Using Internet Search Engines for Effective Information Retrieval

Gwarzo, Sani Mohammed(Ph.D)
Department of Library and Information Sciences
Bayero University Kano, Nigeria
sanigwar@gmail.com

Abstract
The paper examined the use of Internet search engines for information retrieval; the paper discussed the distinctions between a web browser and search engines and their types; the needs for the users to have knowledge about search strategies and search techniques were also highlighted. Limiting and expanding search results which were considered very important for a successful internet searching were discussed with practical examples. The paper ends with conclusion that the internet has become an important source for all types of information, but to retrieve important source, the users must have the skills and the knowledge of using the internet to find useful resources that will address their information needs. The paper recommended the need for academic and public libraries to spearhead the training of their users on finding good and reliable information from the Internet.

Keywords: Internet, web browser, Search Engine, Search strategy, Search techniques.

Introduction
Internet is a network of networks. A network is defined as two or more devices connected via a communication device in order to "see" and "talk" to each other. The Internet is built on the principle of one computer connected to another computer in an endless circle. In short the Internet is designed to connect millions of computers, from one organisation to another organisation or from one country to another country, making it possible to communicate and share information on a global basis. According to Internet World Stat (2019) the internet has consolidated itself as a very powerful platform that has changed forever the way we do business and the way we communicate. It has also become the universal source for communications and for sharing information to millions of people, at home, school and at work. Additionally, the Internet World Stat puts the total Internet users, as at 3rd March 2020, to be about 4, 574, 150, 134. This represents more than 59.6% of the world population of 7, 796, 949, 710 (Internet World Stat, 2020).

One distinctive feature of the Internet that makes it different from other organisations handling information, is the fact that Internet thrives on the willingness of individuals to share information. It is this willingness that continues to drive and sustain the Internet. Whoever has information he/she wants to share with the rest of the world will simply publish it on the internet. Singh (2001) described the Internet "as the biggest library in the world in which information is not properly structured and organized, there are no standardised rules of classification or access." This raises a number of fundamental problems, such as the amount of what is publish, lack of a central organisationto control what to and what not to publish.

As a source of Information, the amount of what is published on the Internet, and available to users, is staggering. As at 4th Feb, 2019 the indexed web contains about 5.29 billion web pages (Kunder, 2019), this number is based on pages indexed by just three search engines namely Google, Bing, and Yahoo. It is estimated that together with the other search engines the web contains more than 8 billion web pages (Alkhaja& Mohammed, 2012). With this staggering figure, how does a user find relevant document to address his information needs? Information resources owned by one organisation like a library, will take it upon itself to
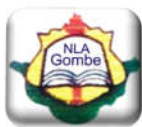
organise them by providing indexes that will assist users to retrieve the resources. But, because the Internet does not have a central authority, its resources are stored unorganized.

Today the Internet has become a house hold name in both developed and developing countries. People, organisations and institutions are using it to publish and to find every type of information they may require to satisfy their information needs. As an information dispenser, the internet contains billions of information resources in different subjects, formats and scattered on different types of computers across the globe. These resources unlike libraries are not organised in any way. To overcome this lack of organisation, search engines and Directories are voluntarily developed, to serve as the index to the contents of the resources publish on the Internet. In order for the search engines and the directories to store the contents of the resource, they break down the contents of the information in to keywords and store those keywords in a central database. Each time a user types his query all the keywords that matched the query are retrieved from the database regardless of whether they are relevant to the user or not. This therefore makes it important for users to develop the skills and the techniques of knowing what to include and what to exclude whenever they are making a request to the search engine for information resources.

Review of Related Literature
This section of the paper discussed literature pertinent to the topic. The literature is specifically, reviewed under the following subheadings: Web browsers, Search Engines, Types of Search Engines, Search Strategy and Search Techniques.

Web browser
A web browser, or internet browser, or just web browser is a software programme that allow users to access and display web pages. A web browser basically displays Hypertext Mark up Language (html) files containing text, audio files, images, still and videos files etc in an ordered lay out on a web page. According to Grauschopf (2019) the main purpose of an internet browser is to 'translate the code that computers use to create websites into text, graphics, and other features of the web pages that we are all used to seeing today". In addition to the display of web pages, web browser also establishes a continuous connection between the sending and the receiving computers so that the transfer is carried out smoothly (Beal, 2020). In other words, the function of the web browser is to connect to the computer that has the resources identified by the search engine and then display the results of what is found. This means that every web browser use the services of a search engine to find the Internet resources. This explains why web browsers are universal in nature; users are given the options to configure their web browsers to work with any search engine. For example, a user who is familiar with Bing or Internet Explorer browsers but is more comfortable using Google search engine can configure his browser to work with his preferred search engine.

Search Engines
Search engine is a component of every web browser, i.e every web browser use the services of search engine to find the resources. Search engines, sometimes referred to as Crawlers, spiders, wanderers or robots, are software programmes that are designed to search for information on the Internet. Wikipedia (2017) described a search engine as software that is designed to assist the users to retrieve information about documents stored in a local computer or on Internet. Using this definition, a search engine can be described as any software or computer programme that searches for information, regardless of whether the database is on a standalone computer or on a network; these networked computers could be near each other or at dispersed locations or from one country to another connected by communication lines. Boswell (2017) discussed the functions of search engines to include

filtering, retrieving information and assisting the users to find what they are looking for on a varied range of subjects and formats. He further pointed out that search engines are fast, easy to use and deliver consistently relevant results, which are the three things any internet user is expecting from a search engine.

Search engines use automated software that crawls through all the computers connected to the Internet and indexed the contents of what it found by keywords. The list of the keywords are then stored in a database on the computer hosting the software of the search engine. When a user issued a query the browser refer the query to the search engine, which in turn search its database and bring out all the keywords that matched the query. This is the reason why results for queries run tended to be very large, running into billions and also very fast. It should also be noted that every word that matched the query is retrieved, whether it is relevant or not. These are some of the facts that made it compulsory for users to know how to use the search engine to retrieve useful resources in accordance with their needs. In other words, an effective search is a search that returned minimum number of hits, which the user can easily browse to find exactly what is needed.

It should be noted that despite search engines effort to retrieve all available resources, there are certain types of resources that cannot be retrieved by search engines or directories. This is because the owners of the sites used different encryption software to mask the sites from being discovered. Scholars such as, Thompson, (2015) and Li, (2018) described the contents of these types of resources as Dark web, Deep web or Invisible web. Dark web, as the name connote is not pleasant to describe, as it stores resources that are against the law and therefore the owners of the web sites use a more powerful encryption tool to prevent search engines and directories from accessing their contents. These types of sites normally contains resources about buying and selling of recreational drugs, pornography, gun running, blackmail and many other vices that require secrecy (Thompson, 2015; Li, 2018).

Deep web and Invisible web on the other hand, offers websites that search engines cannot find. Thus the deep web to some is regarded as a generic name that covers not only the dark web but also web sites that require registration, and also all websites behind pay walls, requiring subscriptions like databases such as Ebsco host, Blackwell and Jstoretc (Guccione, 2020).

Types of Search Engines
There are many types of search engines each has its own characteristics, i.e. strength and weaknesses. Some are particularly good in finding all types of general information/document, while others confine themselves only to a particular type of information in a specialised area. Some search engines offer general information such as Google, while specialised search engine addressed specialised information in a particular subject or group of subject's e.g. Education World. Some are more academic than others, i.e. some presents more authoritative information acceptable to academic community, such as Google, Google chrome and Google scholar. Other search engines provide more commercially oriented information. Some carry a lot of adverts to offset the cost of their existence others carry very little or no advertisements at all. As indicated earlier, each search engine has its own strength and weakness; Users are therefore advised to become familiar with the characteristics of two or more search engines in order to reap the benefit of finding the needed information as complete, accurate and as fast as possible. This is particularly important in Nigeria, where telecommunication services are at the lowest ebb.
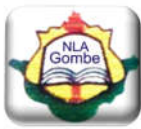
Webnots, (2019) classified search engines into four categories namely Crawler-based search engine, Human powered directories, Hybrid search engine and Other search engines. Crawler based search engines are those that creates their listings by using a piece of software that "crawl" or spider the web and then index what it finds. Human powered directories, on the other hand, depend on human editors to create the listings. Hybrid use both crawler based and manual index for listing the sites. According to Webnots, (2020) most crawler based search engine like Google uses crawler as a primary mechanism and human powered directories as secondary. The other types are the specialised search engines such as email, medical fields etc. However it should noted that the advantage between crawler based search engine and human powered directory is that while crawler based search are good for finding key words or specific known search topic, human powered directories are good for a general search in unfamiliar territory. According to Khiste, Veer and Maske (2011) "the greatest advantage of subject directories is that they are very useful in browsing subject information, especially in cases where highly specialised, selected information is required" This simply means that if a user already knows what he wants he can conveniently use a search engine. But in a situation where he is exploring or just browsing he can use directories to see the extent of what is available in that given area or subject.

Pujar (2011) however, categorized search engines into 6 different classes these included the following:

1. **General Search engine**
   eg. Bing, Google and Gigablast
2. **Meta search engine**
   eg Dogpile, metacrawler, Infospace
3. **Subject/Specific Search engine**
   egScirus, Intute, Entrez, Gopubmed and**VROOSH!**
4. **Intelligent/Specialised Search engine**
   egKasmix, Wolfram Alpha, Searchmedica,

5. **Deep/Invisible Search engine**
   eg Library catalogues, Databases
6. **Scholarly Search engine**
   eg Google scholar, Base, OAISTER

Search Strategy
Before conducting any search and the final decision to use either a search engine or a directory, developing an effective search plan is important. Search strategy is a plan for finding information that will providethe information needed by the users. According to Xie and Joo (2010) a search strategy refers to a plan for a search task. In simple language, a search strategy is the determination of how users intend to carry out the search using the identified keywords. The more thought and care a user puts during this plan, the more productive and more useful will be his search results.

Many scholars have proposed different ways of strategizing a search, for example, Bates (2020) discussed five strategies that can be used to find web resources, and these are:

a. Purpose of Research
b. Find a reliable Search engine
c. Identify keywords
d. Use Boolean Operators to narrow the search
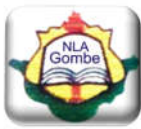
e.   Review the Results

Purpose of Research

According to Bates, (2020) understanding the purpose of the research is the first step in conducting fruitful search. Understanding the purpose, will allow the user to determine the extent, depth and the scope of the topic he wish to find on the internet. Johnson and Lamb (2007) proposed the creation of concept map or visual planning guide. The plan should "include the central topic or problem, a word list and a set of required information". It will also allow the user to determine what type of information is required (ie. Ideas, facts, opinions, options) what form of presentation is preferred (text, graphics, charts)? The user can write different sentences about what he is researching on in order to have a firm grip on the topic. He/she can write down any idea he has regarding the topic, even those harebrained. Poor understanding of the purpose will result in poor determination of the topic and consequently, results in retrieving documents that may not be useful to the research. The user should bear in mind that all information retrieval systems are simply designed to supply and deliver the information not exactly as he wanted it but close enough to enable him decide based on what he has specified in his search terms.

Find a reliable Search engine

The knowledge of when to use a search engine or a search directory is very important. This knowledge becomes handy when conducting a research in an area where the user is not familiar. It should be noted that when conducting a research in an unfamiliar area, using a directory is better than using a search engine. Apart from giving you the latest happening or current information on your subject, the directory also enables you to have a general idea about the topic you have chosen. It also assists you to find relevant references to your topic as well as list of important resources that have been written by experts in that area. In addition, the number of resources to consult is limited to what the editors knew existed and therefore very few compared to the results from a search engine.

Based on the foregoing, if the user decides to use a search engine, knowledge about their types, characteristics, and how each work is very important. According to Boswell (2017) "... not all search engines are created equal. Every search tool out there delivers a different experience and depending on what you are looking for, it might not always be pleasant". Therefore mastery of how to use two or more search engines is very important, more especially in Nigeria, where majority of users connect to the internet through GSM telephone service providers. The knowledge about two or more search engines will allow the user to use the best search engines and get all the information needed.

Identifying keywords From the word list generated the user should identify the key words that make up the topic of his research, if possible, identify the dependent and independent variables of the topic.

Identify the search terms that describe those concepts The user should look at the list of his keywords again. It will be good to determine the synonyms of the words he intends to use; words that have the same meaning with his keywords and also determine words that have alternative spelling, abbreviations and acronyms for the words on the list. The user should also identify words that have broader or narrower meanings to his original terms. This information is very important because as he starts the search he may likely use the words to refine (expand or contract) search terms. For instance, in the event he conducts a search and few hits are retrieved he can easily use the synonym words to expand the search.

Use Boolean Operators to narrow the search
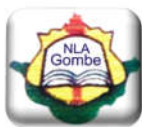
Boolean operators, AND, OR NOT are very effective when search for internet resources. they allow the user to limit his searches when too many results are obtained or to expand his result when few results are retrieved.

Review the Results
After you have conducted the search, look through what has been retrieved. Does the results of what was retrieved matches what you wanted to use for your purpose of the research. If it is contrary to your expectations, you need to re-formulate your search query and run the search again.

Search techniques
Suggestions on how to start a search was proposed by Harris (2000), who discussed five possible scenarios under which Internet resources may be sought:

a. This is the first scenario in which the internet user has a **broad subject** in mind, but don't know exactly what he is interested in or how to look it up. For broad subjects, such as information processing, Tibet, photosynthesis, volcanoes, Samuel Johnson, rope making and so on, Harris (2000) suggested the use of an online encyclopedia or similar reference site.
b. The second scenario is a situation in which the user needs information about a **general subject area.** The user wants to find information about business, law, medicine, old master paintings, government documents and so forth. Harris (2000) suggested the use of targeted (specialized) search engine
c. The User has a **well-known specific subject** in mind and wants to find some relevant sites. To find sites for subjects such as Hawaiian vacations, exercise routines, colleges and universities in California, dietary guidelines, art museums, and the best source for these types of information is the directory.
d. The user has a need for information on a **very specific and unusual subject,** such as tea tree oil, Chihuahua coughing, paradichlorobenzine, or the Airbus 320, Harris (2000) advised the user to use a general search engine.
e. When the user is writing a **research paper** and need substantial information from many sources. Use a combination of several search engines, directories, and other tools.

Having decided you want to use a search engine many Scholars, including Norton, (2011); Collins, (2018) have emphasised the need to narrow down the concept of the search terms even before starting the search. Because, the narrower you can get your search terms from the beginning, the more successful your search will be. To this regards, Collins (2018) advised the search terms to be specific, but with sufficient qualifiers in order to retrieve the exact documents. He cited as an example, searching for coffee, which he pointed out that using the single word, will retrieve much more than we can use. But when sufficiently qualified (roasted Arabica Coffee) the result of the search becomes manageable.

It should be noted that the search technique that can be used to retrieve resources can be grouped into two. The first is when a user runs a query and millions of results are displayed. In this instance the user needs to use the search techniques to limit the result to manageable number. Secondly, when the user runs a query and few results are displayed, in this case he needs to expand the search terms. In other words, there are search commands that a user can use to limit or expand his search result. The command for limiting result include the following:: (a) Use of Quotation marks to locate a Specific phrase (b) Searching within a site (c) Finding words within a Web Address (d) Search within web page titles (e) Limit searches
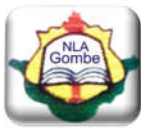
to specific domains (f) Use basic math to narrow the search (g) Finding document by file format (h) Use of wildcard.

1. Limiting Search Results

(a) Use of Quotation marks to locate a Specific phrase

Using specific phrase is the simplest way of narrowing and saving your precious time. By putting quotes around the most important phrases you are forcing the search engine to recognise the importance of the concept and to fish out only documents in which the two or more words appear side by side. According to Collins (2018) when you use quotation marks around a phrase you are telling the search engine to bring back pages that include these search terms exactly how they appeared. Norton (2011) emphasised that using quotation marks become handy especially if you are trying to retrieve document in which you remember only part of a phrase you have seen on a web page or part of a quotation.

(b) Searching within a site

Sometimes a user may have a particular site in mind, or title of the work, or the URL of a document, i.e. he knows that the site contain the exact information he is looking for. For example using the top level domain of .edu will return information from educational institutions. Substituting the formula [**site:.edu "subject I'm looking for"]** with the name of the school and indicating what is being sought will retrieve document from that educational establishment.

site:www.school.edu "subject I'm looking for" Frame your subject within quotations if it is more than two words; this lets the search engine know that you want to find those two words right next to each other.

However, if you have a particular school in mind that you'd like to search, use that school's Uniform Resource Locator (URL)address in your search.You can also use the **inurl** command to tell search engines (this works best with Google) to look only for web addresses, aka URLs, that include your search terms. You are specifically telling the search engine that you *only* want to look within the URL - you don't want to see results from anywhere else but only the specified URL and what is being sought could either be in the basic body of contents, titles, metadata etc .

According to Collins (2018) for the '**inurl**' to work you need to keep the following in mind:

- If you're looking for an exact phrase, say, "three blind mice", you'll need to remember that web addresses typically don't include spaces between words. However, if your query contains a common enough phrases, Google will automatically match results that include common variations. You might have to play with this a little bit in order to make it work how you'd like it to work - remember, while search engines are incredibly advanced, they still require help to understand what you're looking for.

- One of the most relatively unknown benefits of the inurl command within Google is that you can use it to search within directories and sub-directories. For example: say you want to find all possible Contact pages on a popular shopping site. You could try this query in Google: inurl:amazon.com "contact".

- Don't want certain sites to show up in your search results? You can exclude by using a simple Boolean math, the minus sign. For example, say you wanted to search for football information, but you wanted to exclude anything from NFL.com You can try this command: football -inurl:nfl.

Nigerian Library Association, Gombe State Chapter

allintitle is another important command that can be used to retrieve information. **You can use this command to restrict search engine to return only titles of documents with that word/s in the title of the document. However it should be noted that t**he term **allintitle** is a search operator specific to Google.

File type is another command that will retrieve only the requested file extension, eg.**file type:**(.pdf) (.swf, .mp3 etc)

2. Expanding Search Results
Commands that can expand results are (a) Boolean operators (b) Use of Alternate or Synonym words and then (c) Use of Multiple search engines.
   a) Use of Boolean operators (**AND OR NOT or +, -**) are used to connect and define the relationship between search terms. These can be used to expand the search results. In the search strategy stage, the user was advised to determine the synonym and alternative spellings of his keyword before conducting the search.
   b) The user can use the Boolean operators of "OR" and "AND" to include those synonym and alternative spellings expand his search results. For example (valuable OR precious OR priceless) (Catalog OR Catalogue).

   c) Not all search engines are equal. The user can use multiple search engines to expand his search results.

Some of these commands can be combined to make a very precise and powerful search, eg **Intitle:** smithonian **Site:** YouTube.com will return you tube videos. By incorporating the two you make your search more efficient; they yield a greater amount of quality content that accurately meets your expectations (DeCesare, 2014).

Methodology
The methodology used for this paper is purely literature based which includes the use of current knowledge including report of substantive findings and review of other studies, as well as theoretical and methodological contribution to topics related to the subject matter of the paper.

In general, the author relied heavily on the secondary sources of data collection based on the existing and current literature available in textbooks, journals, reports and internet resources. There are a number of existing guidelines for using literature review methodology. However, the present paper adopted the use of systematic and integrated review (Whittemore, and Knafl, 2005).

Conclusion
The Internet has become an important source of all types of information, regardless of the format. However, to retrieve important resource to address users' information needs, users must have the skills to determine the type, depth and breadth of what they need and also have the knowledge on how to retrieve resources to address those needs. Hence the imperative for the users to have knowledge on the types of search engines, search strategy and search techniques.

Recommendations
This study wishes to make the following recommendations:
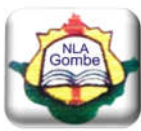
a. Academic and State public libraries should spear head the training of users on best practices for using the internet to find relevant resources that will address their information needs.
b. There is need for users' to know how to determine whether to use a search engine or directories whenever they need pertinent information.
c. Also the users need to have the skills to limit or expand their search results.

References

Alkhaja, A. and M. Mohammed (2012) Search Engines retrieved from *Slide share*

Bates, A. (2020) Elementary Internet Search Strategies: The Information Treasure Hunt Retrieved 23 June 2020, from https://www.slideshare.net/abates1128/elementary-internet-search-strategies-the-information-treasure-hunt

Beal, V. (2020) How do Web Browsers Work? Webopedia Reference Retrieved from https://www.webopedia.com/quick_ref/internet_browsers.asp 21/06/2020

Collins, J. (2018) How to Search Within a Web address: Use the inurl command to quickly find keywords in URls. Retrieved 15/12/18 https://www.lifewire.com/hot-to-search-within-web-address-3482391?print

Collins, J. (2018) The Invisible Web: What It Is, How You Can Find It Retrieved from https://www.lifewire.com/invisible-web-how-to-find-it-3482490

Crossman, A. (2018) Convenience Samples for Research: a Brief Overview of the Sampling Technique https://www.thoughtco.com/convenience-sampling-3026726?print

DeCesare, J. A. (2014) The Expanding Role of Online Video in Teaching, Learning, and Research. *Library Technology Reports* alatechsource.org. February/March

Grauschopf, S. (2019) What's an Internet browser? a simple definition. Internet Browsers: A layman's Guide to how they Work Retrieved from https://www.thebalanceeveryday.com/what-is-internet-browser-892819 21/06/2020

Guccione, D. (2020) What is the dark web? How to access it and what you will find. Available from CSOonline.com/article/3249765/what-is -the- dark- web -how- to- it-what- you-will-find/

Harris, R (2000) Internet Search Tips and Strategies. - Retrieved 1/10/2018 from https://www.virtualsalt.com/howlook.htm

The Internet World Stat (2019) https://www.internetworldstat.com/stats.htm

The Internet World Stat (2020) https://www.internetworldstat.com/stats.htm

Iqbal, M. Z. & Mahmood, A (2011) Factors Related to Low Research Productivity at Higher Education Level *Asian Social Science* Vol. 7, No. 2 Available at www.ccsenet.org/ass

Johnson, L. and Lamb, A (2007) An information skills workout Wikis and collaborative Course Hero available at https://www.coursehero.com

Khiste, G.P, D.K.Veer and D.B. Maske (2011) Role of Search Engines in Library at a Glance *International Journal of Information Dissemination and Technology* vol.1: 2
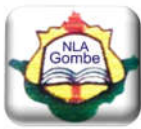
Kunder, M. (2019) the Size of the World Wide Web (the Internet). Retrieved from https://www.worldwidewebsize.com

Li, W. (2018) what is the difference between the dark web and the deep web? http://www.quora.com

Norton, A. (2011) 10 tips for smarter, more efficient Internet searching Retrieved on 1/10/18 from https://www.techrepublic.com/blog/10-things/10-tips-for-smarter-more-efficient-internet-searching/

Pujar, S. (2011) Search Engines retrieved from *Slide share*

Singh, N. (2001) Internet: Importance and Usage for Library and Information Professional *DESIDOC Bulletin of Information Technology* Vol.21no 3

Taherdoost, H. (2016). Sampiling methods in Research Methodology; How to Choose a Sampling Technique for Research. *International Journal of Academic Research in Management (IJARM)* Vol.5:2

Techopedia (2018) Meta Search Engine retrieved from www.techopedia.com

Thompson , C. (2015) Beyond Google: Everything you need to know about the hidden internet http://www.businessinsider.com

Whittemore, R. and Knafl, K. (2005) The integrative review: updated methodology *Journal of Advanced Nursing* 52(5), 546– 553

Wikipedia (2017) Web Search Engine https://en.wikipedia.org/wiki/Web_search_engine

Xie, I. and Joo, S (2010). Transitions in search tactics during the Web-based search process https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21391