

# TECHNICAL SEO On Website

## Robots.txt check-

Check if your website is using a robots.txt file. When search engine robots crawl a website, they typically first access a site's robots.txt file. Robots.txt tells Googlebot and other crawlers what is and is not allowed to be crawled on your site.

In order to pass this test you must create and properly install a robots.txt file.

For this, you can use any program that produces a text file or you can use an online tool (Google Webmaster Tools has this feature).

Remember to use all lower case for the filename: robots.txt, not ROBOTS.TXT.

A simple robots.txt file looks like this:

```
User-agent: *
Disallow: /cgi-bin/
Disallow: /images/
Disallow: /pages/thankyou.html
```

This would block all search engine robots from visiting "cgi-bin" and "images" directories and the page "http://www.yoursite.com/pages/thankyou.html"

## TIPS:

- You need a separate Disallow line for every URL prefix you want to exclude
- You may not have blank lines in a record because they are used to delimit multiple records
- Notice that before the Disallow command, you have the command: User-agent: \*. The User-agent: part specifies which robot you want to block. Major known crawlers are: Googlebot (Google), Googlebot-Image (Google Image Search), Baiduspider (Baidu), Bingbot (Bing)
- One important thing to know if you are creating your own robots.txt file is that although the wildcard (\*) is used in the User-agent line (meaning "any robot"), it is not allowed in the Disallow line.
- Regular expressions are not supported in either the User-agent or Disallow lines

Once you have your robots.txt file, you can upload it in the top-level directory of your web server. After that, make sure you set the permissions on the file so that visitors (like search engines) can read it.

**Forms validation-** Your website is running perfectly or not, need to check forms in your websites. We will check the forms from desktop and mobile as well.

## LET'S GET TO THE NEXT LEVEL TOGETHER

First Name*	Last Name*
Email*	Phone*
Message	
Input this code: <b>WYCL</b>	
<input type="submit" value="Submit"/>	

Input some invalid information just to check that forms are not taking invalid information which will reflect a negative image of your brand.

Also check, while submitting the right information, if a genuine customer is submitting his request and not able to reach us.

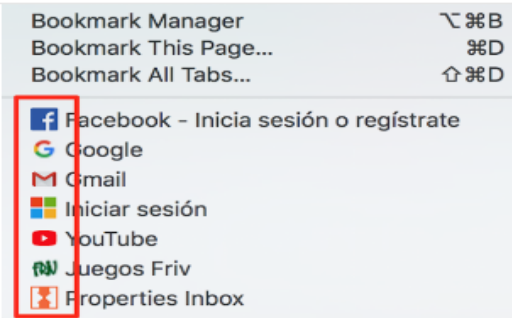
**Canonicalization-** If you have a single page that's accessible by multiple URLs, or different pages with similar content. Google sees these as duplicate versions of the same page. Google will choose one URL as the *canonical* version and crawl that, and all other URLs will be considered *duplicate* URLs and crawled less often.

Even if you aren't actively creating duplicate pages for cross-channel promotion, duplication happens naturally due to the different parts of a web address. Does your website use a secure protocol, like HTTPS? Then you may have an HTTP and HTTPS version of your webpage.

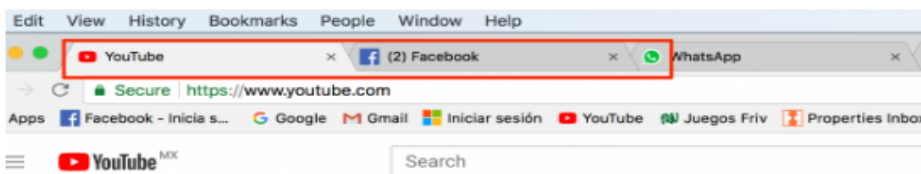
**Favicon-** A favicon is a small 16×16 pixel icon that serves as branding for your website. Its main purpose is to help visitors locate your page easier when they have multiple tabs open. Due to their tiny size, favicons work best as simple images or one-to-three characters of text.

Where can you see the Favicon?

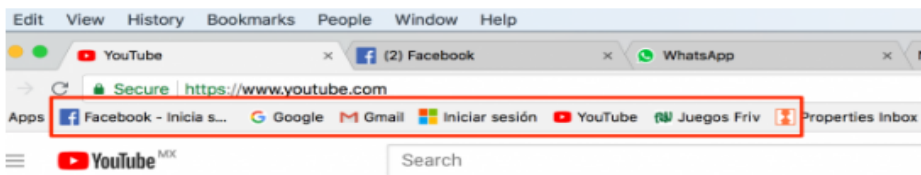
Favicons are found next to anything that identifies your website. This includes bookmarks, tabs, toolbar apps, history results, and search bars. Take a look at the picture examples below:



• **Browser tabs**



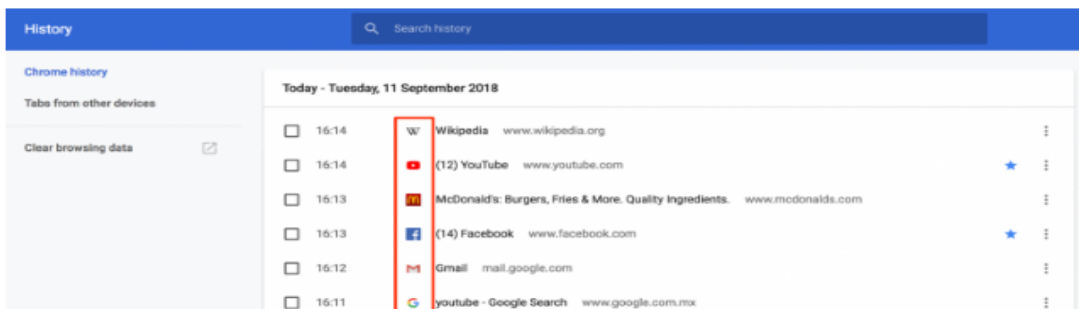
• **Toolbar Apps**



• **History drop down menu**



• **Browser history**



• **Search bar**

