [33]. Наприклад:

*Вийшла звідкись гарна дівка,
На ній стрічка-семицвітка;
А де з річки воду брала,
Там коромисло зламала. (веселка).*

Вітання — усталені вислови-кліше, що вживаються при зустрічі знайомої особи, або ні, з метою привітатися [19]. Відокремлюють також групу святкових вітань, вживання яких обмежується тільки в конкретний період, на свята. Наприклад: вітання: *«Дай боже!», «Доброго дня!», «Як ся маєте?», «Хай вам Бог помагає!», «Ласкаво просимо!»*; святкові вітання: *«Христос ся рождає! – Славимо його!», «Христос Воскрес! – Воістину воскрес!»*;

Побажання — словесні зичення будь-яких хороших, позитивних благ чи то моральних, чи матеріальних [19]. Існує чимало формул побажань, які прив'язані до якогось конкретного явища, події. Наприклад, тости — застільні побажання, які можуть адресуватись конкретній особі, завдяки якій відбувається нагода частування, або усім гостям. Сюди ж належать побажання-подяки, та прощання, які часто формулюються як побажання. Наприклад:

побажання: *«Рости великий!»*, *«Хай йому земля буде пухом»*, *«Щоб у вас і у нас все було гаразд!»*; тости: *«За синєє небо, за жовті поля, за світ у якому нема москаля!»*, *«Вип'єм на дорогу за нашу Перемогу!»*, *«Всі кому налили – п'ють за Збройні сили!»*; прощання: *«Щасливої дороги!»*, *«Всього найкращого!»*, *«Гарного вечора/дня!»*

Прикмети — це вислови, в яких ознаки життєвих, природних явищ пов'язуються з іншими та несуть передбачення чогось, часто якоїсь події [19]. Наприклад: *«Ластівки низько літають – дощ обіцяють»*, *«Не заглядай у піч, бо вуса не ростимуть»*

Прокльони — різкі форми вираження незадоволення, осуду, гніву, обурення до людини, для міфічного заподіяння їй зла [19]. Наприклад: *«А щоб тебе качка копнула!»*, *«Шляк би тя трафив!»*, *«Щоб воно тобі боком вилізло!»*.

Актуальною й по сьогоднішній день постає проблема класифікації паремій. Організація лінгвістичної матерії була практичним питанням впродовж століть.

Головна причина цього, звісно, відсутність якогось єдиного, чіткого алгоритму для встановлення лексичних, семантичних, функціональних т. ін. ознак, щоб упорядковувати за ними закономірності та відносити пареміологічні одиниці до спільних або ж відмінних класів [39]. Проте науковці багато досліджували це питання, й деякі з них все ж висунули свої ідеї, принципи на цю проблему.

Однією з таких робіт і найвдалішою, на наш погляд, була наукова праця Ю. Прадіда «Робоча схема основних знань людини про навколишній світ» [24]. Вона не стосувалася конкретно пареміології, а визначала теоретичні та практичні засади вирішення окремих фразеологічних мікро- та макросистем, але однозначно мала великий вплив на праці пареміографії. Про це свідчать самі назви деяких пізніше виданих пареміологічних робіт «Прислів'я та приказки: Природа і народ» [26]; «Прислів'я та приказки: міжособистісні взаємини» [25].

У цій синоптичній схемі автор ставить найвищим рівнем «Всесвіт», який об'єднує два основні архіполя: «Жива природа» та «Нежива природа». У перший входять семантичні поля «Флора», «Фауна», «Людина». У другий – «Предмет» і «Явище» [7]. А далі, ієрархічна структура класифікації, запропонована науковцем, має такий вигляд: синонімічний ряд → семантична група → семантичне поле → тематична група → тематичне поле → ідеографічна група → ідеографічне поле → архіполе [1].

Ю. Прадід недаремно назвав свою схему "робочою", оскільки вона може бути допрацьованою, уточненою. З цієї нагоди наступники з цього питання висували свої ідеї. Ми обрали схему Ж. Колоїз, позаяк вона адаптувала структуру автора до пареміологічної системи. Відтак структура має такий вигляд: структурно-семантична модель → синонімічний ряд → семантична підгрупа → семантична група → семантичне поле → тематичне поле. Наприклад, поле «Людина» охоплює такі основні семантичні поля й семантичні групи:

1. «Зовнішній вигляд» (семантичні групи: а) «Врода», б) «Зріст», в) «Одяг»);
2. «Вікові, фізіологічні та фізичні особливості» (семантичні групи: а) «Вік», б) «Фізіологічний стан», в) «Фізичний стан»);

3. «Внутрішній світ» (семантичні групи: а) «Розумові здібності», б) «Риси характеру», в) «Емоційний стан»);
4. «Людські взаємини» (семантичні групи: а) «Родинні стосунки» (семантичні підгрупи: I «взаємини між батьками і дітьми», II «дружиною та чоловіком», III «братами і сестрами», IV «невісткою і свекрухою» тощо), б) «Товариські стосунки») [12];

У нашій роботі, синонімом до паремії вважатимемо таку одиницю, яка відмінна за матеріальною формою вираження, але подібна за значенням. Наприклад: «*Який рід, такий і плід*» вважатимемо синонімом паремії «*Який батько, такі й діти*» через їх схоже, однакове тлумачення.

Все ж, питання класифікації викликає чимало дискусій й потребує подальших досліджень. А ми поки зупинимось на цих ідеях, і застосуємо їх у практичній частині нашої наукової праці.

1.2. Українська пареміографія: витoki й сучасність

Вітчизняні здобутки в галузі пареміографії безперечні. Вже існує чимало пареміографічних праць, кількість яких постійно зростає. Чи не найпершою працею, пов'язаною з галуззю пареміології, деякі дослідники вважають «Збірник Святослава» (1073-1076 рр.). Вона була написана на замовлення київського князя і є визначною пам'яткою української писемності, мови, історії. Цей твір є найдавнішою з датованих давньоруських книжок для читання [22]. У ній можна знайти так звані афористичні вислови. Одне з видань містить «Стословець патріарха Геннадія», що включає сто висловів, які стосуються правил християнської віри та морального життя. Важливим є той факт, що автор вже тоді спробував згрупувати афористичні вислови за певними темами: Повчання багатим, Про заклик сильних т. ін. [28].

Появу ранніх робіт з пареміології традиційно пов'язують приблизно з кінцем XVIII– початку XIX століттям, коли основним об'єктом досліджень стали прислів'я і приказки. Одним з найдавніших зразків є рукописна збірка «Приповісті посполиті» Климентія Зіновієва [9], що була зарахована до найкращих надбань

українського письменства. У рукописі зібрано понад півтори тисячі паремійних одиниць, таких як прислів'я, приказки й народні сентенції. Вони були розташовані за алфавітом і не мали перекладу або витлумачення їх змісту.

Значною подією у становленні вітчизняної пареміографії варто вважати «Словник української мови» П. Білецького-Носенка [2]. Цей словник, попри те, що він не є спеціальною пареміографічною працею, мав значний вплив на подальші дослідження в цій науковій галузі, адже у ньому паремії виступають яскравими ілюстраціями в досить багатьох словникових статтях. Крім того, на противагу попереднім працям, він подає наголос для більшості слів в українських пареміях. А ще словник містить тлумачення паремій, що також є важливим для розуміння їх змісту.

Вагомою працею XIX ст. був доробок М. Номиса «Українські приказки, прислів'я и таке инше» [20]. Займаючись пареміографією автор «спирався на великий колектив збирачів прислів'їв і приказок. Сам був активним збирачем фольклору, знав безліч прислів'їв та приказок, пересипав ними свої художні твори, спогади про дитинство і юнацькі роки» [20, 13-14]. У збірці налічується близько 15 000 одиниць, зібраних із різних українських територій і різними людьми. Тому перед укладачем цього «скарбу» постало складне питання класифікації матеріалу, оскільки одержані друковані й рукописні матеріали мали різні системи класифікації або надсилались невпорядкованими. Упорядник зупиняється на тематичному принципі організації матеріалу. Кожна паремія отримує порядковий номер і також містить паспортизацію. Вона може включати скорочену назву території, джерела або прізвище особи, яка записала представлену конструкцію. Всупереч клопіткій роботі Номиса, дослідники дещо розкритикували її за складний для користування спосіб викладення матеріалу (словникових статей) та ту ж тематизацію. Проте вплив цієї праці був колосальним.

На початку XX ст. світ зустрів потужну працю, яка була результатом дослідницької та збирацької діяльності І. Франка, збірник «Галицько-руські народні приповідки» [32]. Цю оригінальну працю, видану шістьма книжками, науковці називають узагальненням усіх доробок української пареміології та

пареміографії того часу. Вона охоплює понад 30 тис. “народного мудрослів’я”. Під час укладання збірки І. Франко дотримувався кількох важливих наукових принципів [13]:

- Застосовував чітку паспортизацію, що означало вказівку на походження, географічне поширення та фіксацію інформації про тих, хто записував паремійні одиниці (якщо не зазначено місцевості);
- Розміщення тем (за основу бралось слово, спільне для багатьох одиниць) було в алфавітному порядку, в яких групувалися приповідки;
- Тлумачення значень паремій не обмежувалося лише семантикою й окресленням ситуації;
- Приписував зразки іншомовних варіантів та зазначав їхню національну специфіку.

На думку Т. Космеди за понад сто років, що минуло з моменту видання «Галицько-руських народних приповідок», історія вітчизняної пареміографії не принесла нічого кращого або оригінальнішого [14]. Ця праця залишається неперевершеним шедевром у галузі збору народних приповідок в Україні.

У 70-90 роки ХХ ст. дослідники відзначилися не лише виданням перероблених і доповнених збірників українських прислів’їв і приказок, як, наприклад: «Народ скаже – як зав’яже (українські народні прислів’я, приказки, загадки)» [33], але й перекладами з інших (слов’янських і неслов’янських) мов: «Грузинські прислів’я та приказки» [8], «Польські прислів’я та приказки» [23], «Іспанські прислів’я та приказки» [10] і багато інших.

У 10-ті роки уже нашого століття в українську пареміографію додалися ще декілька нових досліджень. Серед них яскраво вирізняється праця «Комунікативний кодекс українців у пареміях» [15]. Цей словник є першим в Україні тлумачним словником паремій, що належить до словників активного типу. Розрізняють взагалі два типи словників: традиційного (пасивного) типу – призначений для пояснення, тлумачення, надання дефініцій для розуміння відповідних мовних одиниць; та активного типу – його головна функція полягає в сприянні синтезу мовлення, допомозі у моделюванні мовлення та побудові

комунікації відповідно до конкретних мовленнєвих жанрів, подій та ситуацій [16]. Крім цього автори визначають свою працю як паремійний словник, за характером репрезентованої мовної (реєстрової) одиниці, а за видовим критерієм – комунікативний словник, адже його основною метою є підвищення рівня комунікативної компетенції сучасних українців.

Кожна словникова стаття репрезентує такі структурні компоненти, як: 1) реєстрова одиниця (РО); 2) покликання на джерело, у якому зафіксована РО; 3) кваліфікація комунікативного статусу РО, її функцій (стратегія, тактика, правило, тенденція, закономірність тощо); 4) складники етнокультурної конотації та мовний коментар; 5) потенційні рекомендації щодо використання РО у мовленні. Наприклад:

Добре то кажуть: свій своєму поневолі брат [Ф, III: 89]. Репрезентовано сутність стришневого принципу дії закону дзеркального відображення спілкування, коли партнери підтримують стиль спілкування, зокрема дружній, відкритий, партнерський. Смісл паремії пов'язаний із давніми уявленнями українців, що закріплені поняттями, які базуються на дихотомії “свій” – “чужий”. Паремія впливає на формування відповідного стилю взаємостосунків і сприймається як неписаний закон – свій своєму найперше має допомагати (тлумачення І. Франка), який схвалюється народом і сприяє зміцненню родинних, соціальних і етнічних зв'язків. Цю приказку можна використовувати як повчання, настанову, аргумент, коментар або маркер відповідної комунікативної ситуації.

Автори посібника з української пареміології відзначають «скрупкульозність як у доборі, так і у витлумаченні фактичного матеріалу, чіткість у виробленні методичних принципів аналізу й послідовне їх дотримання» [12, ст. 261].

«Комунікативний кодекс українців у пареміях: тлумачний словник нового типу» є новим пареміографічним зразком, що відзначається ґрунтовним аналізом реєстрових одиниць, зокрема й з лінгвістичних позицій, на відміну від багатьох попередніх збірників, які обмежувалися звичайним (чи то алфавітним, чи то тематичним) реєстром. Поява таких неординарних праць свідчить про інтенсивну розбудову української лексикографії та пареміографії, що нині зосереджені на

масовому створенні словників нового типу, універсалізації та уніфікації традиційних лексикографічних доробків. Проте, попри наявний словникарський арсенал та накопичений досвід, низка теоретичних і практичних завдань потребує прискіпливішої уваги науковців, більш ретельного їх вивчення, неабияких зусиль і творчого підходу.

1.3. Паремії у мовній практиці українців: соціолінгвістичний експеримент

Пареміологічні експерименти, які досі використовуються, не відповідають основним методологічним принципам тестування, таким як стандартизація, об'єктивність, порівнянність, надійність і валідність через низьку соціолінгвістичну репрезентативність, що супроводжується довільним складанням списків досліджуваних прислів'їв чи приказок без достатньо чітко визначених критеріїв відбору [34]. Наше дослідження не виняток. Ми проводимо соціологічне опитування для того, щоб визначити, чи справді актуально на сьогодні збереження та вивчення паремій для українців.

У праці про фразеологічний мінімум з української мови для слов'яномовних студентів, Г. Кузь провела експеримент, де було запропоновано пройти анкетне опитування за 45 пареміями студентам-полякам, що навчались на спеціальності "Україністика". «Відповіді на запитання анкети дали можливість уточнити цей список, виявивши ті паремії, які студенти найбільше А) знають/розуміють; Б) використовують» [18, 101]. Ось невеличкий перелік найбільш відомих/уживаних: «Яблуко від яблуні недалеко падає»; «Робота не вовк – у ліс не втече»; «Чим хата багата, тим і рада»; «Всюди добре, а дома найкраще». Однак різниця між показниками того що вони знають і що використовують, у деяких пареміях велика. Про що це може свідчити? Така закономірність може призвести до поступового "вимирання" вживання прислів'їв і приказок. Проте, спираючись на загальновідомий факт, що мова – це в першу чергу засіб мислення, пізнання всесвіту, найважливіший інструмент спілкування, ймовірність такої події дуже низька, але не дорівнює нулю.

На перший погляд, зараз в Україні зростає тенденція до збереження та використання всього українського. Люди зацікавлені до повернення традицій, шанування того, що колись було зроблене чи придумане нашими предками. Самі ж паремії є важливим елементом культурної спадщини, що відображають мудрість та світогляд народу. Тому наш експеримент може якоюсь мірою підтвердити доречність та необхідність реалізації нашого проєкту.

Соціолінгвістичний експеримент

Проблема: недостатнє вивчення ставлення українського населення до прислів'їв та приказок, а також їхнього місця у сучасній мовній практиці.

Об'єкт дослідження: прислів'я та приказки.

Предмет дослідження: думка українців про використання прислів'їв та приказок у повсякденному мовленні, ставлення до їхнього збереження та популяризації.

Метод дослідження: анкетне онлайн-опитування (див. Додаток Б).

Мета: аналіз ставлення українського населення до прислів'їв та приказок, визначення їхнього використання у повсякденному мовленні, а також бажання використовувати їх частіше за наявності зручних ресурсів з поясненнями.

Вибірка: українці різного віку. Загальна кількість опитаних респондентів – 65 осіб.

Дослідницькі питання:

1. Чи використовують респонденти прислів'я/приказки у повсякденному мовленні.
2. Чи розуміють респонденти значення паремій.
3. Визначення думки респондентів щодо сучасного стану вживання прислів'їв/приказок.
4. Чи бажають респонденти частіше використовувати паремії.
5. Чи вживали б респонденти частіше паремії, якби для цього був зручний ресурс з їх тлумаченнями.

Гіпотези:

1. Люди рідко вживають прислів'я та приказки у повсякденному мовленні.

2. Люди бажають використовувати прислів'я та приказки більше.
3. Люди користувалися б зручним ресурсом, який пояснює значення паремій.

Результати соціолінгвістичного експерименту:

Відповідно до результатів опитування [див. Додаток А], гіпотеза про те, що люди рідко вживають прислів'я та приказки у повсякденному мовленні практично спростувалася. Більша частина респондентів (81,5%) відповіли, що використовують ці види паремій у мовній практиці. Такі результати свідчать про те, що на сьогодні їх використання не є на низькому рівні. Однак на питання – «Чи вважають респонденти, що прислів'я/приказки зараз на межі "вимирання" з вжитку?», 63 % відповіли, що так.

В анкеті опитування учасники відповідали на питання щодо знання і розуміння значення конкретних прислів'їв і приказок. Ми підібрали такі одиниці: *«Або пан, або пропав»*, *«Один за всіх, всі за одного»*, *«Не вір словам, а вір ділам»*, *«В доброму хазяїна і соломинка не пропаде»*, *«Вівсяна каша хвалилася, начебто з коров'ячим маслом народилася»*, *«Одна жарина і в печі гасне, а дві і в полі горять»*, *«Хто спить весною, мерзне зимою»*, *«Утікав від вовка, та впав на ведмедя»*. До перших трьох більш як 90 % респондентів відповіли, що знають і розуміють значення. Усі інші, близько 50 %, не знають їх, але здогадуються що вони означають. Найменш відомим і зрозумілим прислів'ям/приказкою було *«Одна жарина і в печі гасне, а дві і в полі горять»*, 33,8 % відповіли, що не знають його.

Отже, як бачимо, 5 із 8 одиниць є менш впізнаваними й важчими для інтерпретування їх пояснення. Звідси випливає, що деякі паремії за своєю структурою транслюють своє пряме номінативне значення, що випливає із суми значень складників, а інші швидше узагальнено-метафоричне, переносне. Наприклад, над прислів'ям *«Не вір словам, а вір ділам»* не потрібно й розмірковувати, що воно означає, а у *«Вівсяна каша хвалилася, начебто з коров'ячим маслом народилася»* не так просто зрозуміти, про що йдеться.

Наступна пара гіпотез частково підтвердилась. 80 % опитаних хотіли б використовувати ці види паремій частіше. У кінці опитування ми запитали

респондентів, чи вживали б вони частіше прислів'я/приказки, якби для цього був зручний ресурс з їх тлумаченнями (наприклад застосунок, телеграм-бот), 72% респондентів висловили своє погодження.

У цілому опитування показало, що прислів'я та приказки залишаються важливим елементом української мовної культури, але існує занепокоєння щодо припинення їх використання. Багато респондентів бажають вживати ці мовні вирази частіше і підтримують ідею створення зручних ресурсів для пояснення їх значення. Звідси ми робимо висновок про практичну значущість нашого проекту.

Висновки до розділу 1

Встановлення походження і віку пареміологічних мовних засобів з усної народної творчості залишається особливо складною проблемою, адже пареміологія досить нова наука. Її терміносистема ще досі не є усталеною, кожен автор вносить своє бачення того чи іншого поняття. Ми прийняли робочі визначення запропоновані М. Лановик, які розкривають такі поняття: паремія, прислів'я, приказка, загадка, вітання, побажання, прикмети та прокльони. У подальшій роботі нашого дослідження ми використовуватимемо лише терміни прислів'їв та приказок, позаяк вони складатимуть реєстр нашої бази даних.

Паремії стали окремими лінгвістичними матеріями, через які у пареміологів виникають труднощі, щоб утворити універсальний підхід до впорядкування цих одиниць. У роботі ми проаналізували робочу схему Ю. Прадіда, яка доволі влучно описує пареміологічні мікро- та макросистеми та інтерпретовану схему Ж. Колоїз. Наведені приклади класифікацій демонструють багатогранність паремійного фонду та різноманітність підходів до їх систематизації. Для практичної частини нашого дослідження ми залишили лише два складники ієрархічної структури класифікації паремій: синонімічний ряд → семантичне поле.

Вітчизняні досягнення українських науковців у галузі пареміографії вирізняються потужним науковим арсеналом робіт. «Збірник Святослава», «Приповісті посполиті» Климентія Зіновієва [9], «Словник української мови» П. Білецького-Носенка [2], «Українські приказки, прислів'я и таке инше» М. Номиса [20], збірник «Галицько-руські народні приповідки» І. Франка [32] і «Комунікативний кодекс українців у пареміях» Т. Космеди [15] утворили підґрунтя знань для нашого дослідження. В сучасній пареміографії окреслилися новаторські тенденції, які неабияк сприяють загальному розвитку мовознавчої науки.

Проведений соціолінгвістичний експеримент вніс нового бачення щодо використання паремій у мовній практиці українців. Анкетне опитування засвідчило про значущість та актуальність прислів'їв та приказок у сучасному українському суспільстві й показало що паремії залишаються важливим

елементом мовної культури, відображаючи національну специфіку, світогляд та цінності народу.

РОЗДІЛ 2. СТВОРЕННЯ БАЗИ ДАНИХ УКРАЇНСЬКИХ ПАРЕМІЙ

Для створення діалогової системи, яка буде виконувати функції пошуку прислів'їв та приказок й генерування вправ для їх вивчення необхідно створити базу даних паремій.

Перше завдання, що постало перед нами, це підготовка джерельної бази. Необхідно знайти й опрацювати ресурси, на основі яких збиратимуться дані – паремії. Нашою ціллю було відібрати такі джерела, які б містили в собі українські прислів'я та приказки, обов'язково їхні тлумачення та, за можливості, категоризацію.

Другий етап полягав у оцінці доступних програмних методів для автоматичного експорту даних. Нам необхідна технологія, яка дозволить найефективніше та найлегше зібрати паремії зі знайдених ресурсів. Потрібно також прийняти рішення стосовно мови програмування і бібліотек, які будуть залучені.

Третій етап передбачав вибір системи керування базами даних та створення структури даних.

2.1. Експертна оцінка електронних лексикографічних джерел дослідження

Пошук матеріалу, яким ми наповнюватимемо нашу базу даних, здійснювався беручи до уваги такі критерії відбору:

1. Ресурс повинен бути доступним в електронному форматі;
2. Наповнення джерела має містити прислів'я/приказки із їх тлумаченнями;
3. Великою перевагою буде, якщо ресурс також розподіляє паремії за семантичними полями/темами;
4. Отримання дозволу від власника джерела на використання матеріалу у некомерційних цілях.

Відбір джерел ми розпочали із паперових видань, словників, збірок, що є доступними в електронному варіанті. Було висунуто ідею – якщо праця якісно просканована чи перевидана у форматі, який підтримує його обробку, прочитання

комп'ютером, то можливо буде отримати цілісний текстовий матеріал і використати його для нашого завдання. Першим таким джерелом було досліджено словник «Українські приказки, прислів'я і таке інше», укладений М. Номисом [20]. Результати проведення оцінки ресурсу та його відповідності з нашими критеріями дозволяють зробити висновок про те, що він нам не підходить. Попри те, що наповнення джерела високе й у нього є тематичний принцип розподілу, він не задовольняє наш основний еталон у відборі про наявність тлумачення, крім того, формат файлу не дозволяє легкої та безпомилкової екстракції матеріалу. Після цього було проведено ще декілька оцінок над такими ресурсами: «Українські народні прислів'я та приказки» [3], «Народ скаже як зав'яже: Українські народні прислів'я, приказки, загадки, скоромовки» [33], «Паремійний корпус Михайла Стельмаха» [11], «Комунікативний кодекс українців у пареміях» [15]. Однак і вони не відповідали усім нашим вимогам. Крім того, було знайдено декілька робіт уже доволі сучасних дослідників, але вони представлені не у повному доступі. Наприклад: «Тематичний словник популярних українських прислів'їв та приказок з коментарями» [27]. Для отримання всіх сторінок потрібно або придбати цю роботу, або зв'язуватися з автором, просити дозволу. Тому ми вирішили відкласти цю ідею і взятися за пошуки доступних інтернет-джерел.

Серед онлайн-ресурсів, яких було чимало, таки знайшлося декілька, які, на перший погляд, відповідали нашим вимогам. Один із них – «Усна народна творчість. Прислів'я та приказки» систематизує за категоріями великий реєстр прислів'їв та приказок [50]. Користувацькі можливості сайту мають зручну навігацію за категоріями. На сайті, при натисканні на категорію, користувача перенаправляє на сторінку з переліком одиниць і має кнопки переходу на попередню і наступну за списком. Проте мовні одиниці не мають тлумачень, через що ми відклали це джерело на випадок якщо знадобиться допомога з розподіленням на теми.

Наступним ми віднайшли електронний ресурс, згаданої раніше праці І. Франка, збірника «Галицько-руські народні приповідки» [46]. Ця знахідка була дуже цінною, адже у ній, як ми знаємо, зібрано найбільший пареміологічний

матеріал України. Втім, зважаючи на те, що у роботі одиниці репрезентовані з урахуванням діалектних особливостей регіонів, з яких вони походять, а також на те, що сам ресурс розроблений у зручному і доступному форматі для користування, ми вирішили не наповнювати нею нашу базу даних.

Трапився нам також проєкт «Вікіцитати» [45], де зібрано цитати, афоризми, приказки тощо. Він побудований на платформі Вікі, що дозволяє редагувати матеріали будь-кому. На ньому розміщена сторінка, яка присвячена прислів'ям та приказкам народів світу. Серед них є й українські. Дослідивши її наповнення, було виявлено, що деякі прислів'я і приказки взяті з уже згаданих раніше робіт. Через це і через відсутність тлумачень ми відмовились від джерела.

Наступна низка ресурсів, уже крім другорядних, охоплювала і найголовнішу вимогу щодо значень пареміологічних одиниць. На сайті «Мала сторінка» є розділ «Прислів'я та приказки українців Канади, зібрані Володимиром Плав'юком» [48]. На цьому електронному ресурсі пареміологічні одиниці поділені на тематичні поля, такі як: робота, страх, совість, честь, Україна, серце, хліб, імена і т. ін.. Кожне поле містить кілька прислів'їв, приказок, що відображають життєву мудрість і досвід українців. Поруч з одиницями в дужках курсивом подано пояснення. Про авторське право адміністрація сайту надала інформацію тож ми можемо використати це джерело для подальших кроків.

Далі нами був розглянутий інтернет-портал, який є збірником онлайн-словників й енциклопедій. На сторінці «Приповідки або українсько-народня філософія» [47] навігація розроблена за клікабельними словами, які перенаправляють на сторінку з одиницями, в яких те чи інше слово є ключовим, а також є алфавітний покажчик. До кожної приповідки подано тлумачення.

Останнім джерелом, який ми дослідили, був вебсайт «Таємна Січ» [49]. Адміністрація порталу запевняє, що він містить багато захопливої та цінної інформації для відвідувачів із найрізноманітнішими інтересами. На сторінці «Словник прислів'їв та приказок з поясненнями» сайту подано перелік пареміологічних одиниць з тлумаченням. Цей ресурс відповідає усім нашим

критеріям відбору, тому він був обраний основною джерельною базою нашого дослідження.

2.2. Вебскрейпінг: аналіз ефективних технологій

Вебскрейпінг – це процедура автоматичного вилучення даних із вебсайтів за допомогою програмного забезпечення. Така технологія є дуже корисним інструментом автоматичного добору матеріалу. Вебскрейпінг дозволяє витягувати та перетворювати неструктуровані дані у структуровані, які можна зберігати та аналізувати в електронній таблиці або базі даних. До виникнення такої технології отримання інформації з вебсайтів найефективнішим і найпрактичнішим методом зазвичай був спосіб копіювання і вставки та ручного аналізу. Однак коли користувачам потрібно вилучити велику кількість даних, здійснювати це вручну непродуктивно.

Вебскрейпінг був єдиним способом отримати інформацію з інтернету до недавньої появи API (англ. Application Programming Interface), який є спеціальним інтерфейсом, призначеним для полегшення спілкування між програмами та серверами. Цей прикладний програмний інструмент, може передавати дані у структурованому вигляді. API може бути доступним для різноманітних типів даних, таких як статті у Вікіпедії, твіти у Твіттері та багато іншого. Хоча деякі вебсайти надають дозвіл на вільне користування їх API, але не всі вони є безплатними, а також можуть бути дуже обмеженими з погляду на те, скільки даних надають і які саме.

Існує багато інструментів і технологій для автоматичного вилучення даних з вебсторінок. Більшість мов програмування, таких як Python [63] або Java [56], мають спеціальні бібліотеки або розширення, які дозволяють ефективно отримувати інформацію таким шляхом. Beautiful Soup [52], Scrapy [67], і Jsoup [58] — найпоширеніші вебскрейпери. Тому ми розглянули ці популярні інструменти. Для цього послуговувалися двома методами емпіричного дослідження: спостереження та порівняння.

BeautifulSoup - це бібліотека синтаксичного аналізу (ще називають це **пáрсинг**) Python [52]. Вона проста у використанні й може переміщатися по HTML (англ. HyperText Markup Language — мова розмітки гіпертексту) і XML (англ. EXtensible Markup Language — розширювана мова розмітки) документа. Цей інструмент витягує різні дані, такі як зображення, тексти, посилання тощо. BeautifulSoup в основному працює шляхом парсингу HTML або XML файлів і генерує дерево, яке він може пройти для пошуку певних елементів. Він також містить різні функції для пошуку і фільтрації дерева.

Scrapy - це фреймворк на основі Python, призначений для вилучення даних з інтернет-ресурсів [67]. Він містить багатий набір функцій для вебсканування, вилучення та обробки даних, що робить його потужним і гнучким інструментом для вебскреперів. Однією з головних переваг Scrapy є його продуктивність. Він швидкий, ефективний і здатний витягувати велику кількість даних за короткий проміжок часу. Він також може формувати результати в декілька форматів, включаючи CSV, JSON, XML та інші.

Jsoup - це бібліотека для вебскрейпінгу мови програмування Java [56]. За допомогою Jsoup можна здійснювати навігацію та пошук у HTML-кодi вебсайту і витягувати всі необхідні дані. Вона має характеристики, подібні до характеристик BeautifulSoup, а також пропонує широкий спектр методів для роботи з DOM (Document Object Model).

Їлмаз Дикілітаз разом зі своїми колегами провели дослідження [42], в якому, для порівняння цих інструментів, взяли інформацію про 24 продукти на головній сторінці сайту електронної комерції. Цей процес було повторено 100 разів для всіх трьох засобів автоматичного вилучення даних. Таблиця 2.1. була створена шляхом взяття середнього значення з цих епох. Вони також описували характеристики комп'ютера, що використовувався в цьому експерименті.

Розмір і масштаб вебсайту, який використовувався для тестування, були критично важливими факторами для охоплення широкого спектра сценаріїв. Крім того, набір даних повинен містити різні типи та формати, такі як структуровані дані, таблиці, неструктурований текст, зображення та мультимедійний контент,

щоб перевірити можливості інструментів для аналізу та вилучення даних у різних форматах представлення. З цих причин авторами було обрано дуже великий вебсайт електронної комерції [42].

У таблиці 2.1. наведено комплексне порівняння трьох популярних бібліотек для вебскрейпінгу: Scrapy, BeautifulSoup та Jsoup. Оцінка базується на трьох ключових метриках: використання пам'яті, використання процесора та робочий час.

Library	Memory Usage	CPU Usage	Working Time
Scrapy	2400 MB	2.8	8.1 sec
BeautifulSoup	8500 MB	3.7	8.6 sec
Jsoup	2150 MB	1.5	7.3 sec

Таблиця 2.1. Порівняння продуктивності бібліотек вебскрейпінгу.

У результаті, автори зазначають, що для проєктів, де швидкість та економія ресурсів є пріоритетнішими за розширену функціональність, BeautifulSoup та Jsoup є ефективними технологіями. Обидві бібліотеки демонструють низьке використання пам'яті та процесора, що призводить до швидшого виконання операцій скрейпінгу та ефективного використання ресурсів. Scrapy, з іншого боку, пропонує комплексне рішення для складніших сценаріїв витягування даних з вебсайтів, особливо якщо потрібні розширені функції [42].

Отож, нами було прийнято рішення написання програмного коду мовою Java та використовувати бібліотеку jsoup. Цей інструмент було обрано з огляду на результати емпіричного дослідження та з урахуванням набутого досвіду його використання для автоматичного вилучення даних із вебсайту.

2.3. Екстракція та редагування пареміологічних одиниць із вебсайту

Для екстракції пареміологічних одиниць із вебсайту було створено програмний код, який представлено у додатках через покликання на GitHub репозитарій [Додаток В]. Створення проєкту мовою Java [56] відбувалося у два етапи:

- 1) написання програмного коду класу, який відповідає за отримання та структурування вебконтенту для вилучення прислів'їв/приказок та їхніх пояснень;
- 2) написання програмного коду класу, який керує основним потоком виконання програми.

Клас *ProverbsWithDescriptionParser* імпортує стандартні класи утиліт Java та Jsoup для розбору HTML. Цьому елементу програми на вхід подається *URL* адреса (*стандартизована адреса ресурсу в інтернеті*) нашого джерела і він знаходить у ньому *html* теги, приписані прислів'ям і приказкам та їх тлумаченням.

```
"Або пан, або пропав." : "Кажуть про сміливу і відважну людину, яка не сидить і не чекає сторонн  
"Або перемогти, або кістьми лягти." : "Так говорять, ідучи на битву: або перемогти, або загинути  
"Або перемогу здобути, або вдома не бути." : "Говориться як підбадьорювання перед змаганням.",  
"Багатий дивується, чим убогий живиться." : "Заможні люди зазвичай байдужі до життя бідних. Сино  
"Багато дітей у сім'ї – благодать Божа." : "Коли багато дітей – сім'я дружна та працьовита.",  
"Багатому і під гору вода тече, а бідному і в долині треба криницю копати." : "Мова йдеться про
```

Рисунок 2.1. Приклад вмісту файлу після запуску Main.java.

На рис. 2.1 подано фрагмент вмісту файлу після запуску Main.java, у якому прислів'я та приказки представлені як ключі (бузковий колір), а пояснення як їх значення (зелений).

Клас *Main* імпортує також стандартні класи утиліт, пакет jackson [54] для перетворення об'єктів Java у JSON і навпаки, та клас *IOUtils* для простіших операцій вводу-виводу, таких як запис даних у файли. Це точка входу програми, яка викликає метод *parse* з класу *ProverbsWithDescriptionParser* для отримання даних, які він повертає, а далі переносить і зберігає необхідну нам інформацію у файлі формату json за допомогою бібліотеки jackson. При його запуску ми отримуємо файл зі збереженими всіма прислів'ями та приказками, що містив у собі сайт, у зручному вигляді для подальшого маніпулювання (див. рис. 2.1), адже надалі ми будемо працювати тільки із мовою програмування Python [63].

Наступним етапом було проведено мануальне редагування отриманих даних. Причиною цього стала нестандартність структури HTML сторінки нашого джерела. Неточності полягали в тому, що деякі прислів'я та приказки попадали у рядки тлумачень інших, або ж навпаки – тлумачення у рядки до самих одиниць.

Наприклад, на рис. 2.2 видно, як у тлумаченні до ключа з прислів'ям «Не для Гриця паляниця» також збереглась паремія «Не для літа хата рубається, для зими». На цьому рисунку показано, що тлумачення однієї із паремій збереглося у ключ, відповідно це треба було виправляти вручну. Серед правок які ще доводилось робити було видалення символічних пробілів «\n».

"Не для Гриця паляниця." : "Кажуть людині, яка зазіхає на чуже.\nНе для літа хата рубається, для зими.
"Кажуть, щоб підбадьорити когось, як порада не впадати у відчай, не складати руки, а діяти." : "",

Рисунок 2.2. Приклад неточностей отриманих даних із вебсторінки.

Наступним елементом у підготовці даних був етап класифікації матеріалу. Нашим рішенням було систематизувати його за принципом, описаним у теоретичній частині. Проте ми не вдаватимемось у деталізовану структурну ієрархію класифікації, а послуговуватимемось тільки семантичними полями. Відтак, у нашій пареміологічній системі наразі налічується 14 семантичних (тематичних) полів, а саме:

- 1) «Про долю, успіх, удачу»;
- 2) «Про віру»;
- 3) «Про невдачу»;
- 4) «Про стосунки між людьми»;
- 5) «Про людські вади»;
- 6) «Про людські чесноти»;
- 7) «Про боротьбу, відвагу»;
- 8) «Про підступність»;
- 9) «Про соціальну нерівність»;
- 10) «Про цінності»;
- 11) «Про родину й родинні стосунки»;
- 12) «Поради на основі життєвого досвіду»;
- 13) «Висновки на основі життєвого досвіду»;
- 14) «Явища природи: пори року».

2.4. Розроблення структури бази даних та імпорт даних

Для розроблення бази даних було написано програмний код, поданий за покликанням у додатку В. Завдання виконано за допомогою модуля `sqlite3` [71], який містить інструментарій для взаємодії з SQLite [70], що є вбудованою і легкою для розуміння системою керування базами даних. SQLite дозволяє зберігати інформацію у локальних файлах, що робить його практичним рішенням для таких проєктів як наш, із невеликими обсягами даних.

Основним елементом, який визначав етап структуризації, було написання SQL-коду `dbscript.sql` [див. Додаток В], який визначає схеми та команди для створення таблиць для бази даних електронного словника паремій. Проїдемося по ключових рядках коду створення бази даних, реляційна схема якої подана на рисунку 2.3.

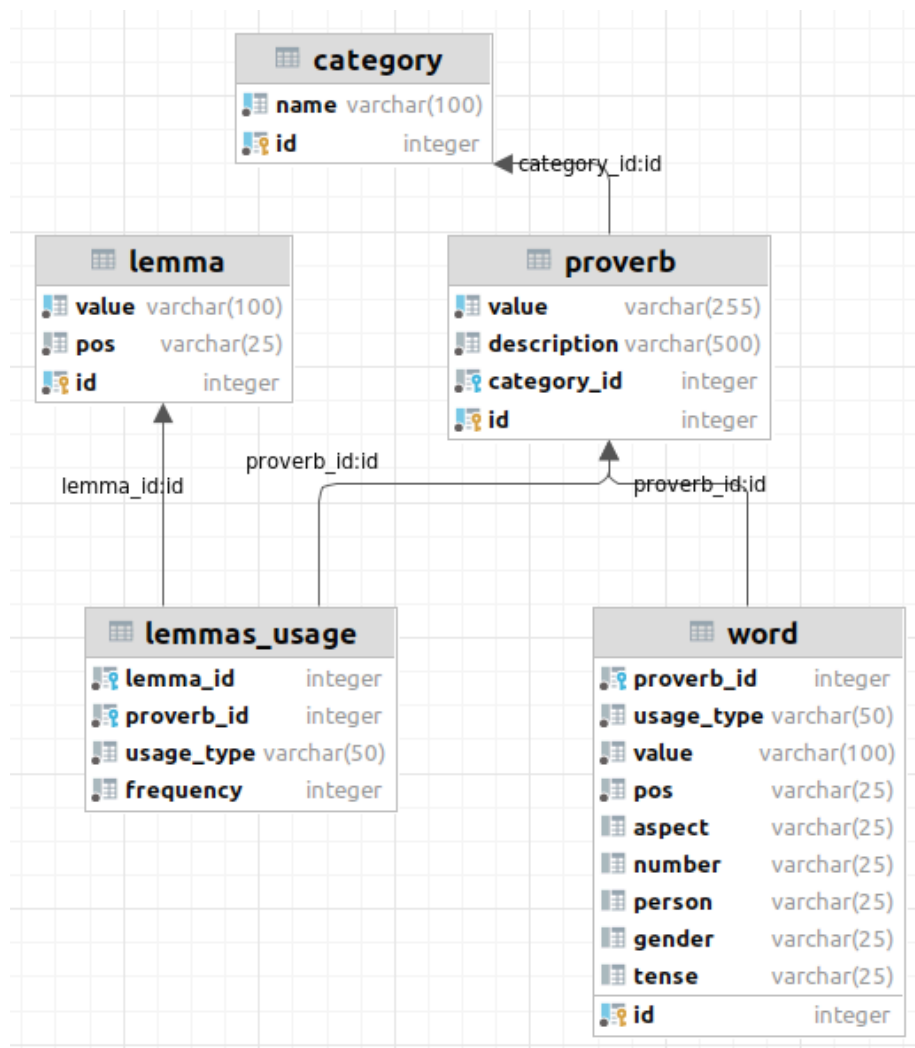


Рисунок 2.3. Загальна реляційна схема бази даних українських паремій.

1. DROP TABLE IF EXISTS: Ці команди використовуються для того, щоб переконатися, що всі створені таблиці з однаковими іменами буде видалено перед створенням нових. Це допомагає уникнути додаткового наповнення таблиці даними, що вже збережені й дозволяє створювати таблиці з нуля.
2. CREATE TABLE: Цей оператор створює самі таблиці:
 - Таблиця *category* – має дві колонки: *id* (ціле число) і *name* (назва семантичного поля).
 - Таблиця *proverb* – має чотири колонки: *id*, *value* (паремія), *description* (тлумачення) і *category_id* (номер поля, до якої одиниця належить).
 - Таблиця *lemma* – має три колонки: *id*, *value* (лема, яка представляє слово чи фразу, що є частиною паремій) і *pos* (тег частини мови, що вказує на її граматичну категорію).
 - Таблиця *lemmas_usage* – має чотири колонки: *lemma_id*, *proverb_id* (вказує на прислів'я, в якому використовується лема), *usage_type* (зазначає, де вживається лема: у самій паремії чи у її тлумаченні) і *frequency* (показує частоту вживання леми у паремії).
 - Таблиця *word* – має десять колонок: *id*, *proverb_id* (вказує на прислів'я, в якому використовується слово), *usage_type* (зазначає, де вживається слово: у самій паремії чи у її тлумаченні), *value* (слово), *pos*, *aspect* (категорія виду дієслова), *number* (категорія числа), *person* (форма особи), *gender* (категорія роду), *tense* (форма часу).

Така характеристика коду забезпечує структуру для зберігання та організації прислів'їв та приказок, лем, категорій, слів та їхніх зв'язків у базі даних.

Наступним завданням було написання коду `processing.py` [див. Додаток В] для імпорту даних: автоматичного введення прислів'їв та приказок у базу даних.

Алгоритм роботи програми для імпорту даних:

1. Імпортування усіх необхідних бібліотек, а саме: `json` [57] – який дозволяє працювати з JSON даними в Python, `rumorphy2` [61] – морфологічний аналізатор української мови, `tokenize_uk` [68] – надає можливості токенизації для української мови, `utils` – написаний нами модуль з єдиною

функцією *read()* для зчитування файлів, та клас *Database*, який забезпечує взаємодію з *SQLite*;

2. Ініціалізація бази даних, викликаючи метод *init*. Він виконує SQL-скрипт для створення необхідних таблиць і налаштування початкової схеми;
3. Створюємо *analyzer* для процесингу слів у пареміях;
4. Зчитування вмісту файлу «*proverbs-with-description-edited.json*» за допомогою функції *utils.read*, у якому зібрані прокласифіковані прислів'я/приказки та їхні значення, й зберігаємо їх у тип даних *dict(словник)*;
5. Ініціалізація основних функцій:
 - *get_lemmas(text)* – приймає текстовий параметр (паремію або її тлумачення), розбиває його на слова за допомогою функції *tokenize_uk.tokenize_words*, а потім використовує об'єкт *analyzer* для аналізу і нормалізації кожного слова. Функція повертає список кортежів, де кожен кортеж містить лему і відповідну їй частину мови;
 - *get_words(text)* – працює аналогічно до попередньої функції, тільки з додатковою умовою, що для дієслів воно також зберігає у кортеж деякі граматичні категорії;
 - *process_lemmas(text, type, proverb_id)* – обробляє лему як в пареміях, так і в їх поясненнях. Для цього вона викликає функцію *get_lemmas*, щоб отримати лему, а тоді кожен знайдений унікальний лему програма перевіряє, чи вона вже існує в базі даних (*db.select_one*), і або отримує ідентифікатор присутньої лему, або вставляє нову лему (*db.insert*);
 - *process_words(text, type, proverb_id)* – виконує таку саму послідовність дій, як і попередня функція, але з унікальними словами;
 - *process_proverb(proverb, description)* – обробляє одне прислів'я/приказку та тлумачення. Спочатку вона перевіряє, чи містить тлумачення інформацію про синоніми, шукаючи підрядок. Якщо синоніми знайдено, вона розбиває його, щоб відокремити синоніми від тлумачення. Далі рекурсивно обробляє кожен синонім, викликаючи *process_proverb* і тоді вставляє синонім як

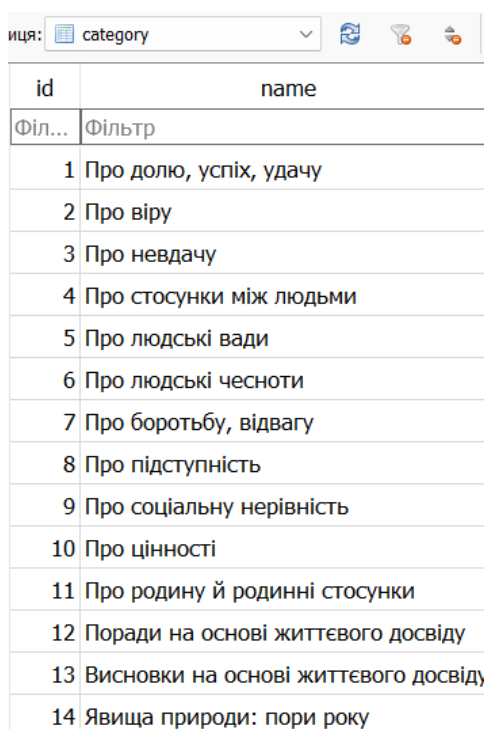
окрему паремію в таблицю *proverb* за допомогою *db.insert* (про синонімію детальніше у підрозділі 2.5);

6. Останнім елементом було створено цикл, що перебирає семантичні поля та пареміологічні одиниці у *dict(словнику)*, отриманому з JSON-файлу, і вставляє поля в таблицю *category*. Потім код проходить по прислів'ях/приказках та їх тлумаченнях в кожному тематичному полі. Цей елемент коду буде виконаний, тільки якщо скрипт запускається безпосередньо, а не імпортується як модуль.

2.5 Лінгвістичний аналіз бази даних українських паремій

Створена й наповнена база даних налічує всього 230 пареміологічних одиниць. Цю кількість повною мірою репрезентовано прислів'ями та приказками. Загальна реляційна схема бази даних зображена на рисунку 2.3. Розглянемо кожну таблицю окремо та визначимо до них їх лінгвістичні картки.

Таблиця *category*. Всього вона містить 14 об'єктів та 2 ознаки (див. рис. 2.4). Об'єктами у ній є семантичні поля, а ознаками – унікальний номер та назва поля.



id	name
Філ...	Фільтр
1	Про долю, успіх, удачу
2	Про віру
3	Про невдачу
4	Про стосунки між людьми
5	Про людські вади
6	Про людські чесноти
7	Про боротьбу, відвагу
8	Про підступність
9	Про соціальну нерівність
10	Про цінності
11	Про родину й родинні стосунки
12	Поради на основі життєвого досвіду
13	Висновки на основі життєвого досвіду
14	Явища природи: пори року

Рисунок 2.4. Наповнена об'єктами таблиця *category*.

Паремії були згруповані за 14 семантичними полями, що були визначені вручну. Така класифікація має перевагу в тому, що вона дає можливість скласти уявлення, як те чи інше явище життя знайшло своє відображення у наших перлах народного слова.

Ієрархічна структура класифікації пареміологічних одиниць має такий вигляд: синонімічний ряд → семантичне поле. Кореляція синонімії використана для доповнення реєстру джерельної бази. Ми беремо інформацію про синонімію із сайту джерела і вносимо синонім, як окрему одиницю, до такого ж семантичного поля, що і паремія, синонімом якої він є. Крім цього, використовуємо тлумачення однієї паремії до ряду синонімів. Так у нашій базі даних знаходиться 35 синонімічних рядів (див. рис. 2.5).

id	value	description
Філ...	Фільтр	Фільтр
30	Плети, плети! Я чув таких, як ти!	Кажуть про теревенів.
31	Брехати — не ціпом махати.	Кажуть про теревенів.
32	Брехач, як деркач: усе дерчить.	Кажуть про теревенів.

Рисунок 2.5. Зразок синонімічного ряду у таблиці *proverb*.

Найбільш наповненим семантичним полем є «Про людські вади»: налічує 49 пареміологічних одиниць. Такий результат може свідчити про те, що, справді, прислів'я та приказки мали й мають популярність, будучи інструментом для влучного, лаконічного висловлення думки про людей. Адже таким способом, мовець може виразити невдоволення, несхвалення чи просто недолюблювання завуальовано, не ображаючи того, кого цей вислів стосується. Наступним полем за кількістю паремій у ній є «Висновки на основі життєвого досвіду» – 43 прислів'я/приказки. Найменш наповненими є «Про віру» – 3 одиниці; та «Про долю, успіх, удачу» – 5 одиниць.

Таблиця *proverb*. Всього вона містить 230 об'єктів та 4 ознаки (див рис. 2.6). Об'єктами у ній є паремії, а ознаками – унікальний номер паремії, паремія, тлумачення паремії та номер семантичного поля, у якому паремія знаходиться.

id	value	description	category_id
Філ...	Фільтр	Фільтр	Фільтр
126	Ми з такого роду, що любимо ...	Любов до свободи, яку відчуває ...	10
127	На свою мірку не міряй.	Кожен має свій особливий погляд ...	10
128	Рідна земля і в жмені мила.	Для кожної людини найдорожче ...	10
129	На чужій сторонці радий своїй ...	Кажуть, коли хочуть висловити ...	10
130	Старий кіт, а масло любить.	Мова йде про те, що навіть старі ...	10
131	Тримайся за землю-матінку, во...	Говориться про те, що ніякі чужі ...	10
132	У кого в голові капуста ...	Люди поважають гідність самої ...	10

Рисунок 2.6. Частина наповненої таблиці *proverb*.

Таблиця *lemma*. Всього вона містить 1190 об'єктів та 3 ознаки (див рис. 2.7). Об'єктами у ній є лема, що є елементами репрезентації зібраних паремій, а ознаками – унікальний номер, лема та тег частини мови.

id	value	pos
Філ...	Фільтр	Філ...
303	брехун	NOUN
304	тісто	NOUN
305	млинець	NOUN
306	навіть	PRCL
307	ворота	NOUN
308	видно	PRED
309	гапка	NOUN

Рисунок 2.7. Частина наповненої таблиці *lemma*.

Таблиця *lemmas_usage*. Всього вона містить 3898 об'єктів та 3 ознаки (див рис. 2.8). Об'єктами у ній є місце використання лема, а ознаками – номер лема, номер паремії та місце використання лема. Під місцем використання мається на увазі, де саме зустрічається ця лема: у самій паремії чи її значенні. Таким чином, можливі всього два рівні застосування лема: *VALUE* – означає що лема зустрічається у паремії; та *DESCRIPTION* – лема зустрічається у значенні. Ця таблиця організована задля імплементування пошукових можливостей за місцем використання лем.

lemma_id	proverb_id	usage_type
Фільтр	Фільтр	Фільтр
1171	223	VALUE
1172	223	VALUE
1145	223	DESCRIPTION
1144	223	DESCRIPTION
234	224	VALUE
1173	224	VALUE
9	224	VALUE

Рисунок 2.8. Частина наповненої таблиці *lemma_usage*.

Таблиця *word*. Всього вона містить 3930 об'єктів та 10 ознак (див рис. 2.9).

Об'єктами у ній є слова, а ознаками:

- 1) унікальний номер слова;
- 2) номер паремії, що вказує на одиницю в якій використовується слово;
- 3) місце вживання слова: у паремії чи у її тлумаченні;
- 4) слово;
- 5) тег частини мови;
- 6) тег категорії виду;
- 7) тег категорії числа;
- 8) тег форми особи;
- 9) тег категорії роду;
- 10) тег форми часу.

У таблиці зберігаються усі слова, що є складниками паремійного реєстру бази даних. Однак 5 останніх ознак, визначаються і зберігаються тільки для слів, що є дієсловами. Для усіх інших частин мови, або коли у дієслова немає такої граматичної категорії, записується інформація *None*. На рисунку 2.9 продемонстровано такий приклад. Таблиця організована для реалізації функціонала вправи із завданням – доповнити прислів'я/приказку дієсловом. Ми зберігаємо й усі інші частини мови у нашій базі даних з перспективою розроблення вправ й на їх опрацювання.

id	proverb_id	usage_type	value	pos	aspect	number	person	gender	tense
Філ...	Фільтр	Фільтр	Фільтр	Філ...	Фільтр	Фільтр	Фільтр	Фільтр	Філь...
3885	227	DESCRIPTION	народна	ADJF	None	None	None	None	None
3886	227	DESCRIPTION	і	CONJ	None	None	None	None	None
3887	227	DESCRIPTION	допомагає	VERB	impf	sing	3per	None	pres
3888	228	VALUE	мерзне	VERB	impf	sing	3per	None	pres
3889	228	VALUE	зимою	NOUN	None	None	None	None	None
3890	228	VALUE	весною	NOUN	None	None	None	None	None
3891	228	VALUE	спить	VERB	impf	sing	3per	None	pres

Рисунок 2.9. Частина наповненої таблиці *word*.

Висновки до розділу 2

Для автоматичного вилучення даних ми використали технологію вебскрейпінгу. Найкращою системою для реалізації вебскрейпінгу зарекомендувала себе бібліотека jsoup [58] для мови програмування Java. За результатами оцінки ефективності вона зайняла одне з передових місць серед конкурентів.

Джерельною базою для наповнення пареміологічними одиницями реєстру нашої бази даних стала вебсторінка «Словник прислів'їв та приказок з поясненнями» [49]. Це джерело містить велику кількість (більше 300 одиниць) прислів'їв та приказок з легкими для розуміння тлумаченнями.

Через нестандартну структуру HTML вебсторінки обраного джерела, нами було здійснено мануальне редагування для усунення неточностей, які перешкоджали точній роботі програми. Вдалося самостійно класифікувати отримані дані за семантичними полями, що значно покращило їх систематизацію.

Завдяки розробленій структурі бази даних ми визначили ієрархію та взаємозв'язки між семантичними полями, прислів'ями, їхніми тлумаченнями, лемами та словами. Кількісне наповнення бази даних складає 230 пареміологічних одиниць. У результаті наша база даних є готовою для імплементування її у наступному практичному етапі та для подальших досліджень, забезпечуючи зручний доступ до великого обсягу пареміологічного матеріалу.

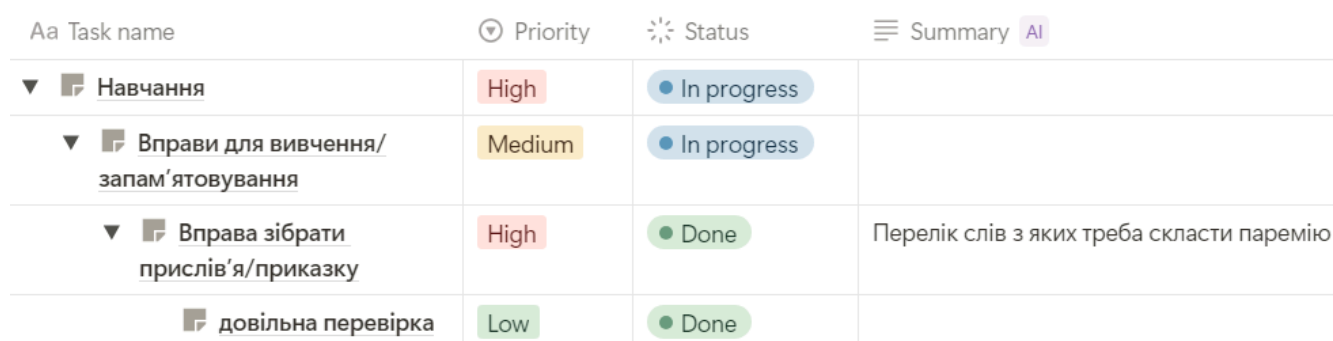
РОЗДІЛ 3. РОЗРОБЛЕННЯ TELEGRAM-БОТА

За Деніелом Джурафскі [38], діалогові системи (dialogue systems), відомі також як розмовні агенти (conversational agents), здатні взаємодіяти з користувачами через текст, усну мову або комбіновані методи з метою надання розваги або досягнення певного результату для допомоги людині. Найпростішими видами діалогових систем є чат-боти – системи які здатні вести тривалі бесіди, імітуючи неструктуровані розмови або "чати", характерні для неформального спілкування між людьми . Такі системи дозволяє будувати Telegram [72] – хмарний додаток обміну повідомленнями для мобільних і десктопних пристроїв. Називають їх Telegram-бот.

Перед початком роботи над розробкою Telegram-бота було вирішено структурувати та відстежувати прогрес виконання завдань. Для цього використовувався застосунок Notion [60], який дозволяє створювати системи управління завданнями, ведення нотаток, управління даними та проєктами.

Завдання були систематизовані у таблицю, яка містила такі колонки:

- Task name – Короткий опис завдання.
- Priority – Визначення пріоритетності завдання (*High, Medium, Low*).
- Status – Поточний статус виконання завдання (*In progress, Done, Not started*).
- Summary – Додаткові деталі або проблеми, які виникли під час виконання завдання й потребують рішення.



Aa Task name	Priority	Status	Summary AI
▼ Навчання	High	In progress	
▼ Вправи для вивчення/запам'ятовування	Medium	In progress	
▼ Вправа зібрати прислів'я/приказку	High	Done	Перелік слів з яких треба скласти паремію
▣ довільна перевірка	Low	Done	

Рисунок 3.1. Фрагмент таблиці застосунку Notion

Самі завдання також піддавались ієрархії за типом, функціоналом (див. Рис. 3.1). Така структура сприяла ефективному управлінню проєктом, значно покращила організацію роботи та дозволила відстежувати прогрес виконання.

3.1. Вибір Python-бібліотеки для розроблення бота

Telegram - це не просто популярний додаток для обміну повідомленнями, а й універсальна платформа для створення ботів. Незалежно від того, чи хочете ви створити просту діалогову систему або складного віртуального асистента зі штучним інтелектом, існує кілька фреймворків, які спрощують процес розробки. Ми ж розглянемо 4 найпопулярніші фреймворки, які розробники можуть використовувати для створення Telegram-ботів різних типів.

Першим із них був `python-telegram-bot` [64]. Ця бібліотека є однією з найпростіших для створення ботів у застосунку Telegram. Її дуже часто рекомендують на просторах інтернету завдяки своїй зручній структурі та повному охопленню Bot API. Окрім базових і ключових можливостей, модуль також включає розширені функції, такі як обробка, завантаження та вивантаження файлів. Також її плюсом є підтримка активної спільноти та велика кількість документації. Однак, `python-telegram-bot` не підтримує асинхронність, що може призвести до значного зниження продуктивності.

Далі було оцінено `Telethon` [74]. На відміну від попереднього, який призначений тільки для створення ботів, цей фреймворк корисний для повноцінних клієнтських додатків Telegram. Це означає, що окрім створення ботів, він дозволяє розробникам керувати каналами, надсилати та отримувати повідомлення, виконувати інші завдання. Проте ця бібліотека для розробників більш високого рівня. У цілому, ми одразу ж відмовились від вибору `Telethon` через відгуки користувачів про його меншу спеціалізованість для роботи з ботами.

Наступним ми розглянули `pyTelegramBotAPI` [62]. Це ще одна бібліотека, яка надає Python-обгортку для Telegram Bot API. Вона має простий для розуміння синтаксис, але має обмежений функціонал порівняно з іншими фреймворками. Її найамбітнішим плюсом є простота.

Останнім був Aiogram [51] – це асинхронний фреймворк для розробки Telegram-ботів. За структурою він є високорівневою бібліотекою, побудованою на основі python-telegram-bot та Telethon. Основна перевага цього фреймворку полягає в обробці паралельних завдань та управлінні декількома екземплярами ботів, що може значно підвищити швидкість та продуктивність одного із. А все-таки, для початківців ця бібліотека може вимагати глибших знань та умінь.

Отож, провівши аналіз Python-бібліотек для роботи з Telegram-ботами, ми обрали той інструмент, який за результатами аналітики визначений не складним для роботи над розробкою, а саме python-telegram-bot. Однак, коли ми розпочали вивчати, дізнаватися як працювати із цим фреймворком, ми зіткнулися з проблемою, що в документації до неї, мало сказано про виконання саме того функціоналу, який був задуманий нами. На просторах інтернету також не могли знайти якісного матеріалу. Тому повернулися у виборі до бібліотеки Aiogram, адже як ми виявили пізніше, саме вона дозволяє утворювати FSM (кінцевий автомат або ще називають машина станів) для створення складних користувацьких сценаріїв. Це так само було нашою необхідною умовою. До того ж документація до Aiogram є українською, бо сам розробник наш земляк. Саме це і зіграло велику роль у кінцевому виборі бібліотеки.

3.2. Реєстрація Telegram-бота та отримання токена

Перед початком роботи над самим функціоналом бота, необхідно його зареєструвати. Для цього у Telegram є спеціальний інструмент з юзернеймом @BotFather. З його допомогою можна створювати нові боти й керувати уже існуючими: він пропонує низку команд, які дозволяють змінити назву, опис, інформацію, фото, зміну налаштувань та багато іншого. Алгоритм реєстрації розпочинається з команди /newbot, після чого бот просить увести назву для свого проєкту та вигадати йому унікальний юзернейм. На рис. 3.2 подано реалізацію цього завдання для нашого проєкту.

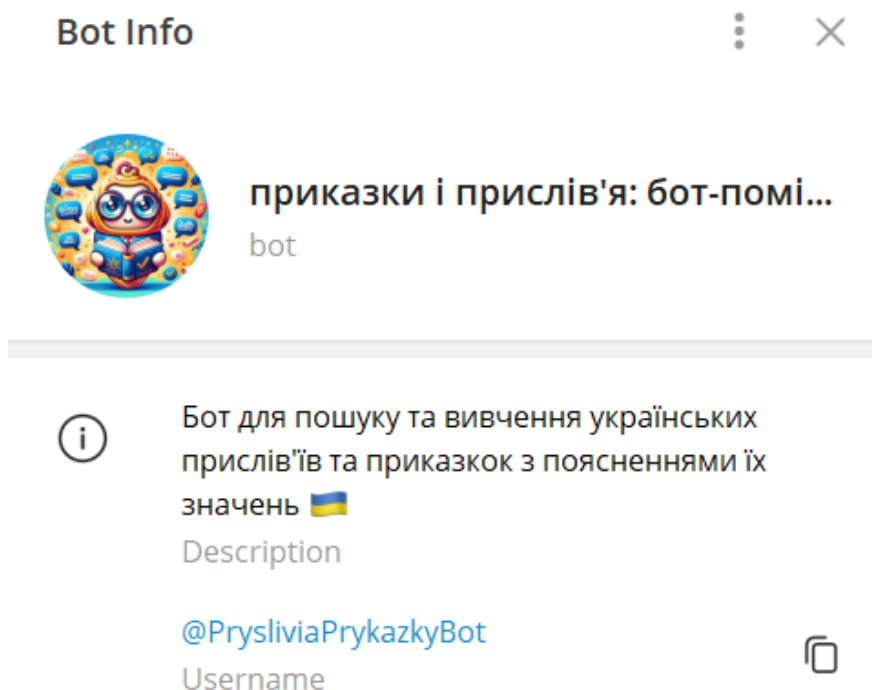


Рисунок 3.2. Назва та юзернейм нашого бота.

Після цього ми також вирішили додати опис (див. рис. 3.3) до зареєстрованого чат-бота, який коротко пояснюватиме, що вміє наш бот. Тож коли користувач знайде наш проєкт за його назвою або юзернеймом, перед початком користування йому буде видно його функціональну характеристику. Сюди ми також додали команду */start* для зручного запуску.

Виконавши попередні кроки, користувачеві приходять повідомлення з привітанням про створення нового бота, а також його токен. Токен — це рядок, який аутентифікує нашого бота в bot API й виглядає він приблизно ось так: "1234567890:AAFD39kkdpWt3ywyRZergyOLMaJhac60qc". Це повідомлення супроводжується попередженням про те, що важливо зберігати його в безпеці, адже він може бути використаний будь-ким для управління нашим проєктом. Тож обов'язково необхідно подбати про обмеження доступу до токена.

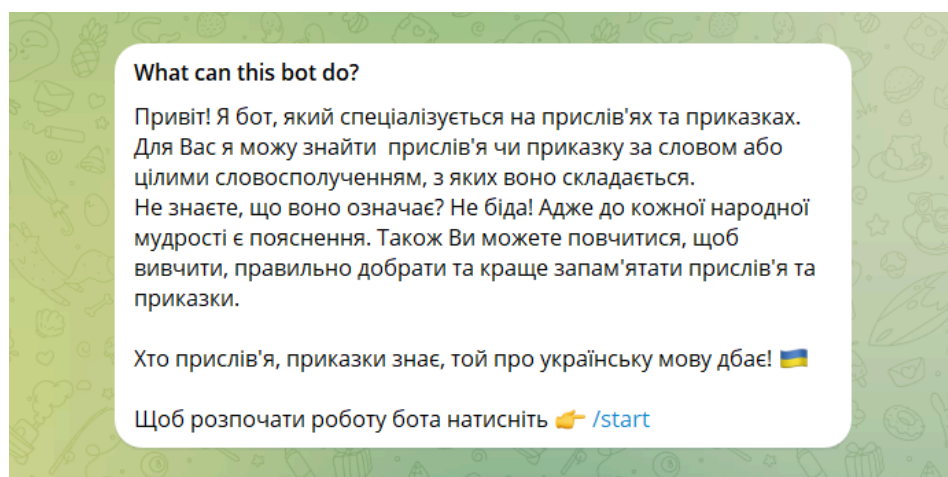


Рисунок 3.3. Опис нашого бота перед його запуском.

3.3. Поетапна реалізація програмного коду Telegram-бота

Отримавши та зберігши відредаговану частину зібраних нами пареміологічних одиниць у базу даних та зареєструвавши нашого бота, приступаємо до подальшого конструювання програмного коду [див. Додаток Г]. Для виконання усіх поставлених цілей ми імпортували необхідні модулі та бібліотеки. У програмному коді вони розділені порожніми рядками на три умовні групи: 1 група – базові бібліотеки Python; 2-га – методи, функції й т.ін. фреймворку aiogram; 3-тя – зовнішні файли та класи, створені нами. Після цього було написано частину, яка відповідала за запуск бота. Спочатку було ініціалізовано такі об'єкти:

- `db = Database()`: Екземпляр об'єкта *Database*, який використовується для взаємодії з базою даних;
- `router = Router()`: Екземпляр об'єкта *Router* – використовуються для визначення способу обробки різних типів повідомлень (наприклад, команд, текстових повідомлень);
- `bot = Bot(token=TOKEN)`: Створення екземпляра бота за допомогою токена;
- `dp = Dispatcher()`: Екземпляр диспетчера – відповідає за отримання оновлень від *Telegram* та надсилання їх відповідному обробнику.

Далі реалізована *main* функція, яка через метод `dp.start_polling(bot)` вводить в дію процес *polling* – це спосіб постійної перевірки Telegram на наявність нових

оновлень. Останнім елементом запускової частини коду є точка входу в скрипт. Це серцевина нашого проєкту, без неї нічого не працюватиме.

3.3.1. Розроблення функціонала пошуку

Не менш вагомою частиною нашої програми є пошукові можливості із нашою тлумачною збіркою прислів'їв та приказок. Проаналізувавши варіанти користувацьких можливостей пошуку серед доступних онлайн-ресурсів словників, ми обрали такі, що на нашу думку, мають найбільший попит та найкраще допоможуть зацікавленим особам отримати той результат, якого вони очікують.

Розглянемо створення окремого коду програми `search.py`, яка відповідає за пошук [див. Додаток Г]. Створений скрипт визначає два класи, *ProverbSearchResult* й *ProverbsFilter*, і дві допоміжні функції, *search_proverbs* і *search_lemmas_for_first_letter*:

- *ProverbSearchResult*: цей клас представляє єдиний результат пошуку паремій. Він має три атрибути: семантичне поле, прислів'я та тлумачення. Таким чином, ми обмежуємо програму, щоб нічого зайвого, непотрібного він не видав користувачеві.
- *ProverbsFilter*: клас, який представляє фільтр, опцію для пошуку прислів'їв і приказок. Він має декілька атрибутів, які можна використовувати для фільтрації результатів пошуку: *lemma*, *usage_types*, *first_proverb_letter* та *substring*. Також у ньому є метод *get_query*, який повертає SQL-запит з параметрами згідно з фільтром, який обрав користувач.
- *def search_proverbs(filter, db)*: ця функція виконує пошук пареміологічних одиниць на основі наданого фільтру у базі даних *db*. Вона викликає метод *get_query* об'єкта *filter* для отримання SQL-запиту та параметрів. Тоді виконує запит за допомогою *db.select_all* і отримує рядки з необхідними даними (пареміями, їх значеннями та категоріями). Потім вона створює об'єкти *ProverbSearchResult* з цих рядків і повертає список результатів.

- `def search_lemmas_by_first_letter(letter, usage_type, db)`: функція, що шукає леми, які починаються з певної літери із зазначенням поля пошуку (у самій паремії у нашому випадку). Виконується SQL-запит, щоб отримати неповторювані леми з таблиць `lemmas_usage` і `lemma`. У кінці, як і попередня функція, за допомогою `db.select_all` отримує рядки, витягує леми з рядків і повертає їх списком.

Усі ці компоненти використовуються для пошуку прислів'їв і приказок на основі різних фільтрів.

Після цього було розпочато роботу над самим ботом, імплементацією у нього функції пошуку. Спочатку був створений обробник (також називатимемо це хендлером) який відповідав на текст "Пошук 🔍" надсилаючи повідомлення «Оберіть вид пошуку:» та відображаючи під ним клавіатуру (`kb.search`), на якій розташовані кнопки «За лемою у паремії», «За лемою у паремії і тлумаченні», «За першою літерою паремії» та «За частинкою у паремії і тлумаченні». Для кожної із цих кнопок потрібні свої хендлери. На цьому етапі нам і згодилася можливість реалізувати скінченний автомат завдяки `aiogram`. Це важливо через те, що так наша програма може здійснювати переходи станів, запам'ятовуючи інформацію, отриману ззовні лишень на момент обробки певного процесу. На рис. 3.4 зображений один з обробників для пошуку «За лемою у паремії». Він обробляє дані зворотного виклику (`CallbackQuery`) `by_lemma_in_proverb`, встановлює стан пошуку в `Search.lemma` і пропонує користувачеві ввести лему для пошуку.

```
@router.callback_query(F.data == 'by_lemma_in_proverb')
async def by_lemma_in_proverb(callback: CallbackQuery, state: FSMContext):
    await callback.answer('Пошук здійснюватиметься за лемою у паремії')
    await callback.message.edit_text('Пошук здійснюватиметься за лемою у паремії.'
                                     '\nУведіть лему для пошуку:')
    await state.set_state(Search.lemma)
    await state.update_data(offset=0)
    await state.update_data(next_search=False)
```

Рисунок 3.4. Обробник для пошуку «За лемою у паремії».

Інші хендлери (`by_lemma_in_proverb_and_meaning`, `by_first_letter_in_proverb`, `by_substring_in_proverb`) працюють аналогічно, встановлюючи різні стани й

запитуючи відповідне пошукове введення.

Далі програма переходить до інших обробників певних типів пошуку, які відповідають за фактичні пошукові запити. Усі вони оновлюють стан отриманих даних, таким чином зберігаючи їх, і крім цього, виконують пошук за допомогою *search_proverbs*, а потім передають результати викликом методу *process_results*. Спеціально розроблений для цього хендлер обробляє та групує паремії за семантичними полями й надсилає їх користувачеві.

3.3.2. Розроблення функціонала навчання

Останнім етапом розробки було написання програмного коду для вивчення паремій [див. Додаток Г]. Ми одразу розуміли, що це буде організовано у вигляді скінченного автомата, адже нам це дозволяє вибрана бібліотека *aiogram*. Спочатку було створено три групи станів: *Test*, *Quiz* і *QuizVerb*, що визначаються за допомогою класу *StatesGroup*. Кожна група містить підстани, що представляють різні етапи виконання вправ. Далі було реалізовано обробник повідомлення Навчатися, який запускає навчальний сеанс, надаючи вступ і демонструючи доступні вправи. Після цього розпочалась реалізація роботи кожної вправи по черзі.

Для вправи – скласти прислів'я чи приказку, ідея полягає в тому, щоб користувач отримував список слів, які утворюють паремію, у випадковому порядку і повинен скласти із них правильно одиницю. Розглянемо будову окремих елементів. Перший хендлер відповідав за ініціацію вправи на складання. Він викликає *choose_category* і встановлює стан *Test.test1*. Перше це асинхронна функція призначена для вибору семантичного поля (у програмі називатимемо це темою, для легшого розуміння користувачами). Вона отримує тему з бази даних із пагінацією, створює вбудовану клавіатуру з кнопками полів та кнопками навігації (наступна/попередня) (див рис. 3.11). Результатом її роботи є надсилання або редагування повідомлення з обробленими параметрами. Стан, у який ми попадаємо після функції, отримує усі прислів'я/приказки вибраного семантичного поля, тоді перемішує слова випадкового прислів'я і надсилає користувачеві для

реконструкції. У кінці переходить у стан *Test.correct_proverb* з вибраною темою та деталями прислів'я. У ньому, наша програма перевіряє, чи відповідає реконструйоване користувачем прислів'я правильному порядку. Якщо використано всі паремії з теми, бот завершує вправу. В іншому випадку користувачеві буде запропоновано наступне прислів'я чи приказку з перетасованого списку.

Наступний обробник запуслав у програмі вправу для встановлення правильного значення. У цьому завданні користувач отримує повідомлення із паремією та варіантами відповідей. Сам обробник виконує ту ж функцію, що і попередній, тільки встановлює інший відповідний стан, який відповідає саме за цю вправу. Його відмінність зводиться до іншої конструкції клавіатури. На ній окрім опцій завершити вправу чи змінити тему ще відображені варіанти відповідей (див. рис. 3.13 або 3.14).

Останньою вправою є – доповнити прислів'я/приказку дієсловом. Знову ж таки, її реалізація нічим не відрізняється від попередніх, окрім як методу відбору дієслів. Для цього завдання ми скористались бібліотекою *rumorphy2* [61]. Вона була використана для тегування частин мови (знаходження дієслів) і низки граматичних категорій: числа, роду, виду дієслова та форми особи й часу. Таке рішення було прийнято для того, щоб до правильного варіанта програма спробувала підібрати хибні, які матимуть близьку граматичну характеристику до першого. Чому ми кажемо спробувала? Це пов'язано із можливостями використаного фреймворку. *Rumorphy2* в першу чергу розроблявся для російської мови, через це його функції для української є дещо обмеженими. Тому щоб отримати вправи кращої якості необхідно підключати якісь інші системи або ж укладати ці вправи вручну, проте це завдання дуже трудомістке і потребує багато часу.

3.4. Демонстрація роботи Telegram-бота

Алгоритм взаємодії користувача із ботом розпочинається на етапі його пошуку у застосунку месенджера Telegram. У полі пошуку вводиться або його назва «приказки і прислів'я: бот-помічник» або юзернейм [@PrysliviaPrykazkyBot](#). Також доступ до бота можна отримати за посиланням

[<https://t.me/PryslivliaPrykazkyBot>]. Перед запуском, користувачеві у полі де відбувається листування видано віконце з описом нашого проєкту (рис. 3.3). Після запуску людині приходить повідомлення «Оберіть опцію на своїй клавіатурі» та на її місці з'являються три кнопки: «Навчатися», «Пошук 🔍» та «Допомога» (див. рис. 3.4).

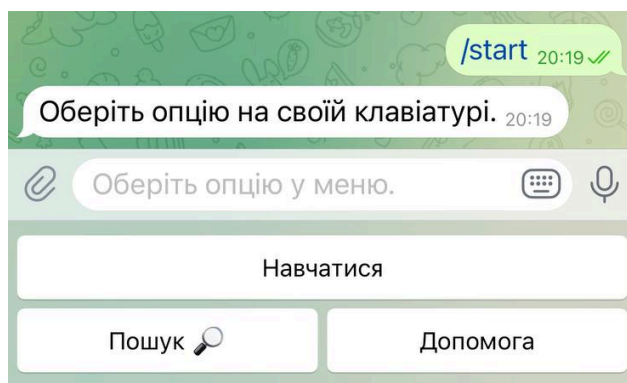


Рисунок 3.4. Запущений вперше Telegram-бот.

Якщо натиснути на кнопку «Допомога» або надіслати команду /help, то користувачеві прийде повідомлення з умовною інструкцією, поясненням, як користуватись ботом (див рис. 3.5). Ми розробили таку функцію беручи до уваги, що не кожен може зрозуміти що і як працює.

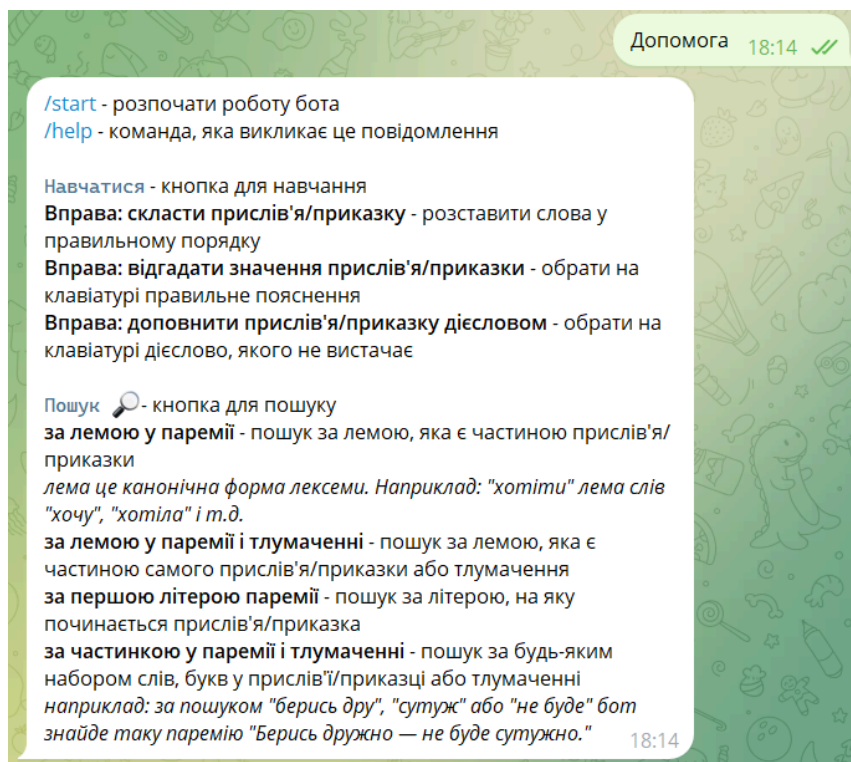


Рисунок 3.5. Текст команди /help і кнопки Допомога.

Обираючи «пошук», користувачеві приходить нове повідомлення з InlineKeyboard – клавіатура прив'язана до повідомлення, що використовує зворотний виклик, замість надсилання повідомлення зі звичайної клавіатури. Це було зроблено для забезпечення кращого користувацького досвіду. На тій клавіатурі є чотири опції типу пошуку (див рис. 3.6).

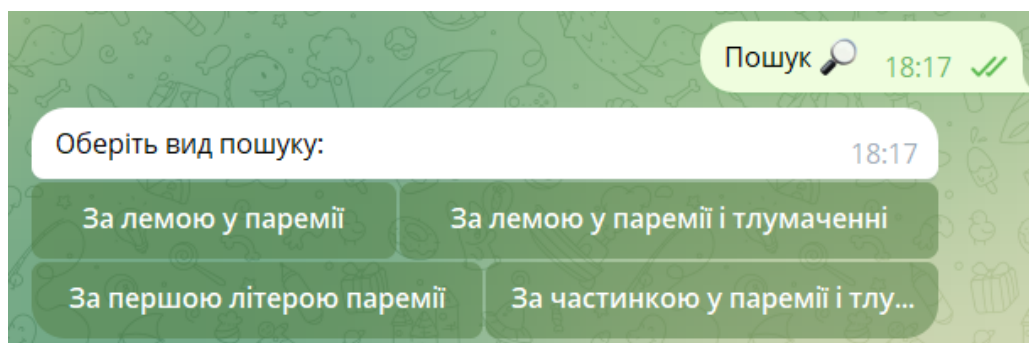


Рисунок 3.6. Повідомлення та кнопки з опціями пошуку.

При натисканні на одну із кнопок, користувачеві приходить спливаюче вікно з інформацією про обраний вид пошуку. За замовчуванням воно відрізняється на десктопній та мобільній версії Telegram. Саме ж повідомлення, яке просило обрати вид пошуку, змінюється і замість його тексту, бот повідомляє додатково, як здійснюватиметься пошук і просить увести необхідні дані (див рис. 3.7).

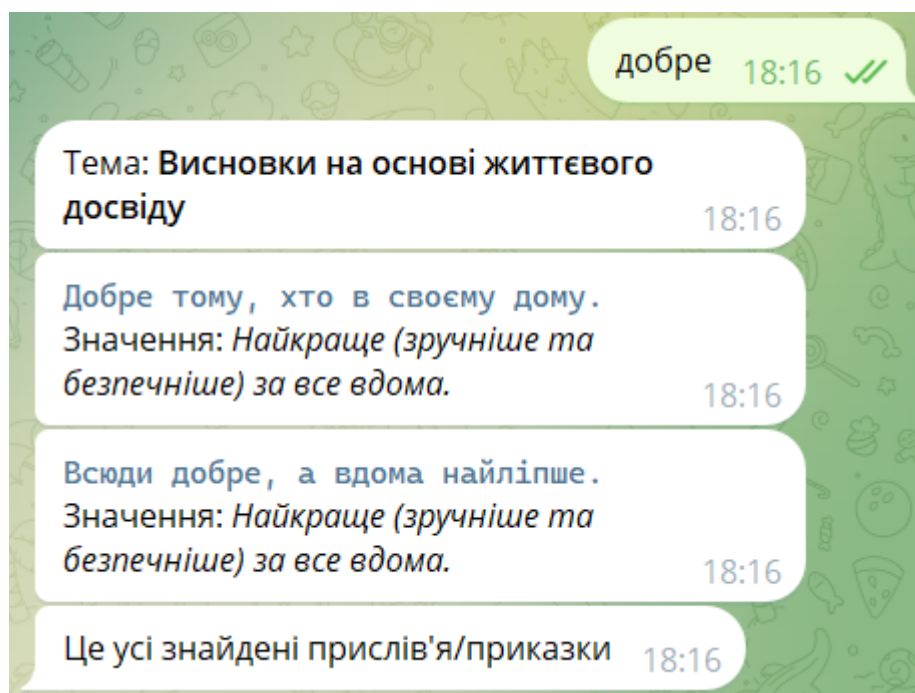


Рисунок 3.7. Результат пошуку за лемою «добре» у паремії.

На рисунку 3.7 — бачимо, що бот спочатку надіслав повідомлення про тему, у якій знайшлися дані отримані на вхід, а далі по черзі він прислав паремії з їх тлумаченнями.

Нами було прийняте рішення у коді виділити різні секції тексту спеціальними маркерами, для візуально кращого виведення результатів. Так, назва теми написана жирним шрифтом, значення паремії – курсивом, а сама одиниця оформлена у стилі, що відображає функцію копіювання. Тобто при натисканні на паремію, як у мобільній, так і десктопній версії месенджера, текст скопіюється у буфер обміну, що дозволяє швидко застосувати прислів'я чи приказку в потрібній функції (у листуванні з кимось, наприклад).

У кінці роботи функції пошуку бот надсилає повідомлення про те, що це усі знайдені паремії. Однак, якщо результати перевищуватимуть 5 одиниць, то їх виведення буде дещо іншим (див рис. 3.8). Під останнім у переліку прислів'ям чи приказкою з'являтиметься кнопка «Далі...», яка при натисканні надсилає наступні одиниці. Таким чином ми уникнули хаосу з надсиланням великої кількості результатів одразу. Якщо ж 5 одиниця все-таки остання, то бот про це сповістить повідомленням.

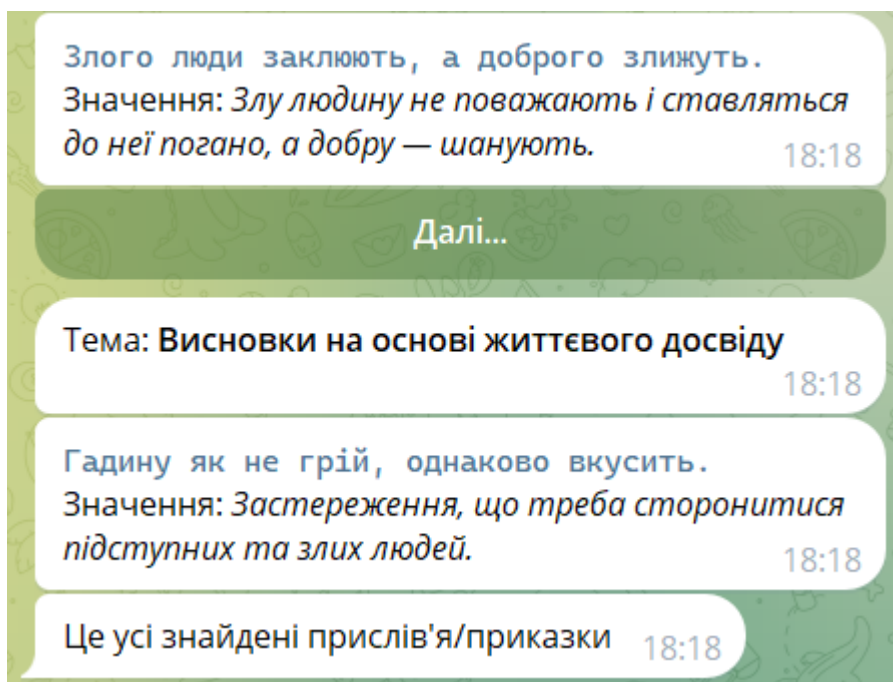


Рисунок 3.8. Результат функціонала кнопки «Далі...» при пошуку за лемою «злий» у паремії й значенні.

Якщо користувач помилився при наборі необхідних дані або бот не знайшов нічого у базі даних, надсилається відповідне повідомлення, яке супроводжується наданням ще однієї спроби, пропустивши крок вибору функції пошуку (див. рис. 3.9).

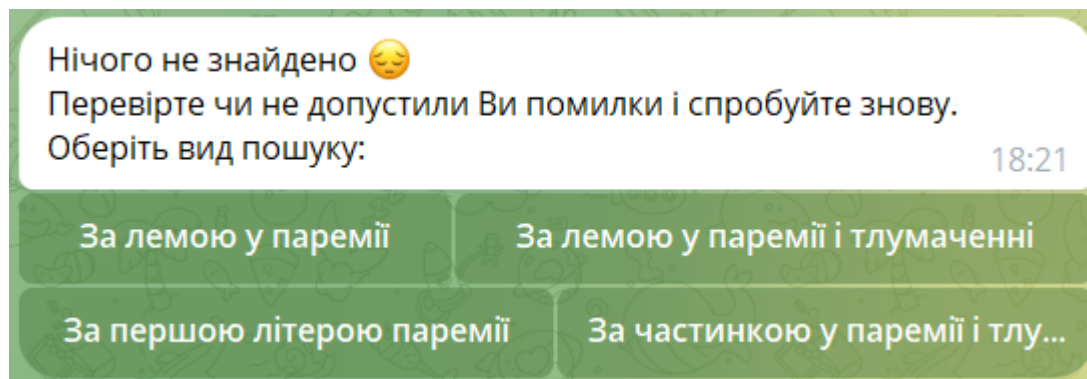


Рисунок 3.9. Повідомлення при неправильно набраних даних або не знайдених одиниць.

Обираючи «Навчання», бот надсилає повідомлення з інформацією про вправи та наступні дії, які необхідно виконати. Замість кнопок меню, у клавіатурі з'являються кнопки, що відповідають завданням (див. рис. 3.10).

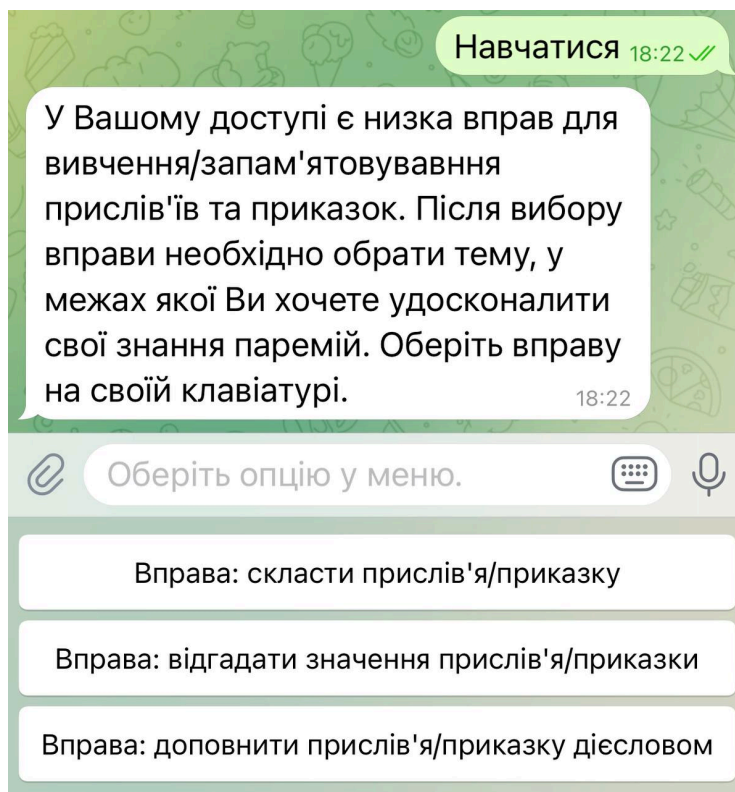


Рисунок 3.10. Повідомлення і клавіатура виклику функції навчання.

Після вибору вправи користувачеві приходить повідомлення з прив'язаними до нього кнопками тем з навігацією для показу інших, та повернення до попередніх (див. рис. 3.11). Обираючи будь-яку з опцій, кнопки зникають і на місці повідомлення з'являється інший текст. Після чого надсилається перше завдання залежно від обраної вправи.



Рисунок 3.11. Повідомлення з вибором теми.

При виконанні вправи із завданням – скласти прислів'я/приказку, бот надсилає випадковий перелік слів, із яких складається паремія і користувачеві необхідно набрати та надіслати їх у правильному порядку (див. рис. 3.12). Ми вирішили зберегти пунктуацію та велику літеру у слова, на яке починається одиниця, для спрощення завдання. Проте незалежно від того чи дотримається цих правил користувач, програма розпізнає паремію та видасть результат. Під час виконання на клавіатурі є опції: «завершити вправу» та «обрати іншу тему».

Коли користувач обирає вправу – відгадати значення прислів'я/приказки, йому приходить повідомлення із паремією та варіантами відповідей (див. рис. 3.13). Клавіатура змінюється на кнопки, що позначають опції й завершення процесу або зміна теми. Результати бот видає такі ж, як і у попередньо описаній вправі.

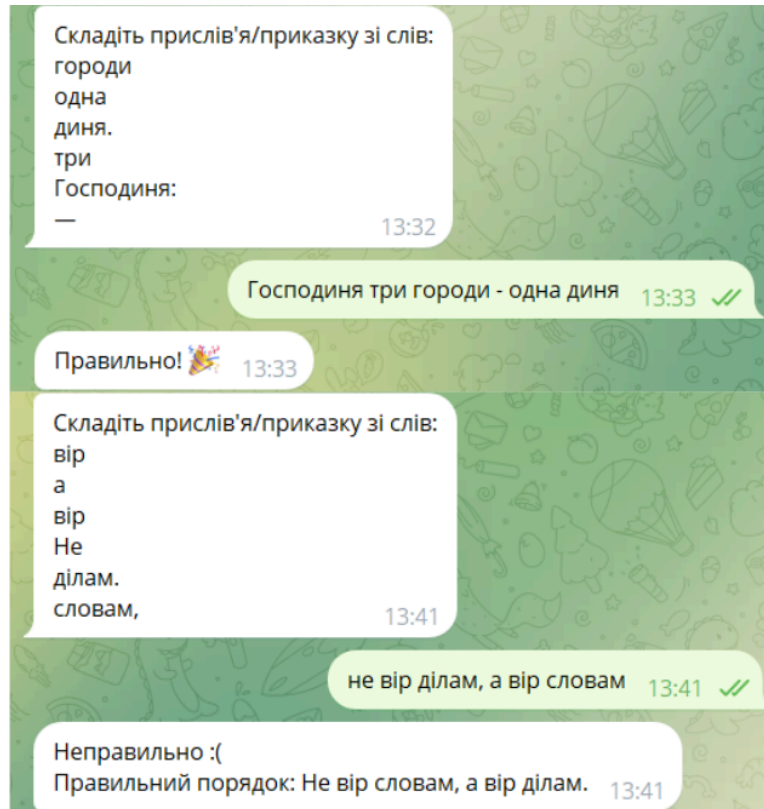


Рисунок 3.12. Приклад виконання вправи скласти прислів'я/приказку.

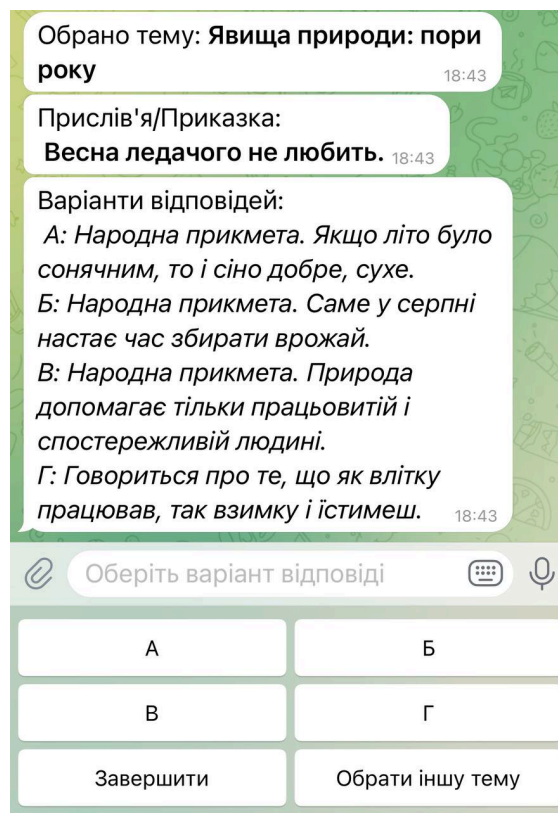


Рисунок 3.13. Приклад запуску вправи відгадати значення прислів'я/приказки.

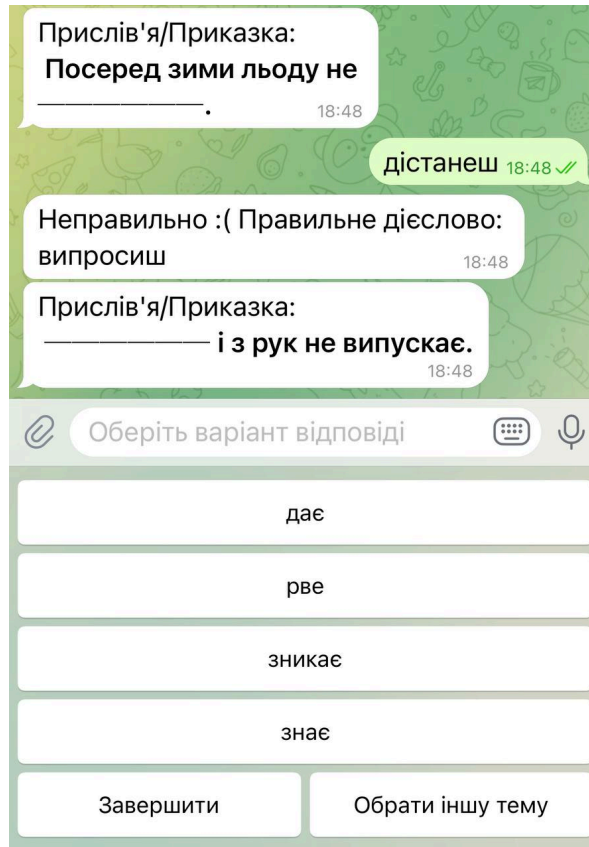


Рисунок 3.14. Приклад вправи доповнити прислів'я/приказку дієсловом.

При виконанні вправи із завданням – доповнити прислів'я/приказку дієсловом, бот надсилає повідомлення із паремією, у якій на місці пропущеного елемента ставиться символічний прочерк (див. рис. 3.14). Клавіатура організована як і в інших, правда варіанти відповідей відображають дієслова.

Коли користувач проходить усі завдання з вправи – згенеровані за певною темою, бот надсилає кількість правильних відповідей та пропонує завершити або продовжити навчання (див. рис. 3.15). Якщо обрано «завершити», то з'являється початкове меню із вибором функції бота.

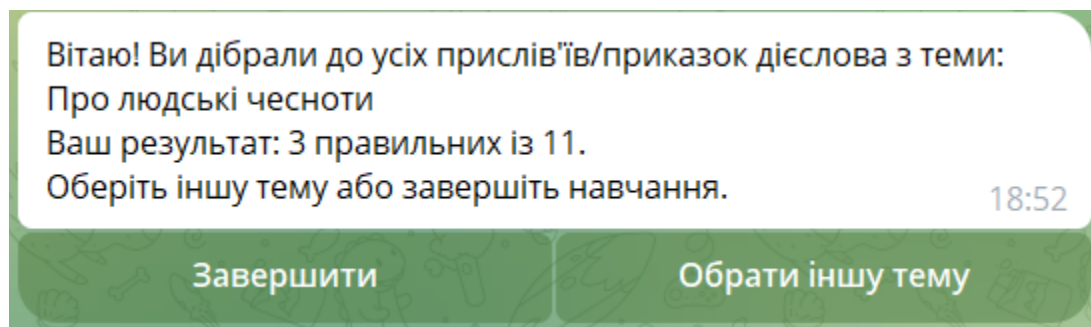


Рисунок 3.15. Повідомлення при закінченні прислів'їв/приказок в обраній темі.

ВИСНОВКИ

У роботі було поставлено мету створити Telegram-бот для правильного добору та вивчення українських паремій з поясненнями їх значень. Цю мету досягнуто. База даних складається із 230 одиниць, прокласифікованих за самотійно встановленими 14 семантичними полями. За покликанням <https://t.me/PrysliviaPrykazkyBot> представлено інтерактивний Telegram-бот «Приказки і прислів'я: бот-помічник», який є ефективним і зручним інструментом, що надає користувачеві можливості для вивчення, правильного добору і розуміння українських паремій у повсякденному мовленні як носіїв української мови, так і іноземців, які виявили інтерес до вивчення української мови. Крім того, такий онлайн-застосунок покликаний популяризувати українські паремії, які відображають ідентичність українського етносу, його мудрість і філософію.

У досягненні мети було виконано всі поставлені завдання, які дозволяють зробити такі висновки:

Пареміографія й пареміологія, хоч і є молодими галузями лінгвістики, але вони мають багатовікову історію та традиції. Вітчизняний пареміографічний доробок є результатом наполегливої праці не одного покоління. На сучасному етапі розвитку українська пареміографія досягла високого наукового рівня. Існує багато оригінальних та професійно виконаних праць, які сприятимуть формуванню методологічних принципів, необхідних для створення словників нових типів.

Проведений експеримент щодо сучасного використання прислів'їв та приказок у мовленні українців показав, що переважна більшість респондентів активно використовують паремії у повсякденному спілкуванні. Однак, водночас значна частина опитаних висловила занепокоєння щодо поступового зникнення цих виразів з активного вжитку. Висловлюючи бажання частіше використовувати прислів'я та приказки, багато респондентів погодились на використання зручних ресурсів для їх пояснення та навчання. Це підкреслює необхідність подальшого розвитку та підтримки таких проєктів, як наш Telegram-бот.

На сьогодні існує чимало програмних технологій, які набагато спрощують такі важливі етапи при укладанні електронних лексикографічних робіт, як вилучення та отримання необхідної інформації з будь-якої джерельної бази. Одна із таких технологій (jsoup) була успішно застосована у нашій роботі.

Розробка повноцінної реляційної бази даних для збереження пареміологічних одиниць, створила надійну основу для легкого подальшого розширення досліджуваного предмета.

Для досягнення кращих результатів роботи програми, доцільно покращити функціональність використаних програмних бібліотек. Можливо, інтеграція інших розробок, таких як sрасу [69], посприяла б удосконаленню встановлення граматичних категорії, що своєю чергою б вплинуло на якість вправи, де необхідно дібрати дієслово до прислів'я чи приказки.

Так само корисним для подальшого розвитку проекту буде більш докладне занурення у функціонал, який надає обрана нами бібліотека для роботи над створенням Telegram-бота, його розширення: додавання нових функцій, таких як нові інтерактивні тести для вивчення паремій (наприклад: співвіднести паремії з їх семантичними полями), можливість додавання користувачами нових паремій та можливість стеження за власним прогресом у вивченні, надсилання щоденного повідомлення із «паремією дня», надання ботові функціонала, який дозволяє викликати його в будь-якому особистому діалозі для швидшого добору прислів'я чи приказки і т. ін.. Такі зміни зроблять розробку більш корисною та привабливою для користувачів. А найголовнішою перспективою розробки є її інтеграція з серверною платформою. Таке покращення дозволить працювати боту цілодобово, підвищить стійкість роботи та швидкодію.

Створена лексикографічна база даних українських паремій дає можливість провести статистичне дослідження лексики, граматики цих одиниць, що можливо виконати у перспективі. Такі дані могли б слугувати вихідною точкою для створення стандартизованих сукупностей, як от паремійний мінімум українців або ж для репрезентації особливостей регіональної приналежності одиниць. Також непогано було б удосконалити базу даних наступним чином: зібрати більшу

кількість пареміологічних одиниць, глибше пропрацювати систему класифікації, додання до реєстру інших видів паремій (наприклад загадок). Таку базу даних можна було б застосовувати в майбутньому для тренування моделей автоматичного класифікування паремій за семантичними полями або розпізнаванню їх у тексті.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Алексахіна Т. О. Методика розроблення ідеографічної класифікації фразеологізмів. Нова філологія. 2014. № 63. С. 6–10.
2. Білецький-Носенко П. Словник української мови / ред. К. К. Цілуйко. Київ : Наук. Думка, 1966. 424 с.
3. Бобкова В. Українські народні прислів'я та приказки / упоряд. ; упоряд., вступ. ст. Ф І Лавров ; упоряд. М. П. Ліждвой ; відп. ред. Максим Тадейович Рильський. Київ : Вид-во АН УРСР, 1955. 447 с.
4. Бондаревська О. Явище апосіопези у прислів'ях і приказках. Рідний край. 2018. № 2. С. 74–79.
5. Бондаренко М. О. Семантика і лінгвокультурний потенціал компаративних паремій (на матеріалі збірки Матвія Номиса «Українські приказки, прислів'я і таке інше»). Українська мова. 2023. № 1. С. 114–124.
6. Гаврилова В. В. Теоретичні засади вивчення паремій. Наукові записки Бердянського державного педагогічного університету. Сер.: філологічні науки. 2015. № 7. С. 11–20.
7. Грозян Н. Ф., Прудникова Т. І. Фразеологічна мікросистема «Людина» як фрагмент мовної картини світу: ідеографічний аспект. Культура народів Причорномор'я. 2005.
8. Грузинські прислів'я та приказки /упоряд., пер. з грузин. Р. Чілачава. – К. : Дніпро, 1975. – 144 с.
9. Зіновійв К.. Вірші. Приповісті посполиті / Підг. тексту І. П. Чепіги. Київ: Наукова думка, 1971. 392 с.
10. Іспанські прислів'я та приказки / упорядн., вступне слово М. Жердинівська; пер з ісп. М. Литвінець. – К. : Дніпро, 1980. 110 с.
11. Колоїз Ж. В., Березовська-Савчук Н. А. Паремійний корпус Михайла Стельмаха : [словник-довідник]. Кривий Ріг : НПП Астерікс, 2015. 326 с.

12. Колоїз Ж. В., Малюга Н. М., Шарманова Н. М. Українська пареміологія : навчальний посібник для студентів філологічних спеціальностей вищих навчальних закладів. Кривий Ріг : КПІ ДВНЗ «КНУ», 2014. 349 с.
13. Колонюк С. М. «Галицько-руські народні приповідки» І. Франка – унікальне джерело фіксації та інтерпретації фатичних комунікативних формул вітання і прощання. *Philological sciences and translation studies: european potential*. 2021. С. 222–225.
14. Космеда Т. А., Колонюк С. М. Багатоликий статус паремії та аспекти її дослідження. *Лінгвістичні студії*. 2022. № 43. С. 20–31.
15. Космеда Т. А., Осіпова Т. Ф. Комунікативний кодекс українців у пареміях: тлумачний словник нового типу. Дрогобич: Коло, 2010. 272 с.
16. Космеда Т. А., Осіпова Т. Ф. Тлумачний словник комунікативного призначення як породження сучасної прагматично зорієнтованої наукової парадигми. *Лексикографічний бюлетень: Зб. наук. пр.* — К.: Ін-т української мови НАН України, 2010. Вип. 19. – С. 7-24.
17. Крикманн А. А. Некоторые аспекты семантической неопределенности пословицы. *Пареміологічний збірник: Пословица. Загадка: Структура, смысл, текст*; ред. Г. Л. Пермякова. Наука, 1978. – С. 82–104.
18. Кузь Г. Фразеологічний мінімум з української мови для слов'яномовних студентів (рубіжний та високий рівні, B1-B2). *Studia ucrainica varsoviensia*. 2020. № 8. С. 97–104.
19. Лановик М. Б., Лановик З. Б. Українська усна народна творчість: підручник / ред. Р. Т. Гром'як. 3-тє вид. Київ, Україна : Знання-Прес, 2005. 591 с.
20. Номис М. Українські приказки, прислів'я і таке інше / упоряд. М. М. Пазяка. Київ : Либідь, 1993. 768 с.
21. Овчиннікова А. А., Некрілова К. В. Народний фольклор про культуру харчування. Актуальні проблеми розвитку харчових виробництв, готельного, ресторанного господарств і торгівлі: тези доп. Всеукр. наук.-практ. конф. молодих учених і студентів. 2011. № 1. С. 40.

- 22.Пешак М., Солонська Н. 925 років визначній давньоруській пам'ятці — Збірнику Святослава 1076. Бібліотечний вісник. 2001. № 6. С. 52–53.
- 23.Польські прислів'я та приказки / упорядн., пер. з польськ. І. Дзюба. – К. : Дніпро, 1970. 176 с.
- 24.Прадід Ю. Ф. Фразеологічна ідеографія (проблематика досліджень). НАНУ, Ін-т української мови; відп. ред. О. О. Тараненко. – К. – Сімферополь, 1997. 252 с.
- 25.Прислів'я та приказки: Взаємини між людьми / Упорядн. М. М. Пазяк. Київ: Наукова думка, 1991. 440 с.
- 26.Прислів'я та приказки: Природа. Господарська діяльність людини. / Упорядн. М. М. Пазяк. Київ: Наукова думка, 1989. 480 с.
- 27.Северинюк В.М. Тематичний словник популярних українських прислів'їв та приказок з коментарями. / Тернопіль: Навчальна книга – Богдан, Вид. 3-тє, доп. і переробл. 2014. 176 с
- 28.Скиба І. Г. Особливості фіксації паремій у словниках за редакцією Б. Грінченка та В. Даля. Слов'янська фразеологія: семантичний, ареальний, історичний та етнокультурний аспекти. 2018. С. 48-51.
- 29.Сокаль М. А. Про можливості реалізації лінгводидактичного потенціалу прислів'їв та приказок на уроках мови в початковій школі. Психолого-педагогічні проблеми сільської школи: зб. наук. пр. – умань: УДПУ ім. павла тичини. 2010. № 33. С. 114–120.
- 30.Соловей Е. С. Українська філософська лірика : Навч. посіб. зі спецкурсу для студ. вузів. Київ: Юніверс, 1999. 366 с.
- 31.Тищенко І. А. Мовленнєві жанри: лакуни української пареміології. Філологія ХХІ століття : зб. наук. пр. студентства й наук. молоді. 2020. С. 38–40.
- 32.Франко І. Я. Галицько-руські народні приповідки: у 3 томах. 2-ге вид. Львів: Вид. центр ЛНУ ім. Ів. Франка, 2006. 2330 с.

- 33.Шумада Н. С. Народ скаже — як зав'яже: українські народні прислів'я, приказки, загадки, скоромовки / ред. В. В. Бичко; іл. А. Г. Гілевич. Київ: Веселка, 1973. 360 с.
- 34.Ďurčo P. Empirical research and paremiological minimum. Introduction to paremiology: a comprehensive guide to proverb studies. Warsaw, Poland. 2015. P. 183–201.
- 35.Firth R. Proverbs in Native Life, with Special Reference to Those of the Maori. Folklore – 1926. №37(2). P. 134–153.
- 36.Grzybek P. SEMIOTIC AND SEMANTIC ASPECTS OF THE PROVERB. Proverbium: Yearbook of International Proverb Scholarship. 38, 1. 2021. P. 67–123.
- 37.Hrisztova-Gotthardt H., Varga M. A. Introduction. Introduction to paremiology: a comprehensive guide to proverb studies. Warsaw, Poland. 2015. P. 1–6.
- 38.Jurafsky D., Martin J. Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech 82 Recognition. 2020.
- 39.Lauhakangas O. Categorization of proverbs. Introduction to paremiology: a comprehensive guide to proverb studies. Warsaw, Poland. 2015. P. 49–66.
- 40.Mieder W. International Proverb Scholarship: An Updated Bibliography. Proverbium: Yearbook of International Proverb Scholarship, 39(1). 2022. 373-410.
- 41.Mieder W. Origin of proverbs. Introduction to paremiology: a comprehensive guide to proverb studies. 2015. P. 28–48.
- 42.Norrick N. R. Subject area, terminology, proverb definitions, proverb features. Introduction to paremiology: a comprehensive guide to proverb studies. Warsaw, Poland. 2015. P. 7–25.
- 43.Performance analysis for web scraping tools: case studies on beautifulsoup, scrapy, htmlunit and jsoup / Y. Dikilitaş et al. Lecture notes in networks and systems. Cham, 2024. P. 471–480.

44. Raji-Oyelade, A. Classifying the unclassified: the challenge of postproverbiality in international proverb scholarship. In O. Lauhakangas & R. Soares (Eds), Acta Proceedings of ICP. 2008 P. 146-155.

ЕЛЕКТРОННІ ДЖЕРЕЛА

45. Вікіцитати. Прислів'я та приказки народів світу. URL: https://uk.wikiquote.org/wiki/Прислів'я_та_приказки_народів_світу (Last accessed: 10.06.2024).
46. Іван Франко. Енциклопедія життя і творчості. Галицько-руські народні приповідки. URL: <https://www.i-franko.name/uk/Folklore/1901/GalRusProverbs.html> (Last accessed: 10.06.2024).
47. Приповідки або українсько-народня філософія URL: <https://slovnyk.me/dict/proverbs> (Last accessed: 10.06.2024).
48. Прислів'я та приказки українців Канади, зібрані Володимиром Плав'юком. URL: https://mala.storinka.org/Прислів'я_та_приказки_українців_Канади_зібрані_Володимиром_Плав'юком.html (Last accessed: 10.06.2024).
49. Словник прислів'їв та приказок з поясненнями. URL: https://kampot.org.ua/ukraine/traducii_ta_zvuchai/zvuchai/2313-slovník-prislvyiv-ta-prikazok-z-poyasnennyami.html (Last accessed: 10.06.2024).
50. Усна народна творчість. Прислів'я та приказки. URL: <https://traditions.in.ua/usna-narodna-tvorchist/pryslivia-ta-prykazky> (Last accessed: 10.06.2024).
51. Aiogram framework. URL: <https://aiogram.dev/> (Last accessed: 10.06.2024).
52. Asyncio – Asynchronous I/O. URL: <https://docs.python.org/3/library/asyncio.html> (Last accessed: 10.06.2024).
53. BeautifulSoup. Python library for pulling data out of HTML and XML files. URL: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/> (Last accessed: 10.06.2024).

- 54.Commons-IO. URL: <https://commons.apache.org/proper/commons-io/> (Last accessed: 10.06.2024).
- 55.Jackson. URL: <https://github.com/FasterXML/jackson> (Last accessed: 10.06.2024).
- 56.Java SE Development Kit 8u391. URL: <https://www.oracle.com/cis/java/technologies/javase/javase8u211-later-archive-downloads.html> (Last accessed: 10.06.2024).
- 57.Json. URL: <https://docs.python.org/3/library/json.html> (Last accessed: 10.06.2024).
- 58.Jsoup: Java HTML Parser. URL: <https://jsoup.org/> (Last accessed: 10.06.2024).
- 59.Logging. URL: <https://docs.python.org/uk/3/library/logging.html> (Last accessed: 10.06.2024).
- 60.Notion: Your connected workspace for wiki, docs & projects. URL: <https://www.notion.so/> (Last accessed: 10.06.2024).
- 61.Pymorphy2. URL: <https://pymorphy2.readthedocs.io/en/stable/> (Last accessed: 10.06.2024).
- 62.PyTelegramBotAPI. URL: <https://pytba.readthedocs.io/ru/latest/> (Last accessed: 10.06.2024).
- 63.Python 3.9.13. URL: <https://www.python.org/downloads/release/python-3913/> (Last accessed: 10.06.2024).
- 64.Python-telegram-bot. URL: <https://python-telegram-bot.org/> (Last accessed: 10.06.2024).
- 65.Random. URL: <https://docs.python.org/uk/3/library/random.html> (Last accessed: 10.06.2024).
- 66.Re. URL: <https://docs.python.org/3/library/re.html> (Last accessed: 10.06.2024).
- 67.Scrapy. Framework for extracting the data you need from websites. URL: <https://scrapy.org/> (Last accessed: 10.06.2024).
- 68.Simple python lib to tokenize texts tokenize_uk. URL: <https://tokenize-uk.readthedocs.io/en/latest/readme.html> (Last accessed: 10.06.2024).

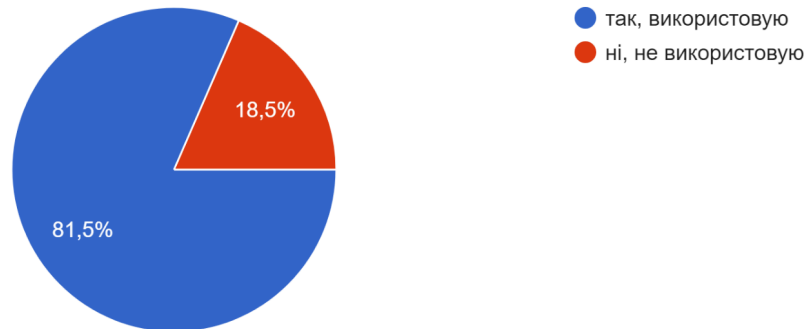
69. Spacy – Industrial-strength in natural language processing in Python. URL: <https://spacy.io/> (Last accessed: 10.06.2024).
70. SQLite. Small. Fast. Reliable. Choose any three. URL: <https://www.sqlite.org/index.html> (Last accessed: 10.06.2024).
71. Sqlite3. URL: <https://docs.python.org/3/library/sqlite3.html> (Last accessed: 10.06.2024).
72. Telegram: a new era of messaging. URL: <https://telegram.org/> (Last accessed: 10.06.2024).
73. Telegram bot «приказки і прислів'я: бот-помічник» URL: <https://t.me/PrysliviaPrykazkyBot> (Last accessed: 10.06.2024).
74. Telethon. URL: <https://docs.telethon.dev/en/stable/#> (Last accessed: 10.06.2024).

ДОДАТКИ

Додаток А. Результати опитування про використання прислів'їв та приказок у мовній практиці українців.

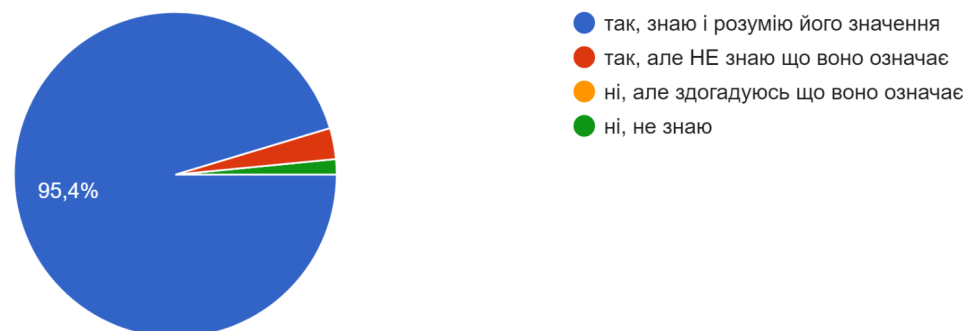
Чи використовуєте Ви прислів'я/приказки у повсякденному мовленні?

65 відповідей



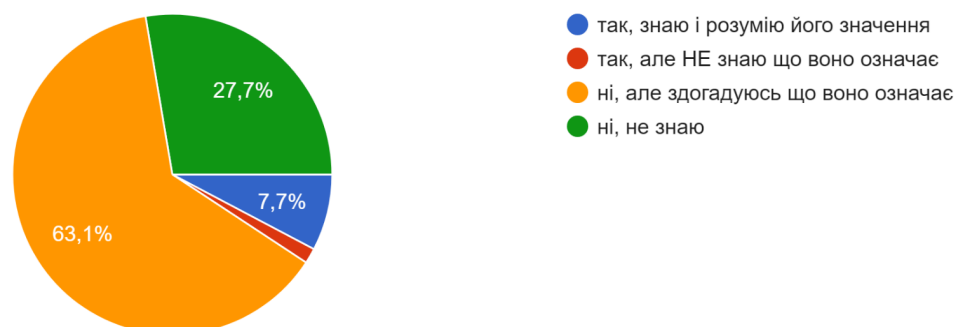
1. Чи знаєте Ви таке прислів'я/приказку? "Або пан, або пропав."

65 відповідей

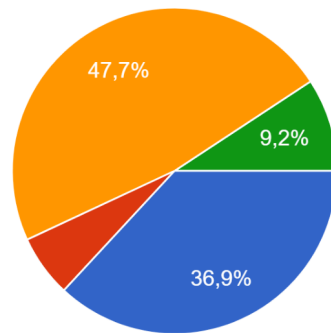


2. Чи знаєте Ви таке прислів'я/приказку? "Вівсяна каша хвалилася, начебто з коров'ячим маслом народилася."

65 відповідей

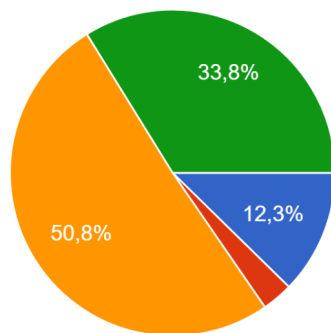


3. Чи знаєте Ви таке прислів'я/приказку? "В доброго хазяїна і соломинка не пропаде."
65 відповідей



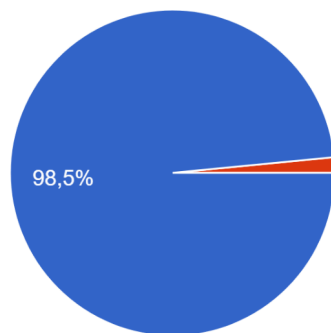
- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

4. Чи знаєте Ви таке прислів'я/приказку? "Одна жарина і в печі гасне, а дві і в полі горять."
65 відповідей



- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

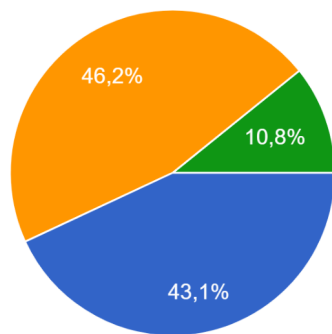
5. Чи знаєте Ви таке прислів'я/приказку? "Один за всіх, всі за одного."
65 відповідей



- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

6. Чи знаєте Ви таке прислів'я/приказку? "Хто спить весною, мерзне зимою."

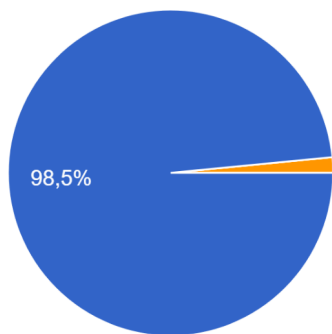
65 відповідей



- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

7. Чи знаєте Ви таке прислів'я/приказку? "Не вір словам, а вір ділам."

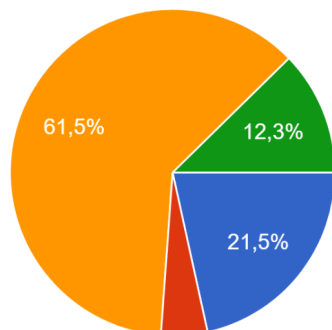
65 відповідей



- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

8. Чи знаєте Ви таке прислів'я/приказку? "Утікав від вовка, та впав на ведмедя."

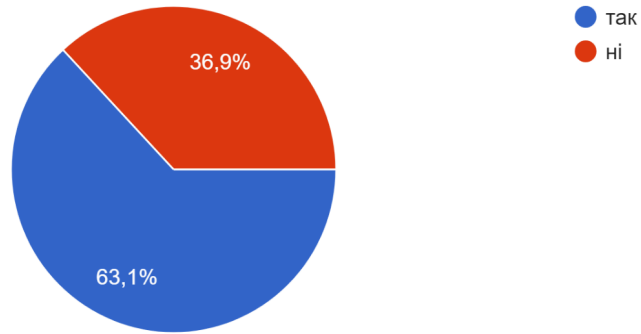
65 відповідей



- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

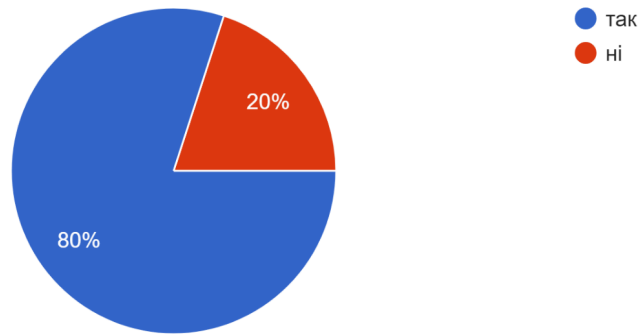
Чи вважаєте Ви, що прислів'я/приказки зараз на межі "вимирання" з вжитку?

65 відповідей



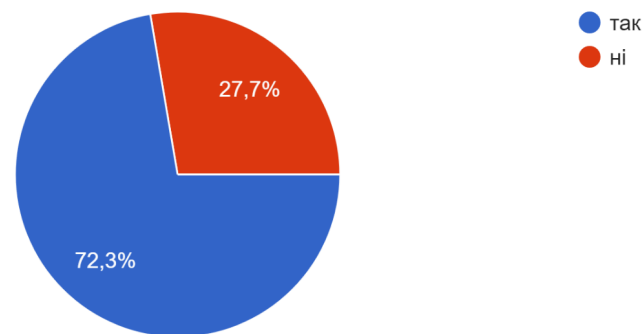
Чи хотіли б Ви використовувати прислів'я/приказки частіше?

65 відповідей



Чи вживали б Ви частіше прислів'я/приказки, якби для цього був зручний ресурс, що пояснює їх значення (наприклад застосунок, телеграм-бот) ?

65 відповідей



Додаток Б. Анкета опитування про використання прислів'їв та приказок у мовній практиці українців.

Чи використовуєте Ви прислів'я/приказки у повсякденному мовленні? *

- так, використовую
- ні, не використовую

1. Чи знаєте Ви таке прислів'я/приказку? *

"Або пан, або пропав."

- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

2. Чи знаєте Ви таке прислів'я/приказку? *

"Вівсяна каша хвалилася, начебто з коров'ячим маслом народилася."

- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

3. Чи знаєте Ви таке прислів'я/приказку? *

"В доброго хазяїна і соломинка не пропаде."

- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

4. Чи знаєте Ви таке прислів'я/приказку? *

"Одна жарина і в печі гасне, а дві і в полі горять."

- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

5. Чи знаєте Ви таке прислів'я/приказку? *

"Один за всіх, всі за одного."

- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

6. Чи знаєте Ви таке прислів'я/приказку? *

"Хто спить весною, мерзне зимою."

- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

7. Чи знаєте Ви таке прислів'я/приказку? *

"Не вір словам, а вір ділам."

- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

8. Чи знаєте Ви таке прислів'я/приказку? *

"Утікав від вовка, та впав на ведмедя."

- так, знаю і розумію його значення
- так, але НЕ знаю що воно означає
- ні, але здогадуюсь що воно означає
- ні, не знаю

Чи вважаєте Ви, що прислів'я/приказки зараз на межі "вимирання" з вжитку? *

- так
- ні

Чи хотіли б Ви використовувати прислів'я/приказки частіше? *

- так
- ні

Чи вживали б Ви частіше прислів'я/приказки, якби для цього був зручний ресурс, що пояснює їх значення (наприклад застосунок, телеграм-бот) ? *

- так
- ні

Додаток В. Повний програмний код вебскрейпінгу та його результатів.

Переглянути всі файли проекту та отриманих даних можна на GitHub репозиторії за посиланням:

https://github.com/KurpitaNazarii/proverbs_scraping

Або у google диску за посиланням:

https://drive.google.com/drive/folders/1LNDpG4sPaJKYAegPsZHxxajHy26riyvy?usp=drive_link

Додаток Г. Повний програмний код Teleram-бота.

Переглянути всі файли проекту можна на GitHub репозиторії за посиланням:

https://github.com/KurpitaNazarii/ukrainian_proverbs_telegram_bot

Або у google диску за посиланням:

https://drive.google.com/drive/folders/1LNDpG4sPaJKYAegPsZHxxajHy26riyvy?usp=drive_link