

Suchmaschinen Workshop

1. Aufgabe: Verschiedene Informationssysteme

- a) Nennen Sie Vor- und Nachteile von Hypertextinformationssystemen.
- b) Charakterisieren Sie die Unterschiede zwischen Faktenretrieval und Information Retrieval.
- c) Wodurch zeichnen sich Expertensysteme als Informationssysteme aus?

2. Aufgabe: Information Retrieval

- a) Erklären Sie den Begriff Information Retrieval und geben Sie einen Überblick der häufigsten Funktionen.
- b) Beschreiben Sie kurz die Aufgabenstellung einer Digitalen Library' als speziellen Fall eines Informationssystems.
- c) Nennen Sie die drei Probleme, die bei der Implementation eines solchen Systems gelöst werden müssen.
- d) Überlegen Sie wie der typische IR-Prozess für diese Art von Systemen aussieht.

3. Aufgabe: Evaluation

Die Ausgabe eines IR-System hat die folgende Hitliste: d0,d1,d2,d3,d4,d5,d6,d7,d8,d9.
Sie bestimmen die Dokumenten {d0,d2,d5,d9} als relevant.

- a) Recall und Precision sind in der Regel invers. Wie lässt sich der Recall berechnen?
- b) Zeichnen Sie den Recall-Precision-Graphen der Ausgabe des IR-Systems, sowohl als interpolierte Kurve (Wahl eines geeigneten Informatik-Hilfsmittel), als auch als einzelne Stützpunkte.

4. Aufgabe: Suchmaschinen

- a) Recherchieren Sie mit den folgenden Suchmaschinen:
 - Google (<http://www.google.ch/>)
 - Bing (<http://www.bing.com/>)
 - AltaVista (<http://www.altavista.com/>)
 - Ask (<http://ask.com/>)
 - Quintura (<http://www.quintura.com/>)
 - Selbst gewählte Suchmaschine

nach dem Begriff "network computer" (gemeint ist hier ein Rechner, dessen Hardware vergleichsweise billig ist und der die meiste Software über das Netzwerk bezieht). Kennzeichnen Sie den Begriff bei der Suche als Phrase. Erstellen Sie jeweils eine Liste, in der Sie die URLs der ersten fünfzehn Trefferangeben und jedem Treffer eine der folgenden Kategorien zuordnen:

- **Relevant (R):** Web-Seite ist für das Thema relevant
- **Irrelevant (I):** Web-Seite hat nichts mit dem Thema zu tun
- **Down (DN):** Host existiert nicht oder ist nicht erreichbar
- **Not found (N):** URL existiert nicht (mehr)
- **Doublette (D):** Genau diese Web-Seite wurde bereits als Treffer angezeigt

Luzern, August 2015
Seite 3/4
Betreff

Füllen Sie dann die folgende Tabelle aus:

	Google	Bing	AltaVista	Ask	Quintura	Eigene
Relevante URLs in % der Treffer 1–5						
Relevante URLs in % der Treffer 1–10						
Relevante URLs in % der Treffer 1–15						
Anzahl Doubletten						
Anzahl Treffer insgesamt						

- b) Interpretieren Sie die Tabelle und geben Sie eine Empfehlung ab, welche Suchmaschine sich für diese Art von Suche am besten eignet. Bewerten Sie die empfohlene Suchmaschine anhand einiger von Ihnen ausgewählten Kriterien.

5. Aufgabe: Indexierung

- Was ist maschinelle Indexierung?
- Beschreiben Sie den Ablauf einer automatischen Indexierung. Die Stoppworte sollen nach einem vorgegebenen Verfahren eliminiert werden.

6. Aufgabe: Heterogenität und Information Retrieval

- Welche Art von Heterogenität gibt es in heutigen Datenbeständen und Informationssystemen?
- Wie lässt sich die Heterogenität überwinden? Nennen Sie Beispiele.

7. Aufgabe: Metadaten und Information Retrieval

Welche Vor- und Nachteile ergeben sich für das Information Retrieval durch die Verwendung von Metadaten?

Luzern, August 2015
Seite 4/4
Betreff

8. Aufgabe: Suchthema, Suchstrategie

Informationsbedürfnis 1:

Welcher Bundesrat hielt anlässlich der Diamant Gedenkfeiern die Rede am offiziellen Gedenktag vom 1. September 1989 auf dem Rütli? Wie gross war der Nachtragskredit, den das Parlament für die Feier bewilligen musste?

Informationsbedürfnis 2:

Sie brauchen Information zu den folgenden beiden Fragen:

1. Wer hat die Firma „National Versicherung“ gegründet?
2. Womit verdiente Franz Kafka seinen Lebensunterhalt?

Versuchen Sie mit verschiedenen Suchmaschinen/Strategien zu recherchieren und beschreiben Sie Ihre dabei gemachten Erfahrungen.

9. Aufgabe: Mark-Up Sprachen und Information Retrieval

- a) Was sind Mark-Up Sprachen?
- b) Wozu werden diese verwendet?
- c) Welche Vor- und Nachteile bringen sie im Information Retrieval Kontext mit sich?