
Kapitel 3

Anfragemethoden für Raum-Zeitliche Daten

Skript zur Vorlesung: Neue Trends zur Suche in modernen Datenbanksystemen
Wintersemester 2013/14, LMU München

© 2012 PD Dr. Matthias Renz, Johannes Niedermayer

3.1 Einführung zur Anfrage auf Raum-Zeitliche Daten

3.1.1 Begriffe und Definitionen

- Raum-Zeitliche Daten: (Spatio-Temporal Data)
 - **Definition: Räumliches Objekt**
Objekt mit Attributen, die sich auf räumliche Informationen (Position, Lage, Geometrie) beziehen.
Position bezieht sich auch auf allgemeine Merkmalsräume.
 - **Definition: Raum-Zeitliches (Spatio-Temporal) Objekt**
Ein räumliches Objekt, dessen räumlicher Zustand (z.B.: Geometrie oder Position) sich mit der Zeit verändert. [EGSV99]
- Raum-Zeitliche Anfragen: (ST Anfragen)
 - Anfragen mit Prädikat über Raum und/oder Zeit die sich auf räumliche und/oder raum-zeitliche Daten beziehen
- Hauptanwendung: Moving Object Data (MOD)
 - Verwaltung von und Anfragen auf bewegliche Objekte (MODB: Moving Object Database)

- Zwei unterschiedliche Blickwinkel auf MODB:
 - Location Management Database: Verwaltung von räumlichen Zuständen (snapshots) über einen Zeitraum hinweg
 - Spatio-Temporal Database: Verwaltung von räumlich-zeitlichen Daten.

- Location Management Database (LMDB)
:= (Snapshot) Datenbank die benutzt wird um
 - (aktuelle) Positionen einer Menge von sich bewegendenden Objekten dynamisch zu verwalten, z.B. eine Menge von Taxis.
 - Anfragen in der Gegenwart oder (sehr) nahen Zukunft zu unterstützen. (Monitoring)
 - Hauptanwendung:
 - *Monitoring*-Anfragen: kontinuierliche Ausgabe des gegenwärtigen Zustandes eines Anfrageergebnisses, das sich auf die Gegenwart bezieht.
 - Szenario: Client-Server Modell
 - Objekte (Clients) senden ihre jeweilige neue Position zum Server
 - Server führt einen Update auf die LM-Datenbank aus.

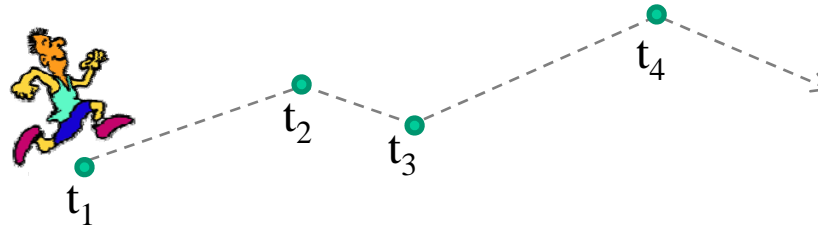
- Spatio-Temporal Database (STDB)
 - Definition
(Räumliche) Datenbank die über einen Zeitraum (häufig eine Historie von) räumliche(n) Zustandsänderungen verwaltet, d.h. räumlich-zeitliche Objekte (ST-Objekte) verwaltet.
 - Hauptanwendung
 - Verwaltung der Änderung des räumlichen Zustandes von (räumlichen) Objekten dessen räumlicher Zustand sich mit der Zeit verändert, z.B. Bewegungen (Trajektorien) von Tieren die über einen Tag beobachtet wurden.
 - Unterstützung von ST-Anfragen die zu einem bestimmten Zeitpunkt ausgeführt werden
:= **Snapshot-ST-Anfragen (Snapshot ST-Queries)**
 - Unterstützung von ST-Anfragen die für eine gewisse Dauer gestellt werden
:= **Kontinuierliche ST-Anfragen (Continuous ST-Queries)**

- Typen von räumlich-zeitlichen Objekten (ST-Objekten):
 - Welche Arten von Objekten werden in einer STDB verwaltet?
 - **(Statisches) Objekt:** Position eines Objekts als Punktobjekt über seine Koordinaten im Vektorraum oder Netzwerkgraph dargestellt.
 - **Bewegliches Objekt:** Bewegung eines Objektes als Trajektorie über eine Linie oder Linienzug (Polylinie) im Vektorraum oder Netzwerkgraph dargestellt.
 - **(Statische) Region:** Objekt mit räumlicher Ausdehnung über Polygon im Vektorraum oder Teilgraph im Netzwerkgraph dargestellt.
 - **Dynamische Region:** (Räumliche) Veränderung eines (räumlich ausgedehnten) Objektes (Region) als dynamisches Polygon / dynamischer Teilgraph dargestellt.
 - Welche Arten von räumlichen Zustandsänderungen unterscheidet man?
 - **Diskrete** Veränderung des räumlichen Zustands:
Beschreibung von Handlungen einer Person die an unterschiedlichen Orten statt finden, wie z.B. Geld abheben, einkaufen, Essen, etc.
 - **Kontinuierliche** Veränderung des räumlichen Zustands:
Beschreibung der Bewegung einer Person, wie z.B. die Autofahrt von der Arbeit nach Hause.

- Wichtigste Abstraktionen von ST-Objekten

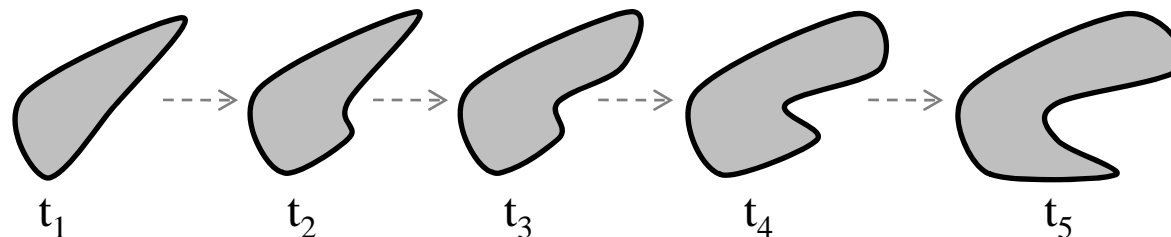
- **Verändernde Position** (z.B. Auto)

Abstrahierung eines „real“ existierenden Objektes (können auch Merkmale eines realen Vorgangs sein) das sich in einem mehr-dimensionalen Raum bewegt und lediglich dessen **Position** im Raum für Anfragen relevant ist.



- **Verändernde Geometrie** (z.B. Fläche eines Waldgebietes)

Abstraktion eines räumlich ausgedehnten Objektes in einem mehr-dimensionalen Raum dessen **Position**, **Ausdehnung** und **Form** zeitlich variiert.



3.1.2 Anwendungsgebiete

– Anfragen von ST-Objekten in modernen Informationssysteme ^[S99]

- Geographic Information Systems (GIS)
 - “Welche Tankstellen befinden sich in der Nähe meines Autos?”
 - “Welches Taxi befindet sich am nächsten am Kunden?”
- Environmental Information Systems
 - “Wie haben sich die polaren Eiskappen in den letzten Jahren verändert?”
 - “In welchem Gebiet hält sich der Luchs normalerweise auf?”
- Multimedia Information Systems
 - Welche Schauspieler treten in mehr als einer Szene gemeinsam auf?
 - Wie bewegt sich Schauspieler X zur Zeit t in Video V?

– Weitere Anwendungen für ST-Objekte [EGSV99]

Anwendungen für bewegliche Punkte	Anwendungen für veränderliche Flächen
<p>Personen:</p> <ul style="list-style-type: none"> • Bewegung von Terroristen/Kriminellen <p>Tiere:</p> <ul style="list-style-type: none"> • Trajektorien von Walen/Luchsen ermitteln • Welche Wege legen sie zurück? Wie schnell? Wie oft pausieren sie? • Wo befinden sich die Wale gerade? • Hat sich ihr Aufenthaltsbereich in den letzten 20 Jahren verändert? <p>Satelliten, Raumschiffe, Planeten:</p> <ul style="list-style-type: none"> • Welche Satelliten werden sich dem Raumschiff in den nächsten 4 Stunden nähern? <p>Autos:</p> <ul style="list-style-type: none"> • Taxis: Welches ist am nächsten zum Kunden? • Lastwagen: welche Routen werden regulär befahren? <p>Flugzeuge:</p> <ul style="list-style-type: none"> • Gab es beinahe-Kollisionen zwischen Flugzeugen? • Fliegen zwei Flugzeuge gerade aufeinander zu? • Sind Flugzeuge über ein bestimmtes Staatsgebiet geflogen? • Wie schnell fliegt ein Flugzeug? Maximalgeschwindigkeit? <p>Schiffe:</p> <ul style="list-style-type: none"> • Drohen Schiffe auf Grund zu laufen? • Sonderbare Routen von Schiffen entdecken (Illegale Machenschaften) • Militärische Analysen 	<p>Länder</p> <ul style="list-style-type: none"> • Was war die maximale Fläche des römischen Reiches? • Wann wurden Staaten vereinigt? • Welche Staaten wurden geteilt? <p>Wälder, Seen</p> <ul style="list-style-type: none"> • Wie schnell schrumpft der Regenwald am Amazonas? • Schrumpft das tote Meer? • Minimale und maximale Ausdehnung von Flüssen über das Jahr? <p>Gletscher</p> <ul style="list-style-type: none"> • Schrumpft der polare Eisschild? Verändert er seine Position? • Rückprojektion: Wie groß war Gletscher X zur Zeit t? <p>Hoch- und Tiefdruckgebiete</p> <ul style="list-style-type: none"> • Wohin bewegen sie sich? • Veränderung von Äquidensiten über die Zeit? <p>Armeen, Truppen:</p> <ul style="list-style-type: none"> • Trajektorie von Hannibals Truppen? <p>Krebs:</p> <ul style="list-style-type: none"> • Wie hat sich der Tumor über die letzten Jahre ausgebreitet? <p>Kontinente:</p> <ul style="list-style-type: none"> • Wie haben sich die Kontinente im letzten Jahr verschoben?

3.1.3 Wissenschaftliche Herausforderungen im Kontext von STDBMS [S99]

1. Ontologie, Struktur und Darstellung von Raum und Zeit
 - z.B. Unterscheidung zwischen Lebensdauer und Bewegung von Objekten
 2. Modelle und Sprachen für STDBMS
 - z.B. feste Integration des räumlich-zeitlichen Aspekts in das STDBMS
 3. GUIs für Spatio-Temporal Information
 4. Architektur eines STDMBS
 5. Speicherstrukturen und Indexierung
 - Erweiterung von Speicherstrukturen auf bewegliche Objekte, Benchmarks
 6. Anfragebearbeitung in Spatio-Temporal Databases
 - Anfragealgorithmen, Anfrageoptimierung
- In dieser Vorlesung beschäftigen wir uns primär mit den Problematiken Anfragebearbeitung und Indexierung (siehe 5. u. 6. oben)

3.1.4 Überblick über Anfragen auf ST-Objekte

– Wichtige Attribute von Anfragen auf ST-Objekte [MDM 07]

- Anfragetyp:
 - Spezifiziert das (räumliche) Anfrageprädikat
 - Z.B.: Bereichsanfrage, NN-Anfrage, RNN-Anfrage, etc.
- Anfragezeit(raum):
 - Spezifiziert den Zeitpunkt auf den sich die Anfrage bezieht
 - z.B.: Vergangenheit, Gegenwart, Zukunft
- Dauer des Anfrageprozesses:
 - Spezifiziert die Periodizität der Anfrage
 - Einmalige Ausführung der Anfrage („Snapshot“) ← → Kontinuierliche Berechnung des „aktuellen“ Anfrageergebnisses („Continuous“)
- Dynamik des Anfrageobjektes:
 - Gibt an ob es sich bei dem Anfrageobjekt um ein dynamisches („bewegliches“) oder statisches („stationäres“) Objekt handelt.
- Dynamik der Datenbankobjekte:
 - Gibt an ob es sich bei den Datenbankobjekten um dynamische („bewegliche“) oder statische („stationäre“) Objekte handelt.

– Beispiel: Varianten von ST-Anfragen

- **Beobachtung (kontinuierlich) der Verkehrsdichte:**



Anfrage-Typ: Bereichsanfrage

Anfragezeit: Gegenwart

Dauer des Anfrageprozesses: kontinuierlich

Dynamik (Anfrageobjekt): stationär

Dynamik (DB): beweglich

- **Was sind meine nächsten Restaurants für die nächste Stunde?**



Anfrage-Typ: NN-Anfrage

Anfragezeit: Zukunft

Dauer des Anfrageprozesses: kontinuierlich

Dynamik (Anfrageobjekt): beweglich

Dynamik (DB): stationär

- **Versende E-coupons zu allen Autos denen ich am nächsten bin.**



Anfrage-Typ: RNN-Anfrage

Anfragezeit: Gegenwart

Dauer des Anfrageprozesses: snapshot

Dynamik (Anfrageobjekt): stationär

Dynamik (DB): beweglich

- **Was war gestern mein kürzester Abstand zum Taxi A?**



Anfrage-Typ: Min.-Dist.-Anfrage

Anfragezeit: Vergangenheit

Dauer des Anfrageprozesses: snapshot

Dynamik (Anfrageobjekt): beweglich

Dynamik (DB): beweglich

– Typisierung von ST-Anfragen nach Dynamik der Objekte:

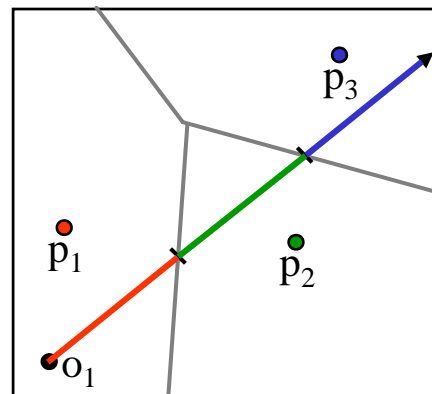
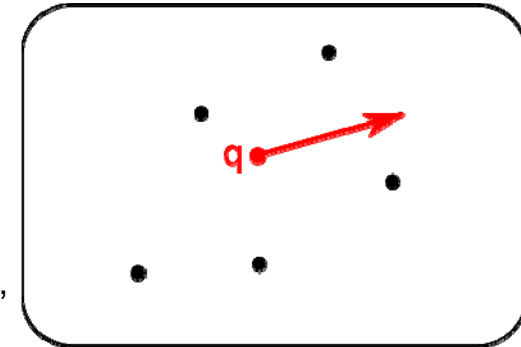
- Nur Anfrageobjekt beweglich:

- Beispielanfrage:

- » „Alle Tankstellen, die sich in den nächsten 5 Minuten in der Nähe meines Autos befinden werden..“

- Eigenschaften:

- » Verwendung von herkömmlichen Methoden zur Verwaltung von statischen Objekten
 - » Anfragen oft durch Vorberechnung von möglichen Anfrageergebnissen (Ergebnisräume) effizient unterstützt. (Z.B.: Voronoizellen für kontinuierliche NN-Anfragen)

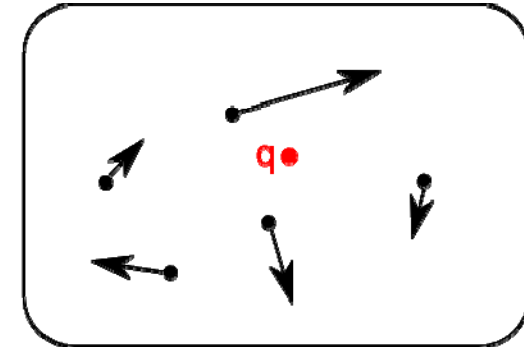


- Indexstrukturen:

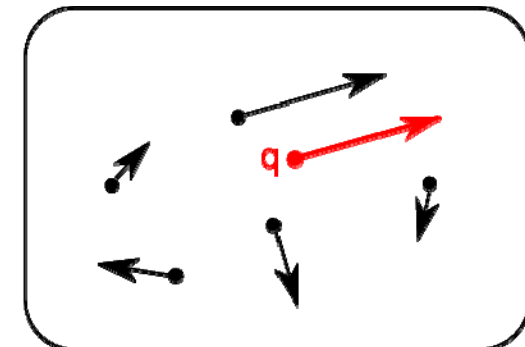
- » R-Baum, Quadtree, etc.

– Typisierung nach Dynamik der Objekte (Cont.):

- Nur DB-Objekte beweglich:
 - Beispielanfrage:
 - » „Alle Autos, die näher an meiner Tankstelle sind als an einer anderen Tankstelle.“
 - Eigenschaften:
 - » Hoch-dynamische Daten, Spezielle Methoden zur Verwaltung der Objekte
Notwendig
 - Indexstrukturen:
 - » Beispiele: STR-Tree, TB-Tree, 3D R-Tree [MGA03]



- DB- und Anfrage-Objekte beweglich:
 - Beispielanfrage:
 - » „Alle Flugobjekte die sich in einem Umkreis von 10km zu meinem Flugzeug befinden.“
 - Eigenschaften:
 - » Hoch-dynamische Daten, Spezielle Methoden zur Verwaltung der Objekte
Notwendig (kann auf den obigen Fall zurückgeführt werden)
 - Indexstrukturen: (s.o.)

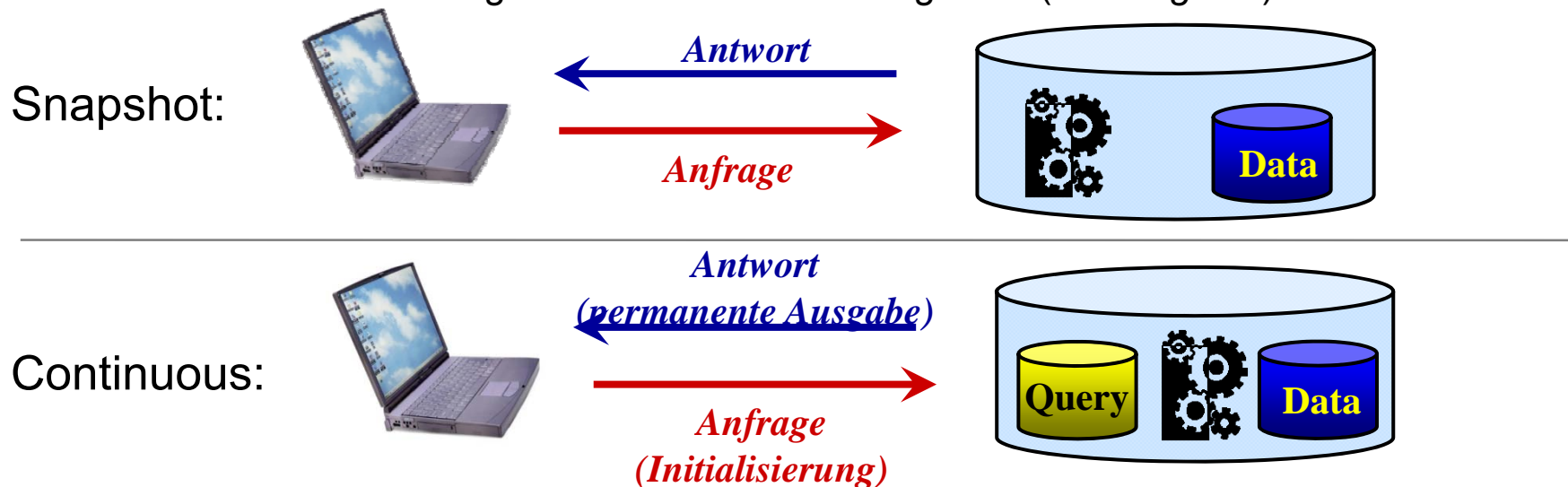


- Typisierung von ST-Anfragen nach Anfragezeit(raum):
 - Anfrage an die Vergangenheit (Past):
 - Beispielanfragen:
 - » „*Welches Auto war zwischen t_1 und t_2 im Gebiet A?*“
 - » „*Welches Auto war zur Zeit t_1 nächster Nachbar von X?*“
 - Eigenschaften:
 - » Historische Daten sind grundsätzlich permanente DB-Einträge (Trajektorien)
 - » Updates nur im Sinne von Löschen (stark) veralteter Daten und Einfügung neuer (aktueller) Daten
 - » Ggf. Approximation der Bewegung, z.B. linear oder polynomiell
 - Indexstrukturen:
 - » Beispiele: STR-Tree, TB-Tree, 3D R-Tree [MGA03]
 - Anfrage an die Gegenwart (Present):
 - Beispielanfragen:
 - » „*Welches Auto ist aktuell im Einflussbereich meiner Tankstelle?*“
 - » „*Welche Autos sind aktuell im Gebiet A?*“
 - Eigenschaften:
 - » Stark dynamische Daten
 - Indexstrukturen:
 - » Grid Indexing [YPK05], Hashing, LUR-Tree [MGA03]

- Typisierung nach Anfragezeit (Cont.)
 - Anfrage an die Zukunft (Future):
 - Beispielanfragen:
 - » „Welche Autos sind in 5 Minuten meiner Tankstelle am nächsten?“
 - Eigenschaften:
 - » Prädiktionsmodell notwendig, z. B. linear oder polynomiell
 - Indexstrukturen:
 - » TPR-Tree, TPR*-Tree, Future Quadtree [MGA03]

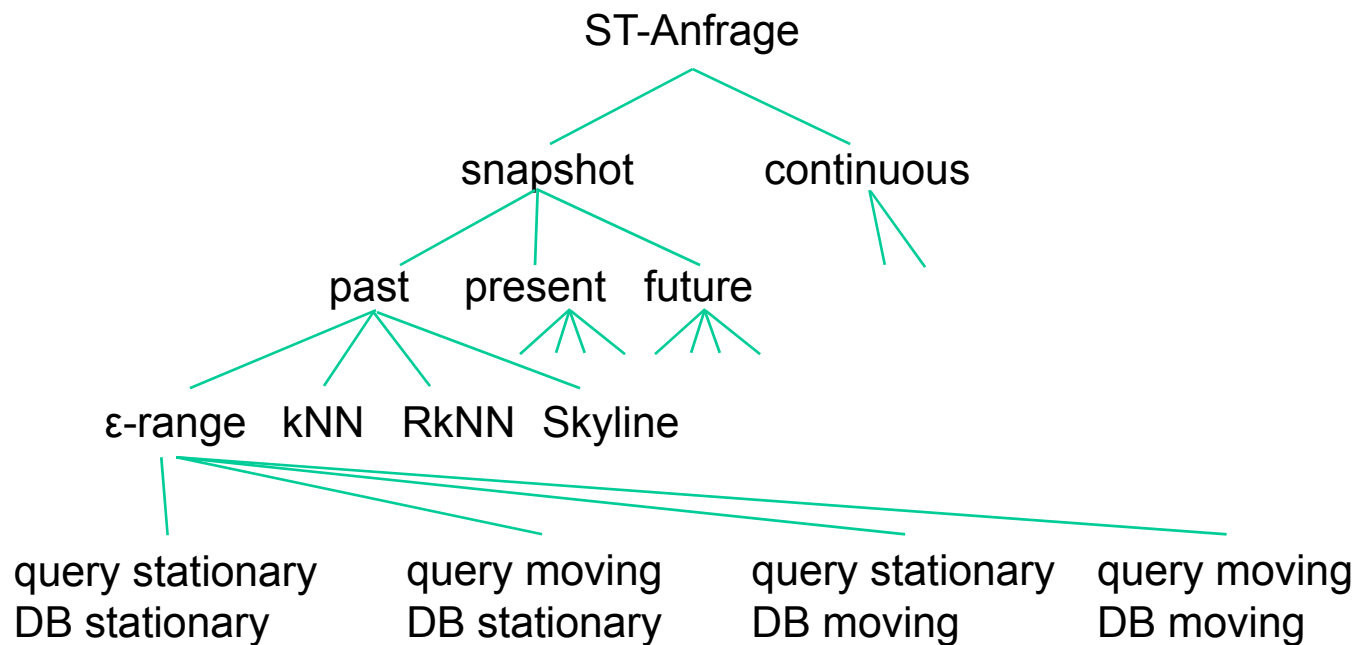
- Typisierung von ST-Anfragen nach Dauer des Anfrageprozesses
 - Snapshot Anfragen:
 - Beispielanfragen:
 - » „Welche Autos waren vor 2 Stunden in der Nähe des Eiffelturms?“
 - Eigenschaften:
 - » Anfrage auf große Menge an ST-Objekten, und persistente Datenbasis
 - Indexstrukturen:
 - » Indexstrukturen für persistente ST-Objekte (Punkte, Trajektorien)

- Typisierung nach Dauer des Anfrageprozesses (Cont.)
 - Kontinuierliche Anfragen (Continuous):
 - Beispielanfragen:
 - » „Permanente Ausgabe der nächsten Raststätte“
 - Eigenschaften:
 - » Initiale Anfrage (für Startzeitpunkt der Anfrage)
 - » Ermittlung WANN und WIE sich das Ergebnis ändert (Update orientierte Anfrage)
 - » Häufig für Client-Server Architekturen ausgelegt
 - Indexstrukturen eher weniger für die Objekte selbst, sondern oft nur zur Verwaltung von vorberechneten Regionen (*safe regions*)



– Zusammenfassung:

- Grundsätzlich viele unterschiedliche Methoden, ST-Objekte effizient zu verwalten und anzufragen
- Wahl der Methode abhängig von der jeweiligen Anwendung (Anfragetyp)
- ST-Anfrage Klassifikation:



3.2 Methoden für Snapshot ST-Anfragen

3.2.1 Allgemeines

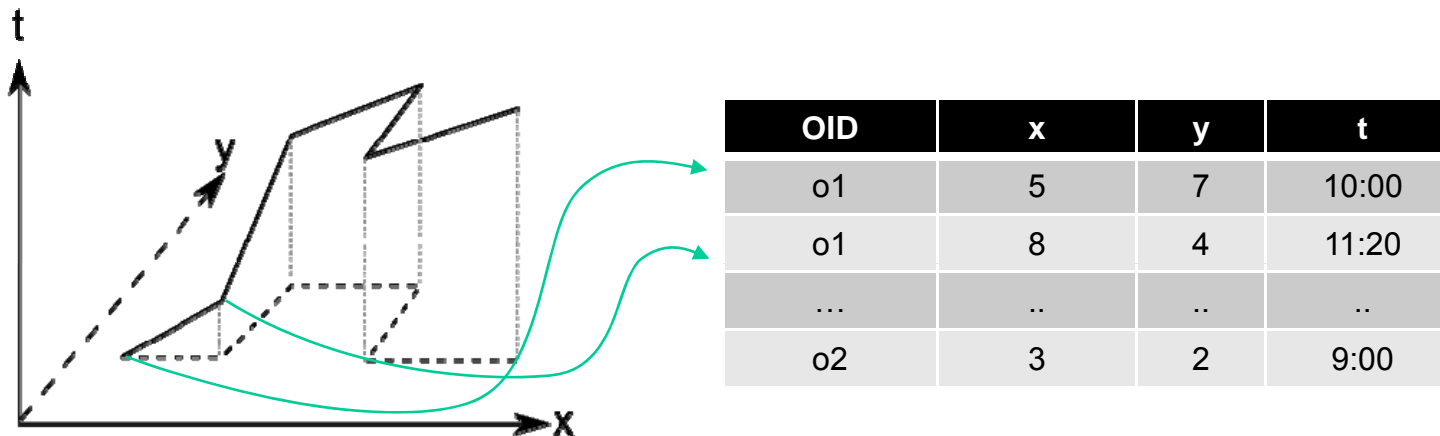
- Snapshot Anfragen beziehen sich auf einen Zeitpunkt bzw. auf einen Zeitraum.
- Grundsätzliche Unterscheidung zwischen
 - Anfragen an die Vergangenheit (Snapshot *Past* Queries)
 - Anfragen an die Gegenwart (Snapshot *Present* Queries)
 - Anfragen an die (nahe) Zukunft (Snapshot *Future* Queries)
- Anfragen werden mittels speziell für ST-Objekte entwickelte Indexstrukturen (sog. ST-Indexe) effizient unterstützt.

3.2.2 Anfragen an die Vergangenheit

– Darstellung von Historischen ST-Daten

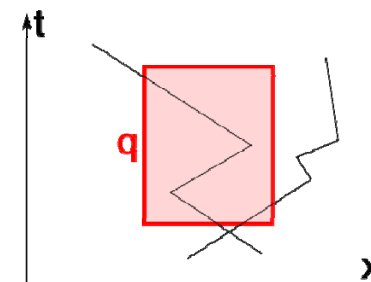
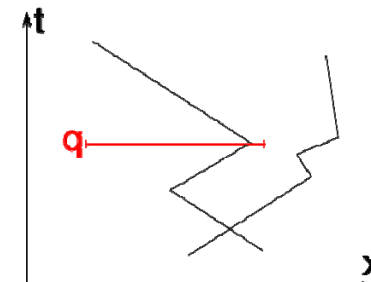
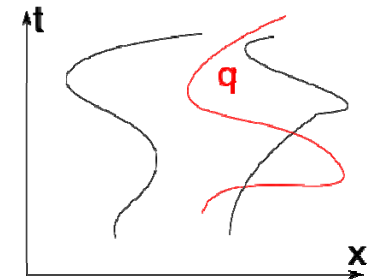
- Abbildung der (vergangenen) Bewegung eines Objektes über Trajektorien (Sequenz von Positionen (Beobachtungen); Verlauf zwischen Beobachtungen oft linear interpoliert)
- Zeit-Attribut wird oft als zusätzliche Dimension betrachtet:
 - Objekte die sich im d -dimensionalen Raum bewegen werden in einen (statischen) $(d+1)$ -dimensionalen Raum transformiert.

=> Trajektorien als Sequenz von $(d+1)$ -dimensionalen Punkten



– Eigenschaften:

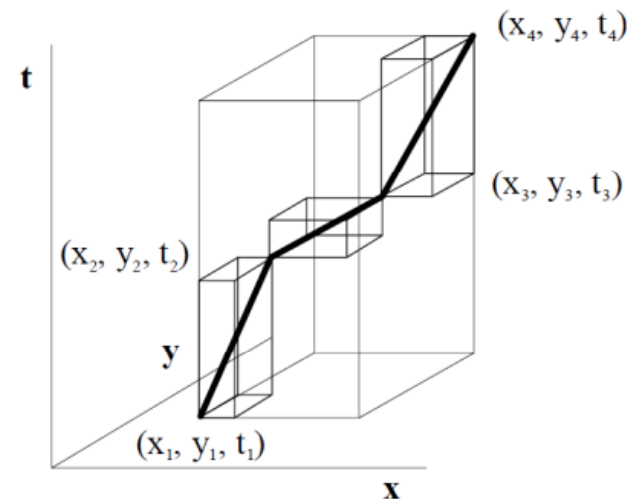
- Verwaltung von sehr großen Menge an Trajektorien
- Daten werden persistent in der Datenbank verwaltet
- Möglichkeit Anfragen auf Raum sowie Zeit zu beschränken:
 - Anfragen entlang der Zeitdimension
 - » Beispiel: „An welchen Ort hat sich Objekt A gestern zwischen 7 Uhr und 10 Uhr vormittags aufgehalten?“
 - » Selektion des Objektes (bzgl. der räumlichen Dimension(en) keine echte räumliche Anfrage)
 - » Auslesen der Teil-Trajektorie im Zeitintervall 7-10 Uhr
 - Anfragen entlang der räumlichen Dimension(en)
 - » Beispiel: „Welche Objekte befanden sich gestern um 8 Uhr in einem bestimmten Bereich?“
 - » Ermittlung der Momentaufnahme um 8 Uhr (Schnitt zwischen Trajektorien und 8 Uhr Zeitscheibe)
 - » Ausführung einer räumlichen Anfrage auf die Momentaufnahme
 - Anfragen entlang zeitlicher und räumlicher Dimensionen (klassische ST-Anfragen)
 - » Beispiel: „Welche Objekte befanden sich gestern zwischen 8 Uhr und 10 Uhr in einem bestimmten Bereich?“



– Verwaltung der Vergangenheit

- Idee:

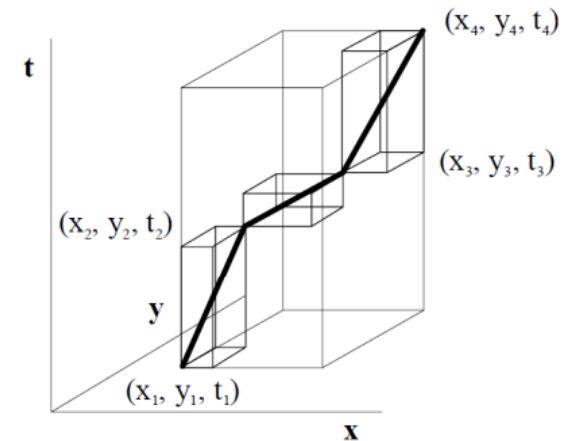
- Erweiterung des R-Baums um eine weitere Dimension, die Zeit t
- Die Trajektorie eines Objektes kann als Sequenz von $(d+1)$ -dimensionalen Liniensegmenten repräsentiert werden
- Die einzelnen Liniensegmente werden in einem R-Baum abgelegt



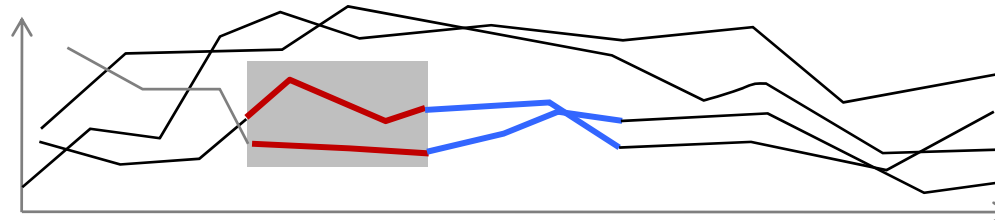
- Beispiele: 3D-R-Tree [MGA03], STR-Tree und TB-tree [PJT00]
- ST-Index-Varianten unterscheiden sich gemäß der Gruppierung der Objekte auf die Index-Seiten (unterschiedliche Splitstrategien bei Überlauf von Seiten)

– Der 3D-R-Tree: [TVS96]

- Prinzip:
 - Verwaltung der $(d+1)$ -dimensionalen Liniensegmente in einem herkömmlichen R-Baum (R^* -Baum).
- Vorteile:
 - Verwendung einer bereits existierenden Indexstruktur
 - Unterstützt alle Snapshot-Anfragen (d.h. räumliche Anfragen zu einem Zeitpunkt oder über einen Zeitbereich)
 - Einfach zu implementieren
- Nachteile:
 - Segmente der Trajektorie eines Objektes werden als „unabhängige“ Objekte/Einträge betrachtet
 - => Eine Trajektorie wird (willkürlich) über mehrere Seiten verstreut
 - => Problem für ST-Anfragen über einen größeren Zeitbereich
 - Seitenregion wachsen zu schnell in der Zeitachse → zu wenig Information für einen Zeitpunkt
 - => Problem bei Anfragen die sich auf einen Zeitpunkt beziehen



- Weiterer Nachteil:
 - Oft werden zusammenhängende Trajektorien-Segmente zusammen angefragt, wie z.B. bei **Trajektorien-basierten ST-Anfragen**.
 - Bei Trajektorien-basierten Anfragen unterscheidet man zwischen:
 - » Teiltrajektorie die für die Erfüllung des Anfrageprädikates relevant ist.
 - » Teiltrajektorie die als Anfrageergebnis ausgegeben wird.
 - Beispiel:
 - » Wo haben sich die Personen die sich **zwischen 8 und 9 Uhr auf dem Kirchplatz** aufgehalten haben **innerhalb der folgenden Stunde** hinbewegt?



- Anfrage wird in 2 Schritten durchgeführt:
 - » 1) **Selektion der Trajektorien** über ST-Anfrage: „...zwischen 8 und 9 Uhr auf dem Kirchplatz ...“
 - » 2) **Ermittlung der Ergebnis-Trajektorien** die auszugeben sind: „...innerhalb der folgenden Stunde hinbewegt.“
- Zugriff auf Zusammenhangskomponenten werden im 3D-R-Tree nicht effizient unterstützt.