

УДК 004.912

ПРИМЕНЕНИЕ АЛГОРИТМОВ СЕМАНТИЧЕСКОГО АНАЛИЗА ТЕКСТОВОЙ ИНФОРМАЦИИ ДЛЯ ОПРЕДЕЛЕНИЯ ЗАИМСТВОВАНИЙ

Мошина О. В., Гордеева О. А.

Самарский национальный исследовательский университет
имени академика С. П. Королёва, г. Самара

Задача семантического анализа текстов на сегодняшний день является востребованной и быстроразвивающейся областью. Семантический анализ способствует классификации текстов, определению предметной области, а также проверке текста на наличие заимствований (плагиат).

На сегодняшний день проблема плагиата имеет высокую актуальность. И для решения задачи по обнаружению плагиата применяют системы, определяющие процент заимствования текстов. Среди систем проверки на плагиат выделяют следующие виды: «онлайновые» системы поиска оригинальных источников в сети Интернет, «оффлайновые» системы поиска дубликатов в пределах локальных коллекций, а также универсальные системы поиска дубликатов на основании текстов собственной коллекции [1].

В основе всех методов лежат эффективные численные алгоритмы поиска и обнаружения заимствований. Главным недостатком многих из автоматических систем поиска плагиата является отсутствие детального лингвистического анализа, в том числе семантического, способствующего решению проблемы автоматизированной обработки исходного текста (естественного языка).

Технологии семантического поиска плагиата предполагают выявление смысловых заимствований, анализ текстов (лингвистический, морфологический, семантико-синтаксический). Среди достоинств семантического подхода выделяют его нечувствительность к сильному перефразированию проверяемого текста.

Цель данной работы – исследование существующих методов и систем семантического анализа текстовой информации для определения неправомερных заимствований.

В данной работе будут рассмотрены модели семантического анализа, предполагающие определение сущностей исходного текста, свойств и отношений между ними [2]. Будут рассмотрены и реализованы в программной части некоторые основные методы и алгоритмы семантической проверки на плагиат, а также в ходе эксперимента проведен анализ полученных результатов (исследований).

Библиографический список

1. Романов, А. С. Модификация метода накопительных сумм для проверки однородности текста и выявления плагиата [Текст]/А.С. Романов//Электронные средства и системы управления. -2013. -№ 2. -С. 30-38.
2. Шелманов, А. О. Исследование методов автоматического анализа текстов и разработка интегрированной системы семантико-синтаксического анализа [Текст]: дисс. канд. техн. наук: 05.13.17: защищена 07.10.2015 – М, 2015. –210 с.