

Проблема атрибуции анонимных и псевдонимных статей в журналах "Время", "Эпоха".

Аннотация. Атрибуция текстов — это достаточно важный аспект науки. С его помощью возможно определить авторов огромного множества анонимных и псевдонимных произведений. В настоящее время существует большое количество текстов, которые необходимо проанализировать и процесс атрибуции будет всегда востребован как для старых, так и для новых сочинений.

Настоящее исследование посвящено ~~к~~ непосредственному изучению атрибуции в ф^{ило}логии и математической лингвистике, ее роли, проблематики, целей и задач.

Основная идея данной работы заключается в разработке специализированной программы, которая на основе вычислений расстояний между зависимыми частями речи может отнести тот или иной текст (в данном случае - псевдонимные и анонимные статьи из журналов «Время» и «Эпоха») к определенной группе, после чего становится известен автор произведения.

Введение. В настоящее время существует широкий набор методик определения авторства текстов. К ним относятся морфологические, синтаксические и лексические. Важную роль в данной области играет объем анализируемого произведения: исследователи говорят о важности длины публикации. Именно поэтому наибольшую достоверность в определении авторства демонстрирует тексты больших объемов. В случае коротких текстов необходимо ~~и дальше~~ синтезировать существующие методики и создавать новые, чтобы добиться соответствующей эффективности.

В работе описаны существующие методики определения авторства, их положительные и отрицательные стороны, определена сложность выбора параметров для анализа текстов, конкретизированы ошибки в работах некоторых исследователей.

Непосредственно цель диссертации заключается в определении авторства произведений из уже вышеупомянутых публицистических работ. Таким образом, для достижения результата были решены следующие задачи:

- Проведен исторический анализ атрибуции: ее понятие и роль в лингвистике;
- Проведен разбор существующих методов атрибуции произведений в математической лингвистике;
- Проведен анализ существующих параметров для определения авторства текстов;

- Проведен разбор существующих положительных и отрицательных сторон в исследованиях различных специалистов в области филологии и математической лингвистики;
- ~~Изучена~~ роль расстояния между зависимыми частями речи в математической лингвистике;
- Была исследована программа TreeTagger: процесс ее работы, характеристики, разбор выводов;
- Были выбраны анонимные и псевдонимные статьи из журналов «Время» и «Эпоха» для проведения практической части исследовательской работы;
- Фиксированы полученные результаты, их анализ.

В данном исследовании также упоминаются существующие ошибки в исследованиях с использованием определения расстояний между зависимыми частями в атрибуции произведений: ошибка разделения целых предложений на более простые, что затрудняет эффективное использование заданного параметра.

Таким образом, структура исследования определяется изучением истории возникновения атрибуции, ее роль в лингвистике, анализом существующих методов и параметров, а также ошибок в исследованиях в области математической лингвистики для определения авторства текстов, разбором и использованием существующих программ, на платформе которых при использовании расстояний между зависимыми частями можно определить автора того или иного произведения.

План-проспект.

1. Обзор существующих методов определения авторства.
Глава описывает работы исследователей по существующим методам определения авторства, в том числе детально разобраны параметры, которые применяются при атрибуции, их положительные и отрицательные стороны, роль компьютерных аппаратов для данной области, история возникновения различных программ, описание их работы.
2. Использование расстояний между зависимыми частями при атрибуции анонимных и псевдонимных статей в журналах «Время» и «Эпоха».
В данном разделе представлено описание выбранных статей из вышеупомянутых журналов, а также результаты исследования на основе TreeTagger. Проведен разбор результатов, описаны положительные и отрицательные стороны работы с данным параметром для определения авторства текстов.

Заключение. Получены следующие результаты:

- ~~Изучена история атрибуции, ее методов;~~
- Сформированы новые пути решения по улучшению качества определения авторства произведений, эффективности и правдивости выходной информации;
- Определены авторы некоторых статей из журналов «Время» и «Эпоха»;
- ~~Была~~ разработана рекомендации по дальнейшим исследованиям в данной области.

В результате исследования были сформированы новые пути решения по улучшению качества определения авторства произведений. Также определены авторы некоторых статей из журналов «Время» и «Эпоха».

Полученные результаты позволяют говорить о возможности и целесообразности применения метода использования расстояний между зависимыми частями речи на основе стилеразличающих синтаксических параметров с учетом параметров индивидуального авторского стиля для определения авторов анонимных и псевдонимных публицистических статей из Петербургских журналов «Время» и «Эпоха».

В целом данный метод представляется весьма перспективным и может использоваться в дальнейших исследованиях при решении аналогичных задач.