

АННОТАЦИЯ

Тема выпускной квалификационной работы: «Моделирование интеллектуальной системы извлечения информации определенного типа из текста»

Автор ВКР: Агрба Диана Анзоровна

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

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

Актуальность темы исследования: В настоящее время наблюдается крайне стремительное развитие интернет технологий, сети Интернет, где знания и информация в огромных количествах представлены преимущественно в текстовой форме и количество неструктурированной информации только увеличивается, при этом в настоящее время анализируется менее 1% текстов, а рост рынка происходит в основном за счет анализа данных социальных сетей, и потому особо актуальной задачей стало автоматизированное извлечение информации из текста. Информация может быть сделана более структурированной посредством преобразования в реляционную форму или добавлением XML разметки. При мониторинге новостных лент с помощью интеллектуальных агентов как раз и потребуются методы извлечения информации и преобразования её в такую форму, с которой будет удобнее работать позже. Выделенная из текста информация может выступать в качестве метаинформации, представляя текстовый документ при решении задач информационного поиска, классификации, кластеризации, аннотирования и реферирования. Для решения этой проблемы в данной работе будет смоделирована интеллектуальная система по извлечению информации из текста на естественном языке.

Цель работы: Цель данной работы состоит в изучении процесса извлечения информации из текста, его организации и моделировании интеллектуальной системы извлечения информации из текста. Результатом данной работы должна стать реализация смоделированной системы.

Задачи: К задачам данной работы относятся сбор, систематизация, обработка фактического материала по теме выпускной квалификационной работы, изучение интеллектуальных систем, исследование проблематики извлечения информации из текста, изучение машинного обучения, разработка модели ИС, рассмотрение способов реализации созданной модели, реализация ИС извлечения информации из текста.

Теоретическая и практическая значимость исследования: Теоретическая значимость полученных результатов обусловлена определением проблемы извлечения информации (в частности ключевых слов) из текста, рассмотрением и выявлением способов решения этой проблемы и поставленных задач, систематизацией имеющихся сведений о проблемной области, разработанных методах и реализующих их алгоритмах, а также обоснованием выбора направлений перспективных исследований, что в итоге привело к созданию интеллектуальной системы извлечения информации из

текста, позволяющей облегчить и ускорить процесс поиска информации и знаний в Интернете и представляющей собой практическую значимость данной выпускной квалификационной работы.

Результаты исследования: Было изучено большое количество информации и материала, были достигнуты цели и выполнены поставленные задачи. Была смоделирована интеллектуальная система извлечения информации определенного типа из текста, и затем реализована с применением высокоуровневого языка программирования Python. В качестве извлекаемой информации были выбраны ключевые слова, которые позволяют извлечь также информацию о теме, которой принадлежат обрабатываемые тексты, при помощи методов машинного обучения. После обучения и тестирования на более чем трех тысячах размеченных текстах был получен результат в 89% точности разработанной программы-классификатора (интеллектуальной системы).

Рекомендации: Для получения лучших, более точных результатов работы разработанной интеллектуальной системы следует пополнить базу данных обучающих текстов новыми материалами. В связи с тем, что ИС еще неточно определяет тему текста, если он недостаточно или чересчур объемный, необходимо доработать ее алгоритм и выбрать лучшие параметры классификации.