



Semantic Web Anwendungen bei der NBI

Malgorzata Mochol
Freie Universität Berlin
Institut für Informatik
Netzbasierte Informationssysteme

[mochol\[at\]inf.fu-berlin.de](mailto:mochol[at]inf.fu-berlin.de)
<http://page.mi.fu-berlin.de/~mochol/>

- Projekt „Wissensnetze“ - Online-Stellenvermittlung mit Semantic-Web-Technologien
 - Online Recruitment Heute
 - Online Recruitment mit Semantic Web Technologien
 - HR-Szenario
 - Klassifikationen + HR-Ontologie
 - Matching
 - Architektur
 - Auswirkungen
- Projekt „Reisewissen“
 - Heutige Probleme bei der online Hotelsuche
 - Lösung mit Semantic Web
 - Komponenten
 - Datenintegration



Wissensnetze

Online-Stellenvermittlung mit Semantic-Web-Technologien

- **Online Recruitment - Hauptkanal der Stellenvermittlung**
 - 47 % der Internet-User (28% der Europäischen Internet User) lesen die Online-Stellenangebote*
- **Vielfalt an Möglichkeiten**
 - Klassische Seiten der Online Vermittler (www.jobpilot.de, www.monster.de)
 - Verschiedene Meta-Job Portale
 - Seiten der staatlichen Arbeitsvermittlungen (www.arbeitsagentur.de, www.ams.se)
 - verschiedenen Seiten der einzelnen Firmen
- **Schwierig, den Überblick zu behalten**



* "Daten für das Online-Recruitment I/2005" wwjGmbH

e-Recruitment - Probleme

- Stellenangebote in Freitext mit unkontrolliertem Wortschatz
- Suche auf Volltext-Basis
- Viele unterschiedliche Taxonomien für Klassifizierung von Jobangeboten
- Unpräzise Indexierung von Stellenangeboten:
 - Probleme in der Erkennung von Jobangeboten im Web
 - Probleme mit der Extrahierung von relevanten Stichworten (Berufsbezeichnung, Qualifikationen, ...)
- Begrenzte Möglichkeit, die Angebote mit den angegebenen Kriterien zu vergleichen

- Suchmaschinen statt Job-Portalen:

- Allgemeine Suchmaschinen wie Google
 - Suchmaschine – Hauptquelle der Suche bei 74% der Internet-User*
 - Google - 41,1% der Internet User benutzt Google*
- Spezialisierte Suchmaschinen wie WorldWideJobs**



* “Daten für das Online-Recruitment I/2005” wwjGmbH

** www.wwj.de

- **Organisatorisch:**
 - Stellenanbieter nutzen gemeinsames kontrolliertes Vokabular für die Annotierung von Stellenangeboten
 - Stellensuchende nutzen gleiches Vokabular für Stellengesuche/Bewerberprofile
- **Technisch:**
 - Einfache Annotation → Reichere Annotation → Ersatz von Freitext durch RDF
 - Stelleangebote direkt auf der Web-Seite des Unternehmens
 - Semantische Suchmaschinen :
 - sammeln Informationen
 - bieten Vergleich auf Basis von semantischen Informationen an (Semantic Matching)

Einfache Annotation

- Mit RDF und Bezug auf gemeinsames Vokabular (z.B. abgeleitet von HR-XML)

```

<html>
  <head>
    <rdf:RDF xmlns:rdf="...#" xmlns:jpp="...#" >
      <jpp:JobPosting
        rdf:about="http://www.example.org/jp1.html"/>
    </rdf:RDF>
  </head>
  <body>...Job posting in free text...
  </body>
</html>

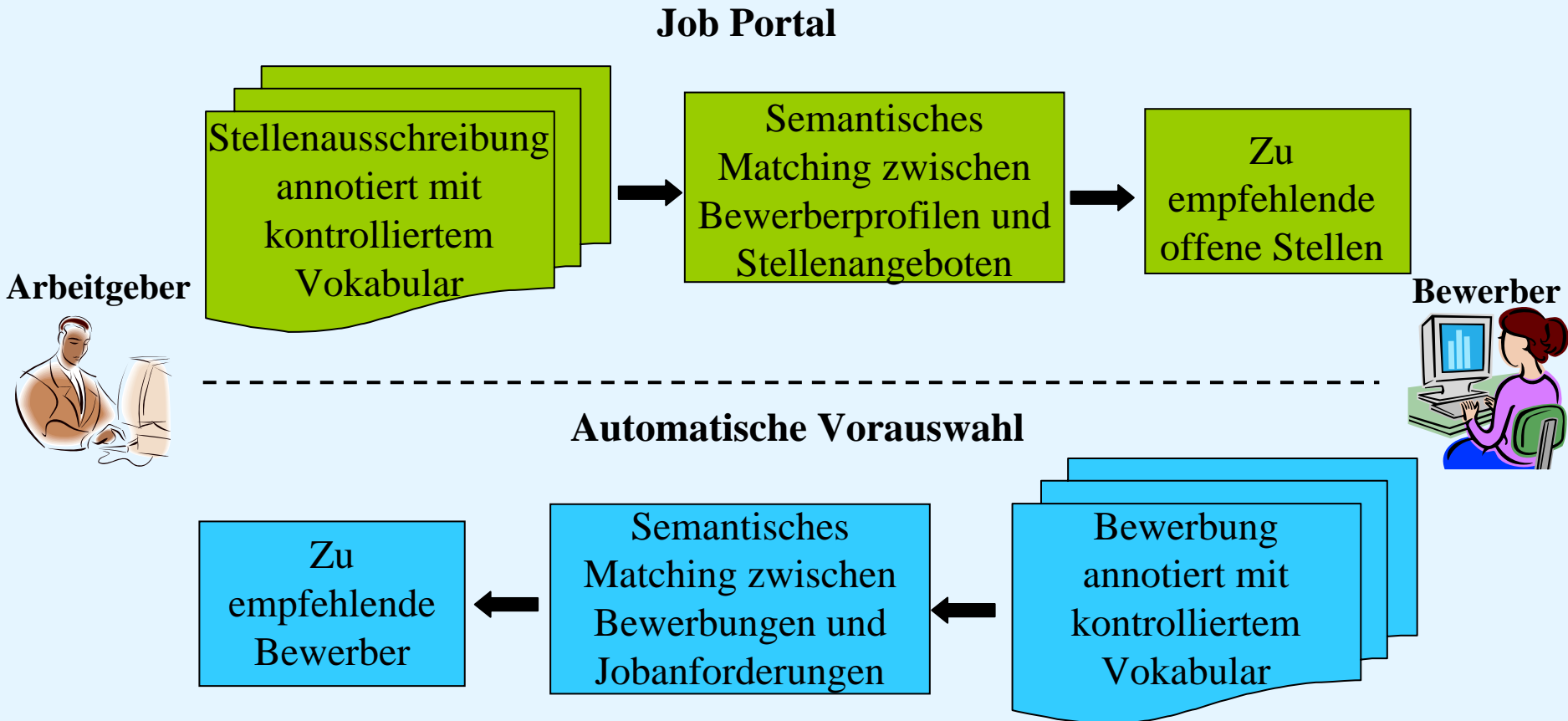
```

- Suchmaschinen können so Stellenangebote identifizieren

Reichere Annotation - Beispiel

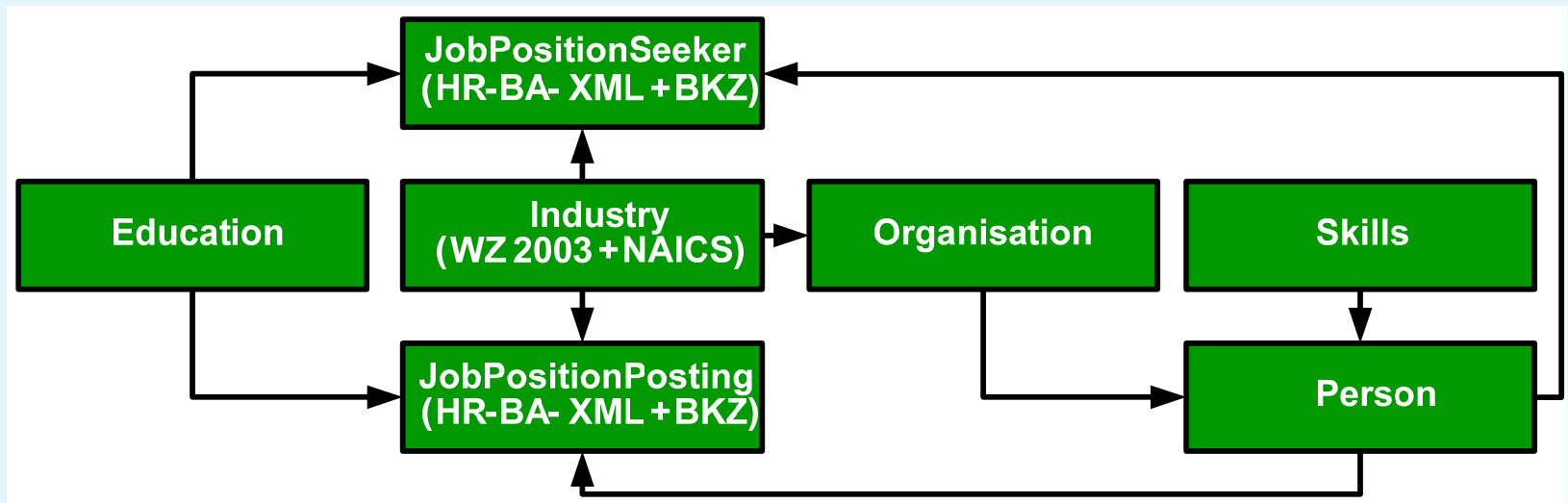
```
<html>
  <head>
    <rdf:RDF xmlns:rdf="...#" xmlns:jpp="...#" xmlns:skills="...#">
      <jpp:JobPositionPosting
        rdf:about="http://www.example.org/jp1.html"/>
      <jpp:requiredCompetence>
        <skills:Java>
          <skills:hasCompetenceLevel rdf:resource="...#expert"/>
        </skills:Java>
      </jpp:requiredCompetence>
    </rdf:RDF>
  </head>
  <body>
    ... Job posting in free text ...
  </body>
</html>
```

```
<?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>
```



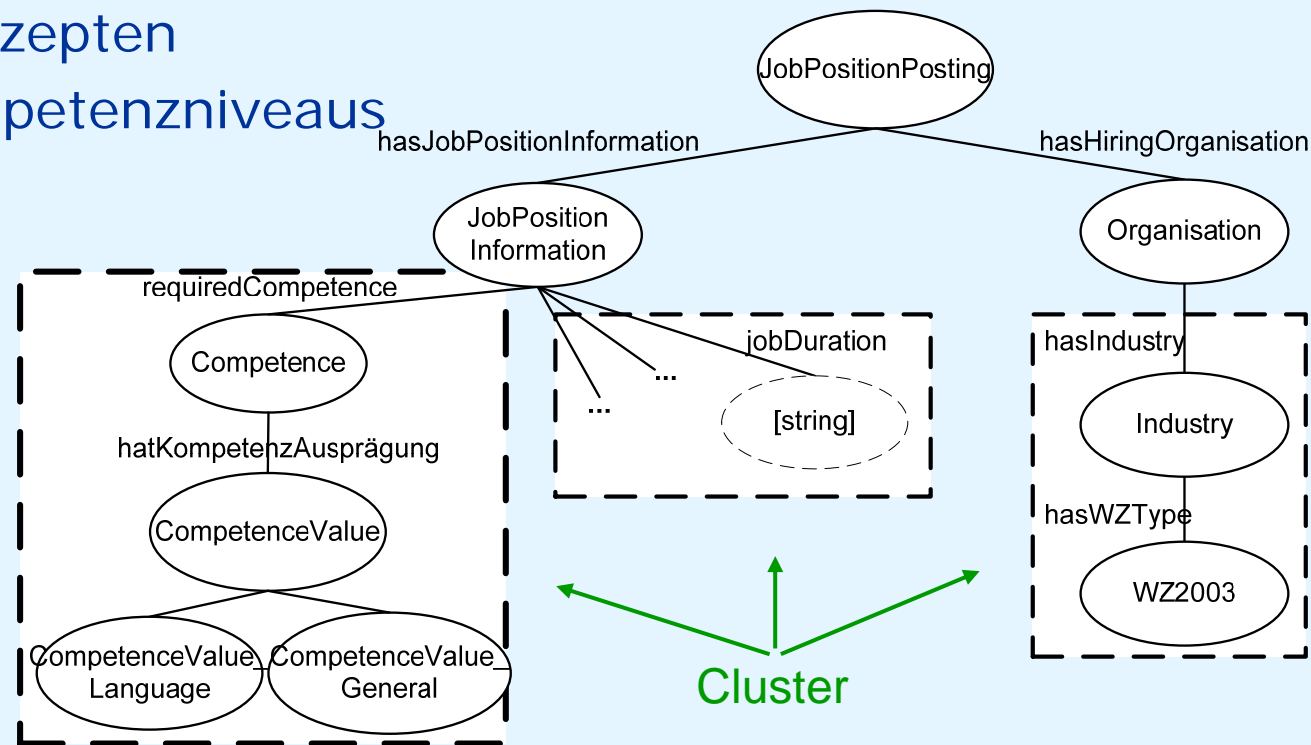
- Existierende Spezifikationen und Ontologien im Human Resource (HR) Sektor
 - Standard Occupation Classification (SOC) System
 - Berufskennziffer (BKZ) – Bundesagentur für Arbeit
 - Wirtschaftszweige 2003 (WZ2003) – Statistisches Bundesamt
 - North American Industry Classification System (NAICS)
 - HR-XML – HR-XML Konsortium
 - HR-BA-XML – deutsche Version der HR-XML von der Bundesagentur für Arbeit
 - HR-XML-SE – Schwedische Version
 - Qualifikations-Ontologie – Projekt KOWIEN

- Abgeleitet (Wiederverwendung) von existierenden Standards
- Definiert Begriffe für Beschreibungen von:
 - Qualifikationen (Skill)
 - Stellenangeboten (JobPositionPosting)
 - Bewerberprofilen (JobPositionSeeker)
 - Firmen (Industry) ...
- Benutzt JobPositionPosting and JobPositionSeeker von [HR-BA-XML](#)



HR-Szenario – Matching

- **Semantisches Matching** – Berechnung der semantischen Ähnlichkeit
(kontrollierter Wortschatz + Hintergrundwissen)
- Gesamtähnlichkeit = Summe von Cluster-Ähnlichkeiten
- Konzeptähnlichkeit bestimmt durch Entfernung zwischen zwei Konzepten
- Vergleich von Kompetenzniveaus



- **Semantic Matching Engine** - Berechnung von Ähnlichkeiten zwischen Bewerberprofilen und Stellenangeboten
- **Input:**
 - HR Ontologie
 - RDF-Bewerberprofil
 - RDF-Stellenausschreibungen
- **Output:**
 - Rankingliste
 - Bewertung (wie gut passt das Stellenangebot zu dem Bewerber)
- **Matching Engine unterstützt:**
 - Verschiedene Matching-Ansätze
 - Kompetenzniveau (Anfänger ... Profi, Grundkenntnisse ... Muttersprachler)
 - Flexible Konfiguration

- Suchmaschinen können
 - zuverlässiger crawlen
 - Sematic Matching benutzen um die Genauigkeit von Vergleichen zu optimieren
- Jobsuchende profitiert von
 - der erhöhten Markttransparenz
 - der effizienteren und schnelleren Jobvermittlung
 - den exakteren Suchergebnissen
- Arbeitgeber kann
 - Semantic Matching für die Automatisierung der Vorauswahl-Phase benutzen
 - Angebote nur auf eigener Firmen-Webseite publizieren

- Projekt „Wissensnetze“
 - <http://nbi.fu-berlin.de/research/wissenanetze>
- Berliner Forschungszentrum „InterVal“
 - <http://www.wiwi.hu-berlin.de/iwi/internetoekonomie>



- „Internetökonomie“
 - <http://www.internetoekonomie.net>





Projekt Reisewissen - Hotelsuche mit Semantic Web



Folien: Mangus Niemann

AG Netzbasierte Informationssysteme

Problemstellung

- Kunde will Hotel buchen → www.ehotel.de
 - Suchergebnis (z.B. London):
 - > 100 Hotels, nach Preis sortiert
 - „Welches Hotel soll ich wählen?“
 - „Bestes Hotel für mich“ ist nicht alleine durch den Preis definiert
 - „Werden meine Wünsche berücksichtigt?“
 - Nähe zu Tagungsort
 - Gute Erreichbarkeit
 - Passende Ausstattung
 - ...
- Übliche Lösung: aufwändige Suche im Internet, Call-Center, Reisebüros, etc.

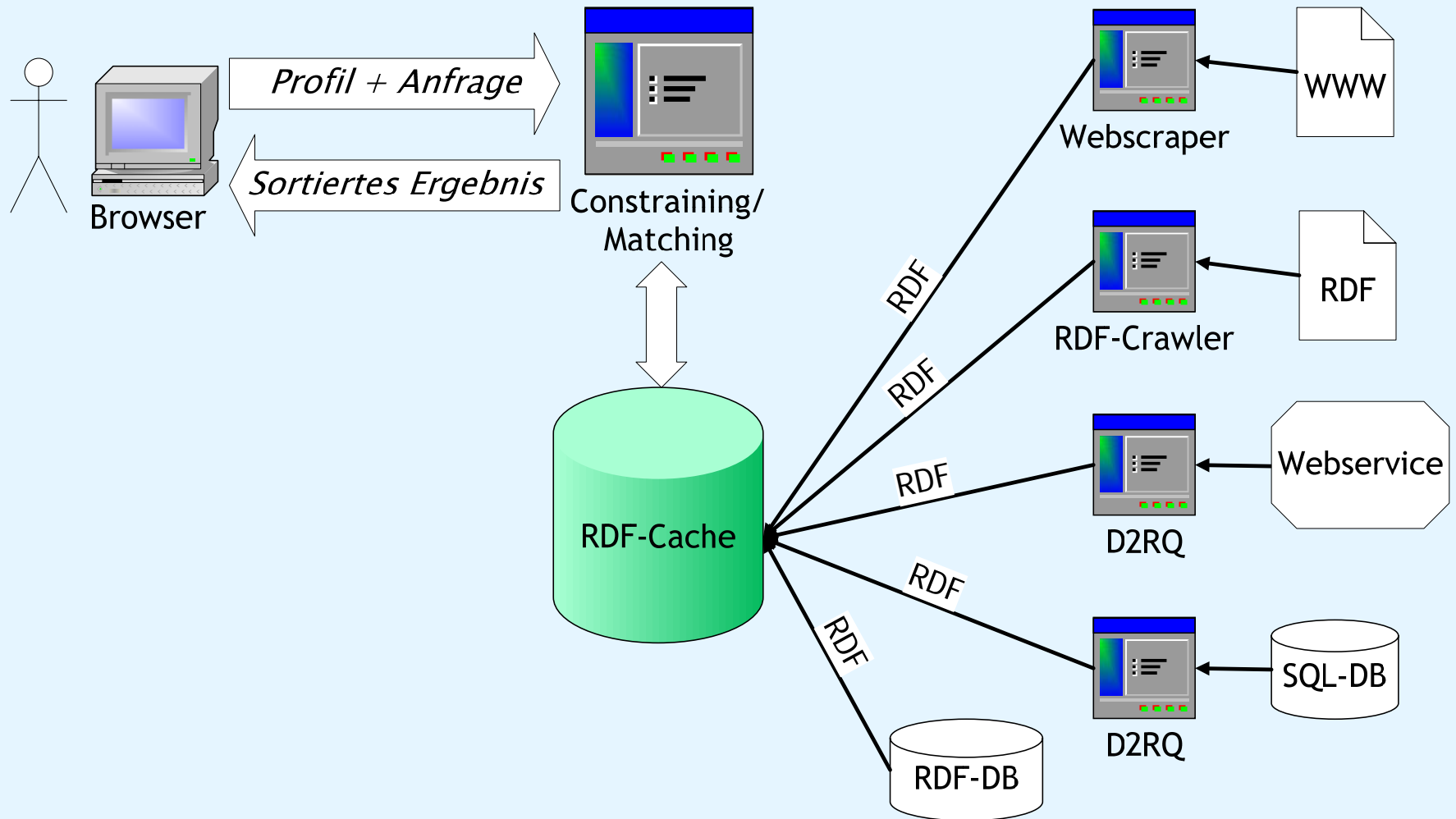
- Entwicklung einer „Hotelbewertungen“-Engine
- Zuhilfenahme von
 - Hotelinformationen
 - Kundenprofilen (Wünsche, Bedingungen)
 - Zusatzinformationen (Reiseführer, Bewertungen)
 - Expertenwissen
- Nutzen:
 - Zeit-/Kostensparnis
 - Qualitätsverbesserung

Ablauf einer Hotelsuche

- Kunde definiert sein Profil bestehend aus
 - Constraints
 - Präferenzen
- Kunde legt Reisedatum und –stadt fest
- Verfügbare Hotels für Datum/Stadt werden gesucht
- Ergebnismenge wird entsprechend der **Constraints** beschnitten
- Ergebnismenge wird entsprechend der **Präferenzen** sortiert

- **Browser-basierte Eingabe** von Kundenwünschen und -Bedingungen
- **Ontologie-basierte Maschine** für die Verarbeitung der Anfrage
- **RDF-Cache** zur Integration der Datenquellen
- Je Datenquelle einen **Übersetzer/Proxy** nach RDF

Schematischer Aufbau



Wo ist das Semantic Web?

- **Ontologien für**
 - Hoteldaten
 - Zusatzinformationen (z.B. POI)
 - Kundenprofile (Bedingungen, Präferenzen)
- **Regeln für Expertenwissen, z.B.**
 - „Ein Zwei-Sterne-Hotel bietet auch ein Frühstücksbüffet an.“
 - „Ein Hotel mit mehr als 200 Betten gilt auch in unmittelbarer Umgebung eines Bahnhofs als seriös“



Hotel-Suche

* Bei mit einem Stern markierten Feld ist eine Eingabe erforderlich.

Land *

Stadt *

Anreise *

Abreise *

Zimmeranzahl/-typ * /

Bedingungen (optional)

Hotel ist als km von entfernt.

LÖSCHEN

Preis ist als €/Nacht.

LÖSCHEN

HINZUFÜGEN

Profil auswählen (optional)

Hauptprofil auswählen

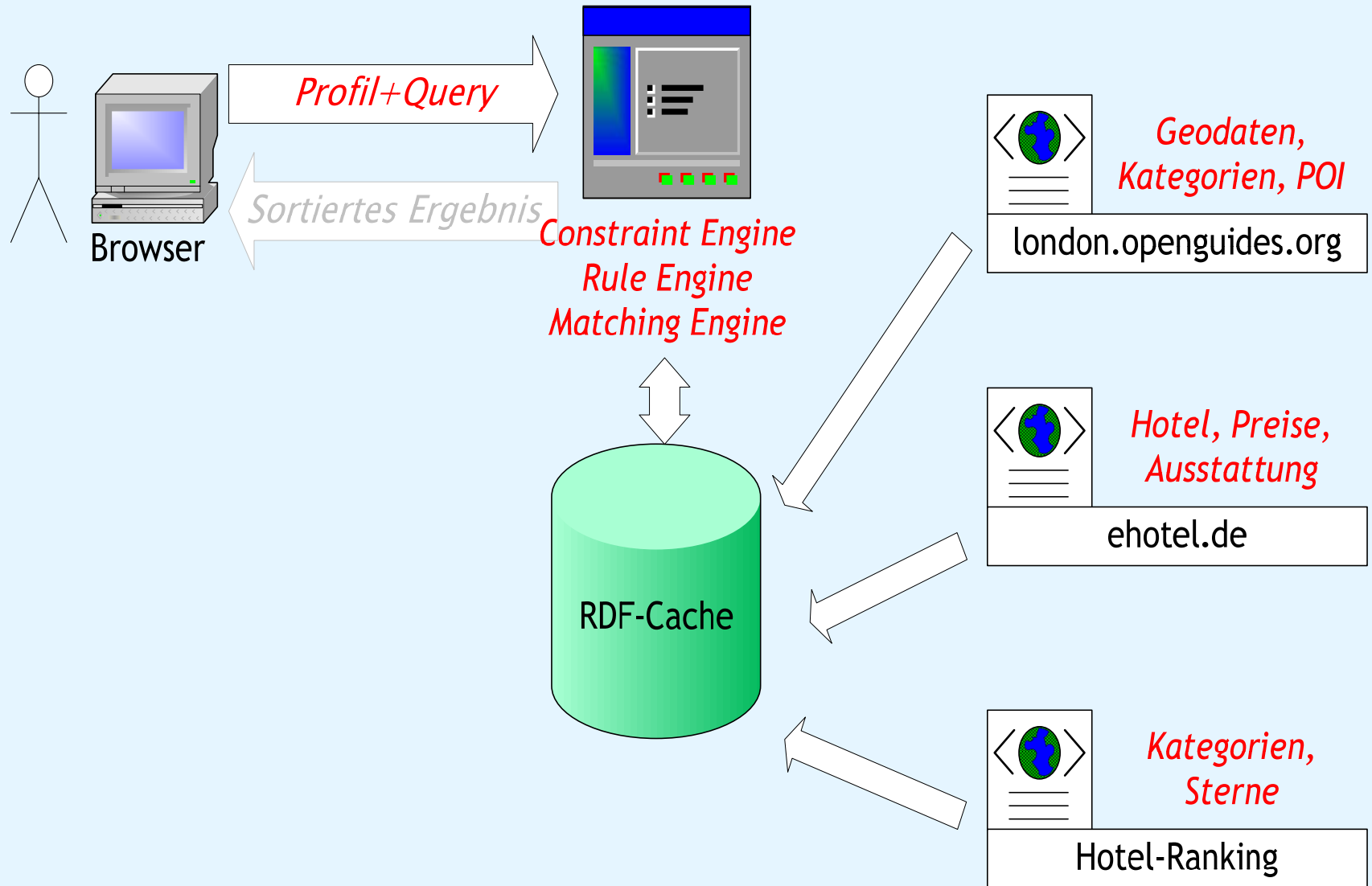
PROFILE ANSEHEN

Unterprofil auswählen
(Mehrfachauswahl möglich)

- Wellness
- Business
- Kommunikation
- Luxus
- Sport

PROFILE BEARBEITEN

SUCHE STARTEN



Datenintegration

- Datenquellen werden nach RDF transformiert
- Entweder auf Anfrage ...
- ... oder gepuffert (Caching)

- Mögliche Datenquellen
 - RDF-Repository
 - Web Services
 - HTML (Webscraping)
 - RDF-annotierte Webseiten
 - SQL-Datenbanken (via D2RQ)

- Regeln prototypisch in Prolog formuliert

- Projekt „Reisewissen“ ist ein praktischer Anwendungsfall für Semantic-Web-Technologien
- Verbesserte Produktsuche
- Integration von Datenquellen
- Transformation in RDF
- Anwendbarkeit lässt sich leicht erweitern auf ...
 - Reise insgesamt
 - Andere Domänen

Das Projekt



Projektpartner



+



AG Netzbasierte Informationssysteme
Lehrstuhl für Wirtschaftsinformatik

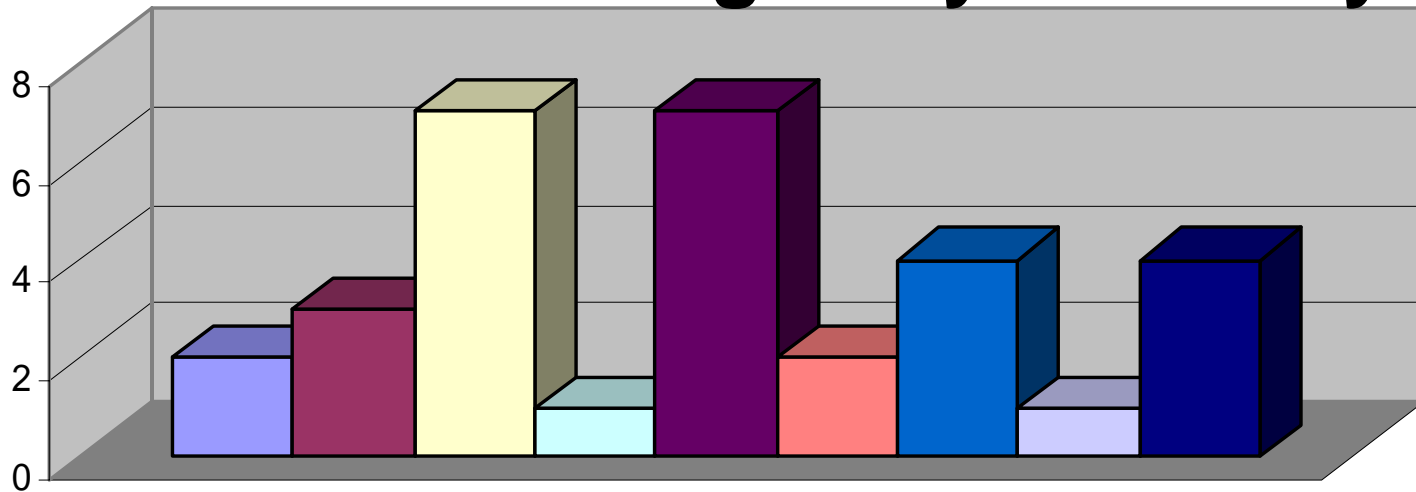
Förderung



<http://nbi.inf.fu-berlin.de/research/reisewissen>

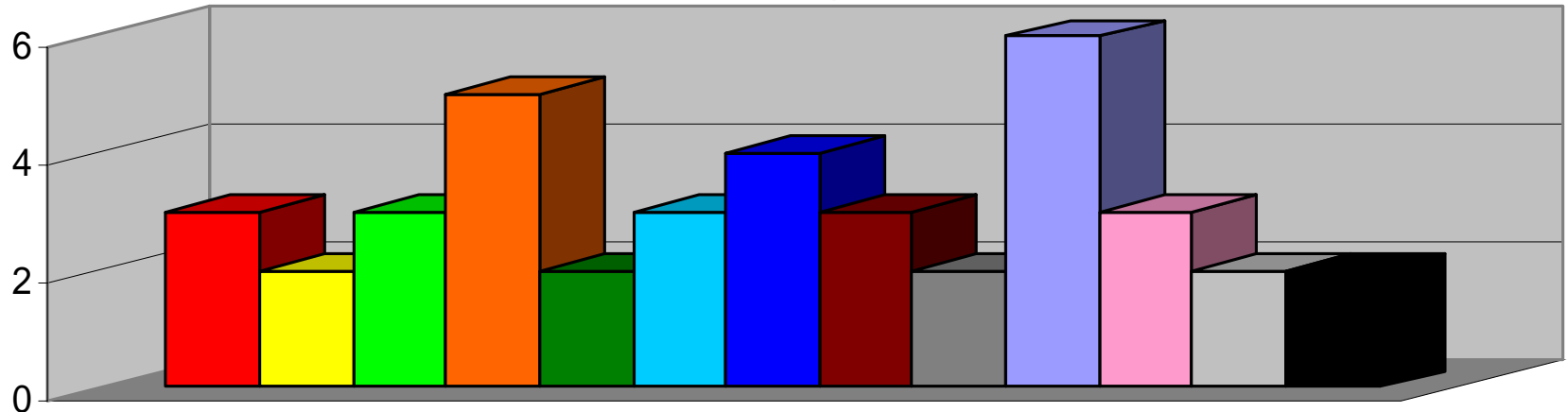
- Projekt „Reisewissen“ – Hotelsuche mit Semantic Web
 - Einsatz von ontologisierten & nicht-ontologisierten Daten
 - Datenintegration
 - Verbesserte Hotelsuche
- Projekt „Wissensnetze“ – Online Recruitment mit Semantic Web
 - Wiederverwendung existierender Standards in der HR-Ontologie
 - Semantisches Matching für Jobsuche
 - Erhöhung der Markttransparenz

Solutions sought by industry



- Matching
- Annotation
- Search
- Navigation
- Integration of data
- Standardization of vocabulary
- Data management
- Consistency checking
- Personalisation

Technology locks



- Semantic query
- Knowledge extraction
- Ontology mapping
- Ontology development
- Ontology maintenance
- Ontology matching
- Ontology-based reasoning
- Semi-automated annotation
- Support for rules
- Storage and retrieval
- Semantic Web Services
- Ontology authoring tools
- Trust

Quelle: L.Nixon, M.Mochol, D1.1.2 Prototypical Business Use Cases EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB, KWEB/2004/D1.1.2/v1.0