

Интернациональность — информационные технологии способствуют глобализации образования и создают возможности для сотрудничества между учебными заведениями разных стран.

Всё это открывает новые возможности для развития образования и повышения его качества в цифровой эпохе.

Список литературы:

1. Информационные и коммуникационные технологии в образовании: учебно-методическое пособие; под ред. И.В.Роберт. — М.: Дрофа, 2008. — 312 с.
2. Полат Е.С. Современные педагогические и информационные технологии в системе образования: учебное пособие для студентов вузов. — 2-е изд., стер. / Е.С. Полат, М.Ю.Бухаркина. — М.: Академия, 2008. — 368 с.

УДК 004.415.2

СИСТЕМА ОБНАРУЖЕНИЯ АКАДЕМИЧЕСКОГО ПЛАГИАТА В ПРОГРАММНЫХ ПРОЕКТАХ

Е. С. Палто, В. С. Разумейчик

Брестский государственный технический университет, г. Брест

Плагиат — умышленное присвоение чужого авторства, выдача произведения науки или искусства, технического решения или изобретения за свое. Плагиат выражается в публикации под своим именем чужого произведения, а также компиляции результатов чужих исследований без указания источников заимствования. Академический плагиат — фальсификация учебных, исследовательских и квалификационных работ. Его совершают не для коммерческой эксплуатации чужого произведения, а для повышения статуса в академическом сообществе, нарушая таким образом этические стандарты академической честности.

Система обнаружения академического плагиата является вспомогательным инструментом, позволяющим сделать экспертную оценку более объективной.

В учебном процессе применение систем обнаружения плагиата имеет свою специфику. В дисциплинах, предполагающих закрепление полученных знаний в ходе выполнения типовых лабораторных работ и/или курсовых проектов, связанных с написанием программ, возникает проблема выявления академической нечестности, связанной с полным или частичным копированием работ других студентов. Проблема актуальна и для дипломных проектов, а также магистерских работ, хоть они и не являются типовыми: существующие системы антиплагиата хорошо применимы лишь к тексту пояснительной записки, в то время как оригинальность разработанного программного кода имеет не меньшую значимость.

К типовым процедурам обнаружения заимствований относятся следующие [1]:

– канонизация (нормализация) текста – исключение некоторых символов и их последовательностей (знаки препинания, цифры, смайлы, скобки, кавычки, пробелы, управляющие символы и т. д.), а также слов, не несущих смысловой нагрузки (предлоги, союзы, междометия, стоп-слова);

– стемминг и лемматизация – приведение словоформ и производных слов к общей форме, что дает возможность распознавать замаскированный плагиат;

– сравнение текстов с использованием различных методов: метод шинглов и его модификации, методы сигнатур, индексирования, контрольных сумм, опорных слов и др.

Приведем основные способы, используемые студентами, чтобы замаскировать (скрыть) в программных проектах факты заимствований фрагментов программного кода:

– написание или редактирование комментариев к фрагментам программы – самый простой и быстрый способ видоизменения программы, не требующий знания языка программирования;

– изменение идентификаторов переменных, функций и т.д. – чуть более сложный для студента путь, требующий минимальных знаний синтаксиса языка программирования;

– изменение порядка размещения фрагментов кода, удаление фрагментов кода – способ, требующий уже специальных навыков программирования;

– смешанный способ – использование двух и более перечисленных выше.

Помимо разработки эффективных алгоритмов обнаружения заимствований фрагментов программных кодов задача заключается также в необходимости создания базы данных программных кодов и организации быстрого поиска в ней.

В основе разработанной программы обнаружения плагиата лежит вычисление расстояния Левенштейна (редакционное расстояние, дистанция редактирования). Расстояние Левенштейна – широко используемая в теории информации и компьютерной лингвистике метрика, измеряющая по модулю разность между двумя последовательностями символов. Программа находит в базе данных программных кодов наиболее похожие строки, определяя наименьшее расстояние Левенштейна для данного кода.

Анализируемые программные коды могут быть представлены в виде исходного текста программы. Включение в алгоритм этапа предварительной обработки – удаление из программного кода всех комментариев – позволяет сделать систему нечувствительной к ним.

Для сокращения временных затрат на анализ строк целесообразно представление исходного текста программы в виде AST-строки.

Представление программного кода в виде AST (абстрактное синтаксическое дерево) позволяет работать с кодом как с деревом (узлами которого являются операторы и выражения, а листья – константы и идентификаторы), автоматически очищая его от комментариев и пустых строк. Таким образом, программный код преобразуется в структуру, которая во много раз меньше по объёму, чем сам

исходный код, что значительно ускоряет работу системы. Кроме того, это даёт возможность проводить глубокий сравнительный анализ.

Для ускорения работы системы был реализован подход обработки AST-строки, когда в ней происходит замена всех идентификаторов на "стандартные" имена переменных (например, a, b, c) и функций (например, f_1, f_2).

Для организации быстрого поиска в базе данных программных кодов разработана ее иерархическая структура, представленная на рисунке 1.

База данных должна проектироваться с учетом возможных направлений ее применения, включая проверку лабораторных, практических и курсовых работ, дипломных проектов и магистерских диссертаций. В качестве примера приведена структура базы данных программных кодов лабораторных работ, разработанная для задач БрГТУ. Одним из ключевых моментов является отсутствие строгого разделения на специальности. Это решение обусловлено тем, что в различных образовательных программах существуют общие дисциплины. При этом различные специализации могут иметь разное время начала и завершения изучения одной и той же дисциплины.

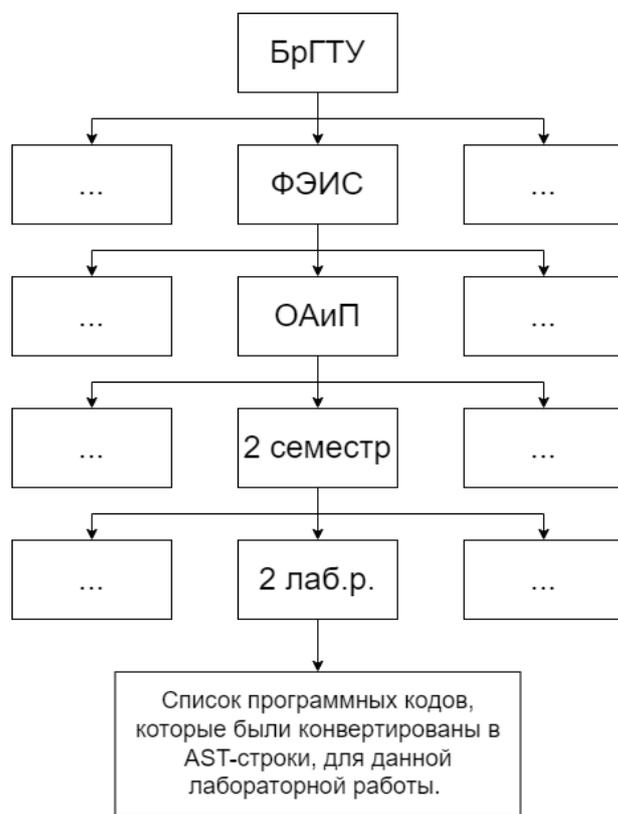


Рисунок 1 – Структура базы данных системы обнаружения академического плагиата в программных проектах

При необходимости, отдельные базы данных могут быть с лёгкостью объединены в одну.

Важно отметить, что линейный список AST-строк для каждой лабораторной работы по мере наполнения базы данных возрастает, алгоритмическая сложность поиска в нем составляет $O(n)$. Но за счёт древовидной структуры представленной выше базы данных снижается скорость нахождения нужного списка с AST-строками, что оптимизирует работу данной системы.

Разработанная система обнаружения плагиата учитывает различные способы видоизменения программного кода, написанного на языке Python. Система протестирована на работоспособность, наилучшую точность обнаружения плагиата показал алгоритм сравнения программных кодов, предварительно преобразованных в AST-строки.

Список литературы:

1. Пунчик, В.Н. Поликонтекстный анализ феномена «плагиат» в информационном обществе / В.Н. Пунчик, З.В. Пунчик // Социология – 2016. – № 1. – С. 83-91.

УДК 004.43+81.33+026.06

БИБЛИОТЕКА ЭЛЕКТРОННЫХ ДОКУМЕНТОВ С ВОЗМОЖНОСТЬЮ ИСПОЛЬЗОВАНИЯ В РЕЖИМЕ БАЗЫ ЗНАНИЙ С САМООБСЛУЖИВАНИЕМ

Я. А. Примак

*Гродненский государственный университет имени Янки Купалы, Гродно
Научный руководитель: А. М. Кадан, кандидат технических наук, доцент*

Базы знаний с самообслуживанием в настоящее время набирают популярность. Под этим термином подразумевают централизованный и структурированный сборник информации определенного направления - о продукте, услуге, отделе или теме. Иногда такую базу называют базой знаний по обслуживанию клиентов или базой знаний службы поддержки клиентов. Часто она имеет возможности поиска и содержит практические советы и инструкции, которые помогают клиентам — внутренним или внешним — решать проблемы, не обращаясь в службу поддержки.

Следуя указанной тенденции, в рамках междисциплинарного проекта в Гродненском государственном университете имени Янки Купалы было решено создать библиотеку электронных документов с возможностью использования в режиме базы знаний с самообслуживанием.

В работе использованы принципы построения баз знаний с самообслуживанием, а также реализован прототип поискового алгоритма. Его особенность в том,