

Бастова В.Е.

*Научный руководитель: к.т.н., доцент К. В. Макаров  
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  
E-mail: 254207095@mail.ru*

### **Разработка информационной системы анализа ВКР на наличие заимствований**

Высшее образование во все времена требовало от студентов оформление всех курсовых и выпускных работ в основном, в текстовой форме. Поэтому актуально создание приложений, которые позволят определять списывание текста.

Основная задача таких приложений не только сообщить заинтересованному пользователю списан ли тот или иной текст, но и показать какой именно фрагмент был заимствован и с каких источников.

Заимствование (плагиат) - вид нарушения прав автора или изобретателя. Состоит в незаконном использовании под своим именем чужого произведения (научного, литературного, музыкального) или изобретения, рационализаторского предложения (полностью или частично) без указания источника заимствования [1]. Так же вместо слова «плагиат» используется термин «нечеткий дубликат».

Разработка приложения по обнаружению заимствований затрудняется по ряду причин, таких как:

- Большой объем данных, который хранится в базах;
- Обеспечение быстродействия системы;
- Предоставление точного анализа текста и др.

В настоящее время существует множество способов нахождения нечетких дубликатов текстовых документов. Все способы можно разделить на два класса. Алгоритмы, которые используют общую статистику о частотных характеристиках всей коллекции документов, называют глобальными, а другие – локальными, принцип которых основан на синтаксическом анализе документа.

Целесообразно представить информационную систему в виде веб-ресурса, с помощью которого имеется возможность с любого компьютера, подключенного к глобальной сети Internet, проверить электронный вариант работы на наличие заимствований.

Разрабатываемая система будет реализована с помощью одного из локальных алгоритмов основанный на представлении текста этих документов в виде совокупности последовательностей определенной длины, состоящей из соседних слов.

Главными функциями разрабатываемой системы являются:

- Сохранение данных о выпускной работе;
- Нахождение процента уникальности текста;
- Выделение цветным маркером места заимствований;
- Определение первоисточника заимствования;
- Подсчет символов и слов в работе.

Так как система рассчитана на работу с базой данных выпускных работ прошлых лет, основным пользователем разрабатываемой системы будет являться преподавательский состав учебного заведения. Внедрение подобной системы в образовательную среду не только облегчит труд преподавателей и в разы ускорит процесс обнаружения скопированных и перефразированных фрагментах в текстах выпускных работ, но и поможет студентам подготовиться к самостоятельному написанию качественной выпускной квалификационной работы.

### **Литература**

1. Сухарев А.Я., Крутских В.Е. Большой юридический словарь (редакция) - Москва, 2002;
2. Зеленков Ю.Г., Сегалович И.В. Сравнительный анализ методов определения нечетких дубликатов для Web-документов, [http://rcdl2011.vsu.ru/doc/full\\_text/Report.7.1.pdf](http://rcdl2011.vsu.ru/doc/full_text/Report.7.1.pdf)