

Облікова картка ДіР



I. Загальні відомості

Державний обліковий номер: 0223U004521

Державний реєстраційний номер: 0123U104013

Особливі позначки: відкрита

Дата реєстрації: 14-11-2023

II. Етап виконання ДіР

Номер етапу: 1

Назва етапу: Розроблення теоретико-методичного забезпечення та формування практичних рекомендацій щодо попередньої обробки та анотування текстової інформації

Початок етапу: 10.2023

Закінчення етапу: 10.2023

Вид звітнього документа: Остаточний звіт

III. Відомості про виконавця ДіР

Повне найменування юридичної особи (або ПІБ фізичної особи): Харківський національний економічний університет імені Семена Кузнеця

Код за ЄДРПОУ: 02071211

Місцезнаходження: проспект Науки, буд. 9-а, м. Харків, Харківський р-н., Харківська обл., 61166, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Розмір організації:

Телефон: 380577020304, 380577020717

IV. Відомості про співвиконавців ДіР

V. Відомості про замовника ДіР

Повне найменування юридичної особи (або ПІБ фізичної особи): Товариство з обмеженою відповідальністю "Друкарня Мадрид"

Код за ЄДРПОУ: 36034689

Місцезнаходження: вул. Гуданова, буд. 18, м. Харків, Харківський р-н., Харківська обл., 61024, Україна

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

Розмір організації:

Телефон: 380800336762

VI. Джерела, напрями та обсяги фінансування ДіР

Підстава для проведення ДіР: 52 - договір з вітчизняною організацією (органами місцевої ради, фондом, асоціацією, концерном тощо)

Напрямок фінансування: 2.2 - прикладні дослідження і розробки

Джерела фінансування

7722 - кошти підприємств, установ, організацій України

Фактичний обсяг фінансування (тис. грн.): 19.000
--

VII. Відомості про ДіР

Назва роботи українською:

Розроблення методів обробки та анотування текстової інформації

Назва роботи англійською:

Development of methods of processing and annotating textual information

Реферат українською:

Розглянуто важливі аспекти аналізу та розробки перспективних методів обробки текстів та їх практичного застосування у видавництвах та на поліграфічних підприємствах. Наведено результати розроблення та практичного дослідження методу нейромережевої класифікації українськомовних текстових новин. Точність побудованої нейронної мережі для класифікації новин за повним текстом становить на тестовій вибірці близько 84%. В якості словника було використано повний набір унікальних слів, які зустрічаються в тренувальному наборі. Наведено результати розроблення та практичного дослідження методу класифікації схожості торгівельних марок за текстовим описом. Запропонований метод дозволив безпомилково класифікувати торгівельні марки, які підлягали тестуванню (з застосуванням нейронної мережі doc2vec). В звіті запропоновано підхід до автоматичного анотування текстів з використанням методу зв'язних словоформ. Основою цього методу є модифікований алгоритм Гінзбурга, що дозволяє виділяти ключові слова тексту, а також зв'язані словосполучення, які мають схожий зміст і загальну тематику. Програмно реалізований метод зв'язних словоформ включено до складу програмного модуля Booker-M, що дозволяє здійснювати анотування текстових документів по категоріям. Отримані анотації є досить точним описом основного сенсу аналізованого тексту.

Реферат англійською:

Considered important aspects of analysis and development of promising methods of text processing and their practical application in publishing houses and printing enterprises. The results of the development and practical research of the method of neural network classification of Ukrainian-language text news are given. The accuracy of the constructed neural network for full-text news classification is about 84% on the test sample. A complete set of unique words found in the training set was used as a dictionary. The results of the development and practical research of the method of classifying the similarity of trademarks based on the textual description are presented. The proposed method made it possible to correctly classify the tested trademarks (using the doc2vec neural network) The report proposes an approach to automatic annotation of texts using the method of connected word forms. The basis of this method is the modified Ginzburg algorithm, which allows you to highlight the key words of the text, as well as related phrases that have a similar meaning and a common theme. The software-implemented method of connected word forms is included in the Booker-M software module, which allows annotating text documents by category. The resulting annotations are a fairly accurate description of the main meaning of the analyzed text.

Індекс УДК: 004.93'1; 004.932

Коди тематичних рубрик: 28.23.15

Керівники роботи

Власне Прізвище Ім'я По-батькові: Удовенко Сергій Григорович

Науковий ступінь: д.т.н.

Вчене звання: професор

Ідентифікатор ORCID ID:

Додаткова інформація:

VIII. Наукова (науково-технічна) продукція (НТП)

IX. Бібліографічний опис

X. Заключні відомості

Керівник юридичної особи

Пономаренко Володимир Степанович
д.е.н.

Перелік осіб-виконавців

Луговий Богдан Владиславович

Тесленко Олег Володимирович

(к. т. н., доц.)

Удовенко Сергій Григорович

(д.т.н., професор, 05.13.03)

Відповідальний за підготовку

Шелкунова Наталія Леонідівна

облікових документів

Телефон

+38 (057) 702-06-70

Реєстратор

**Керівник відділу УкрІНТЕІ, що є
відповідальним за реєстрацію наукової
діяльності**



Юрченко Тетяна Анатоліївна