Платформа хранения, доставки и показа видео YouTube является одной из самых популярных в наше время. Она обладает большим спектром возможностей, что позволяет упрощать процессы работы с видеоматериалами. Одним из таких процессов является поиск по тексту субтитров YouTube, генерируемых автоматически, что предоставляет возможность реализации алгоритма автоматизации этого процесса. Это позволит повысить производительность организаций, активно использующих YouTube в работе.

В работе рассматривается методы YouTube API: авторизация и получения субтитров. Рассмотрены возможные методы создания обратного индекса и поиска для формата получаемых субтитров, которые сравниваются с библиотекой ElasticSearch.

YouTube is one of the most popular video storing, sharing and showing platform. It has many abilities that may simplify video processing. An example of such process is searching over video text without having it. YouTube is creating captions automatically using voice-recognizing algorithms. This ability helps to develop YouTube captions search algorithm.

This article considers methods of authorization, captions retrieval, indexing and searching. YouTube API provides captions and authorization methods. Few searching approaches are described: inverted indexing with cosine similarity and Elastic Search library. There is comparison of these approaches that helps to decide which model to use in some cases.