УДК 004.8
DOI https://doi.org/10.15589/znp2024.1(494).16

ANALYSIS OF MODERN MODELS OF CONTEXT-AWARE QUESTION-ANSWERING SYSTEMS

АНАЛІЗ СУЧАСНИХ МОДЕЛЕЙ КОНТЕКСТ ОРІЄНТОВАНИХ СИСТЕМ НАДАННЯ ВІДПОВІДЕЙ

Mykhailo Yu. Vyshniak
mykhailo.vyshniak@nure.ua
ORCID: 0000-0002-3240-139X
Mykhailo Yu. Pyrozhenko
mykhailo.pyrozhenko@nure.ua
ORCID: 0000-0003-0785-1273

М. Ю. Вишняк,
канд. техн. наук, доцент
М. Ю. Пироженко,
аспірант

Kharkiv National University of Radio Electronics, Kharkiv
Харківський національний університет радіоелектроніки, м. Харків

Abstract. Question-answering systems are information systems designed for finding answers to questions asked by users in natural language. This discipline is one of the main areas of natural language processing and information retrieval, with practical applications in various fields, including culture, art, education, science, medicine, economics, and business. In the standard formulation of the answer retrieval problem, the input to a context-aware question-answering system is a context and a questions about the context. Reliable models and methods allow question-answering systems to provide more accurate responses to the users, save efforts, and enable quick decisions. *Purpose.* Analyzing the state of development of the methodological framework for the development of information systems, enriching the natural language processing tools for the Ukrainian language, and in particular, for building question-answering systems. *Method.* The study was conducted using a systematic approach, methods of abstraction, system analysis, comparison and synthesis. *Results.* We consider the individual components of context-aware question-answering systems designed to rank and extract answers, tools and models for their construction. *Scientific novelty.* The article solved an urgent scientific task, which is to review individual components, namely components for ranking and extracting answers, context-aware answer systems, and to substantiate models and tools for their construction. *Practical importance.* The ability to apply the research findings and conclusions to the development and implementation of answer systems; in the process of teaching natural language processing disciplines in higher education institutions; in writing natural language processing manuals; in conducting applied research.
Key words: question-answering systems; text mining; machine learning; methods; models; natural language processing; Ukrainian language.

Анотація. Системи надання відповіді – це інформаційні системи для пошуку відповідей на запитання, які користувач ставить природною мовою. Ця дисципліна є одним з головних напрямів обробки текстів природних мов та інформаційного пошуку, що має практичне застосування в різних сферах, зокрема, в культурі, мистецтві, освіті, науці, медицині, економіці та бізнесі. У стандартному формулюванні задачі пошуку відповідей, вхідними даними для контекст орієнтованої системи надання відповідей є контекст та запитання до контексту. Надійні моделі та методи дозволяють системі надання відповідей надати користувачу більш точні відповіді, економить його зусилля та дають можливість приймати швидкі рішення. *Мета.* Аналіз стану розвитку методологічної бази розробки інформаційних систем, збагачення інструментарію обробки природної мови для української мови, зокрема, для побудови систем надання відповідей. *Методика.* Дослідження проводилось з використанням системного підходу, методів абстрагування, системного аналізу, порівняння та синтезу. *Результати.* Нами розглянуті окремі компоненти контекстно-орієнтованих систем надання відповідей, що призначені для ранжування та вилучення відповідей, інструменти та моделі для їх побудови. *Наукова новизна.* Було вирішено актуальне наукове завдання, що полягає в огляді окремих компонентів, а саме компонентів для ранжування та вилучення відповідей, контекстно-орієнтованих систем надання відповідей, обгрунтуванні моделей та інструментів для їх побудови. *Практична значимість.* Полягає в можливості застосування наукових положень та висновків дослідження для розробки та впровадження систем надання відповідей; у процесі викладання дисциплін з обробки природних мов у вищих навчальних закладах; під час написання посібників з обробки природних мов; під час проведення прикладних досліджень.
Ключові слова: системи надання відповідей; текстова аналітика; машинне навчання; методи; моделі; обробка природної мови; українська мова.

ПОСТАНОВКА ЗАДАЧІ

Задача надання відповіді є однією з класичних задач з обробки природної мови. У її стандартному формулюванні, вхідними даними для контент орієнтованої системи надання відповідей є контекст та запитання до цього контексту. Контекстом може бути стаття, документ, текст доповіді або інша текстова інформація. Задачею є пошук та представлення кінцевої відповіді користувачу.

Ця задача є досить актуальною на даний час та потребує дослідження можливостей розв'язання для кількох мов, зокрема української, визначення відмінностей між мовами з точки зору доступності наборів даних. Для вирішення складних завдань з автоматизації надання відповідей важливим також є розгляд можливостей щодо використання різних моделей для української мови. Задача потребує розгляду сучасних моделей та методів, що дозволяють зменшити потребу в даних, зосередивши увагу на навчанні моделей для більш загальних завдань.

Відповідно до постановки задачі, у цьому дослідженні ми розглянули окремі компоненти систем надання відповідей та інструменти для їх створення, а також трансформаторні моделі. Окрім цього, були проаналізовані загальні проблеми побудови сучасних систем надання відповідей, в тому числі для української мови.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Останніми роками дослідження систем надання відповідей були зосереджені на покращенні точності відповідей цих систем шляхом підвищення їх здатності розуміти природну мову та опрацьовувати складні питання.

У праці [1] досліджуються різні підходи до розв'язання задачі надання відповідей для української та англійської мови, а також навчальні дані. У дослідженні розглядаються різні моделі машинного навчання для обробки природної мови, засновані на найсучасніших моделях, представлених дослідниками за останні кілька років. Автори порівнюють результати різних моделей, налаштованих на різних даних для підвищення точності прогнозів.

Авторами [2] представлено контекстно-орієнтовану модель відповідей на запитання для української мови на основі статей Вікіпедії з використанням моделі BERT [3]. Результатом роботи моделі є відповідь на питання. У якості даних для навчання та валідації використовується перекладений набір SQuAD [4].

Праця [5] демонструє використання нових моделей для систем надання відповідей, а також моделі для генерації запитань. Задача генерації запитань інтегрована з системою надання відповідей, щоб запропонувати релевантні запитання з вхідного контексту. Запропонована модель генерує запитання з контексту, введеного користувачем, та використовує різні моделі для отримання відповідей з контексту.

У статті [6] досліджуються системи надання відповідей, які мають на меті відповісти на запитання у формі природної мови на основі великих неструктурованих документів та описуються методи машинного розумінням тексту. Розглянуто останні тенденції у цій галузі та представлено сучасну архітектуру. Обговорено ключові проблеми розробки систем та пропонується аналіз найпоширеніших тестів.

Стаття [7] представляє комплексне дослідження алгоритмів ранжування текстових документів на основі контенту, алгоритми ранжування сторінок на основі структури посилань та гібридні алгоритми, які були запропоновані дослідниками в останні роки.

Слід зазначити також дослідження [8–10], що, значною мірою, вплинули на розвиток цієї галузі обробки текстів природної мови.

ВІДОКРЕМЛЕННЯ НЕВИРІШЕНИХ РАНІШЕ ЧАСТИН ЗАГАЛЬНОЇ ПРОБЛЕМИ

Попри важливість проблеми, для української мови вона ще не вирішена належним чином. Не було знайдено жодного публічного результату для української мови, окрім попередньо навченої моделі BERT, яка не є точно налаштованою через відсутність відповідних наборів даних для української мови [2]. Вона використовує набір зроблений шляхом машинного перекладу англомовного набору SQuAD. В ході дослідження не було знайдено жодного практичного прикладу використання україномовних систем надання відповідей для будь-якої галузі діяльності.

МЕТА ДОСЛІДЖЕННЯ

Дослідження було проведено з метою аналізу стану розвитку методологічної бази розробки інформаційних систем, збагачення інструментарію обробки природної мови для української мови, зокрема, для побудови системи надання відповідей.

Методи, об'єкт та предмет дослідження.

Дослідження проводилось з використанням системного підходу, методів абстрагування, системного аналізу, порівняння та синтезу. Предмет дослідження – поточний стан розробки систем надання відповідей. Об'єкт дослідження – моделі та методи система надання відповідей.

ОСНОВНИЙ МАТЕРІАЛ

Контент орієнтовні системи надання відповідей працюють за схемою, що включає в себе аналіз питання, пошук документів та уривків, ідентифікацію сутностей та представлення відповіді [11]. Відповідно до мети роботи, нами розглянуто компоненти ранжування та вилучення відповідей.

Для побудови систем надання відповідей потрібен компонент ранжування, оскільки дозволяє звузити кількість документів для подальшого опрацювання

системою. Компонент ранжування отримує на вхід запит природною мовою та колекцію документів, використовує методи інформаційного пошуку, щоб повернути уривки з коллекції документів, які потенційно можуть містити відповідь на запитання. Компонент можна розробити спираючись на відомі методи пошуку у тексті, наприклад TF-IDF [12] або Okapi BM25 [13]. Можна використовувати існуючи інструменти, що надають можливість пошуку за текстом у колекції документів, такі як ElasticSearch [14] або Lucene [15].

Іншою важливою складовою системи є компонент вилучення відповідей. Кожне питання разом з уривком пропускається через мовну модель для отримання відповідей. Для контекст орієнтованої системи надання відповідей є контекст та запитання до контексту. Надійні моделі та методи дозволяють системі надання відповідей надати користувачу більш точні відповіді, економить зусилля, дають можливість користувачу приймати швидкі рішення.

Неконтрольовані методи до завдання пошуку відповідей базуються на пошуку відповідно до ймовірності правильності відповіді за типами, шаблоном речення, ключовими словами в реченні, відстанню між словами відповіді та словами запиту, розділовими знаками, довжиною послідовності слів із запитання та іншими показниками. Система надання відповідей сортує кандидатів для відповідей відповідно до розрахованої ймовірності їх правильності. Контрольовані методи, своєю чергою, використовують маркований набір даних для навчання системи надання відповідей.

До класичних методів слід віднести логічні методи. Такі методи використовуються для вирішення завдань, оскільки логічні уявлення дозволяють показати більш абстрактні поняття, часові та логічні відношення. Відповідно до методів логістичної регресії [16] та машини опорних векторів [17], контекст розбивається на речення та додається до двійкового вектора. Цільове речення (відповідь) позначається як 1, а всі інші елементи – як 0. Після цього на розмічених даних навчається мультиплікативна логістична регресія або машина опорних векторів. Однією з переваг цього підходу є можливість додавання деяких ознак до моделі.

Перевагами класичних підходів є простота та висока прозорість моделей. Разом з тим, ефективність моделей на таких вибірках є невисокою. Результат погіршиться, якщо збільшити розмір контексту або поставити за мету отримати більш конкретну відповідь (словосполучення). Досліди для української мови показали результати гірші в порівнянні з такими для англійської мови, що пов'язано з особливостями граматики української мови, меншою кількістю текстових корпусів, наявністю дієвідмінювання та відмінювання слів та іншими особливостями [18].

При побудові україномовних моделей розробнику доводиться вирішити певні завдання. Для поширених мов, наприклад, англійської, існує безліч наборів даних, таких як SQuAD, WikiQA [19], SearchQA [20] та інших [21]. Однак, в інших випадках розробникам доводиться використовувати отримувати навчальні набори шляхом перекладу наборів з інших мов. Форма перекладу часто викривляє початковий задум тексту вносячи помилки.

Long Short-Term Memory (LSTM) – це рекурентна архітектура нейронної мережі, яка дозволяє будувати моделі від послідовності до послідовності [22]. Для LSTM моделі розміри вхідного та вихідного векторів не є фіксованими. На вході модель отримує контекст та запитання, на виході – оцінки слів з контексту. Щоб поєднати вектор контексту та вектор запитання додається шар уваги – точковий добуток векторів контексту та вектора запитання. Після цього результат точкового добутку перетворюється на ймовірність того, що це відповідь на запитання.

Text-to-text transfer transformer (T5) – це сучасна модель обробки природної мови, розроблена дослідниками Google Research [23]. Вона базується на архітектурі трансформатора, та, з того часу, стала основою для багатьох інших моделей, включаючи BERT та GPT [24].

T5 – це модель перенесення з тексту в текст, що означає, що вона може бути точно налаштована для виконання спектру завдань з розуміння природної мови, таких як класифікація тексту, переклад мови та надання відповідей на запитання. Однією з ключових інновацій T5 є префіксний підхід до навчання, коли модель налаштовується під конкретне завдання шляхом навчання за допомогою префікса, що додається до вхідного тексту. Прикладами таких префіксів є "binary classification", "generate question", "answer question" та інші.

T5 досягає найсучасніших результатів у широкому спектрі завдань та вважається дуже складною та потужною моделлю обробки природної мови, що демонструє високий рівень можливості тонкого налаштування та ефективний спосіб передачі знань.

Generative Pre-trained Transformer (GPT) – це тип мовної моделі та фреймворк для генеративного штучного інтелекту. Існує декілька версій GPT моделей, але не всі з них є з відкритим кодом та доступом до використання. GPT-2 є сучасною моделлю у задачах мовного моделювання. Особливістю цієї моделі є відсутність контексту. Ця модель може генерувати наступне слово лише на основі попереднього тексту. Отже, щоб вона відповіла на питання, ми повинні перефразувати речення з корпусу питань в розповідне речення з фразою для відповіді. Лише тоді GPT-2 генерує відповідь. З одного боку, це може бути перевагою, якщо немає конкретних даних для отримання відповіді. З іншого боку, точність буде низькою для завдань зі спеціальних галузей, якщо модель не

навчалася на даних з відповідних тем. Разом з тим, попереднє навчання на українських корпусах потребує значних ресурсів та часу.

Bidirectional Encoder Representations from Transformers (BERT) – це модель на основі трансформаторів, яка створена дослідниками з Google. Багатомовна модель BERT була побудована для 100 найпопулярніших мов у Вікіпедії. Процес навчання моделі BERT складається з двох етапів: попереднє навчання на текстових корпусах для задачі моделювання мови (вимоглива до обчислювальних ресурсів) та точне налаштування на наборах даних типу «запитання-відповідь».

Існує кілька модифікацій багатомовних моделей BERT, які відрізняються процесом налаштування, наприклад, налаштовані на різних наборах даних з урахуванням регістру та без урахування регістру текстів. Моделі BERT, налаштовані на перекладених даних, показують гарні результати також для мов з відмінним від латинського алфавітом. На сьогодні найкращий результат для переважної більшості мов, включаючи українську мову [2], дають моделі попередньо навчені на перекладеному наборі даних.

Створення моделі відповідей на основі BERT з нуля займає багато часу. Існують бібліотеки, які надають прості у використанні обгортки для попередньо навчених BERT моделей для завдань з пошуку відповідей на запитання. DeepPavlov [25] та HuggingFace [26] – дві популярні бібліотеки, які забезпечують підтримку попередньо навчених BERT моделей. Цей результат та результати для завдання XNLI [27] є досить вагомими аргументами на користь застосування багатомовної моделі BERT для побудови системи запитань-відповідей для української мови.

RoBERTa [28] – це простий, але дуже популярний наступник BERT, що покращує його завдяки ретельній та розумній оптимізації навчальних параметрів. Кілька простих змін у сукупності підвищують продуктивність RoBERTa та дають змогу перевершити BERT практично у всіх завданнях, для яких він був розроблений. Разом з публікацією RoBERTa виник інший трансформатор, XLNet [29]. Однак зміни XLNet, реалізувати значно складніше, ніж у RoBERTa, що робить його менш поширеним.

RoBERTa використовує ту саму архітектуру, що й BERT. Однак, на відміну від BERT, під час попереднього навчання вона навчається лише генерації пропущеної лексеми замість навчання передбаченню наступного речення.

DistilBERT [30] націлений на оптимізацію навчання завдяки зменшенню розміру та збільшенню швидкості BERT при спробі зберегти продуктивність. Зокрема, DistilBERT важить на 40% менше, ніж оригінальна BERT–модель, вона на 60% швидша за неї та зберігає 97% її функціональності.

DistilBERT використовує майже ту саму архітектуру, що й BERT, але тільки з 6 блоками шифратора замість 12. Ці блоки ініціалізуються простим взяттям 1 з кожних 2 перед навчених блоків шифрувальника BERT. Крім того, з DistilBERT видалено зіставлення токенів та функції пулінгу. На відміну від оригінального BERT, DistilBERT преднавчання відбувається тільки через моделювання мови за маскою.

ALBERT [31] був представлений приблизно в той самий час, що і DistilBERT. Як і DistilBERT, ALBERT зменшує розмір моделі BERT (у 18 разів менше параметрів), а також навчається в 1,7 раза швидше. Але на відміну від DistilBERT, у ALBERT немає компромісу продуктивності. Це відбувається через різницю в тому, як структуровані експерименти. DistilBERT підготовлений так, щоб використовувати BERT як учителя для процесу навчання. ALBERT, як і BERT, навчається з нуля. ALBERT перевершує всі попередні моделі, включно з BERT, RoBERTa, DistilBERT та XLNet. За допомогою цих методів зменшення параметрів, ALBERT досягає результатів з меншою архітектурою моделі. Щоб розмір прихованих шарів та розмірність вбудовування були різними, ALBERT деконструює матрицю вбудовування на 2 частини. Це збільшує розмір шару, не змінюючи фактичного розміру вбудовування. Після розкладання матриці, ALBERT додає лінійний або повно-зв'язний шар після завершення фази вбудовування.

BERT та ALBERT мають по 12 блоків кодування. В ALBERT ці блоки спільно використовують усі параметри. Це зменшує розмір параметрів у 12 разів, а також збільшує регуляризацію моделі (метод калібрування моделей машинного навчання задля уникнення перебору). ALBERT використовує техніку, за якої випадково обрані нейрони ігноруються під час навчання.

ОБГОВОРЕННЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

Наразі моделі на основі трансформаторів є найдосконалішими для більшості завдань обробки природної мови.

Яскравими прикладами моделей, побудованих на трансформаторній архітектурі, є BERT та похідні моделі, такі як RoBERTa, mBERT та ALBERT, разом з оптимізаціями, призначеними для додавання додаткових шарів, які дозволяють моделі вирішувати різні завдання. Початковим завданням при побудові системи, однак, є заповнення масок – навчання моделі шляхом заміни пропусків у навчальному тексті ймовірними пропущеними словами.

Іншими помітними прикладами трансформаторів є моделі сімейства GPT, GPT-2, GPT-3 та наступні. На відміну від BERT, GPT виводить послідовність слів, одне за одним. Кожне згенероване слово додається до вхідних даних, так що наступні слова семантично та граматично пов'язані з ним. Моделі GPT призначені для роботи з великою кількістю параметрів.

T5 є моделлю обробки текстів природної мови, попередньо навченою на багатозадачній суміші неконтрольованих та контрольованих завдань, для якої кожне завдання перетворюється у формат текст у текст.

При навчанні україномовних моделей систем надання відповідей розробник зітхається з проблемою відсутності оригінальних наборів даних. Окрім

цього, існує загальна проблема, що не залежить від мови. Якщо людина може формувати цілісну картину на основі декількох документів, для моделей таке з'єднання потребує додаткового налаштування. Втім існує практичний досвід навчання моделей на перекладеному наборі даних.

ВИСНОВКИ

Були розглянуті окремі компоненти контекстно-орієнтованих систем надання відповідей, що призначені для ранжування та вилучення відповідей, інструменти та моделі для їх побудови. Сфера обробки природної мови є доступною, але все ще значною мірою недослідженою, особливо коли йдеться про багатомовні моделі та обробку текстів загалом. Маючи на увазі завдання автоматичної відповіді на запитання, ми маємо прагнути розширити можливості сучасних моделей та методів як для української мови, так і загалом систем надання відповідей. Хоча для української мови вже є певні напрацювання, їх, безумовно, потрібно вдосконалити.

REFERENCES

[1] Dashenkov, D., Smelyakov, K., & Turuta, O. (2021). Methods of Multilanguage Question Answering. 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), 251–255. doi:10.1109/PICST54195.2021.9772145

[2] Tiutiunnyk, S., & Dyomkin, V. (2019). Context-Based Question-Answering System for the Ukrainian Language. Advances in Data Mining, Machine Learning, and Computer Vision. Proceedings, 81.

[3] Kenton, J. D. M. W. C., & Toutanova, L. K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL-HLT (pp. 4171-4186).

[4] Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016, November). SQuAD: 100,000+ Questions for Machine Comprehension of Text. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (pp. 2383-2392).

[5] Kumari, V., Keshari, S., Sharma, Y., & Goel, L. (2022, January). Context-based question answering system with suggested questions. In 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 368-373). IEEE.

[6] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. Multimedia Tools and Applications, 82(3), 3713–3744. doi:10.1007/s11042-022-13428-4

[7] Sharma, P., Yadav, D., & Garg, P. (2020). A systematic review on page ranking algorithms. International Journal of Information Technology, 12. doi:10.1007/s41870-020-00439-3

[8] Ali, I., & Yadav, D. (2019). Question reformulation based question answering environment model. International Journal of Information Technology. https://doi.org/10.1007/s41870-019-00332-8

[9] Mandar Suryavanshi. (2023). Question Answering System Approaches: A Review. International Journal of Advanced Research in Science, Communication and Technology, 288–296. https://doi.org/10.48175/ijarsct-8301

[10] Tahri, A., & Tibermacine, O. (2013). Dbpedia Based Factoid Question Answering System. International journal of Web & Semantic Technology, 4(3), 23–38. https://doi.org/10.5121/ijwest.2013.4303

[11] Vyshniak, M. Yu., & Pyrozhenko, M. Yu. (2022). Modern Trends In The Development Of Systems For The Natural Language Responses Formation When Searching In The Text Document Databases. Collection of Scientific Publications NUS, 489(2), 60–65. https://doi.org/10.15589/znp2022.2(489).9

[12] Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures. Information Processing & Management, 39(1), 45-65.

[13] Sparck Jones, K., Walker, S., & Robertson, S. E. (2000). A probabilistic model of information retrieval: development and comparative experiments. Information Processing & Management, 36(6), 779–808. https://doi.org/10.1016/s0306-4573(00)00015-7

[14] Elasticsearch, B. V. (2018). Elasticsearch. Software, version, 6(1).

[15] Azizan, A., Mohd Sanusi, N. I. N., Khairuddin, N., & Shafie, A. S. (2022). Lucene Search Engine Development: a beginner's experience. Mathematical Sciences and Informatics Journal (MIJ), 3(2), 80-92.

[16] Sperandei, S. (2014). Understanding logistic regression analysis. Biochemia Medica, 24 (1), 12-18. https://doi.org/10.11613/BM.2014.003

[17] Kwok, J. T.-Y. (1998). Automated Text Categorization Using Support Vector Machine. International Conference on Neural Information Processing. Retrieved from https://api.semanticscholar.org/CorpusID:18431174

[18] Manning, C. A., Luckyj, G., & Rudnyckyj, J. (1950). A Modern Ukrainian Grammar. The Modern Language Journal, 34(1), 80. https://doi.org/10.2307/318971

[19] Yang, Y., Yih, W.-t., & Meek, C. (2015). WikiQA: A Challenge Dataset for Open-Domain Question Answering. У Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics. https://doi.org/10.18653/v1/d15-1237

[20] Dunn, M., Sagun, L., Higgins, M., Güney, V. U., Cirik, V., & Cho, K. (2017). SearchQA: A New Q&A Dataset Augmented with Context from a Search Engine. CoRR, abs/1704.05179. Retrieved from http://arxiv.org/abs/1704.05179

[21] Vyshniak M. Yu., & Pyrozhenko M. Yu. (2023). Publicly available datasets and metrics to advance research on question-answering systems. Taurida Scientific Herald. Series: Technical Sciences. (1), 13-24.

[22] Wang, S., & Jiang, J. (2016, June). Learning Natural Language Inference with LSTM. In K. Knight, A. Nenkova, & O. Rambow (Eds.), Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 1442–1451). doi:10.18653/v1/N16-1170

[23] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., … Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. J. Mach. Learn. Res., 21, 140:1-140:67. Retrieved from http://jmlr.org/papers/v21/20-074.html

[24] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

[25] Burtsev, M., Seliverstov, A., Airapetyan, R., Arkhipov, M., Baymurzina, D., Bushkov, N., … Zaynutdinov, M. (2018, July). DeepPavlov: Open-Source Library for Dialogue Systems. In F. Liu & T. Solorio (Eds.), Proceedings of ACL 2018, System Demonstrations (pp. 122–127). doi:10.18653/v1/P18-4021

[26] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., … Rush, A. (2020, October). Transformers: State-of-the-Art Natural Language Processing. In Q. Liu & D. Schlangen (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 38–45). doi:10.18653/v1/2020.emnlp-demos.6

[27] Conneau, A., Rinott, R., Lample, G., Williams, A., Bowman, S. R., Schwenk, H., & Stoyanov, V. (2018). XNLI: Evaluating Cross-lingual Sentence Representations. In E. Riloff, D. Chiang, J. Hockenmaier, & J. Tsujii (Eds.), Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018 (pp. 2475–2485). doi:10.18653/V1/D18-1269

[28] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., … Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. CoRR, abs/1907.11692. Retrieved from http://arxiv.org/abs/1907.11692

[29] Yang, Z., Dai, Z., Yang, Y., Carbonell, J. G., Salakhutdinov, R., & Le, Q. V. (2019). XLNet: Generalized Autoregressive Pretraining for Language Understanding. CoRR, abs/1906.08237. Retrieved from http://arxiv.org/abs/1906.08237

[30] Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. CoRR, abs/1910.01108. Retrieved from http://arxiv.org/abs/1910.01108

[31] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020. Retrieved from https://openreview.net/forum?id=H1eA7AEtvS

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] Dashenkov, D., Smelyakov, K., & Turuta, O. (2021). Methods of Multilanguage Question Answering. 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), 251–255. doi:10.1109/PICST54195.2021.9772145

[2] Tiutiunnyk, S., & Dyomkin, V. (2019). Context-Based Question-Answering System for the Ukrainian Language. Advances in Data Mining, Machine Learning, and Computer Vision. Proceedings, 81.

[3] Kenton, J. D. M. W. C., & Toutanova, L. K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL-HLT (pp. 4171–4186).

[4] Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016, November). SQuAD: 100,000+ Questions for Machine Comprehension of Text. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (pp. 2383–2392).

[5] Kumari, V., Keshari, S., Sharma, Y., & Goel, L. (2022, January). Context-based question answering system with suggested questions. In 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 368–373). IEEE.

[6] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. Multimedia Tools and Applications, 82(3), 3713–3744. doi:10.1007/s11042-022-13428-4

[7] Sharma, P., Yadav, D., & Garg, P. (2020). A systematic review on page ranking algorithms. International Journal of Information Technology, 12. doi:10.1007/s41870-020-00439-3

[8] Ali, I., & Yadav, D. (2019). Question reformulation based question answering environment model. International Journal of Information Technology. https://doi.org/10.1007/s41870-019-00332-8

[9] Mandar Suryavanshi. (2023). Question Answering System Approaches: A Review. International Journal of Advanced Research in Science, Communication and Technology, 288–296. https://doi.org/10.48175/ijarsct-8301

[10] Tahri, A., & Tibermacine, O. (2013). Dbpedia Based Factoid Question Answering System. International journal of Web & Semantic Technology, 4(3), 23–38. https://doi.org/10.5121/ijwest.2013.4303

[11] Vyshniak, M. Yu., & Pyrozhenko, M. Yu. (2022). Modern Trends In The Development Of Systems For The Natural Language Responses Formation When Searching In The Text Document Databases. Collection of Scientific Publications NUS, 489(2), 60–65. https://doi.org/10.15589/znp2022.2(489).9

[12] Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures. Information Processing & Management, 39(1), 45–65.

[13] Sparck Jones, K., Walker, S., & Robertson, S. E. (2000). A probabilistic model of information retrieval: development and comparative experiments. Information Processing & Management, 36(6), 779–808. https://doi.org/10.1016/s0306-4573(00)00015-7

[14] Elasticsearch, B. V. (2018). Elasticsearch. Software, version, 6(1).

[15] Azizan, A., Mohd Sanusi, N. I. N., Khairuddin, N., & Shafie, A. S. (2022). Lucene Search Engine Development: a beginner's experience. Mathematical Sciences and Informatics Journal (MIJ), 3(2), 80–92.

[16] Sperandei, S. (2014). Understanding logistic regression analysis. Biochemia Medica, 24 (1), 12–18. https://doi.org/10.11613/BM.2014.003

[17] Kwok, J. T.-Y. (1998). Automated Text Categorization Using Support Vector Machine. International Conference on Neural Information Processing.

[18] Manning, C. A., Luckyj, G., & Rudnyckyj, J. (1950). A Modern Ukrainian Grammar. The Modern Language Journal, 34(1), 80. https://doi.org/10.2307/318971

[19] Yang, Y., Yih, W.-t., & Meek, C. (2015). WikiQA: A Challenge Dataset for Open-Domain Question Answering. У Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics. https://doi.org/10.18653/v1/d15-1237

[20] Dunn, M., Sagun, L., Higgins, M., Güney, V. U., Cirik, V., & Cho, K. (2017). SearchQA: A New Q&A Dataset Augmented with Context from a Search Engine.

[21] Вишняк, М. Ю., & Пироженко, М. Ю. (2023). Загальнодоступні набори даних та метрики для сприяння дослідженням систем надання відповідей. Таврійський науковий вісник. Серія: Технічні науки, (1), 13–24.

[22] Wang, S., & Jiang, J. (2016, June). Learning Natural Language Inference with LSTM. In K. Knight, A. Nenkova, & O. Rambow (Eds.), Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 1442–1451). doi:10.18653/v1/N16-1170

[23] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., … Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. J. Mach. Learn. Res., 21, 140:1-140:67.

[24] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

[25] Burtsev, M., Seliverstov, A., Airapetyan, R., Arkhipov, M., Baymurzina, D., Bushkov, N., … Zaynutdinov, M. (2018, July). DeepPavlov: Open-Source Library for Dialogue Systems. In F. Liu & T. Solorio (Eds.), Proceedings of ACL 2018, System Demonstrations (pp. 122–127). doi:10.18653/v1/P18-4021

[26] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., … Rush, A. (2020, October). Transformers: State-of-the-Art Natural Language Processing. In Q. Liu & D. Schlangen (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 38–45). doi:10.18653/v1/2020.emnlp-demos.6

[27] Conneau, A., Rinott, R., Lample, G., Williams, A., Bowman, S. R., Schwenk, H., & Stoyanov, V. (2018). XNLI: Evaluating Cross-lingual Sentence Representations. In E. Riloff, D. Chiang, J. Hockenmaier, & J. Tsujii (Eds.), Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018 (pp. 2475–2485). doi:10.18653/V1/D18-1269

[28] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., … Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach.

[29] Yang, Z., Dai, Z., Yang, Y., Carbonell, J. G., Salakhutdinov, R., & Le, Q. V. (2019). XLNet: Generalized Autoregressive Pretraining for Language Understanding.

[30] Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.

[31] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020.

Дата надходження статті до редакції: 13.02.2024
Дата затвердження статті до друку: 26.02.2024