

УДК 004.912

РАЗРАБОТКА МЕТОДА РАССЛЕДОВАНИЯ УТЕЧЕК КОНФИДЕНЦИАЛЬНОЙ ИНФОРМАЦИИ В ЭЛЕКТРОННЫХ СООБЩЕНИЯХ НА ТЮРКСКОМ ЯЗЫКЕ НА ОСНОВЕ ЛИНГВИСТИЧЕСКОЙ ИДЕНТИФИКАЦИИ АВТОРА С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ

Покасова А.И. (Университет ИТМО, Санкт-Петербург)

Научный руководитель – к.т.н, доцент Воробьева А.А.
(Университет ИТМО, Санкт-Петербург)

Стремительное развитие технологий обмена информацией также повысило риск утечки оной, ввиду возможности анонимизации автора сообщения в данных системах обмена. Современные методы идентификации автора сообщения отлично показали себя на флективных языках, однако малоэффективны для агглютинативных языков, к которым относятся языки тюркской группы. Таким образом, необходимо разработать метод идентификации автора сообщения с учетом особенностей языков данной группы.

Введение. В проанализированных работах, посвященных проблеме идентификации автора электронного сообщения, главным подходом к построению информационно-поискового образа сообщения является анализ стилометрических признаков сообщения. Данная модель формирования информационно-поискового образа сообщения малоэффективна для агглютинативных языков. Следовательно, необходимо выбрать иную модель представления сообщения, а также нейросетевую модель классификации для разработки метода идентификации автора сообщения.

Основная часть. Предложенный метод использует иное формализованное представление сообщения: представление в виде графа. Вершинам графа соответствуют слова сообщений, полученные при токенизации коллекции на словесном уровне, ребра показывают наличие связей между словами. Каждый лингвистический уровень обрабатываемого документа связан с одним из компонентов графа.

Выводы. Предложенный метод позволил увеличить точность классификации по сравнению с методами, использующими анализ стилометрических признаков сообщения. Это позволяет использовать его на практике в таких областях, как информационная безопасность и форензика.

Покасова А.И. (автор)

Подпись

Воробьева А.А. (научный руководитель)

Подпись