

Quality and readability of web-based Arabic health information on COVID-19: Infodemiology study

Esam Halboub

Jazan University

Mohammed Sultan Al-Akhali

Jazan University

Hesham M Al-Mekhlafi

Jazan University

Mohammed Nasser Alhajj (✉ m.n.alhajj@hotmail.com)

Thamar University

Research Article

Keywords: Health information, Quality, Misinformation, public health, Infodemiological, COVID-19

Posted Date: September 16th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-76512/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on January 18th, 2021. See the published version at <https://doi.org/10.1186/s12889-021-10218-9>.

Abstract

Objective: The study sought to assess the quality and readability of the web-based Arabic health information on COVID-19.

Methods: Selected search engines were searched on 13 April 2020 for specific Arabic terms on COVID-19. The first 100 consecutive websites from each engine were obtained. The quality of the websites was analyzed using the Health on the Net Foundation Code of Conduct (HONcode), the Journal of the American Medical Association (JAMA) benchmarks, and the DISCERN benchmarks instrument. The readability was assessed using an online readability calculator tool.

Results: Overall, 36 websites were found eligible for quality and readability analyses. Only one website (2.7%) was HONcode certified. No single website attained a high score based on the DISCERN tool; the mean score of all websites was 31.5 ± 12.55 . Regarding JAMA benchmarks, a mean score of 2.08 ± 1.05 was achieved by the websites; however, only 4 (11.1%) websites achieved all JAMA criteria. The average grade levels for readability were 7.2 ± 7.5 , 3.3 ± 0.6 and 93.5 ± 19.4 for Flesch Kincaid Grade level, SMOG, Flesch Reading Ease, respectively.

Conclusion: Most of the available web-based Arabic health information on COVID-19 doesn't have the required level of quality, irrespective of being easy to read and understand by most of the general people.

Background

Coronavirus disease 2019 (COVID-19) has been a terrifying disease since it appeared in December 2019 in Wuhan, China. The causative pathogen is known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1, 2]. The disease has spread exponentially that almost all countries have been affected, prompting to imposing curfew and lockdown aiming to limit the community spread of the disease. During such outbreaks, people crave news. They are eager to follow everything about the disease: numbers of new and critical cases, and related deaths; the performance of the health systems; the preventive measurements announced by the relevant authorities; the availability of therapeutic remedies and vaccines; innovating new policies to fight the disease and so on [3]. Hence, the access to and use of the World Wide Web -also known as Web (internet or net can be used interchangeably)- increases obviously, where people can find lots of information. Such information ranges from personal opinions and discussion groups to scientific articles in peer-reviewed journals.

In fact, the Web was a giant step of mankind. For example, "Just Google it" has become the first response upon facing unknown information or for answering a question. Theoretically, the Web seems to be a very good tool for the public to obtain additional medical information that they do not know about their conditions or about the current pandemic surrounding them [4, 5]. However, thousands or even millions of websites appear with every single-word search, but few of which are relevant to what we are looking for,

not to mention the quality of those few. Regrettably, the public does not know which websites are trustworthy and which are not, despite the fact that there are certified medical/health websites, but very few [6]. The access to the scientific articles, which are trustful, is limited and for-paid in most instances, along with using scientific terms that are not common and difficult to be understood by the public. The net result might be obtaining misinformation with a subsequent adoption of unhealthy behaviors; using unapproved drugs or harmful herbs; and applying inappropriate preventive measures. The problem of quality of the web-based health information is not language-exclusive, although its impact might be less obvious on English-speaking people due to the fact that most of the scientific production is published in English, and very few -but after a while- are translated to other languages.

In the Arabic world, very few speak English, and no certified Arabic medical websites are available, except for the international organizations websites that translate their contents into different languages [7-11]. In the time of COVID-19, many Arabic medical, educational, social, news, and even sports websites publish materials regarding the disease. The study, therefore, sought to assess the quality and readability of online Arabic health information on COVID-19.

Methods

This was an infodemiological study in which selected search engines were searched for specific Arabic terms on COVID-19.

Search Strategy

The search for websites was conducted on 13 April 2020. The cookies information was erased from the browser prior to starting the search. To prevent any biases arising from preceding searches, browsing was done using Incognito (InPrivate) mode. Using Google Chrome version 81.0.4044, the following engines were searched: "Google (<http://www.google.com>)," "Yahoo! (<http://www.yahoo.com>)," and "Bing (<http://www.bing.com>)." The most widely-used Arabic translations of the following words were used as search keywords: Coronavirus, Corona, and COVID-19. The following combination was used in Google search engine: "Coronavirus-كورونا فيروس" OR "Coronaكورونا-" OR "COVID-19 كوفيد-19". Upon agreement on the search strategy among all, each engine was searched by one of the authors.

The first 100 consecutive websites (the first 10 consecutive pages) from each engine were obtained. These websites from the three engines were checked for duplicates, which, when present, were removed. Websites that presented health information on COVID-19 in the Arabic language were selected for subsequent evaluation. The following criteria were applied to exclude the websites: 1) Language other than Arabic; 2) Information on COVID-19 just by hints, or exclusively audio or visual-based; 3) Complete scientific articles or textbook; 4) Banner advertisements or sponsored links and discussion forums; 5) Blocked sites, or sites with denied direct access (required ID and password); 6) No information about COVID-19; and 7) News and news agency, and social media. The remaining websites were included and assessed for quality and readability, as indicated below. Figure 1 depicts the different stages of the search strategy we followed.

Quality Assessment tools

The quality of the included websites was evaluated using DISCERN [12], the Journal of the American Medical Association (JAMA) Benchmarks [13], and the Health on the Net Foundation Code of Conduct (HONcode) assessment tools [14].

DISCERN tool is a questionnaire that includes 16 questions. It is structured into 3 sections: Questions 1–8 addresses whether the website can be trusted as a source of data about selected therapy; questions 9–15 are about therapy options, and question 16 measures the overall quality score at the end of the evaluation. Each question is scored from 1 to 5, where 1 indicates a poor website, and 5 indicates a good quality website.

The JAMA benchmarks were published by the Journal of the American Medical Association. This tool evaluates the following points: authorship (whether authors, their contributors, affiliations, and relevant credentials were displayed or not); attribution (whether clear references and sources for the content were provided or not); disclosure (whether ownership, sponsorship, advertising, underwriting, commercial funding or support sources and any potential conflicts of interest were displayed or not); and currency (whether dates of initial posting and updating of the content were mentioned or not). For each fulfilled criterion, the website scores 1 point; otherwise it scores 0 point. The range for each site is from 0 to 4 points.

The website that complies with HONcode is granted permission to display a stamp (HON award-like badge) on its website. This is a certificate that stays valid for 1 year only.

Quality assessment using DISCERN and JAMA was conducted by two authors (EH and MSA). To minimize the subjectivity, both authors assessed 5 websites together using these two tools, and they resolved any discrepancies by discussion. Moreover, later on, inter-examiner calibration was calculated for the whole websites. For HONcode, we downloaded its software and incorporated it as an extension into Google Chrome. With each search, a HONcode seal appeared on the certified website. For confirmation, each website with the HONcode seal was further checked for the currency of its certificate in the main HONcode website.

The guidelines for readability as indicated by the American Medical Association (AMA) and the US Department of Health and Human Services (USDHHS) were consulted. These guidelines recommend that the patient reading material, to be more accessible and understandable by the general public, should not be higher than 5th or 6th grade reading level [15, 16]. The readability was assessed using an online readability calculator tool, "http://www.online-utility.org/english/readability_test_and_improve.jsp." Although this tool was primarily designed to analyze the English text, it can be used for other languages, as indicated in the website. Moreover, before commencing the study, the authors tested the validity of this tool using Arabic texts. Three Arabic paragraphs with three different levels of difficulty (simple, medium, and difficult) were analyzed. The results revealed corresponding values based on the difficulty of the text. This website analyzes the text using different common, well-known analyzing tools (Gunning Fog Index

(GFI), Coleman Liau Index (CLI), Flesch Kincaid grade level (FKGL), Automated Readability Index (ARI), Simple Measure of Gobbledygook (SMOG), and Flesch Reading Ease (FRE)). The GFI, CLI, and ARI were not considered in the analyses because these indices use the number of letters to formulate the readability score. This formula is not applicable in the Arabic text as, unlike the English word, the Arabic word is composed of letters linked together. The acceptable readability level was set to be ≥ 80.0 for the FRE and < 7 for the FKGL, and SMOG [15, 16].

Results

The web search revealed a total of 157,086,000 results. Out of the 300 screened websites, 81 websites were excluded as duplicates, resulting in 219 websites analyzed for eligibility. One hundred and eighty-three websites were excluded (no Arabic language, irrelevant information, news and news agency, and audio or video content). Thus, 36 eligible websites were included for quality and readability analyses (Supplementary Table). The inter-examiner agreement (Kappa value) for the DISCERN tool and the JAMA benchmarks were 0.88 and 0.95, respectively.

The initial search showed that only two (5.5%) websites (“who.int” and “mayoclinic.org”) had the HONcode seal. However, with further confirmation in the HONcode website, the certificate of “who.int” website was expired “<https://www.hon.ch/HONcode/Conduct.html?HONConduct982312>”. Hence only one website (2.7%) was HONcode certified.

The high score (≥ 65) based on the DISCERN tool was not achieved by any website. However, 13 (36.1%) websites had moderate score (33-64), and the remaining ($n= 23$, 63.9%) scored low (16-32). In general, the mean score of all websites was 31.5 ± 12.55 .

Regarding JAMA benchmarks, all websites achieved a mean score of 2.08 ± 1.05 . Only 4 (11.1%) websites achieved all JAMA criteria (scored 4 out of 4). Two (5.6%) websites scored 0 (did not fulfill any of JAMA criteria). The majority (41.7%) of the analyzed websites had a score of 2 (achieved 2 criteria). Most websites didn't display information on the authorship and attribution, while they displayed information on the disclosure and currency (Table 1).

The mean grade level based on the Flesch Kincaid grade level was 7.2 ± 7.5 . However, most of the included websites (66.7%) had scores < 7 , reflecting easy content to be understood by the general public. The “unicef.org” website had the most difficult content (FKGL= 46.72). When excluding this website, the mean grade level dropped down to a score of 6.0 ± 3.1 . According to the SMOG Index score, the mean grade level needed to understand the text of the websites was 3.3 ranging from 3 to 5.3. The readability ease according to the Flesch Reading Ease Index revealed a mean score of 93.5 ± 19.4 . Again, the website “unicef.org” had the most complex text (FRE= -9.21). When excluding this website, the mean increased to 96.4 ± 8.2 . More details are presented in Table 2.

Discussion

As COVID-19 is a novel disease, it has been the trending news in all media and websites worldwide recently; and Arab media and websites have not been an exception. The current study sought to assess the quality and readability of the health information on COVID-19 provided by the Arabic websites. We searched the most famous search engines in the Arabic world. The first 100 websites from each engine were obtained, although the users mostly do not go beyond the first 20 websites [17-20]. The small number of included websites is due to the exclusion of news, and news agencies and media. At a time of a pandemic like COVID-19, these are the most frequent sources of information, at least from the users' point of view [21, 22]. However, these sources just broadcast and/or publish what they get from the responsible sources, along with their special (unknown) sources. Apart from the daily reports of the new cases and deaths, the relevant information from the health point of view is that related to the way the virus spreads, the signs and symptoms of the disease, the required preventive measures and guidelines, and the available treatment and vaccines. Unfortunately, such information is hardly, or inappropriately, covered in News and by news agencies and media. Hence, we excluded these websites [20, 23].

Assessing web-based Arabic medical information has been scarcely addressed but in context of few medical conditions other than COVID-19 like oral cancer [11], breast cancer [10], epilepsy [9], autism [8], and oral health [7].

Only one website was HONcode certified. Surprisingly, the HONcode certificate of the WHO website was invalid (expired). As a nonprofit and nongovernmental organization, HONcode aims at promoting transparent and reliable health information online and issues its certificates based on a minimum mechanism to provide good-quality, objective, and transparent medical information to the internet users. The certified websites have the right to display the HONcode seal; this means they agree to comply with the standards listed, and are subjected to random audits for compliance [24].

With regard to the DISCERN tool, no single website scored as high. Most of the shortcoming can be attributed to the second section (questions 9-15) where data about the treatments, the alternatives, the side effects ...etc. of the proposed drugs were scarcely or improperly covered. Even to less extent, the first section (question 1-8) also contributed to the low-quality score: no or scarce data were available on the aims and achieving them; relevance of the topic; source of information; date of publication; being biased or balanced; and area of uncertainty. The net result is a lower quality in the last question (overall). The shortcoming in the second section might be ascribed to the fact that the disease is novel, and no confirmed treatments and alternatives have been available yet. However, the low-quality score of the first section cannot be attributed to the same reason. Hence, the website should fulfill these criteria for any written content.

The mean score of JAMA benchmarks was 2.08 ± 1.05 . Most of the shortcoming comes from not-mentioning information on the authorship and attribution, while information on the disclosure and currency were displayed. It is strange to find health topic on whatever website without author and references.

The quality assessment was not as expected. The analyses revealed that the information was lower than the quality standards required for health information, and hence it was not entirely reliable. Similar results about health information on COVID-19 were concluded by Cuan-Baltazar et al. for the English and Spanish websites, but considering that they included the news agencies and media [5]. As the disease is more serious in Europe and USA, in terms of incident cases and associated deaths, the quality of the English and Spanish health information about COVID-19 is expected to be higher than the Arabic ones. It is not the case, however.

The availability of such poor information is misleading, especially nowadays where the disease is so close to everyone, prompting them to believe in what they read, despite the poor quality, and turn it to a practice that may eventually be harm. The picture is dark and gets darker if we consider the sites that have been excluded (like news agencies and media). Further, the scientific information about COVID-19 seems to be full of flaws owing to the fact that the disease is novel, and no full picture of its etiopathogenesis, clinical manifestations, laboratory findings, preventive and treatment measures have been in hand yet [5]. Hernández-García et al. [25] argued that “It is necessary to urge and promote the use of the websites of official public health organizations when seeking information on COVID-19 preventive measures on the internet.”

With regard to the readability, the analyses revealed simple text in most of the websites that can be read and understood by most of the general public. It is discouraging that most of the websites provide poor quality health information, but it is simple to read and understand; this jeopardizes the readers. It is advantageous to have websites that provide simple topics understood by most of the people, but this is disastrous considering the poor quality of these topics.

Conclusions

In conclusion, most of the available web-based Arabic health information on COVID-19 did not have the required level of quality, irrespective of being easy to read and understand by most of the general people. The internet is powerful, yet a two-edged tool when it comes to the health sector. Hence, the governments, in collaboration with international and national health agencies/organizations, have to adopt initiatives and actions that ensure spread correct and reliable information on the internet. To help achieve this, they have to support more visibility of reliable information, coordinates with scientific institutes or organizations aiming at sharing reliable information, develop simple tools to assess the quality of information on websites, and use these assessments to refine the misinformation and find reliable information.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the findings of this article are available from the corresponding author.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding.

Authors' contribution

ES and MNA participated in designing the study, data collection, interpretation of the data, statistical analysis and drafting the manuscript. MSA and HMA participated in designing the study, data collection, interpretation of the data and drafting the manuscript. All authors read and approved the final manuscript.

Acknowledgement

Not applicable.

Abbreviations

COVID: Corona Virus Disease; HONcode: Health on the Net Foundation Code of Conduct; JAMA: The Journal of the American Medical Association; USDHHS: US Department of Health and Human Services; FKGL: Flesch Kincaid grade level; ARI: Automated Readability Index; SMOG: Simple Measure of Gobbledygook; FRE: Flesch Reading Ease; SD: Standard Deviation.

References

1. [<https://www.who.int/ar/emergencies/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses>]
2. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, Iosifidis C, Agha R: World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International journal of surgery (London, England)* 2020, 76:71-76.
3. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P: Association of the COVID-19 pandemic with Internet Search Volumes: A Google Trends(TM) Analysis. *International journal of*

infectious diseases : IJID : official publication of the International Society for Infectious Diseases 2020, 95:192-197.

4. Tan SS, Goonawardene N: Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review. *J Med Internet Res* 2017, 19(1):e9.
5. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E: Misinformation of COVID-19 on the Internet: Infodemiology Study. *JMIR public health and surveillance* 2020, 6(2):e18444.
6. Kulasegarah J, McGregor K, Mahadevan M: Quality of information on the Internet-has a decade made a difference? *Irish journal of medical science* 2018, 187(4):873-876.
7. Almaiman S, Bahkali S, Alabdulatif N, Bahkaly A, Al-Surimi K, Househ M: Promoting Oral Health Using Social Media Platforms: Seeking Arabic Online Oral Health Related Information (OHRI). *Studies in health technology and informatics* 2016, 226:283-286.
8. Alnemy FM, Alnemy FM, Alamri AS, Alamri YA: Characteristics of Arabic Websites with information on autism. *Neurosciences (Riyadh, Saudi Arabia)* 2017, 22(2):143-145.
9. Alkhateeb JM, Alhadidi MS: Information about epilepsy on the internet: An exploratory study of Arabic websites. *Epilepsy & behavior : E&B* 2018, 78:288-290.
10. Alnaim L: Evaluation Breast Cancer Information on The Internet in Arabic. *Journal of cancer education : the official journal of the American Association for Cancer Education* 2019, 34(4):810-818.
11. Alakhali MS: Quality Assessment of Information on Oral Cancer Provided at Arabic Speaking Websites. *Asian Pac J Cancer Prev* 2020, 21(4):961-966.
12. Charnock D, Shepperd S, Needham G, Gann R: DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *Journal of epidemiology and community health* 1999, 53(2):105-111.
13. Silberg WM, Lundberg GD, Musacchio RA: Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor—Let the reader and viewer beware. *JAMA* 1997, 277(15):1244-1245.
14. Boyer C, Baujard V, Geissbuhler A: Evolution of health web certification through the HONcode experience. *Studies in health technology and informatics* 2011, 169:53-57.
15. Edmunds MR, Barry RJ, Denniston AK: Readability assessment of online ophthalmic patient information. *JAMA ophthalmology* 2013, 131(12):1610-1616.
16. Kher A, Johnson S, Griffith R: Readability Assessment of Online Patient Education Material on Congestive Heart Failure. *Advances in preventive medicine* 2017, 2017:9780317.
17. Borgmann H, Wölm JH, Vallo S, Mager R, Huber J, Breyer J, Salem J, Loeb S, Haferkamp A, Tsaur I: Prostate Cancer on the Web-Expedient Tool for Patients' Decision-Making? *Journal of cancer education : the official journal of the American Association for Cancer Education* 2017, 32(1):135-140.

18. Sacchetti P, Zvara P, Plante MK: The Internet and patient education—resources and their reliability: focus on a select urologic topic. *Urology* 1999, 53(6):1117-1120.
19. Nguyen SK, Ingledew PA: Tangled in the breast cancer web: an evaluation of the usage of web-based information resources by breast cancer patients. *Journal of cancer education : the official journal of the American Association for Cancer Education* 2013, 28(4):662-668.
20. Janssen S, Fahlbusch FB, Käsmann L, Rades D, Vordermark D: Radiotherapy for prostate cancer: DISCERN quality assessment of patient-oriented websites in 2018. *BMC urology* 2019, 19(1):42.
21. Case DO, Johnson JD, Andrews JE, Allard SL, Kelly KM, JotASfIS, Technology: From two-step flow to the Internet: The changing array of sources for genetics information seeking. 2004, 55(8):660-669.
22. Steffl-Mabry J: A social judgment analysis of information source preference profiles: An exploratory study to empirically represent media selection patterns. *Journal of the American Society for Information Science Technology* 2003, 54(9):879-904.
23. Weissenberger C, Jonassen S, Beranek-Chiu J, Neumann M, Müller D, Bartelt S, Schulz S, Möniting JS, Henne K, Gitsch G *et al*: Breast cancer: patient information needs reflected in English and German web sites. *Br J Cancer* 2004, 91(8):1482-1487.
24. HON Code of Conduct for medical and health Web sites. *American Journal of Health-System Pharmacy* 2000, 57(13):1283-1283.
25. Hernández-García I, Giménez-Júlvez T: Assessment of Health Information About COVID-19 Prevention on the Internet: Infodemiological Study. *JMIR public health and surveillance* 2020, 6(2):e18717.

Tables

Table 1: Quality analysis of the included websites (N= 36)		
	Frequency	Percent
HONcode		
Certified	1	2.8
Not certified	35	97.2
DISCERN		
High (≥ 65)	0	0.0
Moderate (33-64)	13	36.1
Low (16-32)	23	63.9
JAMA Benchmark		
No item met	2	5.6
One item met	8	22.2
Two Items met	15	41.7
Three items met	7	19.4
Four items met	4	11.1
Authorship-JAMA		
0 (Not met)	20	55.6
1 (Met)	16	44.4
Attribution-JAMA		
0 (Not met)	23	63.9
1 (Met)	13	36.1
Disclosure-JAMA		
0 (Not met)	12	33.3
1 (Met)	24	66.7
Currency-JAMA		
0 (Not met)	14	38.9
1 (Met)	22	61.1

	Flesch Kincaid Grade level	SMOG	Flesch Reading Ease
Mean	7.2	3.3	93.5
SD	7.5	0.6	19.4
Minimum	2.8	3	-9.2
Maximum	46.7	5.3	105
< 7 score	66.7% (n= 24)	83.3% (n= 30)	NA
≥ 7 score	33.3% (n= 12)	16.7% (n= 6)	NA
≥ 80 score	NA	NA	94.4% (n= 34)
< 80 score	NA	NA	5.6% (n= 2)

NA: Not Applicable.

Supplemental Information

A Supplementary Table was not provided with this version of the manuscript.

Figures

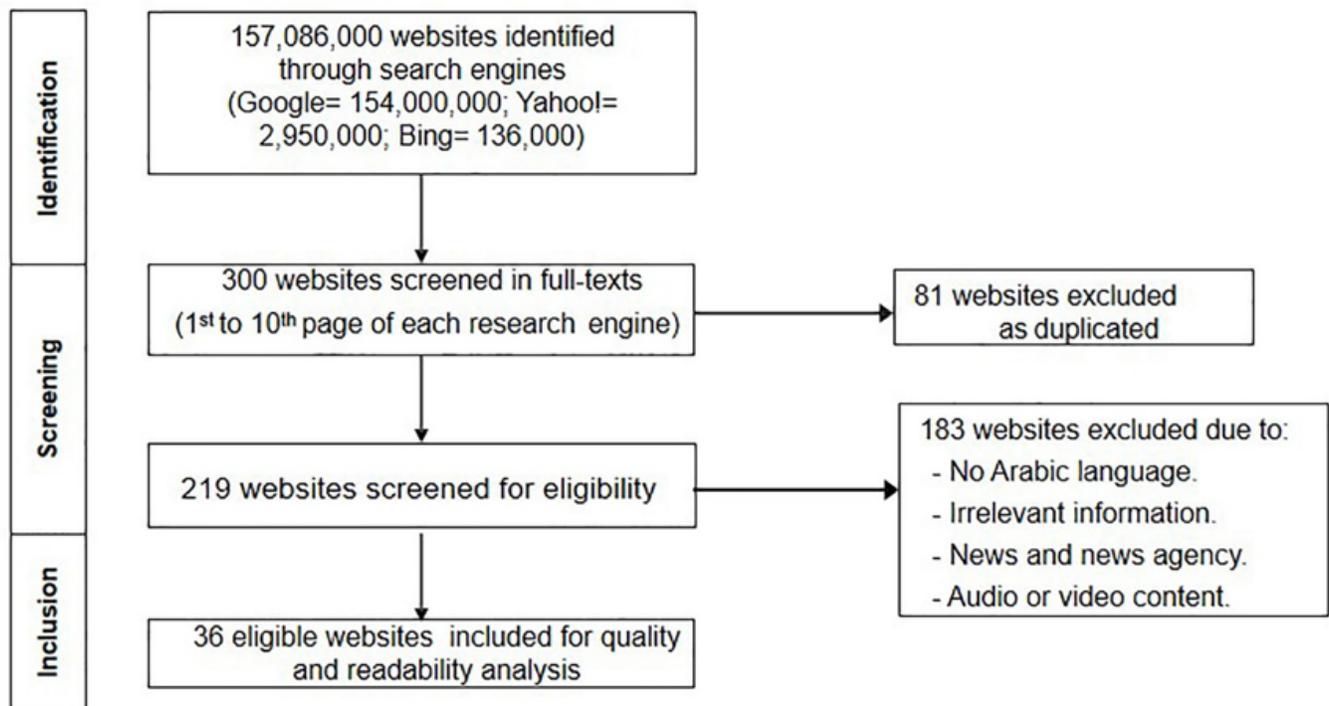


Figure 1

depicts the different stages of the search strategy we followed.