

# DOES IT MATTER WHICH SEARCH ENGINE IS USED?

A user study using post-task relevance judgments

Sebastian Sünkler, [Dirk Lewandowski](#)

Hamburg University of Applied Sciences, Hamburg, Germany

ASIST 2017, Washington D.C., 1 November, 2017

# OVERVIEW OF PRESENTATION

1. Motivation and introduction to the new approach
2. Literature review
3. Research questions (for the user study)
4. Methods
5. Results
6. Summary and conclusion

# MOTIVATION

Search engines are an important means for finding information on the Web, and because of the Web's importance to knowledge acquisition, they are also an important means to what users get to know online.

Users predominantly use Google, especially in the European countries, where Google has a market share of well over 90% (comScore, 2013).

Is Google really better than its competitors?

# APPROACHES TAKEN SO FAR

Approaches taken so far either focus on

- (1) comparing the results of search engines using jurors who are not aware of where each result comes from (for an overview, see Lewandowski, 2015),
- (2) comparing search engines' results ranking to users' rankings of the same results (Bar-Ilan, Keenoy, Yaari, & Levene, 2007),
- (3) measuring task success or user satisfaction in user studies where users use either the one or the other search engine (cf. White, 2016, p. 328ff.).

# NEW APPROACH

**Method extending user-centered approaches by adding system-oriented tests to the model.**

With this method, we can measure:

- (1) The relevance of the results users selected in their search sessions,
- (2) The relevance of all results users were presented with in their search sessions (*i.e., results which they did not click on*),
- (3) The relevance of the results from a competing search engine for the users' queries (*i.e., results they did not see in their search session*).

To test this method: User study (64 participants)

# LITERATURE REVIEW

# LITERATURE REVIEW (1)

- **Users are satisfied with search engine results** (Purcell, Brenner, & Raine, 2012; Westerwick, 2013)
- **Users select results predominantly from the first few positions** (Joachims et al., 2007; Petrescu, 2014).
- **Only 10,000 different websites account for approx. 80% of clicks on the search engine results pages** (Goel et al., 2010)
- Results quality: Google outperforms Bing on navigational queries. **Google's results for informational queries are, on average, only slightly more relevant than Bing's** (Lewandowski, 2015).

# LITERATURE REVIEW (2)

## Approaches to measuring retrieval effectiveness

### System-centred approach (“Cranfield-style”)

- Controlled environment, focus on the “objective” quality of search results
- Allows for comparing results from different search engines
- Major shortcomings: Isolated relevance judgments; does not take the user into account

### User-centred approach (Interactive Information Retrieval)

- Allows for interaction between user and system
- Major shortcoming: Does not allow for systematically comparing search engine results

→ **Need for combining elements from both paradigms:** Letting users judge on the relevance of the results from *their own search sessions* with *their own search queries*.



# RESEARCH QUESTIONS

# RESEARCH QUESTIONS

RQ1: To what extent do users use different search engines when working on search tasks given by the experimenter?

RQ2: Do users judge results they actually selected when performing their tasks as more relevant than results not selected during the task?

RQ3: Would users be equally satisfied with the results of another search engine; i.e., would Google users be equally satisfied with results from Bing?

RQ4: Would users judge search results as more relevant if they themselves judged the retrieved information as correct to solve the task?

RQ5: Do the two search engines perform differently when considering simple vs. complex tasks?

# METHODS

# METHODS

## Nine steps for designing studies combining users' session data and explicit relevance judgments

- (1) Selection of search engines to investigate.
- (2) Definition of user groups.
- (3) Definition of test timeframe and test environment
- (4) Design of questionnaires (allowing users to, e.g., rate difficulty, outcome, effort, and learning success of tasks).
- (5) Design of simulated search tasks.
- (6) Definition of scales and questions about the search results
- (7) Collecting interaction data, taking into account the search engine(s) used, results clicked, and the search queries used.**
- (8) Build result sets based on search engines used, search queries, clicked results and additional search results from other search engines.**
- (9) Analyze user interactions and results judgments.

## **Search-Logger (Singer, Norbistrath, Vainikko, Kikkas, & Lewandowski, 2011)**

- Tool developed to log user interactions in a Web browser
- Provides options to design search tasks

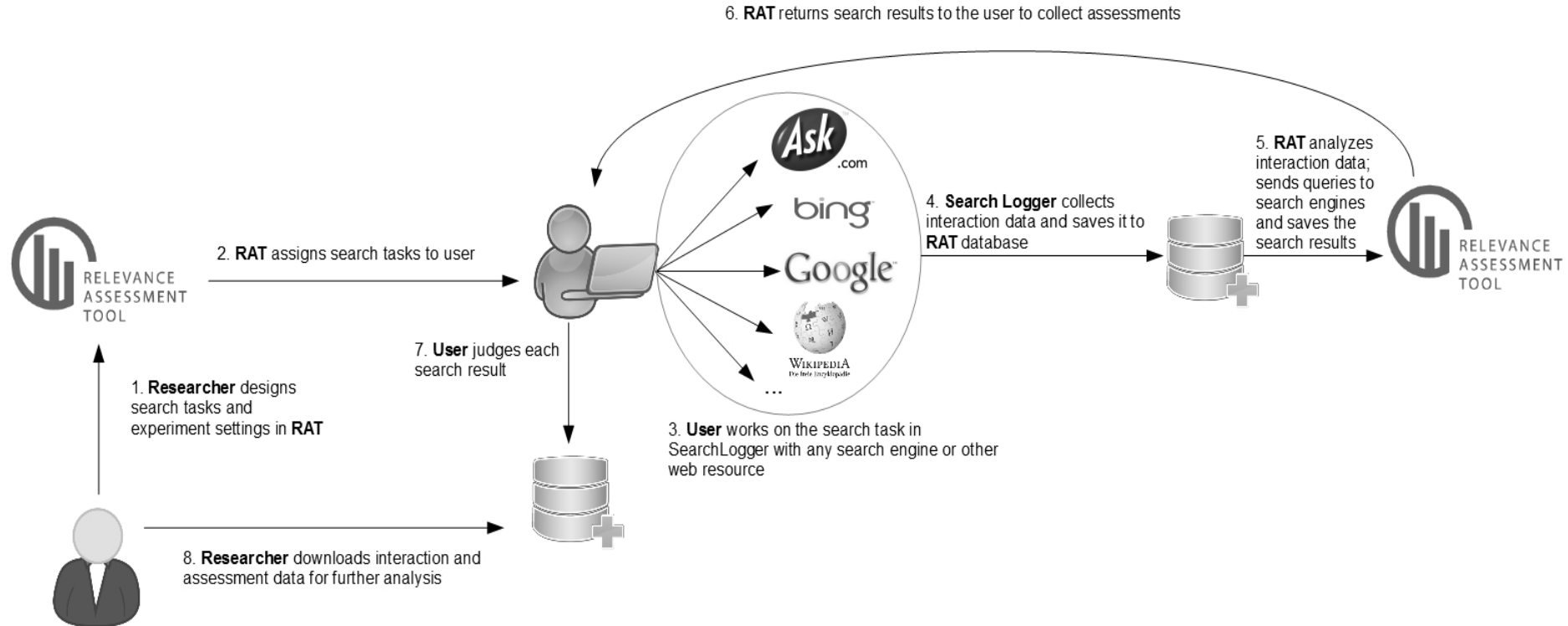
## **Relevance Assessment Tool (RAT; Lewandowski & Sünkler, 2013)**

- Software toolkit that allows researchers to conduct large-scale retrieval effectiveness studies based on results from (commercial) search engines and other information retrieval systems
- Modules for (1) designing studies, (2) collecting data from search systems, (3) collecting judgments on the results, (4) downloading/analysing the results.

### **→ New approach combines RAT and Search-Logger**

- Application analyses the Search Logger output, extracts queries and search engines used from it, and feeds the resulting data into the Relevance Assessment Tool.

# MODEL IN PRACTICE



# SEARCH LOGGER OUTPUT

	User	User Action	Action Description	Timestamp	Queries	Reformulations
1						
2	24	User started a new SC	Displaying pre-SC form for SC index 10;;	09.08.2011 12:20:55	0	
3	24	User opened a tab	User opened a tab;;	09.08.2011 12:20:55	0	
4	24	User submitted pre-SC form	(a=(5))(b=(5))(c=(5))(d=(1))(formID=(search_case_11))	09.08.2011 12:21:48	0	
5	24	User performed a Google WEB search	http://www.google.de/search?hl=de&client=firefox-a&hs=	09.08.2011 12:22:45	Ländernamen in Chemischen Elemente	new
6	24	User viewed a related link	http://de.wikipedia.org/wiki/Etymologische_Liste_der_che	09.08.2011 12:23:21	0	
7	24	User viewed a related link	http://www.cosmiq.de/qa/show/531055/welches-land-wu	09.08.2011 12:24:22	0	
8	24	User viewed a related link	http://de.wikipedia.org/wiki/Etymologische_Liste_der_che	09.08.2011 12:25:29	0	
9	24	User performed a Google WEB search	http://www.google.de/search?q=argentum&ie=utf-8&oe=i	09.08.2011 12:25:46	argentum	new
10	24	User performed a Google WEB search	http://www.google.de/search?hl=de&client=firefox-a&hs=	09.08.2011 12:26:09	argentum chemische elemt	narrowing
11	24	User performed a Google WEB search	http://www.google.de/search?hl=de&client=firefox-a&hs=	09.08.2011 12:26:13	argentum chemische element	change
12	24	User viewed a related link	http://de.wikipedia.org/wiki/Silber;;Silber - Wikipedia	09.08.2011 12:26:20	0	
13	24	User finished a SC	Displaying post-SC form for SC index 10;;User has presuma	09.08.2011 12:27:06	0	
14	24	User opened a tab	User opened a tab;;	09.08.2011 12:27:06	0	
15	24	User submitted POST-SC form	(a=(5))(b=(1))(c=(1))(d=(1))(formID=(search_case_11_2))	09.08.2011 12:27:17	0	

# RAT USER INTERFACE

## Relevance Assessment Tool

Fortschritt: 0%  100%

### Suchaufgabe:

Finden Sie Informationen über die HAW Hamburg

### Suchbegriffe:

haw hamburg

Bitte Bewerten Sie die Relevanz des Dokuments in Bezug auf die oben genannte Suchaufgabe auf einer Skale von 0 - 4?

- 0 - nicht relevant
- 1 - kaum relevant
- 2 - teilweise relevant
- 3 - relevant
- 4 - sehr relevant

Ist das HAW-Logo auf jeder Seite korrekt dargestellt?


- ja
- nein

Hier können Sie die Relevanz kommentieren:

Eingabe:

Nächstes Ergebnis

Überspringen



HAW Hamburg

Deutsch English Français Espagnol  
(language available when flag highlighted)

English  
Aktuell  
Studiengänge

**Bewerbung**

- Bewerbungsfristen
- Deutsche, EU-Bewerber/innen**
  - Voraussetzungen
- Bewerbung**
  - Studienplatzvergabe
  - Hochschulwechsel
  - Studiengangswechsel
  - Gaststudierende
  - Nicht EU Bewerber/innen
  - International Guest Students
  - Masterstudiengänge
  - Vorkurse

International  
Formales  
Beratung  
Campusleben  
Schulcampus  
FAQ  
Team Studierendenzentrum

## Studium

Design, Medien und Information | Life Sciences | Technik und Informatik | Wirtschaft und Soziales

Studium > Bewerbung > Deutsche, EU-Bewerber/innen > Bewerbung

### Online Bewerbung

An der HAW Hamburg bewerben sich Staatsangehörige aus Deutschland, der EU & EWR sowie Ausländer/innen und Staatenlose, die ihre [Hochschulzugangsberechtigung](#) in Deutschland erworben haben, innerhalb der Bewerbungsfristen online für ihre Studiengänge. Folgende Schritte sind (in den meisten Fällen) erforderlich:

- 1. Für einige Studiengänge ist die Bewerbung nur zum Winter- oder nur zum Sommersemester möglich. Informieren Sie sich auf der [Studiengangsübersichtsseite](#)!
- 2. Beginnen Sie mit dem [Selbsttest](#), der Ihnen bei der Entscheidung für ein Studienfach helfen soll. Er ist bis auf zwei Ausnahmen für die Bewerbung aller Studiengänge obligatorisch. [Übersicht](#)

Für die Onlinebewerbung ist eine Registrierung beim Selbsttest notwendig. Bitte bewahren Sie die Zugangsdaten

bis zu Ihrer endgültigen Onlinebewerbung auf.

- 3. Merken Sie sich, welche E-Mail-Adresse Sie beim Selbsttest angegeben haben und geben Sie diese in dem entsprechenden Feld im Kasten (siehe unten) an! Jetzt können Sie mit der Online-Bewerbung beginnen.
- 4. Drucken Sie sich am Ende Ihrer Online-Bewerbung Ihren Zulassungsantrag aus.
- 5. Reichen Sie den Zulassungsantrag zusammen mit den angeforderten Unterlagen im Studierendensekretariat ein, zum Beispiel mit der Begründung Ihres [Härtefallantrags](#) oder mit einer einfachen Kopie Ihrer Hochschulzugangsberechtigung. Für **nichtdeutsche** Zeugnisse ist mit der [beglaubigten Kopie](#) der [Hochschulzugangsberechtigung](#) eine amtliche Übersetzung einzureichen.

### Bewerbungsfristen 2011/2012

Wintersemester 2011/12: 1. Juni - 15. Juli  
Sommersemester 2012: 1. Dezember - 15. Januar

**Ausnahmen:**

Information Engineering (Bachelor):  
1. April – 15. Juli (Bewerbungen nur zum Wintersemester möglich)

Public Health (Master):  
1. April - 30. September (Bewerbungen jeweils für das kommende Sommersemester möglich)

Renewable Energy Systems (Master):  
1. April - 15. Juli (Bewerbungen nur zum Wintersemester möglich)

Health Sciences (Master):  
01. April - 15. Juli (Bewerbungen nur zum Wintersemester möglich)

### Sie haben sich bereits beworben?

[Hier](#) können Sie den Bearbeitungsstand & Bewerbungsstatus abfragen sowie Zulassungs- oder Ablehnungsbescheid anfordern



# METHODS FOR USER STUDY

## Convenience sample of 64 participants

- Average age: 37.2 years (SD=12.29)
- 31 female; 33 male

## Tasks

- Two tasks per user (one simple, one complex)
- Tasks randomly selected from a total of 60

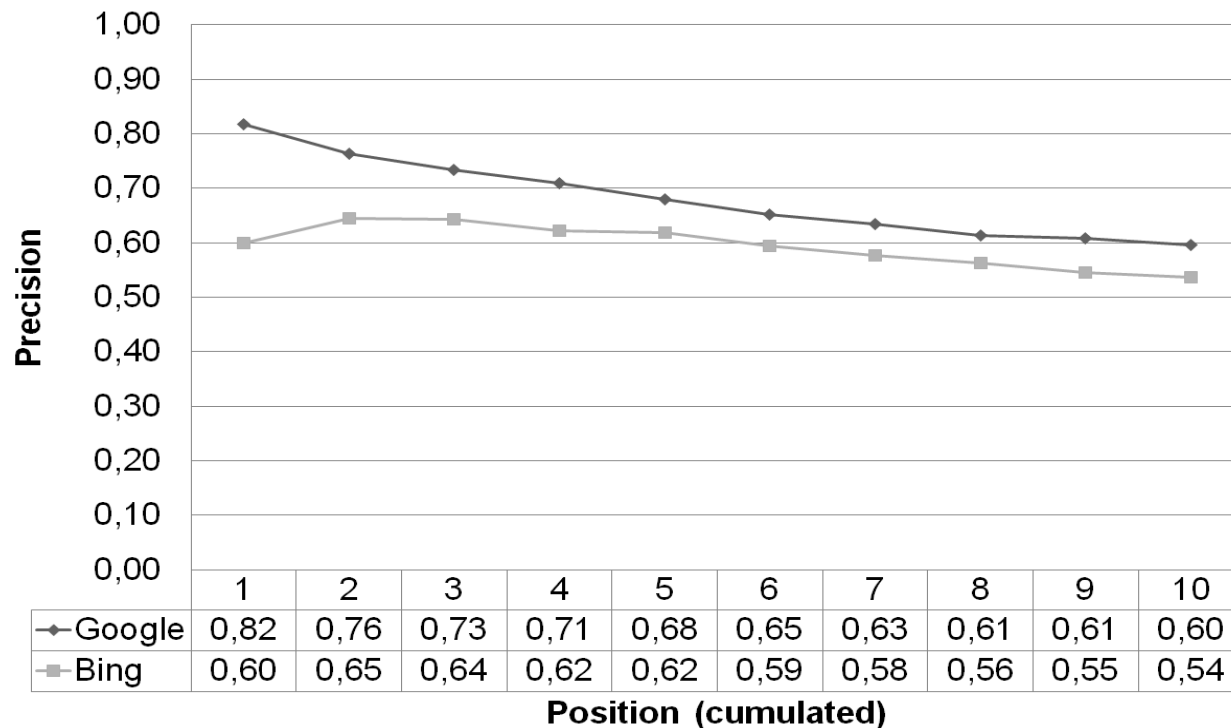
## Data collection

- Lab study in Hamburg, Germany
- 30 mins. per participant.
- Queries extracted from the sessions were sent to Google and Bing; top 10 results collected and judged → 1,156 results from Google; 1,132 from Bing.

# RESULTS

# GOOGLE RESULTS ARE, ON AVERAGE, JUDGED BETTER ON ALL RESULTS POSITIONS\*

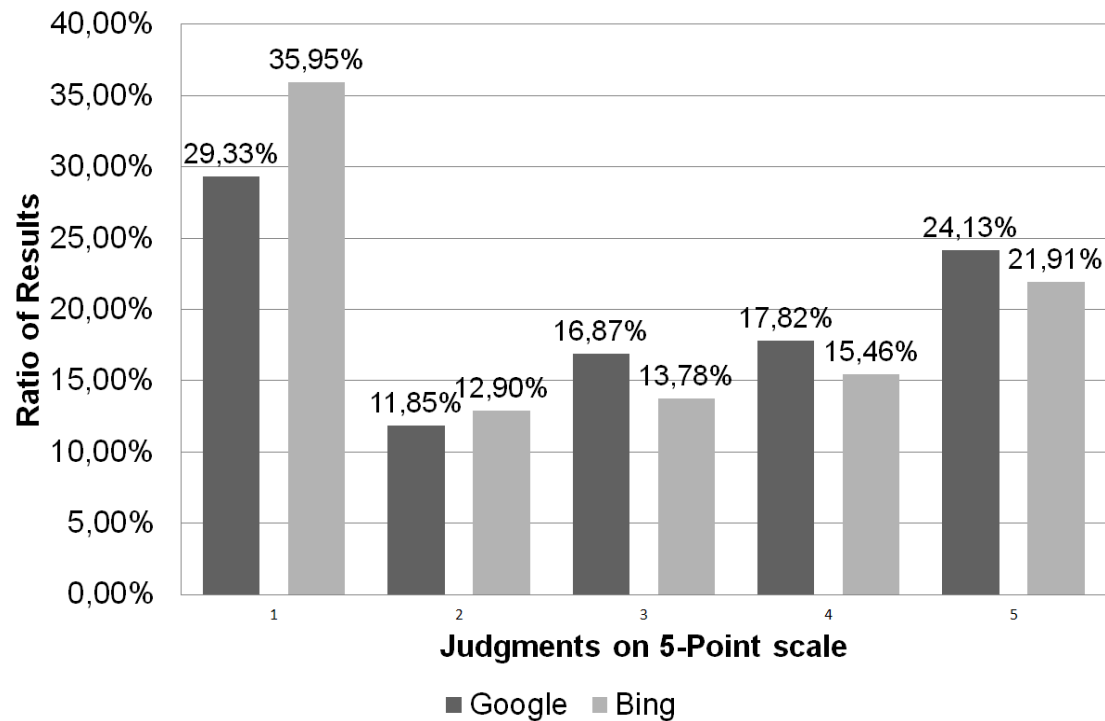
## Precision graph for Google and Bing results



\* No significant differences, though, expect for MAP@5.

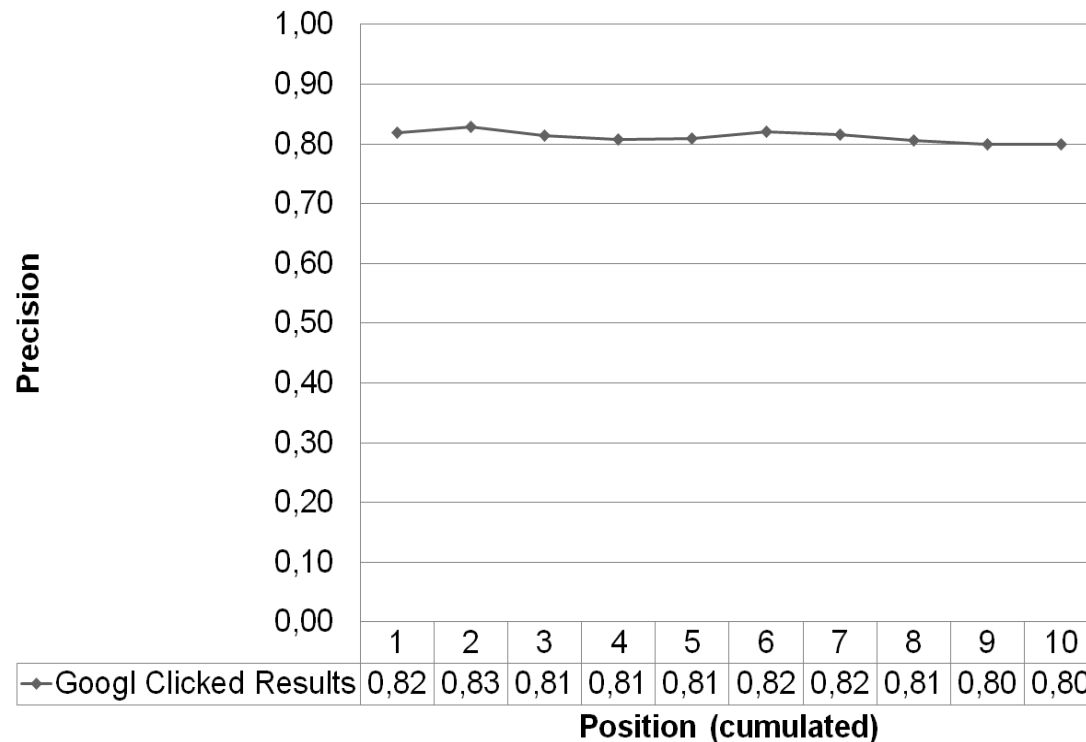
# USING THE MORE DIFFERENTIATED 5-POINT SCALE, GOOGLE OUTPERFORMS BING

## Distribution of judgments on the 5-point scale



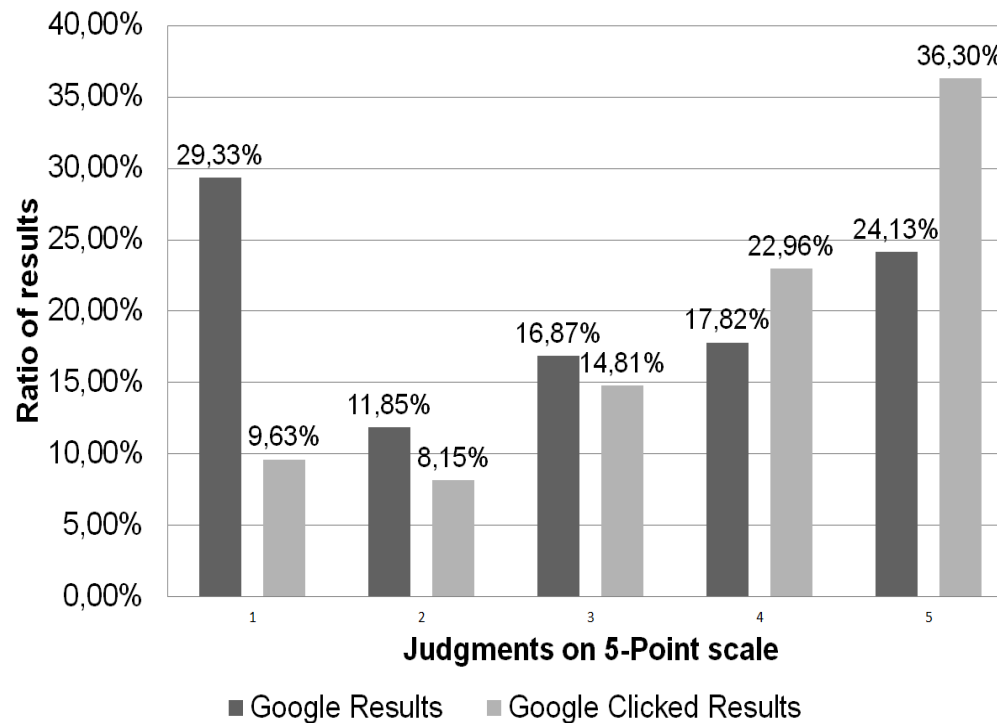
# USERS ARE WELL ABLE TO JUDGE THE RELEVANCE OF THE RESULTS PRESENTED BASED ON SNIPPETS ON THE SERPS

## Precision graph for clicked results (Google)



# USERS ARE WELL ABLE TO SELECT RESULTS THAT THEY LATER JUDGE AS HIGHLY RELEVANT

## Distribution of judgments for clicked results



# RESEARCH QUESTIONS ANSWERED

**RQ1: To what extent do users use different search engines when working on search tasks given by the experimenter?**

→ Users solely used Google.

**RQ2: Do users judge results they actually selected when performing their tasks as more relevant than results not selected during the task?**

→ Results actually selected are judged better than results not selected. Snippets are helpful and users can use them to their benefit.

**RQ3: Would users be equally satisfied with the results of another search engine; i.e., would Google users be equally satisfied with results from Bing?**

→ Results from Google were judged as being more relevant than those from Bing but the differences are not too big.

# SUMMARY AND CONCLUSION



# SUMMARY

With the proposed method, we can measure:

- (1) The relevance of the results users selected in their search sessions,
- (2) The relevance of the results users were presented with in their search sessions *but which they did not click on*,
- (3) The relevance of the results from a competing search engine *which they did not see in their search session*.

The user study confirmed the feasibility of the approach and showed how it can be used to answer research questions that cannot be addressed well using other approaches.

# CONCLUSION AND FUTURE RESEARCH

- It matters to users which search engine is used (they prefer Google). However, the differences in results quality are small and we assume that users would also be satisfied using Bing.
- Approach combines elements from information retrieval evaluation and from information behaviour research.
- While the study presented is a lab study, the same approach could also be used in a natural settings → “Firefox to go”
- While the software toolkit is not available for download, we invite researchers to contact us if they are interested in using it for their research.

# THANK YOU

Sebastian Sünkler, 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

For references, additional readings and further data analysis, see the paper available at <https://arxiv.org/abs/1710.08390>