



What does Google recommend when you want to compare insurance offerings?

A study investigating source distribution in Google's top search results

Sebastian Sünkler, HAW Hamburg

Dirk Lewandowski, University of Duisburg-Essen, HAW Hamburg

Introduction

Search engines act as intermediaries between users and information objects on the Web by collecting, indexing, evaluating, and providing user interfaces to information. They are usually regarded as neutral and evaluate search results objectively (Halavais, 2018; Lewandowski, 2017). However, search engine optimization (SEO) methods significantly influence the ranking of commercial search engines, with both positive and negative effects. The importance of search engines, especially Google, is obvious due to their massive usage and high user trust. However, **little is known about the composition of the search result sets. Search engines are biased** (Tavani, 2012) because they tend to favor certain results, based on assumptions in their algorithms. This bias can lead to **misleading results**, especially when searching for **highly commercialized areas** such as medical treatment, credit, and insurance.

Research question

Our research investigates the distribution of different types of sites within search results sets for insurance comparisons, examining domains and providers for 121 search queries by addressing the following questions:

RQ1: How many different domains can be found in the top positions?

RQ2: How many different providers can be found in the top positions?

RQ3: How often do providers with more than one result appear in the Top10 for the various search queries?

Methods

For our study, we used a mixture of **automated and manual processes** to generate datasets for research and exploratory data analysis. Thematically relevant search queries were filtered from the logfile of a German commercial search engine and screen scraping was performed using the **Relevance Assessment Tool**¹. Other modules were used to process the data, and **KNIME** was used for aggregation and transformation. Descriptive statistics were used to examine the popularity of the top 10 providers and domains to determine how frequently certain websites and providers are used in insurance comparisons.

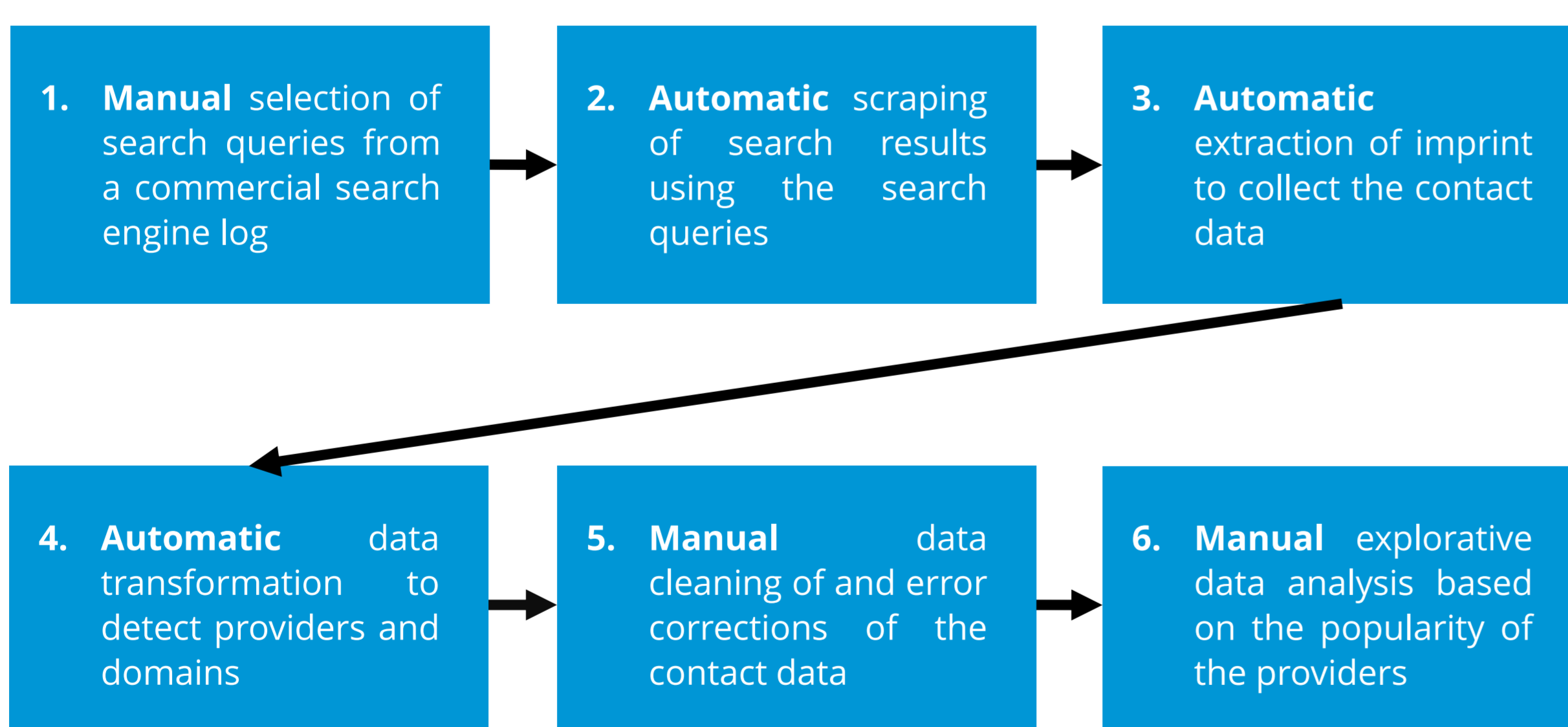


Figure 1: Data collection and analysis workflow

THE STUDY IS PUBLISHED IN:

Lewandowski, D. & Sünkler, S. (2019). „What Does Google Recommend When You Want to Compare Insurance Offerings?: A Study Investigating Source Distribution in Google's Top Search Results“. *Aslib Journal of Information Management* 71, 3 (May 20, 2019): 310–24. <https://doi.org/10.1108/AJIM-07-2018-0172>.

AVAILABILITY OF SOFTWARE AND RESEARCH DATA:

¹<https://rat-software.org/>

²<https://zenodo.org/record/2572691>

REFERENCES:

Halavais, A. (2018), *Search Engine Society*, Second ed., Polity, Cambridge, UK.
Lewandowski, D. (2017), 'Is Google Responsible for Providing Fair and Unbiased Results?', in Taddeo, M. and Floridi, L. (Eds.), *The Responsibilities of Online Service Providers*, Vol. 31, Springer, Berlin Heidelberg, pp. 61–77.
Lewandowski, D. & Sünkler, S. (2019). „What Does Google Recommend When You Want to Compare Insurance Offerings?: A Study Investigating Source Distribution in Google's Top Search Results“. *Aslib Journal of Information Management* 71, 3 (May 20, 2019): 310–24. <https://doi.org/10.1108/AJIM-07-2018-0172>.
Tavani, H. (2012), 'Search Engines and Ethics', 27 August, available at: <http://plato.stanford.edu/entries/ethics-search/>

Results

Our dataset comprises **22,138 search results from 3,278 domains**, with an average of 182.96 results per query². The **top 10 positions are 116 of the 3,278 domains**, with only ten listed in first place. Domain popularity is determined by counting the occurrence of all domains in the top 10 or top 5 results. The **five most popular providers cover 65.6% of all results up to the fifth position**, with the five most popular domains accounting for 42.9% of search results. The share of the most popular providers decreases with increasing position: 31.4% for the three most popular providers and 46.9% for the five most popular providers.

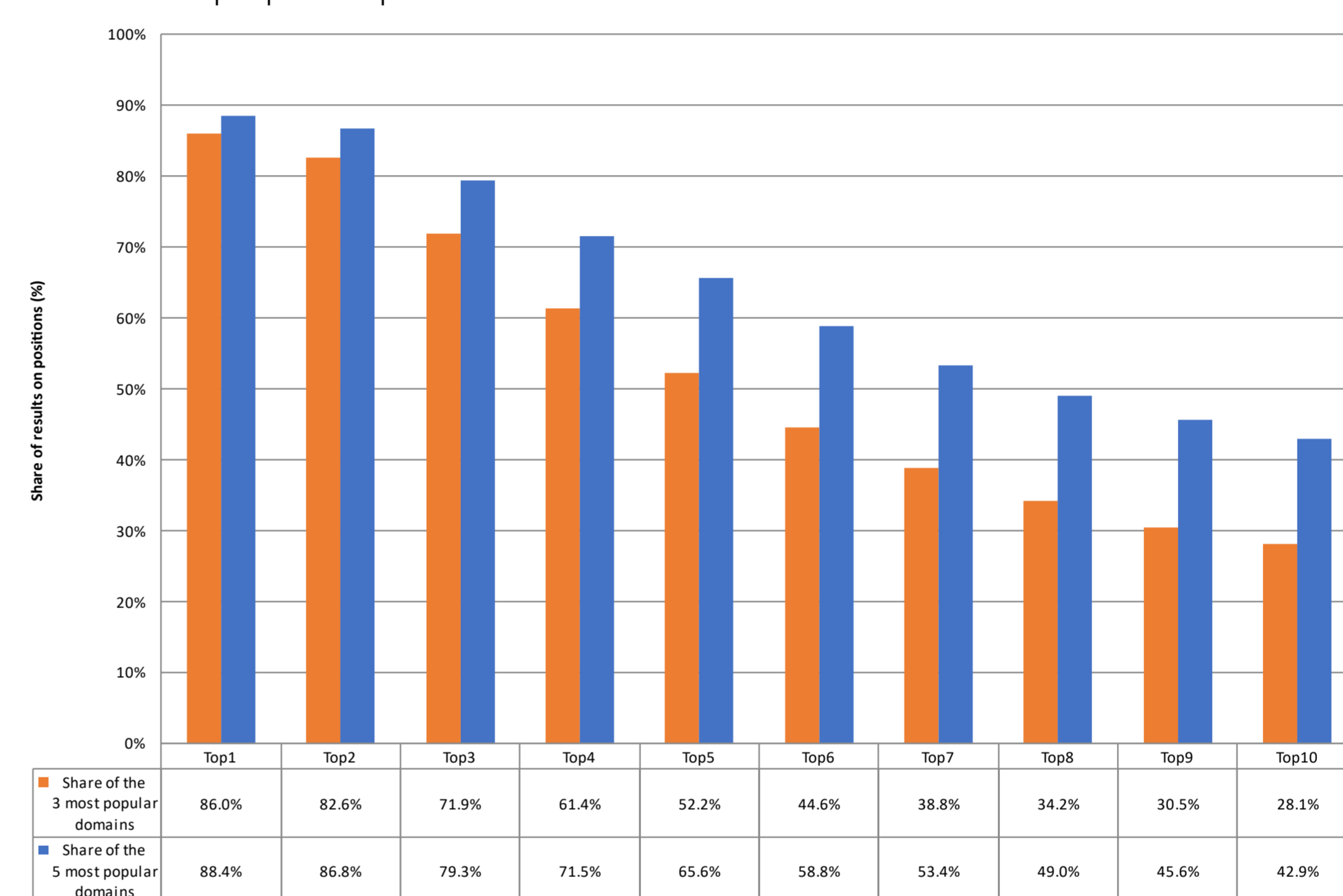


Figure 2: Relative distribution of the domains on the positions 1 – 10

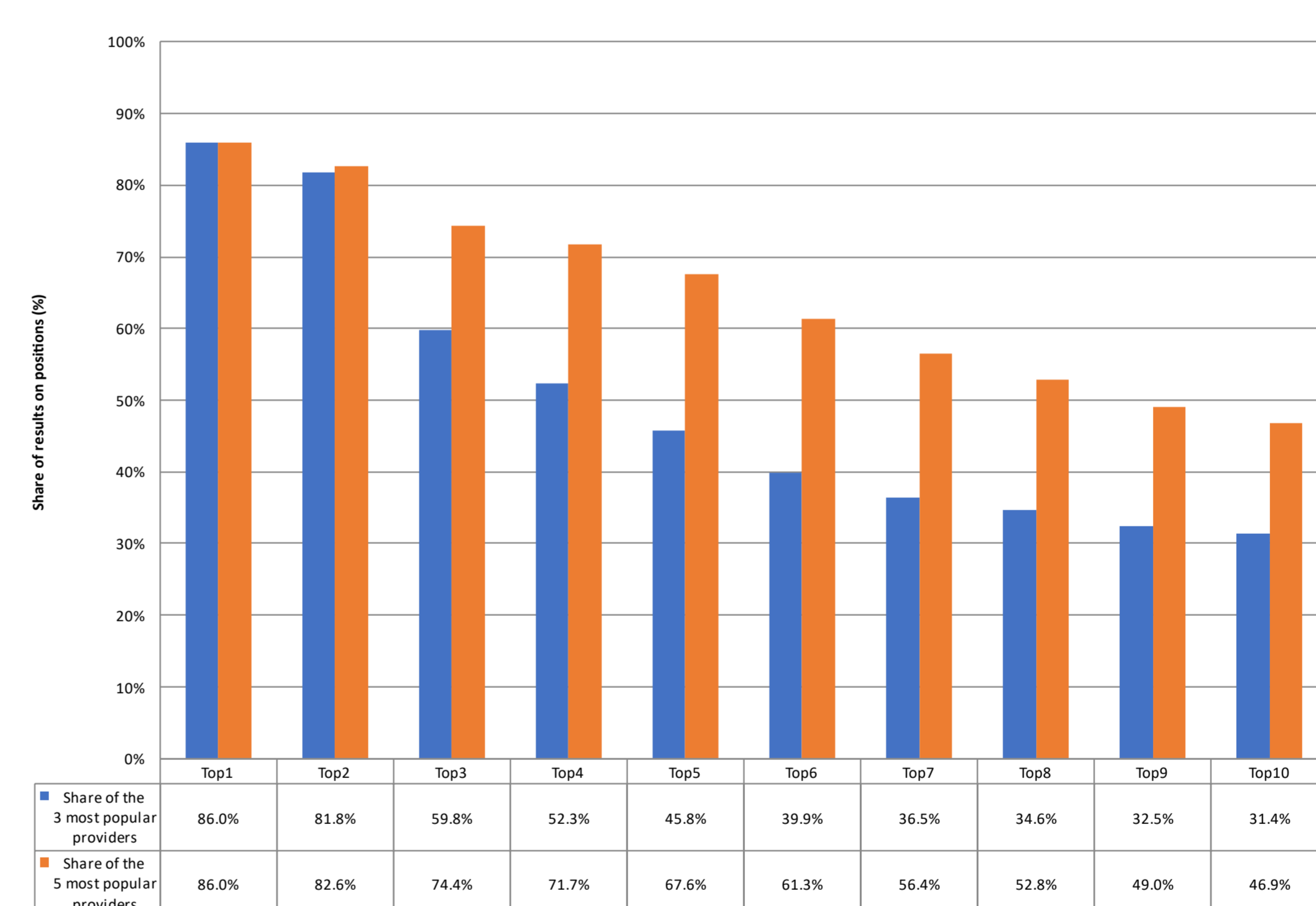


Figure 3: Relative distribution of the domains on the positions 1 – 10

Discussion / Conclusion

Our study of Google search results for **insurance offerings shows that a few providers dominate the top positions**, with only 116 domains and 93 different providers in the top 10 search results. We were able to show that our methodology, which uses automated and manual processes to measure the overlap between top domain results and provider contact information, provides fruitful results for the discussion of search engine bias. Limitations of the study include the small sample size, the use of the German version of Google search, and the failure to account for personalization and contextualization issues. Future studies should create larger data sets and consider search query frequency to determine vendor positions. However, the methodology and software can be used to conduct similar studies on other topics, such as controversial issues like abortion and nuclear energy, and areas where content providers have a strong interest in appearing high in search results, such as credit, travel, and health.

Contact

sebastian.suenkler@haw-hamburg.de
<https://searchstudies.org/team/sebastian>

**HOCHSCHULE FÜR ANGEWANDTE
WISSENSCHAFTEN HAMBURG**
Hamburg University of Applied Sciences