

УДК 004.012.82

И.В. ГРУЗДО

Национальный аэрокосмический университет им. Н.Е. Жуковского "ХАИ", Украина

ПРОБЛЕМЫ АНАЛИЗА ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ТЕКСТОВ ДЛЯ ОБНАРУЖЕНИЯ ПЛАГИАТА В УЧЕБНЫХ РАБОТАХ

В работе проведено уточнение понятия «плагиат» по отношению к студенческим работам. Дана классификация студенческих работ. При этом в качестве классификационных признаков использованы объем текстового документа и состав его структурных частей. Определены критерии эффективности систем поиска заимствований в текстах, этих работ. Приведен обзор и критический анализ существующих программно-инструментальных средств анализа естественно-языковых текстов (на русском и украинском языках) в аспекте их пригодности к выявлению плагиата в студенческих работах. Анализ определяет необходимость создания моделей и методов, для программно-инструментальных средств анализа естественно-языковых текстов и как следствие повышение образовательного процесса в Украине.

Ключевые слова: *студенческий плагиат, виды плагиата, признаки плагиата, предотвращение плагиата.*

Введение

В настоящее время одной из наиболее острых проблем организации образовательного процесса в Украине является проблема нарушения академической добросовестности студентов. При выполнении индивидуальных заданий и итоговых работ, таких как работа бакалавра, дипломный проект (работа) и магистерская работа, значительное число учащихся руководствуется принципом быстрого и простого заимствования текстов из многочисленных источников, размещенных в среде Интернет - чужих дипломных проектов (работ), эссе, курсовых и т.п. [1,2]. Если раньше для написания перечисленных выше документов необходимо было потратить много времени для работы в научной библиотеке, то сейчас готовые работы покупаются или берутся у старших курсов, или свободно загружаются учащимися с Web-сайтов и банков рефератов, которые являются самыми популярными источниками заимствований.

Результаты статистических исследований, приведенных М. Железновой [3], свидетельствуют, что количество академических работ, содержащихся в соответствующих Интернет-коллекциях, постоянно растет, и каждые новые зачетная и экзаменационная сессии побивают предыдущий рекорд по частоте посещений таких сайтов. Таким образом, явление некорректных заимствований в образовательном процессе в настоящее время принимает массовый характер.

Легкость составления академических работ без необходимости проведения самостоятельных иссле-

дований препятствует достижению главной цели учебного процесса, которая заключается в обеспечении высокого уровня остаточных знаний учащихся и навыков использования этих знаний в их практической деятельности.

Недобросовестное отношение к самостоятельному выполнению практических заданий приводит к отсутствию навыков анализа, обобщения и изложения информации в письменном виде, что, во-первых, препятствует подготовке в ВУЗах высококвалифицированных специалистов, а, во-вторых, затрудняет процесс перехода к новой модели развития общества на основе знаний и высокоэффективных технологий, кроме того подобные явления снижают престиж как ВУЗа в целом, так и отдельных специальностей.

В связи с большим количеством систем поиска текстологических зависимостей, предлагаемых рынком в настоящее время, все большую актуальность приобретают вопросы, связанные с оптимальным выбором конкретных программно-инструментальных средств в соответствии с характером и типом исследуемых документов. Большинство систем оказывают значительную помощь в учебном процессе при его контроле и автоматизируют рутинные процессы установления вручную фактов плагиата. В настоящее время в литературе известно мало работ, посвященных вопросам анализа естественно-языковых текстов в аспекте их применимости для выявления заимствований в студенческих работах, что и определяет актуальность данной статьи

Цель данной статьи состоит в критическом анализе программно-инструментальных средств

обработке естественно-языковых текстов в аспекте их применимости для выявления заимствований в студенческих работах.

Постановка задачи исследования

В настоящее время имеется множество программно-инструментальных средств (платных, условно бесплатных, свободно распространяемых), пригодных для выявления текстологических совпадений в различного рода документах. Кроме того, известна номенклатура студенческих работ, поэтому необходимо, сформулировав ряд характерных критериев, установить пригодность каждой из известных систем поиска текстологических совпадений для анализа различных типов студенческих работ с целью установления плагиата.

Решение указанной задачи предполагает реализацию следующей последовательности шагов:

- уточнение понятия «плагиат» и на основе уточненного понятия формулировка качественных критериев оценки плагиата;
- классификация студенческих работ в соответствии с признаками, важными для выявления плагиата;
- анализ возможностей различных программно-инструментальных средств анализа естественно-языковых текстов для выявления плагиата в различных типах студенческих работ.

Виды, формы, признаки студенческого плагиата

В словарь иностранных слов [4] плагиат трактуется как «выдача чужого произведения за свое, а так же использование в своих трудах чужого произведения без ссылки на автора».

В большом юридическом словаре [5] плагиат – вид нарушения авторских прав, состоит в незаконном использовании под своим именем чужого произведения или изобретения, рационализаторского предложения без указания источника заимствования. Принуждение к соавторству также рассматривается как плагиат.

Плагиат (лат. *plagium* – похищение) [6]– 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.

М. Бреннер предлагает классифицировать плагиат на случайный и хорошо продуманный [7]. Если в первом случае студент не ссылается на источник, пытаясь избежать так называемой цепи цитирования, то к хорошо продуманному плагиату прибегает

умышленно, скрывая его всеми возможными способами.

Наряду с этим общим понятием появилось понятие студенческого плагиата, под которым понимается нарушение академических норм и правил, связанных с использованием чужой интеллектуальной собственности в качестве своей, а также отказ от самостоятельной, творческой деятельности в учебном процессе, что дискредитирует сами основы подготовки специалиста в ВУЗе [8].

Компания Turnitin, которая занимается разработкой систем поиска заимствований, случаи плагиата разделяет на две группы (цит. по: [9]). В первую группу попадают заимствования без указания источника. К ним относится: «призрачный автор», который выдает полностью выполненную кем-то другим работу за свою. Такой вид «творчества» распространен в двух формах.

Первая представляет собой использование материалов из Интернета или услуг «консультационных фирм», готовых написать работу на заказ по любой теме. Данная группа представлена следующими модификациями:

- 1) «фотокопия» – копируется значительная часть текста из одного источника без изменений;
- 2) «подвернувшийся материал» – делается попытка скрыть плагиат путем копирования нескольких разных источников, текст которых не меняется, но плагиатор вставляет свои переходные фразы между частями текста;
- 3) «плохая маскировка» – сохраняется смысл текста источника, но некоторые формулировки меняются на близкие по смыслу;
- 4) «труд лени» – информация из различных источников практически полностью перефразируется, пишется в одном стиле;
- 5) «украл у себя» – заимствуется текст из своих ранее написанных работ.

Ко второй группе относятся заимствования с указанием источников, тем не менее являющиеся плагиатом. Среди них выделяют:

- 1) «забытую ссылку» или «дезинформатор», связанные с неправильным или ошибочным оформлением ссылок на источник;
- 2) «слишком идеальное перефразирование», имеющее место, когда цитата не взята в кавычки и у читателя создается неверное впечатление о том, что автор привел свою оригинальную интерпретацию взглядов, изложенных в источнике;
- 3) «идеальное преступление» совершается, когда правильно приводятся некоторые цитаты, а остальные перефразируют, что создает у читателя иллюзию авторского анализа цитируемого;
- 4) «обильное цитирование» происходит с соблюдением всех правил цитирования и перефрази-

рования, но работа практически не содержит оригинальных результатов авторского исследования. Для обозначения последнего случая используется более корректный, чем плагиат, термин - «компиляция».

Формой студенческого плагиата является и списывание на экзаменах со шпаргалок. В американских ВУЗах шпаргалками регулярно пользуются до 98% студентов, которые не видят в этом ничего предосудительного [10].

Так же очень интересна типология студентов, прибегающих в учебной работе к плагиату, разработанная Дж. Уильямсом [11]:

- «ленивые плагиаторы» - как правило, неуспевающие и нецелеустремленные студенты, которые полностью копируют чужие работы, меняя в лучшем случае лишь их названия, используют сайты плагиата;

- «хитрые плагиаторы» прекрасно знают, что такое плагиат и прилагают значительные усилия для того, чтобы он не был обнаружен. Скрыть источники информации стараются путем неполного списка используемой литературы и ссылок;

- «случайные плагиаторы» - студенты с недостаточным уровнем учебных навыков, без опыта самостоятельного написания работ и не знающие академических норм.

Под плагиатом будем понимать незаконное использование под своим именем чужого произведения, рационализаторского предложения без указания источника заимствования, без учета фактов о соответствии законодательству и об умышленности деяния, и как следствие отказ от самостоятельной, творческой деятельности в учебном процессе.

Виды студенческих работ

В современной системе высшего профессионального образования подготовка студентами письменных работ традиционно рассматривается с двух точек зрения, которые не только не противоречат друг другу, но и находятся в диалектическом единстве.

С одной стороны, любая письменная работа – это форма учебно-методического воздействия на обучающегося, направленная на формирование у него практических навыков самостоятельной работы, способности логически оперировать ранее приобретенными профессиональными знаниями, умения творчески применять их на практике. С другой стороны, подготовка и написание студентом той или иной работы, предусмотренной учебным планом ВУЗа, является формой контроля его знаний по определенной дисциплине или ее части, а также группе учебных дисциплин. В этом случае оценивается широта знаний обучающегося, глубина его профессиональных умений и навыков, но, в первую оче-

редь, – умение творчески осмыслить, проанализировать теоретическую или практическую проблему и логически изложить ее, а также способность сделать самостоятельные выводы из обобщенного учебного или научного материала, документальных и иных материальных источников.

Общего и единого стандарта, регламентирующего оформление таких работ нет; наиболее близкими нормативными документами являются: стандарт Украины по оформлению отчета в сфере науки и техники [12], СТП, принятые внутри ВУЗа [13, 14], а так же методические указания по выполнению выпускных работ.

Охарактеризуем более подробно основные виды письменных работ по естественнонаучным образовательным дисциплинам, регламентируемые ДСТУ 3008-95, СТП ХАИ 4.01-95, ГОСТ 7.32-91 и СТП ХАИ 5.01-2007 - это расчетно-графические задания, рефераты, отчеты по лабораторным работам, домашним заданиям, курсовые работы и проекты, работа бакалавра, дипломный проект (работа) и магистерская работа, (рис. 1).

Для более полной картины и для того, чтобы увидеть разницу как в самой структуре работ так и в их объеме выделим обязательные структурные части видов письменных работ и отметим в табл. 1 знаком "+", а рекомендуемые - знаком "р" (табл. 1).

Так как спектр научных работ слишком разнообразен, необходимо четко определиться с какими именно видами текстовых учебных документов, будем работать. Так как наиболее трудно отследить плагиат в работах, которые имеют разветвленную структуру документа и объем более 50 страниц (работа бакалавра, дипломный проект (работа) и магистерская работа), то именно в них и следует искать заимствования.

Современное состояние проблемы применения программно-инструментальных средств анализа естественно-языковых текстов для выявления студенческого плагиата

Существует широкий спектр поисковых систем для определения текстологических совпадений в различного рода документах. С учетом специфики таких систем любую из них можно отнести к одному из следующих классов.

1. Утилиты статистического анализа текста.

2. Утилиты лингвистического анализа текста (морфология, синтаксис).

2.1. Системы обработки естественного языка.

2.2. Системы обработки машинного языка.

Исходя из особенностей объекта нашего анализа, а именно студенческих работ, в дальнейшем бу-

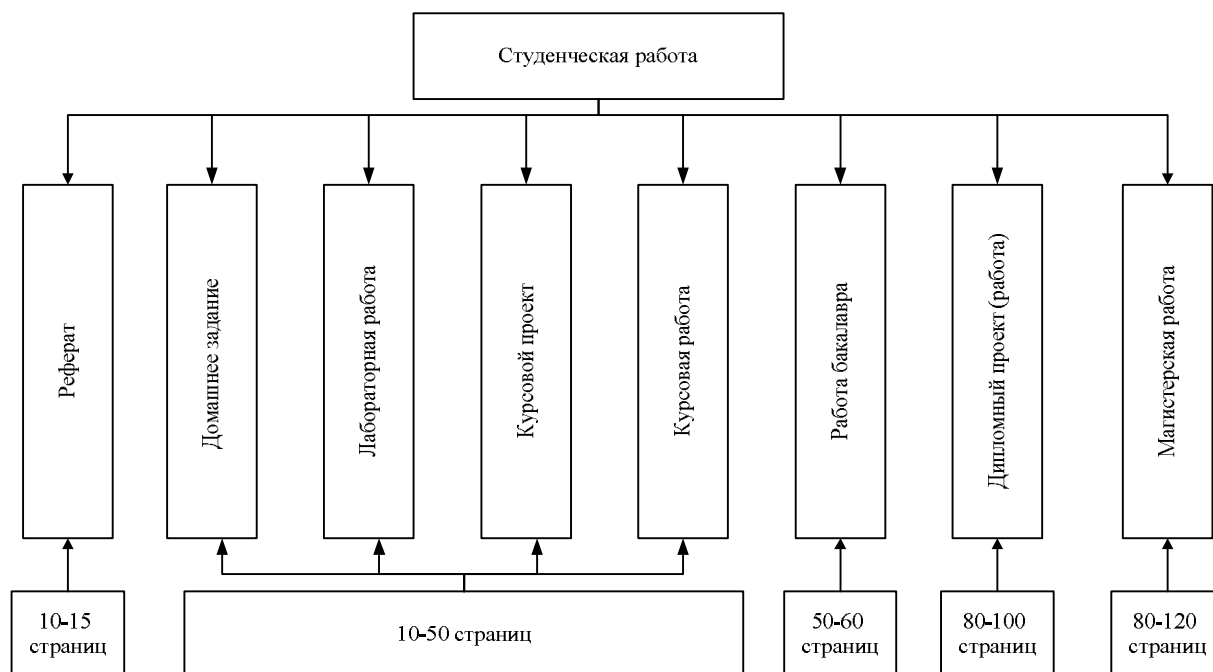


Рис. 1. Классификация видов студенческих работ по объему

Таблица 1

Виды текстовых учебных документов

Структурные части	Работа бакалавра, дипломный проект и магистерская работа	Курсовой проект, курсовая работа	Реферат на заданную (выбранную) тему	Домашние задание	Отчет по лабораторной работе
Титульный лист	+	+	+	+	+
Задание	+	+	р	+	+
Реферат	+	р			
Содержание	+	+	+	р	
Перечень условных обозначений, терминов и сокращений	р	р	р	р	
Введение	+	+	+	р	
Основная часть	+	+	+	+	+
Заключение, выводы	+	+	+	+	+
Список использованных источников	+	+	+	р	
Приложения	р	р	р	р	р

дем рассматривать системы обработки естественно-го языка. Типичным программным средством этого типа является TurnItIn.

TurnItIn - это коммерческая разработка американской компании iParadigms [15]. Система, проверяет документы на наличие в них некорректных заимствований из разнообразных текстов на естественном языке.

Поисковая база данных TurnItIn включает около 8 млрд веб-страниц и более 4,5 млрд рефератов и курсовых, десятки тысяч научных работ и аналитических статей из открытых и закрытых источников, а так же все студенческие работы, когда-либо про-

ходившие проверку. Время обработки запроса на проверку одной работы может составлять несколько дней. По окончании проверки пользователю предоставляется отчет об оригинальности работы с указанием в процентах доли текстологических заимствований.

В настоящее время правомерность использования службы TurnItIn подвергается сомнению [16]. Система включает все проверяемые академические работы в свою базу данных, получая экономическую выгоду без выплаты компенсации авторам. Служба, предназначенная для выявления случаев присвоения авторства, сама становится инструментом наруше-

ния прав интеллектуальной собственности.

Поскольку система TurnItIn оперирует исключительно с электронными текстами, набранными латиницей, проблематична ее адаптация к анализу работ украинском и русском языках.

"Детектор плагиата" [17] – компьютерная система поиска нарушения авторских прав, ориентированных на поиск плагиата в Интернете. После проведения сравнения пользователю предоставляется возможность просмотреть подробную информацию о найденных фрагментах совпадения. Данная система ориентирована на работу с русскоязычными текстами, однако главным ее недостатком является не эргономичный интерфейс.

Компьютерная система Плагиат-Информ [18], по сравнению с другими компьютерными системами рассматриваемого класса обладает тем несомненным преимуществом, перед другими средствами обнаружения плагиата, поскольку специально разработанная для определения наличия факта плагиата в студенческих работах, кроме того информационная база объединяет информационное пространство множества ВУЗов и, соответственно, обеспечивает независимость от территориального фактора, обеспечивая при этом быстрый и удобный доступ к работам, одновременно защищаемым в пределах этого пространства. Одним из недостатков PlagiatInform является не возможность работы с украиноязычными текстами.

Система «Антиплагиат» [19] ориентирована, главным образом, на поиск плагиата в студенческих работах, для наиболее эффективного решения этой задачи разработанная специальная версия системы «Антиплагиат.ВУЗ». Безусловным преимуществом данной системы является возможность предоставления двух режимов - быстрого и подробного. Анализ документа в системе «Антиплагиат» завершается выдачей подробного отчета, в котором имеется возможность просмотреть подробную информацию о найденных фрагментах совпадения [20]. Поддерживаемый формат файлов ASCII.

Недостаточная эффективность рассматриваемой системы проявляется во первых в сравнительно низкой скорости анализа текстов (судя по информации, представленной на официальном сайте [19], скорость проверки, обеспечиваемая программой, составляет 800 страниц (формата А4) в секунду (при неуказанной степени заполнения базы данных)), и во вторых, в существенном риске не распознавания плагиата, если анализируемый текст представляет собой полную копию, одно из документов в информационной базе.

Advego Plagiatus [21] – компьютерная система в которой реализована два метода анализа текста – простой и глубокий. Первый метод используется

программой по умолчанию и работает несколько быстрее. Второй способ более точный, но требует больше времени на поиск похожих фраз и словосочетаний в интернете.

Для рассматриваемой системы характерна высокая скорость обработки текстов в зависимости от применяемого метода (от одной до нескольких минут). Advego Plagiatus в большинстве случаев точно идентифицирует плагиат [22]. Недостатком системы является то, что результат ее работы исчисляется в процентах уникальности контента. Этот показатель лишь косвенно указывает на текстологические заимствования поскольку не несет в себе информации об источнике плагиата.

Лингвоанализатор [23] - компьютерная система, специально разработанная для определения наличия плагиата в тексте, принадлежащему определенному автору. Выбранное произведение сравнивается по своим стилистическим характеристикам со всеми авторами в каждой подвыборке, после чего выдается аргументированное заключение о близости произведения тому или иному автору, а также к произведениям выбранных авторов. После проведения сравнения пользователю предоставляется возможность просмотреть подробную информацию о найденных фрагментах совпадения.

Недостаточная эффективность рассматриваемой системы проявляется во первых в ее сравнительно низкой надежности (сами разработчики указывают [24], что поиск плагиата посредством анализа стиля документа может дать отрицательный результат при безусловном наличии плагиата), во вторых, в не возможности работы с текстами, объем которых менее 500 листов, в третьих, невозможностью работы с украиноязычными текстами.

Программа EVE2 является коммерческим приложением [25] и представляет собой интерфейс к поисковой машине Google. В соответствии с обзором [26] в EVE2 реализована возможность обработки документов разных форматов, в которых хранится значительная часть материала в Интернете: HTML, ASCII. К недостаткам EVE2 относятся, то что поиск выполняется только по Web-контенту в формате HTML и ASCII, а большая часть материала Всемирной паутине хранится в других форматах [27]. Поскольку EVE2 оперирует исключительно с электронными текстами, набранными латиницей, проблематична ее адаптация к анализу работ на украинском и русском языках.

Plagiarism-Finder - приложение, позволяющее проверить текст на предмет текстового совпадения с документами, хранящимися в Интернете [28]. В этой системе процедура поиска достаточно высока и не превышает нескольких минут. После проверки документа, пользователь получает подробный

HTML отчет, в котором представляются результаты, и выделены в тексте подозрительные абзацы так же есть ссылки на источники заимствований.

Одним из недостатков Plagiarism-Finder является возможность работы только с точными предложениями, состоящими из семи и более слов, но при этом он не работает, с меньшими фрагментами текста. Кроме того, существенным недостатком рассматриваемой системы является необходимость подключения к Интернету.

Приложение CopyCatch Gold разработано программистом британской фирмы CFL Software Developments Д. Вулсом [29], оно поддерживает многие популярные форматы, однако, обработка русскоязычных документов в формате MS Word ведется не корректно. Об этом свидетельствуют результаты испытаний демоверсии продукта, доступной для загрузки на официальной Web-странице [30].

Специальная интернет-служба Tumitin – лидер на рынке по обнаружению плагиата [27]. В качестве источников Tumitin использует Интернет-ресурсы. В случае обнаружения текстовых совпадений система выдает предупреждение. При этом программа не делает выводов, является ли проверенная работа плагиатом, а лишь выделяет цветом в анализируемом тексте те места, которые являются повторениями из других источников. При этом в системе ведется учет процента таких повторений по отношению к общему объему анализируемой работы.

Принципиальными недостатками Tumitin является то, что она, будучи способна анализировать документы, составленные на любом из большинства европейских языков, во первых допускает при этом сильные погрешности и, во вторых, не адаптирована для русского и украинского языков.

Автономное приложение (desktop application) WordCHECK [31] ориентировано, главным образом, на поиск плагиата в студенческих работах. Все проверяемые документы загружаются в базу данных, где сравниваются для обнаружения копирования в пределах учебной группы. Анализ документа в WordCHECK завершается выдачей подробного отчета, в котором имеется возможность просмотреть подробную информацию о найденных фрагментах совпадения. Данная система не адаптирована для русского и украинского языков, однако главным ее недостатком является неэргономичный интерфейс.

WCopyFind продукт, разработанный профессором Л. Блумфилдом в университете штата Виргиния (США), является одним из самых ранних инструментов для обнаружения фактов списывания студентами [32], обладает дружественным графическим интерфейсом и простой функциональностью. Анализ документа завершается выдачей подробного

отчета, в формате HTML с подробной информацией о найденных фрагментах совпадения и интернет-ссылками на них. К серьезным недостаткам следует отнести отсутствие возможности поиска совпадений в Интернете и не способность работы с украиноязычными текстами.

Plagiarism Advisory Service является национальной компьютерной системой по определению плагиата в Великобритании. Важная особенность детектора в том, что он дает ссылки на Web-сайты, где находятся оригиналы использованных работ [33]. Поскольку Plagiarism Advisory Service оперирует исключительно с электронными англоязычными текстами, проблематична ее адаптация к анализу работ на украинском и русском языках. Кроме того, существенным недостатком рассматриваемой системы является необходимость подключения к Интернету.

Анализ систем обработки естественного языка приведен в табл. 2. Табл. 2 отражает соответствие каждого программно-инструментального средства анализа естественно-языковых текстов используемому в нем методу и алгоритму. Во всех системах, для которых известны используемые в них методы анализа, эти методы не являются прямым воспроизведением одного из классических методов анализа текстов (Кука, строк Фибоначчи, Кнута - Морриса, Бойера-Мура, Бойера-Мура-Хорспула, Бейза-Ейтс и т.д. [34]), а представляют собой специально разработанные процедуры, составляющие элементы «ноу-хау» разработчика.

Выводы

На основе анализа различных подходов к определению плагиата, было сформулировано особое определение плагиата, отражающие суть текстовых заимствований в студенческих работах.

В работе была проведена классификация студенческих работ по признаку объема текстовых документов, а так же по признаку наличия их структурных частей.

Анализ функциональных возможностей имеющихся на рынке программного обеспечения средств обработки текстов показал, что среди этих средств нет ни одного, которое в полной мере бы удовлетворяло потребностям образовательного процесса в Украине.

Методическое обеспечение анализируемых программных продуктов (TurnItIn, Детектор плагиата, Плагиат-Информ, «Антиплагиат», Advego Plagiatus, Лингвоанализатор, EVE2, Plagiarism-Finder, CopyCatch Gold, Tumitin, WordCHECK, WCopyFind, Plagiarism Advisory Service и др.) не базируется на классических методах анализа текста,

Таблица 2

Системы обработки естественного языка

Детектор	Тип системы	Алгоритм, метод
TurnItIn	Интернет- служба	Собственные
Tumiting	Интернет- служба	Неизвестны
Plagiarism-Finder	Интернет- служба	Неизвестны
Детектор Плагиата	Приложение	Разработка интерфейса к поисковой машине Google™, и собственные
Плагиат-Информ	Приложение	Неизвестны
«Антиплагиат»	Интернет- служба	Собственные
Advego Plagiatus (Адвего Плагиатус)	Приложение требующие доступ к Интернет	Неизвестны
"Лингвоанализатор"	Интернет- служба	Неизвестны
EVE2	Приложение-интерфейс к поисковой машине Интернет	Собственные
Plagiarism Advisory Service	Автономное приложение	Собственные, цветовой код
Copy Catch	Автономное приложение	Разработка интерфейса к поисковой машине Google™
WordCHECK	Автономное приложение	Неизвестны
WCOPYFind	Автономное приложение	Неизвестны

а представляет собой специальные средства, с элементами «ноу-хау».

Указанные выше обстоятельства определяют необходимость создания моделей и методов, а на этой основе – алгоритмов как основы для синтеза эффективных программно-инструментальных средств анализа естественно-языковых текстов (на русском и украинском языках), наиболее полно удовлетворяющие процессам образовательного процесса в Украине.

Литература

1. Груздо И.В. Проблемы машинного анализа естественно-языковых текстов для обнаружения плагиата в научных работах, / И.В. Груздо // Всеукраїнська науково-технічна конференція "Інтегровані комп'ютерні технології в машинобудуванні ІКТМ – 2010", – Х.: НАУ "ХАІ", 22–24 ноября 2010 г. – Т.3. – 220 с.

2. Выполнение письменных работ Пакет «Антиплагиат.ВУЗ» Версия для учебных заведений система обнаружения заимствований в текстах [Электронный ресурс презентация к проекту Антиплагиат] / Презентация пакета "Антиплагиат.ВУЗ". Версия для учебных заведений система

обнаружения заимствований в текстах (ppt, 422 Кб). – Режим доступа к документу: <http://corp.antiplagiat.ru/index.aspx?doc=info>.

3. Железнова М. Поколение Copy-Paste [Электронный ресурс] / М. Железнова // Русский NEWSWEEK. – 2004. – №3. – Режим доступа к документу: http://studentochka.ru/editor_32.html. – Заголовок с экрана.

4. Большой толковый словарь иностранных слов. в 3 т. – Ростов-на-Дону: "Феникс", 1995. – Т. 2. – 452 с.

5. Юридический словарь / Юрид. центр науч. исслед. и правовой информации. – М.: Инфра-М, 2009. – 864 с.

6. Большая Советская Энциклопедия. – М.: Издательство, 1970. – Т. 23. – 766 с.

7. Brenner M. Plagiarism and secure examinations in online distance education University of Gävle. [Электронный ресурс] / M. Brenner. – Sweden, 2004. – 15 p. – Режим доступа к документу: <http://www.hig.se/learningcenter/erfbanklfiler/ukraina040908.pdf>. Заголовок со статьи.

8. Evans J. The new plagiarism in higher education: from selection to reflection [Электронный ресурс] / J. Evans // Learning and Development Centre. – 2000. – Режим доступа к документу: <http://www.warwick.ac.uk/ETS/interactions/vo14no2/evans.htm>.

9. Николаев Е. Что такое плагиат, или О западных стандартах научной этики [Электронный ресурс] / Е. Николаев // Освітній портал – 19.05.2006. – Режим доступа к документу: <http://www.osvita.org.ua/articles/68.html>. – Заголовок с экрана.
10. Altbach Ph.G. The question of corruption in Academe / Ph.G. Altbach // International higher education. – 2004. – №34. – P. 4–7.
11. Williams J.B. Plagiarism: deterrence, detection and prevention. [Электронный ресурс] / J.B. Williams; Edited by John Houston and David Whigham // The Handbook for Economics Lecturers, April 2005. – 20 p. – Режим доступа к документу: <http://www.economicsnetwork.ac.uk/handbook/printable/plagiarysm.pdf>. – Заголовок со статьи.
12. ДСТУ 3008–95. Документація. Отчеты в сфере науки и техники. Структура и правила оформления. – Введ. – 23.02.1995. – К. Госстандарт, 1995. – 37 с.
13. СТП ХАИ 4.01–95. Стандарт предприятия. Студенческая учебная работа. Текстовая часть (пояснительная записка). Титульный лист. – Взамен ГОСТ 7.32–91, СТП ХАИ 4.02–86; введ. 01.09.1995 – Харьковский авиационный институт, 1995. – 11 с.
14. Павленко В.Н. Порядок оформления учебных и научно-исследовательских документов: учебное пос. / В.Н. Павленко, А.С. Набатов, И.М. Тараненко. – Харьков: НАУ ХАИ. – 2007. – 65 с.
15. TurnItIn. Plagiarism prevention service TurnItIn [Электронный ресурс] / Материалы с сайта разработчика TurnItIn. – Режим доступа к документу: <http://www.turnitin.com>. – Заголовок с экрана.
16. Technology & Teaching. Turnitin.com, a Pedagogic Placebo for Plagiarism [Электронный ресурс] / Emailed 6.5.01; Archived on the Web 13.06.01. – Режим доступа к документу: <http://www.bedfordstmartins.com/technotes/techtiparchive/ttp060501.htm>. – Заголовок с экрана.
17. Программа «Детектор плагиата» [Электронный ресурс] / Материалы с сайта разработчика Детектор плагиата. – Режим доступа к документу: <http://www.detector-plagiata.ru/index.html>. – Заголовок с экрана.
18. PlagiatInform Проблемы плагиата в вузах – [Электронный ресурс] / Материалы с сайта разработчика PlagiatInform. – Режим доступа к документу: <http://www.searchinform.com/search-site/ru/main/full-text-search-case-studies-plagiatinform.html>. – Заголовок с экрана.
19. Интернет-сервис AntiPlagiat.ru [Электронный ресурс]. – Режим доступа к документу: <http://www.antiplagiat.ru>. – Заголовок с экрана.
20. Википедия. Антиплагиат [Электронный ресурс]. – Режим доступа к документу: <http://ru.wikipedia.org/wiki/Антиплагиат>. – Заголовок с экрана.
21. Advego Plagiatus [Электронный ресурс] / Материалы с сайта разработчика Advego Plagiatus. – Режим доступа к документу: http://advego.ru/blog/read/plagiatus_news/. – Заголовок с экрана.
22. Проверка контента на плагиат [Электронный ресурс] / Сервисы, программы, алгоритмы. – Режим доступа к документу: <http://blog.negotiant.org/proverka-kontenta-na-plagiat-servisy-programmy-algoritmy/>. – Заголовок с экрана.
23. ЛингвоАнализатор [Электронный ресурс] / Материалы с сайта разработчика ЛингвоАнализатор. – Режим доступа к документу: <http://www.rusf.ru/books/analysis/>. – Заголовок с экрана.
24. Хмельёв Д.В. О Лингвоанализаторе 3-эпсилон [Электронный ресурс] / Д.В. Хмельёв // Электронный журнал "Самиздат". – 2009. – №1. – Режим доступа к документу: http://samlib.ru/h/hmelew_d_w/descrwin.shtml. – Заголовок с экрана.
25. CaNexus. Plagiarism detection system EVE2 [Электронный ресурс] / CaNexus. – Режим доступа к документу: <http://www.canexus.com>. – Заголовок с экрана.
26. Gaither Renoir Plagiarism Detection Services, [Электронный ресурс] / Renoir Gaither // Shapiro Undergraduate Library, University of Michigan 2004. – №11. – Режим доступа к документу: <http://www.lib.umich.edu/acadintegrity/instructors/violations/detection.htm>
27. Neill J. Colin A Web-Enabled Plagiarism Detection Tool / Colin J. Neill, Ganesh Shanmuganthan // IEEE Computer Society IT Professional. – September 2004. – V 6, №5. – P. 19–23.
28. Plagiarism-Finder [Электронный ресурс] / Материалы с сайта разработчика Plagiarism-Finder. – Режим доступа к документу: <http://www.m4-software.com/en-index.htm>. – Заголовок с экрана.
29. CFL Software Limited. Collusion and plagiarism detection program [Электронный ресурс]. – Режим доступа к документу: <http://www.copycatchgold.com>. – Заголовок с экрана.
30. Turnitin. The global leader in addressing plagiarism and delivering rich feedback [Электронный ресурс] / Материалы с сайта разработчика Turnitin. – Режим доступа к документу: www.turnitin.com. – Заголовок с экрана.
31. WordCHECK keyword software [Электронный ресурс] / Материалы с сайта разработчика WordCHECK. – Режим доступа к документу: <http://www.wordchecksyste.ms.com>. – Заголовок с экрана.
32. Plagiarism detection software WCopyfind [Электронный ресурс] / Материалы с сайта разработчика WCopyfind. – Режим доступа к документу: <http://plagiarism.phys.virginia.edu/Wsoftware.html>. – Заголовок с экрана.
33. Савочкина Т.С. Студенческий плагиат в высшем профессиональном образовании / Т.С. Савочкина, И.В. Нератова // Вестник Тверского государственного университета. Серия: Педагогика и психология – 2009. – Вып. 6. – С. 90.– 104.

34. Алгоритмы: построение и анализ. Introduction to Algorithms / Т.Кормен, Ч. Лейзерсон, Р. Ривест, К. Штай; под ред. И. В. Красикова. – 2-е изд. – М.: Вильямс, 2005. – 1296 с.

Поступила в редакцию 21.02.2011

Рецензент: д-р техн. наук, проф., зав. каф. экономико-математического моделирования В.М. Вартамян, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.

ПРОБЛЕМИ АНАЛІЗУ ПРИРОДНО-ЯЗИКОВИХ ТЕКСТІВ ДЛЯ ВИЯВЛЕННЯ ПЛАГІАТУ В НАВЧАЛЬНИХ РОБОТАХ

I.V. Gruzdo

У роботі проведено уточнення поняття «плагіат» стосовно студентських робіт. Дано класифікацію студентських робіт. При цьому як класифікаційні ознаки використаний обсяг текстового документа й склад його структурних частин. Визначено критерії ефективності систем пошуку запозичень у текстах, цих робіт. Наведено огляд і критичний аналіз існуючих програмно-інструментальних коштів аналізу природно-язикових текстів (на російській і українській мовах) в аспекті їхньої придатності до виявлення плагіату в студентських роботах. Аналіз визначає необхідність створення моделей і методів, для програмно-інструментальних засобів аналізу природно-язикових текстів і як наслідок підвищення освітнього процесу в Україні.

Ключові слова: студентський плагіат, види плагіату, ознаки плагіату, запобігання плагіату.

PROBLEMS OF THE ANALYSIS OF NATURAL LANGUAGE TEXTS FOR PLAGIARISM DETECTION IN STUDIES

I.V. Gruzdo

In job specification of concept "plagiarism" in relation to student's jobs is spent. Classification of student's jobs is given. Thus as classification signs the volume of the text document and structure of its structural parts are used. Criteria of efficiency of systems of search of loans in texts, these jobs are defined. (In Russian and an Ukrainian language languages) in aspect of their suitability the review and the critical analysis of existing programming-tool means of the analysis of natural language texts is resulted in plagiarism revealing in student's jobs. The analysis defines necessity of creation of models and methods, for programming-tool means of the analysis of natural language texts and as consequence increase of educational process in Ukraine.

Key words: student's plagiarism, plagiarism kinds, plagiarism signs, plagiarism prevention.

Груздо Ирина Владимировна – аспірантка каф. інженерії програмного забезпечення, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: tigralwovna@rambler.ru.