# Flexible und konfigurierbare Softwarearchitekturen für ontologiebasierte Informationssysteme

**Promotionsvortrag** 

Michael Wessel

Vorsitzender: Prof. ter Haseborg

Gutachter: Prof. Möller, Prof. Mayer-Lindenberg

Institut für Softwaresysteme, TUHH

## Inhaltsangabe

- Generelle Motivation & Hintergrund
- Anwendungsspezifische Fragestellungen → Analyse → Anforderungen an Architektur-Modell
- Entwurf und Vorstellung des Architektur-Modelles
- Evaluation: Prototypen (Anwendungsdomäne)
- Zusammenfassung & Ausblick

## **Motivation & Hintergrund**

- These: Bedeutung ontologiebasierter Informationssysteme nimmt zu (z.B. "Semantic Web")
- Informationssysteme mit Schlussfolgerungsmechanismen
- Kernthema: ontologiebasierte Anfragebeantwortung

## **Motivation & Hintergrund**

- These: Bedeutung ontologiebasierter Informationssysteme nimmt zu (z.B. "Semantic Web")
- Informationssysteme mit Schlussfolgerungsmechanismen
- Kernthema: ontologiebasierte Anfragebeantwortung
- Verdeutlichendes Beispiel (informal)
  - Begriffsdefinitionen (intensionale Komponente):
     Mutter gdw. Frau mit Kind
     Frau gdw. weiblich & Person
  - Fakten (extensionale Komponente):
     Betty ist weiblich Betty ist Person
     Betty hat Kind charles
  - **Anfrage** "Instanzen von Mutter": {Betty}.

- These: Bedeutung ontologiebasierter Informationssysteme nimmt zu (z.B. "Semantic Web")
- Informationssysteme mit Schlussfolgerungsmechanismen
- Kernthema: ontologiebasierte Anfragebeantwortung
- Verdeutlichendes Beispiel (Logik)
  - Begriffsdefinitionen (intensionale Komponente):

$$\forall x.Mutter(x) \leftrightarrow Frau(x) \land \exists y.hat\_Kind(x,y)$$
$$\forall x.Frau(x) \leftrightarrow Weiblich(x) \land Person(x)$$

• Fakten (extensionale Komponente):

• Anfrage  $\{i \mid \Sigma \models mutter(i)\} = \{betty\}$ 

# **Motivation & Hintergrund**

- These: Bedeutung ontologiebasierter Informationssysteme nimmt zu (z.B. "Semantic Web")
- Informationssysteme mit Schlussfolgerungsmechanismen
- Kernthema: ontologiebasierte Anfragebeantwortung
- Verdeutlichendes Beispiel (Beschreibungslogik)
  - Begriffsdefinitionen (TBox):

```
Mutter \stackrel{.}{\equiv} Frau \sqcap \exists hat\_kind. \top
Frau \stackrel{.}{\equiv} Weiblich \sqcap Person
```

• Fakten (ABox):

betty: Weiblich betty: Person (betty, charles): hat\_kind

• Anfrage  $\{i \mid KB \models mutter(i)\} = \{betty\}$ 

### **Motivation & Hintergrund**

- These: Bedeutung ontologiebasierter Informationssysteme nimmt zu (z.B. "Semantic Web")
- Informationssysteme mit Schlussfolgerungsmechanismen
- Kernthema: ontologiebasierte Anfragebeantwortung
- Motivation: Bereitstellung eines flexiblen Rahmens zur Realisierung ontologiebasierter Informationssysteme . . .
  - ... ausdrucksstarke und skalierbare ontologiebasierte Anfragesprachen bietet (Variablen)
  - ... und verschiedene Repräsentationsarten unterstützt (gleich)
  - ...der auch für **räumliche Informationssysteme** (Anwendungsdomäne 1) genutzt werden kann

- Anwendung: ontologieb. Anfragebeantwortung für GIS
- Räumlich-thematische Anfrage:

"In welchen europäischen Ländern mit mehr als 60 Mio. Einwohnern ist Hamburg nicht enthalten?"

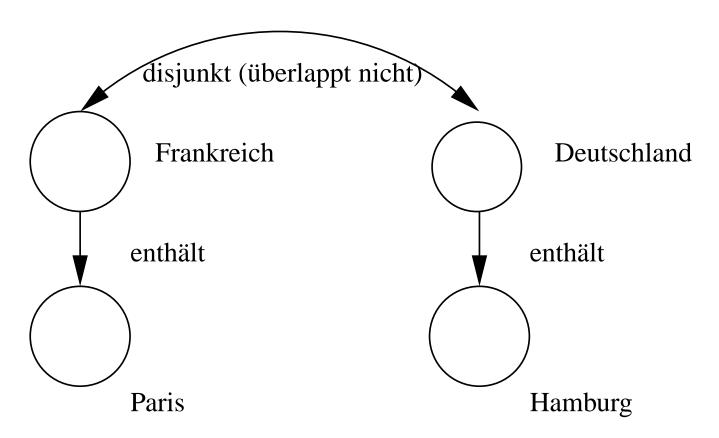
- Anwendung: ontologieb. Anfragebeantwortung für GIS
- Räumlich-thematische Anfrage:

"In welchen europäischen Ländern mit mehr als 60 Mio.

Einwohnern ist Hamburg nicht enthalten?"



- Anwendung: ontologieb. Anfragebeantwortung für GIS
- Räumlich-thematische Anfrage:
   "In welchen europäischen Ländern mit mehr als 60 Mio.
   Einwohnern ist Hamburg nicht enthalten?"
- **F1:** Wie und für welche Aspekte können Ontologien und speziell Beschreibungslogiken genutzt werden? Z.B. zur Def. räumlich-thematischer Begriffe?
- **F2:** Wie kann (u.U. unterbestimmte) räumliche Information logik-geeignet repräsentiert und für räumliche Schlussfolgerungen genutzt werden?



Qualitative Karte mit qualitativen räumlichen Relationen In welcher Beziehung stehen Paris und Hamburg?

- Anwendung: ontologieb. Anfragebeantwortung für GIS
- Räumlich-thematische Anfrage:
   "In welchen europäischen Ländern mit mehr als 60 Mio.
   Einwohnern ist Hamburg nicht enthalten?"
- **F1:** Wie und für welche Aspekte können Ontologien und speziell Beschreibungslogiken genutzt werden? Z.B. zur Def. räumlich-thematischer Begriffe?
- **F2:** Wie kann (u.U. unterbestimmte) räumliche Information logik-geeignet repräsentiert und für räumliche Schlussfolgerungen genutzt werden?
- **F3:** Wie können entsp. Anfragesprachen aussehen?
- **F4:** Wie können Anfragen auf Konsistenz geprüft werden?

Beispiele für inkonsistente Anfragen

 $enth\ddot{a}lt(x,y) \wedge enth\ddot{a}lt(y,z) \wedge disjunkt(x,z)$ räumlich inkonsistent

> $stadt(x) \wedge land(x)$ thematisch inkonsistent

 $naturschutz\_park(x) \land enthält(x, y) \land wohngebiet(y) \land$   $(\forall enthält.(gruen\_flaeche \sqcup wasser\_flaeche))(x)$  räumlich-thematisch inkonsistent

Sowohl Anfragen als auch Beschreibungen können inkonsistent werden

### Antworten auf Fragestellungen

- **F1**: Raumaspekte in Beschreibungslogiken: Beschreibungslogik-Workshops '00-'03
- F2, F3, F4: Räumlich-thematische Anfragesprachen, Potentiale und Defizite von Standard-Beschreibungslogiksystemen: Knowledge Representation meets Databases '03
- **F3:** Erweiterte Anfragemöglichkeiten f. thematische Aspekte = ABox-Anfragesprache: u.a. Beschreibungslogik-Workshop '04
- **F3:** Skalierbarkeit der ABox-Anfragesprache: u.a. Deutsche Jahrestagung für Künstliche Intelligenz, '06

# Antworten auf Fragestellungen (2)

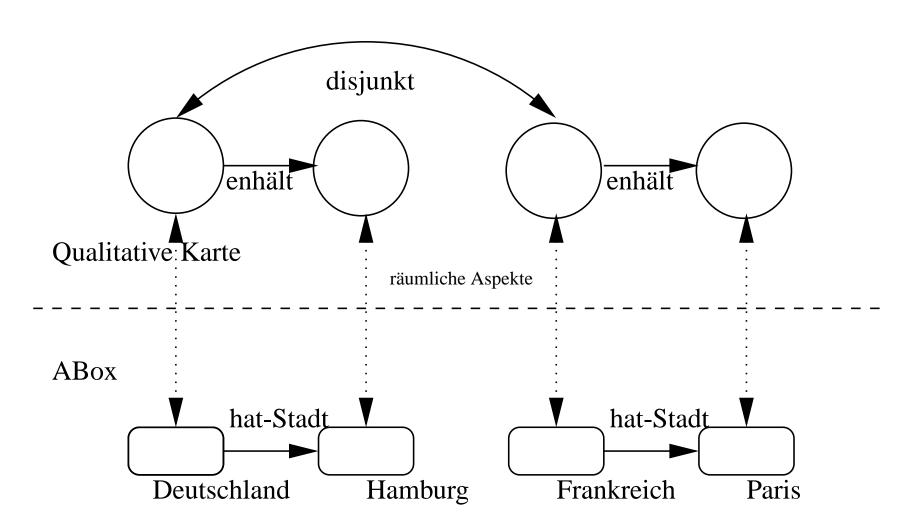
- **F1**: Repräsentation datenwertiger Attribute in Beschreibungslogiken: International Joint Conference on Automated Reasoning '01
- **F2:** Räumlich/thematische Default-Inferenz: International Conference on Spatial Information Theory '99
- . . .
- ⇒ Übersicht über die Architektur (eigentlich
   Architektur-Modell): Journal of Applied Logic Special
   Issue on Emperically Sucessful Systems '07
- → Analysefazit, das zur Anforderungsdefinition an das Architektur-Modell führte

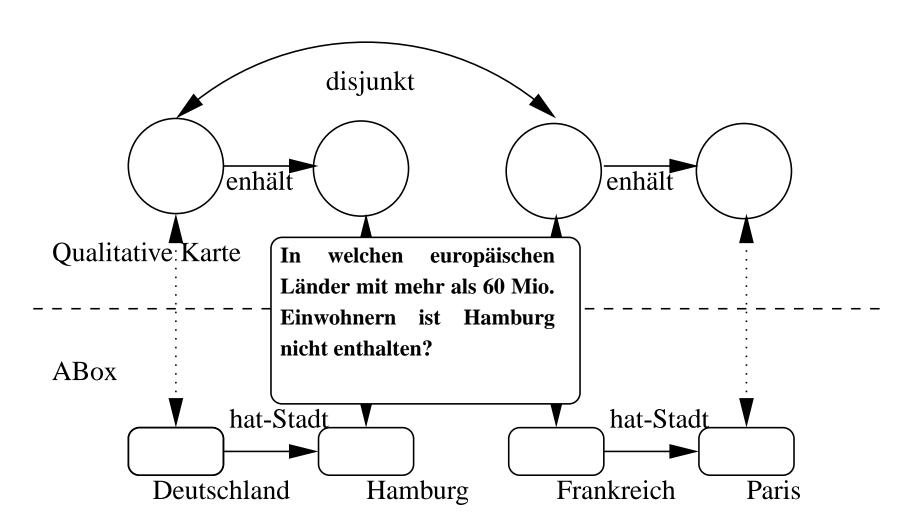
### Analysefazit des Autors

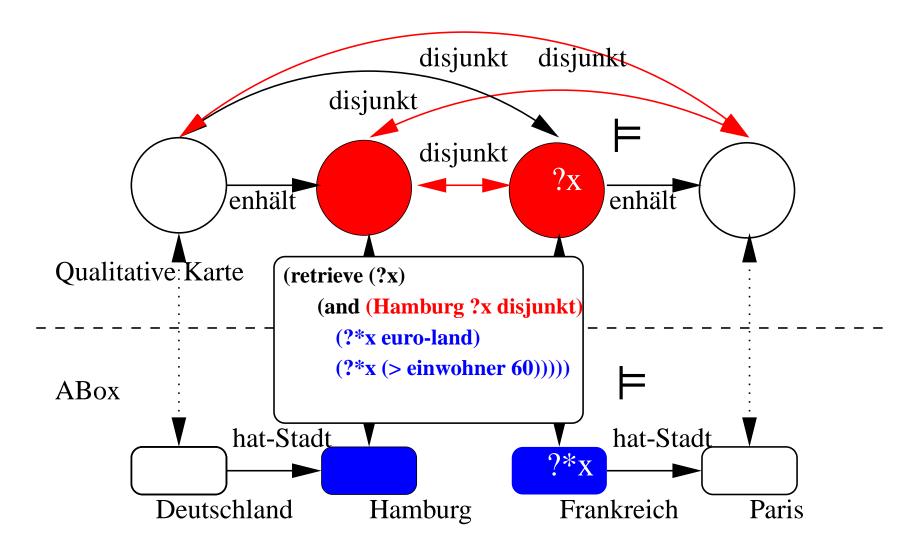
- die **integrierte** Behandlung räumlicher und thematischer Aspekte in einer Beschreibungslogik führt
  - entweder zu in der Praxis nicht nutzbaren, weil stark eingeschränkten Beschreibungssprache,
  - oder zur Unentscheidbarkeit der Logik.
- ⇒ pragmatische Trennung räumlicher & thematischer Aspekte erforderlich, spezialisierte Repräsentationen erforderlich, auch wg. Skalierbarkeit
- ⇒ Re-Kombination der Aspekte durch eine hybride räumlich-thematische Anfragesprache
  - Inferenz (Konsistenzprüfung) für bestimmte Klassen von Anfragen möglich → Mehrwert

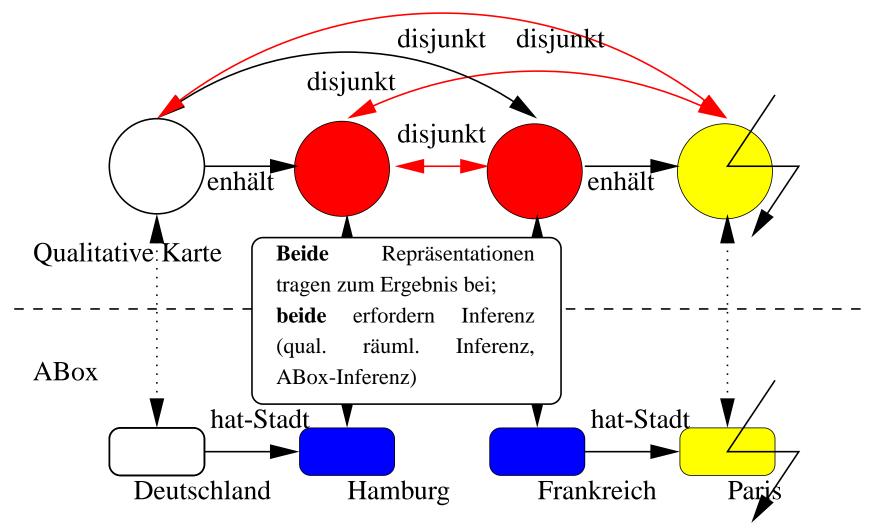
### Gewonnene Anforderungen

- verschiedene spezialisierte Repräsentationen
  - räuml. Repräsentationen: qual. und geometr.
  - thematische Information: **u.a.** ABoxen
  - ⇒ anwendungsspezifisch spezialisierbares generisches (hybrides) Datenmodell
  - ⇒ anwendungsspezifisch spezialisierbare generische (hybride) Anfragesprache für das Datenmodell
- Repräsentations-spezifische Folgerbarkeitsrelationen und Inferenzverfahren erforderlich
  - für die ABox-Anfragebeantwortung: |=
  - qualitatives räumliches Schließen
  - geometrische Inferenz ("Ablesen aus der Karte")





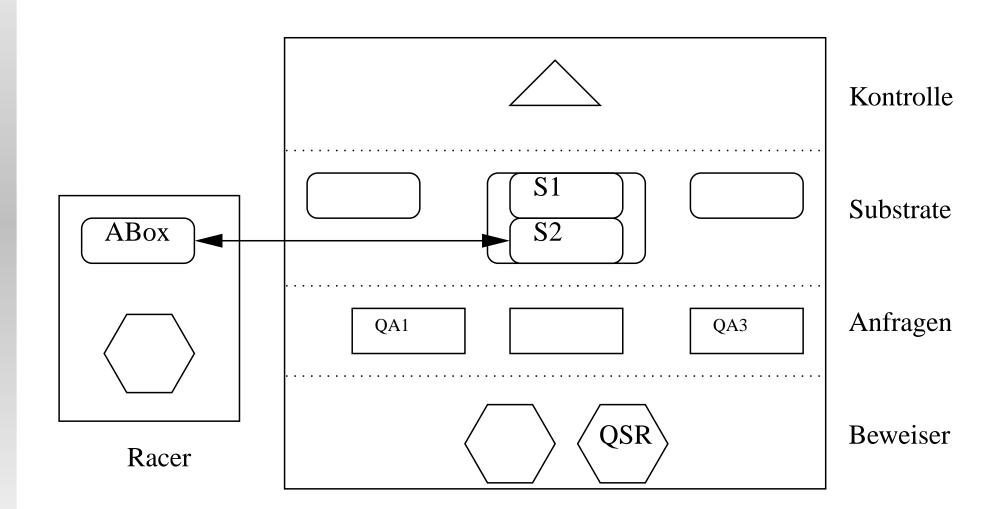


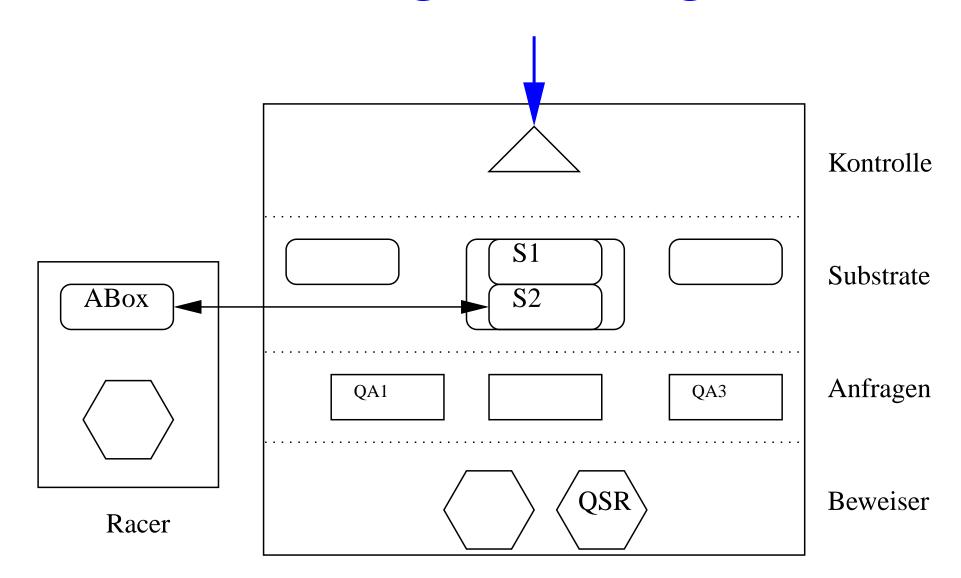


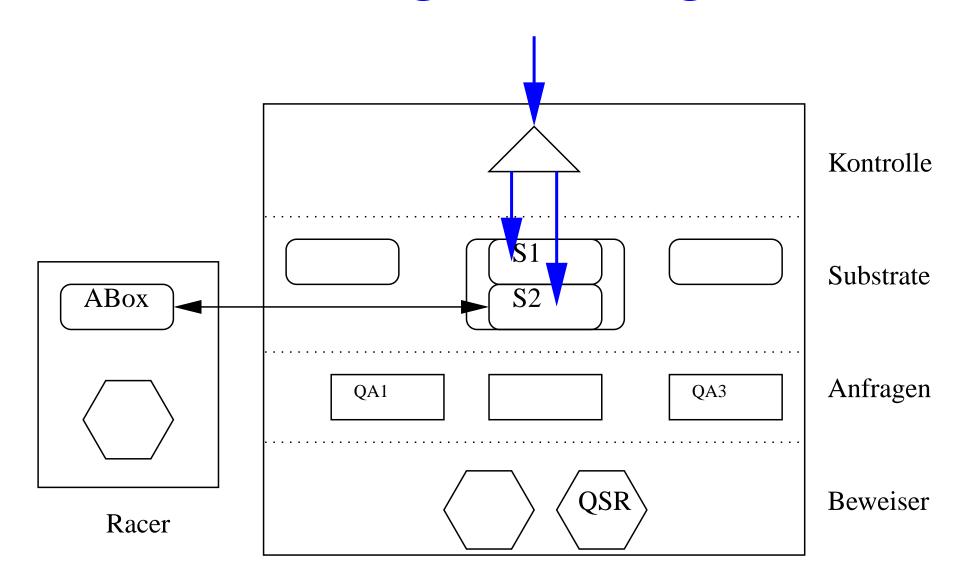
### **Architektur-Modell – Basis-Entwurf**

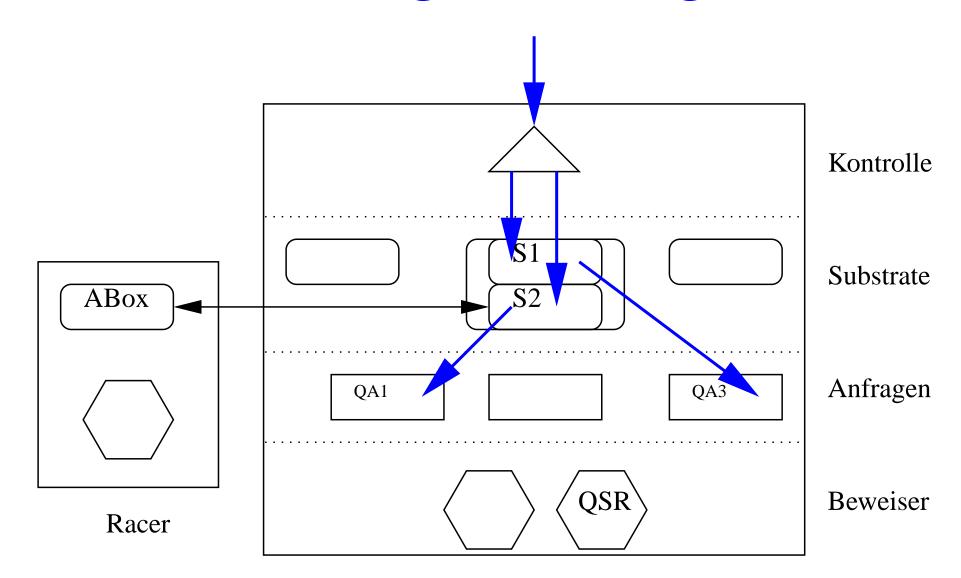
- Hauptkonzepte: Substrat-Datenmodell, generische Substrat-Anfragesprache, Beweiser
- Substrat = **Knoten- und Kanten-beschrifteter Graph** (semi-strukturiert)
- Anwendungsspezifisch spezialisierbar:
   Beschriftungssprachen variabel
- Substrat-Anfragesprache spezialisierbar:
   Substrat-spezifische Anfrage-Atome
- formalisiert in Prädikatenlogik 1. Stufe (**nur** zu Formalisierungszwecken)
- Semantik wohldefiniert: Konsistenz und Folgerbarkeit (=)
- ⇒ "Dienstqualität" anhand Semantik verifizierbar

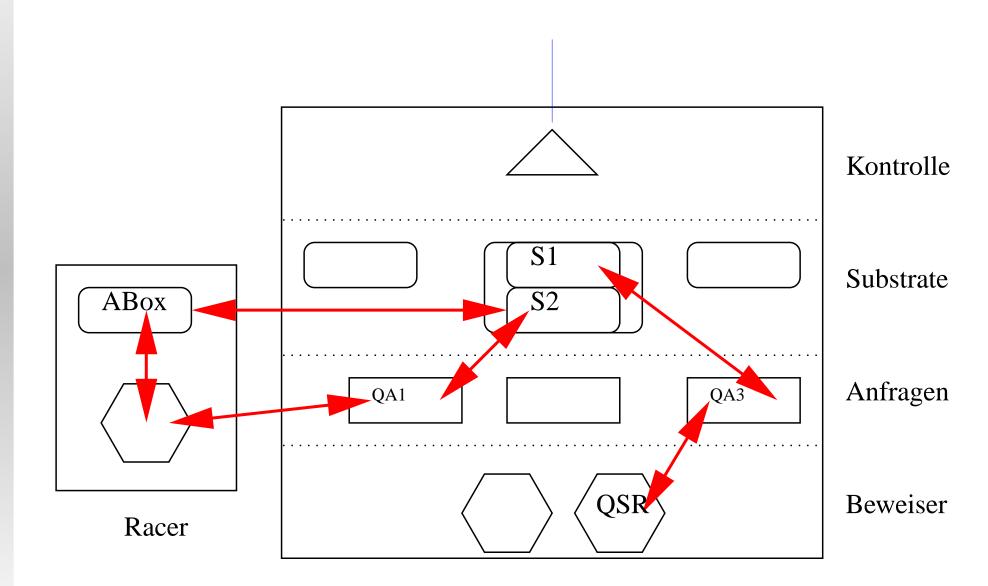
# **Architektur-Modell: Komponenten**

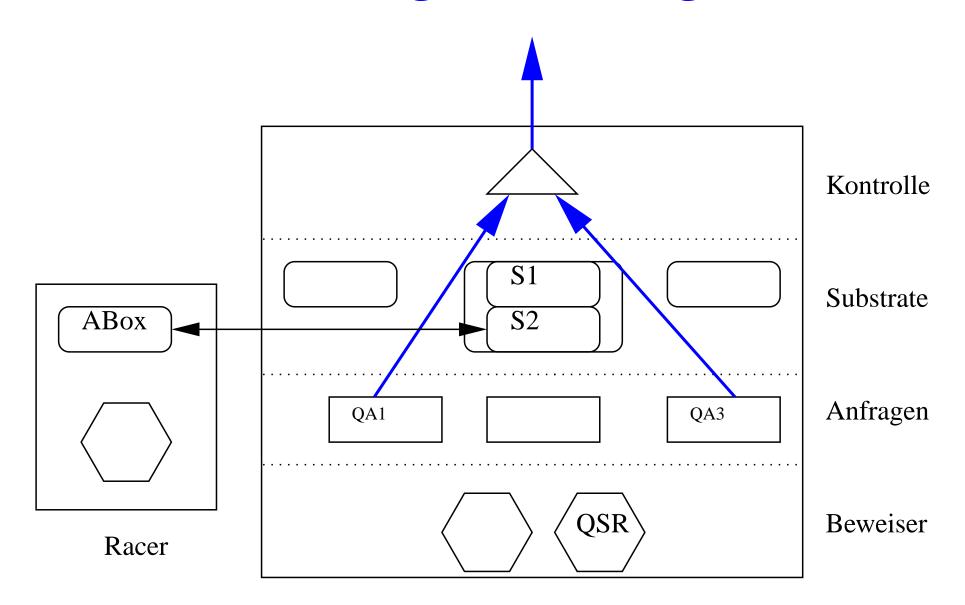












## Architektur: Anfragekonsistenz

- Ausnutzung weiterer wichtiger Eigenschaften qualitativer räumlicher Relationen: paarweise Disjunktheit
- $enth\ddot{a}lt(x,y) \wedge enth\ddot{a}lt(y,z) \wedge disjunkt(x,z)$
- $enth\ddot{a}lt(x,y) \wedge enth\ddot{a}lt(y,z) \models enth\ddot{a}lt(x,z)$
- enthält(x, z) widersprüchlich mit disjunkt(x, z)
- Qualitativer räumlicher Kalkül ("RCC"): zwischen je zwei Objekten gilt genau eine räumliche Relation
- entsprechende Verfahren für thematische Konsistenz, z.B.
   Reduktion auf ABox-Konsistenz
- ⇒ Algorithmen zur Entscheidung von Konsistenz &
   Enthaltensein für best. Klassen räumlich-thematischer
   Anfragen

### Definition des Substrat-Datenmodelles

• Knoten- und kantenbeschrifteter Graph:

$$S = (V, E, L_V, L_E, \mathcal{L}_V, \mathcal{L}_{\mathcal{E}})$$

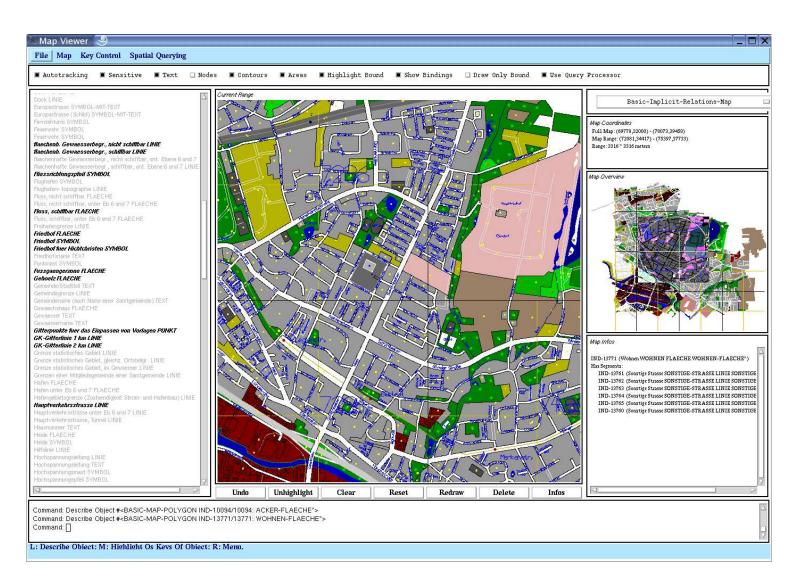
- Beschriftungssprachen  $\mathcal{L}_{\mathcal{V}}$ ,  $\mathcal{L}_{\mathcal{E}}$  variabel
  - ABoxen:  $\mathcal{L}_{\mathcal{V}} =_{def} \{ \text{Beschreibungen} \},$  $\mathcal{L}_{\mathcal{E}} =_{def} \{ \text{Relationsnamen} \}$
  - Qual. Karten:  $\mathcal{L}_{\mathcal{V}} = \emptyset$ ,  $\mathcal{L}_{\mathcal{E}} =_{def} \{ \text{räuml. Relations namen} \}$
  - Geometr. Karten:  $\mathcal{L}_{\mathcal{V}}=_{def}$  "Geometrie-Beschreibungssprache",  $\mathcal{L}_{\mathcal{E}}=\emptyset$
- Hybride Substrate:  $(S_1, S_2, *)$ , \* ist eine partielle injektive Funktion von  $V_1$  nach  $V_2$  (da  $*^{-1}$  benötigt)
- z.B.  $S_1 = RCC5$ -Netz (qualitative Karte),  $S_2 = ABox$

### Definition der Anfragesprache

- Substrat-Anfragesprache: Datalog-ähnlich (nicht-rekursiv mit Negation, sicher, Funktor-frei)
- Atome atom:
  - unäre (?x C) bzw. C(x))
  - binäre (?x ?y R) bzw. R(x,y))
  - C, R "passend" zu  $\mathcal{L}_{\mathcal{V}}$ ,  $\mathcal{L}_{\mathcal{E}}$ , evtl. komplex  $\rightarrow$  Flexibilität!
- Körper k, Kombination von atomen durch Konstruktoren:
  - {and, union, neg, project-to} bzw.  $\{\land (``,"), \lor, \setminus, \pi\}$
- Substrat-Anfragen:  $ans(\vec{x}) \leftarrow k \ (\vec{x} \ \text{in} \ k)$
- **Hybride Anfragen**: mischen Atome für  $S_1$  und  $S_2$

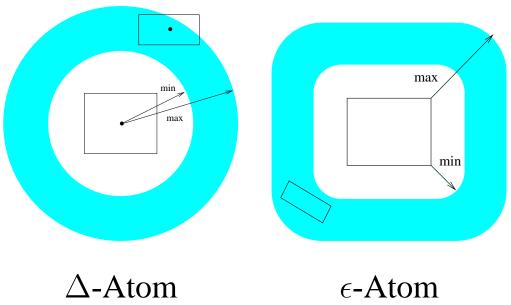
### **Evaluation: Digitale Stadtkarte**

### DLMAPS-System: Karten-Darstellung



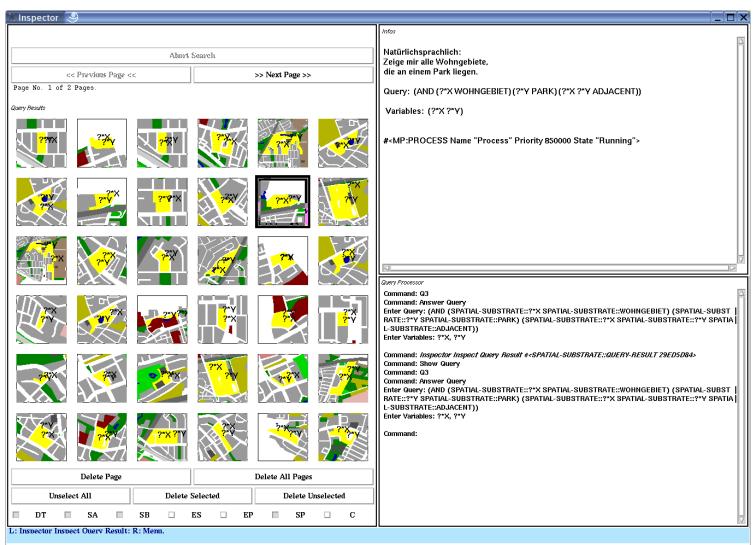
# **Evaluation: Digitale Stadtkarte**

- Illustration einer spez. Substrat-Anfragesprache: DISKQL
- bietet diskutierte thematische & qual. räumliche Atome
- Atome zur Adressierung der Geometrie der Karten-Daten:
  - geometr. Eigenschaften: (?x (:area (<min> <max>)))
  - geometrisch/metrische Relationen:  $\Delta$  und  $\epsilon$ -Atome



### **Evaluation: Digitale Stadtkarte**

DLMAPS-System: Anfrage-Schnittstelle "Welche Wohngebiete grenzen an einen Park"?



### Konkrete DISKQL-Anfrage

"Suche nach einer adäquaten Immobilie für einen Millionär"

```
(retrieve (?villa ?wohngebiet ?golf-club ?kirche)
    (and (?*wohngebiet
              (and wohngebiet
                   (or (all klassifikation villen-gegend)
                       (string= name ''Blankenese''))))
         (?wohngebiet ?villa :contains)
         (?*villa
              (and villa
              (all status zu-verkaufen) (> preis 10000000)
              (some hat-extra swimming-pool)))
         (?kirche ?wohngebiet (:inside-epsilon nil 200))
         (?wohngebiet ?golf-club :adjacent)
         (?*golf-club (and golf-club (all hat-mitglied millionär)))))
```

### Softwaretechnischer Rahmen

- Als Rahmenwerk in CLOS: "Klassen für alles"
  - Substrate,
  - Beschreibungen (Sprachen),
  - Anfragen (Atome, Körper), ...
- Erweiterbar durch Vererbung
- Nutzung von Mehrfachvererbung (z.B. räumliche ABox)
- Nutzung von Multi-Methoden: entails-p
- Baukastensystem zur Beweiser-Konstruktion: MIDELORA
  - Angebot: bis zur Beschreibungslogik SHIN
- Schablonenmethoden (Standard-Funktionalität)
  - komplexe Substrat-Anfragebeantwortungsmaschine

### Wissenschaftlicher Einfluss der Arbeiten

- RACERPRO-Nutzer: NRQL = expressive und skalierbare ABox- und SemWeb-Anfragesprache, wurde hier als Anfragesprache für thematische Aspekte genutzt
- Konsistenzcheck von UML-Diagrammen mittels NRQL-Anfragen
- RACERPRO + NRQL: erstes Beschreibungslogiksystem, das die sogen. LUBM-Anfragen evaluieren konnte
- Optimierungstechniken aus NRQL-Maschine wurden in andere Beschreibungslogik-Systeme übernommen
- NRQL-Regeln für Szenen-Interpretation
- Anwendung der Architektur im BOEMIE-Kontext zur Multimedia-Interpretation: Abduktion mit NRQL-Regeln

# Terminologische Rechtfertigung

- Warum handelt es sich um eine "flexible und konfigurierbare" Softwarearchitektur?
  - kann **Regionen** im Entwurfsraum abdecken (verschiedene QLs, Beweiser, Substrate, ...)
  - nicht-antizipierte Änderungen/Nutzungen waren möglich (Abduktion m. NRQL im BOEMIE-EU-Projekt)
  - ⇒ Flexibel = "Rahmen bricht nicht"
    - hohe Kombinierbarkeit und Orthogonalität der Basis-Abstraktionen
  - ⇒ Konfigurierbarkeit

### Wissenschaftlicher Nutzen

- Framework nicht öffentlich, aber Ausprägungen in RACERPRO verfügbar (NRQL)
- Komplexitätsresultate für einige DLs
- Formale & softwaretechnsiche Abstraktionen
- ⇒ kommuniziert und kommunizierbar (Publikationen) und daher nutzbar von anderen Forschern (unabhängig von COMMON LISP etc.)
  - Entwürfe stellen eine Art "Landkarte für ein intellektuell schwieriges Terrain" oder "Blueprints" (f. "Referenzarchitekturen") dar

### **Ausblick**

- Erweiterung der Arbeiten geplant: WSL, Schweiz (Eidg. Forschungsanstalt für Wald, Schnee und Landschaft)
- DFG-Projekt wird beantragt
- ⇒ Arbeiten werden von Domänen-Experten aufgegriffen

### Was nicht gesagt werden konnte

- Anwendungsdomäne 2: "Information Retrieval im Semantic Web"
- NRQL im Detail, innovative Anfragekonstrukte: project-to, neg (Negation as Failure)
- Umgehung des "Symbolic Bottlenecks": räumliche ABoxen mit virtuellen Assertionen
- verschiedene Ausprägungen der Substrat-Anfragesprache
- Baukastensystem zur Konstruktion von Beweisern:
   MIDELORA

Danke
für Ihre
Aufmerksamkeit!

## Semantik von Anfragen

- Semantik von  $\{\land, \lor, \setminus, \pi\}$  fix: algebraisch,  $\{\cap/\times, \cup, \setminus, \pi\}$
- Semantik von atom durch  $S \models atom$  definiert:
  - $C(x)^{\mathcal{E}} =_{def} \{i \mid S \models C(i)\}$
  - $R(x,y)^{\mathcal{E}} =_{def} \{ (i,j) \mid S \models R(i,j) \}$
- |= definiert f. Grund-Atome
- ⇒ Semantik definiert
  - $S \models grund\_atom$  entscheidbar (hier stets angenommen)
- ⇒ Anfragen entscheidbar
  - aber: aufgrund von Negation as Failure (NAF)-Semantik von "\":  $\models_{NAF}$  statt  $\models$  (nicht-monoton)
  - Idee:  $\models$  f.  $S \times atom$  als Multi-Methode

### Tableaukalkül: Regeln

### **□-Regel:**

#### wenn

- 1.  $x: C_1 \sqcap C_2 \in \mathcal{A}$
- 2.  $\{x: C_1, x: C_2\} \not\subseteq A$

#### dann

$$\mathcal{A}' := \mathcal{A} \cup \{x : C_1, x : C_2\}$$

### **□-Regel**:

#### wenn

- 1.  $x: C_1 \sqcup C_2 \in \mathcal{A}$
- $2. \quad \{x: C_1, x: C_2\} \cap \mathcal{A} = \emptyset$

dann 
$$\mathcal{A}' := \mathcal{A} \cup \{x : C_1\}$$
  
 $\mathcal{A}_1 := \mathcal{A} \cup \{x : C_2\}$ 

### ∃-Regel:

#### wenn

- 1.  $x: \exists R.C_1 \in \mathcal{A}$
- 2. es gibt kein y, sodass

$$\{y:C_1,(x,y):R\}\subseteq\mathcal{A}$$

#### dann

$$\mathcal{A}' := \mathcal{A} \cup \{y : C_1, (x, y) : R\}$$

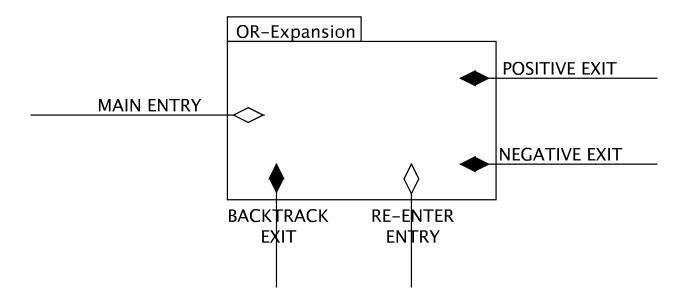
### $\forall$ -Regel :

#### wenn

- 1.  $\{x: \forall R.C_1, (x,y): R\} \subseteq \mathcal{A}$
- 2.  $y:C_1\notin\mathcal{A}$

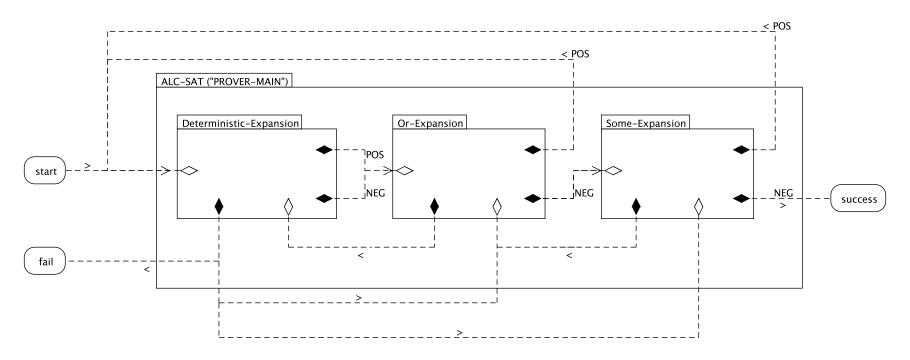
dann 
$$\mathcal{A}' := \mathcal{A} \cup \{y : C_1\}$$

### Modell einer (Tableau-) Regel



- MAIN ENTRY: beim 1. Betreten des Knotens im Suchbaum
- POSITIVE EXIT: Regel war anwendbar
- NEGATIVE EXIT: Regel war nicht anwendbar
- BACKTRACK EXIT: Kontrolle an Eltern-Knoten zurückgeben
- RE-ENTER ENTRY: Kontrolle von Kind-Knoten zurückerhalten

Regelanwendungskette f. ALC-Beweiser



- (define-prover ((<Aufgaben-Klasse><Sprach-Klasse> <ABox-Klasse>)) ...)
- MIDELORA nutzt ein ABox- bzw. Tableau-Substrat
- vollständig objektorientiert
- → Kompensationsoperationen (Historien)
  - bietet räumliche ABoxen
  - "viele kleine spezialisierte Beweiser" statt eines großen
- ⇒ Wartbarkeit, modulare Verstehbarkeit (konzise)
  - dann Wiederverwendung von Teilen notwendig: Regeln
  - Wiederverwendung im MIDELORA-Raum organisiert:

$$\mathcal{T} \times \mathcal{DL} \times \mathcal{K}$$