

openHPI: **Wie funktioniert eine Suchmaschine?**
Wie wird Gesucht?

Dr. Ralf Krestel
Hasso-Plattner-Institut
Universität Potsdam

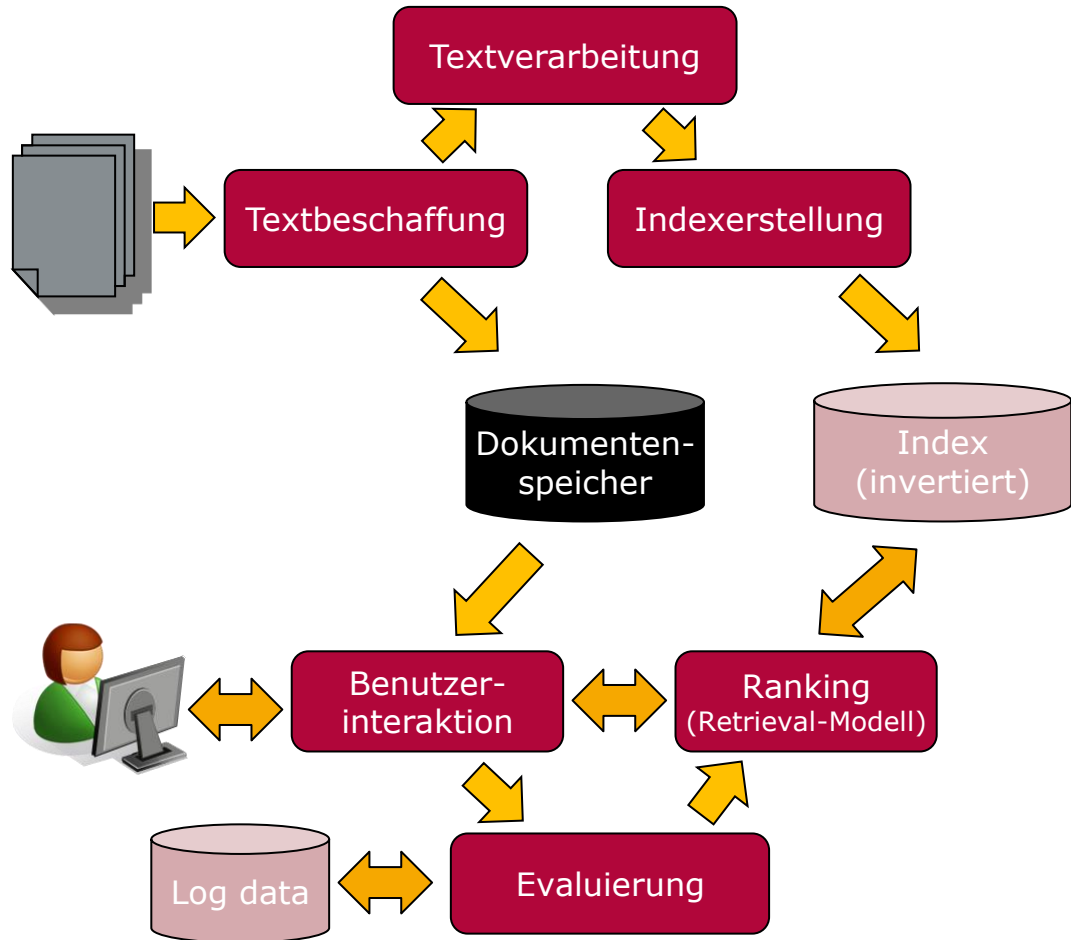
■ Grundlegende Bausteine

□ Indexierung

- Textbeschaffung
- Textverarbeitung
- Indexerstellung

□ Anfrage

- Benutzerinteraktion
- Ranking
- Evaluierung



- Anfragetypen im Web (Broder 2002):
- **Navigationsorientiert** (Wiederfinden) 20%
 - Suche nach einem bereits bekannten Dokument/Seite
 - „Facebook“, „hpi“
- **Informationsorientiert** (Wissen) 50%
 - Suche nach Fakten (konkreter Informationsbedarf)
 - „Hauptstadt von Mali“
 - Suche nach umfassendem Überblick (problemorientierter Informationsbedarf)
 - „Was ist Demokratie“
- **Transaktionsorientiert** (Handlung) 30%
 - Suche nach Seiten, um etwas zu machen (kaufen, anmelden, downloaden)
 - „Flug nach Frankfurt buchen“

**Klassische
Recherche**

- Interaktionen von Nutzern einer Suchmaschine unterscheiden sich basierend auf:
 - **Art** der Suchaufgabe
 - **Fachwissen** des Informationssuchenden
 - **Zeit und Aufwand**, die investiert werden können
- Klassische Recherche (oder **explorative Suche**) (Marchionini 2006)
 - Aneignen von Wissen („**Learn**“)
 - **Mehr nötig als eine** Anfrage und ein Antwortdokument
 - **Zeitintensiv**: Viele Informationen verarbeiten, zusammenführen, verstehen
 - Erforschen eines Themas („**Investigate**“)
 - Kann sich über **längere Zeit** hinziehen
 - **Kritisches** Hinterfragen der gefundenen Informationen
 - **Analysieren**, Zusammenführen, **Auswerten** einzelner Informationsteile
 - Zur Vorhersage, Planung oder Wissenserweiterung
 - **Alle** relevanten Informationen, Erkennen von **Wissenslücken**

- Informationssuche kann klassisch als **vierstufiger Prozess** angesehen werden:
 1. Problemerkennung
 2. Formulierung des Informationsbedürfnisses
 3. Formulierung einer Anfrage
 4. Auswerten der Ergebnisse
- Neuere Modelle betonen mehr den **dynamischen** Charakter des Prozesses
 - Nutzer lernen während sie suchen
 - Informationsbedürfnis wird aufgrund der Ergebnisse angepasst
- „**Berry-Picking**“ (Beerenpflücken)
 - Zusammensuchen von Informationsteilen auf der Ergebnisseite
- Heutige Suchmaschinen erlauben sehr einfach **Umformulierungen** basierend auf Ergebnissen

50% der Nutzer von Websuche

- Navigation: Der Suchende schaut sich die Informationsstruktur an und hangelt sich über Links zu seinem Ziel
- Gute Strategie, falls die Struktur zu dem Informationsbedürfnis passt
 - Einfacher Information wiederzuerkennen, als zu erinnern
 - Voraussetzung: Passende Links verfügbar
 - Ohne Links sehr frustrierend
- Beispiel:
 - Informationsbedürfnis: Treiber für bestimmten Laserdrucker
 - Nutzer klickt auf *printers*, dann *laser printers*, dann:
 1. *HP laser printers*
 2. *HP laser printers model 9750*
 3. *software for HP laser printers model 9750*
 4. *software drivers for HP laser printers model 9750*
 5. *software drivers for HP laser printers model 9750 for the Win98 operating system*

- Viele Studien zum Thema Suchprozess
 - Hinweise für gutes Design
- Nutzer formulieren Anfragen oft **leicht** um
- Nutzer suchen oft Informationen, die sie schon mal gesucht haben
 - Suchstrategien unterschieden sich
 - Suchmaschinen bieten Anfragehistorie und personalisiertes Ranking
- Nutzer können oft schwer einschätzen, ob ein Dokument relevant ist
 - Je unwissender ein Nutzer in Bezug auf ein Thema, desto schwerer
- Nur die ersten paar Ergebnisse werden angesehen
 - Qualität der ersten paar Treffer im Vergleich zum Rest wird stark überschätzt
- Nutzer können ganz schlecht einschätzen, wie viele der relevanten Dokumente sie gefunden haben (Recall)
- Experten suchen anders als Anfänger
 - Geduldiger, sorgfältiger → bessere Ergebnisse