Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0824U002552

Особливі позначки: відкрита

Дата реєстрації: 12-07-2024

Статус: Наказ про видачу диплома

Реквізити наказу МОН / наказу закладу: № НСВС_71_24 від 24.09.2024

II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Міщенко Людмила Дмитрівна
2. Liudmyla Mishchenko

Кваліфікація:

Ідентифікатор ORCID ID: 0000-0002-5101-2517

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 123

Назва наукової спеціальності: Комп'ютерна інженерія

Галузь / галузі знань: інформаційні технології

Освітньо-наукова програма зі спеціальності: Комп'ютерна інженерія

Дата захисту: 04-09-2024

Спеціальність за освітою: Комп'ютерна інженерія

Місце роботи здобувача: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): ДФ 26.002.184; ID 6516

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

V. Відомості про дисертацію

Мова дисертації: Українська

Коди тематичних рубрик: 20.53.19, 20.55, 20.55.01, 20.55.03, 20.59, 20.60

Тема дисертації:

1. МЕТОД РОЗПІЗНАВАННЯ ФЕЙКОВИХ НОВИН У МЕРЕЖІ ІНТЕРНЕТ НА ОСНОВІ ОБРОБКИ ПРИРОДНОЇ МОВИ

2. THE METHOD OF RECOGNIZING FAKE NEWS ON THE INTERNET BASED ON NATURAL LANGUAGE PROCESSING

Реферат:

1. У дисертаційній роботі вирішено актуальну науково-прикладну проблему оперативного розпізнавання та маркування фейкових повідомлень-новин за обмеженою апріорною інформацією для рішення задачі оперативного розпізнавання фейків та нейтралізації їх негативного впливу шляхом маркування в умовах значної кількості, щільності потоку та обмеженості за змістом окремого повідомлення. Проведено аналіз заходів інформаційно-психологічного впливу (ІПсВ). Визначено, що об’єктами ІПсВ можуть бути громадяни; цілями ІПсВ – зміна думок, настроїв, дій. Проаналізовано множини способів реалізації ІПсВ та визначено, що найпоширенішими є розповсюдження оманливої інформації через фейкові новини. Досліджено, що

засобами реалізації фейків є короткі інформаційні повідомлення, що розповсюджуються електронними медіа через сайти новин, соціальні мережі. Домінанта форми фейків належить повідомленням природньою мовою у форматі новин. Проаналізовано, що більшість традиційних підходів базується на комплексному застосування головних груп індикаторів, що дає лише апостеріорний результат на рівні спростування та не відповідає принципам оперативності – реальності масштабу часу. Визначено, що головним показником якості протидії фейкам є оперативність розпізнавання та маркування. Проведено аналіз вимог до оперативності розпізнавання фейків в умовах їх значної кількості, щільності потоку та обмеженості за змістом окремого повідомлення та обрано спосіб виявлення фейків як аналіз контенту повідомлення. Визначено конкретику процесів для оперативного та комплексного вирішення задач розпізнавання та маркування фейкових новин. Проаналізовано використання частотного аналізу токенів тексту коротких інформаційних повідомлень та формування словнику фейків на динамічно-оновлювальній множині повідомлень. Зроблено аналіз статистичного критерія Байєса, адаптованого до лінгвістичної стилістики повідомлень прийняття рішення про фейк та його маркування та доцільність застосування методів машинного навчання без вчителя. Проведено аналіз результатів використання методу оперативного розпізнавання фейкових новин за обмеженою апріорною інформацією, який базується на комплексному синергетичному поєднанні процесів обробки природної мови: частотного аналізу токенів тексту коротких інформаційних повідомлень; удосконаленому спосіб контент-аналізу фейкових новин; бінарної класифікації повідомлень за удосконаленим Naive Bayes класифікатором та функції ранжування BM25. Метод вирізняється реалізацією процесу навчання на динамічно поновлюваній множині коротких інформаційних повідомлень з достовірних джерел. Це забезпечує оперативне розпізнавання фейків із точністю у діапазоні від 85% до 93% для їх бінарного маркування за змістом окремого повідомлення. У порівнянні з відомою стратегією NLP, яка базується на використанні статистичної міри TF-IDF без інтеграції Naive Bayes класифікатора, відсоток точності класифікації тексту становить від 80% до 90%. Отже, запропонований метод дозволив підвищити ефективність розпізнавання фейкових новин в середньому на 2,5%. Удосконалено спосіб бінарної класифікації та маркування фейкових новин, якій базується на використанні Naive Bayes класифікатора та функції ранжування BM25 та відрізняється адаптивним вибором параметрів функції ранжування за експериментальними даними, отриманими як результат частотного аналізу новин з достовірних джерел. Це дозволяє підвищити точність класифікації текстових даних на 14% на динамічно оновлюваній множині коротких інформаційних повідомлень без втрати швидкодії у порівнянні з відомими реалізаціями етапу класифікації фейкових новин в мережі Інтернет. Удосконалено спосіб контент-аналізу фейкових новин, що базується на використання схеми навчання без вчителя. Спосіб вирізняється використанням динамічно змінюваного за переліком джерел і змістом контенту DataSet масиву з достовірних джерел, за яким формується словник для розпізнавання фейків. Це забезпечує оперативне формування ознак фейків з урахуванням динаміки змін їх стилю та сфери. Результати проведених експериментів показали потенціал із високою стійкістю формувати бінарну оцінку із маркування фейків для споживачів з критичними умовами доступності. Також можливість оперативного розпізнавання і бінарного маркування фейків на пристроях із низькою продуктивністю, енергоспоживанням та із тимчасовою відсутністю доступу до глобальних інформаційних мереж. Розроблене комплексне серверне програмне забезпечення дозволить реалізувати моделі NLP для застосування запропонованого методу. Результат роботи комплексу – оперативне маркування фейкових повідомлень користувачам мобільних пристроїв. ПЗ можна інтегрувати в інформаційно-аналітичні ресурси та використати для реалізації хмарних сервісів розпізнавання фейкових новин у мережі Інтернет. Отже, актуальність тематики дисертаційного дослідження полягає у вирішенні задачі оперативного розпізнавання фейкових повідомлень – нових в умовах їх значної кількості, щільності потоку та обмеженості за змістом окремого повідомлення.

2. The dissertation addresses the urgent scientific and practical problem of promptly recognizing and labeling fake news messages using limited a priori information. This is essential for mitigating the negative impact of fake news by marking them effectively, especially given the high volume, dense flow, and limited content of individual messages. The study analyzed information and psychological influence (IPI) measures, identifying citizens as IPI

targets and the goal of IPI as changing opinions, moods, and actions. The research found that the most common IPI method is spreading deceptive information via fake news, primarily through short messages on news websites and social media, usually in a natural language news format. Traditional approaches to fake news detection rely on complex indicator groups, providing only post-facto results and lacking real-time operationality. The study identified speed of recognition and labeling as the key quality indicators in countering fake news. It analyzed requirements for rapid fake news detection under conditions of high volume, dense flow, and limited message content, choosing content analysis as the detection method. The study detailed processes for comprehensive and rapid fake news detection and labeling, analyzing the use of frequency analysis of text tokens from short messages and the creation of a dynamically updated fake news dictionary. The Bayesian statistical criterion, adapted for the linguistic style of messages, and the feasibility of unsupervised machine learning methods were evaluated. The method proposed combines natural language processing (NLP) techniques: frequency analysis of text tokens, an enhanced content analysis method, binary message classification using an improved Naive Bayes classifier, and the BM25 ranking function. This approach ensures rapid fake news recognition with 85-93% accuracy for binary content-based labeling, outperforming the traditional TF-IDF method integrated with Naive Bayes, which achieves 80-90% accuracy, thus improving fake news detection efficiency by 2.5%. The improved binary classification and labeling method using the Naive Bayes classifier and BM25 ranking function features adaptive parameter selection based on experimental frequency analysis data from reliable sources. This method increases text data classification accuracy by 14% on a dynamically updated set of short messages without compromising speed compared to known Internet fake news classification implementations. The content analysis method was enhanced using an unsupervised learning scheme, utilizing a dynamically changing dataset from reliable sources to form a fake news recognition dictionary. This ensures the timely formation of fake news features, considering dynamic changes in their style and scope. Experimental results showed the potential for consistently forming binary evaluations and labeling fakes for users under critical access conditions. The method also allows for prompt recognition and binary labeling of fake news on low-performance devices with low power consumption and temporary lack of access to global networks. The developed server software supports the implementation of NLP models for the proposed method, providing prompt fake news labeling for mobile users. This software can be integrated into information-analytical resources and used to implement cloud services for Internet fake news recognition. In summary, the dissertation's relevance lies in solving the problem of rapid fake news message recognition under conditions of high volume, dense flow, and limited individual message content.

Державний реєстраційний номер ДіР:

Пріоритетний напрям розвитку науки і техніки: Фундаментальні наукові дослідження з найбільш важливих проблем розвитку науково-технічного, соціально-економічного, суспільно-політичного, людського потенціалу для забезпечення конкурентоспроможності України у світі та сталого розвитку суспільства і держави

Стратегічний пріоритетний напрям інноваційної діяльності: Розвиток сучасних інформаційних, комунікаційних технологій, робототехніки

Підсумки дослідження: Теоретичне узагальнення і вирішення важливої наукової проблеми

Публікації:

- Mishchenko L., Klymenko, I. (2023). Recognizing fake news based on natural language processing using the BM25 algorithm with fine-tuned parameters. Eastern-European Journal of Enterprise Technologies, 6 (2 (126)), с. 33–40. DOI: https://doi.org/10.15587/1729-4061.2023.293513.
- L.D. Mishchenko, I. A. Klymenko, A method of accelerated fake news recognition based on natural language processing and removal of vowels in words. Збірник наукових праць «Проблеми інформатизації та управління» 1(73)/2023, 2023-04-28. с. 39-44. ISSN 2073-4751. DOI: https://doi.org/10.18372/2073-4751.73.17643.

- L. Mishchenko, I. Klymenko, METHOD FOR DETECTING FAKE NEWS THROUGH WRITING STYLE. Технічні науки та технології, 4 (34). Чернігів, Україна. DOI: https://doi.org/10.25140/2411-5363-2023-4(34).
- Л. Міщенко, І. Клименко. РОЗПІЗНАВАННЯ ФЕЙКОВИХ НОВИН ІЗ ВИКОРИСТАННЯМ ОБРОБКИ ПРИРОДНОЇ МОВИ И АРХІТЕКТУРИ З НИЗЬКИМ ЕНЕРГОСПОЖИВАННЯМ ДЛЯ ПЕРИФЕРІИНИХ ОБЧИСЛЕНЬ. Збірник наукових праць «Проблеми інформатизації та управління» 4(76)/2023. с. 49-57. DOI: https://doi.org/10.18372/2073- 4751.76.18241
- L. Mishchenko, I. Klymenko, METHOD FOR DETECTING FAKE NEWS BASED ON NATURAL LANGUAGE PROCESSING. The VI International Scientific and Practical Conference «Modern ways of solving the problems of science in the world», February 13 – 15, 2023. Warsaw, Poland. p. 375- 378.
- Людмила Міщенко. СПОСІБ РОЗПІЗНАВАННЯ ФЕЙКОВИХ НОВИН. Science, society, education: topical issues and development prospects: V International Scientific and Practical Conference Kharkiv, Ukraine. 12-14 April 2020
- L. Mishchenko, I. Klymenko, V. Tkachenko. The fake news recognition method based on Naïve Bayes with improved TF-IDF algorithm. Mathematical Modeling and Simulation of Systems (MODS'2023). Chernihiv, Ukraine, November 13-15, 2023.

Наукова (науково-технічна) продукція: технології; методи, теорії, гіпотези; програмні продукти, програмно-технологічна документація

Соціально-економічна спрямованість: підтримка демократичних процесів та підвищення рівня медіаграмотності; зменшення витрат на боротьбу з дезінформацією

Охоронні документи на ОПІВ:

Впровадження результатів дисертації: Впроваджено

Зв'язок з науковими темами: № 0121U108261; №0113U002314

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Клименко Ірина Анатоліївна

2. Iryna A. Klymenko

Кваліфікація: д. т. н., доц., 05.13.05

Ідентифікатор ORCID ID: 0000-0001-5345-8806

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Дорош Марія Сергіївна

2. Mariia S. Dorosh

Кваліфікація: д. т. н., доц., 05.13.22

Ідентифікатор ORCID ID: 0000-0001-6537-9857

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Чернігівська політехніка"

Код за ЄДРПОУ: 05460798

Місцезнаходження: вул. Шевченка, буд. 95, Чернігів, Чернігівський р-н., 14035, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Висоцька Вікторія Анатоліївна

2. Vysotska Viktoriia A.

Кваліфікація: д. т. н., доцент, 10.02.21

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Львівська політехніка"

Код за ЄДРПОУ: 02071010

Місцезнаходження: вул. Степана Бандери, буд. 12, Львів, 79013, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Олійник Юрій Олександрович

2. Yurii Oliinyk

Кваліфікація: к. т. н., доц., 05.13.06

Ідентифікатор ORCID ID: 0000-0002-7408-4927

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Писарчук Олексій Олександрович
2. Oleksii Pysarchuk

Кваліфікація: д.т.н., професор, 05.22.13

Ідентифікатор ORCID ID: 0000-0001-5271-0248

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VIII. Заключні відомості

Власне Прізвище Ім'я По-батькові голови ради	Новотарський Михайло Анатолійович
Власне Прізвище Ім'я По-батькові головуючого на засіданні	Новотарський Михайло Анатолійович
Відповідальний за підготовку облікових документів	Міщенко Людмила Дмитрівна
Реєстратор	УкрІНТЕІ
Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності	Юрченко Тетяна Анатоліївна