



Hasso
Plattner
Institut

IT Systems Engineering | Universität Potsdam

Google™

YAHOO!®



Search Engine Evaluation

Franziska Häger, Lutz Gericke

23.07.2009

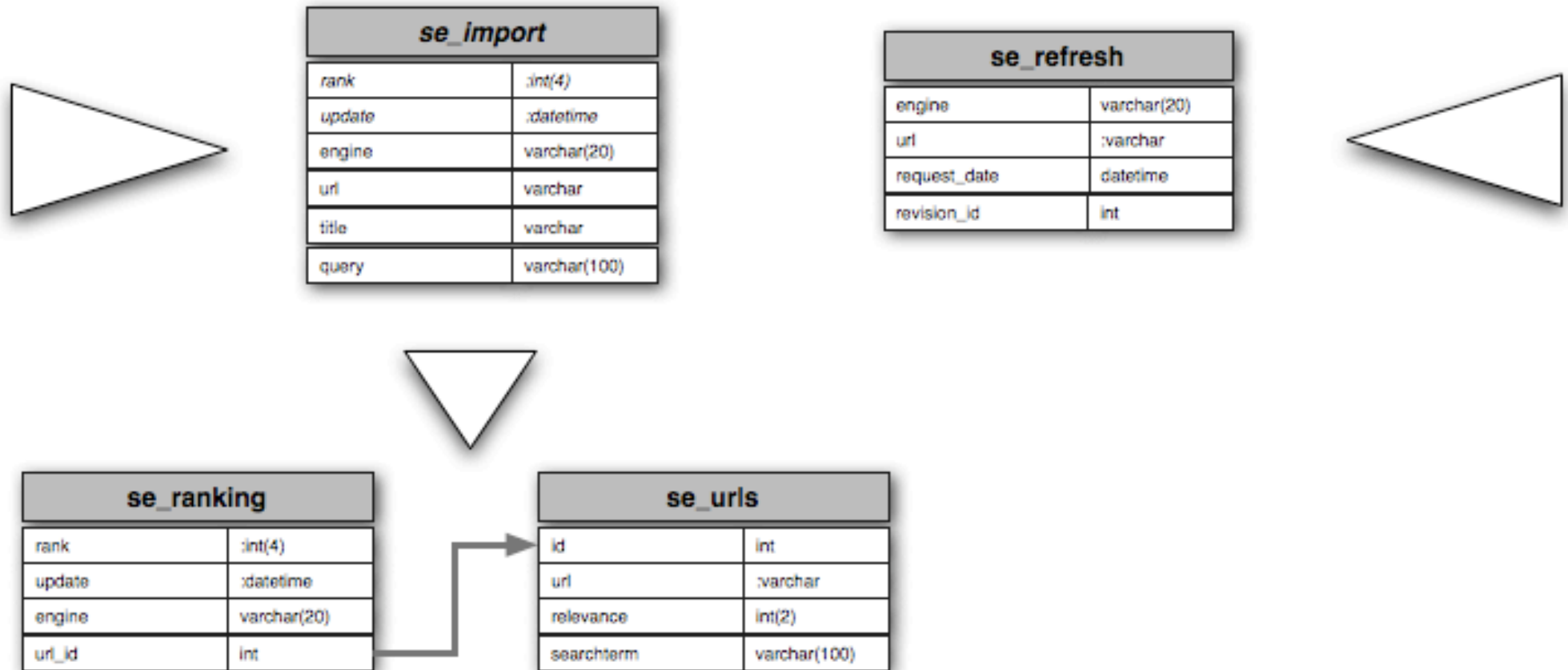
- Tägliche Abfragen mit geringer Ergebnismenge
 - Evaluation von Precision und Recall
- Revisionsabfrage aus Suchmaschinen-Caches
 - Bestimmung der Aktualität
- Einzel-Queries zum Test von speziellen Suchfeatures
 - Stemming, Rechtschreibkorrektur, Stop Words, ...
- Kleiner Engine Test im Kurs
 - Usability
 - Unterschiedliche Herangehensweisen differenzieren

- Anfragen
 - „supermassive black hole“ physics
 - „down quark“ „particle physics“
 - „magical michael“
 - magical michael
- Alle unter 100 Ergebnisse (außer ‚magical michael‘)
- Relevance von Hand bestimmt, Modell dazu:
 - 1-sehr relevant; 2-eher relevant;
3-eher nicht relevant; 4-gar nicht relevant

- Python-Skripte
 - Ranking der Testanfragen abfragen (24h)
 - Aktualität der Caches überprüfen (2h)
- Statistik-Skripte
 - Recall/Precision-Ranking Graphen
 - Cache-Age Development Graph
- Erfassung der Daten in MySQL

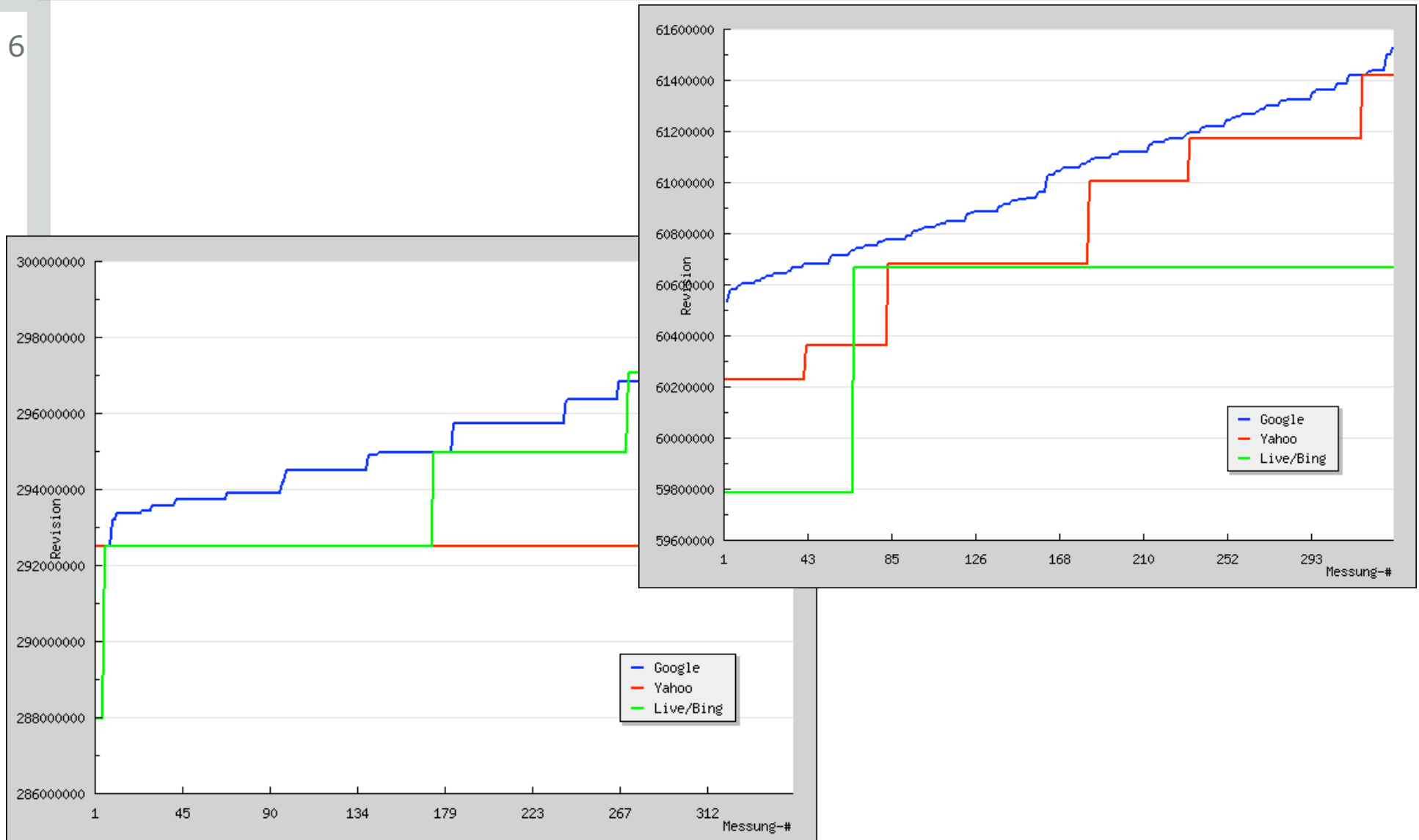
DB-Modell

5



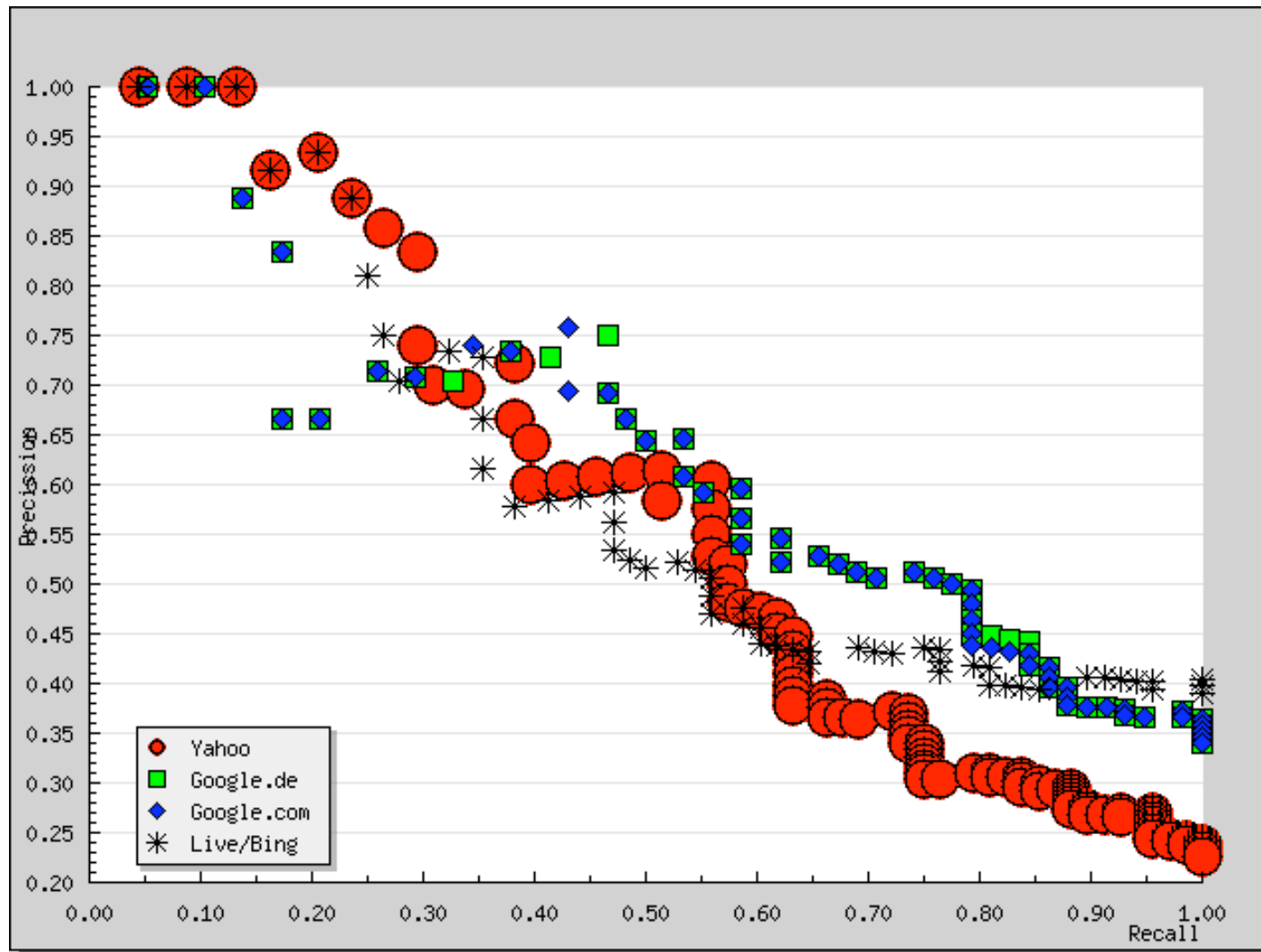
Cache-Age Development Graphs

6



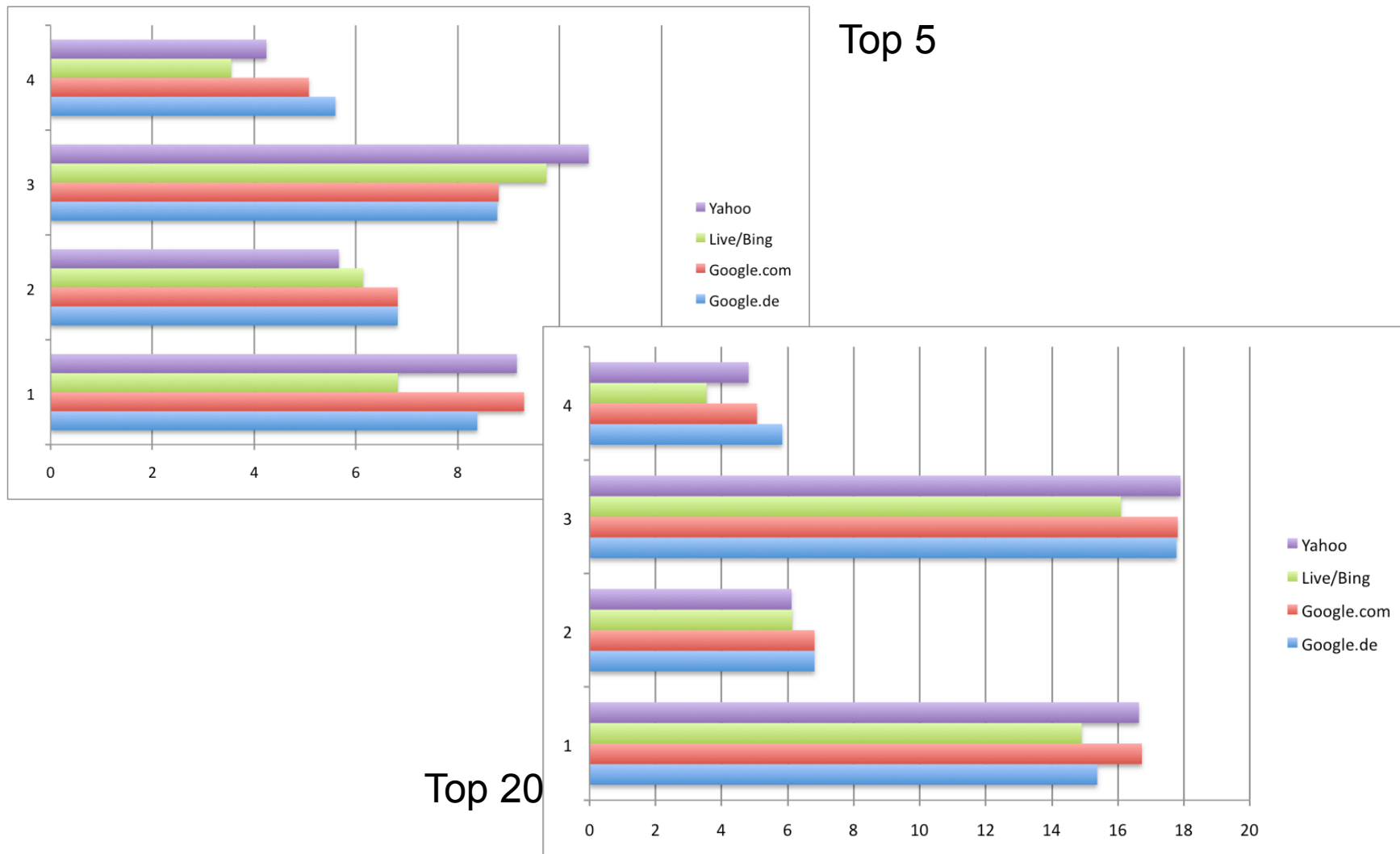
Ranking-Graphs

7



Discounted Cumulative Gain

8



Search Engine Features

Meta Suchworte und short cuts (Bsp: define, site:, url:,....)	+	+	+
Querysprache ()	-	-	+
Querysprache ..	+	-	-
Querysprache "query"	+	+	+
Querysprache +	+	+	+
Querysprache - / NOT	+	+	+
Querysprache & / AND	D	D	D
Querysprache / OR	+	+	+
Hilfe / Erklärungen	+	+	+
Umrechnen von Einheiten (Bsp: 13km in miles)	+	+	-
Rechnen (Bsp: 3*4+2)	+	+	+
	Google	Yahoo	Bing

|+ möglich | D per default | - nicht möglich |

Search Engine Suchmöglichkeiten

	Reisesuche	-	+	-
	Creative Commons Suche	*	+	-
	Lokale Suche	*	+	-
	Personensuche	-	+	-
	Jobsuche	-	+	-
	Suche für Wiss. Dokumente	+	-	-
	Blogsuche	+	-	-
	Buchsuche	+	-	-
	Starsuche	-	-	+
	Codesuche	+	-	-
	Maps	+	+	+
	Produktsuche (Shopping)	+	+	+
	Newssuche	+	+	+
	Bilder- / Videosuche	+	+	+
	erweiterte Websuche	+	+	+
	Standard Websuche	+	+	+
Google		+	+	+
Yahoo		+	+	+
Bing		+	+	+

|+ möglich | * indirekt möglich | - nicht möglich |

- Angebotenen Möglichkeiten sind recht ähnlich, Standardsuche plus weitere Suchen, Querysprache und Metasuchen nahezu identisch.
- Spezial Features und Spezialsuchen sind jedoch nicht überall gleich
- Hier entscheidet was man eigentlich sucht, welche Suchmaschine am geeignetsten ist.

■ Stemming

- Bored blues band
- Swim pool

■ Getrennt zusammenschreibung

- Schiff fahrt
- Imallgemeinen

■ Stop Words

- Nicht krank sein
- Nicht sein
- To do

■ Rechtschreibung

- Apstieg
- Schifffahrt / Schiffffahrt
- nähmlich

- Alle Suchmaschinen verwenden die untersuchten Techniken.
- Einfluss beim Ranking scheint unterschiedlich
- Google scheint semantischer zu arbeiten (siehe „nicht sein“)
- Yahoo und Bing machen Rechtschreibkorrektur nicht nach aktuellem Wörterbuch? (siehe Schiffahrt)
- Bing gibt keine Hinweise wenn Ergebnisse nicht genau die gesuchten Worte enthalten

Auswertung der Umfrage

15

- Leider nur 7 Antworten
- Auswertung pro Suchmaschine nicht wirklich sinnvoll
- Aufgabe 1: Anfragen 1-2, Lösungsseite meistens Berlinale.de, Anfragen-Länge 3-4
- Aufgabe 2: Anfragen 1-2, Lösungsseite meistens Wikipedia, Anfragen-Länge 2-6
- Ausgabe 3: Anfragen 2-9, Lösungsseiten alle verschieden, Anfragen-Länge 3-5, teilw. nicht gelöst
- Aufgabe 4: Anfragen 1-2, Lösungsseiten verschieden, Anfragen-Länge 2-3

- **Auswertung der Fragen zur Nutzung:**
- Häufigste Suchmaschine: Google
- Wenn weitere Suchmaschinen verwendet werden, dann seiten intern oder für spezielle Suchen
- Query-Länge: zwischen 1,5 und 4 (sehr verschieden)
- Query-Sprache wird fast nicht genutzt, wenn dann AND
- Erweiterte Suche wird wenn nötig genutzt

- **Auswertung der Fragen zu Spezialfeatures:**
- **Verwendete Spezialfeatures:** Im Cache, Währungsrechner, Taschenrechner, Übersetzung, ähnliche Seiten, Auf gut Glück!,
- **Verwendete Suchmöglichkeiten:** Bildersuche, Maps, Videosuche

Auswertung der Umfrage -Fazit

18

- Der Kurs hatte keine Lust mehr auf Suchmaschinen ... 😊
- Query-Sprache wird kaum genutzt
- Von den Features werden vor allem Bildersuche, Maps und Übersetzung genutzt
- Man googelt mit Google! 😊

Ende

Unterschied Google.de/.com

20

