Another material for search engine working.

* How Search Engine Works?

Search engines perform several activities in order to deliver search results.

* **Crawling** - Process of fetching all the web pages linked to a website. This task is performed by a software, called a **crawler** or a **spider** (or Googlebot, in case of Google).
* **Indexing** - Process of creating index for all the fetched web pages and keeping them into a giant database from where it can later be retrieved. Essentially, the process of indexing is identifying the words and expressions that best describe the page and assigning the page to particular keywords.
* **Processing** - When a search request comes, the search engine processes it, i.e. it compares the search string in the search request with the indexed pages in the database.
* **Calculating Relevancy** - It is likely that more than one page contains the search string, so the search engine starts calculating the relevancy of each of the pages in its index to the search string.
* **Retrieving Results** - The last step in search engine activities is retrieving the best matched results. Basically, it is nothing more than simply displaying them in the browser.

SEO techniques are classified into two broad categories:

* **White Hat SEO** - Techniques that search engines recommend as part of a good design.
* **Black Hat SEO** - Techniques that search engines do not approve and attempt to minimize the effect of. These techniques are also known as spamdexing.

## White Hat SEO

An SEO tactic is considered as White Hat if it has the following features:

* It conforms to the search engine's guidelines.
* It does not involve in any deception.
* It ensures that the content a search engine indexes, and subsequently ranks, is the same content a user will see.
* It ensures that a web page content should have been created for the users and not just for the search engines.
* It ensures good quality of the web pages.
* It ensures availability of useful content on the web pages.

Always follow a White Hat SEO tactic and do not try to fool your site visitors. Be honest and you will definitely get something more.

## Black Hat or Spamdexing

An SEO tactic, is considered as Black Hat or Spamdexing if it has the following features:

* Attempting ranking improvements that are disapproved by the search engines and/or involve deception.
* Redirecting users from a page that is built for search engines to one that is more human friendly.
* Redirecting users to a page that was different from the page the search engine ranked.
* Serving one version of a page to search engine spiders/bots and another version to human visitors. This is called **Cloaking** SEO tactic.
* Using hidden or invisible text or with the page background color, using a tiny font size or hiding them within the HTML code such as "no frame" sections.
* Repeating keywords in the metatags, and using keywords that are unrelated to the website content. This is called **metatag stuffing**.
* Calculated placement of keywords within a page to raise the keyword count, variety, and density of the page. This is called **keyword stuffing**.
* Creating low-quality web pages that contain very little content but are
* instead stuffed with very similar keywords and phrases. These pages are called **Doorway or Gateway Pages**.
* Mirror websites by hosting multiple websites - all with conceptually similar content but using different URLs.
* Creating a rogue copy of a popular website which shows contents similar to the original to a web crawler, but redirects web surfers to unrelated or malicious websites. This is called **page hijacking**.

Always stay away from any of the above Black Hat tactics to improve the rank of your site. Search engines are smart enough to identify all the above properties of your site and ultimately you are not going to get anything.

Every SEO strategy is focused on ranking as high as possible in the search engines. To do this, we all try to design and develop a website that Google’s secret algorithm will love. That’s basically what SEO is about. The factors in Google’s algorithm can be divided into two categories which will determine the ranking of your website: on-page factors and off-page factors. Here, I’ll discuss the differences between the two, explain the importance of on-page SEO and go over the most essential on-page SEO factors

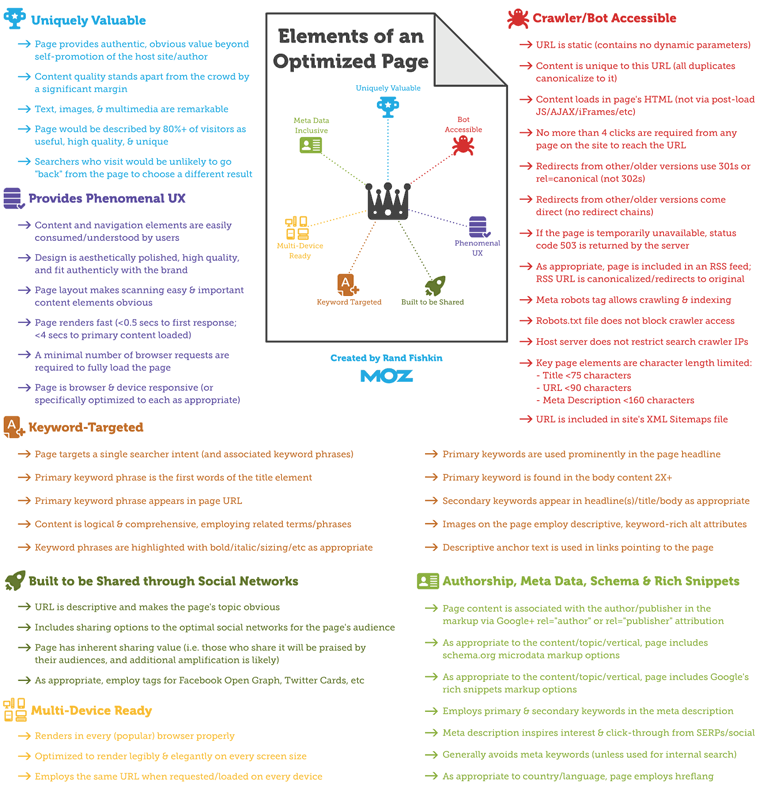
Two types of SEO

* ON page
* Off page

**SEO-On-Page optimization-**

On-site SEO (also known as on-page SEO) is the practice of optimizing elements on a website in order to improve [search engine ranking and visibility](http://moz.com/learn/ranking-visibility). This can involve optimizing both the content and HTML source code of pages on a site.

On page optimization refers to all measures that can be taken directly within the [website](https://www.searchmetrics.com/glossary/websites/) in order to improve its position in the [search rankings](https://www.searchmetrics.com/glossary/rankings/)



## Importance of on-page SEO

On-page SEO consists of all the elements of SEO you can control best. If you own a website, you can control the technical issues and the quality of your content. We believe on-page issues should all be tackled as they’re in your own hands. If you create an awesome website, it will definitely start ranking. Focusing on on-page SEO will also increase the probability that your off-page SEO strategy will be successful. Link building with a crappy site is a very tough job. Nobody wants to link to articles that are badly written or boring.

**Lets now Discuss some On page SEO factors**

1. **WEB SITE DOMAIN**

When you start thinking of doing a business through internet, the first thing that you think about is your website domain name. Before you choose a domain name, you should consider the following:

* Who would be your target audience?
* What you intend to sell to them. Is it a tangible item or just text content?
* What will make your business idea unique or different from everything else that is already avilable in the market?

Many people think it is important to have keywords in a domain. Keywords in the domain name are usually important, but it usually can be done while keeping the domain name short, memorable, and free of hyphens.

Using keywords in your domain name gives you a strong competitive advantage over your competitors. Having your keywords in your domain name can increase click-through-rates on search engine listings and paid ads as well as make it easier to use your keywords in get keyword rich descriptive inbound links.

Avoid buying long and confusing domain names. Many people separate the words in their domain names using either dashes or hyphens. In the past, the domain name itself was a significant ranking factor but now search engines have advanced features and it is not a very significant factor anymore.

Keep two to three words in your domain name that will be easy to memorize. Some of the most notable websites do a great job of branding by creating their own word. Few examples are eBay, Yahoo!, Expedia, Slashdot, Fark, Wikipedia, Google, etc.

1. **Design and layout**

The website design and layout gives the first impression about your site. There are sites which are too fancy and regular net surfers just reach those sites and come out even without creating a single click. Search engines are very smart but after all, they are software and not human being, who can read the content of their interest. If you make your site too complicated, then the search engine would not be able to parse the content of your site properly, and finally indexing would not be efficient, which results in a low rank. The actual page content should have a keyword density(**Keyword density is basically the number of times a specific keyword appearing in the whole page**) of about 10% and should weigh in at about 200 words - but there are as many opinions about this as there are SEO experts. Some say, keyword density should be 5% and some say it should be 20%. You can go with 10% which is good enough.

1. **Optimized Keyword**

A keyword is a term that is used to match with the query a person enters into a search engine to find specific information. Good keyword phrases are specific and descriptive.

## Keyword Frequency

This is calculated as how often does a keyword appear in a website title or description. You do not want to go overboard with frequency, however, since on some engines if you repeat a word too many times, you are be penalized for "spamming" or keyword stuffing.

## Keyword Prominence

It is a measure of how early or high up on a page, the keywords are found. Having keywords in the first heading and in the first paragraph (first 20 words or so) on a page are best.

## Best Places to Put Keywords

Here is a list of places where you should try to use your main keywords.

* Keywords in the <title> tag(s).
* Keywords in the <meta name="description">.
* Keywords in the <meta name="keyword">.
* Keywords in <h1> or other headline tags.
* Keywords in the <a href="http://yourcompany.com">keywords</a> link tags.
* Keywords in the body copy.
* Keywords in alt tags.
* Keywords in <!-- insert comments here> comments tags.
* Keywords in the URL or website address.

1. **Optimized Metatags**

There are two important meta tags:

* Meta description tags
* Meta keyword tags

Some search engines may display the meta description as a part of the search results, but the meta keyword tags should not appear in search results.

The general consensus among SEO experts is that metatags are dead. Even so, many of these same experts continue to use metatags in their own sites.

What Do the Metatags Look Like?

You can add the following in the head section of the web page:

<meta name="keywords"

content="KEYWORD1 KEYWORD2 KEYPHRASE1 etc.

about 30 to 40 unique words">

<meta name="description"

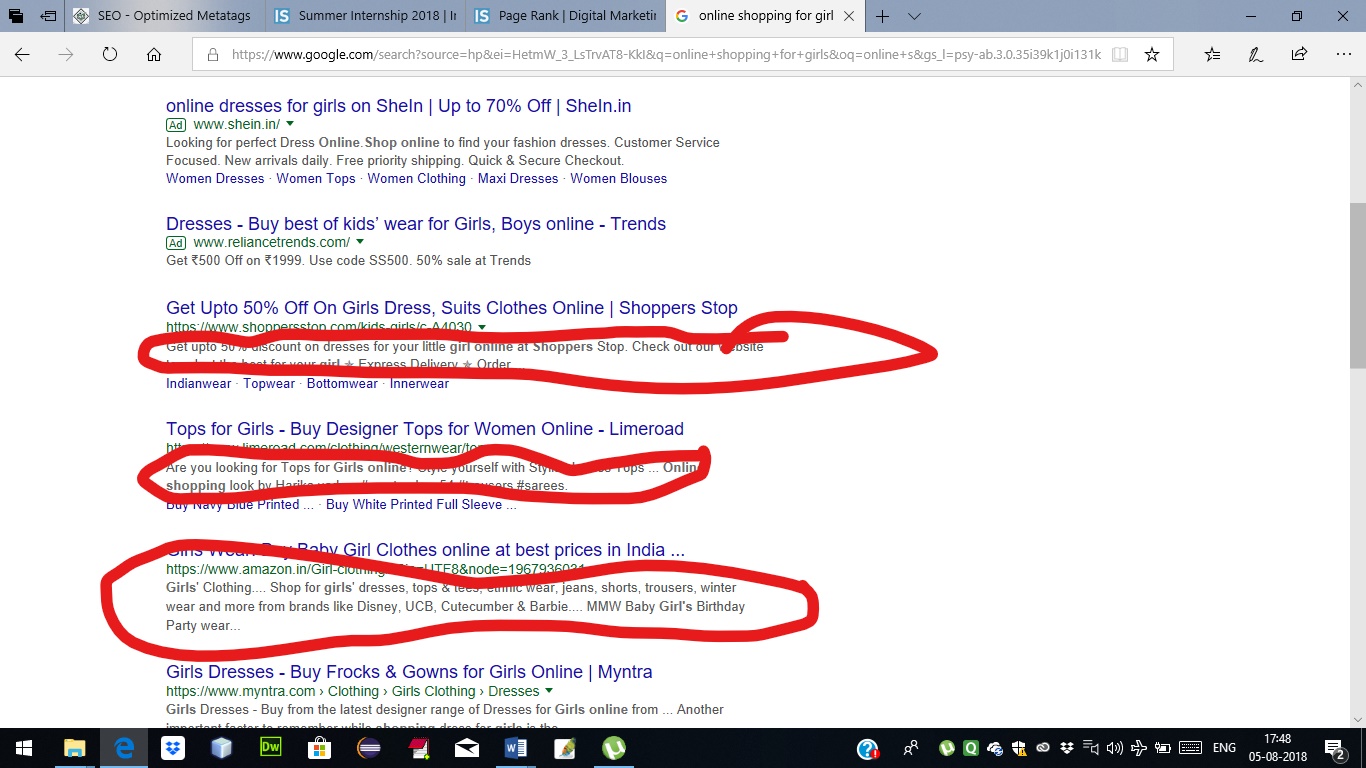
content="An accurate, keyword-rich description

about 150 characters">

## Meta Description Tag Tips

Important tips for good Meta description tags:

* Use keywords in your meta description tag.
* There should not be more than 150 characters in a description metatag of a single web page.
* Use a different meta description tag for each page, as each page is different and stands a better chance of being found if you place a good title and description on it.



1. **Optimized title**

An *HTML TITLE* tag is put inside the head tag. The page title (not to be confused with the heading for a page) is what is displayed in the title bar of your browser window, and is also what is displayed when you bookmark a page or add it to your browser Favorites.

This is the one place on a webpage where your keywords MUST be present. Correct use of keywords in the title of every page of your website is extremely important to Google - particularly for the homepage. If you do nothing else to optimize your site, remember to do this!

Here are some considerations while designing the title of a webpage:

* The title shouldn't consist of more than about 9 words or 60 characters.
* Use keywords at the very beginning of the title.
* Do not include your company name in the title unless your company name is very well known.

Improper or nonexistent use of titles in webpages keeps more websites out of top rankings on Google than any other factor except perhaps for a lack of relevant content on a page or a lack of quality links from other websites that point to your site.

## Best Practices for Creating Titles

Here are some best practices you should follow for creating titles on pages:

* Each page should have a unique title.
* If practical, try to include your Primary Keyword Phrase in every title of every page.
* Begin the title of your home page with your Primary Keyword Phrase, followed by your best Secondary Keyword Phrases.
* Use more specific variations to your Primary Keyword Phrase on your specific product, service, or content pages.
* If you must include your company name, put it at the end of the title.
* Use the best form, plural or singular, for your keywords based on what Word Tracker says is searched on more often.
* Do not overdo it - do not repeat your keywords more than 2 to 3 times in the title.
* Make sure the <title> tag is the first element in the <head> section of your page - this makes it easier for Google to find the page.

1. **Optimized Anchor**

Use descriptive anchor text for all your text links. Most search engines consider anchor text of incoming links when ranking pages. Here is an example of anchor:

<a href="otherpage.htm" title="Anchor Title">Anchor Text</a>

Listed below are some of the important points to note about anchors:

* The *Anchor Title* plays a very important role and is seen by most of the search engines. The anchor title should have appropriate keywords. Anchor title helps the site visitors using a balloon, and displaying written text.
* The *Anchor Text* is another important part, which should be selected very carefully because this text is used not only for search engines but also for navigation purpose. You should try to use the best keywords in your anchor text.
* The *otherpage.htm* is the link to another webpage. This link could be to an external site. Here, you need to ensure that the linked page does exist; otherwise it is called a broken link, which gives a bad impression to search engines as well as to site visitors.

In this case, *Anchor Text* has been replaced by an image. So, while using an image in place of an anchor text, it should be checked that you have put *alt* tag properly. An image alt tag should have appropriate keywords.

1. **Content**

Content basically includes what you see on the site: the text, graphics, and even links to other websites. You should not use excessive graphics because they are not Search Engine Friendly plus heavy graphics normally put the users out when they get downloaded, especially over a slow network.

Thousands of articles, books, and forum entries are available on how to make your website search engine friendly, but ultimately, one rule stands above the rest: *Unique, high-quality, unduplicated content is the king.*

Superior the quality of your content, the higher the ranking you achieve, larger the traffic you gain and greater the popularity of your website. Search engines prefer good quality sites in their index and search results.

Relevant, fresh, and timely content is crucial in attracting visitors to your website. It helps you both draw traffic from search engines and create audience loyalty.

## Unique, High-Quality Content

When people visit a website for information, they want your unique spin on a topic. How is your material or content unique? Is that uniqueness obvious, and easy to find and to understand? Visitors want unique, high-quality site content. It is not only your home page content, but also all the linked pages should have useful and easy-to-understand content.

Now-a-days, search engines have become very smart and they are able to understand complete grammar and complete phrase. Hence while ranking a page against other, the content available on a page matters.

Sites with duplicated, syndicated, or free content are get given red flags by the search engines.

## SEO Content Writing (Copy Writing)

SEO Content Writing (also referred as SEO Copy writing), involves the process of integrating keywords and informative phrases which make up the actual content of your website.

While writing your webpage content, the following tips may help you in keeping it better than others.

* The content should be directed for the specified target audience.
* Keyword density is strictly adhered as per search engine guidelines.
* Titles should always be eye-catching, compelling your visitors to read on and want to know what you offer in your website.
* Do not use confusing, ambiguous, and complex language. Use small statements to make your content more understandable.
* Keep your web pages short.
* Organize and distribute the content on the webpages.
* Divide your web page content also into short paragraphs.

It is not only SEO you need to think about. Many factors contribute to make your site popular.

* If your site is having something really unique, then people like to suggest it to their friends.
* Other webmasters like to create a link of your site on their sites.
* Your site visitors start trusting on your site and they look forward for the next content update and keep coming again and again.
* Although you are listed out by search engine, a but net surfer will click only that page whose content snippet looks more unique and interesting.

Creating, editing, and promoting unique high-quality content is difficult and time consuming. But in the end, the golden rule of SEO is that **Content is the King**. It is not because of a search engine, but it is for your site visitors. A page that is read by people is better than a page that is read by bots.

So, write your content after a serious thought. Keep your title, keywords, link text, metatags up-to-date, unique, and interesting.

1. **Verifying the website**

## Why HTML/XHTML Verification is Required?

There are various reasons to verify your website before hosting it over the internet.

* Any webpage quality depends on how well you have written your webpage. It should be syntactically correct and should pass all the Quality Gates.
* When any search engine does indexing for your web page content, it might get confused if the HTML tags are not written properly, and much of the web page content might not be indexed properly.
* There might be many HTML tags, which you are using in your webpage but then have been depreciated and many of the search engines do not support them.
* Consistency, HTML Code Beauty, Process Compliance are always appreciated by good webmasters.

## What is W3C Compliance?

W3C is the World Wide Web Consortium and since 1994, the W3C has provided the guidelines by which, websites and webpages should be structured and created.

## Rules for W3C Compliance

There following rules, are to be following while developing a webpage.

* Every tag must be closed.
* The head and body tags are now mandatory.
* Empty tags get a terminating slash. An empty tag is a tag that doesn't require an end tag. Examples include <br> and <hr>.
* <BR> is now <br />.
* <HR> is now <hr />.

<IMG SRC="--"> is now <img src="--" />

* All tags must be lower-case. This does not apply to attributes, only tags. For example, both of these formats are acceptable under the XHTML DTD:
* <FONT color="#ffffcc"> is invalid
* <font color="#ffffcc"> is valid

<font color="#FFFFCC"> is also valid

* All the attribute values should be put with in double quotes.
* Tags may not be nested.
* <b><i>Text</b></i> This is invalid

<b><i>Text</i></b> This is valid

* The <pre> tag should not contain: img, object, big, small, sub, or sup.
* One <form> tag cannot be inside another <form> tag.

## What is Word Stemming?(Different Topic in module 2

## )

Google uses a feature called **word stemming** that allows all forms of the word - singular, plural, verb form as well as similar words to be returned for a given search query.

So if someone types in "house plans", not only the pages that are optimized for that phrase but the pages that contain all variations of that phrase are returned. For example, "house plan", "house planning", "house planner".