

# «ЭМОЦИОНАЛЬНОЕ» И «ФОРМАЛЬНОЕ»: ПРОБЛЕМА ВЫДЕЛЕНИЯ КЛЮЧЕВЫХ СЛОВ КОМПЬЮТЕРНЫМИ ПРОГРАММАМИ В СОПОСТАВЛЕНИИ С МЕТОДИКОЙ ИХ ЭКСПЕРТНОГО ВЫЧЛЕНЕНИЯ<sup>1</sup>

Гамзатова А.Ф., СПбГУ, 2 курс бакалавриата  
Научный руководитель: Шерстинова Т.Ю., к.ф.н.,  
доцент кафедры математической лингвистики СПбГУ

Ключевые слова являются основной составляющей любого текста: именно они обеспечивают его целостность. Ключевые слова (иначе: слова-концепты, лексические доминанты) – «важные слова или фразы, дающие высокоуровневое описание содержания текстового документа, позволяющие выявить его тематику» [1, стр. 85]. Несмотря на прозрачность определения термина, вопрос о том, какие слова в тексте являются ключевыми, остаётся спорным.

В практике выделения ключевых слов существует два подхода: экспертный (осуществляемый людьми) и автоматический (компьютерная обработка текста). Каждый из этих подходов обладает сильными и слабыми сторонами.

**Цель работы:** выявить наиболее продуктивную методику выделения ключевых слов путём интеграции методик их экспертного и компьютерного выделения.

**Основные задачи:** изучить методы выделения ключевых слов людьми (в т.ч. предложенных Ворониной, Коршуновым, Гриневой); проанализировать технологические аспекты автоматического извлечения ключевых слов; определить схожие элементы в методиках выделения слов-концептов экспертным и компьютерным способами.

Главным «минусом» в использовании экспертного подхода выступает тот факт, что ключевые слова выделяются людьми интуитивно [1, стр. 86]. Это обстоятельство ставит под сомнение научность такого подхода.

Ещё один недостаток вычленения ключевых слов вручную – отсутствие корреляции между размером текста и количеством ключевых слов. Ключевые слова должны составлять в тексте 25-30% [2, стр.281]. При выделении их человеком это требование зачастую нарушается.

Кроме того, выделенные вручную слова-концепты не всегда присутствуют в тексте. Эта проблема актуальна при работе с текстами художественной литературы. Читатель склонен придумывать слова, позволяющие отразить его субъективное восприятие прочитанного.

При машинном извлечении ключевых слов выбор лексических доминант на основе эмоций исключается. Невозможна при таком подходе и подстановка новых слов.

Компьютерное вычленение лексических доминант представляется более научным: данные о частотности употребления слова и степени его «keyness» свидетельствуют о доказательности и объективности показателей.

В то же время, при использовании компьютерных программ мы сталкиваемся с проблемой чрезмерной «формализации». Так, служебные слова, вследствие их частой повторяемости, определяются более значимыми, чем знаменательные.

Возникает проблема: «эмоциональность» в вычленении ключевых слов, влекущая за собой некорректность получаемых данных при экспертном подходе, и чрезмерная «формализация», выдвигающая на первый план союзы, не способствующие раскрытию тематики текста при компьютерном подходе.

Поискам путей решения этой проблемы и посвящено настоящее исследование.

---

<sup>1</sup> Исследование выполнено при поддержке гранта РФФИ № 17-29-09173 «Русский язык на рубеже радикальных исторических перемен: исследование языка и стиля предреволюционной, революционной и постреволюционной художественной прозы методами математической и компьютерной лингвистики (на материале русского рассказа)».

Основными методами исследования выступают **эксперимент** и **сравнение**: сопоставительный анализ данных, полученных путём экспертного и автоматического выделений ключевых слов текста.

Исследование выполняется на материале русского рассказа. Малый размер и лаконичность изложения основной идеи текста способствуют проведению детального анализа.

Автоматическое выделение ключевых слов происходит при помощи компьютерной программы AntConc. Возможность обработки текстов на русском языке и многообразии предоставляемых опций обосновывают выбор этой программы.

Практическая значимость настоящего исследования состоит в применении разработанной методики при создании корпусов текстов, поиска текста определённой тематики в базах.

### *Литература*

1. Ванюшкин А.С., Гращенко Л.А. Методы и алгоритмы извлечения ключевых слов // Новые информационные технологии в автоматизированных системах. 2016. №19. С. 85-87.
2. Москвитина Т.Н. Ключевые слова и их функции в научном тексте // Вестник Челябинского государственного педагогического университета. 2009. № 11. С. 270-283.