Внедрение системы Антиплагиат в Российской Государственной Библиотеке

© Романов Михаил Юрьевич Житлухин Дмитрий Анатольевич

ЗАО «Форексис»

mromanov@forecsys.ru dzitlukhin@forecsys.ru

Аннотация

В докладе приведено описание внедрения системы Антиплагиат в Российской Государственной библиотеке.

В рамках этого внедрения в РГБ поставлена система поиска заимствований по базе оцифрованных авторефератов и диссертаций, а также реализована поддержка проекта «научный поиск».

Приведено подробное изложение результатов внедрения, состав системы, вопросы безопасности и результирующие технические и скоростные характеристики.

1 Цели и задачи системы

1.1 О системе Антиплагиат

Интернет-сервис www.antiplagiat.ru (см. [1]) был создан в 2005 г. (см. [5]) для проверки текстовых документов на наличие заимствований из общедоступных сетевых источников. Изначально система разрабатывалась для внедрения в крупный коммерческий ВУЗ МИЭМП. Функциональное ядро системы Антиплагиат использует алгоритмы, разработанные сотрудниками ВЦ РАН и компании Форексис (см. [6]).

Стратегической задачей проекта Антиплагиат является повышение качества российского образования и научной деятельности преимущественно в тех случаях, когда требуется творческая работа по написанию рефератов, курсовых и дипломных работ, диссертаций и иных материалов. Эта задача решается путем побуждения обучающихся к самостоятельному написанию текстов, а не к созданию их, например, путем компиляции текстов, найденных в интернете и других источниках, касающихся заданной тематики.

В сентябре 2005 года была запущена в эксплуатацию тестовая версия интернет-сервиса www.antiplagiat.ru. Впервые использование системы Антиплагиат было включено в обязательный учебный процесс вуза в ноябре 2005 года. В 2006 году на пятом Конкурсе русских инноваций разработка получила приз Минсвязи РФ «За лучший проект в области телекоммуникаций». Летом 2006 года Советом Ассоциации негосударственных вузов РФ принято решение рекомендовать членам Ассоциации применение сервиса. В июне 2007 года использование в ВУЗах интернет-сервиса www.antiplagiat.ru было рекомендовано Советом по качеству образования при Рособрнадзоре РФ. К июлю 2007 года была разработана адаптированная под нужды Высшей аттестационной комиссии (ВАК) Минобрнауки РФ система для обязательной проверки на плагиат всех диссертаций и авторефератов и начата ее эксплуатация.

Система Антиплагиат постоянно развивается и расширяет функциональность. Постоянно происходит совершенствование алгоритмов поиска для выдачи более корректных результатов и для учёта ситуаций и особенностей, не обрабатывавшихся ранее. В 2006 г. введена возможность развёртывания системы на оборудовании держателей документов. В 2007 г. реализована поддержка распределённой проверки по совокупности коллекций документов. В 2008 г. введено использование для ВУЗов механизма «сигнальных статистик», подсказывающих преподавателю, что в проверяемом документе есть подозрение на попытку «обхода» системы. Постоянно развивается и дорабатывает пользовательский интерфейс.

В настоящий момент систему Антиплагиат используют такие вузы, как Высшая школа экономики, Московский институт экономики, менеджмента и права, Московский городской психолого-педагогический университет, Московский государственный педагогический университет, Современная гуманитарная академия и другие. Создан специальный сайт именно для работы с ВУЗами (см. [2]). А публичный интернет-сервис используют около 100 тысяч пользователей в России и за рубежом.

Внедрение системы в Российскую Государственную Библиотеку началось в первой половине 2008 г. Во второй половине 2008 г. открыт сайт для поиска по базе электронных документов РГБ (см. [3]). В настоящий момент продолжается активное расширение функциональности системы и построение новых сервисов на её основе.

Труды 11[й] Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2009, Петрозаводск, Россия, 2009.

Система Антиплагиат разработана и поддерживается компанией Форексис.

1.2 Задачи системы в РГБ

Российская государственная библиотека (РГБ) является уникальным хранилищем диссертаций, защищенных в стране с 1944 года по всем специальностям, кроме медицины и фармации (см. [4]). Всероссийский фонд диссертационных работ был создан в 1944 году. Сейчас в фонде диссертаций хранятся свыше 900000 томов диссертаций, причём ежегодно поступает около 30000 диссертаций (20000 кандидатских и 10000 докторских). В 2003 году была начата работа по оцифровке этого хранилища документов.

Перед системой Антиплагиат стояла задача организации эффективного поиска заимствований по оцифрованной базе диссертаций, а также построения различных сервисов на основе поискового движка. К числу этих сервисов относится:

- предоставление ВУЗам и другим пользователям системы возможности проверки текстов по базе диссертаций Российской Государственной Библиотеки;
- реализация ядра системы «научный поиск»;
- предоставление сотрудникам РГБ аналитических средств, обеспечивающих работу с метаданными хранилища диссертаций.

2 Результаты внедрения

Первой и ключевой задачей системы было предоставление сервиса по проверке

2.1 Состав системы

Система Антиплагиат.РГБ состоит из следующих модулей.

- Хранилище документов (коллекция). Содержит набор документов, необходимые индексы и позволяет осуществлять поиск по ним.
- Модуль импорта из систем РГБ. Позволяет по расписанию либо вручную загружать документы из источников РГБ в коллекцию Антиплагиат. Поддерживает библиографический стандарт MARC. Обнаруживает и корректно обрабатывает изменения в уже загруженных документах (то есть позволяет работать в режиме обновления метаданных).
- Модуль доступа к внешним экземплярам системы Антиплагиат. Позволяет извне проверять документы по базе диссертаций, а также из системы Антиплагиат.РГБ проверять по внешним коллекциям.
- Пользовательский интерфейс в виде сайта antiplagiat.rsl.ru. Предоставляет функциональность, сходную с общедоступным сервисом www.antiplagiat.ru для пользователей читальных залов.
- Модуль предоставления функций системы Антиплагиат через WebService для внутреннего использования сотрудниками РГБ.
- Модуль для аналитической обработки метаданных документов. Предоставляет OLAP-куб с метаданными на Microsoft SQL Server 2005 для внутреннего использования сотрудниками РГБ. У них появляется возможность построения сводных отчетов по любым имеющимся в системе атрибутам с различными вариантами агрегации. Пример: вывести средний процент цитирования у всех загруженных в систему статей по теме «Психология» с недельной агрегацией по дате публикации.
- Модуль администрирования системы (веб-интерфейс). Позволяет управлять правами доступа и настройками системы (такими, как импорт).
- Модуль поддержки научного поиска. Предоставляет расширенную функциональность доступа к ядру системы для потребностей научного поиска.
- Модуль просмотра отчётов о проверке offline (Antiplagiat Report Viewer). Позволяет экспортировать результаты проверки в специальный файл (контейнер), включающий в себя исходный документ, отчёт о проверке, а также специальные ключи, обеспечивающие подлинность просматриваемого файла.

2.2 Качество и скорость проверки

Система Антиплагиат изначально разрабатывалась для проверки больших объёмов текста, загружаемых пользователями сети internet, в связи с чем она с самого начала оптимизировалась по времени работы - время проверки среднего документы ещё в 2005 году составляло несколько секунд. Со временем, благодаря открытой проверке на прочность системы в internet'e, открывались некоторые способы её "обхода" - способы автоматической или полуавтоматической переработки текста, после которых он не распознавался системой как списанный. Система шла в ногу со временем - при выявлении новых методов обхода они оперативно закрывались, что порой приводило к некоторому снижению производительности. В последних версиях был внедрён новый поисковый движок, уникальный индекс которого позволил решить проблемы производительности. Также теперь имеется возможность вручную управлять соотношением

«цена/качество» - можно сделать индекс маленьким, при этом проверка текстов будет выполняться быстро, но некоторые мелкие фрагменты могут быть не замечены, а можно большим, обеспечив высокое качество проверки, но потребуется больше дискового пространства и времени на проверку. Наконец, применение современных SSD-накопителей повысило скорость работы системы практически на порядок. Теперь время проверки составляет в среднем менее 1 секунды.

2.3 Интеграция различных коллекций документов

С развитием системы появилась возможность хранить отдельные независимые базы текстов - коллекции. Документ может проверяться как по одной, так и по нескольким коллекциям одновременно, что обеспечивает дополнительную гибкость при проверке. При подключении к системе различных организаций, для каждой из них заводится своя коллекция. После этого держатель коллекции может распоряжаться этой коллекцией по своему усмотрению - предоставлять доступ к ней другим пользователям, определять правила работы со своей коллекцией - можно ли будет им скачивать из неё документы или же использовать только для поиска цитирования без явной загрузки документов.

В Российской Государственной Библиотеке система работает на выделенном сервере, который находится в ведении ИТ-отдела РГБ. Таким образом, любой запрос на проверку по базе диссертаций автоматически направляется пользовательской системой на центральный шлюз Антиплагиат, который уже разбирает запрос и переадресует его на систему, расположенную в РГБ. Результаты проверки аналогичным образом через центральный шлюз доставляются до конечного пользователя.

В следующих трёх параграфах рассматриваются вопросы безопасности системы.

2.4 Сохранность текстовых баз

Система не реплицирует текстовые коллекции. При установке локальной версии у заказчика вся его текстовая база хранится на его сервере и никогда не реплицируется на другие машины в сети. Все проверки по данной коллекции обрабатывают сервера заказчика, что обеспечивает сохранность текстовой базы.

Систему можно настроить таким образом, чтобы она не предоставляла во вне целиком найденные источники заимствования. В этом режиме производится процедура «сокращения источника», то есть из него оставляются только те фрагменты, по которым найдено совпадение, либо можно запретить выдачу источников совсем. Это позволяет защитить коллекцию документов от распространения.

2.5 Безопасность протоколов

Взаимодействие систем www.antiplagiat.ru и Антиплагиат.РГБ осуществляется по шифрованному каналу с защитой по протоколу SSL. Авторизация системы проводится по специально подготовленным идентификационным ключам. Это гарантирует, что только один экземпляр системы может функционировать, как «библиотека документов РГБ» и никакая внешняя система не может получить несанкционированный доступ к пересылаемым данным.

Доступ к проверке по базе диссертаций имеют только те пользователи, которым администраторы системы явным образом открыли доступ к такой проверке. В настоящий момент поддерживается ограниченное предоставление доступа к проверке, причём ограничение может задаваться по числу проверяемых документов, по их суммарной длине и по времени использования доступа.

2.6 Подлинность отчётов

У каждого отчёта имеется несколько цифровых подписей. Необходимость нескольких подписей вызвана распределённостью системы и тем, что один отчёт строится сразу несколькими коллекциями, находящимися на разных серверах у разных подписчиков в сети.

Каждая коллекция имеет свои ключи и подписывает результаты своей работы - фрагмент отчёта, в дальнейшем называемый ревизией. Ревизия включает в себя контрольную сумму текста, вычисленную по алгоритму MD5 вместе со списком найденных источников и блоков цитирования. Опционно в ней могут содержаться сокращённые тексты самих источников (если включено сокращение источников, см. п. 2.4), а также атрибуты как отдельных источников, так и коллекции в целом. Наличие в ревизии контрольной суммы проверяемого документа гарантирует, что ревизия строилась именно для проверяемого документа, а не для какого-то другого текста (что исключает подмену текста проверяемого документа на другой, например пустой, с целью получения отчёта о полной оригинальности), а цифровая подпись самой ревизии гарантирует, что ревизия построена проверяющей коллекцией (исключается искажение самой ревизии).

После того, как все фрагменты отчёта построены, они через систему шлюзов отправляются к инициатору проверки. Например, если проверяемый документ находился в РГБ, то все фрагменты отправляются в РГБ, если с сайта www.antiplagiat.ru, то в хранилище данного сайта. Инициатор проверки собирает их в единый отчёт и подписывает его своим ключом, что гарантирует, что в отчёт включены все необходимые ревизии. Помимо этого, подпись отчёта позволяет однозначно определить создателя данного отчёта - при его просмотре посредством бесплатной

программы ReportViewer создатель отображается в нижней части экрана.

Таким образом, подделки отчётов полностью исключены.

2.7 Редактирование и просмотр отчётов

Несмотря на то, что система фильтрует незначащие заимствования, некоторые из них остаются в отчётах, т.к. значимость или незначимость некоторого фрагмента может оспариваться. Если невозможно точно определить, что фрагмент незначим, то он остаётся в отчёте и уже пользователь должен принять решение, значим он или нет. Для упрощения работы по анализу текста пользователю предоставлена возможность самостоятельно удалять некоторые найденные участки из отчёта, а также сохранять изменённые отчёты прямо в своей коллекции. Для защиты от злоупотребления удалением найденных блоков, при применении данной функции в отчёте появляется предупреждение, что некоторые блоки были удалены, а также возможность вернуть все блоки обратно. Также предусмотрена возможность экспортировать отчёты, в том числе и с удалёнными блоками, в виде отдельных файлов и просматривать их с помощью отдельного приложения, не требующего подключения к сети - AntiPlagiat ReportViewer.

2.8 Научный поиск

Сервис представляет собой средство сравнения более-менее обширного текста пользователя с каждым из текстов, имеющихся в хранилище. В отличии от системы "Антиплагиат" система "Научный поиск" в основном предназначена для обнаружения небольших совпадений, таких, как общие цитаты, имена лиц и учреждений, заглавия литературных произведений, устоявшиеся фразеологизмы и обороты речи. Предполагается, что получившаяся подборка документов позволит исследователю получить представление о направлении работ в смежных, параллельных и вообще как-то ассоциированных областях.

Сейчас область поиска системы «Научный поиск» совпадает с областью поиска системы Антиплагиат.РГБ, то есть поиск осуществляется по коллекции авторефератов и диссертаций по всем отраслям знаний с 1998 г. защиты.

Система возвращает список документов, которые похожи (содержат полные или частичные заимствования) на исследуемый.

Для каждого из них можно снова провести поиск - и получить список связанных уже с ним документов.

Список проверенных документов отражается в верхней части экрана. Всегда можно вернуться к любому из них. Если текущий документ имеется в электронном каталоге - можно перейти туда.

Отчёт содержит отрывки текущего документа, содержащие заимствования. Совпадающие фрагменты выделены жёлтым цветом. Стрелочки слева позволяют перемещаться от одного совпадающего фрагмента к другому. Номер в квадратных скобках - это порядковый номер документа в списке источников, из которого заимствован выделенный фрагмент.

2.9 Технические характеристики

Формирование хранилища

- Исходный объем PDF файлов – 2.5 TB, общий объем ANSI текстов - 57 GB
- Число документов – 260000, средний размер текста – 222 KB
- Общее время создания хранилища – 26 ч.
- Объем получившегося хранилища системы – 46 GB

Характеристики

- Оборудование: 2 двуядерных процессора Xeon 1.6 GHz, 4 GB Ram
- Реально используется только один жесткий диск емкостью 135 GB
- Время проверки документа по локальному хранилищу РГБ – не более 3 сек.
- Время проверки документа по 2-хранилищам РГБ и Антиплагиат одновременно – не более 5 секунд.

3 Особенности устройства системы

В данном разделе будет кратко описан основной модуль системы - коллекция, отвечающая за хранение документов, отчётов, а также поддержку индексов и выполнение поиска.

3.1 Хранение документов и атрибутов.

Коллекция хранит все данные в виде архивов - больших бинарных файлов, внутри которых подряд сохранены отдельные документы коллекции. Данных подход существенно экономит дисковое пространство за счёт отсутствия неиспользуемой ёмкости в конце кластеров файловой системы диска, а также существенно ускоряет поиск нужного текста по сравнению с хранением текстов в отдельных файлах. Для дополнительной экономии места поддерживается сжатие текстов различными архиваторами, на данный момент используются алгоритмы deflate (также применяется в архиваторе ZIP) и BWT (применяется в архиваторе BZ2). Для хранения данных определённого типа заводится свой архив, на данный момент их 6:

- Архив текстов документов. Обязателен, хранит тексты документов;
- Архив нормализованных текстов. Необязателен, в нём сохраняются видоизменённые тексты документов, подготовленные к поиску цитирования - все слова нормализованы (это и дало название данному архиву), т.е. приведены к начальной словоформе, буквы ё заменены на е и т.п.. Может быть перестроен по

архиву текстов, в случае его отключения нормализация текстов выполняется по мере необходимости, что замедляет процесс проверки документов;

- Архив документов. Необязателен, хранит исходные двоичные файлы документов (например *.doc; *.pdf). Нужен только в случае если пользователь захочет получить исходный двоичный файл документа, для поиска цитирования не используется;
- Архив ревизий. Обязателен, хранит построенные фрагменты отчётов;
- Архив атрибутов документов. Обязателен, хранит атрибуты каждого документа вместе с историей их изменения - имеется возможность проследить историю правок пользователями атрибутов указанного документа;
- Архив кэширования. Необязателен, хранит дополнительные двоичные данные для поискового ядра по часто используемым источникам. В случае его отключения ядро перестраивает эти данные при каждом обращении к источнику, что увеличивает время поиска.

Все архивы могут размещаться на отдельных физических дисках, что позволяет оптимизировать дисковый ввод/вывод и достигать оптимальной производительности дисковой подсистемы сервера.

3.2 Индексы.

Индексы - это структуры данных, позволяющие существенно ускорить поиск за счёт организации системы навигации по текстам и отсутствия необходимости перебора всей коллекции текстов для проверки факта наличия заданного фрагмента.

У коллекции есть два индекса, выполняющие одну и ту же задачу, но обладающие разными характеристиками:

- Постоянный индекс. Оптимизирован для хранения больших объёмов данных, обеспечивает быстрый поиск при любых объёмах коллекции, нечувствителен к её размеру. Относительно компактен, занимает мало места на диске. Добавление новых записей требует существенных временных затрат - приходится полностью пересматривать весь индекс.
- Временный индекс. Позволяет быстро добавлять новые документы, но с ростом объёма время поиска увеличивается. Предназначен для временной индексации документов, пока они не занесены в постоянный индекс.

Оба индекса поддерживают усечение, позволяющее обменивать качество проверки на снижение времени поиска и объёма индексов. Усечение индекса сделано таким образом, что существенное снижение его объёма приводит к несущественному снижению качества поиска. В частности, при принудительном усечении индекса в 4 раза средняя оценка оригинальности по тестовому корпусу, сформированному из загруженных пользователями документов выросла всего на 1% и составила 63%. При усечении индекса более, чем в 8 раз, качество начинает существенно снижаться. При усечении в 64 раза средняя оценка того же корпуса выросла на 14% и составила 76%. При применении неусечённого индекса оценка составила 62% оригинальности.

Потеря качества проявляется в основном на заимствованиях небольшой длины, поэтому если нужно отлавливать копирование только больших блоков (например, сразу по несколько страниц), можно использовать усечение вплоть до 64 раз, качество будет оставаться приемлемым.

При усечении индекса в N раз время поиска по данному индексу также снижается в N раз.

Все индексы позволяют дополнительно держать в оперативной памяти небольшой объём данных, ускоряющий обработку документов с большим процентом оригинальности. Данная надстройка работает по принципу hash-таблицы и осуществляет быстрый отсев фрагментов текста, заведомо отсутствующих в индексах на диске. Соответственно, отсеянные фрагменты искать на диске бесполезно, что позволяет экономить на обращениях к внешней памяти.

Существует возможность отключения индексов у коллекции. Отключение временного индекса позволяет очень быстро добавлять много документов в индекс, но поиск по ним начнётся только после перестройки постоянного индекса. Отключение обоих индексов делает невозможным поиск цитирования по данной коллекции. Целесообразно, если коллекция используется исключительно как хранилище документов, позволяет сэкономить немного памяти.

4 Перспективы системы

В ближайшее время планируется внедрение следующих компонентов:

- Приоритетов коллекций и отдельных документов с целью выявления первоисточников цитирования;
- Возможность задания параметров проверки для каждого документа индивидуально;
- Разработка дополнительных средств мониторинга использования текстовых баз и контроля отсутствия несанкционированного доступа;
- Добавление третьего промежуточного индекса в коллекцию;

4.1 Приоритеты коллекций

С ростом числа коллекций, а также их объёма, возникла проблема поиска первоисточников цитирования. Один и тот же фрагмент текста может содержаться в большом количестве источников, их число может достигать сотни и даже тысячи

документов. С целью экономии времени каждая коллекция (при конфигурации по умолчанию) ищет только один источник для каждого фрагмента текста. Если о документах ничего не известно, то выбрать документ-источник из множества документов, содержащих нужный фрагмент, можно только псевдослучайным образом - например, взять документ, попавший в коллекцию раньше других. К сожалению, данная стратегия иногда приводит к некорректным результатам - например, при цитировании текста закона или общеизвестного литературного произведения источником может быть объявлена другая диссертация, цитирующая тот же самый текст. Конечно, данное недоразумение будет разрешено при просмотре отчёта пользователем, но это заставляет пользователя затратить дополнительное время на редактирование блоков цитирования с целью переквалифицирования данного фрагмента.

При наличии у документов приоритетов, система сможет автоматически выбирать подходящий источник. Задание приоритетов потребует больше времени на грамотное составление хранилища, но существенно упростит работу с системой в дальнейшем. Вероятно, будет целесообразно назначить высокие приоритеты тем документам, которые заведомо являются первоисточниками текста, на их отбор и потребуется дополнительное время.

Помимо приоритетов документов вводятся также приоритеты целых коллекций, что позволяет регулировать значимость источников из различных хранилищ на этапе сборки отчёта из ревизий. При обнаружении фрагмента текста в источниках из разных коллекций, будет выбран источник из более приоритетной.

Например, для кого-то более важно заимствование из документов, находящихся в открытом доступе в сети internet, а для кого-то - из хранилища РГБ. Для достижения желаемых результатов достаточно будет поднять приоритет нужной коллекции.

Приоритеты коллекций могут также использоваться для исключения легитируемого цитирования из общей оценки документа - весь текст, найденный в высокоприоритетной коллекции не учитывается при вычислении итоговой оценки документа.

4.2 Индивидуальные параметры проверки

Каждая коллекция имеет множество настроек, определяющих, как она будет выполнять поиск. Сейчас все эти параметры задаются в конфигурационном файле коллекции и применяются при проверке любых документов.

Иногда возникает необходимость задания параметров индивидуально для каждого документа - например, для проверки одних документов достаточно грубой оценки оригинальности, но очень важна скорость проверки, для других, наоборот, в первую очередь важно качество. В частности, может потребоваться полный список источников для каждого фрагмента текста, построение которого может занять существенное время.

Предполагается сделать возможным включение в запрос на проверку дополнительных параметров, с их последующим сохранением в ревизии, с целью осуществления более гибкого применения системы и расширения сферы её применимости.

4.3 Контроль передаваемых данных

Несмотря на то, что система препятствует распространению текстовых баз, остаются некоторые вопросы по поводу невозможности извлечения текстов из коллекции сторонними людьми через её интерфейс, используемый для проверок документов. Действительно, невозможно дать гарантию того, что в кодировании системы безопасности не было допущено ошибок. С другой стороны, имеется гипотетическая возможность сознательного создания дырок для воровства чужих текстовых баз. Для контроля честности протокол системы сделан таким образом, что имеется возможность сохранять все передаваемые между модулями системы данные и обеспечить тем самым контроль за передачей текстов и отчётов.

Реализации контроля мешает сравнительно большое количество команд в протоколе системы, т.к. он рассчитан на широкую сферу применения. Даже в случае реализации модуля для анализа передаваемых данных с открытым кодом, в нём будет достаточно сложно разобраться.

Для решения данной проблемы предполагается выпустить модуль туннелирования, который отвечает за передачу данных между коллекцией подписчика и его шлюзом в упрощённой форме, с поддержкой необходимого минимума команд и без шифрования. В результате можно будет проанализировать каждый переданный системой байт данных и убедиться, что ничего лишнего передано не было.

4.4 Добавление промежуточного индекса в коллекцию

Как было описано ранее (см п.3.2) у коллекции имеется два индекса. К сожалению, при большой нагрузке по добавлению документов в индекс производительность коллекции падает - она либо будет постоянно занята перестройкой постоянного индекса, либо временный индекс станет большим и превратится в "узкое место" при поиске цитирования. Предполагается, что промежуточный индекс будет устроен аналогично постоянному, но по объёму будет существенно меньше его, что позволит выполнять его перестройку существенно быстрее.

Литература

[1] Публичный сайт системы Антиплагиат. http://www.antiplagiat.ru.

[2] Сайт системы Антиплагиат для ВУЗов. http://corp.antiplagiat.ru.

[3] Сайт системы Антиплагиат в РГБ. http://antiplagiat.rsl.ru.

[4] Сайт Электронной Библиотеки Диссертаций РГБ. http://diss.rsl.ru/.

[5] Ю.И. Журавлёв и др. «Система распознавания интеллектуальных заимствований «Антиплагиат» // Доклады 12-й всероссийской конференции «Математические методы распознавания образов» (ММРО-12). Москва, 2005. С. 329-332.

[6] Ю.И. Журавлёв и др. «О проекте «Антиплагиат» // Доклады международной конференции «Интеллектуализация обработки информации» - 2006. Симферополь, 2006. С. 92-94.

On integration Antiplagiat system in Russian State Library

Romanov Mikhail Yurievich
Zhitlukhin Dmitriy Anatolievich

In the article the description of the integration of the Antiplagiat system in the Russian State Library is given.

Within this integration the system of searching for adoptions over the dissertations and abstracts base is implemented in the RSL, and also the "scientific search" project support is developed.

The detailed exposition of the integration results, the system structure, the security matter and final technical and speed characteristics are given.