

Дипломная работа

Разработка стратегии поискового продвижения портала «**FigaSebe.ru**»

Дипломный руководитель: _____

Выполнил: **Захарченко Дмитрий**

Москва 2012

Оглавление

1. Постановка задачи	4
2. Анализ спроса	5
2.1 Семантическое ядро сайта	5
2.2 Обзор конкурентов.....	9
2.2.1 Обзор сайта magic-class.ru	11
2.2.2 Обзор сайта goodmagic.ru	11
2.2.3 Обзор сайта micromagic.ru.....	13
2.2.4 Обзор сайта what-focus.ru.....	13
2.3 Резюме	13
3. Технология поискового продвижения.....	16
3.1 Чем хорошо продвижение	16
3.2 Структурная схема работы поисковой системы	16
3.3 Особенности работы поисковиков	18
3.3.1 Способы определения спама	19
3.4 Индексация	20
3.4 Типы объектов	23
3.5 Факторы ранжирования.....	23
3.5.1 Текстовые факторы.....	24
3.5.2 Ссылочные факторы	25
3.5.3 Хостовые.....	29
3.5.4 Региональные	29
3.5.5 Поведенческие	29
3.5.6 Социальные.....	32
3.6 Машинное обучение	33
4. Аудит сайта figasebe.ru.....	34
4.1 Анализ индексации	34
4.2 Анализ текстовых факторов.....	38
4.3 Ссылочные факторы	39
4.4 Хостовые факторы	45
4.5 Региональные факторы.....	45
4.6 Поведенческие факторы	45
4.7 Социальные факторы	47
5. Сравнение с конкурентами.....	48
5.1 Анализ сайта goodmagic.ru	48
5.2 Анализ what-focus.ru	49

5.3 Анализ micromagic.ru	50
5.4 Анализ magic-class.ru.....	51
5.5 Резюме	52
6. Развитие сайта figasebe.ru	53
6.1 План работ.....	53
6.3 Перспективы	55
7. Заключение	55
8. Список использованной литературы	56

1. Постановка задачи

В рамках дипломной работы разрабатывается стратегия продвижения для сайта figasebe.ru.

Figasebe.ru – ресурс, посвященный обучению фокусам. Обучение может быть трех видов: онлайн (просмотр видео-роликов), через посещение тренингов и просмотр dvd дисков с обучающими записями.

Для того чтобы разработать стратегию продвижения, необходимо составить максимально полную картину происходящего в данной нише Рунета. Для этого необходимо:

- 1) Определить целевую аудиторию сайта, оценить уровень спроса, составить максимально подробный список потребностей и запросов пользователей
- 2) Оценить конкурентную среду, выявить достоинства и недостатки других сайтов подобной тематики. Как с маркетинговой точки зрения, так и особенности технической реализации.
- 3) Разработать алгоритм увеличения видимости сайта figasebe.ru в поисковиках по необходимым запросам.
- 4) Предложить способы удержания аудитории и увеличения количества посещений сайта figasebe.ru. Разработать план по улучшению проекта.

2. Анализ спроса

Для реализации успешной стратегии продвижения проекта, необходимо четко представлять уровень спроса. Это является ключевым фактором.

На сайте figasebe.ru представлены следующие услуги:

- 1) Покупка видео с обучением фокусам. Стоимость ролика от 9 до 15\$.
- 2) DVD курсы с полным списком.
- 3) Магазин реквизита
- 4) Тренинги

Целевая аудитория ресурса – молодые люди 10-25 лет, которые хотят привлекать внимание на вечеринках и уметь удивлять людей. Часто школьники и студенты. Большинство мужчины.

2.1 Семантическое ядро сайта

Попробуем описать потребности, которые возникают у целевой аудитории. Для этого составим семантическое ядро сайта.

Семантическое ядро – совокупность словесных описаний потребностей пользователей.

Для этого, представим себя на месте пользователя. Предположим, что бы он ввел, если бы хотел научиться делать фокусы. Чем больше вариантов запросов мы придумаем, тем полнее будет семантическое ядро.

Очевидно, что нельзя предусмотреть абсолютно все виды написания запросов и придумать все варианты формулировок. На каком-то этапе фантазия заканчивается и необходимо использовать качественно другой подход. Его называют « подбор неявных запросов». К таким запросам относят:

- запросы, по которым подвигаются конкуренты;
- что еще искали пользователи, когда вводили основные слова;

Инструменты для подбора неявных запросов:

- поисковые подсказки
- подсказки систем контекстной рекламы
- Neiron

Так же существуют способы предсказания запросов. Условно их можно разделить на 2 вида:

Зависимое предсказание – подбор запросов с учетом сезонности и известных последовательности действий (например, выборы президента или покупка колясок после рождения ребенка).

Независимое предсказание – прогноз для одноразовых событий (например, первый полет на Марс). Такой способ актуален для новостных сайтов и других событийных тематиках.

В данной работе способы предсказания спроса рассматриваться не будут.

После получения большого списка запросов, необходимо проверить их ценность. Действительно ли такие слова вводят пользователи в поисковую систему и как часто они это делают.

Эту оценку можно сделать при помощи статистических инструментов:

- статистика ключевых слов (yandex.wordstat, adwords, rambler adstat)
- счетчики посещений (как чужие, так и собственные)
- сторонние базы и системы статистики (semrush, База Пастухова, liveinternet, neiron)

Чтобы упростить работу, зачастую эти 2 процесса совмещают.

Одновременно придумываются запросы, оценивается их частота и собираются «похожие слова».

Для этого в статистике ключевых слов вводят самый очевидный запрос (например, «фокусы»), и смотрят, какие запросы с этим словом задавали пользователи.

При таком подходе «от общего к частному» часто возникает проблема омонимии. В нашем слово «фокус» может означать:

- автомобиль Форд Фокус
- физическая характеристика оптической системы (например, объектива)
- группы людей и наборы тестов для исследований
- учебник английского языка
- ТЦ в челябинске
- велосипед марки стелс

Т.е. люди, которые вводили слово «фокус» в поисковую строку не обязательно хотели научиться угадывать карты в колоде. Все эти запросы являются нетематическими для нашего сайта.

Так же можно выделить следующие группы неподходящих запросов:

- Нецелевой (сайт не хочет удовлетворить потребность пользователя)
- Мусорные (со служебными символами, абракадабры и т.п.)
- Пустые (запросы с низкой полнотой)

Полнота запроса – характеристика запроса, равная отношению запроса только из этих слов ко всем запросам, где есть этим слова.

Использовать параметр «полнота запроса» необходимо с учетом конкуренции. В низком показателе полноты нет ничего плохого. Однако, он очень часто сопровождается высокой конкуренцией. Поэтому продвигаться по таким запросам экономически нецелесообразно.

После получения большого списка запросов и их частотностей, необходимо их логически сгруппировать и объединить синонимичные формулировки в однозначные потребности. Т.к. страница сайта прежде всего должна помогать удовлетворять потребность пользователя.

Для проекта figasebe.ru были выделены следующие типы запросов и созданы соответствующие категории:

- по предметам (с картами, монетками, зажигалками и т.д.)
- по видам (детские, домашние, для начинающих, простые)
- информационные (известные фокусники, телепередачи)

Всего было обработано: ~10000 запросов

Отфильтровано

Нетематических ~9000 (большая часть относится к автомобилю)

Мусорных ~100

Пустых ~600

Итого подобрано: **406 запросов**

Пример группы правильных тематических целевых запросов см. Табл. 1.

Фокусы с деньгами			
Фокусы с деньгами			
фокусы с деньгами	1769	483	27%
фокусы с деньгами обучение	316	207	66%
как сделать фокус с деньгами	106	85	80%
секреты фокусов с деньгами	339	80	24%
секреты фокусов с деньгами видео	109	26	24%
как научиться фокусам с деньгами	132	66	50%
как делать фокусы с деньгами	109	40	37%
фокус с купюрой	460	106	23%
видео фокусы с деньгами	247	31	13%

Фокусы с монетами			
фокусы с монетами	7873	1777	23%
фокусы с монетами обучение	2362	1743	74%
фокусы с монеткой	1977	611	31%
фокусы с монетами обучение видео	362	286	79%
обучение фокусам с монеткой	424	279	66%

Монета в бутылке			
секрет фокуса монета в бутылке	283	173	61%
фокус с бутылкой и монетой	217	126	58%

Как делать фокусы с монетами			
как делать фокусы с монетой	175	95	54%
как научиться фокусам с монеткой	166	87	52%
как научиться фокусам с монетой	152	75	49%
как делать фокусы с монеткой	165	64	39%
как сделать фокус с монетой	255	169	66%

Секреты фокусов с монетами			
секреты фокусов с монетами	1471	234	16%
секреты фокусов с монеткой	162	23	14%

Табл. 1 Запросы для категории «фокусы с деньгами»

2.2 Обзор конкурентов

Для проверки правильности подбора семантики, а так же для оценки конкурентности необходимо проанализировать сайты, оказывающие услуги в той же области.

Чтобы составить список конкурентов можно использовать 2 подхода:

- Ручной подбор через выдачу поисковиков
- Использование рейтингов (Мегаиндекс, Seorate.ru)

Ручной подбор:

- + максимальная полнота
- низкая технологичность

Использование рейтингов:

- + удобство и быстрота
- маленький охват

Основным недостатком рейтингов является ограниченное количество категорий. Конкретно для категории «фокусы» в этих системах статистики нет, поэтому нельзя посмотреть рейтинг сайтов.

Поэтому для начальной выборки используем выдачу поисковика. Но для большей точности используем обе подхода сразу. Сначала по нескольким основным запросам получим список сайтов из ТОП10. Затем проверим каждый сайт в системе Мегаиндекс и сравним по параметру сайты по параметру «эффективные показы». Эффективные показы является характеристикой того, сколько раз пользователи поисковых систем увидят описание сайта в результатах поиска по поисковому запросу.

$$\text{Количество эффективных показов} = (\text{Количество Запросов в месяце}) \times (\text{Видимость [\%]} / (100\%)).$$

Были выбраны следующие начальные запросы:

- обучение фокусам
- карточные фокусы
- фокусы с монетками
- фокусы для начинающих

С учетом пересечений по запросам было получено 32 сайта. Затем были отфильтрованы сайты, содержащие вирусы, неработающие сайты, видеохостинги (youtube.com, video.yandex.ru, rutube.ru, video.mail.ru) и не прямые конкуренты. К прямым конкурентами были отнесены сайты, полностью посвященные фокусам и

обучению. К непрямым – общетематические сайты, где фокусам посвящен только раздел и несколько страниц.

После фильтрации осталось 14 сайтов, которые были отсортированы по количеству эффективных показов. Сайт figasebe.ru находится на 5ом месте по популярности (Табл. 1).

#	Сайт	Эфф. Показов
1	http://magic-class.ru/	24911
2	http://goodmagic.ru/	21425
3	http://www.micromagic.ru/	17305
4	http://what-focus.ru/	12973
5	http://figasebe.ru/	11856
6	http://micromagic.narod.ru/	5570
7	http://cardtrick.ru/	5123
8	http://www.fokus-pokus.info	4535
9	http://fokusy.kak-nauchitsya.ru/	3619
10	http://elementrick.ru/	3522
11	http://russiamagic.ru/	3469
12	http://fokusmagic1.narod2.ru/	2313
13	http://www.hotmagic.ru/	1862
14	http://ridsmagic.ru/	310

Табл. 2. Выбор конкурентов по популярности.

Проведем предварительный анализ сайтов конкурентов.

2.2.1 Обзор сайта magic-class.ru

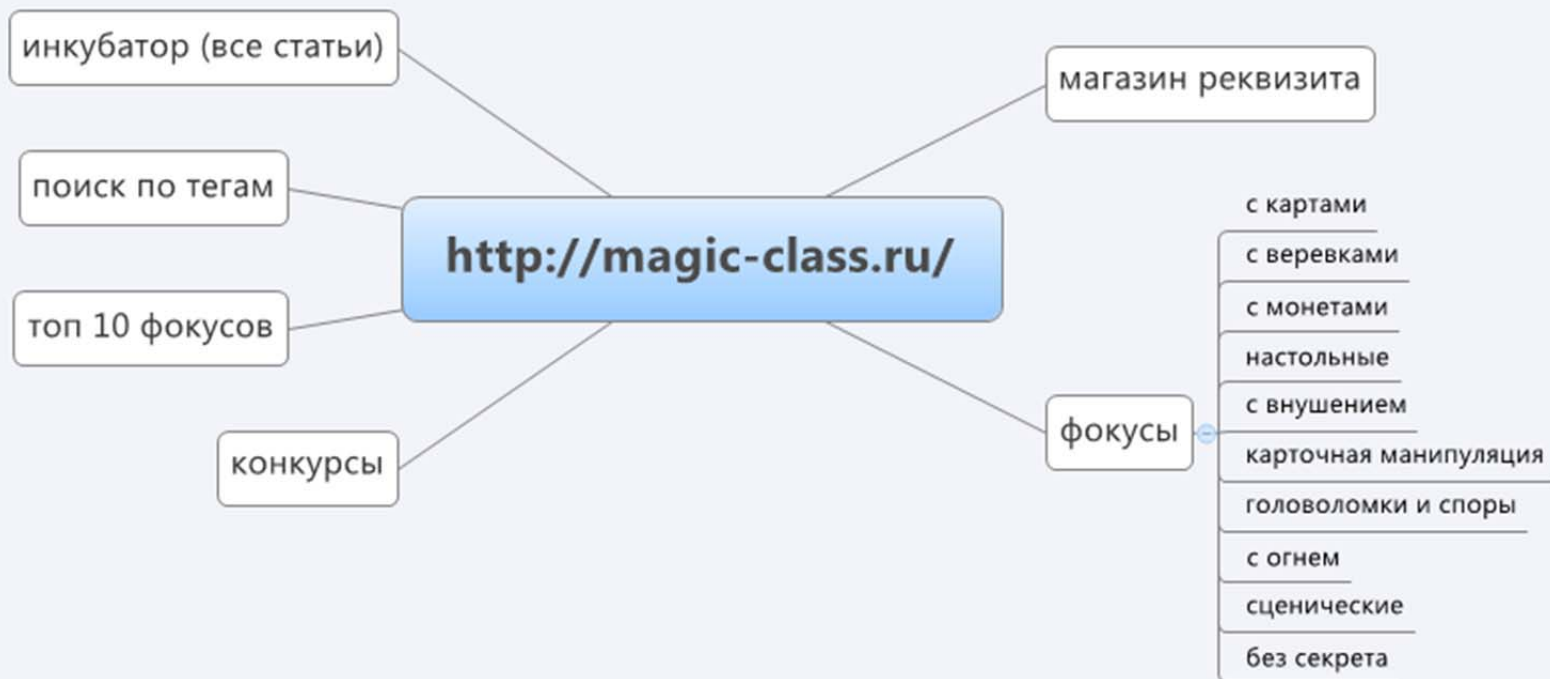


Рис. 4.1 Структура сайта magic-class.ru

Сайт magic-class.ru представляет из себя блого-социальную сеть, посвященную фокусам. Пользователи могут публиковать собственные видео с фокусами, комментировать ролики. Всего написано ~9000 постов. Монетизируется через продажу контекстной рекламы.

2.2.2 Обзор сайта goodmagic.ru

Сайт является модерлируемым каталогом фокусов. Пользователь может добавить описание фокуса или видео ролик, и, после одобрения администратора, фокус появиться на сайте. Добавленные записи могут обсуждаться пользователями.

Из отличительных особенностей – удобная и красивая категоризация фокусов. Имеет базу подписчиков: ~1200 человек получают новости через RSS.

Сайт зарабатывает на продаже контекстной рекламы.

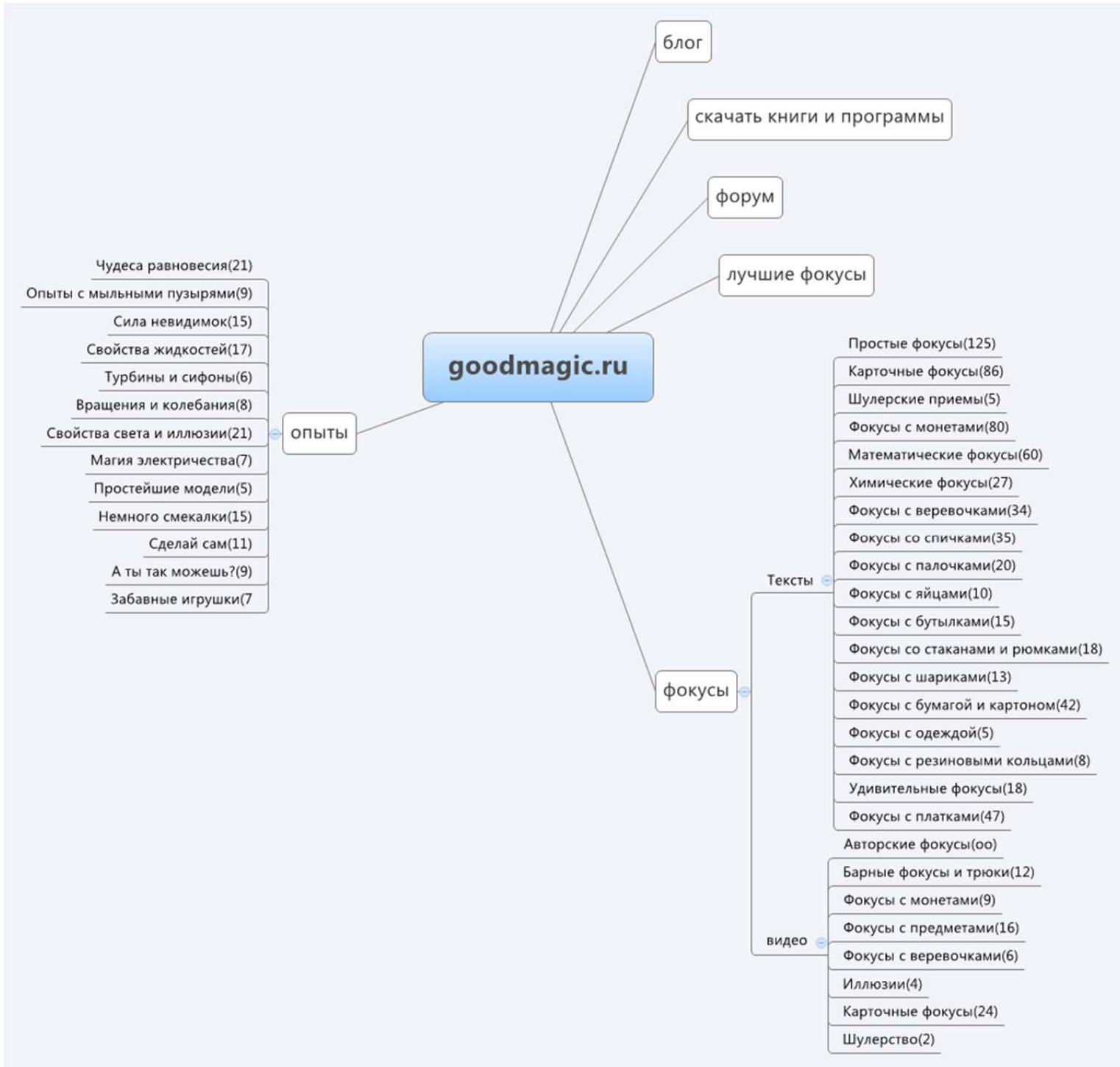


Рис. 4.2 Структура сайта goodmagic.ru

2.2.3 Обзор сайта micromagic.ru

Закрытый каталог фокусов, новости по теме.

Особенность - достаточно большой и обновляемый форум.

Продает баннерную, контекстную рекламу, размещает ссылки.



Рис. 4.3 Структура сайта micromagic.ru

2.2.4 Обзор сайта what-focus.ru

Авторский блог без ярко выраженной структуры. Неопрятный дизайн, часть роликов не загружается. На первый взгляд удивительно, как он может занимать первые места в рейтинге. Возможно, причина будет установлена при подробном техническом анализе.

2.3 Резюме

При осмотре сайтов конкурентов, было выявлено, что основные страницы – категории фокусов. Большая часть материалов – UCG. Возможно, эти схемы придется использовать при развитии сайта figasebe.ru

У части сайтов семантика расширена математическими фокусами, опытами и другими трюками. Для проекта figasebe.ru эти запросы на данном этапе не рассматриваются.

С учетом обзора конкурентов и категоризации семантического ядра составим оптимальную структуру сайта, посвященного фокусам, с точки зрения максимального охвата:

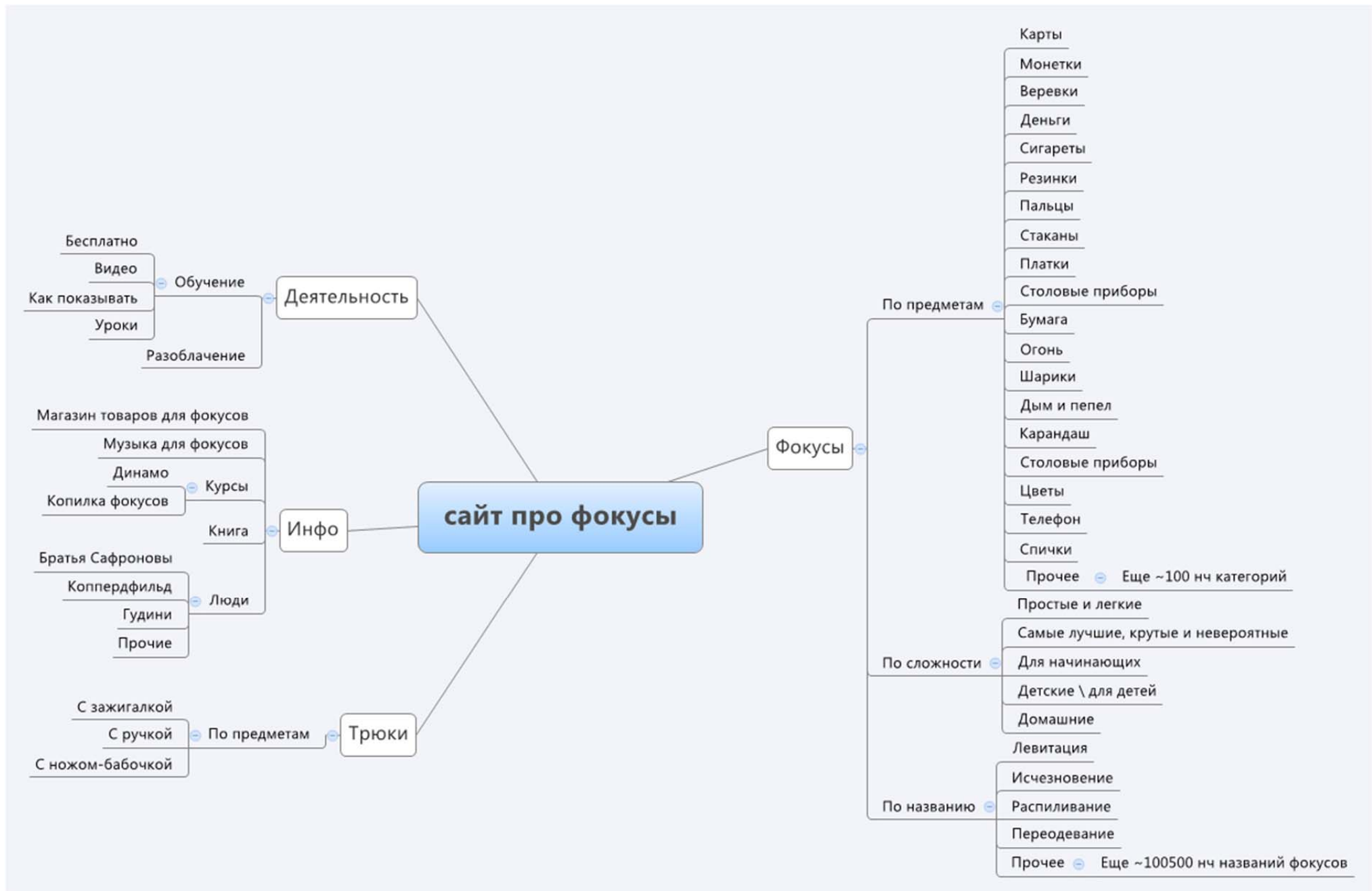


Рис. 4.4 Оптимальная структура сайта про фокусы

Если наложить на существующий сайт figasebe.ru, то необходимо будет увеличить список трюков с предметами, изменить способы подбора трюков, «подтянуть информационную составляющую».



Рис. 4.5 Структура сайта figasebe.ru

3. Технология поискового продвижения

3.1 Чем хорошо продвижение

...

Для того, чтобы продвинуть выбранный сайт в поисковых системах, прежде всего необходимо проанализировать принцип работы этих поисковых систем.

3.2 Структурная схема работы поисковой системы

... добавить 2 абзаца про историю ПС ...

Реальная схема каждой из поисковых систем является коммерческой тайной и не подлежит разглашению. Однако, при разработке поисковика существуют общие принципы и наработанные практики. По этим общим принципам и построена нижеприведенная структурная схема.

В ней не разбираются подробно конкретные реализации алгоритмов. Показаны только основные логические узлы современной системы.

Условно, работу ПС можно разделить на 4 больших блока:

- 1) Сбор и обработка информации
- 2) Хранение обработанной информации (индексы)
- 3) Ранжирование
- 4) Взаимодействие с пользователями

WEB

Поисковик

СБОР ДАННЫХ:

- много роботов. Классификации по видам контента\скорости обхода\названию
- учет параметров. Robots.txt,nofollow

КЭШ:

- сохранение копии материала
- оптимизация ранжирования. Для популярных запросов ответы подготовлены и лежат в «оперативке», что позволяет экономить ресурсы

ФИЛЬТРАЦИЯ:

- спам
- вирусы
- запрещенные материалы
- нарушения поисковой лицензии

ИНДЕКСАЦИЯ:

- определение языка, формата документа, региона, параметров для ранжирования
- инвертированный индекс (поиск по терминам), k-граммный (для поиска по маске), индекс частых биграмм (бритни спирс), координатный индекс (позволяет учитывать расстояние м\у словами), параметрический и др.?
- индексы могут динамически обновляться, для ускорения сжимаются, и предварительно сортируются для быстрого поиска

НАСТРОЙКИ ЮЗЕРА:

- регион
- подсказки, список запросов

ОБРАБОТКА:

- подсказки
- спел-чекер
- определение языка
- определение региона
- морфологический разбор и снятие омонимии
- расширение запроса (добавление синонимов)
- выделение объектов (люди, географические названия)
- категоризация (спектр): купить\обзор\отзывы
- классификация запроса : термин,цитатный,навигационный,коммерческий

РАНЖИРОВАНИЕ:

- много факторов. Классификации: статические\динамические. Текстовые\ссылочные\возрастные\кликовые и т.д.
- итоговая формула большая, постоянно меняется и никто не знает точных коэффициентов? (=

ФОРМИРОВАНИЕ ВЫДАЧИ:

- заголовок
- сниппет
- выделение областей
- фавикон
- уточнения по категориям
- структура сайта
- контактная информация, прочее содержимое сайта
- колдунчики (контент от ПС)
- предпросмотр страниц
- контекстная реклама

ОБУЧЕНИЕ:

- машина на обучающих выборках пытается определить коэффициенты для факторов, чтобы сортировка результатов по алгоритму совпала с ручной сортировкой. Проверка происходит на контрольных выборках.

ОЦЕНКИ АССЕСОРОВ:

- работают по инструкции. Инструкцию пишут работники ПС. Их представление о «правильности» интернета косвенно определяет ранжирование
- паре «запрос-документ» выставляют оценку

ЗАПРОС

РЕЗУЛЬТАТ

ВСЕ ПОД КОЛПАКОМ

ПС собирает статистику по ВСЕМУ. Поведение пользователей, запросы, документы, оценки. Для улучшения качества поиска (=

3.3 Особенности работы поисковиков

Свою работу поисковые машины, могут оценить по следующим метрикам:

Полнота – отношение найденных документов ко всем существующим.

Точность – отношение найденных релевантных ко всем найденным.

rFound – правильность сортировки результатов поиска.

Цель поисковой системы – улучшить качество поиска. Основные проблемы, с которым сталкиваются веб-поиска – это огромные объемы информации и большой процент спама.

Из-за больших объемов возникают технические сложности: множество серверов, скорость обработки данных, объемы хранимой информации. Ограниченность в ресурсах вызывает необходимость идти на компромиссы и оптимизировать работы алгоритмов, что приводит к ухудшению качества поиска.

Спам мешает и пользователю, и поисковой системе. Из-за спама в результаты поиска попадают страницы, где нет ответа на вопрос. В итоге, качество поиска снижается, пользователь тратит больше времени на поиск нужной информации, а поисковая система – ресурсы на индексирование и хранение бесполезных данных.

Чтобы такого не происходило, поисковики стараются не допускать попадания спама в результаты поиска. Для этого используется система поискового антиспама — набор алгоритмов, позволяющих отделить спам от качественных веб-страниц. Страницы, содержащие спам, могут быть понижены при ранжировании или исключены из поиска.

Это должен учитывать оптимизатор при работе с веб-сайтом и проверять все действия на возможную их классификацию в качестве спама.

3.3.1 Способы определения спама

Т.к. поисковые машины обрабатывают огромное количество данных, они имеют достоверную статистику и могут на основе нее сортировать спам.

Подходы к определению спама:

- Статистические
- Лингвистические
- Временные срезы
- Анализ логов (поведение пользователей)
- Поведение спамеров (оптимизаторов)

Санкции за наличие спама:

- Штраф на документ-запрос
 - Штраф на документ
 - Штраф на HOST
 - Удаление из индекса документа (-ов)
 - Удаление из индекса HOSTа
- (см. 3.5 Типы объектов)

Штрафы проявляют себя в понижении ранжирования сайта и документов.

Удаление из индекса означает, что документ вообще не будет показываться в результатах поиска.

Основная задача оптимизатора: избегая санкций за спам, максимально оптимизировать каждый из факторов ранжирования.

Способы определения спама по различным параметрам сайта будут указаны в разделе 3.6 Факторы ранжирования.

3.4 Индексация

Поисковая машина отвечает на вопросы пользователей, находя нужные документы в интернете. Размеры современного интернета исчисляются в миллиардах миллиардов байтов. Не существует технической возможности обходить весь интернет каждый раз, когда пользователь задает вопрос. Поэтому поисковая система заранее «скачивает» все страницы, определенным образом обрабатывает их, т.е. подготавливает поисковый индекс и сохраняет его у себя на серверах.

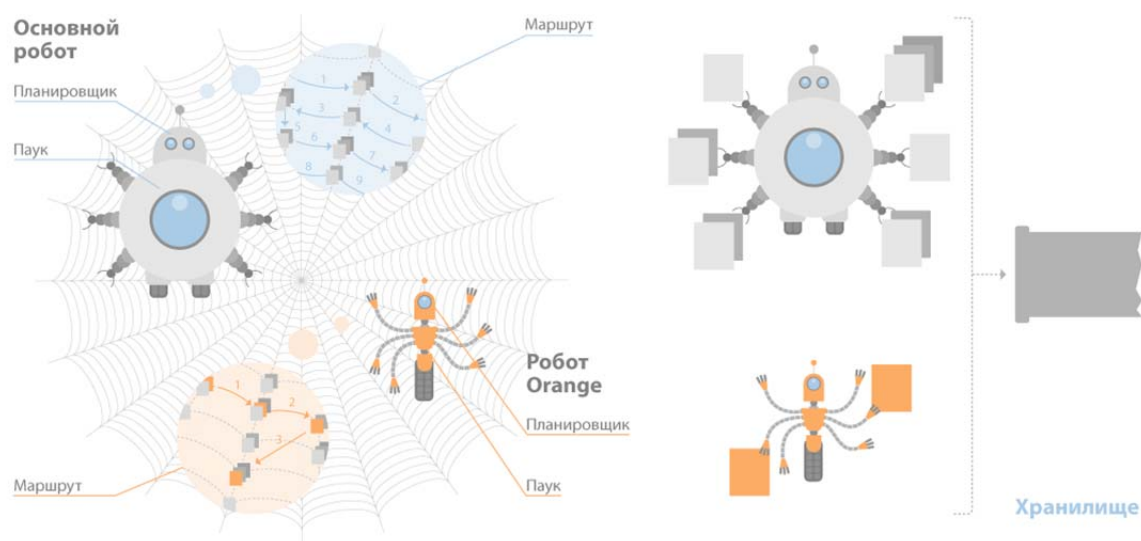


Рис. 3.1. Схема получения страниц для индексации

Индексация – первоочередное действие в работе поисковых систем, поэтому ему нужно уделить особое внимание. Если у поисковика в индексе не будет необходимого сайта, то он не сможет показать его в качестве ответа на вопрос пользователя.

Т.к. ресурсы поисковой машины ограничены, то каждому сайту она уделяет определенной количество «внимания», т.е. выделяет краулинговый бюджет.

Рассмотрим стандартные способы влияния на индексацию.

Способы ускорения и улучшения индексации больших сайтов (более 10`000 страниц), у которых часто бывают проблемы с индексацией, рассматриваться не будут.



Рис. 3.2. Схема составления индекса поисковой системы

Способы влияния на индексацию:

Robots.txt – специальный файл, который содержит указания для поисковых систем. Можно указывать страницы, запрещенные к индексированию, разрешенные к индексированию,

Sitemap.xml – файл, содержащий ссылки на страницы сайта, рекомендуемые к индексации.

Заголовки сервера:

Коды ответов:

200 – ОК, страницу существует

301 – страница перемещена навсегда

404 – страница не найдена

X-Robots-Tag – можно указать дату, после которой страница будет недоступна.

Метатеги – html теги на странице, которые рекомендуют запрет\разрешение на индексацию страницы и ссылок.

Дубликаты страниц

Т.к. поисковая система стремится дать наиболее полный и подробный ответ на вопрос пользователя, алгоритм старается не выдавать несколько

одинаковых результатов по одному запросу. Это относится как к дублям страниц на разных сайтах, так и на одном.

Возможные проблемы при наличии дублей на сайте:

- распыление краулингового бюджета (проиндексируется меньше или не то, что надо)
- возможен выбор неосновного документа, как наиболее вероятного (при ранжировании не будут учитываться факторы, которые относятся к основному документу)
- вероятность неиндексации основного документа (в случае нахождения дубля раньше основной страницы. Проблемы см. пунктом выше)

Поэтому при разработке интернет-сайта, стоит обращать внимание на возможные дубли страниц и стараться их устранять. Если это невозможно, то оптимизатор для успешного продвижения ресурса должен закрывать неосновные страницы от индексации.

3.4 Типы объектов

Работа поисковой машины может быть условно сведена к 3 видам объектов.

- Хост (весь сайт)
- Документ (страница сайта)
- Запрос-документ (связка вопроса пользователя и результата выдачи)

Каждый из факторов ранжирования применим к одному или нескольким объектам. При оптимизации параметров необходимо различать эти объекты и уделять большее внимание тому объекту, который имеет большее влияние.

Тип запроса	Объект и влияние		
	хост	документ	запрос-документ
высокочастотный	низкий	средний	высокий
среднечастотный	средний	средний	средний
низкочастотный	высокий	средний	низкий

Табл. 3. Зависимость влияния объекта на вклад в ранжирование

3.5 Факторы ранжирования

Условно все параметры сайтов можно поделить на следующие типы:

- Поведенческие
- Ссылочные
- Текстовые
- Региональные
- Хостовые
- Социальные

Степень влияния факторов определяется достоверностью данных. Чем больше информации у поисковой машины по группе факторов, и чем меньше способов влияния на эту группу (спама), тем больше ПС доверяет этому типу параметров. См. табл.4

Тип запроса	Степень влияния на ранжирование					
	поведенческие	ссылочные	текстовые	региональные	хостовые	социальные
высокочастотный	крайне высокая	средняя	средняя	высокая	низкая	средняя
среднечастотный	средняя	высокая	средняя	высокая	средняя	низкая
низкочастотный	низкая	низкая	высокая	средняя	высокая	низкая

Табл. 4. Зависимость влияния типа факторов на вклад в ранжирование

Далее подробно разбираются каждый тип факторов. Типы перечислены в хронологическом порядке: вначале те, которые раньше стали учитывать ПС.

3.5.1 Текстовые факторы

Встречаемость слов из текста запроса в материалах, относящихся к объекту документа.

Способы определения спама:

- размер документа в словах (длинные - спамные)
- длина области (для заголовка естественно 2-10 слов. Больше – спам)
- доля видимого текста (соотношение содержимого к разметке. У спама 80% содержимого. У естественных – много разметки.)
- средняя длина слова (тексты с длинными словами спамны)
- кол-во слов в анкерах (много слов атрибут спамных документов)
- сжимаемость текста(спамные тексты сжимаются существенно лучше естественных)
- дисперсия длин предложений (средняя длина предложений в русском языке 5-10 слов. У спамных текстов больше)
- по частоте слов (в спамных документах мало упоминаются самые частотные слова языка)
 - частотные биграммы ()
 - орфография (документы с больших количеством ошибок спамные)
 - плотность неуникальных n-грамм (в спамных меньше)

Подробнее [6],[7]

Способы определение дублей:

- шинглы
- частотные биграммы
- распределение частотностей слов

Чтобы создавать новые правильные тексты нужно ставить ограничения: «меньше воды, больше мяса». Не использовать вводные слова, деепричастные и причастные обороты, пустые выражения вроде «всем

известно что». Тексты должны быть полезны пользователю. Например для кондиционеров: не что такое кондиционеры и зачем они нужны. А какие они бывают и каких выбирать.

Оптимизация по текстовым факторам:

Один из простых алгоритмов ранжирования по текстовым факторам является алгоритм BM25.

$$W_{single} = \log(p) * \frac{TF}{TF + k_1 + k_2 * DocLength}, \quad (2)$$

$$k_1 = 1, k_2 = 1/350$$

В этой формуле TF – число вхождений леммы в документ, DocLength – длина документа в словах.

Особенностью данной функции является сильная зависимость от числа вхождений, и слабая от длины документа. Достаточно быстро уходит в насыщение. Нет смысла продолжать оптимизировать тестовую составляющую, т.к. пропадет естественность текста.

Это алгоритм может применяться отдельно по различным зонам (заголовки, содержание, анкорлист).

3.5.2 Ссылочные факторы

Статические - не зависят от запроса.

PageRank (постраничный, подоменный)

TrustRank

Кол-во ссылок на документ\хост

Кол-во ссылающихся хостов

Распределение по страницам\тиц\рг

Динамические - зависят от запроса. Это текстовые, которые применяются к объекту «запрос-документ». Имеют достаточно быструю скорость влияния.

Эти два типа ссылочных факторов обладают разным временем учета и могут быть оптимизированы отдельно.

Способы борьбы со спамом:

Распределение ссылок по

- страницам (у спамных сайтов мало, у естественных много входных страниц)
- титц и рг (у спамных смещение в сторону «прокаченных» ресурсов)
- длина анкора (у естественных большинство однословных запросов, у спамных больше слов)
- динамике появления ссылок на доноре \ на акцепторе
- динамике появления SEO ссылок
- время появления ссылки и документа-источника
- отношение исходящих ссылок \ исходящим SEO
- отношение страниц \ исходящим ссылкам
- распределение по уровням вложенности доноров (морд << 2ой << Зий << суммы остальных)
- распределение по кол-ву ссылок с домена (сквозные ставят партнеры. Пользователи ставят 1)

Яндекс в 2009 году опубликовал работу, известную как «Мадридский доклад», в котором пояснил, как определяет сео-ссылки.

SEOText – показатель оптимизированности текста

К типичным слова *SEOText* относят: купить, продажа, заказ, москва, недорого, дешево, оптом, магазин, от производителя, каталог, прайс, доставка, цена и т.д.

К типичным естественным слова в ссылках относят: прилагательные, название компании, цифры, Цитирование (источник, подробно) и т.д.

SEOin - показывает вероятность того, что сайт продвигается с помощью платных ссылок

SEOout – вероятность того, что документ продает ссылки

SEOLink – вероятность того, что конкретная ссылка сео-ссылка

$$SEOout = k1 * AvgSEOin + k2 * AvgSEOText + k3 * NTh + ... (1)$$

$$SEOLink = l1 * SEOText + l2 * SEOin + l3 * SEOout + ... (2)$$

(1) в (2)

$$SEOLink = I1 * SEOtext + I2 * SEOin + I3 * k1 * AvgSEOin + I3 * k2 * AvgSEOtext + I3 * k3 * NTh + I3 * \dots + \dots$$

Этот алгоритм может быть улучшен за счет использования сегментатора документов в алгоритме microHITS для блоков ссылок. [8]

Определение бирж статей:

- Кластерность (все ссылки размещаются в одном разделе. Например, «партнеры»)
- Тип документа
- Длина документа (у бирж статей есть минимальный допустимый объем текста. Большинство сео-текстов незначительно превышают этот объем)
- avgSEOout по кластеру

Определение единичной ссылки:

- avgSEOout HOSTa
- Скорость появления ссылок

Способы учета алгоритма антиспама.

Т.к. поисковая машина умеет распознавать покупные ссылки, они могут занижать степень влияния для таких ссылок. Поэтому, чтобы увеличить передачу статического веса на документы, желательно маскировать покупные ссылки.

Маскировка SEOtext:

использовать в качестве анкоров url. Меньше сео-слов, больше типичных естественных слов.

Маскировка SEOout :

- учитывать расположение на странице (не ставить в подвал и ссылочные блоки)
- тематичность соседей (желательно, чтобы донор ссыллся на документы одной тематики)
- Низкий avgSEOin соседей
- Низкий avgSEOtext соседей
- Низкий avgSEOout HOSTa

- Не попадание документа в подозрительный кластер. (биржа статей)

Маскировка SEOin :

- Естественные статистические распределения
- Естественная динамика прироста
- Низкий avgSEOtext анкор-листа
- Низкий avgSEOout документов-источников ссылок

При естественном развитии событий, любой инфоповод начинается обсуждаться людьми и журналистами. Для соблюдения естественности динамики прироста ссылок можно использовать маскировку под вирусную активность. Т.е. публиковать ссылки в новостных ресурсах, блогах и социальных сетях.

Оптимизация статических ссылочных факторов.

Основная задача: повысить статический вес документов, не попадая под санкции. Для этого необходимо соблюдать естественность распределений и в качестве доноров отбирать документы, который передают как можно больше веса.

- Получение ссылок с низким SEOLink
- Улучшение SEOin
- Обеление динамики роста ссылок

Оптимизация динамических ссылочных факторов.

Т.к. основным атрибутом динамических ссылочных факторов является анкор ссылки. Необходимо оптимизировать анкорлист по алгоритму VM25. И учесть естественную динамику появления ссылок

- VM25 по анкору / анкор-лист
- кол-во ссылок с запросом
- время появления ссылок
- отношение всех ссылок к ссылкам с запросом

3.5.3 Хостовые

- возраст сайта (с момента первой индексации. Спамные отваливаются в течение первого года)
- ключевые слова в url (как в домене, так и в документе. У спамных - злоупотребление)
- нахождение в модерлируемых каталогах (Яка, Дmoz, Википедия. Спамных в них почти нет.)

3.5.4 Региональные

Существуют геозависимые и геонезависимые запросы.

У геозависимых сортировка результатов в выдаче зависит от региона, из которого был задан запрос. У геонезависимых нет. Если сайт принадлежит тому же региону, что выбрал пользователь (автоматически определяется по местоположению), то по геозависимым запросам сайт получают существенный бонус в ранжировании.

Задать сайту регион в Яндексе можно через панель вебмастера, яндекс.каталог, или автоматически через страницу контактов на сайте.

В нашей тематике «фокусы» большинство запросов геонезависимые, поэтому в дальнейшей работе этот фактор подробно разбираться не будет.

3.5.5 Поведенческие

Кликовые – учитывают поведение пользователя на выдаче.

Равнозначно могут считаться по документу и хосту.

Запросонезависимые:

- число показов
- числов кликов
- CTR документа\хоста
- число запросов, по которым были клики
- число уникальных посетителей за временной интервал
- число посещений за временной интервал
- среднее время просмотра
- среднее кол-во активных действий на документе

Запросозависимые:

- число показов по запросу
- число кликов по запросу

- CTR, когда документ кликается первым
- CTR, когда документы кликается последним
- CTR, когда кликается только этот документ по всем запросам
- доля кликов по документу в выдачах по запросу
- доля кликов с учетом региона

Временные:

- среднее время просмотра документа

Номер клика:

- средняя позиция
- средний номер клика
- средний номер клика с конца
- среднее число документов в выдаче перед документом, которые были кликнуты до этого
- среднее число документов в выдаче после документа, которые были кликнуты после
- число документов, стоящих в выдаче выше, но по которым кликнули позже

Вероятностные факторы:

- вероятность быть последним кликнутым документом
- вероятность клика на документ на позицию выше
- вероятность клика на документы на позицию ниже
- вероятность клика 2 раза подряд
- вероятность того, что к документу вернулись после клика ниже
- вероятность того, что после клика, были клике выше

Активность пользователей:

- число уникальных посетителей за интервал по запросу
- число посетителей за интервал по запросу
- среднее время просмотра, если перешел по запросу
- среднее кол-во активных действий. Если перешел по запросу

Зависят от запроса и выдачи:

- кол-во показов по запросу
- число кликов по выдачам запроса
- средняя позиция кликнутых документов
- среднее время работы с запросом
- средняя позиция документа, по которому кликают первым
- среднее время до первого клика

Сводные запросные факторы:

- число запросов без кликов
- доля запросов без кликов

- число запросов только с одним кликом
- доля запросов только с одним кликом
- среднее число кликов по выдаче
- среднее число выдач в сессии
- среднее число кликов в сессии
- % наиболее кликовых документов
- CTR в зависимости от места документа в выдаче
- кликовая энтропия
- среднее время до первого клика
- среднее время до последнего клика
- среднее время между двумя кликами

Полный список представлен в работе [5].

Сессионные – учитывают поведение пользователя на документе.

- длительность просмотра страницы (можно нормировать на размер документа)
- длительность просмотра сайта (средняя по всем страницам)
- количество действий
- граф переходов
- порядок документов в сессии. Последний – важно!
- тип перехода (инпут \ клик)
- время перехода
- доля перехода не по ссылкам

Способы борьбы со спамом:

- увеличение показателей факторов без изменений документа

Оптимизация кликовых факторов:

- Не заспамленный тайтл
- Лаконичный сниппет
 - Привлекательный / интересный
 - Где нужно, понуждающий к действию
- Расширения
- Социальные кнопки
- Быстрые ссылки

- Хлебные крошки
- Фавиконка
- Организации

Оптимизация сессионных факторов:

- Получение посещаемых ссылок
- Трафик из соц. сетей
- Туреin трафик
- Поведение на документах
- Улучшение навигации под продвигаемые документы
- Удержание на продвигаемом документе

3.5.6 Социальные

Возможные сигналы от соц. Сетей:

- упоминаемость в блогах, соц. Сетях
- наличие аккаунтов в соцсетях (контакт, твиттер, facebook, youtube)
- кол-во лайков

3.6 Машинное обучение

Поисковая система должна научиться строить правило, которое определяет для каждого запроса, какая страница является хорошим ответом на него, а какая — нет. Для этого поисковая машина анализирует свойства веб-страниц и поисковых запросов. У всех страниц есть какие-то признаки. Некоторые из них — статические — связаны с самой страницей, например, количество ссылок на эту страницу в интернете. Некоторые признаки — динамические — связаны одновременно с запросом и страницей — например, присутствие в тексте страницы слов запроса, их количество и расположение.

Кроме факторов ранжирования поисковой системе необходимы образцы — запросы и страницы, которые люди считают подходящими ответами на эти запросы. Оценкой того, насколько та или иная страница подходит для ответа на тот или иной запрос, занимаются специалисты — ассессоры. Они берут поисковые запросы и документы, которые поиск находит по этим запросам, и оценивают, насколько хорошо найденный документ отвечает на заданный запрос. Из запросов и хороших ответов составляется обучающая выборка. Она должна содержать самые разные запросы, причем в тех же пропорциях, в которых их задают пользователи. На обучающей выборке поисковая система устанавливает зависимость между страницами, которые ассессоры посчитали релевантными запросам, и свойствами этих страниц. После этого она может подобрать оптимальную формулу ранжирования — которая показывает релевантные запросу сайты среди первых результатов поиска. [9]

Критерии ассесорских оценок для коммерческой выдачи:

Доверие: Контакты, Телефоны, Гарантия, Информация о компании (Сертификаты и Реквизиты), Лица компании и фото офиса или склада, Примеры выполненных работ, Отзывы, Клиенты, Статьи, Грамотность текстов

Качество услуги: Цены, Сравнение, Сервисы, Время работы, Доставка

Удобство: Организация каталога, Поиск по сайту, Фак, Корзина, Сравнение и подбор

4. Аудит сайта figasebe.ru

Проведем анализ сайта figasebe.ru с учетом всех групп факторов и найдем ошибки

4.1 Анализ индексации

Анализируем Robots.txt. Код ответа сервера 200 (все правильно).

Содержание:

*User-agent: **

Host: www.figasebe.ru

Прописано зеркало с www, сайт склеен. Все страницы разрешены к индексации.

Правильно настроен 301 редирект с сайта без www На сайт с www.

Проверяем наличие спама, вирусов, порно и прочих материалов, противоречащих поисковой лицензии Яндекса. Если такой контент присутствует на сайте, ПС будет накладывать существенные штрафы на сайт (см. 3.3). Запрещенных материалов не найдено.

Сравниваем индексацию в различных поисковиках, чтобы найти возможные дубли страниц.

Индекс Яндекса (ИЯ): 336 страниц.

Индекс Гугла (ИГ): 439 страниц.

Разница в индексации есть, необходимо проверить почему. Т.к. страниц в индексе обеих поисковых систем меньше 1000, можно просто визуальное просмотреть все страницы в индексах и сравнить. В качестве помощи можно использовать программу Xenu. Получить весь список страниц и покластерно сравнить индексацию.

Анализ страниц показал, что существуют следующие типы контентных страниц:

<http://www.figasebe.ru/Trickfree.php?id=sj6> – бесплатные трюки, ИЯ 28, ИГ 6

<http://www.figasebe.ru/Trick.php?id=94> – платные трюки, ИЯ 149, ИГ 189

Особенность: в индекс ПС **УЖЕ** есть страницы вида

http://www.figasebe.ru/Trick.php?id=99&post=1185801_816 и

http://www.figasebe.ru/Trick.php?id=24&utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+figasebe%2FMZSA+%28%D0%A4%D0%B8%

[D0%B3%D0%B0%D1%81%D0%B5%D0%B1%D0%B5.%D1%80%D1%83+-+%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5+%D1%84%D0%BE%D0%BA%D1%83%D1%81%D0%B0%D0%BC%29](http://www.figasebe.ru/trening.php?tren=3&pred=1)

Которые являются дублями страницы без параметра post, utm_source, utm_campaign, utm_medium. Параметры необходимо убрать через директиву clean-param в robots.txt для яндекса и через панель вебмастера в гугле.

<http://www.figasebe.ru/trening.php?tren=3&pred=1> – тренинг, нет и индекс!

<http://www.figasebe.ru/questions.php?type=2&tren=1> – вопрос, ИЯ 1, ИГ 2

<http://www.figasebe.ru/in.php?fromtrick=91> – страница покупки, ИЯ 3, ИГ 1

<http://www.figasebe.ru/blogpost.php?id=31> – пост блога, ИЯ 36, ИГ 36

Есть еще прочие служебные страницы, которые не несут смысловой нагрузки для поисковиков, и их необходимо закрыть от индексации, например:

<http://www.figasebe.ru/invitefriend.php?ref=top> ИЯ 6, ИГ 1 (ссылка на партнерку. См. 4.3)

<http://www.figasebe.ru/inside.php?logout=1> (вход в личный кабинет) ИЯ 2, ИГ 2

<http://www.figasebe.ru/mailremind.php?smartemail=voinu%40mail.md&smartname=&smartreferer=http%3A%2F%2Ffigasebe.ru%2F> (восстановление пароля) ИЯ 7, ИГ 4

<http://www.figasebe.ru/maillsuccess.php?smartemail=paha2%40one.lv&smartname> (подписка) ИЯ 1, ИГ 1

Так же к таким страницам относятся registration.php, login.php, indexshop.php, bonusshow.php (просмотр приводит к бану по IP на 1 час)

Анализ индекса гугла также выявил кластер <http://www.figasebe.ru/digishop/>, где расположен сам магазин по продаже фокусов, реализованный через систему digiseller. Это служебный раздел, с множеством возможных дублей, поэтому его желательно закрыть от индексации. На данный момент ИЯ 20, ИГ 17.

Кластер <http://www.figasebe.ru/oplata/> и <http://www.figasebe.ru/sms/> используемые для оплаты через СМС. И кластер /konkurs/ с конкурсами.

Были найдены страницы вида *.shtml, на которых слегка другой дизайн. Возможно, это старая версия сайта, ее необходимо закрыть.

На сайте существует множество параметрических дублей. Доступных как по ошибке программистов, так и просто не отличающиеся контентом. **ВСЕ** смысловые страницы обладают параметрическими дублями.

Примеры:

<http://www.figasebe.ru/buydvd.php?dvd=1>
<http://www.figasebe.ru/buydvd.php?dvd=1&a=b>

<http://www.figasebe.ru/questions.php?type=2&tren=1>
<http://www.figasebe.ru/questions.php?type=2&tren=2>

Так же неправильно обрабатываются несуществующие страницы. Вместо 404 ошибки возвращают код ответа 200.

Пример:

<http://www.figasebe.ru/buydvd.php?dvd=1000>

У сайта есть магазин реквизита для фокусов, расположенный на поддомене: shop.figasebe.ru. Его анализ не производился.

Резюме:

На сайте существует много служебных страниц, много существующих параметрических дублей и еще больше потенциальных дублей. Эти страницы перетаскивают на себя часть краулингово бюджета и статического веса. Для исправления этой ситуации необходимо закрыть эти страницы от индексации.

Анализ html кода:

После анализа html-код страниц сайта были найдены следующие ошибки:

- 1) У некоторых страниц отсутствуют обязательные закрывающие теги (на главной body и html). Это усложняет разбор страницы поисковым роботом.
Необходимо закрыть эти теги.
- 2) На страницах присутствует по несколько тегов h1. Теги заголовков пересекаются с другими тегами. Это уменьшает релевантность страницы по запросам.
Необходимо оставить только 1 заголовок первого уровня.

- 3) На другие страницы сайта прописаны относительные ссылки. При случайном изменении структуры сайта или навигации, относительные ссылки сбьют навигацию.
Необходимо изменить ссылки на абсолютные.
- 4) Большие куски javascript кода размещаются напрямую в странице. Это увеличивает размер каждого документа и тормозит загрузку сайта. Все повторяющиеся куски js кода необходимо вынести в отдельные файлы.
- 5) В текстах, которые присутствуют на сайте, очень много лишних тегов: b и font. Усложняет разбор страницы и увеличивает вес.
- 6) На страницах присутствуют одинаковые блоки (шапка, футер, правая колонка с баннерами, регистрацией и ссылками) . Это уменьшает уникальность каждой страницы в рамках сайта. Чтобы этого избежать, желательно подгружать эти блоки при помощи скрытого через robots.txt JS

Т.к. основное содержание сайта – это видео ролики, имеет смысл улучшить поиск по видео. Для этого у ПС существуют инструменты под названием микроформаты. Каждое видео можно специальным образом описать через html теги, чтобы поисковой машине было его легче находить и быстрее определять его смысл. [9]

Так же Google умеет показывать скриншот ролика на странице выдачи. Это часто увеличивает CTR. Для части страниц с сайта figasebe.ru скриншоты отображаются, а для части нет. Имеет смысл определить принцип, по которому происходит выделение ролика и добиться, чтобы у всех страниц он показывался.

<p>Люберцы, Московская область Изменить место</p>	<p>Фигасебе.ру - фокус Влияние - Фокусы с деньгами, с картами, с ... www.figasebe.ru/Trick.php?id=105 Фокус Влияние. Идеальное предсказание с листком газеты. Получается в 100% случаев, можно показывать несколько раз подряд. Этот трюк не должны ...</p>
<p>Любая длительность Короткие (0–4 мин.) Средние (4–20 мин.) Длинные (от 20 мин.)</p>	<p>Фигасебе.ру - фокус Триумф - Фокусы с деньгами, с картами, с ... www.figasebe.ru/Trick.php?id=104 2 май 2012 - 3 мин. Фокус Триумф. Зритель мешает колоду карт, выбирает одну и подписывает. Вы хаотично смешиваете карты лицом вниз/вверх, а в итоге всё ...</p>
<p>Весь Интернет Только на английском и русском</p>	<p>Фигасебе.ру - фокус Призрачная монета - Фокусы с деньгами, с ... www.figasebe.ru/Trick.php?id=96 26 мар 2012 - 1 мин. Фокус Призрачная монета. Монета исчезает в ваших руках, чтобы появиться внутри колоды и указать на карту зрителя. Фигасебе.ру - это лучшие видео ...</p>

Рис. 4.1 Скриншоты видео в выдаче Гугла

4.2 Анализ текстовых факторов

1) Заголовки созданы по шаблонам:

Страница фокуса: «Фигасебе.ру – НАЗВАНИЕ_ФОКУСА - Фокусы с деньгами, с картами, с сигаретами, романтические фокусы, обучение фокусу фокус видео»

Страница категории: «Фигасебе.ру - НАЗВАНИЕ_КАТЕГОРИИ | обучение фокусам, секреты фокусов, карточные фокусы, видео обучение, фокусы бесплатно»

Проблема: длинные заголовки, большая часть которых повторяется. Необходимо переписать заголовки в соответствии с семантикой. Заголовок на каждой странице должен быть уникальный, лаконичный и привлекать внимание. Повторяющиеся слова (например, название сайта), необходимо переместить в конец. Это улучшает ранжирование в ПС и необходимо для удобно навигации, если открыто несколько вкладок.

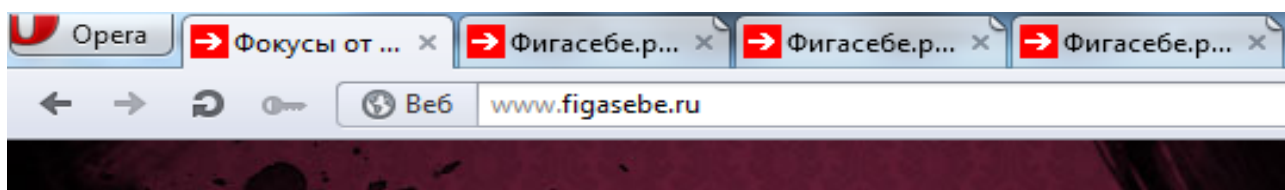


Рис. 4.1 Малоинформативные заголовки сайта

2) Метатег description для всех категорий одинаковый.

Это уменьшает уникальность каждой страницы и не дает преимущества в кликовых факторах в выдаче. Необходимо сделать уникальное, бросающееся в глаза описание, чтобы привлечь внимание в выдаче.

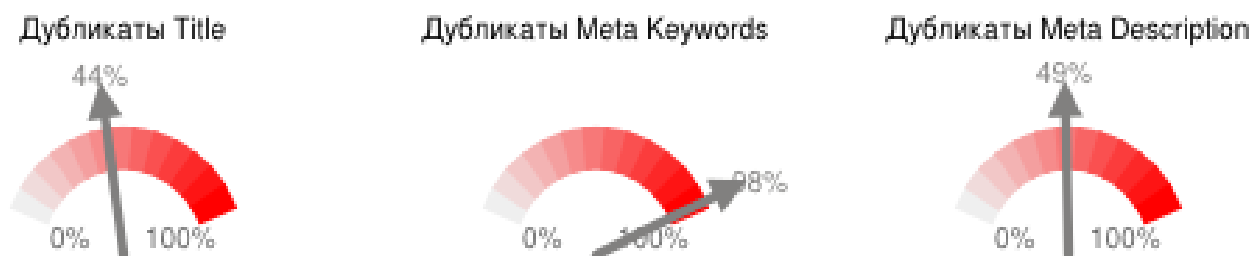


Рис. 4.1 Дубликаты заголовков и описания по сервису сайтрепорт.рф

3) На сайте существует всего 3 страницы, подходящие для привлечения трафика из поисковых систем: главная (запросы обучения фокусам), бесплатные фокусы (бесплатное обучение фокуса) и магазин (магазин реквизита).

Для увеличения трафика с поисковых систем необходимо создать страницы под структуру, описанную в разделе 2.1. Эта структура достаточно хорошо подходит к существующему сайту, нужно только увеличивать количество материалов.

Особенностью проекта является то, что основное содержание сайта – это не тексты, а видео-ролики. Т.к. поисковик не умеет распознавать образы на видео, а ищет в основном по тексту, то для этих страниц необходимо составить текстовые описания. Это можно сделать по следующему принципу:

Для страниц карточек фокусов – транскрибация видео. Если такой способ окажется слишком затратным, можно попробовать генерировать описания по шаблонам. Для этого необходимо выделить возможные свойства фокусом Например:

Сложность: простой, средний, сложный
 Сколько времени требует на изучении: несколько минут, часов, дней
 С каким предметом: монетка, сигарета, карты, веревки ...
 Где показывать: на улице, на вечеринке, на представлении
 На кого рассчитан: для детей, для девушек, для группы людей,
 Сколько длится по времени: несколько секунд, несколько минут
 Производящее впечатление: удивляет, шокирует, настораживает,
 срывает крышу
 Кем придуман: ФИО
 Кто показывает: ФИО

Далее эти характеристики последовательно в разных комбинация подставляются в предложения и составляют текста. Подробнее методика описано в [11].

Для страниц категории достаточно 2-3х предложений о разделе. Остальную текстовую релевантность дадут краткие описания фокусов.

4.3 Ссылочные факторы

Внутреннее ссылочное:

1) На сайте figasebe.ru отсутствует внутренняя ссылочная оптимизация. Анализ сайта при помощи программ PageWeight показал, что вес равномерно распределен между категориями и служебными страницами сайта.

Для повышения эффективности продвижения, статический вес необходимо сосредоточить только на продвигаемых страницах (категориях фокусов и главной странице).

Для этого можно закрыть существующие меню от поисковиков при помощи скрытого от поисковиков JS. И разместить дополнительные статические блоки ссылок. Чтобы повысить текстовую релевантность анкорлиста, ссылки можно делать разными словами.

Если владельцы ресурса хотят оставить текущее меню, то блок seo-ссылок можно сделать появляющимся по клику на слове «еще».

2) На сайте размещены внешние сквозные ссылки на ресурсы Рамбле ТОП 100, Вебмани Сток. Этим ссылкам необходимо прописать атрибут nofollow, чтобы сохранить статический вес в рамках сайта.

3) Существует большое количество ссылок на страницы, которые не отличаются содержанием для анонимных пользователей. Например при клике на ссылку «оплата трюка», происходит перенаправление на страницу оплаты с номером трюка. Но на этой странице находится только одно окно регистрации. Такие ссылки необходимо закрыть в robots.txt

Внешнее ссылочное:

Для анализа внешних ссылок сделаем выгрузку из сервиса Solomono. И посмотрим возможные распределения.

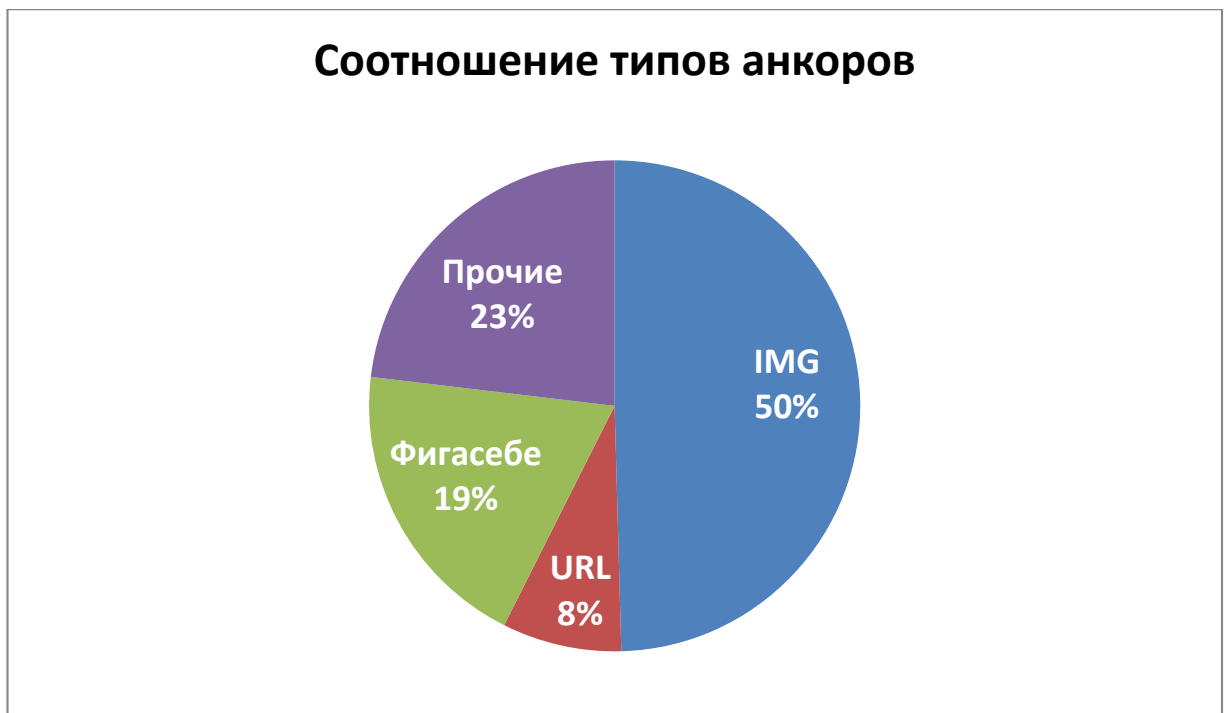
Всего на сайт проставлено 3905 ссылок с 194 сайтов.



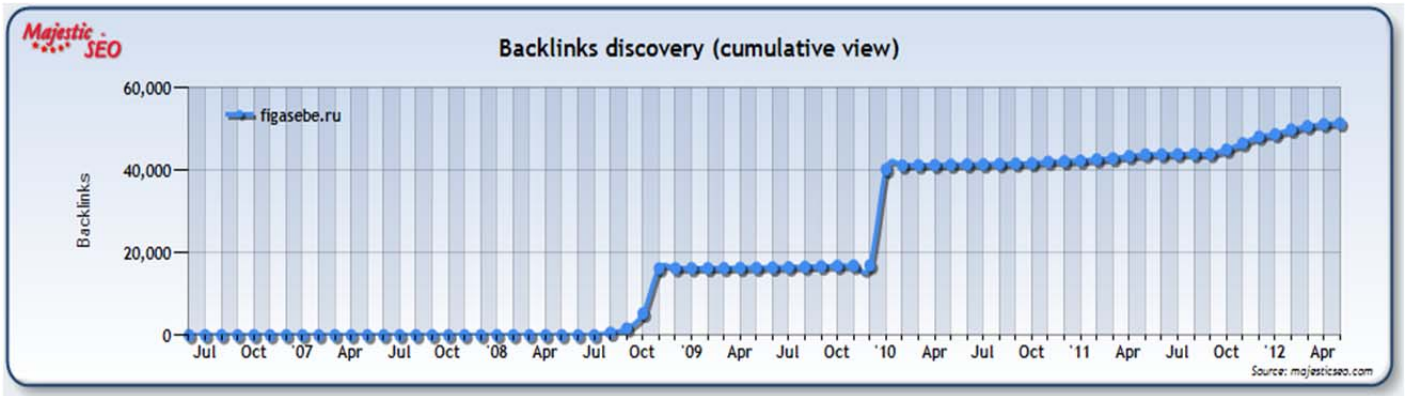
Граф. 1. Распределение количества ссылок по длине слов в анкоре



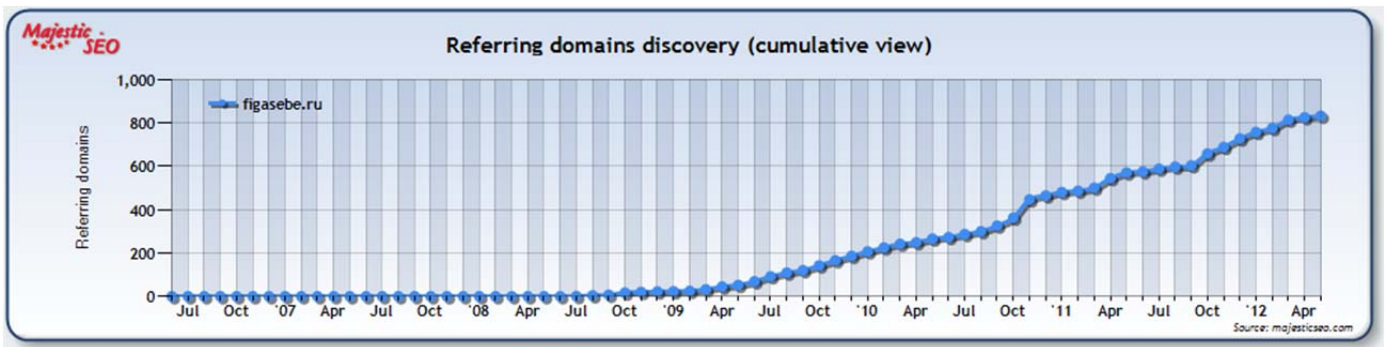
Граф.2. Распределение количества ссылок по ТИЦ



Граф.3 Соотношение типов анкоров



Граф. 4 График роста кол-ва внешних ссылок по времени

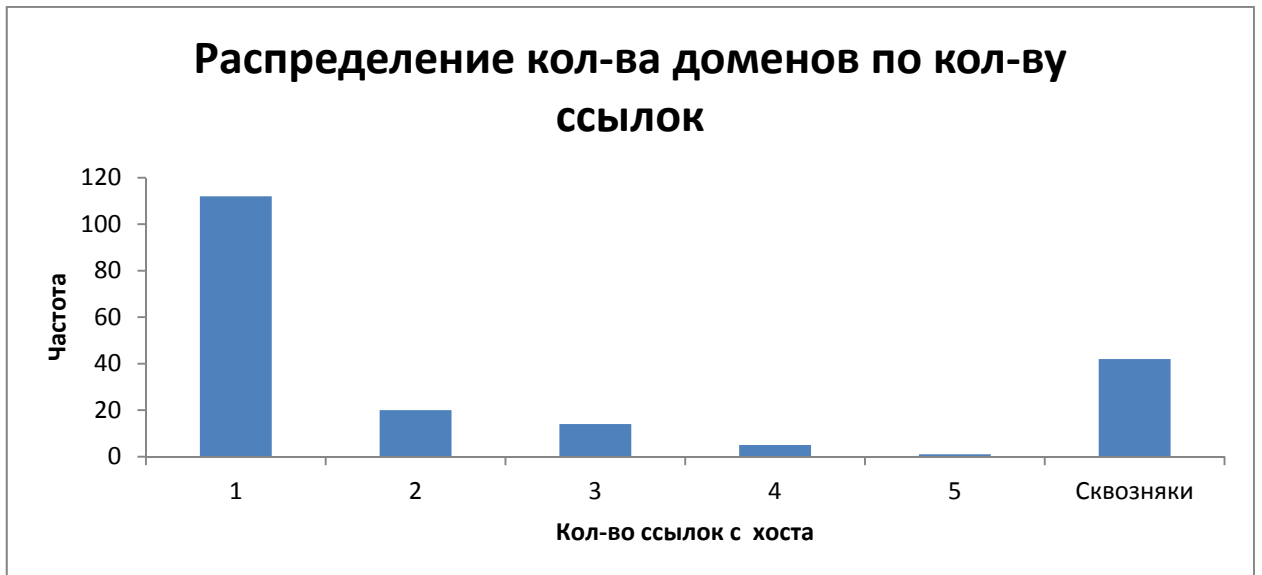


Граф. 5. График роста кол-ва ссылающих доменов по времени

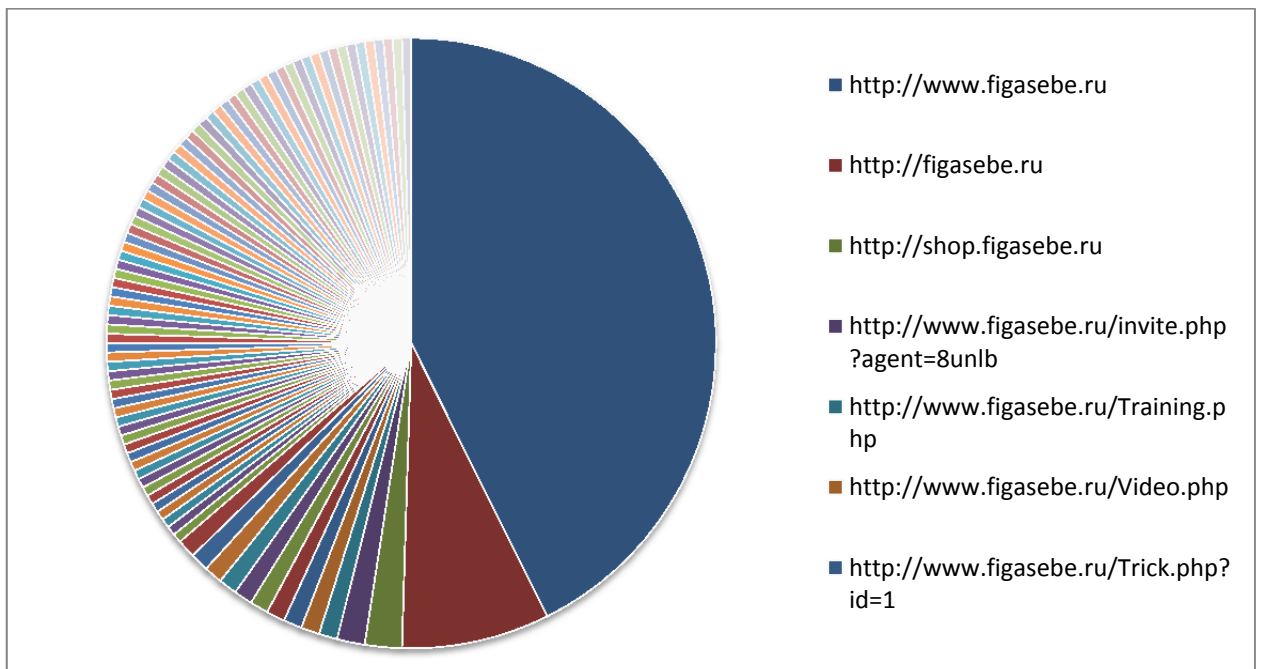
На графика количества внешних ссылок есть скачки. На графике роста количества доменов плавный рост. Можно сделать вывод, что на части доменов проставлены сквозные ссылки.



Граф. 6 Распределение количества ссылок по уровням вложенности страниц-доноров



Граф. 7 Распределение кол-ва доменов по кол-ву ссылок



Граф. 8 Распределение ссылок по входным страницам

Распределение по времени появления ссылок относительно времени появления документа не представляется возможным, т.к. нет подходящего инструмента.

Оценим SEOtext анкоров. В качестве сео-слов будет использовать запросные слова: бесплатно, огонь, монеты, карты, деньги, сигареты. Отношение естественных анкоров к покупным ~ 0.09

Оценим SEOout доноров по параметру

avgLinkOut = отношению кол-ва исходящих к проиндексированным страницам.

avgLinkOut = 1,26

Все распределения похожи на распределения естественного вида. Важно сохранить эту естественность при дальнейшем продвижении.

Примеры неестественного продвижения см. в п.5

Теперь составим стратегию продвижения и определим конкретные способы оптимизации ссылочных факторов.

Исходя из п. 3.5.2 ясно, что ссылочные факторы бывают статическими и динамическими.

Оптимизировать *динамически факторы* можно достаточно быстро и просто – нужно купить ссылок с анкором в виде необходимых запросов. Т.к. это будут явно SEO-ссылки, то имеет смысл их покупать более дешевыми, чтобы получить меньшую потерю статического веса.

Подробнее остановимся на способах получения *статического веса*.

У сайта существует особенность: партнерская программа. Она предоставляет вебмастерам ссылки вида <http://figasebe.ru/invite.php?agent=CODE>. Эта страница не содержит полезного и уникального контента, однако возвращает код ответа 200. Это приводит к тому, что скапливается статический вес на страницах, которые не привлекают поисковый трафик.

Правильная реализация – поставить 301 редирект на главную страницу, а код агента записывать в куки. Таким образом можно сохранить ссылочный вес и определять партнера, привлекающего трафик.

Партнерская программа – хороший пример привлечения естественного статического веса.

Так же хорошим способом является ручная простановка ссылок на сайтах партнеров и непродávующих ссылки площадках.

Рассмотрим способы покупки ссылок на существующих биржах. Не секрет, что в них существует достаточно большой процент спама и сайтов созданных только для продажи ссылок.

Наша задача – получить ссылки, которые передают большой статический вес. Это могут быть либо сайта в 2х случая: мало исходящих ссылок и/или много входящих. Так же необходимо помнить про алгоритм антиспама, который занижает передач статического веса

... тут еще кусок на 2-3 страницы ...

4.4 Хостовые факторы

Дата регистрации домена 2008.03.12, сайту более 4х лет. Возрастной фактор учитывается почти полностью. (По мнению А.Калинина, после 5 лет санкций за возраст уже нет).

На сайте отсутствует ЧПУ. Желательно сделать.

Сайт зарегистрирован в яндекс.каталоге, что может давать дополнительный бонус при ранжировании.

4.5 Региональные факторы

Т.к. запросы геонезависимые, то данный фактор не влияет на качество продвижения.

4.6 Поведенческие факторы

Для улучшения CTR в выдаче необходимо поменять заголовки и описания страниц (см. пункт 4.2).

Так же необходимо прописать адрес организации и привязать группы из социальных сетей. Это увеличит площадь, занимаемую сайтом в выдаче и увеличит CTR.

Сравнение с конкурентами по сервису Алекса показал, что у проекта уже достаточно хорошие поведенческие показатели. Возможно этого добились благодаря уже существующим способам удержания посетителя на сайте:

- видео нельзя промотать вперед
- при закрытии страницы возникает всплывающее окно
- при заходе на сайт всплывает окно с предложением подписаться на рассылку

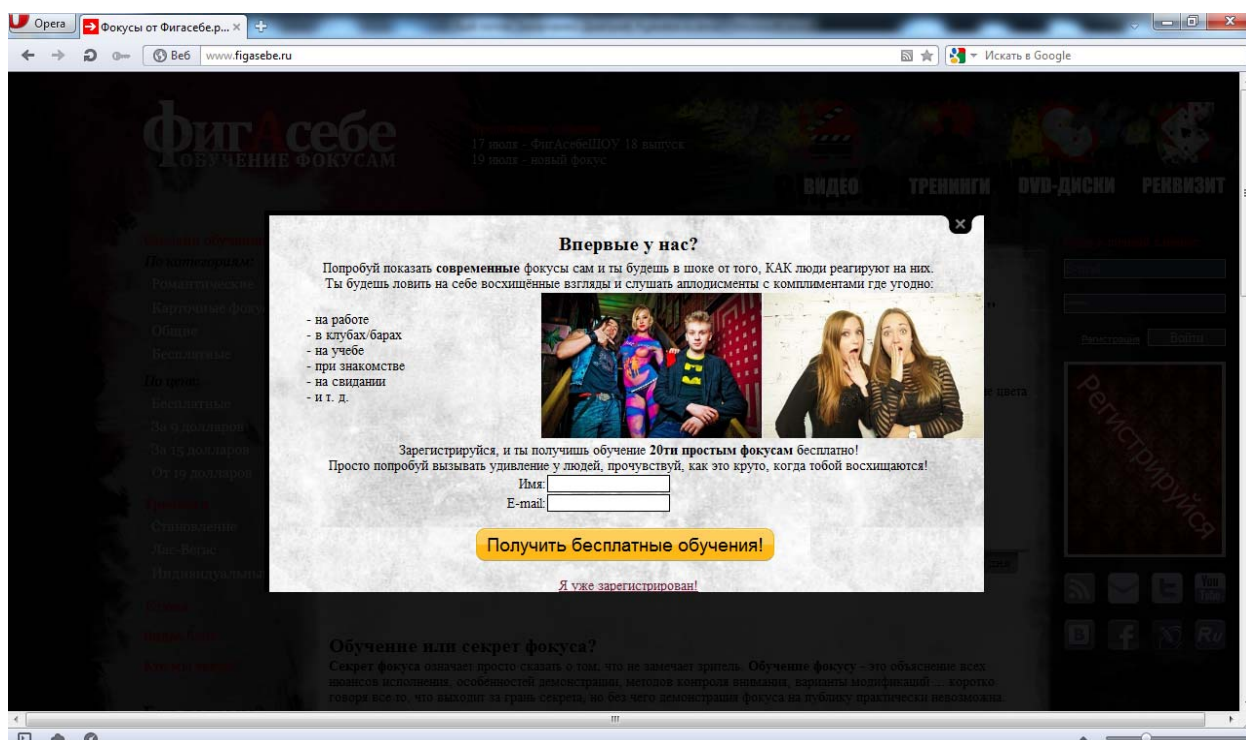


Рис. 4.2 Окно регистрации

Правильность этого окна необходимо проверить на a/b тесте. Есть предположение, что оно существенно увеличивает количество отказов.

Для улучшения поведенческих показателей сайта можно использовать различные «завлекалки». Например:

- Найти спрятанный фокус на сайте (увеличивает кол-во просмотров)
- Собери пазл и узнай секрет фокуса (увеличивает время просмотра и кол-во активных действий)

Так же необходимо подправить верстку: на текущий момент сайт оптимизирован под «странный» размер экрана. При разрешении 1280x800 присутствует горизонтальная полоса прокрутки.

4.7 Социальные факторы

У сайта есть группы в контакте и фейсбуке. Группы вконтакте имеет ~15000 подписчиков. Свой канал на youtube. Большая часть трафика приходит через эти социальные ресурсы. Так же имеется блог в жж и rutube, но последнее обновление на них было в 2010 г. Если этими каналами не занимаются, то лучше убрать на них ссылку

#	Month	Views	AvgViewMonth	good	bad	likeRank
1	10,00	427669	42767	2878	500	0,852
2	10,00	319221	31922	2099	136	0,939
3	10,00	405893	40589	3535	126	0,966
4	9,00	335703	37300	10293	346	0,967
5	9,00	300538	33393	6000	224	0,964
6	8,00	235305	29413	2628	199	0,93
7	8,00	204383	25548	2294	203	0,919
8	7,00	203281	29040	9788	204	0,98
9	9,00	178921	19880	2073	338	0,86
10	8,00	221921	27740	6422	230	0,965
11	5,00	211933	42387	8256	302	0,965
12	4,00	360865	90216	12381	348	0,973
13	3,00	35738	11913	9	1	0,9
14	2,00	123764	61882	9	1	0,9
15	1,00	118037	118037	3821	206	0,949
16	0,50	57338	114676	9	1	0,9
17	0,25	89745	358980	5361	141	0,974
18	0,10	47000	470000	2178	73	0,968
AVG		215403	88094			0,937

Табл. 5. Статистика просмотров видеороликов с фокусами на youtube

Анализ статистики просмотров показал, что большая часть просмотров видео приходится на сайт вконтакте и карамбатв. Т.к. ФигаСебе является проектом студии КарамбаТВ, поэтому он тесно интегрирован с другими крайне популярными проектами: +100500 и некоторыми другими.

Резюмируя, социальные факторы «подняты до небес». Этот большой потенциал может существенно повлиять на поисковое продвижение.

Так же, на страницах используются кнопки «поделиться с друзьями». В качестве эксперимента можно попробовать поменять положение этих кнопок.

5. Сравнение с конкурентами

Сравнение посещаемости по alexa.com

	Pageviews/User	Bounce %	Time on Site	Alexa Traffic Rank	in ru	Reputation
what-focus.ru	1,10		1:54	3,473,722		98,00
goodmagic.ru	3,00	26.7%	3:51	853,873	67,599	52,00
micromagic.ru	4,50	22.9%	3:45	295,947	20,131	88,00
magic-class.ru	3,40	37.6%	4:00	223,545	16,753	212,00
figasebe.ru	4,40	37.7%	4:38	170,878	6,684	93,00

Табл. 3. Сравнение посещаемости конкурентов по alexa.com

5.1 Анализ сайта goodmagic.ru

Ошибки:

- в листингах фокусов несколько h1. H1 содержит ссылку на страницу фокуса.
- на карточке фокуса h1 = ссылке на текущую страницу.

Хорошие решения.

ЧПУ, лаконичные title.

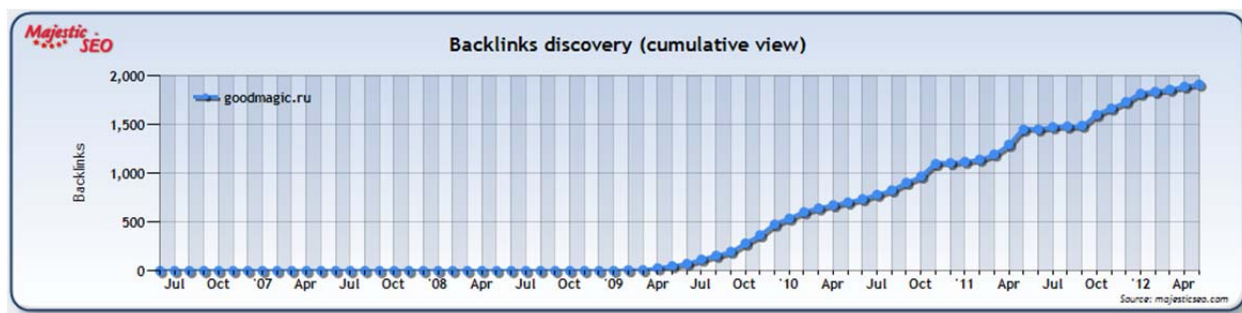
Внутренняя перелинковка: анализ сайта через pageweight, показал, что основные веса имеют категории фокусов и несколько конкретных страниц. Все сделано правильно.

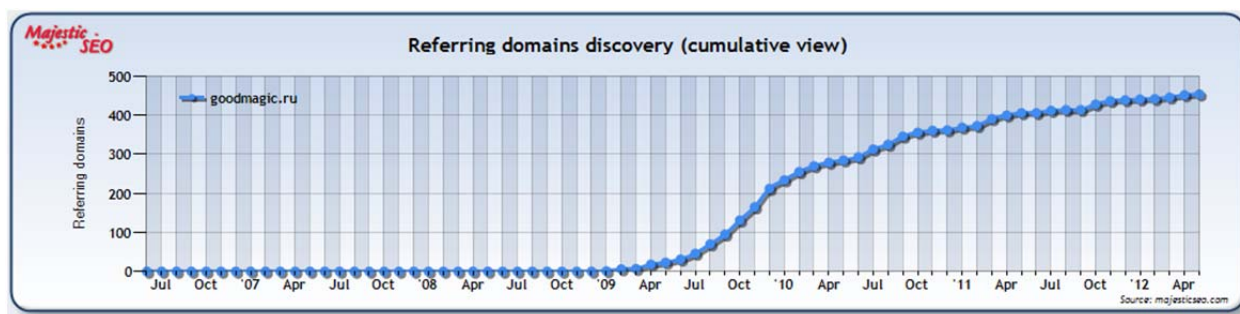
Индекс яндекс 998, гугл 1000.

Внешние ссылки: 761 с 88 доноров. Тиц 20, пр2.

Посещаемость по LI ~1700 человек в день.

Вывод: сайт хорошо оптимизирован по внутреннему ссылочному. Нет проблем с индексацией.





5.2 Анализ what-focus.ru

Title в норме. Пустые h1 на всех страницах.

Внутренняя перелинковка не оптимизирована для поискового трафика. Вес собран на категориях, служебных страницах, листингах постов по месяцам и нескольких карточках фокусов.

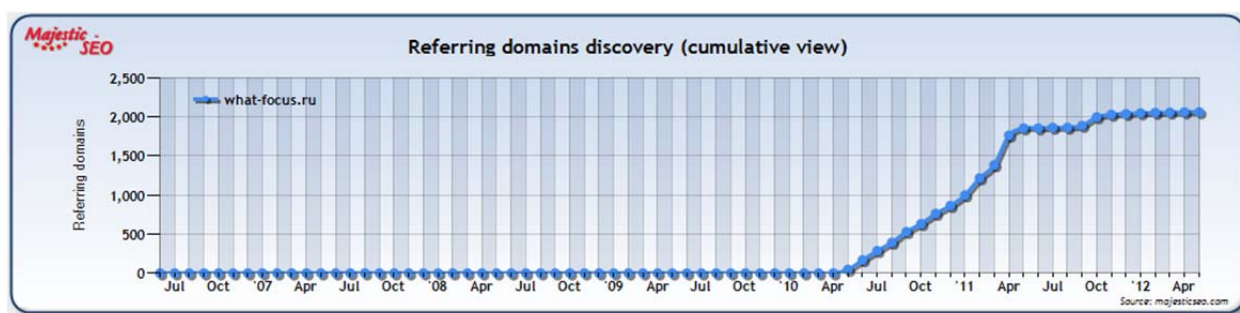
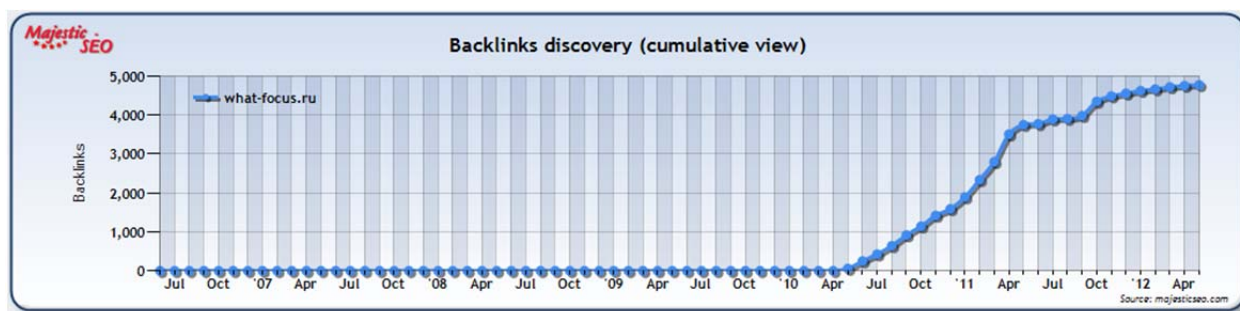
Индекс яндекса 911, гугл 2300 стр.

Внешние ссылки: 1511 с 770 доноров. Тиц 10, пр0.

На сайте не установлены системы статистики – нет возможности проверить посещаемость.

Хорошая видимость по Мегainдексу не совпала с хорошими показателями посещаемости через alexa. На этот сайт не стоит ориентироваться как на конкурента.

Проверим динамику прироста ссылок:



Достаточно резкая. Есть подозрения, что ссылки покупаются. Проверим естественность других параметров.



Как и предполагалось, распределение смещено в сторону увеличения кол-ва слов.

Оценим SEOtext анкоров. В качестве сео-слов будет использовать запросные слова: бесплатно, огонь, монеты, карты, деньги, сигареты. Отношение естественных анкоров к покупным **~0.2**

Оценим SEOout доноров по параметру

avgLinkOut = отношению кол-ва исходящих к проиндексированным страницам. **avgLinkOut = 1,26**

5.3 Анализ micromagic.ru

Несколько h1; h1 содержит большие блоки html .

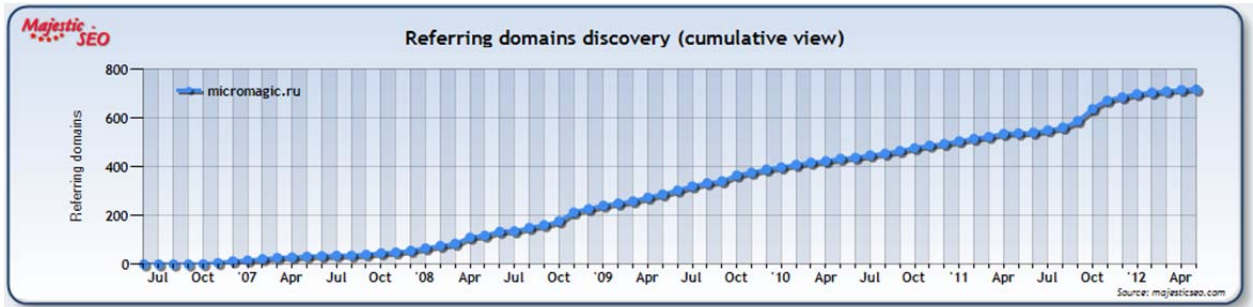
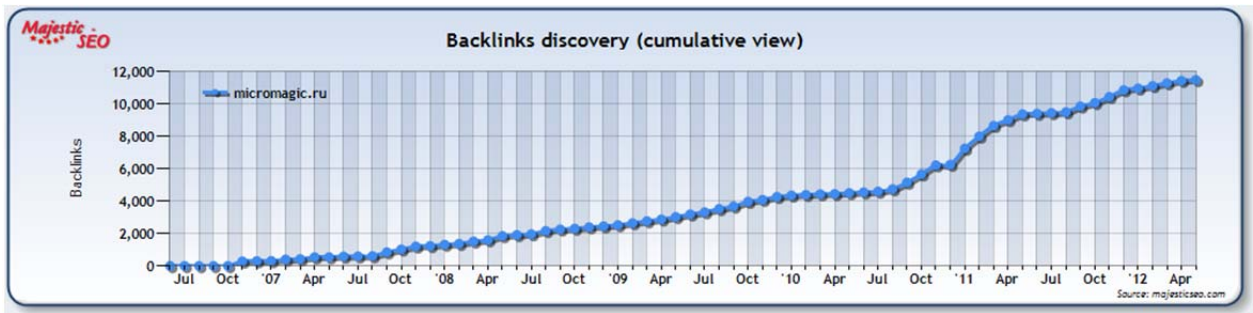
Нет ЧПУ.

Сделан на джумле: очень много дублей страниц. В индексе яндекса 19000, гугла 50000. Большие проблемы с индексацией. Не рационально используется потенциал внутренней перелинковки.

Плюсы – очень большой и насыщенный форум. Проблемы для форумов стандартные.

Внешние ссылки: 1989 с 158 доноров. Тиц 50, пр 3.

Посещаемость mail.ru: ~1300 человек в день



5.4 Анализ magic-class.ru

Сайт по каким-то причинам не парсится программами page-weight, хену и sitetracer.

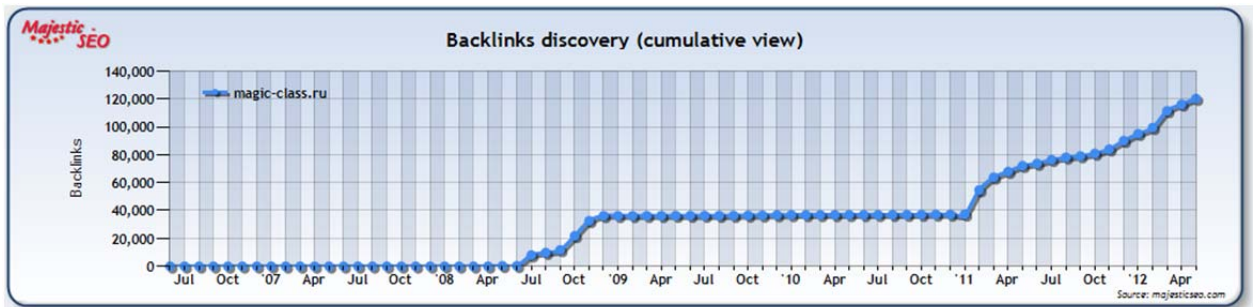
Содержит множество конструкций вида «<h1>#{title}</h1>» на одной странице.

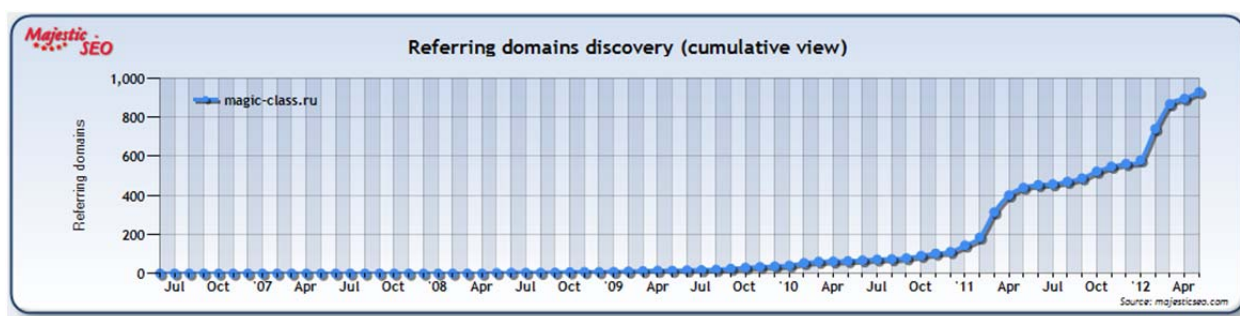
Индекс яндекса 50000, гугл 39000.

Внешних ссылок: 18860 с 502 доноров. Тиц 50, пр 3.

Посещаемость по LI ~3500 человек в день.

Проверим динамику появления ссылок.





Достаточно резкая. Есть подозрения, что ссылки покупаются. Проверим естественность других параметров.



Как и предполагалось, распределение смещено в сторону увеличения кол-ва слов.

Оценим SEOtext анкоров. В качестве сео-слов будет использовать запросные слова: бесплатно, огонь, монеты, карты, деньги, сигареты. Отношение естественных анкоров к покупным ~ 0.19

Оценим SEOout доноров по параметру

avgLinkOut = отношению кол-ва исходящих к проиндексированным страницам. **avgLinkOut = 0.95**

5.5 Резюме

Анализ конкурентов показал, что в тематике нет хорошо оптимизированных сайтов. Есть 1 качественный форум и 3 сайта, которые продвигаются при помощи ссылок, 2 из которых делают это с недостаточной эффективностью. И только на одном сайте правильно подобрана структура и сделана внутренняя перелинковка.

Исходя из этих данных, можно сделать вывод, что конкуренция низкая.

Чтобы занять высокие позиции достаточно будет хорошей внутренней оптимизации и небольшой внешней ссылочной массы.

6. Развитие сайта figasebe.ru

6.1 План работ

При внесении изменений на сайт необходимо учитывать скорость учета влияния и важность факторов.

		Объект	Важность	Время
текстовые	вхожд. важные зоны	запрос-документ	критично	быстро
текстовые	вхождения текст	запрос-документ	важно	быстро
текстовые	уникальность	документ	полезно	быстро
url-вые	ключ в url	запрос-документ	полезно	быстро
url-вые	ключ в домен	запрос-документ	-	-
url-вые	ЯК, Дмоз	host	полезно	-
ссылочные	вхождения анкор-лист	запрос-документ	критично	быстро
ссылочные	динамика появления	host/документ	важно	долго
ссылочные	стат. вес	документ	критично	долго
ссылочные	стат. вес	host	важно	долго
ссылочные	естественность	host/документ	важно	долго
поведенческие	ctr в выдаче	запрос-документ	критично	долго
поведенческие	возвраты в выдачу	запрос-документ	критично	долго
поведенческие	browserank	host/документ	важно	долго
поведенческие	повед. ранки сайт	host/документ	важно	долго
социальные	активность	host	полезно	долго
социальные	лайки	host/документ	полезно	долго
региональные	регион	host	критично	быстро
региональные	организация	host	критично	быстро
региональные	текст, ссылки	host	полезно	быстро

Табл. 6.1 Важность и время учета факторов

		Объект	Насыщение	Время
текстовые	вхожд. важные зоны	запрос-документ	быстро	быстро
текстовые	вхождения текст	запрос-документ	довольно быстро	быстро
ссылочные	вхождения анкор-лист	запрос-документ	довольно быстро	быстро
ссылочные	стат. вес	документ	долго	долго
ссылочные	стат. вес	host	долго	долго
ссылочные	естественность	host/документ	долго	долго
поведенческие	ctr в выдаче	запрос-документ	относительно быстро	долго
поведенческие	browserank	host/документ	долго	долго
поведенческие	повед. ранки сайт	host/документ	долго	долго
социальные	активность	host	долго	долго
социальные	лайки	host/документ	долго	долго
региональные	регион	host	супер быстро	быстро
региональные	организация	host	супер быстро	быстро

Табл. 6.2 Пределы оптимизации факторов

	1 этап	2 этап	3 этап	далее
Настройка	запрет индексации системных страниц, оптимизация кода, seo-hide	-		
Текстовая оптимизация	зоны title, h1	генерация текстов для фокусов, наполнение согласно структуре	оптимизация description	
Внутреннее ссылочное	распределение веса на категории			
Внешнее ссылочное	оптимизация текстовых	наращивание статического веса		
Региональные	не требуется			
Хостовые	настройка ЧПУ	-		
Поведенческие кликовые	-	улучшение CTR в выдаче и LastClick		
Поведенческие сессионные	постоянная работа над улучшением поведения пользователя			

Табл. 6.3 План работ над сайтом figasebe.ru

Временная оценка:

1 этап – 10 часов

2 этап – 1-2 месяца

3 этап – 2-3 месяца

Далее ∞

Оценка ссылочного бюджета через агрегаторы не дает результата, т.к. тематика некоммерческая. Предположительно, будет достаточно нескольких тысяч рублей.

6.3 Перспективы

После расширения сайт в соответствии со структурой, предложенной в п. 2.1 возможные следующее развитие сайта:

- форум фокусников
- обучающих видео
- трюки и слейтинг, жонглирование, другие трюки

7. Заключение

Сайт ФигаСебе предоставляет действительно востребованный и интересный контент.

Если говорить на языке ассессорской оценки, figasebe.ru витальный сайт в тематике «фокусы». Именно о таких сайтах говорит Яндекс: «Делайте качественные сайты, создавайте уникальные сервисы и он займет достойное место в поиске».

Владельцы сайта уже решили сложную задачу по привлечению и удержанию посетителей на сайте. Занять позиции в поисковой системе ему мешает отсутствие внутренней оптимизации и низкие ссылочные показатели. Эти недостатки устраняются достаточно быстро.

После оптимизации сайт займет первые позиции в поисковой системе, владельцы будут получать трафик и зарабатывать деньги, посетители будут удовлетворять потребности, а Яндекс улучшит качество поиска.

8. Список использованной литературы

1. <http://www.google.ru/intl/ru/help/features.html>
2. <http://company.yandex.ru/technologies/search/>
3. Введение в информационный поиск, презентация, А.Калинин
4. Введение в информационный поиск, Кристофер Д. Маннинг, Прабхакар Рагхаван, Хайнрих Шютце
5. Поведенческие (кликочные) факторы в информационном поиске, <http://www.altertrader.com/publications36.html>,
6. Detecting Spam Web Pages through Content Analysis, Alexandros Ntoulas, Marc Najork, Mark Manasse, Dennis Fetterly, 2006
7. Лексическая статистика в оценке качества коммерческих текстов, Ирина Борисова (Wikimart), 2012
8. «Мадридский доклад Яндекса» <http://events-old.yandex.ru/public/conferences/www2009.xml>,
9. <http://shema.org>
10. Журнал SEODigetst. Выпуск 16, стр 66-67, Л. Гороховский
11. Журнал SEODigetst. Выпуск 16, стр 64-65, А. Чекушин