

openHPI: **Wie funktioniert eine Suchmaschine?**
Interaktion: Anfragen

Dr. Ralf Krestel
Hasso-Plattner-Institut
Universität Potsdam

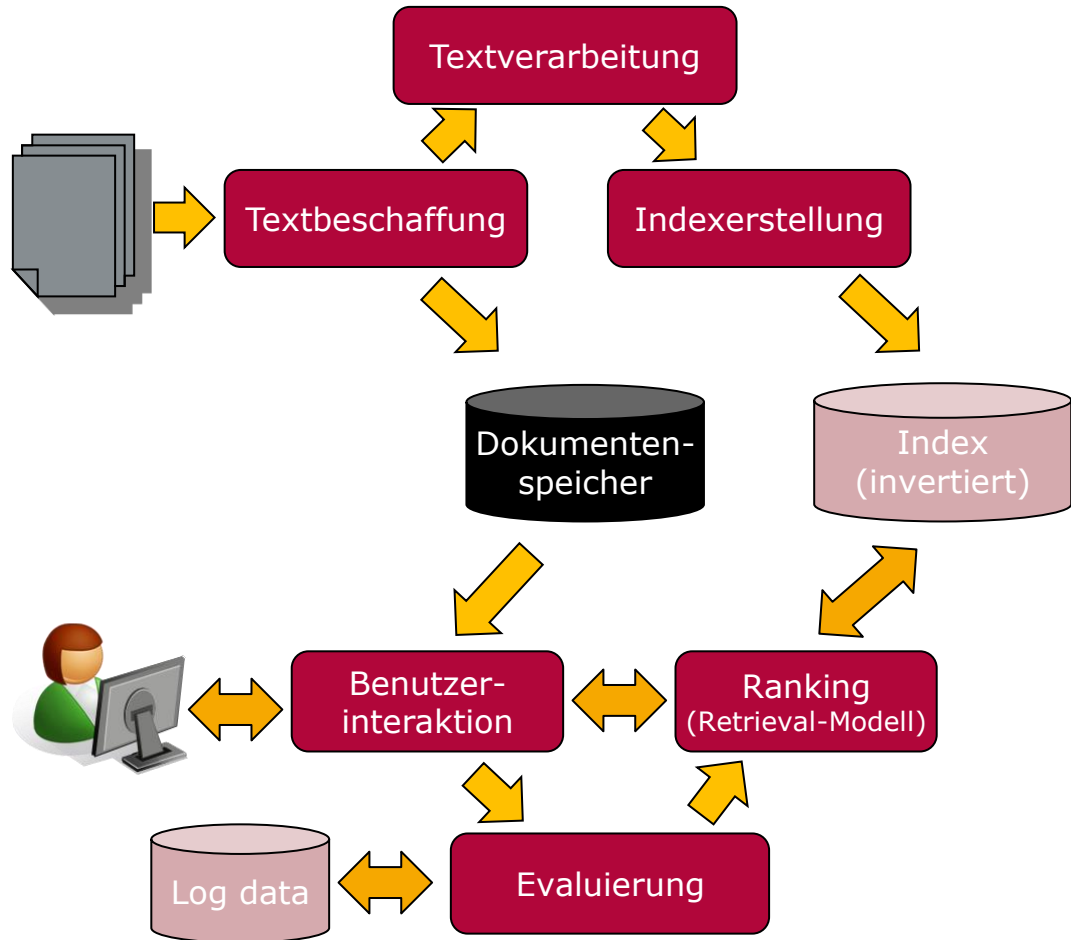
■ Grundlegende Bausteine

□ Indexierung

- Textbeschaffung
- Textverarbeitung
- Indexerstellung

□ Anfrage

- Benutzerinteraktion
- Ranking
- Evaluierung



- **Benutzerinteraktion** vollzieht sich mit Hilfe der **Benutzerschnittstelle**
 - Oberfläche dient als **Vermittler** zwischen Nutzer und Suchmaschine
- Benutzerschnittstelle soll den Suchenden beim **Verstehen** und **Formulieren** ihrer Informationsbedürfnisse helfen
- Weitere konkrete Aufgaben: Nutzer helfen
 - **Anfragen** zu formulieren
 - aus **Informationsquellen** auszuwählen
 - **Ergebnisse** zu verstehen
 - **Suchfortschritt** festzuhalten

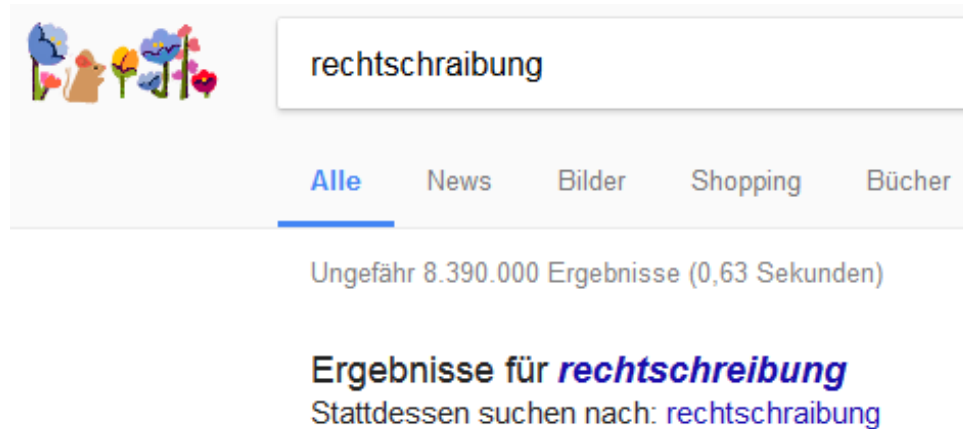
- Anfrage = **Übersetzung** des Informationsbedürfnisses eines Benutzers für die Suchmaschine
- ASK-Hypothese (Belkin et al. 1982):
 - „Es fällt Benutzern schwer, ihr **Informationsbedürfnis** genau zu artikulieren, da die gesuchte Information eine **Lücke** in ihrem Wissen darstellt.“
- **Benutzerinteraktion** und **Kontext** wichtig, um Anfrage zu verstehen
- **Benutzerinteraktion** passiert
 - bei der Formulierung und Umformulierung der Anfrage
 - beim Durchsehen/Anklicken der Ergebnisse
- Benutzer können nicht den **Rankingalgorithmus** einer Suchmaschine ändern, aber die **Ergebnisse** durch **Interaktion**
- Informationsbedürfnis kann durch Benutzerinteraktion spezifiziert werden
 - Gleiche Ausgangsanfrage, unterschiedliche Informationsbedürfnisse

- Einfach, natürlichsprachige Schnittstellen ermöglichen Suche auch für nicht Experten
- Suchmaschinen verstehen meist nur **Stichwörter**, nicht Fragen
- Die meisten Nutzer haben sich angepasst und stellen keine Fragen, sonder verwenden Stichwörter
 - Durchschnittliche Länge einer **Webanfrage**: 2,3 Wörter
 - Durchschnittliche Länge einer Forumsfrage (**community-based question answering** (CQA)): 30 Wörter
 - Auswahl der Stichwörter nicht immer einfach
 - Anfragespezifizierungsmethoden können helfen



- Textverarbeitung der Anfrage wie der Dokumente vor dem Indexieren
 - Ein paar Unterschiede bei Stopwortentfernung und Stemming
 - Ein paar Dinge nicht notwendig (keine Struktur)
- **Anfragebasierte Stopwortentfernung**
 - Entfernen der Stoppwörter in der Anfrage
 - Behalten von Stoppwörtern im Index
 - Flexibler bei Anfragen, die Stoppwörter enthalten
 - Stoppwörter in der Anfrage können
 - wie normale Wörter behandelt werden
 - entfernt werden
 - unter Umständen entfernt werden (falls kein + davor steht)
- **Anfragebasiertes Stemming**
 - Dokumente werden nicht gestemmt, Anfrage wird mit Wortvarianten erweitert
 - Alternativ: Indexieren von Wörtern und ihrer Wortvarianten

- Hilfe durch die Suchmaschine bei der Formulierung der Anfrage in Form von automatischer **Rechtschreibkorrektur**



- Wichtiger Teil beim Verarbeiten von Webanfragen
 - 10-15% aller Webanfragen haben Rechtschreibfehler

- Allgemeiner Ansatz: Suchmaschine schlägt Korrektur für Wörter vor, die nicht in einem **Wörterbuch** vorkommen
 - Hilft nicht immer: „Information Retrieval Kurz“
- Korrektur durch Vergleich der Anfragewörter mit Wörtern im Wörterbuch mit Hilfe eines **Ähnlichkeitsmaßes**
- Am häufigsten verwendet: **Editierdistanz**
 - Minimale Anzahl von Operationen auf Buchstabenebene, um von einem Wort zum anderen zu kommen
 - Löschen, Ersetzen, Einfügen, Vertauschen (Damerau-Levenshtein-Distanz)
- 80% aller Rechtschreibfehler bei Webanfragen haben DLD=1
 - Fehleer → Fehler (Löschen)
 - Fehlär → Fehler (Ersetzen)
 - Feler → Fehler (Einfügen)
 - Felher → Fehler (Vertauschen)

Interaktion: Anfragen

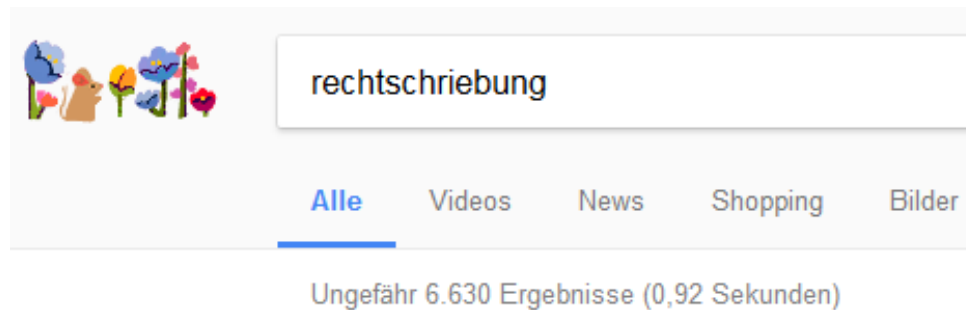
Rechtschreibkorrektur III

488941 britney spears	147 breatny spears	54 britneay spears	26 britneyt spears	17 brittanie spears
40134 brittany spears	147 brittiney spears	54 britner spears	26 brittan spears	15 brinney spears
36315 brittney spears	147 britty spears	54 britney's spears	26 brittne spears	15 britten spears
24342 britany spears	147 brotney spears	54 britnye spears	26 btittany spears	15 briterney spears
7331 britny spears	147 brutney spears	54 britt spears	24 beitney spears	15 britheny spears
6633 briteny spears	133 britteney spears	54 brttany spears	24 birteny spears	15 britneny spears
2696 britteny spears	133 briyney spears	48 bitany spears	24 brightney spears	15 brittamy spears
1807 briney spears	121 bittany spears	48 briny spears	24 brintiny spears	15 brittmey spears
1635 brittny spears	121 bridney spears	48 brirney spears	24 britanty spears	15 brytnei spears
1479 brintey spears	121 britainy spears	48 britant spears	24 britenny spears	15 btirney spears
1479 britanny spears	121 britmey spears	48 britnety spears	24 britini spears	15 rittney spears
1338 britiny spears	109 brietney spears	48 brittanny spears	24 britnwy spears	14 brinet spears
1211 britnet spears	109 brithny spears	48 brttney spears	24 brittni spears	14 britneyy spears
1096 britiney spears	109 britni spears	44 birttany spears	24 brittnie spears	14 britten spears
991 britaney spears	109 brittant spears	44 brittani spears	21 biritney spears	12 beritney spears
991 britnay spears	98 bittney spears	44 brityney spears	21 birtany spears	12 bretiny spears
811 brithney spears	98 brithey spears	44 brtitney spears	21 biteny spears	12 briatny spears
811 brtiney spears	98 brittiany spears	39 brienty spears	21 bratney spears	12 brieney spears
664 birtney spears	98 btitney spears	39 brritney spears	21 britani spears	12 brinany spears
664 brintney spears	89 brietny spears	36 bbritney spears	21 britanie spears	12 britaany spears
664 briteney spears	89 brinety spears	36 briitany spears	21 briteany spears	12 britan spears
601 bitney spears	89 brintny spears	36 britanney spears	21 brittay spears	12 britine spears
601 brinty spears	89 britnie spears	36 briterny spears	21 brittinay spears	12 britnea spears
544 brittaney spears	89 brittey spears	36 britneey spears	21 brtany spears	12 britnes spears
544 brittnay spears	89 brittnet spears	36 britnei spears	21 brtiany spears	12 britnez spears
364 britey spears	89 brity spears	36 britniy spears	19 birney spears	12 britnny spears
364 brittiny spears	89 ritney spears	32 britbey spears	19 brirtney spears	12 brittenay spears
329 brtney spears	80 bretny spears	32 britneu spears	19 britnaey spears	12 brittneys spears
269 bretney spears	80 britnany spears	29 britent spears	19 britnee spears	12 brittony spears
269 britneys spears	73 brinteney spears	29 brittnany spears	19 britony spears	12 brteny spears
244 britne spears	73 brittainy spears	29 britttany spears	19 brittanty spears	12 btitany spears
244 brytney spears	73 pritney spears	29 btiney spears	19 britttney spears	12 btittney spears
220 breatney spears	66 brintany spears	26 birttney spears	17 birtny spears	12 btriney spears
220 britiany spears	66 britnery spears	26 breitney spears	17 brieny spears	12 btritney spears
199 britnney spears	59 briitney spears	26 brinity spears	17 brintty spears	12 rittany spears
163 britnry spears	59 britinay spears	26 brittenay spears	17 brithy spears	

Interaktion: Anfragen

Ähnliche Anfragen Vorschlagen

- Hilfe durch die Suchmaschine bei der Formulierung der Anfrage in Form von **Query Suggestion**
 - **Vorschlagen** von Anfragen
 - **Ähnliche** Anfragen



Ähnliche Suchanfragen zu openHPI

openhpi **app**

openhpi **lösungen**

openhpi **erfahrungen**

hasso plattner institut email prüfen

openhpi **python**

hpi open course design thinking

hasso plattner institut design thinking

opencourseworld

- Beste Quelle: **Anfragen-Logdatei** (query log)
 - Protokolliert alle **Anfragen**
 - Speichert auch angezeigte **Ergebnisseiten**
 - Und Speicherung der **Klicks** (click log)
- Beispielwörter, welche häufig zusammen mit “Zierfische” gesucht werden: kaufen, bilder, lebend, arten, süßwasser, blau, aquarium, zubehör,...
- **Anfrageerweiterung** durch hinzufügen von Wörtern (“Zierfische kaufen”)
- Gruppieren von Anfragen basierend auf **Angeklickten Ergebnissen**:
 - Jede Anfrage wird durch die Seiten dargestellt, die die Benutzer bei dieser Anfrage angeklickt haben
 - Anfragen sind ähnlich, wenn die angeklickten Seiten stark überlappen
 - „hpi” und „Hasso Plattner Institut”

Bester Treffer für beide Anfragen: www.hpi.de