Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0420U102237

Особливі позначки: відкрита

Дата реєстрації: 14-12-2020

Статус: Захищена

Реквізити наказу МОН / наказу закладу:

II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Голуб Тетяна Василівна

2. Golub Tetiana

Кваліфікація:

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: кандидат наук

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 05.13.05

Назва наукової спеціальності: Комп'ютерні системи та компоненти

Галузь / галузі знань: Не застосовується

Освітньо-наукова програма зі спеціальності: Не застосовується

Дата захисту: 26-11-2020

Спеціальність за освітою: Радіотехніка

Місце роботи здобувача: Національний університет "Запорізька політехніка"

Код за ЄДРПОУ: 02070849

Місцезнаходження: вул. Жуковського, 64, м. Запоріжжя, Запорізький р-н., Запорізька обл., 69063, 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 11.052.03

Повне найменування юридичної особи: Державний вищий навчальний заклад "Донецький національний технічний університет"

Код за ЄДРПОУ: 02070826

Місцезнаходження: пл. Шибанкова, буд. 2, м. Покровськ, Покровський р-н., Донецька обл., 85300, 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 університет "Запорізька політехніка"

Код за ЄДРПОУ: 02070849

Місцезнаходження: вул. Жуковського, 64, м. Запоріжжя, Запорізький р-н., Запорізька обл., 69063, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

V. Відомості про дисертацію

Мова дисертації:

Коди тематичних рубрик: 28.23.33, 28.23.15

Тема дисертації:

1. Програмно-апаратні засоби скорочення часу процесу класифікації текстів з використанням мікросхем програмованої логіки

2. Hardware and software means for time reducing of the text classification process using microchips of programmable logic

Реферат/анотація:

1. Дисертаційна робота присвячена вирішенню актуальної наукової задачі розробки засобів комп'ютерної класифікації текстових документів, орієнтованих на скорочення часових витрат, з комплексним використанням стандартних програмних засобів та ПЛІС FPGA при аналізі текстів. В роботі удосконалено математичну модель класифікації текстів шляхом додаткового видалення незначущих для категорій спільної тематики ознак, що дозволяє опосередковано скоротити часові витрати при класифікації. Модифіковано алгоритм стемінга для обробки україномовних текстів. Розроблено уточнений метод формування простору ознак категорій, використання якого дозволяє зменшити зазначений простір в середньому на 24%, що

призводить до скорочення часу подальших обчислень на 19%. Розроблено програмний спосіб попередньої підготовки даних у форматі, придатному для обробки за допомогою апаратної платформи, а також спосіб апаратного прискорення в схемі ідентифікатора, що дозволяє виконувати одночасну обробку декількох однотипних задач. Розроблено новий комплексний метод скорочення часу класифікації, в результаті використання якого досягається як скорочення часових витрат на програмну обробку одного окремого тексту, так і прискорення етапу ідентифікації декількох текстів апаратними засобами. Вирішена науково-практична задача скорочення часових витрат при класифікації текстів на основі програмно-апаратної реалізації.

2. The thesis is devoted to solving the topical scientific problem of developing the methods of text document automatic classification focused on reduction of time expenses and increasing of calculations productivity with using of complex standard software and FPGA at the analysis of the Ukrainian texts. The linguistic analysis of Ukrainian-language texts has been performed and language's morphological features have been emphasized. The methods of text pre-processing including tokenization, removal of "stop words", stemming, lemization have been reviewed. The expediency and conditions of their using in the construction of classifiers including the possibility of hardware implementation have been analyzed. The methods of a category feature space formation have been considered in this work. An analysis of ways to reduce this space and index its elements for reducing the time spent on its further processing has been performed. A mathematical model of the text classification process is given. It is determined that the process of test documents classification is complex. It consists of two main stages: the formation of the category features space and text identification. Each of them contains internal components. Thus, the text classification is a complex process that requires significant time because it has a large number of accompanies calculations. In this regard it is established that at almost each of the internal classification stages it is possible to achieve a certain reduction of the task execution time. In particular, in the dissertation work the software solution includes development and improvement of preliminary text preparation methods, reducing the space of classification features. Hardware solution includes the acceleration using the conveyor organization of the internal adder, as well as parallelization of calculations at the stage of text identification by category. Pre-processing of the input text depends on the peculiarities of word formation of the language that was used to write the text. Therefore, the dissertation algorithm for Ukrainian-language texts is modified in the dissertation, taking into account some features. This improvement has reduced the time costs when performing text stamping by an average of 22%, when performing classification - by an average of 5%. In dissertation the author has developed the modified mathematical model of formation of space of signs of separate categories of the general subject taking into account additional filtering of terms on the basis of threshold values. In this case, the threshold values are formed separately for each category on the basis of statistical data of the weights of the terms (part of the word after stemming without endings and suffixes). The weight coefficients of the terms are determined by the TF-SLF. Performing filtering allows to exclude from the feature space terms that are common to all categories under consideration, that is, do not contribute to their separation. Such removal reduces the feature space of a certain category in the range from 18% to 35%. A refined method for forming the space of category attributes based on a modified stemming algorithm for the Ukrainian language and a method for refining threshold values for filtering the category attribute space based on a modified mathematical model have been developed by the author in the dissertation work. Using the refined method in experiments allowed reducing the time spent on the classification of text documents by an average of 20%. A software implementation of the text classifier has been developed on the basis of the refined method of forming the category attributes space. Using this method allowed processing four data sets of 715 μs, while one data set was processed for 714 μs. To transfer information from software to hardware, data must be presented in an appropriate format. The peculiarity of the existing data format standards in computer technology lies in their redundancy, which leads to significant resource costs when using it on the FPGA platform. Therefore, a data format that works with Cyrillic characters and implements the optimal size of the input data by minimizing the size of the character code, has been proposed. In accordance with the format suggested by the author, the size of the character code is 6 bits, the size of the word code corresponds to the number of characters in the word multiplied by the size of the character code. Using this method reduces the

volume of processed binary arrays up to 25%. A software and hardware method for classifying texts has been developed on the basis of the developed algorithms and methods. On the basis of this method, a software and hardware complex has been developed. Experimental studies have been confirmed the effectiveness of using the developed methods in terms of reducing the time spent on the text classification.

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. Зеленьова Ірина Яківна
2. Zeleniova Iryna Iakivna

Кваліфікація: к. т. н., 05.13.05

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. Кривуля Геннадій Федорович
2. Kryvulia Hennadii F.

Кваліфікація: д. т. н., 05.13.00

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. Ситніков Валерій Степанович
2. Sitnikov Valery S.

Кваліфікація: д. т. н., 05.13.05

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 Башков Євген Олександрович

Власне Прізвище Ім'я По-батькові головуючого на засіданні Башков Євген Олександрович

Відповідальний за підготовку облікових документів

Реєстратор

Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності



Юрченко Т.А.