Необходим инструмент обработки и анализа контента с целью нахождения потенциальных возможностей для перелинковки внутри сайта.

## Териминология

**Перелинковка** - расположение ссылок на одних страницах сайта ведущих на другие страницы сайта с целью

1. Информирования пользователя о наличии сопутствующих материалов релевантных тематике изучаемой им страницы
2. Распределение входящего внешнего “веса” на целевые страницы сайта
3. Повышения релевантности страниц акцепторов за счет добавления целевых ключевых слов в анкоры ссылок на страницах донорах.

**Страницы Доноры** - страницы на которых была размещена ссылка ведущая на другие страницы сайта и страницы других сайтов.

**Страницы Акцепторы** - страницы сайта на которые указывают ссылки размещенные на страницах донорах.

**Пассаж** (предложение) - единица текста ограниченный дефинитивным знаком препинания (“.” или “!” или “?”) и пробелом с одной стороны (начало) и дефинитивным знаком препинания в конце

**Лемматизация** - это метод морфологического анализа, который сводится к приведению словоформы к ее первоначальной словарной форме (лемме).

**Типы вхождений**

1. Сверх точное - ключевое слово найдено в той же форме, с тем же порядком слов и не разбитое другими словами и артиклями
2. Точное - прямое вхождение ключа (порядок слов и отсутствие других слов между ними) в пассаж но допускается изменение словоформ составляющих его слов.
3. Разбитое - полное (все слова) вхождение ключа в пассаж но допускается разбиение ключа другими словами и изменение словоформ.
4. Частичное - вхождение части ключа в пассаж, допускается изменение словоформ.

## Вводные

Доступные данные.

1. Единица контента-текста
2. Список ключевых слов с указанием релевантных страниц акцепторов

Пример Входящих данных - [ТУТ](https://docs.google.com/spreadsheets/d/16SaoSvJ-_Fm2qmXMXcyZ_3oRqur2qb8J/edit?usp=sharing&ouid=100546750707529639281&rtpof=true&sd=true)

Предварительный список операций.

1. Разбить текст на список пассажей
2. Осуществить поиск ключевых слов в пассажи в сверх точном вхождении
3. Привести к лемме пассажи
4. Привести к лемме ключи
5. Осуществить проверку вхождения приведенных к лемме ключей по следующим типам вхождений
   1. Точное
   2. Разбитое
   3. Частичное

## Формат отчета

**Вкладка 1 (Ключи)**

Колонка 1 - Список ключевых слов

Строка 1 - Список типов вхождений

В ячейках пересечений отдавать порядковый номер пассажа (если больше одного, то через запятую) по типу вхождения.

**Вкладка 2 (Пассажи)**

Колонка 1 - Нумерация пассажей

Колонка 2 - Пассажи расположенные в порядке их очередности в представленном тексте

Пример отчета (схематичный) - [ТУТ](https://docs.google.com/spreadsheets/d/16TdgS44WRPdLtMCxxyJ7eSIkYtALIeSG/edit?usp=sharing&ouid=100546750707529639281&rtpof=true&sd=true)