

## **ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА**

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**

Одним из элементов научного, а также учебного процесса (по крайней мере, в высших учебных заведениях) является выполнение научно-исследовательских и квалификационных работ, основной отличительной характеристикой которых является свойство новизны. При этом любая работа строится на результатах предшествующих, что предполагает обязательное и достаточно объемное использование информационных ресурсов.

Современное состояние информационных технологий значительно облегчило процессы поиска информации и подготовку документов, но при этом спровоцировало широкомасштабное, в том числе и прямое заимствование текстов: фактически сложился метод написания работ, который получил название «copy-paste».

Такое использование информационных ресурсов можно расценить как противоправное заимствование или плагиат. Под плагиатом, в целом, понимают дословное изложение чужого текста или изложение чужого текста с заменой слов и выражений без изменения содержания заимствованного текста.

Контроль на предмет плагиата был и есть объективной необходимостью и, кроме того, он несет свой положительный вклад и как элемент воспитания научной добросовестности.

Необходимо отметить, что в последние годы с этим явлением ведется борьба и с использованием средств автоматизации. Однако, имеющиеся на сегодняшний день средства выявления заимствований не всегда позволяют обнаруживать неявные заимствования в измененных текстах. Авторы приспосабливаются к особенностям работы имеющихся систем контроля и редактируют свои работы, ориентируясь на особенности и недостатки этих систем. Таким образом, существует необходимость в создании универсаль-

ного инструмента, который позволит выявлять все случаи заимствований, включая неявно выраженные.

Исходя из вышеизложенного, **актуальность** диссертационного исследования, посвященного проблеме выявления неявно выраженных заимствований в научно-технических текстах, не вызывает сомнений.

Разработанные автором методы, модели, алгоритмы и экспериментальное программное обеспечение автоматического выявления неявно выраженных заимствований в научно-технических текстах позволяет более объективно оценивать научные и квалификационные работы различного уровня. А это, в свою очередь, поможет повысить уровень подготовки научных и профессиональных кадров и, в конечном итоге, улучшить качество, как научных исследований, так и высшего образования.

### **Степень обоснованности научных результатов, выводов и рекомендаций**

Автор достаточно корректно использует известные научные методы, адекватные природе изучаемых процессов.

Им изучены и проанализированы известные подходы и теоретические положения других работ.

Результаты, полученные теоретическим путем, не противоречат практике. Научные положения и выводы, сформулированные в диссертационной работе, базируются на достаточном экспериментальном материале.

Предлагаемое диссертантом решение задачи по автоматическому выявлению неявно выраженных заимствований позволяет определять в массиве документов близкие по содержанию, выявляя в них текстовые фрагменты, которые наиболее вероятно имеют некоторую, формально оцениваемую смысловую идентичность.

### **Оценка новизны и достоверности**

В качестве новых научных результатов диссертантом выдвинуты:

1. Модель процесса выявления заимствований в документах (включая неявно выраженные) на основе анализа их смысловой структуры.
2. Метод установления смысловой близости и смысловой схожести фрагментов текста на основе анализа их смысловой структуры.
3. Алгоритм выявления наименований понятий в научно-технических текстах.



4. Алгоритм автоматического установления смысловых отношений между наименованиями понятий.

5. Алгоритм выявления заимствований в документах (включая неявно выраженные).

6. Экспериментальный программный комплекс выявления заимствований в научно-технических текстах (включая неявно выраженные).

7. Результаты исследования по автоматическому выявлению заимствований, подтверждающие достоверность и эффективность предложенных методов.

К основным результатам работы, отличающимся научной новизной, относятся:

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

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.

Достоверность научных результатов, выводов и рекомендаций, приведенных в диссертационной работе, достигнута за счет: корректного применения апробированного математического аппарата, использования методов, адекватных природе изученных процессов, верификации результатов в рамках известных теоретических конструкций, проведения экспериментальной оценки адекватности разработанной модели. Полученные теоретические и экспериментальные результаты обладают свойством непротиворечивости и воспроизводимости.

Практическая значимость работы заключается в том, что научные и практические результаты диссертационных исследований были использованы в Федеральном государственном автономном научном учреждении «Центр информационных технологий и систем органов исполнительной власти» (ФГАНУ ЦИТиС) в рамках государственного задания на НИР в 2012-2014 гг. по теме «Исследование и разработка методов семантической экспертизы структуры и содержания научно-технических документов, а также наличия регламентированных для данного типа документов разделов и выявления несанкционированных заимствований (включая неявные заимствования)» при создании подсистемы выявления заимствований в текстах.



Также практические результаты, полученные автором, были использованы в рамках создания промышленной системы «Мониторинг СМИ» для Ситуационно-кризисного центра Госкорпорации Росатом (ФГУП «СКЦ Росатома»), реализующей функции сбора, консолидации, оперативной обработки поступающих документов для решения задачи формализации смыслового содержания и установления смысловой близости документов.

Результаты работы достаточно полно изложены в 14-ти публикациях, в том числе в 5 научных статьях в ведущих рецензируемых изданиях, апробированы на научных конференциях, соискатель является соавтором 6-ти официально зарегистрированных программ.

### **Замечания по диссертационной работе**

По существу работы можно сделать некоторые замечания:

1. В работе приводится сравнение эффективности разработанного метода и распространённого метода выявления заимствований – метода «шинглов» с точки зрения полноты и точности выявления заимствования. Но при этом не приводятся сравнительные характеристики по их быстрдействию и затратам вычислительных ресурсов.
2. В работе приводится три метода формализации наименований понятий, но из текста работы не совсем понятно для каких задач и в каких случаях необходимо использовать каждый из предлагаемых методов.
3. Пороговое значение смысловой близости автором выбрано экспериментально, однако неясно надо ли и как будет он определяться для других предметных областей или массивов документов другого вида.
4. В работе имеются некоторые, на мой взгляд, неудачные формулировки. Некорректно представлено выражение для меры смысловой близости.

Также представляется, что предложенные модели и алгоритмы инвариантны по отношению к предметной области (а не специфичны для научно-технических текстов): по крайней мере, из изложенного материала диссертации не следует явных ограничений на область применения.

Отмеченные недостатки не являются принципиальными и не влияют на ценность полученных в диссертационной работе научных результатов.

### **Заключение**

Представленная диссертация является законченным научно-исследовательским трудом, выполненным автором самостоятельно и на высоком научном уровне. Рассмотренная в диссертационной работе проблема

выявления неявно выраженных заимствований в научно-технических текстах соответствует специальности 05.13.17 – «Теоретические основы информатики».

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».

Доктор технических наук, профессор,  
Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Национальный исследовательский ядерный университет «МИФИ», профессор кафедры системного анализа



Н. В. Максимов

«17» ноября 2015 г.

Максимов Николай Вениаминович

E-mail: nv-maks@yandex.ru

Тел: +7 903 683-12-83

Адрес: 114409, г. Москва, Каширское ш., д. 31.

Подпись удостоверяю  
Заместитель начальника отдела  
документационного обеспечения  
НИЯУ МИФИ



*Хорошилова А.А.*