

Book II - Chapter 3 - Eliminating SE Roadblocks

183

Unfortunately, bus. owners often block SEs and ∴
cut off a huge segment of audience

If SE can't find content,

it can't index it, which means

it can't determine relevance. When this happens

Result - you don't get ranked!

I. Ensuring Search Engine Visibility

meta name="robots" content="noindex, nofollow"

Many developers use this when building site

Result - SE ignores page + every link on it

184 A. Checking your robots.txt file

Go to haselb5.org/robots.txt

May get Page Not Found error

Or, may see User-agent: *

Disallow: /blog.html

What you don't want to find is Disallow: /

This tells SE crawler to ignore every page!

Disallow can be used to hide pages that change a lot
hide duplicate content, or

keep SE out of what you don't want them crawling

Go to www.robotstxt.org to learn more

B. Checking for meta robots tag

= another way to hide pages from SE

View source code on any page

nofollow

If robots tag is there + contains noindex or n
remove both

If you don't know reason it's there, delete it
Don't depend on developer to remove them!

185 C. Eliminating registration forms

Form hurts in two ways

1. Greatly reduces # of people who read content
2. SE + fill out forms, so they'll never see info arguments in favor of forms - and rebuttals
3. We won't get any leads! They'll just read it
Will 1 # who read it by ID +
Because they're looking for info = good prospects
Make sure contact info is there!
2. Someone will steal our content
Will happen anyway!
Place a link on every page to the article itself
If page is stolen, will link back to site!
3. We need a return from this
SE generate most of the traffic online
By removing form → 1 customer + leads than lose
4. I won't be able to reach those who read it,
any I won't have their contact info
Reach them by posting another great article
Build reputation as an authority,
reach more readers, 1 marketing opp.

D. Eliminating login forms - for same reasons

Reasons owners use logins

1. Feel they need to add value for registered members
2. Want to force registration
3. Don't want company secrets getting out
Then don't put it on Internet!

II. Providing a Way to Browse

187 SEs move from pg. to pg. on site by following links
+ use search box!

+ use navigation dropdown box

★ Make sure there's a way to click thru + browse
the entire database using regular HTML links

1. Location database that shows list of local ^{street +} addresses
 2. Searchable directory of products
 3. Address book of company's staff
 4. Any other data stored in db + searched using form
- add a link that says: Browse all products
add a link that lets visitors click instead of search
by choosing their state + then city
Make sure all your content is browseable!

III. Using AJAX and DHTML

188 AJAX allows page to provide new info w/o reloading
= faster, smoother, makes customers happy

DHTML = used to create drop-down menus, simple
animations, + special effects

Problems: + DHTML

Ajax loads content using Javascript, but
SE ≠ read JS!

= effectively hides the page from SE

Make sure Web page degrades gracefully

Provides basic navigation + all content for non-JS

189 Test it - Use Web Developer Toolbar - disable CSS + JS
If you see the links from drop down, ok!

IV. Avoiding All-Flash Pages

189

SE ≠ read Flash animations! don't have plug-in
Flash compiles links, fonts, + structural info differently
Flash loads multiple pages of text, video motion graphics
into 1 pg. of HTML - creates same problems as Ajax
Google + Adobe announced they're working to make it
more search friendly
Not there yet

SE might be able to crawl a Flash animation
but can't organize + categorize content
+ may ignore much of the text copy

Warning signs

1. Visitors click from pg to pg within Flash -
but ≠ actually go to a new page
2. Yellowpipe Lynx Viewer shows blank pg or ≠ text
3. Web designer can't swear Web site = SE ready

V. Avoiding Client-side Redirects

190

Why - many Web developers used pgs. of just keywords
+ then used re-directs for human readers

How - javascript - SE ≠ support
meta refresh tag - SE ≠ understand

191

Cloaking = redirection tactic + a big No-No
= cardinal sin of SEO - use server side or 301 redirect

VI. Avoiding Duplicate Content Checking site with Web Dev

1. Disable JS, meta redirects, + CSS Toolbar
2. Click around site - if can go everywhere: ok!
3. If not, provide an alternative way for SE
ie links in footer

VII. Avoiding Duplicate Content

191 Reasons Why it's Bad

1. Duplication = tool used to fool SE's:
ex. replicating 1 web site on diff. domains so = linked
2. Creates confusion - hard for SE to rank pages
3. Other sites may link to either version - splits vote
Each link to your Web site = a vote

B. How caused? It's usually a mistake!

Created by inconsistent linking.

bad pagination scripts

sloppy Web site building practices

A. Finding duplicate content using Google site.narfe65.org

1. Go to Google + type in your site's main pg. ^
2. Click thru all the results pages

Message may say they've omitted similar pages

3. Repeat the search with omitted pages → duplicates

Another method:

1. Copy a sentence from somewhere on site,
one that others + likely to use
2. Search for sentence on Google in quotes
3. Results - show sites that copied your pg
+ pages on your site that = duplicates

B. Linking to your homepage the right way

www.narfe65.org

narfe65.org

narfe65.org/index.html

} all 3 point to same pg,
but SE sees each = unique!

∴ creates duplicate pages!!

splits the relevance vote which competes for authority

194

solution - always use same method for links
<http://www.narf65.org/>

Then set up a 301 redirect from the others

Make narf65.org → www.narf65.org

Reduces link confusion + prevents home pg. duplication

C. Using consistent URLs

Dynamic, database driven sites can create thousands of duplicate pages, as they automatically create pages upon user request

ex. <http://www.mysite.com?pageid=1&content=234&category=category>

vs. <http://www.mysite.com?pageid=1&content=234>

VIII Dealing with Broken links

195

Broken links STOP indexing of your pages
 + may reduce relevance of your site

A. Finding Broken links - <http://home.snafe.de/Hilman/xerulink.html>

Use Xenu (on PC) or Integrity (on Mac) pp. 167-8

Run the tools, put results in a list

Using list, fix all the broken links

B. Using a 301 redirect - link refers to page # there!

What it does - 301 redirect says: Page # here

Go to this other page instead

Don't Use a 302 redirect - says page will be back soon!

Search engine # pass the link relevance from old → new

+ doesn't replace old page with new one

Set up a 301 redirect on either web server

Internet Information Server 1.55 (Microsoft)

or Apache

IX Removing Code Bloat

195 SEs look for a high ratio of content to code -
so take these steps to maximizing the ratio

A. Coding using standards

196 Using XHTML and CSS2.0 → minimizes code bloat
keep styling code in a diff. document
put all ~~XHTML~~ content on XHTML page
Biggest benefit - stop using tables for layout!

B. Removing inline Javascript and CSS

Move them into separate files!

197 `<script language="javascript" type="text/javascript" src="thescript.js">`
`<link rel="stylesheet" href="style.css" type="text/css">`

Additional benefit:

Visiting Web browsers will cache the info on user's HD
speeds site performance

You use less bandwidth

More importantly - site-changes + additions - more easily made