

technology Mar 3

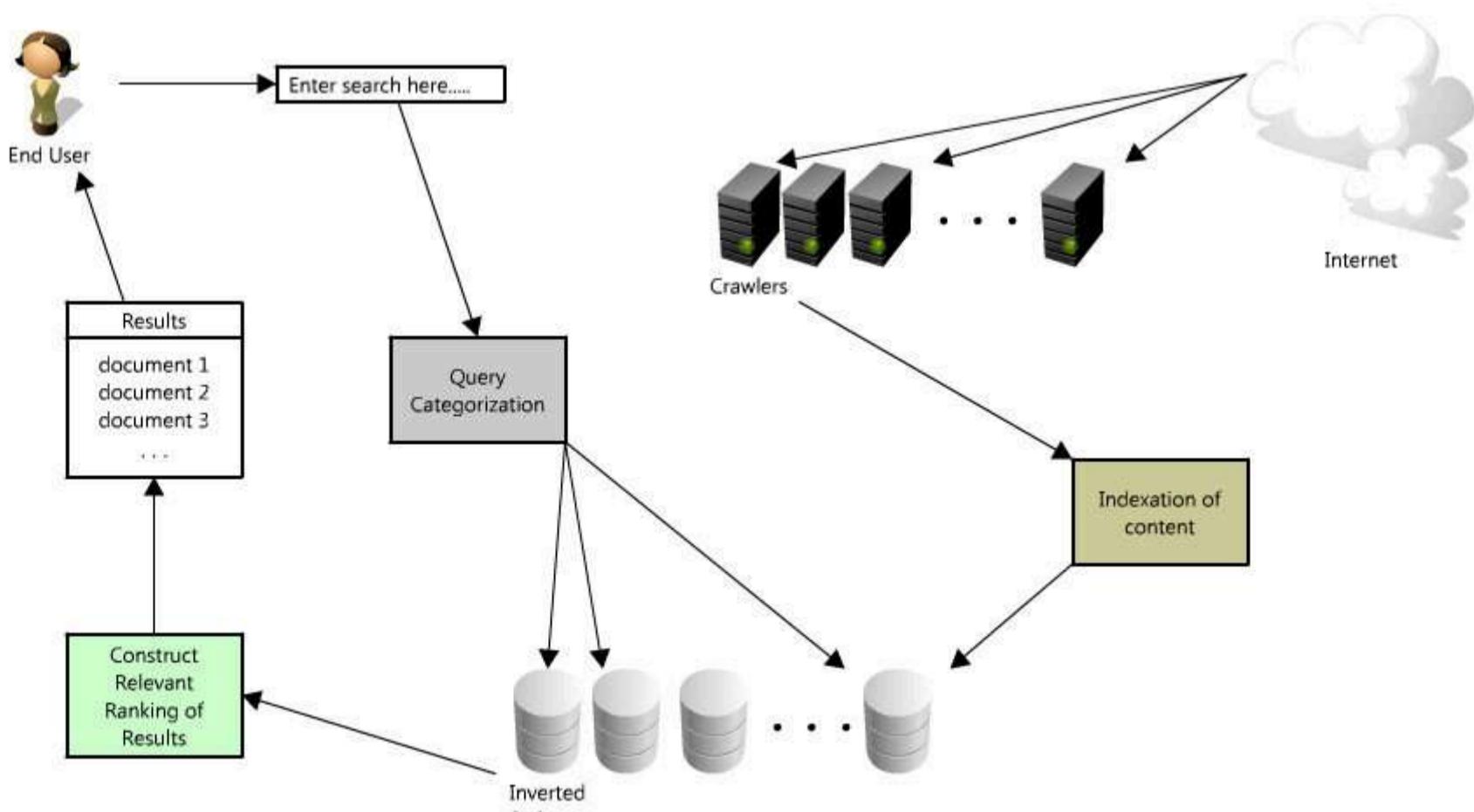
What Every Programmer Should Know About SEO

Over the last few years since I started paying attention to SEO I have noticed many "developer guides to SEO". Largely, these guides are written by developers and not SEO experts. At this point, I am neither of those, but since I spent the last few years working for an SEO tools company, I managed to garner quite a bit of knowledge on the topic. Of course, the things I love about SEO, are the things I love about search - big data, fast performance, and interesting algorithms. Below I have compiled the minimum of amount of SEO knowledge a developer should have to properly optimize a website for search engine discovery (ask for help [website developers in Vegas](#)). I largely wrote this for my team and couple of other people who have asked, but hopefully you will find it useful. Oh and special thanks on this post goes to Bryce Howard - I am very lucky to have someone so brilliant to help me edit.-----In order to understand SEO, I think it helps to start with the basics:

How do search engines work?

If you have the time, Google's architecture paper makes for great reading: <http://infolab.stanford.edu/~backrub/google.html> It is an older, and somewhat dated paper, but very approachable to even the non-whitepaper readers among us :) And they definitely explain concepts in more detail than I only mention in passing. So what happens?

1. User inputs a query
2. Query is categorized (this is a nontrivial information retrieval problem - since the category really matters - you can read a little bit more here: http://en.wikipedia.org/wiki/Web_query_classification)
3. Scans over the collection of documents that are organized similarly to inverted indexes (http://en.wikipedia.org/wiki/Inverted_index) to match the query to a set of relevant documents
4. Results are returned to the user, sorted by relevancy (with relevancy also being another hard problem - [http://en.wikipedia.org/wiki/Relevance_\(information_retrieval\)](http://en.wikipedia.org/wiki/Relevance_(information_retrieval)))



1. A web crawler has to crawl your site (and every other site on the web)
2. Indexes all the pages (meaning that a search engine needs to be able to "parse" and store the contents of every web page on the Internet)
3. Organizes the indexed collection and stores alongside corresponding metadata to provide extremely fast and relevant search results (and there are two interesting problems in this step - document classification (http://en.wikipedia.org/wiki/Document_classification) and how to create a performant search engine index ([http://en.wikipedia.org/wiki/Index_\(search_engine\)](http://en.wikipedia.org/wiki/Index_(search_engine)))

The order in which results are returned is determined by degree of relevancy. Roughly, relevancy is a product of the number of query keywords found on a given page, and the authority of that page's domain. Relevancy = # query keywords * domain's authority. Authority has many factors, but it largely relates to links. Obviously a search engine cares about what you are saying on your site, but they care even more about what other people are saying about you (or in other words how they link to you). So authority is a combination of incoming links with the corresponding anchor text. PageRank (<http://searchengineland.com/what-is-google-pagerank-a-guide-for-searchers-webmasters-11068> - definition) is another important part of authority. At its simplest form PageRank is popularity – and every site that links to your site counts as a vote to your popularity. As the more popular a site is, the larger their vote in the popularity score. Of course such an algorithm is easy to game (i.e. with link farms, circular linking, etc.) so Google has evolved the factors that impact any site's authority – including domain diversity, how important the domains are that link to your site, etc. Using authority to filter out spammy or malicious sites makes a lot of sense – it's difficult to fake the 100s or 1000s of links that make a page relevant to a search. The anchor text (<http://www.seomoz.org/learn-seo/anchor-text>) used to link back to a site is a strong signal on the relevancy to various queries (and this also helps explain why having a keyword in your site name is helpful for ranking for that keyword – so much of the incoming anchor text to your site will use your brand or website name). I have glossed over a ton of really interesting and important work here, however, you can definitely read up on more of this if you are interested.-----

The Basics

As a developer when you think about the basics of SEO (or what **you** need to know about SEO – which mostly means not fucking things up) is as follows:

Make sure your site is crawlable.

There are two key parts to consider when it comes to crawling: making your site (and your pages) discoverable and making sure your content is indexed properly (indexation). This entails making sure that all pages are reachable by a robot, and that when a robot views the pages it sees all of the **relevant** content. *Ensure pages render without JavaScript enabled* – browse like a crawler would! The cool thing about all that fancy Ajax and JavaScript is that you can selectively render things and dynamically generate content – the downside is that you need to make sure you do it right for the people (or in this case robots) without JavaScript. Browse your site with JavaScript off and make sure all of the links and pages are reachable and the content renders (Firefox developer tools allow you to easily browse without JavaScript). In order for crawlers to register your page as relevant you need to make sure they can see all the content. So if you load a bunch of text dynamically with Ajax, or JavaScript, create a non-JavaScript version that will show the same information. In addition to making sure the web crawlers can reach your pages and see your content, you should also pay attention to how you construct links and anchor text to internal pages on your site. Anchor text is a key part of the search engine's algorithm so if you don't have any pages elsewhere on the Internet linking to the pages on your site (like you just started your blog and no one has linked back to yet, etc) then the best thing you can do is make sure that your own internal links use relevant keywords in the anchor text. *Use descriptive anchor text to pages – even on your own site.* Often times this is a great case for breadcrumb navigation – it helps with anchor text and provides relevant internal links and anchor text on your site. *Limit the number of links on the page* (there are lots of opinions on this, but generally if you have too many links then your site could be considered **spammy** by a search engine). *A crawler can't index a page that it can't see.* Search results consist of the content attributed to a page. If you are using Ajax, make sure that the content for each page on your site has its own URL. Sometimes you see Ajax sites where the user can interact and render the page without a new URL ever showing up – that may look awesome from a usability experience, but it can really hinder your search engine rankings if you haven't made a version that allows a user to access all that content without Ajax. *Follow URL best practices.* If you use URL conventions in "new" ways, it may hurt your SEO. For example, adding a # as part of the URL like "<http://www.yourdomain.com/p#product7>" – Google treats that as an anchor not a unique URL, so if it is a unique page consider a standard query parameter.

Use the right keywords in all the right places

There are lots of places that list and cover these – but just make sure you hit all the bases:

- Put them in the URL (and even better if your domain name has your keywords)

- The title of the page
- Have an h1 tag (and you can use CSS to style it to be smaller than your h2 if you'd like)
- Put alt text on images (and other objects like video, etc), and use descriptive image file names. When in doubt look at accessibility standards (<http://www.w3.org/standards/webdesign/accessibility>) - where there is alt text for screen readers there is alt text for search engines (and this means audio transcripts, yo!).

Have contextually relevant text content on the page. If you read the papers about TFIDF (<http://web4.cs.ucl.ac.uk/staff/jun.wang/blog/2009/07/08/tfidf/>) and LDA (<http://www.cs.princeton.edu/~blei/papers/BleiNgJordan2003.pdf>) you can learn a little bit about some of the algorithms that serve as the basis for the ones used to classify your site. If you understand these papers, then the importance of having good - robot readable - text content on your pages should be obvious. Just make sure it isn't duplicate content....*Order of the words is said to matter;* so put the most relevant ones in front, and towards the top of the page. This is why good SEO dictates "page title | site name" and not the reverse for titles. And finally a last takeaway on this topic - don't ever go overboard; too many keywords (or keyword stuffing) is a spammy signal, so choose something reasonable.

Avoid duplicate content

Google (and other search engines) using duplicate detection algorithms like shingling (you can read more about it in this textbook chapter: <http://infolab.stanford.edu/~ullman/mmds/ch3.pdf>) Avoid duplicating content from the web (unless you aggregate it with a lot of other content to make it appear different - the way we do with headlines and snippets of news on our product pages). This of course strongly applies to pages within your own site as well. Content duplication can confuse a search engine about which page is the authority (it can also result in penalties if you just cut and paste other people's content too) and then you can have your own pages competing with each other for ranking! If you must have duplicate content, make use of rel=canonical to let the search engines know which URL is the one they should be considered authoritative. But what if your page is a copy of another found on the web? Well then start coming up with some strategies to add more text and information to differentiate your pages, because such duplicate content is never likely to rank well.

Use smart meta descriptions

These are the little snippets that show up on search result pages beneath your link. These are actually not as important for SEO, but are super important if you want users to actually click on your links (and isn't that the whole reason you want to rank well anyway?) Proper meta descriptions allow a user to quickly determine if your page is really what they were looking for, something that can drastically improve click through rates (and therefore traffic) from a page of search results. Here is a link to some best practices on meta descriptions: <http://www.seomoz.org/learn-seo/meta-description> For the advanced users - get Google to showcase your site navigation: <http://support.google.com/webmasters/bin/answer.py?hl=en&answer=47334>

Freshness!

Google crawls sites that are updated more frequently (with quality content) more often. Fresh sites also tend to rank higher - so make sure at least part of your site is being updated regularly; corporate blog is a great way to do this (plus can give you a place to add contextually relevant content for your users - yay!). If you blog, try not to show the whole post on a page. Why? Well, then there is duplicate site on your page until that page moves into the archives (so if you have the option, only show an excerpt on the homepage). More on freshness here: <http://www.seomoz.org/blog/google-fresh-factor>

Fast - site speed

Google has stated the page load speed matters in their algorithm, so make sure you have tuned your site and are obeying best practices to make things speedy. <http://searchengineland.com/google-now-counts-site-speed-as-ranking-factor-39708> This is also good for your business as well, faster page loads have been shown to increase conversions. <http://www.wordstream.com/blog/ws/2011/08/23/page-speed-conversion-rate-optimization> (nice infographic on the topic at this link.)

301s vs. 302s and site errors (like 404s)

As I mentioned earlier in this article links are a big part of any search engine's algorithm. You want to make sure that any links you have continue to help drive your traffic and rankings. You should setup a Google webmaster tools account (<https://www.google.com/webmasters/tools/home?hl=en>) and dive into the crawler errors section. If you have a site that has been around for a while, chances are you may have some 404s. For example, this can

happen a lot for commerce sites that have products come and go - the old product pages are no longer relevant so those URLs may return a 404 when a user types them (or clicks on them in the search results) to view the product. Make sure you properly 301 redirect any ranking URLs that have been moved (a 301 represents a permanent redirect - <http://support.google.com/webmasters/bin/answer.py?hl=en&answer=93633>) to another relevant page or at least a friendly page explaining the URL is no longer valid (a page that returns a 200 status OK message). This way any links coming into the old page will be redirected, so hopefully the users will have a better experience and you will preserve any SEO karma you have built up to the old page. There is a great resource on this technique here: <http://www.seomoz.org/learn-seo/redirection>, including why you shouldn't use 302s (in most cases.)

Be patient

It may take a while to see the results of your SEO modifications. This makes sense if you factor in the time it takes for google bot to crawl the updated pages, then process each page and update all the corresponding indexes with the new content. And that can be quite a while when you are dealing with petabytes of content. Just imagine how long it creates to create an index of the web and to keep it updated and the right things stored in the inverted index.. When I was at my previous company we would crawl the web and even though we have optimized our whole system for speed, it could still take over a week to crawl 10Billion pages - especially if you respect robot politeness

(http://en.wikipedia.org/wiki/Web_crawler#Politeness_policy) and exponential back off. Plus since Google wants search queries and the results to be fast, all the indexes are pre-computed, and it takes time - sometimes a lot of time. And even if your site was in the index, you have to wait for people to query those keywords, and then come to your site - and if you aren't yet authoritative it may be a while! Sometimes this whole process can take over a month, so be patient.

Picking keywords

There are lots of strategies and work that can go into picking the right keywords - too much for the scope of this post. However, here are some key things you should consider when choosing:

- Understand what your users are actually searching for - what is their intent?
- Some keywords are competitive, find ones you can rank for by targeting the most relevant keywords better, or building links with the relevant anchor text
- Look at traffic volume - it does no good to target keywords that don't convert or don't have enough volume. You can get this information using Google Analytics via the SEO Optimization section and queries, or in Google Webmaster Tools. Here is a link with some more strategies: <http://www.seomoz.org/ugc/3-awesome-ways-to-leverage-google-analytics-in-e-commerce-seo>
- Target content at keywords (like blog posts to specific terms, or questions, your target users are likely to type into a search box. I know some people test this quantitatively by buying ads, although I have less experience on this particular strategy myself and some reservations - so do your research.)

If you want to learn more, here's some links that may help:<http://www.searchenginejournal.com/how-to-do-keyword-research-infographic/40437/><http://www.searchenginewizard.com/lisa-barone/five-steps-to-effective-keyword-research/><http://www.seospeedsight.com/seoforum/general-seo-discussion/how-to-build-increase-web-traffic-part-1-long-tail-keyword-strategy/><http://www.prelovac.com/vladimir/how-to-do-clever-keyword-research>

Building links

Since links and anchor text are such a key part of SEO, at some point you may want to consider getting more links to your site. As a developer, you probably don't want to talk to a bunch of people to get links (or at least I don't - I hate sales-y interactions of any kind). Typical link building is just as it sounds - you do what you can to get links pointing to your site - this can be through deals, partnerships, PR pitches, link exchanges - even paying for links (although I would not recommend this tactic). Although generally since search engines try to strip out the "chrome" of websites then if your link looks like a banner ad it may matter less, so "paid" is not always the best strategy. As one commenter pointed out on [Hacker News](#) - the best strategy for most sites when it comes to link building is to build interesting content that people will want to reference. While seeking more links is more beneficial more relevant links can help just as much. If you have a distribution partner for your content, or you build a widget, or anything else people will put on the web that links back to your site - you can dramatically improve link relevancy by ensuring all links have the optimal keywords in the anchor text. You should also ensure all links to your site point to your primary domain (<http://www.yourdomain.com>, and not a subdomain like <http://widget.yourdomain.com>). Additionally you want as many links to contain appropriate alt text. You get the idea. :) Another, perhaps less obvious way, to build links (or at least traffic) is to use social media - so setup your facebook, twitter, and google+ and whenever you have new links be sure to

share them. These channels can also work as an effective channel to drive more traffic to your site.*[also see the comments for more opinions on link building - there are lots of strong opinions on this topic, although in general it is the one you have the least control over so I didn't spend as much time on it here]*

In conclusion

Modern day search engine algorithms have been extensively optimized to detect cheaters, spam sites, or others looking to abuse the system, but most importantly to return only the most relevant results to the user. Take a look at SEOMoz's search engine ranking factors to see how many things engines consider as inputs to relevancy and authority - <http://www.seomoz.org/article/search-ranking-factors>. For example, a lot of SEOs talk about the importance of having diverse domains linking to your site that aren't in the same C-block of IPs - and others will tell you that circular linking ("you link to me and I'll link to you") is less valuable. Of course, though, those are nuances to the system - so if I were you I wouldn't spend my energy worrying about it and instead focus on building a great site with awesome content - unless you are going to be doing SEO full time it is probably a better use of your efforts (especially if you can write code to solve real problems). Related: Why do you need to use specialized SEO, like [Godrank casino SEO](#)? So sure, there are lots of ways to game the system, but generally you are best off following best practices and spending your energy on building real value, great content, and products. The search engines keep improving and it takes more and more sophistication to get around things and your risk being dropped out of the index all together (<http://www.seo.com/blog/black-hat-seo-infographic/>). But SEO isn't just about building a great site that is crawlable and has the right content - it is also about converting users. So make sure you have lots of metrics and know how to use things like Google Analytics. That way you can track what is working for you, and what isn't, and optimize your efforts. And if you have all this mastered, then check out Google Website Optimizer (it is free) and improve the colors, text and layout to take your conversions up another notch. And finally, if you built your site right, and optimized your pages, the most defensible SEO strategy is links, so think about ways to get widgets or links from other sites. Hopefully this is enough to get you started though! And if you want more information there are quite a few great SEO sites where you can continue your reading. [Video Reviews of Web 20 Ranker](#) may be a great help, too. P.S. Google has a pretty good [SEO basics guide](#) you can also check out - it covers some things I didn't mention here, namely **rel=nofollow links** and **robots.txt** - two important topics.

developers • programmers • search engines • seo • webmasters • websites

⟨ Questions for Candidates to Ask in an Interview

Systems Engineers Interview Questions
(also for System Administrators and Network Engineers) >

 Search

[linkedin // contact](#)