

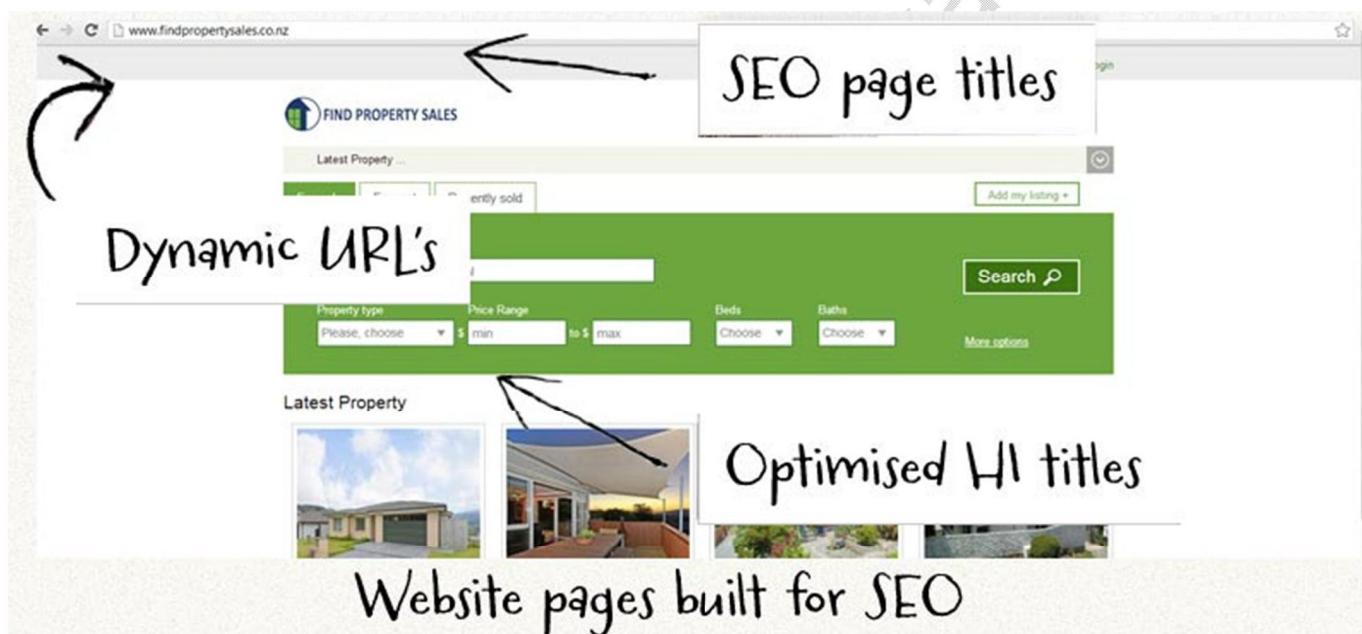
prakashgajera-----

CH6- Developing an SEO-Friendly Website

prakashgajera-----

Making Site Accessible to Search Engines

- We will examine the major elements of how to assess the **search engine friendliness of your site**.
- **Making** your site content accessible to search engines is the first step toward creating visibility in **search results**.
- **Once your website content** is accessed by a search engine, it can then be considered for relevant positioning within the **SERPs**.
- **Developing an SEO-friendly site architecture** requires
 - A. **Significant** amount of thought,
 - B. **Planning**, and
 - C. **Communication**,
 - D. Due to the **large number** of factors that influence the ways a search engine sees your site and the large number of ways in which a website can be put together.



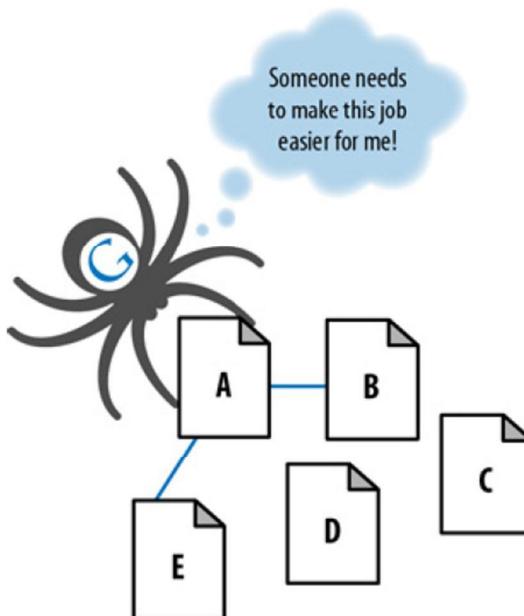
- The first step in the SEO design process is to ensure **that your site can be found and crawled by the search engines**.
- This is not as simple as it sounds, **as there are many popular web design and implementation constructs that the crawlers may not understand**.

1. Indexable Content

- To rank well in the search engines, your site's content—that is, the material available to visitors of your site—should be in HTML text form.
- **For example**, while the search engines do crawl images and Flash files, these are content types that are difficult for search engines to analyze, and therefore they do not help them determine the topical relevance of your pages.

2. Spiderable Link Structures

- **Search engines** use links on web pages to help them discover other web pages and websites.
- For this reason, we strongly recommend taking the time to build an internal linking structure that spiders can crawl easily.
- **Many sites make** the critical mistake of hiding or obfuscating their navigation in ways that limit spider accessibility, thus impacting their ability to get pages listed in the search engines' indexes.

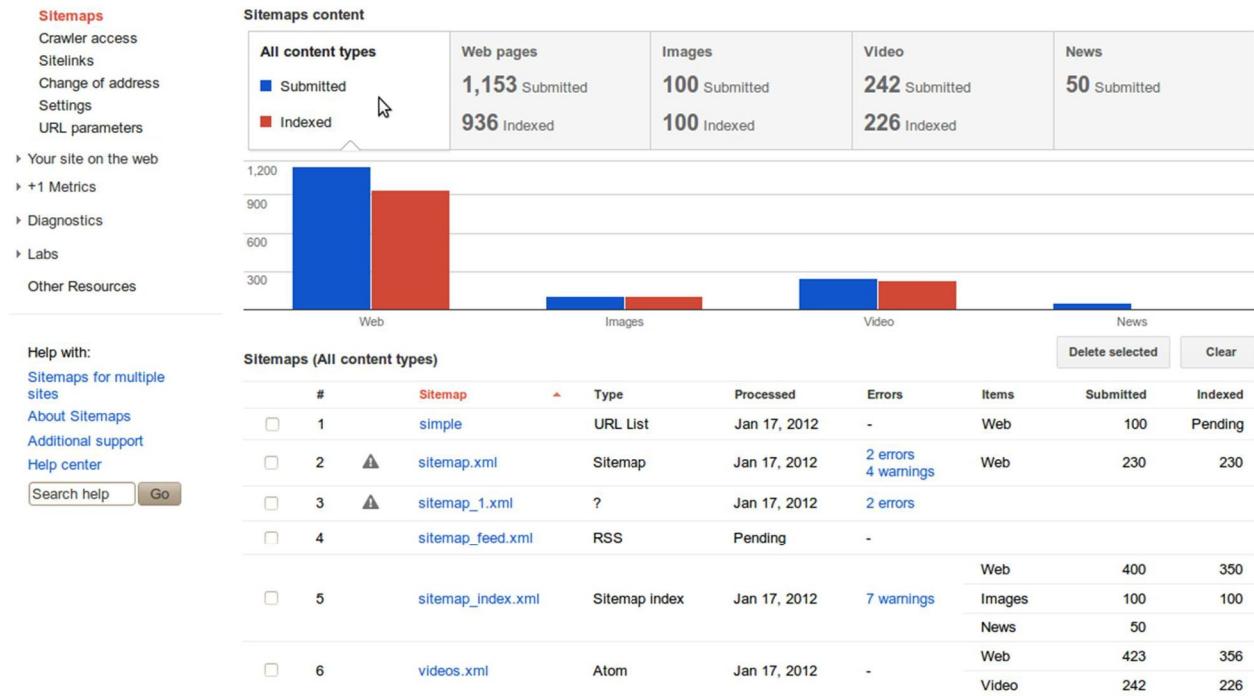


2.1 Here are some common reasons why pages may not be reachable:

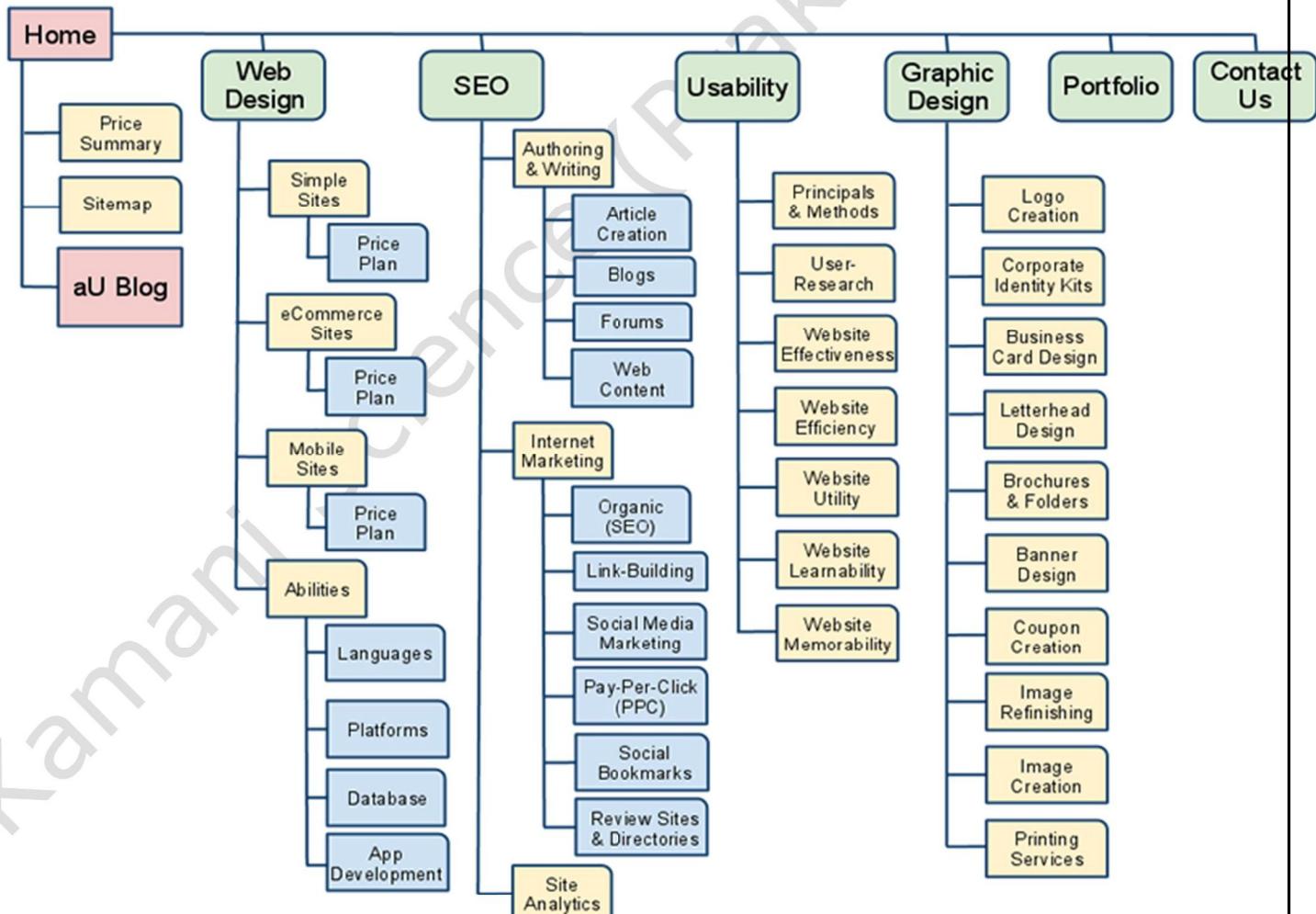
1.	Links in submission-required forms
2.	Links in hard-to-parse JavaScript
3.	Links in Flash, Java, or other plug-ins
4.	Links in PowerPoint and PDF files
5.	Links pointing to pages blocked by the meta robots tag, rel="NoFollow", or robots.txt
6.	Links on pages with many hundreds or thousands of links
7.	Links in frames or iframes

3. XML Sitemaps

- Google, Yahoo!, and Bing (from Microsoft, formerly MSN Search, and then Live Search) all **support a protocol known as XML Sitemaps**.
- Google first announced it in 2005, and then Yahoo! and MSN Search agreed to support the protocol in 2006.
- Using the Sitemaps protocol you can supply the search engines with a list of all the pages you would like them to crawl and index.



❖ Layout of an XML Sitemap



❖ What to include in a Sitemap file

- | | |
|----|-----------------|
| 1. | Mobile Sitemaps |
| 2. | Video Sitemaps |

3.	Image Sitemaps
4.	Managing and updating XML Sitemaps
5.	Updating your Sitemap with Bing.
6.	Updating your Google Sitemap

Google and the other major search engines discover and index websites by crawling links. Google XML Sitemaps are a way to feed to Google the URLs that you want crawled on your site.

----- prakashgajera -----

Creating an Optimal Information Architecture (IA)

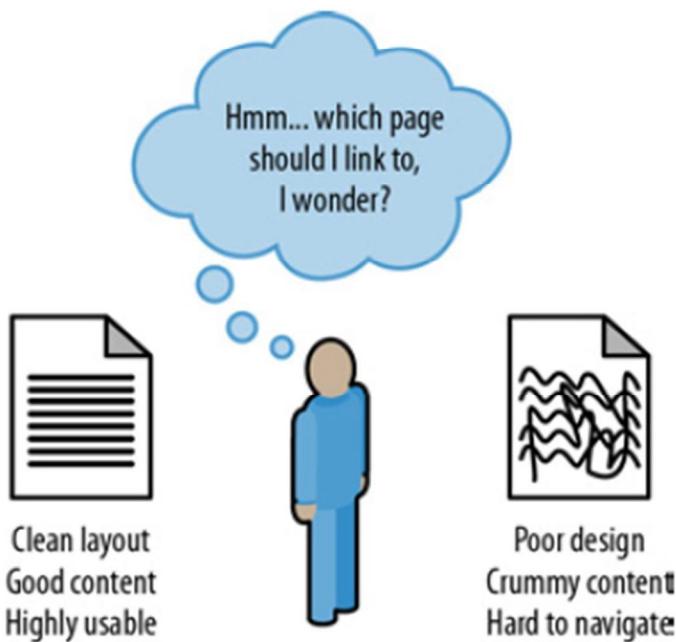
- Making your site friendly to **search engine crawlers** also requires that you put some thought into your site **information architecture**.
- A **well-designed architecture** can bring many benefits for both users and search engines.

1. The Importance of a Logical, Category-Based Flow

- The search engines face myriad technical challenges in understanding your site.
- Crawlers are not able to perceive web pages in the way that humans do, and thus significant limitations for both accessibility and indexing exist.
- Logical and properly constructed website architecture can help overcome these issues and bring great benefits in search traffic and usability.

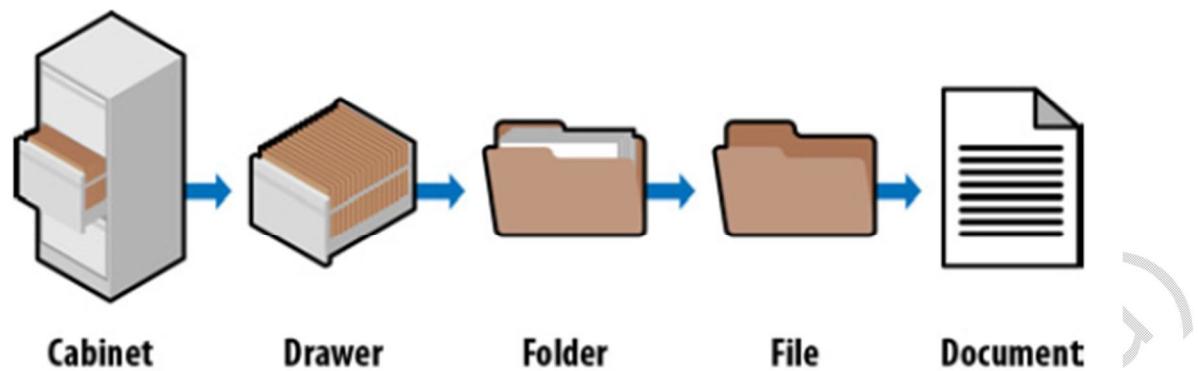
2. Usability and search friendliness

- Search engines are trying to reproduce the human process of sorting relevant web pages by quality.
- If a real human were to do this job, usability and the user experience would surely play a large role in determining the rankings.



3. An analogy

Look at how a standard filing cabinet is organized. You have the individual cabinet, drawers in the cabinet, folders within the drawers, files within the folders, and documents within the files

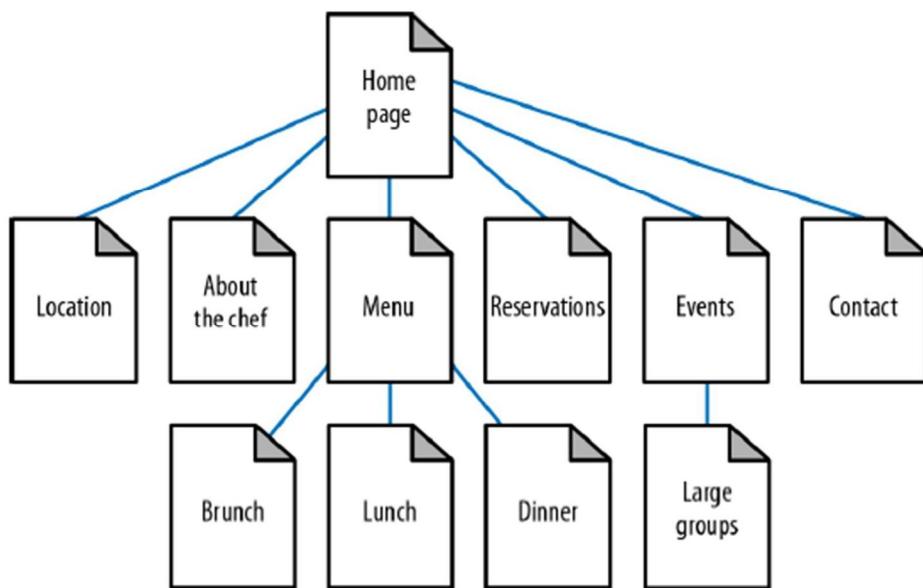


4. Site Architecture Design Principles

- When planning your website, remember that nearly every user will initially be confused about where to go, what to do, and how to find what he wants.

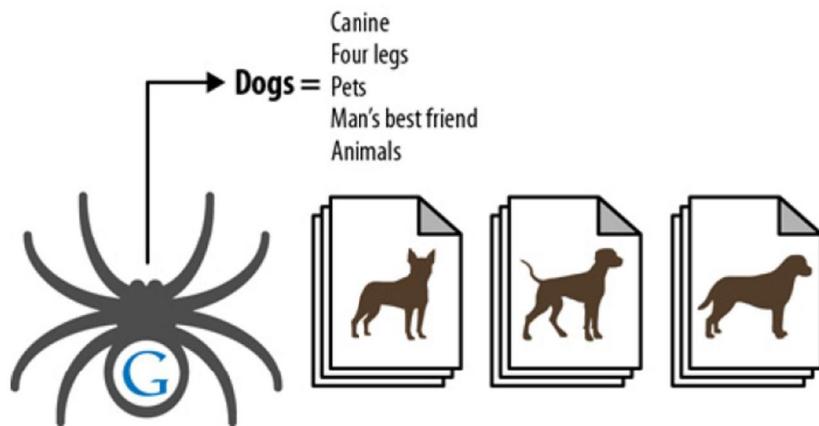
5. Designing site architecture

- | | |
|-----------|--|
| A. | List all of the requisite content pages (blog posts, articles, product detail pages, etc.). |
| B. | Create top-level navigation that can comfortably hold all of the unique types of detailed content on the site. |
| C. | Reverse the traditional top-down process by starting |
| D. | Once you understand the bottom, fill in the middle. Build out a structure for subnavigation to sensibly connect top-level pages with detailed content. |
| E. | Include secondary pages such as copyright |



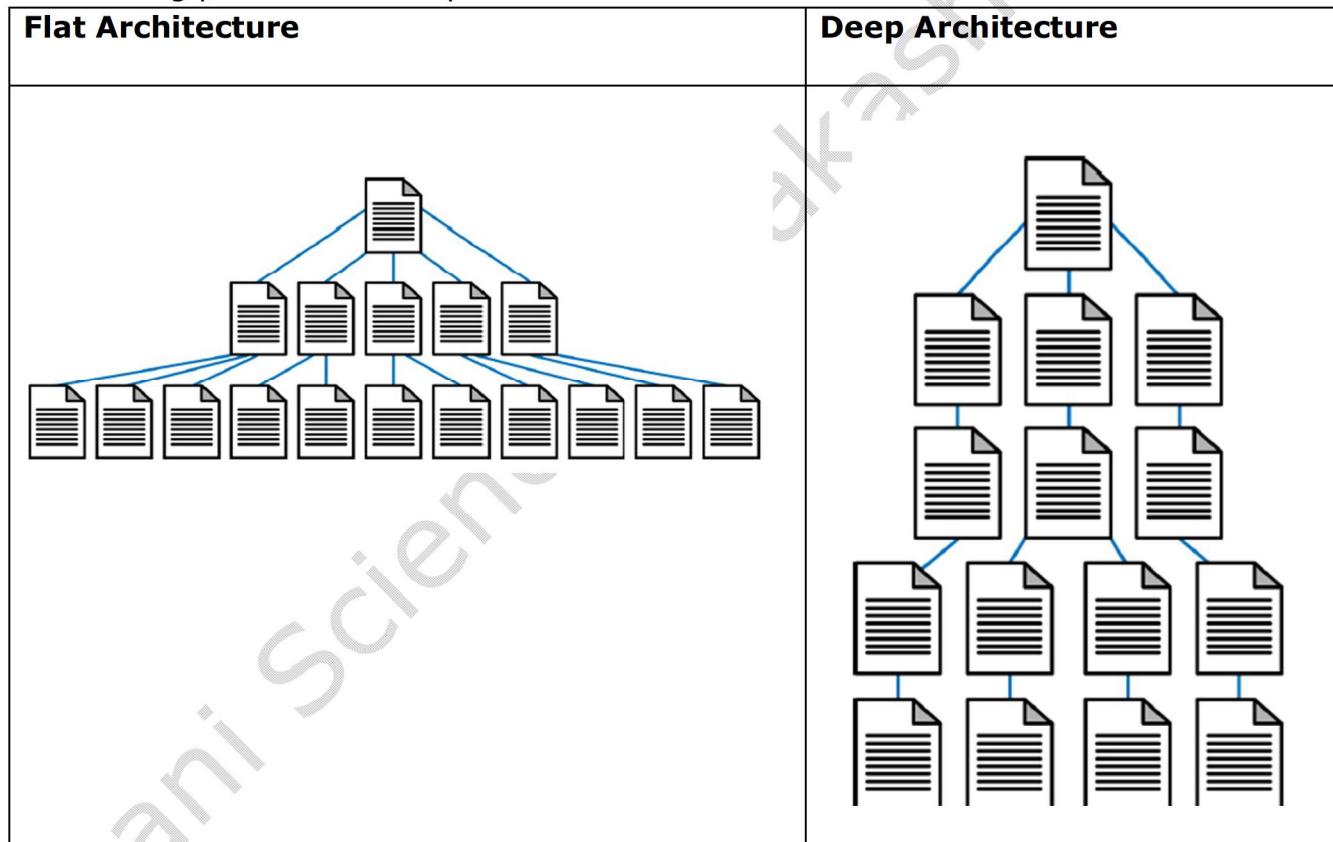
6. Category structuring

- As search engines crawl the Web, they collect an incredible amount of data (millions of gigabytes) on the structure of language, subject matter, and relationships between content.



7. Flat Versus Deep Architecture

- One very strict rule for search friendliness is the creation of flat site architecture. Flat sites require a minimal number of clicks to access any given page, whereas deep sites create long paths of links required to access detailed content.



8. Search-Friendly Site Navigation

Basics of search engine friendliness

Site elements that are problematic for spiders

Search and web forms.

Java, images, audio, and video

AJAX and JavaScript.

Root Domains, Subdomains, and Microsites

Root Domain is main domain of website.

- Individual pages/URLs**

These are the most basic elements of the Web: filenames, much like those that have been found on computers for decades, which indicate unique documents.

http://www.yourdomain.com/page.html.

- Subfolders**

The folder structures that websites use can also inherit or be assigned metrics by search engines. In the URL **http://www.yourdomain.com/blog/post17.html**

- Subdomains/fully qualified domains (FQDs)/third-level domains**

In the URL **http://blog.yourdomain.com/page.html**, three kinds of domain levels are present. The top-level domain (also called the TLD or domain extension) is .com, the second-level domain is yourdomain, and the third-level domain is blog. The third-level domain is sometimes referred to as a subdomain.

- Complete root domains/host domains/pay-level domains (PLDs)/second-level domains**

The domain name you need to register and pay for, and the one you point DNS settings toward, is the second-level domain (though it is commonly improperly called the "top-level" domain). In the URL **http://www.yourdomain.com/page.html**, yourdomain.com is the second-level domain.



When to Use a Subfolder

- If a subfolder will work, it is **the best choice 99.9% of the time**.
- Keeping content on a single root domain and single subdomain (e.g., **http://www.yourdomain.com**) gives the maximum SEO benefits,
- as the engines will maintain all of the positive metrics the site earns around links, authority, and trust and will apply these to every page on the site.

When to Use a Subdomain

- If your marketing team decides to promote a URL that is completely unique in content or purpose and would like to use a **catchy subdomain** to do it, using a subdomain can be practical.

- **Maps.google.com** is an example of where the marketing considerations make a subdomain an acceptable choice.
- One good reason to use a **subdomain** is in a situation where doing so can look more authoritative to users, as a result of creating separation from the main domain.

When to Use a Separate Root Domain

- If you **have a single**, primary site that has
 - earned links,
 - built content, and
 - attracted brand attention and awareness,
 - it is very rarely advisable to place any new content on a completely separate domain.
- There are rare occasions when this can make sense, and we'll walk through these, as well as explaining how singular sites benefit from collecting all of their content in one root domain location.

Microsites

- There is a lot of debate about microsites,
- And although we generally recommend that you do not saddle yourself with the hassle of dealing with multiple sites and their SEO risks and disadvantages,
- it is important to understand the arguments, even if there are only a few, in favor of doing so.

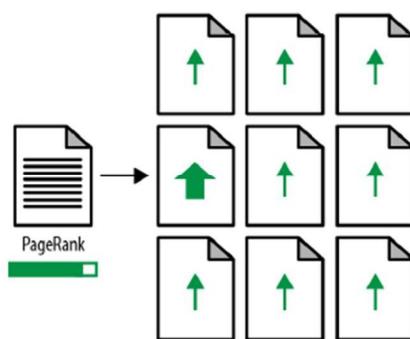
When should you consider a microsite?

- When you own a specific keyword search query domain
- When you plan to sell the domains
- if you're a major brand building a "secret" or buzz-worthy microsite

Here are the reasons for not using a microsite:

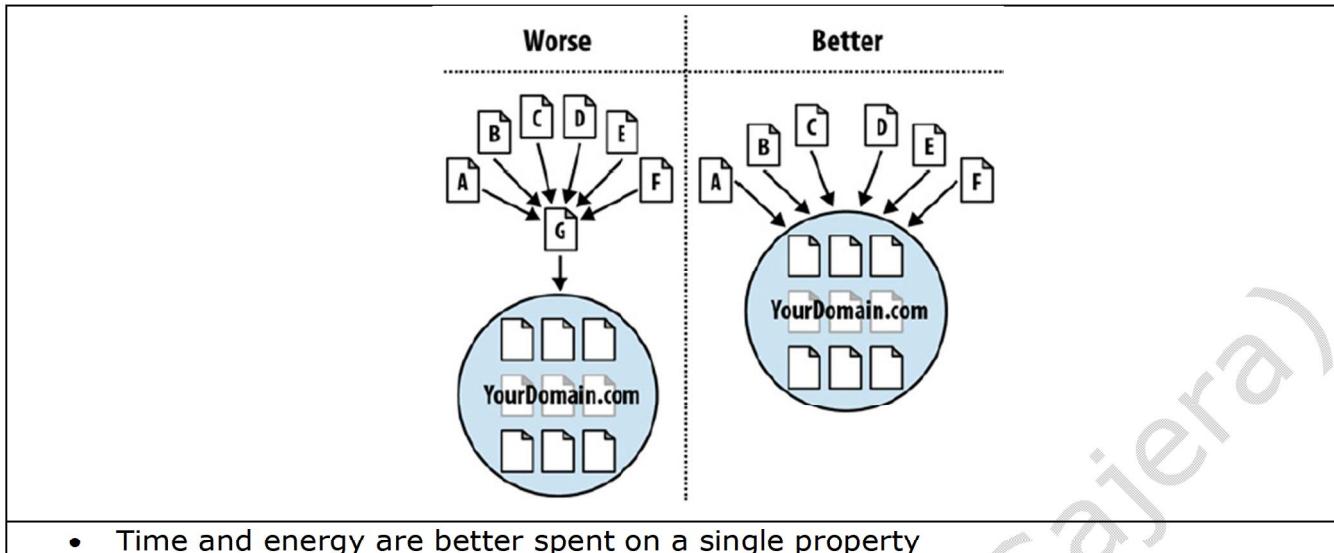
- Search algorithms favor large, authoritative domains
- Multiple sites split the benefits of links

**The rising tide (of a link)
lifts all boats (on the domain)**



Not only does the individual page
gain rankings, the entire domain
(and every page on it) receives
some ranking benefit

- A large, authoritative domain can host a huge variety of content



prakashgajera-----

Optimization of Domain Names/URLs

Two of the most basic parts of any website are the **domain** name and the **URLs** for the pages of the website.



❖ Optimizing Domains

Here are 12 tips that will be indispensable in helping you select a great domain name:

- 1** Brainstorm five top keywords
- 2** Make the domain unique
- 3** Choose only .com-available domains for your country
- 4** Make it easy to type.
- 5** Make it easy to remember
- 6** Keep the name as short as possible
- 7** Create and fulfill expectations.
- 8** Avoid trademark infringement
- 9** Set yourself apart with a brand
- 10** Reject hyphens and numbers
- 11** Don't follow the latest trends
- 12** Use an AJAX domain selection tool.

❖ Picking the Right URLs

What follows are some guidelines for selecting optimal URLs for the pages of your site(s).

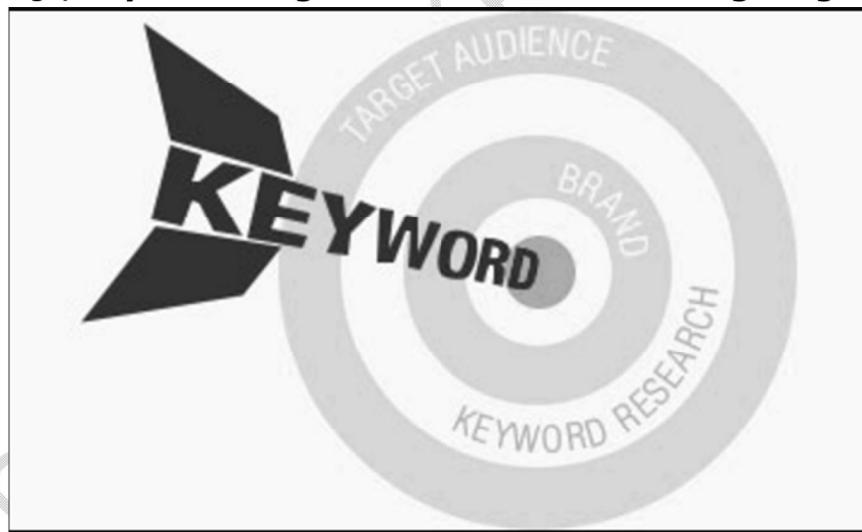
- | | |
|-----|-------------------------------------|
| 1. | Describe your content |
| 2. | Keep it short |
| 3. | Static is the way |
| 4. | Descriptive are better than numbers |
| 5. | Keywords never hurt |
| 6. | Subdomains aren't always the answer |
| 7. | Use fewer folders |
| 8. | Hyphens separate best |
| 9. | Stick with conventions |
| 10. | Don't be case-sensitive |
| 11. | Don't append extra data |

prakashgajera-----

Keyword Targeting

- The search engines face a tough task:

- Based on **a few words in a query** (or sometimes only one),
- They must return a list of relevant results, order them by measures of importance, and hope that the searcher finds what she is seeking.
- This practice has long been a critical part of search engine optimization, and although other metrics (**such as links**) have a great deal of value in the search rankings, **keyword usage is still at the core of targeting search traffic.**



This section explores some of the more prominent places where a publisher can place those keywords.

A. Title Tags

- For keyword placement, **title tags** are the most critical element for search engine relevance.
- The title tag is in the **<head>** section of an **HTML document**, and it is the only piece of "**meta**" information about a page that influences relevancy and ranking.
 1. Incorporate keyword phrases.
 2. Place your keywords at the beginning of the title tag.
 3. Limit length to 65 characters (including spaces).
 4. Target longer phrases if they are relevant.

5. Use a divider
6. Focus on click-through and conversion rates
7. Target searcher intent

B. Meta Description Tags

Meta descriptions have three primary uses:

- To describe the content of the page accurately and succinctly
- To serve as a short text “advertisement” to click on your pages in the search results
- To display targeted keywords, not for ranking purposes, but to indicate the content to searchers

- **Here are seven good rules for meta descriptions:**
 - Tell the truth.
 - Keep it succinct
 - Author ad-worthy copy.
 - Test, refine, rinse, and repeat
 - Analyze psychology
 - Include relevant keywords
 - Don’t employ descriptions universally

C. Heading (H1, H2, H3) Tags

- The **Hx** tags in HTML (<h1>, <h2>, <h3>, etc.) are designed to indicate a headline hierarchy in a document.
- Thus, an <h1> tag might be considered the headline of the page as a whole, whereas <h2> tags would serve as subheadings, <h3>s as tertiary-level headlines, and so forth.

D. Document Text

- The **HTML** text on a page was once the center of **keyword optimization activities**.
- Metrics such as **keyword density** and keyword saturation were used to measure the perfect level of keyword usage on a page.
- As far as the **search engines are concerned**, however, the text in a document—and particularly the frequency with which a particular term or phrase is used—has very little impact on how happy a searcher will be with that page.

E. Image Filenames and alt Attributes

-

F. Boldface Text (or)

G. Avoiding Keyword Cannibalization

- One of the nastier problems that often crops up during the course of a **website’s information architecture**, **keyword cannibalization** refers to a site’s targeting of popular keyword search phrases on **multiple** pages, forcing the engines to pick which one is most relevant.

H. Keyword Targeting in CMSs and Automatically Generated Content

- **Large-scale publishing systems**, or those that produce automatically generated content, present **some unique challenges**.

- If **hundreds of pages** are being created every day, it is not feasible to do independent keyword research on each and **every page**, making **page optimization** an interesting challenge.

I. SEO Copywriting: Encouraging Effective Keyword Targeting by Content Creators

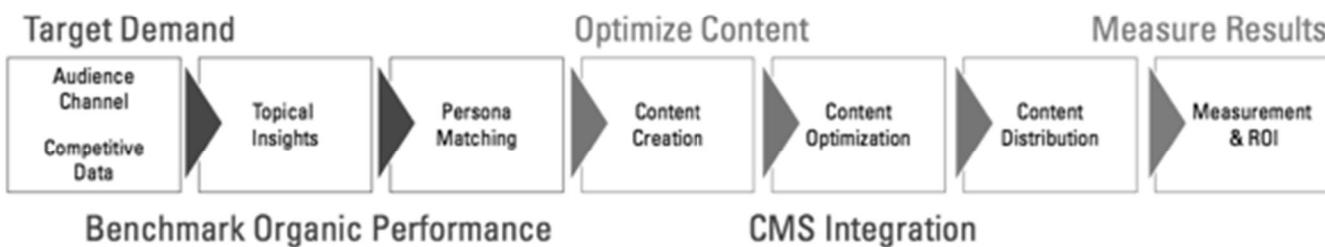
J. Long-Tail Keyword Targeting

prakashgajera-----

Content Optimization

- o **Content optimization** relates to how the presentation and **architecture of the text, image, and multimedia content** on a page can be optimized for search engines.
- o Many of these **recommendations** are second-order effects.

Content Performance Marketing - The Optimization Process



1. Content Structure

- Because **SEO** has become such a holistic part of website development and improvement,
- It is no surprise that **content formatting**—the presentation, style, and layout choices you select for your content—are a part of the process.
- **Choosing browser-safe sans serif** fonts such as **Arial and Helvetica** is a wise choice for the Web;
- Verdana in particular has received high praise from usability/readability experts, such as that Web offered in an article posted at <http://webaim.org/techniques/fonts/>.

2. Content length and word count

3. Visual layout

4. CSS and Semantic Markup

5. Schema.org and Micro formats

a. Markup in search results

b. Supported types of markup

c. Impact of rich snippets

6. Content Uniqueness and Depth

7. Content Themes

prakashgajera-----

Duplicate Content Issues Controlling Content with Cookies and Session IDs

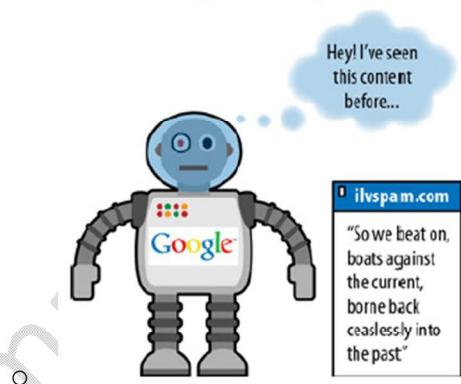
- o **Duplicate content can result from many causes**, including licensing of content to or from your site, site architecture flaws due to **non-SEO-friendly CMSs, or plagiarism**.

- Over the past five years, however, spammers in desperate need of content have begun the now much-reviled process of scraping content from **legitimate sources, scrambling the words** (through many complex processes), and repurposing the text to appear on their own pages in the hopes of **attracting long-tail searches** and serving contextual ads (and various other nefarious purposes).
- Thus, today we're faced with a world of "**duplicate content issues**" and "**duplicate content penalties**."
- Here are some definitions that are useful for this discussion:
 1. Unique content
 2. Snippets
 3. Shingles
 4. Duplicate content filter
 5. Duplicate content penalty

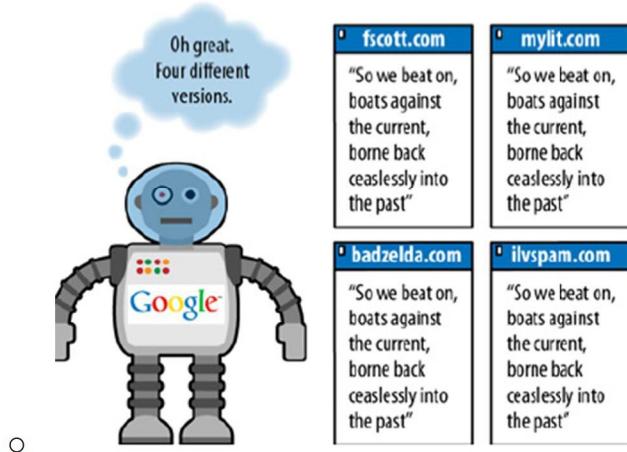
• How Search Engines Identify Duplicate Content

- Google as it finds duplicate content on the Web.
- The page with text is assumed to be a page containing duplicate content (**not just a snippet, despite the illustration**).
- Each page of **duplicate content** is presumed to be on a **separate domain**.
- The steps that follow have been simplified to make the process as easy and clear as possible.
- This is almost certainly not the exact way in which Google performs (but it conveys the effect).

Phase I: Google finds duplicate content



Phase II: Google checks comparable docs



- **There are a few facts about duplicate content that bear mentioning,**

1. Location of the duplicate content
2. Percentage of duplicate content
3. Ratio of code to text
4. Ratio of navigation elements to unique content
5. Licensed content

Controlling Content with Cookies and Session IDs

- **Sometimes** you want to more carefully dictate what a search engine robot sees when it visits your site.
- In general, **search engine representatives** will refer to the practice of showing different content to users than crawlers as cloaking, which violates the engines' **Terms of Service (TOS)** and is considered **spammy** behavior.

What's a Cookie?

- **A cookie is a small text file** that websites can leave on a visitor's hard disk, helping them to track that person over time.
- **Cookies are the reason Amazon.com remembers** your username between visits and the reason you don't necessarily need to log in to your **Hotmail** account every time you **open your browser**.
- **Cookie data typically** contains a short set of information regarding when you last accessed a site, an ID number, and, potentially, information about your visit.

What Are Session IDs?

- Session IDs are virtually identical to cookies in functionality, with one big difference: **when you close your browser (or restart)**, session ID information is (usually) **no longer stored** on your hard drive.

Use Cookies or Session IDs to Control Search Engine Access?

- Showing multiple navigation paths while controlling the flow of link juice
- Keeping limited pieces of a page's content out of the engines' indexes
- Granting access to pages requiring a login
- Avoiding duplicate content issues

prakashgajera

Content Delivery and Search Spider Control

- It can be valuable to show search engines one version of content and show humans a different version.
- This is **technically called "cloaking,"** and the search engines' guidelines have near-universal policies restricting this behavior.
- In practice, **many websites, large and small,** appear to use some forms of cloaking without being penalized by the search engines. However, use great care if you implement these techniques, and know the risks that you are taking.

Websites that perform some level of cloaking:

Google.com
NYTimes.com
Wine.com
Yelp.com
Trulia.com

When to Show Different Content to Engines and Visitors

Here are some of the most common ones:

1. Multivariate and A/B split testing
2. Content requiring registration and First Click Free
3. Navigation unspiderable by search engines
4. Duplicate content
5. Different content for different users

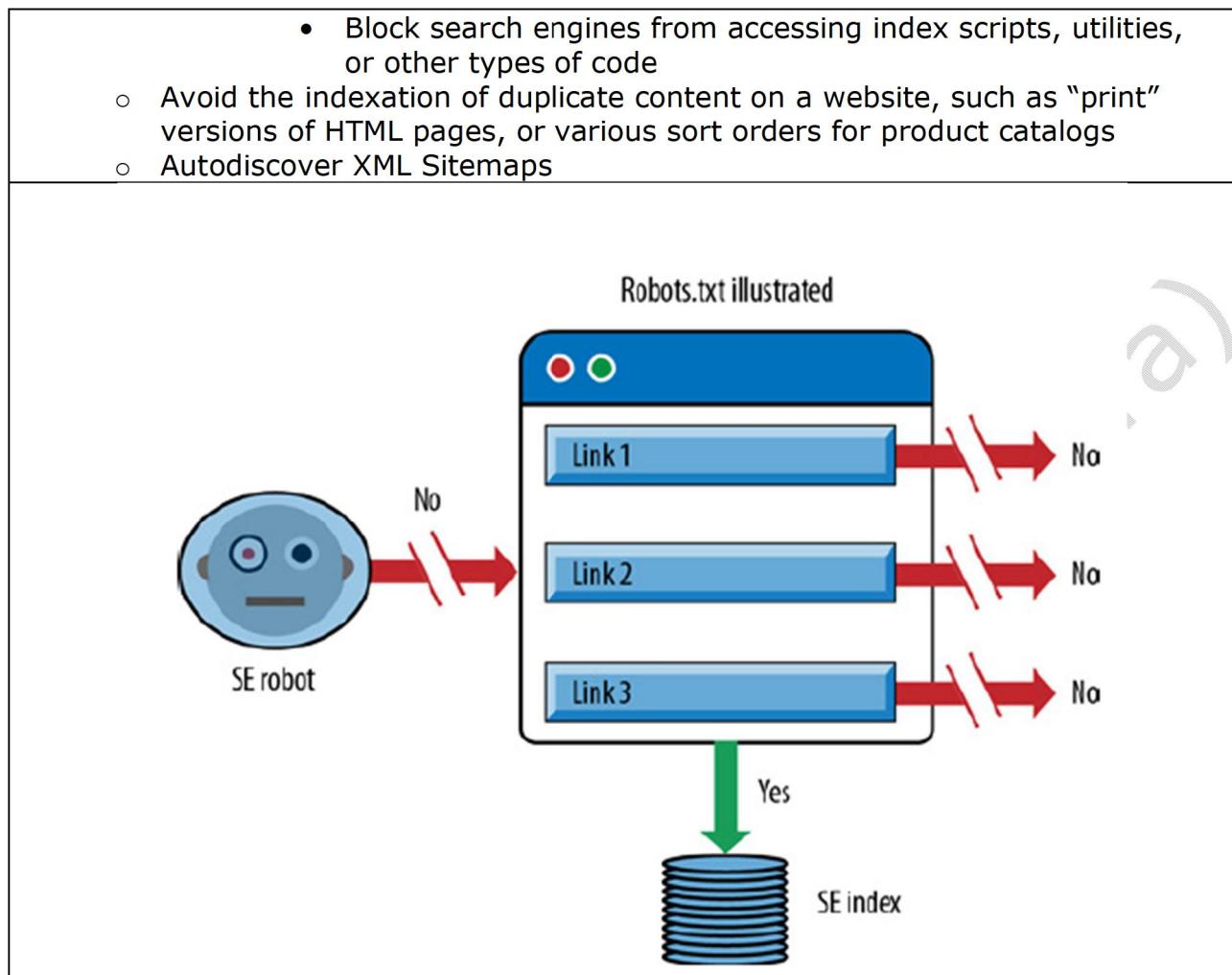
How to Display Different Content to Search Engines and Visitors

- A variety of strategies exist to segment content delivery.
- The most basic is to serve content that is not meant for the engines in unspiderable formats (e.g., placing text in images, Flash files, plug-ins).
- You should not use these formats for the purpose of cloaking.
- You should use them only if they bring a substantial end user benefit (such as an improved user experience).
-

The robots.txt file

- This file is located on the root level of your domain
- (**e.g., <http://www.yourdomain.com/robots.txt>,**)
- It is a highly versatile tool for controlling what the spiders are permitted to access on your site. You can use **robots.txt** to:
 - Prevent crawlers from accessing nonpublic parts of your website

- Block search engines from accessing index scripts, utilities, or other types of code
- Avoid the indexation of duplicate content on a website, such as “print” versions of HTML pages, or various sort orders for product catalogs
- Autodiscover XML Sitemaps



prakashgajera-----

Redirects and Content Management System (CMS) Issues

- **Redirects**

A redirect is used to indicate when content has moved from one location to another. For example, suppose you have some content at

<http://www.yourdomain.com/old.html>, and you decide to restructure your site. As a result of this restructuring, your content may move to

<http://www.yourdomain.com/critical-keyword.html>.

- **Why and When to Redirect**

Here are some scenarios in which you may end up needing to implement redirects:

- You have old content that expires, so you remove it.
- You find that you have broken URLs that have links and traffic.
- You change your hosting company.
- You change your CMS.
- You want to implement a canonical redirect (redirect all pages on <http://yourdomain.com> to <http://www.yourdomain.com>).
- You change the URLs where your existing content can be found for any reason.

- **Good and Bad Redirects**

“301 moved permanently”

This status code tells the browser (or search engine crawler) that the resource

has been permanently moved to another location, and there is no intent to ever bring it back.

"302 moved temporarily"

This status code tells the browser (or search engine crawler) that the resource has been temporarily moved to another location, and that the move should not be treated as permanent.

- The default document redirect solution**

You will need to plug in the appropriate default filename for your own web server:

Copy the contents of index.php to another file. For this example, we'll be using sitehome.php.

Create an Apache DirectoryIndex directive for your document root. Set it to sitehome.php.

Do not set the directive on a server-wide level; otherwise, it may cause problems with other folders that still need to use index.php as a directory index.

put this in your httpd.conf:

```
<Directory /your/document/root/examplesite.com/>
    DirectoryIndex sitehome.php
</Directory>
```

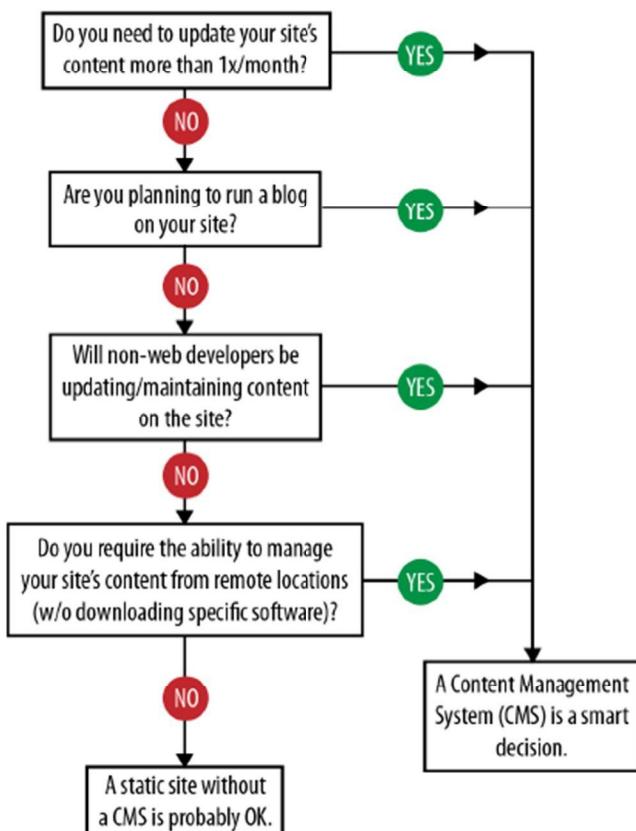
Clear out the contents of your original index.php file and insert this line of code:

```
<? header("Location: http://www.example.com"); ?>
```

Content Management System (CMS) Issues

When looking to publish a new site, many publishers may wonder whether they need to use a CMS, and, if so, how to ensure that it is SEO-friendly.

Do You Need A CMS For Your Site?



- Due to the inexpensiveness of customizable,
 - free platforms such as
 - **Drupal,**
 - **Joomla,**
 - **WordPress, and**
 - **Mambo,**
 - It is increasingly rare for a publisher to develop a static site, even when a CMS isn't required.

By dealing with these, you will ensure a relatively smooth platform for content delivery:

1. Title tag customization and rules
2. Static, keyword-rich URLs
3. Meta tag customization
4. Enabling custom HTML tags
5. Internal anchor text flexibility
6. Intelligent categorization structure
7. Pagination controls
8. 301-redirect functionality
9. XML/RSS pinging
10. Image-handling and alt attributes
11. CSS exceptions
12. Static caching options
13. URLs free of tracking parameters and session IDs
14. Customizable URL structure
15. Static-looking URLs
16. Keywords in URLs
17. Multilevel categorization structure
18. Paraphrasable excerpts
19. Meta NoIndex tags for low-value pages
20. Keyword-rich intro copy on category-level pages
21. NoFollow links in comments
22. Customizable anchor text on navigational links
23. XML Sitemap generator
24. HTML4, HTML5, or XHTML validation
25. Comments, and antispam mechanisms

----- prakashgajera -----

Optimizing Flash

- As referenced several times earlier in this chapter, Flash is popular on the Web, but it presents challenges to the search engines in terms of indexing the related content.
- This creates a gap between the user experience with a site and what the search engines can find on that site.

Here are some reasons why Flash is still not fully SEO-friendly:

Different content is not on different URLs

This is the same problem you encounter with AJAX-based pages. You could have unique frames, movies within movies, and so on that appear to be completely unique portions of the Flash site, yet there's often no way to link to these individual elements.

The breakdown of text is not clean

Google can index the output files in the .swf file to see words and phrases, but in Flash, a lot of your text is not inside clean `<h1>` or `<p>` tags; it is jumbled up into half-phrases for graphical effects and will often be output in the incorrect order. Worse still are text effects that often require "breaking" words apart into individual letters to animate them.

Flash gets embedded

A lot of Flash content is only linked to by other Flash content wrapped inside shell Flash pages. This line of links, where no other internal or external URLs are referencing the interior content, leads to documents with very low PageRank/link juice.

Flash doesn't earn external links like HTML

An all-Flash site might get a large number of links to the home page, but interior pages almost always suffer.

SEO basics are often missing

Anchor text, headlines, bold/strong text, image alt attributes, and even title tags are not simple elements to properly include in Flash. Developing Flash with SEO in **mind is just more difficult than doing it in HTML**.

A lot of Flash isn't even crawlable

Google has indicated that it doesn't execute external JavaScript calls (which many Flash-based sites use) or index the content from external files called by Flash (which, again, a lot of Flash sites rely on). These limitations could severely impact what a visitor can see versus what **Googlebot** can index.

----- prakashgajera -----

Best Practices for Multilanguage/Country Targeting

- Many businesses target multiple countries with their websites, and for such businesses, various questions arise. Do you put the information for your products or services all on the same domain?

Targeting a Specific Country

- Starting with the basics of international targeting, it is important to let the search engines know where your business is based in as many ways as possible.
- These might include:
 - Using a country-specific domain (e.g., .co.uk)
 - Hosting your site locally, not abroad
 - Including the physical local address in plain text on every page of your site
 - Setting Google Webmaster Central geotargeting to your country of interest
 - Verifying your address with Google Maps
 - Getting links from in-country websites
 - Using the local language on the website

-----Best Of Luck-----

----- prakashgajera -----