

INDEX

1 Was fehlt im Query Server:.....	1
3 Boostfaktoren.....	2
4 Sortierung der Ergebnisse (Berechnung des Scores).....	2
Formel für google:.....	2
Formel für seekport:.....	2
Formel für msn.....	2
Formel für Lycos.....	3
Formel für Fireball.....	3
5 Ausgabe der Scores.....	3
6 Sortierung der Treffer.....	3
7 Allgemeine Grundsätze.....	3

1 Was fehlt im Query Server:

- es fehlt noch die URL zum Treffer in der Trefferliste
- msn einbauen (niedrige priorität)

2 Result Clustering

- title und summary aller Treffer vergleichen
- Also wie bei sortieren: Wort 1 und 2 nehmen (außer Stoppwörter) und alle anderen treffer vergleichen, ob diese vorkommen ... dann Wort 1+3 dann Wort 1+4 ...
- Wenn ein Cluster gebildet, dann rechne Gesamtpunkte aus: für jedes Wort im titel 2 pkt. Für jedes Wort im summary 1. pkt.
- Cluster für den user bilden ab Gesamtpunktzahl:6
- Cluster nach Gesamtpunktzahl sortieren

3 Boostfaktoren

Jeder Treffer der in mehreren Suchmaschinen vorkommt soll einen Boostfaktor bekommen. Dieser pusht diesen Treffer bei der Berechnung des Scores. Sprich bei eliminieren der doppelten Treffer ordnest du jedem doppelten Treffer einen Wert(Boostfaktor) zu. Dieser errechnet sich wie folgt:

Treffer 1x vorhanden :	1.0
2x	1.2
3x	1.3
4x	1.4
5x	1.5
6x	1.6

4 Sortierung der Ergebnisse (Berechnung des Scores)

Für die Berechnung des Score ist es wichtig die Bewertung immer in Reihenfolge vorzunehmen: als erstes google dann seekport, msn, lycos, fireball.

Jeder Treffer einer Suchmaschine bekommt Trefferpunkte:

1. Treffer	10 Punkte
2. Treffer	9 Punkte
...	...
10. Treffer	1 Punkt

Es soll jeder Treffer einen Score bekommen. Diese Errechnen sich wie folgt:

Formel für google:

Relevanz=Trefferpkt*Trefferpkt/10 * Boostfaktor

Formel für seekport:

Relevanz=Trefferpkt*Trefferpkt/11 * Boostfaktor

Formel für msn

Relevanz=Trefferpkt*Trefferpkt/12 * Boostfaktor

Formel für Lycos

Relevanz=Trefferpkt*Trefferpkt/15 * Boostfaktor

Formel für Fireball

Relevanz=Trefferpkt*Trefferpkt/20 * Boostfaktor

Die Treffer der Result Cluster stehen aber trotzdem als erstes.

5 Ausgabe der Scores

Es werden keine scores (Relevanz) ausgegeben!!!

6 Sortierung der Treffer

Es wird nach Score Sortiert, auch innerhalb von Result Clustern.

7 Allgemeine Grundsätze

- möglichst für alle Zahlen Variablen in einer Konfigurationsdatei setzen
-