Finding Reliable Information on the Web Should and Can Still Be Improved

Mathias Glatz[1], Hermann Maurer[1] and Muhammad Tanvir Azfal[2]

[1]Graz University of Technology, Austria
[2]Capital University of Science and Technology, Pakistan

This paper addresses two major weaknesses of locating rather specific information on the Web. First, to find information for a specific topic is still quite difficult. Second, if located, the degree of reliability of the information is not clear. We explain that much progress has been made concerning the first issue, yet the real issue is never explicitly mentioned: the interaction between a user and the search engine has to be good enough so that the search engine really knows what the user wants. We will discuss a new approach to solve this. Concerning the second aspect, we will show that no serious large-scale attempts have been made to help users judge the reliability of the information found. We propose a set of measures that would change the situation dramatically.

ACM CCS (2012) Classification: Human-centered computing → Interaction design → Interaction design theory, concepts and paradigms

Keywords: search engine; user interaction; information presentation

1. Introduction

The unabated growth of the number of Websites in the last 20 years (e.g. from about 2.4 million in 1998 to more than 860 million in 2015 [8]) has been a challenge to all search engines. Even if all servers are crawled and terms looked for are found in many places, how can one rank the pages found so that likely candidates appear early on the resulting list? The idea to define a PageRank by mainly considering the number of links to and from a page, as introduced by Lawrence Page and Sergey Brin in 1998 [1], was the basis for early Google searches. By now many additional factors are being considered when the PageRank is calculated by Google. A fairly detailed description of the developments of search engines and some of the material presented in the sequel is found in Access to Knowledge on the Web [4] (and can be freely accessed at the URL listed in References), hence the discussion will not be repeated here. Due to the dominance of Google as search engine, with 90.62 % of all searches world-wide (`http://gs.statcounter.com/search-engine-market-share`), we will restrict attention to Google which is the leading search technology except in some small niches.

In the following section, we will discuss some aspects of Google searches and propose an alternative. Section 3 will discuss the issue of reliability. Section 4 reports on some experiments. For a long list of references consider the Bibliography in [4].

2. Searching with Google

Historically, searching with Google meant entering a search term and Google would respond with a long list of URLs of pages with some relevance to the search term, the order of entries defined by PageRank and modified by many other parameters. There are several companies trying to analyze Google's ranking factors by implementing empirical tests to offer search engine optimization for customers. A lot of websites offer freely accessible lists of ranking factors such as [5] and [3]. Despite contin-

uing improvement of the ranking mechanism, a linear ordering of results does not really make sense, not surprisingly stated [6]:

Different persons are likely to look for different aspects associated with the same search term.

This is not just due to the fact that one word or search term can mean completely different things. In [4] the example "Galaxy" is used: It can be interpreted in the context of astronomy, in the context of smartphones, or in the context of names of clubs, etc. Much more widely spread than ambiguous terms are terms that relate to many different facts. Searching for "energy" does not make sense if the search engine does not know whether the user wants a definition, or types of energy, the ways how to generate energy in a certain way, effects on climate, etc.

Hence our main credo, and the point we want to get across in this section is that a dialogue between user and search engine has to take place. Only in this way can the search engine really know what the user wants.

For the authors, it comes as a surprise that Google has reacted to this obvious fact very late, yet it has slowly reacted to it in a number

energy **radio**	energy **examples**
energy **physics**	energy **definition science**
energy **types**	**kinetic** energy
what is energy **in science**	**potential** energy

Figure 1. Related searches for "energy" on Google.

of ways. First, the SERP (Search Engine Result Page) has been improved by offering a "Knowledge Graph", a knowledge base containing information about a topic gathered from various sources such as Freebase and Wikipedia [7]; by offering alternatives at least at the end of a page; by having a section "People also ask" and/or "People also search for".

If you try out the query "energy", you can see a strong deviation from just a long list of URLs. Instead, at the end of the page you find a list of a suggestions as seen in Figure 1.

In some cases many alternatives are offered before the "real" search starts. For instance, the search "Schwarzenegger" on May 1, 2018 provided also quite a bit of topical material on the very first result page. After all, when searching, one would traditionally expects a list of URLs of sites dealing with the topic. In this particular

Figure 2. Searches for "Schwarzenegger" on Google.

case, a short general contribution is shown, plus pictures, plus reviews of movies by the actor, and a few outstanding contributions, but not at all a systematic list of entries from different sources.

However, a big step forward in the direction we believe in is that Google is using better and better natural language processing methods, so there is no need to type one or two words, but one can be much more specific by inputting longer and more complex queries. For instance, a query like "dangers of coal power versus nuclear power", produces quite good results. Yet, as long as language understanding is far from perfect, we will still have erroneous hits. For example, "provide arguments why giving up on nuclear energy is a mistake" includes a pointer to an entry named "10 Reasons to Oppose Nuclear Energy".

The real question is:

Are there alternatives to allowing more and more complex natural language queries.

We believe that there is one approach that may be worth pursuing: Use a step by step simple dialogue between user and search engine, and allow the search engine multiple times to suggest alternatives (based on preliminary searches). Note that this will work particularly well if either language understanding to understand what a site report has to offer becomes better, or if information providers build site-maps according to the defined rules that can be understood by the search engine.

We have conducted a few experiments in this direction, but it would require a major effort that our group is not capable of, to show whether this approach works or not.

To show what we mean, here is a conceivable part of a dialogue, as shown in Figure 3, where the regular font means user action, italics means search-engine action, while the text in brackets is there just for explanation purposes.

It is interesting to note that the Google query "History of religion in Graz between 1600 and 1680" gives a number of results, none of them fitting well, although one 114 page document has a few references to the Thirty Years War (religious war). Whether a superb language analysis could have picked out only this document and even mentioned relevant pages is not clear.

Summarizing, search engines will get continuously better with better natural language understanding. Without conclusive proof, but by conducting a number of tests, we believe that a dialogue between user and search engine will be more efficient and/or successful, than a single user query, no matter how complexly one is allowed to formulate it.

3. Reliability

Once a result has been delivered by a search engine, how can one judge its reliability? There is a surprising set of methods that could be used, but they are not available or used much at the moment.

3.1. Quality of Servers

How come we have numeric ratings of hotels or restaurants, but none for servers? It is reasonable to assume that search engines have ratings, yet they are not accessible to the public. Our challenge: introduce an App that allows, for any URL, to retrieve its reliability rating (say between 0 and 20 points) and also allows users to contribute to the rating by reducing or increasing it a bit.

Graz.

What do you want to know about Graz? (The engine offers some 6-12 alternatives among them "History")

History.

History since 1918 or before?

History between 1600 and 1680. (At this point the search engine works away identifying contributions that seem to deal with this period).

Interested in science, economy, politics, or request other alternative.

Religion.

Ok, wait a moment. (And then the search engine comes back with a number of pointers)

Figure 3. Illustration of dialogue.

3.2. Source

Whenever statements are made, the source must be mentioned. When e.g. Google displays a Knowledge Graph, it should be stated clearly where the information comes from.

3.3. Answers Must Come with Definitions

When an answer is provided, it should be made clear what definition it is based on. This reasonable rule is usually totally ignored. Here are some examples:

You search for the number of Austrian Nobel prize winners. You will get an answer – but what does it mean? The number of such winners born in Austria (and what if where they were born in a place that is not in Austria any more?), or the number of winners who got the award while living in Austria, or those having done the work for which they got the award in Austria, but getting the award itself much later while living in another country, etc?

What does is mean that the Mulu caves on Borneo are the largest in SE Asia? Largest in volume, largest in length of corridors, largest in height difference, etc?

How many persons were killed by the Chernobyl nuclear accident? Do you mean directly due to massive radiation (then the answer is 32), within the next year (how do you know some would not have died anyway), how many more will still develop deadly cancer due to radiation received (nobody knows), etc?

How big is a country? Do you include or exclude lakes or glaciers? (Austria includes, France excludes anything more than one square-kilometer); the mouth of big rivers – how much is river (part of the country), how much of it is already part of the ocean; when do you measure: at low tide or high tide?

The list can be continued arbitrarily. But the fact is that answers do not mean anything without knowing what they are based on. And, really, we want more: we want a number or fact from the search engine based on some rules, and maybe answers to the same question based on other rules. Like: When I ask for the population of Nigeria, I may get 38 million (1950). Maybe I want the figure in 2015, a stunning 180 million; or the forecast for 2050, an incredible 440 million (more people than all of Europe).

3.4. Is an Answer Correct or Incorrect, How Can You Tell?

There are statistics revealing that search engines have the highest ratings for trust [2], but the answer is: You can't trust them either. But you have a bit of help if it is a server with high reliability rating. Better yet: if the answer can easily be checked (like a figure) why does the search engine not tell you: "By server A the figure is R, by server B it is S, by server C it is T, etc."?

Main conclusion: Let us put pressure on information providers and on search engines to make sure that the problems in 3.1 – 3.4 are prevented.

4. Experiments

In our group we have looked at some of the issues mentioned. In [4] an algorithm was developed that allows a user to choose one (of a number of cities), and in a dialogue choose a number of parameters (such as pictures, size, history, etc.). The algorithm provides the answer. This corresponds to the issues discussed in Section 2. But the algorithm works in an unusual way: it visits a number of servers that we have assessed to be reliable (see Section 3.1), and prides the user with the information requested as obtained from the "best" server. If desired it can also provide small tidbits of information which can lead to other servers in the background. The algorithm prefers servers satisfying requirements listed in Sections 3.2 or 3.3. The interested reader can reproduce a small version of the experiment via `https://austria-forum.org/ search-assistant`. Figure 4 illustrates a skeleton of a SERP introduced in the experiment.

There are some obvious problems: Concerning 3.1 we have no list of reliable servers, but have just hand-picked a few. It would be nice when showing tidbits to avoid redundancies, but we do not have adequate natural language algorithms to do so. Concerning 3.2 and 3.3, we cannot really use preferences because the properties desired are satisfied by almost no site. We are ignoring 3.4 since we tried it in another test as follows:

Figure 4. Illustration of SERP in the experiment conducted.

We are running a major server global-geography.org and there we have conducted further experiments. We checked the size of countries, cities and mountains, and found incredible differences. Thus, out of the 193 UN countries in the world we cannot tell you the size of 40 (!). Some reasons are political: is the Crimean peninsula now Russia or Ukraine? Do some of the Kuril Islands belong to Russia or Japan (they are all under Russian jurisdiction right now, but Japan claims Itururp und Kunashir), etc. But the issues go much deeper: In Austria-Forum, we cannot even tell you how large France (!) is, since databases present figures from 543965 km^2 to 674843 km^2, i.e. differ by up to 20% and they do not tell us how they have arrived at those figures. See Figure 5 for the variety of sizes of France in Austria-Forum.

total: 643,801 sq km; 551,500 sq km (metropolitan France)
land: 640,427 sq km; 549,970 sq km (metropolitan France)
water: 3,374 sq km; 1,530 sq km (metropolitan France)
note: the first numbers include the overseas regions of French Guiana, Guadeloupe, Martinique, Mayotte, and Reunion
We are unable to verify those figures from Factbook, since the 6 main sources we used for checking give different figures as follows:
Factbook↗: 643801
DBpedia↗: 674843
Geoname↗: 547030
Infoplease↗: 547030
Britannica↗: 543965
Wolfram↗: 551500

Figure 5. Different sizes of France in Austria-Forum.

5. Conclusion

Search engines require the best language processing facilities available. Whether good PageRank facilities and a (possibly complex) query with natural language interpretation will be the final answer, or more structured dialogue would be better, we cannot judge yet, but would bet on the second alternative.

Independent of this, there are three major issues, as stated in the paper, that have to be taken into account, if the information we find on the Web is to be considered credible.

References

[1] S. Brin and L. Page, "Reprint of: The Anatomy of a Large-scale Hypertextual Web Search Engine", *Computer Networks*, vol. 56, no. 18, pp. 3825–3833, 2012.
https://doi.org/10.1016/j.comnet.2012.10.007

[2] M. Bush, (2016), Edelman Trust Barometer Finds Global Trust Inequality is Growing [Online]. Available:
http://www.edelman.com/news/2016-edelman-trust-barometer-release/

[3] P. Grabowski, (2014), 30 Most Important Google Ranking Factors A Beginner Should Know. [Online]. Available:
https://unamo.com/blog/seo/30-important-google-ranking-factors-beginner-know

[4] M. Glatz, (2018), Access to Knowledge on the Web [Online]. Available:
https://austria-forum.org/web-books/accessknowledge00en2018isds/

[5] M. Maksimava, (2016), Google 9 main ranking factors: the ultimate cheat sheet. Available:
https://www.link-assistant.com/news/ranking-factors.html

[6] H. Maurer, "A Linear Ordering of a Multi-parameter Universe is Usually Nonsense", in *Theoretical Computer Science*, vol. 429, pp. 222–226, 2014.
http://doi.org/10.1016/j.tcs.2011.12.042

[7] P. Singhal, (2012), Introducing the Knowledge Graph: things, not strings [Online]. Available:
https://googleblog.blogspot.co.at/2012/05/introducing-knowledge-graph-things-not.html

[8] Internet Live Stats, (2018), Anzahl der Webseiten weltweit in den Jahren 1992 bis 2015 [Online]. Available:
https://de.sta3sta.com/sta3s3k/daten/studie/290274/umfrage/anzahlder-webseiten-weltweit/

Contact addresses:
Mathias Glatz
Graz University of Technology
Graz, Austria
e-mail: mathias.glatz@gmail.com

Hermann Maurer
Graz University of Technology
Graz, Austria
e-mail: hmaurer@iicm.edu

Afzal Tanvir
Capital University of Science and Technology
Pakistan
e-mail: mafzal@cust.edu.pk

Mathias Glatz studied computer pedagogy at the Graz University of Technology and completing it with an MSc degree in this area in early 2018. He is now pursuing an advanced degree in Web applications under the supervision of Professor Hermann Maurer and at the same time has a leading position as software developer at AVL, one of the biggest motor and machine manufacturers in Austria. His research focuses on improving user interfaces for Web applications, particularly on improving search on the Web and trying to assess the reliability of information hence obtained.

Hermann Maurer has been Professor of Computer Science at the Graz University of Technology for many years. He received his PhD from the University of Vienna in 1965, and has held professorial positions in Canada, the USA, Germany, Brazil and New Zealand. His research interest was originally in the theory of programming languages, but has since shifted much to multimedia techniques and WWW related issues. He has supervised some 50 PhD and 400 MSc theses, was foundation Dean for Computer Science in Graz, Austria, and is author of over 20 books and 700 other publications. He served as officer of the Academia Europaea for over ten years and also as member of the board of trustees for this institution.

Muhammad Tanvir Afzal got his MSc in Pakistan, but did his PhD in the area of semantic text analysis and plagiarism detection under the supervision of Professor Hermann Maurer at the Graz University of Technology, obtaining his degree in 2005. He is now Professor of Computer Science at the Capital University of Science and Technology, Islamabad, Pakistan. His research focuses on automatic paper classification using both text analysis and analysis of classical classification systems reference counts. Dr. Azfal is currently supervising a substantial group of computer science students pursuing the MSc and PhD degrees.