

УДК 004.891

Бычкова Мария Владимировна

Студент
Кубанский государственный технологический
университет
Краснодар, Россия
mar.bychkova@gmail.com

Maria V. Bychkova

student
Kuban State Technological University
Krasnodar, Russia

**СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ
ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА ДЛЯ
ОПТИМИЗАЦИИ БИЗНЕС-ПРОЦЕССОВ И
МАРКЕТИНГОВЫХ СТРАТЕГИЙ**

**COMPARATIVE ANALYSIS OF NATURAL
LANGUAGE PROCESSING METHODS FOR
OPTIMIZING BUSINESS PROCESSES AND
MARKETING STRATEGIES**

Аннотация

Статья посвящена сравнительному анализу различных методов NLP на основе существующих решений, таких как контекстно-свободные грамматики (Томиита-парсер), модель BERT и чат-боты на базе GPT. Исследование акцентирует внимание на применимости этих технологий для малых и средних предприятий, где важны простота внедрения, доступность и эффективность. В ходе анализа выявлены основные преимущества и недостатки каждого метода, что позволило определить наиболее перспективные направления для реализации в условиях ограниченных ресурсов. Выводы исследования показывают, что чат-боты являются наиболее подходящим решением для небольших компаний, тогда как BERT обеспечивает высокую точность при анализе сложных текстов, а Томиита-парсер полезен для обработки стандартных документов.

Ключевые слова:

обработка естественного языка, анализ тональности, чат-бот, машинное обучение

Abstract

This article is devoted to a comparative analysis of various NLP methods based on existing solutions, such as context-free grammars (Tomita parser), the BERT model and GPT-based chatbots. The study focuses on the applicability of these technologies to small and medium-sized enterprises, where ease of implementation, accessibility and efficiency are important. The analysis revealed the main advantages and disadvantages of each method, which made it possible to identify the most promising areas for implementation in conditions of limited resources. The findings of the study show that chatbots are the most suitable solution for small companies, whereas BERT provides high accuracy in analyzing complex texts, and Tomita parser is useful for processing standard documents.

Keywords:

natural language processing, sentiment analysis, chatbot, machine learning

Введение

В условиях современной конкуренции бизнес нуждается в инструментах, которые помогают лучше понимать и реагировать на потребности клиентов. Одним из таких инструментов являются технологии обработки естественного языка (NLP), которые позволяют компаниям извлекать полезную информацию из текстовых данных, таких как отзывы, комментарии и публикации в социальных сетях. В данной статье рассматриваются и сравниваются различные методы NLP на примере существующих решений, такие как контекстно-свободные грамматики, BERT и чат-боты, в контексте их применения для оптимизации бизнес-процессов и маркетинговых стратегий.

Это исследование приобретает особую значимость, поскольку NLP уже перестал быть прерогативой только крупных компаний — малый и средний бизнес также начинает активно использовать его возможности для повышения эффективности. Однако для небольших компаний особенно важно найти удобный и доступный инструмент, который можно легко внедрить и адаптировать под их нужды.

Анализ решений для обработки естественного языка

Рассмотрим три существующих решения для анализа текста: Томита-парсер, основанный на контекстно-свободных грамматиках и словарях ключевых слов, модель BERT и чат-боты с использованием GPT. Каждый из этих методов имеет свои плюсы и минусы (табл. 1), что делает их подходящими для различных задач и условий использования.

Таблица 1 – Плюсы и минусы существующих решений

Решение	Плюсы	Минусы
Томита-парсер	Высокая точность при четко определенных правилах. Низкие требования к ресурсам.	Трудоемкость настройки и поддержки. Ограниченная гибкость. Не учитывает контекст.
BERT	Глубокое понимание контекста. Высокая точность и адаптивность.	Высокие требования к вычислительным ресурсам. Сложность внедрения и предобучения.
Чат-боты с GPT	Интерактивность и удобство использования. Мощная обработка естественного языка. Гибкость. Простота внедрения	Высокие вычислительные затраты. Зависимость от внешних API. Потенциальные ошибки при специфических текстах.

Томита-парсер

Томита-парсер — это инструмент для извлечения структурированных данных из текста на естественном языке, разработанный компанией Яндекс [1]. Основой метода являются контекстно-свободные грамматики (CFG) и словари ключевых слов, которые позволяют строить правила для анализа текстов. С помощью Томита-парсера можно анализировать тональность предложений, сопоставляя слова с лексиконом и применяя заранее заданные правила.

Метод обеспечивает высокую точность при четко определенных правилах, однако требует значительных усилий для настройки и не учитывает контекст, что ограничивает его гибкость.

Технология BERT

BERT (Bidirectional Encoder Representations from Transformers) — это модель, разработанная Google [2], которая представляет собой один из самых мощных инструментов для обработки естественного языка. Особенность BERT заключается в его способности учитывать контекст слов в обоих направлениях, что значительно улучшает понимание текста.

BERT позволяет добиться высокой точности при анализе текста, особенно при определении тональности и выделении аспектов текста. Однако модель требует значительных вычислительных ресурсов и специализированных знаний для ее внедрения и адаптации [3].

Чат-боты Telegram

Чат-боты в Telegram [4] предоставляют удобный интерфейс для анализа текстов. Они используют технологии обработки естественного языка и модели машинного обучения, такие как GPT, для автоматического анализа и определения тематики. Чат-боты могут собирать данные с различных платформ, анализировать их тональность и представлять результаты пользователям в удобном формате.

Несмотря на удобство и интерактивность, такие решения требуют значительных вычислительных ресурсов, особенно при использовании моделей типа GPT, и могут зависеть от внешних API [5].

Результаты исследования

Сравнительный анализ методов обработки естественного языка показывает, что наиболее подходящим решением для малого и среднего бизнеса, стремящегося оптимизировать свои процессы и маркетинговые стратегии, являются чат-боты. Они предлагают простоту внедрения и удобство использования, что особенно важно для компаний с ограниченными ресурсами. Модель BERT, благодаря своей высокой точности, идеально подходит для задач, требующих глубокого анализа контекста, и может быть использована крупными компаниями для более сложных бизнес-целей. Томита-парсер, с его возможностями обработки стандартных документов, также имеет свои преимущества, однако его ограниченная гибкость делает его менее подходящим для быстро меняющихся маркетинговых задач.

Список использованных источников

1. Томиита-парсер, документация. Основные понятия [Электронный ресурс]: Режим доступа - <https://yandex.ru/dev/tomita/doc/dg/concept/about.html> (дата обращения 10.07.2024).
2. Google Blog — Understanding searches better than ever before [Электронный ресурс]: Режим доступа - <https://blog.google/products/search/search-language-understanding-bert/> (дата обращения 13.08.2024).
3. Находим главное в отзывах. Опыт разработчиков геопоиска Яндекса [Электронный ресурс]: Режим доступа - <https://habr.com/ru/companies/yandex/articles/665132/> (дата обращения 13.07.2024).
4. Telegram Blog —Bots: An introduction for developers [Электронный ресурс]: Режим доступа - <https://core.telegram.org/bots> (дата обращения 17.07.2024).
5. Telegram Marketplace Bot [Электронный ресурс]: Режим доступа - <https://github.com/qkonstantin/ReviewerBot> (дата обращения 17.07.2024).