

# SEARCH STUDIES: INFORMATION SCIENCE, COMPUTER SCIENCE, AND BEYOND

**Dirk Lewandowski**

Hamburg University of Applied Sciences, Germany

[dirk.lewandowski@haw-hamburg.de](mailto:dirk.lewandowski@haw-hamburg.de)

# AGENDA

1. Intro: The Search Studies research group
2. How search engines produce results
3. Contextual advertising and how it may mislead users in their information seeking
4. The influence of search engine optimization (SEO)
5. Conclusion

# INTRO: THE SEARCH STUDIES RESEARCH GROUP

## OUR TEAM

[Home](#) / [Our Team](#)

### Researchers



**Dirk Lewandowski**  
*Professor*



**Ulrike Spree**  
*Professor*



**Sebastian Sünkler**  
*Research Assistant*



**Christiane Behnert**  
*Doctoral student*

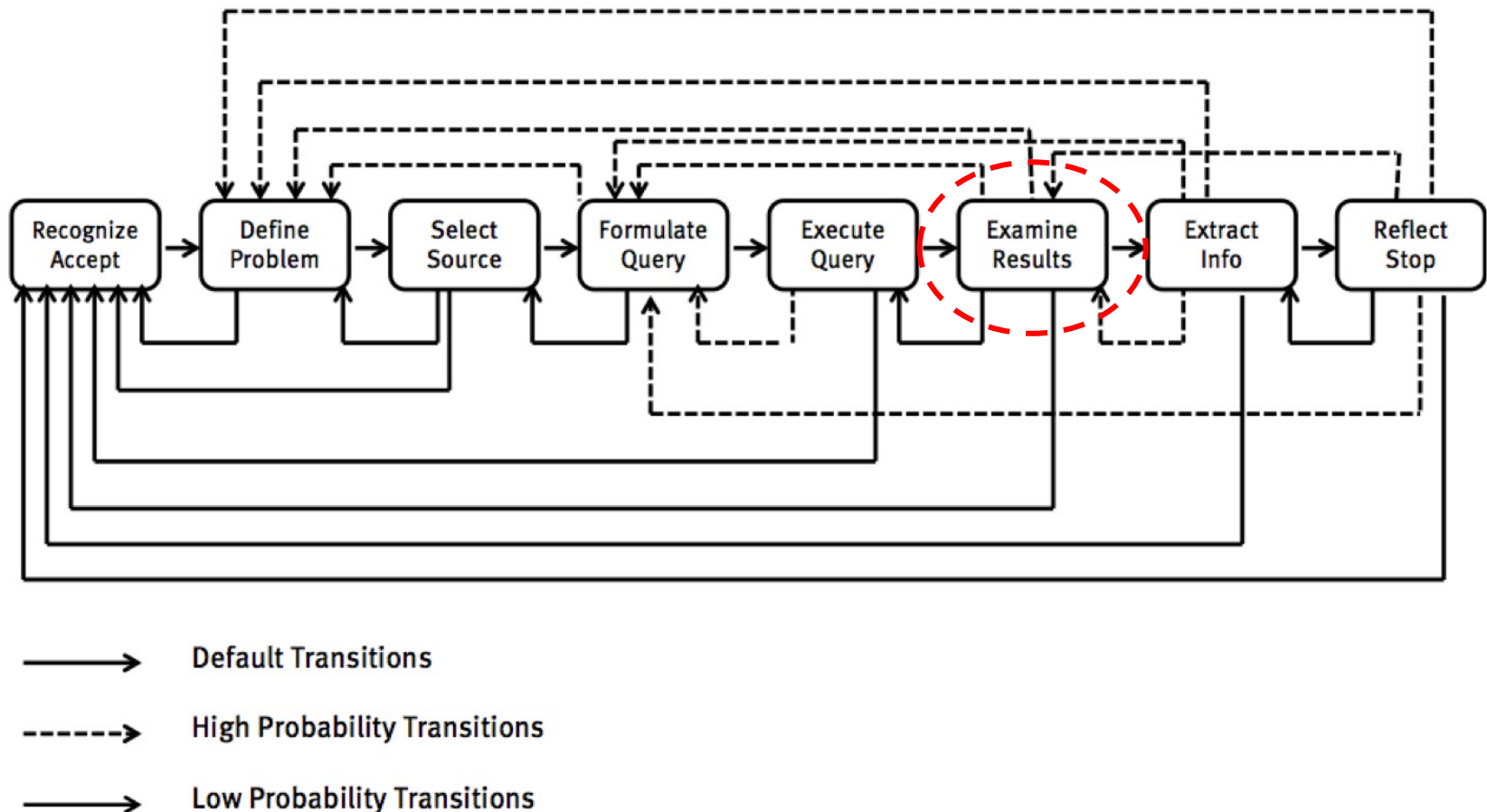


**Sebastian Schultheiß**  
*Research Assistant*



**Helena Häußler**  
*Research Assistant*

# SUB-PROCESSES IN INFORMATION SEEKING



(Marchionini, 1995)

# TWO PILLARS

## Search engine data analysis

What results are shown to users?

- Automatic data collection (screen scraping SERPs)
- (Semi-)automatic analysis of URLs and result documents
- Software development

## User studies

What results do users perceive and select?

- Task-based surveys, lab studies (incl. eye tracking)
- Focus on representative samples (surveys), diverse samples (lab studies)

# DEVELOPING RESEARCH SOFTWARE

## Our journey so far

- We have been developing software since 2009.
- Software development came from a lack of suitable software for our research

## Software

- Relevance Assessment Tool (RAT) – juror-based studies
- AAPVL – finding non-compliant food products on the web
- SEO-Effekt – identifying search engine optimization (SEO) in search results

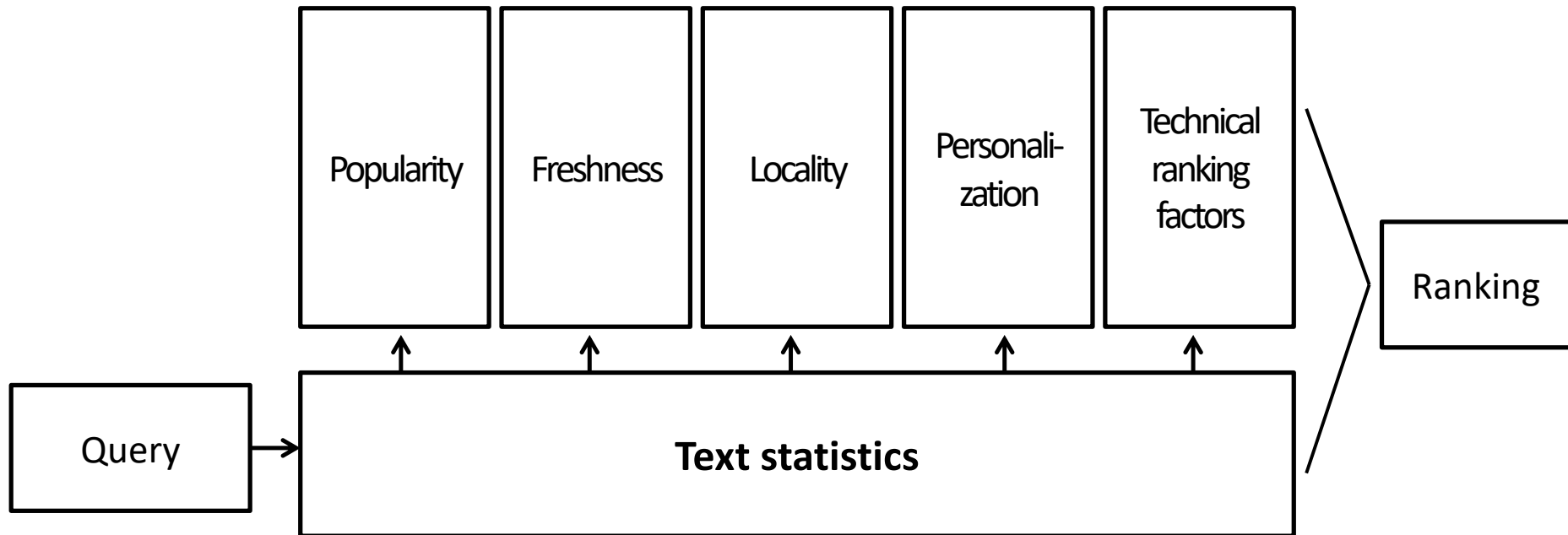
## Research software for

- Collecting results from commercial search engines (screen scraping)
- Making results available for assessments (studies with jurors)
- Automatic analysis of results (extracting indicators, machine learning)

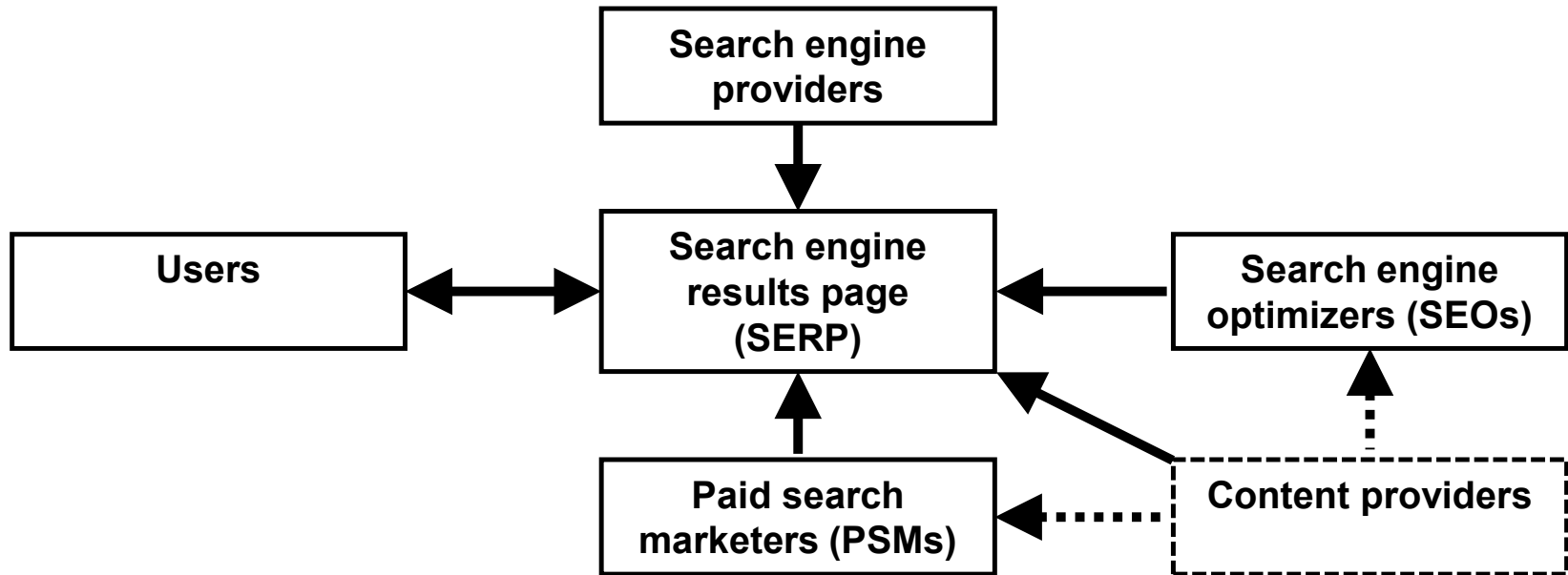
# HOW SEARCH ENGINES PRODUCE RESULTS



# SEARCH ENGINE RANKING (SIMPLIFIED)



# INFLUENCES ON SEARCH RESULT PAGES



(Schultheiß & Lewandowski, 2021b)

# SOME PROBLEMS RESULTING FROM THE “POPULARITY APPROACH”

Query: Sportlerinnen  
[female athletes]

Web results only



sportlerinnen



rp-online.de › Sport

## Die schönsten Sportlerinnen der Welt - RP Online

Ihr sportliches Metier beherrschen sie nahezu perfekt. Aber Frauen wie Michelle Kwan oder Carolina Klüft haben noch andere Qualitäten. Sehen Sie selbst -

www.sport1.de › boulevard › 2016/09 › die-heissesten-... ▼

## Die heißesten Sportlerinnen der Welt | Sport1 Bildergalerie

So sexy kann Sport sein. Egal ob Speerwerferin aus Paraguay, französische Fußballerin oder sexy Surferin: SPORT1 zeigt die heißesten **Sportlerinnen** der Welt.

www.t-online.de › sport › die-schoensten-sportlerinnen-d...

## Die schönsten Sportlerinnen der Welt - T-Online

Die schönsten **Sportlerinnen** der Welt (Quelle: imago images). (Quelle: imago images). Die US-amerikanische Golferin Paula Creamer sieht nicht nur gut aus, ...

www.pinterest.de › fr\_schubert › sportlerinnen

## Die 100+ besten Bilder zu Sportlerinnen in 2020 | sportler ...

29.07.2020 - Erkunde Franz Schuberts Pinnwand „**Sportlerinnen**“ auf Pinterest. Weitere Ideen zu Sportler, Tennisspieler, Sport frauen.

www.freenet.de › Sport › Abpiff ▼

## Die schönsten Sportlerinnen der Welt Bild 1 - Freenet

Ob Leichtathletik, Tennis oder auch Kampfsport: Wir zeigen die attraktivsten **Sportlerinnen** der Welt! 2 von 67. © Imago. Die schönsten **Sportlerinnen** der Welt.

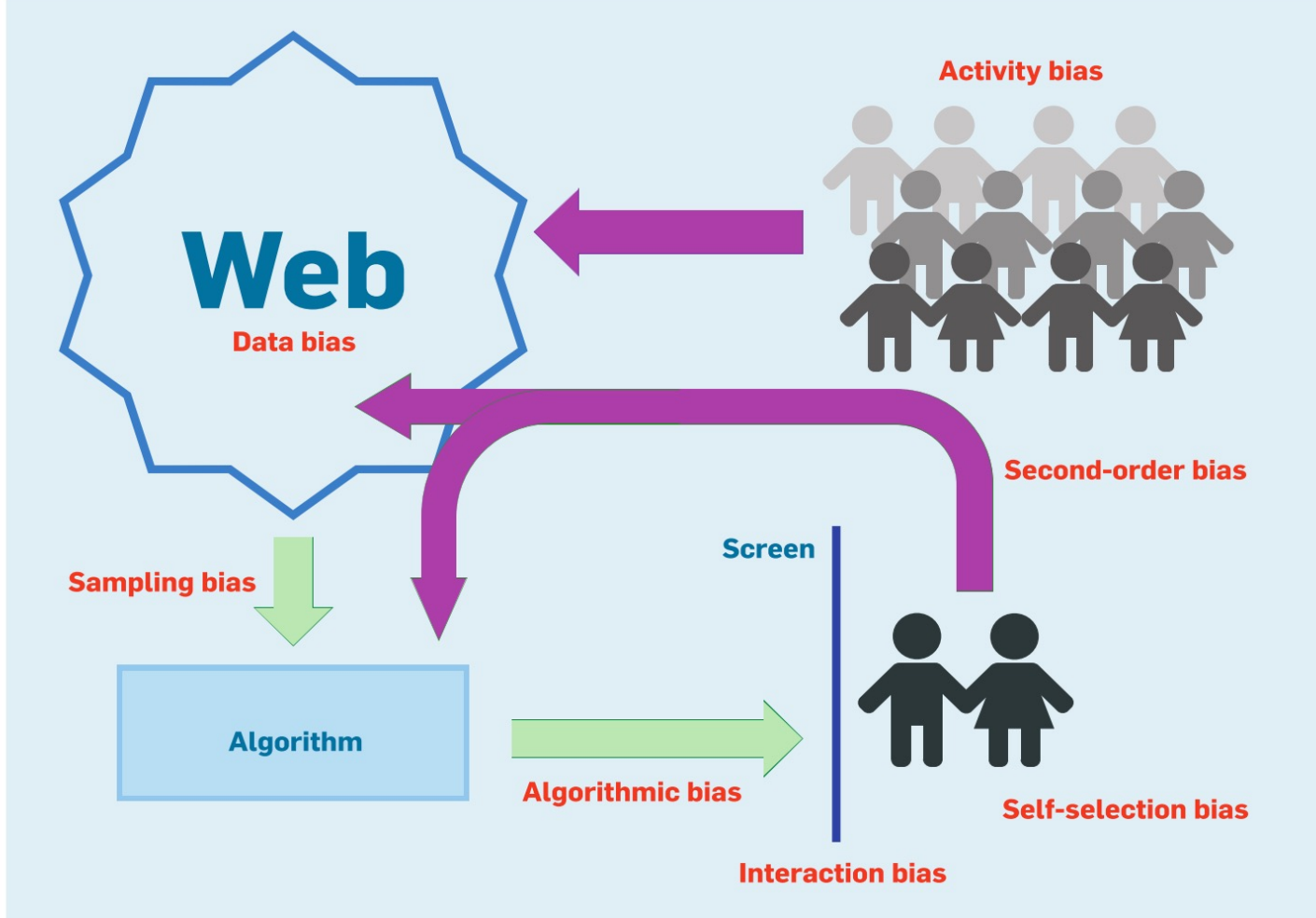
www.freenet.de › Sport › Abpiff ▼

## Das ist Deutschlands schönste Sportlerin Bild 1 - Freenet

Deutschland schönste **Sportlerin** ist Mareen Apitz, Volleyballerin beim SC Dresden. Bei einer Umfrage des Playboy landete sie auf dem ersten Platz. 2 von 73.

# BIAS ON THE WEB

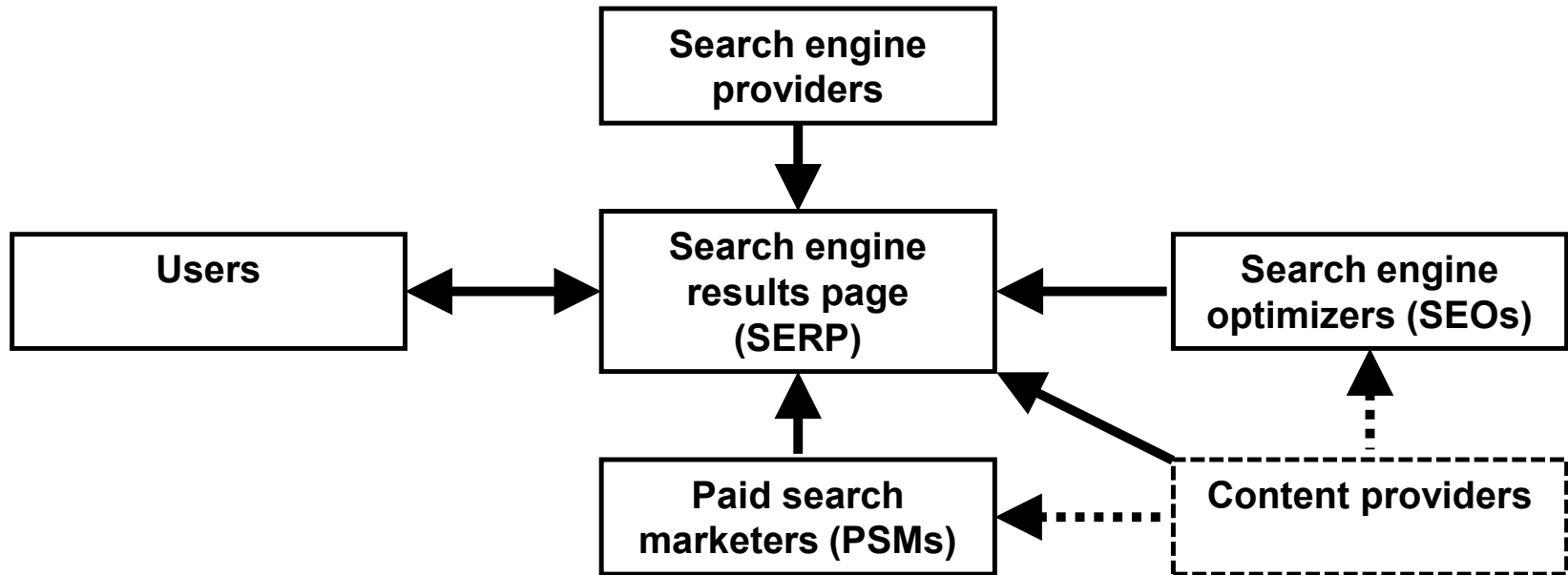
Figure 1. The vicious cycle of bias on the Web.



(Baeza-Yates, 2018)

11

# INFLUENCES ON SEARCH RESULT PAGES



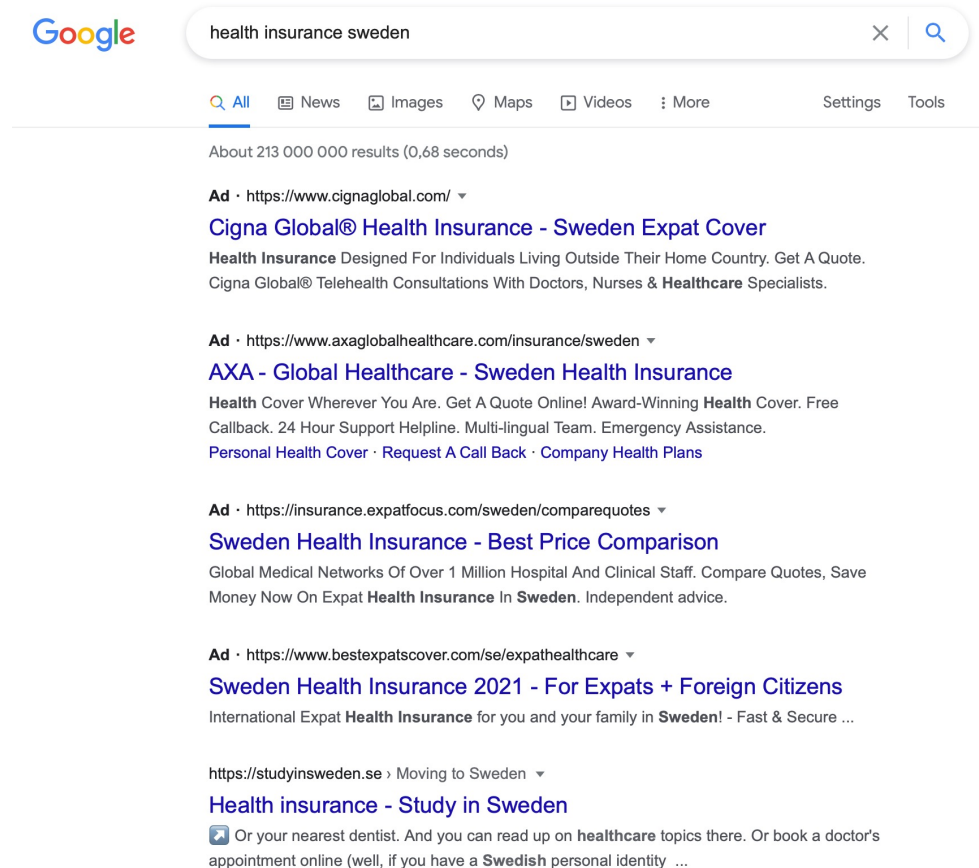
(Schultheiß & Lewandowski, 2021b)

12

# CONTEXTUAL ADVERTISING AND HOW IT MAY MISLEAD USERS IN THEIR INFORMATION SEEKING

# IMPORTANCE OF ADVERTISING TO SEARCH ENGINE COMPANIES

In 2019, 83% of Google's annual revenue of \$162 billion was generated through advertising (Alphabet Inc., 2020).



# MEASURING USER UNDERSTANDING OF ADVERTISING AND ITS EFFECTS ON SELECTION BEHAVIOUR

## Two large-scale studies representative of the German online population (data from 2014, 2020)

1. Users' knowledge on Google's business model
2. Four tasks where users had to mark all ads on SERPs
3. Experiment with manipulated SERPs (2014 only)

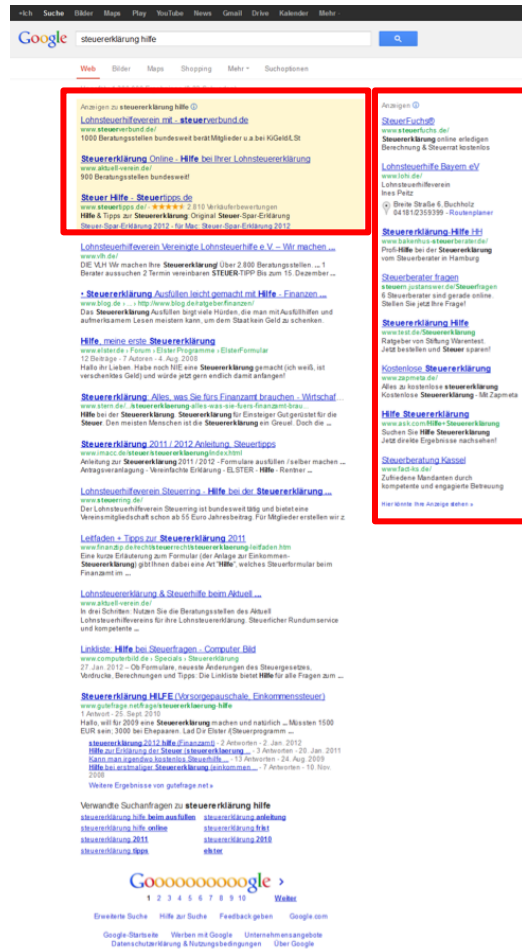
→ Results from the 2020 study (as yet unpublished) confirm the older results.

## Eye-tracking study (data from 2019)



# SAMPLE TASK: MARK ALL THE ADVERTISEMENTS ON THIS SEARCH ENGINE RESULTS PAGE

N=1,000



Only 35% marked all ads correctly.

18% marked at least some organic results as ads.

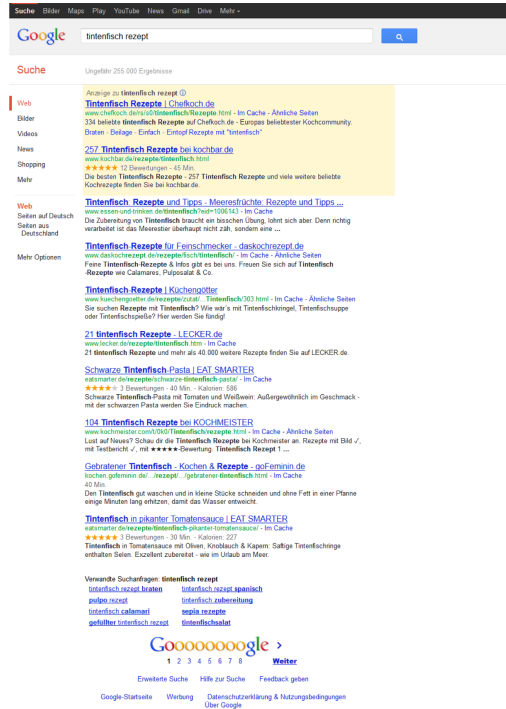
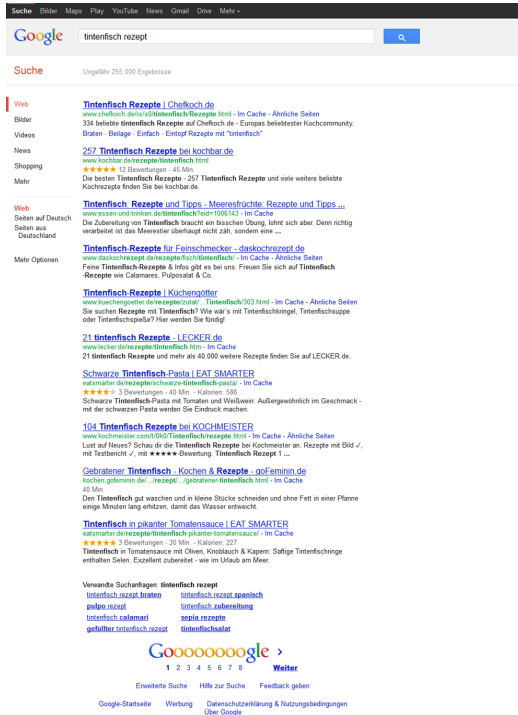
Still, more than 90% of users regard themselves as competent when it comes to using search engines.

(Lewandowski et al., 2018)

16

# RESULTS FROM THE EXPERIMENT

N=1,000



Users not able to distinguish ads from organic results clicked on the first ad about twice as often (40.3% vs. 21.6%).

(Lewandowski, 2017)

17

# UNDERSTANDING OF ADS INFLUENCES USERS' VISUAL BEHAVIOUR

N=100

Participants with a low level of knowledge on search advertising are more likely to click on ads than subjects with a high level of knowledge.

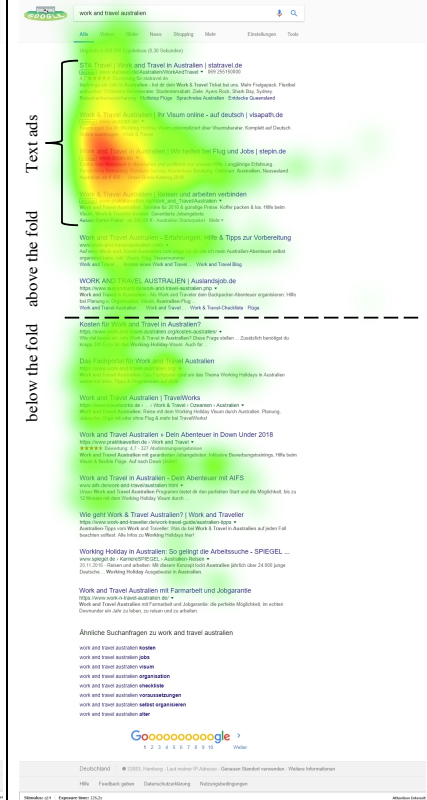
Subjects with little knowledge show less willingness to scroll down to organic results.

Larger effect on the mobile device

little knowledge of ads



comprehensive knowledge of ads



(Schultheiß & Lewandowski, 2021a)

# THE INFLUENCE OF SEARCH ENGINE OPTIMIZATION (SEO)

# SEARCH ENGINE OPTIMIZATION

Search engine optimization (SEO) is „the practice of optimizing web pages in a way that improves their ranking in the organic search results” (Li et al., 2014)

The market for SEO has reached 80 billion dollars in the U.S. alone (McCue, 2018).

SEO not only for products and services but also for informational content:

- Health
- Politics
- Energy (e.g., nuclear power)
- ...

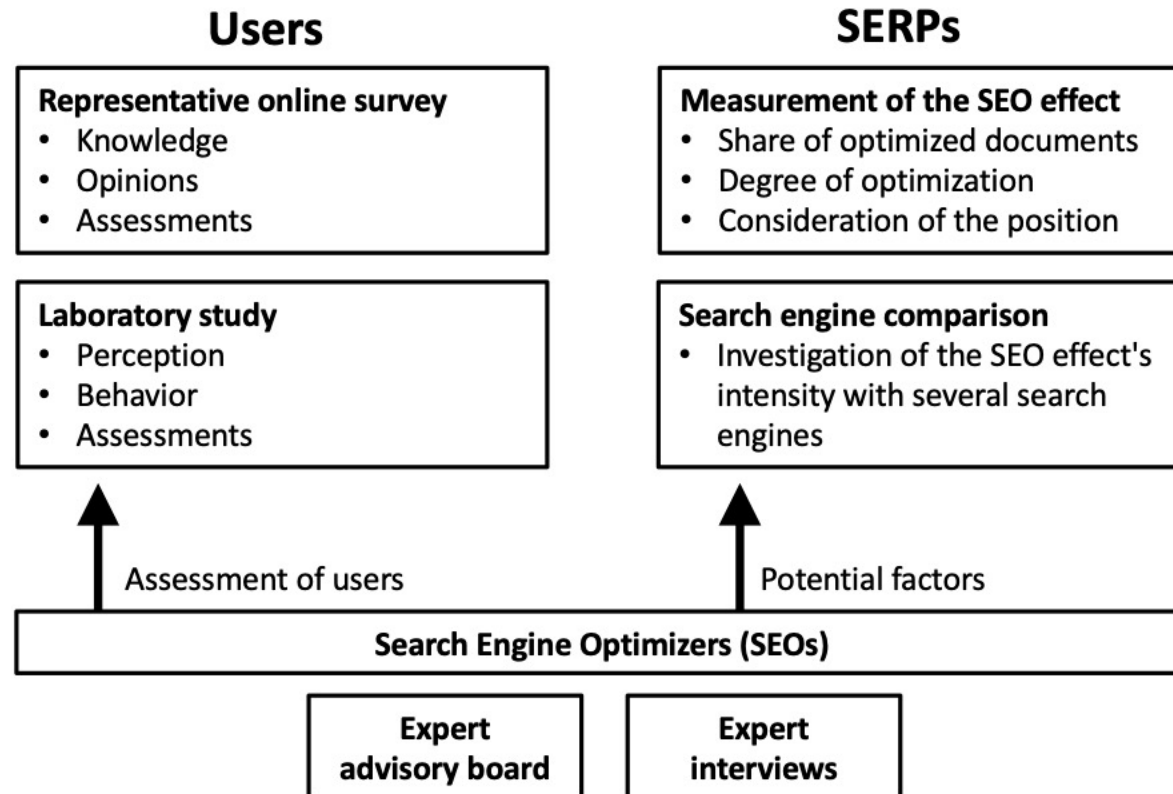
# THE “SEO EFFEKT“ PROJECT

Funded by



Deutsche  
Forschungsgemeinschaft

German Research Foundation



# SEO-EFFEKT: SOME FINDINGS (1)

## **Expert interviews** (Schultheiß & Lewandowski, 2021c)

- Search engine optimizers regard SEO as a means to providing users with better results.
- SEO is considered necessary for content providers to ensure visibility.

## **Representative user study** (publication in preparation)

- Users have a very limited understanding of SEO.

## **Trust in Google** (results from representative user study; Schultheiß & Lewandowski, 2021b)

- Users with little search engine knowledge are more likely to trust and use Google than users with more knowledge.

## **Data analysis** (Lewandowski, Sünkler & Yagci, 2021)

- Tool to extract information from HTML code + machine learning.
- Three datasets, more than 250,000 results analysed



 Elemente
  Netzwerk
  **Quellen**
 Speicher
  Konsole



- ▼ Breakpoints
  - D Debugger-Statements
  - Ex Alle Ausnahmen
  - Ex Nicht abgefangene Ausnahmen
  - A Zuweisungsfehler

**Nach Typ** Nach Pfad

The-Swedish-School-of-Library-and-Informati...

- ▶ Bilder
- ▶ Schriften
- ▶ Skripts
- ▶ Stylesheets
- ▶ Erweiterungsskripte

```

3
4 <!DOCTYPE html>
5 <html lang="en">
6 <head>
7   <meta charset="utf-8">
8   <meta http-equiv="X-UA-Compatible" content="IE=edge">
9   <title>The Swedish School of Library and Information Science - University of Borås</title>
10  <link rel="stylesheet" href="/assets/build/css/main.css?v=1.6.6.0">
11
12  <link rel="apple-touch-icon" sizes="180x180" href="/assets/img/favicons/apple-touch-icon.png">
13  <link rel="icon" type="image/png" sizes="32x32" href="/assets/img/favicons/favicon-32x32.png">
14  <link rel="icon" type="image/png" sizes="16x16" href="/assets/img/favicons/favicon-16x16.png">
15  <link rel="manifest" href="/assets/img/favicons/site.webmanifest">
16  <link rel="mask-icon" href="/assets/img/favicons/safari-pinned-tab.svg">
17
18  <link rel="stylesheet" href="https://use.typekit.net/ckl2ezo.css">
19
20  <meta name="msapplication-TileColor" content="#2f2f2f">
21  <meta name="theme-color" content="#ffffff">
22
23  <meta name="viewport" content="width=device-width, initial-scale=1">
24  <meta name="description" content="The Swedish School of Library and Information Science (SSLIS) is one of a kind in Sweden. Within Library and Information Science, our research, prog...">
25
26  <meta property="og:title" content="The Swedish School of Library and Information Science" />
27  <meta property="og:type" content="website" />
28  <meta property="og:url" content="https://www.hb.se/en/the-swedish-school-of-library-and-information-science-sslis/" />
29  <meta property="og:image:height" content="2000" />
30  <meta property="og:image:width" content="2000" />
31
32
33  <meta property="og:site_name" content="Högskolan i Borås">
34  <meta property="og:description" content="The Swedish School of Library and Information Science (SSLIS) is one of a kind in Sweden. Within Library and Information Science, our research, prog..." />
35
36  <meta name="twitter:card" content="summary">
37  <meta name="twitter:site" content="@hogskolaniboras">
38  <meta name="twitter:creator" content="@hogskolaniboras">
39  <meta name="twitter:description" content="The Swedish School of Library and Information Science (SSLIS) is one of a kind in Sweden. Within Library and Information Science, our research, prog...">
40  <meta name="twitter:title" content="The Swedish School of Library and Information Science">
41
42

```



# HOW CAN WE DETERMINE WHETHER A RESULT IS OPTIMIZED?

## Manually compiled lists

- Customers of SEO agencies
- News websites
- ...

## Positive indicators extracted from the HTML code (and website information)

- Use of SEO tools
- Use of analytics tools
- Use of microdata and schema.org markup
- SEO-related commands in robots.txt
- ...

## Negative indicators extracted from the HTML code

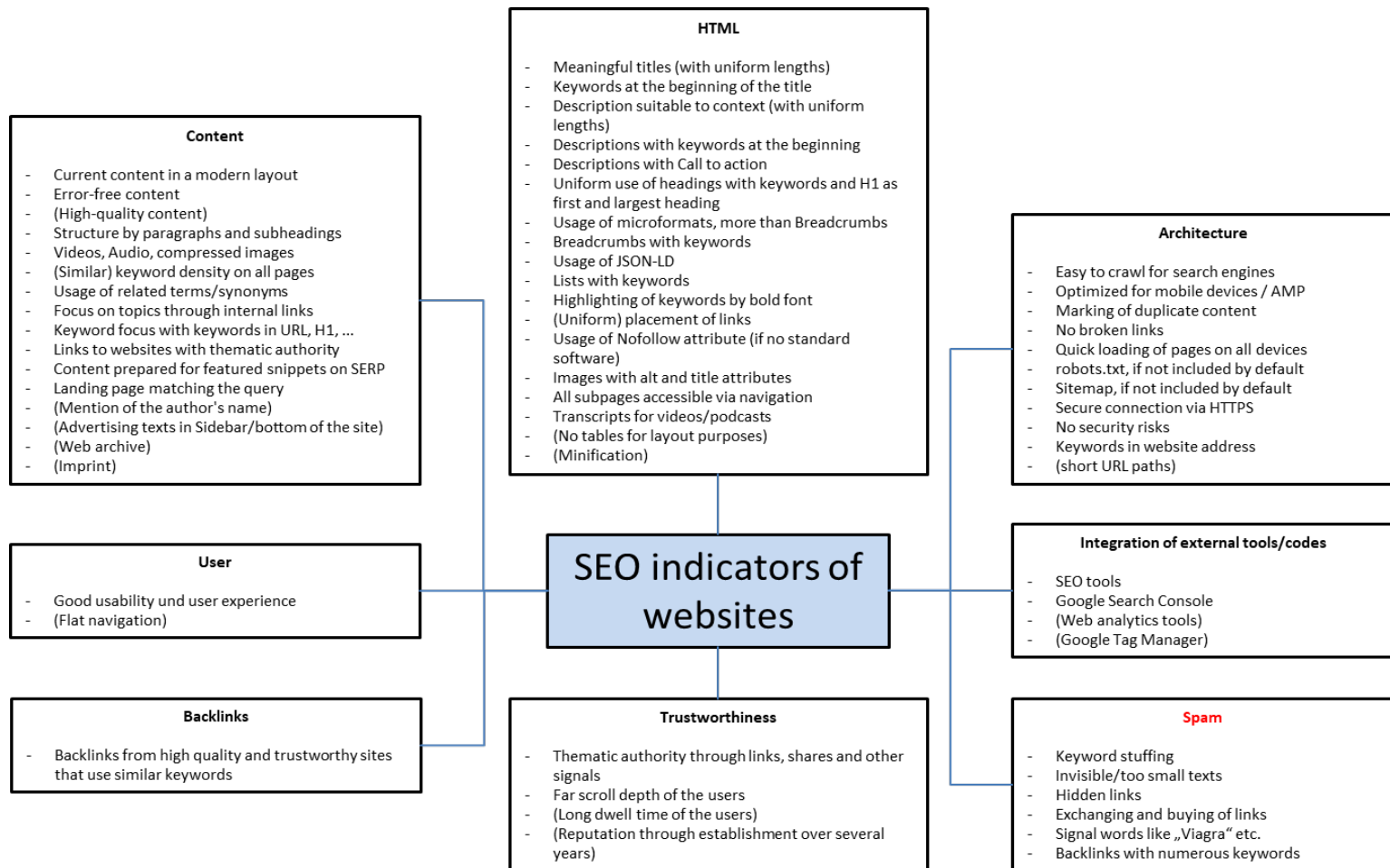
- Empty meta description or title tag
- Same title for every page of a website
- ...

<!--This site is optimized with the Yoast SEO plugin v12.4 - <https://yoast.com/wordpress/plugins/seo/>-->

<!-- Google Analytics -->

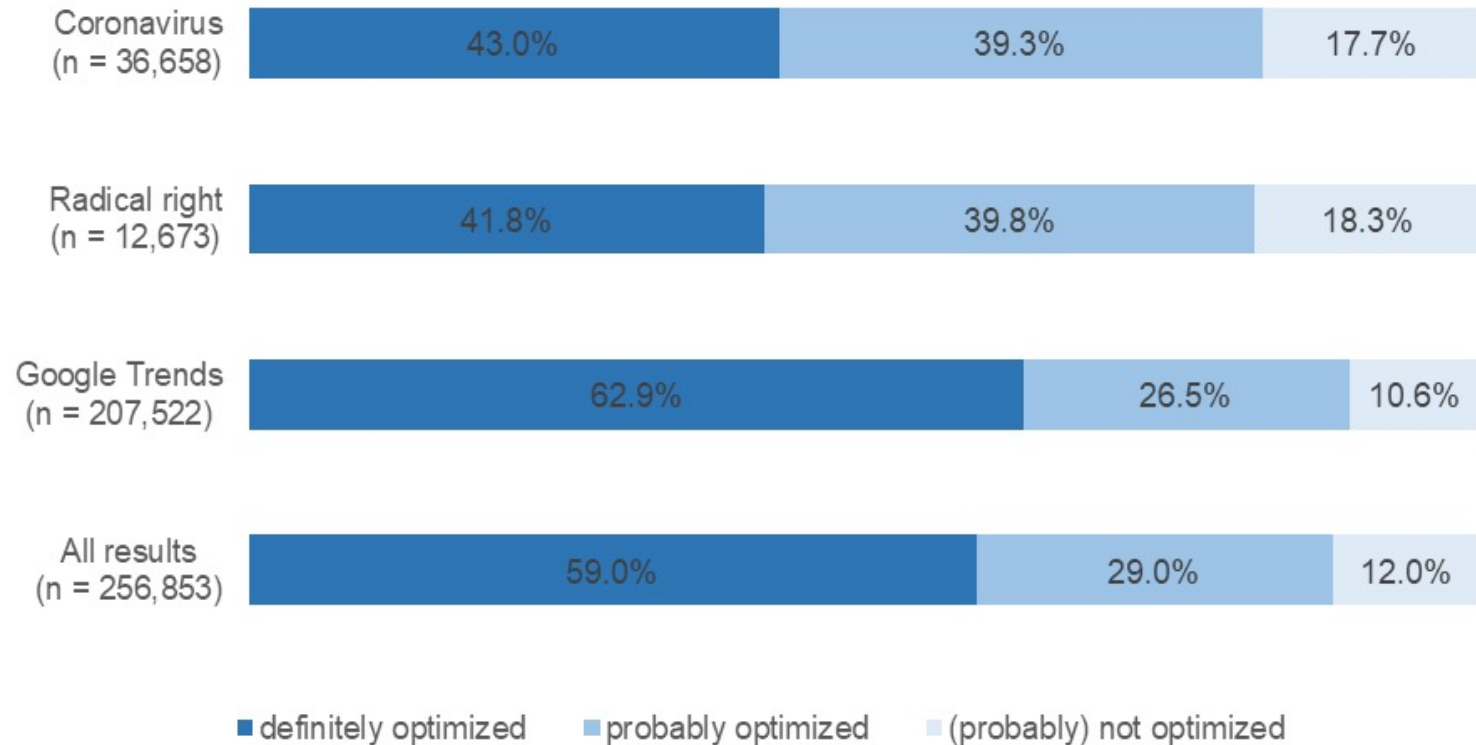
<script>(function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;...

# RANKING FROM THE PERSPECTIVE OF SEARCH ENGINE OPTIMIZATION (SEO)



**Explanation:**  
Indicators (or parts thereof) that are in brackets were controversial (agreement by one interviewee, rejection by another)

# RESULTS (1): WHAT PROPORTION OF RESULTS HAS BEEN OPTIMIZED?



(Lewandowski, Sünkler & Yagci, 2021)

27

## RESULTS (2): ARE THERE LESS SEO-OPTIMIZED PAGES THE FURTHER YOU GO DOWN THE RESULT LIST?

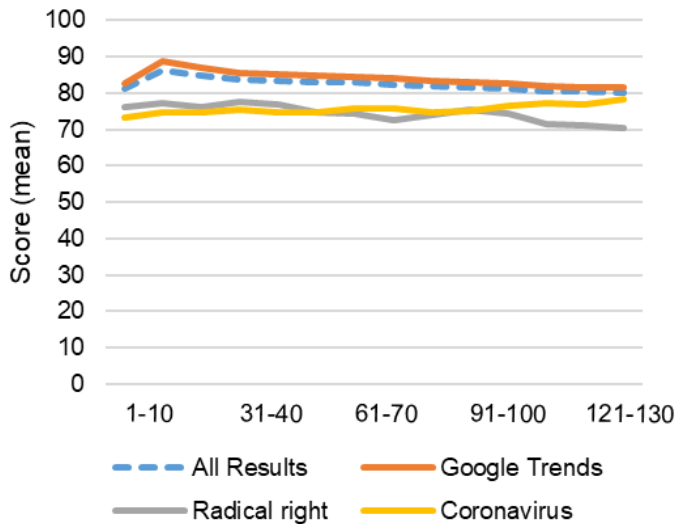


Figure 3: Score up to position 130

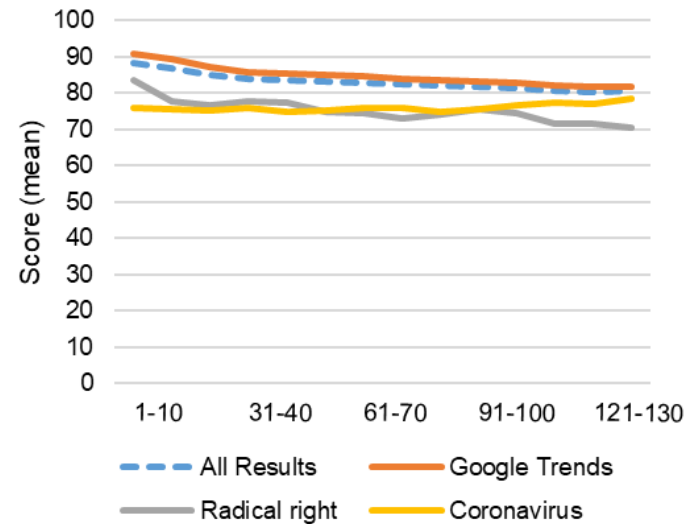


Figure 4: Score without Wikipedia up to position 130

(Lewandowski, Sünkler & Yagci, 2021)

# DEMO TOOL RESULTS

Results for <https://www2.bui.haw-hamburg.de/pers/ulrike.spree/>

## Tools & Plugins

SEO Tools: ✗  
Analytics Tools: ✗

## URL Classification

Online Marketing Customer: ✗  
News Service: ✗  
Ads on Website: ✗  
Company Website: ✗  
Online Shop: ✗  
Not optimized: ✗

## Indicators for SEO

https: ✓  
Description: ✗  
Title: ✓  
robots.txt: ✗  
Sitemap: ✗  
nofollow: ✗  
speed: ✓  
canonical: ✗  
viewport: ✗  
Micros: ✗

## SEO Assessment

Most probably: ✗  
Most probably not: ✗  
Probably: ✗  
Likely not optimized: ✓  
Uncertain: ✓

Home | Biographie | Veröffentlichungen |  
WS 2014/15 | Veranstaltungsarchiv | Lehrmaterialien nach Themen | Bachelor-/Masterarbeiten |  
Projekte | Auslandspraktikum | Studienberatung / BAföG-Dozentin |



## Prof. Dr. Ulrike Spree

### Lehrgebiete:

Hochschule für Angewandte Wissenschaften Hamburg  
Fakultät Design Medien Information  
- Department Information -  
Mediencampus Finkenau  
Finkenau 35  
22061 Hamburg

Informationsdienstleistung und Informationsvermittlung  
Wissensorganisation und Records Management  
Information Research

Telefon: 040/42875/3607  
Mail: [ulrike.spree@haw-hamburg.de](mailto:ulrike.spree@haw-hamburg.de)

**Sprechstunde:**  
mittwochs 15:00 bis 16:30  
Am 03.12.2014 muss die Sprechstunde leider ausfallen, da ich dienstlich unterwegs bin.

**Aktuelle Informationen:**  
Kurse in Moodle  
Themenvorschläge Bachelor- und Masterarbeiten  
Infos zum BAföG-Antrag

### Aktuelle Veröffentlichungen und Vorträge:

- Alexander, Fran; Spree, Ulrike (Hgg.): Aslib Journal of Information Management. Special Issue: Semantic Search. Vol. 66, 2014, Nr. 5 (Beginn ab S.473)  
<http://www.emeraldinsight.com/toc/ajim/66/5>
- Spree, Ulrike: How Readers Shape the Content of an Encyclopedia: A Case Study Comparing the German Meyers Konversationslexikon (1885-1890) with Wikipedia (2002-2013). In: Culture Unbound : Journal of Current Cultural Research; Vol. 6, Article 29, 2014  
<http://www.cultureunbound.ep.liu.se/v6/a29/cu14v6a29.pdf>
- Spree, Ulrike: Kulturelle Teilhabe ermöglichen : Möglichkeiten und Grenzen zielgruppenorientierter Bibliotheksarbeit für Menschen mit Demenz. In: Medien & Altern 2013; 3. S. 77-93  
[preprint]
- Spree, Ulrike: Wörterbücher und Enzyklopädien. In: Grundlagen der praktischen Information und Dokumentation / hrsg. von Kuhlén, Rainer. 6., völlig neu gefasste Ausgabe. - Berlin, Boston, Mass. : De Gruyter Saur, 2012, S. 550-559. [preprint]
- Ulrike Spree: Mutually dependent : Encyclopedias and their audiences. Vortrag gehalten am 09.11.2012 auf der Konferenz "Encyclopedias and beyond" in der National Library Oslo
- Ulrike Spree; Jutta Lindenthal; Andje Knaack: Wortnetz Kultur - ein Thesaurusprojekt zur kollaborativen Erschließung von Fachinformationen des kulturellen Erbes. In: Information, Wissenschaft & Praxis 2012; 63(1). S.23-36. [preprint]

... weitere Publikationen

### Mein Motto:

When action grows unprofitable, gather information; when information grows unprofitable, sleep.

(Demo tool available at <http://5.189.155.20:5000/>)

29

# CONCLUSION

# CONCLUSION

## Summary

- Billions are spent on influencing on what users see on result pages of commercial search engines.
- Users have a very limited understanding of commercial influences on search results (advertising, SEO, and search engine companies' self-interests).
- Combining social science / information science with computer science methods provides a fruitful basis for conducting research on commercial search engines.

## Consequences for information science

- An obvious solution, of course, is a need for more information literacy. However, IL alone is a too simple solution.
- Information science research has always assumed that relevance is the sole aim of search system providers but now commercial interests influence what users get to see. IS as a discipline should position itself against this background.
- To gain a better understanding of the impact commercial search engines have on knowledge acquisition, we need a wider theoretical and methodological lens, i.e., moving beyond just result selection.



# THANK YOU

Dirk Lewandowski

Hamburg University of Applied Sciences, Hamburg, Germany

[dirk.lewandowski@haw-hamburg.de](mailto:dirk.lewandowski@haw-hamburg.de)

[www.searchstudies.org/dirk](http://www.searchstudies.org/dirk)



# REFERENCES

- Alphabet Inc. (2020). *Alphabet Announces Fourth Quarter and Fiscal Year 2019 Results*, [https://abc.xyz/investor/static/pdf/2019Q4\\_alphabet\\_earnings\\_release.pdf](https://abc.xyz/investor/static/pdf/2019Q4_alphabet_earnings_release.pdf) (accessed 11 December 2020).
- Baeza-Yates, R. (2018). Bias on the web. *Communications of the ACM*, 61(6), 54–61. <https://doi.org/10.1145/3209581>
- Lewandowski, D. (2017). Users' Understanding of Search Engine Advertisements. *Journal of Information Science Theory and Practice*, 5(4), 6–25. <https://doi.org/10.1633/JISTaP.2017.5.4.1>
- Lewandowski, D., Kerkmann, F., Rümmele, S., & Sünkler, S. (2018). An empirical investigation on search engine ad disclosure. *Journal of the Association for Information Science and Technology*, 69(3), 420–437. <https://doi.org/10.1002/asi.23963>
- Lewandowski, D., Sünkler, S., & Yagci, N. (2021). The influence of search engine optimization on Google's results: A multi-dimensional approach for detecting SEO. In *WebSci '21: Proceedings of the 13th ACM Conference on Web Science*. New York: ACM.
- Li, K., Lin, M., Lin, Z., & Xing, B. (2014). Running and chasing - The competition between paid search marketing and search engine optimization. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 3110–3119. <https://doi.org/10.1109/HICSS.2014.640>
- Marchionini, G. 1995. *Information Seeking in Electronic Environments*. Cambridge Series on Human-Computer Interaction.
- McCue, T. (2018). SEO Industry Approaching \$80 Billion But All You Want Is More Web Traffic. Retrieved from <https://www.forbes.com/sites/tjmccue/2018/07/30/seo-industry-approaching-80-billion-but-all-you-want-is-more-web-traffic/>
- Schultheiß, S., & Lewandowski, D. (2021a). How users' knowledge of advertisements influences their viewing and selection behavior in search engines. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.24410>
- Schultheiß, S., & Lewandowski, D. (2021b). Misplaced trust? The relationship between trust, ability to identify commercially influenced results, and search engine preference. *Journal of Information Science*. <https://doi.org/10.1177/01655515211014157>
- Schultheiß, S., & Lewandowski, D. (2020c). "Outside the industry, nobody knows what we do" SEO as seen by search engine optimizers and content providers. *Journal of Documentation*, 77(2), 542–557. <https://doi.org/10.1108/JD-07-2020-330127>