

Музалев Сергей Владимирович

канд. экон. наук, доцент, ведущий научный сотрудник
Фонд информационного обеспечения науки,
ФГБОУ ВО «Финансовый университет
при Правительстве Российской Федерации»
г. Москва

Кузнецов Игорь Владимирович

научный сотрудник
Фонд информационного обеспечения науки
ассистент кафедры
ФГБОУ ВО «Московский государственный
технический университет им. Н.Э. Баумана»
г. Москва

Тузова Светлана Юрьевна

канд. хим. наук, начальник отдела научно-
технического сопровождения экспертизы
Фонд информационного обеспечения науки
г. Москва

Дивненко Ольга Владимировна

канд. пед. наук, доцент, профессор
НОУ ВО «Национальный институт бизнеса»
директор
Фонд информационного обеспечения науки
г. Москва

**ОСОБЕННОСТИ ВЫЯВЛЕНИЯ ТЕКСТОВЫХ ЗАИМСТВОВАНИЙ
В ХОДЕ ПРОВЕДЕНИЯ ЭКСПЕРТИЗЫ В РАМКАХ РЕАЛИЗАЦИИ
ФЦП «ИССЛЕДОВАНИЯ И РАЗРАБОТКИ ПО ПРИОРИТЕТНЫМ
НАПРАВЛЕНИЯМ РАЗВИТИЯ НАУЧНО-ТЕХНОЛОГИЧЕСКОГО
КОМПЛЕКСА РОССИИ НА 2014–2020 ГОДЫ»**

Аннотация: в работе рассмотрены основные системы специального программного обеспечения, позволяющие проанализировать научно-технические проекты, для проведения качественной экспертизы на некорректные заимствования. В работе также представлены функции, которыми обладают системы.

Ключевые слова: информация, культура заимствований, программы, сервисы, системы.

В рамках реализации Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы» (далее – Программа, ФЦП) экспертиза проводится на стадиях формирования тематики лотов, конкурсного отбора проектов по заявленным тематикам, непосредственно исполнения проектов.

Простота и скорость обмена данными в рамках информационного пространства дала возможность человеку получать доступ к огромному объему информации, что привело к обострению проблем культуры заимствований. Не является исключением и сфера деятельности ученых и исследователей по подготовке заявок и отчетных материалов в рамках Программы.

Проект не должен являться повторением научно-исследовательских работ (ПНИЭР, проектов), выполненных в предшествующие периоды и (или) выполняемых в сроки реализации проекта. Таково одно из требований конкурсной документации, выполнение которого позволяет, в том числе, исключать возможности двойного финансирования одних и тех же работ.

В рамках подготовки экспертных заключений эксперт обращает внимание непосредственно на проверку заявок и отчетных материалов на некорректные заимствования и их допустимость.

На современном этапе для проведения качественной экспертизы на некорректные заимствования в Дирекции научно-технических программ (далее – Дирекция) используются следующие основные системы специального программного обеспечения, позволяющие проанализировать научно-технические проекты:

– TextAppliance;

- Система «Антиплагиат»;
- «4science»;
- Система Анализа проектов [6].

Так, TextAppliance представляет собой программно-аппаратный комплекс, состоящий из сервера и интеллектуальных сервисов поиска и анализа больших коллекций текстовых документов. Основными сервисами являются:

- формирование и индексация больших массивов текстовых документов из интернета, баз данных, корпоративных хранилищ;
- семантический, фразовый и эксплоративный поиск;
- поиск тематически и семантически похожих документов;
- семантический поиск текстовых заимствований, в том числе сильно перефразированных;
- быстрая кластеризация и классификация документов и коллекций;
- формирование, сопоставление и анализ пользовательских коллекций документов;
- тематический анализ коллекций документов – выявление динамики публикаций по заданным темам в разных коллекциях на временной шкале;
- автоматическое формирование ключевых слов для документов и коллекций;
- автоматическое реферирование документов;
- анализ качества научных текстов – проверка на соответствие формальным требованиям к научным публикациям.

В большей степени, TextAppliance используется для выделения ключевых слов в тексте с целью дальнейшего анализа представленных заявок и тематики в таких базах данных как WOS и Scopus.

Система «Антиплагиат» также дает возможность проверять текстовые документы на наличие заимствований. Структура и технические характеристики

«Антиплагиата» обусловлены главным функциональным требованием: обеспечить пользователей удобным, быстрым и доступным web-сервисом для высококачественной проверки документов на наличие заимствований.

В данной системе предлагается полный отчет о проверке на заимствования, содержащий ранжированный список обнаруженных источников заимствований и полный текст проверяемого документа, в котором особым образом выделены заимствованные фрагменты текста. Полный отчет содержит функциональность, позволяющую исследовать и квалифицировать каждый заимствованный фрагмент текста.

Важно отметить, что область поиска системы «Антиплагиат» включает несколько частей:

- проиндексированные страницы сети интернет;
- коллекция полных текстов диссертаций и авторефератов Российской государственной библиотеки;
- коллекция текстов юридических и нормативных документов Lexpro;
- коллекция полных текстов статей научной электронной библиотеки Elibrary.ru.

Таким образом, «Антиплагиат» использует уникальные алгоритмы, обеспечивающие быстрый и эффективный поиск заимствованных фрагментов.

Третья система специального программного обеспечения, позволяющая анализировать научно-технические проекты на предмет заимствований – это «4science». Эта независимая площадка для взаимодействия науки и бизнеса создана для развития кооперации между учеными, предпринимателями и руководителями госструктур, для создания площадки для обмена опытом и формирования сообщества агентов влияния на экосистему.

Заметим, что «4science» позволяет сократить время на работу с документами, дает возможность сравнивать отчеты с другими на прямой и косвенный плагиат, а также отследить источники вплоть до конкретных страниц.

«Система Анализа проектов» создана с целью противодействия прямым и косвенным заимствованиям в проектах, конкурсных заявках, заявках на формирование тематики, а также для совершенствования процедуры отбора проектов, предлагаемых к реализации в рамках ФЦП. Она применяется как на стадии анализа конкурсных заявок, так и на стадии анализа отчетных материалов. «Система Анализа проектов», в отличие от классических систем, позволяет сравнивать документы на семантическую близость, независимо от того, произведена ли замена слов или поменялся их порядок. Другими словами, есть возможность увидеть косвенный плагиат, когда документы близки по смыслу. Также система позволяет не учитывать формальные разделы, шаблонную и несодержательную информацию. Заметим, что классические системы такую предварительную фильтрацию не производят.

В заключение важно отметить, что все представленные выше системы дают возможность проводить анализ прямого копирования (при этом используется разный механизм поиска и сравнения). Каждая из представленных систем обладает рядом следующих функций:

- выявление прямого копирования документ-документ;
- выявление прямого копирования объект-объект (контракт-контракт);
- выявление семантической близости документ-документ;
- выявление семантической близости объект-объект.

В качестве отличительных характеристик можно выделить различное наполнение баз данных и разный уровень интерпретации и визуализации информации.

Важно отметить, что главной ценностью использования специального программного обеспечения для интеграции сервисов выявления заимствований в электронных документах и сервисов, предоставляющих информацию о научно-технических проектах и наукометрических данных, является выявление и недопущение прямых и косвенных заимствований в заявках и отчетной документации, а также возможность поиска похожих или семантически близких работ и результатов исполнения контрактов.

Список литературы

1. Журавлев Ю.И. Система распознавания интеллектуальных заимствований «антиплагиат» / Ю.И. Журавлев [и др.] // Математические методы распознавания образов. – Т. 12. – 2005. – №1. – С. 329–332.
2. Осипов Г.С. TextAppliance – новое решение для интеллектуального поиска и анализа больших массивов текстов / Г.С. Осипов, И.В. Смирнов, И.А. Тихомиров, И.В. Соченков // Материалы второго международного профессионального форума «Книга. Культура. Образование. Инновации» («Крым-2016»).
3. Официальный сайт «4science» [Электронный ресурс]. – Режим доступа: <https://4science.ru/>
4. Официальный сайт «Антиплагиат» [Электронный ресурс]. – Режим доступа: <http://www.antiplagiat.ru/>
5. Официальный сайт «TextAppliance» [Электронный ресурс]. – Режим доступа: [https://](https://textapp.ru/) <http://textapp.ru/>
6. [Электронный ресурс]. – Режим доступа: <http://cap.sstp.ru/>