FACULTY OF DESIGN, MEDIA AND INFORMATION
Department of Information

RAT
RESULT ASSESSMENT TOOL

HAW HAMBURG

Googling for diseases

Exploring source types, SEO measures, and information quality of health-related search results

Sebastian Schultheiß[1], Helena Häußler[1], Dirk Lewandowski[1,2], and Sebastian Sünkler[1]

[1] Hamburg University of Applied Sciences, [2] University of Duisburg-Essen

Background

For individuals and society, searching for health-related information is of great importance. In 2020, about **one in two EU citizens searched online for health-related information** (Eurostat, 2021) and used search engines to make important **medical decisions** (Ray, 2020).

When searching, people usually only scan a few top-placed results (e.g., Granka et al., 2004). To achieve a top position, health information providers can use search engine optimization (SEO) measures to **have the best chances of being selected by users**. Information providers carry out SEO for various reasons, primarily for commercial interests such as increasing product sales (Röhle, 2010, p. 81). Indicators like commercial interests are perceived by the users and influence **how they think about the quality of a web page** (Sun et al., 2019). Hence, SEO can also bring content into users' focus that is *not* perceived as high-quality information.

The focus of this study is thus an exploratory **investigation of source diversity** when searching for diseases and **what information quality** is mediated to users **through SEO measures**.

Research Questions

(1) What source types are users facing when searching for diseases?

(2) To what extent do the providers use search engine optimization (SEO) measures?

(3) Is there a correlation between SEO measures and the quality of web pages?

Methods

For *N* = 318 health-related queries taken from **gesund.bund.de**, a website run by the German Federal Ministry of Health, we collected ***N* = 15,927 search results** from ***N* = 4,949 unique domains**.

We first manually classified the domains by **source type**. Second, we analyzed the **SEO probability** of the domains using our SEO classification tool (Lewandowski et al., 2021). Third, we conducted a user study (see Schultheiß et al., 2022). We had *N* = 61 participants **evaluate the quality** of random samples of optimized and non-optimized web pages (*N* = 100 pages each). The subjects were asked to assess the page quality using criteria such as trustworthiness and objectivity, which are among the most frequently chosen criteria by users when assessing the quality of online health information (Sun et al., 2019). The subjects also commented on their actions by **thinking aloud**.

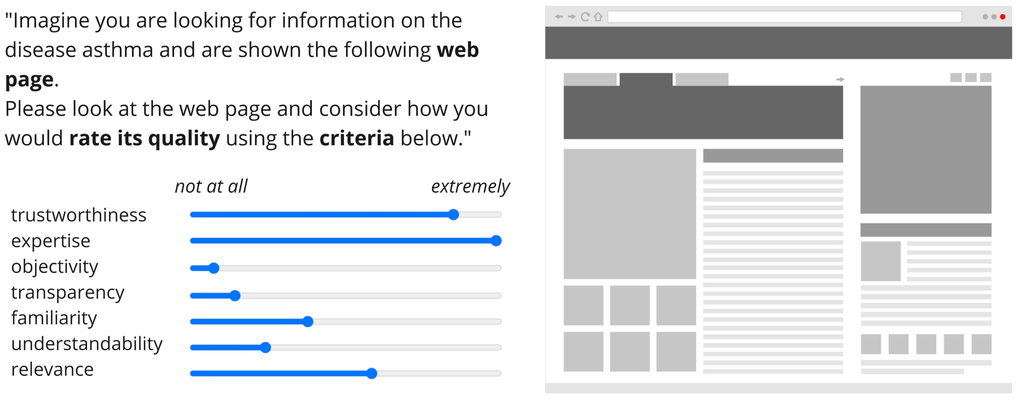

Figure 1: User evaluation of web page quality

References:

Eurostat. (2021). *One in two EU citizens look for health information online*. https://ec.europa.eu/eurostat/de/web/products-eurostat-news/-/edn-20210406-1

Granka, L. A., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. *Proceedings of the 27th Annual International Conference on Research and Development in Information Retrieval - SIGIR '04*, 478–479. https://doi.org/10.1145/1008992.1009079

Lewandowski, D., Sünkler, S., & Yagci, N. (2021). The influence of search engine optimization on Google's results. *13th ACM Web Science Conference 2021*, 12–20. https://doi.org/10.1145/3447535.3462479

Ray, L. (2020). *2020 Google Search Survey: How Much Do Users Trust Their Search Results?* https://moz.com/blog/2020-google-search-survey

Röhle, T. (2010). *Der Google-Komplex: Über Macht im Zeitalter des Internets*. Transcript. https://doi.org/10.14361/transcript.9783839414781

Schultheiß, S., Häußler, H., & Lewandowski, D. (2022). Does Search Engine Optimization come along with high-quality content? *ACM SIGIR Conference on Human Information Interaction and Retrieval*, 123–134. https://doi.org/10.1145/3498366.3505811

Sun, Y., Zhang, Y., Gwizdka, J., & Trace, C. B. (2019). Consumer Evaluation of the Quality of Online Health Information: Systematic Literature Review of Relevant Criteria and Indicators. *Journal of Medical Internet Research*, *21*(5), e12522. https://doi.org/10.2196/12522

Funding information:
This work is funded by the German Research Foundation (DFG Deutsche Forschungsgemeinschaft), grant number 417552432.
Data availability: All files from the user study are available From the OSF repository (https://osf.io/uk4eq/).

Try out our SEO classification tool

Results

Users are most frequently confronted with results by **clinics and practices**. This is followed by information from **commercial providers**, such as pharmaceutical companies.

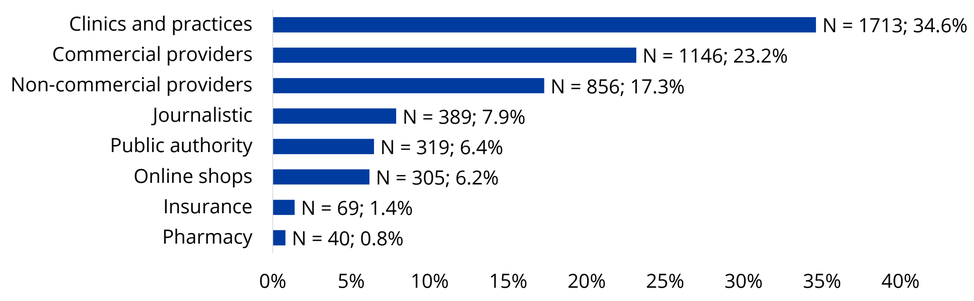

Figure 2: Source types

SEO is **widely used** by **online shops**, **journalistic sites**, and **commercial providers**, as the results of the SEO classification tool show. In contrast, **non-commercial providers** and **public authorities** use **SEO rarely**.

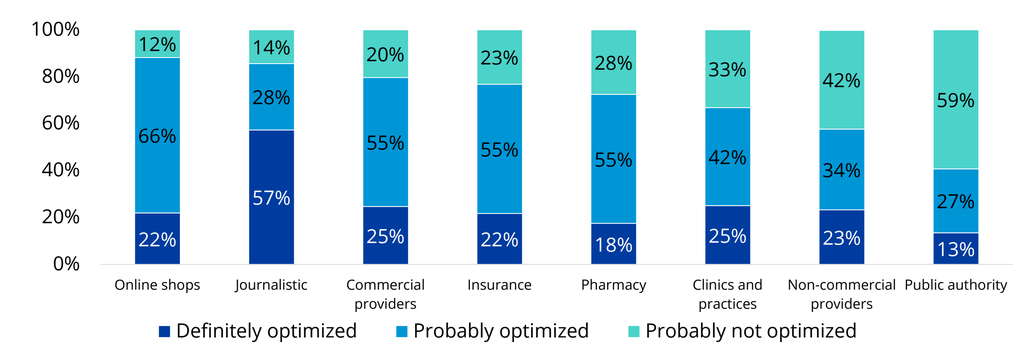

Figure 3: SEO measures by source types

The results of the user study revealed significant results for the criterion **expertise**. **Non-optimized** pages were considered **more competent** than optimized pages [$F(1, 59) = 4.217$, $p = 0.044$]. Subjects mainly justified their assessments by making statements directly related to the website operator.

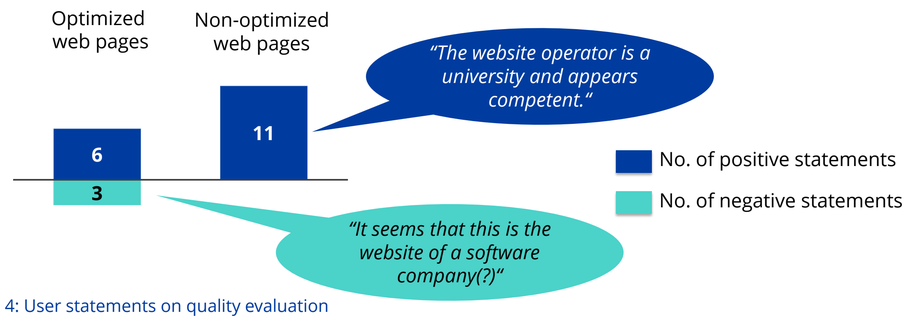

Figure 4: User statements on quality evaluation

Discussion

In this study on searching for diseases, we see a **high proportion of commercial providers** in the search results. These providers make **intensive use of SEO**, which contrasts with non-commercial providers and public authorities, who rarely promote the visibility of their content through SEO.

When asked about web page quality, users attribute **higher expertise to non-optimized pages**. They justify their assessments with a more competent and reputable appearance of the non-optimized websites.

Therefore, there is a risk that high-quality but non-optimized content could be **outranked by optimized content of lower quality**. This could lead to search engine users overlooking relevant information, resulting in uninformed or even harmful decisions.

Our study has implications for non-commercially motivated content producers to **strengthen their focus on SEO** as well as for search engine operators, as they should give **priority in their rankings to sources that possess expertise**.

One limitation is the small sample size of the user study. Starting points for future research include repeating the user study with a larger sample and a more sophisticated SEO classification and investigating the transferability of the results to other topics relevant to society, for example, political topics.

Email addresses
sebastian.schultheiss@haw-hamburg.de
helena.haeussler@haw-hamburg.de
dirk.lewandowski@uni-due.org
sebastian.suenkler@haw-hamburg.de

HOCHSCHULE FÜR ANGEWANDTE WISSENSCHAFTEN HAMBURG
Hamburg University of Applied Sciences