



Для воздействия на ученицу и ее одноклассников, адресант использует прием информирования о негативных последствиях неподчинения.

Так как нам известны только фрагменты речевого поведения участников буллинга, мы не можем оценить семиотические приемы вариативной интерпретации события (аннулирующий, фингирующий, идентифицирующий или модальный способ преобразования условно нейтрального повествования). В рассказе-воспоминании приемы вариативной интерпретации события не используются.

Описывая речевые особенности конкретного проявления вербальной агрессии, лингвист-эксперт помогает правосудию.

Литература

1. Кожина, М.Н. Речевой жанр и речевой акт (некоторые аспекты проблемы): сб. науч. работ / М.Н. Кожина. – Саратов: ГосУНЦ «Колледж», 1999. – С. 52–61.
2. Анцупов, А.Я. Словарь конфликтолога / А.Я. Анцупов, А.И. Шипилов. – СПб.: Питер, 2006. – 528 с.
3. Ожегов, С.И. Толковый словарь русского языка / С.И. Ожегов, Н.Ю. Шведова. – Изд. 4-е, доп. – М.: ИТИ Технологии.
4. Колосов, С.А. Агрессия в языке и дискурсе / С.А. Колосов // Вестник ТвГУ, Серия: Филология (14). – С. 198–208.
5. The village Беларусь: Буллинг который всегда с тобой [Электронный ресурс]. – URL: <http://www.the-village.me/village/culture/personal/263325-bullying>. – Дата доступа: 14.01.2018.

АНАЛИЗ СТИЛЯ ТЕКСТА ДЛЯ ЦЕЛЕЙ ЕГО АТРИБУЦИИ С ПРИМЕНЕНИЕМ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

С.В. Чащин

Как показали многочисленные публикации, объектом проводимых в интересах суда и/или следствия экспертиз (исследований) может выступать текст. Исследование текста может проводиться с целью установления авторства или получения каких-либо сведений об авторе и условиях создания текстового документа (атрибуция текста) [1].

Задачи атрибуции можно разделить на идентификационные и диагностические [1].

Идентификационные задачи позволяют осуществить проверку авторства:

- подтвердить авторство определенного лица;
- исключить авторство определенного лица;
- проверить тот факт, что автором всего текста был один и тот же человек;
- проверить тот факт, что написавший текст является при этом его настоящим автором.

В качестве примеров (хотя и вне судебного расследования) можно отметить решение вопросов об авторстве: «Тихого Дона» М. Шолоховым или М. Шолоховым и Ф. Крюковым [2, с.22], анонимных и псевдонимных произведений, приписываемых А.С. Пушкину [3], «Элегии» Вильямом Шекспиром [4, с. 244–245] и др.

Диагностические задачи позволяют определить личностные характеристики автора (образовательный уровень, родной язык, знание иностранных языков, происхождение, место постоянного проживания и др.) и / или факт сознательного искажения письменной речи.

Методы атрибуции позволяют исследовать текст на пяти уровнях: пунктуационном, орфографическом, синтаксическом, лексико-фразеологическом, стилистическом [2].



Под «авторским стилем» обычно понимаются последние три уровня. Анализ именно синтаксического, лексико-фразеологического и стилистического уровней представляет наибольший интерес и наибольшую сложность.

Методы анализа стиля можно разделить на две большие группы – экспертные и формальные [1]. Экспертные методы предполагают исследование текста профессиональным лингвистом-экспертом. К формальным относятся приемы из теории вероятностей и математической статистики и алгоритмы машинного обучения. Классификация основных формальных методов атрибуции текстов дана, например, в работе [6] и, в более развернутом виде, в [8].

Обзор основных формальных подходов к атрибуции текстов дан в работах [1; 2; 4–6]. Анализ как русскоязычных, так и иностранных текстов показал растущую актуальность в применении алгоритмов машинного обучения при проведении исследований в области атрибуции текстов [2; 4; 6] – класса методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач.

При этом, с учетом результатов опубликованных в [1; 2; 4–8] исследований, представляется целесообразным при построении таких систем применять количественные признаки, характеризующие стилистические особенности автора спорного текста. Обзор основных признаков, используемых в атрибуции текста, дан в [4–6].

В условиях роста популярности формальных методов исследований высокую значимость приобретают системы поддержки принятия решений, позволяющие исследователю сосредоточиться на анализе полученных результатов без необходимости самостоятельно проводить промежуточные расчеты. Такие системы позволяют, во-первых, избежать ошибок при выборе соответствующих алгоритмов, проведении расчетов и оценке результатов, во-вторых, – ускорить производство исследований за счет высокой скорости расчетов. Важнейшими этапами разработки таких систем являются определение признаков, по которым проводится исследование, и определение конкретных алгоритмов с учетом ограничений, накладываемых высокой социальной значимостью выводов экспертиз (исследований) и повышенными требованиями к их объективности.

Проведенным исследованием (с применением специально разработанной информационной системы, реализующей отбор проб, токенизацию (разбиение на слова) текстов, и использующей алгоритмы машинного обучения для построения модели) на текстах русских авторов конца 19 – начала 20 веков (Л.Н. Толстой, М. Горький, А.П. Чехов) установлено следующее. В настоящее время единого универсального алгоритма и универсального набора признаков, обеспечивающих высокую (более 80%) точность проводимых исследований вне зависимости от избранного стиля и объемов текста, не существует. Минимальный объем художественного текста, при котором точность модели превышала 60%, составлял 5.500 слов. В зависимости от применяемых алгоритмов, размера проб (либо всего текста) и избираемого пула признаков удалось получить точность в пределах 60–100% (мода – 87%) при чувствительности системы 0,6–0,85.

Так, в качестве алгоритмов использовались:

- метод k-ближайших соседей;
- метод простейшего байесовского классификатора;
- метод логистической регрессии.

В целом, модели с использованием каждого из трех алгоритмов показали примерно равные результаты точности и чувствительности при достаточном их обучении. Оптимальные результаты были получены для текстов объемом не менее 150.000 символов с пробами от 6,500 слов, при проведении кросс-валидации модели методом 10 сложений. С ростом количества итераций кросс-валидации повышается точность результатов, что может обеспечить снижение объемов исследуемых текстов и проб без снижения точности модели.



Представляется целесообразным проводить дальнейшее изучение возможности применения алгоритмов машинного обучения в практике автороведческой экспертизы с подбором оптимальных признаков, алгоритмов и объемов текста для различных стилей авторских текстов.

Литература

1. Батура, Т.В. Формальные методы определения авторства текстов / Т.В. Батура // Вестник НГУ. Серия: Информационные технологии / под ред. А.М. Федотова. – Новосибирск, 2012. – Том 10, вып. 4. – С. 81–94.
2. Гоголева, В.А. Математический подход к установлению авторства и времени создания текста на основе исследования его энтропии / В.А. Гоголева, А.П. Шкарапута // Вестник Пермского университета, разд. «Математика. Механика. Информатика» / под ред. О.Г. Пенского. – Пермь, 2014. – Вып.4 (27). – 2014. – С. 22–28.
3. Хозяинов, С.А. Атрибуция публицистических произведений, приписываемых А.С. Пушкину: дис. ... канд. филол. наук: 10.02.21 / С.А. Хозяинов; СПбГУ. – СПб., 2008. – 264 л.
4. Juola, P. Authorship Attribution / P.Juola // Foundations and Trends in Information Retrieval. – Hanover, 2006. – № 1(3). – С. 233–334
5. Романов, А.С. Методика и программный комплекс для идентификации автора неизвестного текста: автореф. дис. ... канд. техн. наук: 05.13.18 / А.С. Романов; ТУСУР. – Томск, 2010. – 26 с.
6. Романов, А.С. Методика идентификации автора текста на основе аппарата опорных векторов / А.С. Романов / Доклады ТУСУРа; под ред. А.А. Шелупанова. – Томск, 2009. – Вып. № 1 (19), часть 2. – С.36–42.
7. Шумская, А.О. Идентифицирующие признаки текстовых сообщений при определении автора / А.О. Шумская // Ползуновский вестник/ под ред. Е.С. Ананьевой. – Барнаул, 2013. – С. 265–266.
8. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / Е.И.Большакова [и др.] – М.: МИЭМ, 2011. – 272 с.

К ВОПРОСУ О ПРОВЕДЕНИИ СУДЕБНОЙ РЕЛИГИОВЕДЧЕСКОЙ ЭКСПЕРТИЗЫ В РЕСПУБЛИКЕ КАЗАХСТАН

Н.А. Муканова

Как и на всем постсоветском пространстве, на территории Республики Казахстан с возрастанием религиозной активности наблюдается и рост распространения материалов религиозного содержания, в том числе пропаганды религиозно мотивированного экстремизма и терроризма, информации агитационного характера, призывающих к вступлению в ряды религиозных организаций деструктивного характера.

Если несколько лет назад на исследование представлялись книги, брошюры, аудиокассеты, то в настоящее время поступают такие объекты, в которых заметен значительный технический прогресс в отношении передачи информации посредством усовершенствования технологий.

Анализ экспертной практики показывает и изменение содержания объектов. Так, если в 2001–2003 годах наблюдался рост поступления экспертиз в отношении таких религиозных организаций, как: «Свидетели Иеговы», «Саентология», «Алля-Аят», то с 2004 года исследование проводилось в основном по материалам религиозно-политической органи-