

# Search Engine bias

## 188.484 - Information Search on the Internet

Simão Jorge da Silva Costa<sup>[12202234]</sup>

Technische Universität Wien, Vienna, Austria

**Abstract.** In this paper I try to get a better understanding how a search engine works, and I try and look into detail to find out if its algorithm is vulnerable to different types of biases, mostly related to user activity and information. I then test some cases in various search engines to test my theories and to check how different they are in the process of searching for different types of information.

**Keywords:** Search Engine bias · Search Engine · Search bias · Engine bias · bias

## 1 Introduction

In today's interconnected world, search engines have become indispensable tools for accessing information and navigating the vast landscape of the internet. However, beneath the surface of these seemingly objective platforms lies a concept that is increasingly garnering attention and concern: search engine bias. The understanding of search engine bias goes beyond its mere definition, delving into the intricate challenges and implications associated with the biases inherent in search results. This issue holds significant weight, as it directly influences the information we consume and shapes our access to diverse perspectives, ultimately impacting our understanding of the world around us.

The rise of search engines has revolutionized the way we seek and consume information. With a few simple keywords, we can instantly retrieve an array of web pages, articles, images, videos, and more, tailored to our queries. However, what may appear as an objective and comprehensive compilation of information is not immune to underlying biases that can subtly manipulate the content and order of search results.

## 2 Search Engine

A search engine is an online tool or software application that allows users to search for information on the internet. It uses algorithms to index and organize vast amounts of web content, such as websites, documents, images, videos, and more. Users can enter keywords or phrases related to their desired information,

and the search engine retrieves relevant results from its database. The results are typically presented as a list of web pages or resources that match the search query, ranked based on their perceived relevance.

## 2.1 Definition of Search Engine bias

Search engine bias refers to the inherent favoritism or partiality in search engine algorithms that may affect the ranking, presentation, and visibility of search results. While search engines strive to deliver the most relevant and accurate information, several factors can introduce bias into the search experience, shaping users' perception of reality and influencing their decisions.

Different search engines can exhibit various types of biases, which can influence the search results presented to users.

## 2.2 Types of Search Engine bias

The understanding of different types of search engine bias sheds light on the challenges and implications associated with biased search results, given its complexity and multi-faceted nature. This issue has a significant impact on the information landscape and users' access to diverse perspectives.

1. **Personalization Bias:** Personalization bias arises from search engines tailoring search results based on individual user behavior, preferences, and demographic information. While personalization can enhance user experience by delivering relevant content, it can also create filter bubbles, where users are exposed to a limited range of perspectives that align with their existing beliefs. This bias can restrict the discovery of alternative viewpoints and impede critical thinking.
2. **Algorithmic Bias:** Algorithmic bias refers to the unintentional biases that may be embedded within search engine algorithms. Algorithms are designed to process and rank information based on a variety of factors, such as relevance, popularity, and user signals. However, biases can emerge from the training data used to develop algorithms, leading to uneven representation or unequal treatment of certain groups or topics.
3. **Political Bias:** Political bias in search engines refers to the tendency to favor or prioritize certain political ideologies or viewpoints in search results. This bias can sway public opinion, shape political discourse, and potentially undermine the democratic process. Examples of political bias range from the promotion of specific news sources to the suppression or de-ranking of opposing viewpoints.

4. **Commercial Bias:** Commercial bias emerges from the influence of advertising and paid results on search engine rankings. Search engines often display paid advertisements alongside organic search results, and these paid listings can sometimes overshadow more relevant or authoritative content. This bias can impact consumer choices, as businesses with higher advertising budgets may dominate search results, potentially distorting market dynamics.
5. **Socio-Cultural Bias:** Socio-cultural bias reflects the inclination of search engines to reinforce existing social stereotypes and inequalities. Search algorithms may inadvertently prioritize or marginalize certain demographics or underrepresented groups, perpetuating biases and limiting equitable access to information. This bias can further entrench societal inequalities and hinder efforts towards inclusivity and diversity.

Understanding these types of search engine bias is crucial for recognizing the potential impact on information retrieval, knowledge acquisition, and societal discourse.

Despite various types of biases can occur, I will put my focus on Personalization and Algorithmic Bias, as they are more general than the rest and it is easier to gather information. Also, the other types of biases can also be aggregated in the ones I selected: *eg.* Cultural and Political bias can be influenced by user location, due to various factors as:

- **Cultural and Societal Influences:** People living in different regions or countries often have distinct political leanings based on the dominant political parties, historical events, social norms, and local issues. Exposure to these factors can shape individuals' political biases.
- **Media and Information Sources:** Users in different locations may have access to different media outlets and information sources. Media organizations, both traditional and digital, can have varying political biases themselves, which can influence the information and narratives users are exposed to. Users in different locations may consume news from sources that align with the dominant political ideology in their region, potentially reinforcing or shaping their own biases.
- **Social and Peer Influences:** Local communities and social networks can exert influence on political beliefs. People often discuss and debate political topics with friends, family, and neighbors, which can contribute to the formation / reinforcement of political biases.
- **Policy and Governance:** The policies and governance structures in a specific location can shape citizens' political beliefs. Different regions may have varying political systems, parties in power, and policy priorities. These factors can influence how people perceive political issues and develop their biases based on the local political landscape.

### 2.3 Decoding Crawling Search Engines

After examining the various types of search engine bias, it is crucial to dig deeper into the implications and challenges that arise as a result of these biases. The impact of search engine bias extends beyond individual users and has broader societal consequences that affect the information landscape and the accessibility of diverse perspectives.

I will put my focus on Crawling Search engines, as they are the most common and used nowadays.

Having gained a foundational understanding of search engine bias, it is essential to analyse deeper into the inner workings of search engines and explore them in detail to uncover potential leads that can provide further insights into the biases present in search results. Understanding the mechanisms and algorithms employed by search engines is a key step towards comprehending the factors that may contribute to information bias.

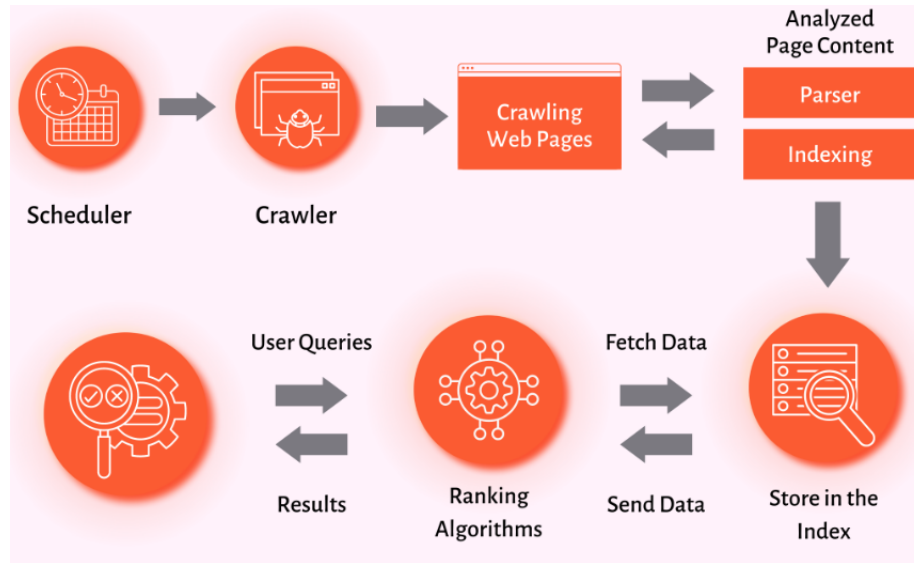
Search engines operate through complex algorithms that analyze and index vast amounts of web content, aiming to provide users with the most relevant and useful results for their queries. These algorithms take into account various factors, such as keyword relevance, website authority, user behavior, and contextual information, to determine the ranking and presentation of search results. However, it is within the intricacies of these algorithms that biases can emerge.

Search engines work through three primary functions:

1. **Crawling:** Scour the Internet for content, looking over the code/content for each URL they find.
2. **Indexing:** Store and organize the content found during the crawling process. Once a page is in the index, it's in the running to be displayed as a result to relevant queries.
3. **Ranking:** Provide the pieces of content that will best answer a searcher's query, which means that results are ordered by most relevant to least relevant.

*Crawling* is the discovery process in which search engines send out a team of robots (known as crawlers / spiders) to find new and updated content. Content can vary — it could be a webpage, an image, a video, a PDF, ... — but regardless of the format, it is discovered by links.

After the indexing step, where search engines process and store information they find in an index, a huge database of all the content they've discovered and deem good enough is set to serve up to searchers.



**Fig. 1.** Working process of a search engine

When a user performs a search, search engines scour their index for highly relevant content and then orders that content in the hopes of solving the searcher's query. This ordering of search results by relevance is known as ranking. In general, you can assume that the higher a website is ranked, the more relevant the search engine believes that site is to the query.

### 3 Exploring possible biases

In this chapter, I will dive into the intriguing realm of search engine biases, exploring the complexities that shape the digital information landscape. I will explore different search engines, but the idea is not only to compare the results between different search engines but also to try and see if they are internally affected by user behaviour, location and underlying algorithmic frameworks, leading to possible variations in the search results presented to users.

#### 3.1 User location

User location plays a pivotal role in shaping search engine results, catering to the specific needs and preferences of individuals in different geographical contexts.

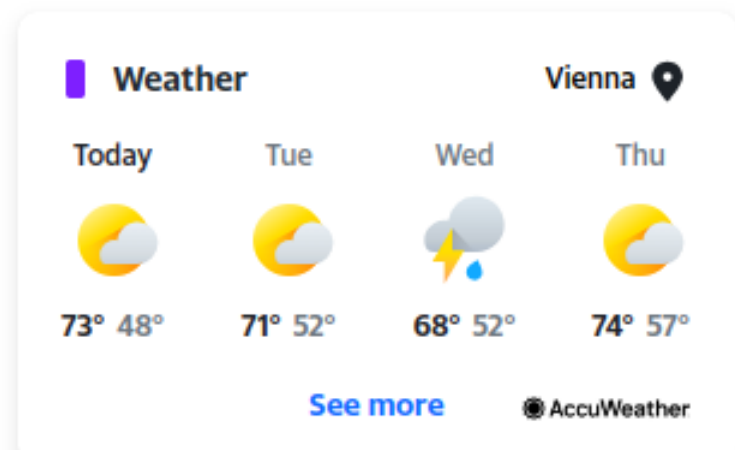
So in order to test the results, I explored 3 different search engines: *Google*, *Bing* and *Yahoo*.

In order to test something simple first, I just did a simple search, just writing "restaurants" on the search tab on the 3 search engines, but in 2 different locations, being the first in Vienna, Austria and the second in Lisbon, Portugal.

After doing the search, I got surprised by some results.

Before doing the search, and knowing that search engines collect my IP address when performing a search (as IP addresses are necessary for search engines to route the search results back), I was expecting for the results to be based on my location, meaning that the results would be biased by it. And that is exactly what happened in Google and Bing. I even noticed that before even completing the search, as the search engine was offering ideas for auto-completing the search.

But, to my surprise, it was different in Yahoo. Despite tracking my IP address just by getting into the site, as it showed me the temperature prediction for the next days in the location I was searching from, as we can see on the right corner in the image below.



**Fig. 2.** Yahoo showing me the temperature prediction for the next days in my IP address location.

But, despite tracking my IP address, the results were far from I expected. During the process of searching explained before, yahoo algorithm did not take into account my location, so it just provided me with restaurants all over famous places in the world and links to websites where I would be able to find restaurants if I introduced the location I wanted to. Even after scrolling down and exploring the next pages, I was not able to find evidence of considering my location in order to do a better search.

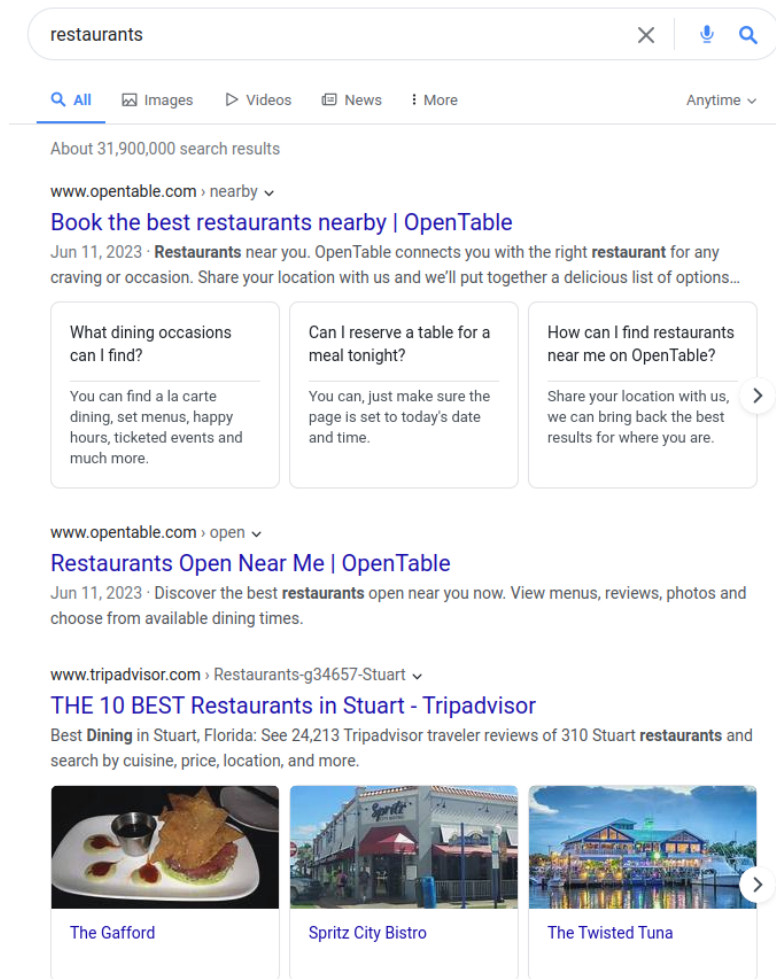


Fig. 3. Yahoo results after searching, from Vienna, for "restaurants"

After some unexpected results, I decided to continue exploring, and even went to do the same search in DuckDuckGo, but despite saying in the main page that they do not track user data, they showed biased results based on my location, despite not being so obvious as the other 2.

### 3.2 User Activity

**Past user history** Regarding results being biased by user activity, I already had some ideas on possible results, as me being someone who has been working with laptops for many years, and so I have been doing a lot of searching in search engines for the past years, I already had experienced some moments when my activity influenced my results, mostly using data from my search history:

- After searching for a specific topic constantly for a relevant amount of time, when I was writing on a search engine some keywords it would propose some auto-completion related to my past history involving a close topic (not necessarily the same), and after the search was complete, similar websites to/or the ones I visited the most were appearing, even though they should not be the first ranked links of that search request.

**Example:** about a year ago I was a lot of the time looking on google for cooking recipes, as I started to cook on my own. And after such a long time doing research about recipes and cooking related-websites, the search engine started to prioritize recipe websites or cooking blogs in the search results when I search for generic terms like "dinner ideas" or "cooking tips". Although it seems related, after doing it on other accounts I noticed that after searching for "dinner ideas" it would prioritize ideas of different dishes to eat, and not cooking websites explaining how to make them!

This is a small difference, but it is certainly visible the bias here, not only for showing different websites but for prioritizing websites that should not be ranked higher than the ones clearly appropriated for the search!

**User Language** As I knew that search engines tracks some of the user data, I started to check if some other data was extracted from the user, so I decided to use Portuguese accounts on the 4 different search engines: *Google*, *Bing*, *Yahoo*, *DuckDuckGo*.

I did the search without using any **Language-Specific Sub-directories**: directories or folders within a file system or a website structure that are used to organize content based on different languages. These sub-directories are commonly used in multilingual contexts to separate content by language, allowing users to access the information that matches their preferred language.

As I am exploring possible biases on search engine results, there is no point on doing my search on these sub-directories, as by that websites can structure their content in a way that facilitates language-specific navigation and improves the user experience for multilingual visitors. When a user selects a particular language, the website can redirect them to the appropriate sub-directory, where they can access content in their preferred language.

After performing a search "Latest news", in none of the 4 websites there appeared any Portuguese news / websites that provide news in Portuguese language, despite doing the search with an account with the country "Portugal"



stated as country of origin and language.

After this results, I reached the conclusion that search engines favour the search query at the most, prioritizing English-based newspapers, and some user relative data, in this case, user language, is not taken into an account in the search process.

As expected, the tested Search engines take keywords as a major step to realize the search.

These results could be predicted, as by looking at how the search engine process works for this specific case there is no evidence of extracting that kind of data by performing the search.

If we take a close look to the usual algorithmic process of doing a search we can pretty much reach the conclusion that the search engine does not access that type of data from the user (despite not knowing every little detail of the search, we can see the effects of a major part of a search request):

1. **Query Parsing:** The search engine analyzes the search query to understand its intent and the specific keywords used. In this case, it recognizes that I was looking for the latest news.
2. **Index Lookup:** The search engine looks into its index, which is a vast database of web pages and other content it has crawled and stored. It searches for web pages and documents that are relevant to, in this case, the latest news.
3. **Ranking:** The search engine applies complex algorithms to rank the indexed content based on various factors, such as relevance, quality, and user signals. It aims to provide the most relevant and authoritative results at the top of the search results page.
4. **Retrieval:** The search engine retrieves the top-ranked web pages from its index, considering factors like their relevance and credibility. This retrieval process aims to provide users with the most trustworthy and authoritative information available on the searched topic.
5. **Presentation:** The search engine displays the search results on the search results page. The top-ranked results appear first, followed by subsequent pages. For a query like "Latest news," a list of news articles or news websites providing the most recent and relevant news updates are shown.

As we can see, there is no evidence of extracting such user data, while, for example, when taking into account user location (Section 3.1), the process of gathering that information would be between steps 2-4, after analysing the search query.

## 4 Conclusion

After the exploration of search engines, their work process and their possible biased results, I was able to get a better understanding about the intricate interplay between algorithms, user behavior, and societal factors in shaping the information landscape we encounter online. While search engines strive to provide relevant and diverse results, inherent biases can emerge due to algorithmic design, personalization, and user location. These biases have the potential to influence the information we access, the perspectives we encounter, and the reinforcement of existing beliefs.

Acknowledging and understanding search engine bias is essential for promoting transparency, fairness, and inclusivity in the digital age. As a users, it is crucial to critically evaluate the information presented to us, seek alternative viewpoints, and advocate for search engine practices that uphold the principles of objectivity and impartiality. Additionally, continued research and dialogue surrounding search engine bias are vital for the ongoing improvement of search algorithms and the cultivation of an informed and diverse online ecosystem.

From my personal experience, I find it useful sometimes for search engines to try and get to know the user and his behaviour, in order to present results that would fit more into his ideas than to the others, but know I acknowledge even more how easy is to learn user behaviour and get to its data, and I am curious how can this evolve, also realizing that what I think in some situations it's beneficial for me may not be to the others, due to user privacy. I was not able to explore the whole possible biases that search results can be affected, but from testing some of the possible cases I was able to get a better understanding of which type of data does the search engine gets from the user and how does it classify it's importance.

## References

1. How Search Engines Work, <https://ahrefs.com/blog/how-do-search-engines-work/>
2. How Google search engine really works?, <https://www.dashclicks.com/blog/how-google-search-engine-really-works/>
3. Search Engine Bias and the demise of search engine utopianism, <https://yjolt.org/sites/default/files/goldman-8-yjolt-188.pdf>
4. Freedman D (2006, April 6): Why Privacy Won't Matter. Newsweek
5. CPA sues google over 'misleading' search results. WebCPA., <http://www.webcpa.com/article.cfm?articleid=193pg-acctodayprint=yes>
6. Understanding searcher behavior. Search Engine Watch., <http://searchenginewatch.com/showPagehtml?page=3613291>