



## Netzbasierte Informationssysteme **Rückschau**

Prof. Dr.-Ing. Robert Tolksdorf  
Freie Universität Berlin  
Institut für Informatik  
Netzbasierte Informationssysteme  
mailto: [tolk@inf.fu-berlin.de](mailto:tolk@inf.fu-berlin.de)  
<http://www.robert-tolksdorf.de>

- „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.“

Information Retrieval  
**Pagerank**  
**Suchmaschinen**  
*Web caches*  
**Mehrsprachlichkeit**  
*Deep Web*  
**Web**  
**Semantic Web**  
**Ganz große Server**

# Veranstaltungsinhalt

---

- 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
- Die Veranstaltung befasst sich mit drei Bereichen
  - *Informationsquellen*: Wie sind sie repräsentiert (HTML/XML), wie findet man sie (Crawling, Deep Web), wie kann man auf sie zugreifen (Internet Protokolle)
  - *Informationsintegration*: Repräsentation von aufbereiteten Informationen (DOM, Indexing), Anfragen (Indexing, Ranking), Ermittlung semantischer Informationen (Metadaten), Wissensverarbeitung (Semantic Web). Systemarchitekturen von Suchmaschinen, Metasuchmaschinen.
  - *Informationsauslieferung*: Darstellungserzeugung, Internationalisierung



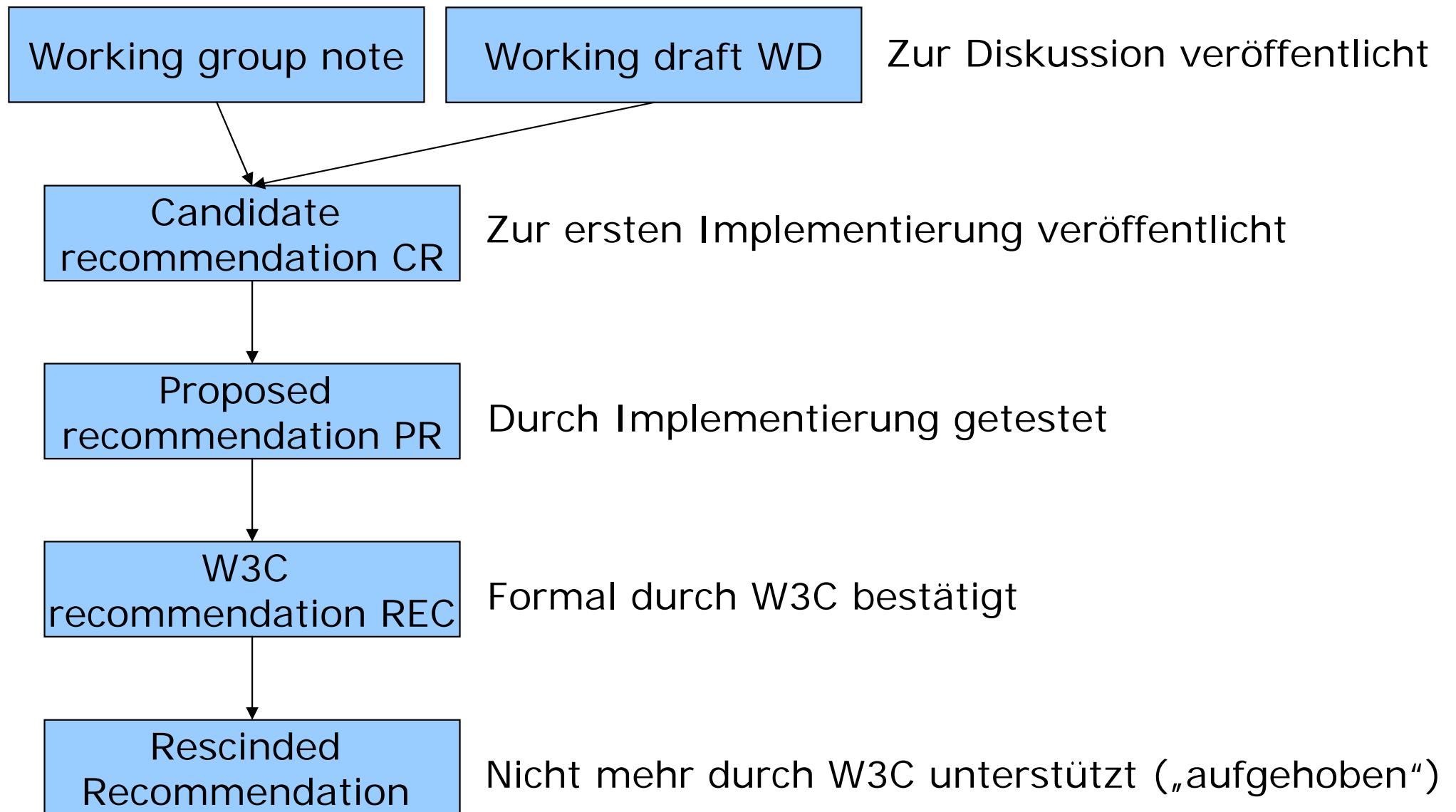
## Architektur des Web

# Die Standardisierung des Web

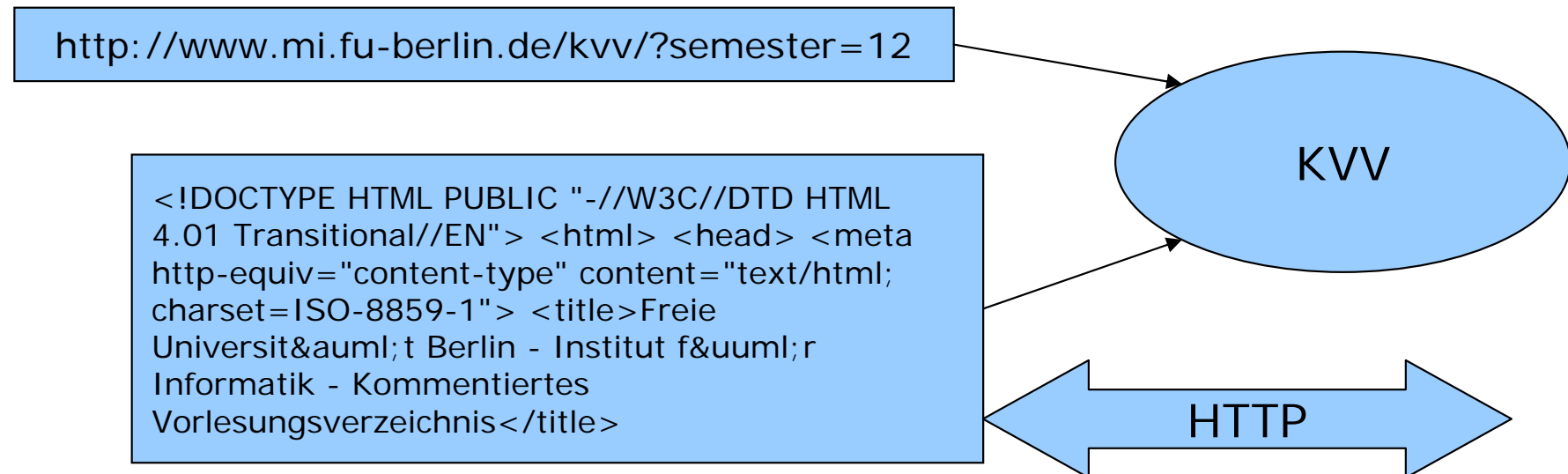
---

- Das World Wide Web Consortium W3C treibt die Entwicklung des Web voran:
  - „W3C's mission is: To lead the World Wide Web to its full potential by developing protocols and guidelines that ensure long-term growth for the Web“  
[<http://www.w3.org/Consortium/>]
- Prinzipien
  - Konsens
  - Offenheit
  - Lizenzfreiheit
- Ca. 400 Mitglieder

# Prozess der Standardentwicklung



- Drei Grundlagen der Web Architektur:
  - *Identifikation* durch Uniform Resource Identifiers URI
  - *Interaktion* durch Protokolle wie HTTP
  - *Datenformate* wie HTML oder XML
- (Informations-)Ressourcen werden identifiziert und durch Interaktion sind sie durch Übermittlung einer Nachricht in einer Repräsentation zugänglich



# URI, URL, URN

---

- Uniform Resource Identifier URI: „A Uniform Resource Identifier (URI) is a compact string of characters for identifying an abstract or physical resource“ [RFC 2396]
- Lediglich Syntax: *schema:schemaspezifischer\_Teil*
- URI-Schema typisiert URIs:
  - <ftp://ftp.is.co.za/rfc/rfc1808.txt>
  - <gopher://spinaltap.micro.umn.edu/00/Weather/Los%20Angeles>
  - <http://www.math.uio.no/faq/compression-faq/part1.html#sec1>
  - <mailto:mduerst@ifi.unizh.ch>
  - <news:comp.infosystems.www.servers.unix>
  - <telnet://melvyl.ucop.edu/>
  - <urn:isbn:n-nn-nnnnnn-n>
  - <fon:+49-838-0>

# URI, URL, URN

- Uniform Resource Locator URL: „[...]a compact string representation for a resource available via the Internet.“ [RFC 1738]
  - Ist ein URI, dessen Schema auf die Zugreifbarkeit der Ressource im Netz hinweist
  - z.B. `ftp://ftp.is.co.za/rfc/rfc1808.txt`
- Uniform Resource Name URN: „[...] intended to serve as persistent, location-independent, resource identifiers and are designed to make it easy to map other namespaces“ [RFC 2141]
  - Ist eher URI, der Eigenschaft der Ressource beschreibt
  - `urn:isbn:n-nn-nnnnnn-n`
  - URN-Namensraum strukturiert URNs (isbn,....)

- Prinzip: Global Identifiers
  - URIs haben sind global eindeutige Bezeichner
  - Damit können sie auch global verwendet werden
  - Netzwerkeffekt: Je mehr das URI Schema genutzt wird umso wertvoller wird es
- Gute Praxis: Identify with URIs
  - Ressourcen sollte durch URIs zu bezeichnen sein
  - Damit erhöhen Anbieter den Wert des Web

- Beschränkung: URIs Identify a Single Resource
  - Unterschiedliche Ressourcen sollen mit unterschiedlichen URIs bezeichnet werden
- Gute Praxis: Avoiding URI aliases
  - Nicht unterschiedliche URIs für die gleiche Ressource verwenden
- Gute Praxis: Consistent URI usage
  - Erhaltene URIs identisch weiterverwenden

# Hypertext Transfer Protocol

- Aufgabe:  
Transfer von Informationen zwischen Web-Servern und Clients
- Port:  
80 ist für HTTP reserviert
- Transportprotokoll:  
TCP
- Protokoll:  
R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach und T. Berners-Lee. *Hypertext Transfer Protocol - HTTP/1.1*. RFC 2616,  
<http://www.ietf.org/rfc/rfc2616.txt>

# MIME Typen

---

- MIME-Typen werden von der Internet Corporation for Assigned Names and Numbers IANA verwaltet
- <http://www.iana.org/assignments/media-types/>
- Verarbeiten eines bestimmten Medientyps nach Erhalt:
  - Teil der Anwendung (siehe auch: `javax.mail.internet.MimeMessage`)
  - eventuell Unterstützung durch Betriebssystem
- Ermittlung des MIME-Typs für eine Datei:
  - Ableitung aus Endung (`javax.activation.MimetypesFileTypeMap`)
  - Ableitung aus Inhalt der Datei

- Gute Praxis: Reuse representation formats
  - Protokolle sollten MIME als Typisierung ausgetauschter Datenströme verwenden
- Gute Praxis: Available representation
  - Unter einer URI sollten sich Repräsentationen der bezeichneten Ressource abrufbar sein

# Software Architektur des Web

---

- Was ist die Software-Architektur des Web?
- REST: Representational State Transfer
  - Minimierung von
    - Netzwerklatenz und
    - Netzwerkverkehr
  - Maximierung von
    - Unabhängigkeit und
    - Skalierbarkeitvon Komponenten

# Wahl des Architectural Style

- Code on demand
  - Funktionalität des Klienten kann durch nachgeladenen Code erweitert werden
  - Nachteil: Visibility sinkt

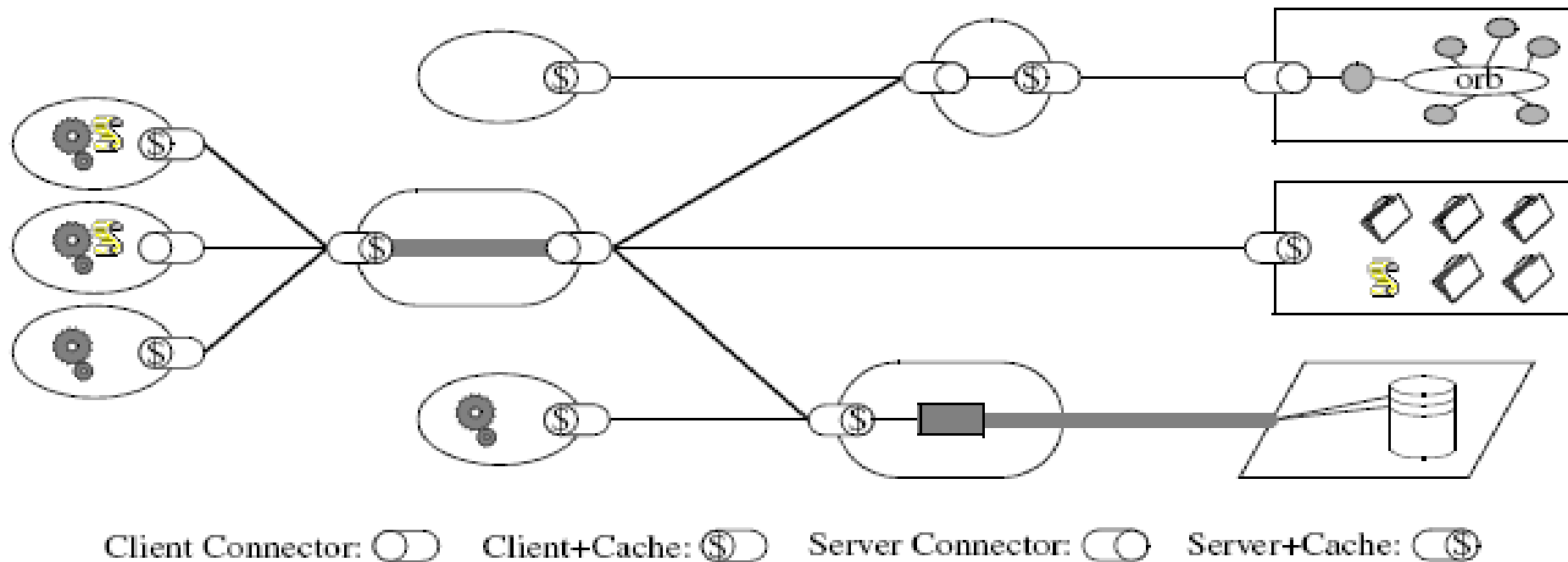
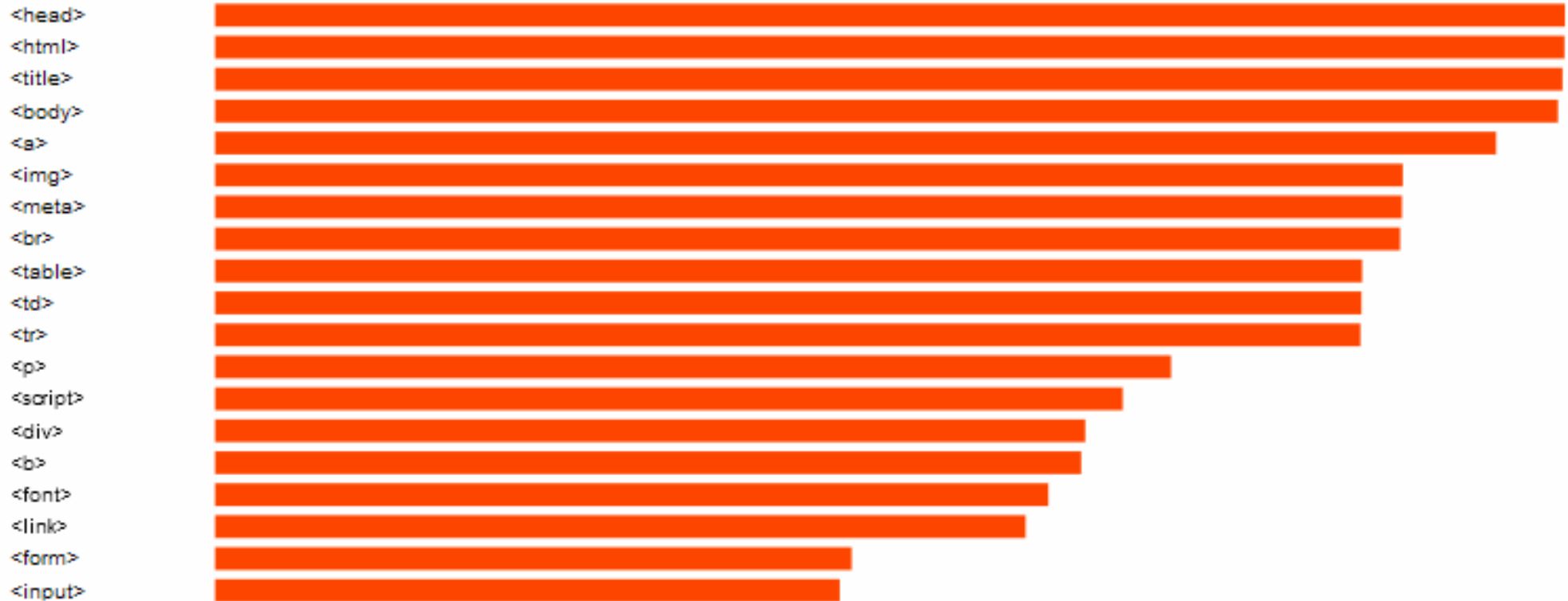


Figure 5-8. REST

# Verwendung von HTML Elementen nach Google Erhebung [<http://code.google.com/webstats>]

What are those elements? Well, the nineteen elements used on the most pages are:



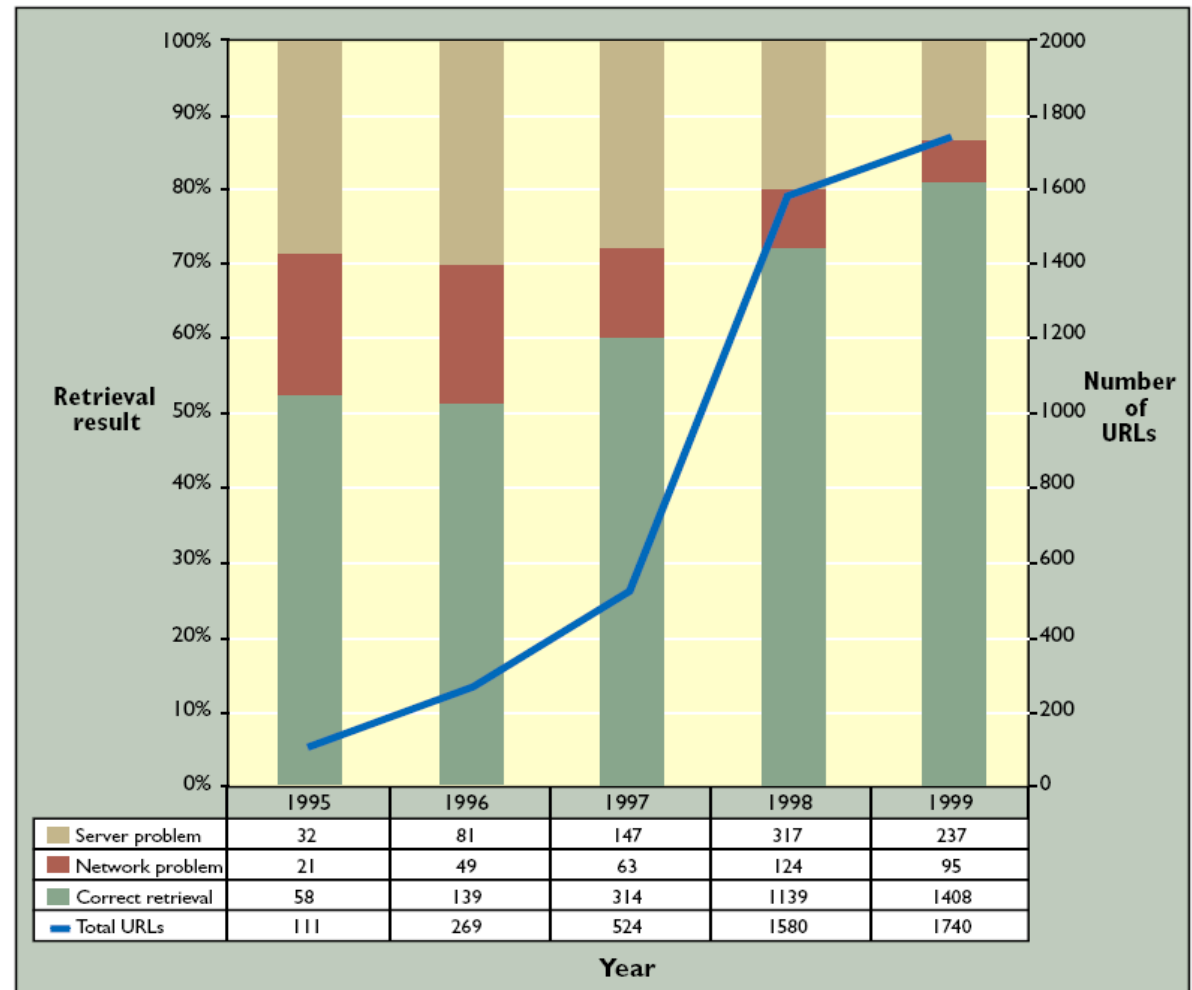
# Lebensdauer von Web-Referenzen

- Wie lange sind URLs in der Regel nutzbar?
- Studie  
Diomidis Spinellis. The decay and failures of web references. Communications of the ACM, Volume 46, Number 1 (2003), Pages 71-77
- Untersucht
  - URL Referenzen in Artikeln der Zeitschriften
    - Communications of the ACM
    - IEEE Computer
- Befürchtung:  
Zunehmende Verwendung von URLs in der wissenschaftlichen Literatur ist bedrohlich für die wissenschaftliche Qualität wenn URLs nicht mehr auflösbar sind.



# Lebensdauer von Web-Referenzen

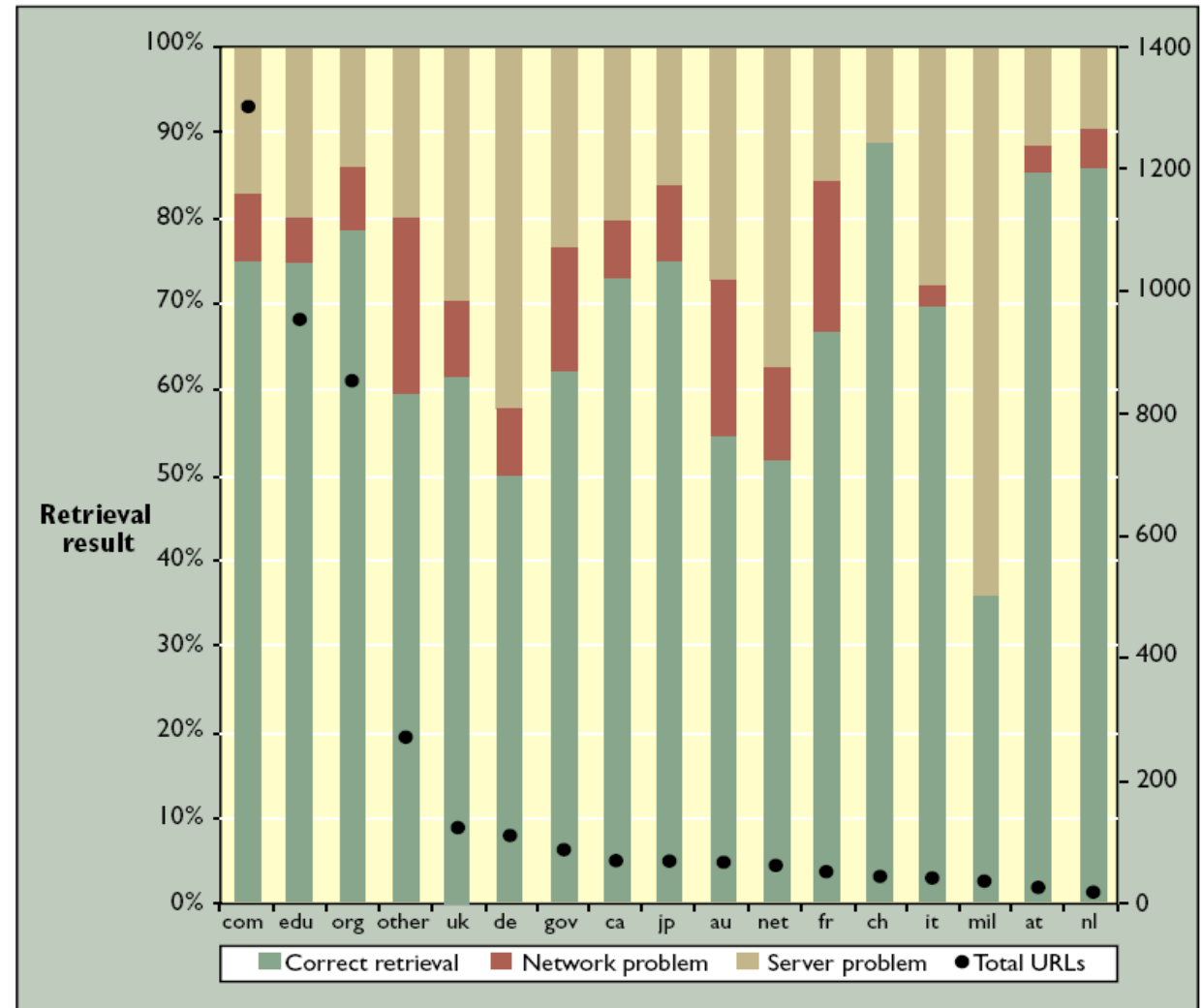
- Halbwertszeit einer URL: 5 Jahre
- Ca. 10% URL-Verfall pro Jahr
  - Vorlaufverzögerung bis zum Erscheinen des Artikels
  - Stabilisierung auf gut gepflegten Servern
- Wiederholung der Untersuchung zeigt
  - Referenzen in neueren Artikeln sind etwas stabiler
    - Bessere Auswahl der Quellen
    - Bessere Wartung der Server



# Lebensdauer von Web-Referenzen

## Weitere Erkenntnisse

- Tiefe Referenzen (lange Pfade) haben kürzere Erreichbarkeit als kurze Referenzen
- Referenzen auf Dateien haben schlechtere Erreichbarkeit als Referenzen auf Verzeichnisse
- Inhalte auf anderen Top-Level-Domains als .edu haben ähnliche Erreichbarkeit



# Anforderungen

- SGML hat Defizit:
  - Komplexität
- XML versucht die Defizite von HTML und SGML auszugleichen und so die Anforderungen Skalierbarkeit und Einfachheit zu erfüllen
  - Definition eigener Elemente und Attribute
  - Beliebig verschachtelbare Dokumentenstruktur
  - Durch Offenlegung der Grammatik Typisierung möglich
  - Erweiterte Link-Strukturen möglich
  - Style-Sheets Anbindung
- Siehe auch
  - Jon Bosak. XML, Java, and the Future of the Web. World Wide Web Journal. Volume 2, number 4, pages 219-227, 1997.  
[http://citeseer.ist.psu.edu/cache/papers/cs/1263/http:zSzzSzwww.cis.udel.eduzSz~deckerzSzcourseszSz889czSzxmlapps.pdf/bosak97xml.pdf](http://citeseer.ist.psu.edu/cache/papers/cs/1263/http%3A%2F%2FzSzzSzwww.cis.udel.edu%2FdeckerzSzcourseszSz889czSzxmlapps.pdf/bosak97xml.pdf)
  - Jon Bosak. The Birth of XML. Sun Developer Network.  
[http://java.sun.com/xml/birth\\_of\\_xml.html](http://java.sun.com/xml/birth_of_xml.html)



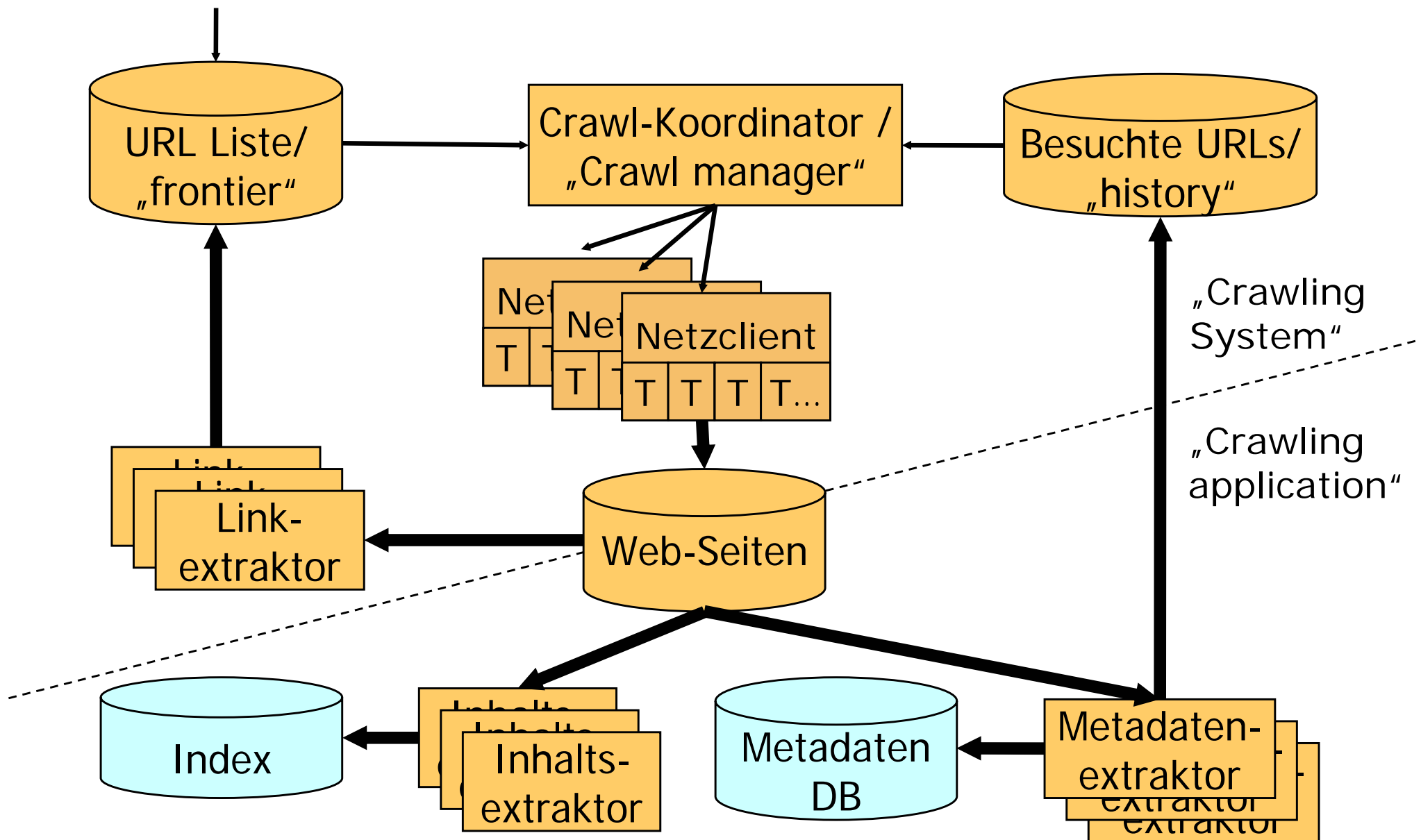
## Crawling und Web Struktur

# Crawling Algorithmus

1. URL-Liste mit unbesuchten URLs initial füllen
2. Nehme URL aus Liste und teste
  - schon besucht?
  - passender Medientyp (html/ps/pdf/gif/...)?
  - andere Kriterien (Ort/...)?
3. hole Seite
4. extrahiere URLs und schreibe sie in URL-Liste
5. extrahiere und indexiere Seiteninhalt
6. extrahiere und speichere Metadaten
7. gehe nach 2

„Crawling loop“

# Einfache Architektur



# Crawler Last

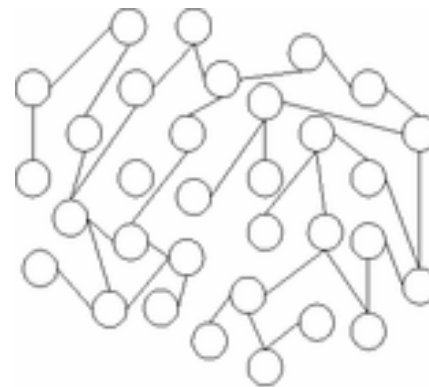
---

- Crawler erzeugen Last beim Server
  - Verarbeitung der Anfragen
  - Auslieferung der Ergebnisse
- “Freundliche” Crawler versuchen das zu vermeiden
  - Keine fortlaufenden Anfragen zum Indexieren einer gesamten Site auf einen Schlag
  - Beachtung des Robot Exclusion Protokolls
  - Beachtung der <meta>-Tags zum Steuern von Robotern

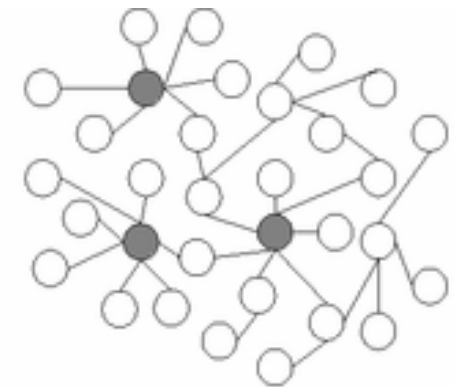
# Power Laws

- Power Laws / Potenzgesetze beschreiben in verschiedenen Gebieten Verhältnisse zwischen Variablen:

- Ökonomie (Pareto 1897)
- Literaturanalyse (Yule 1944)
- Soziologie (Zipf 1949)
- Natur: Lawinenstärke
- Web Charakteristiken



(a) Random network

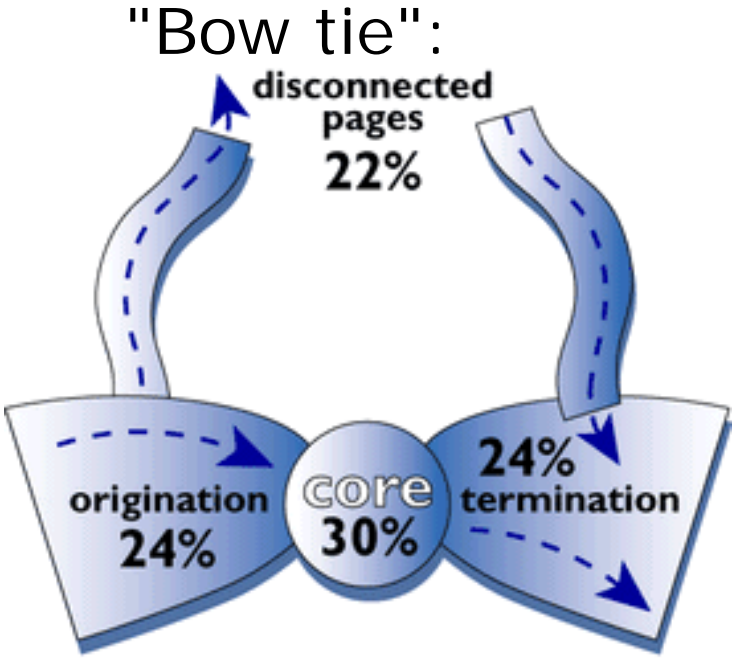
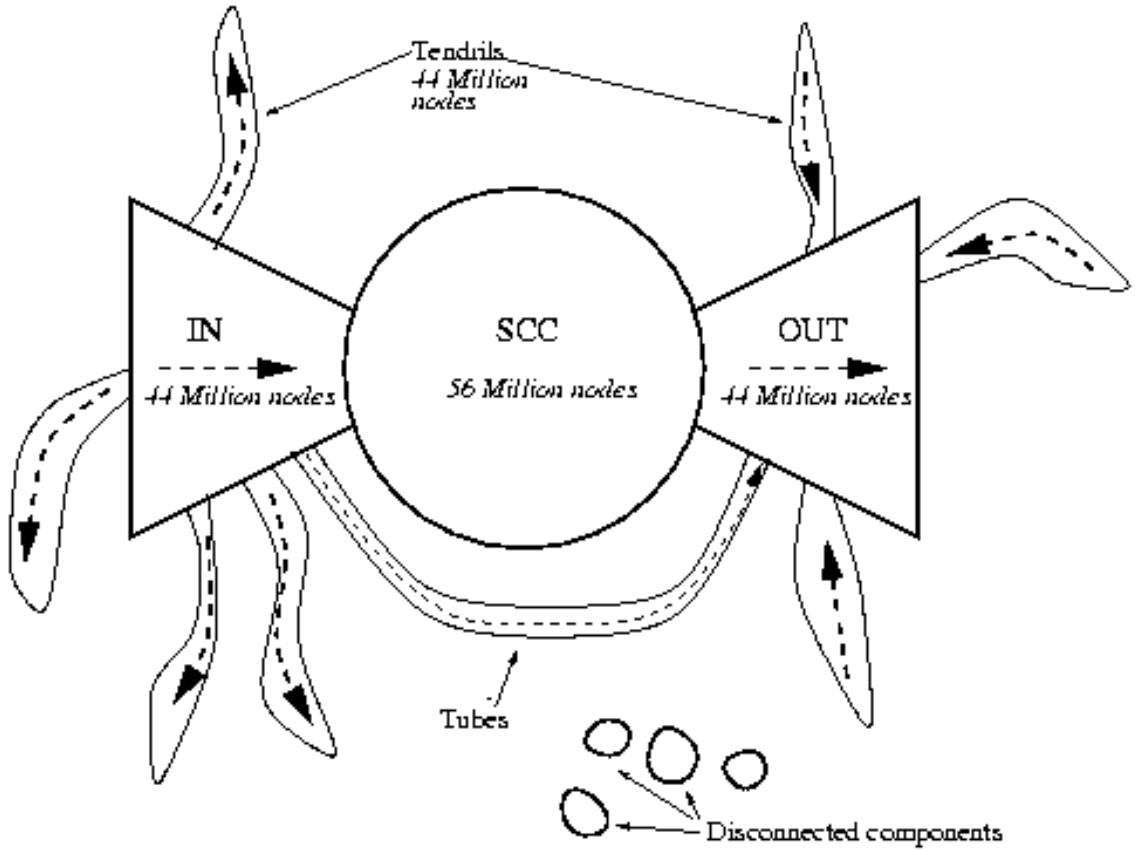


(b) Scale-free network

[Abb. wikipedia.org]

- Form:  $y \propto x^a$  für festes  $a > 1$
- Monotone strukturlose Verteilung
- Verhältnis ändert sich nicht entlang der Größenskalen  
-> Skalenfreiheit
- Tritt als Phänomen an verschiedenen Stellen bei Web-Maßen auf (Topologie, Nutzerverhalten etc) auf

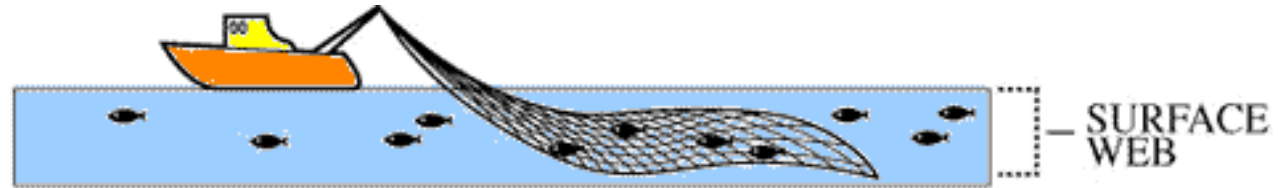
# Struktur des Web



Region	SCC	IN	OUT	Tendrils	Disc.	Total
Grösse	56463993	43343168	43166185	43797944	16777756	203549046
Anteil	28%	21%	21%	22%	8%	100%

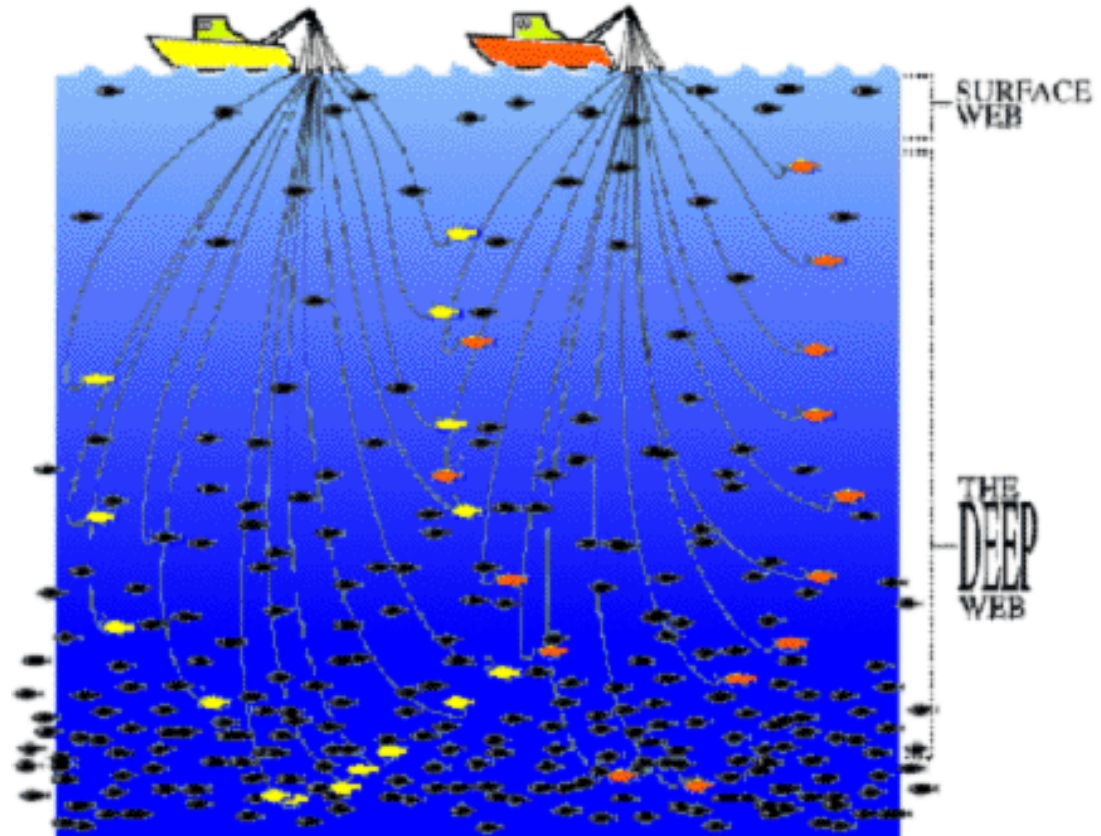
# "Deep Web"-Argumentation

- Traversierung des Web über Links



führt nur zu einem Bruchteil der Informationen

- "Deep Web" wird von Datenbankinhalten gebildet
- Umfang 400-500 mal größer als "normales" Web
- 500 Mrd Dokumente vs. 1 Mrd Dokumente
- Zugriff aber nur durch Datenbankanfragen möglich





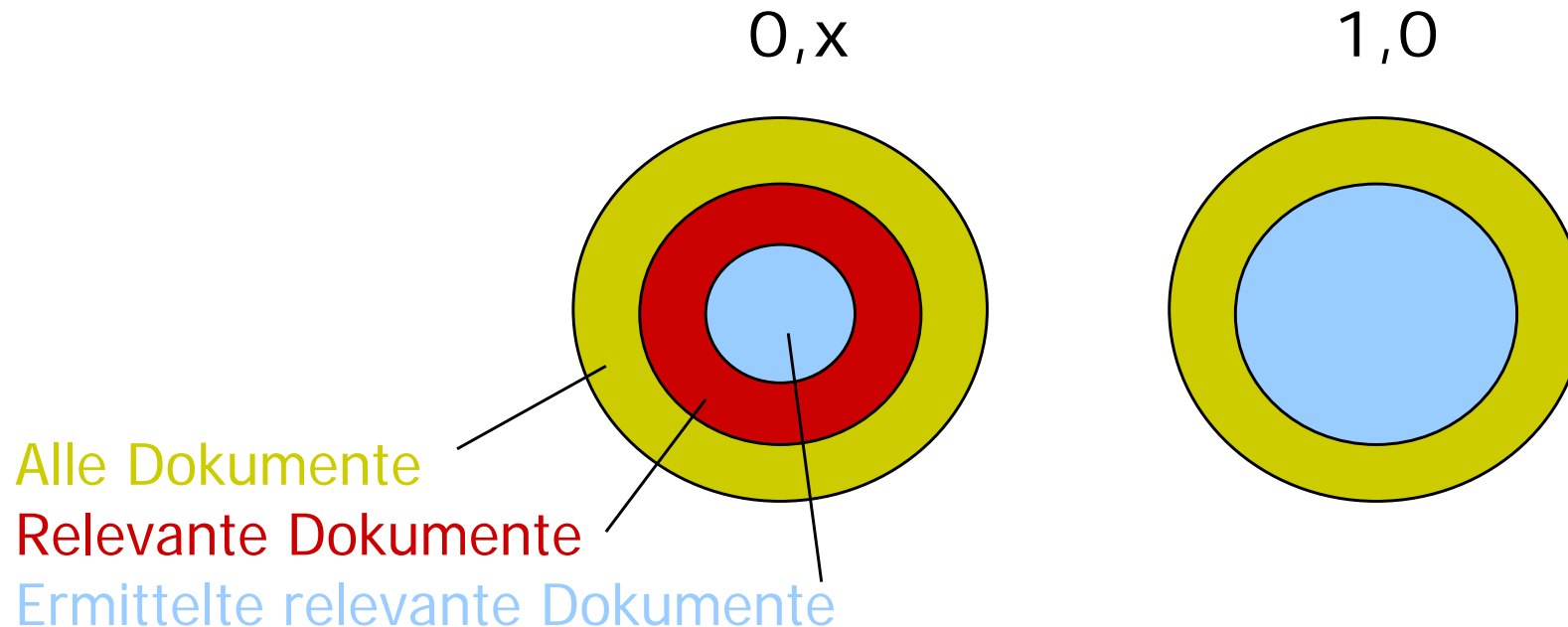
## Information Retrieval

- Aufgabe des Information Retrievals:  
  
Technologien bereitstellen, die für eine Anfrage relevante Dokumente aus einer Sammlung von Dokumenten heraussuchen
- Dokumente werden üblicherweise so vorverarbeitet und repräsentiert dass Anfragen einfach zu beantworten sind
- Bei Suchmaschinen üblich:
  - Volltextindex gesammelter Seiten erstellen
  - Anfragen an den Volltextindex weiterleiten
  - Ergebnisse ordnen
  - Verweise auf Ursprungsdokumente an Nutzer ausliefern

# Recall und Precision

- *Recall (Nachweisquote)*:  
Wie gut findet das System relevante Dokument wieder?

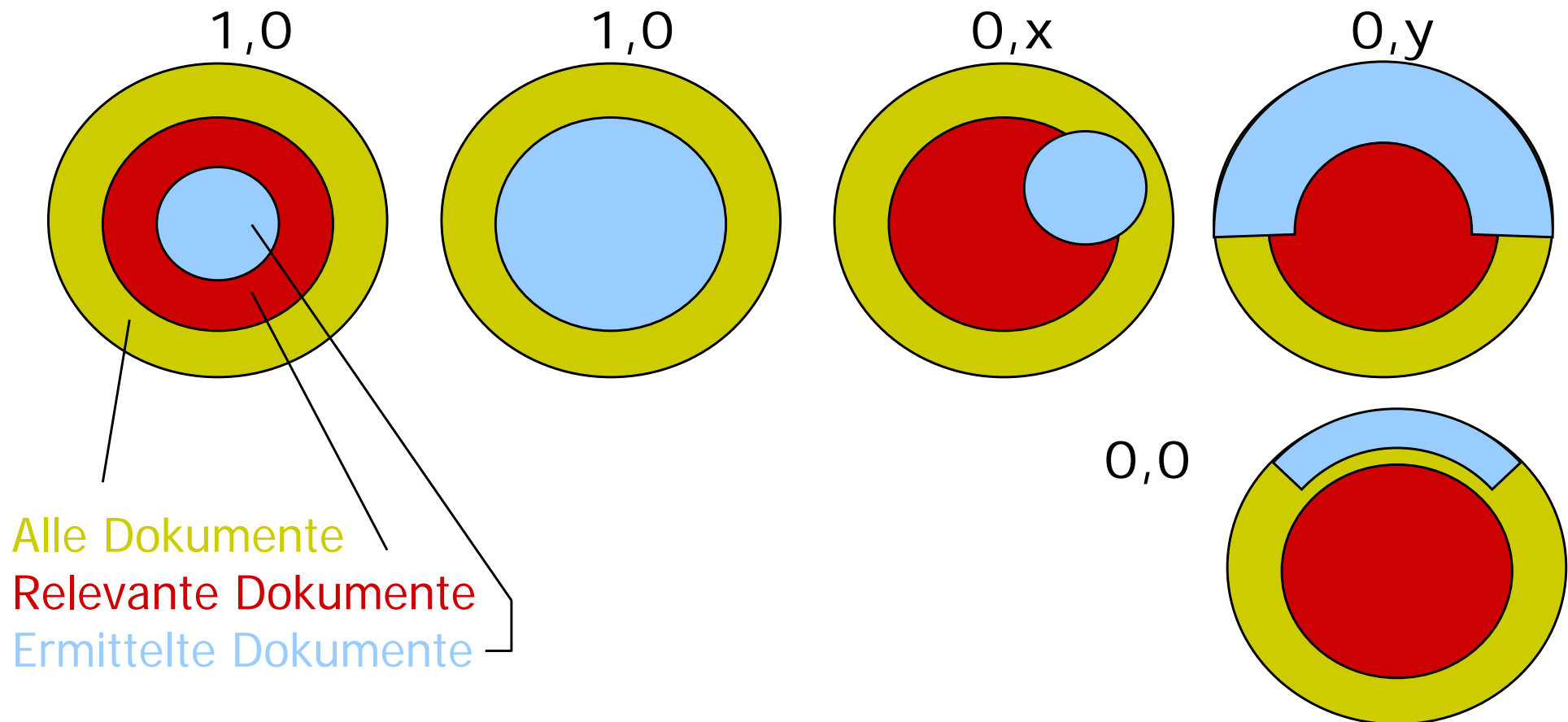
$$\text{recall} = \frac{\text{Anzahl ermittelte relevante Dokumente}}{\text{Anzahl relevante Dokumente}}$$



# Recall und Precision

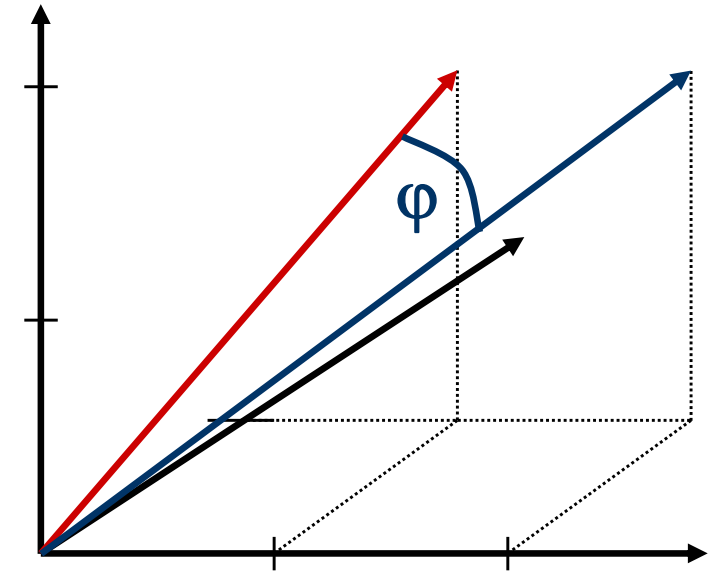
- Precision (Präzision)*: Wie gut ist die Antwortmenge

$$\text{precision} = \frac{\text{Anzahl ermittelte relevante Dokumente}}{\text{Anzahl ermittelte Dokumente}}$$



# Ähnlichkeit im Vektorraum

- Cosinusmaß: Nimmt Unterschied der Richtung der Vektoren, also den Winkel zwischen Dokument und Anfrage



$$\begin{aligned}
 \cos\varphi * |d_j| * |q| &= w_{1,j} * q_1 + \dots + w_{n,j} * q_n \\
 \text{sim}(d_j, q) &= \cos\varphi \\
 &= \frac{d_j \bullet q}{|d_j| * |q|} \\
 &= \frac{\sum w_{i,j} * q_i}{\sqrt{\sum w_{i,j}^2} * \sqrt{\sum q_i^2}}
 \end{aligned}$$

# Automatische Gewichtung

- *Termfrequenz*: Ein in einem Dokument häufiger Term ist charakteristischer als ein seltener Term  
 $tf_{ij}$ : Häufigkeit von Term  $T_j$  im Dokument  $i$
- Führt zu hohem Recall
- Leicht zu beeinflussende Dokumentenbewertung:  
Wiederholung eines Wortes
- *Document frequency (Dokumentenfrequenz)*:  
Seltener Term ist für eine Dokument charakteristisch in dem er häufig auftritt
- $df_j$ : Anzahl des Auftretens von  $T_j$  in  $N$  Dokumenten
- *Inverse document frequency*:  
Wie stark ist  $T_j$  charakteristisch
- $idf_j = \log(N/df_j)$
- auch:  $idf_j = \log(df_{\max}/df_j)$  und andere

# Automatische Gewichtung

- *tfidf* Regel:  
Anzahl des Vorkommens eines Terms gewichtet mit  
dessen Charakterisierungsfähigkeit  
 $w_{ij} = tf_{ij} * \log(N/df_j)$
- Es existieren sehr viele weitere  
Gewichtungsmöglichkeiten
- Gewichtung wählen
  - für Dokumentenvektoren
  - für Anfragevektoren

# Normalisierung

---

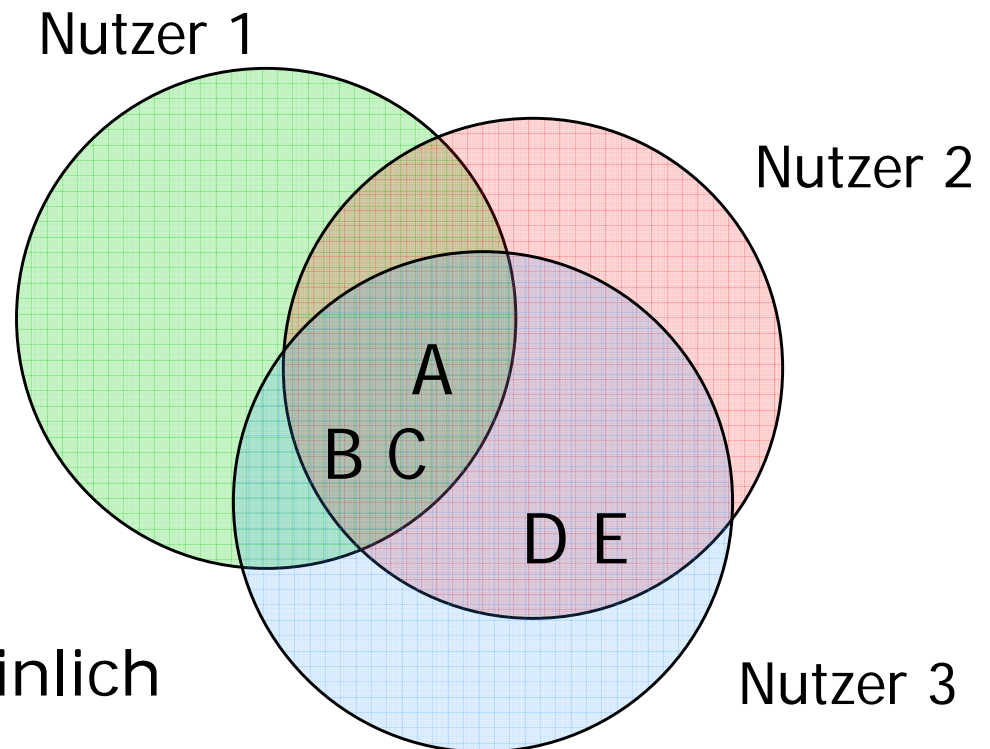
- Dokumentenvorbereitung
  - Verarbeiten und Entfernen von HTML
  - Ermittlung indexierungsrelevanter Informationen
    - alt Attribut bei <img>
    - <meta>-Tags
    - lang Attribut
    - ...
  - Umgang mit Zeichenkodierungen
  - Entitäten expandieren
  - Interpunktion entfernen
- Aufteilen in Token
- Stop words entfernen
- Stemming

# Multimedia Indexing

- Große Anteile der im Netz verfügbaren Informationen sind kein Text und können nicht in einem *Volltextindex* gespeichert werden
- Beispiele:
  - Bilder – Fotos, Zeichnungen
  - Audio – Musik, Sprache
  - Video – Filme, Nachrichten
- Oftmals ist Information in mehreren Medien verteilt
  - Beispiel: Nachrichten, Bild und Ton
- Problem
  - Ermittlung von Indextermen
  - Entwurf von Ähnlichkeitsmaßen
- Unser Beispiel: Audio

# Grundidee

- Ähnliche Nutzer haben ähnliche Vorlieben
- Vorlieben eines Nutzers können genutzt um einem anderen einen Vorschlag zu machen
- Beispiel: A...E sind Produkte, Informationen etc.
- Nutzer 1, 2 und 3 ähneln sich, da sie alle A, B und C mögen/haben
- Für Nutzer 1 sind wahrscheinlich auch D und E relevant



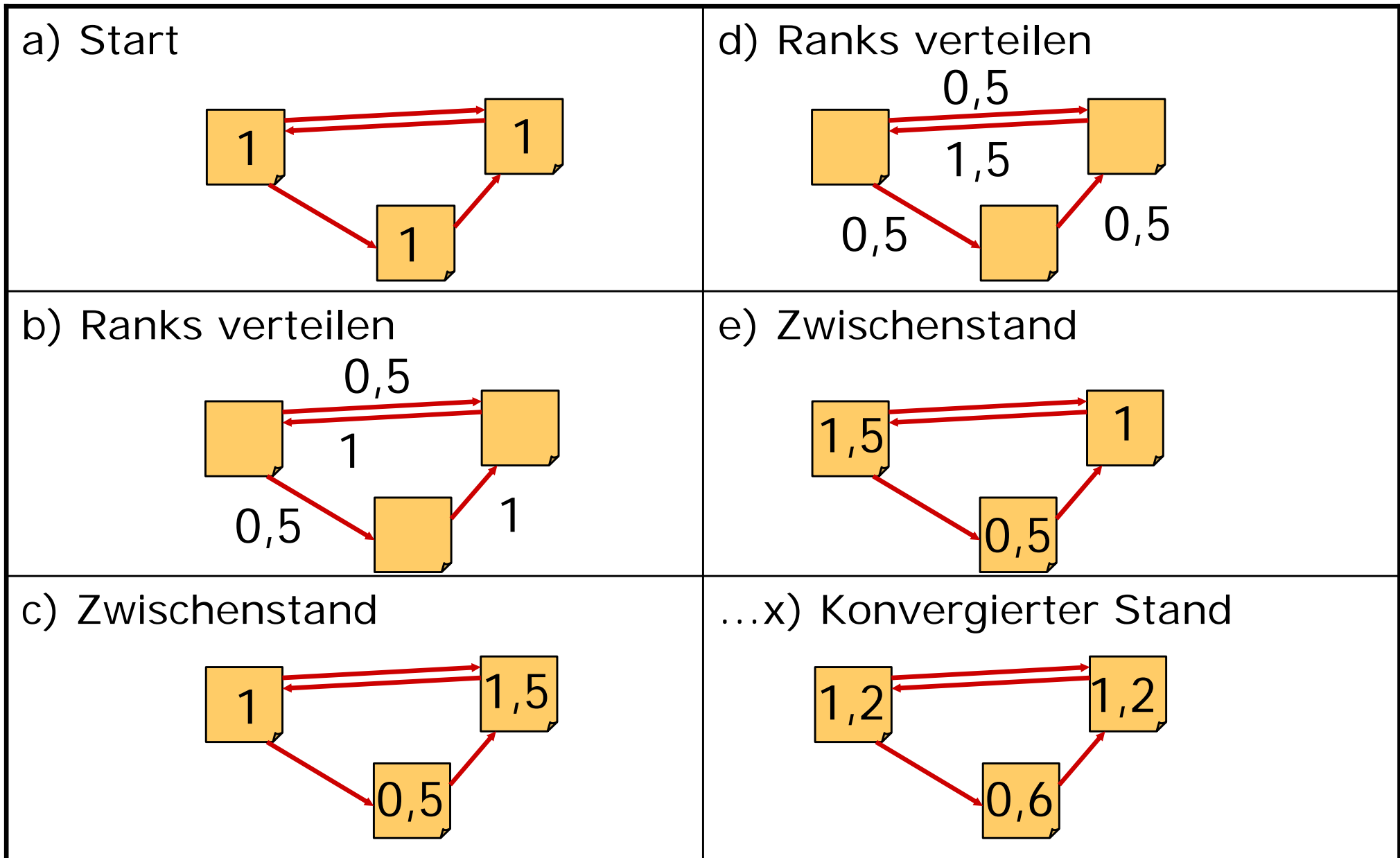


## Suchmaschinenverfahren

# PageRank

- PageRank Verfahren: Bewertung aller Web-Seiten nach ihrer relativen Popularität
- Kerntechnologie von Google
- Viele Verweise auf eine Seite legen nahe, dass die Seite wichtig ist
- Setzen eines Links ist Einschätzung der Wichtigkeit der referenzierten Seite (ähnlich einem Zitat)
  - Sie ist aber eigentlich nur populär!
- Links von wichtigen Seiten erhöhen Wichtigkeit
  - Eine Seite hat hohen PageRank, wenn die PageRanks der Seiten, die auf sie verweisen, hoch sind
- Seiten mit hohem Pagerank werden in Ergebnismengen zuerst gelistet

# Beispieliterationen

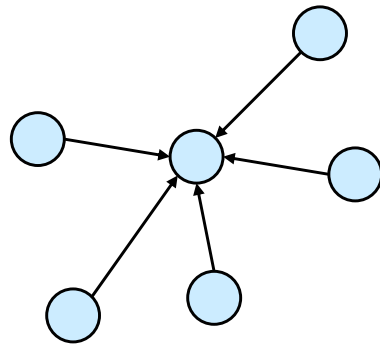


# Vollständiger PageRank

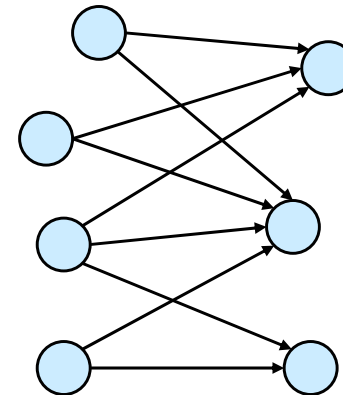
- Intuition: Einfache Version modelliert Zufalls-Surfer und die Wahrscheinlichkeit, dass Seiten von ihm durch zufälliges Verfolgen von Links besucht werden
- In einer Schleife springt der Surfer zu einer beliebigen andern Seite.  $E$  modelliert die Verteilung dieser zufälligen Auswahl
- PageRank ist die Verteilung der Wahrscheinlichkeit eine bestimmte Seite durch zufällige Navigation zu erreichen
- Bestimmung aus globaler Sicht

# Autoritäten und Hubs

- Autoritative Quellen in  $G_\sigma$  sollten sich dadurch auszeichnen, dass die Mengen der Seiten, die auf sie zeigen überlappen
- *Hubs* verweisen auf mehrere Autoritäten



grosser in-degree



Hubs

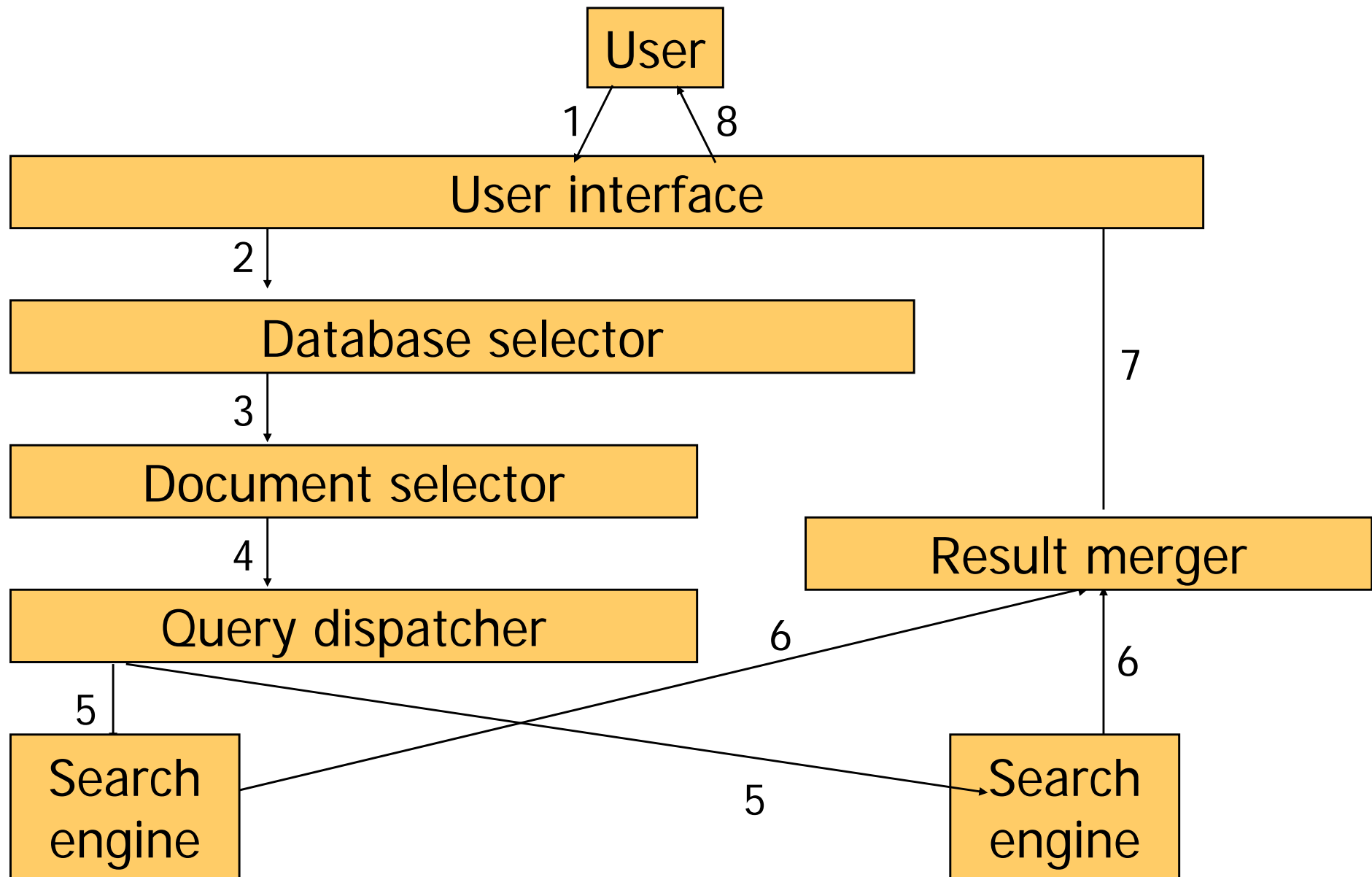
Autoritäten

- Guter Hub zeigt auf gute Autoritäten
- Gute Autorität wird von vielen guten Hubs referenziert
- PageRank ermittelt nur populäre Autoritäten

# Vorteile

---

- Bessere Abdeckung des Suchraums Web:  
Suchmaschinen decken immer nur einen Teil des Web aus, Metasuchmaschinen deren Summe
- Bessere Skalierbarkeit  
Suche wird auf Suchmaschinen verteilt
- Bessere automatische Absuche mehrerer Suchmaschinen  
Keine manuellen Abfragen notwendig
- Effektivität der Suche verbessern  
Spezialsuchmaschinen gruppieren Dokumente thematisch





## Web Nutzung und Betrieb

# Messgrößen

- Hits
  - Anzahl der Abrufe von Informationen
  - Summe der Anzahl der Requests mit 200 und 304 Antwort
  - Nicht sehr aussagekräftig, weil nicht jede Datei eigenständige Informationseinheit
- Pageviews/Page Impressions
  - Anzahl der abgerufenen HTML-Seiten
  - Anzahl der Hits mit HTML Dateien als Antwort
  - Beschränkt auf einen Medientyp

# Messgrößen

---

- Visits / Sessions
  - Zusammenhängende Abrufe in einem Zeitraum
  - Navigationspfade aus Logfile
  - Nicht zuverlässig identifizierbar
  - Problem: Wann ist Visit beendet?
- Heuristiken
  - Zeitorientiert:
    - Gesamtdauer einer Visit ist nach oben begrenzt
    - Verweildauer auf einer Seite ist nach oben begrenzt
  - Navigationsorientiert
    - Topologische Begrenzung: Sitzungsende, wenn Seite nicht von vorherigen Seiten aus erreicht werden konnte
    - Begrenzung durch Referrer: Sitzungsende, wenn Seite nicht durch Navigation von vorheriger Seite erreicht wurde

# Aus [www.spiegel.de/index.html](http://www.spiegel.de/index.html)

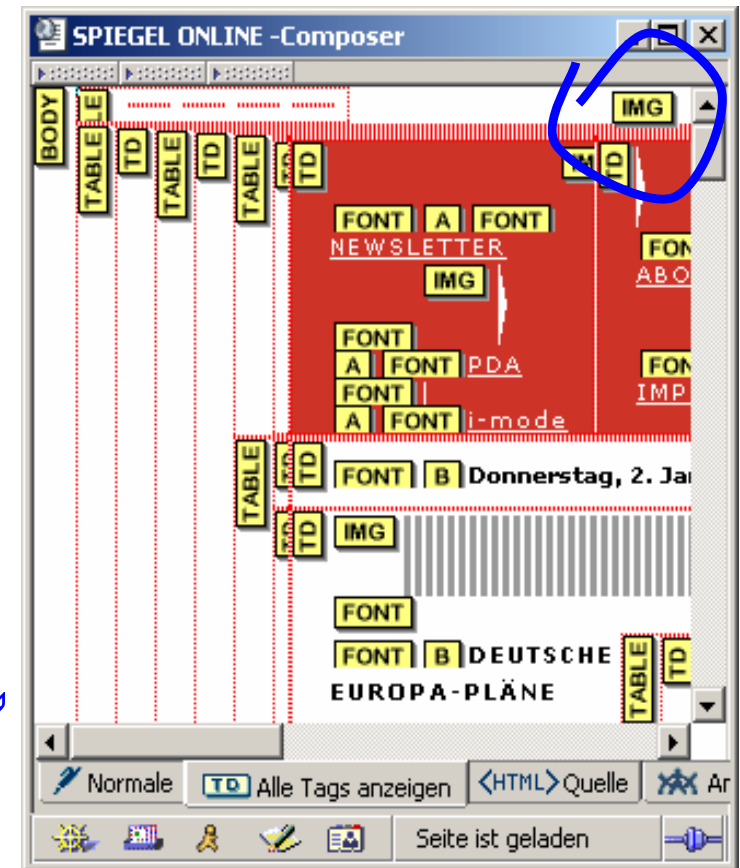
```

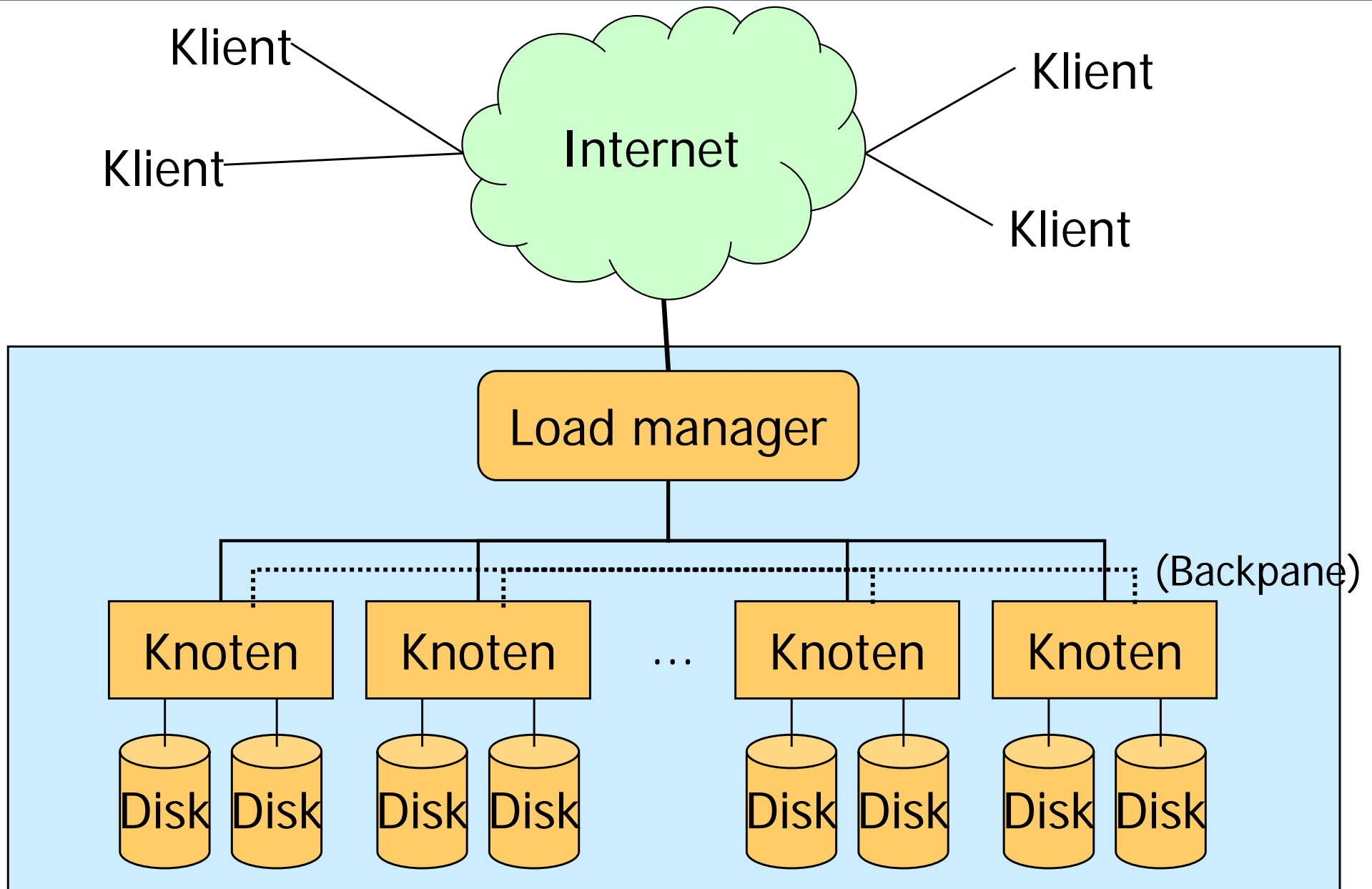
<body bgcolor="#ffffff" text="#000000"
  link="#b20a15" vlink="#b20a15" alink="#ff0000"
  marginheight="0" marginwidth="4" leftmargin="4"
  topmargin="0" rightmargin="4" bottommargin="0">
<!-- I W V E R S I O N = " 1 . 2 " -->
<script language="JavaScript">
<!--
  var IWV="http://spiegel.iwbox.de/cgi-bin/iw/CP/
    spiegel;/home/c-18/be-PB64-aG9tZXBhZ2UvY2VudGVy";
  document.write('<IMG SRC="'+IWV+'?r='+
    escape(document.referrer)+'" WIDTH="1"
    HEIGHT="1" BORDER="0" ALIGN="RIGHT">');
// -->
</script>

<noscript>
  <IMG SRC="http://spiegel.iwbox.de/cgi-bin/iw/CP/spiegel;
    /home/c-18/be-PB64-aG9tZXBhZ2UvY2VudGVy"
    WIDTH="1" HEIGHT="1" BORDER="0" ALIGN="RIGHT">
</noscript>
<!-- /I W V -->
<!-- I W V E R S I O N = " p r e v " -->

<!-- /I W V -->

```





# Load management

- Layer-4-switches
  - Hardware die TCP "versteht"
  - Switch leitet Pakete aufgrund von TCP-Dienste-Feld und Portnummern an unterschiedliche Server weiter
- Layer-7-switches
  - Hardware die HTTP "versteht"
  - Können URLs mit Netzbandbreite parsieren und leiten Pakete entsprechend weiter
- Meistens als Paar vorhanden
- >20Gbits/s Durchsatz
- Automatisches Monitoring von Knoten

# Verfügbarkeit

---

- Zentrale Anforderung an große Dienste:

## Verfügbarkeit (Availability)

- Gemessen in "Neunern":
  - Vier Neuner: 0,9999 Verfügbarkeit (<60 Sek. Ausfall/Woche)
  - Fünf Neuner: 0,99999
- Ähnlich geleitete Systeme:
  - Telefonsystem
  - Zugverkehr
  - Wasserversorgung

## Weitere Maße

- MTBF: Mean-time-between-failures, mittlerer zeitlicher Abstand zwischen Fehlern
- MTTR: Mean-time-to-repair, mittlere Dauer der Fehlerbehebung

- $$\text{uptime} = \frac{(\text{MTBF} - \text{MTTR})}{\text{MTBF}}$$

- MTBF = 6 Tage (518400s), MTTR = 20 Min (1200s)

$$\text{uptime} = \frac{(518400\text{s} - 1200\text{s})}{518400\text{s}} = 0,997685$$

# Yield und Harvest

- Weiteres Maß: Yield – wie viele Anfrageergebnisse?
  - $\text{yield} = \frac{\text{bearbeitete Anfragen}}{\text{gestellte Anfragen}}$
  - entspricht Nutzererfahrung
  - gewichtet Uptime-Sekunden
- Weiteres Maß: Harvest – welcher Teil der Datenbank ist nutzbar?
  - $\text{harvest} = \frac{\text{zugreifbare Daten}}{\text{gesamte Daten}}$
  - Erweiterbar zum Anteil der nutzbaren Dienste



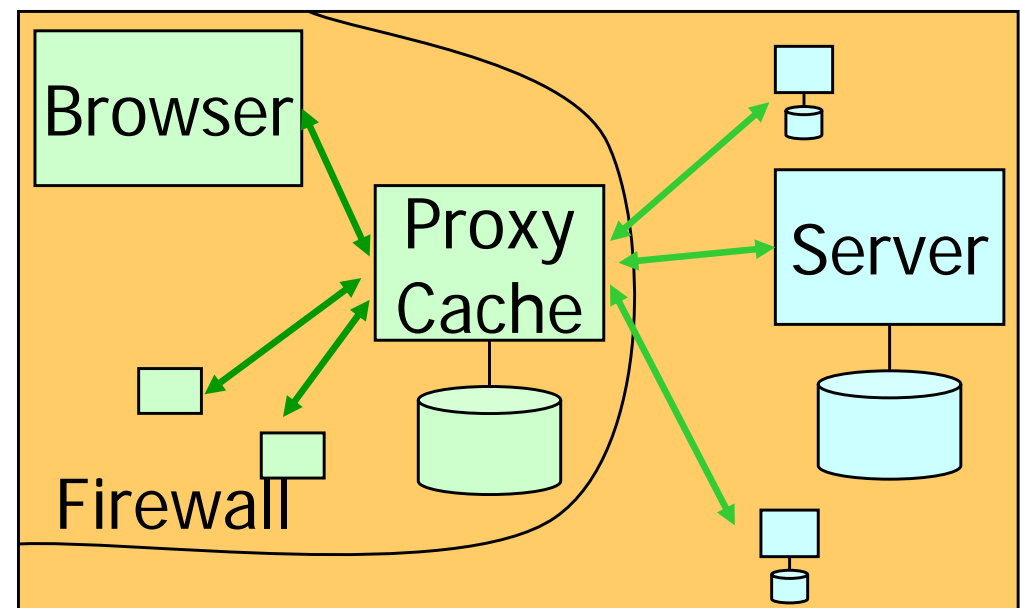
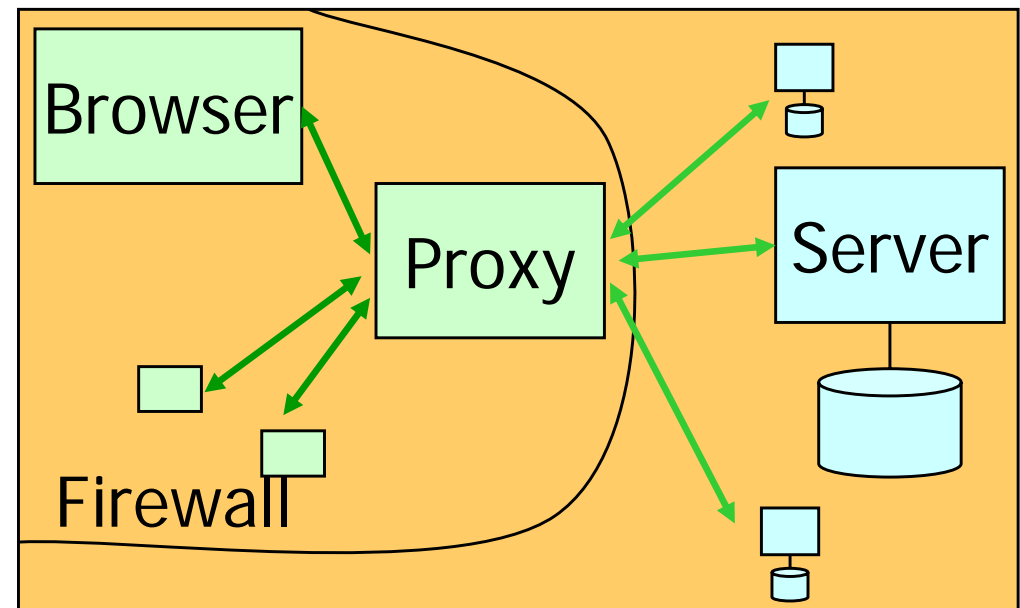
## Caching

# Caching

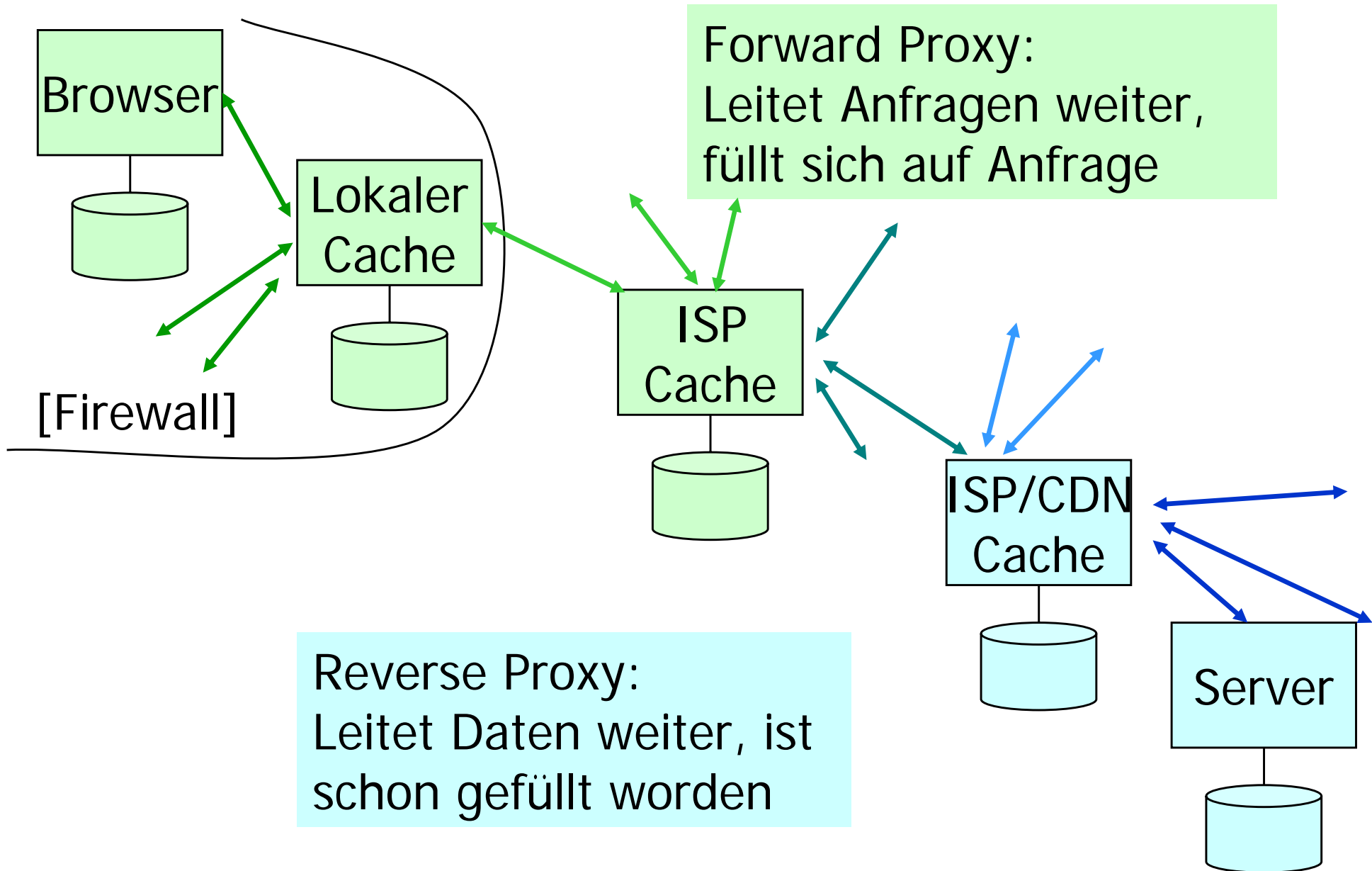
- Ursprünglich aus Rechnerarchitektur:
  - CPU schneller als Hauptspeicher
  - → Daten in schnellem Zwischenspeicher, dem *Cache* halten
- Ziel im Web: Netzwerklatenz kaschieren
  - Klient schneller als Netz (+ Server)
- Grundbegriffe:
  - Gesuchte Daten zwischengespeichert vorgefunden: *Hit / Treffer*
  - Gesuchte Daten nicht gefunden: *Miss / Fault / Fehler*
  - Bei Fehler nachgeladenes Originaldatum in Cache gespeichert
  - → Annahme: Mehrere Zugriffe zeitlich gruppiert
  - Oft: Block um gesuchte Daten in Cache geholt
  - → Annahme: Zugriffe örtlich gruppiert

# Proxies und Caching

- *Proxy/Stellvertreter* anstelle vom Klienten
  - Leitet HTTP Anfrage von Klienten an Server / andere Proxies weiter
  - Tritt für den Server als Klient auf
  - Leitet Antwort an Klienten weiter
- *Proxy Cache*
  - Agiert auch als Cache
  - ➔ Annahme: Zugriffe organisatorisch gruppiert



# Caches im Web allgemein



# Prefetching

---

- Auswirkung
  - zwischen Browser und Server
    - 45% weniger Latenz beim Browser
    - 200% Netzlast
  - zwischen Proxy und Server
    - 60% weniger Latenz beim Browser
  - zwischen Klient und Proxy
    - 23% Weniger Latenz
    - Großer Browser-Cache notwendig
- Push-Ansätze
  - Server/Proxy verteilt Dokumente zu Klienten
  - nicht im HTTP Modell
- Prefetching nicht verbreitet / nicht als sinnvoll angesehen

- Wenn Cache voll ist, muss Platz durch Löschung von Objekten geschaffen werden
- Drei Ansätze
  - Traditionelle Ansätze, basiert auf Objektnutzung
    - Least Recent Used LRU:  
Objekt mit ältester Nachfrage wird gelöscht
      - Populärste Strategie
      - Großes neues Objekt verdrängt viele kleine alte
    - Least Frequently Used LFU:  
Objekt mit wenigsten Nachfragen wird gelöscht
      - Größe von Objekten nicht berücksichtigt
      - Da die meisten Objekte gleich oft genutzt werden (1) ist Auswahl eigentlich zufällig
    - Pitkow/Recker:  
Wie LRU, aber: falls alle Objekte am selben Tag nachgefragt wurden, wird größtes gelöscht

# Cache Kohärenz

- Nutzer können veraltete Seiten vom einem Cache erhalten
- Ähnlich Cache-Kohärenz in Verteilten Systemen, aber
  - andere Zugriffsmuster
  - andere Dimensionen
  - Web-Objekte werden nur an einem Ort geändert
- HTTP Unterstützung
  - Header Expires: *Datum* liefert Ungültigkeitsdatum
  - GET mit If-Modified-Since: *Datum* Header liefert Seite nur bei Änderungen nach einem Datum
  - Header Pragma: no-cache verhindert Caching
  - Header Last-Modified: *Datum* liefert Änderungsdatum
  - Header Date: *Datum* enthält Datum des letzten Tests auf Aktualität
  - Header ETag: *Signatur* liefert eine Quersumme des Objekts

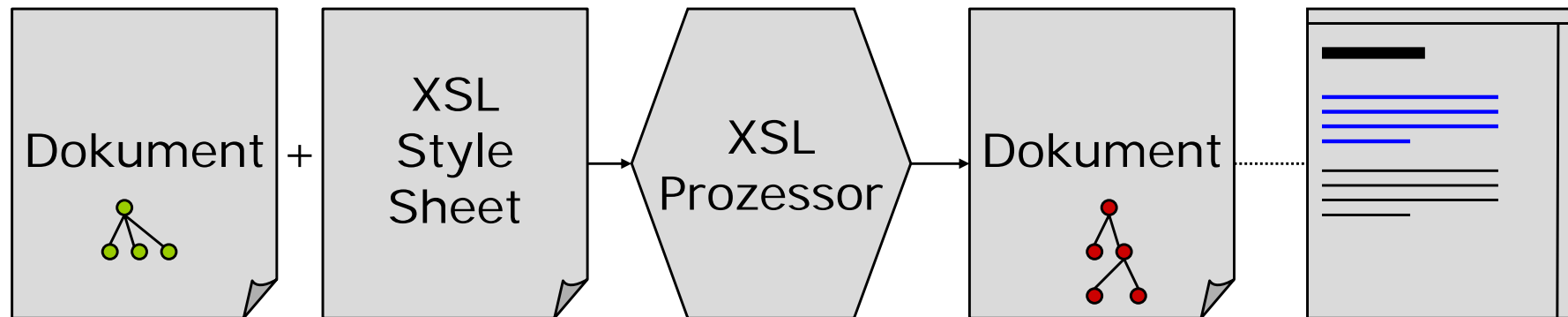
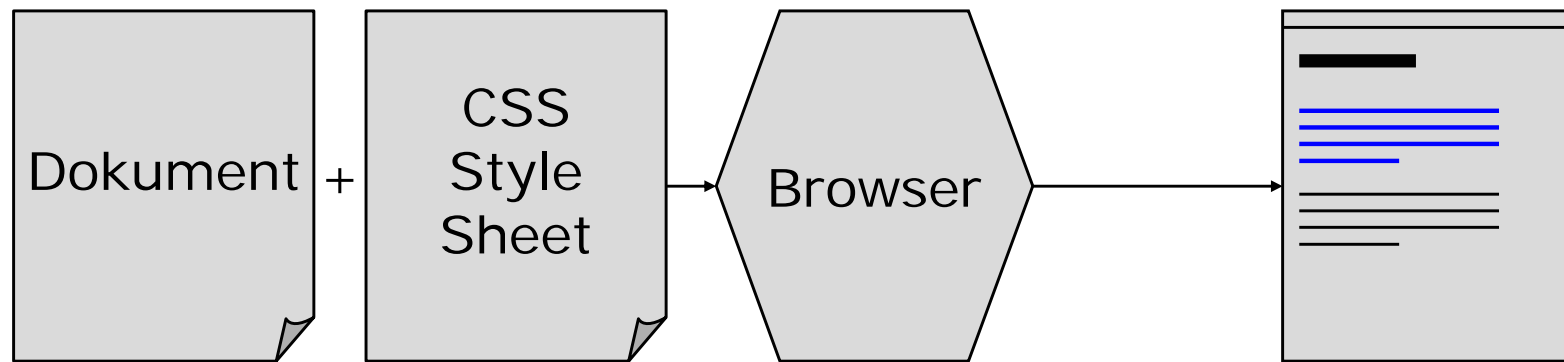


## Inhaltsdarstellung

- Trennung von Inhalt und Darstellung ist aber wichtig für
  - Geräteunabhängigkeit von Informationen (Handy vs. PC)
  - Medienunabhängigkeit von Informationen (Grafik vs. Sprache)
  - Sprachunabhängigkeit von Informationen (" vs. „ vs. »)
  - Mehrkanal Veröffentlichungen (WAP und Web)
  - Verarbeitbarkeit von Informationen
- Notwendig:
  - Darstellungsmodelle
  - Sprachen zur Festlegung von Darstellungseigenschaften
  - Verarbeitungsmodelle für diese Sprachen

# Style Sheets

- W3C Style Activity entwickelt
  - Cascading Stylesheets CSS
  - XML Stylesheet Language XSL



# CSS Grundidee

- Grundidee:  
Zu HTML Tags werden definierte Attribute für Darstellungseigenschaften gesetzt
- CSS-Datei getrennt von HTML-Datei gehalten
- Beispiel: Um Überschriften in großer blauer Schrift darzustellen:  

```
h1 {color: blue; font-size: 22pt; }
```
- CSS definiert
  - Rahmensyntax zur Notation
  - Menge von Attributen
  - Menge von Werten
  - Bedeutung
  - Mechanismen zur Anbindung von Stilinformationen an und in HTML Seiten

# CSS2: Medienarten

- Darstellungsstil ist abhängig vom Ausgabemedium
  - Bildschirm
  - Papier
  - Spache
  - Braille
  - ...
- CSS erlaubt getrennte Stildefinitionen:

...

```
a:link {  
  color: #000099;  
  text-decoration : none ;  
}
```

```
@media print {
```

```
  a:link, a:visited, a: hover, a: active, a: focus {  
    text-decoration: none;  
    color: blue  
  }
```

```
}
```

# XSL Regeln

- Eine XSL-Regel definiert Muster und dazugehörige Transformationen
- Sie wird auf alle passenden Knoten im Quelldokument angewandt
- Beispiel:

```
<xsl:template match="ANSCHRIFT">
  <tabelle>
    <xsl:apply-templates select="NAME" />
  </tabelle>
  <tabelle>
    <xsl:apply-templates select="ORT" />
  </tabelle>
</xsl:template>
```



## Mehrsprachigkeit im Web

# Internationalisierung

---

- *Internationalisierung* ist die Planung und Implementierung von Diensten und Produkten so dass sie einfach an lokale Sprachen und Kulturen anpassbar sind, was *Lokalisierung* ist
- Internationalisierung
  - „I18N“ - „I - eighteen letters –N“ – „Internationalization“
  - Voraussetzung für Lokalisierung
  - Beispiele
    - Platzgestaltung in GUIs läßt Raum für Sprachen die mehr Zeichen benötigen
    - Verwendung internationaler Zeichenrepertoires und -codes, z.B. Unicode
    - Vergabe leicht übersetzbarer Beschreibungen für Graphiken
    - Verwendung allgemeinverständlicher Beispiele (Social Security Number ...)
    - Vorausplanung der Übersetzung in Sprachen mit Kodierungen mit mehr als einem Byte pro Zeichen in Software

# Lokalisierung

---

- *Lokalisierung* ist die Anpassung eines Produktes oder Dienstes an eine Sprache, Kultur und lokales "look-and-feel" was durch *Internationalisierung* vereinfacht wird
- Lokalisierung
  - „L10N“ – „L - ten letters –N“ – „Localization“
  - Übersetzung
  - Aber auch: Anpassung an Zeitzonen, Währung, Feiertage, Farbkonventionen, Namen, Geschlechterrollen etc.
  - Ziel: Lokalisiertes Produkt oder Dienst soll so aussehen, als sei er/es lokal entwickelt worden

# Sprachkürzel nach RFC 4646

- RFC 3066
  - Sprachkürzel:  
Ländercode-Sprachcode (en-US)
  - Bezüge auf ISO Standards
- RFC 4646
  - Sprachkürzel:  
language-script-region-variant-extension-privateuse
  - <http://www.iana.org/assignments/language-subtag-registry>:
    - Type: language  
Subtag: fr  
Description: French  
Added: 2005-10-16  
Suppress-Script: Latn
    - Type: region  
Subtag: CA  
Description: Canada  
Added: 2005-10-16
  - fr-CA ist gültiges Sprachkürzel

- Alle HTML Elemente können Sprachbezogene Attribute tragen
  - lang-Attribut: Wert ist Sprachcode
    - Wird vom umgebenden Element „geerbt“
    - Kann jeweils überschrieben werden
    - Default ist durch Content-language HTTP Header gegeben
  - dir-Attribut: (Horizontale) Schreibrichtung der Schrift
    - ltr: Left-to-Right
    - rtl: Right-to-Left
    - Wird vom umgebenden Element „geerbt“
    - Kann jeweils überschrieben werden

# Zeicheneigenschaften

- Zeichenrepertoire (Character Set, Abstract Character Repertoire, ACS)
  - Eine Menge von Zeichen
  - Definiert durch Namen und Beispiele
  - {Pfund (£), Zett (Z), Ypsilon (Y), Herz (♥)}
  - Keine Ordnung, keine Codierung
- Zeichencode (Coded Character Set, CCS)
  - Abbildung(en) Zeichen → Zeichenposition
  - Z → 5A, ç → FEA5 (Khah)
  - z.B. UNICODE, ISO 8859-1

# Zeicheneigenschaften

- Zeichenkodierung (Encoding)
  - Character Encoding Form (CEF)
    - Abbildung einer Zeichenfolge auf Strom gleichgroßer Codes
    - z.B. 

005A	FEA5
------	------
  - Character Encoding Scheme (CES)
    - Abbildung einer Zeichenfolge auf einen Bytestrom
    - z.B. 

5A	00	A5	FE
----	----	----	----
- Zeichensatz
  - Bedeutung unklar, kann Repertoire, Code oder Kodierung meinen
- „charset“
  - meint Encoding!

# Zeicheneigenschaften in HTML

- HTML Spezifikation arbeiten immer auf UNICODE Basis
- UNICODE muss nicht benutzt werden, Software muss aber so tun als benutze sie UNICODE
  - Klare Spezifikationen
  - Erlaubt Internationalisierung
  - Unterstützt Lokalisierung
  - Rückwärtskompatibel (ISO 8895-1 gleich unterstem UNICODE Zeichencode)
  - Abstrahiert von Repräsentation der Zeichen in Byteströmen
- UNICODE ist Abstraktionslevel
  - Abstrahiert in der Spezifikation von interner Zeichenkodierung
  - Abstrahiert von Transportrepräsentation

# Content Negotiation

- Auswahl passender Information bezüglich der Dimensionen
  - Medienart (Accept: text/html, text/plain)
  - Sprache (AcceptLanguage: en-us; q=0.75, en; q=0.5; \*; q=0.25)
  - Encoding (Accept-Encoding: compress; q=0.5, gzip; q=1.0)
  - Charset (AcceptCharset: iso-8859-1, utf-8; q=0.75, \*; q=0.5)
  - Angegebene Qualitätsmaße
- Server-abhängige Implementierungen
  - z.B. Schema über Dateinamen:
    - foo.en.html
    - foo.html.en
    - foo.en.html.gz

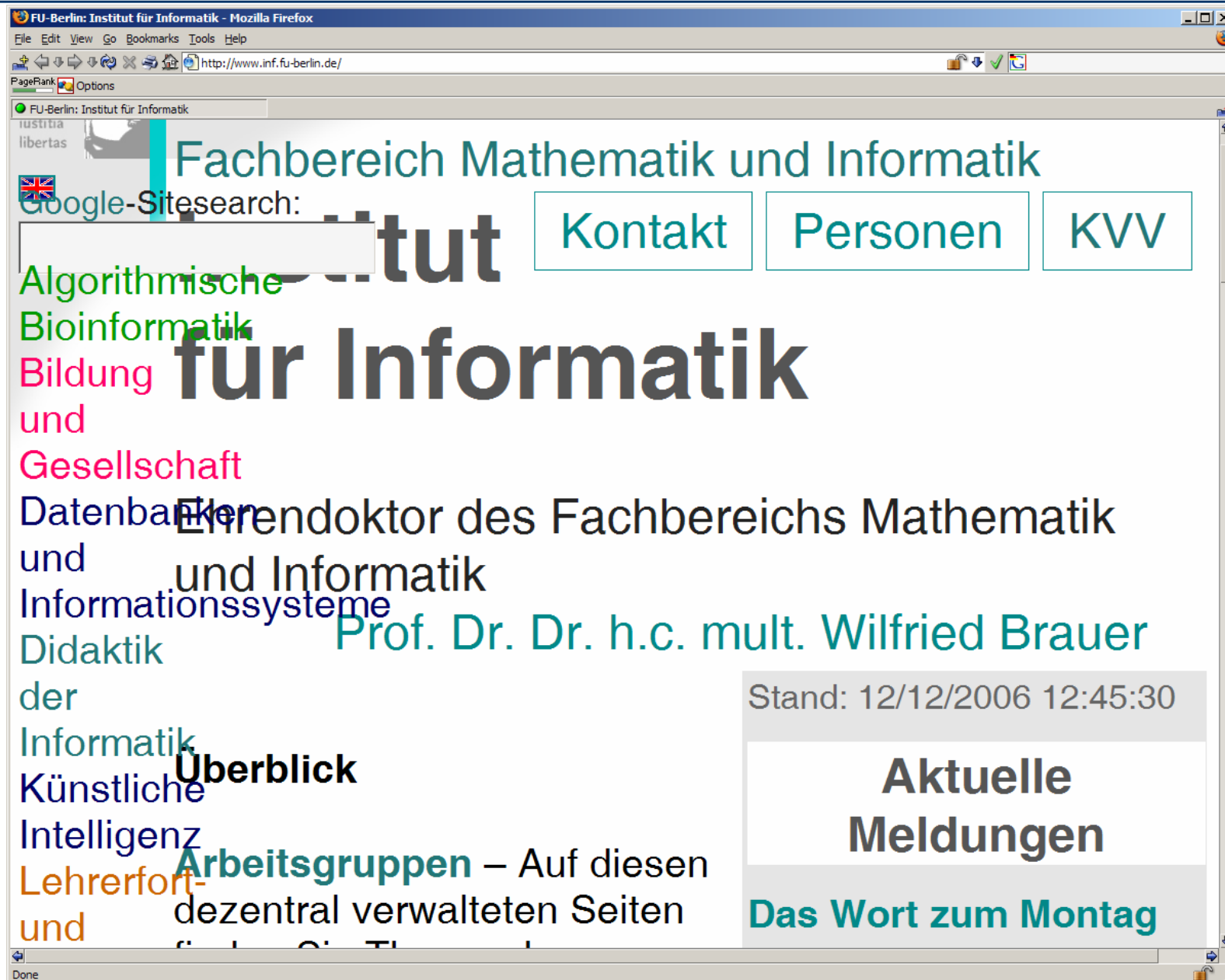


## Web Accessibility / Barrierefreiheit

# Barrierefreiheit (Accessibility)

- Barrierefreiheit für Behinderte, Ältere, ... bedeutet
  - Wahrnehmen von Informationen
  - Verständnis von Informationen
  - Navigation zwischen Informationen
  - Interaktion mit Informationsquellen
- Szenarien
  - Farbenblinder Online-Käufer
    - Stylesheet-Wahl!
  - Journalist mit RSI (repetitive strain injury)
    - Tastatureingabe!
  - Tauber Student
    - Untertitel für Audioströme!
  - Blinder Buchhalter
    - Andere Ausgabe (Braille), angemessene Formatierung!
  - Schüler mit Leseschwäche
    - Zusätzliche Grafiken!
  - Rentner
    - Vergrößerungsmöglichkeit!

# Institutshomepage vergrößert...



# WAI Web Accessibility Initiative

- <http://www.w3.org/WAI/>
- W3C Initiative zur Entwicklung von Anleitungen zur Etablierung von Barrierefreiheit mit Web Technologien
- Bereiche
  - Web Content Accessibility Guidelines (WCAG)
    - Beschäftigt sich mit den Informationen auf Web Seiten und deren Zugangsmöglichkeiten
  - User Agent Accessibility Guidelines (UAAG)
    - Eigenschaften von Klienten in Bezug auf Barrierefreiheit
  - Authoring Tool Accessibility Guidelines (ATAG)
    - Beschäftigt sich mit Werkzeugen zum Inhaltsmanagement und deren Fähigkeiten zur Ermöglichung von Barrierefreiheit

- Web Content Accessibility Guidelines (WCAG)
  - W3C Recommendation 5-May-1999
  - <http://www.w3.org/TR/WCAG10/>
  - WCAG 2.0 vor Verabschiedung
  - Curriculum for Web Content Accessibility Guidelines 1.0  
<http://www.w3.org/WAI/wcag-curric/>
  - Viele Beispiele dort entnommen
- Ziele
  - Transformierbarkeit
    - Trennung von Struktur und Darstellung
    - Textverwendung
    - Audio- und Video als Alternative
    - Keine Annahmen über Geräte
  - Verständlichkeit und Navigierbarkeit
- Aufbau
  - Guidelines
  - Checkpoints mit Prioritäten
  - Technologiehinweise

## Guideline 4 - Sprache markieren

- Checkpoint 4.1, Priorität 1
  - Clearly identify changes in the natural language of a document's text and any text equivalents (e.g., captions)
    - Mother, he's asking you to go.  
He's saying, <SPAN lang="fr"> "Allons, Madame plaisante!" </SPAN>
- Checkpoint 4.2, Priorität 3
  - Specify the expansion of each abbreviation or acronym in a document where it first occurs
    - When in Boston, be sure to visit the <ACRONYM TITLE="Museum of Fine Arts">MFA</ACRONYM>, <ACRONYM TITLE="Massachusetts Institute of Technology">MIT</ACRONYM> and, of course, the <ACRONYM TITLE="World Wide Web consortium">W3C</ACRONYM>. These destinations are easily reached via <ABBR TITLE="Massachusetts Avenue">Mass. Ave.</ABBR> or <ABBR TITLE="Memorial Drive">Mem. Dr.</ABBR>.
- Checkpoint 4.3, Priorität 3
  - Identify the primary natural language of a document
    - <HTML lang="it">



## Rich Web Clients

# Rich Web Clients

---

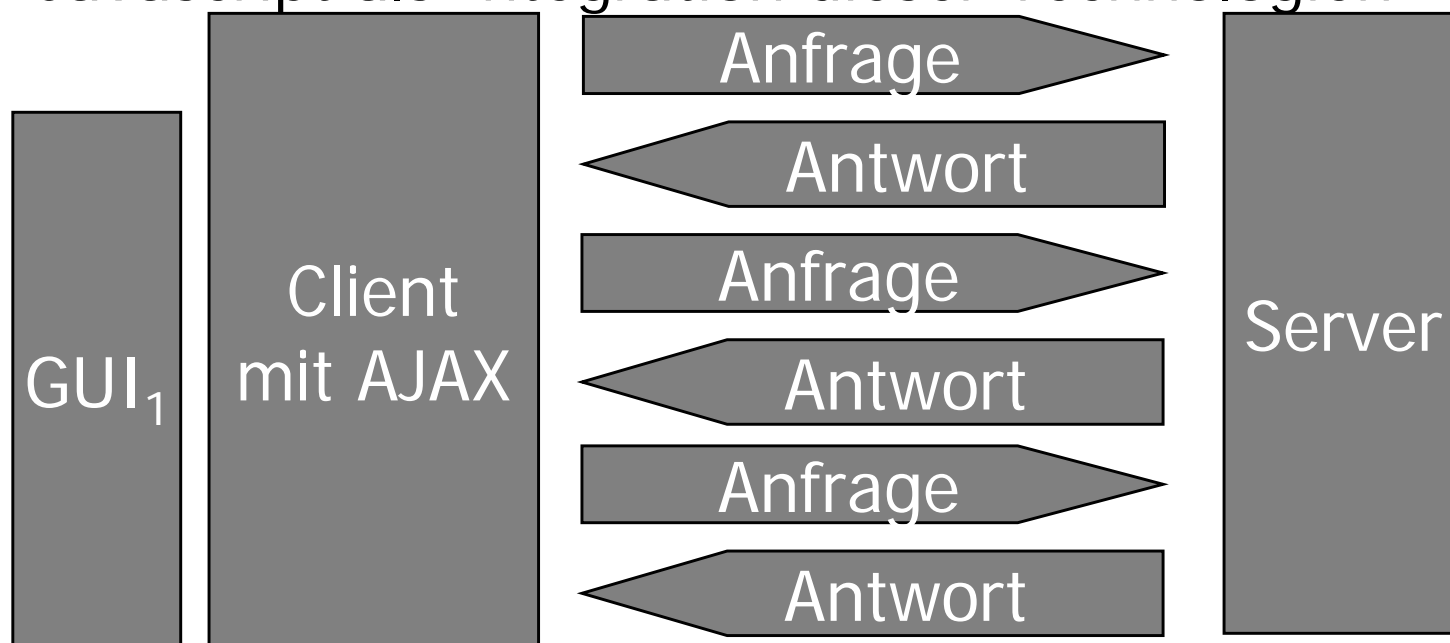
- Verlagerung von Anwendungsteilen zusammen mit GUI zum Browser/Clients
  - Applets
  - Javascript
  - ...
- Teile der Anwendung „laufen“ im Browser
- Sie kombinieren verschiedene Medien zu einer Oberfläche

# Zusammengesetzte Dokumente

---

- Compound Document by Reference Framework 1.0
  - W3C Working Draft 22 November 2006
  - <http://www.w3.org/TR/CDR/>
- „Compound Document“: Dokument das aus mehreren Teildokumenten in unterschiedlichen Auszeichnungssprachen zusammengesetzt ist
  - XHTML + SVG + MathML
  - XHTML + SMIL
  - XHTML + XForms
  - XHTML + VoiceML
- Problem: Jeweilige Spezifikationen decken nicht alle Aspekte solcher Kompositionen ab
- Compound Document by Reference Framework, CDRF ist generischer Rahmen für die Verarbeitung solcher Dokumente

- Asynchronous JavaScript and XML (AJAX) realisiert dies durch Kombination von
  - Präsentationssprachen XHTML und CSS
  - Interaktion und Modifikation im Browser mit DOM
  - Datenaustausch mit XML
  - Datentransfer durch asynchrone HTTP-Anfragen
  - Javascript als Integration dieser Technologien





## Metadaten, Semantic Web

# Metadaten in HTML: Dublin Core

---

- "Dublin Core" ([http://purl.org/metadata/dublin\\_core](http://purl.org/metadata/dublin_core)) ist der Versuch, ein verbreitetes Schema für Metadaten zu etablieren:
- "The Dublin Core Metadata Initiative is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. DCMI's activities include consensus-driven working groups, global workshops, conferences, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices."
- Dublin Core Metadata Element Set, 1.1: Reference Description
  - Quelle: <http://dublincore.org/documents/1999/07/02/dces/>
  - Status: DCMI Recommendation 1999-07-02
- Dublin Core Metadata for Resource Discovery
  - Status: IETF RFC 2413, September 1998
- Encoding Dublin Core Metadata in HTML
  - Status: IETF RFC 2731, December 1999

## Weitere Informationen

- Geschachtelte Instanz
  - „Brun Hilda is a visiting lecturer here from Germany who doesn't have her own web page. However, because I am such a nice person, I have agreed to let part of my web page space belong to her. She is 23.“
  - <INSTANCE  
 KEY="http://www.cs.umd.edu/users/george/#BRUNHILDA">  
 <CATEGORY NAME="cs.Lecturer">  
 <RELATION NAME="cs.name">  
   <ARG POS=TO VALUE="Brun Hilda">  
 </RELATION>  
 <RELATION NAME="cs.age">  
   <ARG POS=TO VALUE="23">  
 </RELATION>  
 </INSTANCE>

# Semantic Web

---

- "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation"

[Tim Berners-Lee, James Hendler und Ora Lassila: The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, Scientific American, May 17, 2001]

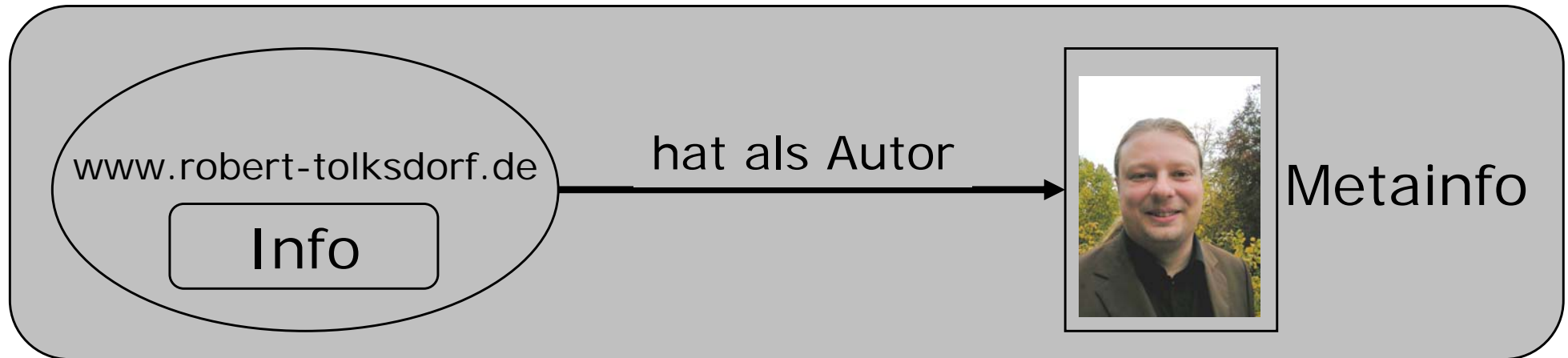
- Explizite Repräsentation von Semantik mit Sprachen
- Genauer:  
Weniger Missverständnisse wegen besserem Kontextbezug
- M2M vs. M2H Kommunikation

# Web Sprachen

- XML und XML Schema
  - DTD definiert Vokabular
  - Rein syntaktische Festlegung, keine Semantik
- RDF
  - Modelliert Daten nach E/R, Objekt/Attribut
  - Keine semantischen Festlegungen
- RDF Schema
  - Festlegungen durch Typisierung (Subklassen, Domain, Range), erweiterbar
  - IS-A Hierarchie ableitbar
- OWL
  - Erweiterte Beschreibungsmöglichkeiten
  - In unterschiedlichen Mächtigkeiten/Komplexitäten (OWL-Lite, OWL-DL, OWL-Full)
- Alle Sprachen sind Standards des W3C
- + Regelsprachen (ORL, SWRL, RuleML,...)
- + Prozessbeschreibungssprachen (OWL-S, WSMO,...)

# RDF Sätze

- Informationen und Metainformationen:



- In RDF als Satz ausgedrückt:

"www.robert-tolksdorf.de	Subjekt
hat als Autor	Prädikat
Robert Tolksdorf"	Objekt

## In RDF definiert

- `<?xml version="1.0"?>`

`<RDF xmlns=`

`"http://www.w3.org/1999/02/22-rdf-syntax-ns"`

`xmlns:s="http://description.de/schema/">`

`<Description about=`

`"http://www.robert-tolksdorf.de">`

`<s: Autor>Robert Tolksdorf</s: Autor>`

`</Description>`

`</RDF>`

Subjekt

Prädikat

Objekt

- Aus so explizit gemachten und maschinenverständlich repräsentierten Aussagen können Tools und Dienste inhaltliche Schlüsse ziehen

# Datenmodell

---

- „RDF-Welt“: Gerichteter Graph
  - Knoten (Ressourcen)
  - Kanten (Properties)
- Ressourcen (RDF Resource)
  - Alles worüber man Aussagen machen kann
  - Identifiziert durch URIs (qualified URIs = URI + fragment identifier)
  - Aussagen sind auch Ressourcen
- Eigenschaften/Beziehungen (RDF Property)
  - Verbinden Ressourcen miteinander oder Ressourcen zu Werten (RDF Literal)
- Aussagen (RDF Statement)
  - (Subjekt, Prädikat, Objekt)
  - “Resource has Property with Value”

# Container-Typen

- 3 Typen von Containern
  - Bag                      ungeordnete Liste              rdf: Bag
  - Sequence              geordnete Liste              rdf: Seq
  - Alternative              eindeutiger Wert              rdf: Alt
- Containers sind auch RDF Ressourcen
- Semantik: offene Mengen
  - unbekannt ob weitere Elemente zu der Menge gehören

# RDF Collections

---

- Ähnlich zu Containern, aber geschlossen:
  - *Alle* Elemente einer Menge sind spezifiziert
- Zugriff auf einzelne Mengenelemente
  - rekursiv
  - first (erstes Element)
  - rest (restliche Elemente)
  - nil (leere Menge)

# Beispiel XML/RDF

```
<rdf:RDF>
```

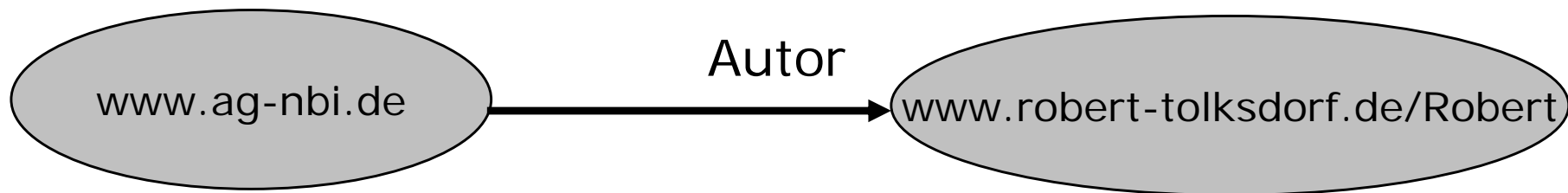
```
<rdf:Description about="http://www.ag-nbi.de">
```

```
<Autor
```

```
  rdf:resource="http://robert.tolksdorf.de/Robert" />
```

```
</rdf:Description>
```

```
</rdf:RDF>
```



# RDF Schema

---

- Mit den grundlegenden RDF Mechanismen lassen sich einfache Aussagen auf vielfältige Weise treffen
- Mit RDF Schema werden einige Typen von Aussagen eingeführt, mit denen Schemas möglich werden, mit denen nützliche Modellierungsaussagen getroffen werden können:
  - "Jede Webseite hat einen Autor"
  - "Webseiten sind elektronische Dokumente"

# RDF Schema

- Elektronischen Dokumente bilden eine Klasse:  

```
<rdf: Description rdf: ID="el ectroni cDocument" >  
<rdf: type rdf: resource=  
    "http: //www. w3. org/2000/01/rdf-schema#Cl ass" />  
</rdf: Descri pti on>
```
- Web-Seiten sind elektronische Dokumente  

```
<rdf: Descri pti on rdf: ID="webPage" >  
<rdf: type rdf: resource=  
    "http: //www. w3. org/2000/01/rdfschem a#Cl ass" />  
<rdfs: subCl assOf rdf: resource="#el ectroni cDocument" />  
</rdf: Descri pti on>
```
- Web-Seiten haben eine URL  

```
<rdf: Property rdf: ID="URL" >  
<rdfs: domai n rdf: resource="#webPage" />  
<rdfs: range rdf: resource=  
    "http: //www. w3. org/2001/XMLSchema#stri ng" />  
</rdf: Property>
```

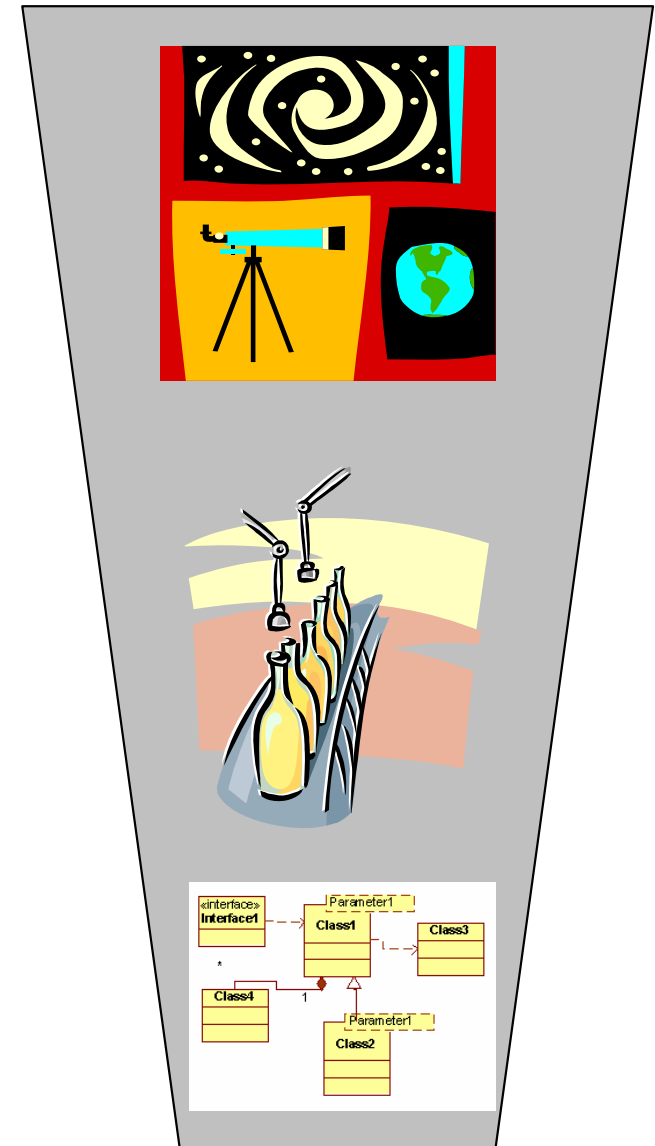
# Ontologie

---

- Ist Beschreibung einer Wissensdomäne mit
  - Standardisierter Terminologie (Klassen, Axiome etc) für Konzepte
  - Beziehungen zwischen Konzepten
  - Ableitungsregeln
- Ziel:  
Einschränkung der Interpretationsmöglichkeiten von Symbolen
- Dadurch: Gemeinsame Sprache
- Dadurch: Wissensaustausch möglich

# Arten von Ontologien

- Top Level Ontologien
  - Domänenüberschreitend
  - Allgemeine Konzepte
  - Person, Mensch, Tätigkeit, Artefakt
  
- Domain Ontologien
  - Auf bestimmten Bereich bezogen
  - Dozent, Veranstaltungsbesuch, Schein
  
- Als Ontologien verpackte Daten- und Klassenmodelle



# Klassenbildung

- Klasse **Car**:

```
<owl:Class rdf:ID="Car">
  <rdfs:comment>no car is a person</rdfs:comment>
```

- Ist Unterklasse einer anonymen Klasse und erfüllt auch deren Eigenschaften:

```
<rdfs:subClassOf>
  <owl:Class>
    <owl:complementOf rdf:resource="#Person"/>
  </owl:Class>
</rdfs:subClassOf>
</owl:Class>
```

# OWL Varianten

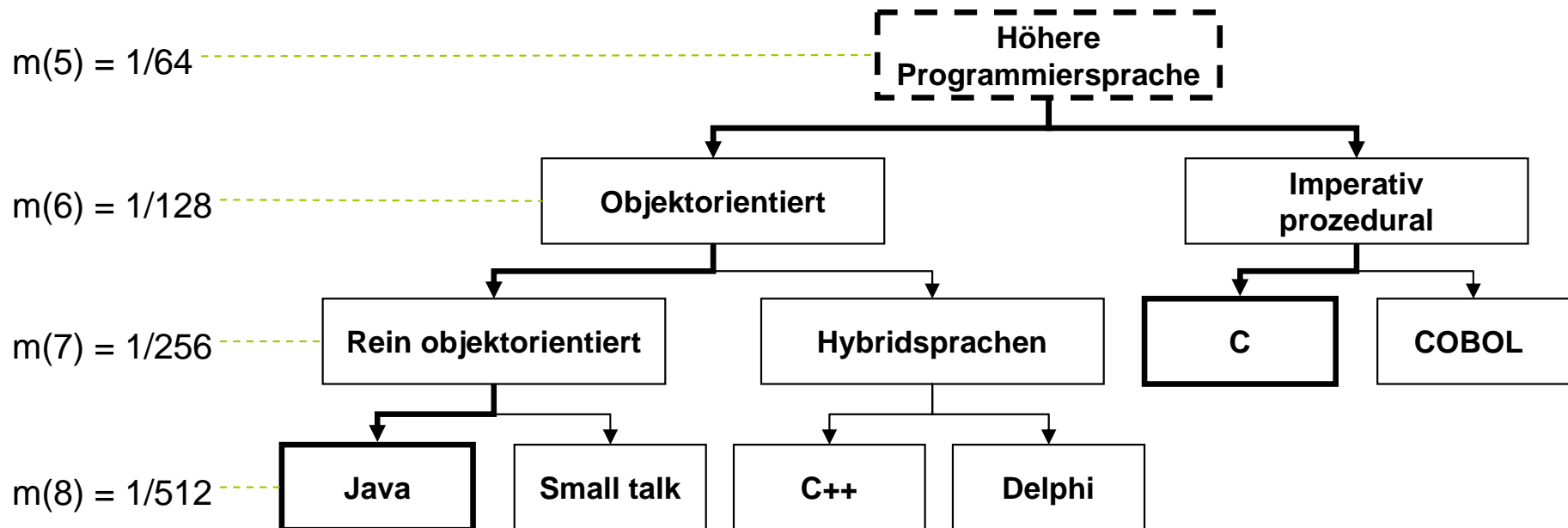
- OWL-Full
  - Volle Modellierungsmächtigkeit  
(z.B. Individuen können gleichzeitig als Klassen gelten)
  - Nicht entscheidbar!!
- OWL-DL
  - OWL-Full + Trennung zwischen Individuen und Klassen + Einschränkung des RDF-Vokabulars + Einschränkungen auf Kardinalitäten + ...
  - Entscheidbar in nicht-exponentieller Zeit
- OWL-Lite
  - OWL-DL + weitere Einschränkungen auf Kardinalitäten + ...
  - Entscheidbar in exponentieller Zeit

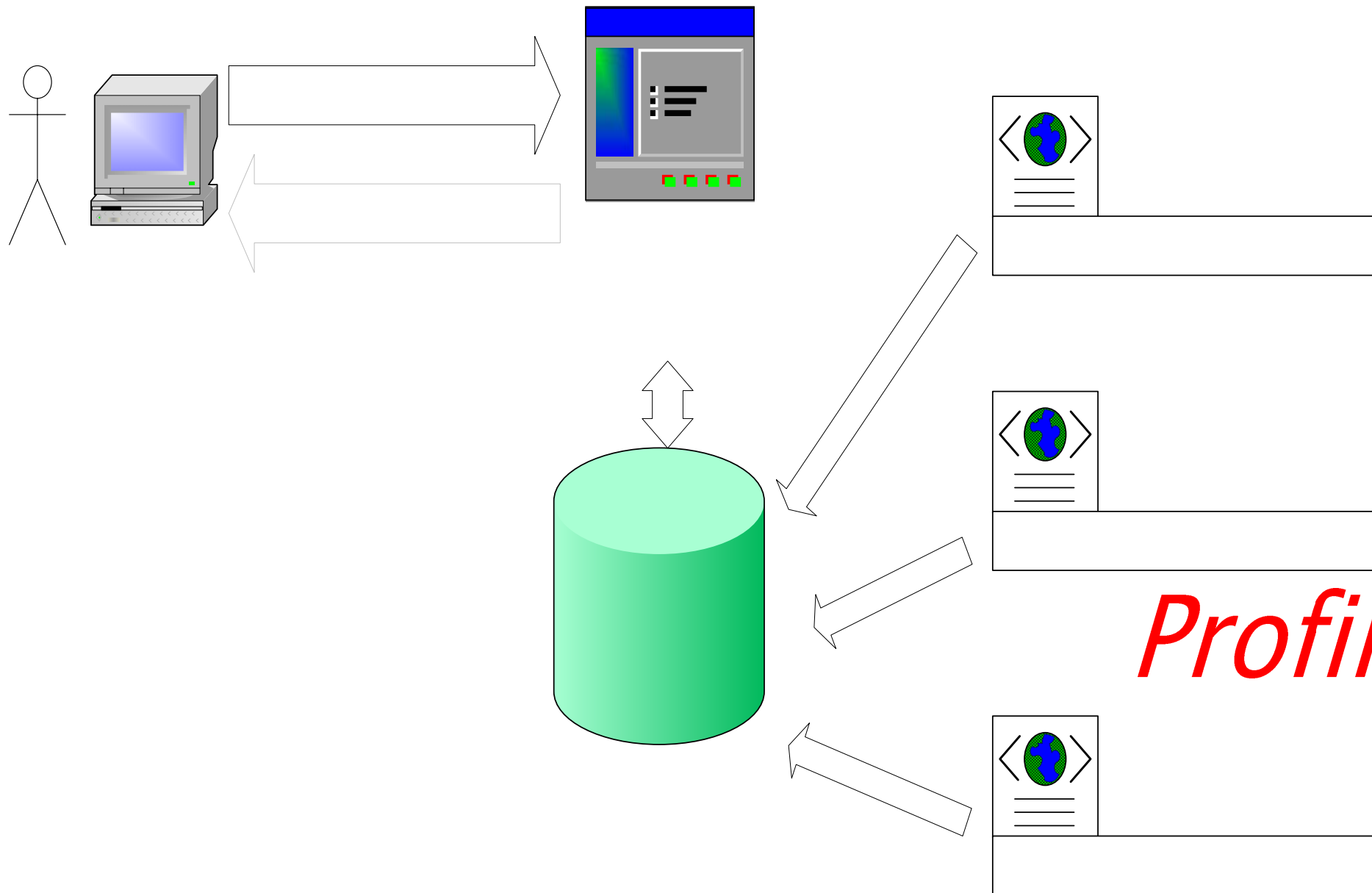
- In Konsequenz Ersetzung des Freitexts des Stellenangebots durch RDF Beschreibung:

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:rdf="..#" xmlns:jpp="..#" xmlns:skills="..#">
  <jpp:JobPositionPosting
    rdf:about="#JobPositionPostingId-inf-44">
    <jpp:hasHiringOrganisation>
      <org:Organisation>
        <org:name>Freie Universität Berlin</org:name>
      </org:Organisation>
    </jpp:hasHiringOrganisation>...
    <jpp:requiredCompetence>
      <skills:Java>
        <skills:hasCompetenceLevel rdf:resource="..#expert"/>
      </skills:Java>
    </jpp:requiredCompetence>...
  </jpp:JobPositionPosting>...
</rdf:RDF>
```

# Prototyp: Semantic Matching

- Taxonomische Struktur als Basis zur Errechnung semantischer Abstände
- Ähnlichkeit zwischen zwei Konzepten  $c_1$  and  $c_2$  ist bestimmt durch ihren Abstand  $d_c(c_1, c_2)$ . Dieser drückt ihre Position in der Konzepthierarchie aus
  - $d_c(\text{Java}, \text{C}) = (1/64 - 1/512) + (1/64 - 1/256) = 13/512$
  - $\text{sim}_c(\text{Java}, \text{C}) = 1 - 13/512 = 0,97$

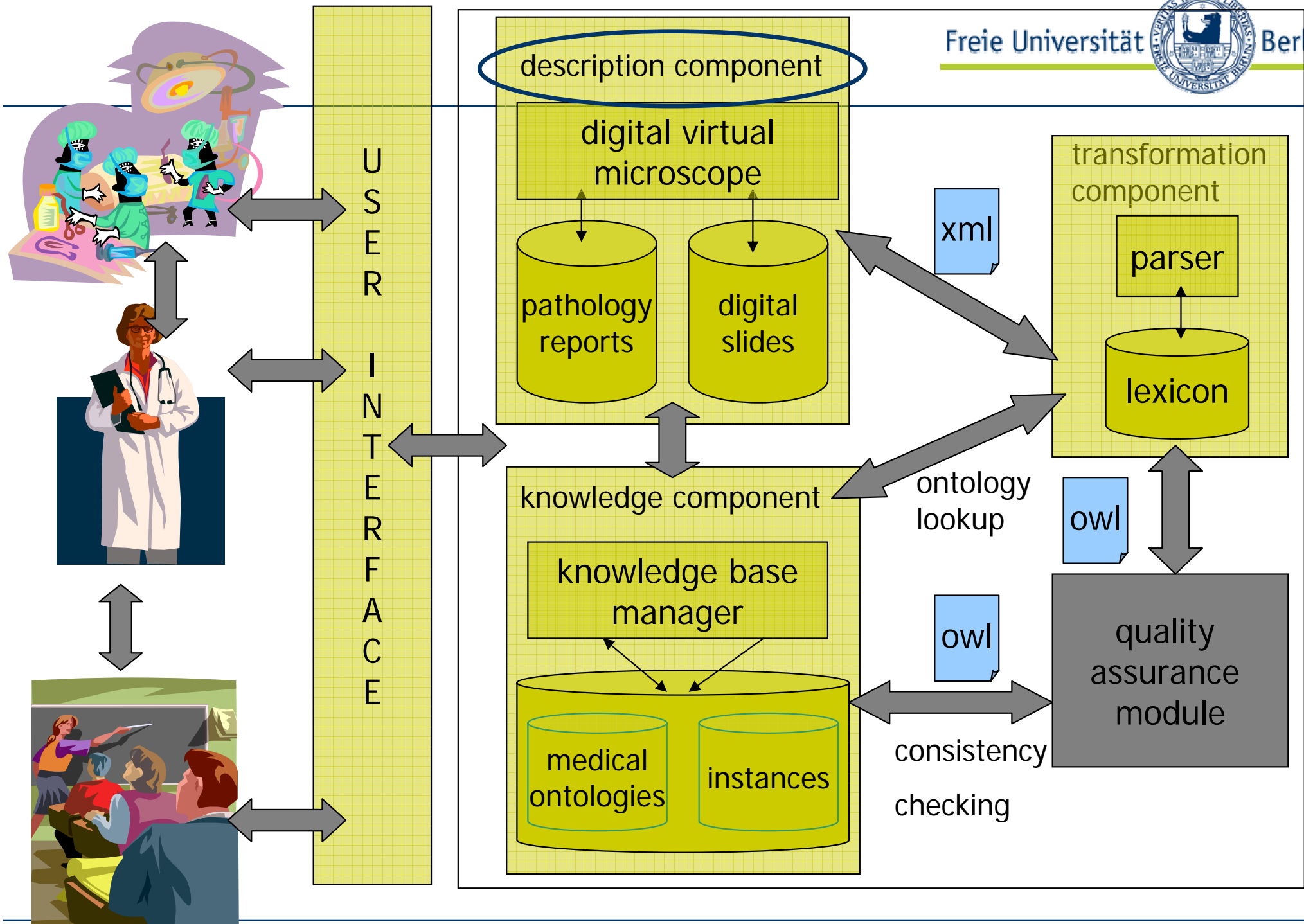




# Passende Hotels: nicht der Preis alleine bestimmt

## Gefundene Hotels

Hotel	<u>Preis EZ/Nacht</u>	<u>Gesamtwertung</u>	<u>Wellness</u>	<u>Sport</u>
<b>Melia White House</b> Albany Street, Regents Park NW1 3UP LONDON <a href="#">DETAILS</a>	99.0 €	90,0 %	80,0 %	100,0 %
<b>Chelsea Village Hotel</b> Stamford Bridge SW6 1HS London <a href="#">DETAILS</a>	89.0 €	75,0 %	100,0 %	50,0 %
<b>Somerset Bishopsgate</b> 196 Bishopsgate EC2M 4NR London <a href="#">DETAILS</a>	150.0 €	65,0 %	80,0 %	50,0 %
<b>Jurys Clifton Ford</b> 47 Welbeck Street WIG 8DN London <a href="#">DETAILS</a>	104.0 €	62,5 %	100,0 %	25,0 %
<b>Hilton London Islington</b> 53 Upper Street N1 0UY London <a href="#">DETAILS</a>	109.0 €	57,5 %	40,0 %	75,0 %
<b>Jurys Inn London</b>				



Suchbegriffe:

lunge

Konzepte suchen

Konzept passt zu:

mind. 1 Begriff

allen Begriffen

genauem Wortlaut

Ergebnisse

lunge lunge lunge

passende Konzepte:

- bulloeses\_lungenemphysem
- Lunge
- Lungen-PE
- Lungenabszess
- Lungenarterienembolie
- Lungenbiopsat
- Lungenemphysem
- Lungenerkrankung
- Lungenfibrose
- Lungenfilia
- Lungengefaess
- Lungengewebe
- Lungeninfarkt
- Lungenkarzinom
- Lungenlappen

Verknuepfung

Oder

Und

Sektionen

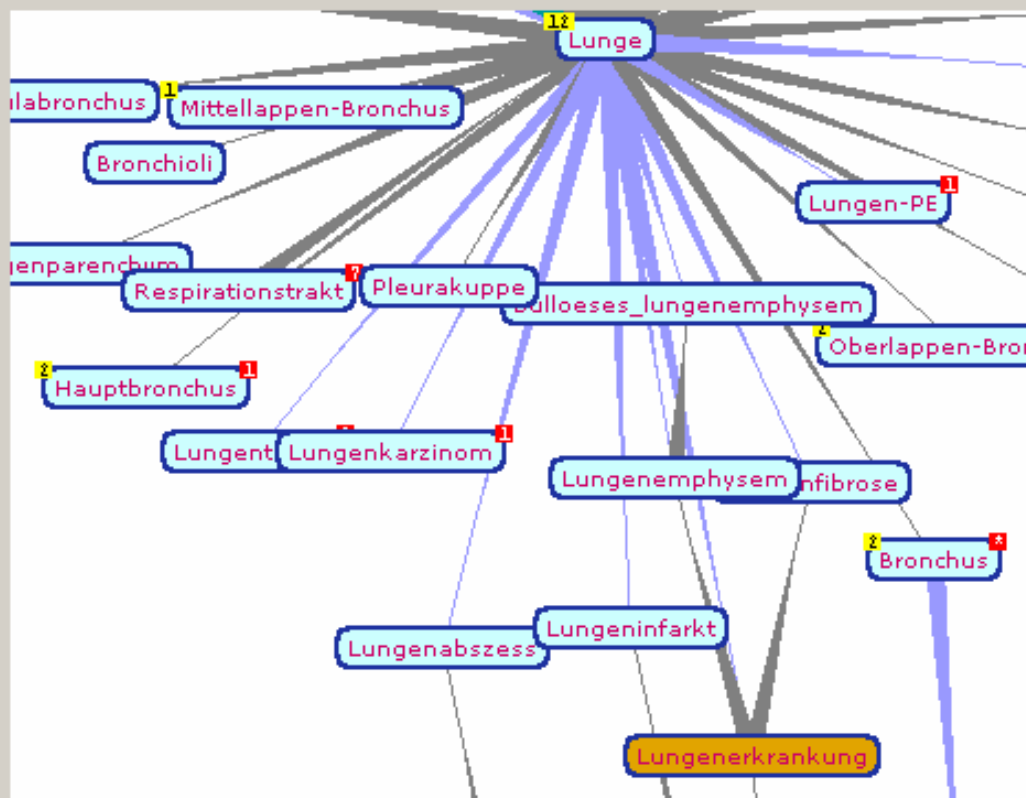
Mikroskopie

Makroskopie

Diagnose

Kommentar

Befunde suchen



logien

WPatho2

Befundbericht

Immunhistologie

MLSSN

WPatho UpperLevel

Beziehungen

Is a

Physical part of

Conceptual part of

Location of

Has location

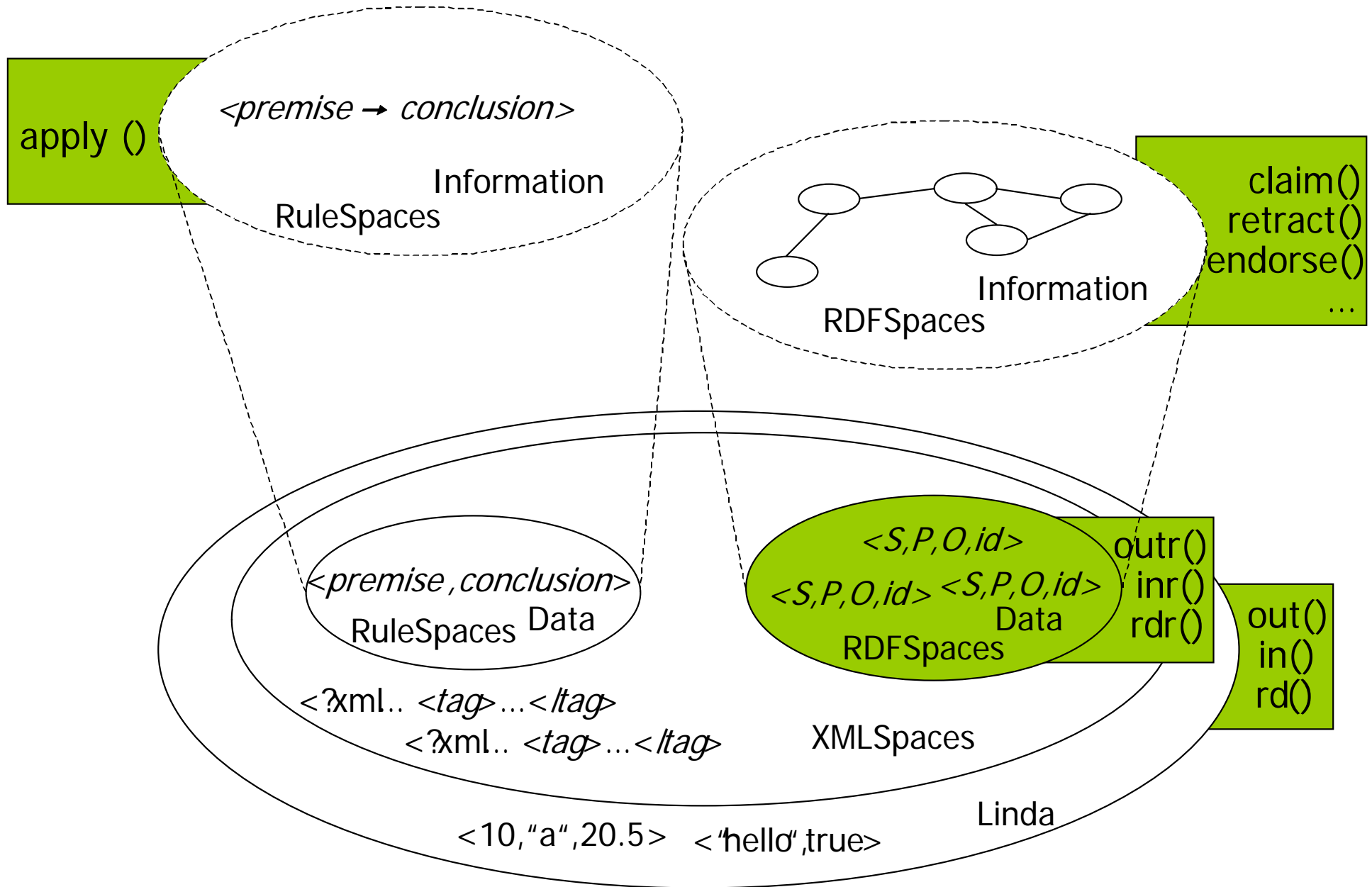
Steuerung

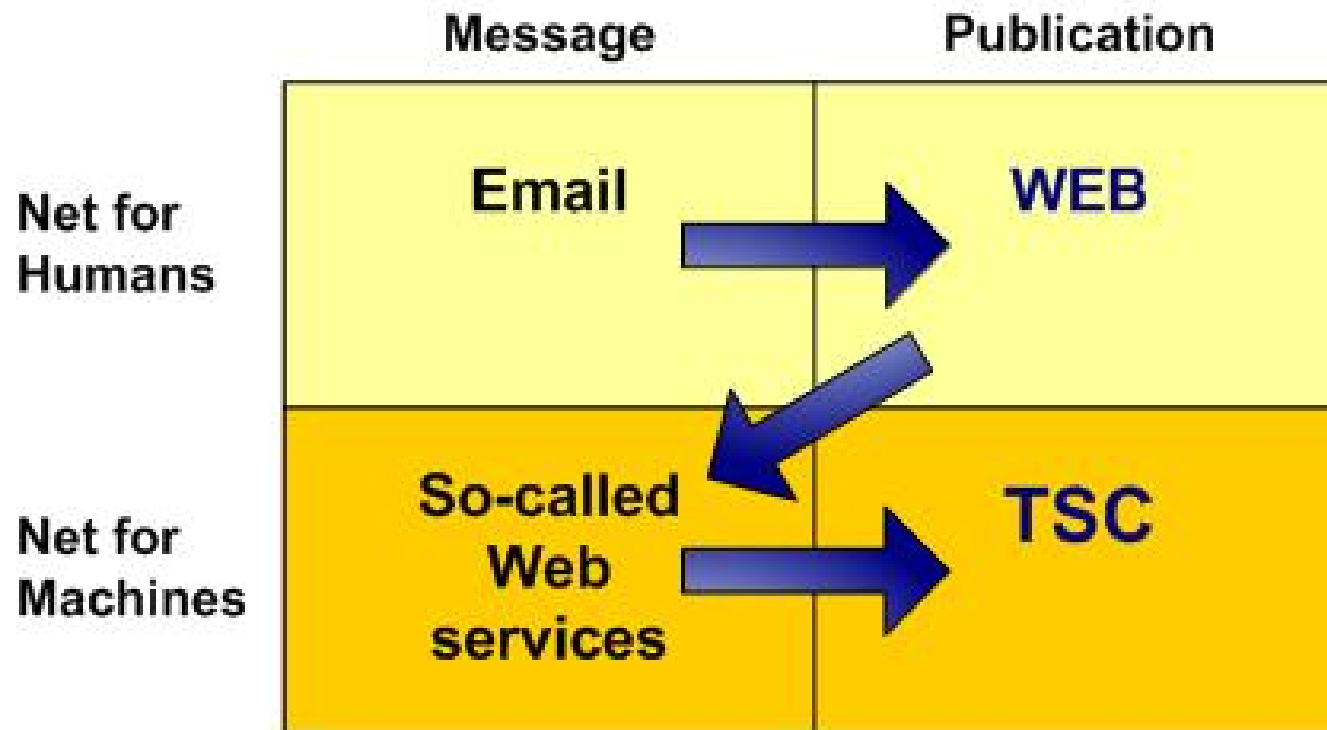
Zoom

Rotation

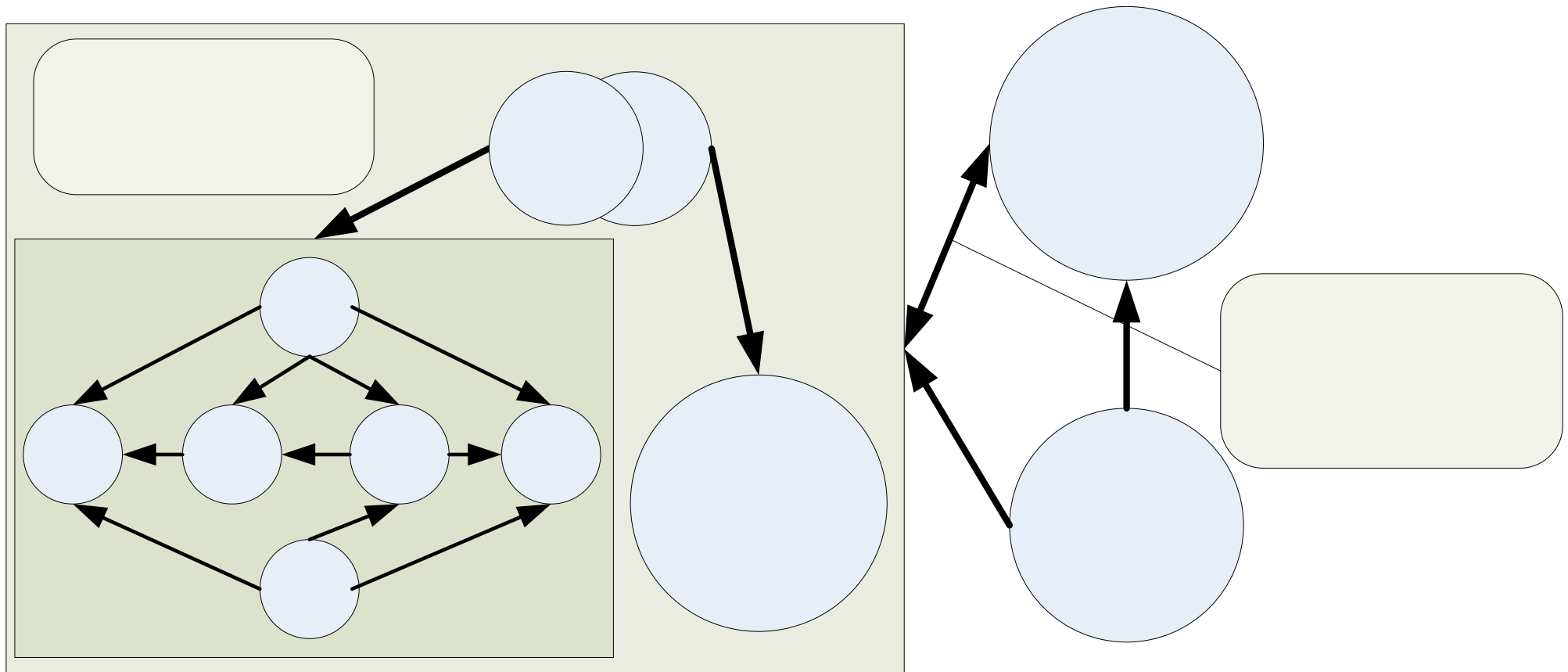
Hyperbolik

# Data and information views





"Triple Space may become the web for machines as the web based on HTML became the Web for humans"

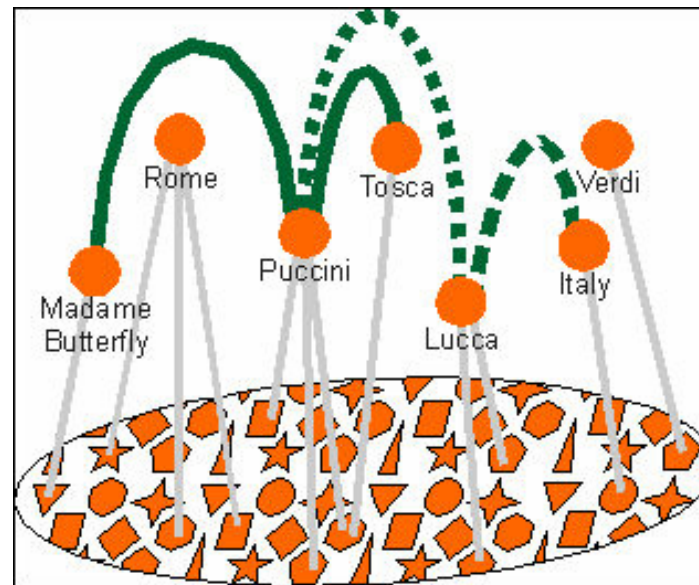
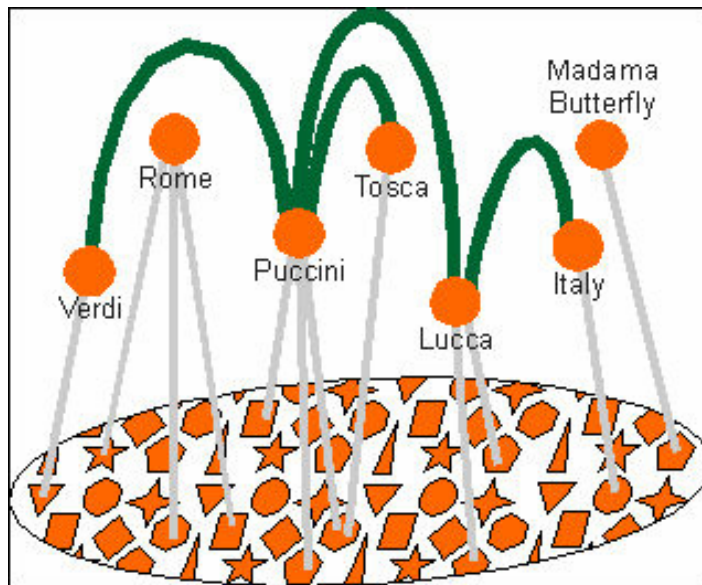


# Topic Navigation Maps...

- ... Repräsentation von Semantik von Informationen
- Information technology - SGML Applications - Topic Maps
  - Ziel: Adressierbare Information semantisch zugänglich machen
  - Status: ISO/IEC 13250:2000
  - Quelle: <http://www.topicmaps.net/>,  
<http://www.y12.doe.gov/sgml/sc34/document/0129.pdf>
- XML Topic Maps (XTM) 1.0
  - Ziel: XML-basiertes Austauschformat für Topic Maps
  - Status: TopicMaps.Org
  - Quelle: <http://www.topicmaps.org/xtm/1.0/>

# Associations

- Sind Beziehungen zwischen Topics
  - "Tosca was *written by* Puccini"
  - "Tosca *takes place in* Rome"
  - "Puccini was *born in* Lucca"
- Können typisiert sein



- Association role: Welche Rolle spielt Topic in Association

- Gültigkeitsbereiche durch `<scope>`:  
`<scope>`  
  `<subjectIndicatorRef xlink:href=`  
    `"http://www.topicsmaps.org/xtm/1.0/language.xtm#en" />`  
`</scope>`
- Namen von Topics:  
`<topic id="denmark" >`  
  `<baseName>`  
    `<scope><topicRef xlink:href="#en" /></scope>`  
    `<baseNameString>Denmark</baseNameString>`  
  `</baseName>`  
  `<baseName>`  
    `<scope><topicRef xlink:href="#da" /></scope>`  
    `<baseNameString>Danmark</baseNameString>`  
  `</baseName>`  
`</topic >`

- Viel Glück in der Klausur