Project Report **ClusterMe!**

Team	Marina Able
	Julian Albers
	Rishith Ellath Meethal
	Anna-Lena Fackler
	Daniel Hernández
	Kyra Kleine
	Tilman Knopp
	Mohammad Youssef Mahfouz
	Sebastian Mair
	Janna Nikonov
	Lea Sophie Seier
	Nicola Stadler
	Julian Trummer
Tutor	Thomas Bickel Haase
	Michael Vetter
Supervisors	Dr. Hannes Petermeier
	Prof. Dr. (em.) Jürgen Scheurle

Preface ClusterMe by Jürgen Scheurle

Modern media provide an excellent opportunity to gather information about almost everything. In this regard the internet (world wide web) has become one of the most popular media. In fact, nowadays it is common to use some search engine on the internet to find out what to do next, where to go, what to buy or rent, what to consume, etc.. Even knowledge traditionally to be found in some encyclopedia or in books is available on the internet.

However, several search engines are run by advertising companies such as Google, which make a huge profit by collecting and storing personal data including search preferences. Using these data, they present personalized search results as well as commercials selected and ranked by means of sophisticated algorithms. So, often the information provided on the internet is actually biased in one way or another and can strongly depend on collected personal characteristics (or rather on characteristics related to the individual internet access device used).

Obviously, this has certain advantages for the users, but there is the drawback of possibly not becoming informed objectively and comprehensively. This is a crucial issue. Therefore, the project team ClusterMe of the academic year 2017/II of the „TUM: Junge Akademie" has developed an online-tool in order to examine to what extent Google search results are personalized and to raise people's awareness concerning that issue. Everybody is invited to support and to take part in this initiative by visting the website www.cluster-me.com.

The ClusterMe online-tool allows to compare the Google search results received by different people (by means of different internet access devices, respectively). In particular, in the case of certain prescribed search items, it automatically determines and visualizes clusters formed by participants who receive similar (in some well- defined quantitative sense) search results. So, participants can find out to which cluster they belong. Also, they are provided

with the search results of all the other clusters. These clusters are supposed to reflect Google's clustering based on personal user profiles. Last but not least, participants are informed about possibilities of how to get unbiased search results on the internet.

The project team ClusterMe did a great job. Being a group of thirteen students from various faculties at the Technical University of Munich, it is not easy to agree upon a topic for a collaborative scientific project and to coordinate the cooperation of the team members in an efficient and goal-oriented way. The team ClusterMe successfully managed to achieve all that and to obtain interesting results during the project period of about eighteen months. Being a mathematician, I especially appreciate, that the team decided to choose a number of quantitative rather than just empirical methods to analyse differences between Google search results. As a mentor of the team, I helped the members of the team to stay motivated and focused over the whole project period, and I offered ongoing academic advice.

Of course, the impact of the ClusterMe initiative depends on the size of the sample of participants. Unfortunately, due to time constraints, only a relatively small group of students participated so far. Hence, the results obtained up to now are not very likely to be representative for general users of Google's search engine. Having more participants could make quite a difference. So, the project is worth to be continued and to be further developed. In any case, general internet users will benefit a lot from participating.

How old is Helene Fischer, where can I find the next best pizzeria and why can we drink water while doing a handstand? The answer to such questions is usually the same: Google.

No matter which question is on our mind, we simply pull out our mobile phones or start the laptop in front of us, open the internet browser and "google" it. Within seconds, we get a series of search results proposed. But the websites we end up on, are usually already at the top of the list. And that is no coincidence. The team behind the website called "ClusterMe" has therefore set itself the task of shedding more light on the search giant Google. Their aim is to find out whether and to what extent the suggestions of the search engine depend on our personal characteristics. Through their online tool, they want to enable Internet Users to compare the results of google queries with other users.

Let's assume that two friends google the same keyword, for example "vegan", both using their private mobile phone. Against expectations, the two friends do not get the same results suggested. Imagine friend A, who has been vegan convinced for years and who has used Google before to find new vegan recipes. His friend, on the other hand, has a taste for meat and he has let Google know that in the past. The hypothesis is, that looking at friend A's first Google suggestions, they might include tips for vegan restaurants or organic supermarkets in his area, completed by new vegan recipes. Among the list of search results for friend B you might find a Wikipedia article about veganism and articles that demonstrate and weigh up the pros and cons of a vegan diet. To put it somewhat exaggeratedly, the theory is, that the search giant Google, knows exactly what the two friends are interested in and on which page they are most likely to click on.

But how does Google know? How would the friends end up with such different, but somehow accurate results? One thing must be clear when using the services of the search giant: Google is an advertising group. They use the search terms that its users enter in the search form to present personalized advertising to each one of them. The company collects and stores data, including users' search history and account information. This way the company might know exactly what our two friends' likes and dislikes are and what their attitude is towards a certain topic. The theory is that one will most likely get search results that resemble rather preferences and less likely conflictive points of view. So you could say, one googles in a filter bubble. In other words, one would be trapped in a cluster. It is precisely these clusters, that the online tool "ClusterMe" wants to prove and make visible for Internet users.

But doesn't life become a bit easier, because the search engine already shows a pre-selection of search results tailored to ones needs? Yes, partly. After all, getting information that is relevant to oneself makes life much faster and less complicated. For example, looking for the weather forecast, it is helpful to get results that are related to the current location or hometown. The situation is different, however, with controversy topics. Think about, for example, the recent European elections. A user is looking for information to help him make the right choice. Suddenly it becomes quite dangerous, if he should only get a certain selection of party homepages suggested. The voter could make his decision without having to deal with the views of other parties.

For this reason, ClusterMe was invented. The online tool examines the personalization through Google. It tries to verify and visualize the clusters, into which searchers are put. It further questions, whether the different clusters are also connected to the personal background of each user, such as sex, age or special diet. The ClusterMe website has been online since the beginning of May and is freely available at www.cluster-me.com. After the user has filled out a short questionnaire and installed a plug-in on their device, the tool is ready to operate. In the background the search query for different keywords, such as "vegan" or "HIV", runs automatically. In the following step, the results of the user as well as those of oth-

er participants are illustrated as simple dots. The distance between these dots reflects the difference between the results the users are proposed to by Google. Similar search results will therefore gather and be defined as a cluster. The user has the possibility to compare the results of different participants and study an analysis of the resulting clusters.

But now that one knows about these potential clusters, the question arises, how to avoid them and how to bypass Google's personalization to get a broader selection of search results. There are multiple answers to these questions. First of all, deleting one's search history and setting the device to private browsing mode, can help. It is also useful to be logged out of the Google Account while searching. However, it remains unclear to what extent this effectively reduces personalization. Alternatively, other search engines, like DuckDuckGo or Startpage, can be considered. These two search engines have set themselves the task of setting new standards for data protection when searching the Internet. They explicitly advertise that they do not store or pass on any information about their users.

Consequently, the question comes up, how Google manages to be the indomitable No. 1. among search engines, despite its sometimes dubious handling of data protection guidelines. According to statistics from Statcounter market researchers, Google handles almost 93 percent of all search queries globally. In addition, competitors such as Yahoo (1.9%) or DuckDuckGo (0.37%) do not reveal any serious challenger. How does Google manage to displace all competitors or at least keep them at distance? Anyone looking for answers to these questions must first study the history of the company. Unlike its competitor Yahoo, Google did not need any human intervention to search the web early on and relied exclusively on an algorithm. Further, the Google founders were convinced that it was not important in the search engine business to produce content themselves – it is sufficient to organize the content of others as well as nobody else does. While others overslept further development, Google is still working on improving its algorithm every day. And as we know, this pays off for the billion-dollar group.

In the end, the web is a uniquely huge and versatile information and communication space, that has shaped and changed our lives in many ways. But it is also a place to be enjoyed carefully. It is the ultimate goal of ClusterMe to raise peoples' awareness towards the personalization effect by Google and to provide a stimulus to question one's current behaviour on the Internet. It is important to be aware of such threats when entering the web through the one door with the big G even tough it is so familiar and comfortable. Other approaches can open up new perspectives that are worth a try. ■

ClusterMe

1. Abstract

Web search engines strongly influence the information users get by filtering the search results and even the order in which they are displayed. Google Search is the main search engine provider globally. While the search results are mainly adapted to the localization of the search, the device used and the timing of the search, it is unclear to what extent search results are additionally personalized according to the user's profile. Although this could lead to better search results, some critics fear it could provoke imbalanced information for some search queries. Receiving information that is likely to be weighted in a particular direction can become a problem when it comes to sensitive topics like politics.

The question is, how this bias – if it exists – could be detected and if possible also be measured. To assess this, the ClusterMe team developed the online tool www.cluster-me.com in order to assess to what extent Google search results are personalized. In addition, the differences in search results should be visualized. Furthermore, the tool gives users the possibility to individually assess how strong their own search results differ from those other users get displayed and to find out which information other users receive that they do not.

The analysis of the data collected demonstrates that search results users received differ both in respect to the links displayed and the

order in which they are ranked. Furthermore, for some search queries groups with similar search results that differ from other users can be seen and can thus be interpreted as clusters. However, those different clusters can primarily be attributed to different geographical backgrounds and therefore language preferences of the users.

2. Background

2.1 Google's position on personalized search results

When people are interested in a particular topic they often choose to use the internet as a source of information. As shown in Figure 1, the internet is the second important source of information for Germans apart from asking family, friends or acquaintances and used more often than television or newspapers.[1] For internet searches Google Search is the search engine used by more than 92% worldwide, and in Germany the number is even higher than that.[2,3] Therefore, Google strongly influences the information people get if they are looking for something on the internet, and the results Google selects to display are of great importance.

To obtain search results for a specific search query Google carries out three things: it sweeps the web to find websites, it indexes the sites found by crawling and it ranks the indexed sites for a specific search query.[4] The underlying algorithms of how the scanning, indexing and ranking of the sites is carried out is a corporate secret that makes the web service so useful and valuable.

In 2005 Google announced that the results displayed to users on the search engine result page (SERP) would from there on not be universally identical for a particular search item, but rather personalized to the Google user.[5] By tailoring the ranking of the results to the user's last searches they want to prioritize information Google assumes is more relevant to the user. In 2009 personalized search was expanded to searches that were conducted while users were not signed into their Google accounts.[6]
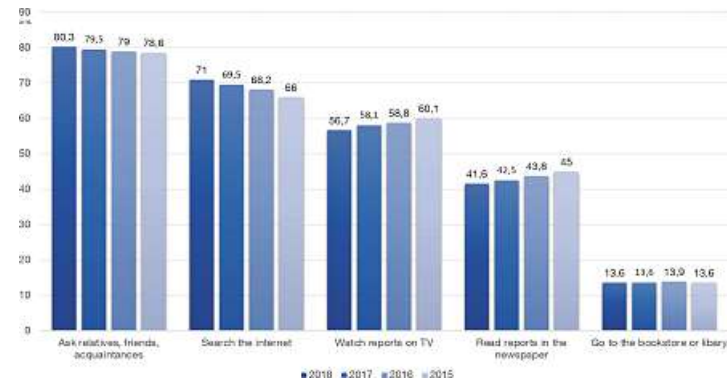

Figure 1: Information sources most used by the population in Germany in 2018 (extract)

In 2011 Google defined personalization as "[…] a special kind of context; it's the context of you. For example, what are you interested in, who do you care about, and what do you search for regularly?"[7]

While being very optimistic about the advantages of personalized search, Google was aware of potentially one-sided information if the results were completely tailored to the user's interest and preferences.

"The science of search is not advanced enough yet to provide a purely personal experience. We aren't confident enough, for example, to say that you're interested in the New York Times and not the Wall Street Journal. However, even if our systems improved so much that we could return only a single source, and it would be the source you like the most, we'd still want to provide a variety of sources and opinions. Our users value diverse viewpoints and serendipitous discovery in search results."[8]

---

1 Statista 2019.
2 Statcounter 2019b.
3 Statcounter 2019a.
4 Google 2019.
5 Google 2005.
6 Google 2009.

7 Google 2011.
8 Google 2011.

However, Google's optimism towards personalization seems to have faded over the years. In 2018 Pandu Nayak – the Vice President of Search – told CNBC that "Right now, there is very little search personalization and what exists is focused on a user's location or immediate context from a prior search."[9] They decided against extensive personalization as "Google has found that it seldom

Most participants saw 10 unique domains   But some saw 9 or 11 domains

Figure 2: Google filter bubble domain variation – Partial results of the DuckDuckGo study

actually improves results."[10] Further he argues, that "a query a user comes with usually has so much context that the opportunity for personalization is just very limited."[11]

2.2 Opposing points of view from the public

In 2011 the publicist and internet activist Eli Pariser claimed in a much-noticed TED talk, "Beware online Filter Bubbles", and in his book, The filter bubble – what the internet is hiding from you, that personalization on the internet is not only ubiquitous but also harmful as it creates filter bubbles. In his talk he defined filter bubbles as "[…] kind of your own personal, unique universe of information that you live in online."[12] He claimed such filter bubbles exist on social media platforms such as Facebook, search engines such as Google Search and also everywhere else on the internet where recommender systems are in use that tailor the information displayed to the individual user and optimize the probability that the user clicks on the results. According to Pariser such filter bubble effects are worrisome as they do not display a well-balanced selection of information, but rather a biased one that is skewed towards the information individual users are are most probably interested in.[13]

Or as Pariser describes it: "Instead of a balanced information diet, you get surrounded by information junk food".[14]

On the other hand, he criticized the fact that the selection process that forms such filter bubbles is not transparent and users are not informed about information that is not displayed to them.[15] "The thing is: you don't decide what gets in and you don't actually see what gets edited out."[16] Pariser's claims were heavily supported by Google rivals, such as the search engine providers DuckDuckGo.com and Startpage.com.[17,18,19]

In 2018 DuckDuckGo published the non-peer-reviewed study "Measuring the 'Filter Bubble': How Google is influencing what you click". In this study the search results of 87 users for the terms "gun control," "immigration," and "vaccination" – three highly debated political topics – were compared. Temporal influences on search results (information relevant to a topic changes over time) were controlled for by asking the participants to run the searches at the same time. Local differences in search results that could be attributed to local relevance of a search query were examined by

 9 D'Onfro 2018.
10 D'Onfro 2018.
11 D'Onfro 2018.
12 Pariser 2011.
13 Pariser 2011.

14 Pariser 2011.
15 Pariser 2011.
16 Pariser 2011.
17 DuckDuckGo 2013.
18 Startpage 2019a.
19 Startpage 2019b.

checking the results for local relatedness. The result of their study is shown in Figure 2. The study was heavily debated and while some were convinced, that the data presented pointed towards a highly relevant problem others criticized the low number of participants and the methodology of the study.[20,21]

87 people searched Google for „gun control" at the same time from across the U.S., logged out and in private browsing mode. They saw 19 domains ordered in 31 ways.

Google strongly disagreed with the interpretation of the results and attributed the differences in search results to the user's "location, language settings, platform and the dynamic nature of search."[22]

In 2017 researchers of the organization AlgorithmWatch conducted a large study to assess the variability of search results users saw during the German Bundestagswahl (German parliamentary elections) when searching for the names of big German parties or famous politicians. In their analysis they did not see strong differences when they controlled for time and location of the search.[23] However, it is unclear how much one can extrapolate these results to other search queries.

2.3 Project objective of ClusterMe
So far there is no strong and unambiguous proof that personalization of search results happens to such an extent that it leads to the formation of filter bubbles. However, if personalization of search results and the subsequent formation of filter bubbles was indeed as strong as suggested by Pariser and others, this could have a massive impact on decision-making of search engine users in general and political decision-making in particular. This is due to the fact, that users trust the relevance-ranking of search engines and that users are strongly influenced by the information that is displayed to them.[24,25]

Also, the question of how often personalization of Google search results occurs, how big the differences between users are and the

Figure 3: Web presence of www.cluster-me.com

resulting consequences on the balance of information users receive is not fully answered. Therefore, the project ClusterMe strives to further elucidate this question. Its goal is to raise awareness of the fact that search results are not static but rather dynamic collections of links, whose ranking is influenced by several different factors. Most of all, it aims to give users the opportunity to compare their search results to those others have gotten and to decide for themselves, whether they receive the well-balanced collection of information they want to see.

3. Goals and methods
3.1 ClusterMe web application
The aim of ClusterMe is to assess the personalization effect of Google through a programmed web application. To increase the awareness of clustering this effect will be displayed as a visualization of the previously explained clusters.

As mentioned before, it is especially young people who get their information from Google searches. In fact, 87% of the adolescents and younger adults in Germany use this search engine several times a week.[26] The links in Google search results are thus the

---

20 DuckDuckGo 2018.
21 Tiku 2018.
22 Google 2018.
23 Spiegel Online 2018.
24 Pan et al. 2007.
25 Epstein and Robertson 2015.

26 Luther 2017.

main source of knowledge for the decision-making process of the younger generation in Germany. Due to this fact, the project focuses on the target group of students of the Technical University of Munich (TUM).

The web tool had been published at the beginning of May in 2019. The web application could easily be reached via the internet site www.cluster-me.com. (Figure 3). Besides the data collection through participation and the display of clusters to its users, the website also pursues the goal of educating its users. It contains explanations on the topic as well as simple tips to avoid filter bubbles. To motivate students to visit the website and use the on-line-tool, ClusterMe accompanied the launch of the website with a kick-off event at the TUM main campus. At this event, a substantial part (60 student) of all website users up to now was recruited. A total of about 130 students took part in the project and thus contributed to sufficiently big data pool for subsequent evaluation within the first three weeks.
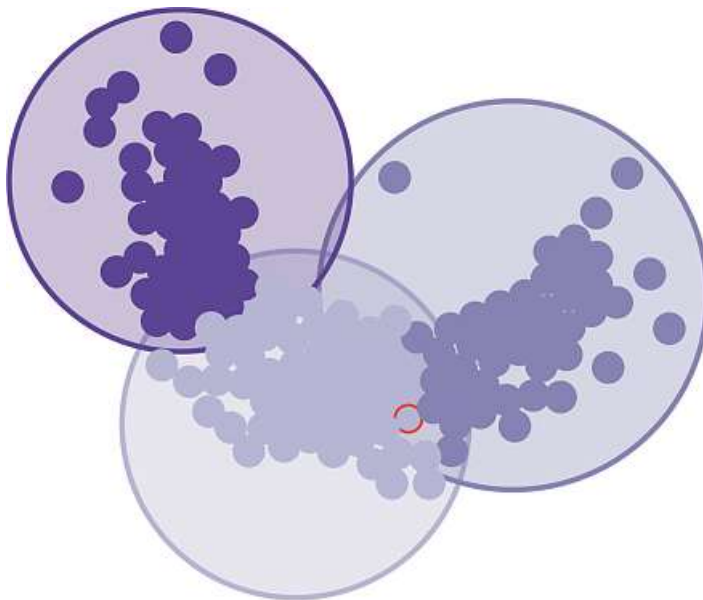
Figure 4: Visualization of the clusters for the search word "Fridays for Future"

The users represented a diverse group in terms of nationality. They came from 25 different countries. German users represented 65% of the participants and the rest of the users are come from China, Turkey, Italy, Egypt, Colombia, Israel, Taiwan, Bosnia, Czech Republic, Macedonia, Luxembourg, Ukraine, Mexico, Peru, Pakistan, Switzerland, Iran, Venezuela, India, Austria, Romania, Laos, Afghanistan and Brazil. The average age of the tool's users was around 26 years. Moreover, most of the users are university students. Also the male to female ratio was skewed, as 70% of the participants were males. Therefore, the sample of users that ClusterMe was able to reach does not represent a broad part of German society. However, it can be viewed as a representative sample of students at the TUM main campus.

3.2 Visualization of Google clusters
For "making clusters visible to everyone" the ClusterMe team developed a clustering tool as a web application. This web application assesses which user gets which results for a certain search query and visualizes the similarity and differences between the users. As seen in Figure 4, in this simulation each user is represented by a single point. Points that are located closely to each other indicate similar search results, while points further away from each other differ in more results. If several users are allocated closely next to each other, they are highlighted as bubbles. Each bubble represents a user-cluster in Google with highly similar results.

The visualization used on the website gave users an impression on how strong the results between users differed. Furthermore, users could click on individual points to see the result page of the respective other user.

Nevertheless, the ClusterMe team concluded that this 2D form of visualization was not well suited for a thorough further analysis of the data. Therefore, the decision was made to not use this format for the final data analysis and rather introduce other forms of analysis that are closely describe in the paragraph 4. Evaluation of the Data.

Hereafter, the steps of the development process of the clustering tool are listed. In addition to the technical development of the tool, the use of the tool in the form of an experiment with 127 participants was also part of the development process:
■ Proof for existing clusters in Google searches
■ Decision on a method to visualize clustering

- Development of a clustering tool
- Small scale testing of the clustering tool
- Refinements following testing and bug fixing
- Selection of interesting/controversial topics for which clustering is visualized as well as relatively neutral words as counter-control
- Preparation of a large scale experiment with approximately 60 participants (TUM students)
- Evaluation of the testing results

3.3 Technical elaboration of the web-tool ClusterMe

3.3.1 Survey and collection of data

With the help of a short questionnaire before using the tool, information about the participant was collected. This was used to estimate how representative the group of users was. Afterwards, the website searches 20 keywords using Google search in the background. Among these search words one can find words on current topics such as "Impfen ja oder nein?" (vaccination yes or no?), "Fridays for Future" or "EU-Wahl" (European elections), but also less discussed words such as "Geburtstag" (birthday) and "TUM" were introduced as counter control. The URLs of the first ten results of each Google search were then stored in a list of a database. Thereby it was possible to collect the necessary data without the user typing every single search query. This database forms the basis for determining and visualizing the results of our analysis as seen before.

3.3.2 Algorithms

The differences in the URL lists of all users form the basis of the clustering bubbles. To determine the differences between them the Levenshtein Distance is used.

- Definition of Levenshtein Distance:
  Minimum number of Paste, Delete, and Replace operations to turn one list of items into another.[27]

To enhance the influence of inserting but also deleting items in contrast to just reordering them, corresponding contributions to the distance for these operations were multiplied with 2.
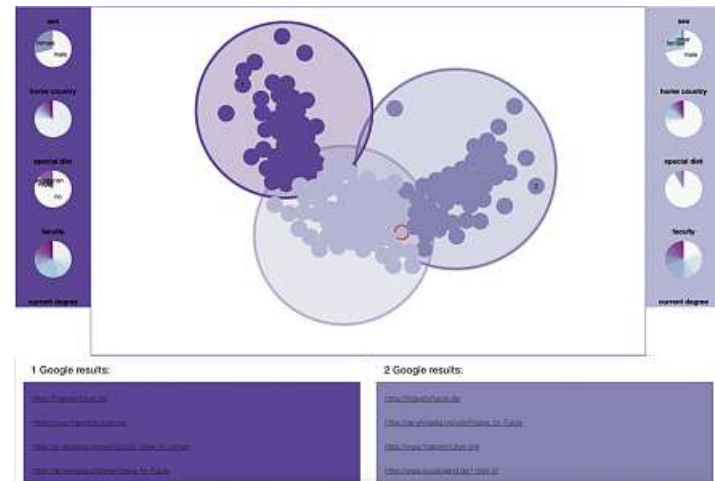
Figure 5: Visualization of a comparison between two clusters and between the results of two users for the search word "Fridays for Future"

- Definitions of clusters
  Mapping each user to points on a two-dimensional plane, so that the distance between two points (users) is approximately proportional to the calculated Levenshtein distance between these two users. This was done using a force simulation, where mutual forces are iteratively applied to the points, and each force is based on the Levenshtein Distance. A two-dimensional representation has been used, as it can be directly displayed to the users as described in Figure 4. For running the force-simulation the Javascript-library d3.js was applied.[28]

As cluster algorithm an iterated k-means algorithm was used, where k is incremented by 1 in each step starting from k=2 until the error loss from k to k+1 relatively to the error in k decreases by less than 10%. For executing the k-means algorithm for a specific value of k the Javascript-library ml-kmeans was applied.[29]

---

27 Centrum für Informations- und Sprachverarbeitung.

28 GitHub.
29 GitHub.

3.3.3 Display of clusters

In the front-end page of the web application a user can inspect the clusters and results of the users for each keyword as follows. By selecting a specific search-keyword, an overview of the users and clusters is displayed: a user is represented by a small point, where the position of that point is the one calculated by the force simulation mentioned above. That means, the further the points – symbolizing users – are apart, the greater the difference in their results. The computed clusters are then shown by a big circle around all points of those users, that are in this cluster. When the user clicks on a point (user), the results (i.e. the first ten URLs) are shown. Clicking on a cluster shows the user statistics for that cluster, which means the aggregated data that had been collected about the users by the questionnaire at the registration using pie diagrams (i.e. age, sex, diet).

4. Outcome and discussion

4.1 Data analysis

After three weeks of data acquisition the data was analyzed using Matlab. Two main factors influencing search results are the time of the search and the location of the user when doing the search. On the first day after launching the tool online a kick-off event at the TUM main campus was organized, where students were recruited to use the tool directly on-site. Therefore, the focus of the analysis is on the data obtained on this first day to minimize effects due to different timing or location of the search.

Further, the analysis was limited to the first five search results, as users are most likely to click on one of those (Figure 7) and changes in the search results at positions further down the site would presumably not have such a relevant impact on the user.[30]

This analysis was conducted by the Search Engine Optimization (SEO) tool provider Sistrix. The evaluation of more than 120.000.000 clicks on Google search result pages demonstrates that search results ranked on top of the result page are much more likely to be clicked on than search results displayed further down.

4.2 Results

4.2.1 Homogeneous groups of results

As an initial analysis, the aim was to assess, whether there are groups of people who get exactly the same search results. For this analysis the order of the results was neglected and only examined whether they were displayed within the first five ranks.

Number and relative size of user groups that get the same search displayed on the ranks 1-5. The order in which the search results were displayed was neglected.

---

30 Beus 2015.

Figure 6: Temporal distribution of the usage of the online tool ClusterMe
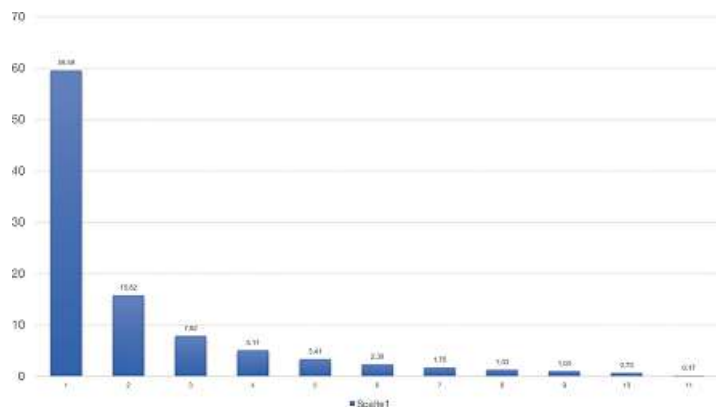
Figure 7: Probability of users clicking on links displayed by Google at different ranking positions

Three main patterns became apparent when taking a look at Figure 8. Although no hard cut-off could be determined, some search queries could be broadly attributed to one of the three patterns while others were placed between them.

(1) For some search queries a large number of users got the same search results, while other users formed several small homogeneous groups. This is the case for search queries such as EU-Wahl, Fridays for Future, Master, HIV or Pizzeria. Especially notable is the result for the search query Atomausstieg as all 60 users received exactly the same search results and form a homogeneous group.

(2) For other search queries several bigger groups with different sizes appeared. This could be observed for Industrie 4.0, Praktikum, Wachstum or Obergrenze.

(3) For the search query TUM almost every user got individual results.

Upon closer inspection the discovery was made, that for TUM, the search result page included a search bar that enabled the user to search directly on the TUM website. This search function itself had an URL that was unique to almost every user. This explains why no bigger homogenous groups could be observed for this keyword and why this search query also displayed a unique pattern in the subsequent analyses.

4.2.2 Differences between homogeneous groups
The comparison of homogeneous groups of users already indicates that there were subsets of users that received unique results, but it was unclear how much those groups differed from one another. To address this question the number of different URLs occurring by

Figure 8: Uniform groups of users with exactly the same search results.
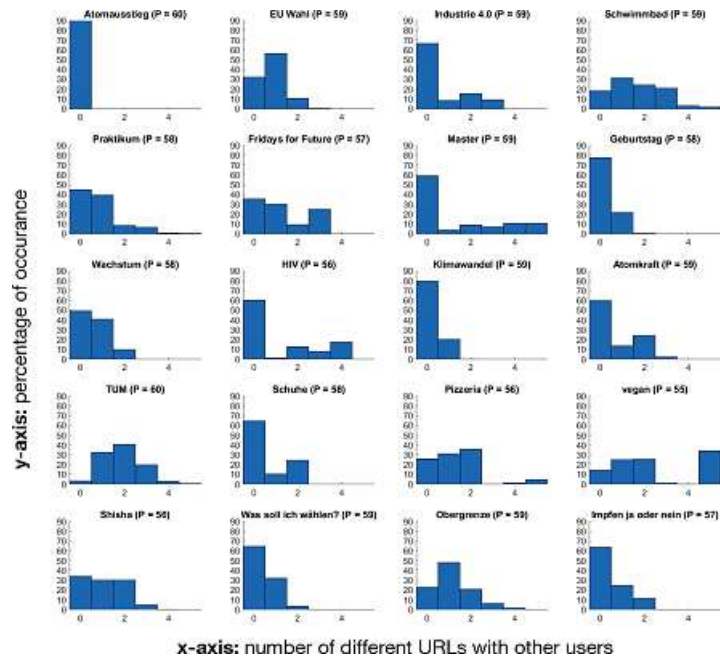
Figure 9: How many times did search results differ in a specific number of URLs?

comparing the results of all users against each other. Subsequently it was counted how many times a certain number of different URLs occurred (Figure 9). Number of different URLs among the results of all users compared against each was calculated and the number of times a certain difference occurred was counted. The order in which the search results were displayed was neglected.

It is reasoned, that if the search results for a specific query that users received could be grouped in homogeneous groups that differed from one another we should observe a high number of cases where search results differed by 0 URLs and additionally a high number of cases where the URLs differed by a specific value. Such a V-shaped distribution could be interpreted as "search-result clustering."

On the other hand, if the results were more heterogeneous between users and no clear clusters formed but everyone differed from everyone else to a certain degree instead, more cases where the number of different URLs was above 1 should be observed, but no clear peak should be visible. Interestingly the V-shaped pattern emerged for several search queries: Industrie 4.0, Fridays for Future, Master, HIV, Atomkraft, Schuhe and vegan. This could be interpreted as a hint towards the existence of clusters of search results that are in themselves homogeneous, but differed from each other.

However, to assess whether clustering of search results actually exists, a visualization of the differences between every user and every other user in a distance heatmap as shown in Figure 10.

Figure 10: Distance heatmap visualizing the differences between the results of every user compared to every other user.
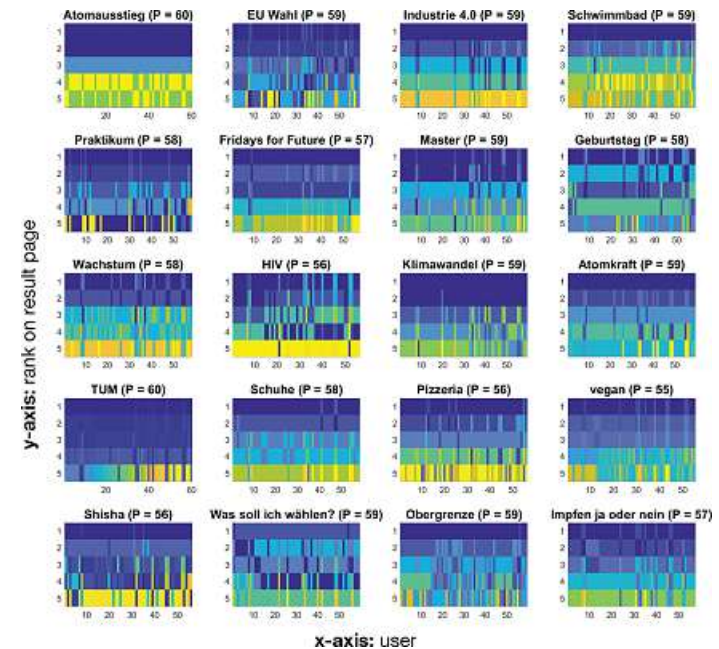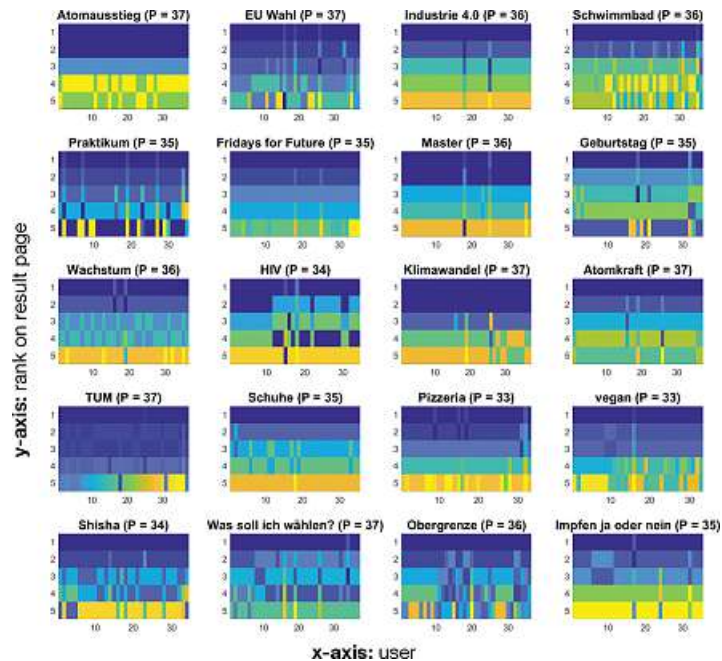

Figure 11: Distance heatmap considering the ranking of search results

The number of different URLs that users get displayed within their first 5 search results are compared. The number of URL mismatches (ranging from 0 to 5) is visualized by color (dark blue to yellow). The order in which the search results were displayed was neglected.

This distance heatmap demonstrates at the same time the formation of homogenous groups (dark blue squares) and enables the assessment of how strong those groups differ from one another (colour of the overlap with other groups). Asymmetries in the heatmap arise if the search results of one of the users contains the same link several times.

It becomes apparent that some search queries lead to "Search Result Clusters" that in themselves are homogeneous, but differ to some extent compared to other "Search Result Clusters." The difference between big clusters is mainly small (1 different URL) and big differences are almost exclusively observed between small clusters and the other users, which speaks against the formation of big "search result clusters" and rather indicates that these users were outliers. However, for the search queries vegan, Master, Atomkraft, Industrie 4.0 and Schuhe differences of two URLs out of the first five URLs could be observed between reasonably sized clusters.

4.2.3 Consideration of the ranking of the search results
In the previous analysis the order in which the search results were displayed was neglected. However, given that the probability a user clicks on a search result link is largely dependent on the rank of the search result, which means how high on the website the result is displayed (Figure 7) it could have an impact on the information users receive.[31] Previous research demonstrates that this effect is not only a result of Google's ability to rank the most relevant links on top of the result page. Instead participants trusted the ranking to represent relevance and even clicked on the top search results after the ranking was modified and links less relevant to the query were placed on top.[32] This emphasizes that it is not only relevant which results are displayed to a user, but also in which order they are ranked. Differences in ranking between users could therefore have a strong impact on the information they get after searching for a specific topic.

To assess this question, we compared the ranking of search results users received after searching for the different queries as seen in Figure 11.

Tile plot visualizing the ranking of the search results that were displayed to different users when they searched for the indicated search query. Each color represents one URL. Similarity of individual colors does not represent similarity of individual URLs.

The visualization of the individual rankings emphasizes that the results users get displayed not only differ in the URLs displayed (as indicated in Figure 8,9,10) but also strongly vary with respect to the order in which they are displayed (Figure 11). While the first ranked result is rather similar in most cases, variation of ranking increases strongly in the lower ranks. The appearance of unique results that only one individual or a homogenous group gets displayed within the first 5 results is primarily restricted to the fourth and fifth rank.
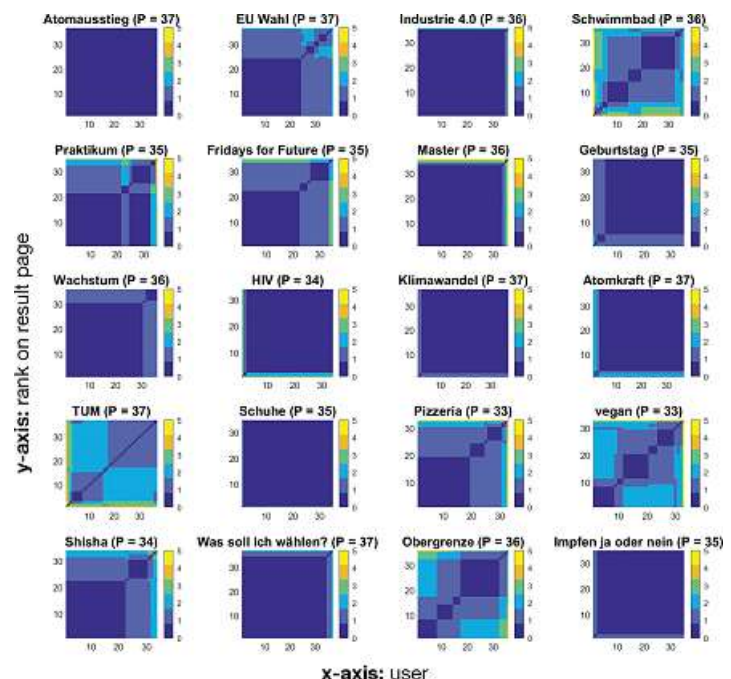
Figure 12: Distance heatmap visualizing the differences between the results depicting only German participants

4.2.4 Differences observed among German students

After a detailed review of the tile-plots visualizing individual results (Figure 8), the ClusterMe team suggests that some users seemed to have unique results differing from the majority of other users for several different search queries. These persons could be outliers. A reason might be that these users had a different home country than the German majority and therefore might receive results in a different language. Thus, it was assessed how strong the observed clustering effects could also be observed within the majority-subgroup of German users (Figure 12,13). In fact, the previously observed clustering effects were less evident within the German user group. In most cases most users got the same results and in the cases when clusters were formed they mainly differed only by one search result. However, for the search terms vegan and Obergrenze there still were clusters with homogenous results of reasonable size that differed in two out of five search results, which were regarded as a relevant number.

The number of different URLs that users get displayed within their first 5 search results are compared. The number of URL mismatches (ranging from 0 to 5) is visualized by color (dark blue to yellow). The order in which the search results were displayed was neglected.

Tile plot visualizing the ranking of the search results that were displayed to different users when they searched for the indicated search query. Each color represents one URL. Similarity of individual colors does not represent similarity of individual URLs.

Nevertheless, the fact that clustering effects mainly disappeared after selecting all German participants points toward the fact that these effects were mainly attributed to home country related differences or another factor such as language preferences that correlate with home countries. These data speak against strong search engine clustering and the subsequent formation of "Filter Bubbles".

4.3 Illustration of the results using the example "Impfen ja oder nein?"

It was reasoned that the ranking of search queries could only have a significant effect for the information balance of users, if search results with different messages or information were ranked differently. Then, the probability of clicking the result ranked on the top and therefore receiving this information should be higher than receiving a potentially contradicting view that is displayed further down the results page.

Surprisingly, such a case has been discovered while analysing the specific URLs displayed to users. This was the case for the search
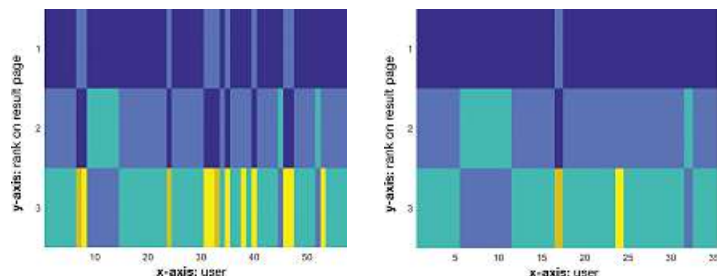
Figure 13: Distance heatmap considering the ranking of search results depicting only German participants

Figure 14: Ranking of the first three search results for the search query "Impfen ja oder nein?"

query "Impfen ja oder nein?" ("vaccination yes or no?") – a polarizing topic.

When searching for this query, users got mainly the same results with no significant differences in the URLs displayed (Figure 9,10,11). In fact, in the previous analysis it was one of the most inconspicuous examples. However, the messages of the search results displayed on the top ranks massively varied: about 82% of users got the URL www.impfen-nein-danke.de and about 17% got the URL www.krankenkassenzentrale.de on the first rank (Figure 14).

Tile plot visualizing the ranking of the search results that were displayed to different users when they searched for "Impfen ja oder nein?". Each color represents one URL. Similarity of individual colors does not represent similarity of individual URLs.

While www.impfen-nein-danke.de strongly opposes vaccination in a very emotion-driven manner, www.krankenkassenzentrale.de emphasizes the health benefits of this life-saving prevention method. Yet, as we performed the same analysis only with users that stated Germany as their home country, most users saw the same result in the first position.

Figure 14. Ranking of the first three search results for the search query "Impfen ja oder nein?" depicting only German participants Tile plot visualizing the ranking of the search results that were displayed to different users when they searched for "Impfen ja oder nein?". Each color represents one URL. Similarity of individual colors does not represent similarity of individual URLs.

Nevertheless, this case study demonstrates that a user searching for health-related information could end up with very different advice depending on the ranking of his search results, even though the URLs displayed to all users are very uniform in total.

4.4 Discussion
Of course, there are several limitations which have to be kept in mind when interpreting both, the data and the results. Given the low number of users analysed for this study, the temporal and local dispersion of searches and the strong focus on TUM students, we can not exactly assess the degree of personalization of web results that can not be attributed to timing and location of the search. However, the results show that there are in fact differences between

the results different users get. Depending on the search query users can be grouped in homogeneous groups that differ from one another. This can be interpreted as evidence for the existence of "Search Engine Clustering" even though the differences are subtle and the claim of the existence of "Filter Bubbles" can neither be confirmed nor contradicted. However, when we restricted our analysis to German users and thereby limited language influences on the results, differences between users greatly diminished and the results are much more homogeneous. This also speaks against the real-life relevance of search engine personalization and the formation of filter bubbles.

Nevertheless, we presented a case study for the search query "Impfen ja oder nein?", that demonstrated that the different rankings of search results can indeed shift the impression a user gets even though all URLs are displayed.

Finally, the fact that users could compare their own results with those of other users met with great interest and enabled individual users to assess on their own how representative their received search results were.

5. Summary and future goals
The web-tool based approach that was used to assess the extent of search engine personalization could estimate the effect, that can now be assumed as fairly low. However, the approach was restricted in several ways that might have prevented the project from discovering "Search Engine Personalization" and potentially subsequent "Filter Bubble Formation". This is mainly due to the fact that the ClusterMe team could only recruit a fairly low number of participants to use the tool within a sufficiently short period of time. Furthermore, the recruitment strategy was targeted towards students. Therefore, differences between search results TUM students received and results other persons of a different age or social-cultural background stay hidden from our view so far.

Further research should aim to target a bigger pool of more diverse users who are more spread across generations, equal male/female representations, and professions. This could be achieved with a different marketing strategy and further improvement of the website-interface that makes it easier to use. Moreover, developing the tool into a mobile application will increase its spread especially across younger generations. In general, making the tool available

to run over mobile phones will give a huge boost to its spread. Furthermore, improvements of the current visualization used on the website could enable the user to more intuitively understand how strong differences between users are. Additionally, the form of analysis that was eventually used to interpret the data should be implemented on the website. Finally, more search terms should be added to the tool in order to understand which categories of words are more affected by the clustering algorithm. This could then help spreading awareness among Google users to be more critical about their search results in these categories.

Together these steps could enable ClusterMe to get a representative data sample which will help us to better understand if and under which circumstances "Search Engine Clusters" form. ◼

Bibliography

– Beus, Julian (2015): Klickwahrscheinlichkeiten in den Google SERPs. Sistrix. Available online at https://www.sistrix.de/news/klickwahrscheinlichkeiten-in-den-google-serps/, checked 5/29/2019.

– Centrum für Informations- und Sprachverarbeitung: Minimale Editierdistanz nach Levenshtein. Ludwig-Maximilians-Universität, München. CIS. Available online at https://www.cis.uni-muenchen.de/~micha/praesentationen/rechtschreib-korrektur/Levenshtein.html, checked on 5/26/2019.

– D'Onfro, Jillian (2018): We sat in on an internal Google meeting where they talked about changing the search algorithm. CNBC. Available online at https://www.msn.com/en-us/money/markets/we-sat-in-on-an-internal-google-meeting-where-they-talked-about-changing-the-search-algorithm/ar-BBNse4y, updated on 9/17/2018, checked on 5/16/2019.

– DuckDuckGo (2013): There are no "regular results" on Google anymore. Film. Available online at https://vimeo.com/channels/duckduckgo/51181384, checked on 5/26/2019.

– DuckDuckGo (2018): Measuring the "Filter Bubble": How Google is influencing what you click. Available online at https://spreadprivacy.com/google-filter-bubble-study/, updated on 12/4/2018, checked on 5/26/2019.

– Epstein, Robert; Robertson, Ronald E. (2015): The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. In National Academy of Sciences. Available online at https://www.pnas.org/content/112/33/E4512/tab-article-info, checked on 5/26/2019.

– GitHub: Force-directed graph layout using velocity Verlet integration. Available online at https://github.com/d3/d3-force, checked on 5/26/2019.

– GitHub: K-Means clustering. Available online at https://github.com/mljs/kmeans.

– Google (2005): Personalized Search Graduates from Google Labs. Available online at http://googlepress.blogspot.com/2005/11/personalized-search-graduates-from_10.html, updated on 11/10/2005, checked on 2/22/2019.

– Google (2009): Personalized Search for everyone. Available online at https://googleblog.blogspot.com/2009/12/personalized-search-for-everyone.html, updated on 12/4/2009, checked on 2/22/2019.

– Google (2011): Some thoughts on personalization. Available online at https://search.googleblog.com/2011/11/some-thoughts-on-personalization.html, updated on 11/23/2011, checked on 2/22/2019.

– Google (2018): Statement on the Study "Measuring the 'Filter Bubble'". Twitter. Available online at https://twitter.com/searchliaison/status/1070027261376491520, checked on 5/26/2019.

– Google (2019): Wie funktioniert die Google-Suche? Available online at https://www.google.com/intl/de/search/howsearchworks/, checked on 2/22/2019.

– Luther, Lisa (2017): JIM-Studie 2016: Google ist Informationsquelle Nummer 1. Medienbewusst.de. Available online at https://medienbewusst.de/internet/20170406/jim-studie-2016-google-ist-informationsquelle-nummer-1.html, checked on 5/26/2019.

– Pan, Bing; Hembrooke, Helene; Joachims, Thorsten; Lorigo, Lori; Gay, Geri; Granka, Laura (2007): In Google We Trust: Users' Decisions on Rank, Position, and Relevance. Available online at https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1083-6101.2007.00351.x, checked on 5/26/2019.

– Pariser, Eli (2011): Beware online Filter Bubbles. Film: TED. Available online at https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles?language=de, checked on 3/27/2019.

– Spiegel Online (2018): Sehe ich andere Suchergebnisse als meine Nachbarn? Googeln nach Politikern und Parteien. Available online at https://www.spiegel.de/netzwelt/netzpolitik/google-projekt-von-algorithmwatch-filterblase-welche-filterblase-a-1219981.html, checked on 5/26/2019.

– Startpage (2019a): Warum erhalte ich über Startpage teilweise andere Suchergebnisse als bei Google? Available online at https://support.startpage.com/index.php?/de/Knowledgebase/Article/View/601/0/warum-erhalte-ich-ber-startpage-teilweise-andere-suchergebnisse-als-bei-google, updated on 2/5/2019, checked on 5/27/2019.

– Startpage (2019b): Wie du mit Startpage.com der „Filterblase" entkommst. Available online at https://support.startpage.com/index.php?/de/Knowledgebase/Article/View/1039/0/wie-du-mit-startpagecom-der-filterblase-entkommst, updated on 2/5/2019, checked on 5/27/2019.

– Statcounter (2019a): Search Engine Market Share Germany. Apr 2009 - Apr 2019. Available online at http://gs.statcounter.com/search-engine-market-share/all/germany/#monthly-200904-201904, checked on 5/16/2019.

– Statcounter (2019b): Search Engine Market Share Worldwide. Apr 2018 - Apr 2019. Available online at http://gs.statcounter.com/search-engine-market-share#monthly-200901-201901, checked on 5/16/2019.

– Statista (2019): Information sources most used by the population in Germany in 2018. Available online at https://www.statista.com/statistics/966748/information-sources-most-used-by-the-population-in-germany/, checked on 5/16/2019.

– Tiku, Nitisha (2018): Study revives debate about Google's role in filter bubbles. In WIRED April 2018. Available online at https://www.wired.com/story/study-revives-debate-about-googles-filter-bubbles/, checked on 5/26/2019.

Self Reflection

Looking back on the past two years as scholarship holders of the TUM: Junge Akademie it can be said that our project was an elaborate but nevertheless very instructive experience. In addition to the purely scientific work, the project became some kind of crash course in terms of project management and time coordination for all of us. During the progress of the project, unexpected obstacles came up. Nevertheless, we are more than satisfied with the outcome of our project ClusterMe and can proudly look back on our time at TUM: Junge Akademie.

Right from the start, we were a comparatively large project group with 13 scholarship holders, which shaped and co-determined the way we worked. At the same time, we created an enormously diverse group with members from various faculties and countries of origin. Thus the project profited from the great manpower and tasks could easily be distributed among several participants. At the beginning, however, it turned out to be a bit tricky to bring all the ideas, the different mind-sets and the different views to a common denominator. Therefore, it was particularly important to open up and talk about different view points in order to avoid problems arising in the first place. We also needed to find a way to keep all members up to date and involved in the decision-making process. We agreed to make short videos at the end of each meeting, in addition to taking standard minutes, that would summarize the things and tasks discussed. This idea turned out to be a good way to keep everyone, including the members who could not be taking part at meetings, on the same page. Ultimately, with many discussions and visualizations we managed to get our ideas on a mutually agreed path in a surprisingly short time. The enrichment by the diversity of the members was in the end greater than the challenge.

After we had found a common denominator, it was time to structure the teamwork. In order to avoid a hierarchy within the group, the different tasks varied each week. The role of the project spokesperson changed each time during the weekly meetings. For the coordination with the mentors as well as with the office of the TUM: Junge Akadmie, however, there were fixed deputies of the group. In order to maintain an overview of the numerous deadlines and different accessibility of the individual members, a group calendar was set up. This helped distributing the tasks throughout the entire working process and planning ahead. In the course of time, 13 different students developed into a well-rehearsed team.

After the topic – the ClusterMe web tool – was quickly determined, most of the work consisted in programming the tool. In addition, we needed a good marketing strategy in order to spread our message and encourage as many students as possible to participate. Therefore, we divided our group into two smaller subgroups – one taking care of the technical aspects of the project, the other one concentrating on marketing and design strategy. Trough the division of the group, the project benefited from the diversity and individuality of each of the members. Even within the subgroups, every member of the team was able to get involved in their area of expertise and thus contribute to the launch of the website.

In addition to the weekly meetings, it was above all the seminar weeks that drove the project forward. We not only had plenty of time for the project work itself, but also benefited from the lectures and seminars given by external specialists. The exchange with other people also helped to change our own perspective on the project and to rethink approaches. We were also able to support each other in discussions with members of other groups, who often faced similar challenges.

But apart from all the hard work, of course the fun was not to be neglected. Therefore, small internal team events and excursions were organized, which kept the motivation high and helped us to put the project aside for some time. Especially at times when we were

struggling with the progress of the project, such activities helped to return to the same enthusiasm and energy we had experienced at the beginning of the project.

In the meantime, some of us are now scattered in different cities of Germany. Nevertheless, we hope that our project ClusterMe will be able to make an impact by introducing the website. We are proud to have shown some people that there is more than one way of searching the internet. Despite minor challenges and setbacks, working together was an incredibly instructive and exciting experience that we would not have wanted to miss.

Acknowledgement

Of course, we would not have been able to set up our project ClusterMe on our own. Therfore, we would like to offer a special thanks to our mentors – Dr. Hannes Petermeier and Prof. Dr. Jürgen Scheurle - as well as to our tutors - Thomas Bickel Haase, Michael

Vetter and Phillip Hulm - who supported us throughout the way. Not only did they give valuable support and assisted when the working progress temporarily came to a standstill. They also helped finding a good structure to the regular meetings and redirected the focus on the essential topics whenever needed. They provided us with their experience in technical matters as well as in organization and planning.

In addition, many thanks go to the TUM: Junge Akademie, who gave us a direction on the long way and still entrusted us with the necessary freedom. Thank you, for helping us during the time-consuming process of the conclusion of the contract with Heroku. Last but not least, we would like to thank all those who participated in our project and thus brought it to a successful conclusion.

The ClusterMe Team

ClusterMe

150 Jahre culture of excellence

TUM

ABSTRACT

Bias are part of everyday life and necessary to simplify daily tasks but sometimes they can also lead us unconsciously into wrong directions. Among this the bias from digital world is one of the major concern of modern society. So, our aim is to create a tool to quantify and visualize our digital bias at the example of google result clusters. Secondly, the outcome of our tool shall be used to increase awareness about the bias among the users.
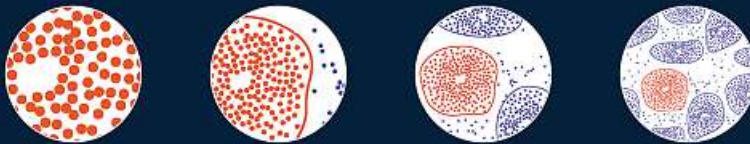
GOAL

ClusterMe aims at clustering search-engine results and develop a web-tool that enables users to be aware of the degree their results are personalized.

METHODOLOGY

- Creating a tool to gather information about users
- Clustering search results of users into groups
- Finding which trait (sex, nationality, etc.) has the highest effect on the clustering process
- Using these results to make people aware about clustering

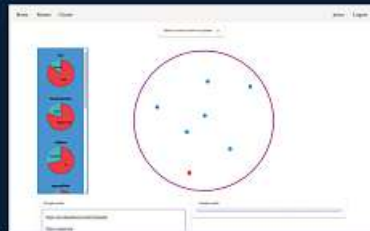## HYPOTHESIS

Digital bias effect on human objective behavior.

MAY 2018

MEMBERS: Marina Able, Julian Albers, Anna-Lena Fackler, Daniel Hernández, Kyra Kleine, Tilman Knopp, Mohammad Youssef Mahfouz, Sebastian Mair, Rishith Eilath Meethal, Janna Nikonov, Lea Sophie Seier, Nicola Stadler, Julian Trummer
TUTORS: Philipp Hulm, Thomas Bickel Haase, Michael Vetter
MENTORS: Prof. Dr. Jürgen Scheurle, Dr. Hannes Petermeier

inspired by
TUM: Junge Akademie

POSTER 1: The first weeks of our 20-month journey at the TUM: Junge Akademie were filled with wild discussions about the very broad topic "Truth and Lies." From time to time concrete projects were thought of but no agreement could be achieved due to the size of the group and the great variety of ideas and goals. However, we came to the conclusion that the topic can be found in almost every aspect of life. To find relevant material on it, most of us started with a simple Internet query, and here the first problem arose. With the same Google search word, for example "truth," we observed big differences between the results of our team members. This was the way the vision for the project was born: the vision of creating a tool to make people aware of the bias that search engine queries can generate. A part of that vision that we soon agreed on was the idea of particularly visualizing the clustering effect. A common perception emerged in the team of visualizing those clusters as clouds of dots with different distances between them to represent the variability of Google search results. Therefore, this sort of visualization of the clusters was already depicted in our first poster.

But we also faced two major challenges. The first challenge was to figure out the best way of implementing those ideas through an appropriately programmed web tool. Associated with this was the task of finding out which trait (sex, nationality, etc.) has the highest effect on the clustering process, and this was a problem that we would later discuss a lot with members of other groups and especially our Mentor, Prof. Scheurle. Our vision of finding out more about Google's search engine algorithm was constrained by the fact that our methodology was only descriptive. So, while exploring the correlations between clustering and the user's traits seemed achievable, causalities would be very hard to talk about. Secondly, organizing a subproject, focusing on working with school kids using the tool, was considered. But as we obtained more information on how difficult it is to organize such an event that includes the collaboration with schools, we decided to set our focus on university students at TUM.

POSTER 2: By the time the second poster was designed, we had already made great advances in our project, and this progress can be seen on Poster 2. We had focused on the essentials and created a first version of our online tool that was able to visualize the clusters in which users are trapped. As can be seen, the methodology is already much more detailed and so are the images that show the visualization of the clustering. Also the description of how the clusters are generated is presented at the end. There were still some challenges to face concerning programming the tool itself, but everyone agreed on the output. Further, the part about investigating the backgrounds of the clustering became more concrete and realistic – we now wanted to find out about the "most dominant trait in each cluster" which is an exclusively descriptive proceeding. At that time, we already had quite a clear image of how we wanted to show the results to the user also.

Again, two major issues were to be solved. We had to decide on the exact search words that the tool should cover. We agreed that they should be interesting for the users, but also polarizing so we could actually see differences between the search results. At the same time, we needed some "negative controls" from which we expected to give quite homogeneous results to all members of our collective TUM students. The second issue was to manage the marketing of the whole project, since the tool would only show relevant data if enough students used the tool to genuinely create the clusters. To deal with negative effects on the neutrality of data collection like differences of time and location, we had to coordinate the dates of the marketing events with the finalization of the tool. Our discussions on that led to a quite concrete timeline that can also be seen on the poster.

POSTER 3: In order to work as efficiently as possible, we decided to split up the group into two subteams – one taking care of the technical issues of the tool, the other one working on a marketing strategy and organizing an event for the launch of the tool. This division of the group can clearly be seen on the third poster where we show the tool itself on the left side and marketing ideas on the opposite side.

At that point the ClusterMe tool was already ready to operate, with some final alterations left to implement. We had already had a few test rounds with friends and family and we were able to put small inconsistencies aside. Also, some legal issues remained to be solved. The image on the right side represents the efforts of the marketing subgroup, which mainly consisted of designing advertisements for the tool and planning an event on the TUM main campus where we would motivate students to use the tool right away. In addition to that event, a poster and a flyer were designed and distributed at the main campus, at Garching campus and at nearby student facilities. So, our plan was very concrete at that point and only few things were left to discuss. ◼

POSTER 4: After solving the last legal issues, we were finally ready to launch the website in May 2019. This was accompanied by a kick-off event we organised at the TUM main campus. There, we presented our project to TUM students and motivated them to participate. Students could win sweets and TUM products by directly using the tool. Additionally, we distributed posters and flyers at sites all over Munich and used social media to reach as many students as possible.

After the website had been online for three weeks, we started on analysing the data. We discussed a lot about how to properly display the data and interpret the results. This was probably the most interesting and enjoyable phase of the project.

As a result of the analysis we developed several graphs and visualizations, such as the displayed heatmaps, which illustrate the results in a powerful and vivid way. Although we were ultimately unable to prove the existence of the cluster or filter bubbles for a variety of reasons, the graphs show that there are indeed differences in the display of search results and that these can have an enormous influence on our decisions.