

---

# Introduction to Data and Knowledge Engineering Tutorium 10 - EBL & Semantic Web

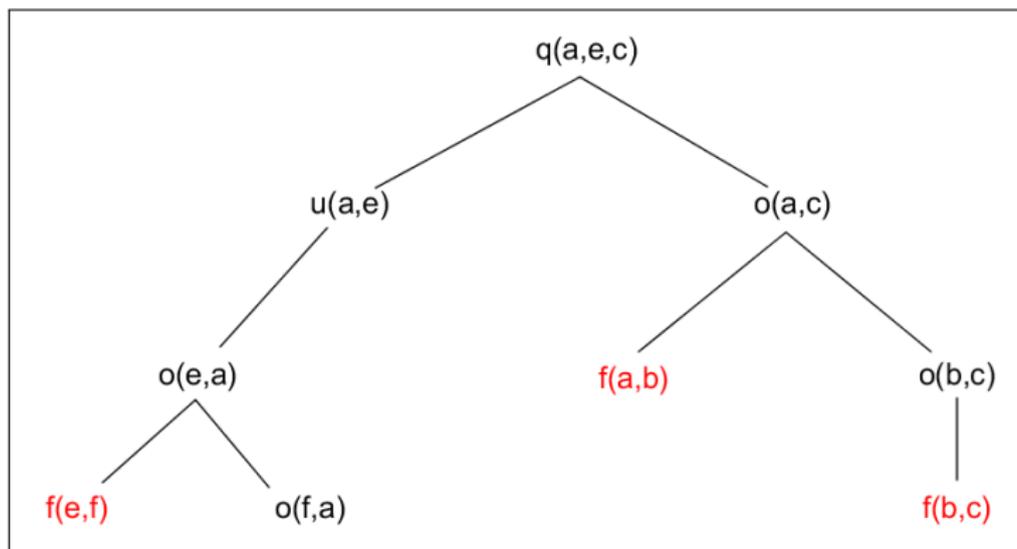
---



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

## Aufgabe 10.3 Explanation-Based Learning (EBL)

a.1) Die Konstruktion eines Beweisbaums für die Query:  $q(a, e, c)$  misslingt...



## Aufgabe 10.3 Explanation-Based Learning (EBL)



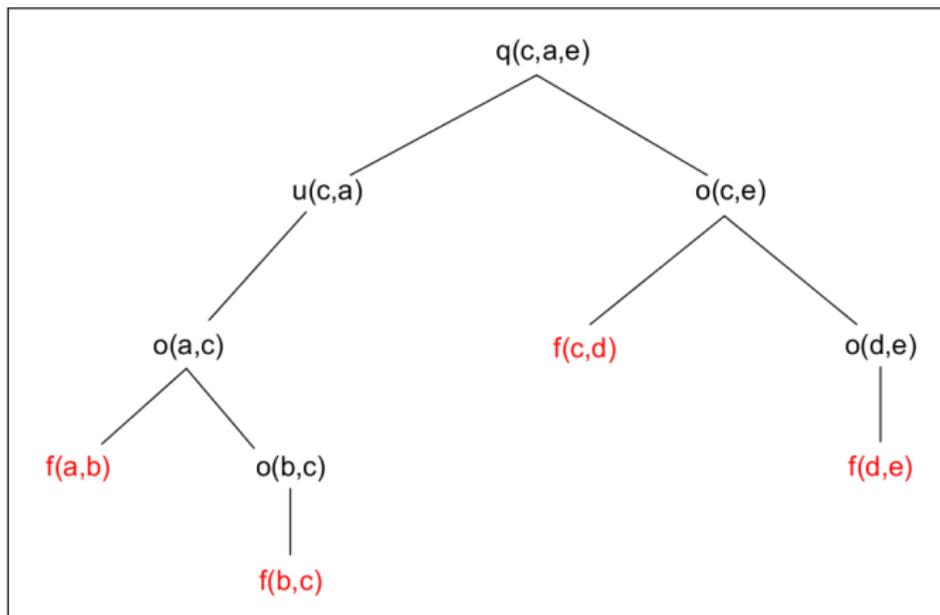
TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

### a.1) Schlussfolgerung:

Die Query:  $q(a, e, c)$  lässt sich nicht beweisen, da  $o(f, a)$  nicht hergeleitet werden kann.

## Aufgabe 10.3 Explanation-Based Learning (EBL)

a.2) Beweisbaum für die Query:  $q(c, a, e)$



### a.2) Schlussfolgerung:

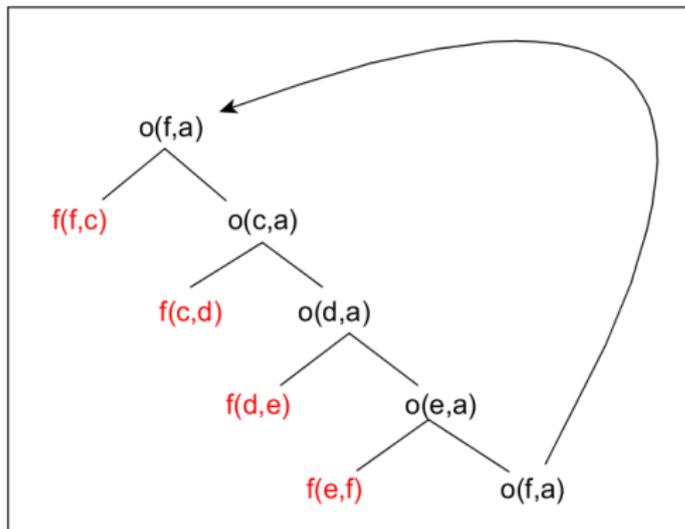
Die Query:  $q(c, a, e)$  lässt sich beweisen, da sämtliche Blätter des Baumes:

1.  $f(a, b)$
2.  $f(b, c)$
3.  $f(c, d)$
4.  $f(d, e)$

...aus Fakten bestehen, die bekanntlich stets wahr sind.

## Aufgabe 10.3 Explanation-Based Learning (EBL)

b.) Nein. Das zusätzliche Faktum  $f(f, c)$  führt zu einer Endlosrekursion. Auch das Ändern der Reihenfolge bei der Regel  $o(X, Y)$  würde nicht ausreichen, um  $o(f, a)$  zu einem Faktum herzuleiten.



## Aufgabe 10.3 Explanation-Based Learning (EBL)



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

c.) Die generalisierte Regel lautet:

$q(X,Y,Z) :- f(Y,B), f(B,X), f(X,D), f(D,Z).$

## Aufgabe 10.3 Explanation-Based Learning (EBL)



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- c.) **Hinweis:** Die erzeugte Regel kann alleine aus den gegebenen Fakten und Regeln gewonnen werden. Insbesondere kommt kein neues Wissen hinzu, jedoch wird der Beweis deutlich verkürzt, da die neue Regel nur aus Fakten besteht.
- d.) Die Generalisierung des Beweises erfolgt natürlich deduktiv, da wir von Fakten auf eine "neue" Regel schließen.

## Aufgabe 10.4 Semantic Web - Theoriefragen



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

a.) Es handelt sich hierbei um einen Ansatz des Semantic Web's, da die Daten welche gecrawlt werden, eine maschinenlesbare Struktur besitzen, welche den typischen Web-Mining Ansatz überflüssig macht (dieser zielt für gewöhnlich auf unstrukturierte Daten im Web).

b.)

**Data Mining:** Die Methoden dieser Disziplin werden auf strukturierte Daten angewendet, also auf Daten, wie sie typischerweise in relationalen Datenbanksystemen in Tabellenform gespeichert sind.

Typisches Problem im Alltag: selten liegen Daten strukturiert vor!

Es gibt Schätzungen die besagen, dass ca. 85% aller Informationen im Web unstrukturiert sind. Data-Mining wäre hier also nicht ohne weiteres denkbar.

**Web Mining:** Da die Analyse von Web-Inhalten eigene Problematiken beinhaltet, hat sich hier eine weitere Variante des Data Mining entwickelt, das sogenannte Web Mining. Diese wendet Methoden des Data Mining an, um Datenstrukturen im WWW zu untersuchen, wobei neben dem eigentlichen Seiteninhalt und der Struktur von Websites auch das Nutzerverhalten zum Gegenstand gemacht wird.

c.)

- ▶ **XML:** ist eine Klasse von Auszeichnungssprachen

Es existieren zahlreicher "prominente" XML-Sprachen, wie z.B.: RSS, GraphML, XHTML, SVG, etc.

XML kann benutzt werden:

- ▶ für standardisierte Datendarstellung in Web
- ▶ zur Syntaxdefinition andererer Markupsprachen (XHTML)
- ▶ zur Trennung von Form und Inhalt auf Webseiten

c.)

- ▶ **XML Schema:** ist eine Sprache mit der man XML-Sprachen definieren kann. So erlaubt XML Schema beispielsweise eine genaue Typisierung der Daten als Zeichenketten, Ganzz- bzw. Fließzahlen, Datum, etc.

Von einer Reihe vordefinierter Grundtypen können nach bestimmten Regeln sogar eigene Datentypen abgeleitet werden.

Es sollte bedenklich werden, dass durch die vielfältigen Möglichkeiten, die XML Schema bietet, auch ein großer Nachteil besteht. XML Schema gilt wegen seiner Komplexität als außerordentlich schwierig zu erlernen und ist leider noch dazu sehr unübersichtlich, was besonders die Fehlersuche bei der Modellierung erschwert.

c.)

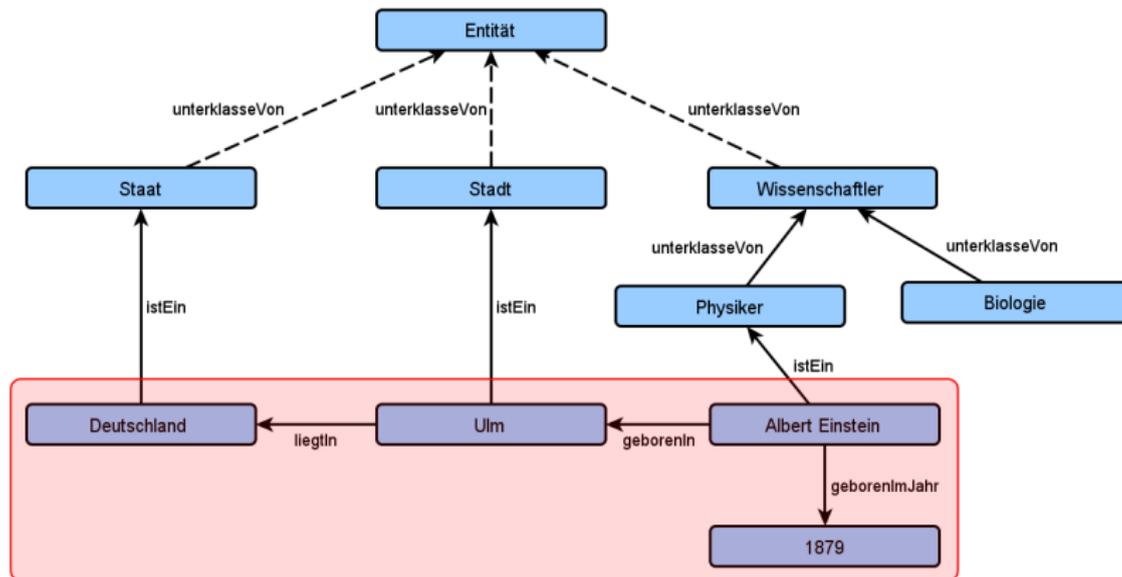
- ▶ **RDF:** Ist ein Datenmodell für Objekte und Beziehungen zwischen konkreten Objekten.
- ▶ **RDFS:** Beschreibt die grundlegenden Elemente, die für die Modellierung der Ontologien benutzt werden können.
  - ▶ Man kann Klassen, Unterklassen, Merkmale und Untermerkmale definieren.
  - ▶ RDFS gibt keine syntaktischen Bedingungen an.
  - ▶ Es gibt Schlüsselwörter, wie: `Class`, `subPropertyOf`, `subClassOf`, etc.

RDFS kann nicht für komplizierte Ontologiedarstellungen benutzt werden, so kann man beispielsweise folgendes nicht beschreiben:

- ▶ ein Buch hat nur einen Preis aber vielleicht mehrere Autoren
- ▶ Datentypen für verschiedene Teile des Buchs (Titel, Preise usw.)
- ▶ ein Buch darf nicht z.B. Hardcover und Softcover gleichzeitig sein können.
- ▶ **OWL:** Ist eine vollständige spezifische Sprache zur Ontologie-Darstellung

## Aufgabe 10.5 Ontologien (OWL, RDF & RDFS)

Mittels RDF können lediglich nur spezifische Fakten kodiert werden, daher...



## Aufgabe 10.5 Ontologien (OWL, RDF & RDFS)



Mit RDFS wiederum wird allgemeines Wissen über die Domäne kodiert, also alle "abstrakteren" Entitäten, bzw. Relationen zwischen diesen.

Beispiele für zweistellige Relationen in dem Ontologie-Graph:

- ▶ `liegtIn(Ulm, Deutschland)`
- ▶ `geborenIn(Albert Einstein, Ulm)`
- ▶ `geborenImJahr(Albert Einstein, 1879)`



**Viel Glück für die Klausur !!!**