

1.4 Motivation - Datenbanken

- Warum spezielle Datenbanken für komplex-strukturierte Objekte?
 - Komplex-strukturierte Objekte
 - Irreversibler Trend in der IT Objekte mit räumlicher, zeitlicher und multimedialen Inhalten (STMM-Objekte) zu betrachten
 - Neue Qualität von Informationen
Beschreibung von komplexen Strukturen, dynamischen Vorgängen, etc.
 - i.A. geringerer Informationsverlust wenn Verarbeitung im jeweiligen „Medium“ statt Umsetzung in ein anderes Medium
 - Warum Datenbank-Technologie für komplex-strukturierte Objekte?
 - Sehr große Mengen an Daten vorhanden
 - Speicherplatzintensive Daten
 - In vielen STMM-Anwendungen ist Mehrbenutzerbetrieb erwünscht
 - Daten sollen (effizient und effektiv) recherchierbar sein

1.4 Motivation - Datenbanken

- Was bieten Standard-DBS (z.B. relationale Datenbanken)
 - Konsistenzerhaltender Mehrbenutzerbetrieb
 - Physische und logische Datenunabhängigkeit
 - Effiziente Anfragebearbeitung durch geeignete Speicherungsstrukturen
 - Unterstützung von Transaktionen
 - Concurrency: Isolation gleichzeitiger Updates verschiedener Benutzer
 - Recovery: konsistentes Wiederaufsetzen im Fehlerfall
 - Überwachung der Datenintegrität
 - Datensicherheit, Datenschutz

(Themen aus der Vorlesung zu Informationssystemen)

1.4 Motivation - Datenbanken

- STMM-DBS (STMM = Spatial, Temporal and Multimedia)
 - Verschiedene Verwendungen des Begriffs, z.B. für:
 - CD Sammlungen, die Infos mittels Stichwortsuche zugreifbar machen
 - Systeme zur Organisation und Sichtung von Informationen mittels Browser (z.B. Wiki)
 - Video-on-demand-Systeme
 - CAD-Systeme, die DBS nutzen
 - In dieser Vorlesung ganz allgemein ein DBS
 - mit hoher Kapazität und Performanz
 - das Spatial-, Temporal-, und Multimedia-Datentypen sowie alphanumerische Datentypen unterstützt
 - das effizient mit sehr großen Datenvolumina umgehen kann

1.5 Recherche in STMM-DBS (Überblick)

■ 1.5 Recherche in STMM-DBS (Überblick)

□ Unterschiede zu traditionellen DBS

- In Standard-DBS spezifiziert Benutzer Bedingungen, die Ergebnisse erfüllen müssen (bestimmte Attributswerte); deklarative Anfragen in SQL
- In STMM-DBS sind Anfragen nach bestimmten Attributswerten eher die Ausnahme
- Typisch: Recherche auf Basis von Ähnlichkeit
- Alternative Spezifikationen einer Anfrage:
 - Konkretes Anfrageobjekt, das durch den Benutzer zur Verfügung gestellt wird (z.B. durch URL, Datei, ...)
 - Vereinfachte Approximation eines Anfrageobjektes (Skizze, Summen, ...)
 - Kein Anfrageobjekt; stattdessen: Datenobjekte werden hierarchisch organisiert, Benutzer navigiert durch diese Hierarchie

1.5 Recherche in STMM-DBS (Überblick)

□ Ähnlichkeitsmodelle

- Modellierung/Repräsentation von Datenobjekten
- Datenspezifisch (Bilder, Sequenzen, ...)
- Anwendungsspezifisch („Personen“, „Abendrot“, ...)

□ Ähnlichkeitsmaße

- Messen der Ähnlichkeit zweier Objektrepräsentationen
- Meist werden *Distanzfunktionen* verwendet
- Oft bildet Distanzfunktion D eine *Metrik*

□ $D(x,y) = 0 \Leftrightarrow x = y$

Reflexivität

□ $D(x,y) \geq 0$

Positiv-Definitheit

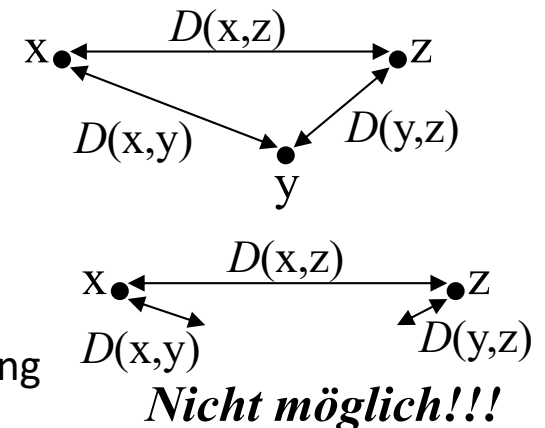
□ $D(x,y) = D(y,x)$

Symmetrie

□ $D(x,y) + D(y,z) \geq D(x,z)$

Dreiecksungleichung

- Metrikeigenschaften **wichtig!!!** für effiziente Ähnlichkeitsanfrage (Beispiel folgt)



1.5 Recherche in STMM-DBS (Überblick)

□ Dreiecksungleichung sehr wichtig für die Effizienz der Suche

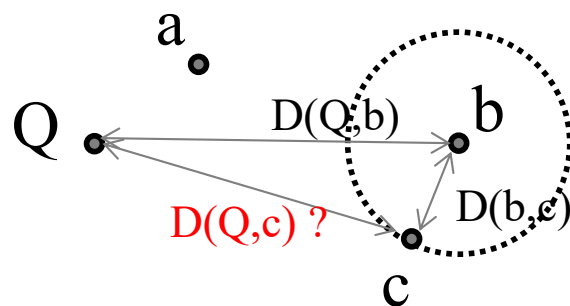
- Gegeben: $DB=\{a,b,c\}$,
 paarweise Distanzen in einer Tabelle
- Gesucht: ähnlichstes Objekt bzgl. Anfrage Q
- Ablaufbeispiel:

	a	b	c
a		6.70	7.07
b			2.30
c			

Zuerst wird a mit Abstand 2 von Q gefunden; a ist aktueller Kandidat.

Dann wird b mit Abstand 7.81 von Q gefunden.

Distanz von Q zu c muss nicht mehr berechnet werden, um c auszuschließen, da



$$D(Q,b) \leq D(Q,c) + D(b,c)$$

$$D(Q,b) - D(b,c) \leq D(Q,c)$$

$$7.81 - 2.30 \leq D(Q,c)$$

$$5.51 \leq D(Q,c)$$