

Aufgabenblatt 8: Unscharfe Mengen, erweiterte Boolesche Verfahren

Aufgabe 1 Unscharfe Mengen

Über dem Intervall $[0, 10]$ sei durch die Eckpunkte $(0.5, 0)$, $(2.5, 1)$, $(5, 1)$, $(7, 0.4)$, $(8, 1)$ und $(9, 0)$ ein Poligonzug als Zugehörigkeitsfunktion einer unscharfen Menge X gegeben.

Die unscharfe Menge Y sei durch die Zugehörigkeitsfunktion $y = 0.1 \cdot x$ gegeben.

- Geben Sie den Träger und die Höhe der Mengen $X \cap Y$ und $X \cup Y$ an.
- Geben Sie die Größe der Menge X an.
- Geben Sie die α -Schnitte der Mengen $X \cap Y$ und $X \cup Y$ für $\alpha = 0.25$, $\alpha = 0.5$, $\alpha = 0.9$ an.

Aufgabe 2 Vektorraummodell und unscharfe Mengen

Beschreiben Sie den Zusammenhang zwischen dem Vektorraummodell des IR und der Theorie der unscharfen Mengen.

- Unter welchen Bedingungen können Dokument- und Anfragevektoren als unscharfe Mengen aufgefasst werden? Wie sind diese unscharfen Mengen in diesem Fall definiert?
- Wie können die Ergebnisse einer Anfrage im Vektorraummodell als unscharfe Mengen dargestellt werden?
- Wie können die Ergebnismengen einer elementaren Anfrage im Booleschen Retrieval mit dem unscharfen Ansatz modelliert werden?
- Wie hängt das MMM Verfahren mit dem Booleschen Retrieval zusammen?

Aufgabe 3 MMM-, Paice- und p-Norm- Methode

Bestimmen Sie zu den Dokumenten aus Aufgabe 4, Blatt 5 die Abstände zu den folgenden Anfragen

ROSA AND LILA AND GRÜN AND SCHWARZ
BLAU OR GRÜN OR WEISS OR BEIGE

- mit der MMM Methode mit Parametern $c = 0.1$ für die AND Anfrage und mit $c = 0.7$ für die OR Anfrage.
- mit der PAICE Methode mit $c = 1.0$ für die AND Anfrage und mit $c = 0.7$ für die OR Anfrage.
- Berechnen Sie die P-Norm-Abstände für $p = 1$ und $p = 2$ zu den folgenden Anfragen.

ROSA:0.6 AND LILA:0.4 AND GRÜN:0.3 AND SCHWARZ:0.1
BLAU:0.3 OR GRÜN:0.6 OR WEISS:0.4 OR BEIGE:0.5

Aufgabe 4 Statistische Unabhängigkeit

- Wie ist die Unabhängigkeit von Ereignissen in einem Wahrscheinlichkeitsraum definiert?
- Wie geht die Unabhängigkeitsannahme bei der Berechnung des Retrieval Status Wertes eines Dokuments im probabilistischen Modell nach Fuhr ein?
- Beschreiben Sie die Annahme inhaltlich!
- Warum wird sie getroffen?