

Кроссязычный поиск дубликатов

Е. В. Тищенко¹, К. В. Воронцов²

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.

¹ 1 Введение

² ...