

## Алгоритмы сравнительного анализа фрагментов кода на основе теории компиляции

*Тямгин Иван Анатольевич*

КРЫМСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ ИМЕНИ В.И. ВЕРНАДСКОГО  
ТАВРИЧЕСКАЯ АКАДЕМИЯ  
ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАТИКИ  
КАФЕДРА ИНФОРМАТИКИ (ГРУППА 601И)

e-mail: tyamgin@mail.ru

*В работе рассматриваются методы обнаружения плагиата в исходном коде программ. Суть плагиата заключается в неправомерном использовании чужого исходного кода программы. Проблема в том, что обнаружить плагиат в этом случае, более трудная задача, нежели обнаружить плагиат в текстовом материале. При сравнении исходных кодов программ необходимо анализировать суть исходного кода, ведь перестановка строк и замена имен переменных в коде программы приведут к значительным изменениям во внешнем виде кодов, но не приведут к изменениям в структуре кода и функциональном устройстве программы.*

Рассмотрим самые распространенные приёмы скрытия плагиата на следующих двух программах (Рис. 1, 2), вычисляющих биномиальные коэффициенты:

```
public class Program
{
    static int Factorial(int n)
    {
        var result = 1;
        for (var i = 2; i <= n; i++)
            result *= i;
        return result;
    }

    static int Binomial(int n, int k)
    {
        return Factorial(n) / Factorial(k) / Factorial(n - k);
    }

    static void Main(string[] args)
    {
        for (var i = 0; i < 10; i++)
        {
            for (var j = 0; j <= i; j++)
                Console.Write(
                    Binomial(i, j) + "_");

            Console.WriteLine();
        }
    }
}
```

Рис. 1. Код программы 1

```

class Solution
{
    static void Main(string[] args)
    {
        for (int i = 0; i < 10; ++i)
        {
            for (int j = 0; j <= i; ++j)
            {
                Console.Write(C(i, j));
                Console.Write("_");
            }
            Console.WriteLine();
        }
    }

    private static int C(int n, int k)
    {
        return F(n)/F(k)/F(n - k);
    }

    private static int F(int n)
    {
        var res = 1;
        for (var i = 2; i <= n; ++i) {
            res = res*i;
        }
        return res;
    }
}

```

Рис. 2. Код программы 2

Чем отличается вторая программа от первой:

- Используется спецификатор класса *public* в первой программе, во второй - нет.
- Классы именованы по-разному. (*Program* против *Solution*).
- Методы классов именованы по-разному. Функция факториала в первом примере - *Factorial*, во втором - *F*. Тоже самое у *Binomial* и *C*.
- Во втором классе методы определены как приватные.
- Порядок методов по второй программе отличен от порядка в первой.
- Первая программа использует неявно типизированные локальные переменные (*var*), в то время как вторая - нет.
- Первая программа использует постинкремент для итерации цикла, вторая - преинкремент.
- Различие в выводе в консоль. Вторая программа разбила вывод на 2 вызова *Console.Write*
- Переименованы локальные переменные

- В первой программе используется  $result *= i$ , во второй -  $res = res * i$
- Различие в форматировании и в использовании фигурных скобок.

Для этих и других различий не должна быть чувствительна система проверки на плагиат.

Таким образом можно сформулировать основные требования к алгоритму сравнения:

- алгоритм должен распознавать хотя бы основные ключевые слова языка;
- переменные не должны считаться разными, если они отличаются только именем;
- при обмене местами двух функций код не должен распознаваться, как совершенно другой;
- аналогичный принцип применим для строк внутри функций: существуют строки, перестановка которых не приводит к изменению функциональности программы;
- при анализе все комментарии в коде должны игнорироваться, аналогично игнорируются пробельные символы и переводы строк;
- алгоритм должен быть расширяемый на другие языки программирования;
- иметь возможность сравнивать коды программ, написанных на разных языках.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] П.А. Хаустов, Ю.Я. Кацман *Алгоритм нечеткого сравнения исходных кодов программ.* – <http://www.eco-oos.ru/biblio/konferencii/prioritetnye-napravleniia-razvitiia-nauki-i-tehnologii-ix/33/>