EASY URLS IN THE CONTENT MANAGEMENT SYSTEM WITH CRAWLERS FOR ADDED SECURITY

Apoorva Ganapathy, Adobe Systems, USA
Alim Al Ayub Ahmed, Jiujiang University, China
Md. Nur-E-Alam Siddique, i-Proclaim, Malaysia

ABSTRACT

Websites are becoming busier by the hour, and URL is what differentiates a single one from the rest. Additionally, a record of bootlickers would require metadata of the page to give classification. This study is aimed at describing the role of friendly URLs in the content management system (CMS) and the power of global ranking with added security. Some selected scientific databases to discover a piece of literature about URLs (search engine optimization) in the content management system and global ranking with crawler with added safety was searched in order to achieve the objective of this study,. The finding shows that keywords and the length of URL are the major factor in CMS and the global ranking of the search engines. It was further revealed that a friendly URL is significant by weighing in to give developed probabilities for sycophants to creep over site pages. It also circumvents the probabilities of uncovering the authentic and real URLs where the content resides, thereby decreasing the chances of hacking. Hence, a concise and understandable URL should be chosen during website development.

Keywords: Friendly URLs, Content Management System, Global Ranking, URL Keywords and Length.

INTRODUCTION

Of recent, the internet is exceedingly active than before with several websites across different network and currently, the number of websites is far above 1 billion (Netcraft, 2017), and its users are near 4 billion as well as its saturation rate of about 59.0%. Observing its trending progress currently, a conclusion could be drawn that this advent is at its final stage that is a parallel rather than exponential growth. Moreover, several statistics indicate that small and medium organizations are operating without a business website (Jose-Manuel et al., 2018).

A content management system offers a graphical interface between users with an instrument to build, collaborate on, revise, broadcast, and pile digital content short of the requirement of writing code from the scratch. A content management system is made up of 2 mechanisms; a content management application and a content delivery application. It is the content management application that builds the graphical user interface to generate, revise, design, and confiscate content from the web short of HTML being aware (Ganapathy, 2015). Whereas a content delivery application mechanism offers the back-end facilities that support organization and conveyance of the content immediately a user generates it using the first mechanism (Jose-Manuel et al., 2018; Hannonhil, 2010).

A URL is the letters you type in the address bar, for instance, a URL for a marketing firm is clinicmatters.io/topsales. A forward slash signifies the segmentation, in the case of a URL every

time there is a slash, it shows a different subfolder. URL contains a domain name which is always the first part of the URL, URL is significant to users for two motives, a short and clear URL structure that brands sense to humans will be dependable and easier to use. Improving your URL system, you provide a pre-information of what is coming up next and offer trust that the outcome see and it is precisely a reflection of where they are going to end up. The second reason is a clear URL provides better navigation and orientation to your user straight away. Friendly URLs help in the content management system and control of worldwide standing with sycophants with added safety (Satinder, 2012).

An Internet crawler is an application that starts with one or more URLs that create a germ set. It identifies URLs from a germ set, as well as procures the Website from such URL. The gotten webpage is then analyzed, to isolate both the link and the text from the page (respective of which points to additional URL). The extracted or isolated text is supply to a text indexer. The haul-out links or URLs are then added to a URL borderline, which at all periods contains URLs whose matching pages have yet to be gotten by the flatterer. At first, the URL borderline consists of the germ set; as pages are hauled out, the matching URLs are removed from the URL borderline. For incessant flatter, the URL of the haul-out page is added back to the borderline again in the forthcoming. This guileless watching recursive traversal of the internet is virtually much more intricate owing to many supplies of a crawling structure – the flatter has to be shared, efficient, scalable, polite, extensible, and robust while eye-catching pages of great value. The flatterer is the most significant portion of a 'search engine' since the value outcomes rest on the flatters (Ganapathy, 2016). The flatters have to slog unceasingly to retain the page source up-to-date. Internet flatter application or script does not travel about to diverse computers on the internet, as intelligent agents or viruses do. A crawler is inherent in a particular machine. The crawler merely directs HTTP application for documentation in order to keep other computers on the internet, the same way a web browser does when the user clicks on the link (Amin & Vadlamudi, 2021, Satinder, 2012; Christos et al., 2019)

Problem Statement

Record of bootlickers would require metadata of the page to give classification. It was to this end that friendly URLs were introduced and they are significant by weighing in to give a developed probability for sycophants to creep over site pages. Friendly URLs also circumvents the probabilities of uncovering the authentic and real URLs where the content resides, thereby decreasing the chances of hacking.

The Objective of the Study

This study is focused on friendly URLs in the content management system and the power of global ranking with crawlers with added security. This article is segmented into five-section, the first section is an introduction, addressing the internet and web user's growth, content management systems, web crawler and URLs, problem statement, and objective of the study. Section two deals the with literature review on the friendly URLs in the content management system and the power of global ranking with crawlers all together with safety.

The method deployed to address the purpose of this paper is section three, while section four handles the finding of this study and the last subsection is the conclusion and recommendation.

LITERATURE REVIEW

Web Content Management System

A WCMS – web content management system is a script or program design to aids maintenance, control, change and reassemble the content on the web pages. Internet content is often stored in a databank and assembled utilizing a flexible language like .Net or XML. The user intermingles with the structure at the facade through a standard web browser. From there the web pages can be edited while maintaining control on parts of the layout (Paruchuri, 2021). A basic CMS has consisted of a programming language, templates, a database, and a dashboard. A subdivision of CM – content management is WCM – web content management. A WCMS is an application that aids in controlling, maintaining, be similar to the content, and changing on a web page. CMS is working on efficient reliance among content (assemblage), supervision, and publication of concluding revised contents (Bergstedt et al., 2003). The framework of WCM is as shown in Figure 1.

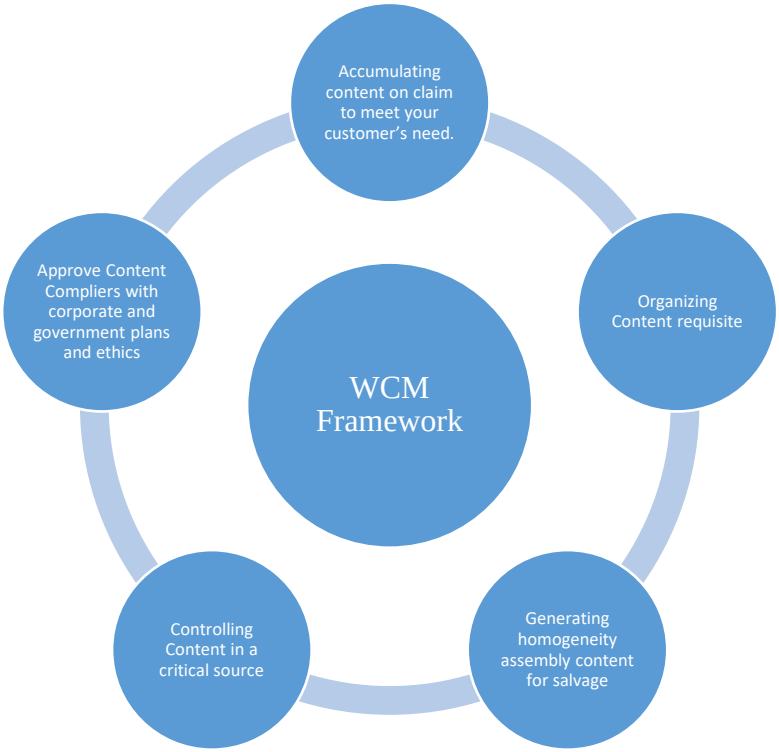

Figure 1
THE FRAMEWORK OF WEB CONTENT MANAGEMENT

Principle of WCMS

A WCMS routines a programming language to collect the content, and deposit it in the database, and then largess it to the users. There are numerous programming languages and scripts that can be utilized in the context of a website. A WCMS is typically a route by means of a

database. Also, out of the several databases available, some of them are only compatible with the certain operating system (OS). Figure 2 shows the working of a web user, database, and web server rapport with WCMS (McDaniel et al., 2017).

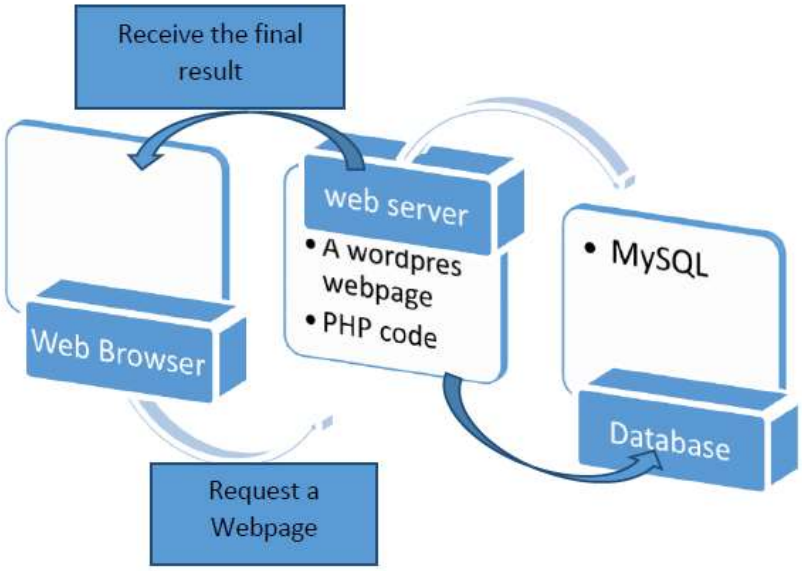

Figure 2
RELATIONSHIP OF WEB BROWSER, WEB SERVER, AND DATABASE

The web content management system has some salient features and there are presented in Table 1.

Tabl4e 1
FEATURES OF WEB CONTENT MANAGEMENT SYSTEM (VADLAMUDI, 2021B)

WCMS Features	Description
Editable content	This feature enables the separation of content from visual appearance site, it becomes much quicker and easier to manipulate and edit. Web content management system software comes with WYSIWYG editing features permitting non-technical consumers to produce and edit content,
Access Control	The maintenance group users feature enable all registered user access to the webpage.
Automated Template	This property aids in creating a template that can be routinely useful to new and present content, allowing the occurrence of all content to change from one crucial place.
Scalable property	Web content management system software contains an element or plug-ins that can be fixed to encompass present coordination functionality.
Web upgrade	This feature enables systematic updates that comprise additional features and keep the system up-to-date.
Workflow management	This feature creates a series and parallel tasks that be practiced in a content management system.

General Approaches Deployed in Wen Content Management System

This subsection discusses general approaches used in web content management systems. These approaches include the following:

Drupal

It is an open-source and free CMS that permits the running, forming, and spreading of the content. This method of WCMS is designed on PHP-based settings and is supported in general public license, which means everyone has a lack of restrictions of copying and sharing it with others (Figure 3 for the Drupal architecture) (Wan et al., 2016).

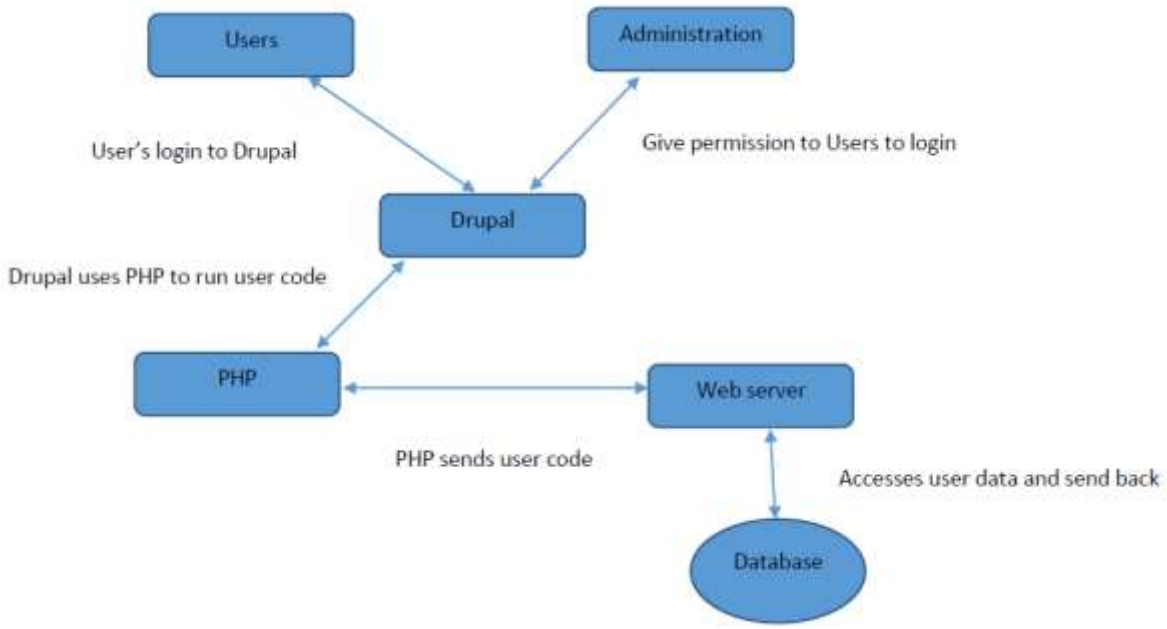

Figure 3
DRUPAL STRUCTURAL DESIGN

Drupal enables stress-free to produce, manage the site and decode whatever in the system with incorporated user interfaces. It connects a website to others and facilities by means of feeds, searches engine linking abilities, etc. As open-source software, no licensing fee is required. It enterprises creative, flexible website to the users and shows more efficiently to raise the number of people visiting the site. It has built-in customizable themes and base themes that are applied to enterprise themes for mounting web applications. Drupal manages content on member, informational and social media sites as well as web applications (Wan et al., 2016).

Drupal advantages range from its flexible content management system which enables handling content types including video, text, audio, real-time statistics, and menu management. It makes available a number of templates for emergent web applications. It saves the stress of starting from the scratch to build complicated or simple web applications and also, it organizes, structure, find and reuse content with over six thousand plug-ins (Bianco & Michelino, 2010).

The setback with a site built with Drupal is the generation of big server loads and on some occasions opens with a dawdling internet connection. Technical-know-how is highly needed and few basic things about the podium to install and modify (Paruchuri, 2020). Drupal is a new CMS,

and it is incompatible with other software. Its performance is relatively low compare to other content management systems. On a whole, Drupal does not have a good user-friendly interface (Maurushat, 2019).

Joomla

This is another open-source CMS that is deployed in building websites and web applications. It is extendable and free, and capable to separate into back-end and front-end templates. Joomla is designed using PHP, MySQL, Software design model, and object-oriented programming (OOP). It comes with a tutorial file that helps the users to learn the basics and how to create sites at ease. The tutorial is sectioned into Joomla Menus, Joomla Basics, and Joomla Modules as well Joomla Advanced (Barker, 2016).

Joomla provides handling the user info like consent to edit, publish, access, create or remove the user, change languages and password. The principal aspect of user handling is validation. In terms of content management, it permits content management via the WYSIWYG editor to edit and create the content in a very modest manner. Joomla adds or edits banners on the websites and manages the designs used on the sites. These designs can be implemented without any alteration of the web content structure. It has managing tools for media files and folders in which you can easily upload, organize and manage your files into your article editor tool. It consists of a link resource that provides useful and sorts them into classes. Joomla permits users to search for the right information on the website. It uses advanced search options, smart indexing, and automatically recommends searches to make Joomla search best. It also creates menus and items on the menus and can be supervised later. The menu can be manipulated in any style of your choice, even in multiple places.

Its advantages range from being freely accessed and it is flexible and compatible with many browsers (Vadlamudi, 2021a). It is user-friendly, that is, it is easy to install and set up. Also, it guarantees the security of the data, and media files, content can upload easily in the piece of writing editor tool. The most noticed setback of this approach is that it is incompatible especially when several modules are set up, plug-ins and extension concurrently. It takes time to load and running and it is not to search engine optimization friendly. Joomla does not support free modules and plug-ins and enhancement are not easily achieved.

Yalcin et al. (2010) documented many parameters that influence search engine optimization and demonstrated the mechanisms behind the procedure of the search engine. Yalcin and co-authors concluded that search engine optimization is a vibrant progression that must be supervised recurrently, tracing negative or positive changes for the enhancement of a site. A fastidious search engine optimization method should be executed at the construction of a website that aspires to enable website visibility on search engines (Wilson et al., 2006). Wilson and co-authors opinion are that keywords should be coined in such a way that it relates to the content of the website so it can be indexed superior in the search outcome for the desired keywords. They buttressed that search engine optimization is a rivalry sector and modifications are made frequently. Cui et al. (2011); Zilincan (2015) analyzed similar topics. The prevailing search engine optimization methods were evaluated by Ahmed et al. (2021) investigating and likening other research whereas stating major facts that could be improved.

Based on the data, Ahmed and co-author suggest some theoretical techniques for search engine optimization. Zhang et al. (2011) examine the subject matter by comparing the search engine optimization parameters utilizing one hundred and sixteen (116) websites. Variables

affecting the ranking of the SERPs and how marketers and web developers can take benefit from them were studied by Killoran (2013). Killoran states that SERPs are made by 3 classes of accomplices; the search engine organizations and programmers, the search engine optimization specialists, and the search engine users. It is asserted that both mark group and rivalry websites have to well-thought-out while selecting keywords. In conclusion, he suggests that the right keyword assignment and connection construction by communication with other content originators are vital for best outcomes. Thekral et al. (2016) followed the same pattern and aimed at the theoretical article of the above features. Kakkar et al. (2015) earlier likened and elucidated numerous algorithms of Google and aims at search engine optimization and by what means they tip to improve ranking, concluding that search engine optimization is an extended period and undercurrents process. Gudivada et al. (2015) examined the mechanisms after search engines and the methods that are applied to rank websites all together within what way it influences the traffic of a website. Gudivada and co-authored conducted research that indicates that seventy percent of the website users with the organic outcomes. Streamlining it, most of the users that are up to sixty percent choose one of the first three organic outcomes. That demonstrates the influence of search engine optimization on website campaigns. Krrabaj et al. (2017) investigated the on-site and off-site search engine optimization features.

METHODS

In order to achieve the focus of this study, which is examining how friendly URLs contribute to the content management system and power of global ranking with crawlers with added security, we study some scientific databases to discover the piece of literature about URLs (search engine optimization) in the content management system and global ranking with crawler with added safety.

A descriptive analysis was carried out based on the pieces of literature selected to determine the relation of URLs to the content management system and global ranking with the crawler.

RESULTS AND DISCUSSION

In the literature review, it was found that some parameters as shown in Table 2 relating to URLs in contributing to the content management system and in the global ranking of the website in search engine indexes.

Table 2
HOW FRIENDLY URLS CONTRIBUTE TO CMS AND RANKING OF WEBSITES

Factors	Cited literature
Keyword in URL	(Krrabaj et al., 2017; Chen et al., 2011; Ganapathy, 2017)
URL Length	(Yalcin and Köse, 2010; Ahmed et al., 2021; Krrabaj et al., 2017; Chen et al., 2011; Patil et al., 2013; Thakur et al., 2011; Ganapathy, 2017)

The prominent variables in the selected works of literature include URLs keywords and URLs length. Apart from the loading time and algorithms of the search engine, the two factors are optimal in the content management system and the global ranking with a crawler.

Despite distinct effects and graphic basics improvement of a website, the undue routine of them may upsurge the loading time. The loading rapidity of a website is posh is not only by graphic

basics but the HTML file and all the components irrespective of their category are responsible for the size of the website and influence the loading speed.

URL keywords are the easier for crawlers to locate a website if its URL contains the desired keyword in relation to their work. Typically, search engines have a tendency to rank higher websites with .gov and .edu domains as these domains are utilized by educational and government websites. However, all types of websites are optimized and can be achieved when the URL length is kept short. The keyword in meta carries a description label; the description label of meta is a concise and summary of the content of a web page. This description label consists of the text that looks like in the search outcomes of the search engines just beneath the link. The description label of the meta is a control on what keywords should the website be indexed for by the search engines. The length of the URL represents the address of the website on the internet or www (World Wide Web). As earlier mentioned under the keyword of URL, the desired keywords ought to be involved in the URL, so it can be map out without difficulty from the search engines. The length of the URL makes it possible for search engine optimization to be user-friendly, thus, they must be short, concise, and comprehensive. Websites with lengthier text have the tendency to rank among the top sites in the search outcomes when likened to sites that use few or fewer text lengths. This signifies that search engines have a preference for content-rich websites. This postulate might relate to the point that in richer content the directed keywords come up more rapidly. Even web users give more preference to longer text websites owing to the fact it is more informative. Also, when the length of the URL is mention, it works in hand with the text-to-code proportion, which is the metric that characterizes the proportion between the front-end portion of the website to the back-end code. The ideal proportion varies from twenty-five percent (25%) to seventy-five percent (75%) this proportion denotes in the graphical text is likened with the HTML component consisting of the image labels and other visual components. Even though, this variable is not in a straight line related to the global ranking of a website. Several variables are centered on this parameter, so it is important to be well-thought-out for a more operative search engine optimization approach (Paruchuri et al., 2021). However, the length of the URL and the use of directed keywords in the title label do not influence the ranking of a website, even though the modification in the outcomes is owing to the fact that the methods of our study are likened to the others. More particularly is the Dean's investigation that reported the sample consisted of fewer search outcomes for more keywords (Thakur et al., 2011).

In the context of URL role in the content management system, it was observed by Jose-Manuel et al. (2018) that web content management system is very flexible if the URL keyword and the length of URL are concise and understandable. This will lead to better and easy use and manage web content (Vadlamudi et al., 2021). The ins and outs for this accomplishment are that they are projected for a very widespread user, short of a need for great computer knowledge and skills, and that they provide a large group of functionalities.

CONCLUSION AND RECOMMENDATIONS

Websites rivalry for the search engine result pages is colossal so an expedient optimization strategy is required, a strategy that consists of an all-inclusive method concerning the search engine optimization variables but relies on the most efficient ones.

This article has been able to address some variables that contribute to the huge success in terms of the content management system and global ranking with crawlers and they have stayed

untouched over a while. Hence, algorithms of search engines have the tendency to modify frequently, and new variables are added while nonoperational ones are removed. This is website developers must pattern the modification of the algorithm and regulate their website in order not to only accomplish but also uphold high content management and global ranking. This study highlighted major variables that contributed or leads to effective search engine optimization. This factor is a friendly URL which has to do the keywords and the length of URLs as the major contributing variable in the content management system and Power of Global Ranking with Crawlers with Added Security.

REFERENCES

Ahmed, A.A.A.; Paruchuri, H.; Vadlamudi, S.; & Ganapathy, A. (2021). Cryptography in Financial Markets: Potential Channels for Future Financial Stability. *Academy of Accounting and Financial Studies Journal, 25*(4), 1–9. https://doi.org/10.5281/zenodo.4774829

Amin, R., & Vadlamudi, S. (2021). Opportunities and Challenges of Data Migration in Cloud. *Engineering International*, *9*(1), 41-50. https://doi.org/10.18034/ei.v9i1.529

Barker, D. (2016). Web Content Management: Systems, Features and Best Practices, 1st ed.; O'Reilly Media: Sebastopol, CA, USA, ISBN 9781491908129.

Bergstedt, S.; Wiegreffe, S.;Wittmann, J. & Moller, D. (2003). Content Management Systems and E-Learning Systems –a Symbiosis? In Proceedings of the 3rd IEEE International Conference on Advanced Technologies, Athens, Greece, 9–11 July 2003; pp. 155–159.

Bianco, F. & Michelino, F. (2010). The Role of Content Management Systems in Publishing Firms. *Int. J. Inf. Manag*., 30, 117–124.

Chen, C.Y. Shih, B.Y. Chen, Z. & Chen, T.H. (2011). The exploration of internet marketing strategy by search engine optimization: A critical review and comparison. *Afr. J. Bus. Manag*., 5, 4644–4649.

Christos, Z., Maro, V., Theodosios, K. & Makrina, K. (2019). Important Factors for Improving Google Search Rank. *Future Internet*, *11*(2): 32, 1-12. https://doi.org/10.3390/fi11020032

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/mjmbr.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

Ganapathy, A. (2017). Friendly URLs in the CMS and Power of Global Ranking with Crawlers with Added Security. *Engineering International*, *5*(2), 87-96. https://doi.org/10.18034/ei.v5i2.541

Jose-Manuel, M.C., Antonio-Jose, A.H., Antonio, G.P., Ramon, S.I. & Maria-Dolores, C. (2018). A Comparative Study of Web Content Management Systems. *Information*, *9*(27): 1-15. doi:10.3390/info9020027.

Killoran, J.B. (2013). How to use search engine optimization techniques to increase website visibility. *IEEE Trans. Prof. Commun*., 56, 50–66.

Krrabaj, S.; Baxhaku, F. & Sadrijaj, D. (2017). Investigating search engine optimization techniques for effective ranking: A case study of an educational site. In Proceedings of the 2017 6th Mediterranean Conference on Embedded Computing (MECO), Bar, Montenegro, 1–4.

Maurushat, A. (2019). Ethical hacking. http://hdl.handle.net/10419/203843.

McDaniel, R. Fanfarelli, J.R. & Lindgren, R. (2017). Creative Content Management: Importance, Novelty, and Affect as Design Heuristics for Learning Management Systems. *IEEE Trans. Prof. Commun.*, 60, 183–200.

Netcraft. Available online: https://www.netcraft.com/ (accessed on 14 May 2021).

Paruchuri, H. (2020). The Impact of Machine Learning on the Future of Insurance Industry. *American Journal of Trade and Policy*, *7*(3), 85-90.

Paruchuri, H. (2021). Conceptualization of Machine Learning in Economic Forecasting. *Asian Business Review*, *11*(1), 51-58.

Paruchuri, H.; Vadlamudi, S.; Ahmed, A.A.A.; Eid, W.; Donepudi, P.K. (2021). Product Reviews Sentiment Analysis using Machine Learning: A Systematic Literature Review. *Turkish Journal of Physiotherapy and Rehabilitation, 23*(2), 2362-2368, https://turkjphysiotherrehabil.org/pub/pdf/322/32-2-316.pdf

Satinder, B.G. (2012). The Issues and Challenges with the Web Crawlers. International Journal of Information Technology & Systems, Vol. 1; No. 1: ISSN: 2277-9825.

Thakur, A.; Sangal, A.L. & Bindra, H. (2011). Quantitative measurement and comparison of effects of various search engine optimization parameters on Alexa Traffic Rank. *Int. J. Comput. Appl*., 26, 15–23.

Vadlamudi, S. (2021). The Economics of Internet of Things: An Information Market System. *Asian Business Review*, *11*(1), 35-40.

Vadlamudi, S. (2021b). The Internet of Things (IoT) and Social Interaction: Influence of Source Attribution and Human Specialization. *Engineering International*, *9*(1), 17-28.

Vadlamudi, S.; Paruchuri, H.; Ahmed, A.A.A.; Hossain, M.S.; & Donepudi, P.K. (2021). Rethinking Food Sufficiency with Smart Agriculture using Internet of Things. Turkish Journal of Computer and Mathematics Education, 12(9), 2541–2551. https://turcomat.org/index.php/turkbilmat/article/view/3738

Wan, S.; Li, D. & Gao, J. (2016). Exploring the Advantages of Content Management Systems for Managing Engineering Knowledge in Product-Service Systems. Procedia CIRP 2016, 56, 446–450.

Yalcin, N. & Köse, U. (2010). What is search engine optimization: SEO? *Procedia Soc. Behav. Sci*., 9, 487–493.

Zilincan, J. (2015). Search engine optimization. In Proceedings of the CBU International Conference Proceedings, Prague, Czech Republic, 25–27 March 2015; Volume 3, p. 506.