

Студ. А.В. Тарасенко

Науч. рук. доц., к.т.н., Н.А. Жилияк

(кафедра информационных систем и технологий, БГТУ)

СЕРВИС ДЛЯ ПОДБОРА НОВОСТНЫХ СТАТЕЙ НА ОСНОВЕ ПРЕДПОЧТЕНИЙ ПОЛЬЗОВАТЕЛЯ

Переход общества в цифровую эру обусловил увеличение спроса на потребление информации в электронном виде. Как результат на данный момент необходимы быстрые и эффективные способы обработки большого количества информации и её структурирование для дальнейшего подбора для конечного пользователя. Всё это привело к увеличению спроса на различные алгоритмы анализа текстовой информации. Идет переход от ручного анализа текста к машинному. В связи с этим нужно создать эффективный алгоритм для анализа текста и выделения ключевых слов, для последующего применения данного алгоритма новостными порталами и агрегаторами.

Для решения поставленной задачи необходимо очистить входные данные [1]. Данный алгоритм разработан и представлен на рисунке 1. После очистки данных необходимо выделить в тексте ключевые слова т.к с их помощью, на основе алгоритма TextRankи будет происходить подбор новостного контента [2].

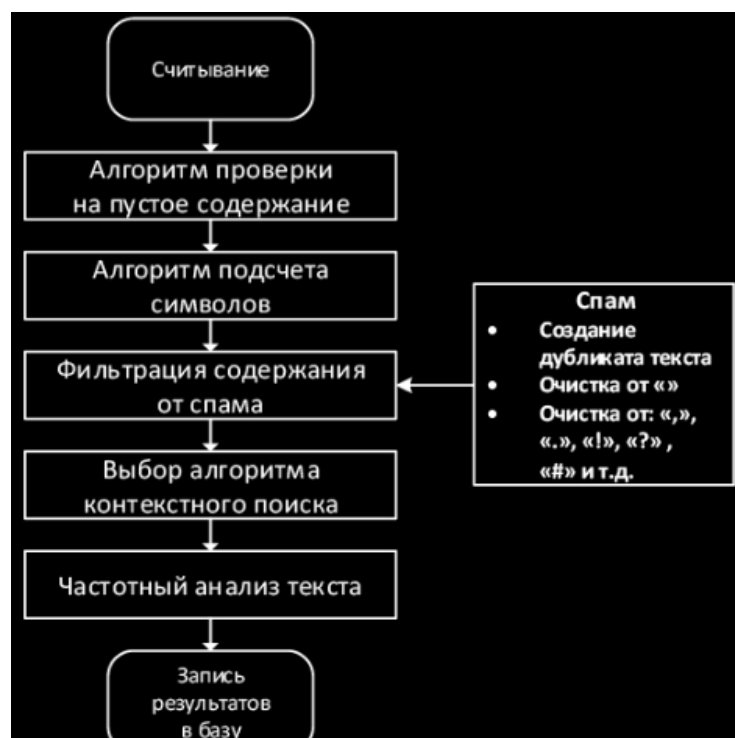


Рисунок 1 – Алгоритм анализа текста

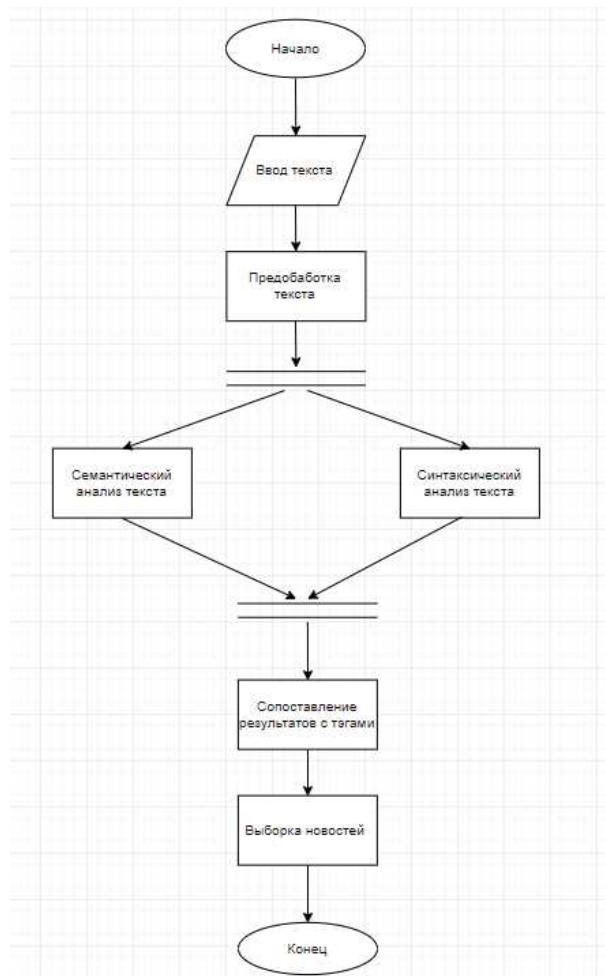


Рисунок 2 – Алгоритм подбора новостного контента

Сервис, созданный на основе алгоритма, разработанного и представленного на рисунке 2, позволяет анализировать входной текст с высокой скоростью и на основе его выдавать схожие статьи. Данный сервис может быть использован новостными порталами и агрегаторами не только для подбора новостей, но и для контекстной рекламы.

ЛИТЕРАТУРА

1. Web Rank/ Shaoping N. [et al] // USA: Google Ltd. – 2018. – P. 36–37.
2. Синтаксический разбор / Буркнцев В.А. // Министерство образования и науки Российской Федерации. Федеральное агентство по образованию Российской Федерации. Еленцкий Государственный университет им И.А.Бунина. – P. 10–15.