

УДК 004.652.8

**Анализ современных решений хранения данных специализированного контекста и разнородных данных социальных медиа**

Вишератин А.А. (НИУ ИТМО), Муратов А.А. (НИУ ИТМО)

**Научный руководитель -- к.т.н., доцент Насонов Д.А. (НИУ ИТМО)**

**Аннотация.** Доклад освещает проблему оптимального хранения сложноструктурированных данных из социальных сетей. После анализа существующих технологических решений, было предложено новое, оптимальным образом удовлетворяющее потребность платформы для интеллектуального анализа подобных данных.

**Введение.** В современной России большое внимание уделяется вопросам применения цифровых технологий в решении социальных и экономических задач, стоящих перед обществом, бизнесом и государством. Создаются программы и средства, способные ускорить достижение целей общественными институтами. Особое место в данной среде занимает концепция умного города -- активное использование цифровых технологий для повышения качества жизни. Это порождает спрос на создание и развитие, а также внедрение таких технологий. Неизбежным этапом на пути к цифровой трансформации городской среды является разрешение проблем, связанных с получением, обработкой и хранением необходимых для этого данных. Этот этап является источником большого количества проблем, которые часто оказываются неразрешимыми для государственных органов и частных компаний. К таким проблемам относятся: сбор разнородных данных из социальных медиа, создание инфраструктуры для их хранения и обработки.

**Основная часть.** Таким образом, формируется потребность в платформе, аккумулирующей сырые данные о городе, а также полученную из них дополнительную информацию - специализированный контекст. Такая платформа должна предоставить возможности собирать разнородные данные из различных социальных сетей, извлекать из них контекст с помощью методов машинного обучения, а также совершать к агрегированным по времени и пространству данным запросы. Последнее необходимо для возможности получения общественными институтами ответов на сложные вопросы, которые помогут решить социальные и экономические задачи, стоящие перед городом на пути к цифровой трансформации. Следует подчеркнуть, что последнее должно быть доступно пользователям без глубокой экспертизы в компьютерных науках и без владения значительными вычислительными ресурсами.

Одним из ключевых препятствий на пути к решению данной проблемы является отсутствие распределённой системы хранения данных, подходящей для этого случая. Характерной особенностью данных из социальных медиа является их слабая структурированность, разреженность, но в то же время высокая сложность, разнообразие и наличие связи между различными сущностями. Такие характеристики данных вкупе с требованиями к конечной платформе, делают графовую модель представления информации привлекательной для подобных данных. В то же время, ввиду отсутствия аналогов данной платформе, на рынке отсутствуют решения, оптимизированные под запросы, которые наиболее вероятно будут поступать от пользователей разрабатываемого сервиса. Вместе эти факторы подтолкнули нас к анализу современных решений хранения подобных данных. Это позволило нам убедиться в необходимости разработать свою систему хранения данных, которая будет сконструирована, принимая во внимание набор специфических требований, стоящих перед разработчиками описанной выше платформы.

**Выводы.** Решением описанной выше проблемы будет создание распределённой базы данных, позволяющей быстро выполнять аналитические запросы к сложноструктурированным данным из различных социальных медиа.

Место для подписей:

\_\_\_\_\_ Вишератин А.А.  
\_\_\_\_\_ Насонов Д.А.  
\_\_\_\_\_ Муратов А.А.