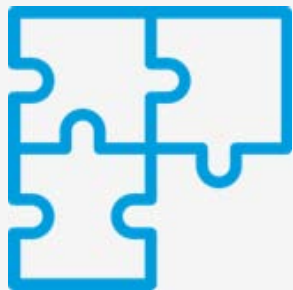


РУКОНТЕКСТ

Сервис поиска плагиата и анализа документов

Перспективные направления развития сервисов обнаружения текстовых заимствований:



Определение перефразированных объектов



Детектор техник сокрытия плагиата



Переводной плагиат



Интеграция с любыми системами ВУЗов



Определение перефразированных фрагментов

Технология TextApplainece:



Поиск заимствований. Анализирует документ с использованием метода многокритериальной оценки и сходства текста.



Визуализация заимствований. Отображает заимствования в удобном для проверяющего виде.



Выявление перефразированных фрагментов текста. За счет применения морфологического, синтаксического и семантического анализа текста.

Алгоритм выявления смыслового содержания текста

Проверяемый текст

Текст докладной записки делится на две части: Констатирующая (описательная), где излагаются имевшие место факты или описывается ситуация, вторая, где излагаются предложения, просьбы.

Сам метод заключается в следующем: на каждом шаге мы выбираем один из исходных элементов и вставляем его на нужную позицию в уже отсортированном списке, до тех пор, пока набор исходных данных не будет исчерпан.

Текст источника

Докладная записка обычно состоит из двух частей: в первой описывается сложившаяся ситуация, во второй излагаются предложения, просьбы, делаются конкретные выводы.

На каждом шаге алгоритма мы выбираем один из элементов входных данных и восстанавливаем его на нужную позицию в уже отсортированном списке, до тех пор пока набор входных данных не будет исчерпан.

* Заимствованием считается всё предложение

Полное сохранение разметки документа с визуализацией

The interface features a top navigation bar with buttons numbered 1 to 11. Below this, a summary bar displays three percentages: 43.53% (green), 7.54% (yellow), and 48.93% (red). A checkbox labeled 'Обозначить заимствования по источникам' is checked. Three buttons are present: 'Оригинальный текст', 'Непроверенный текст', and 'Заимствования из нескольких источников'. Below these are 'Показать все' and 'Скрыть все' buttons. A list of sources is shown, with the third item selected:

- 1. Методологический подх...
http://www.isa.ru/proceedings/images/d...
Г. П. Акимова, Е. В. Пашкина, А. ...
2008 76.23%
- 2. Электронные архивы: м...
http://www.isa.ru/proceedings/images/d...
Г. П. Акимова, М. А. Пашкин, Е. В. П...
2014 54.12%
- 3. Ситуационно-аналитиче...
http://www.isa.ru/proceedings/images/d...
Г. П. Акимова, А. В. Соловьев, Е...
2007 7.54%

The main text area on the right contains a paragraph with blue highlights:

На основании составленных частных моделей рисков, противодействия им, функционирования информационной системы создается общая модель катастрофоустойчивости. На основании данной модели проводятся расчеты и определяются стратегии повышения защищенности элементов ИС.

Идеальным вариантом проведения полноценного моделирования катастрофоустойчивости информационной системы является создание ситуационно-аналитических центров, реализующих модели функционирования ИС (см.). Их использование значительно повышает эффективность функционирования и развития информационной системы в целом.

Заключение

Большие информационные системы все чаще становятся неотъемлемой частью производственного процесса на промышленных предприятиях, коммерческих организациях и в государственных структурах. Чем крупнее организация, тем большая по масштабам информационная система требуется для охвата и управления всем производственным и/или технологическим циклом, но и тем больше риск потери критически важной информации.

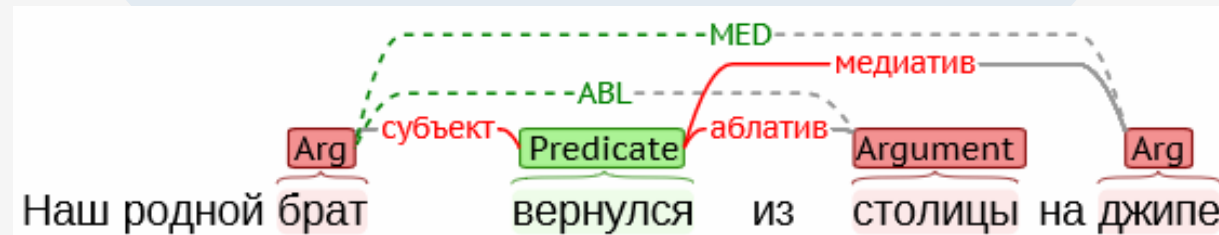
Для больших систем, работа которых связана с оперативной обработкой информации, а простой оборачивается большими материальными (или нематериальными: политические, имиджевые и т. д.) потерями, необходимо уделять большое внимание степени защищенности систем от катастрофических воздействий не только природного, но и техногенного,

Метод семантического поиска текстовых заимствований

Семантическая структура текста источника



Семантическая структура заимствования



При перефразировании семантика текста сохранилась



Определение попыток сокрытия заимствований



Форма представления результатов обходов

1234567891011

3.5%40.77%7.46%

Неизвестные слова

Слова с символами из разных алфавитов

Невидимый текст

Используемые алфавиты в словах с заменами символов:

CYRILLIC

LATIN

UNKNOWN

Интернете **содержания** и **аннотаций** выпусков похоже может быть компромиссом, **на** котором **согласятся** **издатели**, **отказывающие** **предоставлять** в **ддоступ** полные тексты и библиотечные **работники**, **сокращающие** свои **затраты** **на** **комплектование** и **аналитическую** роспись периодики.

Выводы

Главный вывод **заключается** в том, что для обеспечения полного и **качественного** **доступа** к педагогической периодике нужно **развивать** все модели **доступа**, а именно:

- o **Создание** общедоступных электронных полнотекстовых библиотек и открытых **архивов**;
- o **Организация** **доступа** к удаленным коммерческим **базам** **данным** периодики
- o **Развитие** **навигационных** систем и систем электронной **доставки** документов **на** основе **публикации** **содержания** выпуском и **аннотаций** **статей**;
- o Улучшение **комплектования** периодикой **классических** **образовательных** библиотек.

Причем **реализация** всех этих моделей (может быть, кроме **традиционного** **комплектования**) требует новых форм **координации** **библиотечно-информационной** деятельности в сфере **образования**. Именно эту цель **ставит** перед собой **создаваемый** при Российской **академии** **образования** Межведомственный совет по **информационным** **ресурсам** в сфере **образования**.

Любое действие в системе инициируется пользователем, для этого предусмотрено два интерфейса: Интерфейс администратора и Интерфейс пользователя. Под интерфейсом администратора понимается набор программного обеспечения, позволяющий управлять системой и поддерживать ее основные функции, такие как пополнение и редактирование словарей, настройка и/или новые поисковые ресурсы, управление пользователями, управление настройками системы и т.д. Под интерфейсом пользователя понимается WEB-интерфейс, с помощью которого выполняются поставленные задачи по поиску информации в Интернет, а также просмотр и обработка полученных в результате поиска данных.

Прежде чем любая-либо задача будет исполнена, она попадает в очередь задач. Задачи исполняются параллельно несколькими агентами, причем обработка может происходить на нескольких компьютерах локальной сети. Каждый агент записывает результаты своей работы в базу данных, которые потом будут выданы пользователю, разумеется, в уже обработанном виде. Задачи могут выполняться несколькими агентами, причем различный класс обработки агента может повлечь за собой постановку новых задач.

Наибольший интерес в системе представляет семантическая фильтрация найденных документов, основными задачами которой являются оценка семантической близости вопроса и документа (в процентах) и упорядочивание результирующего набора документов в соответствии с этой оценкой. Документы с большим процентом семантической релевантности показываются в первую очередь. [14, 15]. Для оценки релевантности используется лингвистический модуль. Релевантность найденных документов оценивается



Переводной плагиат



Машинный перевод текстов



Смысловой перевод текстов



Метод транзакции





Переводной плагиат

Построение матрицы терминов из массивов документов, которые имеют соответствия из разных языковых групп

Проверяемый текст

Плагиат выражается в публикации под своим именем чужого произведения, а также в заимствовании фрагментов чужих произведений без указания источника заимствования.

Текст источника

Plagiarism is the "wrongful appropriation" and "stealing and publication" of another author's "language, thoughts, ideas, or expressions" and the representation of them as one's own original work.



Интеграция с любыми системами ВУЗов



Для студентов и работников ВУЗа не потребуется никаких дополнительных действий. Для них единственным изменением будет открытие представления работы в нашей системе



Руконтекст не имеет доступа к личной информации, кроме той, что передает система ВУЗа)



Некоторые изменения доступны всем пользователям, как автоматическая индексация и отслеживание изменений отчёта

РУКОНТЕКСТ

Возможности



Единый аккаунт



Структура ВУЗа



Актуальные
ссылки

250 000 000
документов

Используемые базы



Обработка
дублей

РУКОНТЕКСТ

Возможности



Пакетная
загрузка



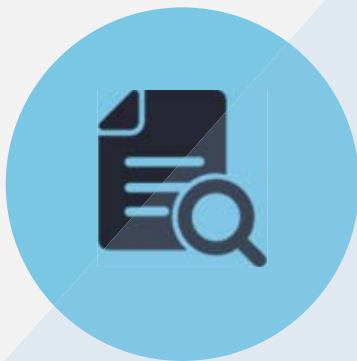
Экспорт
результатов



Интеграция с любыми
системами ВУЗов

РУКОНТЕКСТ

Дополнительные сервисы



Анализ
текста



Поиск похожих
документов



Тематический
анализ

РУКОНТЕКСТ

Анализ качества документа

РУКОНТЕКСТ

Поиск заимствований

Поиск плагиата

История проверок

Создать билет

Выданные билеты

Семантический поиск

Тематический анализ

Анализ текста

Аккаунт

Тарифы

[Проверить файл](#) [Проверить текст](#) [История проверок](#) [Пакетная загрузка](#)

Файл Текст

Загрузите файл с публикацией

Выбрать файл

Распознавать текст

Проверка методик обхода

Принимаемые файлы:

.html, .txt, .doc, .docx, .pdf, .ppt, .pptx, .xls, .rtf, .zip, .odt

Параметры

Год публикации 2016

Автор

Т Заглавие *

Проверяющий

Комментарий

Клиентоориентированность



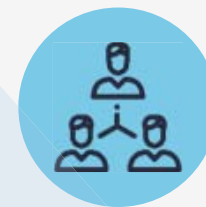
РУКОНТ
РУКОНТЕКСТ



Редактирование
структуры ВУЗа



Проверка по
билетам



Управление
аккаунтами

Новая версия «Руко́нтекст» 3.0


РУКО́НТЕКСТ


Поиск заимствований

Добро пожаловать, [User](#) | [Выход](#)

- Поиск плагиата
- История проверок
- Создать билет
- Выданные билеты
- Семантический поиск
- Тематический анализ
- Анализ текста
- Аккаунт
- Тарифы

[Проверить файл](#) [Проверить текст](#) [История проверок](#) [Приложение Руко́нтекст \(альфа-версия\)](#)


0 файлов в очереди на проверку 

 **Проверяющий ***

Выберите файл

или перетащите документ в это поле

Принимаемые файлы: .html, .txt, .doc, .docx, .pdf, .ppt, .pptx, .xls, .rtf, .odt, .zip, .rar

Распознавание текста 



Проверка методик обхода 



Проверить

Добавить организационную структуру



Новости

Когда проверяют на плагиат?

На базе программных средств, разработанных в ФИЦ ИУ РАН, компания РУКОНТ в 2015 году запустила сервис поиска плагиата РУКОНТЕКСТ.



РУКОНТЕКСТ

Промежуточные успехи и результаты

В настоящее время в РУКОНТЕКСТ:

~**250 млн.** документов размер индексной базы данных

~**20 000** документов - максимальная пиковая нагрузка в сутки
(июнь 2017)

~**600 000** документов проверено за последние пол года, при этом:

~**58%** средняя оригинальность текстов (очень много списанных работ)

~**5%** попыток обойти средства поиска плагиата

РУКОНТЕКСТ

Что дальше?

- Новые способы обхода – улучшенные средства по их определению
- Лучше средства визуализации текстовых заимствований
- Больше базы, расширение поддержки нескольких языков
- Кросс-языковой поиск текстовых заимствований

Плагиата, как массового явления, не будет только тогда, когда обмануть технические средства и преподавателей будет трудозатратнее, чем написать оригинальную работу, а наказание за попытку обмана будет суровым и неотвратимым

Благодарю за внимание!

Семенов Денис



+7(499)129-29-09

+7(499)129-72-12



semenov@akc.ru