

Министерство образования Республики Беларусь  
Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК 004.421:343.533-047.27

Иванов  
Илья Петрович

АЛГОРИТМЫ ПОИСКА ПЛАГИАТА В ИСХОДНОМ КОДЕ

**АВТОРЕФЕРАТ**

на соискание степени магистра технических наук

по специальности 1–40 80 02 «Системный анализ, управление и обработка  
информации»

---

Научный руководитель  
Курулев Александр Петрович  
кандидат технических наук, доцент

---

Минск 2022

## КРАТКОЕ ВВЕДЕНИЕ

Проблема плагиата очень стара и не теряет своей актуальности по сей день. Вопрос плагиата исходного кода является одной из проблем современного образования, научно-исследовательской сферы и сферы информационных технологий в целом. Плагиат - это умышленно совершаемое физическим лицом незаконное использование или распоряжение охраняемыми результатами чужого творческого труда, которое сопровождается доведением до других лиц ложных сведений о себе как о действительном авторе.

Особенно остро проблема плагиата стала видна с ковидными ограничениями, когда большое количество учебных заведений, будь то школы или вузы, оказались вынуждены переносить свои учебные процессы на онлайн-платформы. Также сильно возросла популярность всех онлайн курсов, в том числе связанных с информационными технологиями.

Плагиат так популярен, благодаря тому, что он предоставляет набор определенных преимуществ. Причины могут быть разными:

- сокращение трудозатрат на разработку и/или тестирование программного обеспечения;
- повышение производительности ПО;
- компенсация недостатка времени на переработку исходного кода или недостаточным количеством знаний о разрабатываемой системе у программиста;
- намеренный плагиат и присвоение интеллектуальной собственности.

В образовательном процессе плагиат является отрицательным явлением как для студента, так и для преподавателя. Студенты идут по пути наименьшего сопротивления и плохо усваивают материал и, как следствие, хуже разбираются в предметной области. Преподаватели же, вместо оценки реальных знаний и навыков студента, вынуждены тратить время и силы на выявление плагиата, что влияет на продуктивность занятий в целом.

В коммерческой сфере плагиат, хоть и может быть полезным на ранних этапах разработки, но при вскрытии своего факта может принести большие иски в сторону компании и большие репутационные потери.

Отсюда появляется потребность в защите авторских прав и совершенствовании алгоритмов, которые помогают выявлять факты заимствований.

Существует множество систем для проверки плагиата исходного кода, но ни одна из них не является устойчивой к незначительным изменениям структуры исходного кода, которые не влияют на итоговое поведение программы. Такие изменения производятся умышленно, с целью запутать систему проверки. Как правило, такие изменения возможны благодаря диверсификации инструментов того или иного языка программирования и могут быть автоматизированы, так как не требуют значительных интеллектуальных усилий.

# ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Работа выполнялась в соответствии с научно-техническим заданием и планом работ кафедры «Информационные технологии автоматизированных систем» по теме «Алгоритмы поиска плагиата в исходном коде».

## **Актуальность темы исследования**

Образовательная сфера, в настоящее время, испытывает возрастающую нагрузку. Процессы цифровизации и перехода образования на онлайн рельсы не останавливается и становится только быстрее. Отсюда вытекает потребность в эффективных алгоритмах поиска плагиата для проверки учебных работ. Также существует высокий спрос на защиту авторского права и совершенствования алгоритмов поиска плагиата в сфере коммерческой разработки.

## **Степень разработанности проблемы**

Существует множество систем для проверки плагиата исходного кода, но ни одна из них не является устойчивой к незначительным изменениям структуры исходного кода, которые не влияют на итоговое поведение программы. Такие изменения производятся умышленно, с целью запутать систему проверки. Как правило, такие изменения возможны благодаря разнообразию инструментов того или иного языка программирования и могут быть автоматизированы, так как не требуют значительных интеллектуальных усилий.

## **Цель и задачи исследования**

*Целью* магистерской диссертации является исследование возможностей повышения эффективности алгоритмов для поиска плагиата в исходном коде.

*Объектом* исследования выступают алгоритмы для поиска плагиата в исходном коде.

*Предметом* исследования является повышение эффективности алгоритмов для поиска плагиата в исходном коде.

Для достижения поставленных целей были выделены следующие задачи:

1. Анализ методов и алгоритмов поиска плагиата в исходном коде.
2. Поиск способов повышения эффективности алгоритма с использованием принципов мутационного тестирования.
3. Реализация программного средства и оценка эффективности модифицированного алгоритма.

## **Теоретическая и методологическая основа исследования**

При проведении исследования и написании диссертации использованы научные публикации, техническая документация и интернет-источники, посвященные вопросам поиску плагиата в свободном тексте и исходном коде, а также принципам мутационного тестирования.

Для решения поставленных задач использованы следующие методы исследования: анализ, синтез, обобщение, сравнение, логический и графический методы.

В работе применялись методы компьютерного проектирования.

### **Научная новизна**

*Научная новизна* заключается в модификации алгоритма для поиска плагиата в исходном коде с использованием принципов мутационного тестирования, что позволит противостоять заимствованиям, которые представляют собой незначительные структурные изменения в исходном коде, которые не влияют на работу программы, и, тем самым, повысить его эффективность.

*Теоретическая значимость* работы заключается в анализе алгоритмов для поиска плагиата в исходном коде и принципов мутационного тестирования и модификации алгоритма.

*Практическая значимость* работы заключается в модифицированном алгоритме и построении программного средства.

### **Основные положения, выносимые на защиту**

1. Анализ методов и алгоритмов поиска плагиата в исходном коде.
2. Анализ принципов мутационного тестирования.
3. Реализация нового алгоритма поиска плагиата в исходном коде на основе базового алгоритма и принципов мутационного тестирования.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения, библиографического списка и приложения.

**В первой главе** проведен анализ существующих алгоритмов для поиска плагиата в исходном коде, рассмотрены наиболее популярные системы поиска плагиата в исходном коде. Также были исследованы принципы мутационного тестирования и возможности их использования, для повышения эффективности базового алгоритма.

**Во второй главе** представлен синтез нового алгоритма путем совмещения существующего алгоритма для поиска плагиата в исходном коде и принципов мутационного тестирования.

**В третьей главе** выполнена реализация усовершенствованного алгоритма, описанного во второй главе, с пояснениями и уточнениями.

**В четвертой главе** представлен анализ результатов, полученных с помощью реализованного программного средства, относительно базового алгоритма по временным показателям и эффективности.

Общий объем диссертационной работы составляет 82 страницы. Работа включает в себя 19 иллюстраций, 3 таблицы, 1 приложение, 1 публикацию магистранта. Библиографический список состоит из 32 наименований.

## ЗАКЛЮЧЕНИЕ

Основной целью данной магистерской диссертации является исследование возможностей повышения эффективности алгоритмов для поиска плагиата в исходном коде. А главными задачами можно назвать разработку и реализацию модифицированного алгоритма поиска плагиата в исходном коде.

Результатом выполнения данной диссертационной работы является разработанный модифицированный алгоритм поиска плагиата в исходном коде.

В первом разделе был проведен анализ методов и алгоритмов поиска плагиата в исходном коде. Были детально изучены существующие методы, алгоритмы и системы поиска плагиата в исходном коде. Также были изучены принципы мутационного тестирования ПО. Была представлена классификация плагиата исходного кода. На основе анализа существующих систем анти-плагиат, было установлено, что современные системы обладают низкой устойчивостью к заимствованиям 3 типа. Сами такие заимствования могут быть произведены автоматически. Поэтому было выдвинуто предположение, что процесс установления факта таких заимствований также можно автоматизировать при помощи принципов мутационного тестирования.

Во втором разделе был синтезирован новый алгоритм поиска плагиата исходного кода, основанный на некоем базовом алгоритме и модификации, которая основана на принципах мутационного тестирования. Эта модификация позволит базовому алгоритму эффективнее противостоять заимствованиям 3 типа. Новый алгоритм позволит, за счет мутаций, находить более длинные цепочки плагиата и, тем самым, преодолевать порог чувствительности базового алгоритма.

В третьем разделе было проведено системное и функциональное проектирование разрабатываемого программного средства. Приложение было разбито на крупный блок и входящие в него модули. Для каждого модуля четко отведена роль в разрабатываемом ПС, а также определен набор классов и абстракции для ее реализации. Также был проведен анализ и обоснован выбор ЯП, платформы и средств разработки. На основе представленного системного и функционального проектирования при помощи выбранных средств разработки, было реализовано приложение, отвечающее исходным критериям.

В четвертом разделе были проведены испытания разработанного ПС. В качестве исходных данных использовались программы с разным количеством строк исходного кода с заимствованиями 3 типа. Был проведен сравнительный анализ эффективности разработанного алгоритма относительно базового, а также замерены временные показатели разработанного ПС. Тестовые запуски были проведены многократно и по итогу были вычислены средние значения.

На основе полученных данных видно, что разработанный алгоритм дает существенный прирост эффективности относительно базового алгоритма

против заимствований 3 типа. Также было выяснено, что разработанный алгоритм является ресурсозатратным по времени. Однако, был предложен ряд мер по устранению данного недостатка, которые трудно было реализовать в рамках создания прототипа.

На основании всего изложенного, считаю, что поставленная цель достигнута в полной мере. В процессе разработки эффективность алгоритмов для поиска плагиата в исходном коде была повышена. Выполненная работа принесла мне большой опыт как молодому специалисту и открыла много возможностей для дальнейшего развития в сфере IT-технологий.

Библиотека БГУИР

## **СПИСОК ОПУБЛИКОВАННЫХ РАБОТ**

1-А Иванов, И.П. Анализ методов и алгоритмов поиска плагиата в исходном коде / И.П. Иванов // Материалы 56-ой научной конференции студентов, магистрантов, аспирантов УО «Белорусский государственный университет информатики и радиоэлектроники» Минск, БГУИР, 2019.

Библиотека БГУИР