

РЕФЕРАТ

Повна назва «Дослідження методів аналізу текстової інформації».

Магістерська робота за фахом: 126 «Інформаційні системи та технології».

Студент гр. ІСТ-18-1м ДДМА, Д.Ю. Веприцький. – Краматорськ, 2019.

Робота містить 106 стор.: 28 рис., 13 табл., 25 слайдів.

Перший розділ містить розглянуті особливості текстових даних та підходи до їх аналізу, проведено дослідження методів аналізу текстових даних. У другому розділі проаналізовані методи перевірки унікальності даних та проведено дослідження методів оцінки унікальності текстових даних. У третьому розділі наведена методика аналізу текстових даних та об'єктно-орієнтована модель програмного забезпечення для аналізу текстових даних. У четвертому розділі виконаний опис програмного забезпечення для аналізу текстових даних, наведений розрахунок економічної ефективності розробки та реалізації програмного забезпечення для аналізу електронних текстових даних. У п'ятому розділі наведені заходи щодо забезпечення безпечних і комфортних умов праці на робочих місцях.

Метою кваліфікаційної роботи є теоретичне обґрунтування та розробка підходів до вирішення задач аналізу текстової інформації, в тому числі визначення плагіату, через використання математичних методів та алгоритмів.

Об'єкт дослідження: порівняльний аналіз та оцінювання унікальності електронних текстових даних.

Предмет дослідження: математичні методи та алгоритми аналізу електронних текстових даних, в тому числі методи оцінки унікальності тексту. При виконанні магістерської роботи використовувалися методи аналізу при дослідженні методів та моделей аналізу електронних текстових даних, методи моделювання застосовані при вдосконаленні методики порівняльного аналізу електронних текстових даних та проектуванні інформаційної моделі програмного забезпечення.

Наукова новизна одержаних результатів полягає в тому, що було удосконалено методика аналізу текстових даних на сторінках інтернет-ресурсів через оцінку унікальності текстових даних за допомогою коефіцієнта Жаккара, відстані Дамерау-Левенштейна та методу шинглів.

Практичне значення одержаних результатів дослідження полягає в зменшенні трудомісткості операцій при аналізі електронних тестових даних (перевірка на плагіат) за рахунок застосування розробленого програмного забезпечення.

ТЕКСТОВИЙ АНАЛІЗ, ВИДИ ТЕКСТІВ, ЗАСОБИ ДЛЯ ОБРОБКИ ТЕКСТУ, МЕТОДИ АНАЛІЗУ ТЕКСТУ, ОЦІНКА УНІКАЛЬНОСТІ, ПРОЕКТУВАННЯ, UML-ДІАГРАМИ, ПРОГРАМНИЙ МОДУЛЬ, МЕТОД ШИНГЛІВ, ЕКОНОМІЧНА ЕФЕКТИВНІСТЬ, ОХОРОНА ПРАЦІ

ABSTRACT

Full name «Research of methods of analysis of text information».

Master's work on the specialty: 126 «Information systems and technologies»

Student gr. IST-18-1m DSMA, Vepritsky D.Y. – Kramatorsk, 2019.

The work contains 106 pages: 28 fig., 13 tab., 25 slides.

The first section contains the considered features of text data and going near their analysis, a study of methods of analysis of text data is undertaken. In the second division the analysed methods of verification of unicity of data and a study of methods of estimation of unicity of text data is undertaken. In the third division the brought methodology over of analysis of text data and object-oriented model of software for the analysis of text data. In a fourth division the executed description of software is for the analysis of text data, the brought calculation over of economic efficiency of development and realization of software for the analysis of electronic text data. In a fifth division the brought events over in relation to providing of safe and comfort terms of labour on workplaces. The aim of qualifying work are a theoretical ground and development of going near the decision of tasks of analysis of text information, including determination of plagiarism, through the use of mathematical methods and algorithms.

Conducting a study: benchmarking and evaluating unique electronic text data.

The study of mathematical methods and algorithms of the reported electronic text data, the mentioned methods Coefficient of functionality of the text. When making a master's thesis, the methods used in the study of methods and models used electronic text data, modeling methods used when using the appropriate methods used to use appropriate methods. information model of software protection.

The scientific novelty of the obtained results is that the technique of text data analysis on web pages was improved by assessing the uniqueness of text data using the Jacquard coefficient, the Damerau-Levenstein distance and the shingle method.

The practical significance of the results of the study is to reduce the complexity of operations in the analysis of electronic test data (check for plagiarism) through the use of developed software.

TEXT ANALYSIS, TYPES OF TEXTS, MEANS FOR TEXT PROCESSING, TEXT ANALYSIS METHODS, ESTIMATES OF UNIQUE, DESIGN, UML-DIAGRAMS, MYSTERY, MYSTERY

РЕФЕРАТ

Полное название «Исследование методов анализа текстовой информации».

Магистерская работа по специальности: 126 «Информационные системы и технологии».

Студент гр. ИСТ-18-1м ДГМА, Д.Ю.Веприцкий. – Краматорск, 2019.

Работа содержит 106 стр.: 28 рис., 13 табл., 25 слайдов.

Первый раздел содержит рассмотренные особенности текстовых данных и подходы к их анализу, проведено исследование методов анализа текстовых данных. Во втором разделе проанализированы методы проверки уникальности данных и проведено исследование методов оценки уникальности текстовых данных. В третьем разделе приведена методика анализа текстовых данных и объектно-ориентированная модель программного обеспечения для анализа текстовых данных. В четвертом разделе выполнено описание программного модуля для анализа текстовых данных, приведен расчет экономической эффективности разработки и реализации программного обеспечения для анализа текстовых данных. В пятом разделе приведены меры по обеспечению безопасных и комфортных условий труда на рабочих местах.

Целью квалификационной работы является теоретическое обоснование и разработка подходов к решению задач анализа текстовой информации, в том числе определение плагиата, через использование математических методов и алгоритмов.

Объект исследования: сравнительный анализ и оценка уникальности электронных текстовых данных. Предмет исследования: математические методы и алгоритмы анализа электронных текстовых данных, в том числе методы оценки уникальности тексту. При выполнении магистерской работы использовались методы анализа при исследовании методов и моделей анализа электронных текстовых данных, методы моделирования применены при совершенствовании методики сравнительного анализа электронных текстовых данных и проектировании информационной модели программного обеспечения.

Научная новизна исследования заключается в том, что была усовершенствована методика анализа текстовых данных на страницах интернет-ресурсов через оценку уникальности текстовых данных с помощью коэффициента Жаккара, расстояния Дамерау-Левенштейна и метода шинглов.

Практическое значение полученных результатов исследования заключается в уменьшенных трудоемкости операций при анализе электронных текстовых данных (проверка на плагиат) за счет применения разработанного программного обеспечения.

ТЕКСТОВЫЙ АНАЛИЗ, ВИДЫ ТЕКСТОВ, СРЕДСТВА ДЛЯ ОБРАБОТКИ ТЕКСТА, МЕТОДЫ АНАЛИЗА ТЕКСТА, ОЦЕНКА УНИКАЛЬНОСТИ, ПРОЕКТИРОВАНИЕ, UML-ДИАГРАММЫ, ПРОГРАММНЫЕ МОДУЛИ, МЕТОД ШИНГЛОВ, ЭКОНОМИЧЕСКАЯ ЭФФЕКТИВНОСТЬ, ОХРАНА ТРУДА