

«УТВЕРЖДАЮ»

Директор Федерального  
государственного бюджетного  
учреждения науки  
Библиотеки по естественным наукам  
Российской академии наук (БЕН РАН)

доктор технических наук, профессор

Н.Е. Каленов

«12» ноября 2015 г.



## ОТЗЫВ

**Ведущей организации Федерального государственного  
бюджетного учреждения науки Библиотеки по естественным наукам  
Российской академии наук (БЕН РАН)**

**о диссертации Хорошилова Алексея Александровича  
«Методы, модели, алгоритмы и экспериментальное программное  
обеспечение автоматического выявления неявно выраженных  
заимствований в научно-технических текстах», представленной  
на соискание ученой степени кандидата технических наук  
по специальности 05.13.17 - «Теоретические основы информатики»**

**Актуальность темы исследования.** Совершенствование информационных технологий обеспечило широкие возможности доступа к различным информационным ресурсам. Эти возможности оказали огромное позитивное влияние на развитие современного общества. Но при этом в обществе возникли проблемы, связанные с незаконным использованием информации, в частности, с незаконным заимствованием чужих результатов интеллектуальной деятельности. Это противоправное заимствование, обозначаемое термином «плагиат», сейчас получило широкое распространение в сфере образования. Существующие на IT-рынке

автоматизированные средства обнаружения заимствований могут выявлять в основном только факты явного заимствования, выполненные путем прямого копирования заимствованного текста. Современные средства способны также выявлять факты заимствования, полученные путем незначительного изменения структуры или лексического состава текстов.

Диссертационная работа Хорошилова Алексея Александровича посвящена актуальной теме – решению проблемы достоверного выявления различного рода заимствований, включая заимствования, выполненные недобросовестными авторами путем значительной переработки лексического состава и структуры заимствованного текста. Решение этой проблемы позволит более объективно оценивать квалификационные работы и, в конечном итоге, повысить качество среднего и высшего образования в нашей стране.

**Целью исследования** является решение проблемы выявления неявно выраженных заимствований в текстах документов. В соответствии с указанной целью в работе поставлены следующие задачи:

1. Исследовать и разработать модели представления смыслового содержания текстов документов.
2. Исследовать и разработать методы и алгоритмы выявления наименований понятий в текстах документов и унификации их смыслового содержания.
3. Исследовать и разработать методы, модели и алгоритмы автоматического выявления заимствований в текстах документов, включая случаи неявно выраженных заимствований.
4. Разработать программное обеспечение для решения задачи автоматического выявления заимствований в текстах документов (включая неявно выраженные заимствования).
5. Провести экспериментальное исследование, устанавливающее достоверность теоретических концепций и эффективность разработанных методов выявления заимствований в текстах документов.

**Структура и объем работы.** Диссертация Хорошилова А. А. состоит из введения, четырех глав, заключения, списка используемой литературы (117 наименований) и 5-ти приложений (содержит 159 страниц текста, 24 рисунка, 29 таблиц).

**Во введении** обоснована актуальность темы, определен предмет исследования, сформулированы цель и задачи исследования, изложены основные



результаты и их научная новизна, обоснована теоретическая и практическая значимость полученных результатов.

**В первой главе** исследуется понятие плагиата в рамках действующего законодательства и анализируются основные методы, использующиеся в программных средствах выявления заимствований в текстах научно-технических документов.

**Во второй главе** описываются технологии и процедуры автоматической обработки текстовой информации, используемые автором для решения задач структурирования и формализации смыслового содержания текстов. Также в этой главе описываются процедуры выявления понятийного состава текстов, установления парадигматических, синтагматических и ассоциативных связей между наименованиями понятий и приведения их к унифицированному формализованному представлению, разработанные для решения задачи выявления неявно выраженных заимствований в текстах документов.

**В третьей главе** автором излагается теоретическая концепция, на которой основан предлагаемый метод обнаружения заимствований в текстах. Затем описывается модель процесса выявления всех случаев заимствований, включая случаи неявно выраженных заимствований, при этом для каждого этапа обработки текста предлагаются модели представления смысла текстов документов.

**В четвертой главе** рассматривается разработанный автором программный комплекс автоматического выявления заимствований. Этот программный комплекс используется при проведении экспериментального исследования и для расчёта основных показателей эффективности разработанных методов выявления заимствований в текстах документов.

**В заключении** приводятся основные результаты, полученные в процессе диссертационного исследования.

**В приложениях** приводятся иллюстрации результатов функционирования разработанного программного комплекса и официальные свидетельства на платформенное ПО МетаФраз, а также акты о внедрении результатов диссертационного исследования.



**Научная новизна исследований.** Стоит отметить, что в работе найдено удачное решение по оптимизации процесса выявления заимствований путем его разделения на три этапа, на каждом из которых используются модели разной степени сложности. Это позволяет повысить эффективность решения задачи, но в то же время оставляет возможности для дальнейшей оптимизации вычислительных процессов при обработке больших объемов информации.

В работе получен ряд новых результатов, среди которых отметим следующие:

1. Методы, алгоритмы и экспериментальное программное обеспечение процесса формализации содержания документов с учетом их смыслового представления.

2. Модели, методы, алгоритмы и экспериментальное программное обеспечение процесса автоматического выявления заимствований в текстах документов, включая случаи неявно выраженных заимствований.

3. Комплексное решение задачи автоматического выявления заимствований в текстах документов (включая неявно выраженные заимствования) на основе анализа их смысловой структуры.

**Практическая ценность работы** заключается в том, что научные и практические результаты диссертационных исследований были использованы в НИР, проводимых в ФГАНУ ЦИТиС, а также реализованы в информационной системе «Мониторинг СМИ», которая функционирует в Ситуационно-кризисном центре госкорпорации РосАтом в режиме промышленной эксплуатации. В ее базе данных уже накоплено более 37 млн. документов и ежедневно в нее поступает и оперативно обрабатывается более 100 тыс. документов и новостных сообщений по различным тематикам.

**Рекомендации по использованию результатов и выводов диссертации.** В настоящее время для выявления заимствований в текстах документов используются автоматизированные средства, которые позволяют находить в большинстве случаев только факты прямого заимствования. Для повышения эффективности выявления случаев неявно выраженного плагиата в документах, полученных путем значительного изменения структуры и лексического состава заимствованного текста, можно рекомендовать использование разработанного автором программного обеспечения. Использование этого программного обеспечения как совместно с имеющимися средствами, так и отдельно от них повысит эффективность выявления фактов заимствований в научно-технических текстах.



Также необходимо отметить, что сфера применения разработанных методов не ограничивается поиском заимствований в текстах документов. Созданные в процессе выполнения исследования средства могут быть использованы практически в любых задачах, где требуются качественные семантические инструменты смысловой обработки текстовой информации.

**Достоверность** полученных результатов подтверждена экспериментальными исследованиями, проведенными с использованием созданных средств.

**Апробация полученных результатов.** Результаты работы докладывались на 8 российских и международных конференциях. В открытой печати по теме диссертации опубликовано 14 работ, из них 5 работ в изданиях, входящих в Перечень ВАК Минобрнауки РФ. Получено 6 свидетельств об официальной регистрации программ для ЭВМ в Роспатенте.

Замечания по работе:

1. Автор в работе подробно описывает программно-лингвистический инструментарий, используемый им в дальнейшем для построения модели процесса выявления заимствований. При этом не совсем понятно, возможно ли использование другого инструментария для реализации поставленных автором задач исследования. Например, инструментария, базирующегося на ситуационно-реляционном подходе или расширенных сетях переходов.

2. Общеизвестно, что в зарубежной литературе часто излагаются идеи, методы, требующие решения в отечественной практике. Некоторые недобросовестные авторы иногда выдают такие методы и идеи за свои собственные. Современные системы выявления заимствований не способны выявлять такие заимствования. Автор не указал, возможно ли использование предложенного метода для выявления заимствований из зарубежных источников.

3. Из описания алгоритма выявления заимствований видно, что процесс, выполняемый этим алгоритмом, достаточно сложный и требует значительных вычислительных ресурсов. Следовало бы для обеспечения приемлемой скорости работы описанных алгоритмов предусмотреть распараллеливание обработки при работе с большим числом документов.

4. В тексте диссертации на рисунке 2.1 (стр. 57) допущена опечатка, перепутаны местами надписи «Выявление НП без словаря» и «Выявление НП по словарю».



5. Не описана методика работы экспертов, анализирующих результаты работы программного комплекса по выявлению фактов заимствований в документах, в которых возможны неявно выраженные заимствования.

6. Следует добавить процедуру проверки орфографии анализируемых документов.

Приведенные замечания не снижают общей положительной оценки работы. Диссертация А.А. Хорошилова является законченной научно-технической работой, выполнена на высоком научном уровне. Сформулированные положения подкрепляются экспериментальными исследованиями. Автореферат достаточно полно отображает содержание диссертации.

Диссертационная работа отвечает критериям п. 9, 10 «Положения о присуждении ученых степеней», а ее автор, Хорошилов Алексей Александрович, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.17 – «Теоретические основы информатики».

Отзыв на диссертацию и автореферат обсужден и одобрен на заседании научно-технического совета Федерального государственного бюджетного учреждения науки Библиотеки по естественным наукам Российской академии наук (БЕН РАН) 12 ноября 2015 года, протокол № 2.

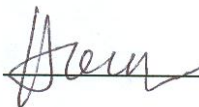
Заместитель директора по научной работе  
Федерального государственного бюджетного  
Учреждения науки Библиотеки по естественным  
наукам Российской академии наук (БЕН РАН),  
доктор технических наук, специальность 05.25.05,

профессор



В.А. Цветкова

Ведущий научный сотрудник  
Федерального государственного бюджетного  
Учреждения науки Библиотеки по естественным  
наукам Российской академии наук (БЕН РАН),  
кандидат технических наук,  
специальность 05.25.05



К.П. Погорелко

119991, г.Москва, ул.Знаменка, д.11/11, (495) 691-22-89, [head@benran.ru](mailto:head@benran.ru),  
Федеральное государственное бюджетное учреждение науки Библиотека по естественным наукам Российской академии наук (БЕН РАН)