

УДК 81:004.8

DOI <https://doi.org/10.24919/2308-4863/82-1-45>

Ірина БАСАРАБА,
orcid.org/0000-0002-3209-9119
доктор філософії з філології,
старший викладач кафедри іноземних мов
Національної академії Державної прикордонної служби України
імені Богдана Хмельницького
(Хмельницький, Україна) irynaborovuk2017@ukr.net

Олег БОРОВИК,
orcid.org/0000-0003-3691-662X
доктор технічних наук, професор,
заступник начальника відділу організації освітньої
та наукової діяльності управління професійної підготовки Департаменту персоналу
Адміністрації Державної прикордонної служби України
(Київ, Україна) bov_nadpsu@ukr.net

Людмила БОРОВИК,
orcid.org/0000-0003-2949-2187
доктор педагогічних наук, професор,
завідувач кафедри загальнонаукових та інженерних дисциплін
Національної академії Державної прикордонної служби України
імені Богдана Хмельницького
(Хмельницький, Україна) blv_nadpsu@ukr.net

ТЕОРЕТИЧНІ ОСНОВИ СТВОРЕННЯ СИСТЕМИ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ ФРАЗЕОЛОГІЧНИХ ОДИНИЦЬ В АНГЛОМОВНИХ ТЕКСТАХ НА ОСНОВІ ВИКОРИСТАННЯ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ

У статті опрацьовано теоретичні основи для розробки ефективної системи автоматичного розпізнавання фразеологізмів в англійських текстах на основі використання сучасних інформаційних технологій, згалом, і штучного інтелекту, зокрема. Здійснено чітку постановку задачі розбудови системи автоматичного розпізнавання фразеологізмів в англійських текстах і спроектовано структуру відповідної системи.

За результатами проведеного дослідження зроблено наступні висновки: на цей час однією з ключових проблем в обробці природної мови є розпізнавання фразеологізмів; актуальним є завдання автоматизації виявлення фразеологізмів в англійських текстах; найбільш поширені існуючі підходи автоматизації виявлення фразеологізмів базуються на застосуванні правил *Rule-based approaches* та машинного навчання; існуючі системи автоматизації виявлення фразеологізмів містять низку недоліків, які не дозволяють ефективно вирішувати завдання їх якісної ідентифікації; актуальним завданням є розробка ефективної системи автоматичного розпізнавання фразеологізмів в англійських текстах, яка була б надійною, застосовною до обробки різноманітних структур речень і різних типів фразеологізмів, включаючи сталі вирази, ідіоми та словосполучення, а також містила б мінімальну кількість недоліків, що характерні для систем, у яких реалізовані методи правил *Rule-based approaches* і машинного навчання; вирішення задачі розбудови системи автоматичного розпізнавання фразеологізмів в англійських текстах може підвищити ефективність виявлення фразеологізмів в англійських текстах; теоретичною основою вирішення сформульованої задачі може слугувати запропонований авторами гібридний метод, основна ідея якого полягає у використанні заздалегідь визначених правил для категоризації фраз, які відповідають певним критеріям, і використання алгоритму машинного навчання для категоризації фраз, які їм не відповідають; ефективним засобом реалізації гібридного методу може бути система, структура якої містить наступні модулі: введення; попередньої обробки; ідентифікації фразеологізмів; класифікації фразеологізмів; виведення; зворотного зв'язку; для розробки системи автоматичного розпізнавання фразеологізмів в англійських текстах, що реалізує гібридний метод, доцільно використати мову Python.

Ключові слова: фразеологічна одиниця, англійський текст, ідентифікація, система автоматичного розпізнавання фразеологізмів, штучний інтелект.

Iryna BASARABA,

orcid.org/0000-0002-3209-9119

Doctor of Philosophy in Philology,

Senior Lecturer at the Department of Foreign Languages Department

Bohdan Khmelnytskyi National Academy of the State Border Guard Service of Ukraine

(Khmelnyskyi, Ukraine) irynaborovyk2017@ukr.net

Oleh BOROVYK,

orcid.org/0000-0003-3691-662X

Doctor of Technical Sciences, Professor,

Deputy Head of the Department of Organization of Educational and Scientific Activities

of Professional Training of the Personnel Department

Administration of the State Border Guard Service of Ukraine

(Kyiv, Ukraine) bov_nadpsu@ukr.net

Liudmyla BOROVYK,

orcid.org/0000-0003-2949-2187

Doctor of Pedagogical Sciences, Professor,

Head of the Department of General Science and Engineering Disciplines

Bohdan Khmelnytskyi National Academy of the State Border Guard Service of Ukraine

(Khmelnyskyi, Ukraine) blv_nadpsu@ukr.net

THEORETICAL FOUNDATIONS OF CREATING A SYSTEM FOR AUTOMATIC RECOGNITION OF PHRASEOLOGICAL UNITS IN ENGLISH TEXTS BASED ON THE USE OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES

The article deals with the theoretical foundations for the development of an effective system for automatic recognition of phraseological units in English texts based on the use of modern information technologies in general and artificial intelligence in particular. A clear statement of the task of developing a system for automatic recognition of phraseological units in English texts is made and the structure of the corresponding system is designed. Based on the results of the study, the following conclusions have been drawn: currently, one of the key problems in natural language processing is the recognition of phraseological units; the task of automating the detection of phraseological units in English texts is relevant; the most common existing approaches to automating the detection of phraseological units are based on the use of Rule-based approaches and machine learning; existing systems for automating the detection of phraseological units contain a number of shortcomings that do not allow to effectively solve the problem of their qualitative identification; the current task is to develop an effective system for automatic recognition of phraseological units in English texts that would be reliable, applicable to the processing of various sentence structures and different types of phraseological units, including fixed expressions, idioms and phrases, and would contain a minimum number of shortcomings that are typical for systems that implement rule-based approaches and machine learning; solving the problem of developing a system for automatic recognition of phraseological units in English texts can increase the efficiency of detecting phraseological units in English texts; the theoretical basis for solving the formulated problem can be the hybrid method proposed by the authors, the main idea of which is to use predefined rules to categorize phrases that meet certain criteria and use a machine learning algorithm to categorize phrases that do not meet them; an effective means of implementing the hybrid method can be a system whose structure contains the following modules: input; pre-processing; phrase identification; phrase classification; output; feedback; to develop a system for automatic recognition of phrases in English-language texts that implements the hybrid method, it is advisable to use the Python language.

Key words: *phraseological unit, English text, identification, automatic phraseological unit recognition system, artificial intelligence.*

Постановка проблеми. Вивчення фразеологізмів (багатослівних виразів, які мають фіксоване значення і вживаються в певному контексті) є важливою частиною лінгвістики, оскільки допомагає зрозуміти складну структуру мови та способи її використання у спілкуванні. Останніми роками дослідженню обробки природної мови, загалом, і фразеологізмів, зокрема, приділяється багато

уваги. Це пов'язано з удосконаленням дослідницького інструментарію. Зокрема, розвиток інформаційних технологій і штучного інтелекту уможливив розробку систем, здатних аналізувати і розуміти людську мову. Обчислювальні спроможності зазначеного інструментарію збільшують кількість лінгвістичних задач, які можуть бути розв'язаними, а також поглиблюють рівні їх опрацювання.

Однією з ключових проблем в обробці природної мови є розпізнавання фразеологізмів. Необхідність автоматизації вирішення цього завдання пояснюється тим, що багато програм обробки природної мови, наприклад інтелектуальний аналіз текстів, пошук інформації, машинний переклад, потребують попереднього вирішення задачі розпізнавання в текстах фразеологічних одиниць. Автоматичне розпізнавання фразеологізмів є складним завданням, яке вимагає поєднання лінгвістичних знань, обчислювальної техніки та алгоритмів.

Аналіз останніх досліджень й публікацій.

В останній період питання розпізнавання фразеологізмів, загалом, та автоматичного розпізнавання, зокрема, було предметом досліджень ряду науковців-філологів і фахівців сфери інформаційних технологій (Басараба, 2020: 9; Meyer, 2017: 190; Bird, 2020: 288; Brown, 2021: 95). У рамках їхніх досліджень аналізувалися різні підходи та формувалися вимоги до програмно-апаратних засобів автоматичного розпізнавання фразеологічних одиниць (ФО).

Зокрема, досліджувалися методи ідентифікації і категоризації ФО в обробці природної мови, що базуються на застосуванні правил Rule-based approaches (Jurafsky, 2019: 240; Drury, 2015: 934). Ці методи покладаються на набір заздалегідь визначених правил для ідентифікації і вилучення фраз з тексту на основі синтаксичних і семантичних шаблонів. Існує кілька різних типів підходів на основі правил, які можна використовувати для розпізнавання фразеологізмів, зокрема зіставлення зі зразком, індукція правил і дерева рішень.

Ключовою перевагою зазначених підходів є їх застосовність до конкретного завдання або галузі та можливість більш точної ідентифікації ФО з їх застосуванням. Однак, незважаючи на переваги, підходи, засновані на правилах, мають і певні обмеження. Однією з головних проблем є створення всеосяжного набору правил, що можуть точно ідентифікувати та класифікувати ФО в певній мові. Це потребує значних затрат часу і може бути складним у підтримці. Ця проблема особливо характерна для мов зі складною граматичною структурою або великим словниковим запасом. Іншим обмеженням є складність урахування контекстно-залежних варіацій у використанні фразеологізмів. Багато фразеологізмів мають кілька значень або можуть вживатися по-різному залежно від контексту, що ускладнює розробку правил, які можуть точно ідентифікувати їх у всіх випадках.

Підходи на основі правил для розпізнавання фраз переважно реалізуються у вигляді програм-

них, а не апаратних засобів. Ці програми можуть бути інтегровані в різні інструменти та платформи NLP, зокрема такі, як: Natural Language Toolkit (NLTK); Stanford CoreNLP; Apache OpenNLP; General Architecture for Text Engineering (GATE); spaCy.

Ще одним поширеним підходом, що використовується для ідентифікації ФО в обробці природної мови, є машинне навчання (Kurdi, 2010: 6226; Mitchel, 2019: 956). Цей підхід використовує спеціальні алгоритми, що базуються на статистичних моделях і використовують великі набори даних. Перші підходи машинного навчання для розпізнавання фраз були засновані на статистичних моделях, таких як приховані марковські моделі та марковські моделі з максимальною ентропією. Ці моделі були розроблені для навчання на великих масивах анованого тексту, що дозволяло системі розпізнавати закономірності та ідентифікувати поширені фрази й ідіоматичні вирази. В останні роки розвиток алгоритмів глибокого навчання, таких як нейронні мережі, призвів до значного прогресу в підходах машинного навчання для розпізнавання фраз. Алгоритми глибокого навчання базуються на штучних нейронних мережах, які призначені для імітації структури та функцій людського мозку. Ці алгоритми здатні навчатися на дуже великих масивах даних і можуть автоматично ідентифікувати та виявляти ознаки з необроблених даних, що робить їх високоефективними для задач обробки природної мови. При використанні підходів машинного навчання можуть застосовуватися принципи: навчання під керівництвом (Supervised Learning); неконтрольоване навчання (Unsupervised Learning); нейронні мережі; глибоке навчання (Deep Learning).

Однією з ключових переваг підходів машинного навчання для розпізнавання фраз є здатність адаптуватися до складності та мінливості природної мови. Алгоритми машинного навчання також можуть автоматично навчатися на основі даних і виявляти закономірності, які можуть бути не одразу очевидними для аналітиків-людей.

Однак зазначені підходи машинного навчання мають і певні недоліки. До числа таких можна віднести високу вартість їх впровадження, потребу у великих обсягах даних, залежність від якості даних, упередженість даних або алгоритмів, брак прозорості у прийнятті рішень.

Зважаючи на недоліки існуючих підходів виявлення ФО, актуальності набуває завдання розробки ефективної системи автоматичного розпізнавання фразеологізмів в англійських текстах. Ефективність системи передбачає її надійність,

застосовність до обробки різноманітних структур речень і різних типів фразеологізмів, включаючи сталі вирази, ідіоми та словосполучення, а також мінімізацію недоліків, що характерні для систем, у яких реалізовані методи правил Rule-based approaches і машинного навчання.

Тому **метою статті** є опрацювання теоретичних засад для розробки ефективної системи автоматичного розпізнавання фразеологізмів в англійських текстах на основі використання сучасних інформаційних технологій, загалом, і штучного інтелекту, зокрема.

Виклад основного матеріалу дослідження.

Для досягнення визначеної мети вбачається за доцільне здійснити чітку постановку задачі розбудови системи автоматичного розпізнавання ФО в англійських текстах та спроектувати структуру відповідної системи.

Постановка задачі розбудови системи автоматичного розпізнавання ФО в англійських текстах

Вхідними даними системи має бути масив англійського тексту, що складається з речень, абзаців або великих текстових сегментів.

Результатом роботи системи має бути список фразеологізмів, знайдених у тексті, що представляє ідентифіковані фрази.

Система повинна опрацювати вхідні дані та автоматично генерувати результат.

Система повинна демонструвати наступні ключові можливості:

1. Комплексне сприйняття: система повинна бути здатна обробляти велику кількість тексту, що охоплює різні жанри, реєстри та лінгвістичні контексти. Вона повинна мати можливість сприймати ФО в складних структурах речень, таких як складносурядні або складнопідрядні, а також питальні та окличні речення. Система повинна бути здатна розпізнавати широкий спектр мовних одиниць, включно зі сталими виразами, ідіоматичними фразами, словосполученнями та іншими лексичними комбінаціями.

2. Точна ідентифікація: система повинна забезпечувати високу точність розпізнавання ФО в тексті. Вона повинна використовувати лінгвістичні моделі, синтаксичний аналіз, семантичну інформацію та контекстуальні підказки, щоб відрізнити ФО від звичайного мовного вжитку. Система повинна застосовувати методи, засновані на правилах і машинному навчанні, щоб підвищити точність і запам'ятовуваність ідентифікації ФО.

3. Масштабованість та ефективність: система повинна бути здатна ефективно обробляти великі обсяги тексту. Вона повинна оптимізувати обчис-

лювальні ресурси, щоб надавати своєчасні результати навіть при обробці великих корпусів або потоків тексту в реальному часі.

4. Подальший аналіз: система повинна створювати список ідентифікованих ФО як вихідний результат, що забезпечує чітке представлення розпізнаних фраз. Вихідні дані повинні бути придатними для подальшого аналізу, такого як моделювання мови, дослідження корпусної лінгвістики або інших завдань з обробки природної мови.

Проектування структури системи автоматичного розпізнавання ФО в англійських текстах

Теоретичною основою розроблюваної системи автоматичного розпізнавання ФО в англійських текстах має бути удосконалений метод автоматичного розпізнавання ФО, якому були б притаманні сильні сторони методів, заснованих на правилах і машинному навчанні. У подальшому називатимемо зазначений метод гібридним. Гібридний метод має забезпечити більш точне виявлення ФО в англійських текстах.

Основна ідея гібридного методу полягає у використанні заздалегідь визначених правил для категоризації фраз, які відповідають певним критеріям, і використання алгоритму машинного навчання для категоризації фраз, які їм не відповідають.

Загальний підхід до реалізації гібридного методу для категоризації фраз може бути таким (Mitchel, 2020: 80; Sarkar, 2018: 1310):

1. Категоризація на основі заздалегідь визначених правил (Rule-based categorization): першим кроком є визначення набору правил, які класифікують вилучені фрази на основі певних критеріїв, таких як їхня синтаксична структура, слова, які вони містять, або їхня частота в тексті. Ці правила можуть ґрунтуватися на лінгвістичних знаннях або знаннях предметної області і можуть бути зроблені вручну або напівавтоматично.

2. Категоризація на основі машинного навчання (Machine learning-based categorization): наступним кроком є навчання алгоритму машинного навчання для категоризації решти фраз, які не відповідають попередньо визначеним правилам. Це передбачає підготовку маркованого набору даних фраз і відповідних їм категорій, вилучення відповідних ознак з фраз і навчання алгоритму машинного навчання передбачати категорії нових фраз на основі цих ознак. З іншого боку, метод, заснований на машинному навчанні, використовує статистичні моделі та алгоритми для вивчення закономірностей і взаємозв'язків у даних і прогнозування нових екземплярів на основі цих вивче-

них закономірностей. Цей метод часто використовується, коли цільові фрази є більш варіативними і не можуть бути легко визначені за допомогою фіксованого набору правил або шаблонів.

3. Гібридна категоризація (Hybrid categorization): після розробки компонентів на основі правил і машинного навчання наступним кроком є об'єднання їх у гібридну систему категоризації. Це передбачає застосування заздалегідь визначених правил до виявлених фраз і їхню категоризацію на основі цих правил. Решта фраз передаються до компоненти, що реалізує машинне навчання, який класифікує їх на основі встановлених шаблонів і взаємозв'язків у даних. Гібридний підхід до розпізнавання та категоризації фразеологізмів може поєднувати ці два методи, використовуючи їхні сильні сторони та долаючи їхні недоліки. Наприклад, метод на основі правил можна використовувати для розпізнавання та категоризації більш фіксованих і легко встановлюваних типів фразеологізмів, тоді як метод на основі машинного навчання можна використовувати для розпізнавання та категоризації більш мінливих і залежних від контексту типів фразеологізмів.

4. Оцінка та доопрацювання (Evaluation and refinement): після того, як гібридна система категоризації розроблена, наступним кроком є оцінка її продуктивності та доопрацювання (за необхідності). Це передбачає використання набору тестових даних для вимірювання точності категоризації. За результатами оцінки може знадобитися доопрацювання або оновлення правил і алгоритму машинного навчання.

5. Розробка моделі на основі машинного навчання, яка може навчитися розпізнавати і класифікувати ФО на основі їхнього контексту та інших релевантних характеристик. Для цього можуть бути використані такі методи, як функціональна інженерія, коли відповідні лінгвістичні ознаки виділяються з тексту і використовуються

як вхідні дані для моделі машинного навчання, або глибоке навчання, коли нейронні мережі використовуються для вивчення репрезентацій тексту і здійснення прогнозів на основі цих репрезентацій.

Для реалізації гібридного методу важливою є попередня обробка досліджуваних текстів, яка має вирішальне значення для підготовки текстових даних до аналізу та зменшення шуму і варіативності.

Найпоширеніші методи попередньої обробки англійських текстів включають токенизацію, позначення частин мови та лематизацію. Токенизація передбачає розбиття тексту на окремі слова або токени, тоді як тегування частинами мови призначає кожному слову граматичну категорію, наприклад, іменник, дієслово або прикметник. Лематизація передбачає скорочення слів до їхньої базової форми, наприклад, скорочення «gunning» до «gun».

Для попередньої обробки англійських текстів у гібридному методі пропонується застосовувати такі загальні прийоми: перетворення в нижній регістр (Lowercasing); токенизація (Tokenization); видалення стоп-слів (Stop word removal); стеммінг або лематизація (Stemming/Lemmatization); видалення пунктуації; видалення чисел.

Наведені теоретичні основи гібридного методу дозволяють запропонувати структуру інформаційної системи автоматичного розпізнавання ФО у вигляді, що наведений на рис. 1.

Структура системи складається з таких модулів.

1. Модуль введення: модуль відповідає за отримання системою вхідних текстових даних англійською мовою. Вони можуть бути у вигляді текстових файлів, URL-адрес тощо.

2. Модуль попередньої обробки: модуль відповідає за попередню обробку тексту (токенизацію, тегування частин мови, лематизацію та синтаксичний розбір). Результатом роботи цього модуля



Рис. 1. Структура інформаційної системи автоматичного розпізнавання фразеологічних одиниць з потоком даних між різними модулями

буде попередньо оброблений корпус текстових даних.

3. Модуль ідентифікації ФО: модуль відповідає за ідентифікацію фразеологізмів у попередньо обробленому корпусі. В основу його роботи покладено гібридний метод.

4. Модуль класифікації ФО: модуль класифікує ідентифіковані фразеологізми за різними категоріями, такими як ідіоми, словосполучення, фразові дієслова тощо.

5. Модуль виведення: модуль відповідає за відображення результатів процесу ідентифікації та класифікації фразеологізмів. Результати можуть бути представлені у вигляді звіту, списку ідентифікованих одиниць або в будь-якій іншій формі, яка буде корисною для кінцевого користувача.

6. Модуль зворотного зв'язку: модуль дозволяє користувачу залишити відгук про ідентифіковані ФО.

Подальше проектування системи автоматичного розпізнавання ФО передбачає вибір апаратних засобів. Для вирішення цього завдання необхідно врахувати таке. Розробка нової системи з використанням моделей машинного навчання вимагає значного досвіду у відповідній сфері, в області обробки природної мови, а також доступу до великих обсягів даних для навчання та перевірки. Крім того, продуктивність моделей глибокого навчання залежить від різних факторів, таких як якість і розмір навчальних даних, вибір гіперпараметрів і архітектура моделі, які вимагають ретельного налаштування та оптимізації. Зважаючи на це, для розробки системи доцільно використати мову Python (Bird, 2019: 195; Thompson, 2023: 15).

Python має потужну бібліотеку під назвою NLTK (Natural Language Toolkit), яка надає різні інструменти та ресурси для обробки тексту, включаючи токенізацію, стеммінг, лематизацію, теґування та синтаксичний аналіз. NLTK також включає готові корпуси та моделі для різних завдань NLP, що робить його зручним інструментом для побудови систем обробки мови. Крім того, Python має широку підтримку фреймворків глибокого навчання, таких як TensorFlow, Keras та PyTorch, які можна використовувати для навчання та впровадження моделей машинного навчання. Ці готові моделі можна доопрацьовувати та адаптувати для виконання конкретних завдань, таких як автоматичне розпізнавання фразеологізмів.

Загалом поєднання простоти використання, потужних бібліотек і широкої підтримки фреймворків глибокого навчання робить Python ідеальним вибором для побудови системи автоматичного розпізнавання фразеологізмів.

Для автоматичного розпізнавання та категоризації ФО з використанням корпусу NLTK та бібліотеки Gensim у скрипті на Python можуть бути необхідні бібліотеки:

1. NLTK: для завдань обробки природної мови, таких як токенізація, стеммінг і теґування частин мови.

2. Gensim: для тематичного моделювання та обчислення подібності за допомогою вбудованих слів.

3. NumPy: для числових обчислень та операцій з масивами.

4. Pandas: для маніпулювання даними та аналізу.

5. Matplotlib: для візуалізації даних.

6. Scikit-learn: для алгоритмів машинного навчання, таких як кластеризація та класифікація.

Однією з готових моделей, яка може бути застосованою для реалізації досліджуваного проєкту, є модель Word2Vec. Word2Vec – це модель нейромережі, яка широко використовується для обробки природної мови, зокрема для класифікації текстів та пошуку інформації.

Висновки. За результатами проведеного дослідження можна зробити такі висновки:

на цей час однією з ключових проблем в обробці природної мови є розпізнавання фразеологізмів;

актуальним є завдання автоматизації виявлення фразеологізмів в англомовних текстах;

найбільш поширені існуючі підходи автоматизації виявлення фразеологізмів базуються на застосуванні правил Rule-based approaches та машинного навчання;

існуючі системи автоматизації виявлення фразеологізмів містять низку недоліків, які не дозволяють ефективно вирішувати завдання їх якісної ідентифікації;

актуальним завданням є розробка ефективної системи автоматичного розпізнавання фразеологізмів в англійських текстах, яка була б надійною, застосовною до обробки різноманітних структур речень і різних типів фразеологізмів, включаючи сталі вирази, ідіоми та словосполучення, а також містила б мінімальну кількість недоліків, що характерні для систем, у яких реалізовані методи правил Rule-based approaches і машинного навчання;

вирішення задачі розбудови системи автоматичного розпізнавання фразеологізмів в англомовних текстах, постановка якої здійснена в статті, може підвищити ефективність виявлення фразеологізмів в англомовних текстах;

теоретичною основою вирішення сформульованої у статті задачі може слугувати запропонований авторами гібридний метод, основна ідея якого полягає у використанні заздалегідь визначених правил для категоризації фраз, які відповідають певним критеріям, і використання алгоритму машинного навчання для категоризації фраз, які їм не відповідають;

ефективним засобом реалізації гібридного методу може бути система, структура якої містить наступні модулі: введення; попередньої обробки; ідентифікації фразеологізмів; класи-

фікації фразеологізмів; виведення; зворотного зв'язку;

для розробки системи автоматичного розпізнавання фразеологізмів в англomовних текстах, що реалізує гібридний метод, доцільно використати мову Python.

Напрямом подальших досліджень вбачається опрацювання алгоритмів побудови системи автоматичного розпізнавання фразеологічних одиниць та оцінювання моделі, що реалізує гібридний метод, а також безпосередньо програмна реалізація досліджуваної системи.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Басараба І. О. Фразеологія та фразеологізм у скопусі лінгвістичних студій. Науковий вісник ЧНУ: серія "Германська філологія". Чернівці, 2020. Вип. 823. С. 9-15.
2. Басараба І. О. Англomовні фразеологічні одиниці: проблема класифікації. «Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Філологія. Соціальні комунікації». Том 31 (70) № 4, 2020. С. 1-8.
3. Iryna Basaraba, Olha Lemeshko Correlation of cognitive abilities and translation skills of phraseological units. SKASE. Journal of Theoretical Linguistics. Košice, Slovak Republic, 2021. Volume 18, № 2. P. 34-50. URL: http://www.skase.sk/Volumes/JTL49/pdf_doc/03.pdf
4. Drury B. An Introduction to Natural Language Processing. 2015. 934 p.
5. Kurdi M. Z. Handbook of Natural Language Processing and Machine Translation. DARPA Global Autonomous Language Exploitation. 2010. P. 6225-6232.
6. Weikum M., Kramler G. Applied Natural Language Processing. Identification Investigation and Resolution. 2010. P. 7162-7166.
7. Dan Jurafsky James H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing Computational Linguistics and Speech Recognition. 2019. P. 175-195.
8. Thompson L. Advances in Natural Language Processing Techniques. (Part of the publication: Conference Materials) In: The Annual Conference of the Association for Computational Linguistics. Vancouver, 2023. P. 22-35.
9. Mitchell L. Foundations of Deep Learning: Building and Training Neural Networks. 2019. Vol. 31, P. 955-965.
10. Daniel Jurafsky, James H. Martin. Foundations of Statistical Natural Language Processing. 2019. P. 240.
11. Brown M. Information Extraction from Text using Machine Learning Methods. (Part of the publication: Conference Materials): International Conference on Data Mining. Tokyo, 2021. P. 95-110.
12. Bertrand Meyer. Object-Oriented Software Construction. Prentice Hall, 2017. P. 190.
13. Mitchell P. Word Sense Disambiguation using Neural Networks. (Part of the publication: Conference Materials): Meeting of the Association for Computational Linguistics. Seattle, 2022. P. 78-91.
14. Bird S., Klein E., Loper E. Natural Language Processing with Python. 2020. P. 288.
15. Sarkar D. A Practical Guide to Text Analytics with Python: Analyzing Text with Natural Language Processing. Foundations of deep learning fundamental concepts architectures and techniques related to deep learning algorithms. 2018. Vol. 26, P. 1307-1332.
16. Steven Bird, Ewan Klein, Edward Loper. Natural Language Processing with Python. Analyzing Text with the Natural Language Toolkit. O'Reilly Media. 2019. P. 191-218.

REFERENCES

1. Basaraba I. (2020) Anhlomovni frazeolohichni odynytsi: problema klasyfikatsii. [English phraseological units: the problem of classification]. Vcheni zapysky Tavriiskoho natsionalnoho universytetu imeni V. I. Vernadskoho. Seria: Filolohiia. Sotsialni komunikatsii. Scientific Notes of Vernadsky Taurida National University. Series: Philology. Social Communications, 31 (70) № 4. 1-8. [in Ukrainian].
2. Basaraba I. (2020) Frazeolohiia ta frazeolohizm u skopusi linhvistychnykh studii. [Phraseology and Phraseologism in the Scope of Linguistic Studies]. Naukovyi visnyk ChNU: serii "Hermanska filolohiia". Scientific Bulletin of the Chernivtsi National University: Germanic Philology Series, 823. 9-15. [in Ukrainian].
3. Bertrand Meyer (2017). Object-Oriented Software Construction. Prentice Hall. 190.
4. Bird S., Klein E., Loper E. (2020) Natural Language Processing with Python. 288.
5. Brown M. (2021) Information Extraction from Text using Machine Learning Methods. (Part of the publication: Conference Materials): International Conference on Data Mining, Tokyo. 95-110.
6. Dan Jurafsky, James H. Martin (2019) Speech and Language Processing: An Introduction to Natural Language Processing Computational Linguistics and Speech Recognition. 175-195.
7. Daniel Jurafsky, James H. Martin (2019) Foundations of Statistical Natural Language Processing. 240.

8. Drury B. (2015) An Introduction to Natural Language Processing. 934.
9. Iryna Basaraba, Olha Lemeshko (2021) Correlation of cognitive abilities and translation skills of phraseological units. *SKASE. Journal of Theoretical Linguistics*, Vol. 18, № 2. 34-50. URL: http://www.skase.sk/Volumes/JTL49/pdf_doc/03.pdf
10. Kurdi M. Z. (2010) Handbook of Natural Language Processing and Machine Translation. DARPA Global Autonomous Language Exploitation. 6225-6232.
11. Mitchell L. (2019) Foundations of Deep Learning: Building and Training Neural Networks. 955-965.
12. Mitchell P. (2022). Word Sense Disambiguation using Neural Networks. *(Part of the publication: Conference Materials): Meeting of the Association for Computational Linguistics*, Seattle. 78-91.
13. Sarkar D. (2018) A Practical Guide to Text Analytics with Python: Analyzing Text with Natural Language Processing. *Foundations of deep learning fundamental concepts architectures and techniques related to deep learning algorithms*. Vol. 26. 1307-1332.
14. Steven Bird, Ewan Klein, Edward Loper (2019) Natural Language Processing with Python. Analyzing Text with the Natural Language Toolkit. O'Reilly Media. 191-218.
15. Thompson L. (2023) Advances in Natural Language Processing Techniques. *(Part of the publication: Conference Materials) In: The Annual Conference of the Association for Computational Linguistics*, Vancouver. 22-35.
16. Weikum M., Kramler G. (2010) Applied Natural Language Processing. Identification Investigation and Resolution. 7162-7166.