

АННОТАЦИЯ

Тема выпускной квалификационной работы: «Разработка интеллектуальной системы автоматического реферирования текстов (на примере русскоязычных текстов)»

Автор ВКР: Буркадзе Нана Гочаевна

Научный руководитель ВКР: канд. экон. наук Тимченко Ольга Викторовна

Сведения об организации-заказчике: ООО «Рекламное агентство КАКАДУ»

Актуальность темы исследования: Современное развитие общества самым дорогим ресурсом является время. Человек постоянно сталкивается с большим количеством информации разного вида и представления, и несмотря на обилие и разнообразие информации вынужден своевременно ее обрабатывать. Большая часть такой информации представлена текстами на естественном языке. В случае, когда документов оказывается слишком много и человек не способен внимательно прочитать их в отведенное для этого время, на помощь приходят системы автоматического реферирования текстовых документов. Автоматическое реферирование и аннотирование текстов является одним из важнейших направлений развития современных информационных технологий. Количество информации, с которой приходится сталкиваться человеку, постоянно растет и приходит время, когда проработать весь необходимый материал становится просто невозможным. Разработка новых алгоритмов автоматического реферирования и аннотирования текстов на естественном языке по-прежнему не теряет своей актуальности, и становится все более необходимой в связи с постоянно возрастающим объемом текстовых данных.

Цель работы: Целью данной работы является исследование задачи автоматического реферирования текста и поиск новых подходов к решению данной задачи с применением технологий на основе интеллектуальных систем. Проектируемая система автоматического реферирования текста должна улучшить смысловое качество реферата и повысить эффективность процессов обработки данных и знаний в компьютерных системах, а также позволить хорошо обрабатывать тексты различных жанров, разного уровня сложности терминологии и объемов.

Задачи: Основными задачами исследования являются обзор современного состояния проблемы автоматического реферирования текстов, исследование методов реферирования текстов, обзор существующих систем реферирования текстов, разработка архитектуры проектируемой системы, разработка алгоритма автоматического реферирования текстов, разработка программного продукта, реализующего разработанный алгоритм автоматического реферирования текстов, выполнение сравнительного анализа предложенного алгоритма и известных алгоритмов решения данной задачи. В результате выполнения работы должна быть разработана структура системы автоматического реферирования текста, выбраны методы и

разработаны алгоритмы, которые необходимо реализовать в ее модулях, а также намечены пути улучшения качества работы разработанной системы.

Теоретическая и практическая значимость исследования: Теоретическая и практическая значимость исследования состоит в апробации предложенного в рамках выпускной квалификационной работы алгоритма автоматического реферирования текстовой информации.

Результаты исследования: В рамках выпускной квалификационной работы была достигнута поставленная цель – спроектирована интеллектуальная система автоматического реферирования текстов на русском языке. В системе использовались статистические методы для предварительной обработки текстов, такие как алгоритм стемминга, основанный на алгоритме Mystem, алгоритм TF*IDF для определения ключевых слов, алгоритм определения вектора весовых признаков (при этом количество выделенных признаков принято равным пяти). Статистические методы обработки текста сформировали входные данные для использования методов нечеткой логики. В качестве функции принадлежности использовался нечёткий центроид. Правила нечеткого вывода сформулированы в форме «если - то». Используемый алгоритм нечеткого вывода позволил для всех предложений текста получить значения функции принадлежности множества «важное предложение». С помощью алгоритма быстрой сортировки предложения реферируемого текста отсортированы по убыванию значений функции принадлежности. Коэффициент сжатия, задаваемый пользователем, определяет количество предложений, которые должны быть включены в реферат. Была проведена оценка эффективности интеллектуальной системы автоматического реферирования текста.

Рекомендации: Для повышения эффективности работы системы в целом необходимо продолжать обучение, включить на стадию предварительной обработки текста процедуру очистки текста от стоп-слов, служебных слов. Кроме того, в процедуру нечеткого вывода необходимо добавить правила типа «если - то» для учета лингвистических правил естественного языка.