

Оценка влияния социальных доменов в результатах поисковой выдачи на стоимость контекстной рекламы

Кузнецов П.А., к.э.н, доцент кафедры экономики и финансов, Пермский национальный исследовательский политехнический университет, г. Пермь, Россия

Плотников А.В., к.э.н, доцент кафедры менеджмента и маркетинга, Пермский национальный исследовательский политехнический университет, г. Пермь, Россия

Бестужев А.А., студент, Пермский национальный исследовательский политехнический университет, г. Пермь, Россия

Аннотация. В работе рассматриваются особенности результатов поисковой выдачи и предлагается новый подход к анализу данных. Выдвигаемая гипотеза заключается в том, что наличие не тематических некоммерческих социальных сетей или социальных доменов в поисковой выдаче говорит о том, что "ключевой запрос" – информационный, а следовательно, обладает пониженной стоимостью клика в рекламном блоке поисковой выдачи в сравнении с коммерческими запросами. Для проверки полученных результатов был выполнен регрессионный анализ средствами Excel.

Ключевые слова: социальная сеть, цифровая экономика, контекстная реклама, поисковая оптимизация, поисковый маркетинг, интернет-маркетинг.

Impact Estimating of Social Domains in the Search Engine Results Pages on the Value of Contextual Advertising

Kuznetsov P.A., candidate of Economic Sciences, Associate Professor, Economics and Finance Department, Perm National Research Polytechnic University, Perm, Russia

Plotnikov A.V., candidate of Economic Sciences, Associate Professor, Management and Marketing Department, Perm National Research Polytechnic University, Perm, Russia

Bestuzhev A.A., student, Perm National Research Polytechnic University, Perm, Russia

Annotation. The paper considers the features of search engine results pages and proposes a new approach to queries data analysis. The proposed hypothesis is that the presence of non-thematic non-profit social networks or social domains in search results indicates that the «queries» is informational. Information queries have a lower cost per click in contextual advertising search engine results pages in comparison with commercial queries. To verify the results obtained, a regression analysis was performed using Excel tools.

Keywords: social network, digital economy, contextual advertising, search engine optimization, search marketing, internet marketing.

В последние годы растущая доступность огромного количества информации, создаваемой в каждом секторе на высокой скорости и в самых разных формах и форматах, является беспрецедентной. Возможность использовать большие данные – это возможность получить более точный анализ и улучшить процесс принятия решений в хозяйственной деятельности. Однако обработка больших данных может быть сложной задачей, и правильная интеграция данных является ключевым аспектом в достижении высокого качества информации. В этой статье мы предлагаем новый подход к анализу данных в результатах поисковой выдачи, полученных в онлайн. Методы семантического поиска используются поисковыми системами на многих этапах потребительского поиска. Было много дискуссий [1], [2], [3], [4] о том, как использовать семантический поиск для лучшего понимания намерений пользователя, или для преобразования запроса, или для извлечения релевантной информации со страницы.

SERP (Search engine results page) – это страница результатов поиска, сгенерированная поисковой системой (Google, Яндекс) [5]. SERP состоит из ссылок на текстовые документы, фото- и видео- материалы, иллюстрации и т.д.,

соответствующие запросу, с кратким их описанием. SERP может включать в себя социальные домены (stack-overflow, youtube, wikipedia, vk, facebook, twitter, slideshare, tumblr, linkedin, vimeo и др.). Помимо собственно ответа на поисковый запрос, SERP содержит и другую в той или иной мере полезную информацию (в том числе рекламную), а также предоставляет некоторые возможности для уточнения поиска.

Целью настоящего исследования является определение влияния наличия в поисковой выдаче социальных доменов (СД) на стоимость клика. Поскольку выдвигаемая гипотеза заключается в том, что наличие не тематических некоммерческих социальных сетей или социальных доменов (youtube, facebook, vk и др.) в поисковой выдаче говорит о том, что "ключевой запрос" – информационный, а следовательно, обладает пониженной стоимостью клика в рекламном блоке SERP в сравнении с коммерческими запросами.

В качестве выборки выступили поисковые запросы по маске 'javascript' в московском регионе за январь 2019 г. Исходный файл был преобразован в формат для регрессионного анализа. Колонка 'Social domains' была разделена на отдельные элементы (колонки Soc0 - Soc5), добавлена колонка 'Soc_qnt' кодирующая количество таких доменов. Следующим шагом колонки Soc0 - Soc5 были преобразованы из категориального типа в бинарный через процедуру 'one-hot encoding'.

Таблица 1

Средняя стоимость клика в рекламе в зависимости от количества социальных доменов в SERP

Названия строк	Среднее по столбцу cost_ads_msk, долл. США	Число элементов в столбце Soc_qnt
0	0,09	5 805,00
1	0,31	2 055,00
2	0,28	1 767,00
3	0,86	336,00
4	0,24	37,00
5	0,97	4,00
6	0,01	2,00

Общий итог	0,19	10 006,00
------------	------	-----------

Результаты содержат статистические данные по присутствию СД в представленной выборке. Записей, в которых присутствовал хотя бы один СД оказалось порядка 42% и средняя стоимость клика (ССК) составила 0,45 USD. Записей без СД – 58%, где ССК = 0,09 USD (рис.1). Из этого факта можно сделать уверенный вывод о пятикратном увеличении ССК при появлении СД в выдаче.

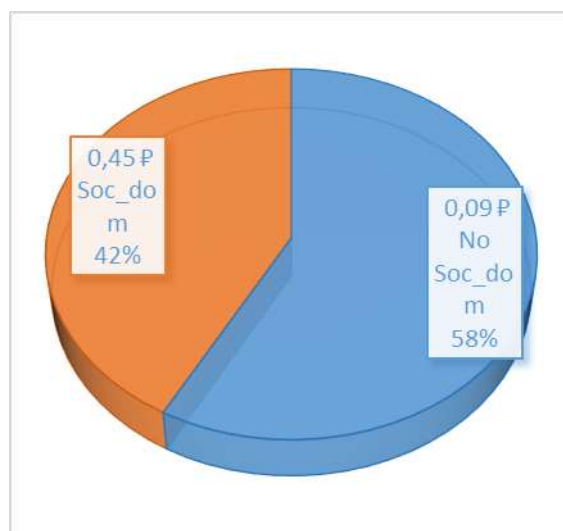


Рис. 1 – Соотношение SERP с социальными доменами и без

В таблице 1 приведена информация о ССК в зависимости от числа СД и распределении количества записей в выборке по количеству СД в записи. Наблюдаемые пики ССК на 3 и 5 СД в записи (зеленый график, рис.2) статистически не особенно значимы, т.к. наблюдаются при незначительном количестве таких записей (синий график).

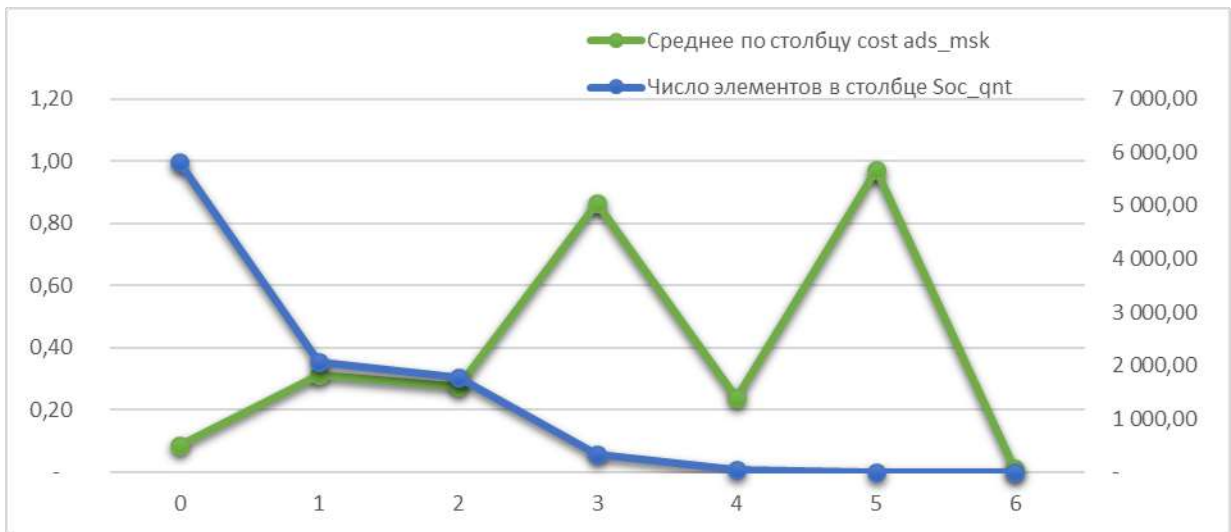


Рис. 2 – Средняя стоимость клика в рекламе в зависимости от количества социальных доменов в SERP

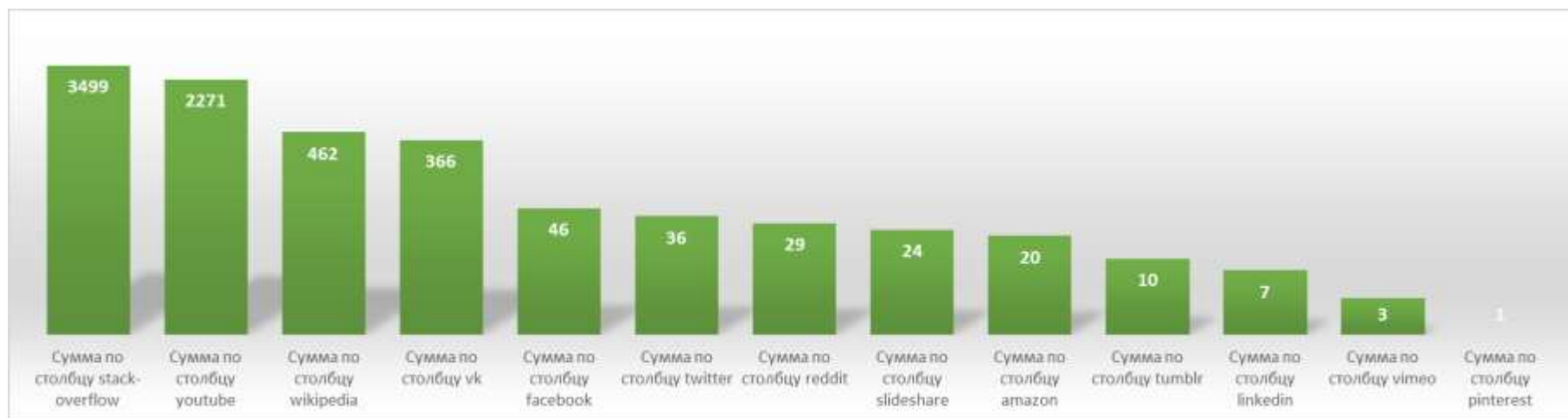


Рис. 3 – Количество записей в выборке по каждому СД

Таблица 2

Количество записей в выборке по каждому СД

Значения СД	Количество
Сумма по столбцу stack-overflow	3499
Сумма по столбцу youtube	2271
Сумма по столбцу wikipedia	462
Сумма по столбцу vk	366
Сумма по столбцу facebook	46
Сумма по столбцу twitter	36
Сумма по столбцу reddit	29
Сумма по столбцу slideshare	24
Сумма по столбцу amazon	20
Сумма по столбцу tumblr	10
Сумма по столбцу linkedin	7
Сумма по столбцу vimeo	3
Сумма по столбцу pinterest	1

На графике в нижней части вкладки приведена информация по количеству записей в выборке по каждому СД. Обращает внимание недостаточное количество записей в выборке по СД, начиная с 'facebook' и правее. График приведен в логарифмическом формате, поэтому нужно обращать внимание на абсолютные значения, а не только на высоту столбиков на нем (рис. 3).

СД	0/1	Среднее по столбцу cost ads_msk	СД	0/1	Среднее по столбцу cost ads_msk
amazon	0	0,193	stack-overflow	0	0,150
	1	0,626		1	0,276
facebook	0	0,191	tumblr	0	0,194
	1	0,844		1	0,042
linkedin	0	0,193	twitter	0	0,194
	1	1,979		1	0,135
pinterest	0	0,194	vimeo	0	0,194
	1	0,010		1	1,227
reddit	0	0,194	vk	0	0,174
	1	0,381		1	0,717
slideshare	0	0,194	wikipedia	0	0,173
	1	0,393		1	0,628
			youtube	0	0,130
				1	0,412

Рис. 4 – Оценка влияния каждого социального домена на среднюю стоимость клика

Для оценки влияния каждой СД на ССК был сделан расчет ССК по каждому социальному домену. При этом в строчке «1» рассчитывается ССК для записей с присутствующим СД, а для строчки «0» - с отсутствующим. По их соотношению можно сделать вывод о степени влияния этого СД на ССК. Чем больше перевес в сторону «1» тем большее влияние этот СД оказывает на ССК. Для удобства соотношение представлено в виде цветowych гистограмм. Интересным представляется результат с запросами, в которых SERP содержит коммерческий сайт Amazon дают более низкую цену в сравнении с бесплатным видеохостингом Vimeo.

Для проверки полученных результатов был выполнен регрессионный анализ средствами Excel. Значение коэффициента влияния каждого из СД приведено в таблице, а на круговом графике коэффициенты обозначены синей линией.

Таблица 3

Регрессионная статистика

Множественный R	0,159329502
R-квадрат	0,02538589
Нормированный R-квадрат	0,024117878
Стандартная ошибка	1,058696596
Наблюдения	10006

Таблица 4

Регрессионная статистика по всем социальным доменам в SERP

	Коэффициенты	Стандартная ошибка	Количество
Y-пересечение	0,110448004	0,013437372	
amazon	-0,025020821	0,250545079	20
facebook	0,220573522	0,162593859	46
linkedin	1,378025655	0,408588756	7
pinterest	-0,103578961	1,058946646	1
reddit	-0,044749009	0,198159522	29
slideshare	-0,011213394	0,217755833	24
stack-overflow	0,003130957	0,025198665	3499
tumblr	-0,36011157	0,336749645	10
twitter	-0,422543502	0,179907603	36
vimeo	0,774858452	0,633912607	3
vk	0,412842762	0,059394358	366
wikipedia	0,380172137	0,051833504	462
youtube	0,219382852	0,029047959	2271

Оранжевом цветом представлено количество записей в выборке по каждому СД. Например, для 'LinkedIn' при самом большом коэффициенте влияния (1,38) количество записей всего 7, что говорит о статистической ненадёжности этого показателя. Таким образом, по предоставленной выборке можно сделать достаточно статистически значимый вывод о положительном влиянии на ССК

только одного СД - 'Youtube', по остальным СД или количество записей недостаточно для достоверности вывода ('linkedin'), либо коэффициент влияния близок к 0 (stack-overflow). Результат регрессии были проверены регрессором с регуляризацией Lasso из библиотеки sklearn и дали практически тот же результат.

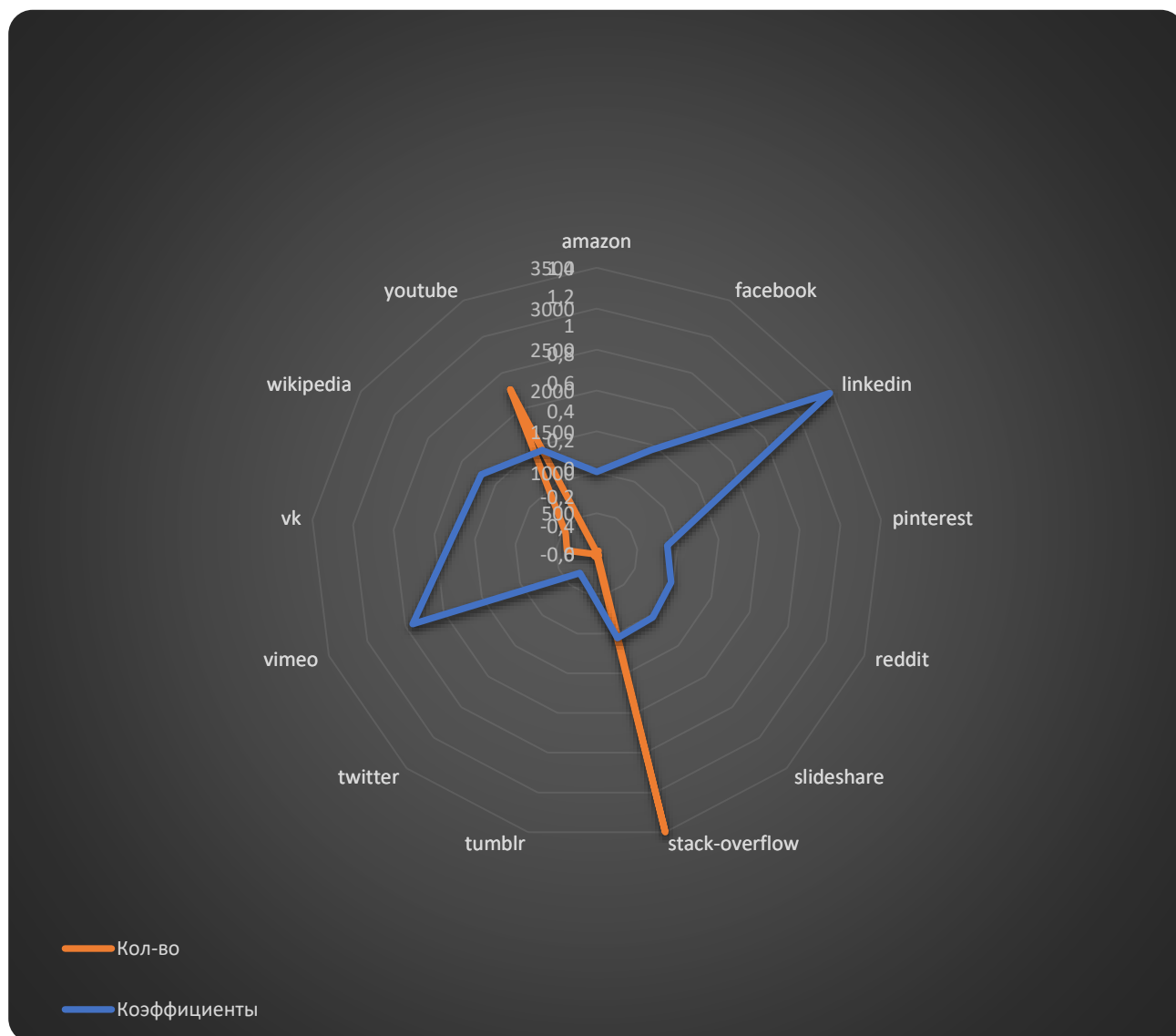


Рис. 5 – Значения коэффициентов влияния каждого социального домена

Для оценки одновременного влияния нескольких СД на ССК была создана матрица влияния, представленная в приложении. Однако ее результаты также статистически не слишком значимы по описанным выше причинам.

Исследование выполнено при финансовой поддержке РФФИ и Пермского края в рамках научного проекта № 18-410-590007.

Библиографический список

1. Dou W. et al. Brand positioning strategy using search engine marketing //Mis Quarterly. – 2010. – С. 261-279.
2. Pan B. et al. The dynamics of search engine marketing for tourist destinations //Journal of Travel Research. – 2011. – Т. 50. – №. 4. – С. 365-377.
3. White R. W., Dumais S. T. Characterizing and predicting search engine switching behavior // Proceedings of the 18th ACM conference on Information and knowledge management. – ACM, 2009. – С. 87-96.
4. Jansen B.J., Zhang M., Zhang Y. The effect of brand awareness on the evaluation of search engine results //CHI'07 Extended Abstracts on Human Factors in Computing Systems. – ACM, 2007. – С. 2471-2476.
5. SERP [Электронный ресурс] <https://www.cy-pr.com/wiki/SERP/>

Приложение. Матрица влияния связей социальных доменов в SERP на ССК

twitter	vimeo	linkedin	reddit	facebook	pinterest	slideshare	wikipedia	vk	youtube	tumblr	stack-overflow	Avg click cost
---------	-------	----------	--------	----------	-----------	------------	-----------	----	---------	--------	----------------	----------------
