

Friendly URLs in the CMS and Power of Global Ranking with Crawlers with Added Security

Apoorva Ganapathy

Senior Developer, Adobe Systems, San Jose, California, USA

Corresponding Contact:

Email: apoorva.kangodu.ganapathy@gmail.com

ABSTRACT

The web space has the capacity for constant evolution in a dynamic world like ours; hence keeping up to date with the amounts of data that keeps trickling into its space per second requires an apt introspection of the fundamental tool designed for such task. The URL is saddled with the duty to safe keep most of the informational resources. So a functional and reliable one always gives users reasons to bookmark and return to it on a future date. This very writes up gave an in-depth definition of what a URL entails and the appropriate way of using it, which we argued should be: precise, comprehensive, explicit, impressionable, and long-lasting. The relevance of this fine internet tool was also explained in much detail, ranging from its ability to enhance user's experience, facilitate good ranking for the web owner, and ensure smooth linking, which others would give a favorable rating. Conclusively, the importance of a URL is that it enables the SEO and other websites linking to your site and your visitors to appreciate the fact that you care for their informational needs and have taken it upon yourself to provide a very dependable platform to provide solutions.

Key words

Uniform Resource Locator (URL), Crawlers, Ranking, Search Engine Optimization (SEO), Metadata

12/31/2017

Source of Support: None, No Conflict of Interest: Declared

This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. Attribution-NonCommercial (CC BY-NC) license lets others remix, tweak, and build upon work non-commercially, and although the new works must also acknowledge & be non-commercial.



INTRODUCTION

The Uniform Resource Locators commonly referred to as URLs, are simply the addresses of the various pages on the internet space. Strangely, these URLs offer apparent meanings of the design of your site and support the connections and affiliations existent between assets.

Likewise, one can see them as the methods by which different locales discover the enablement to connect to your content and access the material data within it. It's additionally a method through which you will urge clients to visit explicit pages inside your site. It is mainly advised that one chooses its URLs very wisely.

WHAT IS URLS

The web space can be called a home for an unimaginable amount of resource materials and information that are important to people? Gaining entrance to these online resources requires a tool that enables smooth navigation and surfing (Donepudi, 2017). This vital tool is primarily understood to be a popular acronym, especially to those conversant with the basics of the Internet, and it is called URL. This acronym stands for Uniform Resource Locator. This Internet surfing tool is nothing more than the address of a given peculiar informational resource on the Web. It will interest you to know that theoretically that each valid URL refers to a unique resource. No doubt these resources can be an HTML page, a CSS document, an image and even many more. Be that as it may, in practical terms, this is usually not the case as there exist some significant exceptions. The most common amongst them is a URL pointing to a resource that no longer exists or has moved.

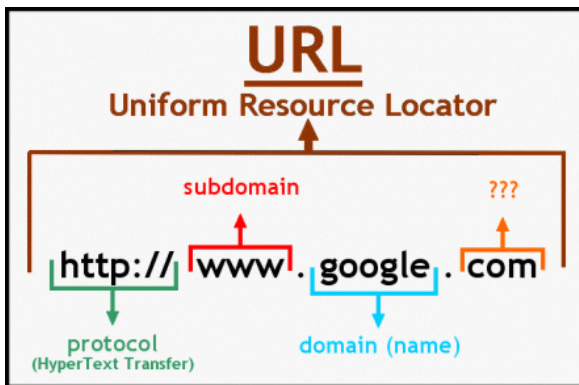


Figure 1: Uniform-resource-locator-(URL), Source: codebridgeplus.com

It should be noted that the URL represents this very resource, and the Web server handles the URL itself. Consequently, it becomes the primary responsibility of the webserver owner to manage that resource and its associated URL cautiously.

In common parlance, it is often said that when the purpose of a thing is not understood, abuse of it becomes inevitable. There are fundamental ways through which one can use the URL and they include:

Make it very Simple and Readable

You should ensure that the URLs are readable and straightforward for the thousands of potential users that would want to access information on the Internet through it. You should even utilize very day to day language words in any possible way and desist from using unfamiliar acronyms and unnatural abbreviations. Whilst trying to maintain this very golden rule, it is essential that, at the same time, you keep things as short as practicable.

Orderly Structure

Having a devised an orderly structure, it is expedient that for fair usage, the index levels suggested by the utilization of forwarding cuts need to count to the course of action of segments and sub-segments in your site marine construction (Ganapathy, 2016). This supports the precision and connections in the client's psyche and adds to your site's evident dependability. Eliminate whatever recommends auto redirection on your site as not exclusively does such redirection look amateurish. Yet, individuals will surely bookmark the resultant URL. When you change the back finish of your site, there's a danger that those bookmarks will quit working and make great uncertainty in clients' psyche with regards to its unwavering quality.

Uniqueness and Originality

Originality is essential in the usage of your Uniform Resource Locator. Situations whereby the same content or resource can be seen on two or more different addresses, should be avoided. Two compelling reasons can be adduced for this. First, you will discover that search engines will decidedly punish duplicate content, or better still, decide to share your ranking across those multiple addresses, and this will have a negative toll on your so much desired ranking. Also, there is the 'issue of trust' for clients. Peradventure circumstances emerge where various URLs return a similar material from a similar record; how can one anticipate that the users should comprehend this? There is an extraordinary chance that they would be left pondering, legitimate and genuine, and climate one form may quit working or not be refreshed (Neogy & Paruchuri, 2014). In any case, it doesn't abandon the way that you can't permit individuals to get too similar content utilizing at least two separate locations.

Permanence and Apparent Changes

It is not strange that internet users have developed the attitude of bookmarking pages, and of course, other sites will link to your content. Now imagine that your URL links are not durable, and yet someone had bookmarked it because of a critical resource discovered? Such lack of permanence would breed distrust for any links associated with your URL. Hence, if these links are to be of any use, they need to continue to function for months or possibly even years into the future because of a likely reference to it.



Figure 2: Importance of URL (Source: www.savvysites.com)

Online Resource with Dates and Archives

Many resource materials exist on the web space in the same fashion as Uniform Resource Locators; this is deemed intelligent on your part to include standard dates to enable adequate differentiation (Vadlamudi, 2016). This same date includes ensuring simplicity of arranging and language autonomy, for it is considered a decent practice to put together any 'dated' URLs concerning the global standard data. The registry levels ought to have appropriate list pages permitting clients to get to those data even after months and pages under that specific level. Then again, it is significant that you ought to try not to utilize dated structures for things where individuals will anticipate the most recent form and would have practically no incentive for the more seasoned renditions.

3 THE IMPORTANCE OF URLs FOR USERS

The importance of the URL to the internet user or its user can never be over-emphasized because of its indispensability in guaranteeing access to the informational resource. As it is often said, the first impression is the last. This beautifully apt and witty aphorism can be likened to the website as the very first impression of the website is undoubtedly the Uniform Resource Locator. In any situation whatsoever in which the URL appears to be confusing and shrouded with ambiguities, likely, people will not take any genuine interest in the website, and the bounce rate will be so alarmingly high. Consequently, several benefits are derivable by the users when they find a highly effective and efficient URL; the importance of the URL to its users includes the following:

Enhancing User Experience

One would be tempted to quiz in his mind, how does a mere tool enhance the user experience? It is straightforward; for instance, if a particular URL's design happens to be correct, it certainly would be beneficial for both humans and search engines. It, as a rule, helps simple and less complicated comprehension of what is the issue here. This will eventually give the client such a lot of clearness about what is inside, and the odds of a client tapping the connection would increment.

Good Ranking

Attaining a desirable ranking is essential to web owners. Achieving such a feat requires a lot more than just having a popular URL, as the URL assumes an exceptionally minor part in deciding the position of the page in the search motor and the significance of the question. In any case, a keyword in the URL can upgrade the page ranking in a highly competent manner. Regardless, one should not neglect to recollect that while URL improves the site visibility, the URL's isolation doesn't fundamentally influence the situation of the page. Subsequently, it will be a pointless activity to make URL fundamentally to recall keywords for them. But then it contributes to attaining such feat, hence why most users consider it an essential tool.

Linking

A feasible and legitimate working URL can fill in as a connection while sharing data in different online media channels and different stages, which certainly aids information sharing amongst various individuals, groups and even corporate bodies within a space of time. This linkage capabilities of the URL has ensured smooth content management and optimization within resource-based entities within the walls of the Internet. It turns out to

be simple for you to reorder the connection on different channels and drive more traffic to the site from various media, leading to benefit amplification.

IMPROVING URLS RANKING

Most content managers desire to ensure that they have a very formidable, proper, friendly and functioning Uniform Resource Locator with an excellent speed optimization that provides that users get value for their time whilst using it. The science that assurances having an exceptionally improved and utilitarian Uniform Resource Locator relies on a couple of things that are needed to be remembered while choosing the URL of the site. The importance of these factors and how they aid ranking cannot be overlooked. Hence, the need to do the following:

Be Very Meticulous While Choosing a Domain Name: Most web owners usually adopt a top-level domain name which can help optimally. Though it may not directly impact the ranking, it will undoubtedly build and strengthen trust among users and clients alike. This top-level domain is known as the 'dot.com' domain instead of dot.biz, 'dot.pro,' and so many others. It should, however, be borne in mind that it isn't practical to have the option to arrange your image name with a dot.com' area (in light of the fact that measurably, there were somewhat more than 125 million dot.com' areas in the year 2016), so it becomes very imperative that one keeps this in mind when brainstorming on a domain in the future.

Adding a Keyword in the URL: Another potent way of effectively improving the rating and eventual ranking of a Uniform Resource Locator is by intelligently adding a Keyword in the URL. This same keyword connotes the unique term which the user optimizes to search for content. Keywords on a page must remain very relevant to what most users and visitors are looking for, which would give them a superior possibility of tracking down your content among the probable outcomes, which would be advanced in the end. Remembering the keyword for URL is a highly savvy thought as it will ensure the improvement of the page's ranking. There is no doubt that these are the essential elements most searchers consider when deciding which site to click.

Keep the URL Short and Simple: Many users will get instant discouragement if they discover how lengthy and complex your Uniform Resource Locator appears to be, hence bearing a negative toll on your ranking ultimately. It is along these lines entirely prudent that in having a URL Structure, let it be under 50-60 characters. This is a decent practice, and nothing to stress over except for a URL with 100+ character is certifiably not a brilliant thought for content administration advancement because the vast majority don't have the persistence nowadays and would hastily lock on the most oversimplified elective URL. This is not to say that adopting a very long URL is difficult for search engines to process; the significant concern, however, lies with the users. Long and complex URL hampers client's experience. On the other side, it has been found that more limited URLs are not challenging to reorder, share via online media, and insert.

Use of Hyphens and Lower case letters: The usage of Alpha-numeric and lower case letters certainly have a significant role in ensuring the favorable ranking of a Uniform Resource Locator. Hence, the act of common usage of these characters is highly discouraged. It is significant that while making the URL structure, one ought to develop the propensity for utilizing hyphens for giving spaces between words as opposed to utilizing underscores. One can admirably utilize the spacebar, which can work (Ganapathy, 2015). However, they

render clumsily in URLs at 20%, which takes away from the lucidness of your pages. This has to be avoided at all cost if possible (This is achievable easily in a modern Content Management System). To appreciate an improved URL ranking, one should likewise figure out how to adhere to a lowercase organization while drafting a URL. That is the best system and try not to utilize capitalized as it prompts different blunders.

Simultaneous Connection of Title with URL: It is good that the website's title is not shrouded in ambiguities but very clear and same too for the URL. It is acceptable practice to coordinate with the page's title with the fitting Uniform Resource Locator to keep away from disarray. The title is the feature of the page, and, normally, it should give a reasonable plan to the client about what's in store on the page, and asynchronous association of the title with the URL builds the opportunity of the client tapping on the URL. Framing the URL Structure requires a great deal of diligent exactness and difficult work, and when this is meticulously put in place, certainly it would have a good rating by the users who would think it very helpful and friendly.

FRIENDLY URLS

One would begin to wonder about the relationship between an internet-based tools like Uniform Resource Locator with a concept like friendliness. The truth is some Uniform Resource Locators are instant turn off for most users and researchers because of the way they are designed. When it is dubbed friendly, it means a Web address that is so simple to read and comprises relatable words that describe the content of the webpage. This type of URL is considered "friendly" for two major reasons-It can enable Users to remember the Web address, and; it also explains the page to search engines.

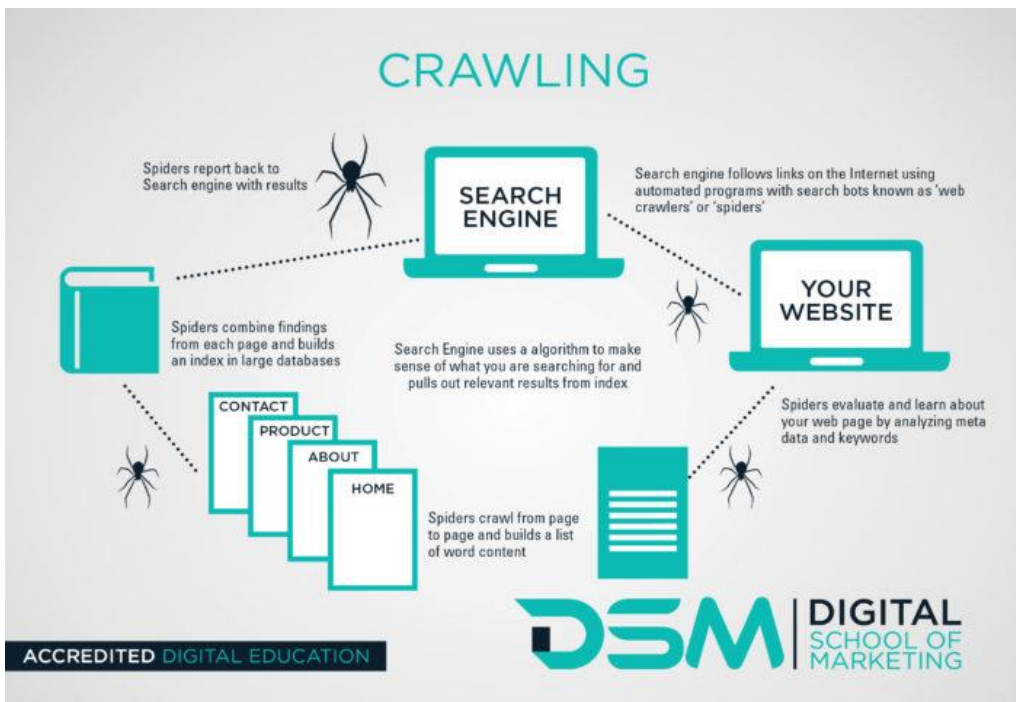


Figure 3: Understanding Crawling (Source: digitalschoolofmarketing.co.za)

Dynamic destinations regularly load specific content depending on the URL factors, making it easy to use to be extremely troublesome. Regardless, generally intelligent and web specialists presently utilize a "URL reworking" system to make less difficult URLs. This very philosophy guarantees that the Web worker stacks an unexpected URL compared to the one in the location bar. This assumes that a basic Web address can highlight a more mind-boggling URL with bunches of factors. It is a fact that user-friendly URLs are helpful for visitors, but then over the years, it has been discovered that most web admins concentrate more on the creation of search engine-friendly URLs. Consequently, these URLs comprise essential keywords that describe the content of the page.

UNDERSTANDING CRAWLING AND SEARCHING

Crawling is an operational process that is usually leveraged by search engine web crawlers (bots or spiders as they are sometimes called). It ought to be noticed that at whatever point a search motor recognizes changes to a page after crawling a page, it can promptly refresh its base in response to the modifications that have been identified. Emblematically, a web crawler bot can be compared to somebody who goes through every one of the books in a not so coordinated library and assembles a card list. Any person who visits the library can rapidly and effectively discover the data need. Presently to help in the arrangement, one would then need to record the books by point; the coordinator will peruse the title, outline, and a portion of the content within each book to discover more about it. It can also be likened to the index usually at the back of a book, which gives point by point blueprint of the relative multitude of spots in the book where a specific theme or expression is referenced. Significantly, you realize that searching focuses, for the most part, on the content that shows up on the page and the metadata about the page that clients don't see.

To secure classified archives or sensitive data, most sites require verification before getting to the important web content. Although this layer of assurance can effectively obstruct outside malignant crawlers, approved insiders can, in any case, crawl the whole site. For example, with full admittance to the NSA's records, Edward Snowden utilized economical and generally accessible web crawler device to scratch at any rate 1.7 million secret documents (To keep from being handily identified, subtle insiders may modify their crawlers to more readily mirror the entrance conduct of genuine clients. To reimburse the deliberately debased download effectiveness, insiders have the persistence to dispatch a determined download in quite a while period. Moreover, numerous insiders may plot and receive a gap and-vanquish procedure to accelerate the crawling cycle.

To recognize and oblige noxious crawlers, analysts have fostered various enemy crawler components, which can be by and large separated into two classes, specifically, **heuristic detection and machine learning detection**. Heuristic detection instruments depend on examining obvious highlights (e.g., visiting pace of individual guest) to characterize unusual site access conduct from the outset and characterize some other conduct as typical conduct. However, they can't identify crawlers that can compel their practices and control those noticed highlights. Machine learning detection components can distinguish malignant crawlers dependent on the distinctive visiting designs between typical clients and vindictive crawlers. The first model the ordinary site access conduct and afterwards characterize some other conduct as unusual. The latest enemies of crawlers systems consolidate these two methods to all the more likely loss of malevolent crawlers. In any case, it is as yet a test to identify and compel shielded inside crawlers that can plot in a relentless scratching.

METADATA

Metadata characterizes other data. It gives itemized realities about a specific thing's content. For example, a picture may comprise metadata that portrays how enormous the image is, the shading profundity, the picture goal, when the picture was made, and other significant data. Additionally, a book archive's metadata may comprise realities about how long the report is, who created it, when the record was composed, and a short rundown of the report.

Data Warehouse Metadata



Figure 4: Metadata (Source: www.theedgearch.com)

Types of Metadata

Meta Data comes in various types, and they include the following:

- Structured metadata:** This is metadata that gives insights regarding information holders and graphically shows how compound articles are collected together, for example, how pages are requested to frame parts. It ordinarily clarifies the sorts, renditions, connections and other fundamental highlights of computerized materials.
- Administrative metadata:** This specific Metadata generally gives out data to empower the administration of an asset, similar to asset type, authorizations, and when and how it was made.
- Reference metadata:** It ought to be noticed that this specific type of metadata contains definite realities about the substance and nature of factual information.
- Statistical metadata:** Another name for this same metadata is process information; it can depict the cycles that gather, interact or produce measurable details.
- Legal metadata:** This specific metadata gives data about the maker, copyright holder, and public authorizing and legitimately related ideas or topic whenever gave.

Importance and Purposes of Metadata

It is worth knowing that metadata serves an avalanche of purposes, with resource discovery one of the most common. Some of the beneficial benefits of metadata has been summarized in this manner. Metadata is a potent means of arranging internet resources, especially with the skyrocketed growth of Web-based resources.

Furthermore, the utilization of metadata is pointed toward working with interoperability and coordinating assets. Improving to portray assets empowers its agreement and cognizance by people just as machines. This ensures the best degrees of how information is mindfully divided between numerous frameworks with particular functional stages, information constructions and interfaces.

RANKING FACTORS: WHAT MATTERS FOR SEO

Some of the factors that matter so much to Search Engine Optimization and boost ranking includes the following:



Figure 5: Ranking-factors (Source: data-flair.training)

Protection of Web Site: Google took it as a mandate upon itself in 2014 to discuss site security and hint that it is a ranking factor. This web space protection relates to the use of HTTPS coding as it is clear that web space with such encoding guarantees a secure connection between a website and its users.

Crawlability: It is very difficult for Search engines to rank a site if they cannot access it. This gives credence to the fact why site crawlability is an essential Search Engine Optimization ranking factor. It has been established that Crawlability permits search engines to scan a website and review its content; this is to able ascertain what the page is about and how it should rank (Narayana et al., 2012).

User-friendly Mobile: This has been highlighted as one of the important factors that will sustain traffic for you your website and influence users decision to visit such again. Mobile-accommodating destinations guarantee an improved client experience by receiving a responsive plan that adjusts the content, so it shows up generally excellent on each screen size (Vadlamudi, 2015). This is because a larger number of searches are led on cell phones than work areas.

High-Speed Rate: Another very core factor for Search Engine Optimization ranking is the high-speed rate of the page. When it turns out that slow loading sites are associated with your URL, it gives an awful client experience. Search engines realize that individuals need to discover replies as fast as could be expected, so they like to show destinations that will stack rapidly for clients. No client would want to invest the entire of his energy on a prolonged page.

User Engagement: It isn't unusual that Search motors regularly look to clients to assist them with figuring out which pages they ought to advance in search rankings. This is because they consider how clients interface with results to decide which pages are better and more

valuable to searchers. Google does this through an AI apparatus called Rank Brain (Paruchuri, 2015).

High-Quality Content: Most people neglect the fact that users or individuals would always opt for high-quality content of whatever information in the content management system. Using high-quality content has a way of boosting client commitment on your site and appeal to web crawlers is by reliably distributing high-quality content on your webpage. At the point when this is noticed consistently, it helps in SEO ranking. Search engines are determined to proffer the best solutions if possible, so they usually offer favorable rankings to web spaces with thoroughly grounded content.

The Right Target Keywords: It is very wrong to embark very opaquely on the creation of content. You ought to capably make content by streamlining research to show you what keywords to target and subjects to cover. It should be borne as a main priority that center research words empower distinguishing proof of famous keywords that can order lucky traffic to your site. This implies that you need to understand keyword search purpose. This Search aim suggests, by and large, the motivation behind why somebody searches. It is likewise recommended that you essentially target terms inside your site's cutthroat reach instead of getting carried away.

CONCLUSION

This article has effectively appraised how the Uniform Resource Locator functions and maximizes great benefits for web owners. It also revealed how important this very internet tool helps navigation and surfing through the Internet, allowing rank high search engine optimization.

REFERENCES

- Donepudi, P. (2017). AI and Machine Learning in Banking: A Systematic Literature Review. *Asian Journal of Applied Science and Engineering*, 6(3), 157-162. <https://doi.org/10.5281/zenodo.4109672>
- Ganapathy, A. (2015). AI Fitness Checks, Maintenance and Monitoring on Systems Managing Content & Data: A Study on CMS World. *Malaysian Journal of Medical and Biological Research*, 2(2), 113-118. <https://doi.org/10.18034/mjmb.v2i2.553>
- Ganapathy, A. (2016). Speech Emotion Recognition Using Deep Learning Techniques. *ABC Journal of Advanced Research*, 5(2), 113-122. <https://doi.org/10.18034/abcjar.v5i2.550>
- Narayana, S. L., Suneetha Devi J., Bhargav Reddy I., Harish Paruchuri. (2012). Optimizing Voice Recognition using Various Techniques. *CiiT International Journal of Digital Signal Processing*, 4(4), 135-141
- Neogy, T. K., & Paruchuri, H. (2014). Machine Learning as a New Search Engine Interface: An Overview. *Engineering International*, 2(2), 103-112. <https://doi.org/10.18034/ei.v2i2.539>
- Paruchuri, H. (2015). Application of Artificial Neural Network to ANPR: An Overview. *ABC Journal of Advanced Research*, 4(2), 143-152. <https://doi.org/10.18034/abcjar.v4i2.549>
- Vadlamudi, S. (2015). Enabling Trustworthiness in Artificial Intelligence - A Detailed Discussion. *Engineering International*, 3(2), 105-114. <https://doi.org/10.18034/ei.v3i2.519>
- Vadlamudi, S. (2016). What Impact does Internet of Things have on Project Management in Project based Firms?. *Asian Business Review*, 6(3), 179-186. <https://doi.org/10.18034/abr.v6i3.520>

--0--