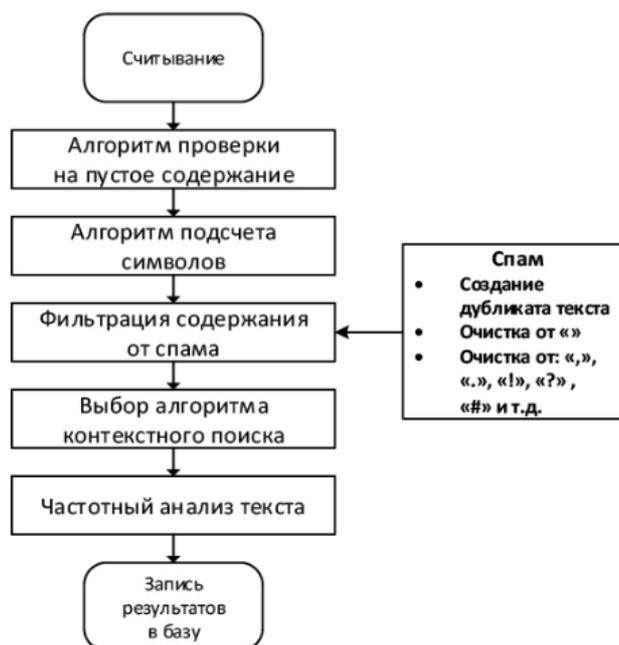


## **СЕРВИС ДЛЯ СЕМАНТИЧЕСКОГО И СИНТАКСИЧЕСКОГО АНАЛИЗА СЛОВ**

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 целесообразно 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] и выделить в нём ключевые слова т. к. с их помощью, на основе алгоритма TextRank и будет происходить анализ исходного текста.

Сначала машина находит в своем основном индексе все документы, в которых встречаются слова из запроса пользователя. Затем уже среди них выбираются те, где выше плотность ключевых слов (из запроса пользователя) в тексте документа, в его заголовке TITLE, в тегах акцентирования (H1 -H6, STRONG, EM) и т.п. Так же принимается в расчет текст ссылок (анкор), ведущих с других проектов на эти документы. В этом тексте тоже анализируется плотность ключевых слов из запроса пользователя. Называется эта штука ссылочным ранжированием.



**Рисунок 1 – Алгоритм анализа текста**

Ну, и в конце концов полученные результаты немного могут быть подкорректированы с учетом показателя Page Rank. В результате чего, будут расставлены по своим местам в поисковой выдаче документы с одинаковой релевантностью по данному запросу. Яндекс тоже использует свой ВИЦ (взвешенный индекс цитирования) для корректировки своей выдачи

Сервис созданный на основе данного алгоритма позволит быстро анализировать входной текст 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.