

Фомичев Ю.С.

Научный руководитель: А.Д. Варламов

**Муромский институт (филиал) федерального государственного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирская обл., ул. Орловская, 23
Россия, 602264, Владимирская область, г. Муром, ул. Орловская, д.23
E-mail: foma532.yurok@yandex.ru**

Поиск сюжетных событий в литературных произведениях на основе анализа частот упоминания персонажей

В настоящее время автоматическая обработка текста – это стремительно развивающаяся область научных исследований, которая направлена на разработку алгоритмов и методов обработки больших объемов неструктурированной информации. Основные трудности, возникающие при решении задач обработки текстов, связаны с необходимостью работы с неструктурированными данными [3, 4]. Поэтому еще не удалось достичь универсального подхода к построению алгоритма, так как каждый конкретный алгоритм определяется строем языка.

На практике анализом больших текстов занимаются литературные критики и рецензенты. Статьи, написанные русскими и зарубежными критиками, продолжают и сегодня оказывать большое влияние на духовную и нравственную жизнь общества [2]. В процессе анализа произведений критикам необходимо не только прочесть большой объем текста, но и многократно возвращаться к его эпизодам для более детального их анализа, оценивать своего рода статистику произведения и так далее. Поэтому много времени ими тратится на поиск интересных их частей произведения.

Современным критикам можно существенно упростить задачу исследования произведений с помощью информационных систем, быстро и эффективно выполняющих поиск по тексту [1, 5], что позволит, в общем, и исключить рутинный поиск в тексте при работе рецензента, чтобы сосредоточиться на сути анализа и не отвлекаться по мелочам.

В процессе исследовательской работы было установлено, что большой информативностью обладает визуализация периодичности появления героев в произведении. Поэтому был разработан алгоритм визуализации активности героев в литературных текстах, который состоит из следующей последовательности операций:

1. Предварительная обработка произведения: разбиение текста на предложения, удаление лишних символов в начале и конце каждого предложения;
2. Поиск слов, начинающихся с заглавной буквы, кроме тех, что стоят в начале предложений, и оценка частот их встречаемости в тексте;
3. Отсевание редких слов из результатов поиска
4. Объединение однокоренных слов в списке найденных, и суммирование их частот;
5. Пересчет частот слов из списка найденных с учетом первых слов в предложениях;
6. Визуализация результатов поиска: построение графика частоты упоминания выбранного героя (или группы героев) во временном срезе.

Предложено использование этого алгоритма для быстрого поиска в литературных текстах сюжетных событий, сцен, актов, опираясь на участников этих событий – персонажей произведения.

В ходе проделанной работы был достигнут результат в виде функционирующей программы, которая облегчит литературным критикам и рецензентам анализ произведений и позволит быстро и продуктивно, перемещаясь в ту или иную часть произведения, достигать более точных и конкретных результатов в своей работе, затрачивая на это в разы меньше времени, чем при классическом поиске в произведении.

Литература

1. Алексеев А.А., Катасёв А.С., Кириллов А.Е., Кирпичников А.П. Классификация текстовых документов на основе технологии text mining // Вестник Казанского технологического университета, 2016, Т19, №18, с. 116 – 119.
2. Канунова Е.Е. Вопросы автоматизации музейного дела // Алгоритмы, методы и системы обработки данных. 2014. № 4 (29). С. 72-76.
3. Ленкин А.В., Баженов Р.И. Исследование систем для Text Mining // Постулат. 2017. № 1 (15). С. 3.
4. Шарапова Е.В., Шарапов Р.В. Универсальная система проверки текстов на плагиат "автор.net" // Информатика и ее применения. 2012. Т. 6. № 3. С. 52-58.
5. Щербатов И.А., Беляев И.О. Многоагентная поисковая система: применение фибоначчиевх куч // Алгоритмы, методы и системы обработки данных. 2015. № 2 (31). С. 86-92.