Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0419U004383

Особливі позначки: відкрита

Дата реєстрації: 11-10-2019

Статус: Захищена

Реквізити наказу МОН / наказу закладу:

II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Сергеєв Данило Сергійович

2. Serheiev Danylo Serhiyovych

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: 04-10-2019

Спеціальність за освітою: 8.05020102 - "Комп'ютеризовані та робототехнічні системи"

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

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Перемоги, 37, м. Київ, Київська обл., 03056, 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 26.002.29

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

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Перемоги, 37, м. Київ, Київська обл., 03056, 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: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Перемоги, 37, м. Київ, Київська обл., 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

V. Відомості про дисертацію

Мова дисертації:

Коди тематичних рубрик: 28.23.15

Тема дисертації:

1. Інформаційна технологія обробки природномовних текстів на основі інтеграційного підходу
2. Natural language texts processing technology based on the integrational approach

Реферат/анотація:

1. Дисертація присвячена вирішенню актуальної науково-технічної задачі розробки інформаційної технології обробки природномовних текстів на основі інтеграційного підходу. На основі аналізу актуальних проблем у галузі обробки природної мови показано, що прикладні технології обробки природної мови виконують поставлені задачі, але є можливим їх удосконалення для вирішення комплексних задач, зокрема машинного перекладу та природномовного пошуку. З цією метою створено формальну модель представлення знань у природномовній базі знань та моделі її основних елементів, якими є квант знань, або найменший елемент знань, та відношення, яке описує зв'язок між квантами знань. Розроблено метод обробки природномовних текстів на основі запропонованої моделі. На основі створених моделей та методу розроблено процедури записування та пошуку природномовних знань для технологій обробки природної мови, які дозволяють встановити зв’язки на структурному рівні між синтаксичною структурою тексту та довільною структурою

метаданих. Теоретично показано, що складність природномовного пошуку з використанням розроблених процедур не перевищує таку для аналогів, і в середньому є меншою ніж в аналогів для складних пошукових запитів. В рамках роботи розроблено інформаційну технологію обробки природномовних текстів на основі інтеграційного підходу та експериментально показано, що її використання дозволяє підвищити середню релевантність природномовного пошуку на 14%.

2. In the recent years, the research in the field of natural language processing (NLP) has achieved significant practical results, including natural-language voice user interface for mobile devices, significant progress in machine translation technologies, handwriting and voice recognition, etc. At the same time, the task of improving the performance of these systems remains relevant. This study focuses on developing information technology for processing natural language texts based on the integrational approach, aimed to increase efficiency of natural language processing technologies. The subject of the research is models, methods, algorithms and information technologies for natural language processing. Based on the analysis of actual problems in the field of natural language processing, it is shown that applied technologies of natural language processing are successful in fulfilling the intended specific tasks, but it is determined that there is a room for improvement in the area of solving complex problems, in particular, machine translation and natural language search. The role of knowledge bases in information technologies of natural language processing is determined as a necessary component for the interaction of different systems. Existing approaches to developing natural-language knowledge bases are characterized and analyzed. A conclusion is made that existing technologies behind natural-language knowledge bases separately allow to achieve high levels of completeness, consistency and flexibility for practical purposes, but no technology combines high scores on all of the aforementioned qualities. A formal model of knowledge representation for a natural-language knowledge base is created, including models of its main elements, namely the quantum of knowledge, or the smallest element of knowledge, and the relation objects that describe connections between quanta of knowledge. A method for processing natural language texts based on this model of knowledge is developed, including procedures for using information technology in applied natural language processing problems. On the basis of the created models and the method, the procedures for writing and searching natural language skills for natural language processing technologies are developed, which allow to establish links at the structural level between the syntactic structure of the text and the arbitrary structure of the metadata. It is theoretically shown that the complexity of the natural-language search using the developed procedures does not exceed the complexity of the analogues, and on average is less than that of the analogues for complex search queries. Examples of use of the developed information technology for processing natural language texts in practical problems, namely natural language search and machine translation are provided. Writing and searching methods are created based on the knowledge representation model, allowing to establish links at the structural level between syntactic structure of the text and arbitrary structure of the metadata in natural language processing technologies. Information technology for processing natural language texts based on the integrational approach is developed, for which it is theoretically proven that the search complexity does not exceed that of the existing alternatives, and is on average 5-12% lower for complex search queries. Subsystems and operations of such system are defined, and database scheme is developed. Computational complexity of natural language knowledge search in the information system is analyzed and compared with the existing alternatives. Experimental testing of the information system is conducted and the acquired data are analyzed, demonstrating increased relevance of search results of natural language search. Within the framework of the work, information technology for the processing of natural language texts has been developed on the basis of the integrational approach. Based on experimental data acquired from measuring relevance of natural language search results, it has been shown that the developed information technology can increase relevance of search results. Specifically, relevance was increased by 14% on average for the whole set of experimental queries and search results, with no significant increase in relevance detected for the top quartile of results sorted by original relevance, and major increase detected for the lower quartile of original results. The information technology for the processing of natural language texts can be used to improve performance of various natural language processing technologies, in particular natural language search systems, machine translation systems and natural language user interfaces.

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. Кисленко Юрій Іванович
2. Kyslenko Yuri I.

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

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Бармак Олександр Володимирович
2. Barmak Oleksandr V.

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. Ланде Дмитро Володимирович

2. Lande Dmytro V.

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 Новіков Олексій Миколайович

Власне Прізвище Ім'я По-батькові головуючого на засіданні Новіков Олексій Миколайович

Відповідальний за підготовку облікових документів

Реєстратор

Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності Юрченко Т.А.