**Контекстная обработка данных социальных сетей.**

**Аннотация**

**Введение**

В настоящее время явление социальных сетей достаточно распространено. Социальные сети уверенно вошли в жизнь современного человека и занимают в ней значимую часть. Главным образом они оказывают влияние на поведение, предубеждения, ценности и намерения человека, что отражается во всех сферах его деятельности. Оказываемое влияние, быстрый рост популярности и открытый доступ к контенту привлекли к социальным сетям внимание правительства, финансовых организации и исследователей. Преобразование контента социальных сетей в текстовую информацию и выделение ключевых концепций стало важным условием для порождения знаний и формулирования стратегий. Анализ полученной информации помогает исследователям улучшить понимание об информационных потоках, о формировании и распространении мнений, о связи ценностей и предубеждений пользователя и генерируемого им контента. Тем не менее число качественных и количественных исследований в данной области слишком мало. Самый значительный барьер при использовании социальных сетей - это отсутствие методологии для выбора, сбора, обработки и анализа информации, полученной с сайтов социальных сетей. Однако, существуют компании по производству программного обеспечения, разработавшие проприетарные системы сбора информации для визуализации данных, и исследователи, разработавшие экспертные системы для анализа настроений[1].

Пользователи социальных сетей активно публикуют данные о своей ежедневной активности, чувствах и мыслях, выражая свое мнение и позицию. Благодаря этому в социальных сетях образуются группы пользователей (сообщества), имеющих общие интересы. Для выявления ключевых концепций и тематик присущих группе пользователей используется контекстная обработка данных, генерируемых участниками группы. В данной работе контекстная обработка данных основана на идеях и принципах тематического моделирования. Результаты такой обработки данных могут использоваться для мониторинга заболеваний, например гриппа [2], или для предсказания поведения рынка.

**// Возможно стоит добавить про тематическо моделирование (во введении не нужно)**

**// Примеры социальны сетей Twitter, Facebook, VK, mail.ru, livejournal — можно добавить графики — популярность сетей и количество активных пользователей с 2000 — 2016 года.**

**Постановка задачи**

Целью данной работы является изучение методов контекстной обработки данных, полученных из социальных сетей, основываясь на принципах и идеях тематического моделирования. Для того чтобы достичь поставленной цели предлагается выполнить ряд задач, связанных с анализом данных и реализацией программных модулей. Ставятся следующие задачи:

1. Анализ предметной области
   1. Сравнительный анализ социальных сетей
   2. Выбор социальной сети в качестве источника данных
   3. Выбор данных в социальной сети
2. Реализация программного модуля для сбора данных из социальной сети
   1. Анализ и выбор технологий
   2. Реализация программного модуля
3. Выбор тематической модели
   1. Обзор по тематическом моделированию
   2. Сравнительный анализ тематических моделей
   3. Выбор тематической модели
4. Реализация программного модуля для предобработки данных
   1. Анализ и выбор технологий
   2. Определение основных этапов предобработки данных
   3. Реализация программного модуля
5. Реализация и обучение тематической модели
   1. Анализ и выбор технологий
   2. Выбор способа обучения тематической модели
   3. Реализация программного модуля, реализующего выбранную тематическую модель
   4. Обучение тематической модели на имеющихся данных
6. Анализ полученных результатов
   1. Определение качества тематической модели
   2. Сравнительный анализ с другими тематическими моделями

**Обзор литературы**

**Выводы(Результаты)**

**Заключение**

**Список литературы**

1. Arturas Kaklauskas. Biometric and Intelligent Decision Making Support // Springer, 2015. 220 p.

2. Paul, MJ. and M. Dredze. 2011. You Are What You Tweet: Analyzing Twitter for Public Health. In Proc. of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM).

**Приложение.**