

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФГБОУ ВО «МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ  
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»

# АЛГОРИТМ ВЫЯВЛЕНИЯ ПЛАГИАТА В ИСХОДНОМ КОДЕ ПРОГРАММ

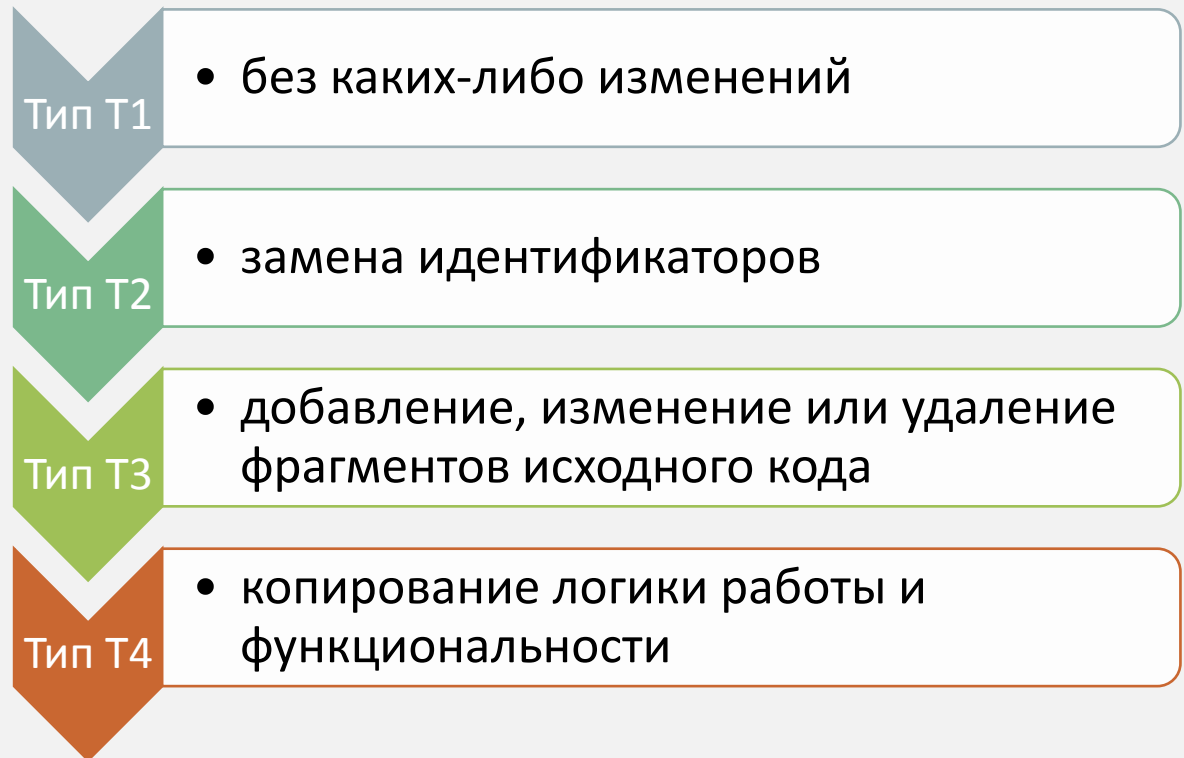
Автор: Морозов П. А., студент каф. 319

E-mail: morozov\_p.a@bk.ru

**Москва 2021**

# ПЛАГИАТ В ИСХОДНОМ КОДЕ ПРОГРАММ

- Плагиат – это использование чужого труда без указания источника с присвоением авторства.
- Плагиат исходного кода программ производится путем копирования с возможными изменениями.



## АКТУАЛЬНОСТЬ РАБОТЫ

**Плагиат во время учебной деятельности приводит к снижению качества получаемого образования и становится привычкой, которая переносится на рабочие процессы после учебы.**

- Копирование текстов работ
- Копирование чужих учебных программ
- Плагиат кода при разработке программных продуктов

# РЕШАЕМЫЕ ЗАДАЧИ

- Цель работы - разработать алгоритм, обеспечивающий автоматическое выявление некорректных заимствований исходного кода программ.
- Для выполнения поставленной цели:
  - Рассмотрены основные типы заимствований
  - Проанализированы способы представления исходного кода и алгоритмов выявления заимствований
  - Изучены алгоритмы анализа текстов
  - Предложен алгоритм определения заимствований

# МЕТОДЫ ПРЕДСТАВЛЕНИЯ ПРОГРАММ

- Текстовое представление
- Представление в виде абстрактного синтаксического дерева
- Токенизированное представление
- Представление в виде метрик программы
- Представление в виде ориентированного графа

# ВЫБОР МОДЕЛИ

Поддержка  
нескольких языков  
программирования

Высокая точность для  
небольших по объему  
исходного кода  
программах

Поддержка  
нахождения как  
можно большего  
количества типов  
заимствования

## ПРИМЕР ФРАГМЕНТ А СЛОВАРЯ ТОКЕНОВ ДЛЯ ЯЗЫКА C++

Идентификатор	Описание	Примеры
CYCLE	Циклы	for, do, while
CONDITION	Условные операторы	if, else
NUMBER	Числовые значения и константы	26, 4.2, -2.2e-8
TYPE	Типы переменных	bool, char, int, double
OPERATOR	Математические операторы и равенство	=, +, -, *, /, %
COMPARE	Операторы сравнения	==, >, <, >=, <=, !=
LOGIC	Логические операторы	&&,
BITS	Побитовые операторы	&,  , ^

# АЛГОРИТМ ВЫЯВЛЕНИЯ ПЛАГИАТА В ИСХОДНОМ КОДЕ ПРОГРАММ

- Для алгоритма Хескела:

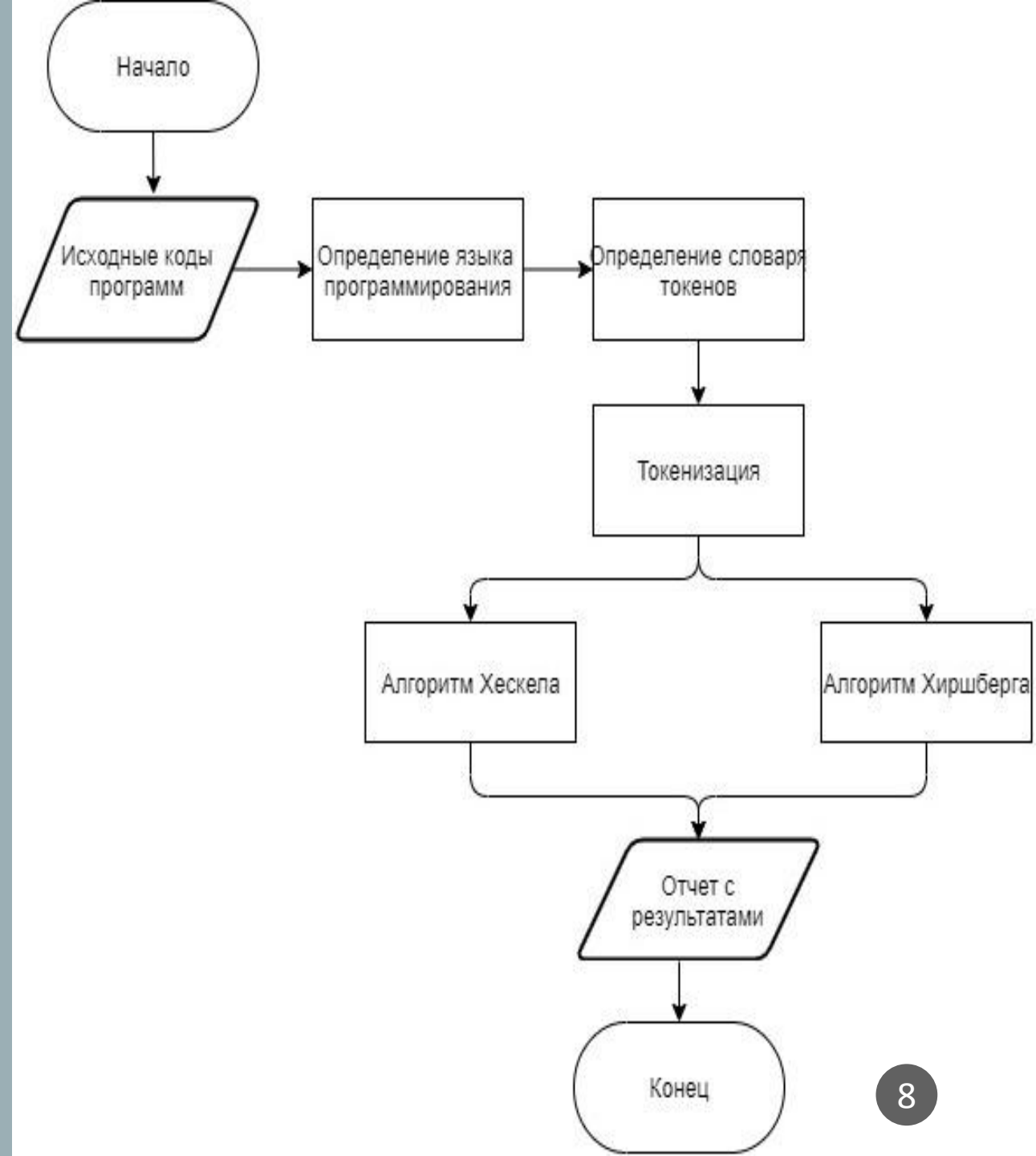
Вероятностная мера Иверсена, где  $A, B$  – множества выявленных непересекающихся подстрок в сравниваемых программах:

$$K(A, B) = \frac{|A \cap B|}{|A \cup B|} * 100$$

- Для алгоритма Хиршберга:

Процент схожести двух программ, где  $A, B$  – длины проверяемых строк,  $LD$  – расстояние Левенштейна:

$$K(A, B) = \left( 1 - \frac{LD}{\text{Max}(A, B)} \right) * 100$$





# ПРИМЕНЕНИЕ АЛГОРИТМА И ЕГО РАЗВИТИЕ

Реализация системы  
выявления плагиата для  
различных языков  
программирования.

Адаптация системы для  
применения к проверке  
учебных задач

Определение процента  
схожести двух программ,  
который может быть  
использован для оценки  
присутствия плагиата.

Поддержка нескольких  
языков программирования

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФГБОУ ВО «МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ  
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»

СПАСИБО ЗА ВНИМАНИЕ

Составитель: Морозов П. А.

E-mail: morozov\_p.a@bk.ru

**Москва 2021**