

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

Факультет прикладної математики

Кафедра програмного забезпечення комп'ютерних систем

«До захисту допущено»
завідувач кафедри

_____ І.А. Дичка

(підпис)

“ _____ ” _____ 2016 р.

Дипломний проект

на здобуття ступеня бакалавра

з напрямку підготовки 6.050103 “Програмна інженерія”

на тему ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ПІДТРИМКИ
ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ АВТОМАТИЗОВАНОГО
РЕФЕРУВАННЯ ПРИРОДНОМОВНИХ ТЕКСТОВИХ ДАНИХ

Виконав: студент 4 курсу, групи КП-21

Пастушенко Андрій Сергійович

_____ (підпис)

Керівник доц., доц., к.т.н. Заболотня Т.М.

_____ (підпис)

Консультант з нормоконтролю старший викладач Онай М.В.

_____ (підпис)

Рецензент доц. кафедри ММСА Інституту прикладного
системного аналізу, доц., к.т.н. Дідковська М.В.

_____ (підпис)

Засвідчую, що у цьому дипломному
проекті немає запозичень з праць інших
авторів без відповідних посилань.

Студент _____

(підпис)

АНОТАЦІЯ

Дипломний проект присвячений створенню програмного забезпечення для підтримки дослідження ефективності автоматизованого реферування природномовних текстів.

Розроблене програмне забезпечення є Windows-застосунком, що слугує платформою для реалізації алгоритмів автоматизованого реферування текстових даних, підтримує гнучке налаштування параметрів цих алгоритмів, одночасну обробку декількох наборів текстових даних та генерацію звітів за результатами роботи.

Запропоновано програмний інтерфейс взаємодії структурних модулів створеного ПЗ. Таким чином кожен сторонній розробник, який має програмно реалізовані нові алгоритми автоматизованого реферування, може інтегрувати їх до системи. Також в рамках дипломного проекту реалізовано модулі завантаження до застосунку текстових даних у популярних форматах (.docx, .doc, .txt, .pdf); модулі, які працюють за класичними алгоритмами автоматизованого реферування природномовних текстових даних (такі, як ACSI-Matic та TextRank), а також імплементовано метод оцінки їх ефективності з використанням метрики Rouge-N.

Розроблений інтуїтивно зрозумілий інтерфейс користувача, що дозволяє користувачеві задавати алгоритм роботи ПЗ, змінювати порядок роботи модулів, зберігати конфігурацію для повторного використання, проводити аналіз результатів роботи системи та генерувати звіт у форматі Excel-документу.

ABSTRACT

This diploma project deals with the development of the software that supports measuring the effectiveness of the automated text summarization algorithms.

The software is a Windows desktop application that serves as a platform for implementing automated summarization algorithms for text data, supports flexible configuration of these algorithms, simultaneous processing of multiple sets of text data and generating reports using results of the processing.

Within the system the API for modules communication is proposed. Using this API, a third-party developer can implement algorithms for automated summarization and integrate them into the system. In the given diploma project modules for the text data I/O in popular formats (such as .txt, .docx, .doc, .pdf), modules that implement classical algorithms for automated summarization (ACSI-Matic and TextRank) and method for evaluating their effectiveness using Rouge-N metric are implemented.

Also user-friendly, intuitive interface is created, that allows user to set algorithm for text processing, change the order of running each module in this algorithm, save program configuration into file for further usages, evaluating the effectiveness of text summarization algorithms and generate reports using results of text processing.