Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0416U003873

Особливі позначки: відкрита

Дата реєстрації: 25-07-2016

Статус: Захищена

Реквізити наказу МОН / наказу закладу:

II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Тодоріко Ольга Олексіївна

2. Todoriko Olha Oleksiivna

Кваліфікація:

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: кандидат наук

Аспірантура/Докторантура: ні

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: 30-06-2016

Спеціальність за освітою: 8.040301

Місце роботи здобувача: Державний вищий навчальний заклад "Запорізький національний університет"

Код за ЄДРПОУ: 02125243

Місцезнаходження: 69600, м. Запоріжжя, вул. Жуковського, 66

Форма власності:

Сфера управління: Міністерство освіти і науки, молоді та спорту України

Ідентифікатор ROR: Не застосовується

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): Д 64.050.07

Повне найменування юридичної особи: Національний технічний університет "Харківський політехнічний інститут"

Код за ЄДРПОУ: 02071180

Місцезнаходження: вул. Кирпичова, 2, м. Харків, Харківський р-н., Харківська обл., 61002, 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: Державний вищий навчальний заклад "Запорізький національний університет"

Код за ЄДРПОУ: 02125243

Місцезнаходження: 69600, м. Запоріжжя, вул. Жуковського, 66

Форма власності:

Сфера управління: Міністерство освіти і науки, молоді та спорту України

Ідентифікатор ROR: Не застосовується

V. Відомості про дисертацію

Мова дисертації:

Коди тематичних рубрик: 50.37.19

Тема дисертації:

1. Моделі та методи очищення та інтеграції текстових даних в інформаційних системах

2. Models and methods of cleaning and integration of text data in information systems

Реферат/анотація:

1. Об’єкт дослідження – процеси очищення та інтеграції текстових даних в інформаційно-довідкових та пошукових системах. Мета дослідження - удосконалення технології очищення та інтеграції текстових даних в довідкових і пошукових інформаційних системах за рахунок використання моделі словозмінної парадигми та методу побудови лексемного індексу при організації пошуку за схожістю. Методи дослідження: методи математичного моделювання, методи об’єктно-орієнтованого аналізу та синтезу програмного забезпечення з використанням уніфікованої мови системного моделювання UML, методи створення лінійних та нейронних класифікаторів, методи теорії ймовірності та статистичного аналізу експериментальних даних. Теоретичні і практичні результати: створення програмного забезпечення у вигляді бібліотеки класів на мові Java для організації словникового пошуку за схожістю та інтеграції наборів даних. Наукова новизна одержаних результатів: вперше: - розроблено моделі словозмінної парадигми, які відрізняються способом

представлення слів та обчислення приблизної міри схожості між представленнями, способом врахування форм слів та позицій символів у словах, що дозволило створити підґрунтя для побудови лексемного індексу, реалізації методу пошуку за схожістю, очищення та інтеграції наборів даних; - запропоновано метод побудови лексемного індексу, який відрізняється від існуючих зменшенням кількості записів за рахунок відображення всіх словоформ слова в один запис індексу, що дозволяє при попередній фільтрації скоротити кількість обчислень міри схожості між зразком та всіма формами слів; удосконалено метод пошуку за схожістю текстової інформації в довідкових та пошукових системах за рахунок використання моделей словозмінної парадигми та лексемного індексу, що дозволяє збільшити точність та повноту попередньої фільтрації; дістала подальшого розвитку інформаційна технологія очищення та інтеграції наборів даних, яка за рахунок вдосконалення методу пошуку за схожістю дозволяє спростити операцію обчислення міри схожості. Ступінь упровадження: результати дисертаційної роботи застосовано у процесі документообігу приймальної комісії для очищення даних в «Системі реєстрації абітурієнтів» Державного вищого навчального закладу «ЗНУ» та для зіставлення записів даної системи та системи електронної подачі заявок для вступу в вищі «Електронний вступ» МОН України, а також використані у навчальному процесі на кафедрі інформаційних технологій Державного вищого навчального закладу «ЗНУ». Галузь використання - очищення та інтеграція інформаційних систем.

2. The object of study is processes of cleaning and integration of data in information reference system and information retrieval system. The purpose of the research is to improve technology of cleaning and integration of text data in information reference systems and information retrieval systems, through the use of model of an inflectional paradigm and methods of creation of a lexeme index at the organization of a tolerant retrieval. Research methods include methods of mathematical modeling, methods of object-oriented software analysis and synthesis using the unified modeling language, methods of creation linear and neural classifiers, methods of probability theory and statistical analysis of experimental data. Theoretical and practical results: the creation of software in the form of a class library in Java for organizing lexical similarity search and integration of data sets. Scientific novelty of the results: for the first time: - a model of the inflectional paradigm was developed, which differs from the existing ones in a way it represents words and calculates an approximate measure of similarity between the representations; the model has the special method of accounting forms of the words and attitudes of characters in words, thereby creating the basis for constructing a lexeme index, the implementation of methods to search for similarities, cleaning and integration sets data; - a method of constructing lexeme index that unlike known analogues has reduced number of entries because of mapping all forms of one word in the index entry, which allows pre-filtering to reduce the amount of hard computation of similarity measures between the sample and all forms of words; improved method of tolerant retrieval of text in the reference and search information systems through the use of model of the inflectional paradigm and lexeme index, which improves the accuracy and completeness of the pre-filtering; further developed the information technology of cleaning and integration of data sets that by improving method of tolerant retrieval simplifies the operation of calculating the similarity measure. Degree of implementation: the results of the thesis are applied in practice of an entrance committee for the data cleaning in "System of registration of applicants" State institution of higher education «ZNU» and to link records of the system and the system of online application for admission to "United state electronic database on education" MES of Ukraine, as well as used in the educational process at the department of information technologies of the State institution of higher education «ZNU». Application is in cleaning and integration of information systems.

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. Гоменюк Сергій Іванович
2. Gomenyuk, Sergiy Ivanovich

Кваліфікація: д.т.н., 05.13.12

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. Шостак Ігор Володимирович
2. Шостак Ігор Володимирович

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. Хайрова Ніна Феліксівна

2. Хайрова Ніна Феліксівна

Кваліфікація: д.т.н., 10.02.21

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. Шаронова Наталія Валеріївна

2. Шаронова Наталія Валеріївна

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

Рецензенти

VIII. Заключні відомості

Власне Прізвище Ім'я По-батькові голови ради Куценко Олександр Сергійович

Власне Прізвище Ім'я По-батькові головуючого на засіданні Куценко Олександр Сергійович

Відповідальний за підготовку облікових документів

Реєстратор

Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності



Юрченко Т.А.