**ДК 004.91**

ПОДХОДЫ к РАЗРАБОТКЕ СЕРВИСА ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ТЕКСТОВЫХ ДАННЫХ ДЛЯ ВЫЯВЛЕНИЯ СОЦИаЛЬНО-политических трендов

Панфилов Г.О. (Национальный исследовательский университет ИТМО)

Научный руководитель –к. полит. н., доцент Чугунов А.В.

(Национальный исследовательский университет ИТМО)

В данной работе приводится анализ подхода к разработке сервиса интеллектуального анализа текстовых данных для выявления социально-политических трендов. В докладе обозначаются текущие проблемные области и возможности решения существующих барьеров.

**Введение.**

Интеллектуальный анализ текста предполагает автоматизированную обработку текстовых массивов с выделением нужной информации и базируется на таких технологиях как извлечение информации (information retrieval), машинное обучение (machine learning) и компьютерная лингвистика. Методы интеллектуального анализа текста сегодня активно развиваются и применяются для решения исследовательских задач в области общественных наук - социологии, политологии, культурологии и др. В связи с непрерывным увеличением производства текстовых данных и их цифровизации, проблема изучения неструктурированных текстов обретала все новую и новую актуальность. Обработка такого рода информации вручную при помощи классических подходов и методов является невероятно трудоемкой, требует больших ресурсов и не исключает вероятности ошибок, сделанных людьми при обработке.

Выше обозначенные проблемы приводят к необходимости разработки и апробации новых методов и алгоритмов анализа текстов с целью выявления социально-политических трендов.

**Основная часть.**

Специалисты в области интеллектуального анализа текста отмечают, что его современные техники имеют преимущества перед классическими подходами, такими как количественный контент-анализ и некоторыми другими подходами, подразумевающими атомарное восприятие текста. На данный момент существует множество примеров использования методов интеллектуального анализа текста в социальных науках. В качестве примеров направлений исследований с использованием интеллектуального анализа текстов можно привести исследования данных, публикуемых пользователями в социальных сетях для анализа общественного мнения, составления психологических особенностей пользователей социальных сетей, мониторинга и определения тональности публикаций в СМИ, изучения тенденций в научных направлениях по текстам научных публикаций и другие.

Возможность быстрой обработки больших массивов узко-специализированных текстов и внедрение методов, исключающих долгую и трудоемкую ручную обработку, неспособную охватить все генеральную совокупность имеющихся данных, позволит каждому исследователю производить такого рода операции с любыми текстами по интересующей его теме и, в результате, оперативно получать новую информацию по изучаемым тематикам.

Одно из таких направлений - исследование социально-политических процессов, в частности разных направлений государственной политики и управления.

В рамках текущего исследования проводится интеллектуальный анализ корпуса текстов, представляющих собой публикации в новостных электронных ресурсах, посвященные общественно-политическим событиям, связанным с цифровым развитием Российской Федерации. Корпус текстов был собран в 2018-2021 г. Собранные материалы были использованы для подготовки еженедельного новостного бюллетеня «Электронное правительство и электронные услуги». Подборка корпуса производилась по принципу «1 событие = 1 публикация», что исключает возможность дублирования текстов.

Данный корпус является объектом исследования с применением методов интеллектуального анализа для изучения тенденций цифрового развития в России, определения тематической направленности, их классификации, выявления ключевых событий, сущностей и акторов, таких как:

- отраслевое направление внутренней или внешней политики (здравоохранение / образование / транспорт и т.д.),

- упоминаемая технология (искусственный интеллект / BigData и т.д.)

- упоминаемый электронный ресурс (Портал государственных и муниципальных услуг, портал «Наш Санкт-Петербург», ГИС «ЖКХ» и т.д.)

- упоминаемый орган власти, МСУ или организация (Минстрой / Минэкономразвития / Правительство Санкт-Петербурга / АНО «Диалог. Регионы» / Ростелеком и т.д.)

- упоминаемые участники процесса цифрового развития (Министр цифрового развития / Губернатор Санкт-Петербурга / Глава IT-компании и т.д.)

**Выводы.**

 Для решения данной задачи представляется возможным как использовать готовое программное обеспечение, так и разработать принципиально новое решение. Программные средства, реализующие методы интеллектуального анализа текста и предполагающие работу с неструктурированной текстовой информацией сегодня представлены на рынке программного обеспечения (например, RapidMiner, GATE и другие). Данные программные продукты потенциально могут быть задействованы для выполнения задач исследования. Также существуют исследования упомянутых программных продуктов, в которых делаются выводы об эффективности последних. Тем не менее, для более углубленного понимания соответствия их функциональных возможностей задачам исследования требуется произвести углубленный анализ существующего программного обеспечения с обработкой тестового массива данных, являющегося частью корпуса текстов.

|  |  |
| --- | --- |
| Панфилов Г.О. (автор) |   |
|  |  |
| Чугунов А.В. (научный руководитель)  |   |