Unit – 3

Developing an SEO Friendly Website

By Swati Lathia

Introduction

Friendly

- It is necessary to have your website such that search engine can easily access it.
- We already know that search engine crawlers are basically software programs.
- They have their own strengths & weakness.
- So make your website search engine friendly so that crawlers can find your website and make it in index for ranking.

Introduction

- Some basic guidelines necessary to create fruitful SEO progress.
- *Each page should target – one to three phrase*
- *Remove duplicate content*
- *Relevant keywords in image alt & title*
- *Avoid frames & iframes*
- *Create a sitemap.xml doc*
- *Add a robot.txt to your root….etc*

Site Accessible to Search Engines

- The first step is to ensure that your site can be found and crawled by search engines.
- This is not as simple as it sounds

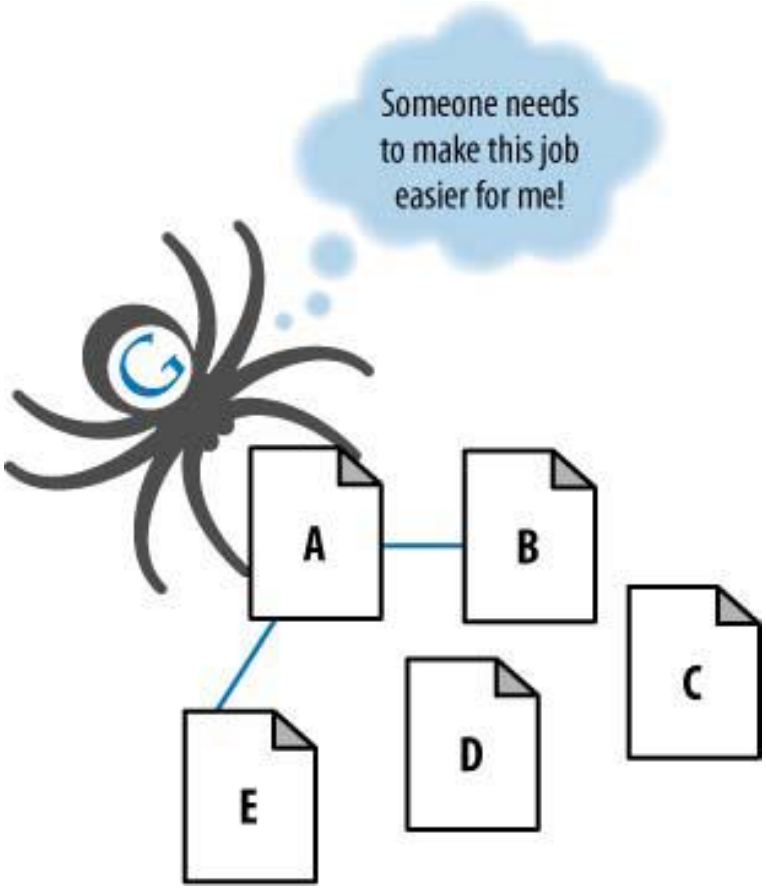
Site Accessible to Search Engines

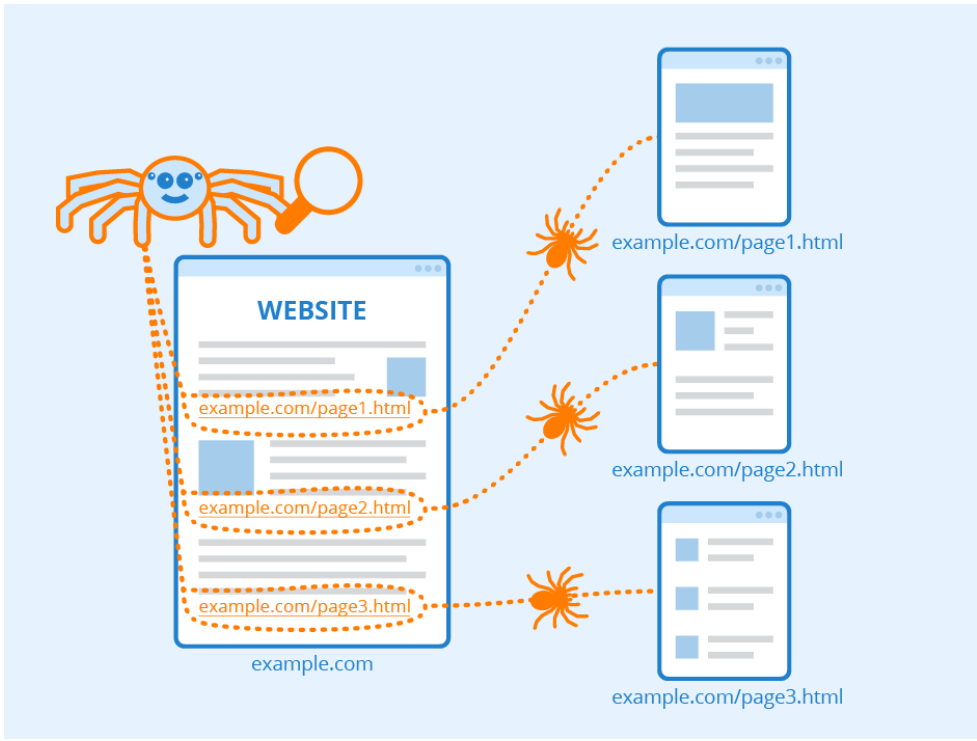
- **Indexable Content**
- To rank well in the search engines, your site's content — must be in HTML text form.
- Images, Flash files, Java applets, and other nontext content - invisible to search engine spiders despite advances in crawling technology.

Site Accessible to Search Engines

- **Spiderable Link Structures**
- Search engines use links on web pages to help them discover other web pages and websites.
- For this reason, website developers should invest the time to build a link structure that spiders can crawl easily.

Site Accessible to Search Engines





Site Accessible to Search Engines

- In above figure, Google's spider has reached Page A and sees links to pages B and E.
- However, even though pages C and D might be important pages on the site, the spider has no way to reach them (or even to know they exist) because no direct, crawlable links point to those pages.

Site Accessible to Search Engines

- **Some common reasons why pages may not be reachable:**

1. Links in submission-required forms
2. Links in nonparsable 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 Robots.txt or meta robot tag

Site Accessible to Search Engines

- **Some common reasons why pages may not be reachable:**

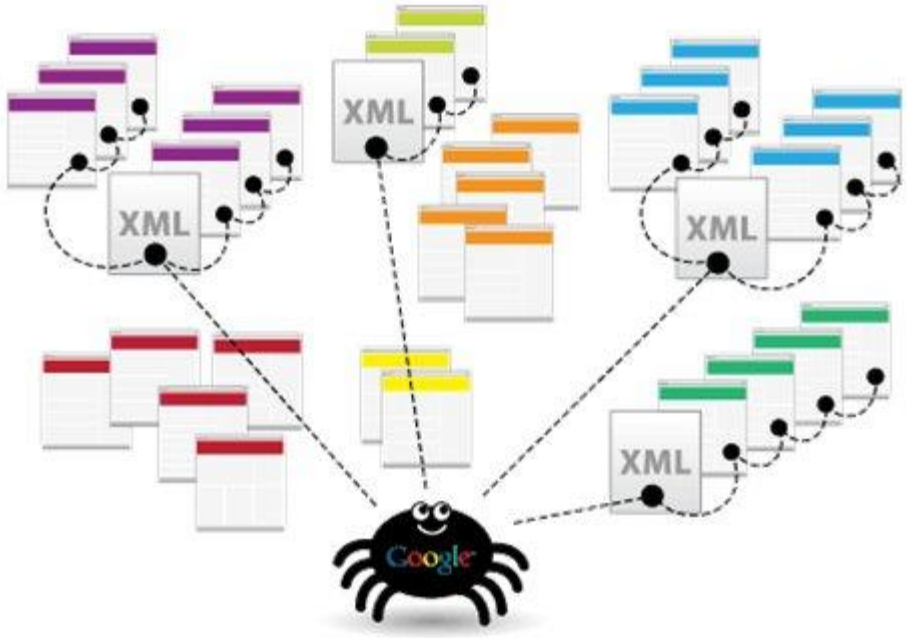
6. Links on pages with many hundreds or thousands of links
7. Links in frames or iframes

XML Sitemaps

- Google first announced it in 2005, and then Yahoo! and Microsoft agreed to support the protocol in 2006.
- Using the Sitemaps protocol you can supply the search engines with a list of all the URLs you would like them to crawl and index.
- Adding a URL to a Sitemap file does not guarantee that a URL will be crawled or indexed.

XML Sitemaps

- However, it can result in pages that are not otherwise discovered or indexed by the search engine getting crawled and indexed.



Benefits : XML Sitemaps

1. For the pages the search engines already know about through their regular spidering, they use the metadata you supply, such as the last date the content was modified (lastmod date) and the frequency at which the page is changed (changefreq), to improve how they crawl your site.

Benefits : XML Sitemaps

2. For the pages they don't know about, they use the additional URLs you supply to increase their crawl coverage.
3. For URLs that may have duplicates, the engines can use the XML Sitemaps data to help choose an original version.
4. Verification/registration of XML Sitemaps may indicate positive trust/authority signals.
5. The crawling/inclusion benefits of Sitemaps may have second-order positive effects, such as improved rankings or greater internal link popularity.

XML Sitemaps

- When you create a Sitemap file you need to take care to include only the canonical version of each URL.
- In other words, in situations where your site has multiple URLs that refer to one piece of content, search engines may assume that the URL specified in a Sitemap file is the preferred form of the URL for the content.

Layout of an XML Sitemap

```
<?xml version="1.0" encoding="UTF-8"?>

<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">

   <url>

      <loc>https://www.example.com/</loc>

      <lastmod>2018-08-24</lastmod>

      <changefreq>weekly</changefreq>

      <priority>0.5</priority>

   </url>

</urlset>
```

Layout of XML Sitemap

- **<loc> : Location**
- This compulsory tag contains the absolute, canonical version of the URL location.
- It should accurately reflect your site protocol (http or https) and if you have chosen to include or exclude www.

Layout of XML Sitemap

- **<lastmod> : Last modified**
- An optional but highly recommended tag used to communicate the file's last modified date and time.
- Google does use the lastmod metadata to understand when the page last changed and if it should be crawled.

Layout of XML Sitemap

- **<Changefreq> : Change Frequency**
- Once upon a time, this optional tag hinted how frequently content on the URL was expected to change to search engines.

Layout of XML Sitemap

- <priority>
- This optional tag that ostensibly tells search engines how important a page is relative to your other URLs on a scale between 0.0 to 1.0
- At best, it was only ever a hint to search engines and Now google says that “They ignore it”

An XML sitemap generator

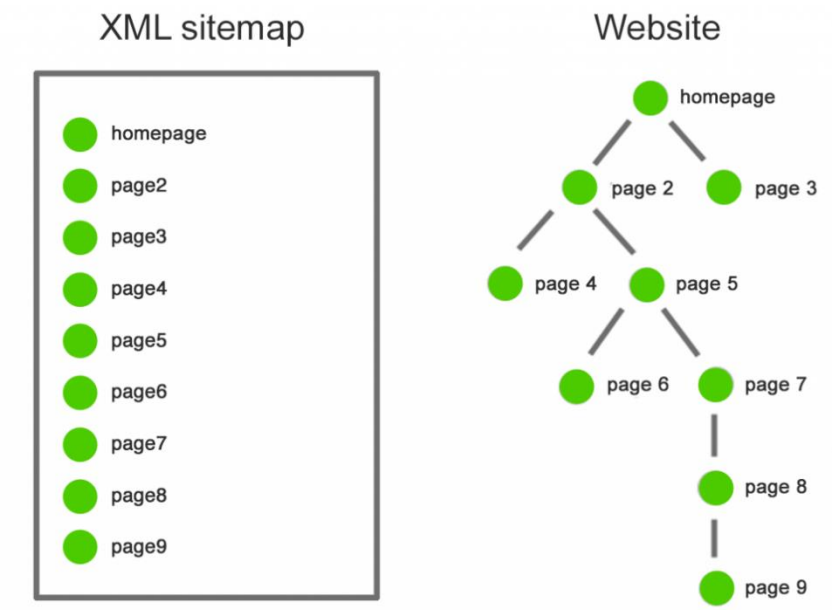
Some examples of XML Sitemap generators:

1. SourceForge.net's google-sitemap_gen
2. ROR Sitemap Generator
3. XML-Sitemaps.com Sitemap Generator
4. Sitemaps Pal
5. XML Echo

- **Simple text**
- You can provide Google with a simple text file that contains one URL per line.

XML Sitemaps

- **Where to upload your Sitemap file**
- When your Sitemap file is complete, upload the file to your site in the highest-level directory you want search engines to crawl (generally, the root directory).



Creating an Optimal Information Architecture

- A well-designed architecture can bring many benefits for both users and search engines.

Creating an Optimal Information Architecture

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

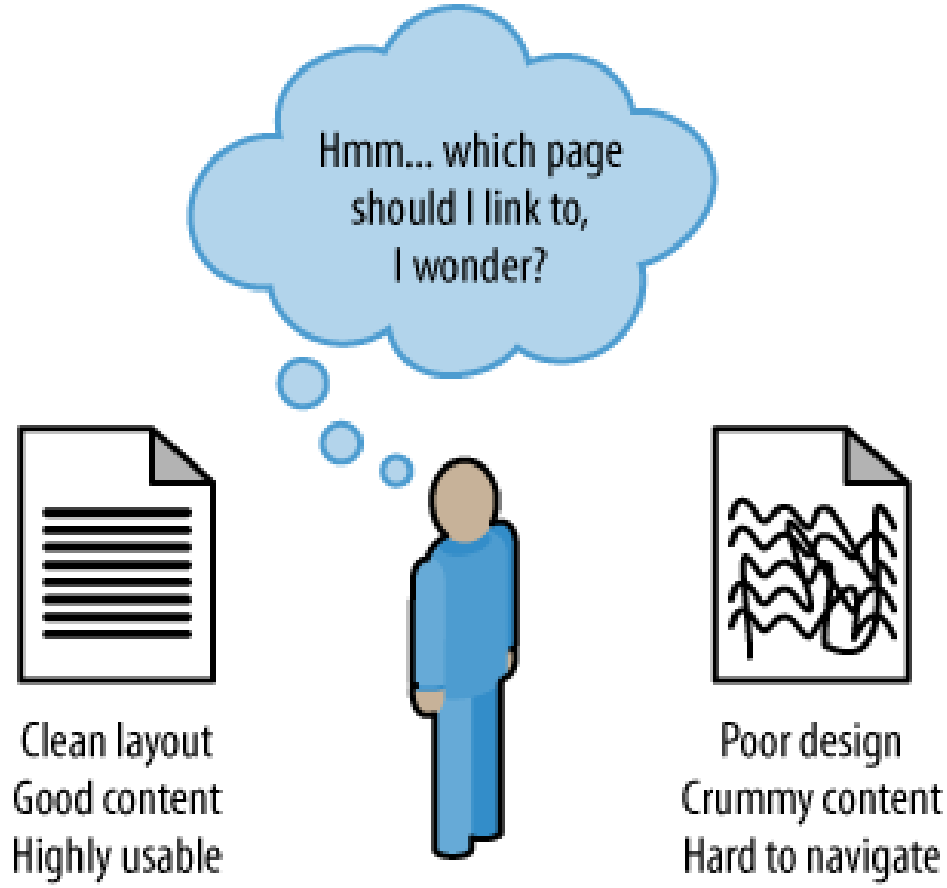
- SEs face countless technical challenges in understanding your site
- Crawlers - not able to perceive web pages in the way that humans do - accessibility and indexing limitation
- A logical and properly constructed website architecture - overcome issues - bring great benefits in search traffic and usability

Creating an Optimal Information Architecture

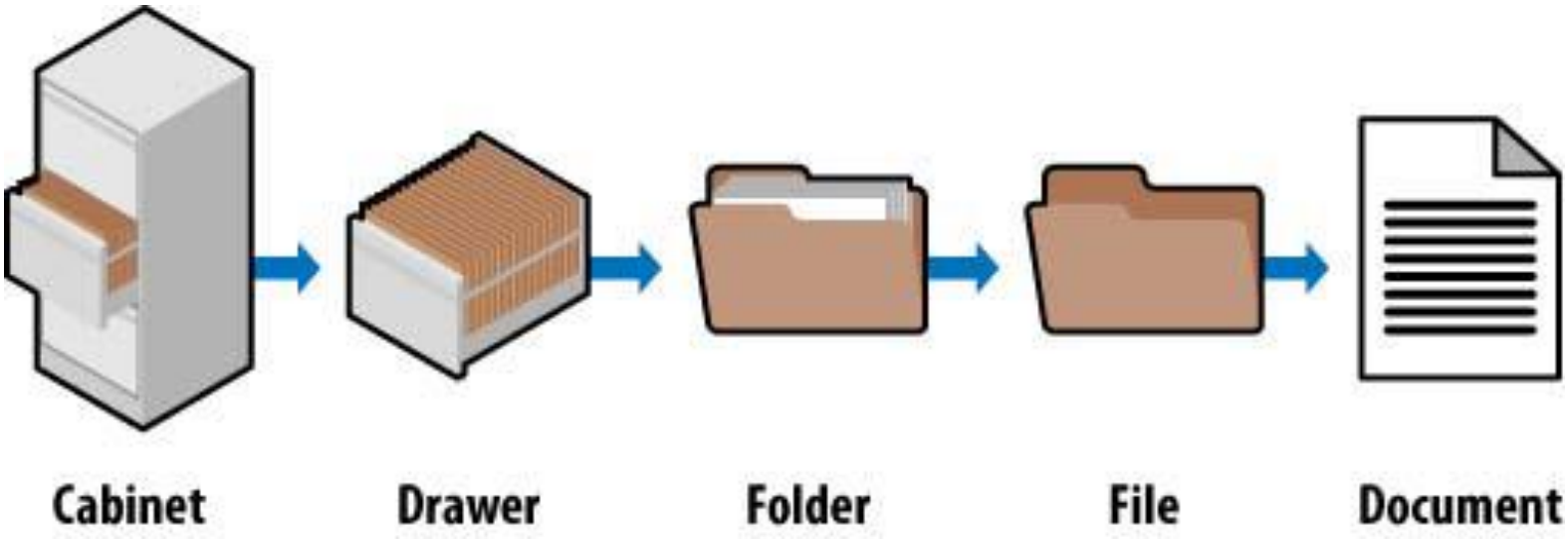
- Important principles of website organization are:

1. Usability, or making a site easy to use;
2. Information architecture or logical hierarchical structure for content.

Usability and search friendliness



Filing cabinet example

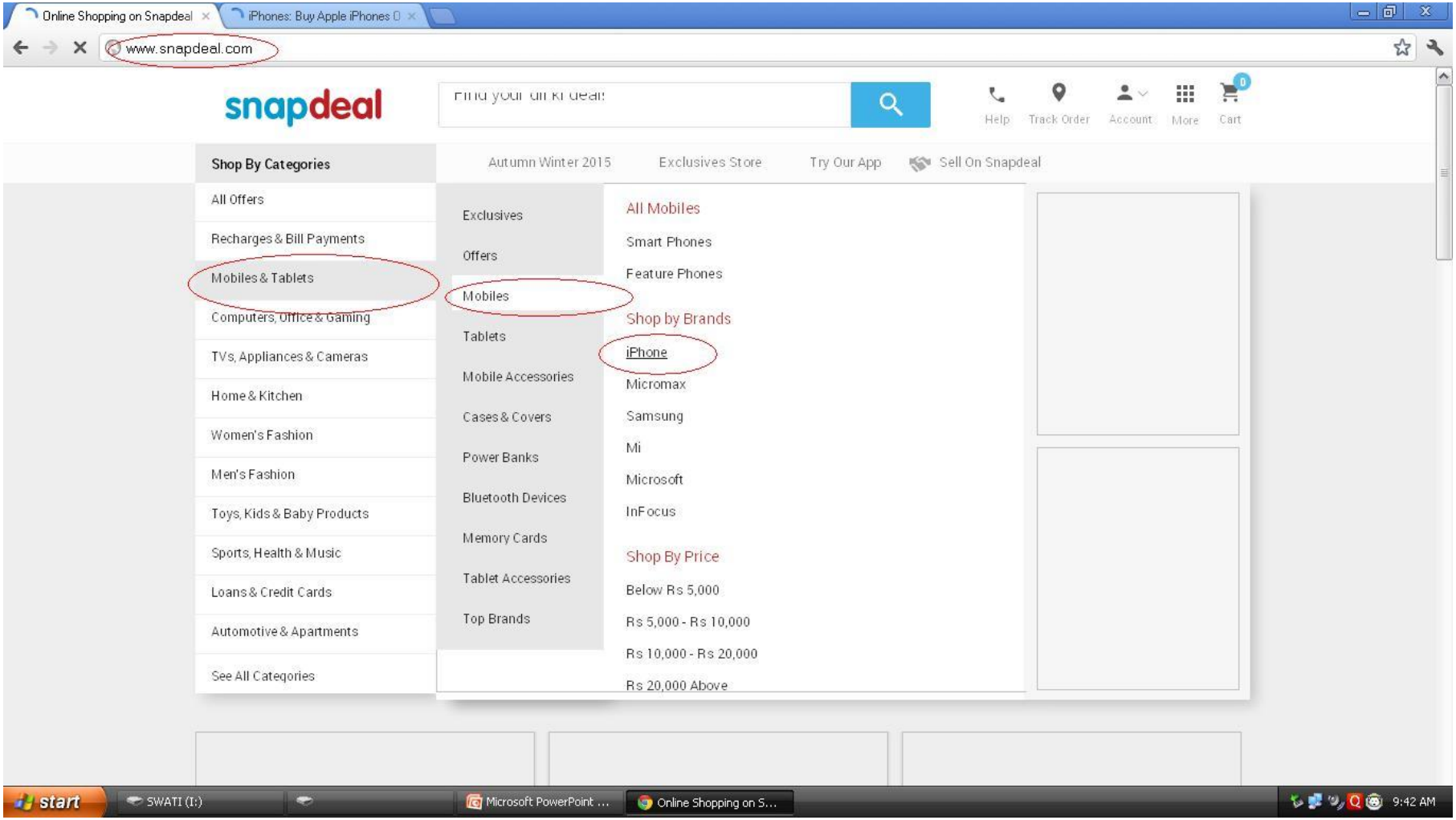


Filing cabinet example

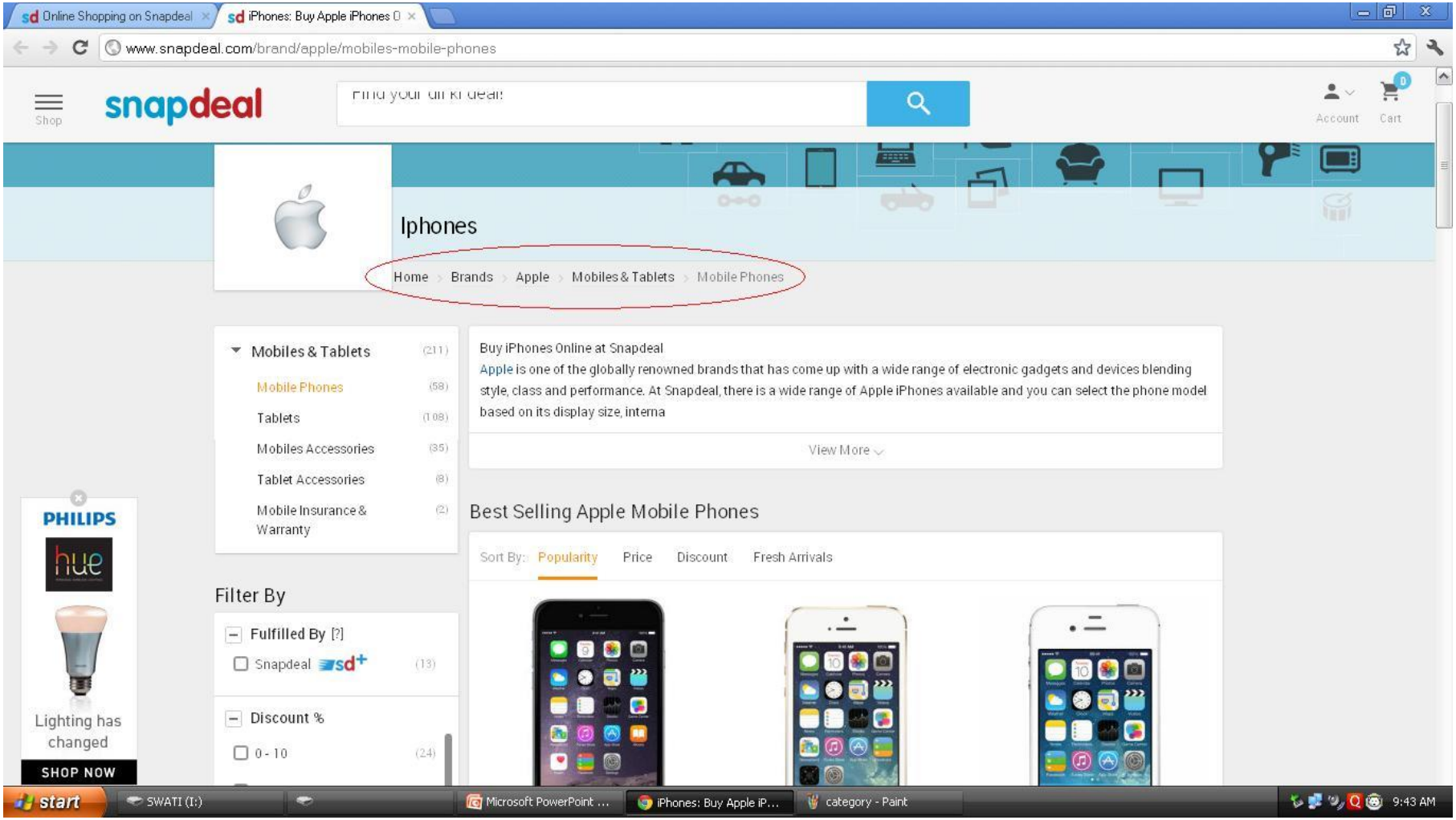
- When you apply this logic to the website, it will look like this :

Filing Category		Site Analogue
Cabinet	→	Fully Qualified Domain
Drawer	→	Primary Site Category
Folder	→	Secondary Site Category
File	→	Tertiary/third Site Category
Document	→	Individual Page/ Item

Snapdeal.com example



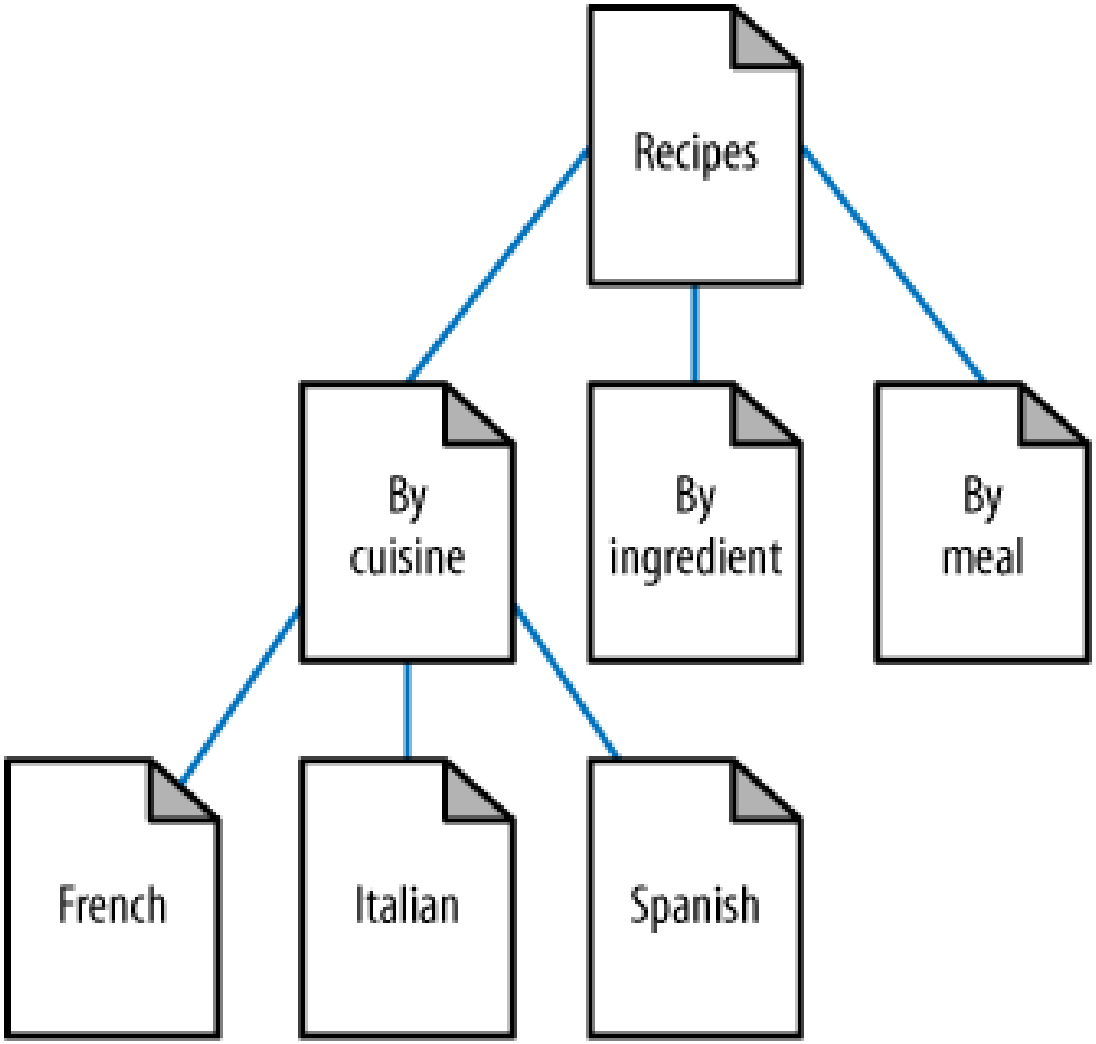
Snapdeal.com example



Site Architecture Design Principles

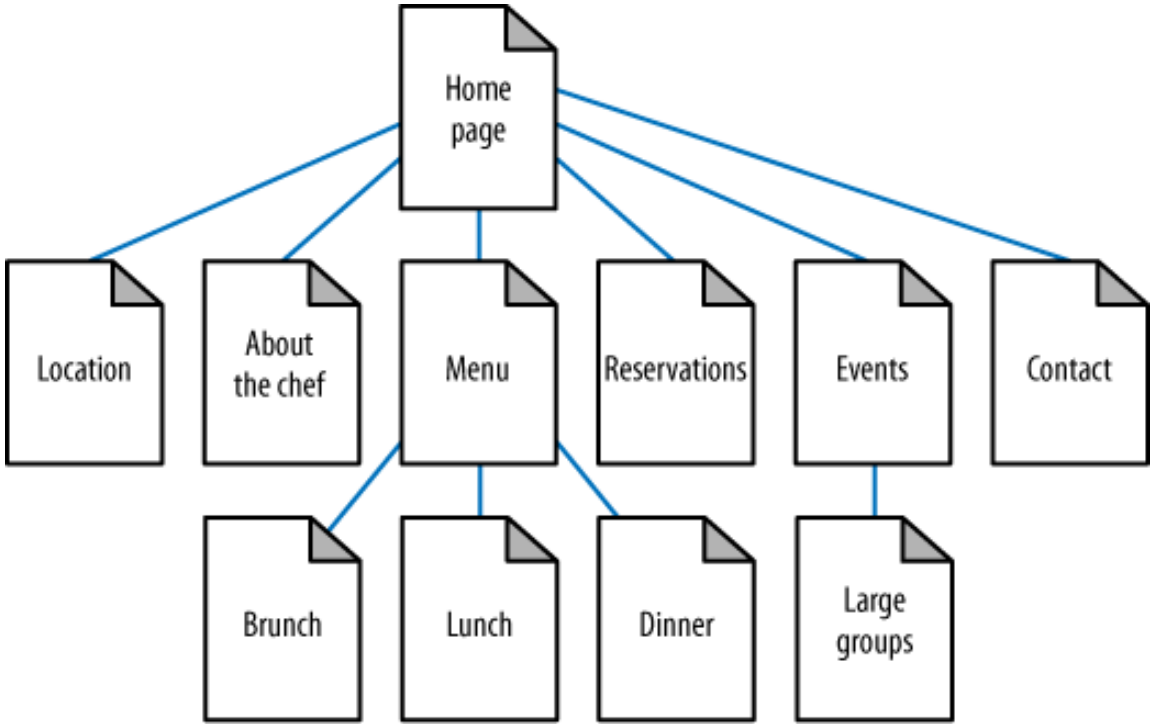
- In conducting website planning, remember that nearly every user will initially be confused about where to go, what to do, and how to find what he wants.
- A well-organized site architecture helps solve these problems and provides usability benefits to both users and search engines.

Recipes Website Example



Designing site architecture

- Site architecture—the creation structure and flow in a website's topical hierarchy



Taxonomy and ontology

- Taxonomy – two - dimensional hierarchical model of the architecture of the site (classification)
- Ontology - the way the human mind thinks about a topic area – time consuming though high ROI

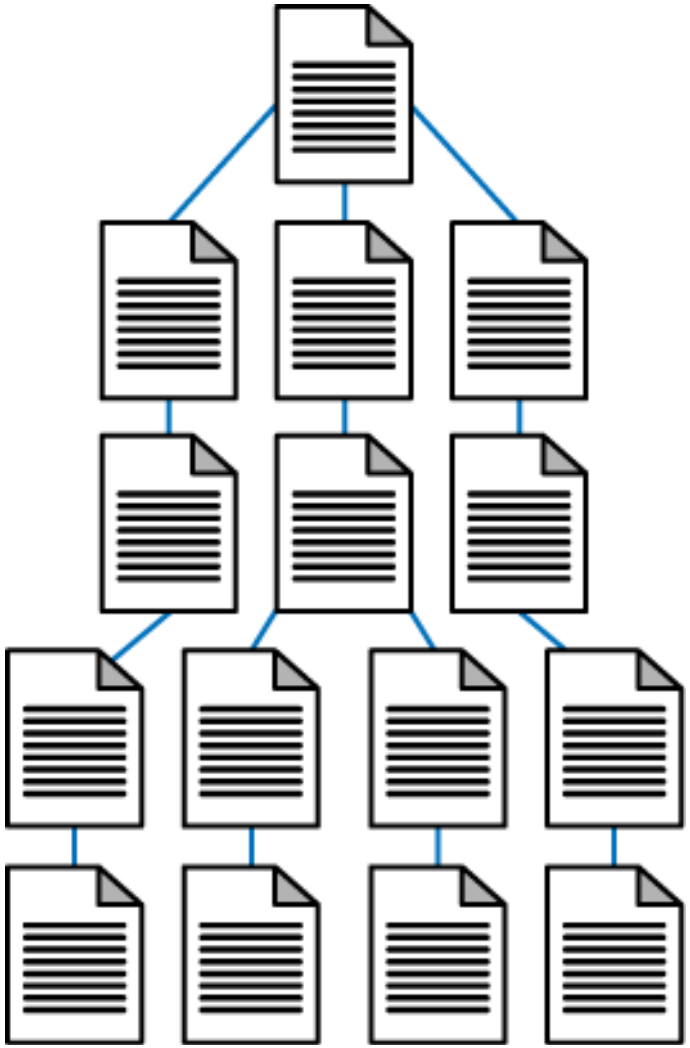
Flat Versus Deep Architecture

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

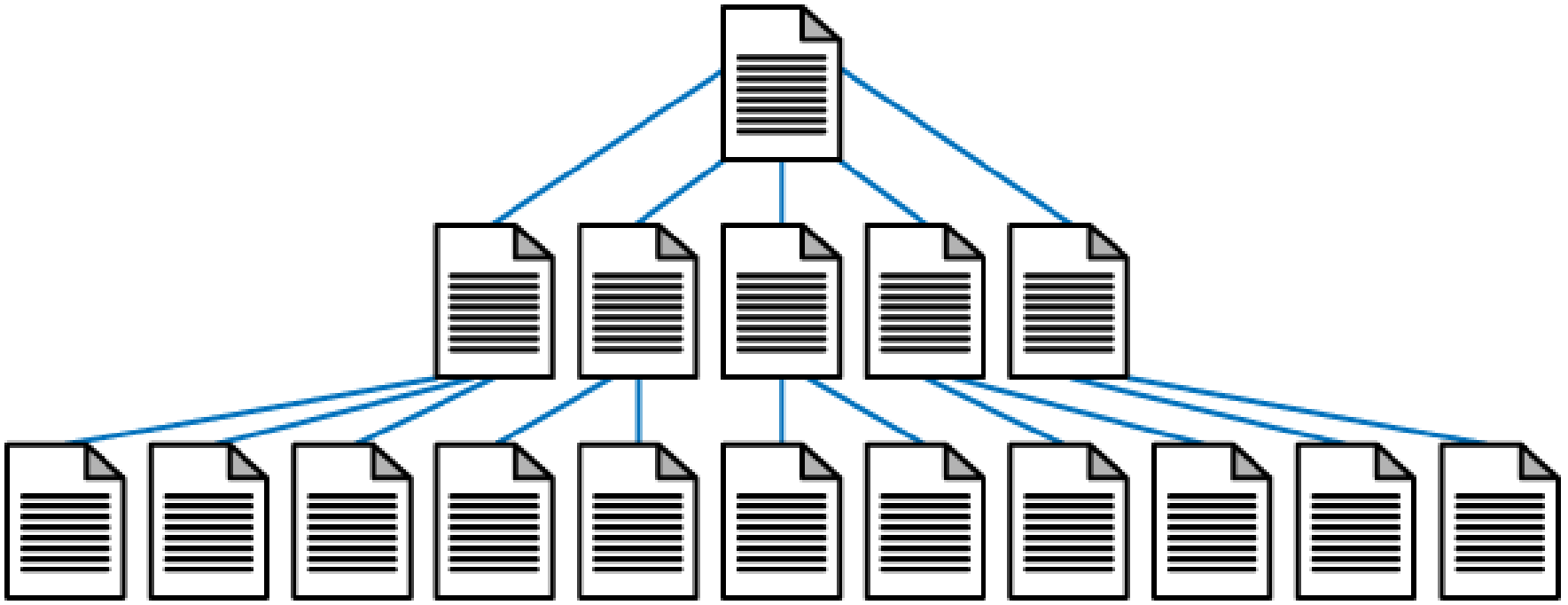
Flat Versus Deep Architecture

- All content should be accessible - three clicks from the home page and/or sitemap page
- If a site is not built to be flat, it can take too many clicks to reach the desired content.

Deep Site Architecture



Flat Site Architecture

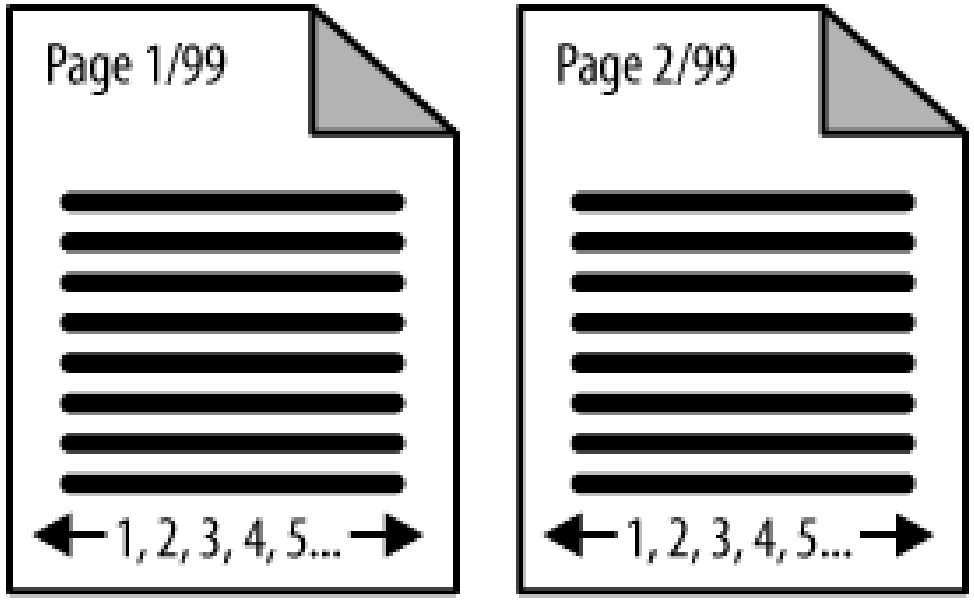


Avoid Pagination

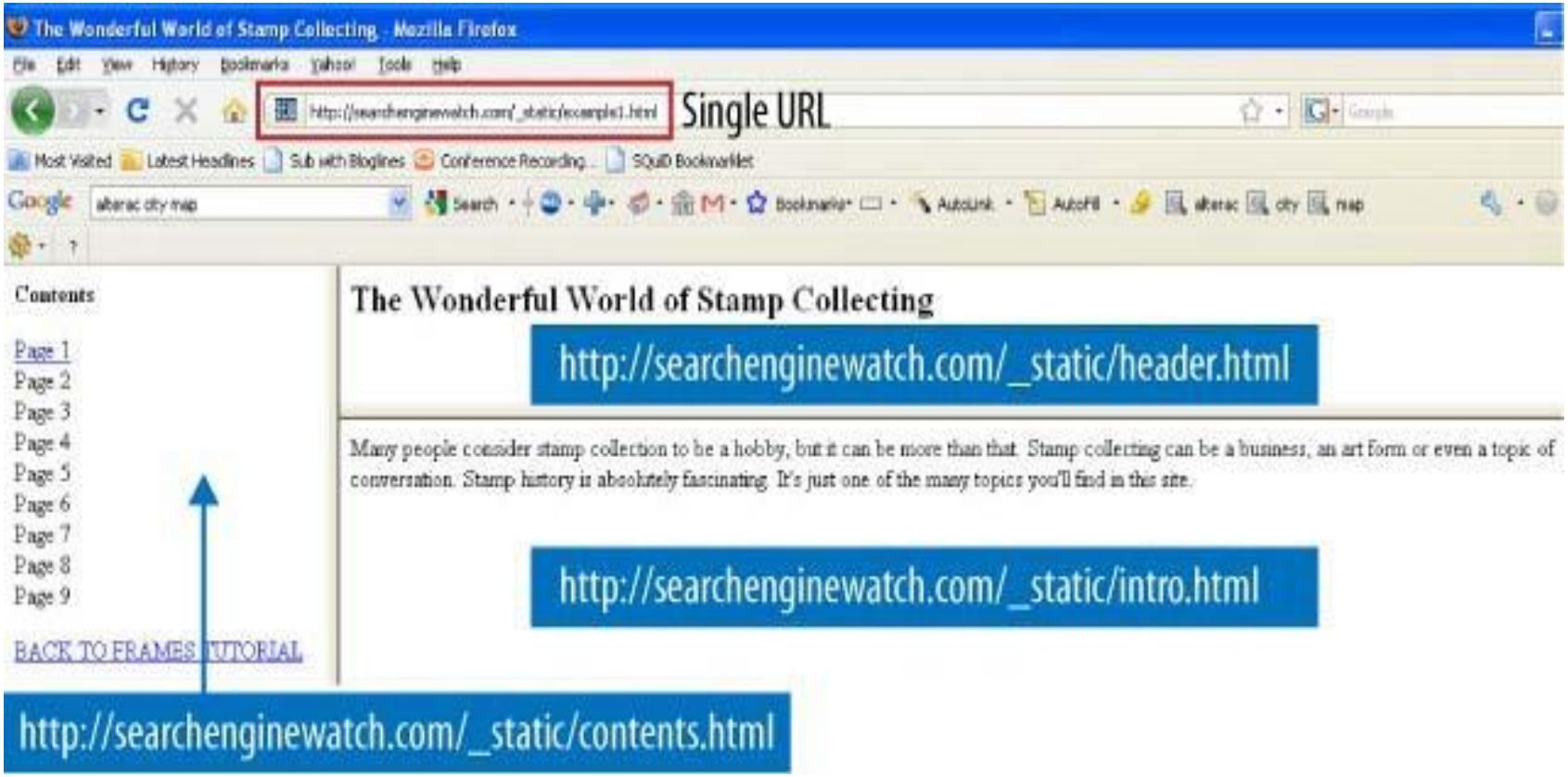
- Pagination - the practice of creating a list of elements on pages separated only by numbers
- Some e-commerce sites use pagination for product catalogs that have more products than they wish to show on a single page

Avoid Pagination

- Pagination is problematic for many reasons.



Search-Friendly Site Navigation



Root Domains, Subdomains & Microsites

- As search engines crawl the Web, they identify four kinds of web structures on which to place metrics :
- **Individual pages/URLs**
- A typical URL might look like http://www.yourdomain.com/page.html.

Root Domains, Subdomains & Microsites

- Subfolders
- In the URL http://www.yourdomain.com/blog/post17.html, “/blog/” is the subfolder and “post17.html” is the name of the file in that subfolder.

Root Domains, Subdomains & Microsites

- **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".

Root Domains, Subdomains & Microsites

- The third-level domain is sometimes referred to as a subdomain.

Root Domains, Subdomains & Microsites

- **Complete root domains/host domain/pay-level domains (PLDs)/second-level domains**
- The domain name you need to register and pay for is the second-level domain.
- In the URL http://www.yourdomain.com/page.html, “yourdomain.com” is the second-level domain.
- Sometimes – Pay Level or Root Domain

Root Domains, Subdomains & Microsites

- Use **sub domains** for larger topics - i.e. if you have a healthy site, it makes sense to create separate domains for each major group of diseases and then create a sub folder for the separate diseases in this group (i.e. heart attack, flu, etc.)

Root Domains, Subdomains & Microsites

- Also, if your site has a forum and or/a a blog in addition to your corporate pages, you should use **sub domains** for the forum and the blog - i.e. *forum.domain.com* and *blog.domain.com*.

What is Microsite?

- A microsite is a brand-specific website (or a single web page) that companies use to promote their individual products, events, or campaigns.
- These are hosted on their own domain or company's subdomain and have a different URL than their company websites.
- Examples: dominosdxp.com | dominos.co.in

Making the case for microsites

- **When should you consider a microsite?**
- When you own a specific keyword search query domain
- For example, if you own “usedtoyotatrucks.com”, you might do very well to pull in search traffic for the specific term used toyota trucks with a microsite.

Making the case for microsites

- **The reasons for not using a microsite:**
- Search algorithms favor large, authoritative domains
- Multiple sites split the benefits of links.
- Time and energy are better spent on a single property
- For more detail : https://vwo.com/blog/what-is-a-microsite/

When to Use a TLD Other Than .com

- When you own the .com and want to redirect to an .org, .tv, .biz, and so on, possibly for marketing/branding/geographic reasons
- When you can use a .gov, .mil, or .edu domain

When to Use a TLD Other Than .com

- When you are serving only a single geographic region and are willing to permanently give up growth outside that region (e.g., .co.uk, .de, .it, etc.).
- When you are a nonprofit and want to distance your organization from the commercial world, .org may be for you.

Optimization of Domain Names/URLs

- **Optimizing Domains**

1. Brainstorm five top keywords
2. Make the domain unique
3. Choose only dot-com available domains
4. Make it easy to type & easy to remember
5. Keep the name as short as possible
6. Avoid trademark infringement
7. Set yourself apart with a brand
8. Reject hyphens and numbers

Keyword Targeting

- **Title tags**
- Place your keywords at the beginning of the title tag
- Limit length 70 characters
- Target longer phrases if relevant
- Focus on CTR
- Target searcher intent
- Meta description tags

Document Text

- A page laden with repetitive keywords attempting to please the engines will provide a very poor user experience; thus, although some SEO professionals today do claim to use term weight.
- Two to ten times, depending on length

Image Filenames and Alt Attributes

- Incorporation of images on web pages - enrich the user experience
- However search engines cannot read the images directly
- Two elements that you can control to give the engines context for images :
- *The filename*
- *Image alt text*

The filename

- Clues to the content of the image
- Don’t name your image example.com/img4137a-b12.jpg, as it tells the search engine nothing at all about the image
- Keyword-rich text
- If it is a picture of the new launched iPhone7, name the file iphone7.jpg and/or have the SRC URL string contain it, as in example.com/iphone7/portrait.jpg.

Image alt text

- Can provide more information about what is in the image
- Can use your targeted keywords
- <img alt=“iPhone 7 photo" src="http://example.com/iphone7.jpg" >
- Use the quotes if you have spaces in the text string of the alt content!

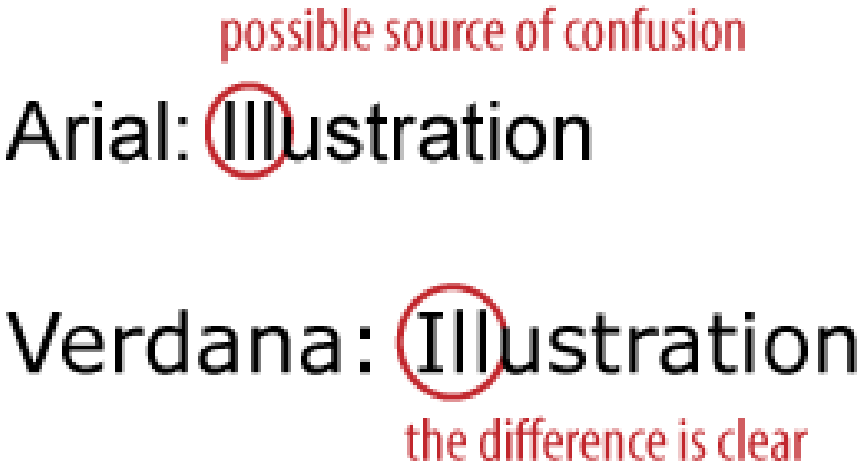
Avoid keyword stuffing

- "Keyword stuffing" refers to the practice of loading a webpage with keywords or numbers in an attempt to manipulate a site's ranking in Google search results. Often these keywords appear in a list or group, or out of context (not as natural prose).

Content Optimization

- **Content Structure**
- The presentation, style, and layout choices you select for your content—is a part of the process
- Choose sans serif fonts such as Arial and Helvetica
- Verdana – more usability/readability , simple, straightforward design, and the characters are not easily confused.

Content Optimization



Verdana: Notice the length of this sentence.

Arial: Notice the length of this sentence.

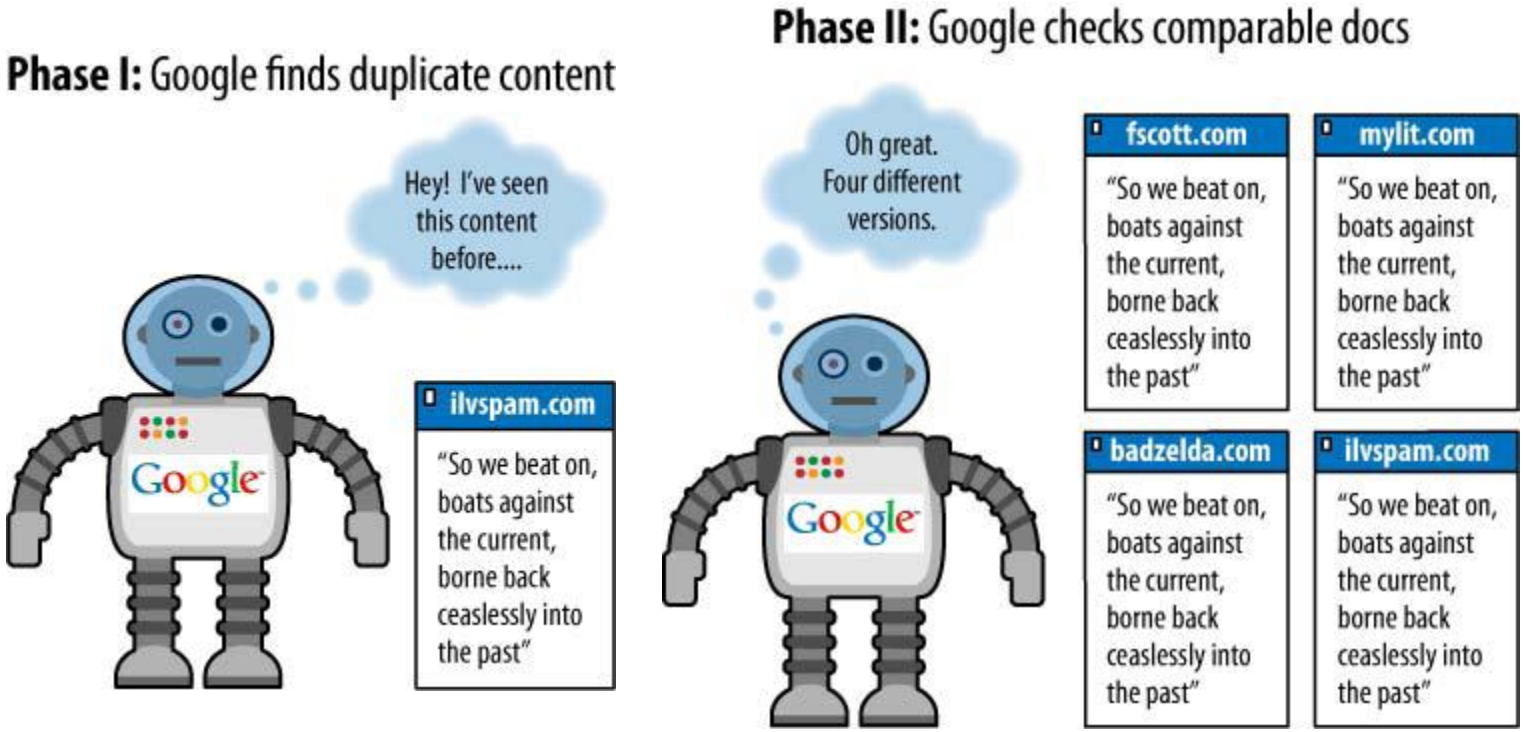
Content Optimization

- **Content length and word count –** Short forms & unique words
- **Visual layout -** Beautiful, simplistic, easy-to-use, and consumable layouts
- **CSS –** 101 kb page size, keeping file size low means faster load times, a higher probability of being fully

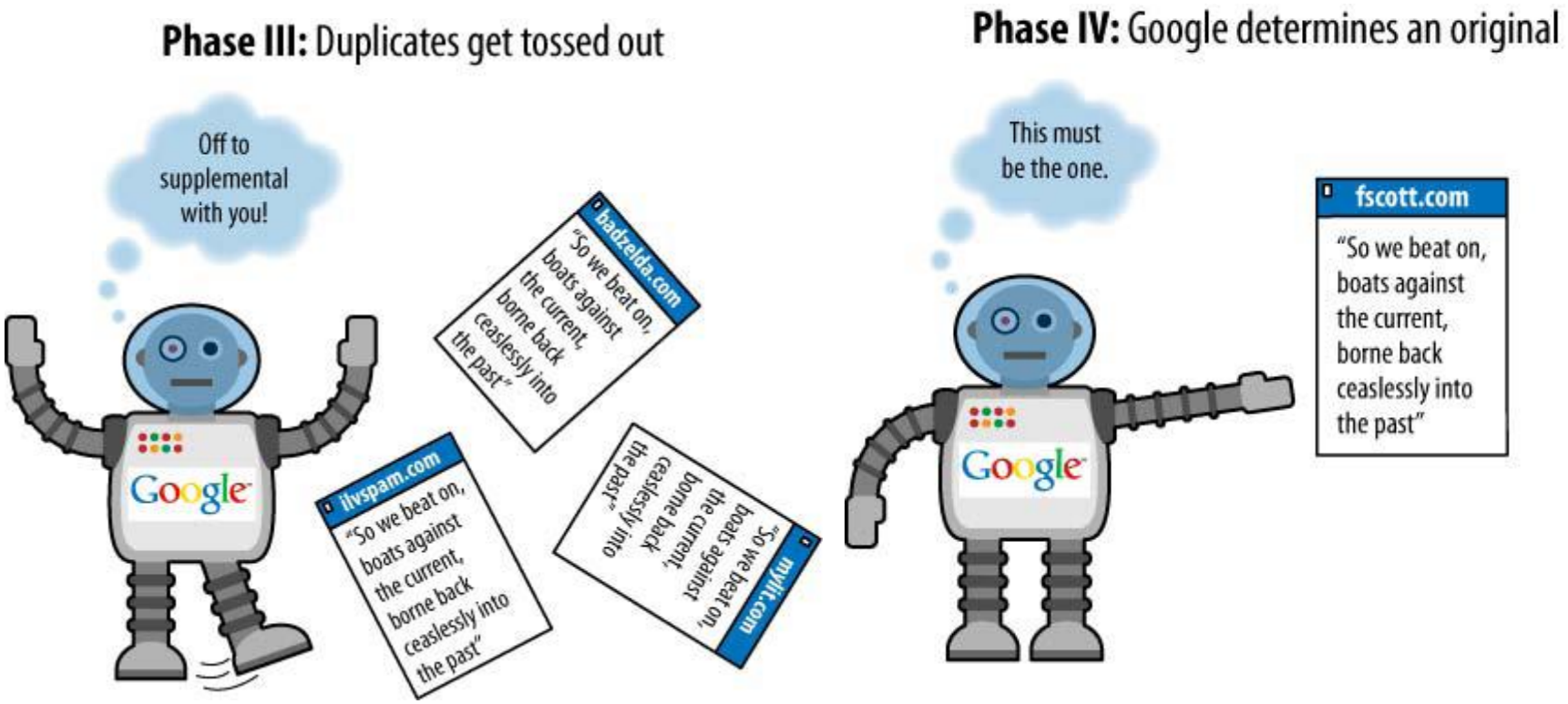
Content Optimization

- **Content Uniqueness and Depth -** Thirty to 50 unique words/sentences that other sites/pages do not have.
- Unique HTML text content, different from other pages on the site in more than just the replacement of key verbs and nouns
- Unique titles and meta description elements
- Don’t simply republish something that’s found elsewhere on the Web

How Search Engines Identify Duplicate Content

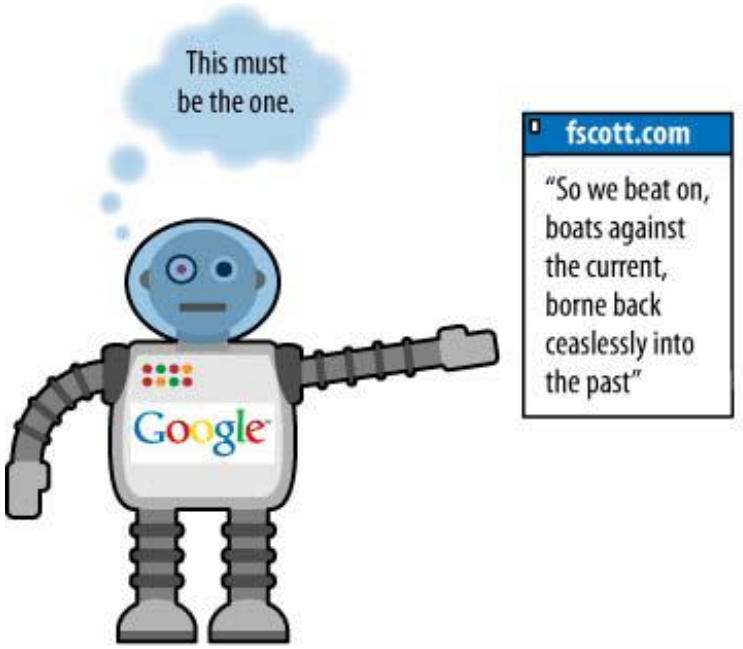


How Search Engines Identify Duplicate Content



How Search Engines Identify Duplicate Content

- Google first saw the content
- Trust
- Most links
- Link to original
- Pagerank



Identifying and Addressing Copyright Infringement

- Use CopyScape.com, a site that enables you to instantly view pages on the Web that are using your content

How to fight with duplicate content issue

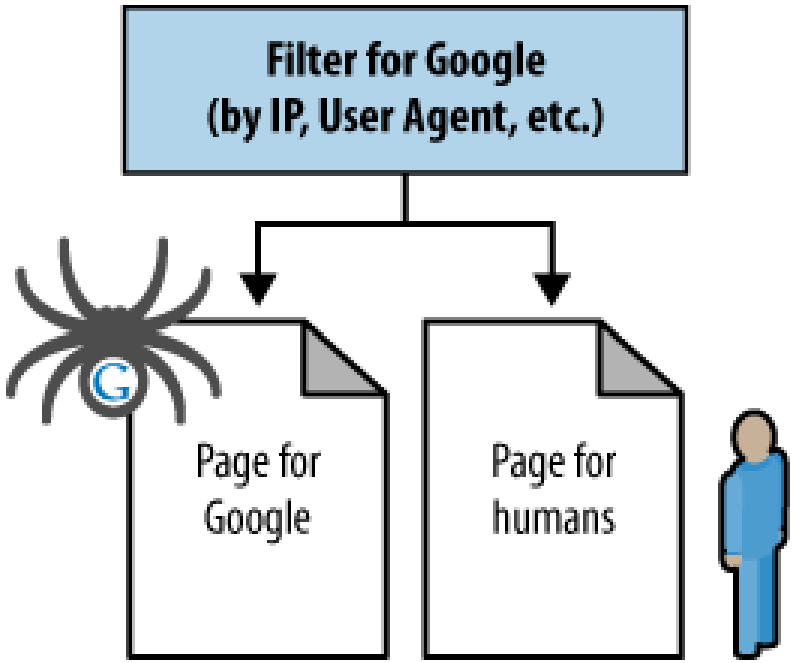
- To file a DMCA infringement request with Google, with Yahoo!, and with Bing also with site's hosting company
- To file a legal suit against the website

How to Avoid Duplicate Content on Your Own Site

- Use the canonical tag
- Use robots.txt to block search engine spiders from crawling the duplicate versions of pages on your site
- Use the Robots NoIndex meta tag to tell the search engine to not index the duplicate pages
- Pagination

Content Delivery and Search Spider Control

- Cloaking - to show search engines one version of content and show humans a different version



Content Delivery and Search Spider Control

- If the engines feel you are attempting to manipulate their rankings or results through cloaking, they may take adverse action against your site.
- The intent of your content delivery doesn't interfere with their goals, you're less likely to be subject to a penalty

When to Show Different Content to Engines and Visitors

- Multivariate and A/B split testing
- Content requiring registration and First Click Free
- Navigation unspiderable to search engines
- Duplicate content

The robots.txt file

- Robots.txt is a text file webmasters create to instruct web robots (typically search engine robots) how to crawl pages on their website.
- Robots.txt files indicate whether certain user agents (web-crawling software) can or cannot crawl parts of a website. These crawl instructions are specified by “disallowing” or “allowing” the behavior of certain (or all) user agents.
- This file is located on the root level of your domain (e.g., http://www.yourdomain.com/ robots.txt)

The robots.txt file

- Thus it is the first document that crawlers open when visiting your site.
- Avoid the indexation of duplicate content on a website
- Using this simple text file, you can easily exclude entire domains, complete directories, one or more subdirectories or individual files from search engine crawling.

The robots.txt file

- **Basic Syntax:**
- User-agent: [user-agent name]
- Disallow: [URL string not to be crawled]
- Crawl-delay: 120 [microsecond]

- Together, these two lines are considered a complete robots.txt file — though one robots file can contain multiple lines of user agents and directives

The robots.txt file

- **Robots.txt file**
  URL: www.example.com/robots.txt
- **Blocking all web crawlers from all content**
- User-agent: *
- Disallow: /
- Using this syntax in a robots.txt file would tell all web crawlers not to crawl any pages on www.example.com, including the homepage.

The robots.txt file

- **Allowing all web crawlers access to all content**
- User-agent: *
- Disallow:
- Using this syntax in a robots.txt file tells web crawlers to crawl all pages on www.example.com, including the homepage.

The robots.txt file

- **Blocking a specific web crawler from a specific folder**
- User-agent: Googlebot
- Disallow: /example-subfolder/
- This syntax tells only Google's crawler (user-agent name Googlebot) not to crawl any pages that contain the URL string www.example.com/example-subfolder/.

The robots.txt file

- **Blocking a specific web crawler from a specific web page**
- User-agent: Bingbot
- Disallow:/examplesubfolder/blockedpage.html
- This syntax tells only Bing's crawler (user-agent name Bing) to avoid crawling the specific page at www.example.com/examplesubfolder/blocked-page.

The robots.txt file

- **Syntax of the robots.txt file**

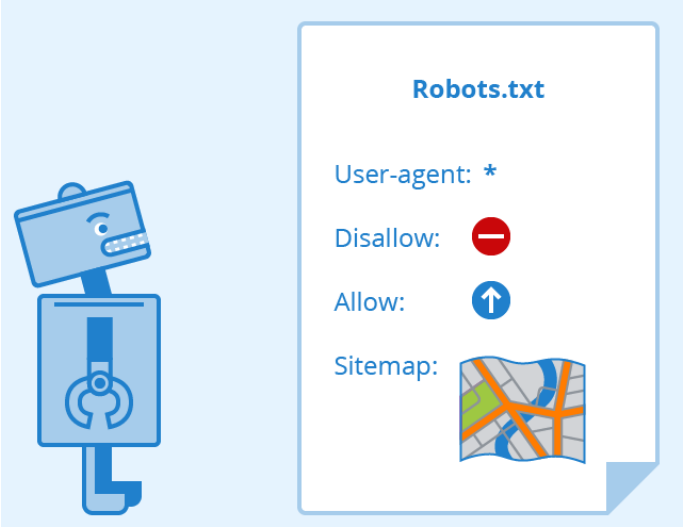
  User-agent: Googlebot

  Disallow:

  User-agent: msnbot

  Disallow: /tmp/

  Disallow: /logs # for directories and files called logs



The robots.txt file

- The preceding example will do the following:
- Allow “Googlebot” to go anywhere.
- Block msnbot from visiting the /tmp/ directory or directories or files called /logs (e.g., /logs or logs.php).

Robots meta tag

- Robots meta directives (sometimes called “meta tags”) are pieces of code that provide crawlers instructions for how to crawl or index web page content.
- **Syntax :**
- <meta name=“robots” content=“noindex, nofollow”>
- To put it in simple terms, this meta tag tells search engines what actions they can take (or not take) on a certain page.
- This meta tag can be included anywhere between the <head> and </head> tags in a page’s header.

Robots meta tag

- Below are the parameters that search engine crawlers understand and follow when they're used in robots meta directives.
- The parameters are not case-sensitive, but do note that it is possible some search engines may only follow a subset of these parameters or may treat some directives slightly differently.

Robots meta tag

- **Indexation-controlling parameters:**
- **Noindex:** Tells a search engine not to index a page
- **Index:** Tells a search engine to index a page. Note that you don't need to add this meta tag; it's the default
- **Follow:** Even if the page isn't indexed, the crawler should follow all the links on a page and pass equity to the linked pages
- **Nofollow:** Tells a crawler not to follow any links on a page or pass along any link equity

Robots meta tag

view-source:https://www.snapdeal.com

```
<meta http-equiv="X-UA-Compatible" content="IE=edge" />
<meta name="google-site-verification" content="60mpvzRXYFjBBgiS8n1nJQ_NF9Sr4mcp08nhnwKQ8qc" />
<meta name="google-site-verification" content="VdxwQvNuNQsYX_-qiUq6XWdt3IVlWB7UE7rjuojiAF4" />
<meta name="msvalidate.01" content="431A1F370506446F4F451D7B16EE166C" />
<meta name="y_key" content="b8b20c3f440781c8" />
<meta name="format-detection" content="telephone=no" />
<meta charset="utf-8"><title>Online Shopping Site India - Shop Electronics, Mobiles, Men & Women C
</title>

        <link rel="canonical" href="https://www.snapdeal.com/" />

        <meta property="fb:admins" content="100002805349917" />
            <meta name="description" content="Online shopping site - Shop Electronics, Mobile, Men
Next Day Delivery ☆ Zero Shipping Charges on SD Gold ☆ www. snapdeal. com"></meta>
                    <meta name="keywords" content="online shopping site, online shopping india, on
                    <meta name="robots" content="index, follow
"/>
                    <meta name="twitter:card" content="app">
        <meta name="twitter:app:country" content="IN"/>
        <meta name="twitter:app:name:iphone" content="Snapdeal"/>
        <meta name="twitter:app:id:iphone" content="721124909"/>
        <meta name="twitter:app:name:ipad" content="Snapdeal"/>
        <meta name="twitter:app:id:ipad" content="721124909"/>
        <meta name="twitter:app:name:googleplay" content="Snapdeal"/>
        <meta name="twitter:app:id:googleplay" content="com.snapdeal.main"/>
        <meta name="twitter:app:url:iphone" content="snapdeal://m.snapdeal.com">
        <meta name="twitter:app:url:ipad" content="snapdeal://m.snapdeal.com">
        <meta name="twitter:app:url:googleplay" content="snapdeal://m.snapdeal.com">
```

The canonical Tag

- If you have a single page accessible by multiple URLs, or different pages with similar content (for example, a page with both a mobile and a desktop version), Google sees these as duplicate versions of the same page.
- Google will choose one URL as the canonical version and crawl that, and all other URLs will be considered duplicate URLs and crawled less often.
- If you don't explicitly tell Google which URL is canonical

What is canonical URL?

- A canonical URL is the URL of the page that Google thinks is most representative from a set of duplicate pages on your site.
- For example, if you have URLs for the same page (for example: example.com?dress=1234 and example.com/dresses/1234), Google chooses one as canonical.
- Note that the pages do not need to be absolutely identical; minor changes in sorting or filtering of list pages do not make the page unique (for example, sorting by price or filtering by item color).

How to Use Canonical Tag

- Mark all duplicate pages with a rel="canonical" *link element*. Add a <link> element with the attribute rel="canonical" to the <head> section of duplicate pages, pointing to the canonical page, like this one:
  <link rel="canonical" href="https://example.com/dresses/green-dresses" />

Why does canonicalization matter?

- Duplicate content is a complicated subject, but when search engines crawl many URLs with identical (or very similar) content, it can cause a number of SEO problems.
- First, if search crawlers have to wade through too much duplicate content, they may miss some of your unique content.
- Second, large-scale duplication may dilute your ranking ability.

Why does canonicalization matter?

- Finally, even if your content does rank, search engines may pick the wrong URL as the "original."
- Using canonicalization helps you control your duplicate content.

The problem with URLs

- You might be thinking "Why would anyone duplicate a page?" and wrongly assume that canonicalization isn’t something you have to worry about.
- The problem is that we, as humans, tend to think of a page as a concept, such as your homepage. For search engines, though, every unique URL is a separate page.

The problem with URLs

- For example, search crawlers might be able to reach your homepage in all of the following ways:
- http://www.example.com
- https://www.example.com
- http://example.com
- http://example.com/index.php
- http://example.com/index.php?r...

The problem with URLs

- To a human, all of these URLs represent a single page. To a search crawler, though, every single one of these URLs is a unique "page.“
- Even in this limited example, we can see there are five copies of the homepage in play. In reality, though, this is just a small sample of the variations you might encounter.

The problem with URLs

- Modern content management systems (CMS) and dynamic, code-driven websites make worse the problem even more.
- Many sites automatically add tags, allow multiple paths (and URLs) to the same content, and add URL parameters for searches, sorts, currency options, etc.
- You may have thousands of duplicate URLs on your site and not even realize it.

Redirects

- Two types of redirects : 301 & 302 both send a human or search engine crawler to the new location but in a different way

Why & When to Redirect

- Redirects are also important for letting search engines know when you have moved content.
- After doing so, the search engines will continue to have the old URL in their index and return it in their search results.
- The solution to this is to implement a redirect. Here are some scenarios in which you may end up needing to implement redirects:

Why & When to Redirect

- 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.

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.

Good and Bad Redirects

- Both forms of redirect send a human or a search engine crawler to the new location, but the search engines interpret these two HTTP status codes in very different ways.
- When a crawler sees a 301 HTTP status code, it assumes it should pass the historical link juice (and any other metrics) from the old page to the new one.
- When a search engine crawler sees a 302 HTTP status code, it assumes it should not pass the historical link juice from the old page to the new one.

Good and Bad Redirects

- In addition, the 301 redirect will lead the search engine to remove the old page from the index and replace it with the new one.
- The preservation of historical link juice is very critical in the world of SEO.
- For example, imagine you had 1,000 links to http://www.yourolddomain.com and you decided to relocate everything to http://www.yournewdomain.com.

Good and Bad Redirects

- If you used redirects that returned a 302 status code, you would be starting your link-building efforts from scratch again.
- In addition, the old version of the page may remain in the index and compete for search rankings in the search engines.

Redirecting a Home Page Index File

- Many websites link to their own home page in a form similar to http://www.yourdomain.com/index.html.
- The problem with that is that most incoming links to the site's home page specify http://www.yourdomain.com, thus dividing the link juice into the site.

Redirecting a Home Page Index File

- Once a publisher realizes this, they will want to fix their internal links and then 301 redirect http://www.yourdomain.com/ index.html to http://www.yourdomain.com/, but there will be problems with recursive redirects that develop if this is not done correctly.

Redirecting a Home Page Index File

- When someone comes to your website by typing in http://www.yourdomain.com, the DNS system of the Internet helps the browser locate the web server for your website.
- How, then, does the web server decide what to show to the browser?
- It turns out that it does this by loading a file from the hard drive of the web server for your website.

Redirecting a Home Page Index File

- When no file is specified, the web server loads a file that is known as the default file. This is often a file with a name such as index.html, index.htm, index.shtml, index.php, or default.asp.
- The filename can actually be anything, but most web servers default to one type of filename or another.
- Where the problem comes in is that many CMSs will expose both forms of your home page, both http://www.yourdomain.com and http://www.yourdomain.com/index.php.

Redirecting a Home Page Index File

- Perhaps all the pages on the site link only to http://www.yourdomain.com/index.php, but given human nature, most of the links to your home page that third parties give you will most likely point at http://www.yourdomain.com/.
- This can create a duplicate content problem if the search engine now sees two versions of your home page and thinks they are separate, but duplicate, documents.
- Google is pretty smart at figuring out this particular issue, but it is best to not rely on that.

Redirecting a Home Page Index File

- Since you learned how to do 301 redirects, you might conclude that the solution is to 301-redirect http://www.yourdomain.com/index.php to http://www.yourdomain.com/. Sounds good, right?
- Unfortunately, there is a big problem with this.

Redirecting a Home Page Index File

- What happens is the server sees the request for http://www.yourdomain.com/index.php and then sees that it is supposed to 301-redirect that to http://www.yourdomain.com/, so it does.
- But when it loads http://www.yourdomain.com/ it retrieves the default filename (index.php) and proceeds to load http://www.yourdomain.com/index.php.
- Then it sees that you want to redirect that to http://www.yourdomain.com/, and it creates an infinite loop.

The default document redirect solution

1. Copy the contents of index.php to another file say sitehome.php
2. Create an Apache DirectoryIndex directive for your document root. Set sitehome.php in it.
3. Put this in an .htaccess file or httpd.conf file in your document root:

```
<Directory      /root/example.com/>
    DirectoryIndex sitehome.php
</Directory>
```

4. Clear the content from index.php & insert this in it

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

The default document redirect solution

- This sets it up so that index.php is not a directory index file (i.e., the default filename).
- It forces sitehome.php to be read when someone types in the canonical URL (http://www.yourdomain.com).
- Any requests to index.php from old links can now be 301-redirected while avoiding an infinite loop.

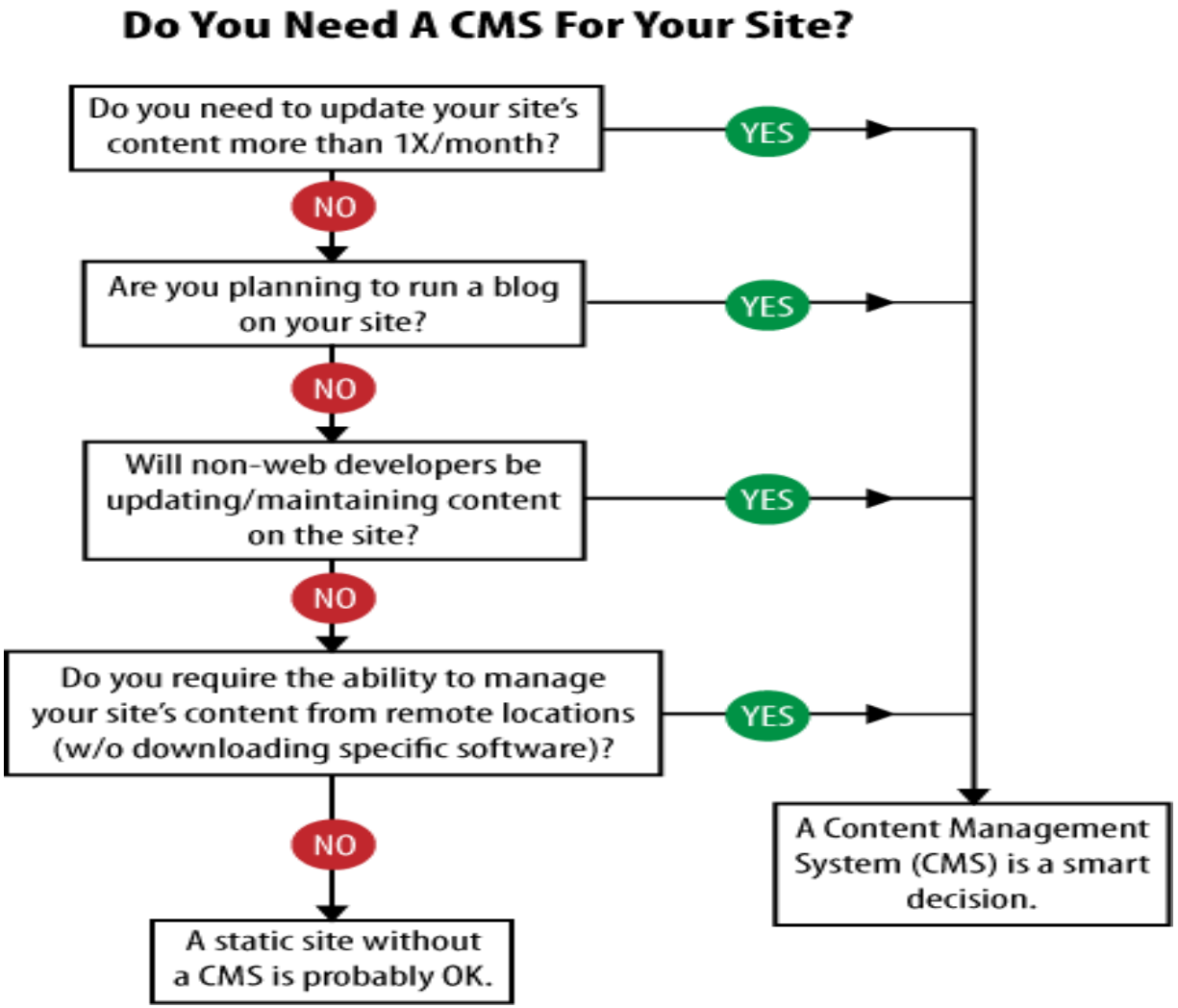
Methods for URL Redirecting

- Server : Apache .htaccess file , Microsoft IIS
- Directly in Programming Language – Perl, PHP, ASP
- Page level redirect - Meta refresh tag, which looks something like this:
- <meta http-equiv="refresh" content="5;url=http://www.yourdomain.com/newlocation.htm" />

Methods for URL Redirecting

- The first parameter in the content section in the preceding statement, the number 5, indicates the number of seconds the web server should wait before redirecting the user to the indicated page.
- This gets used in scenarios where the publisher wants to display a page letting the user know that he is going to get redirected to a different page than the one he requested.

Content Management System (CMS) Issues



How to ensure a CMS is search engine friendly

- Title tag customization & URLs

How to ensure a CMS is search engine friendly

- Meta tag customization & HTML custom tags
- Customized navigational structure
- Pagination controls
- 301-redirect functionality – sadly disallow by many CMS

Content Management System (CMS) Issues

- CSS exceptions
- URLs free of tracking parameters and session IDs or

  Customizable URL structure
- Static-looking URLs
- Tagging and tag clouds
- Breadcrumb navigation
- XML sitemap generator
- Pingback, Trackback, Comments

Optimizing Flash

- Flash is popular on the Web, but each 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.

Flash

- It used to be that search engines did not index Flash content at all.
- In June 2008, Google announced that it was offering improved indexing of this content.
- It indicates that Google can index text content and find and follow links within Flash files.
- However, Google still cannot tell what is contained in images within the Flash file.

Best possible results through Flash

- If Flash is a requirement for whatever reason, there are best practices you can implement to make your site more accessible to search engine spiders.
- **Flash meta tags**
- Beginning with Adobe/Macromedia Flash version 8 - support for the addition of title and description meta tags to any .swf file (ShockWave Format file or Small Web Format) .

Best possible results through Flash

- **Adobe Flash search engine SDK**
- Testing of .swf files to see what search engines are extracting from a given file
- The SDK includes an application named swf2html that extracts text and links from a Macromedia Flash SWF file and outputs it to an HTML document.

Best possible results through Flash

- Internal Flash coding
- Only character-based text will be read (not shape based)
- Search engine spiders do not see dynamically loaded content (text added from an external source, such as an XML file)
- The font size of text does not affect search engines; they read any size font.
- Special characters such as <, >, &, and " are converted to HTML character references (< > & and ") and should be avoided.
- links in Flash should always point to HTML pages, not other .swf files.

Best possible results through Flash

- SWFObject is a small Javascript file used for embedding Adobe Flash content.

- **SWFObject(Javascript library file)**

```
<script type=text/javascript src='swfobject.js'>
</script>
<div id='flashcontent'>
    This text is replaced by the Flash movie.
</div>
<script type=text/javascript>
    var so =new
    SWFObject("file.swf","themovie","200","100","8","#cccccc");
    So.write("flashcontent");
</script>
```

Best possible results through Flash

- **NoScript tags**

  <script type="text/javascript“ src="YourFlashFile.swf">
  </script>

<noscript>

<H1>Mirror content in Flash file here.</H1>

<p>Any content within the NoScript tags will be read by the search engine spiders, including links http://www.sitename.com, graphics, and corresponding Alt attributes.

</noscript>

Best possible results through Flash

- Scalable Inman Flash Replacement (sIFR)
- sIFR is a technique that uses JavaScript to read in HTML text and render it in Flash instead.

Multilanguage/Country Targeting

- 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:

- A country-specific TLD (ccTLD) for your domain (e.g., .co.in)
- Physical local address in plain text on every page of your site
- Google Webmaster Central geotargeting setting
- Verified address with Google Maps
- Links from in-country websites
- Use of the local language on the website

Multilanguage/Country Targeting

- Problems with Using Your Existing Domain
- The Two Major Approaches (.com or .co.in)

Multilanguage/Country Targeting

- Multilanguage Issue
- Here are best practices for targeting the search engines as of this writing, using Spanish and English content examples:

Multilanguage/Country Targeting

- **Content in Spanish and English serving the same country:**
- Create a single website with language options that change the URL by folder structure; for example, http://www.yourdomain.com versus http://www.yourdomain.com/esp/.
- Build links from Spanish and English language sites to the respective content areas on the site.
- Host the site in the country being served.
- Register the appropriate country domain name (for the United States, .com, .net, and .org are appropriate, whereas in Canada using .ca or in the United Kingdom using .co.uk is preferable).