Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0418U003920

Особливі позначки: відкрита

Дата реєстрації: 03-12-2018

Статус: Захищена

Реквізити наказу МОН / наказу закладу:

II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Прищепа Сергій Вікторович
2. Pryshchepa Serhii Viktorovych

Кваліфікація:

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: кандидат наук

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 05.13.06

Назва наукової спеціальності: Інформаційні технології

Галузь / галузі знань: Не застосовується

Освітньо-наукова програма зі спеціальності: Не застосовується

Дата захисту: 29-11-2018

Спеціальність за освітою: Системи технічного захисту інформації, автоматизація її обробки

Місце роботи здобувача: Київський національний університет будівництва і архітектури

Код за ЄДРПОУ: 02070909

Місцезнаходження: просп. Повітрофлотський, 31, м. Київ, Київ, 03680, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): Д 26.861.05

Повне найменування юридичної особи: Державний університет телекомунікацій

Код за ЄДРПОУ: 38855349

Місцезнаходження: вул. Солом'янська, 7, м. Київ, Київська обл., 03110, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

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

Повне найменування юридичної особи: Інститут проблем реєстрації інформації НАН України

Код за ЄДРПОУ: 03771755

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

Форма власності:

Сфера управління: Національна академія наук України

Ідентифікатор ROR: Не застосовується

V. Відомості про дисертацію

Мова дисертації:

Коди тематичних рубрик: 27.35.25, 50.39.21, 50.07.03

Тема дисертації:

1. Інформаційна технологія екстрагування нових подій на основі лінгвомережевого аналізу у глобальних мережах

2. Information Technology of New Events Extraction Based on Linguistic Network Analysis in Global Networks

Реферат/анотація:

1. У дисертаційній роботі вирішується актуальне наукове завдання розробки інформаційної технології екстрагування нових подій на основі лінгвомережевого аналізу інформаційних потоків соціальних мереж, яка здатна працювати з великими масивами слабоструктурованих текстових даних та великою кількістю шуму, що є притаманним сучасним соціальним мережам, з метою підвищення ефективності автоматичного екстрагування подій з інформаційних потоків соціальних мереж в умовах великих обсягів інформації та значного інформаційного шуму. Проведено аналіз сучасного стану інформаційних технологій екстрагування подій і виявлення їх новизни та існуючих наукових методів для виконання даної задачі. Встановлено, що вимоги до екстраговуванної інформації та недоліки наявних підходів до екстрагування інформації не задовольняють сучасні потреби, таким чином, виникає протиріччя між потребою державних та бізнес структур своєчасно виявляти нові події з динамічно зростаючих обсягів інформації у глобальних мережах та обмеженістю науково-методичного апарату для ефективного їх виявлення. Одним із шляхів вирішення цієї

суперечності є розробка ефективної технології екстрагування нових подій з певних інформаційних потоків. Для цього в роботі розроблено метод виявлення нових подій із текстів повідомлень соціальних мереж. Сутність методу полягає в тому, що з розподілених масивів документів певних категорій в яких наявні ключові слова зі словника тригерів події приймається рішення про приналежність документу до події чи не події, використовуючи наївний Байєсів класифікатор, а використання технік text mining для задачі виявлення понять та сутностей, дає змогу проводити автоматичний аналіз документу та порівняння з наявними в базі даних подіями за рахунок порівняння компонентів події, які представлені векторами події (тригери події, заголовки, ключеві слова, рейтинг джерела) за косинусною мірою та мірою входження для інших компонент (сутності, локації). Призначення методу – виявляти нові події з текстових повідомлень в слабоструктурованих інформаційних потоках, шляхом інтелектуальної обробки текстів. Для автоматичного наповнення словника тригерів події, що використовується в розробленому методі виявлення подій, розроблено новий метод виявлення дескрипторів у текстових масивах. Призначення методу полягає в виявленні та використанні дескрипторів для формування запитів і пошуку за ними релевантних документів по конкретній новинній темі та використання їх в якості можливих тригерів для подієвої класифікації. Для вдосконалення ефективності моніторингу та виявлення документів про подію від першоджерел розроблено метод рейтингування джерел, що здійснює комплексне рейтингування джерел на основі оцінки подієвості та зв'язків адаптованих під спеціальні теги соціальної мережі для визначення авторитетності авторів певних повідомлень та активних посередників розповсюдження інформації, а для виявлення найбільш ймовірного першоджерела події використовується граф горизонтальної видимості. При цьому, методи дослідження базуються на використанні теорії математичного аналізу, теорії ймовірностей і математичної статистики, теорії графів, теорії складних мереж, обробки природних мов, машинного навчання та комп'ютерної лінгвістики. Розроблено інструментально-програмні засоби, що реалізують розроблену технологію в системах контент-моніторингу InfoStream та X-SKIF, що підтверджують акти впровадження. Проведено оцінку ефективності технології на базі тестової вибірки документів українською та російською мовою з соціальної мережі Twitter по темі Автокатастрофи (ДТП) у розмірі 1000 документів, яка показала на 3-5% кращі результати по сбалансованій F-мірі у порівнянні з іншими підходами екстрагування нових подій з інформаційних потоків глобальних мереж.

2. In the dissertation work the actual scientific task of development of information technology of extraction of new events is solved on the basis of linguistic network analysis of information flows of social networks that is capable of working with large masses of poorly structured text data and a large amount of noise, which is inherent in modern social networks, in order to increase the efficiency of automatic extraction of events from information flows of social networks in the conditions of large volumes of information and considerable informational noise. The analysis of the current state of information technologies of extraction of events and the detection of their novelty and existing scientific methods for the fulfillment of this task was carried out. It has been established that the requirements for extracted information and the shortcomings of existing approaches to extracting information do not meet current needs, thus, there is a contradiction between the need for state and business structures to timely identify new events from the dynamically increasing volumes of information in global networks and the limited scientific and methodical apparatus for their effective detection. One way to resolve this contradiction is to develop an effective technology for extraction of new events from certain information flows. To do this, a method for identifying new events from the texts of social networking messages has been developed in the research. The essence of the method lies in the fact that from distributed arrays of documents of certain categories in which there are the key words from the dictionary triggers of the event a decision is made on whether the document belongs to the event or not, using the naive Bayesian classifier, and the use of the technique of text mining for the task of identifying concepts and entities, allows you to automatically analyze the document and compare it with the events available in the database by comparing the components of the event, which are represented by event vectors (event triggers, headers, key words, source rating) by cosine level and the degree of entry for other components (entity, location). Appointment of the method - to detect new events from text messages in poorly structured information flows through the intellectual processing of texts. To automatically fill the trigger

dictionary of an event used in the developed event detection method, a new method for detecting descriptors in text arrays has been developed. The purpose of the method is to identify and use descriptors to generate queries and search for relevant documents for a specific news topic and use them as possible triggers for event classification. In order to improve the effectiveness of monitoring and detecting documents about an event from primary sources, a source rating methodology is developed which performs a comprehensive ranking of sources based on the assessment of the event and the links adapted for special social networking tags to determine the credibility of the authors of certain messages and the active mediators of the dissemination of information, and for the detection of the most probable source of the event, a graph of horizontal visibility is used. In this case, research methods are based on the use of the theory of mathematical analysis, probability theory and mathematical statistics, graph theory, the theory of complex networks, natural language processing, machine learning and computer linguistics. Instrumental and software tools implementing the developed technology in content monitoring systems InfoStream and X-SKIF, which confirm implementation acts, are developed. An assessment of the effectiveness of technology based on the test sample of documents in Ukrainian and Russian from the social network Twitter on the topic of Auto accident (Road Traffic accident) in the size of 1000 documents was carried out, which showed 3-5% better results on a balanced F-measure compared with other approaches extraction of new events from information flows of global networks.

Державний реєстраційний номер ДіР:

Пріоритетний напрям розвитку науки і техніки:

Стратегічний пріоритетний напрям інноваційної діяльності:

Підсумки дослідження:

Публікації:

Наукова (науково-технічна) продукція:

Соціально-економічна спрямованість:

Охоронні документи на ОПІВ:

Впровадження результатів дисертації:

Зв'язок з науковими темами:

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Ланде Дмитро Володимирович
2. Lande Dmytro Volodymyrovych

Кваліфікація: д. т. н., 05.13.06

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Кучеров Дмитро Павлович

2. Kucherov Dmytro Pavlovych

Кваліфікація: д. т. н., 05.13.06

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

Власне Прізвище Ім'я По-батькові:

1. Фурашев Володимир Миколайович

2. Furashev Volodymyr Mykolaiovych

Кваліфікація: к. т. н., 05.13.06

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

Рецензенти

VIII. Заключні відомості

Власне Прізвище Ім'я По-батькові голови ради	Вишнівський Віктор Вікторович
Власне Прізвище Ім'я По-батькові головуючого на засіданні	Вишнівський Віктор Вікторович
Відповідальний за підготовку облікових документів	
Реєстратор	
Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності	Юрченко Т.А.