

**УДК 004.89**

**НЕЙРОННЫЕ МОДЕЛИ ДЛЯ ВЫДЕЛЕНИЯ НАБОРОВ ТЕМАТИК ИЗ ДИАЛОГОВ  
СЛУЖБЫ ПОДДЕРЖКИ**

**Мамаев Н. К.** (Университет ИТМО)

**Научный руководитель – д. т. н. Матвеев Ю. Н.** (Университет ИТМО)

Данная работа посвящена экспериментам по выделению наборов тематик из диалогов службы поддержки. Выделение набора тематик и дальнейшая классификация диалогов на тематики — важный этап оценки качества обслуживания. Многие команды контент-менеджеров до сих пор выполняют эту работу вручную, несмотря на высокий потенциал автоматизации. Цель работы — показать эффективность работы методов кластеризации для решения упомянутых задач.

Основным результатом работы службы поддержки, который может подвергаться дальнейшему анализу, являются аудиозаписи диалоги оператора с клиентом. Эти фонограммы редко используются в исходном виде — как правило, они преобразуются в текстовые транскрипции с помощью систем распознавания речи, поскольку работа с текстовыми данными эффективнее. Тем не менее, анализ текстов до сих пор часто выполняется вручную, что является трудоёмкой и изнуряющей работой, т. к. формулировки и проблематика повторяются из диалога в диалог. По этой причине целесообразно использовать методы автоматического анализа текста, которые ускоренно развиваются и показывают результаты, сравнимые с человеческими.

Для демонстрации эффективности автоматических методов анализа текста, основанных на глубоких искусственных нейросетях, были применены выбранные методы кластеризации на базе моделей BERT, word2vec, TF-IDF. Полученные кластеры были сравнены с кластерами, верифицированными экспертами, были вычислены оценки Rand index и Mutual information. В качестве данных была использована выборка диалогов, составленная в результате работы службы поддержки одной из популярных банковских компаний. Анализу подвергались как целые диалоги, так и отдельные реплики.