

УДК 004

JEL коды: D89

08.00.13

Нечеткие множества и логики, их применение в SEO и веб-поиске Fuzzy sets and logic, their application in SEO and web search

Воробьев Владимир Игнатьевич,

к.э.н., Санкт-Петербургский университет технологий управления и экономики, Россия.

Vorobev Vladimir Ignatievich,

Ph.D., St. Petersburg University of Management Technologies and Economics, Russia.

Аннотация

В статье рассматривается формализация, модельное решение задачи релевантного SEO-поиска с учетом потребительских свойств искомого поискового объекта, его оценок экспертами (форумами). Приведены также практические правила поиска.

Abstract

The article deals with the formalization, model solution of the problem of a relevant SEO search, taking into account the consumer properties of the desired search object, its ratings by experts (forums). Practical search rules are also provided.

Ключевые слова: поисковая, оптимизация, SEO, модель, релевантность, тематичность, региональность, запрос, нечеткое, множество.

Keywords: search, optimization, SEO, model, relevance, topicality, regionality, query, fuzzy, set.

Постановка задачи, ее формализация

Для различных сложных веб-запросов используются различные показатели: релевантность, тематичность, региональность и др. Информация о количественном выражении (мере) этих показателей часто нечеткая. Информационный запрос, анализа такой информации требуется при разработке предметно-ориентированных интеллектуальных систем, экспертных систем (баз знаний).

Пусть $R = \{r_m : m = 1, 2, \dots, M\}$ – множество атрибутов, показателей запросов. Для m -го показателя есть ориентировочная количественная информация, значение, которым ставят в соответствие термы нечеткой логики. Значением лингвистической переменной могут стать, например, показатели «стоимость невысокая», «предпочтительный вариант», «узкого применения» и др.

Для определения значений лингвистической переменной необходимо идентифицировать, задать четкие значения термов лингвистической переменной, функцию принадлежности, по которой определяется мера «принадлежности» точных значений термам. В качестве стандартных форм функций принадлежности берут « Λ -функции», « Π -функции», « Z -функции», « S -функции».

Переход от нечетких к определенным значениям осуществляется дефазификацией, например, методом центра максимума, суть его, на основе Λ -функции, – в нижеследующем.

Значения m -го переменной (показателя) u описаны Λ -функцией, за центр максимума используем u_{mj}^* . Средневзвешенное u_m определим как:

$$y_m = \frac{\sum_{j=1}^{n_m} y_{mj}^*}{n_m}.$$

В связи с разнообразием ключевых слов (мощностью семантического ядра) с одинаковой базой (например, «купить», «недорого», «Москва»), необходимо выбрать такой элемент, который отвечал бы множеству требований R посетителя.

Для заданных термов поиска $K = \{k_l: l = 1, 2, \dots, L\}$ сформируем, например, среднечастотное подмножество $T^d \subset T$, где T – множество всех термов поиска. Сформировать T^d можно с помощью семантического ядра.

После этого, в T^d выбирается терм, удовлетворяющий потребителю требованиям (показателям) $r \subset R$. Для заданного запроса с потребительскими параметрами K , показателями $r \subset R$ отыскивается такой терм, для которого $t_o = \arg \min_{t \in T^d} F(t)$ при условии: $y_m(t) \neq 0$, $m \in R^z$, $t \in T^d$, $y_m(t)$ – количественная мера m -го показателя для t -го терма.

Получена задача многокритериальная с критерием оптимальности $F(t)$, в которой необходимо найти $t \in T^d$, не являющийся оптимальным для $F_i(t)$, но приемлемым для всего запроса, интегрального критерия $F(t)$. Задача решается известными методами.

Но остается нерешенной задача формирования семантического ядра и правил SEO-поиска. Формальных решений быть пока не может, но рассмотрим практические, 2018 года, советы.

Правила SEO-2018 и онлайн-поиск

Основы SEO-2017, разумеется, стабильны. Но добрая сотня (сотни) правил алгоритмов индексации, ранжирования, реагирования Google динамично улучшаются (по сравнению с первоначальным PageRank, например), в смысле, интеллектуализируются, интегрируются.

Большинство старающихся получить SEO-результат в 2018 году, только начинают, веб-сайт только индексировали, начали исследовать конкурентные ключевые слова. PageRank просматривал веб-страницы, от исходной ссылки до финишной, исследуя виртуальные сети. Чем больше ссылок перейдет на веб-страницу, тем выше релевантность, индекс, приписываемый этой странице, выше ранг.

Все изменилось, речь не только о ссылках теперь. Появились «три кита» SEO:

1. полномочия, выше авторитет – больше можно ссылок передавать (проверить «накопленные» полномочия можно на спецдоменах, специнструментарием, например, MozBar);
2. содержание, контент должен быть «просто супер», ценным, уникальным, грамотным, релевантным, интригующим иногда (контентное продвижение – самое длительное, самое надежное, самое окупаемое);
3. возраст индексации, «послужной список» с качеством ссылок (они взаимосвязаны).

Основные правила-доминанты SEO-2018:

1. идентифицировать скорость (следовательно, и потенциальную видимость) страницы, улучшить ее используйте данные Google (есть специнструментарий, например, Pingdom), добиваясь сокращения реакции сервера на запрос (пользователь пошел деловой, ждать не будет), устраняя блокировку

визуализации (CSS, JavaScript, например), кеширования браузера, оптимизируя мультимедиа (Comhressor, GZip, например);

2. использовать сети доставки контента (CDN домена), DNS, «зеркала», «облака», где возможно, быстро «обращая» контент по миру;
3. создать полезный контент, увеличивающий ценность, уменьшающий зависимость от SEO («человеческий», не для ботов), увеличивающий трафик (основные признаки такого контента – менее 2000 слов, внутренние ссылки, перелинковка в разумными внешними, факты, статистика, таблицы, «важности», выводы, структурированность, качество минимизированной графики, релевантность ключевых слов);
4. обеспечить мобильность, пользовательскую комфортность, мультиплатформенность (мобильные запросы – более 67% запросов), отличный интерфейс, простой, интуитивно внятный, быстрый, удобный, с поиском по сайту, каталогом, категориями, с минимумом «всплывший» окон;
5. дополнить контент аудио-видео, подкастами (только после этого длинным, длительным контентом);
6. использовать технологию LSI от Google, для семантического индексирования (латентного), разнесения ключевых слов (принцип: «релевантен тот контент, который ищет пользователь, а не тот, который возвращается по ключевому слову);
7. создать исходящие ссылки, внутреннюю перелинковку (2-3 раза) контента, обеспечивая переход на другие страницы, задержку пользователя на сайте;
8. обеспечить качество, разумное количество ссылок и их естественное вхождение;
9. продавать контент, используя доверенные домены, независимо от SEO;
10. быть осторожным «рекламистом», Google много рекламы не нравится определено, как и длительная ее загрузка с платформы;
11. используйте лидогенерацию, четкие призывы к действию;
12. не пытаться обманывать и не обманывайтесь, «черно-серое» продвижение – с кратковременным эффектом, сосредоточьтесь лучше на добавленной ценности, стратегии;
13. использовать соцсети (вирусный маркетинг, контент), управление трафиком, но правильно, т.е. ориентированное на ценность трафика, а не на ценность бизнес-процесса, продукта;
14. привлекать (заинтересовывать) трастовые ресурсы, влиятельных лиц в содействии (например, ссылки).

Для стартапа, новичков SEO можно акцентировать более короткий, простой, общий список проблем:

1. улучшать ROI, «реальную» рентабельность стратегии продвижения, конверсии, помня, что SEO – стратегия, надолго, а ROI должно «отозваться» через месяц-два, это тактика;
2. временные инвестиции, даже простая пара структурных изменений на сайте (его видимости) дадут основу для длительного развития;
3. технический SEO;

4. ошибок, штрафов критически бояться не стоит, хотя они могут ухудшить рейтинг, все исправляемо обычно следует бояться преднамеренного, неоднократного манипулирования ранговыми показателями;
5. таргетинг, выбор SEO-слов, релевантного семантического ядра;
6. сложность SEO уменьшается временем, инвестициями, мотивацией, релевантной стратегией.

Узнаем о SEO больше, попробуем его практически, анализируем результаты, эволюционируем.

Рекламный сервис Google.Adwords позволяет просмотреть объявления будущим работодателем, имеет ниже перечисляемые преимущества.

1. Таргетинг. Если поисковая оптимизация (семантическое ядро запросов, СЯ) отвечает требованиям региональной, геовременной ориентированности запроса, то предпочитают Google. Полнее используют частотные запросы СЯ, скомпилированные для SEO. Рекламодатель устанавливает самостоятельно бюджет, минимум ставок (ежедневно). Автоматически регулируется рекламная площадка, необходимые объявления в результатах.

2. Управление. Рекламная компания Google.Adwords дает возможность контроля рекламной кампании, групп объявлений, например, тексты могут иметь общий язык, еженедельные бюджеты, распределения и даты завершения, группа – несколько обнародований.

3. Редактирование. В Google.AdWords размещаемы баннеры различного формата (468x60; 728x90; 300x250; 120x600; 160x600). Анимированные картинки представимы лишь в доступных графических форматах.

4. Анализ. Аналитика контекстной рекламы зависит от настроек кампании, цены. Может учитывать денежные компенсации контекстной рекламы, количество посетителей, стоимость клика, частотность и др.

Для Google.AdWords результат - контекстная реклама, Google-SEO, возможность редактирования формы текста объявления, уместного во всех рекламируемых компаниях.

Например, для ценового диапазона товаров (услуг), можно настроить самые важные слова в соответствующей рекламе.

Заключение

Предложенная формальная модель решения поисковой SEO-задачи и практические правила позволят эффективно решать поисковые задачи, повысить релевантность с учетом потребительских (регионально-тематических) требований к искомому образцу.

Литература

1. Адамович И.М., Заикин М.Ю., Заусаев О. В., Пешков А. Н. Технологии работы в WEB-пространстве Интернет // Электронный журнал «Исследовано в России». <http://zhurnal.apecelarn.ru/articles/2000/044.pdf>
2. Ашманов И.С., Иванов А.А.. Оптимизация и продвижение сайтов в поисковых системах. - СПб.: Питер, 2008.
3. Голик В.С. Эффективность Интернет-маркетинга в бизнесе // Маркетинг в России и за рубежом. - 2007. - № 1. - С. 124- 133.
4. Гридина Е.Г., Иванников А.Д., Булгаков М.В., Чиннова И.И., Сигалов А.В.. Система федеральных образовательных порталов: 3 года в Интернете // Открытое Образование. - 2005. - № 1.
5. Колисниченко Д.Н.. Поисковые системы и продвижение сайтов в Интернете. - М.: Диалектика-Вильямс, 2007.

6. Левшин Д.В., Марков А. С. Алгоритмы интеграции СУБД PostgreSQL с семантическим веб. <http://sp.cmc.msu.ru/datalog/levshin-swps.pdf>
7. Макарец А.Б. Анализ образовательных услуг с позиций маркетингового управления // Вестник Саровского физтеха. - 2008 - №13-14. - С. 128- 135.
8. Макарец А.Б., Беляева Г. Д. Интернет-сайт вуза - эффективный инструмент маркетинговых коммуникаций // Сб. ст. Всеросс. научн.-практич. конф. «Маркетинг: современные реалии и вызовы времени» / Мордовский университет. - Саранск, 2008. - С. 86-93.
9. Неретина Е.А., Макарец А.Б. Рыночная ориентация процессов формирования и продвижения образовательных услуг высшего учебного заведения // Интеграция образования. - 2008. - № 2. - С. 18-24.
10. Образовательные ресурсы в Интернет. Тестирование сайтов. Сайты российских вузов. // Сотри-терВид. - 2006. - № 12. - С. 45- 46.
11. Покровский Н. Usability-методы исследования Web-сайта [Электронный ресурс] // Сайт Usability.ru. - Режим доступа: <http://www.usability.ru/Articles/um.htm>.
12. Рыжиков С.В. Способ учета рекламных компаний для интернет-проектов и событийный анализ сайта // Интернет-маркетинг. - 2006. - № 1. - С. 45-53.
13. Стивенс У.Р. Протоколы TCP/IP. Практическое руководство, Монография. Перевод с английского и комментарии А. Ю. Глебовского. — Санкт-Петербург: БХВ - Петербург: Невский диалект, 2003. — 671 с.: ил. — (В подлиннике). — ISBN 5-7940-0093-7. — ISBN 5-94157-300-6
14. Bemers-Lee Tim, Chen Yuhsin, Chilton Lydia, Connolly Dan, Dhanaraj Ruth, Hollenbach James, Lerer Adam, Tabulator David. Exploring and Analyzing linked data on the Semantic Web SWUI 2006 The 3rd International Semantic Web User Interaction Workshop November 6, 2006 — Athens, Georgia, USA collocated with ISWC 2006 <http://swui.semanticweb.org/swui06/>
15. Cho J., Garcia-Molina H. The evolution of the web and implications for an incremental crawler. 26th International Conference on Very Large Databases 2000. Cairo, Egypt.
16. Kobayashi Mei, Takeda Koichi. Information retrieval on the web ACM // Computing Surveys. Vol. 32. Issue 2 (June 2000). <http://portal.acm.org/citation.cfm?id=358934>
17. Risvik K.M., Michelsen R. Search Engines and Web Dynamics// Computer Networks. Vol. 39. № 3. 21 June. 2002. P. 289-302. <http://www.ingentaconnect.com/content/els/13891286/2002/00000039/00000003/art00213>

References

1. Adamovich I.M., Zaikin M.Yu., Zausayev O.V., Peshkov A.N. Technologies of work in the WEB-space of the Internet // Electronic journal "Researched in Russia". <http://zhurnal.ape.relarn.ru/articles/2000/044.pdf>
2. Ashmanov I.S., Ivanov A.A. Optimization and promotion of sites in search engines. - Spb.: Peter, 2008.
3. Golik V.S. The effectiveness of Internet marketing in business // Marketing in Russia and abroad. - 2007. - № 1. - p. 124- 133.
4. Gridina E.G., Ivannikov A.D., Bulgakov M.V., Chinnova I.I., Sigalov A.V. System of Federal Educational Portals: 3 years on the Internet // Open Education. - 2005. - № 1.
5. Kolisnichenko D.N. Search engines and website promotion on the Internet. - M.: Dialectics-Williams, 2007.
6. Levshin D.V., Markov A.S. Algorithms for integrating PostgreSQL DBMS with the semantic web. <http://sp.cmc.msu.ru/datalog/levshin-swps.pdf>
7. Makarets A.B. Analysis of educational services from the point of view of marketing management // Bulletin of the Sarov Physics and Technology Institute. - 2008 - №13-14. - pp. 128-135.
8. Makarets, A.B., Belyaev, G. D. The website of the university is an effective tool for marketing communications. st. Vseross scientific-practical conf. "Marketing: modern realities and challenges of the time" / Mordovia University. - Saransk, 2008. - p. 86-93.
9. Neretina, E.A., Makarets, A.B. Market orientation of the processes of formation and promotion of educational services of higher education // Integration of education. - 2008. - № 2. - p. 18-24.
10. Educational resources on the Internet. Testing sites. Sites of Russian universities. // wipe-terBId. - 2006. - № 12. - p. 45- 46.
11. Pokrovsky N. Usability-methods for researching a Web site [Electronic resource] // Website Usability.ru. - Access mode: <http://www.usability.ru/Articles/um.htm>.
12. Ryzhikov S.V. The method of accounting advertising companies for Internet projects and event analysis of the site // Internet marketing. - 2006. - № 1. - p. 45-53.

13. Stephens U.R. TCP / IP protocols. Practical Guide, Monograph. Translation from English and comments by A. Yu. Glebovsky. - St. Petersburg: BHV - Petersburg: Nevsky Dialect, 2003. - 671 pp., Ill. - (In the original). - ISBN 5-7940-0093-7. - ISBN 5-94157-300-6
14. Bemers-Lee Tim, Chen Yuhsin, Chilton Lydia, Connolly Dan, Dhanaraj Ruth, Hollenbach James, Lerer Adam, Tabulator David. SWUI 2006 Exploring and Analyzing Web Site November 6, 2006 - Athens, Georgia, USA collocated with ISWC 2006 <http://swui.semanticweb.org/swui06/>
15. Cho J., Garcia-Molina H. The evolution of the web and implications for an incremental crawler. 26th International Conference on Very Large Databases 2000. Cairo, Egypt.
16. Kobayashi Mei, Takeda Koichi. Information retrieval on the ACM web // Computing Surveys. Vol. 32. Issue 2 (June 2000). <http://portal.acm.org/citation.cfm?id=358934>
17. Risvik K.M., Michelsen R. Search Engines and Web Dynamics // Computer Networks. Vol. 39. № 3. 21 June. 2002. P. 289-302. <http://www.ingentaconnect.com/content/els/13891286/2002/00000039/00000003/art00213>