

Лабораторная работа

«Реализация алгоритмов автоматического морфологического и лексико-грамматического анализа текста естественного языка»

Цель работы – закрепить знания морфологического и лексико-грамматического анализа текста.

Основные **задачи** работы следующие:

1. Изучить теоретический материал, необходимый для решения задачи о редакционном расстоянии.
2. Закрепить навыки программирования естественно-языковых систем и вспомогательных прикладных программ на одном из языков программирования.

Порядок выполнения работы

Задание 1. Создать программную реализацию решения задачи о редакционном расстоянии методом динамического программирования между словами, подаваемыми на вход системы и словами текста, полученными в результате выполнения лабораторной работы №1.

Задание 2. Спроектируйте внешний интерфейс автоматизированной системы осуществляющей решения задачи о редакционном расстоянии между словами, подаваемыми на вход системы и словами текста, полученными в результате выполнения лабораторной работы №1.

Задание 3. Разработайте алгоритмы сервисных функций, предоставляемых Вашей системой.

Задание 4. Реализуйте разработанные алгоритмы в виде программ на языке программирования.

Задание 5. Протестируйте разработанную систему, используя различные входные тексты.

Задание 6. Разработайте и внедрите в систему средства помощи пользователю.

Задание 7. Протестируйте систему, включающую средства помощи пользователю.

Задание 8. Оформите отчет по работе, используя такие программные средства, как Microsoft Visio или Draw.io и включив в него следующие основные пункты:

- структурно-функциональную схему разработанного приложения;

- описание структур хранения данных, алгоритмов их обработки, необходимых для реализации базовых требований к разработанной программе;
- оценку быстродействия приложения;
- выводы по работе и по перспективам использования приложения.

Отчет предоставить для проверки в электронном виде.

Требования к разрабатываемой системе

- ✓ на входе – слова естественно-языкового текста содержащие орфографические ошибки;
- ✓ на выходе – перечень слов, расстояние до которых от входного слова не превышает некоторого заданного Вами значения, упорядоченный по убыванию величины расстояния редактирования;
- ✓ интерфейс системы должен быть предельно простым и доступным для пользователей любого уровня, содержать понятный набор инструментов и средств, а также help-средства;
- ✓ наличие средств распечатки полученной на выходе информации.

Варианты

№ варианта	Язык текста	Формат документа
1	Русский	Html
2	Английский	Html
3	Китайский	Html
4	Арабский	Html
5	Немецкий	Html
6	Русский	docx
7	Английский	docx
8	Китайский	docx
9	Арабский	docx
10	Немецкий	docx