

Neuere Konzepte von Informationssystemen - Teil II

Kapitel 5: Imperfektion in Datenbanken

Neuere Konzepte von Informationssystemen - Teil II

Kapitel 5: Imperfektion in Datenbanken
- Motivation -

3

Szenario: Verkehr

- „Staumelder“ per Telefon an Rundfunkanstalt
- Daten mit verschiedenen Arten der Imperfektion
 - natürlichsprachliche Zustandsbeschreibung (*fuzzy*)
 - frei, lebhaft, stockend, Stau
 - ungenaue Angaben (*imprecise*)
 - Staulänge
 - Stauposition
 - keine Angaben (*null*)
 - Dauer der Störung
 - Staumelder genießen unterschiedliches Vertrauen (*uncertain*)

SS 2004

Heiko Schepperle: Neuere Konzepte von Informationssystemen - Teil II

4

Arten der Imperfektion

- Unsichere Daten (*uncertain*)
- Unscharfe Daten (*vague, fuzzy*)
- Ungenaue Daten (*imprecise*)
- Fehlende Daten (*null-values*)

SS 2004

Heiko Schepperle: Neuere Konzepte von Informationssystemen - Teil II

5

Theorien zur Abbildung von Imperfektion

- Wahrscheinlichkeitstheorie
- Unsichere Daten sind mit Wahrscheinlichkeiten behaftet.
- Possibilitätstheorie
- Möglichkeit eines Ereignisses wird modelliert.
- Fuzzy-Theorie
- Unscharfe Begriffe werden mit Hilfe der Fuzzy-Theorie abgebildet.

SS 2004

Heiko Schepperle: Neuere Konzepte von Informationssystemen - Teil II

Neuere Konzepte von Informationssystemen - Teil II

Kapitel 5: Imperfektion in Datenbanken
- Fuzzy-Theorie -

Fuzzy-Menge #1

- Erweiterung der klassischen Menge
- klassische Menge
 - charakteristische Funktion (Zugehörigkeitsfunktion)

$$\mu_S(x) = \begin{cases} 1 & \text{falls } x \in S \\ 0 & \text{sonst} \end{cases}$$
 - Vereinigung, Schnitt und Komplement

$$S_1 \cup S_2 = \{x \mid x \in S_1 \text{ or } x \in S_2\}$$

$$S_1 \cap S_2 = \{x \mid x \in S_1 \text{ and } x \in S_2\}$$

$$\bar{S} = \{x \mid x \notin S\}$$

Fuzzy-Menge #2

- Fuzzy-Menge
 - charakteristische Funktion

$$\mu_S(x) = \begin{cases} 1 & \text{falls } x \text{ vollständig in } S \\]0, 1[& \text{falls } x \text{ teilweise in } S \\ 0 & \text{falls } x \text{ nicht in } S \end{cases}$$
 - Vereinigung, Schnitt und Komplement

$$\mu_{S_1 \cup S_2} = \max\{\mu_{S_1}(x), \mu_{S_2}(x)\}$$

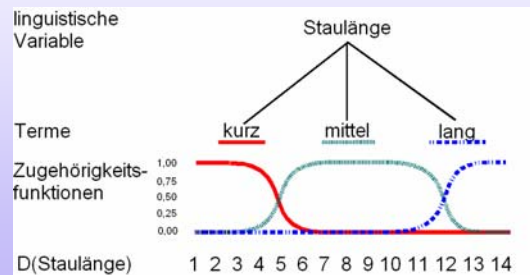
$$\mu_{S_1 \cap S_2} = \min\{\mu_{S_1}(x), \mu_{S_2}(x)\}$$

$$\mu_{\bar{S}} = 1 - \mu_S(x)$$

Fuzzy-Theorie #1

- Linguistische Variablen
 - beschreiben natürlich sprachliche Konzepte
 - besitzen verschiedene Terme
- Terme
 - Fuzzy-Mengen mit Zugehörigkeitsfunktion
- Beispiel
 - Linguistische Variable „Staulänge“
 - Terme „kurz“, „mittel“, „lang“

Fuzzy-Theorie #2



Neuere Konzepte von Informationssystemen - Teil II

Kapitel 5: Imperfektion in Datenbanken
- Relationale Umsetzung -

Relation mit Termen

| Datum | Autobahn | Staulänge | Zugehörigkeit |
|------------|----------|-----------|---------------|
| 27.07.2003 | A81 | kurz | 0,3 |
| 27.07.2003 | A81 | mittel | 0,8 |
| 27.07.2003 | A81 | lang | 0,1 |
| 27.07.2003 | A5 | kurz | 0,1 |
| 27.07.2003 | A5 | mittel | 0,5 |
| 27.07.2003 | A5 | lang | 0,6 |
| 28.07.2003 | A81 | kurz | 0,0 |
| 28.07.2003 | A81 | mittel | 0,3 |
| 28.07.2003 | A81 | lang | 0,9 |

13

Relationale Operatoren

- Selektion
- Projektion
- Vereinigung
- Schnitt
- Kartesisches Produkt

SS 2004

Heiko Schepplinger: Neuausgabe Konzepte von Informationssystemen - Teil II

14

Selektion

- Selektion auf allen Tupeln (Zeilen) einer Relation
- Gruppierung aller daten-identischen Tupel
- gruppierte Tupel erhalten Maximum der Zugehörigkeitswerte

SS 2004

Heiko Schepplinger: Neuausgabe Konzepte von Informationssystemen - Teil II

15

Projektion

- Projektion auf alle Spalten
- Gruppierung aller daten-identischen Tupel
- gruppierte Tupel erhalten Maximum der Zugehörigkeitswerte

SS 2004

Heiko Schepplinger: Neuausgabe Konzepte von Informationssystemen - Teil II

16

Vereinigung und Schnitt

- Vereinigung und Schnitt ausführen
- Gruppierung aller daten-identischen Tupel
- gruppierte Tupel erhalten bei der Vereinigung das Maximum und beim Schnitt das Minimum der Zugehörigkeitswerte

SS 2004

Heiko Schepplinger: Neuausgabe Konzepte von Informationssystemen - Teil II

17

Kartesisches Produkt

- Für jede Kombination von Tupeln werden alle Datenspalten (d. h. ohne Zugehörigkeitswert) aneinander gefügt und eine Zugehörigkeitsspalte angehängt, in der das Minimum der Zugehörigkeitswerte aus den ursprünglichen Tupeln eingetragen wird.

SS 2004

Heiko Schepplinger: Neuausgabe Konzepte von Informationssystemen - Teil II