**ВВЕДЕНИЕ**

В настоящее время в эпоху быстро развивающихся интернет технологий всё большее влияние на жизнь современного человека оказывает компьютеризации. В связи с этим в сфере организации мероприятий происходят существенные изменения направленные на автоматизацию процессов управления событиями, публикацию информации о них, сбор данных об участниках и их отзывах о мероприятиях.

Крупные компании, обладающие большим бюджетом могут позволить себе заказать услугу сопровождения своего мероприятия у сторонней организации, которая позаботится о приглашении гостей, сборе информации, продаже билетов, рекламной акции, размещении информации о событии на специализированных ресурсах. Однако есть большое число компаний, чья бизнес-модель предполагает проведение различны, часто бесплатных мероприятий – конференции, митапы, организация вечеринок с рекламной целью. При этом у данных компаний зачастую отсутствует достаточный объём средств на оплату комплексных услуг сторонних компаний.

Перспективным направлением видится создание сервиса, обеспечивающего автоматизированное сопровождение мероприятий. При этом, данный сервис должен быть оптимизирован не для управления событием, а для привлечения участников событий, так как это есть основная цель организаторов события. Основные функции приложения помимо непосредственного управления событием (создание, изменение, удаление, публикация): фильтрация, поиск, регистрация различных типов (буду участвовать, возможно буду участвовать, не интересует), сбор отзывов, отслеживание информации об отменённых регистрация, интеграция с календарями (Google, Outlook).

Emoticons

Инструменты анализа семантической окраски текста полезны в маркетинге и управлении репутацией [10], то их эффективность часто оценивается на сборниках обзоров, которые обычно содержат мнения людей, выраженные на естественном языке, часто вместе с соответствующей (числовой) оценкой, определяющей суждение. Наиболее часто используемым набором данных для оценки качества алгоритмов является сборник из двух тысяч отзывов о фильмах на английском языке, которые были разделены экспертами в области психологии на 3 группы: негативные, нейтральные и позитивные[25]. Широкое распространение получил подход под названием «Пакет слов». Основная идея которого заключается в поиске специфических слов, которые наиболее сильно влияют на эмоциональную окраску текста. Эффективность этого подхода на упомянутой ранее коллекции отзывов составляет 87.2%. На современном этапе развития технологий в данной области исследования фокусировались на различных векторных представлениях текста, включая векторные представления с дополнительными признаками, представляющими семантические различия между словами [36] или векторные представления с весами на основе tf-idf [24]. Такие подходы обычно дают точность более 90% на наборе данных одной предметной области для которой изначально был настроен алгоритм, при на произвольных наборах данных качество оценки значительно падает. Альтернативные подходы, основанные на словарях, как правило, демонстрируют более низкую точность в наборе данных одной предметной области, но, как правило, более надежны во всех на наборах, составленных из случайно отобранных текстов [34]. Кроме того, основанные на словарях подходы могут быть легко обобщены относительно других языков с помощью словарей [19]. Было показано, что довольно простая структура анализа настроений на основе лексики имеет точность до 59,5% на полном наборе данных обзора фильмов [10]. Более сложный подход к анализу настроений на основе анализа словосочетаний имеет среднюю точность 68.0% на 1900 документов из данных обзора фильмов [33]. Приблизительно 5-ти процентный прирост точности по сравнению с анализом словосочетаний дают алгоритмы, анализирующие структуру текста: наличие сложных предложений, вопросительных, восклицательных, размер предложений. Несмотря на то, что современные подходы к анализу настроений на основе словарей рассматривают новые перспективные направления: анализ структурных и семантических аспектов контента, они обычно не получают информацию из потенциально важных маркеров настроения, которые широко используются в сегодняшнем пользовательском контенте. В качестве таких маркеров выступают смайлы (эмодзи, эмотиконы). Отличительной особенностью смайло является то, что их эмоциональная окраска не зависит от предметной области, к которой принадлежит текст. Смайлы являются важными ориентирами для оценки настроений в современно пользовательском контенте. Ключ к сбору информации из смайликов заключается в понимании того, как они соотносятся с общим настроением текста. В публичных источниках не было найдено исследований, в которых анализировалось совместное использование лексически словарей и словарей смайлов для анализа стилистической окраски текста. Одной из целей данной дипломной работы является разработка подобного алгоритма и применение его для анализа отзывов о мероприятиях.