

НЕинтеллектуальная система «АНТИПЛАГИАТ»



В последнее время на фоне модернизации ВАК, реформирования диссоветов, скандалов в отношении неправомерных заимствований в диссертациях всё чаще говорят об использовании системы «Антиплагиат». Основная тема дискуссий – может ли система подобного типа с высокой долей точности определить, является ли научная или квалификационная работа самостоятельной.

Ситуацию прояснили разработчики и ведущие пользователи системы.

Юрий ЧЕХОВИЧ: «Антиплагиат» — независимый коммерческий проект, направленный на создание программных средств поиска и оценки текстовых документов. Это полностью российская разработка, созданная при участии ведущих отечественных математиков. «Антиплагиат» — инструмент, который доступен каждому для проверки любых текстовых документов на наличие заимствований.

Работу над системой мы начали восемь лет назад, развиваем её и не собираемся останавливаться на достигнутом. Наша цель в том, чтобы к 2015 г. системой оценки самостоятельности учебных и научных работ был оснащён каждый российский вуз. «Антиплагиат» должна стать общим местом, как сайт вуза, как система управления вузом. Другой нашей целью является развитие системы, так что мы не просто будем «под-

свечивать» в тексте цитаты и заимствования, а позволим обнаруживать безусловного автора цитаты, находить документы сходной тематики, выявлять «деревя заимствований» и т.п.

Константин РУДАКОВ: «Антиплагиат» — это не интеллектуальная система, это поисковик, такой же, как и «Яндекс». Вызов для нашего интеллекта при разработке этой системы был такой: надо, чтобы за небольшое время текст приличного размера можно было полностью сверить с несколькими сотнями миллионов документов на более-менее пристойной технике. Здесь решаются проблемы объёма, быстрого поиска, качественной индексации. Текст рассматривается как последовательность символов, никаких тайных механизмов, никакой семантики в программе не заложено. Это некий стандарт де-факто для первичного поиска

совпадений символьных последовательностей.

Система сделана достаточно грамотно с инженерной точки зрения. Других систем подобного назначения пока нет, остальные — идеологически иные, их нужно использовать все вместе, никакой монополии быть не может. Это хорошая, быстрая машина, правильно читающая то, что ей положено. Не нужно требовать от неё интеллектуального анализа, и это, вообще говоря, серьёзный вопрос — надеяться ли «Антиплагиат» каким-то интеллектом. Сейчас система выдаёт одну-единственную цифру — процент цитирования. Эта функция по интеллекту ровно такая же, как статистика в Word. Плохо, когда подобные системы начинают заменять мозги преподавателей. Если вы купили навигатор, это не значит, что вы должны закрыть себе ветровое стекло и ездить по навигатору. Может быть абсолютно оригинальная работа, которая на 100% будет состоять из цитат. Сам по себе набор цитат может быть серьёзной научной работой. Да, как говорят многие учёные, в том числе В.А. Садовничий, «Антиплагиат» несовершенен, но это абсолютно вспомогательная вещь. Если с помощью системы преподаватель обнаружил, что существенная часть работы была опубликована, когда аспирант был ещё в третьем классе — вот это криминал.

Какой процент плагиата допустим? Конечно, ноль! В работе не может быть плагиата. Если в ней есть существенный фрагмент текста, в котором суть работы, и он откуда-то взят, эту работу нельзя принимать. Это воровство. Эту цифру — процент заимствований — никому нельзя показывать, тем более выставлять в качестве критерия. То, что пишут блогеры — абсолютно неквалифицированные высказывания. Почему в докторской человек цитирует свою кандидатскую? Диссертация по определению — компилятивная работа. А основные результаты работы по положению ВАК должны быть опубликованы. Широкая научная общественность должна посмотреть ваши статьи. А потом, что, их нужно переписывать другими словами? Когда человек берёт фрагменты своего собственного текста — это совершенно нормальная ситуация.

Александр ВИСЛЫЙ: Один из крупнейших текстовых массивов, по которым осуществляется проверка — это электронные тексты диссертаций, которые в РГБ представлены с 2000 г. по настоящее время. Этот

объём, конечно, не на 100% достаточен, чтобы обнаруживать совпадения. Кроме того, из-за проблем с текущим законодательством нет текстов учебников, монографий, статей, и если заимствования были из этих источников, то их обнаружить практически невозможно. По большей части все основные совпадения, которые находятся, — это совпадения одной диссертации с другой. Если кто-то хочет получить при анализе этих совпадений хоть какое-то значимое заключение, необходимо, чтобы после того как проработала система «Антиплагиат», эксперт проанализировал каждое отдельно взятое

«**Сейчас система выдаёт одну-единственную цифру – процент цитирования. Эта функция по интеллекту ровно такая же, как статистика в Word. Плохо, когда такие системы начинают заменять мозги преподавателей**»

совпадение. Единственным заключением, позволяющим принять решение, является примерно такое: вот восемь страниц одной и другой диссертаций, совпадающих от начала до конца. Однако нужно иметь в виду, что можно перевести российскую диссертацию, скажем, на молдавский язык, и никакой «Антиплагиат» и никакая другая система ничего не поймут.

Мы проверяем диссертации, такая услуга есть в РГБ, сейчас у нас работают семь экспертов. Что мы проверяем? Во-первых, когда к нам приходит человек со своим собственным текстом, если такую задачу перед ним поставил диссертационный совет, и нужно представить справку. Во-вторых, к нам обращаются диссертационные советы. В-третьих, бывают случаи, когда к нам приходят запросы из Минобрнауки. Но к нам нельзя придти и проверить чужую диссертацию.

В прошлом году мы проверили 173 работы, и в 11 были обнаружены существенные заимствования.

Что касается «Антиплагиата», изначально мы договаривались немного о другом. Очень хотелось получить механизм, который позволял бы по тексту статьи, написанной учёным, совершая проверку по большому массиву данных, выдавать похожие работы, чтобы мы понимали, кто ещё этой работой занимается.



Юрий ЧЕХОВИЧ
исполнительный директор
ЗАО «Анти-Плагиат»



Константин РУДАКОВ
член-корреспондент РАН,
заведующий кафедрой
«Интеллектуальные
системы» МИФИ, научный
руководитель проекта
«Антиплагиат»



Александр ВИСЛЫЙ
Генеральный директор РГБ



Сергей РОЩИН
проректор Высшей
школы экономики



Юрий ЖУРАВЛЁВ
академик РАН

Сергей РОЩИН: Наш университет достаточно давно осознал, что в современную информационную эпоху вопрос заимствований текстов — это серьёзная проблема. Существовая в современном информационном пространстве, работая с текстами на компьютере, современная молодёжь совершенно по-другому воспринимает проблемы авторства, интеллектуальной собственности, и тому масса примеров.

В середине 2000-х годов это стало «захлёстывать», и мы приняли положение о том, что плагиат будет караться. А тут уже подросла и система, и мы с 2008 г. начали работать в полутестовом режиме, а с 2009 г. перешли к тотальному использованию в университете системы «Антиплагиат». Все дипломные и курсовые работы в обязательном порядке загружаются в эту систему, без этого человек не может выйти на защиту. Также могут загружаться и другие письменные работы по желанию преподавателя.

Безусловно, я поддерживаю коллег в том, что никакая система не избавляет от содержательного, серьёзного анализа любого кейса, связанного с заимствованием. Система — это инструмент, который позволяет выявить некоторые технические совпадения, а дальше уже смотреть, являются ли они наказуемыми. У нас даже есть такое понятие, как «неоригинальный текст». Это нормативные, законодательные документы, наименования учреждений, таблицы данных Росстата и т.п. Такой текст должен просто быть грамотно оформлен, без него нельзя сделать множество работ. А вопрос неправомерного заимствования — уже предмет для разбирательств. Это должны профессионально делать те коллеги, которые связаны с этой областью деятельности.

Очень важна система ответственности. В нашем университете любое неправомерное заимствование ведёт к занулению оценки работы. Существуют дисциплинарные взыскания — выговор, а затем отчисление из университета. В год речь идёт о десятках таких случаев на университет, в особенности среди курсовых работ. По отношению к масштабам университета это не очень много, но и не мало, при этом динамика более-менее устойчива, я здесь вижу очень серьёзные проблемы с ценностями студентов, которые приходят к нам учиться.

У нас нет массового опыта проверки диссертаций, да это и не столь актуально. Как научный руководитель я хорошо представляю, что когда я вижу процесс написания работы ежемесячно, то её даже абсурдно

проверять. Это целесообразно по отношению к тем работам, которые появляются «вдруг» и «ниоткуда».

У нас большой раздел прикладных работ, и мы также стали использовать систему «Антиплагиат» с учётом большого массива накопленной в ней информации, для того чтобы ограничивать самоцитирование в различных исследовательских работах. Так как такие работы выполняются под заказ, получение одного и того же результата не всегда правомерно. Опять-таки, здесь не может быть никакого автоматического, механического суждения. Любой инструмент без профессионалов — это ничто. А вот когда он встроен в систему правил и регламентов, он становится действенным и эффективным.

Юрий ЖУРАВЛЁВ: Любое новое техническое средство, как правило, сначала встречается «в штыхы». Когда я был студентом, общественность очень ругала компьютеры, которые «применяются там, где не надо, и пытаются считать то, что принципиально не считается». Любая система такого типа, как бы качественно она ни была сделана, может быть очень хорошей в хороших руках и плохой — в плохих. Прежде, чем применять любую систему, нужно понимать, как она работает. Никому не придёт в голову подходить к современному компьютеру, не умея с ним работать.

«Антиплагиат» — принципиально новая система, которая в наш век, когда рейдерство встречается не только в промышленности и сельском хозяйстве, но и в науке и литературе, является инструментом, который мог бы с этим бороться, как минимум, контролировать. Не хочу сказать, что эта задача полностью решена, сделан первый шаг.

Мне приходится последние 15 лет работать в ВАК, и в работах по физике и математике случаи заимствования — редчайшие, но и они тоже бывают, мы их снимаем как правило, особенно в последние годы.

Итак, «Антиплагиат» — не панацея, но и не бесполезная трата бюджетных средств вуза. Любой инструмент эффективно работает, если его правильно применять. И, в любом случае, нельзя заменять ни одной, даже самой продвинутой, технической системой интеллектуальный потенциал учёного, научного руководителя, преподавателя — такие выводы можно сделать по результатам встречи. ■



АНТИПЛАГИАТ

СТАТИСТИКА ПРОЕКТА

по состоянию
на 3 июня 2013 г.:

К системе «Антиплагиат» подключено **185** российских и зарубежных вузов, в которых функционируют **1075** диссертационных советов.

Интернет-коллекция состоит из **44 074 490** документов с общим объёмом текста **454,60 Gb**.

За всю историю было выполнено **21 574 222** проверок.

Количество зарегистрированных пользователей — **652 036**.

Пользователями загружено на проверку **19 784 514** документов с общим объёмом текста — **552,3 Gb**.