

## Обзор предметной области

Научный стиль - наиболее строгий стиль речи, используемый для написания научных статей. Характеризуется использованием научной терминологии, исключая жаргонизмы. Научный стиль не допускает личного изложения [1]. Проверять текст на соответствие научному стилю, следует в первую очередь реализовать и базовую проверку на качество текста. К такого рода анализу можно отнести SEO-анализ. SEO (search engine optimization) анализ [2-3] популярен и актуален в связи с необходимостью продвижения ресурсов, товаров и услуг в сети Интернет. SEO анализ текста дает возможность понять, насколько часто употребляются ключевые слова в тексте, как много в тексте слов, не имеющих смысловой нагрузки и т.д. SEO-анализе вводит следующие термины для двух критериев, которые проверяются в данной работе: Тошнота – это показатель повторений в текстовом документе ключевых слов и фраз. Синонимом тошноты является термин плотность [2]. Вода - процентное соотношение стоп-слов и общего количества слов в тексте [2]. Так как эти критерии вычисляемы, то можно автоматизировать их получение, и существуют веб-сервисы, проверяющие текст по этим критериям - сервисы, позволяющие провести SEO-анализ текста. Рассмотрим некоторые из них.

### **1y.ru**

Анализатор качества контента [4]. Анализ проводится на базе закона Ципфа, то есть качество текста в данном случае определяется на основании соответствия частоты употребления слов в естественной речи и тексте. Результат выдается в двух окнах: в одном — график, в другом — частота использования отдельных слов и рекомендации по корректировке.

### **text.ru**

Сервис проверки текстов по многим параметрам, включая уникальность, проверку орфографии, выделение ключевых слов [5].

### **contentmonster.ru**

Сервис, осуществляющий поиск стоп-слов и подсчет их процентного соотношения к общей длине текста [6]. Определяет стоп-слова как всё то, что не

несет самостоятельной смысловой нагрузки, но без чего не бывает связных текстов: предлоги, частицы, междометия, причастия, союзы, а также некоторые наречия, существительные и глаголы. Слишком большое количество таких слов затрудняет восприятие текста и увеличивает его водность.

### **Критерии сравнения аналогов**

Сравнение аналогов будет проводиться по следующим критериям:

- Многокритериальная проверка - как много критериев проверки использует сервис;
- Ограничение длины текста - отсутствие ограничения длины текста, поступающего на проверку;
- Проверка научного стиля - проверка текста на соответствие научному стилю.

В табл.1 представлено сравнение аналогов.

Таблица 1 - Сравнение аналогов

Аналог	Многокритериальная проверка	Ограничение длины текста	Проверка научного стиля
ly.ru	-	-	-
text.ru	+	+	-
content-monster.ru	+	-	-

Результаты сравнения показывают, что часть существующих сервисов предлагает многокритериальную проверку текста, при этом, не ограничивая его по длине. Но все аналоги осуществляют SEO-проверку, ни один из них не реализует проверку статьи на соответствие научному стилю.

Так же существует эмпирическая закономерность распределения частоты слов естественного языка - Закон Ципфа: если все слова языка или достаточно длинного текста упорядочить по убыванию частоты их использования, то частота  $n$ -го слова в таком списке окажется приблизительно обратно пропорциональной его порядковому номеру  $n$  [7-8]. Соответствие распределения слов в тексте закону Ципфа говорит об уровне его естественности. Расчет этого критерия так же можно автоматизировать. Также важными

показателями научной статьи являются её экспертность и полезность. На данный момент это может проверить только специалист в данной области, но разработки подобных инструментов ведутся [9].

#### Список литературы

1. Демидова А. К. Пособие по русскому языку: научный стиль, оформление научной работы. – Рус. яз., 1991.
2. Davis H. Search engine optimization. – " O'Reilly Media, Inc.", 2006.
3. Словарь терминов семантического анализа. // URL: [seopult.ru/library](http://seopult.ru/library)
4. Сервис оценки качества текста. // URL: [1y.ru](http://1y.ru)
5. Сервис оценки качества текста. // URL: [text.ru](http://text.ru)
6. Сервис оценки качества текста. // URL: [contentmonster.ru](http://contentmonster.ru)
7. Newman M. E. J. Power laws, Pareto distributions and Zipf's law //Contemporary physics. – 2005. – Т. 46. – №. 5. – С. 323-351.
8. Lelu A. Jean-Baptiste Estoup and the origins of Zipf's law: a stenographer with a scientific mind (1868-1950) //Boletín de Estadística e Investigación Operativa. – 2014. – Т. 30. – №. 1. – С. 66-77.
9. Dong X. L. et al. Knowledge-based trust: Estimating the trustworthiness of web sources //Proceedings of the VLDB Endowment. – 2015. – Т. 8. – №. 9. – С. 938-949.