Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0824U001387

Особливі позначки: відкрита

Дата реєстрації: 01-04-2024

Статус: Захищена

Реквізити наказу МОН / наказу закладу:

II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Дмитренко Олег Олександрович
2. Dmytrenko Oleg Oleksandrovych

Кваліфікація:

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: доктор філософії

Аспірантура/Докторантура: ні

Шифр наукової спеціальності: 122

Назва наукової спеціальності: Комп'ютерні науки

Галузь / галузі знань: інформаційні технології

Освітньо-наукова програма зі спеціальності: Комп'ютерні науки

Дата захисту: 22-04-2024

Спеціальність за освітою: Прикладна математика

Місце роботи здобувача:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): 5098

Повне найменування юридичної особи: Інститут проблем реєстрації інформації Національної академії наук України

Код за ЄДРПОУ: 03771755

Місцезнаходження: вул. М. Шпака, буд. 2, Київ, 03113, Україна

Форма власності: Державна

Сфера управління: Національна академія наук України

Ідентифікатор ROR:

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

Повне найменування юридичної особи: Інститут проблем реєстрації інформації Національної академії наук України

Код за ЄДРПОУ: 03771755

Місцезнаходження: вул. М. Шпака, буд. 2, Київ, 03113, Україна

Форма власності: Державна

Сфера управління: Національна академія наук України

Ідентифікатор ROR:

V. Відомості про дисертацію

Мова дисертації: Українська

Коди тематичних рубрик: 20.19.27

Тема дисертації:

1. Інформаційні технології формування та аналізу мережевих моделей предметних галузей на основі лінгвостатистичного підходу

2. Information technologies for formation and analysis of network models of subject domains based on linguostatistical approach

Реферат/анотація:

1. Дмитренко О.О. Інформаційні технології формування та аналізу мережевих моделей предметних галузей на основі лігвостатистичного підходу. – Кваліфікаційна наукова праця на правах рукопису. Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки». – Інститут проблем реєстрації інформації, НАН України, Київ, 2024. У дисертаційній роботі представлені результати проведених здобувачем досліджень, які виконують актуальне наукове завдання формування та аналізу мережевих моделей предметних галузей на основі лінгвостатистичного підходу обробки тематичних текстових даних та інформаційних потоків. Актуальність цього дослідження пов'язана з тим, що з початком стрімкого розвитку інформаційно-комунікаційних технологій та глобалізацією інформаційного простору розпочалося стрімке

збільшення інформаційних ресурсів, що розподілені у вебмережі. Їх розвиток відбувається куди швидше, ніж коли-небуть раніше. І як наслідок, це призвело до збільшення динамічних інформаційних потоків і, відповідно, стрімкого збільшенням об'ємів даних, представлених у електронному вигляді. Важливо зазначити й той факт, що обсяг вищезгаданих даних подвоюється приблизно кожні 18 місяців. І наразі у глобальній мережі Інтернет налічується більше сотні трильйонів документів, і частина з них – це величезні масиви текстових даних, аналіз яких може дати критично важливу інформацію. Та з експоненційним збільшенням інформаційних потоків зростає і частка неструктурованих або слабоструктурованих даних, що, безперечно, ускладнює пошук необхідної та релевантної інформації. Наприклад, основна частина таких даних (близько 95%) є неструктурованими, і лише зовсім мала (близько 5%) – це різні бази даних, де зберігається структурована інформація, яка може бути використана під час прийняття рішень. Тож перед інформаційним суспільством постає також і ряд специфічних проблем, пов'язаних зокрема з критичною невідповідністю між розвитком сучасних інформаційних систем і збільшенням динамічних інформаційних потоків у глобальних комп'ютерних мережах. А тому питання подальшої комп'ютеризованої обробки текстових даних з метою екстрагування знань та подальшої їх структуризації є важливим та актуальним у сучасному інформаційному середовищі. Метою дисертаційної роботи є розробити нові методи формування мережевих моделей предметних галузей на основі текстових корпусів і лігвостатистичного аналізу текстів та розробити нові методи аналізу побудованих мереж для того, щоб приймати ефективні рішення у відповідних предметних галузях, з якими змістовно пов'язані тексти. Дисертаційна робота спрямована на вдосконалення і розширення існуючих підходів до моделювання мережевої структури предметних галузей на основі лінгвістичних даних. Основні завдання включають розробку алгоритмів побудови мереж, які враховують специфіку текстових даних, а також розробку методів аналізу отриманих мереж з метою виявлення ключових зв'язків та характеристик, які допоможуть у прийнятті обґрунтованих рішень у відповідних галузях заснованих на досліджуваних текстових даних. Об'єктом дослідження є процес структуризації у вигляді мережевих моделей текстових інформаційних потоків, розподілених у вебмережі. Предметом дослідження є лінгвостатистичні методи формування та аналізу мережевих моделей предметних галузей на основі текстових корпусів. Для вирішення проблеми та поставлених задач для досягнення мети були використані наступні наукові методи: методи автоматичної обробки та аналізу природної мови та методи комп'ютерної лінгвістики, завдяки яким проводилась попередня комп'ютеризована обробка природномовних текстів, лексичний аналіз та виявлення семантичних зв'язків; методи статистичного аналізу, що застосовувались для виокремлення ключових термінів (слів та словосполучень) із текстових даних; та методи дискретної математики, зокрема, методи теорії графів та складних мереж, завдяки яким здійснювалась побудова мережевих моделей предметних галузей та подальше дослідження й аналіз отриманих моделей. У дисертаційній роботі проведено огляд та аналіз сучасних лінгвостатистичних методів, що застосовуються для структуризації текстових даних шляхом побудови мережевих моделей предметних галузей. Також описано методи та алгоритми аналізу мережевих структур і підходи до комп'ютеризованої обробки та аналізу текстових документів. Крім цього було акцентовано увагу й на проблемах, які можуть виникати під час використання методів статистичного зважування. Детально розглянуто основні рівні лінгвістичної обробки тестових даних. Розглянуто основні ідеї семантичного пошуку, як одного із найперспективніших видів автоматизованого повнотекстового інформаційного пошуку.

2. Dmytrenko O.O. Information technologies for formation and analysis of network models of subject domains based on linguostatistical approach. – Qualifying scientific work on the manuscript rights. Doctor of Philosophy dissertation under 122 «Computer Science» specialty. – Institute for Information Recording of the National Academy of Sciences of Ukraine, Kyiv, Ukraine, 2024. In the dissertation, the author presents the results of conducted research aimed at addressing the current scientific task of forming and analyzing network models of subject domains based on linguostatistical methods for processing thematic textual data and information flows. The relevance of this research is associated with the rapid development of information and communication technologies and the globalization of the information space, which has led to a significant increase in informational resources distributed across the web. Their development is occurring much faster than ever before. As a result,

this has led to an increase in dynamic information flows and, consequently, a rapid increase in the volumes of data presented in electronic form. It is important to note the fact that the volume of the aforementioned data roughly doubles approximately every 18 months. Currently, the global Internet comprises over a hundred trillion documents, and a portion of these are massive collections of text data, the analysis of which can provide critically important information. But with the exponential increase in information flows, the proportion of unstructured or weakly structured data is growing, undoubtedly complicating the search for necessary and relevant information. For instance, the major part of such data (approximately 95%) is unstructured, with only a very small percentage (about 5%) being various databases containing structured information that can be used in decision-making processes. Therefore, the information society also faces a series of specific problems, particularly related to the critical discrepancy between the development of modern information systems and the increase in dynamic information flows within global computer networks. Therefore, the issue of further computerized processing of text data for knowledge extraction and subsequent structuring in the form of a certain ontology remains important and relevant in the modern information environment. The purpose of the dissertation work is to develop new methods for the construction network models of subject domains based on text corpora and linguostatistical analysis of texts and to devise new methods for analyzing the formed networks to make effective decisions within the respective subject domains that are semantically related to the texts. The dissertation aims to improve and expand existing approaches to modelling the network structure of subject domains based on linguistic data. The main tasks include developing algorithms for the construction networks that consider the specifics of text data, as well as developing methods to analyze the resulting networks to identify key connections and characteristics, aiding in making informed decisions within respective domains based on the researched text data. The object of the research is the process of structuring in the form of network models of text information flows distributed across the web. The subject of the research is linguostatistical methods for the construction and analyzing network models of subject domains based on the text corpora. To solve the problem and achieve the set goals, the following scientific methods were used: natural language processing and computational linguistics methods were employed for preliminary computerized processing of natural language texts, lexical analysis, and identification of semantic relationships; statistical analysis methods were applied to extract key terms (words and phrases) from text data; and discrete mathematics methods, particularly graph theory and complex networks, were used for the construction network models of subject domains and further research and analysis of the obtained models. The dissertation provides a review and analysis of modern linguostatistical methods used for structuring text data through the formation of network models of subject domains. The methods and algorithms for analyzing network structures and approaches to computerized processing and analysis of text documents are also described. Additionally, attention was focused on the problems that may arise when using statistical weighting methods. The main levels of linguistic processing of text data are considered in detail. The main concepts of semantic search, as one of the most promising types of automated full-text information search, are considered.

Державний реєстраційний номер ДіР:

Пріоритетний напрям розвитку науки і техніки: Інформаційні та комунікаційні технології

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

Підсумки дослідження: Нове вирішення актуального наукового завдання

Публікації:

- Ланде, Д. В., & Дмитренко, О. О. (2018). Створення мереж слів на основі текстів з використанням алгоритмів графів видимості. Information Technology and Security: Ukrainian research papers collection, 2018, Vol. 6, Iss. 2 (11). 5-18.
- Lande, D. V., & Dmytrenko, O. O. (2019). Визначення вагових значень зв'язків у мережі термінів. Реєстрація, зберігання і обробка даних, 21(4), 40-48.

- Lande D. V., Dmytrenko O. O., Andriichuk O. V., Tsyganok V. V., & Porplenko Y. V. (2020). Building of directed weighted networks of terms for decision-making support during information operations recognition, In: Mathematical Modeling and Simulation of Systems (MODS'2020). MODS 2020. Advances in Intelligent Systems and Computing, vol 1265, (pp. 197-208). Springer, Cham. Pages.
- Lande, D., Andriichuk, O., Dmytrenko, O., Tsyganok, V., & Porplenko, Y. (2020). Побудова баз знань систем підтримки прийняття рішень з використанням направлених мереж термінів при дослідженні інформаційних операцій. Information Technology and Security, 8(2), 153-163.
- Дмитренко, О. О. (2020). Побудова направлених зважених мереж термінів із застосуванням Part-of-speech tagging. Реєстрація, зберігання і обробка даних, 22(4), 47-55.
- Zgurovsky, M. Z., Boldak, A. O., Lande, D. V., Yefremov, K. V., Pyshnograiev, I. O., Soboliev, A. M., & Dmytrenko, O. O. (2021). Enhancing the Relevance of Information Retrieval in Internet Media and Social Networks in Scenario Planning Tasks, IEEE International Conference on System Analysis & Intelligent Computing SAIC 2021: System Analysis & Intelligent Computing. Studies in Computational Intelligence, vol 1022. Springer, Cham, 187-199
- Lande, D., Soboliev, A., & Dmytrenko, O. (2022). Intelligent technologies in information retrieval systems. Artificial intelligence, 27(1), 260-268.
- Lande, D. V., Dmytrenko, O. O., Shevchenko, A. I., Klymenko, M. S., & Vakulenko, M. O. (2023). Spoken language identification based on the transcript analysis. Digital Scholarship in the Humanities, 38(2), 586-595.
- Dmytrenko, O. (2022). Formation Networks of Terms for Identifying Semantic Similarity or Difference Degree of Texts in Cybersecurity. Theoretical and Applied Cybersecurity, 4(1).

Наукова (науково-технічна) продукція: технології

Соціально-економічна спрямованість:

Охоронні документи на ОПІВ:

Впровадження результатів дисертації: Впроваджено

Зв'язок з науковими темами:

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Ланде Дмитро Володимирович
2. Dmytro Lande

Кваліфікація: д. т. н., професор, 05.13.06

Ідентифікатор ORCID ID: 0000-0003-3945-1178

Додаткова інформація:

Повне найменування юридичної особи: 02070921Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

Код за ЄДРПОУ: 02570927

Місцезнаходження: пр-т Перемоги, 37, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Власне Прізвище Ім'я По-батькові:

1. Циганок Віталій Володимирович

2. Vitaliy Tsyganok

Кваліфікація: д. т. н., с.н.с.

Ідентифікатор ORCID ID: 0000-0002-0821-4877

Додаткова інформація:

Повне найменування юридичної особи: Інститут проблем реєстрації інформації Національної академії наук України

Код за ЄДРПОУ: 03771755

Місцезнаходження: вул. М. Шпака, буд. 2, Київ, 03113, Україна

Форма власності: Державна

Сфера управління: Національна академія наук України

Ідентифікатор ROR:

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Стрижак Олександр Євгенійович

2. Oleksandr Stryzhak

Кваліфікація: д.т.н., професор, 05.13.06

Ідентифікатор ORCID ID: 0000-0002-4954-3650

Додаткова інформація:

Повне найменування юридичної особи: Центральний науково-дослідний інститут озброєння та військової техніки Збройних Сил України

Код за ЄДРПОУ: 24292711

Місцезнаходження: проспект Повітрофлотський, буд. 28, Київ, 03049, Україна

Форма власності: Державна

Сфера управління: Міністерство оборони України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Рогушина Юлія Віталіївна

2. Julia Rogushina

Кваліфікація: д.філософ, професор

Ідентифікатор ORCID ID: 0000-0001-7958-2557

Додаткова інформація:

Повне найменування юридичної особи: Інститут програмних систем Національної академії наук України

Код за ЄДРПОУ: 05540149

Місцезнаходження: проспект Академіка Глушкова, буд. 40, корп. 5, Київ, 03187, Україна

Форма власності: Державна

Сфера управління: Національна академія наук України

Ідентифікатор ROR:

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Крючин Андрій Андрійович
2. Andriy Kryuchyn

Кваліфікація: д. т. н., чл-кор.НАН України, 05.13.05

Ідентифікатор ORCID ID: 0000-0002-5063-4146

Додаткова інформація:

Повне найменування юридичної особи: Інститут проблем реєстрації інформації Національної академії наук України

Код за ЄДРПОУ: 03771755

Місцезнаходження: вул. М. Шпака, буд. 2, Київ, 03113, Україна

Форма власності: Державна

Сфера управління: Національна академія наук України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Балагура Ірина Валеріївна
2. Iryna Balagura

Кваліфікація: д. т. н., с.н.с.

Ідентифікатор ORCID ID: 0000-0001-9627-2091

Додаткова інформація:

Повне найменування юридичної особи: Інститут проблем реєстрації інформації Національної академії наук України

Код за ЄДРПОУ: 03771755

Місцезнаходження: вул. М. Шпака, буд. 2, Київ, 03113, Україна

Форма власності: Державна

Сфера управління: Національна академія наук України

Ідентифікатор ROR:

VIII. Заключні відомості

Власне Прізвище Ім'я По-батькові голови ради	Додонов Олександр Георгійович
Власне Прізвище Ім'я По-батькові головуючого на засіданні	Додонов Олександр Георгійович
Відповідальний за підготовку облікових документів	А.А.Крючин
Реєстратор	УкрІНТЕІ
Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності	Юрченко Тетяна Анатоліївна