

Аннотация

В данной работе рассматривается проблема контекстной обработки данных, полученных из социальных сетей. В качестве данных рассматриваются текстовые сообщения, генерируемые пользователями в сообществах социальных сетей. Контекстная обработка подразумевает выделение тематик и концепций в коллекции текстовых сообщений, что позволяет проводить анализ сообществ, из которых эти сообщения были взяты. Результаты такого анализа имеют существенное прикладное значение, например, мониторинг заболеваний или политических взглядов групп пользователей социальных сетей.

Целью данной работы является исследование и применение методов тематического моделирования для выполнения контекстной обработки данных. Для достижения указанной цели в работе решаются задачи получения данных из социальных сетей и их предобработки, выбора и построения тематической модели, оценки качества тематической модели и анализа полученных результатов. Выполнение этих задач включает в себя реализацию соответствующих программных модулей. Экспериментальная часть работы связана с построением тематической модели на основе данных, полученных из сообществ социальной сети.

Результаты проведенного исследования показывают, что использование идей тематического моделирования для контекстной обработки текстовых сообщений сообществ позволяет выделить ключевые концепции и тематики этих сообществ.