Why Is This First? Understanding and Analyzing Internet Search Results

Chris D. Ham

University of North Texas

Primarily due to their convenience, online search engines such as Google and Bing are becoming a central location for obtaining information. As a result, societies give search engines tremendous control over the spread of information to the public. Through a highschool-level sample lesson plan, the article was written to promote dialogue with teachers on the importance of teaching the intricacies of search engines. The sample lesson plan begins with fundamental knowledge on the functionality of search engines with emphasis on algorithms. With this instruction, students can understand not only search engines, but also their manipulation potential, which leads to ramifications. Using the manipulation potential as a catalyst, many societal concerns are raised, such as spread of misinformation, oppression of certain groups, and impact on behavior. Through this instruction and dialogue, practitioners will have a resource to integrate search engines into their curriculum in response to this new concern.

Keywords: search engine, algorithms, databases, indexing, crawling, PageRank, search engine optimization, metrics, teacher, secondary education, mathematics

Introduction

With the massive amounts of information contained in the Internet, algorithms are an essential tool to organize and process online data. However, users can manipulate algorithms to spread or contain information. For example, the use of bots by Russians in social media platforms was used to spread misleading information about political candidates, thus endangering the democratic process of the 2016 U.S. Presidential election (Bessi & Ferrara, 2016). Another example is that the spread of misinformation through antivaccine websites in Google search results in various countries has resulted in public health risks (Arif et al., 2018). These are just two instances of how algorithms are given tremendous control over the flow of information in an online medium. With the growing youth population (Colby & Ortman, 2015) and their high usage of search engines (Purcell, Brenner, & Rainie, 2012), it is critical for educators to teach students about the intricacies of search algorithms, including their potential to be manipulated and the subsequent ramifications. In detail, students need to have an understanding of the societal impact of algorithms, including their ability to spread or censor information. In response, a cross-curricular lesson plan was designed towards high school students to inform and engage students on the fundamentals of search engine algorithms along with its societal impacts.

Understanding Google's Search Algorithm

Searching information on Google is often the starting point for users to find curated information. Consequently, the search results displayed on Google play a critical role in the flow of public information. Alphabet Inc., the parent company of Google, is a business that profits from behavioral

I would like to express my great appreciation to Dr. Dan Krutka for his valuable and constructive suggestions during the development of this work.

Please address queries to: Chris D. Ham, University of North Texas. Email: hamchris815@gmail.com

data where it uses computational analysis to predict human behavior (Zuboff, 2019). These predictions are used within Alphabet and sold to other companies often herding human behavior, leading to behavioral modification that Zuboff (2019) calls *surveillance capitalism*. As part of search engine optimization (SEO), Google allows companies' websites appear earlier in search results through Google AdWords and sponsorships. This results in two separate search outcomes: sponsored and organic results (Berman & Katona, 2013). Sponsored results stem from the business partnership between the marketing company and Google. These results are clearly labeled as sponsored results for consumers. However, organic results can also be manipulated by taking advantage of Google's search algorithm. Because there is no business partnership, manipulated organic results are curated with all other nonsponsored results. This can be concerning because consumers are not aware of this manipulation.

Even though Google does not publicly disclose its entire ever-changing search algorithm, there are numerous key fundamental components within the algorithm that can be optimized for companies to maximize their marketing potential. For example, one of the most significant factors within the Google search algorithm is PageRank (Yang, Huang, & Luh, 2016), which is a metric designed to calculate and compare the importance of websites. Marketers can study the intricacies of PageRank and implement a framework that will maximize their PageRank scores through manipulation (Qin, Zhuo, Tan, Xie, & Ye, 2018). With higher PageRank scores, websites are organically displayed earlier in search results, which can increase its exposure.

Beyond marketing strategies, there are societal issues to be discussed. In conjunction with PageRanks, research has shown that Google's search algorithm personalizes search results for individual Google users based on those users' interests (Simpson, 2012). This introduces issues where users become ignorant of information that Google found to be less tailored to their previous searches. Consequently, users will further develop their own personalized search profiles. This not only reinforces and narrows users' buying tendencies, but can also narrow and solidify their political and social outlook due to this tailoring of information exposure. Furthermore, Google has a feature called autocomplete where users type keywords into Google search, and it instantly suggests additional keywords to complete the search. For example, in 2013, searching for "women should" first suggests "women should [stay at home]," then "women should [be slaves]" and "women should [be in the kitchen]" (Ogilvy & Dubai, 2013). The brackets denote Google's autocomplete suggestions. With this feature, a gender stereotype is pushed, which can lead to digital oppression. Due to criticism that autocomplete suggested problematic keywords. Google has made changes to the autocomplete algorithm where the results can no longer be replicated. Nevertheless, Google's autocomplete has the potential to attack and damage certain groups of people due to users taking advantage of Google's autocomplete algorithm (Miller & Record, 2016).

According to the Pew Research Center, 66% of search engine users believe that their search results are a fair and unbiased source of information (Purcell et al., 2012). As briefly demonstrated already, search results are not fair and unbiased. In 2016, Oxford Dictionary named *posttruth* the word of the year, stemming from political influences based on emotions and beliefs, instead of facts ("Word of the year 2016 is...," 2016). This illustrates a disconnect where users believe that their sources from search engines are not biased when they are being influenced with biased information grounded on emotion and beliefs. Influence in thought and behavior from search engine results are one of many concerns that students need to be educated in.

Students have learned about algorithms in various ways in computer science classroom settings (Barendsen et al., 2015; Giordano & Maiorana, 2015; Grivokostopoulou, Perikos, & Hatzilygeroudis, 2016). Furthermore, societal issues in a digital medium have been taught, typically in the form of digital citizenship and social media (Gleason & Von Gillern, 2018; Goh & Sun, 2015). However, there are few resources for teachers who seek to improve students' understanding of search engine

algorithm and the associated personal and societal impact. In the following lesson plan, access to lesson objectives, description of lesson interactive activities, and discussion prompts are provided, which can be used to teach high school students about search engine algorithms and the societal implications.

Teaching Method

There are three key elements that can help students understand search engines: (a) search algorithm basics, (b) understanding algorithms, and (c) societal impact. The lesson is intended for general education high school classrooms. It is ideal for social studies, humanities, mathematics, and computer science courses. But elements of this lesson plan can be adopted to other subjects. The implementation and depth of content are open to the teacher, depending on the classroom setting and teacher's knowledge. Digital downloads of all materials can be requested from the author.

Search Algorithm Basics

Learning Objective 1: Students will understand the fundamental design and purpose of search algorithms.

Learning Objective 2: Students will understand the necessity of search algorithms in terms of Internet data.

Educators can begin the lesson by asking the students, what search engines do you use to find information? Alternatively, the students could be asked, how do you find information online? Students are likely to produce answers such as Google, Bing, Yahoo, and DuckDuckGo. Teachers can then further question where the students make connection between search algorithm and search engines. For example, students could be asked, "How do you think search engines organize and gather information on the Internet?" Lastly, students can be asked to guess the total data size of the Internet. After revealing the estimated answer found in slide 6, students should be directed to make the connection between the massive amount of data contained in the Internet and the need to organize all of the data through search algorithms. A sample presentation is shown in Figure 1.

Understanding Algorithms

Learning Objective 1: Students will understand the process of crawling and indexing webpages.

Learning Objective 2: Students will understand multiple variables involved in search algorithms.

Learning Objective 3: Students will understand the calculations involved in PageRank.

Due to the massive amounts of information found on the Internet, search engines use a computational tool called crawling and indexing to find and organize as much of the Internet as possible. To demonstrate the process of crawling and indexing, students should be instructed to go to a sample website (www.sites.google.com/view/coffee-website/). The site contains a general homepage with six sample pages. Students will crawl and index the sample website to understand the process. Teachers are encouraged to use other websites, such as the school's website, to create an index. To check for accuracy, teachers can type "site:www.yoursamplewebsite.com" into Google search to display all of the index created by Google's algorithm. Figure 2 demonstrates sample questions and definitions of crawling and indexing.



Figure 1. Sample Presentation for Understanding the Basics of Search Algorithms



Figure 2. Sample Activity Teaching the Process of Crawling and Indexing Through a Sample Website

The next portion of the lesson is designed to raise concerns on the curation process of Google search results. Students are given a sample scenario where a Google search user types in "benefits of coffee" into Google search. Teachers will prepare ahead of time by printing and cutting out the top 10 results. The individual results are shuffled and placed in an envelope for the students. As the students open up the envelope, the students are asked to place the 10 results in order. The students will make an educated guess using their previous experiences with search engines. To further guide the students, teachers can point out date of publication, creditability of sources, relevance to the keywords, and domain endings such as .com, .org, and .edu. Once all of the students have made their educated guess, the teacher will reveal the order produced by Google's search algorithm. Students can discuss how their answers compare to the results generated by Google. Students should be encouraged to question or decipher how the algorithm made its decision. It is very likely for students to predict incorrectly due to the bias created through Google's algorithm. For example, students often expect the article from *Harvard Health* to be first, but Google's algorithm placed it second. The reason for this sequence is further discussed in the next portion of the lesson. In this portion of the lesson, students should be encouraged to predict the reason and method for Google's search results. In addition, students should be encouraged to predict any potential ramifications. These predictions will be reviewed in the next portion of the lesson. Figure 3 includes guiding questions, Google's search results, and introduction to the complexity of search algorithms.

With student predictions of algorithm methods and its potential ramifications, the lesson begins to investigate the intricacies of search algorithms. The lesson continues into the calculations involved in PageRank. This portion of the lesson has the most flexibility in terms of depth of content. Students with knowledge of algebra, including sigma notation and matrices, can go through the calculations involved in PageRanks. Students with knowledge in computer science, including coding, can write series of codes to process and organize sample data. For courses where the calculation of PageRanks is not relevant or out of scope, there are two alternative options. Students can conceptually understand PageRanks without any calculations. Based on Figure 4, each arrow given can be understood as a vote. As one website gives more votes to other websites, the value of all the votes from that website goes down. For example, the three arrows given from Page A has less value per arrow than the single arrow given from Page C. Ultimately, the website that receives the most combined value from the votes is the most popular and valuable. As a result, it receives the highest PageRank score. Alternatively, PageRanks can simply be instructed as a tool made by Google that is used to determine the most valuable website. As the value of the website goes up, the earlier it will show up on a research result, which will increase its exposure.

The sample lesson explores the mathematical calculations of PageRanks by solving a problem. The problem consists of four different webpages with various links to other sites. Furthermore, the problem is solved only using a simplified algorithm, without any coding. It does not include calculations on damping factor, computation methods, and implementation methods, which would reflect a truer PageRank algorithm (Langville & Meyer, 2011). In slides 16–21, the PageRank calculations are outlined step by step. For computer science courses, teachers are encouraged to pursue these concepts more deeply through coding.



Figure 3. Sample Activity Where Students Implement Their Own Search Algorithm With Sample Index



Figure 4. Example Problem Used to Demonstrate PageRank

Once students understand the methodology of PageRanks, students now understand one of the biggest contributing factors in search algorithms. With this knowledge, teachers can refer back to the previous portion of the lesson to apply the concept of PageRanks to the search results on the benefits of coffee. The students should be able to predict that the first article in the search results most likely had the highest PageRank score. This means that many websites referred to the first article. With this concept, students should be guided to conclude that websites with the highest PageRank score will appear first in the search results, regardless of the creditability or quality of the content. Generally, creditable websites are commonly referred, which results in a high PageRank score. However, as demonstrated in the example, this is not always the case. Teachers should continue to promote discussion on the potential ramification of this concern. The ramifications will be fully explored in the last section of the lesson.

PageRanks are considered to be one of the most significant contributing factors to Google's search algorithm. However, there are other predicted contributing factors involved in Google's search algorithm (Peters, 2015b). The next portion of the lesson found in slides 22–27 is designed to briefly further explore other contributing factors in Google's search algorithm to illustrate additional methods of manipulation. For example, agnostic factors, such as website security and load speed, play a minor role in Google's algorithm. With this factor, marketers are encouraged to keep their website safe and fast. A brief description of the factors is shown in Figure 5. The complete analysis including the subcategories of each ranking factor can be found on Peters's (2015a) study. Teachers are encouraged to use these factors to promote additional dialogue on the manipulability of search algorithms with its ramification. Exploring the details of each factors are not necessary. The next section of the lesson will gather all of the information, predictions, and concerns discussed to explore the societal ramification through examples.



Figure 5. A Brief Description of Each Ranking Factor From Sample Lesson. Adopted from "Search Engine Ranking Factors 2015" by M. Peters, 2015. Copyright 2015 by Moz.

Societal Impact

The last portion of the lesson is designed to encourage students to connect their new knowledge of search engine mechanics to an understanding of the larger social ramifications within search algorithms through classroom discussion.

Learning Objective 1: Students will apply their knowledge on search algorithms with its societal impacts, including, but not limited to, SEO, autocomplete, and spread of misinformation or biased information.

The lesson contains three components to illustrate societal impact: SEOs, autocomplete, and biased information. Each component demonstrates a unique impact. This lesson uses SEOs to demonstrate how search results can be modified to reach a target audience. Autocomplete is used to demonstrate how women can be oppressed through projection of stereotypes. Lastly, biased information is used to portray how search results can create a false reality, which can influence behavior. Discussion of all three components are recommend but not essential. Teachers can choose to exclude components to meet the needs of the students. The components are driven by class discussion where students are encouraged to generate issues and solutions. Alternatively, students can be engaged through a class debate, group investigation, inquiry, or a research activity.

Slides 29–36 of the sample lesson explain a variety of background information on SEOs. These slides explain how companies will sponsor their content with search engines, such as AdWords, to increase visibility. Furthermore, they explain how even organic results can be modified to be ranked higher by adapting to the search algorithm. PageRanks can be revisited to understand one significant method to increase webpage ranking. Lastly, the slides explain how search engines will change the search algorithm and sponsored content depending on the location of the user. It demonstrates how users from the United States and the United Kingdom receive different sponsored and organic results. All three aspects of the algorithms raise concerns about the trustworthiness and ethics of search content. With the goal of personalization, users can see different results just from location data. Teachers can include additional variables that are involved in personalization such as user history and interests (Simpson, 2012). This further raise concerns about personal bias and personalized search results. The two questions at the end of the first section in slides 35–36 should be asked to encourage students to have a discussion on the bias found within search results. Figure 6 provides screenshots of the sample lesson to visually demonstrate sponsored, organic, and location results.



Figure 6. Examples of Sponsored, Organic, and Location Results

In the second section, students investigate the ramification of autocomplete. With autocomplete, the search engine attempts to predict or suggest the user's search query. For example, when a user types in "when is," Google may suggest "when is Memorial Day?" If the prediction is correct, typing is minimized by the user. However, as shown in Figure 7, the autocomplete feature can make suggestions that press on female stereotypes, which can lead to oppression of women. Even though Google claims the autocomplete algorithm has been changed to no longer make these suggestions ("Autocomplete policies," 2019), Figure 8 demonstrates how autocomplete can still make questionable suggestions even after the changes made to the autocomplete algorithm. Students should be encouraged to have a discussion on the benefits and drawbacks of the autocomplete feature. In addition, students should discuss how and why autocomplete can lead to societal issues.



Figure 7. Example From the Sample Lesson Demonstrating Sexism. Adapted from "U.N. Women Ad Series Reveals Widespread Sexism" by M. Ogilvy and M. Dubai, 2013. Copyright by 2013 by U.N. Women.



Figure 8. Sample Search Demonstrating Autocomplete Feature Within Google and Bing. Search conducted on March 2019 based on the United Kingdom.

Finally, ramification of biased information is demonstrated through a horrific example. In 2015, Dylann Roof murdered nine African Americans at a church in Charleston, South Carolina. Based on his manifesto, the tailoring of search results based on Dylann's past interest in White supremacist websites may have helped further reaffirm his racist ideology (Noble, 2018). When Google's search engine directed Roof to White supremacist websites, instead of trustworthy government data, it potentially confirmed racist views Roof held instead of disrupting his misunderstandings. This example can raise serious concerns on the spread of biased information involving hate or explicit content. It illustrates how biased information from search results can influence human behavior. Slide 47 of the sample lesson contains an insightful excerpt from Dylann Roof's manifesto that illustrates how Roof was affected by search results. With this, students should have discussions on potential impact from biased information. Guided discussion questions are found on slides 48 and 49. At this point, students should be able to connect the potential manipulation opportunities to appear first in search engines with the ramifications like with Roof. This discussion serves as a conclusive opportunity to spiral and connect all the learning objectives.

As the lesson concludes, the students are expected to have an understanding of the general functionality of search algorithms and its affordances and constraints on our society. Students should be encouraged to reflect on what they have learned. Furthermore, the overall concept of the lesson is to understand that search results are not free from bias. At a minimum, consumers need to be aware of this information, especially on its ramifications. As a result, the teacher can conclude the lesson with a call to action. What actions can students take to inform the public and make a change? Students should be encouraged to educate others, spread awareness, conduct research, or reach out to corporations and representatives to ensure our search engines provide valuable information to users.

Conclusion

Teachers play a critical role in educating students to prepare for their future. Due to their high usage, search engines play an integral role in modern society, especially for students. Students need to be educated on the functionalities, benefits, and drawbacks of search engines to prepare for their future. There are opportunities for search engines to be integrated into high school curriculum. The functionalities integrate with mathematics and computer science courses, especially with search engine algorithms. The benefits and drawbacks integrate with social studies or humanities courses. The lesson plan outlined in this study serves as a model for curriculum implementation.

References

- Arif, N., Al-Jefri, M., Bizzi, I. H., Perano, G. B., Goldman, M., Haq, I., ... Ghezzi, P. (2018). Fake news or weak science? Visibility and characterization of antivaccine webpages returned by Google in different languages and countries. *Frontiers in Immunology*, 9, 1215. doi:10.3389/fimmu.2018.01215
- Autocomplete policies. (2019). Retrieved from https://support.google.com/websearch/answer/7368877?hl=en
- Barendsen, E., Mannila, L., Demo, B., Grgurina, N., Izu, C., Mirolo, C., ... Stupurienė, G. (2015). Concepts in K-9 computer science education. In V. Dagienė (Ed.), *Proceedings of the 2015 ITiCSE on Working Group Reports* (pp. 85–116). New York, NY: ACM.
- Berman, R., & Katona, Z. (2013). The role of search engine optimization in search marketing. Marketing Science, 32, 644–651. doi:10.1287/mksc.2013.0783
- Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday*, 21, 11. doi:10.5210/fm.v21i11.7090

- Colby, S. L., & Ortman, J. M. (2015). Projections of the size and composition of the U.S. population: 2014 to 2060: Population estimates and projections (Current Population Reports P25-1143). Washington, DC: U.S. Census Bureau.
- Giordano, D., & Maiorana, F. (2015). Teaching algorithms: Visual language vs flowchart vs textual language. In T. Rüütmann & M. Auer (Eds.), *Proceedings of the 2015 IEEE Global Engineering Education Conference* (EDUCON; pp. 499–504). Tallinn, Estonia: IEEE. doi:10.1109/EDUCON.2015.7096016
- Gleason, B., & Von Gillern, S. (2018). Digital citizenship with social media: Participatory practices of teaching and learning in secondary education. Journal of Educational Technology & Society, 21, 200–212.
- Goh, T. T., & Sun, P.-C. (2015). Teaching social media analytics: An assessment based on natural disaster postings. *Journal of Information Systems Education*, 26, 27–36.
- Grivokostopoulou, F., Perikos, I., & Hatzilygeroudis, I. (2016). An educational game for teaching search algorithms. In B. McLaren (Ed.), *Proceedings of the 8th International Conference on Computer Supported Education* (pp. 129–136). Rome, Italy: SCITEPRESS.
- Langville, A. N., & Meyer, C. D. (2011). *Google's PageRank and beyond: The science of search engine* rankings. Princeton, NJ: Princeton University Press.
- Miller, B., & Record, I. (2016). Responsible epistemic technologies: A social-epistemological analysis of autocompleted web search. New Media & Society, 19, 1945–1963. doi:10.1177/1461444816644805
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York, NY: New York University Press.
- Ogilvy, M., & Dubai, M. (2013). U.N. Women ad series reveals widespread sexism. Retrieved from http://www.unwomen.org/en/news/stories/2013/10/women-should-ads
- Peters, M. (2015a). Search engine ranking factors 2015: Correlations. Retrieved from https://moz.com/search-ranking-factors/correlations
- Peters, M. (2015b). Search engine ranking factors 2015. Retrieved from https://moz.com/search-ranking-factors
- Purcell, K., Brenner, J., & Rainie, L. (2012). Search engine use 2012: Pew Research Center Research Report. Retrieved from https://www.pewinternet.org/2012/03/09/search-engine-use-2012/
- Qin, Z., Zhuo, C., Tan, W., Xie, J., & Ye, J. (2018). Large-scale targeted marketing by supervised PageRank with seeds. In P. Perner (Ed.), *Machine learning and data mining in pattern* recognition (pp. 409–424). Cham, Switzerland: Springer International.
- Simpson, T. W. (2012). Evaluating Google as an epistemic tool. Metaphilosophy, 43, 426-445.
- Word of the year 2016 is... (2016). Retrieved from https://en.oxforddictionaries.com/word-of-theyear/word-of-the-year-2016
- Yang, S.-A., Huang, T.-L. D., & Luh, C.-J. (2016). Estimating Google's search engine ranking function from a search engine optimization perspective. Online Information Review, 40, 239– 255. doi:10.1108/OIR-04-2015-0112
- Zuboff, S. (2019). The age of surveillance capitalism: The fight for the future at the new frontier of power. New York, NY: Hachette Book Group.

The *Journal of Educational Research and Practice* provides a forum for studies and dialogue that allows readers to better develop social change in the field of education and learning. Journal content may focus on educational issues of all ages and in all settings. It also presents peer-reviewed commentaries, book reviews, interviews of prominent individuals, and additional content. The objectives: We publish research and related content that examines current relevant educational issues and processes aimed at presenting readers with knowledge and showing how that knowledge can be used to impact social change in educational or learning environments. Additional content provides an opportunity for scholarly and professional dialogue regarding that content's usefulness in expanding the body of scholarly knowledge and increasing readers' effectiveness as educators. The journal also focuses on facilitating the activities of both researcher-practitioners and practitioner-researchers, providing optimal opportunities for interdisciplinary and collaborative thought through blogging and other communications. <u>Walden University Publishing</u>: http://www.publishing.waldenu.edu