GOGGLES: Democracy dies in darkness, and so does the Web

Brave Search Team
Brave
Munich, Germany
search@brave.com

Abstract

This paper proposes an open and collaborative system by which a community, or a single user, can create sets of rules and filters, called *Goggles*, to define the space which a search engine can pull results from. Instead of a single ranking algorithm, we could have as many as needed, overcoming the biases that a single actor (the search engine) embeds into the results. Transparency and openness, all desirable qualities, will become accessible through the deep re-ranking capabilities *Goggles* would enable. Such system would be made possible by the availability of a host search engine, providing the index and infrastructure, which are unlikely to be replicated without major development and infrastructure costs. Besides the system proposal and the definition of the *Goggle language*, we also provide an extensive evaluation of the performance to demonstrate the feasibility of the approach. Last but not the least, we commit the upcoming Brave search engine to this effort and encourage other search engine providers to join the proposal.

Keywords

search, open ranking, algorithmic transparency

1 Motivation

Democracy dies in darkness, a line recently adopted by the Washington Post as their slogan, warns us that unless people are informed with facts and truth, no true democracy is possible. Those who benefit from darkness have always tried to control media in order to control and manipulate public opinion with propaganda. Until recently, propaganda has been the exclusive domain of nation-states or state-sponsored actors through mass media [19]. With the mass popularization of the Web in the last two decades and the subsequent privatization of it by big platforms like Google, YouTube and Facebook, the paradigm has changed. Propaganda is no longer a tool of an elite, but it has been commoditized to the extent that it is as accessible as advertisement, becoming a weapon that too many actors have access to. One must appreciate the irony that those most vocal about the risks of propaganda are those who controlled it in the past. Nevertheless, the risk of fake-news—a neologism created to mitigate cognitive dissonance—cannot be ignored [5, 6, 30, 33, 36]. It is dangerous for a society if people living in it cannot distinguish between facts, opinions and outright misinformation. Although this danger has always existed, today the situation is dire if only because quantitative becomes qualitative and although all information is theoretically available, in practical terms it is not.

1.1 A Single Point of Failure

Like never before, all the information (and misinformation) of the world is available upon request. But the way to access this information has narrowed to become a quasi-monopoly. The abundance of information has led to a significant transfer of power from creators to aggregators. Access to information has been monopolized by companies like Google and Facebook [27]. While everything is theoretically still retrievable, in practice we are looking at the world through the biases of a few providers, who act, unintentionally or not, as gatekeepers. Akin to the thought experiment about the tree falling in the forest [3], if a page is not listed on Google's results page or in the Facebook feed, does it really exist?

The biases of Google and Facebook, whether algorithmic, data induced, commercial or political dictate what version of the world we get to see. Reality becomes what the models we are fed depict it to be [24]. And a reality defined by Google's search ranking algorithm, is one that does not and cannot capture the intricacies and variety of human knowledge and opinion.

Traditionally, the role of media was to serve as the middleman separating the chaff from the grain, of course with their respective biases. Journalists and editors were the curators and the publishing house was responsible by reputation and by law. Furthermore, every country had tens or hundreds of, to a certain degree, independent firms. Media consolidation in the 90s somewhat killed the field [37], reducing the number of firms able to filter information. But the real impact came with the consolidation of the big Internet platforms, basically Google and Facebook. The role of curation has been eliminated as the majority of value is captured by the platforms so it is no longer economically viable [10, 18, 25]. With fewer and weaker intermediaries, we also reduce the amount of independent points of views or windows to the world.

We have been forced to trust that the worldview of a few internet platforms is non-partisan while it clearly cannot be. The public space has been privatised by a handful of private corporations. Such concentration of access to information is a single point of failure, and it has failed.

2 Proposal

Let us start with a disclaimer; there is no technical solution that solves the aforementioned problem once and for all. The issues derived from monopolies are well understood and fall well beyond the reach of any technical solution.

However, what we could do, is to acknowledge that market dynamics coupled with freemium models tend to produce a winner-takes-all scenario [4], the prelude of monopolies. Under these market constraints, we propose to increase the number of options, windows through which reality is made sense of. While it would be desirable to achieve that goal through independent actors (platforms), in lieu of that we can achieve the same effect within the same platform. The proposal presented in this paper can be portrayed as a fail-safe to prevent any platform from becoming a single window to the world. If Brave or any other company were to displace Google, the ranking algorithm would still be the one dictating

the way the world is perceived. We would have changed actors, but the problem would remain.

In this paper we introduce ***Goggles***, which is meant to provide people with a way to access information according to their **explicit biases**. In layman's terms, to put *Goggles* on, to see a different version of reality.

Search engines are free to incorporate user-defined *Goggles*, specified in an open language drafted in Section 5, and modify their ranking so that the user's explicit preferences take precedence over the ranking of the search engine itself.

Such system would have the potential to pierce a hole in the single-window effect produced by the search engine's ranking algorithms. In a way, it is opening the ranking algorithm to the people using the search engine.

Goggles go beyond personalization. As a matter of fact, they are orthogonal. The rationale is not to customize the ranking according to the implicit interests of the user, but to offer a mechanism to define multiple rankings, plural, open and explicit, for only if it is so, can it be trusted. The benefit for the users is that they would be empowered to explore multiple realities in a straight-forward way. The point is to offer people the freedom to choose their own biases while being conscious of them. The benefit for the content creators is that they have multiple options to expose their content, by increasing their potential audience, which will reduce the need to optimize for the single set of biases implicitly encoded in the search engine's ranking [17].

The point is not to create an even stronger echo-bubble, which is what happens under personalization. Rather, the aim is to promote plurality and let people proactively and consciously choose. Confirmation bias exists; people tend to only acknowledge information that fits their own bias [26]. However, a large fraction of people are interested in exploring alternative viewpoints [14]. Current platforms, however, do not facilitate such exploration process [22], seeking alternative options (*for better or worse*) implies a cost. The costlier it is, the less likely it becomes for people to break from the single-window effect exacerbated by the ranking algorithms.

It is also not the point of *Goggles* to mitigate the fake-news phenomena, at least not directly. While having more plurality opens the space for wacky theories, it also opens the space for rational and informed ones. The way to fight fake news is to rebate them, not to ban or bury them [11]. Otherwise we will have no instrument left to control those who decide what qualifies as fake [1].

We envision a scenario where a community of people create and curate *Goggles* like,

- **"Tech Blogs".** Imagine searching through a collection of personal and company blogs curated by the community.
- **"Product Reviews without commercial intent".** Get rid of all sites with price comparisons, affiliate links, etc. Basically, to browse over product descriptions and reviews.
- **"Independent Media for any country".** Would demote major newspaper and promote minor outlets.
- **"Exclude top 1000 domains".** Would remove results from most popular domains on the Web to surface less prominent ones.
- **"Recipe search that my mom likes".** Only searches recipes on tasteofhome.com, nowhere else being considered, would become a site search.
- **"Nature lovers in the Pyrenees".** An extremely curated list of high-quality sites for hiking/trekking in the area. Excluding the more generic sites not specialized in that area.
- **"Wikipedia / Reddit / <Any site> search".** Site search is just an instance of what *Goggles* can be. The other way round also works; results that exclude results from a given site (e.g. Facebook).
- *We recently observed the tech community discussing the shortcomings of search engines [9], particularly in surfacing content by some spaces in the web. It was exciting to see how almost all the use-cases in the discussion could be addressed by* Goggles.[2]

Each of these *Goggles* is fully owned, controlled and maintained by its creators according to their own terms and services. *Goggles* can be shared, extended, and modified to fit anyone's particular needs. The most likely scenario, however, is that the great majority of users will rely on *Goggles* maintained by others because of their coverage, quality, and most importantly, because of the trust of the maintainers' integrity. Trust is an important aspect of *Goggles*. There is no way to guarantee that a particular Goggle fulfils its promise, but any Goggle can be forked, and their users vote with their feet. The fact that the list of rules composing a Goggle is open and can be copied/extended by anyone will prevent the creation of a lock-in by the original authors/creators, mimicking the ecosystem lock-in of the likes of Apple, Google and Facebook [28]. Of course, for such system to work, people must trust that the search engine serving as host applies the rules defined by the Goggle against their index without alteration. Besides the language definition, which must be standard to allow integration with the search/retrieval algorithms, a search engine should stay out of the *Goggles* ecosystem to maximize trust and variety.

The contributions of this paper are:

(1) To propose the concept of *Goggles* for open/collaborative ranking. Note that the proposal/definition alone, is not entirely novel (as will be discussed in the Background Section 3).
(2) To define the *Goggles* language, which allows people to define their own ranking preferences in a simple way, using a grammar inspired by the ad-blocking community (proven to be both easy to write and maintain and to be expressive enough.)
(3) The commitment that the Brave search engine will implement and apply user-defined *Goggles*. Which means modifications on the ranking algorithms (details in Section 4). We encourage other search engines to follow. *Goggles* is in no way owned by or exclusive to Brave search engine. It belongs only to its creators and users.
(4) To show that search engines can serve an additional role to the community by exposing their infrastructure and index. Allowing public and open access to such privileged resources.

[1] *Quis custodiet ipsos custodes?* Who will guard the guards themselves?

[2] Note that *Goggles* project started late 2019 but was put on hold due to the shutdown of the Cliqz search engine. Happily, the project will continue as part of Brave from 2021 onward.

Let us emphasize once again that this proposal, *Goggles*, does not fix the problems of misinformation, echo-chambers, confirmation biases, etc. These problems are very human in nature, and no technology can solve them. At most, it can only exacerbate or mitigate them, the latter being the case of the system presented in this paper. What we propose in this paper is a way to decrease the single-window effect created by the search engines such as Google, Bing, and of course, Brave. By opening the ranking from one(s) to many we open the possibility of having many different rankings, serving different biases and intents. Needless to say, that search engines must collaborate on that effort by providing the infrastructure and index to back it up.

Goggles intends to offer multiple perspectives to the same query and to be explicit about it. So that people choosing liberal media *Goggles* are free to do so, but this is a conscious and deliberate choice. If they want, they can explore the opposite *Goggles* to expand their perspective. Something as simple as this is not easy, as systems are not designed to that purpose [7, 32]. Allow us to stress that the biases embedded on a Goggle do not need to be "positive". There will be *Goggles* created by creationists, anti-vaccination supporters or flat-earthers. However, the biases will be explicit, and therefore, the choice is a conscious one. We do not anticipate any need for censorship in the context of *Goggles*. Clearly illegal and sensitive content like child pornography or extreme violence should already be filtered out by the host search engine at the index layer. Consequently, such content should not be surfaced by any Goggle.

We would like to stress out that biases do not need to exist only on highly polarizing issues such as politics, religion, language, etc. Non-partisan topics like strong localization, advertisement or commercial intent removal are likely to have a strong presence. *Goggles* can just be ways to increase plurality and open niches for content that is otherwise buried under the rule of a single source of ranking.

3 Background

To the best of our knowledge *Goggles* is the first attempt to open up the ranking component of a search engine to the community.

Perhaps the most related system to *Goggles* is personalization [23], the ability to alter ranking according to the user's interests or intents. Note that this comparison, although reasonable, is deceptive. Personalization, outside the realm of faceted search [2, 34], is not actionable for the user, at most they can opt out from it. The aim of *Goggles* is not to have a single ranking fitting better the user's interests, but to offer users a wide range of possible rankings and let them choose. The same rationale applies to rankings subjected to locales, either language or geography.

We mention faceted search, which shares with *Goggles* that ability to provide external information to the query to help the search engine refine the results the user was looking for. In the case of faceted search, the user does not provide an external rule for ranking, but additional metadata, typically in a structured form. For instance, named entities, reference codes, dates, etc. Information provided by the user to facilitate the retrieval. This approach is useful on many verticals like flights, trips, books, movies, products, but is not the most convenient for general purpose, as it demands from the user a) knowledge of the domain, and b) extra burden on the input query. *Goggles* also imposes these constraints at creation time, but not while using them. Thus, the extra effort is not paid by the end-user but by the Goggle's creator/maintainer.

Goggles also share similarities with collaborative efforts for content discovery and classification, for instance, social bookmarks systems [20, 29] or curated lists [31]. However, such systems are designed for sharing and not suitable for search both because of the limited coverage and the lack of a proper search infrastructure.

Another area where *Goggles*' contribution is relevant is algorithmic transparency. We are not aiming to make the Brave search engine ranking transparent, but rather to allow people to modify and alter it *a posteriori*. Transparency of the ranking would provide explainability and accountability for the results and it would help to detect unfairness or illegitimate biases (e.g. gender, race, religion). We could achieve similar results with *Goggles*, but in an indirect manner. Note that full transparency on the ranking (the main ranking algorithm that is) would introduce challenging problems. Intellectual property aside, which is not a small thing, we would further open the search engine to the harmful effects of SEO (search engine optimization). SEO, especially when invasive, is one of the biggest headaches search engines have, giving access to the particularities of the main ranking would immediately result in a boost of those sites that rely on SEO to be on top, which are usually not the ones with the best content.

A similar argument can be made on the topic of open search. This proposal does not open the full search engine, but it provides the ability to modify the most important constituent, the results. Building, maintaining and operating a search engine is neither easy nor cheap. Something along the lines of our proposal could become a suitable middle ground. Traditional search engines could act as hosts, providing their index and computational resources. The final ranking, however, could be driven by a community of people maintaining a large and open collection of *Goggles*.

The underlying idea behind *Goggles* is simple, borderline trivial. As a matter of fact, related concepts have been proposed in the past [12], however, unless it is coupled with a search engine infrastructure, the chances of success are small. Custom rerankers are only one side of *Goggles*. Performing a rerank, depends both on the rules of reranking but also on the original result-set where the rules will be applied. Hence, the effectiveness of the system is predicated on obtaining a large set of results on which the rules can be applied. Without the active collaboration of a search engine provider, such large result-set is not available. Top 10 results or top 50 in the case of Bing API [13] are not nearly big enough. Of course, scraping is always a possibility, but latency will become an unsolvable issue. It would take a few seconds to scrape the first 100 results out of a search engine, if we manage to not get blocked. And still, a result-set of 100 results, while better than 10, is still way too small. The only way to efficiently implement something like *Goggles* is with the collaboration of a search engine which allows the user to send a custom re-ranking function to be applied to the first set of results (typically in the tens of thousands) rather than on the final steps where the candidate result-set has already been reduced enough to have a poor overlap with the user custom re-ranking. In Section 5 we briefly describe how the *Goggles* language is applied to Brave's search ranking algorithm.

4 Integrating with existing search engines

Modern search engines have strict latency requirements, usually less than a second, in which they need to respond to the user query. A common way to architect a search engine to address this issue is to split the process into multiple phases. The recall phase involves matching the user query against billions of (in some cases, a lot more) pages with simple features to help reduce a candidate set to a reasonable size for further processing, typically in the order of few thousands. Subsequent phases, usually known as precision phases, narrow down the candidate set using a stack of increasingly sophisticated and costly models. The last phase of this process, the ranking, involves a very small candidate-set and is the one responsible for the final ordering of results given to the user.

The effectiveness of *Goggles* increases the earlier they are integrated into the search process so that more pages can be subjected to the rules being applied. Consider the Goggle "Filter out the results from the top 1000 domains on the internet", which could be an interesting way to explore the internet. Applying this on the final result set for most queries would lead to very few results, if any, due to the inherent bias in most search engines to surface content from popular domains. The rules defined by *Goggles* are better applied to the largest candidate-set possible, so that the intersection between candidates and rules to be applied is not empty. Only when intersection is large enough, will the re-ranking introduced by *Goggles* be noticeable.

Deep integration between *Goggles* and the host search engine is needed for the system to work. However, such integration poses different issues: 1) *Efficiency:* applying the rules against all elements of the candidate set (typically URLs) has to be extremely fast to minimize the overhead. In the following section we will present our solution to this issue. And 2) *Independence:* the host search engine needs to have total control over their index. This trait is given on search engines running their own fully-fledged index, e.g. Google, Bing, Yandex, Baidu and Brave. However, other search engines that rely totally or in part on external indexes might not have the ability to pull a large enough candidate-set to perform the user re-rank defined on his Goggle. DuckDuckGo, Qwant and Ecosia, which rely on the Bing API, are limited to whatever the API offers.

In this paper we lay down the language and the supporting matching engine, however, integrating such system into the code of a large-scale search engine is non-trivial. We commit Brave search engine to do so, to be a host for *Goggles*. We believe and welcome other search engines to also be hosts, after all, the more choices of *Goggles* and of hosts search engines, the better.

5 Language for *Goggles*

For the purpose of *Goggles*, we created a DSL (Domain Specific Language) which will allow users to express rules able to capture flexible filtering logic applied on a large set of search results. This DSL needed to be plain text and self-contained to ease hosting and sharing, flexible enough to express fine-grained filtering logic of URLs and page features, yet sufficiently constrained so that filtering can be implemented in a very efficient way (as mentioned previously, this system needs to be able to match thousands of candidate results against thousands of rules for each user query, without impacting latency in a perceivable way). Finally, it needed to be accessible enough so that even people without a technical background could quickly grasp its syntax and write rules, which would also encourage collaboration around the creation and curation of *Goggles* (e.g. communities).

After considering all these requirements, we realized that we could leverage prior work, addressing a totally different use-case but sharing similar challenges. We decided to base our DSL upon a subset of the syntax used by content blockers to perform "network filtering" (i.e. ads- and trackers-blocking): the so-called "Adblock-Plus filters syntax" [1, 21]. This language already proved in the past that it, 1) allows to express logic to target URLs in a powerful way, 2) can be implemented extremely efficiently [38], and 3) is friendly to contributors and gave rise to numerous communities maintaining lists with a robust open collaboration model [15, 16, 35].

The language is also already widely documented, is flexible enough to allow custom extensions while maintaining backward compatibility (e.g. new options can be added without breaking other engines). This last point is especially important since we hope that other search engines will follow suit and also adopt support for *Goggles*. It was observed in the content-blocking communities that, in practice, maintainers have an incentive to keep compatibility with a maximum number of engines, and will thus use the features which are widely supported in priority (common denominator) and rely on engine-specific features only if they cannot do otherwise; this allows some flexibility for engines implementing custom extensions to the language.

We now give a brief overview of this language, the draft spec of which will be hosted publicly and open for participation in the future.

A list of filters, or Goggle, is a self-contained text file where each line can contain a filter (empty lines or comments—line starting with a '!' character—are ignored). Ranking of search results will be altered based on the filters contained in the file. Each filter is composed of two parts: a *trigger* and an *action*, separated by a *$* character: *<trigger>$<action>*. The trigger part is a pattern which needs to match a result candidate. It can leverage the following features:

- **Plain Patterns**—allow targeting a URL (or another result attribute like its title) based on a string of characters which it should contain. The filter "/coronavirus-" would trigger on any URL containing this specific string of characters (e.g. https://example.com/coronavirus-update.html).
- **Wildcard Patterns**—extend plain patterns with globbing capabilities: the special symbol "*" can be used to match any number of characters. Filter "/health/*/coronavirus-" would match any URL containing the substring "/health/", followed by zero or more characters, then "/coronavirus-" (e.g. https://example.com/health/2020/coronavirus-update.html).
- **Left and Right Anchors**—introduce a special "|" character which, when appearing at the start or end of a filter, forces a pattern to match the beginning or end of a URL. Filters "|https://" and ".html|" would match URLs starting with *https://* or ending with *.html*, respectively.

Each filter can also be annotated with additional options (following the "$" character). Multiple options can be specified at the same time, and separated by comas. We leverage this syntax to

add ways to further fine-tune the behaviour of *Goggles*; either to specify which features of a result candidate should be considered (i.e. *target*), or how the ranking should be affected (i.e. *action*). For example:

- **$boost=XX**—is used to alter the ranking of specific results by *XX* (e.g. *$boost=1* would not alter the ranking, while *$boost=2* would make a result two times more important).
- **$discard**—completely drops candidates from the list of results.
- Filtering based on specific attributes of the result page can be achieved with:
  - **$lang=XX**—to target the language.
  - **$inurl**—to target the URL.
  - **$inquery**—to target queries leading to a candidate.
  - **$intitle**—to target the title.
  - **$indescription**—to target the description.
  - **$intext**—to target the full content.

Last but not the least, these features can all be combined to form complex filters. For example, the filter: */news/*/covid.html|$inurl*, would match candidates based on their URL.

This description is by no mean complete or final, and we will release a specification of the language once it is stabilized.

5.1 Protocol

To allow users and communities to create and curate *Goggles* over time, we propose the following protocol, inspired by the most successful filters maintainers from content-blocking communities.

We propose two modes operations for maintainers: 1) A development setup implemented as a Web User Interface which allows to quickly get feedback over newly created filters, by showing which results end up in the final result set in real time. This setup is intended to speed-up the process of creating filters, reducing friction and offering a seamless workflow. The resulting filters can then be hosted publicly on a platform such as GitHub and made available to a wider public. And 2) The production setup which is directly integrated into any search engine prepared to be a host for *Goggles*. The end user could specify a link (or identifier) to the *Goggle* in the form of network accessible URI. The search backend is then responsible for fetching the Goggle definition from the URI (or use a cached version of it), compiling it to an efficient representation optimized for matching speed, and applying it at the recall-phase to the search results to produce a resulting candidate set.

5.2 Privacy Considerations

It is important to consider the potential privacy implications of sending a *Goggles* URIs together with the query. The URI can become a unique user identifier, especially for those people using non-popular *Goggles*. Therefore, there is a risk of a host search engine building a partially complete user profile in some circumstances. This should not be a problem for all host search engines, though; Google and Bing for instance, link all queries to the users' accounts and consider it a desirable feature. However, for privacy preserving search engines like Brave, this becomes a hurdle.

Note, however, that the URI only doubles as a user identifier under certain conditions: 1) when a user is consistently using it for all queries, and 2) when the URI is only used by that user (or a very small group of users). None of these conditions should be the default *modus operandi* of *Goggles*. We would expect *Goggles* to be used only for a fraction of queries. Also, we expect users to rely on multiple *Goggles* for different tasks. And finally, we expect a great majority of users to rely on popular *Goggles*, for which the URI is not a valid user identifier. Reality, however, does not need to conform with expectations. We should provide an additionally mechanisms to protect privacy for those niche cases. One proposal would be to allow sending multiple *Goggles* URIs on a single query, so that the true *Goggle* is obfuscated on a larger set. The host search engine would return results for all the *Goggles* and on the client-side the results for the padding *Goggles* would be dropped. This approach, however, imposes a serious overhead on the host search engine. The final solution to this problem is left for future work.

Number of URLs	Number of filters	Time (ms)
1000	1	0.17
1000	10	0.20
1000	100	0.24
1000	1000	0.33
10000	1	1.56
10000	10	1.78
10000	100	2.08
10000	1000	3.10

Table 1: Summary of the performance evaluation (time in milliseconds) for different number of URLs and filters.

6 Performance evaluation

As previously discussed, *Goggles* can only shine when applied to a very large candidate set of results (thousands of URLs). For this reason, the filtering logic can only take place in the search backend, during the recall phase. Consequently, we operate under a very tight time budget (few milliseconds) to ensure that the overall search latency is still acceptable and that the backend remains able to handle many concurrent requests from users.

To assess the viability of *Goggles* from a performance perspective, we first implemented a prototype leveraging our in-house high-performance JavaScript content blocking library [8], then a custom Rust re-implementation of a similar engine, tuned for performance. The following figures were obtained by sampling 10k results with query "coronavirus" from our search index. The filters used were a selection of 1000 domains from the most popular domains, which we use as a "trustworthy list of domains"-*Goggle*. We run the measurements with varying number of URLs and filters to get insights into how the total time evolves as a function of the input size. Results are summarized in Table 1. These measurements were performed using our Rust prototype, compiled with rustc 1.43.1, on a reasonably fast ultrabook CPU (i7 U6600) using two cores (4 logical threads using hyper-threading).

From these results we can conclude that our initial Rust prototype is already delivering good performance on a reasonably large set of candidate URLs (note that recall phase is typically sharded across multiple servers, so the aggregated candidate set could be

much larger). The figures obtained from our reference implementation give us confidence about the feasibility of the approach, even on the rare case of a single server. Secondly, we observe that the processing time per-request is almost constant thanks to the efficient dispatching data-structure used in the filtering engine [38]; this shows that *Goggles* could be handling many more filters while still meeting our time budget; the runtime being almost exclusively impacted by the number of URLs in the initial result-set (assuming the filtering runs on a single CPU). Digging further, we observed that pre-processing of URLs, which consists of extracting the hostnames as well as tokenizing the URL, is the current bottleneck with a total of 70% of the overall time spent, whereas looking up filters from the index only takes around 10% of the total time. This shows that we could improve the performance drastically by focusing our effort on these two functions.

7 Conclusion

We believe that the system/framework proposed in this paper would be beneficial to maintain a healthier Web. *Goggles* would foster openness and diversity thanks to the community maintenance and ownership. The later being very important as the added value created should not exclusively be in control of the host search engine, or else we might end up on the current status-quo. Besides, community *Goggles* also requires the active participation on a host search engine, which would provide access to its index and infrastructure. We are happy to commit Brave search to this endeavor, as we did with the now defunct Cliqz search[3].

Needless to say that *Goggles* will be open to any other search engine or institution that is enticed by this proposal.

Acknowledgments

The bulk of *Goggles* can be attributed to the Cliqz search engine team, which was shutdown in March 2020. Fortunately, parts of the team and core intellectual property is now part of Brave, which is happy to continue the mission of building an alternative search outside of Big Tech.

References

[1] AdblockPlus. [n.d.]. AdblockPlus filters explained. https://adblockplus.org/filter-cheatsheet

[2] Ori Ben-Yitzhak, Nadav Golbandi, Nadav Har'El, Ronny Lempel, Andreas Neumann, Shila Ofek-Koifman, Dafna Sheinwald, Eugene Shekita, Benjamin Sznajder, and Sivan Yogev. 2008. Beyond basic faceted search. In *Proceedings of the 2008 international conference on web search and data mining*. 33–44.

[3] George Berkeley. 1881. *A treatise concerning the principles of human knowledge*. JB Lippincott & Company.

[4] K Boudreau, Lars Bo Jeppesen, and Milan Miric. 2019. Competing on free(mium): digital competition with network effects. *SSRN Electron. J* 10 (2019).

[5] Pew Research Center. 2016. *Many Americans Believe Fake News Is Sowing Confusion.* https://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/

[6] Simone Chambers. 2021. Truth, Deliberative Democracy, and the Virtues of Accuracy: Is Fake News Destroying the Public Sphere? *Political Studies* 69, 1 (2021), 147–163. https://doi.org/10.1177/0032321719890811 arXiv:https://doi.org/10.1177/0032321719890811

[7] Allison J. B. Chaney, Brandon M. Stewart, and Barbara E. Engelhardt. 2018. How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, British Columbia, Canada) *(RecSys '18)*. Association for Computing Machinery, New York, NY, USA, 224–232. https://doi.org/10.1145/3240323.3240370

[8] Cliqz. [n.d.]. *Efficient embeddable adblocker library.* https://github.com/cliqz-oss/adblocker

[9] Hacker News Community. 2020. *We can do better than DuckDuckGo.* https://news.ycombinator.com/item?id=25129431

[10] Australian Competition and Consumer Commission. 2019. Digital Platforms Inquiry. https://www.accc.gov.au/publications/digital-platforms-inquiry-final-report

[11] Saoirse De Paor and Bahareh Heravi. 2020. Information literacy and fake news: How the field of librarianship can help combat the epidemic of fake news. *The Journal of Academic Librarianship* 46, 5 (2020), 102218.

[12] Google Developers. 2020. *Custom Ranking.* https://developers.google.com/custom-search/docs/ranking

[13] Microsoft Docs. 2017. *Custom Search API v7 reference.* https://docs.microsoft.com/en-us/rest/api/cognitiveservices-bingsearch/bing-custom-search-api-v7-reference

[14] Elizabeth Dubois and Grant Blank. 2018. The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society* 21, 5 (2018), 729–745. https://doi.org/10.1080/1369118X.2018.1428656 arXiv:https://doi.org/10.1080/1369118X.2018.1428656

[15] EasyList. [n.d.]. EasyList filter subscription. https://github.com/easylist/easylist

[16] EasyList. [n.d.]. EasyList forum. https://forums.lanik.us/

[17] Eric Goldman. 2005. Search engine bias and the demise of search engine utopianism. *Yale JL & Tech.* 8 (2005), 188.

[18] Ruth A. Harper. 2010. The Social Media Revolution: Exploring the Impact on Journalism and News Media Organizations. *Inquires Journal* (2010).

[19] Edward S. Herman and Noam Chomsky. 1988. *Manufacturing Consent: The Political Economy of the Mass Media.* Pantheon Book.

[20] Paul Heymann, Georgia Koutrika, and Hector Garcia-Molina. 2008. Can social bookmarking improve web search?. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*. 195–206.

[21] Raymond Hill. [n.d.]. uBlock Origin static filter syntax. https://github.com/gorhill/uBlock/wiki/Static-filter-syntax

[22] Ryan Holmes. 2016. *The Problem Isn't Fake News, It's Bad Algorithms—Here's Why.* https://observer.com/2016/12/the-problem-isnt-fake-news-its-bad-algorithms-heres-why/

[23] Stephen Lawrence. 2005. Personalization of web search. US Patent App. 10/676,711.

[24] Leonard Mlodinow and Stephen Hawking. 2010. *The grand design.* Random House.

[25] Efrat Nechushtai. 2018. Could digital platforms capture the media through infrastructure? *Journalism* 19, 8 (2018), 1043–1058. https://doi.org/10.1177/1464884917725163 arXiv:https://doi.org/10.1177/1464884917725163

[26] Raymond S Nickerson. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology* 2, 2 (1998), 175–220.

[27] House Committee on Judiciary. 2020. *Investigation of Competition in Digital Markets.* https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf

[28] Jean-Christophe Plantin, Carl Lagoze, Paul N Edwards, and Christian Sandvig. 2018. Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society* 20, 1 (2018), 293–310.

[29] Christian Bauckhage Robert Wetzker, Carsten Zimmermann. 2008. Analyzing social bookmarking systems: A del. icio. us cookbook.

[30] Dietram A. Scheufele and Nicole M. Krause. 2019. Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences* 116, 16 (2019), 7662–7669. https://doi.org/10.1073/pnas.1805871115 arXiv:https://www.pnas.org/content/116/16/7662.full.pdf

[31] Sindre Sorhus. [n.d.]. *Awesome lists about all kinds of interesting topics.* https://github.com/sindresorhus/awesome

[32] Miriam Steiner, Melanie Magin, Birgit Stark, and Stefan Geiß. 2020. Seek and you shall find? A content analysis on the diversity of five search engines' results on political queries. *Information, Communication & Society* 0, 0 (2020), 1–25. https://doi.org/10.1080/1369118X.2020.1776367 arXiv:https://doi.org/10.1080/1369118X.2020.1776367

[33] Joshua Tucker, Andrew Guess, Pablo Barbera, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. *SSRN Electronic Journal* (01 2018).

[34] Daniel Tunkelang. 2009. Faceted search. *Synthesis lectures on information concepts, retrieval, and services* 1, 1 (2009), 1–80.

[35] uBlockOrigin. [n.d.]. Resources for uBlock Origin, uMatrix: static filter lists. https://github.com/uBlockOrigin/uAssets

[36] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151. https://doi.org/10.1126/science.aap9559 arXiv:https://science.sciencemag.org/content/359/6380/1146.full.pdf

[37] Paul Wellstone. 1999. Growing media consolidation must be examined to preserve our democracy. *Fed. Comm. LJ* 52 (1999), 551.

[38] WhoTracks.Me. 2019. Adblockers Performance Study. https://whotracks.me/blog/adblockers_performance_study.html

[3] See Acknowledgments