IJICTR
International Journal of Information & Communication Technology Research

Volume 16- Number 4 – 2024 (44 -56)

Performance Evaluation of Internet Domain Ranking Methodologies

Pejman Goudarzi*
ICT Research Institute
Tehran, Iran
pgoudarzi@itrc.ac.ir

Davood Maleki
ICT Research Institute
Tehran,Iran
dmaleki@itrc.ac.ir

Alireza Mansouri
ICT Research Institute
Tehran, Iran
amansuri@itrc.ac.ir

Received: 1 July 2024 – Revised: 4 October 2024 - Accepted: 3 November 2024

Abstract—**Internet domain ranking is one of the important tools for demonstrating a domain popularity level. To evaluate and rank Internet/web domains according to their referrals, popularity, and traffic, domain rankings are extracted in various methods. This study investigates and analyzes these methods and their relevant platforms. For this purpose, the main ranking methods and their inherent characteristics, including similarity, stability, responsiveness, and the degree of benignity (e.g., the low possibility of changing the lists), are examined, and their potential effects on the conclusions are determined. Furthermore, this study specifies the domains ranking indicators used by the main ranking methods. Finally, as the main conclusion of this study, using the Cloudflare radar and combined Tranco ranking are recommended to rank the Internet domains. Moreover, in domains ranking, for each domain of the extracted list, it is necessary to check their IP and Name Server (NS) and delete those that do not have an IP address or their NS are expired. We have used the multi-criteria decision making (MCDM) methodology to obtain an overall ranking score between different competing ranking scenarios/criteria. Based on the results of this paper, we can conclude that the Cloudflare radar and Tranco with overall ranking efficiency score of 81.8% and 79.9% are the most efficient ranking methodologies based on mixture of different ranking metrics, respectively**.

Keywords: Domain ranking, Likert scale, MCDM, Top-Level Domains, Ranking platform

Article type: Research Article

[DOI: 10.61186/itrc.16.4.44]

I. INTRODUCTION

According to Siteefy statistics, currently, there are around 1.11 billion websites (Each website is associated with an Internet domain name) in the world, and about 10,500 new websites are created every hour [1]. Due to this huge number of web pages and the rapidly growing rate, automatic search engines like Google and Yandex are needed to provide reliable information for users.

The size and extent of web-related contents over the Internet has exponentially grown due to increasing level of interactions between web users over the Internet. Due to numerous challenges in manual analysis, an automatic system is needed to give reliable information from such a large collection of contents. In this context, search engines like Google, Yandex, etc. are considered as tactical information retrieval tool over the Internet.

An important methodology which is adopted by search engines to demonstrate relevant and efficient query results is the domain ranking. Ranking methods for the most popular active websites are being used by

* Corresponding author

research communities, security analysts and industry activists to demonstrate the importance, timeliness and relevance of the Internet domain names.

The objective of a domain ranking is to arrange web sites in an ordered list based on visual metrics such as page views or unique visitors over some time period [2]. Such rankings are typically produced by third parties such as Alexa, Cloudflare radar, Majestic, Umbrella or Quantcast, and the rankings are inferred from measurements of user panels, i.e., a group of people who are compensated to allow their web browsing behavior to be tracked.

In this article, we will discuss and analyze the ranking methodologies of the Internet domains (the ranking of the most visited domains) for all service provider domains. Furthermore, we evaluate the ranking of domains based on parameters such as domain name query delay, IP Owner, NS[1], location, etc.

The main contributions of the paper are:

- We have introduced different and important domain ranking methods and their associated evaluation metrics in an integrated manner.
- We have compared the performance of different domain ranking methods based on specific metrics in different scenarios.
- Using a single mixed score which has derived from analytic hierarchy process (AHP) methodology, we rate and specify the most effective domain ranking methods based on a weighted mixture of 17 different metrics which are investigated in four different and independent scenarios.

The rest of the paper is organized as follows:

In Section II, some definitions and background are presented. In Section III, related work have been discussed. In Section IV, ranking of domain popularity of websites in different evaluation platforms is introduced. In Section V, we have introduced some important domain ranking evaluation metrics. In Section VI, domain ranking performance analysis and comparison has been proposed. Finally, in Section VII, some concluding remarks and open research opportunities have been proposed.

II. Background

Before going into the main background, we introduce some definitions regarding this area of research.

2.1 Some Definitions:

Domain Owner: The entity which the domain has been registered to

NS: Domain name server

Domain Ranking: Domain ranking/rating is one of the important indicators for checking and analyzing the sites, its purpose is to check the size and quality of the links that are provided for the sites [1-2].

URL: Uniform Resource Locator

To explain how to extract the domain name, URL components are introduced in Fig.1. In this figure, the part used as the domain name has been displayed.

As can be shown in Fig. 1, different parts of a URL are: communication protocol (same as http or https), sub-domains, second level domain, top-level domain (TLD), Internal branches and routes.

The ranking of domains includes all important domains of service providers. For example, to identify Iranian domains, at least the following sources should be used:

- Collection of domains registered with nic.ir
- Domains registered in eNamad.ir[2]
- Domains featured on https://trends.netcraft.com/topsites?c=IR
- Domains provided by the Communication Infrastructure Monitoring Center

DNS: Domain Name System

Is a core protocol/technology which is used in the Internet for providing flexible decoupling of a service’s domain name and the hosting IP addresses. It has been widely used in network communications, e-business, and multimedia services such as content delivery networks (CDN). Almost all Internet applications need to use DNS to resolve domain names and achieve accurate resource location.

Domain names are an essential part in Internet engineering. Regarding this fact, there are various software vendors that their product virtually divides the network into several pieces based on the domain name for different purposes.

As an example, Hillstone's security manager platform is one of the software tools which is related to managing virtual domains. This software enhances network security by allowing businesses to divide their networks into multiple virtual domains [3-4].

Usually, domains can be based on geography, business unit, or security functions. It provides the flexibility needed to manage Hillstone's infrastructure while simplifying configuration, speeding up deployment cycles and reducing management costs. Hillstone's Security Manager allows the administrator to divide security management into multiple virtual domains [3].

Domain SEO[3] is the practice of optimizing owned domains to make them more accessible to human visitors and search engines. It involves picking a simple phrase, an optional subdomain, and a top level domain (TLD) to create the perfect web identity.

Some elements of the domain name may play into the overall SEO success. These include elements such as memorability, length, keyword usage, brandability, etc.

[1] Name server
[2] A digital trust indicator in the country
[3] Search Engine Optimization

[DOI: 10.61186/itrc.16.4.44]

One of the factors to measure the success of SEO in the long term is to check the validity of the domain or the strength of the site's domain, which itself is obtained from the performance in other SEO factors [5].

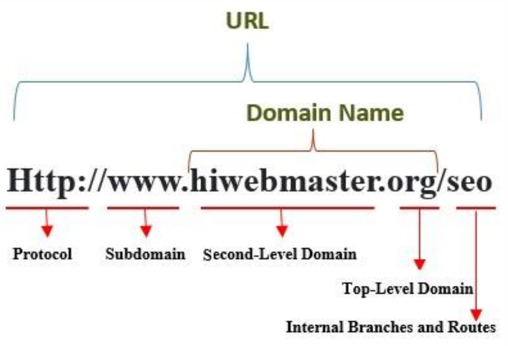

Figure 1. Different parts of a URL

In this paper, the methods /platforms for evaluating and ranking of Internet domains in the world have been analyzed. To do so, different methods have been examined and evaluated. For the main ranking methods, their inherent characteristics, including similarity, stability, responsiveness, and the degree of benignity (the low possibility of manipulating the lists), have been examined and their potential effects on the conclusions have been determined.

Furthermore, the indicators used in the ranking of the domains that have been used by the main and major ranking methods have been specified. Finally, this paper concludes that it is more convenient to use the novel and combined Tranco ranking to weight the Internet domains [6]. Moreover, in ranking the domains, for each domain from the extracted list, it is necessary to check the IPs and NS of the domains and to delete the domains that do not have an IP address or their NS has been expired.

In summary, the main contributions of the current paper are summarized as follows:

Different Internet ranking methods/platforms have been compared and their strengths/weaknesses have been specified.

It is specified that the Tranco ranking method is the most efficient one in terms of benignity and other ranking efficiency features.

2.2 Types of internet domains in terms of extension

Today, there are more than millions of domains on the Internet that have different extensions such as .com or .ir. For each country there is an internet link according to its name, such as .uk for England. There is another category of domains on the Internet that are specific to enterprises and organizations. The domain name is of great value and importance, but the user should also pay attention to the effectiveness of the extension when registering the domain.

The subject to be investigated here is the domain extension which is called Top-Level Domain (TLD). Top Level Domains (TLDs) can be a generic top level domain (gTLD) or a country extension (ccTLD) or a combination of both [7].

2.3 The process of registering domains in ICANN

All domains are registered under the supervision of ICANN [4] (the world's only domain registration authority) and through registrars approved by this organization.

All the registrars have valid and active representatives who communicate with them for domain registration, renewal, DNS change and whois change of users' domains. These agents, called registrars, are often domain and hosting service companies that offer different domain rates and services. ICANN is a domain registration authority that allows users to register a domain with a desired extension. Due to the possibility of searching for domains, online and real-time registration, a large number of domains with different extensions are registered daily. There are four main categories for TLDs introduced by ICANN, and each category is defined based on the content of websites [7]

The different types of top-level domains (TLD) are:

- Generic Top Level Domains (gTLD[5])
- Supported Top Level Domains (sTLD[6])
- Country Code Top Level Domains (ccTLD[7])
- Advanced infrastructure (ARPA[8])

➢ gTLD

gTLD is an abbreviation of the term and is the most common type of domain that allows different types of users to register. The most famous gTLD examples are: .com, .org, .net, .name, .info and .biz.

➢ sTLD

sTLD is a type of public top-level domain provided by public organizations. Users who want to register their site under these types of domains must follow certain rules. Some examples of sTLDs are: .gov, .edu, .int, .mil, .tel, .post and .asia.

➢ ccTLD

Country code TLD refers to the ISO code of a specific geographical location or territory. ISO code is a two-letter code that represents the name of certain regions. This type of domain extension is useful for sites that want to offer their services or products in a specific country. Common examples of ccTLDs are Spain's .es, Russia's .ru, United States' .us, and Iran's .ir.

➢ High level infrastructure domain

The only existing ARPA infrastructure top-level domain extension is the infrastructure top-level domain. These suffixes are reserved by the IANA for the IETF

[4] The Internet Corporation for Assigned Names and Numbers
[5] Generic Top-level Domains
[6] sponsored top-level domains
[7] Country Code Top-level Domains
[8] Address and Routing Parameter Area

[DOI: 10.61186/itrc.16.4.44]

or Internet Engineering Task Force. Therefore, they are only used for technical infrastructure issues [8].

Also, in addition to the examples of TLD domains above, a complete list of TLDs is available on the IANA's (Administration of Assigned Numbers) website.

III. Related Work

Internet domain and web page ranking have been considered in some previous researches as described below.

The authors in [9] have performed a comprehensive survey regarding web page ranking using web mining techniques. The authors have presented different approaches/techniques, algorithms and evaluation approaches in previous researches and identified some critical issues in page ranking and web mining, which provide future directions for the researchers working in the area.

In [10], the authors have proposed an architecture for web domain ranking which includes processing capability required for handling Big Data available on the web. The proposed architecture presents a new method for web domain ranking that is independent of the link structure of the web graph. The proposed method provides web domain ranking based on the number of unique visitors, the number of user sessions, and session duration.

Most of the Internet domain rating methods are designed based on the context of user queries, for improving the search engine results, [11-13].

Among the proposed methods, link analysis is the most widely used one which is tailored for measuring the web page importance which is derived using the web link graph. Two important link analysis techniques that have been considered as the basis of lots of developed web rating methodologies are HITS [14] and PageRank [15] algorithms.

Kline et al., have proposed a triangulated rank-ordering of web domains. They have considered the surprisingly challenging problem of generating consistent and reliable web site rankings based on unique visitors per day. They illustrate the challenge this represents using data from three large and independently-sourced Internet user panels [16].

Malicious domain name detection is another active research area in domain ranking. Some related work are listed below.

Zhao et al. in [17] proposed a malicious domain names detection algorithm based on lexical analysis and feature quantification. In the mentioned paper, the potential malicious domain name is determined to be malicious or normal based on its reputation value. The effectiveness of the proposed detection method has been demonstrated by experiments on public available data.

The authors in [18] proposed a supervised machine learning approach based on keyword density which is sensitive to detect malicious webpage. They analyze the domain name features such as keyword frequency and length attribute, for detecting malicious websites.

The algorithm proposed in [19] mines patterns in a dynamic manner for identifying malicious URLs generated by some harmful programs without using any pre-specified item or element.

Work [20] uses supervised learning methodology (support vector machine) and some important URL properties such as number of dots, hyphens, numeral characters, URL length, similarity index and an indicator for representing the existence of IP address in URL, for identifying malicious websites/domains.

Some other authors proposed a specific classifier which is called DGA (Domain Generation Algorithm) and use linguistic features to identify the malicious domain names in a real-time manner [21].

The main difference between the current paper and the mentioned ones is the fact that in this study, we have done a comprehensive analysis and evaluation to assess the performance of different Internet domain ranking methodologies in term of pre-specified ranking quality indices (such as benignity, availability, stability, etc.). As far as we know, this is the first time that Internet domain ranking is evaluated from this point of view.

IV. Domain Ranking Platforms

In this section we have investigated six popular domain ranking methods/platforms which are Alexa, Cisco Umbrella, Majestic, Quantcast, Tranco and Cloudflare radar.

In [6] and [22], four different site ranking methods, including Alexa, Cisco Umbrella, Majestic and Quantcast rankings, have been studied because these ranking methods are free and are mostly used by researchers and analysts. All of these rankings use different data sources and scoring methods to calculate their rankings.

In order to assess the prevalence of security and privacy practices in a sample of the web, researchers rely on website popularity ratings such as the Alexa list. While the validity of these ratings is rarely questioned, research findings suggest contrary to this. Website popularity ratings like Alexa's top 1 million sites, which many use in their research, are unreliable [6].

For the four main rankings methods, their intrinsic properties (similarity, stability, responsiveness, and benignity) are examined and their potential effects on the conclusions are determined. The possibility of manipulating the composition of these lists and changing the results have also been investigated.

Researchers, security analysts, and companies often use popular sites to evaluate security and privacy practices on a sample of the web. Additionally, popular sites are often assumed to be benign and therefore whitelisted.

Commercial providers such as Alexa had been published daily rankings of the most popular websites, but so far their credibility has rarely been questioned. To gain more insight into how reliable this data is, we seek to answer the following three questions:

- How do such rankings affect research?
- Can malicious actors abuse rankings?

[DOI: 10.61186/itrc.16.4.44]

- How to improve such rankings?

Intrinsic features of site popularity ranking methods can have an impact on research.

Internet top lists are widely used in research. But their quality and stability are questionable and they are subject to manipulation and can be manipulated easily. In order to protect research from manipulated lists, the top lists provided by Alexa, Cisco Umbrella, Majestic, and QuantCast have been logically and legally combined in Tranco in order to rank the inclusion of a site in the output list [23][9].

4.1 Alexa

Alexa ranks websites based on a combined measure of page views and site users, and creates a list of the most popular websites based on this ranking averaged over quarterly periods. Only the top-level domain of the site is registered and includes any sub-domains [24].

Alexa primarily relies on end users installing a browser extension that sends all visited URLs to Alexa. Alexa, as a subsidiary of Amazon, publishes a daily list of one million websites since December 2008. Only paid-level domains are ranked, except for subdomains of certain sites that offer "home pages or personal blogs" (e.g. tmall.com, wordpress.com).

4.2 Cisco Umbrella

Cisco Umbrella counts the number of IPs that access a domain through OpenDNS.

Cisco Umbrella has been published a daily list of one million entries since December 2016. Each domain name may be ranked based on the amount of traffic which is collected by it and all its subdomains. The rankings calculated by Cisco Umbrella are based on DNS traffic (called OpenDNS) and claim to exceed 100 billion requests per day from 65 million users. Domains are ranked based on the number of unique IP addresses which they export DNS queries to [6].

4.3 Majestic

Majestic counts the number of subnets that host a web page and link it to a domain [6].

4.4 Quantcast

Quantcast mainly ranks sites that report the number of visitors through an analytics script. These different ranking methods indicate that rankings may represent diverse features and may be combined in different manners [6].

4.5 Tranco

In order to resolve the challenges associated with the previous domain ranking methods, in this section, we introduce a new and improved ranking method called Tranco. This method is an approach that researchers and industry activists can use to obtain lists with desirable and more appropriate features. Tranco is a research-based method that is resistant to manipulation by top ranking sites. Tranco allows the research community to work with reliable and repeatable rankings. This is an improved rating provided through an online service available at https://tranco-list.eu.

Tranco combines all data from existing rating Tranco (including Alexa, Cisco Umbrella, Majestic, and Quantcast) with the aim of improving popularity rating features for research, while addressing the shortcomings of existing ratings with averaging and filtering techniques. In this method, you can average the existing ratings in a period of time and select a set of providers. Apply additional filters to it. For example, for use of Tranco method, standard lists can be provided by filtering out unresponsive or malicious sites that can be easily used. It is also possible to create mixed lists with high adjustment capability. In the new method, to improve the stability of the hybrid listings and agree on which domains are really popular, Tranco's default setting uses the rankings of all four other methods provided for a 30-day period.

In the domain ranking based on Tranco, for each domain from the extracted list, it is necessary to check the IPs and NS of the domains and delete the domains that do not have an IP or their NS has expired. After finalizing the list, each domain is evaluated and ranked using Tranco ranking, and a certain number of top sites are selected [6].

4.6 Cloudflare radar

Cloudflare radar is a new cloud-based ranking method which ranks the websites based on the website popularity level in terms of estimated relative size of the user population that accesses a domain over some period of time. It uses privacy-enabled 1.1.1.1 resolver and machine learning (ML) techniques to filter bot-generated traffic and only focuses on real user generated traffics [25]. It uses these DNS data to calculate the top and trending domains found on both the global and country pages on Cloudflare radar.

In summary, Cloudflare creates two types of domain rankings which are listed as follows [25]:

- A global per country-based ordered list of the top 100 most popular domains which updates every 24 hours daily.
- An unordered global dataset which updates every week and includes most popular domains which is partitioned into the following number of domains: 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, 50,000, 100,000, 200,000, 500,000, 1,000,000.

There is no standard definition of popular and less popular sites, and different lists of top sites have different methods of calculating their popularity, which is not entirely transparent. However, the number of searches and queries for a site provides a clear measure of popularity. If a host appears in the results of a search engine for a popular search term or more, it is likely to be seen by the user [26].

Top lists, especially Alexa's top sites list, have been used in many studies. Also, different methods are used by providers to generate lists of popular and top sites. Alexa ranks popular sites based on the number of visits measured by a browser-based panel. While Cisco

[9] https://kb.builtwith.com/general-questions/what-is-tranco-quantcast-majestic-and-umbrella-numbers

[DOI: 10.61186/itrc.16.4.44]

Umbrella uses user behavior, list of top sites is based on the number of DNS requests for hosting.

Majestic provides a list of top sites based on a link graph, which is based on the links discovered in Majestic's own crawl, that is, not based on user behavior, but based on the number of links from pages that point to other pages. They do, it is created.

Common Crawl platform follows a similar approach in its web graphs [27]. Common Crawl data is widely used for research in various fields. In this method, crawling is usually done on a monthly basis, updating known pages but also crawling new pages. In this method, the focus is on the breadth instead of the depth of the hosts, that is, Common Crawl tries to get a wide sample of hosts and more pages from higher ranked domains.

In order to generate a list of popular and top sites by Tranco, it offers a combination of Alexa, Quantcast, Cisco and Majestic data. The two primary data sources, (i.e. Alexa and Quantcast), are no longer available today. The Tranco list produced on July 31, 2021, includes a combination of rankings provided by Alexa, Umbrella, and Majestic [26].

V. Domain Ranking Evaluation Metrics

In this section, several features/metrics of ranking methods that can be effective for evaluating their suitability are analyzed and investigated [24].

5.1 Similarity

In Fig. 2, we can conclude that between January 2018 to November 2019, four different ranking methods (Alexa, Umbrella, Quantcast and Majestic) have only 2.48% agreement and similarity in ranking popular web sites. This means only 70000 from 2.82 million sites. We can conclude that in similarity index point of view, there is very little intersection and agreement between these popular ranking methods.

5.2 Stability and consistency

The mentioned four ranking methods have still big difference in consistency and stability index. As shown in Fig. 3, between the four ranking methods Quantcast and Majestic lists are the most stable and consistent ones. These two lists has changed around 1% in each day, but another ranking method (Umbrella) changes around 10 percent in each day. About 500000 top rankings associated with Alexa has been changed in each day from 30 January 2018 which was due to a sudden variation in Alexa's website ranking procedure (about 50 percent daily change based on Fig. 3).

A highly stable list gives a pool of efficient domains, while may miss those domains which their associated popularity level changes in an instant manner. This results in the fact that some newly added websites or peaks may not be tractable. However, an unstable list that varies greatly due to the more popular websites may produce huge changes in time-based analyses and deductions.

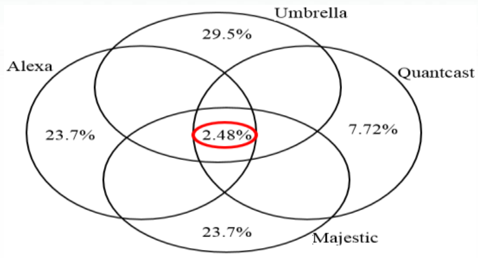

Figure 2. Average daily crossover between the four lists from January 2018 to November 2019 [24]

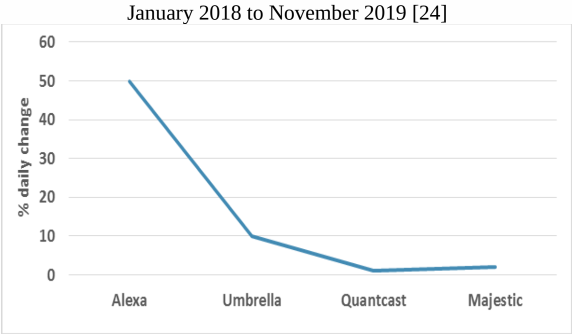

Figure 3. The percentage of intersection of each provider's list in two consecutive days

5.3 Responsiveness and accessibility

Fig. 4 depicts the responsiveness and accessibility for the four ranking methods mentioned above. From accessibility point of view, about 11 percent of Majestic ranking and 5 percent of Quantcast and Alexa rankings are not accessible. In Umbrella ranking, the value is about 28 percent. The reason behind most of the Umbrella's faults is that it has not a filtering mechanism for removing invalid subdomains or domains. Those domains which are not accessible result in the fact that we cannot demonstrate the whole Internet by taking a specific amount of samples and may lead to incorrect conclusions.

5.4 Benignity

It was verified based on Google Safe Browsing in 2018 that around 0.22 percent of Majestic websites were possibly not benign and may be malware domains. Almost all of the ranking methods have been ranked malicious websites. As an example, in top 10000 Alexa's domains, four domains have a form of manipulation. A domain from 10000 top Majestic ones offers malicious software. The existence of these malicious domains in the list of Quantcast and Alexa is approximately high and this results in their lower benignity score in Fig.5. Majestic has the best benignity score.

5.5 Google's ranking metrics

Google uses 200 different ranking parameters for efficient ranking of Internet domains. From these parameters, 8 parameters/factors are of paramount importance. Based on [28], these factors are as follows:

Quality Content: This is the most important SEO factor. Google tends to present to users high-quality, informative, and relevant content.

[DOI: 10.61186/itrc.16.4.44]

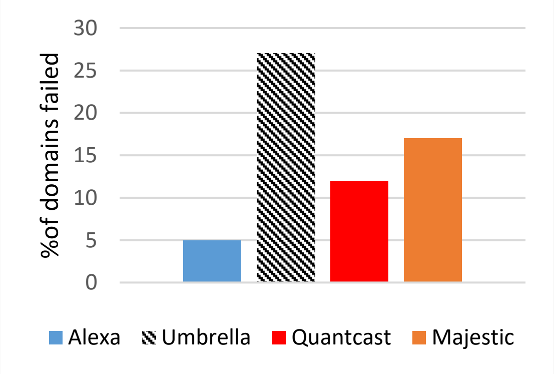

Figure 4. Response and HTTP status code reported in lists

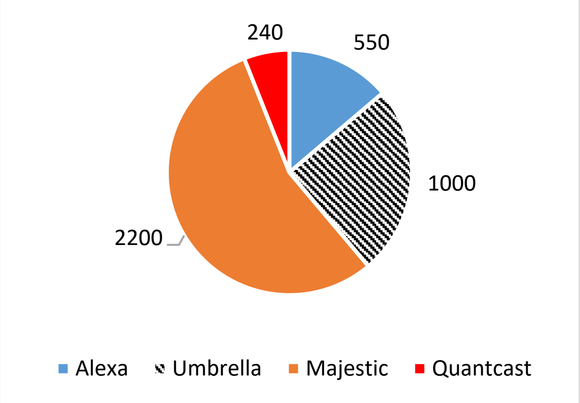

Figure 5. Benignity score of various ranking methods

Backlinks: These are Links from other websites to our website. By increasing the number of high-quality backlinks, the website rank increases.

Technical SEO: This includes features like mobile-friendliness, download speed and the ability of crawling in a specific website to ensure easily indexing by search engines.

Keyword Optimization: In order to help search engines to know what is the content of our website, we must use relevant keywords in it.

User Experience (UX): This factor shows how users are satisfied using a specific website.

Schema Markup: Some structured data for helping search engines better understand the website data.

Social Signals: The social emotional signs such as likes, shares, and other social interactions that a specific domain's content receives from its users.

Brand Signals: The overall reputation score of a website or domain amongst its users.

VI. DOMAIN RANKING PERFORMANCE ANALYSIS AND COMPARISON

This section is subdivided into five subsections. In subsection 6.1, we investigate the performance of six different ranking methods (Tranco, Quantcast, Umbrella, Majestic, Cloudflare radar and Alexa) from stability, benignity and availability point of view. In subsection 6.2 we compare these methods based on five important criteria that should be applied when evaluating any web site (i.e. authority, accuracy, objectivity, currency, and coverage). In subsection 6.3, we design a new evaluation scenario with more effective google-based ranking parameters for comparing the performance of different domain ranking methodologies. In subsection 6.4, we compare the performance of these ranking methods based on TLD validity and mean website volatility performance. Finally in subsection 6.5, we compare the performance of different ranking methods based on a unified mixed score which has been derived from scenarios A to D.

It must be mentioned that in performance comparison between different methods in scenarios A to C, we have used the Likert scale [29].

Numerous types of rating scales have been created to assess attitudes directly, meaning that individuals are aware that their attitudes are under examination. The Likert scale is the most commonly utilized among these ratings. The Likert scale, in its completed version, consists of either five (or even seven) points, enabling individuals to convey their level of agreement or disagreement with a specific statement (see Fig. 6). This scale generally offers five response options, allowing participants to express the intensity of their agreement or disagreement concerning the posed statement or question.

A Likert scale is a psychometric measurement tool commonly employed in surveys to gauge participants' preferences or their level of agreement with a particular statement or a series of statements. Participants evaluate quality on a continuum ranging from high to low or from best to worst, typically utilizing five or seven response options.

The response categories within Likert scales possess a hierarchical order; however, the intervals separating the values cannot be assumed to be uniform. The metrics such as mean, median, mode or other statistical distributions can be used for interpreting the results. Experts have similarly argued that the analysis can also employ frequencies (the percentages of responses within each category), contingency tables, $\chi 2$ tests, the Spearman rho evaluation, or the Mann-Whitney U test, rather than relying only on parametric tests.

Likert-type scales are commonly employed in the fields of medical education and research. They are often utilized for purposes such as gathering feedback from trainees at the conclusion of their rotations, evaluating trainees by faculty members, and assessing performance following educational interventions.

To assigning appropriate Likert scale, we have investigated related web content and scientific papers related to each domain ranking method and verified the results by consulting with experts in the field. Similarly, for scenario D, we have used a similar approach for rating each domain ranking method.

6.1 Scenario A

Popularity ranking data collection processes relies on a limited and small range of the Internet sites, either focusing on a specific criterion or obtaining information from a small population. It means that sampling of a small amount of targeted traffic can be considered significant on the scale of the entire Internet. Additionally, ranking in the previous methods

[DOI: 10.61186/itrc.16.4.44]

presented providers typically don't filter out automated or fake traffic or domains that don't represent real websites which further reduces the number of domains with real traffic in their listings.

Therefore, malicious agents can have incentives to manipulate the ranking of domains. They can do this by:

Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
(1)	(2)	(3)	(4)	(5)

Figure 6. Different levels of Likert scale (5-level case)

- Whitelisting their own malicious domains.
- Hiding malicious actions in other domains.
- Influencing broader policy decision-making by changing the perceived prevalence of security issues and practices.

Also, result of the investigations shows that Alexa's most used list was vulnerable to manipulation. Because Alexa method, rely more on their "traffic rate" browser extension that reports all page views, for example, in this method, traffic can be feigned by installing an extension on the Chrome browser and then automatically visiting our domain [6], [30].

In Table I, the indicators used in the ranking of the domains that are used by the main and major ranking analytical methods and platforms are presented. For the main ranking methods, the score of their inherent features, including similarity, stability, responsiveness, and degree of benignity, have been specified. In this table, the lowest score of the index and characteristic of stability, availability and benignity is 1 and the highest score is 5 according Likert scale [31].

Similar to Fig.6, the methodology for filling the Table 1 is based on designing a form which has 5 sentiment levels (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree) for each indicator of an specific ranking method. Then, we have consulted with the experts in the field to fill the forms and then taken the average result and rounded it to the nearest natural number (between 1 to 5) to represent its Likert scale.

The scores in Table I, are according to the figures and diagrams of the previous section. As indicated in the Table I, the Tranco and Cloudflare radar methods have the best characteristics, including stability rating, accessibility rating, and the site's benignity rating against manipulation [32].

TABLE I. INDICATORS AND METRICS OF DIFFERENT METHODS (PLATFORMS) FOR RANKING WEBSITE DOMAINS IN SCENARIO A

Ranking Method	Ranking indicators	Stability score	Benign score	Availability
Cloudflare radar	Using DNS data to calculate the top and trending domains found on both the global and country pages. Uses privacy-enabled 1.1.1.1 resolver and machine learning (ML) techniques to filter bot-generated traffic and only focuses on real user generated traffics	5	5	5
Alexa	Measurement is based on a combination of page views and site users and end users. Through the browser extension, the visited URLs are recorded	1	2	4
Cisco Umbrella	Counting the number of IPs of a domain based on OpenDNS traffic. The number of unique IP addresses to which DNS requests are exported.	2	3	1
Majestic	By counting the number of subnets a web page hosts to link to a domain Ranking is based on the referring subnets.	3	4	2
Quantcast	Counting site visitors is through an analytical script	4	1	3
Tranco	It provides an intelligent and combined collection of all valid data lists of available ranking methods and is resistant to list manipulation.	5	5	5
Common Crawl	This method is performed by crawling and counting the number of subnets hosted by each domain on a monthly basis.	1	1	1
Moz Pro	Its range and rating scale is small. It is used to rank sites locally.	1	1	1
imilarweb	Similar to Alexa, they provide tools for data collection, analysis and ranking.	1	1	1
Serpstat	It acts as an SEO platform with various tools to help webmasters analyze their situations and make decisions.	1	1	1
Spyfu	It is a powerful tool for competitor analysis, which is one of Alexa's main services that is very useful.	1	1	1
Comscore	Similar to Alexa, the website ranking list, which ranked by page-wide web traffic and provided a list of the most visited sites online.	1	1	1
Watch Them Live	It's an analytics program for websites that basically records user behavior as it happens and performs general analytics and more.	1	1	1

In the above table, ranking methods in 12 platforms presented that are very different. Researchers in the field of web security or Internet measurement use the ranking of popular and top websites. However, these rankings have different opinion about which domains are the most popular, as they can change significantly on a daily basis and can also be manipulated by malicious agents.

Fig. 7 shows the score of the ranking indicators in the main and major platforms including Tranco, Majestic, Quantcast, Alexa, Cisco Umbrella and other sub-platforms (according to the Table I). According to Fig. 7, Tranco and Cloudflare radar methods are the best method and Cisco Umbrella and other sub-platforms are the worst method in ranking sites in terms of stability, access and manipulation. Total scores of indicators of stability, availability and benignity for Tranco and cloudflare radar platforms is 15 scores, Majestic platform is 9 scores, Quantcast platform is 8 scores, Alexa platform is 7 scores, Cisco Umbrella platform is 6 scores and other platforms is 3 scores.

6.2 Scenario B

In this subsection, we investigate the performance of five Internet domain ranking methods based on five important criteria that should be applied when evaluating any web site. These criteria are: authority, accuracy, objectivity, currency, and coverage. These

[DOI: 10.61186/itrc.16.4.44]

criteria are described in the Table II and are scored based on [33].

In Fig. 8, we have compare the performance of six ranking methods (Tranco, Umbrella, Majestic, Alexa, Cloudflare radar and Quantcast) from averaging the viewpoint of 10 independent experts based on the scoring system parameters presented in [33] using Likert scale.

Based on metrics used in Fig.8, The sorted mean score of Cloudflare radar, Tranco, Alexa, Cisco Umbrella, Majestic and Quantcast are 4.042, 3.87, 3.47, 3.22, 2.6 and 2.45 respectively.

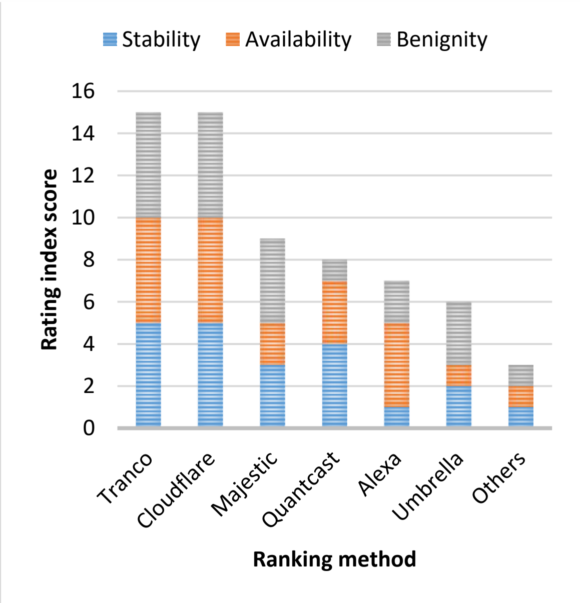

Figure 7. The score of ranking indicators in different platforms in scenario A

As can be verified in Fig. 8, the Cloudflare radar ranking outperforms other ranking methods in currency, accuracy and objectivity criteria. Tranco and Alexa have the best performance in coverage criterion. The overall mean worse coverage performance is associated with the Quantcast ranking.

TABLE II. IMPORTANT SCORING SYSTEM FOR WEB SITE EVALUATION [33] IN SCENARIO B

Metric	Sub-criteria/sub-metric	Score
Authority	• The author(s) first and last name is clearly identifiable (6 points) • The author's credentials are present (8 points) • The author is qualified to address the topic (6 points) • The publisher/sponsor of the website is identifiable (6 points) • Contact information is provided (email, phone, address(4 points)	30
Accuracy	• Information appears as accurate (5 points) • Information can be verified elsewhere (5 points) • Spelling and grammar are correct (4 points) • The website contains links to reputable outside sources for additional information (3 points) • The site has citations for borrowed materials (3 points)	20
Objectivity	• The purpose of the site is clear (6 points) • The site is unbiased or a bias(es) are easy to identify (8 points) • The site is free of advertisements (3 points) • The site is not a blog or wiki page (3 points)	20
Currency	• The creation date is provided (4 points) • The latest revision date is provided (4 points) • The revision date is appropriate for the subject (not out of date) (6 points) • All links provided are current (6 points)	20
Coverage	• The site is well organized (3 points) • Contains relevant information (4 points) • The site can be viewed completely (no fees, special browser requirements, or "under construction" signs) (3 points)	10
Total score		100

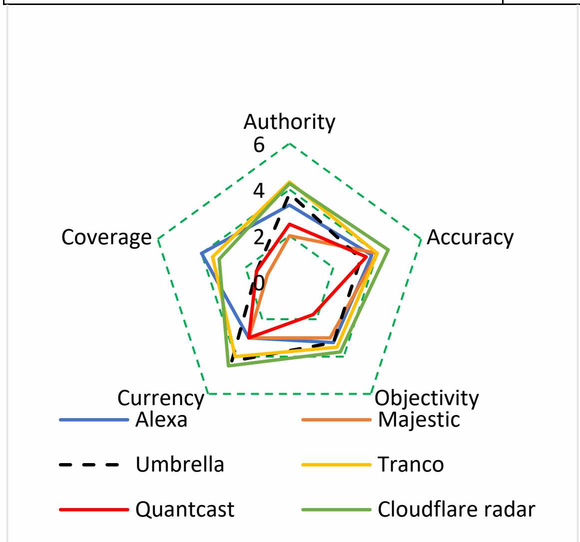

Figure 8. The score of different ranking methods in scenario B

6.3 Scenario C

In another scenario, we have compared the six proposed ranking methods (Alexa, Umbrella, Tranco, Quantcasy, Majestic and Cloudflare radar) with google's top 8 ranking factors [28].

We have ranked the capability of different ranking methods based on these 8 parameters using Likert scale (see Fig. 9).

If we define a *similarity measure* as the closeness of the average score of a ranking system in all of the 8 metric in this chart to google's ranking top score (5), we can conclude that the similarity measures associated with Cloudflare radar, Tranco, Cisco Umbrella, Alexa, Majestic and Quantcast are 3.17, 3.04, 2.27, 1.87, 1.5 and 1.22 respectively. Hence, the Cloudflare radar ranking methodology is the most similar approach with the google ranking methodology and Alexa and Quantcast are the most different ones.

6.4 Scenario D

In another comparison scenario, we have compared the five proposed ranking methods from TLD validity point of view for top 1000 and top 1000000 sites based on the latest IANA TLDs list (see Table III).

TABLE III. TLD VALIDITY PERFORMANCE COMPARISON FOR DIFFERENT RANKING METHODS IN SCENARIO D

	Top 1000	**Top 1000000**

[DOI: 10.61186/itrc.16.4.44]

Ranking method	valid	invalid	valid	invalid
Tranco	400	1	1500	100
Cloudflare	860	0	2410	1
Majestic	57	0	810	12
Umbrella	22	0	778	2536
Alexa	162	2	1001	3212
Quantcast	34	0	901	45

It can be verified in the Table III that the Tranco still has the largest amount of valid TLD in top 1k and 1M web sites and Umbrella has the worst performance from this perspective. In another viewpoint, we have compared the mean volatility/changeability of ranking with respect to cumulative distribution of domains for 1, 3 and 12 months periods in Fig. 10 for different five ranking methods [34]. It can be verified that Alexa and Quantcast has the worst 12-months volatility performance and Tranco and Cloudflare radar are the best ones in this point of view.

6.5 Unified mixed score

Based on the results of scenarios A, B, C and D, the scenario/metric weighting/importance factors as depicted in Table IV have been presented. These weighting factors have been derived using consulting with experts in the field.

TABLE IV. IMPORTANCE LEVEL ASSOCIATED WITH EACH SCORING SCENARIO (A, B, C, D) AND THEIR ASSOCIATED METRICS

Scenario	Scoring metric	Weight (%)	No.
A (30%)	Stability	50	1
	Benignity	20	2
	Availability	30	3
B (20%)	Authority	30	4
	Accuracy	20	5
	Objectivity	20	6
	Currency	20	7
	Coverage	10	8
C (40%)	Quality content	20	9
	Backlists	15	10
	Technical SEO	20	11
	Keyword optimization	15	12
	UX	15	13
	Schema markup	7	14
	Social signal	5	15
	Brand signal	3	16
D (10%)	Non-volatility	100	17

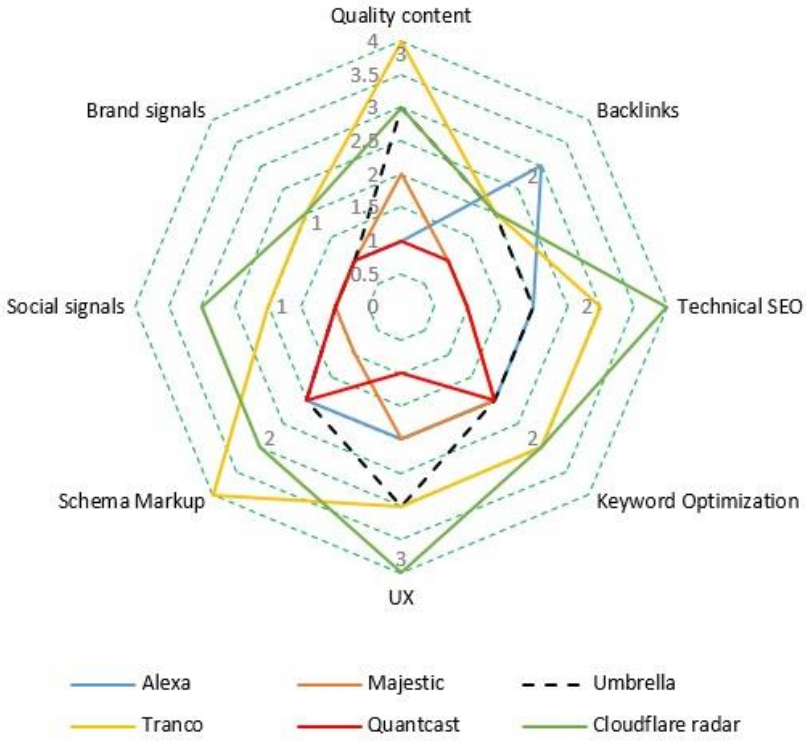

Figure 9. Performance comparison for different ranking methods based on Google's 8 top ranking parameters in scenario C

[DOI: 10.61186/itrc.16.4.44]

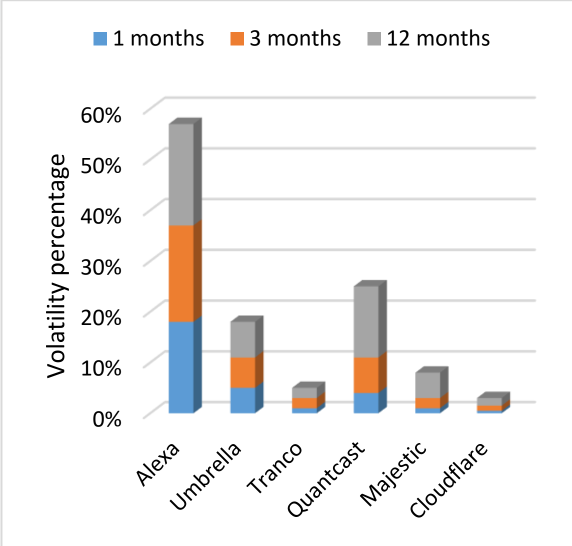

Figure 10. The mean website volatility performance for different ranking methods in scenario D

We have used multi-criteria decision making (MCDM) for calculating the overall ranking score between different ranking scenarios/criteria.

Multi-criteria decision-making represents a significant challenge in the realm of decision-making, focusing on identifying the optimal alternative by evaluating multiple criteria during the selection process. MCDM encompasses a variety of tools and methodologies that can be utilized across diverse domains, ranging from finance to engineering design.

The analytic hierarchy process (AHP) -as a sub-category of MCDM- is a systematic method employed in decision-making theory for organizing and analyzing complex choices, integrating principles from both mathematics and psychology. This method provides a reliable means of quantifying the weights assigned to various decision criteria. It leverages the experiences of individual experts to assess the relative significance of factors through pair-wise comparisons. Respondents evaluate the importance of each pair of items using a specifically designed questionnaire. By utilizing AHP, decision-makers can ascertain the relative importance of criteria and, where relevant, sub-criteria, ultimately leading to the identification of the most favorable alternative.

A simple typical example of using AHP is choosing a leader in an organization. The objective of this decision is to identify the most appropriate leader among three candidates. The criteria for evaluation include experience, education, charisma, and age. Assume that the weight associated with these criteria be 0.1, 0.4, 0.3 and 0.2 respectively. Assume that each criterion has a value between 1 and 5 based on the Likert scale. Hence, based on AHP method, a person with the highest weighted mean score will be chosen as the leader of organization.

We have selected AHP for calculating the overall ranking score between different ranking scenarios/criteria and employed the results of Table IV to obtain an overall normalized domain ranking efficiency score for each ranking method S_j (j=1,2,…,6) based on Eq. 1 for each of the six different ranking methods (Cloudflare radar, Tranco, Cisco Umbrella, Majestic, Quantcast and Alexa) as follows:

$$S_j = \frac{1}{5}\left\{\sum_{i=1}^{3} S_{ji}^{A} w_i^{A} \times 0.3 + \sum_{i=1}^{5} S_{ji}^{B} w_i^{B} \times 0.2 + \sum_{i=1}^{8} S_{ji}^{C} w_i^{C} \times 0.4\right\} + S_j^{D} \times 0.1 \quad (1)$$

in which, w_i^M is the weight associated with scoring metric i in scenario M ($M \in \{A, B, C, D\}$). Multiplier 1/5 is the normalization factors associated with Likert scale scoring in scenarios A, B and C respectively. Importance levels 0.3, 0.2, 0.4 and 0.1 are associated with scenarios A, B, C and D respectively according to Table IV. S_{ji}^M is the score of ranking method j for i^{th} metric of scoring scenario M ($M \in \{A, B, C, D\}$).

It must be mentioned that for filling S_j^D in Eq. 1, we have used the mean 12-months site non-volatility percentage in place of mean 12-months site volatility percentage in scenario D (as depicted in Fig. 10) because it better represents the domain ranking performance of this scenario (non-volatility percentage=1- volatility percentage).

In Fig.11, we have depicted the overall normalized ranking score for each of the six different methodologies. As can be deduced from this figure, the Cloudflare radar and Tranco are the best domain ranking methods respectively from this point of view. In order of importance, Cisco Umbrella, Majestic, Alexa and Quantcast are placed in other positions in terms of the proposed mixed metric domain ranking scoring system (Eq. 1).

VII. CONCLUSION AND FUTURE RESEARCH AREAS

Nowadays, ranking methods for the most popular active websites are being used by researchers, security analysts and industry activists to demonstrate the importance and relevance of the Internet domain names.

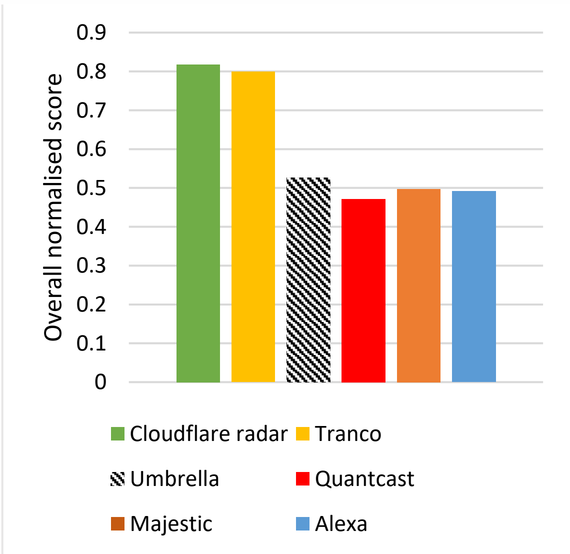

Figure 11. Overall normalized ranking efficiency score (between 0 and 1) of different domain ranking platforms using AHP methodology

[DOI: 10.61186/itrc.16.4.44]

However, the features and properties of the adopted methodologies are largely unknown. In this article, a review and analysis has been done on different analytical methods and platforms used for ranking, and their characteristics and also their impacts on the list of results have been evaluated.

The findings of this study show that all rankings methods can be manipulated easily (even on a large scale) with fake requests like the Alexa method. In particular, Alexa's comprehensive list can be easily manipulated. Considering the security implications in decisions and policies, by placing sites in whitelists, attackers can use manipulation techniques to make malicious domains look harmless and impose security issues/treats.

In this paper, standard options are evaluated in improving the important characteristics of ranking methods, including stability, responsiveness, and degree of benignity, and show that no list has a disproportionate impact on the combined Tranco list, and the Tranco list changes in average less than 0.6% per day. It means that the Tranco list can be used even in different applications, since the set of domains does not change significantly.

We have employed the MCDM and specially the AHP methodology to obtain an overall ranking score between different competing ranking scenarios/criteria. Based on the results of this paper we can conclude that the Cloudflare radar and Tranco are the most efficient ranking methodologies and also the Cloudflare radar ranking can better resemble the google's website ranking behavior in terms of 8 important google's ranking parameters.

Future research areas include the following disciplines:

As new and data-oriented ranking methods evolve according to big data property of web-based content, new data-oriented and machine learning-based ranking methods can be proposed that take into account the user perception and rank the search results in more efficient and user-friendly manner.

The use of fuzzy inference system (FIS) is also recommended as future research direction in place of the proposed AHP methodology.

In order to continue the important work of measuring the current state of website rankings, Tranco's method can be developed and extended. Because this method provides a new approach to creating rankings that intelligently aggregates existing lists and improves features that are important for conducting valid and reliable research. This method can also be shared publicly and its regularly updated lists can be used as an accessible and verified source for obtaining future popularity ratings in future domain ranking approaches.

Another important research area is incorporating Internet domain name allocation policies (such as location requirements, application limits, trademark policies, information availability, etc.) in evaluating the performance of different ranking methods [35].

Furthermore, leveraging domain ranking log correlation and data fusion techniques can be adopted for analyzing the performance of multiple web ranking methodologies.

Incorporating other emerging Internet domain ranking strategies such as Amazon Quicksight,, SERanking [32] or Semrush ranking [36] which rank the website based on multiple different metrics such as new traffic statistics derived from organic searches, visibility level within Google's search results and other artificial intelligence (AI) and local SEO tools in the performance comparison will be another open research area.

Another research direction is considering other MCDM methodologies such as technique for order of preference by similarity to ideal solution (TOPSIS), aggregated indices randomization method (AIRM) and treatment of the alternatives according to the importance of criteria (TACTIC) as alternative solution for ranking efficiency evaluation of ranking platforms [37].

Finally, the application of artificial intelligence and deep learning (DL) techniques on big data logs generated by different ranking methods can give more advanced insights regarding the ranking efficiency of each particular ranking methodology.

REFERENCES

[1] "How Many Websites Are There in the World?" [online] Available at: https://siteefy.com/how-many-websites-are-there/ [Accessed: 24 August 2023

[2] K., Talattinis, Z. Christina, and S. George. "Ranking domain names using various rating methods", ICCGI 2014.

[3] https://www.hillstonenet.com/wp-content/uploads/Hillstone_HSM4.7.0_EN_0321.pdf

[4] L. H. Orans, J. D'Hoinne, and J. Chessman. "Market Guide for Network Detection and Response", 2020.

[5] D. Sharma, R. Shukla, A.K. Giri and S. Kumar, "A brief review on search engine optimization." 9th international conference on cloud computing, data science & engineering, 2019.

[6] V. L. Pochat, T. V. Goethem, S. Tajalizadehkhoob, M. Korczyński and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation.", arXiv preprint arXiv:1806.01156, 2018.

[7] D. Giomelakis and A. Veglis, "Investigating search engine optimization factors in media websites: The case of Greece." Digital journalism, pp. 379-400, 2016.

[8] E. Cooper, "A guide to internationalized top-level domains." No. 178 Managing Intell. Prop., 2008.

[9] P. S. Sharma, D. Yadav and R. N. Thakur, " Web Page Ranking Using Web Mining Techniques: A Comprehensive Survey", Mobile Information Systems, pp.1-19, 2022.

[10] L. Rabiei, M. Mazoochi and M. Bagheri,"Web Domains Ranking with Real User Traffic Based on the Big Data Platform", International Journal of Information and Communication Technology Research, vol. 12, no. 1, pp. 32-41, 2020.

[11] A. Signorini, A survey of Ranking Algorithms. Department of Computer Science, University of Iowa, 2005.

[12] D. K. Sharma and A Sharma, "A comparative analysis of web page ranking algorithms". International Journal on Computer Science and Engineering, vol. 2, no. 8, pp. 2670-2676, 2010.

[13] A. Borodin, G. O. Roberts, J. S. Rosenthal and P. Tsaparas, "Link analysis ranking: algorithms, theory, and experiments". ACM Transactions on Internet Technology (TOIT), vol.5, no.1, pp. 231-297, 2005.

[14] J. M. Kleinberg, "Hubs, authorities, and communities". ACM computing surveys (CSUR), 1999.

[15] L. Page, S. Brin, R. Motwani and T. Winograd, The PageRank citation ranking: Bringing order to the web. In. Stanford InfoLab, 1999.

[16] J. Kline, A. Aelony, B. Carpenter and P. Barford, " Triangulated Rank-ordering of Web domains", 32nd International Teletraffic Congress (ITC 32), 2022.

[17] H. Zhao, Z. Chang, W. Wang and X. Zeng, "Malicious Domain Names Detection Algorithm Based on Lexical Analysis and Feature Quantification", IEEE Access, 2019.

[18] B. Altay, T. Dokeroglu, and A. Cosar, ''Context-sensitive and keyword density-based supervised machine learning techniques for malicious Web-page detection,'' Soft Comput., vol. 23, no. 12, pp. 4177–4191, Jun. 2019.

[19] D. Huang, K. Xu, and J. Pei, ''Malicious URL detection by dynamically mining patterns without pre-defined elements,'' World Wide Web, vol. 17, no. 6, pp. 1375–1394, Nov. 2014.

[20] M. Zouina and B. Outtaj, ''A novel lightweight URL phishing detection system using SVM and similarity index,'' Human-Centric Comput. Inf. Sci., vol. 7, no. 1, pp. 1-13, Jun. 2017.

[21] S. Schiavoni, F. Maggi, L. Cavallaro, and S. Zanero, ''Phoenix: DGA-based botnet tracking and intelligence,'' in Proc. 10th GI Int. Conf. Det. Int. Malware, Vulnerability Assessment (DIMVA), pp. 192–211, 2014.

[22] S. Englehardt and A. Narayanan, "Online tracking: A 1-million-site measurement and analysis", Proceedings of the ACM SIGSAC conference on computer and communications security. 2016.

[23] T. Alby and R. Jäschke, "Analyzing the Web: Are Top Websites Lists a Good Choice for Research? in Linking Theory and Practice of Digital Libraries", 26th International Conference on Theory and Practice of Digital Libraries, TPDL Padua, Italy, 2022.

[24] https://labs.ripe.net/author/samaneh_tajalizadehkhoob_1/the-tale-of-website-popularity-rankings-an-extensive-analysis/

[25] https://blog.cloudflare.com/radar-domain-rankings/

[26] D. Prantl and M. Prantl, "Website traffic measurement and rankings: competitive intelligence tools examination", International Journal of Web Information Systems, vol.14, no. 4, pp. 423-437, 2018.

[27] J. M. Patel. "Introduction to common crawl datasets." Getting Structured Data from the Internet: Running Web Crawlers/Scrapers on a Big Data Production Scale, pp. 277-324, 2020.

[28] https://backlinko.com/google-ranking-factors

[29] https://www.simplypsychology.org/likert-scale.html

[30] G. E. Rodríguez, et al. "Cross-site scripting (XSS) attacks and mitigation: A survey." Computer Networks vol. 166, 2020.

[31] K. L. Wuensch, What is a likert scale? and how do you pronounce 'likert?'. East Carolina University, 2005.

[32] https://www.contentpowered.com/blog/alexa-com-dead-alternatives/

[33] https://www.studocu.com/en-us/document/the-university-of-texas-health-science-center-at-houston/introduction-to-applied-health-informatics/authority-accuracy-objectivity-currency-and-coverage-evaluation-for-websites/16159240

[34] https://www.profound.net/pages/resources/DomainRank_Whitepaper.pdf

[35] https://www.oecd-ilibrary.org/docserver/237020717074.pdf?expires=1730707550&id=id&accname=guest&checksum=30F6D7F79DC356EDC3388ACE2899FBFB

[36] https://www.semrush.com/kb/27-rank

[37] H. Taherdoost and M. Madanchian, Multi-Criteria Decision Making (MCDM) Methods and Concepts. *Encyclopedia*,.vol. 3, no.1, pp. 77-87, 2023.

Pejman Goudarzi received his B.Sc. degree in Electronics from Sharif University of Technology Tehran-Iran in 1995. He also received his M.Sc. and Ph.D. degrees in Telecommunications and Electrical Engineering both from Isfahan University of Technology, Isfahan-Iran in 1998 and 2004 respectively. Dr. Goudarzi is an associate research professor at ICT Research Institute (ITRC). His main research interests are: wireless video communication, distributed resource allocation algorithms, game theory, soft computing, data science and quality control in data networks.

Davood Maleki is a faculty member of ICT Research Institute (ITRC). He received his M.Sc. degree in computer software engineering from Ferdowsi University of Mashhad. Currently, he is working as a colleague and supervisor in fundamental, practical and strategic projects in Information Technology department of ICT Research Institute.

Alireza Mansouri is an Assistant Professor of ICT Research Institute (ITRC). He received his B.Sc. and M.Sc. from Sharif University of Technology, both in Computer Engineering-Software and his Ph.D. in Computer Engineering-Information Technology from University of Tehran. He is currently deputy of "IT faculty" and head of "Data Analysis Research Group". His research interests include data science, computational social science, cloud computing, and agent based modeling and simulation.

[DOI: 10.61186/itrc.16.4.44]