

Файл на вход – имеет ключевое слово, количество показов ключевого слова, набор урл (обрабатывается как текстовая строка).

#### ШАГ 1 создание групп.

Если у разных ключевых слов есть общие урл, то такие ключевые слова объединяются в одну группу. Порог объединения задается пользователем. Это целое число от 2 до 30.

Например, порог стоит «4» все ключевые слова имеющие (больше или равно) 4 или более общих урл объединяются в одну группу.

Ключевые слова внутри созданной группы сортируются по убыванию показов ключевого слова.

Все урл на основе которых ключевые слова объединились в группу записываются в файл на выход.

В группе может быть от 1 до бесконечности ключевых слов.

В конце файла идут единичные ключевые слова без общих урл, сортировка по убыванию показов ключевого слова.

В зависимости от порога на выходе получается от 150 до 250 групп.

#### ШАГ 2 приоритизация созданных групп.

На этом шаге идет работа только с группами созданными на 1 шаге.

Каждой группе необходимо присвоить id.

id 1 берется группа с самым большим числом «Показов всех ключей группы» ей присваивается id 1. Далее смотрится, с какими группами у нее есть минимум 1 общий урл.

Далее из всех групп с минимум 1 общим урл, берутся только группы с **3\*** общими урл, среди групп с 3 общими урл идет сортировка по убыванию показов группы и присваивается id 2, id 3, id 4 и т. Далее берутся группы только группы с 2 общими урл, среди групп с 2 общими урл идет сортировка по убыванию показов группы и присваивается id 5, id 6, id 7 и т. Далее берутся группы только группы с 1 общими урл, среди групп с 1 общими урл идет сортировка по убыванию показов группы и присваивается id 8, id 9, id 10 и т.

Далее берется следующая группа с самым большим количеством показов группы ей присваивается id 11. Далее смотрится, с какими группами у нее есть минимум 1 общий урл.

Далее из всех групп с минимум 1 общим урл, берутся только группы с 3 общими урл, среди групп с 3 общими урл идет сортировка по убыванию показов группы и присваивается id 12, id 13, id 14 и т. Далее логика идентична описанной выше.

В конце идут группы (и / или одиночные ключевые слова) не связанные общими урл, они сортируются просто по убыванию к-ва показов ключевого слова / группы.

**\*3** – это не константа. В первом шаге мы задаем параметр «порог объединения» это число всегда минус – 1, минус – 2, минус – 3 от порога объединения. Например, если порог объединения «4», то это число будет 3, 2, 1. Если порог объединения «7» число будет 6, 5, 4, 3, 2, 1.