

# Кроссязычный поиск дубликатов

*Е. В. Тищенко<sup>1</sup>, К. В. Воронцов<sup>2</sup>*

В данной статье рассматривается задача кроссязычного поиска текстового плагиата. Современные методы векторизации документов и последующего поиска совпадений преимущественно основываются на одном языке, что приводит к возникновению проблемы возникновения однообразных мультязыковых коллекций документов.

Целью работы является получение модели, кодирующей информацию о распределении слов в тексте независимо от их языковой принадлежности, при этом ограниченной по размеру и времени обучения для ее практического использования.

## <sup>1</sup> 1 Введение

<sup>2</sup> ...