Chapter 04 Researching and Using the Web

Shambhavi Roy

Clinton Daniel
*University of South Florida*

Manish Agrawal
*University of South Florida*

Follow this and additional works at: https://digitalcommons.usf.edu/dit_tb_eng

Scholar Commons Citation
https://digitalcommons.usf.edu/dit_tb_eng/6

This Book Chapter is brought to you for free and open access by the The Modernization of Digital Information Technology at Digital Commons @ University of South Florida. It has been accepted for inclusion in *FUNDAMENTALS OF INFORMATION TECHNOLOGY: Textbook – English* by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.
## CHAPTER CONTENTS

**Overview** 38

**Information Sources** 38
- Search Engines 38
- Specialized Review Platforms 41
- Public Wikis/Encyclopedias 42

**Online Training** 43
- Informal Training Sites 44
- Massive Open Online Courses (MOOCs) 44
- Training by Industry-Specific Organizations 45

**Using the Web: Risks, Pitfalls, and Strategies** 45
- Is the Source Reputable? 46
- Check the URL 46
- Look at Multiple Sources 46
- What’s Beyond the Headline? 47
- Is the Author an Expert in the Field? 47
- Check Your Biases 47
- Bounce Your Ideas With Others 47
- What Motivates the Source? 48
- Search Tracking 48

**Chapter Terms and Definitions** 49

**Chapter Case: Christopher’s Google Search** 50
Overview

Never in history have we had so much knowledge stored, organized, and ready to be consumed with the click of a mouse or the tap of a button. If we step back for a moment and think about what is available on the web, we will be amazed by the information close at hand, all thanks to computers and the Internet. You can find the score of any weekend football game, price of grain in Nebraska, weather in Antarctica, tide schedules in South Australia, and the outcome of an election in Britain by typing your question in any search engine. If you are so inclined, you could have someone explain to you the Theory of Relativity, how to lay a brick wall, how to plan a party, or how to edit genes. If you want to learn to play the piano, understand road regulations in Florida, or learn about mountaineering, you can do it all from the comfort of your couch.

IT gives you more power in terms of access to knowledge than kings and emperors of yesteryears ever had. However, as with any powerful tool, you should know how to handle it with care and caution to avoid getting into trouble and not harm yourself or others. In this chapter, we will introduce some of the common sources of information online and evaluate their pros and cons.

Information Sources

Search engines are the most common source of online information. However, depending upon your needs, there are also other information sources including review sites, multimedia sites, and educational sites. We will briefly introduce them here.

Search Engines

Search engines are software that allow users to search for information of interest. Search engines have been with us ever since the Internet became popular and have evolved over time to yield super-accurate results. Google, in particular, as well as Microsoft have built their reputations and fortunes by providing answers to your questions, accurately and reliably. Both Google Search and Microsoft Bing constantly refine their algorithms and can even predict and propose suggestions to autofill your queries. Their search engines are integrated with extensive external databases to answer all sorts of questions in the most helpful way possible.
The basic capability of a search engine is to find documents on the Internet that correspond most closely to the search term entered by the user. For example, if you search for “USF,” search engines locate all pages related to USF, sort the pages by their relevance to the search term, and display the results, with the most relevant results on top. Results for “USF” from two popular search engines—Google and Bing—are shown in Figure 12. We see that the results are not identical, reflecting differences in the algorithms used by the two search engines to process searches.

FIGURE 12 — Search engine results will populate differently depending on the search engine used.
The search results for USF show how search engines make money. In the results from Bing, the first result is labeled as an “Ad,” and encourages users to check out Keiser University. Organizations pay search engines for placement in search results, and this is one of the most profitable businesses in modern times. If a few visitors searching for USF register for courses at Keiser University, the investment in the search ads can be profitable for Keiser.

Search engines are expanding their capabilities to offer a single point of entry for any information-based capability. For example, if you search for the price of a ticket to fly from Tampa to San Francisco, search engines can fetch the prices from different airline databases and show them to you in a comparative list. If you search for sneakers, they will not only bring you information about sneakers but also show you the prices and availability in nearby stores. When you search for the weather in New York, the search engine may not have the information stored directly in its database, but it will query external services such as weather.com on your behalf and present the information. All these capabilities make search engines very powerful as we become more and more dependent on information.

Recent developments in artificial intelligence have enabled powerful voice recognition capabilities in inexpensive consumer devices. See Figure 13. With the popularity of these hands-free devices and increased accuracy of voice recognition, you don't even need to type your question. You can just ask Siri for the local weather, Google for a stock price, and Alexa to order cereals for the family.

**FIGURE 13** — Smart speakers.
Search Engine Evolution

Search engines have evolved through four primary stages: (1) manually curated table of contents (e.g. Yahoo!; starting in 1994); (2) keyword-based indexes of webpages (e.g. Lycos, Excite, Alta Vista; starting in 1996); (3) link-based ranking of webpages (e.g. Google; starting in 1998); and (4) embedded goal-specific search engines (e.g. Amazon, TikTok; starting in 2003). In 1994, Jerry Yang and Dave Filo began manually creating a hierarchical directory of websites to help users find interesting sites on different topics. Yahoo! is credited with giving Amazon its first boost. Three days after Amazon was founded, Jerry Yang emailed Jeff Bezos asking for permission to list Amazon on the “What’s cool” section of Yahoo! Amazon received $12,000 worth of orders in the first week of being listed on Yahoo!

The National Science Foundation started the Digital Library Initiative (DLI) in 1994 to simplify information finding on the nascent Internet. Several projects to index webpages based on keywords emerged from this and related projects and became popular between 1996–1999. Eventually in 1998, one of the DLI projects led to the basic technology used by Google. Instead of ranking webpages based on keywords, Google’s technology relies on the judgment of website authors to link to other websites. Such links are considered reliable indicators of a webpage’s relevance and are used to create the page rank.

In recent years, sites have begun to develop search engines customized to their needs. Social media sites such as TikTok have developed search engines to show media that a user is most likely to be interested in watching next. Shopping sites such as Amazon have developed search engines that show the most profitable products that a user is likely to buy next.

Specialized Review Platforms

General search engines like Google and Bing are very good at pulling up relevant information from the web in response to a query. However, search engines do not generate information on their own. In recent years, several companies have recognized business opportunities in helping users generate content that meets some specialized information needs of other users. Typically, such User Generated Content (UGC) helps other users decide which products and services to buy. Companies that help users add their reviews to products and services are known as specialized review platforms. See Figure 14. Examples include Yelp for restaurants, and TripAdvisor for travel destinations. Specialized review platforms also help businesses. G2 for example, is a popular specialized review platform for business software. Search engines have also begun facilitating such reviews. If you want to know

what others have to say about something, a specialized review platform might be a good starting point. For example, if you want to buy a car, check out a restaurant, research the quality of a business, find a good dentist, get more information about your current medication, or buy a new refrigerator, these review platforms can be of great help.

Specialized review platforms can be categorized into two broad types: those offering expert opinion and those publishing crowd-sourced reviews.

Expert opinion platforms like WebMD, Consumer Reports, Edmunds.com, and Nolo.com hire experts in specific fields (health, household appliances, cars, laws) to write articles and reviews offering in-depth information about products and services in their areas of specialty.

Crowd-sourced review platforms gather information from users and use voting algorithms to identify the most relevant and useful reviews. One of Amazon’s secret recipes to success has been its extensive collection of user reviews on products. Yelp, a crowd-sourced review platform, collects diner feedback on restaurants and can quickly recommend a restaurant you might like. Since these sites aggregate a community’s feedback, you might end up with a greater diversity of opinion than those provided by experts.

As with any recommendation, you must keep an eye out for potential fraud and misdirection. The review platforms may be more favorable towards the products of their business associates. Another problem is fake reviews. As positive ratings lead to increased sales, product manufacturers are tempted to influence reviews. Be aware of the possibility that even Amazon and Yelp reviews might be written by ghost review writers compensated by product manufacturers.

Public Wikis/Encyclopedias

Wikipedia, a great knowledge base on the Internet, is an open encyclopedia that allows users to edit content. Wikipedia is one of the most popular sites on the Internet with over 59 million articles,
and it has a large and passionate community constantly editing and updating articles to keep them up to date and accurate. On the other hand, paid encyclopedias, like Encyclopedia Britannica, are commercial products that curate articles written by subject matter experts. Investigations suggest that there is little observable difference in reliability of information between the two.

Wikipedia’s Origins in St. Petersburg, FL

Wikipedia was founded in St Pete, FL in 2003. The company combined founder James Wales’ love for encyclopedias with the Wiki technology created by Ward Cunningham in Portland in 1995. The wiki technology enabled users to edit pages on their own, which helped Wikipedia to grow and find a niche in the market between Yahoo!’s manual hierarchy and Google’s automated search.

However, Wikipedia has had its own share of issues and controversies. In 2005, the biography of an American journalist, John L. Seigenthaler, Jr., on Wikipedia falsely identified him as a conspirator in the assassinations of John F. Kennedy and Robert F. Kennedy. These claims survived Wikipedia’s community policing for 132 days.

Since Wikipedia is based on the wisdom of the community, the less commonly accessed and reviewed portions of the site may have less accurate information than the more common ones. There is also potential for “mob activity” where a group of biased and motivated individuals influence the tone and content of an article.

Online Training

Information sources are abundant on the Internet, which opens many opportunities for learning. Educational opportunities are found in a variety of formats on the Internet. Online training has become very popular and can fit many learning styles.

Informal Training Sites

In recent years, popular media and entertainment sites like YouTube have also become useful sources of information, particularly for tutorials on specific topics. Whether you want to troubleshoot your dishwasher, learn how to program in python, or prepare for a job interview, you are pretty much guaranteed to find a suite of videos tailored to your specific need. As the videos are rated by the community (with likes, dislikes, and subscribes), the YouTube algorithm combines the search capabilities of a search engine with an understanding of the viewing habits of similar users. This enables YouTube to constantly tune its suggestions for videos most likely to solve your problem. You can also interact with the content provider by commenting on the videos and contributing further to the content.

Apart from the benefits of curations and ratings by the community, videos have a distinct advantage as a learning medium. You can rewind and watch the difficult parts multiple times, increase the speed if you are already familiar with the content, skip portions you understand, and watch content at your own convenience on a device of your choice.

However, there are several caveats to using YouTube as a learning medium:

- identifying the right training yourself,
- having the time and discipline to see the training through, and
- not receiving any certification after the training.

Massive Open Online Courses (MOOCs)

Massive Open Online Courses (MOOCs) are online educational courses open to anybody in the world. These educational platforms are provided by individuals, organizations, and firms. Some MOOCs, like Khan Academy, focus on free courses and some, like Coursera, offer a mix of free and paid courses. Some MOOCs like EdX even offer proctored examinations and certificates of completion.

Many leading universities are experimenting with MOOCs to bring the expertise of their faculty to students around the world. MIT offers the vast majority of their class materials for free on MIT Open Courseware. Several leading universities including MIT and Harvard are collaborating on a shared learning platform called EdX to offer free and low-cost courses. These universities are also exploring

---


how the EdX MOOC platform can be used to offer students low-cost credentials such as micro-bachelors and micro-masters.

Most MOOC classes tend to be structured and take you on a step-by-step journey from novice to expert. Whether you are interested in researching, furthering your career, or just learning, MOOCs are a great way to go and have, not surprisingly, exploded in popularity in recent years.

Training by Industry-Specific Organizations

Once you enter your chosen professions, your industry or vendor-specific platform will usually offer several training options. One such training that is currently quite popular is offered by Amazon for their Amazon Web Services (AWS) technologies. These trainings are available for free and are very popular among folks eager to become AWS experts. Similarly, other leading technology companies including SAP (for large business operations), Oracle (for Data management), or Cisco (for Network Administration), offer extensive training globally both online and in person. There are also certification programs offered by industry-specific organizations for general technology expertise such as Scrum Master (for product development), Program Manager, and Business Analyst, and these can be extremely helpful in acquiring the relevant skills and finding jobs in these roles.

Using the Web: Risks, Pitfalls, and Strategies

Although the web makes it easy to find information, not every bit of data obtained online is equally reliable. Before you act on information received online, you should be aware of the deceptions and traps. Here are some things to keep in mind when evaluating information received from online sources.

Is the Source Reputable?

Print publications with patterns of incorrect information usually do not survive for long. Therefore, print publications with wide circulation (e.g. popular magazines and newspapers) generally put in considerable effort to ensure correctness of information. This is part of their gatekeeping role. Gatekeeping is the process that publications follow to select and present information to their readers. Mechanisms that increase reliability of information include experienced editors and reliable sources. When you read something in a print publication, you can be reasonably confident that the publisher put in their best effort to ensure that the information is correct.

However, many people who post information online are not trained to verify information before publishing it. Therefore, online information has a greater likelihood of being incorrect than print information. It also does not cost money to post information on blogs, social media, and other online platforms, which leads to vast amounts of information getting posted. Online information can reach wide audiences through search engines and social media even if incorrect. Social media platforms such as Twitter are widely considered the “Digital Town Square.” What should a reader do to get reliable information from online sources?

One mechanism is reputation. As every high schooler knows, reputations take a long time to earn, and no time to lose. A reputation for correctness is therefore an indicator that the online source has mechanisms in place to validate information before publishing it. Information posted by reputable sources is more likely to be correct. Although it is hard to judge the reputability of an online source, you can take some precautions. Look up the business name of a source on search engines to check reviews in other places on the web. On social media, you can look up a user’s list of followers. Users followed by other reputable users are likely to be reputable themselves.

Check the URL

Unethical businesses often create look-alike websites to confuse people and grab user credentials and sales. This is called website spoofing. Even if you think you are getting information from university websites, banks, or other well-known private institutions, you should double check the URL to make sure you are where you think you are.

Look at Multiple Sources

Checking out information on several websites will help you look at the problem/solution from different perspectives. One source might focus on the ease of use of a device while another may


32 Twitter has a mechanism to verify user accounts, indicated by 🏅. You can read about the verification procedure at https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts (accessed June 2023).

33 There are many variants of spoofing. You can look these up online including at this article: Sagar Joshi, “What is spoofing? How to Protect Yourself Against It?” https://www.g2.com/articles/spoofing (accessed June 2023).
bring attention to its price. If you get the same information from multiple reputable sources, it is a good indication of its authenticity. For example, if several car review websites agree that the model you want to buy is safe, reliable, and a good value for your money, then you can feel assured about having made the right choice.

What’s Beyond the Headline?

As users are reluctant to pay subscription fees for online publications, there is increasing pressure on websites to earn revenues by drawing traffic and showing ads to visitors. A popular way to do this is by using attention-grabbing headlines that may have little to do with the content of the article. These headlines serve as clickbait to draw readers to the site, not to summarize the information on the page or explain nuances. Therefore, when consuming online information, do not rely on the headlines alone and make it a point to read all the available information in the article, not just the first few sentences. It is common to see subjective conclusions at the top with caveats and drawbacks buried near the end of the article.

Is the Author an Expert in the Field?

Online ad revenues have created a business model of influencers. Influencers are people who are able to encourage potential buyers of a product or service by recommending items online, usually on social media. Influencers make money by ensuring their content is visible and by having a lot of followers. When you search for product recommendations, you are likely to come across content that is popular or promoted by influencers, but not necessarily accurate or relevant to you. What works for the influencer’s specific circumstances may not work for you. It is therefore useful to verify the expertise of the person handing out advice. For example, if you want information about a type of diet, a nutritionist is more likely to have studied the impact of the diet under a variety of conditions compared to a popular sports star who may have followed the diet under strict supervision of a team of experts.

Check Your Biases

The ease of finding information we like makes the Internet great at confirming our biases and dragging us into an echo chamber that magnifies our beliefs. As the World Wide Web has dramatically increased our ability to connect with others, even the most outlandish ideas and conspiracy theories can have large groups of enthusiastic subscribers, egging each other on.

This is an issue particularly with subjective opinions. There is a reason why we maintain our biases: either we aren’t aware of their existence or wholeheartedly believe in their authenticity. Before reaching out to the web to validate your opinions, you should take a moment to be honest with yourself and understand if you are actually willing to change your initial opinion. Are you looking for new information just to prove your point? Are you only seeking out people who’ll agree with you? Can you argue the issue from the perspective of those holding the opposing view?

Bounce Your Ideas With Others

It is always a good idea to discuss what you have found on the web with friends, family, and teachers. The power of community to sift out good ideas from bad ones can never be underestimated.
What Motivates the Source?

Web sites and search engines increasingly make money by encouraging you to buy things. They are therefore motivated to prioritize displaying information for which they get paid. When you search for a particular laptop for example, the first few search results are likely to be advertisements by competing merchants and not necessarily the best/cheapest places to buy. The burden is on you to check the results and pick the one that makes the most sense.

Similarly, if you search for insurance agents on a search engine, the first results are mainly advertisements from various insurance companies. When you click on any of the links, the advertiser pays Google a “click-through” fee. Google stands to gain by showing advertisers who pay the most, not necessarily those with the best products and services.

Therefore, it is useful to be on the lookout to understand if your source of information makes money by influencing your decision. Reputable sites clearly separate advertisements from other content, but others tend to blur the difference.

Search Tracking

Another pitfall of searching for things online is search tracking. Most search engines tend to hold on to queries and use them to gauge your potential needs, often showing targeted advertisements on other sites you visit in the future. Your information might be shared or sold to other Internet merchants, who will likely be storing the information in databases around the world. You might be surprised with related ads long after you have abandoned the primary search engine. We will cover more about this in Chapter 17 on “Ethical Issues in Information Technology.”
## Chapter Terms and Definitions

<table>
<thead>
<tr>
<th><strong>Algorithm:</strong></th>
<th>A systematic and logical sequence of steps designed to solve a problem</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Influencers:</strong></td>
<td>Individuals who have a large audience through a social media platform</td>
</tr>
<tr>
<td><strong>Massive Open Online Courses (MOOC):</strong></td>
<td>Open access educational courses available to participants using an online platform</td>
</tr>
<tr>
<td><strong>Scrum Master:</strong></td>
<td>In the Scrum approach to agile project management, the team member whose responsibility is to effectively get the team closer to the goals and keep everyone on track</td>
</tr>
<tr>
<td><strong>Search Engine:</strong></td>
<td>An algorithm designed to find resources related to what is input into a search interface</td>
</tr>
<tr>
<td><strong>Social Media:</strong></td>
<td>A digital platform for interaction between people</td>
</tr>
<tr>
<td><strong>Spoofing:</strong></td>
<td>Disguising the true or trusted identity of a person or device</td>
</tr>
<tr>
<td><strong>Uniform Resource Locator (URL):</strong></td>
<td>The website/address of a resource on the Internet.</td>
</tr>
<tr>
<td><strong>User Generated Content (UGC):</strong></td>
<td>An individual's content creation on platforms; platforms include Instagram, Twitter or YouTube</td>
</tr>
<tr>
<td><strong>Wiki:</strong></td>
<td>User generated content on a digital platform that facilitates collaboration</td>
</tr>
</tbody>
</table>
Chapter Case

Christopher’s Google Search

Christopher stepped off the school bus and started to walk home as he thought about the homework assignment his teacher had assigned him earlier that day. His teacher asked him to use a search engine to find information about the Florida state bird. He was to research the topic and think about how the search engine works.

Christopher sat down at his home computer, opened up an Internet browser, and navigated to google.com. While on the Google search engine page, he typed the search terms, “Florida state bird”. The top result displayed the following:

“Wow!”, Christopher yelled. “144,000,000 results in .84 seconds!”. “It looks like the Northern mockingbird is the Florida state bird”. Christopher thought to himself, “How did Google know to put this reference to the Florida state bird in front of 144,000,000 other results?”. To understand how the Google search engine worked to display the results, Christopher looked over Google's Documentation website:

https://developers.google.com/search/docs/fundamentals/how-search-works

After he reviewed the information on the website, he realized that the Google search engine works in three stages, and not all pages make it through each stage. The stages include crawling, indexing, and then finally serving the search results.

Question 1: Research the terms “crawling” and “indexing” as they relate to the topic “search engine.” Based on the results of your research, how do these terms help search engines work?

Question 2: The results of a search using Google can vary depending on how relevant the information is to the search engine user. According to Google, relevancy can be determined by many factors to include information such as the user’s location, language, and device (desktop or cell phone). Why do you think Google considers this type of relevant information when displaying the results of a search?