

Мешкова Екатерина Владимировна,
кандидат технических наук, преподаватель,
Колледж экономики и сервиса института
сферы обслуживания и предпринимательства (филиала)
Донского государственного технического университета
(КЭС ИСОиП (ф) ДГТУ), город Шахты

Кривошеков Владимир Александрович, студент,
Колледж экономики и сервиса института
сферы обслуживания и предпринимательства (филиала)
Донского государственного технического университета
(КЭС ИСОиП (ф) ДГТУ), город Шахты

ОПРЕДЕЛЕНИЕ АВТОРСКОГО СТИЛЯ: КТО ВЫ – ГОРЬКИЙ ИЛИ ГОГОЛЬ? DEFINING THE AUTHOR'S STYLE: ARE YOU GORKY OR GOGOL?

Аннотация: В работе объединены два подхода к анализу произвольного текста: статистический метод и метод морфологического анализа. Рассмотрена идея о том, что вектор морфологических признаков текста достаточно уникален для каждого автора. Следовательно, имея базу данных векторов морфологических признаков текстов различных авторов, возможно определение совпадения (или близости) авторского стиля для произвольного текста.

Abstract: The paper combines two approaches to the analysis of arbitrary text: the statistical method and the method of morphological analysis. The idea that the vector of morphological features of a text is quite unique for each author is considered. Therefore, having a database of vectors of morphological features of texts by different authors, it is possible to determine the coincidence (or proximity) of the author's style for any text.

Ключевые слова: авторский стиль, морфологический анализ, классификация текстов.

Keywords: author's style, morphological analysis, text classification.

В работе [1] была изложена идея о том, что вектор морфологических признаков текста несет информацию об авторском стиле произведения. Следовательно, имея базу данных векторов морфологических признаков текстов различных авторов, возможно определение авторского стиля (или близости стиля произвольного произведения) тому или иному известному писателю [2].

Рассмотрим кратко данный подход.

В представленной работе объединены два подхода к определению авторского стиля: статистический метод и метод морфологического анализа. В качестве основной характеристики стиля рассматривается количественный состав и соотношение частей речи в тексте.

В исследовании были получены индивидуальные характеристики различных писателей. Например, в показанной ниже таблице представлены относительные характеристики (частоты) результатов анализа трех произведений Н.В. Гоголя.

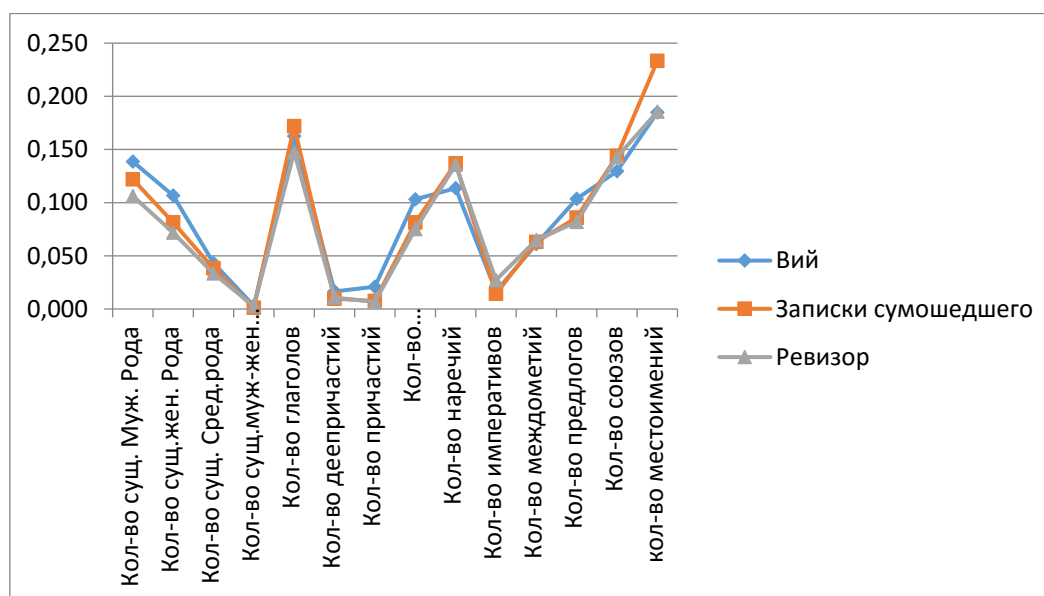
Таблица 1

<i>Относительное кол-во (частота)</i>	<i>Вий</i>	<i>Записки сумасшедшего</i>	<i>Ревизор</i>
<i>Кол-во сущ. Муж. Рода</i>	0,139	0,122	0,106
<i>Кол-во сущ.жен. Рода</i>	0,107	0,081	0,072
<i>Кол-во сущ. Сред.рода</i>	0,043	0,038	0,033



<i>Кол-во сущ.муж-жен рода</i>	0,003	0,001	0,003
<i>Кол-во глаголов</i>	0,163	0,172	0,147
<i>Кол-во деепричастий</i>	0,016	0,010	0,011
<i>Кол-во причастий</i>	0,021	0,008	0,007
<i>Кол-во прилагательных</i>	0,103	0,081	0,075
<i>Кол-во наречий</i>	0,114	0,137	0,136
<i>Кол-во императивов</i>	0,017	0,014	0,027
<i>Кол-во междометий</i>	0,061	0,063	0,065
<i>Кол-во предлогов</i>	0,104	0,086	0,082
<i>Кол-во союзов</i>	0,130	0,144	0,143
<i>Кол-во местоимений</i>	0,185	0,233	0,186

На рисунке приведены соответствующие гистограммы:



Из рисунка видно, что количественные характеристики всех трех произведений близки.

Проведен эксперимент по определению авторского стиля на основе полученных характеристик путем сравнения усредненных характеристик произведений различных авторов (определение близости векторов методом Декарта или Чебышева [2]).

Очевидно, что рассмотренный подход может быть применен и на практике.

Например, в одном из исследований была рассмотрена методика использования данной идеи для формирования функционально-ролевых коммуникационных кластеров на основе морфологических особенностей вербального контекста.

Мы же предлагаем на основании рассмотренного метода построение web ресурса для реализации следующего функционала.

1. Определение степени близости вашего стиля к какому-то автору из базы данных ресурса.
2. Сравнение с конкретным автором.



3. Коррекция своего произведения, путем изменения количества тех или иных морфологических параметров, с целью приблизить текст к стилю выбранного автора.

Отметим, что близость вашего стиля к стилю Льва Толстого не делает вас Львом Толстым.

Первоначально для разработки приложения словари устанавливаются вместе с программой RML (Русская морфология) от компании АОТ (поставщика морфологического модуля) [4], и во время разработки программа обращается к словарям, как части программы RML [3]. Однако в дальнейшем словари интегрируются в проект.

Разрабатываемое приложение строится как сайт. Анализируемый текст загружается либо как текстовый файл, либо фрагмент текста копируется в окно интерфейса. В перспективе это может быть реализовано и как мобильное приложение

В конце исследования отметим, что вектор морфологических признаков текста несет информацию об авторском стиле произведения. Следовательно, имея базу данных векторов морфологических признаков текстов различных авторов, возможно определение авторского стиля (или близости стиля произвольного произведения) тому или иному известному писателю.

Список литературы:

1. Мешков В.Е., Мешкова Е.В. Определение авторского стиля на основе статистическо-морфологического анализа произведения. Теория операторов, комплексный анализ и математическое моделирование: Тезисы докладов XIII Международной научной конференции (пос. Дивноморское, 7-14 сентября 2016г.). – Владикавказ: ЮМИ ВНЦ РАН, 2016. – 257с.

2. Мешков В.Е., Мешкова Е.В. Статистическо-морфологический метод анализа для определения авторского стиля. В сборнике: Science: discoveries and progress Proceedings of articles II International scientific conference. Editors F.I. Kevlja, M.A. Derho, T.F. Kosyreva, S.S. Kugaevskij. Karlovy Vary – Moscow, 28–29 апреля 2017 года.

3. Основные интерфейсы компонента морфологического анализа [Электронный ресурс]: Компания АОТ, 2016. URL: <http://www.aot.ru>

4. Русский морфологический словарь Диалинг [Электронный ресурс]: Компания АОТ, 2017. URL: <http://www.aot.ru>

