

# Vorlesung Netzbasierte Informationssysteme

## Rückblick

**Prof. Dr. Adrian Paschke**

Arbeitsgruppe Corporate Semantic Web (AG-CSW)  
Institut für Informatik, Freie Universität Berlin

[paschke@inf.fu-berlin.de](mailto:paschke@inf.fu-berlin.de)

<http://www.inf.fu-berlin.de/groups/ag-csw/>



# Ziel der Lehrveranstaltung

---

- „**Netzbasierte Informationssysteme** stellen mit der Verbreitung des Web im weltweiten Maßstab Informationen bereit. Die Vorlesung soll Kenntnisse um die wichtigsten Technologien, Probleme und Lösungsansätze solcher Systeme vermitteln. Im Übungsteil wird das Verständnis vertieft.“
- Integrierte Sicht auf relevante technische Aspekte mit Schwerpunkt **(Corporate) Semantic Web Technologien**

# Zusammenfassung

---

- Internetbasierte Informationssysteme als tragende Infrastruktur für die Durchführung von Kollaborationen, Geschäftsprozessen,
  - Workflow, Groupware, ...
- Technisches Verständnis ist wichtig,
  - um das Potential neuer Techniken einschätzen zu können,
  - und dieses gestaltend einsetzen zu können
- Moderne Netzbasierte Informationssysteme erfordert ein Verständnis für Zusammenhänge verschiedener Technologien
  - Netzwerkprotokollen
  - Softwarearchitekturen verteilter netzbasierter IS
  - XML, Web Services und neue Web-Standards
  - Semantic Web Technologien und Complex Event and Rule Processing

# Gliederung

---

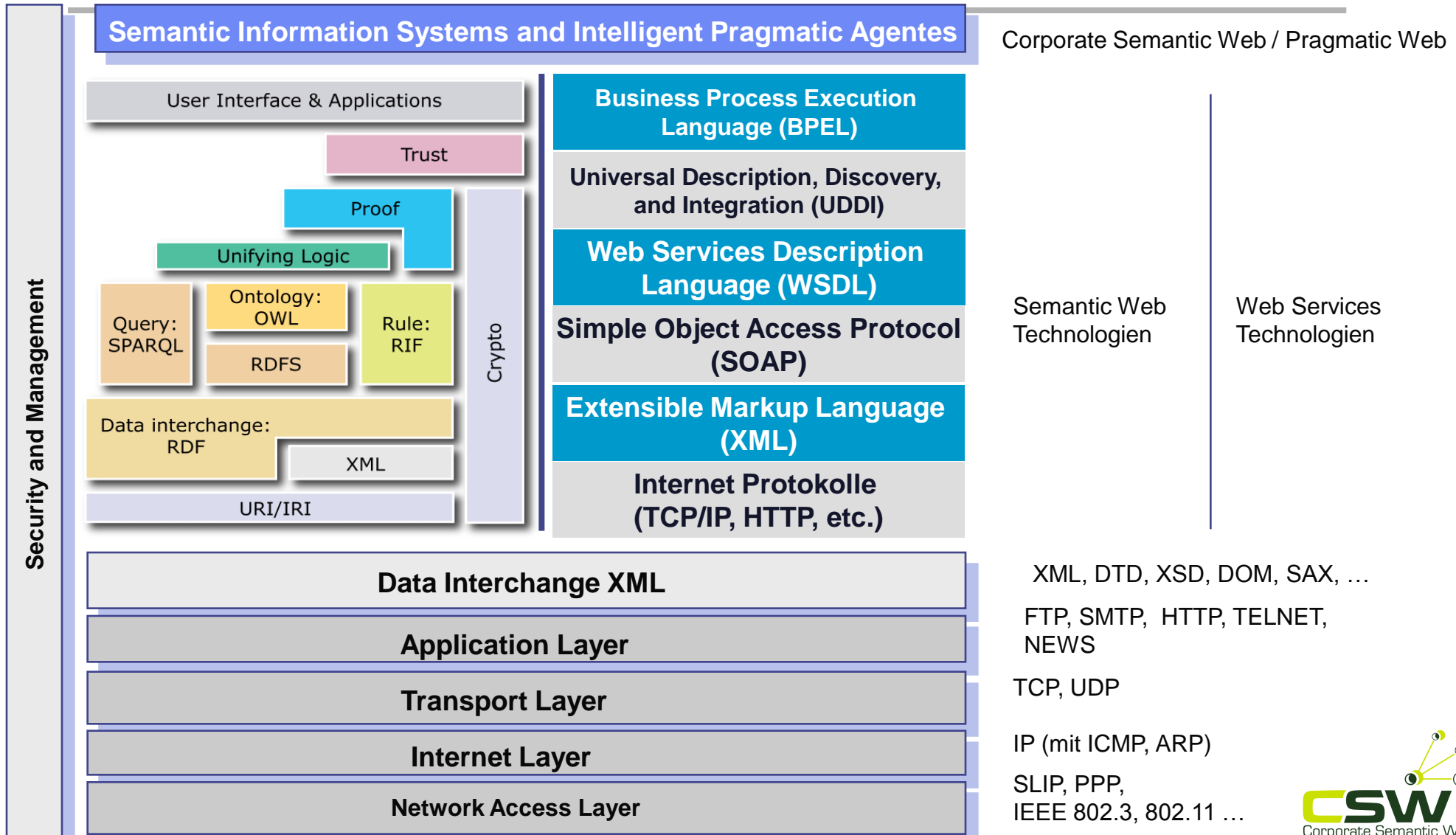
## **Technologien und Standards**

- **Netzwerkprotokolle**
- **Sichere Internetprotokolle**
- **Web-basierte Informationssysteme I**
- **Web-basierte Informationssysteme II**
- **XML**
- **Service Oriented Architectures und Web Services Computing**
- **Semantic Web I**
- **Semantic Web II**
- **Semantic Web III**

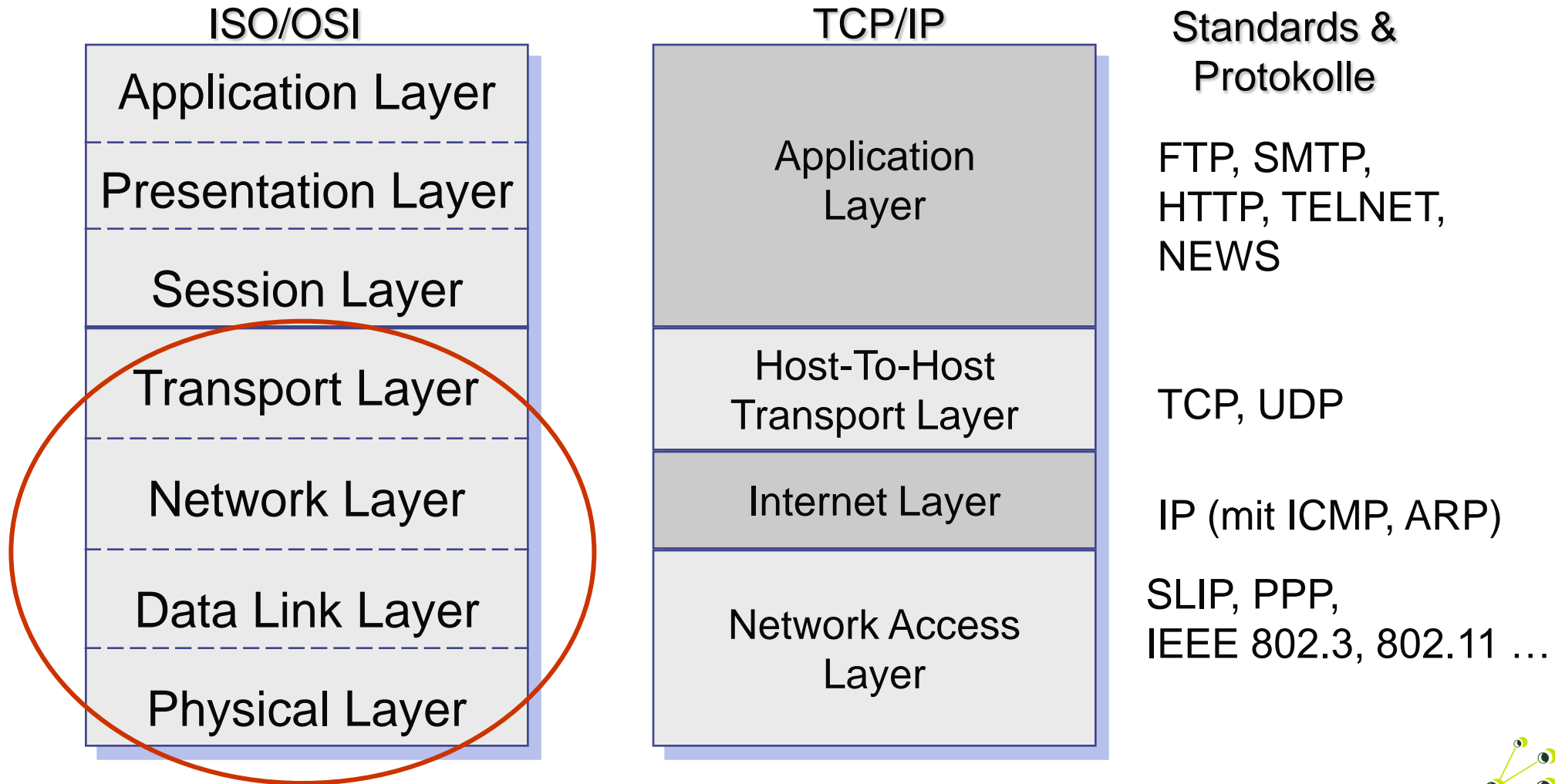
## **Methoden und Anwendungen in & für (Semantische) Netzbasierte Informationssysteme**

- **Struktur und Erschließung des Webs**
- **Information Retrieval and Filtering**
- **Information Discovery - Text Mining**
- **Web Mining und Suchmaschinenverfahren**
- **Ontology Engineering**
- **Corporate Semantic Web**

# Überblick Technologien und Standards



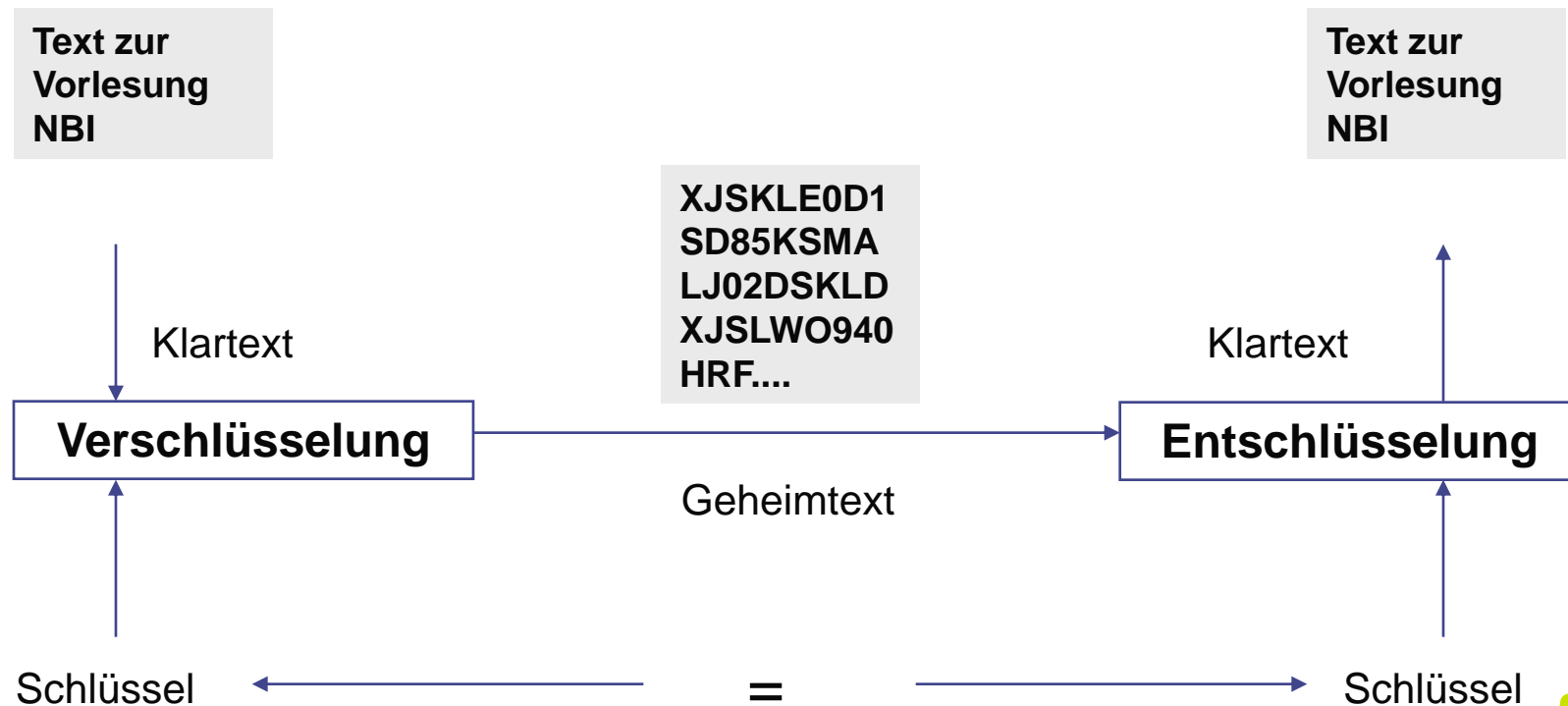
# Einheiten 2+3: Internetprotokolle + Sicherheit



- 
- **Wozu dient die Netzmaske beim Routing-Vorgang (5 P) ?**
    - Zur Entscheidung, ob eine IP-Adresse zu einem Netzwerk gehört, wird sie mit der Netzmaske bitweise AND-verknüpft
    - Mithilfe der Netzmaske lassen sich mehrere IP-Adressen zusammenfassen
    - Mehrere zusammengefasste IP-Adressen nennt man ein Subnetzwerk
    - Alle Bits des Netzwerkteils sind auf 1 und alle Bits des Geräteteils auf 0 gesetzt.
  - **Nennen Sie 3 Probleme bei der Übertragungssicherheit von IP Version 4. (3 P)**
    - Inhalt der Pakete ist einsehbar und modifizierbar
      - Übertragung der Protokollelemente der Anwendungsebene im Klartext
    - Routingvorgang ist unsicher
      - Netzwerke außerhalb der unternehmensinternen LANs unterliegen weder der eigenen Kontrolle noch der einer zentralen Autorität
      - Keine Zustellungsgarantie
    - IP Adresse kann nicht zur Authentifikation von Personen verwendet werden

# Ver- und Entschlüsselung mit symmetrischer Kryptographie

**Erklären Sie das Prinzip der symmetrischen Verschlüsselung und erläutern Sie dieses anhand eines symmetrischen Verschlüsselungsverfahrens? (4 P)**



# Klassische symmetrische Verfahren

---

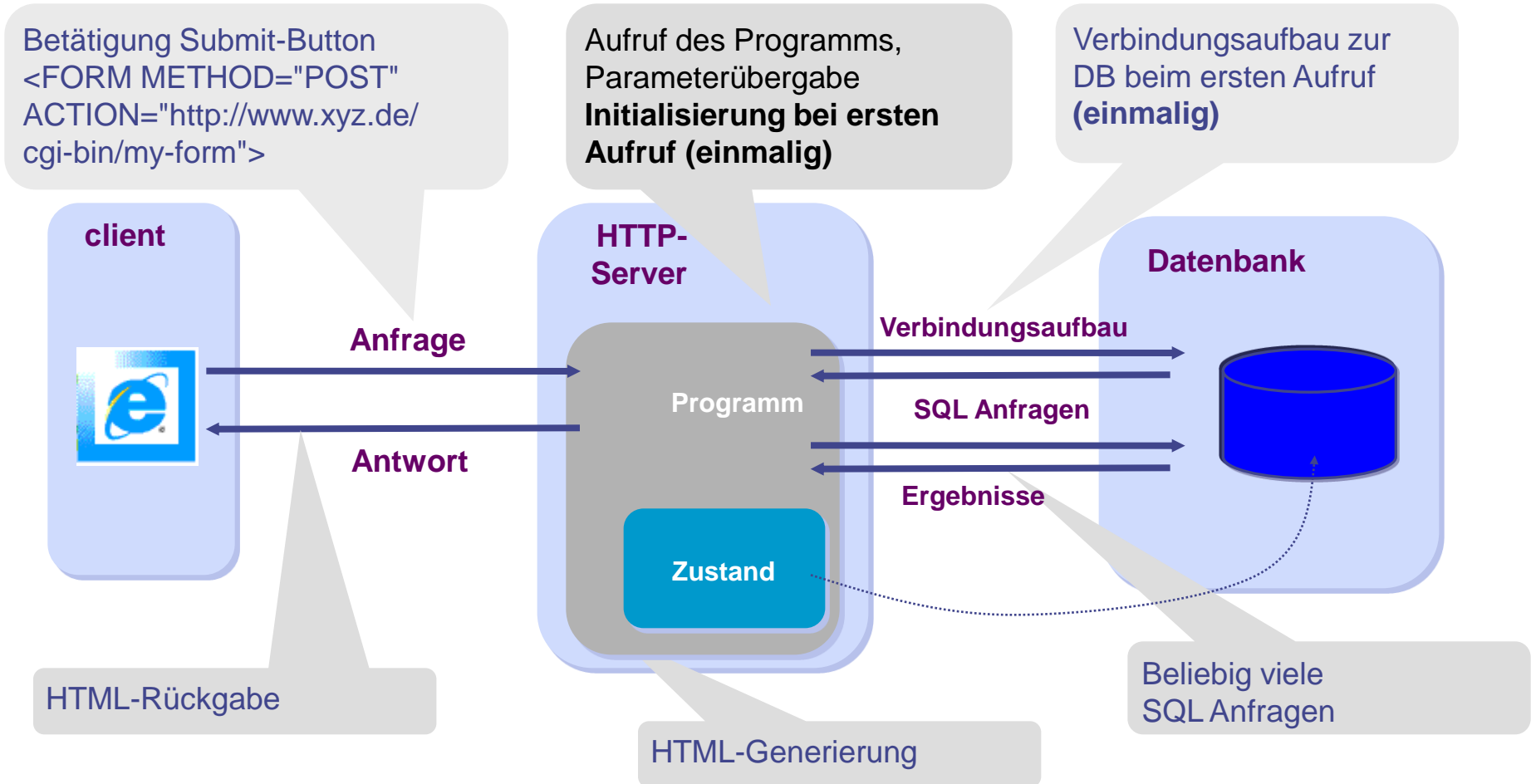
## ■ Cäsar-Verfahren

- $c = (p + s) \bmod 26$ , bei  $s = 3$
- **A** → **D**
- **B** → **E**
- **C** → **F**
- . . . .

## ■ Vigenère-Verfahren

- abhängig von der Position im Text (zusätzliches Verschlüsselungswort), z. B. **HALLO**
- **HALLOHALLOHALLOHALLO**
- **DIESISTEINENACHRICHT**
- Abbildung: Caesar-Verschiebung um den Wert des zugeordneten Buchstabens :
  - **D** → **D** + **H** = **K**
  - **I** → **I** + **A** = **S**
  - **E** → **E** + **L** = **P**
  - . . .

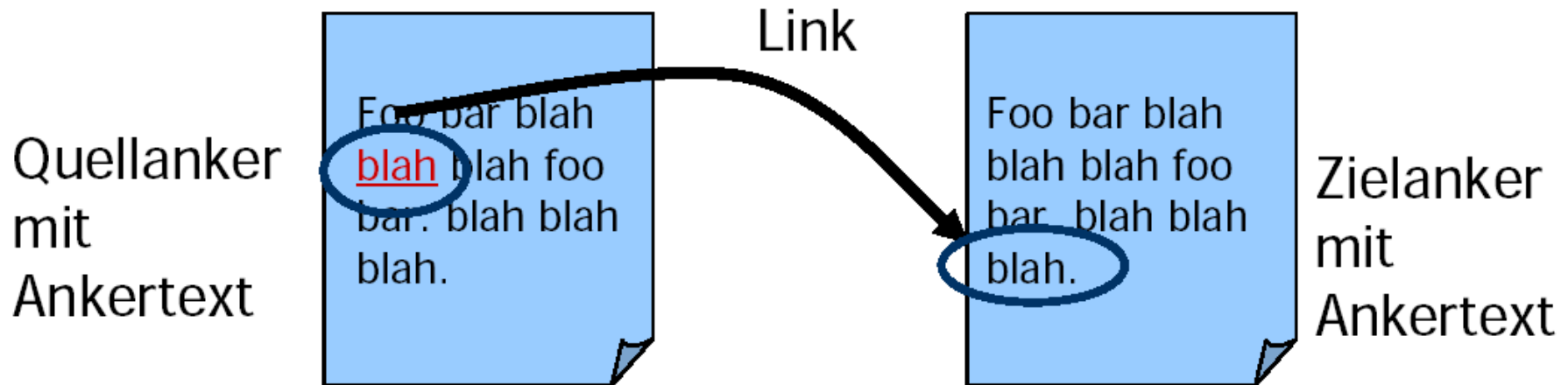
# Einheit 4: Web Informationssysteme I



# Einheit 5: Web Informationssysteme II

---

- Architektur des Webs und Webanwendungen



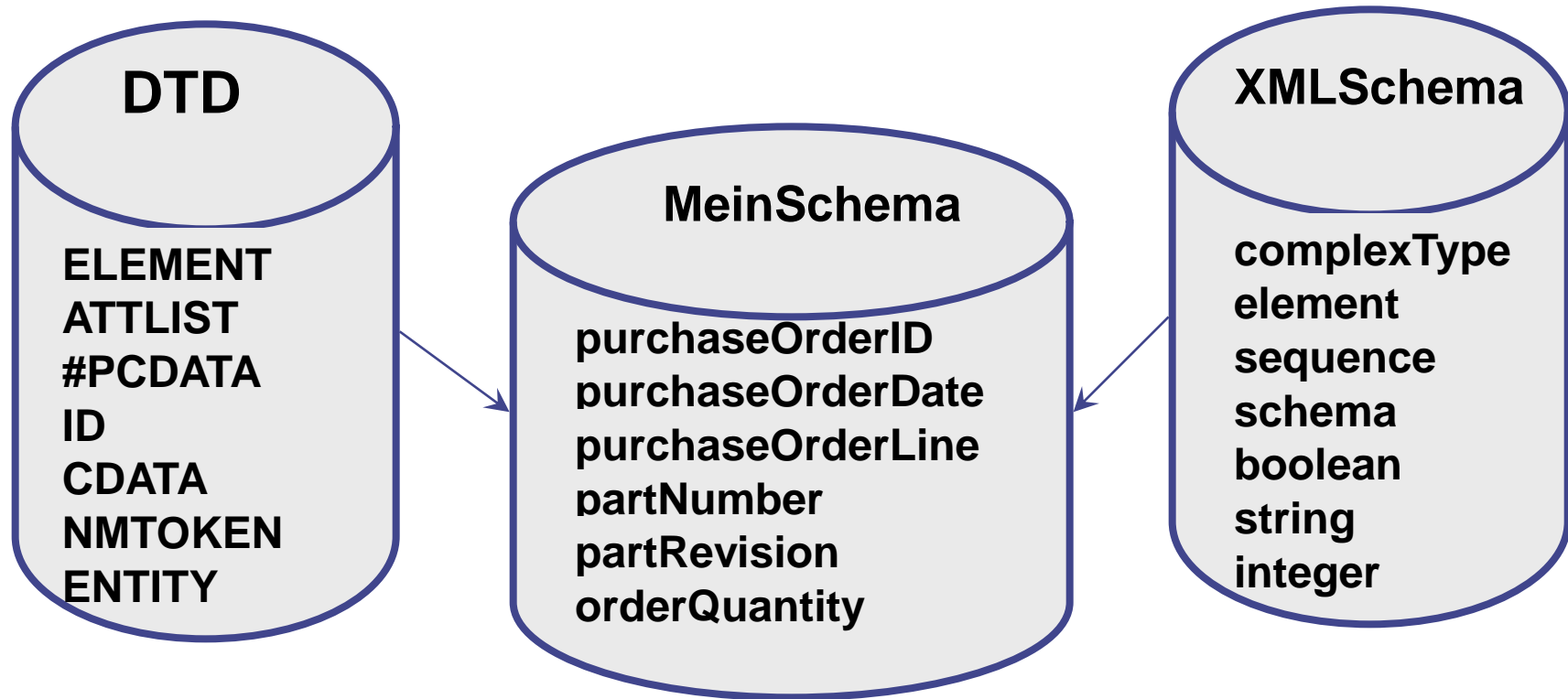
- 
- Erläutern Sie die Funktionsweise von CGI?
    - Serverseitiges Verfahren
    - Server stellt fest: Anfrage bezieht sich auf Programm/Skript (kein Dokument)
    - Server startet das Programm/Skript und übergibt etwaige Argumente (z. B. aus HTML-Formular)
    - Server gibt Output des Programms/Skripts (HTTP + HTML) über Web Server zurück an den Client
    - Im Falle eines Fehlers übergibt der Server eine Fehlermeldung an den Client zurück

- 
- Nennen Sie drei Beispiele für sprachabhängige Darstellung von Inhalten
  - Was ist Lokalisierung?
  - Was ist Internationalisierung?
  - Wie bezeichnet man Sprachen und Länder nach ISO und den neuem RFC Standard?
  - Wie kann man in HTML Spracheigenschaften markieren?
  - Wie kann man sprachspezifische Darstellungseigenschaften in CSS notieren?
  - Welche http Mechanismen unterstützen mehrsprachige Sites?
  - Was ist ein Zeichenrepertoire?
  - Was ist Unicode?

- 
- Nennen Sie zumindest 4 Object-Services, die im CORBA-Standard der OMG definiert sind. Erläutern Sie in jeweils einem Satz die Aufgaben der jeweiligen Services. (4 P)
  - Worin liegt der Unterschied zwischen Session Beans und Entity Beans? (3 P)
  - Beschreiben Sie die Schritte, die ein Klienten bei einem Zugriff auf eine EJB durchführen muss. (6 P)
  - Nennen und erläutern Sie die drei Grundlagen der Web-Technologie.
  - Was sind URIs und URNs?
  - Nennen Sie zwei Prinzipien oder Gute Praxis zur Verwendung von URIs
  - Wie sind Anforderungsmittelungen in http aufgebaut?
  - Wie sind Antwortmittelungen in http aufgebaut?
  - Welche Ziele verfolgt die REST Architektur des Web?
  - Nennen und erläutern Sie zwei Anforderungen bei der Wahl der Web-Architektur
  - Nennen und beschreiben Sie zwei Aspekte des REST Architectural Style
  - Wie ist es um die Lebensdauer von URLs bestellt? Warum kann das problematisch sein?
  - Nennen Sie die Grundkonzepte von Auszeichnungssprachen in der SGML/HTML/XML Welt

# Einheiten 6 : XML und XML Processing

---



Definiert in the XML 1.0 Spezifikation  
<http://www.w3.org/TR/REC-xml>

Definiert in XMLSchema Spezifikation  
<http://www.w3.org/2001/XMLSchema>

- 
- Beschreiben Sie in einem Satz die Hauptfunktionalität folgender XML Standards: XML, XML Schema, DTD, Namespaces, XSLT, XPath, XQuery
  - Was ist eine DTD? Erläutern Sie den Sprachumfang von DTD?
  - Übersetzen Sie folgendes DTD-Element in einen äquivalenten Ausdruck eines XML Schemas:  
<! ELEMENT Person (ID|(Name, Lastname)) >. Wie ändert sich die XML Schema Elementbeschreibung, wenn die beiden Elemente „Name“ und „Lastname“ in beliebiger Reihenfolge vorkommen dürfen?

# DTD – Document Type Definition

- Beschreibt die logische und physikalische Struktur eines XML-Dokuments (Grammatik)
- Nicht XML-Syntax
- Kennt 6 Typen, Hauptsächlich
  - Elemente
  - Attributlisten
  - Kommentare
  - Entities
  - Notationen
  - Prozess-Instruktionen
- Rudimentäre Datentypen
  - PCDATA | CDATA
- In- oder Outline eingebunden

## Beispiel DTD

```
<!ELEMENT Bestelldokument (Lieferadresse, Bestellung, Zahlung)>
<!ATTLIST Bestelldokument datum CDATA #REQUIRED
                                prio ( hoch | niedrig ) "niedrig"
>
<!ELEMENT Lieferadresse (#PCDATA)>

<!ELEMENT Bestellung (Artikel+)>
  <!ELEMENT Artikel (Artikelnummer, Beschreibung, Anzahl)>
    <!ELEMENT Artikelnummer (#PCDATA)>
    <!ELEMENT Beschreibung (#PCDATA)>
    <!ELEMENT Anzahl (#PCDATA)>

<!ELEMENT Zahlung (Zahlungsart | Zusatzangaben?)>
  <!ELEMENT Zahlungsart (#PCDATA)>
  <!ELEMENT Zusatzangaben (#PCDATA)>
```

\* + ?

---

Gegeben sei folgendes XML-Dokument:

```
<?xml version="1.0" encoding="utf-8" standalone="yes" ?>
<notizen>
  <notiz title="Training">
    <termin>Termin ist der 20.07.2008</termin>
    <ort>Ort ist die Sporthalle Sued</ort>
    <detail>Frau Mitkommerin abholen</detail>
    <detail>vorher tanken</detail>
  </notiz>
  <notiz title="ILS Klausur">
    <detail>Lerngruppe jeden Dienstag und Mittwoch </detail>
    <detail>20.00 Uhr bei Frau Mitlerner</detail>
  </notiz>
</notizen>
```

a) Formulieren Sie eine Grammatik (XML-DTD oder XML-Schema), welche obiges XML-Dokument erfolgreich validiert

b) Welche Elemente werden durch nachfolgende XPath-Ausdrücke identifiziert?

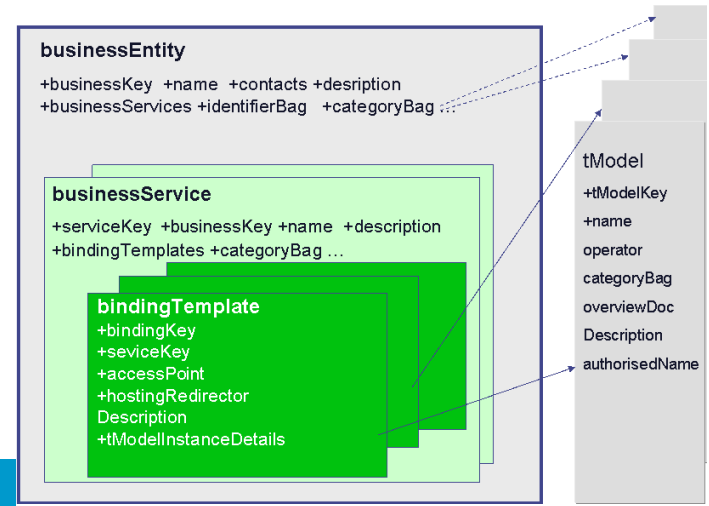
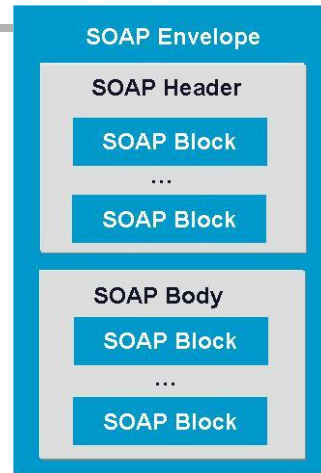
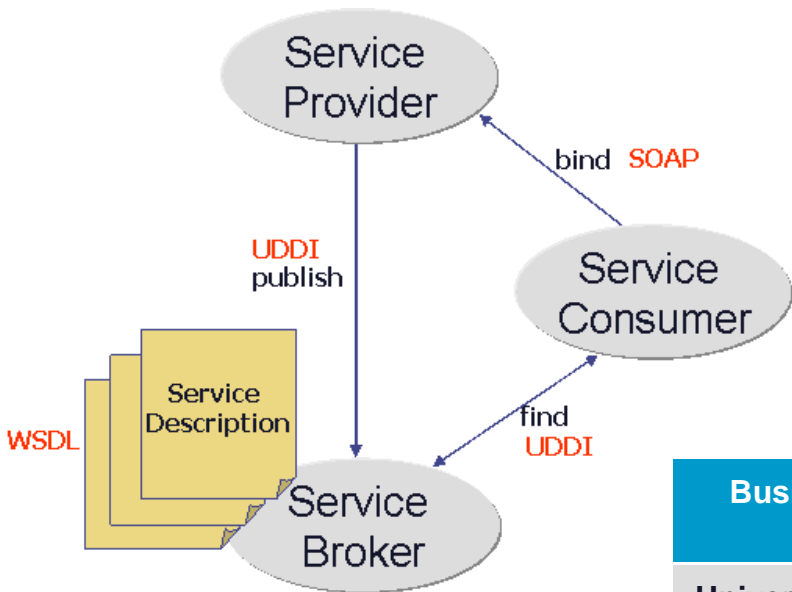
```
//notiz
```

```
//notiz[@title="Sport"]/detail
```

```
/notizen/notitz[2]/detail[2]
```

```
//notiz[@title="Training"]/ort
```

# Einheit: 7: Web Services Computing



- Business Process Execution Language (BPEL)
- Universal Description, Discovery, and Integration (UDDI)
- Web Services Description Language (WSDL)
- Simple Object Access Protocol (SOAP)
- Extensible Markup Language (XML)
- Internet Protokolle (TCP/IP, HTTP, etc.)

```

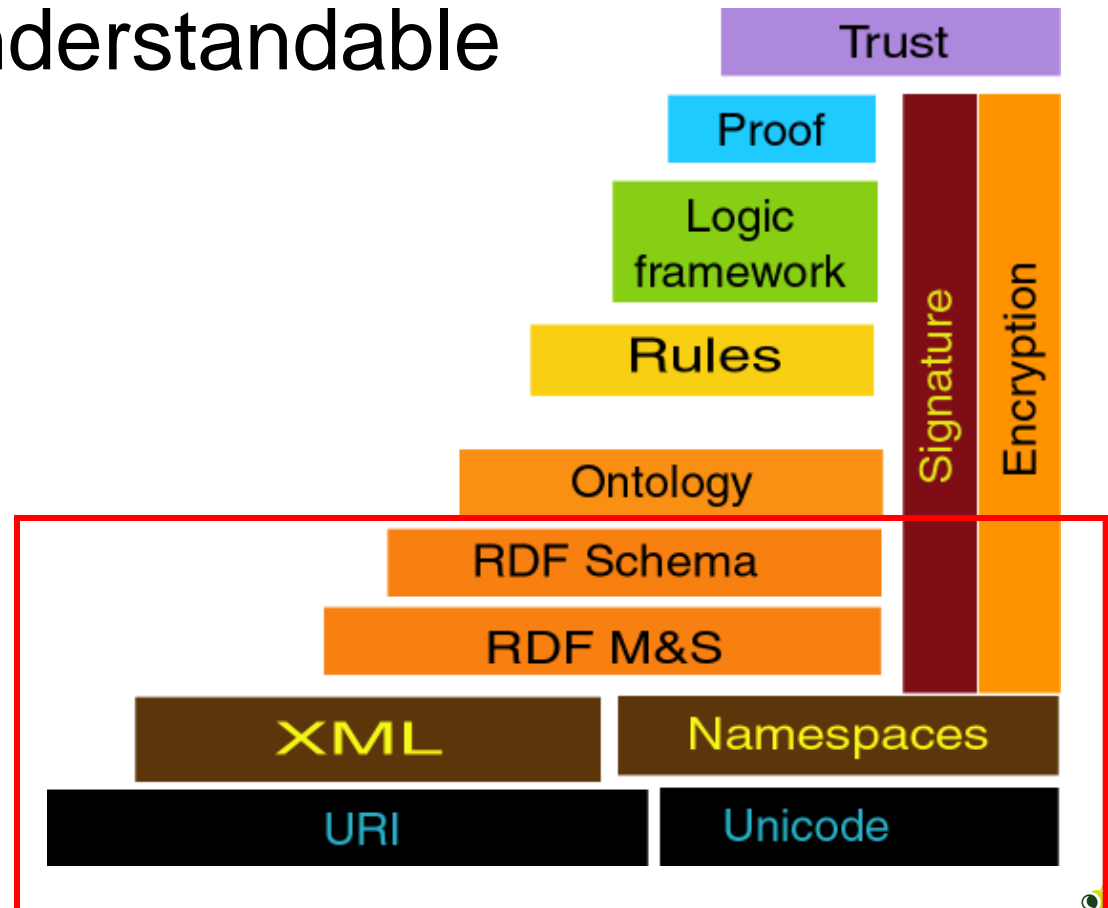
<?xml version="1.0" encoding="utf-8" ...>
<definitions ...>
  <types/>
  <message/> <portType/>
  <binding/>
  <service/>
</definitions>
    
```

- 
- **Welche Eigenschaften eines Web Service werden durch ein <service> Definitionselement in einer WSDL Datei spezifiziert? (5 P)**
    - **Dienstadresse (URL) wird angegeben - Wo befindet sich der Dienst?**
    - **Enthält einen oder mehrere Ports**
    - **Port spezifiziert die Adresse einer einzelnen Bindung.**
    - **Service ist eine Sammlung von Ports.**

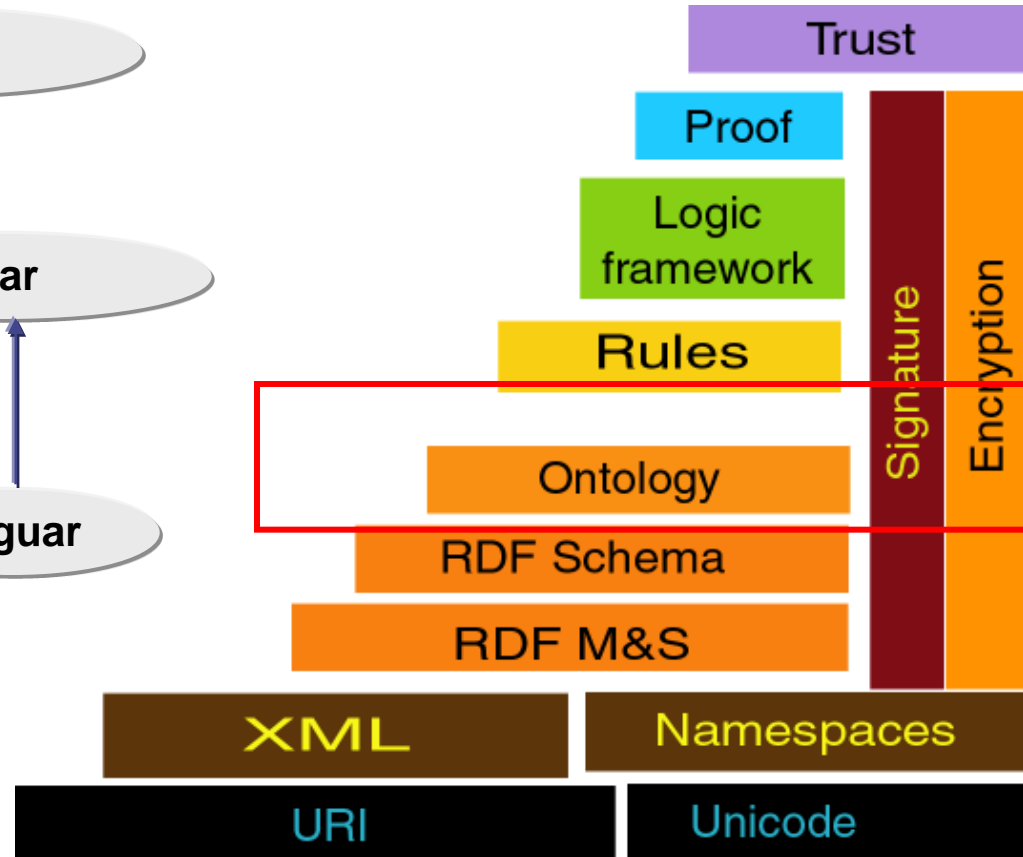
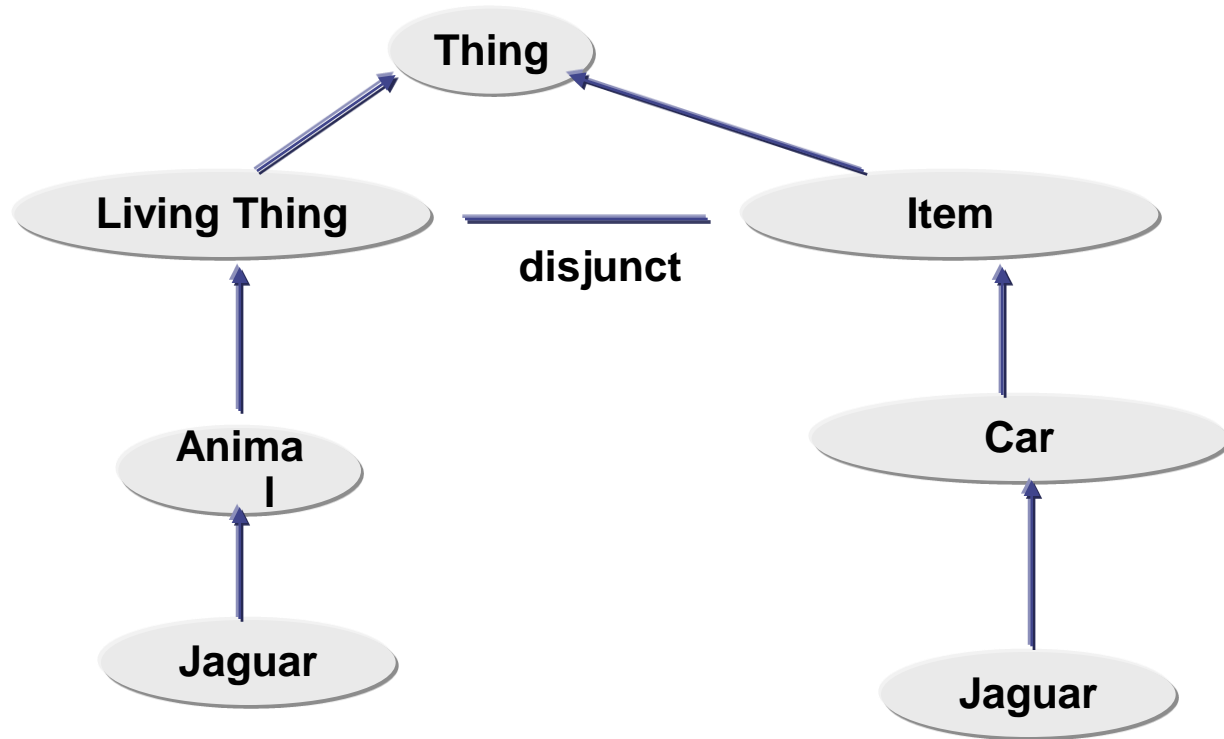
# Einheit 8: Semantic Web I

---

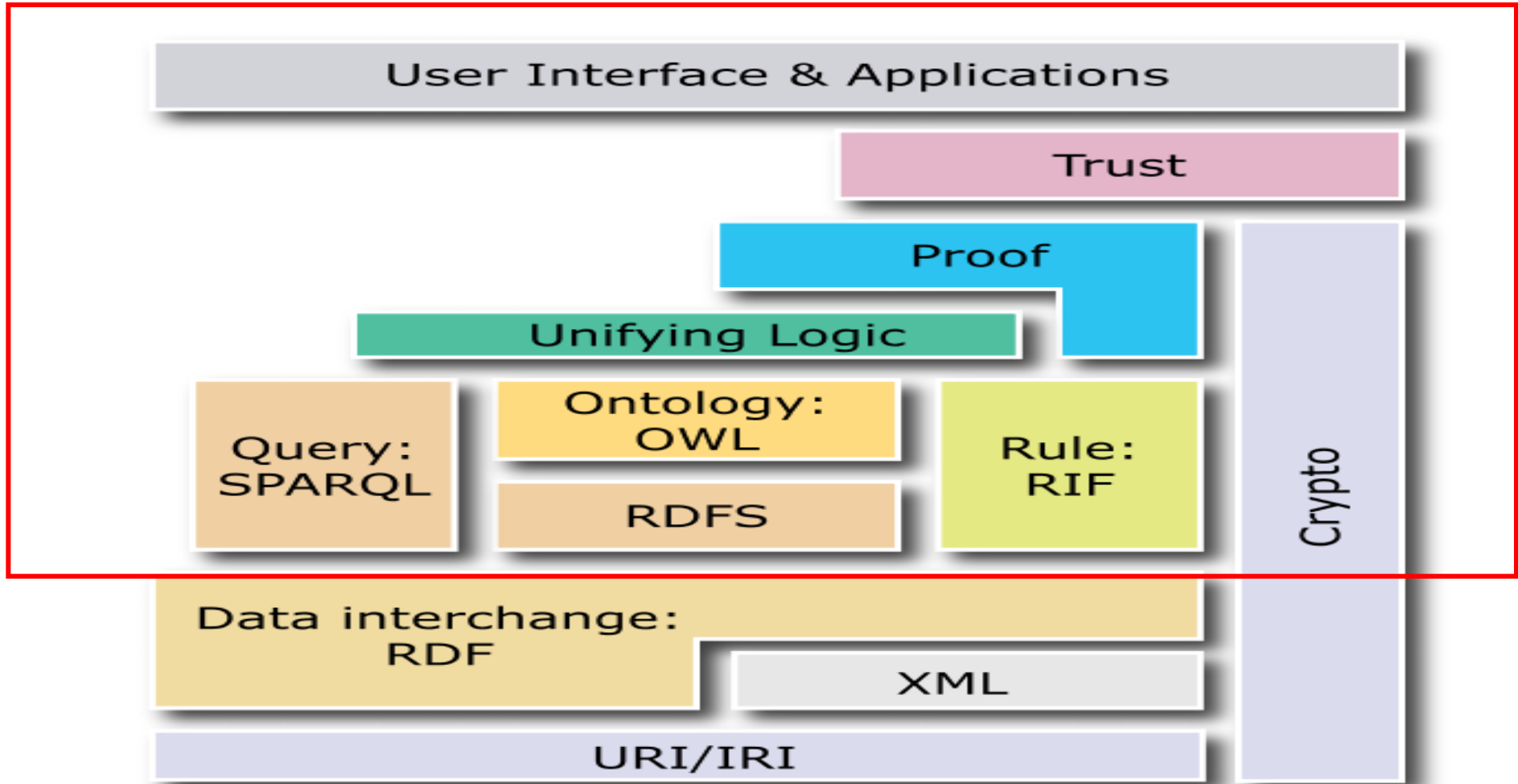
- „Make the Web understandable for machines“



# Einheit 9: Semantic Web II



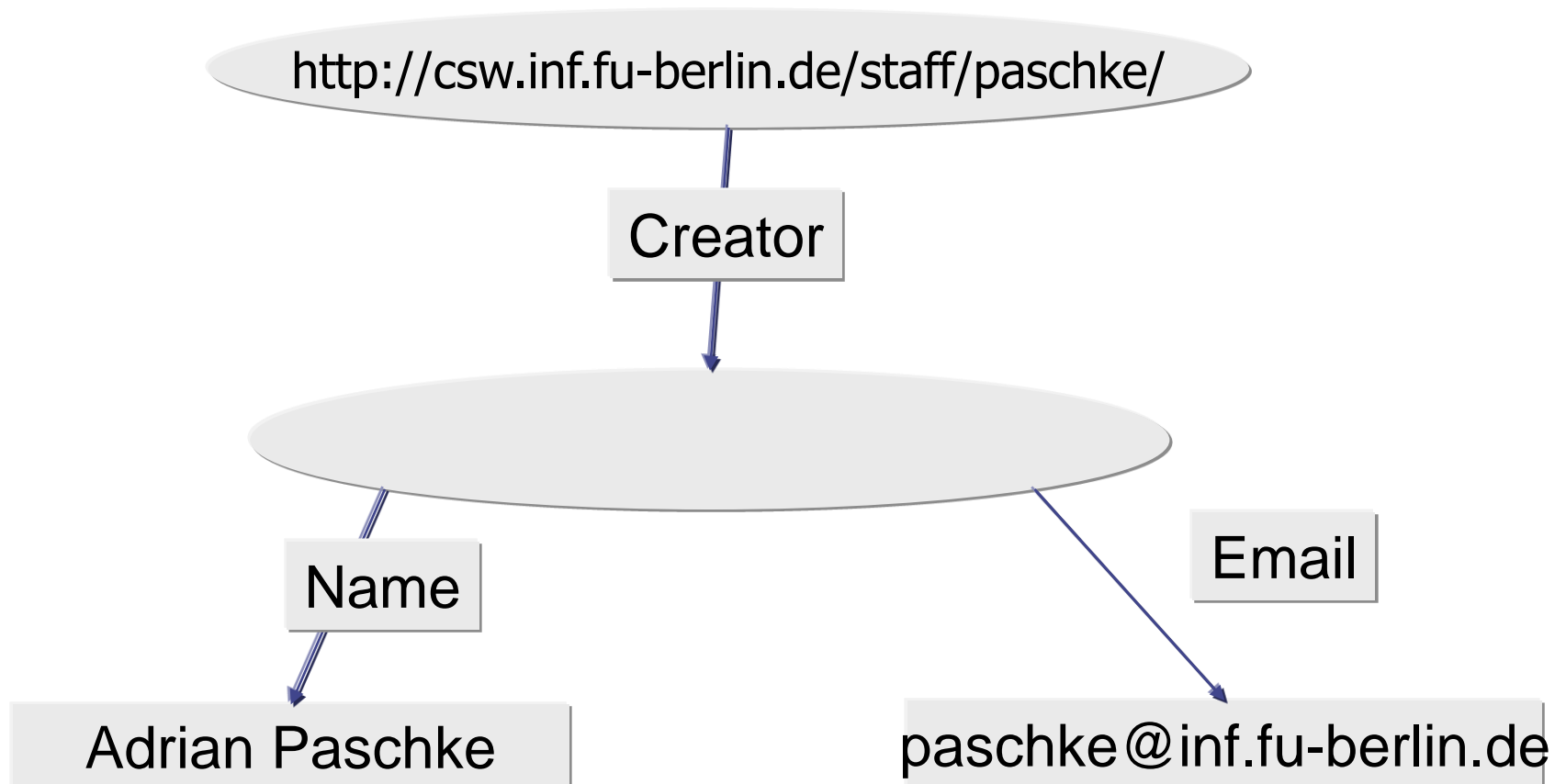
# Einheit 10: Semantic Web III



- 
- Was ist RDF und wofür wird es eingesetzt?
    - Resource Description Framework - ein Standard zur Beschreibung von Ressourcen
    - Allgemeine Metadaten und deren Struktur können mit RDF abgebildet werden
  - Nennen sie drei RDF Containertypen und beschreiben sie die Unterschiede?
    - Bag:
      - An unordered collection
    - Sequence:
      - An ordered collection
    - Alternative:
      - Unordered set of alternatives
  - Stellen Sie die Beziehung zwischen der „Webseite“ einer „Person“, die durch ihren „Namen“ und ihre „eMail Adresse“ charakterisiert wird als RDF Graph dar

# Erweitertes RDF-Diagramm

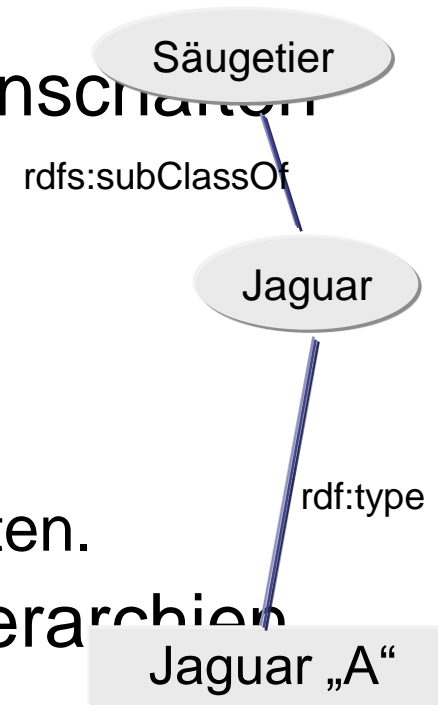
---



- 
- Was ist eine Ontologie?
    - Eine Ontologie ist eine explizite und formale Spezifikation einer Konzeptualisierung
    - Ontologien beschreiben das gemeinsame Wissen einer Domäne (Semantik):
      - Semantische Interoperabilität zwischen (verknüpften) Vokabularen
    - Typische Komponenten
      1. Klassen (Konzepte) der Domäne
      2. Eigenschaften (Rollen) der Klassen
      3. Einschränkungen (Constraints) der Eigenschaften
      4. Individuen (Ausprägungen) der Klassen → Wissensbasis
  - Wozu dient RDFS? Beschreiben sie die Ausdrucksmächtigkeit zur Repräsentation von Ontologien. Welche Zusätzlichen Möglichkeiten bietet OWL?

# RDF Schema

- Sprache zur Spezifikation von Schemata
  - Definiert Klassen und deren Eigenschaften, die in einer RDF Description mit Werten belegt werden können.
- Ein Schema legt die Klassen und Eigenschaften eines Anwendungsbereiches fest
- Definition von Beziehungen
  - zwischen Klassen,
  - zwischen Eigenschaften (*Properties*),
  - sowie zwischen Klassen und Eigenschaften.
- Ermöglicht den Aufbau von Konzepthierarchien
  - Jaguar erbt Eigenschaften von Säugetier



# Slotted Un-positional Object Oriented Representation

---

- Erklären Sie kurz die Bedeutung der einzelnen Sprachkonstrukte dieser RuleML Regel. Wozu dienen die Slots? Was bedeutet hier Stripe-Skipped Syntax?
  - Positionsunabhängige Benutzer-definierte role -> filler Paare
  - Stripe-Skipped = Weglassen der Role Tags

<Implies>

<Atom>

<Rel>spending</Rel>

<slot><Ind>spender</Ind><Var>customer</Var></slot>

<slot><Ind>amount</Ind><Ind>min 5000 euro</Ind></slot>

<slot><Ind>period</Ind><Ind>previous year</Ind></slot>

</Atom>

<Atom>

<Rel>premium</Rel>

<slot><Ind>client</Ind><Var>customer</Var></slot>

</Atom>

</Implies>

# Weitere Fragen

---

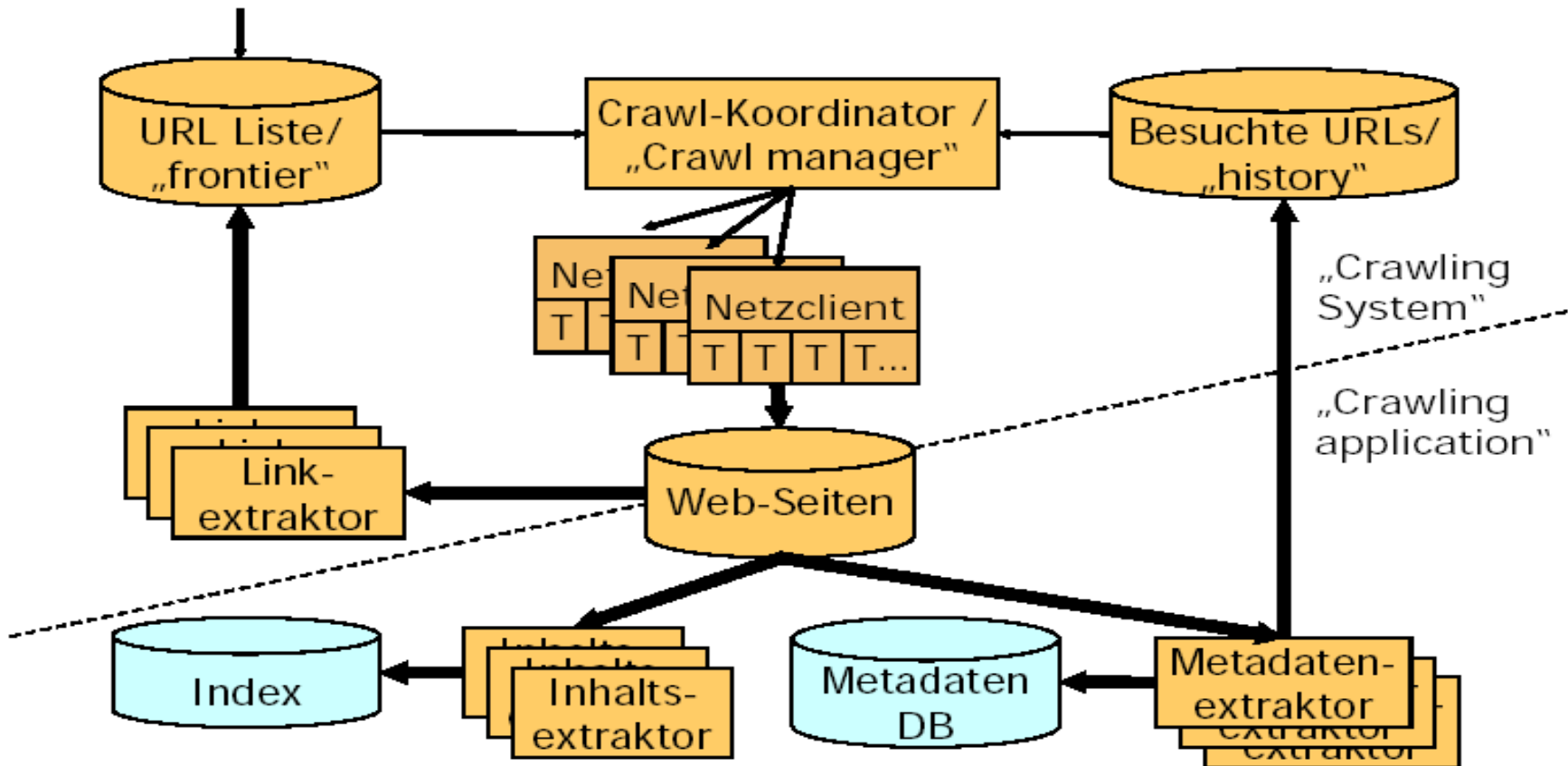
- Wie sind Aussagen in RDF formuliert?
- Warum bilden RDF Aussagen ein "Web"?
- Was sind RDF Properties?
- Was sind Blank Nodes in RDF?
- Was sind RDF Container und welche gibt es?
- Was sind RDF Collections?
- In welcher Syntax kann man RDF notieren?
- Was führt RDFS zusätzlich zu RDF ein?
- Geben Sie ein Beispiel für "Subproperties" in RDF?
- Wo entsteht in dem vorgestellten Projekt Reisewisse Mehrwert durch Semantische Technologien im Gegensatz zur normalen, datenbankbasierten Realisierung?

# Weiter Fragen

---

- Was sind Ontologien (aus Informatiksicht)?
- Was ermöglichen Ontologien?
- Welche Arten Ontologien kann man unterscheiden?
- Was ist OWL?
- Was erlaubt OWL zusätzlich zu RDF/S?
- In welchen Varianten ist OWL definiert und wie unterscheiden diese sich?
- Benennen Sie anhand der vorgestellten Anwendungen des Semantic Web, wo Mehrwert durch diese Technologien entsteht, den man nicht mehr herkömmlichen Informationssystemen haben kann.
- Beschreiben Sie die Begriffe Taxonomie, Thesaurus und Ontologien
- Was sind Metadaten?
- Was ist der Dublin Core Standard?
- Welche HTML Elemente verwendet Dublin Core?
- Welche Metadaten sind im Dublin Core Schema definiert? Nennen Sie zwei Beispiele.

# Einheit 11: Struktur und Erschließung des Web



- 
- Beschreiben Sie die Architektur eines Web Crawlers. Welche Design Optionen haben Sie – erklären sie diese?
    - Siehe Folie „Design Optionen / URL Liste“
    - Siehe Folie „Design Optionen / Link Extraktion“
    - Siehe Folie „ Entnahme/Erweiterung der URL-Liste“
    - Siehe Folie „ Crawl-Koordinator“
    - Siehe Folie „Inhaltsextraktion“
    - Siehe Folie „ Metadaten“

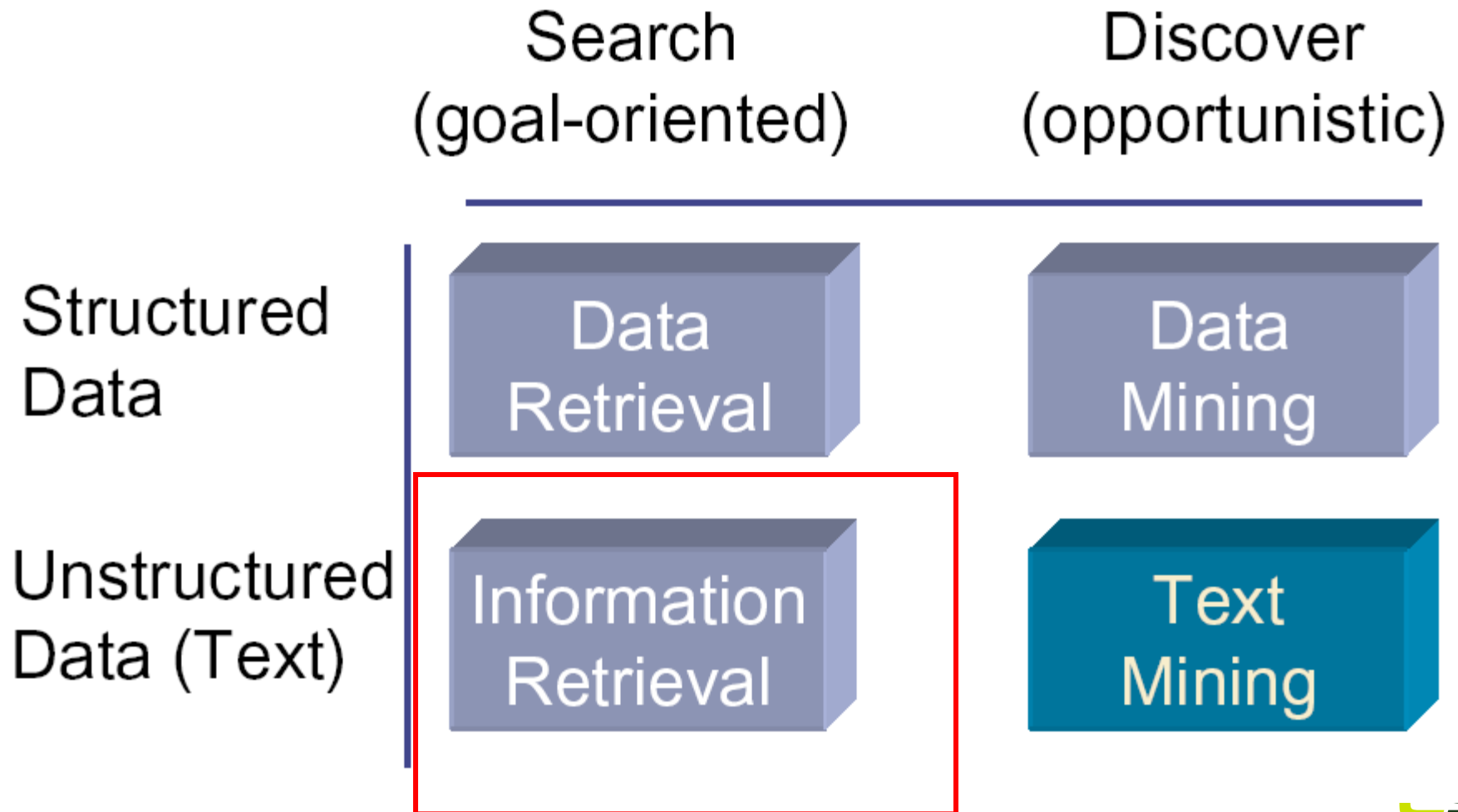
# Weitere Fragen

---

- Was macht ein Crawler?
- Wie ist der Grundlegende Crawler-Algorithmus?
- Wie ist ein Crawler aufgebaut?
- Welche möglichen Crawling-Strategien gibt es?
- Was macht man um effizient zu Crawlen?
- Wie kann ein Server das Crawler-Verhalten beeinflussen?
- Kann eine Suchmaschine auch ohne Crawler arbeiten?
- Welche Aussagen kann man über die Struktur des Web machen?
- Was ist das „Deep Web“?

# Einheit 12: Information Retrieval und Filtering

---



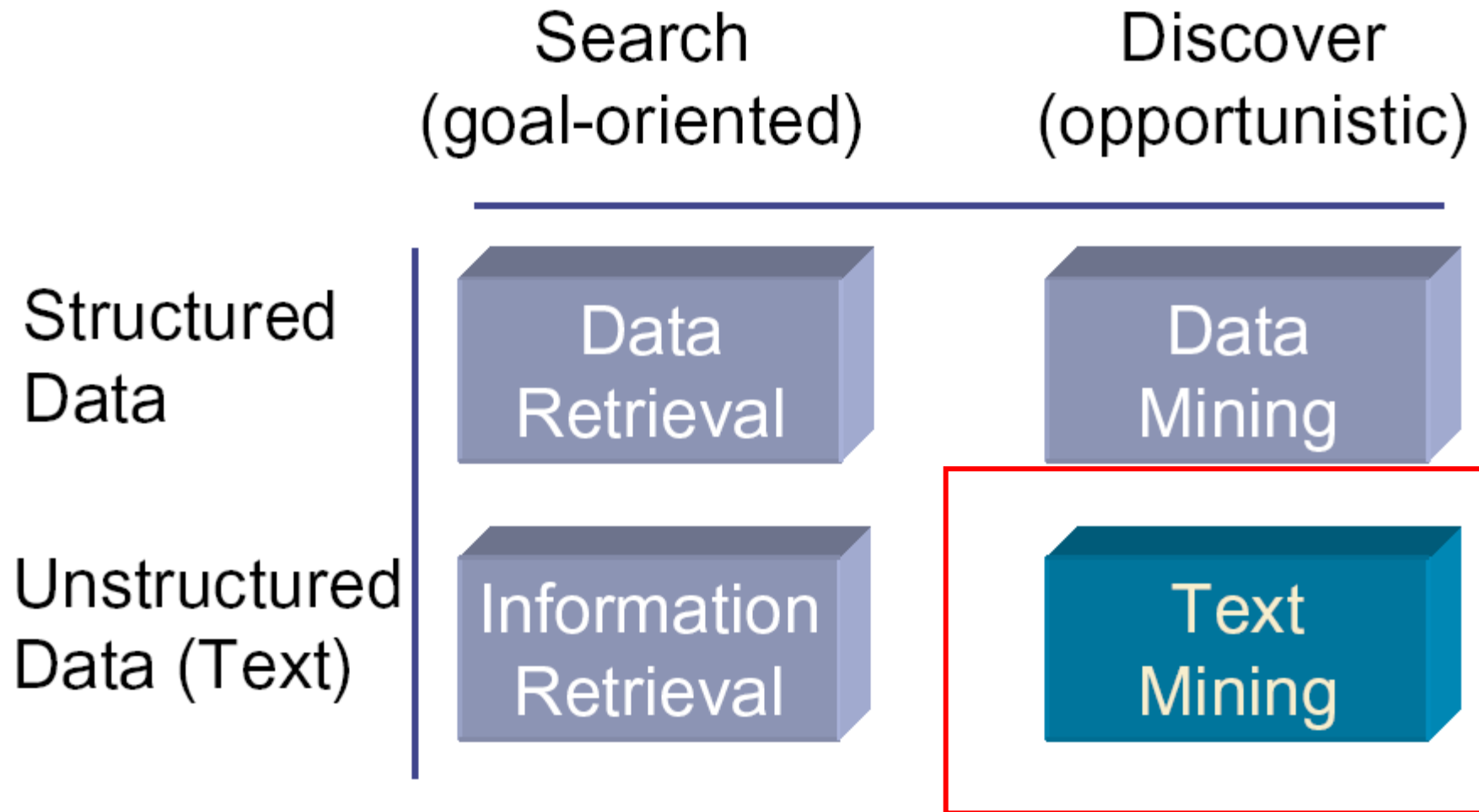
# Weitere Fragen

---

- Was macht Information Retrieval
- Wie mißt man die Güte eines IR Verfahrens?
- Triviales Verfahren für perfektes Recall?
- Triviales Verfahren für perfektes Precision?
- Was ist das Vektorraummodell?
- Wie kann man Terme in Vektoren gewichten?
- Welche wichtigen Ähnlichkeitsmasse im Vektorraummodell gibt es?
- Wie bereitet man Dokumente vor?
- Wie sieht der Porter-Stemming Algorithmus aus?
- Ist Porter-Stemming universell für alle Dokumente im Netz?
- Wie funktioniert das Collaborative Indexing im Harvest System?
- Wie geht man beim Indexing von multimedialen Informationen vor? Illustrieren Sie anhand des Indexings von Audio-Daten.
- Was ist Collaborative Filtering?
- Auf welcher Grundannahme beruht es?
- Kann man alles, was in einem Collaborative Filtering System enthalten ist auch in einem herkömmlichen Index halten? Warum?

# Einheit 13: Information Discovery - Text Mining

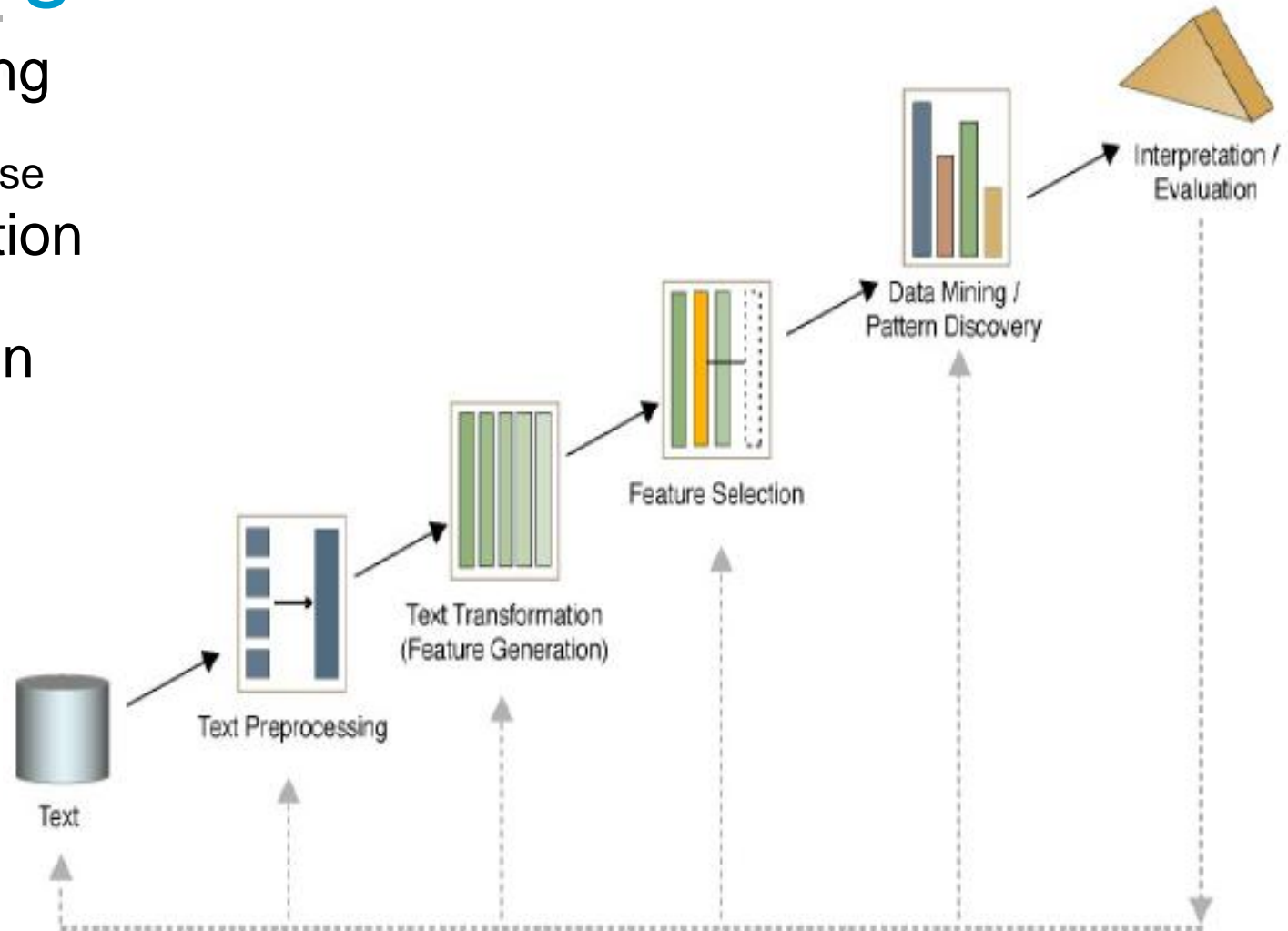
---



- 
- Beschreiben Sie den Text Mining Prozess und beschreiben Sie zu jedem Schritt kurz ein mögliches Verfahren.
    - Verfahren siehe Text und Web Mining Vorlesung

# Text Mining Prozess

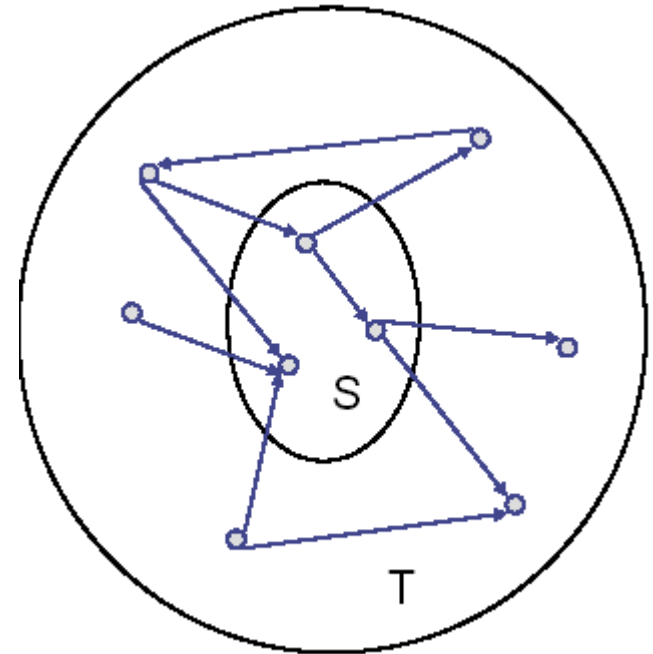
- Text Preprocessing
  - Syntaktische und Semantische Analyse
- Features Generation
  - Menge an Worten
- Features Selection
  - Einfaches zählen
  - Statistik
- Text/Data Mining
  - *Classification* von Dokumenten
  - *Clustering* von Dokumenten
- Analyse der Ergebnisse



# Einheit 14: Web Mining

---

- Web Mining
- Page Rank
- HITS
- ..



- 
- Grenzen Sie voneinander ab, Web Structure Mining, Web Usage Mining, Web Content Mining
    - Web Content Mining
      - Entdeckung (discovery) der nützlichen Informationen von Webinhalten, inklusive text, image, audio, video, etc.
      - Finden von Webressourcen
      - Dokumentkategorisierung und Clustering
      - Information Extraktion von Webseiten
    - Web Usage Mining
      - Fokussiert die Analyse von Logs wie Search Logs, User Activity Logs
      - Finden von interessanten Patterns der Webnutzung
    - Web Structure Mining
      - Studiert das Model, welches der Linkstruktur des Web unterliegt;
      - Normalerweise auf Basis der In- und Out-Link Informationen einer Webseite

# Weitere Fragen

---

- Was ist die Grundidee von Verfahren, die die Struktur des Web verwenden?
- Wie funktioniert PageRank?
- Was braucht man um PageRank anzuwenden?
- Wie funktioniert HITS?
- Was braucht man im Gegensatz zu Pagerank für HITS nicht?
- Was sind Metasuchmaschinen?
- Warum haben Metasuchmaschinen überhaupt einen Vorteil?
- Was sind die Schritte und Probleme bei Metasuchmaschinen?

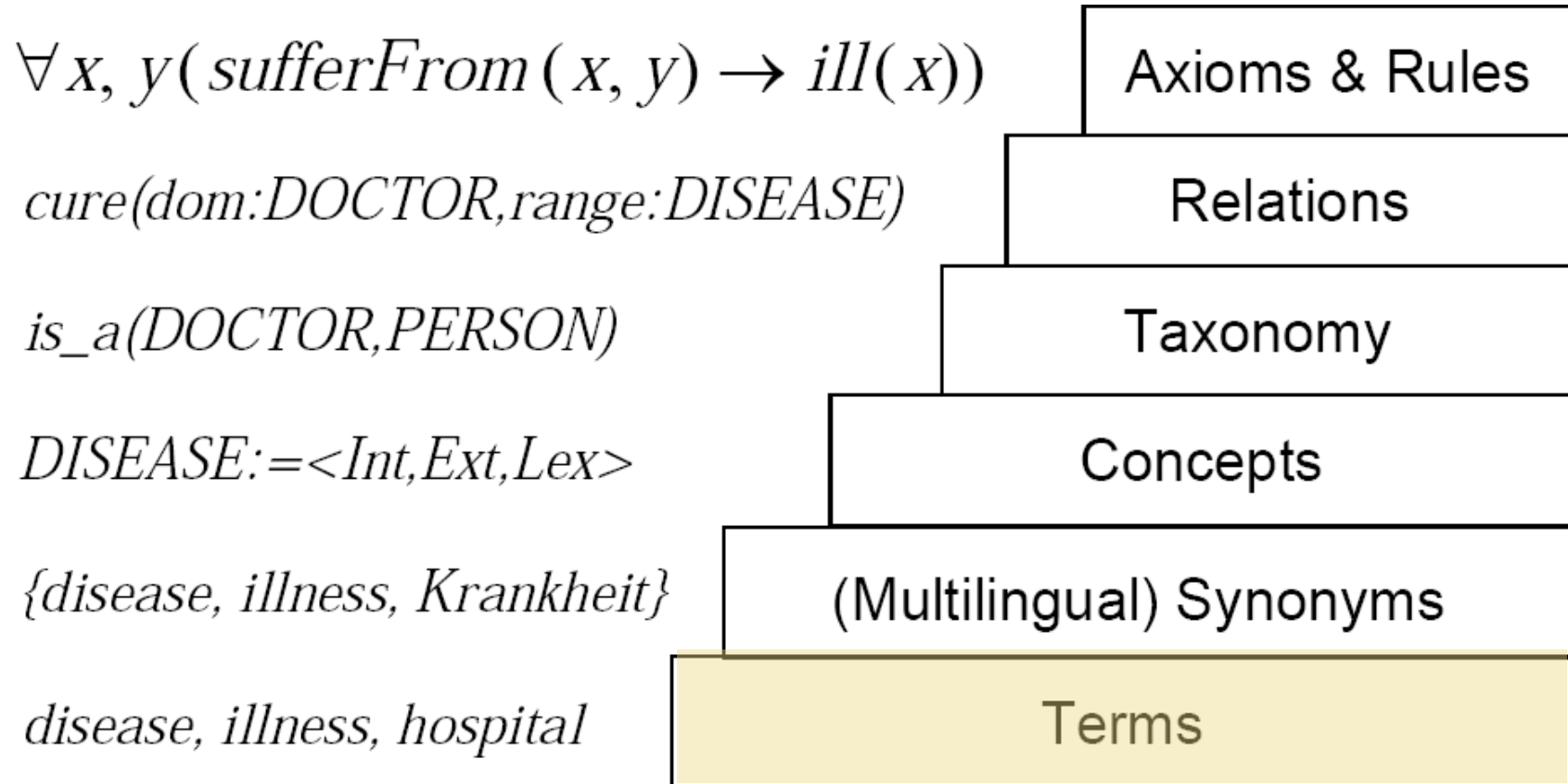
# Weitere Fragen

---

- Was ist Web Usage Mining
- Was steht in Web-Logfiles drin?
- Was ist das Problem der Logfile-Daten?
- Welche Maße für Nutzungsdaten gibt es?
- Was ist eine Session?
- Wie sieht ein Messmechanismus durch Dritte aus?
- Was sind typische Fragen, die durch Web-Usage Mining beantwortet werden sollen?

# Einheit 15: Ontology Engineering

---



- 
- Beschreiben Sie das TFIDF Verfahren. Wozu kann es eingesetzt werden? Wo liegen die Nachteile des Verfahrens?

most popular weighting schema

(normalized word frequency)

$$tfidf(w) = tf \cdot \log\left(\frac{N}{df(w)}\right)$$

The word is more important if it appears several times in a target document

The word is more important if it appears in less documents

*tf(w)* term frequency (number of word occurrences in a document)

*df(w)* document frequency (number of documents containing the word)

*N* number of all documents

*tfidf(w)* relative importance of the word in the document

# ■ Erklären Sie Probleme und Verfahren der Entity Recognition

## Lexicons

... Buffalo Company in Denver, CO.

↑ member?

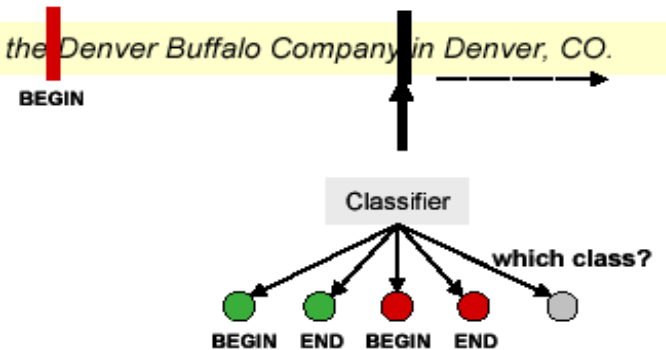
Alabama  
....  
Denver  
....  
Wyoming

## Regular Expressions

... Denver Buffalo Company in Denver, CO. ... at the Denver Buffalo Company in Denver, CO.

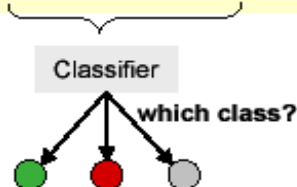
$[A-Z][a-z]+( [A-Z][a-z]+)^* \text{Company}$

## Boundary Classifier



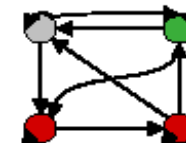
## Sliding Window Classifier

He will be playing at the Denver Buffalo Company in Denver, CO.



## Finite State Machines

Abraham Lincoln was born in Kentucky.



Most likely state sequence?

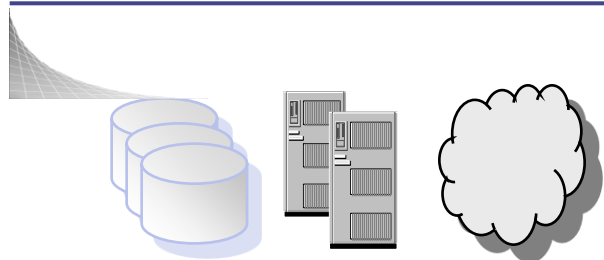
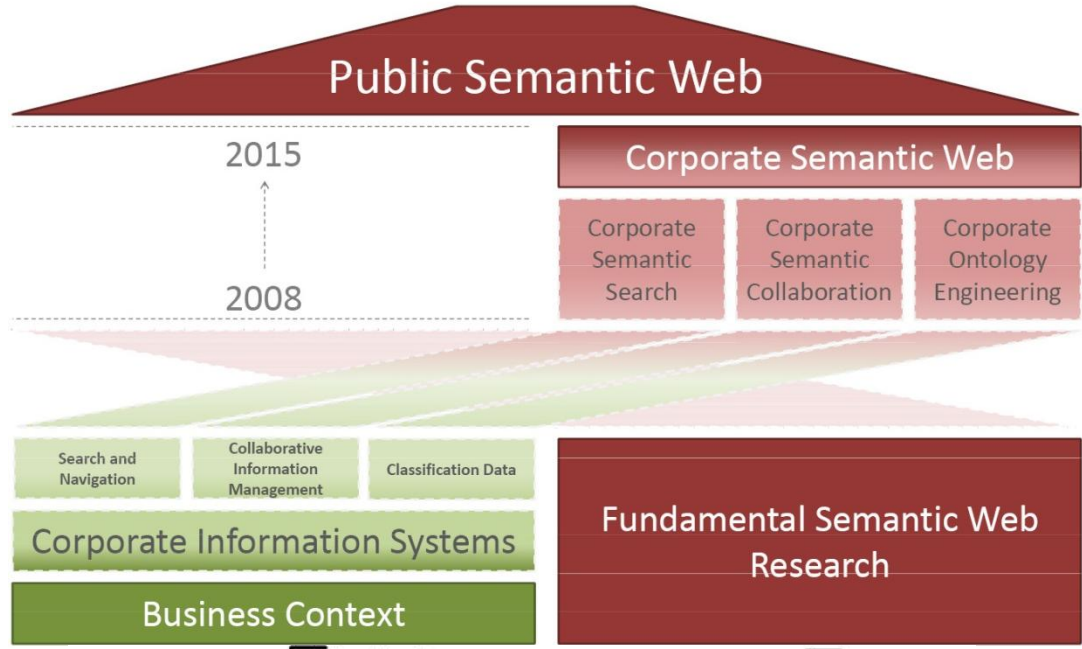
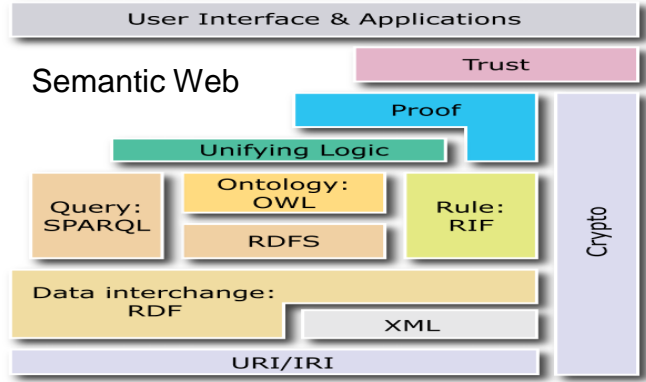
- 
- Erläutern Sie die vier Phasen des Ontology Engineerings
    - Siehe Folien zu
      - Anforderungen und Analyse
      - Design und Implementierung
      - Testen und Validierung
      - Wartung und Pflege

# Weitere Fragen

---

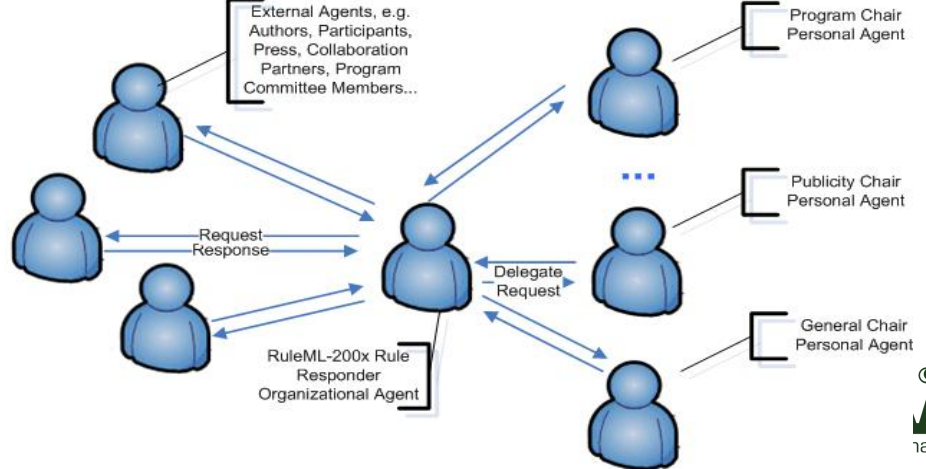
- Beschreiben Sie drei Methoden der Ontologieentwicklung
  - Siehe Folien zu
    - Formale Methoden
    - Benutzer-fokusierte Methoden
    - Kollaborative Methoden
    - Agile Methoden
    - Automatisierte Lernmethoden
- Beschreiben Sie COLM, SCRUM, OntoCom
- Briefly explain an approach how to extract relationships from text. Illustrate your approach on extracting triples
- What are the advantages and disadvantages of your approach?
- Erklären Sie den Ontology Learning Layer Cake und geben Sie zu jeder Schicht ein Beispielfahrer/methode an

# Einheit 16: Corporate Semantic Web und Ubiquitous Pragmatic Web



Storage with virt. Images

Blades, Cluster Network, Load Balancer, Switches



- 
- Grenzen Sie das Corporate Semantic Web vom Public Semantic Web ab. Wo liegen die wesentlichen Unterschiede?
  - Beschreiben Sie die Idee Personalisierter Ontologie-basierter Suche. Wann macht Personalisierung Sinn? Wo liegt der Nutzen? Erklären Sie wie diese mit Hilfe von Semantische Ähnlichkeit zwischen Objekten realisiert werden könnte und geben Sie ein selbstgewähltes Beispiel
    - Lösung siehe Folien zu SemF Framework

- 
- Geben Sie Beispiele für Kollaborationswerkzeuge
    - Groupwaresysteme
      - Kommunikationssysteme
      - Kollaborationssysteme
      - Koordinationssysteme
    - Social Software
      - Wikis
      - Weblogs
      - Social Bookmarking
    - Inhaltsorientierte Systeme
      - Kollaborative (Web) Dokumentenmanagementsysteme / Contentmanagementsysteme
      - Learningcontentmanagementsysteme
      - Portalsysteme
      - Lernmanagementsysteme
    - Systeme der künstlichen Intelligenz
      - Kollaborative Agentensysteme
      - Regel-basierte Workflow / intelligente Event Processing Systeme

- 
- Was versteht man unter CSCW?
    - CSCW beschäftigt sich mit dem Verstehen sozialer Interaktion sowie der Gestaltung, Implementierung und Evaluierung von technischen **Systemen zur Unterstützung sozialer Interaktion**

# Welche Aspekte der Suche können unterschieden werden?

Form der Suchanfrage

---

Qualität der Suchergebnisse

User Involvement

Suchziel/ Suchobjekt

Suchbestände: Daten, Texte,  
etc.

Visualisierungsaspekte

---

Worin liegt der Unterschied zwischen Web 2.0 und Web 3.0 (Semantic Web + Web 2.0 )? Geben Sie Beispiele.

	Web 2.0	Web 3.0
Tagging	<ul style="list-style-type: none"> <li>▪ annotieren mit <b>ambiguen</b> Stichwörtern</li> <li>▪ Singular/Plural-Problem</li> <li>▪ Synonyme</li> <li>▪ Keinerlei Intelligenz</li> </ul>	<ul style="list-style-type: none"> <li>▪ annotieren mit <b>eindeutigen</b> Stichwörtern</li> <li>▪ Inferenz (Tag „Hund“ folgert Tag „Tier“)</li> </ul>
Rekombination von Daten verschiedener Quellen	<ul style="list-style-type: none"> <li>▪ Mesh-Ups vorab von Hand programmiert</li> </ul>	<ul style="list-style-type: none"> <li>▪ Spontan durch End-Nutzer (siehe Piggybank)</li> </ul>
Suche	<ul style="list-style-type: none"> <li>▪ Stichwortsuche oder Tag-Suche <i>findet</i> Dokumente</li> </ul>	<ul style="list-style-type: none"> <li>▪ Semantische Suche</li> </ul>

# Hinweise zur Klausur

---

- Klausur 17.2.
- 14 Uhr in SS05
  
- 90 Minuten
  
- Anmeldung Campus Management + KVV !

**Viel Erfolg !**

# Feedback ?

---

- Feedback zur Vorlesung + Übung
  
- Hinweis
  - Lehreevaluation
  - <http://lehreevaluation.fu-berlin.de/>

# Studenten-Arbeiten zu vergeben!

---

- Diplom-/Masterarbeiten
- Werkverträge

Arbeiten überwiegend in Zusammenarbeit mit und bei Industriepartnern z.B. FIZ Chemie, Bundesverband Deutscher Industrie, Semtation, EsPresto, Mister Wong, HPI 3D Geo, ...

- Bei Interesse bitte bei mir melden
  - [paschke@inf.fu-berlin.de](mailto:paschke@inf.fu-berlin.de)