

УДК: 004.02

EDN: [BMYPVI](#)



Методы кластеризации в SEO проектировании

А.В. Усачёв¹, А.А. Яблокова^{2*}

¹Сибирский федеральный университет, Гуманитарный институт, кафедра информационных технологий в креативных и культурных индустриях, пр. Свободный, д. 82А, г. Красноярск, 660041, Россия

²Красноярский краевой Дом науки и техники Российского союза научных и инженерных общественных объединений, ул. Урицкого, д. 61, г. Красноярск, 660049, Россия

* E-mail: alena.yabl@yandex.ru

Аннотация. В статье рассмотрены такие методы кластеризации семантического ядра, как логическая группировка, группировка по топам поисковой выдачи, группировка по семантической схожести, и три способа кластеризации в рамках метода группировки по топам поисковой выдачи: hard, middle, soft.

Ключевые слова: семантическое ядро, кластеризация, логическая группировка, группировка по топам, группировка по семантической схожести

Keyword clustering in SEO

A.V. Usachev¹, A.A. Yablokova^{2*}

¹Siberian Federal University, School for the Humanities, Department of Information Technologies in Creative and Cultural Industries, 82A Svobodny pr., Krasnoyarsk, 660041, Russia

²Krasnoyarsk Regional Science and Technology City Hall of the Russian Union of Scientific and Engineering Public Associations, 61 Uritskogo Street, Krasnoyarsk, 660049, Russia

E-mail: alena.yabl@yandex.ru

Abstract. This paper deals with methods and ways of clustering keywords, describes their advantages and disadvantages. Identified factors to consider for choosing a method of keyword clustering and recommendations are given for grouping keywords into clusters.

Keywords: semantic core, clustering, logic-based keyword clustering, SERP-based keyword clustering, semantic-based keyword clustering

1. Введение

Внедрение цифровых технологий в различные аспекты деятельности человека, общества и государства предопределило становление и развитие информационного общества, в котором информационные технологии рассматриваются в качестве одной из движущих сил развития [1]. Парадигма информационного общества, затрагивающая цифровую экономику, предусматривает существенные изменения не только в способах производства товаров и услуг, но и их реализации. Для эффективного ведения бизнеса в современных условиях рекомендуется внедрять IT – стратегии в бизнес – процессы. Одним из направлений IT – стратегии выступает реализация товаров и услуг через веб – сайты.

2. Цель исследования

Про высокие показатели эффективности работы веб – сайтов можно говорить только в том случае, если они закрывают потребности пользователей и бизнеса. Чтобы сайт выполнял возложенные на него задачи, необходимо производить разработку в соответствии с рядом требований, в том числе и с учетом требований поисковых систем [2].

Параметры, по которым поисковая система оценивает эффективность сайта и определяет его позицию в органической выдаче в ответ на запрос пользователя, а также сама формула ранжирования, являются коммерческой тайной, но ряд основных факторов ранжирования известен. Одной из групп факторов ранжирования, которые учитываются поисковыми системами, является внутренняя оптимизация сайта, которая соотносится с таким этапом разработки сайта, как SEO – проектирование [3]. В рамках данной работы рассмотрены методы кластеризации в SEO – проектировании и даны рекомендации по применению.

3. Результаты

Этап SEO – проектирования позволяет подготовить сайт для дальнейшего продвижения через поисковые системы. От того, как разработчик подойдет к реализации этапа, зависит будет ли сайт проиндексирован веб – краулером, добавлен в поисковый индекс и в результатах поисковой выдачи находиться в топах, чтобы пользователь перешел на сайт и выполнил целевое действие [4].

Итак, этап SEO – проектирования – это подготовка ресурса для последующей индексации и продвижения на более высокие позиции в поисковой выдаче для

привлечения целевой аудитории на сайт. Без оптимизации ресурса под требования поисковых систем в результатах поиска сайт будет отсутствовать, т.к. не проиндексирован, или находиться на низких позициях, что приведет к снижению целевого трафика и отсутствию конверсии.

Первый шаг SEO – проектирования – сформировать семантическое ядро. В семантическое ядро сайта должны входить ключевые слова, характеризующие тематику ресурса, товары или услуги заказчика, для реализации которых разрабатывается ресурс и по которым пользователи будут искать сайт через поисковые системы.

Формирование семантического ядра начинается со сбора базовых ключевых слов, которые определяют тематику сайта. Подбор направлений для продвижения и базовых ключевых слов осуществляется на основе информации о товарах или услугах заказчика, также учитывается семантика конкурентов. Список базовых ключей дополняется при помощи сервисов подбора слов, поисковых подсказок и спецификаторов. Полученное семантическое ядро чистят, удаляя нецелевые вхождения, чтобы в ядре остались только релевантные ключи. [5].

После очистки семантического ядра, проводится его кластеризация: вся семантика делится на кластеры для последующего распределения по страницам сайта.

Метод группировки слов подбирается с учетом объема семантического ядра. Для наиболее точного результата рекомендуется использовать комбинацию разных методов (логической группировки, группировки по семантической схожести и группировки по топу поисковой выдачи) и способов кластеризации (soft, moderate (middle), hard) [6].

Метод логической группировки используется для семантических ядер, в которых находится до нескольких сотен ключевых слов. При логической группировке оценивается смысловая близость: среди ключевых слов предполагаемой группы должно быть ведущее слово, которое будет общим свойством и под которое будут подбираться остальные ключевые слова.

Логическая группировка проводится вручную в одну или несколько итераций: ключевые слова группируют на кластеры по одному или нескольким общим признакам. Как правило, первоначально семантическое ядро разделяют по коммерческому и информационному интену (определяя интент каждого ключа вручную, но в случаях, когда интент самостоятельно определить невозможно, используют сервисы для определения процента коммерческой), внутри полученных кластеров проводят повторную группировку ключевых слов по одному или нескольким общим признакам,

разделяя кластеры на подгруппы. Иногда наоборот, первоначально разделяют ключевые слова на кластеры по общим признакам (в одну или несколько итераций), а внутри кластеров делят на подгруппы по информационному и коммерческому интенту.

Логическая группировка дает достаточно точный результат и возможность группировки по нескольким признакам, но чем больше ключевых слов в семантическом ядре, тем выше вероятность допустить ошибки, поэтому метод логической группировки применим к ресурсам с небольшими семантическими ядрами.

Чтобы кластеризовать семантическое ядро по методу группировки по топам поисковой выдачи, необходимо проанализировать сайты со схожей тематикой, которые находятся в топе поисковой выдачи и найти пересечения ключевых слов на веб – страницах. Ключевые слова из списка поочередно проверяются онлайн – сервисом или программой, по каждому ключу составляется топ поисковой выдачи. Результаты поиска по каждому из ключей сравниваются между собой на предмет одинаковых адресов страниц, и в зависимости от установленного порога пересечений, ключевые слова объединяются в кластер. Ключевые слова, по которым совпадения в поисковой выдаче не были найдены, объединяются в отдельную группу, которые обрабатываются вручную.

Существуют три способа кластеризации по топам поисковой выдачи: soft, moderate (middle), hard [7].

При soft (мягкой) кластеризации вспомогательные ключевые слова сравниваются с высокочастотным ключом по количеству общих URL в топе выбранной поисковой системы. Ключевое слово будет включено в кластер, если количество одинаковых URL соответствует установленному порогу пересечений. В полученной группе все вспомогательные ключи связаны с высокочастотным ключом, но между собой ключи могут не пересекаться по общим URL в поисковой выдаче. Поэтому в кластер добавляется много ключевых слов со слабыми связями, из – за чего страдает точность группировки.

При hard (строгой) кластеризации вспомогательные ключевые слова сравниваются с высокочастотным ключом по количеству общих URL в топе поисковой выдачи, дополнительно проводится сравнение вспомогательных ключевых слов между собой. Для включения в кластер вспомогательные ключевые слова должны быть связаны с главным ключом и друг с другом общими URL, количество которых соответствует установленному порогу пересечений. При строгой кластеризации отсеивается много

ключевых слов, однако удастся разбить семантическое ядро на кластеры с максимально релевантными ключевыми словами.

При moderate / middle (умеренной) кластеризации вспомогательные ключевые слова сравниваются с высокочастотным ключом и между собой по количеству общих URL в топе поисковой выдачи. Если количество общих URL между главным и вспомогательными ключевыми словами соответствует степени группировки, ключи объединяются в кластер. В полученной группе все вспомогательные ключевые слова будут попарно связаны друг с другом, но в разных парах количество общих URL могут быть разными. Умеренная кластеризация обеспечивает большую точность, чем при мягкой кластеризации, и не позволяет отсеивать большинство ключевых слов, как при строгой группировке.

Метод группировки по топам поисковой выдачи (комбинация разных способов) является популярным и востребованным, т.к. позволяет кластеризовать тысячи ключевых слов за короткое время в автоматическом режиме с достаточно высокой точностью и с минимальными проверками вручную после.

Метод кластеризации по семантической схожести основан на применении нейросетей. Суть кластеризации заключается в разделении ключевых слов с учетом весовых коэффициентов на кластеры по смысловому совпадению. В кластеры попадают семантически близкие ключи (помимо смыслового совпадения ключевые слова могут быть схожи по морфологическому признаку), если есть пересечение с ключевым словом с тяжелым весом. Кластеризация по семантической схожести используется только для крупных семантических ядер, т.к. требуется разработка формул и обучение нейросети для обработки ключевых слов в автоматическом режиме.

4. Заключение

Каждый из рассмотренных методов и способов кластеризации обладает достоинствами и недостатками. Поэтому при выборе метода кластеризации или их комбинации в первую очередь следует учитывать объем семантического ядра сайта. Если семантического ядро небольшое, то подойдет логическая группировка в ручном режиме, т.к. приобретать подписку платных программ ради группировки нескольких сотен слов нецелесообразно. Группировка по топам поисковой выдачи подходит для сайтов с семантическими ядрами, которые содержат несколько десятков тысяч ключевых слов, которые нужно кластеризировать в автоматическом режиме за короткий срок. В

зависимости от тематики и типа сайта выбирается способ: если это коммерческий ресурс и высококонкурентная тематика, то выбирают способ hard, чтобы распределение ключевых слов на кластеры обеспечивало высокую релевантность страниц сайта. Если для сайта точность кластеризации не важна (для сайтов с неконкурентными тематиками) или необходимо привлечь трафик на страницы по большому количеству запросов (информационные ресурсы) применяется способ soft. Способ middle применяется, если нужно найти золотую середину между способом hard и soft: отсеять часть ключевых слов, которые не релевантны посадочной странице и оставить те, которые хотя бы частично входят в тематический кластер. К таким сайтам можно отнести маленькие и средние интернет – магазины. Метод группировки по семантической схожести применим только для крупных коммерческих проектов, реализация которых долговременна и дорогостояща.

Список литературы

1. О Стратегии развития информационного общества в Российской Федерации на 2017 – 2030 годы: указ Президента РФ от 09.05.2017 г. № 203. URL: <https://base.garant.ru/71670570> (дата обращения: 01.05.2022 г).
2. Ашманов, И. С. Оптимизация и продвижение в поисковых системах / И. С. Ашманов. – Санкт-Петербург: Питер, 2019. – 512 с.
3. Шамина, И. Total SEO часть 1. Полное практическое руководство по продвижению сайтов / И. Шамина, А. Носаченко. – Вологда: Инфра-Инженерия, 2021. – 604 с.
4. Анисимов, С. Поисковое продвижение сайтов / С. Анисимов. – РнД: Книжкин Дом, 2021. – 416 с.
5. Энж, Э. SEO: искусство раскрутки сайтов / Э. Энж, С. Спенсер, Д. Стрикчиола. – Спб.: БХВ-Петербург, 2017. – 816 с.
6. Давиденко, А. Кластеризация семантического ядра / А. Давиденко // kokoc.com: [сайт], 2021. – URL: <https://kokoc.com/blog/klasterizaciya-zaprosov> (дата обращения: 01.05.2022 г.).
7. Полушкин, Н. Кластеризация семантического ядра / Н. Полушкин // <https://www.webdius.ru> [сайт], 2021. – URL: https://www.webdius.ru/blog/kak_sdelat_klasterizaciyu_semanticheskogo_yadra (дата обращения: 01.05.2022 г.).