



Международный журнал информационных технологий и
энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 659.441.32

GOOGLE KNOWLEDGE GRAPH API КАК ИНСТРУМЕНТ SEO

¹ Агеев Н.А., ² Пастернак А.В.

ФГБОУ ВПО Смоленский государственный университет, Смоленск, Россия (21400,
г. Смоленск, ул. Пржевальского, 4), e-mail: ¹ nikita.ageev17@gmail.ru,

² artempasternak1@yandex.ru

В данной работе приводится описание поисковой оптимизации, отмечается значимость подбора ключевых слов, анализируется структура программного интерфейса приложения Google Knowledge Graph API. Целью нашего исследования является изучение отдельных особенностей алгоритмов выдачи похожих запросов в поисковой системе Google. Для выявления особенностей работы алгоритмов Google Knowledge Graph используются методы обратного инжиниринга. В частности, реконструирован орграф структуры, основанной на запросе «Уильям Шекспир» (глубина запроса – 3 уровня). Осуществлен формальный и семантический анализ построенного орграфа. Формальный анализ основан на ранжировании узлов графа по полустепеням исхода и захода. Семантический анализ проведен для верхнего квартиля ранжированного списка узлов. Актуальность данной темы обусловлена необходимостью поиска путей решения проблемы «низкого» расположения сайта в списке выдачи поисковой системы, что напрямую влияет на количество пользователей, посетивших этот сайт. Основным результатом проведенного исследования являются установленные некоторые закономерности выдачи похожих запросов на основе Google Knowledge Graph API.

Ключевые слова: поисковая оптимизация, генерация ключевых слов, граф знаний, программный интерфейс приложений

GOOGLE KNOWLEDGE GRAPH API AS AN SEO TOOL

¹ Ageev N.A., ² Pasternak A.V.

Smolensk State University, Smolensk, Russia(21400, Smolensk, Przhevalsky st., 4), e-mail:

¹ nikita.ageev17@gmail.ru, ² artempasternak1@yandex.ru

This article provides a description of search engine optimization, notes the importance of keyword selection, and analyzes the structure of the Google Knowledge Graph API application programming interface. The purpose of given article is to study certain features of the algorithms for issuing similar queries in the Google search engine. To identify the features of the Google Knowledge Graph algorithms, reverse engineering methods are used. In particular, the digraph of the structure based on the query "William Shakespeare" was reconstructed (the depth of the query is 3 levels). A formal and semantic analysis of the constructed directed graph has been carried out. Formal analysis is based on ranking the nodes of the graph according to the degrees of outgoing and entering. Semantic analysis was carried out for the top quartile of the ranked list of nodes. The relevance of this topic is due to the need to find ways to solve the problem of the "low" location of the site in the list of search engine results, which directly affects the number of users who visited this site. The main result of the study is the established certain patterns of issuing similar queries based on the Google Knowledge Graph API.

Keywords: search engine optimization, active word generation, knowledge graph, application programming interface.

Что такое SEO

Поисковая оптимизация SEO (Search Engine Optimization) — это процесс, который

улучшает количество и качество трафика, поступающего на веб-сайт через поисковые системы [2, 3, 4, 5]. Так же SEO можно рассматривать как процесс создания мета-описания веб-сайта, который приводит к эффективному ранжированию по выбранным ключевым словам в результатах обычного поиска.

Существует несколько способов использования SEO, и основные связаны с элементами на странице и отдельными элементами за ее пределами. С помощью SEO на странице измеряется все, что происходит внутри веб-сайта и что контролируется владельцем веб-сайта, например, мета-описание, заголовок, контент и изображения.

Согласно выводам М. М. Хейтала и О. Маршалла [2], чтобы получить более высокий рейтинг и получить значимый трафик от поисковых систем, веб-сайт должен быть удобным для пользователя, предоставлять уникальный контент и иметь хорошо сформулированное мета-описание.

Методы SEO вне страницы в основном связаны с внешними аспектами, которые имеют существенное влияние на рейтинг сайта в поисковых системах. Наиболее эффективным фактором является использование ссылок, которые направляют пользователя на страницу компании.

В нашем исследовании мы остановимся на одном из основных элементов поисковой оптимизации – генерации ключевых слов.

Ключевые слова

Ключевые слова — это слова, которые пользователи вводят в поле поиска в поисковых системах, и эти слова систематизируются Google при сканировании и индексировании веб-страниц.

Подбор ключевых слов является одним из наиболее важных факторов, влияющих на SEO [3]. Ключевые слова являются основным фактором, который «объединяет» поисковую систему и веб-сайт, а это означает, что алгоритмы поисковой системы поднимают страницы с правильными ключевыми словами в списке результатов поиска.

Важность высокого места в результатах поиска очевидна, поскольку пользователи с большей вероятностью нажимают на результаты на первой странице.

Значимость правильного подбора ключевых слов так же очевидна. Однако генерация полного набора ключевых слов «вручную», без использования вспомогательных инструментов, затруднена. Например, к сайту кафе может привести запрос «Чем заняться в вечер пятницы», или ещё более экзотические запросы. Подбор всей совокупности ключевых слов, которые могут привести потенциальных пользователей к сайту, требует анализа огромного массива информации, отражающего: историю реальных запросов, семантику, психологию и прочие области знаний. Совершенно очевидно, что крупная компания, специализирующаяся в определенной области, например, автомобилестроение, не говоря уже о кафе, не в состоянии провести эту работу в полном объёме.

Google Knowledge Graph

Естественно ожидать появления соответствующих сервисов от мировых монополистов в сфере поиска информации и BigData. В частности, в 2012 году компания Google анонсировала открытый информационный ресурс в данной области «Google Knowledge Graph» [5, 6].

Граф знаний Google содержит обобщенную информацию в областях знаний, чтобы дать краткий обзор конкретной темы и дать дополнительные рекомендации по связанным областям.

В наше время графы знаний получили широкое распространение в бизнес-приложениях, и компании стремятся к практичности и извлекают выгоду из их использования.

Х. Цзоу в своём исследовании подробно анализирует 4 основные области использования графов знаний [6]:

- Системы ответов на вопросы за счет повышения семантической осведомленности о социальных чат-ботах и цифровых помощниках (Siri, Alexa и т. д.).
- Системы поддержки точности, разнообразия и принятия решений.
- Усовершенствованные поисковые системы.
- Специфическое использование предметной области, например, в сфере медицины, кибербезопасности, новостей и образования.

Application Programming Interface

Практическое использование Google Knowledge Graph существенно облегчается интегрированным в него API. API (Application Programming Interface) — «программный интерфейс приложения» — описание способов (набор классов, процедур, функций, структур или констант), которыми одна компьютерная программа может взаимодействовать с другой программой.

Описание кода

Программный интерфейс приложения от Google предоставляет пример кода запроса по Taylor Swift.

```
{
  "@context": {
    "@vocab": "http://schema.org/",
    "goog": "http://schema.googleapis.com/",
    "resultScore": "goog:resultScore",
    "detailedDescription": "goog:detailedDescription",
    "EntitySearchResult": "goog:EntitySearchResult",
    "kg": "http://g.co/kg"
  },
}
```



Рисунок 1 – Первый фрагмент кода

1. В строке «vocab» содержится краткая ссылка на словарь семантической разметки Schema.org. На второй строке данного раздела содержится полная ссылка на данный сайт.
2. Раздел 2 начинается с оценки результата поиска и его подробного описания. Далее отображается результат поиска объектов.
3. Строка «kg» выводит связанные запросы.

```
'@type': "ItemList",
'itemListElement': [
  {
    "@type": "EntitySearchResult",
    "result": {
      "@id": "kg:/m/0dl567",
      "name": "Taylor Swift",
      "@type": [
        "Thing",
        "Person"
      ]
    }
  }
],
```



Рисунок 2 – Второй фрагмент кода

1. В первом разделе определяется тип запроса.
2. Далее определяется id и имя запроса.
3. Раздел 3 конкретизирует сущность объекта.

```
"description": "Singer-songwriter",
"image": {
  "contentUrl": "https://t1.gstatic.com/images?q=tbn:ANd9GcQmVDAhjhWnN20Wys2ZM03PGAhupp5tN2LwF_BJmiHqi19hf8Ku"
},
"url": "https://en.wikipedia.org/wiki/Taylor_Swift",
"license": "http://creativecommons.org/licenses/by-sa/2.0"
```



Рисунок 3 – Третий фрагмент кода

1. В 1 разделе продолжается детализация описания объекта.
2. В строке 2 отображается ссылка на изображение по запросу.
3. Далее следует ссылка на википедию и лицензию.

```
"detailedDescription": {
  "articleBody": "Taylor Alison Swift is an American singer-songwriter and actress. Raised in Wyomissing, Pe",
  "url": "http://en.wikipedia.org/wiki/Taylor_Swift",
  "license": "https://en.wikipedia.org/wiki/Wikipedia:Text_of_Creative_Commons_Attribution-ShareAlike_3.0_Un",
},
"url": "http://taylorswift.com/"
},
"resultScore": 4850
```

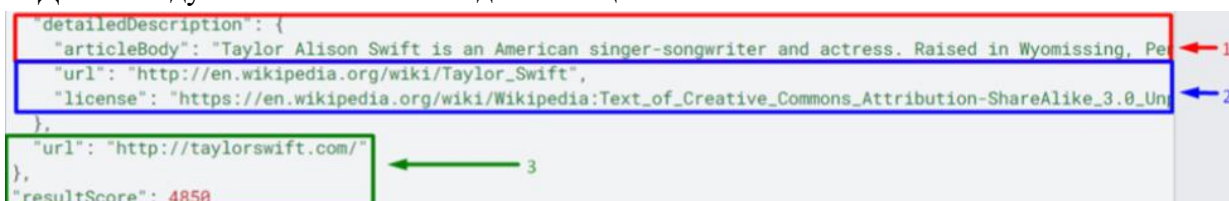


Рисунок 4 – Четвёртый фрагмент кода

1. В первой строке указывается подробное описание объекта запроса.
2. Далее указывается ссылка на википедию и её лицензию.
3. В завершении отображается ссылка на сайт объекта и выдаётся оценка результата.

Обратный инжиниринг

Многие исследователи, занимающиеся изучением практических приложений, жаловались на закрытость, недоступность алгоритмов, которые используются в Google Knowledge Graph [2]. Как отмечают С.-Дж. Лу, С.-А. Янг и Т.-Л. Хуанг: «до сих пор ни одной компании не удалось полностью раскрыть детали ранжирования поисковых систем» [5]. В силу этого единственным возможным

вариантом анализа особенностей функционирования Google Knowledge Graph API является обратный инжиниринг.

Обратный инжиниринг — исследование некоторого готового устройства или программы, а также документации на него с целью понять принцип его работы [1].

В рамках обратного инжиниринга мы построили графовую модель, основанную на запросе «Уильям Шекспир», представляющую собой древовидную структуру глубиной 3 уровня. При этом использовались традиционные подходы к моделированию семантических сетей: узлы орграфа соответствовали объектам, дуги соответствовали бинарному отношению на множестве объектов «быть в списке похожих запросов для данного запроса». Результатом моделирования стал орграф с 45 вершинами и 240 дугами, диаграмма которого представлено на Рисунке 5.

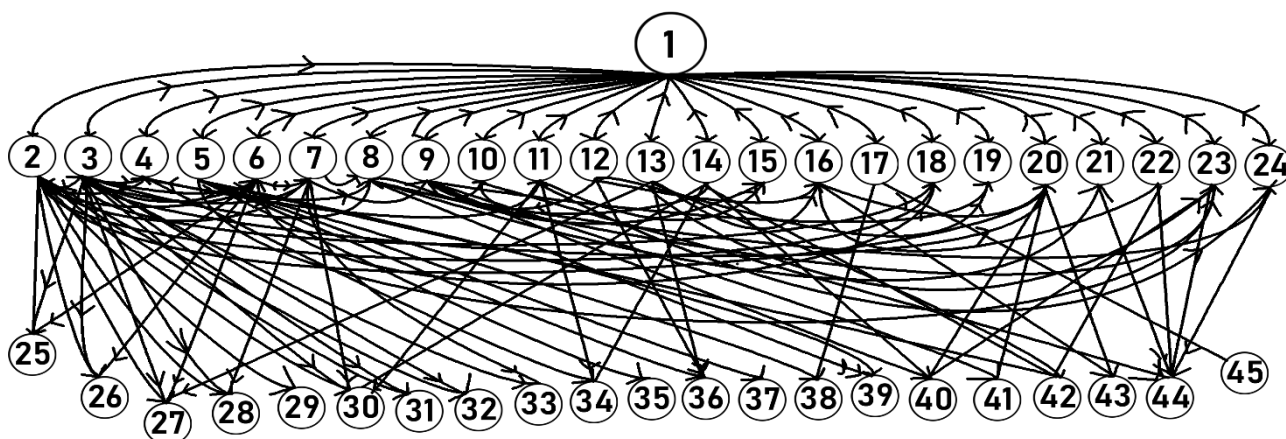


Рисунок 5 – Древовидный граф по ключевому слову "Уильям Шекспир"

Формальный анализ построенного графа основывался на вычислении полустепеней исхода и захода всех узлов и последующим ранжировании узлов по вычисленным полустепеням.

Для содержательного (семантического) анализа были отобраны узлы верхнего квартиля каждого из ранжированных списков (по полустепени исхода и по полустепени захода по отдельности).

Информация о двух верхних квартилях представлена в Таблице 1.

Таблица 1 – Верхний квартиль полустепени захода и исхода.

Полустепени захода			Полустепени исхода		
N	p+	Имя	N	p-	Имя
30	13	Эдмунд Шекспир	2	18	Энн Хатауэй
9	10	Елизавета I	3	18	Хемнет Шекспир
25	10	Элизабет Бернард	7	17	Джудит Шекспир
28	9	Ричард Кини	6	15	Сьюзен Холл
3	8	Хемнет Шекспир	8	15	Чарльз Диккенс
8	8	Чарльз Диккенс	9	14	Елизавета I
14	8	Джефри Чосер	14	14	Джефри Чосер
26	8	Джон Холл	10	13	Уильям Вордсворт
20	8	Джоан Шекспир	13	13	Мольер
2	7	Энн Хатауэй	11	11	Эдгар Аллан По

Результаты семантического анализа

1. Энн Хэтуэй, Хемет Шекспир, Чарльз Диккенс, Елизавета I и Джефри Чосер входят в оба топа. Мы видим, что доминирует культурная идентичность, показывает тех, кто относится к Великобритании. Исторический и личностный контекст преобладают над художественным (творческим).

2. Мы видим, что в верхнем квартиле списка, ранжированного по полустепени захода преобладают родственники У. Шекспира, так как многие родственники были связаны друг с другом. Родственники образуют клику (плотно связанное сообщество). 7 из 10 это родственники, 2 из 10 писатели и последний - это правитель Елизавета I. Она находится высоко в списке за счет многочисленных родственников У. Шекспира, живших в период её правления.

3. В верхнем квартиле списка, ранжированного по полустепени исхода количество родственников в 2 раза меньше, чем в верхнем квартиле списка, ранжированного по полустепени захода. При этом количество писателей увеличилось больше, чем в 2 раза. Мы видим, что происходит размывание чисто британского контекста, появляется Ж.-Б. Мольер (Франция), Эдгар Аллан По (США).

Вывод:

Проведённое нами исследование показало значимость поисковой оптимизации значимость подбора ключевых слов для поисковой оптимизации и эффективность Google Knowledge Graph API для отбора ключевых слов. Проведённое графовое моделирование и формальный и семантический анализ позволили установить, что при выдачи похожих запросов в Google преобладают исторический и личностный контекст над профессиональным и общекультурным. Проведённое нами исследование показало, что метод обратного инжиниринга эффективен при решении проблемы выявления особенностей функционирования исследуемого алгоритма, однако данный метод является весьма трудоёмким, что ограничивает область его применения.

Список литературы

1. Baxter, I. D. and Mehlich, M. (2000). Reverse engineering is reverse forward engineering. Science of Computer Programming, Vol. 36, Issues 2–3, March 2000: 131-147. <https://www.sciencedirect.com/science/article/pii/S0167642399000349>
2. Nietala M. M. and Marshall O. (2021). Search Engine Optimization and the connection with

- Knowledge Graphs. University of Gävle, Faculty of Education and Business Studies, Department of Business and Economic Studies, Business administration: 1-75. <https://www.diva-portal.org/smash/get/diva2:1524391/FULLTEXT01.pdf>
3. Iskandar, M. S. and Komara, D. (2018). Application Marketing Strategy Search Engine Optimization (SEO). IOP Conference Series: Materials Science and Engineering, Vol. 407. <https://iopscience.iop.org/article/10.1088/1757-899X/407/1/012011>
 4. Kritzinger, W. (2013). Search engine optimization and pay-per-click marketing strategies. / Journal of Organizational Computing and Electronic Commerce, Vol. 23: 273–286. http://digitalknowledge.cput.ac.za/bitstream/11189/2849/3/Kritzinger_WT_Weideman_Melius_FID_2013.pdf
 5. Luh, C.-J., Yang, S.-A. and Huang, T.-L. D. (2016). "Estimating Google's search engine ranking function from a search engine optimization perspective", Online Information Review, Vol. 40, No. 2: 239-255. <https://doi.org/10.1108/OIR-04-2015-0112>
 6. Zou, X., (2020). A Survey on Application of Knowledge Graph. *Journal of Physics: Conference Series*, Vol. 1487: 012016. <https://iopscience.iop.org/article/10.1088/1742-6596/1487/1/012016>

References

1. Baxter, I. D. and Mehlich, M. (2000). Reverse engineering is reverse forward engineering. Science of Computer Programming, Vol. 36, Issues 2–3, March 2000: 131-147. <https://www.sciencedirect.com/science/article/pii/S0167642399000349>
 2. Hietala M. M. and Marshall O. (2021). Search Engine Optimization and the connection with Knowledge Graphs. University of Gävle, Faculty of Education and Business Studies, Department of Business and Economic Studies, Business administration: 1-75. <https://www.diva-portal.org/smash/get/diva2:1524391/FULLTEXT01.pdf>
 3. Iskandar, M. S. and Komara, D. (2018). Application Marketing Strategy Search Engine Optimization (SEO). IOP Conference Series: Materials Science and Engineering, Vol. 407. <https://iopscience.iop.org/article/10.1088/1757-899X/407/1/012011>
 4. Kritzinger, W. (2013). Search engine optimization and pay-per-click marketing strategies. / Journal of Organizational Computing and Electronic Commerce, Vol. 23: 273–286. http://digitalknowledge.cput.ac.za/bitstream/11189/2849/3/Kritzinger_WT_Weideman_Melius_FID_2013.pdf
 5. Luh, C.-J., Yang, S.-A. and Huang, T.-L. D. (2016). "Estimating Google's search engine ranking function from a search engine optimization perspective", Online Information Review, Vol. 40, No. 2: 239-255. <https://doi.org/10.1108/OIR-04-2015-0112>
 6. Zou, X., (2020). A Survey on Application of Knowledge Graph. *Journal of Physics: Conference Series*, Vol. 1487: 012016. <https://iopscience.iop.org/article/10.1088/1742-6596/1487/1/012016>
-