

Morgen in meiner Stadt

- Semantische Lern-Plattform

Andreas Winschu

Motivation

- Entwicklung einer Lern-Web-Plattform
- Bereicherung der Plattform durch Methoden des Semantic Web Ansatzes
 - Annotieren der Inhalte
 - Semantische Suche/Navigation

Inhalt

1. Motivation

2. Grundlagen

3. Related Work

1.ActiveRDF

2.Faceted Browsing

3.LarkC

1.webPIE

4.Bottom up Reasoning

4. Abgrenzung

Semantic Web

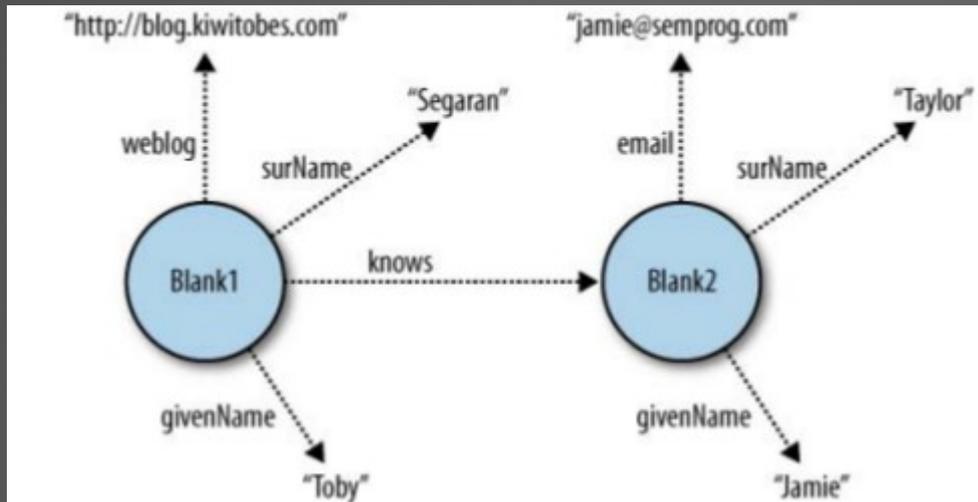
Grundidee:

- Maschinenlesbares web

Database-driven web applications	Semantic Web applications
Centralised	Decentralised
One fixed schema	Semi-structured
One fixed vocabulary	Arbitrary vocabulary
Centralised publishing	Publish anywhere
One datasource	Many distributed datasources
Closed systems	Open systems
Constraint-based semantics	Open-world (DL) semantics

Semantic Web

RDF



RDF(S) / Ontologie

- Implizites Wissen, Rückschluss auf RDF Fakten
- Modellierung der Begriffswelt

SPARQL

- RDF Abfragesprache

ActiveRDF

- OO Semantic Web Programming
 - Eyal Oren
 - Renaud Delbru
 - Sebastian Gerke
 - Armin Haller
 - Stefan Decker

Digital Enterprise Research Institute
National University of Ireland, Galway

ActiveRDF

- Semantic Web APP's erfordern händeln von RDF Data

Datenmodell

OO vs. Triple based RDF

RDF Datamapper ähnlich Hibernate, ADO.NET

- Mappen von RDF Schema Klassen auf Klassen einer Programmiersprache

Person.firstName

anstatt

Resource.getProperty(<http://xmlns.com/foaf/0.1/firstName>)

ActiveRDF

relational web applications:

web application framework

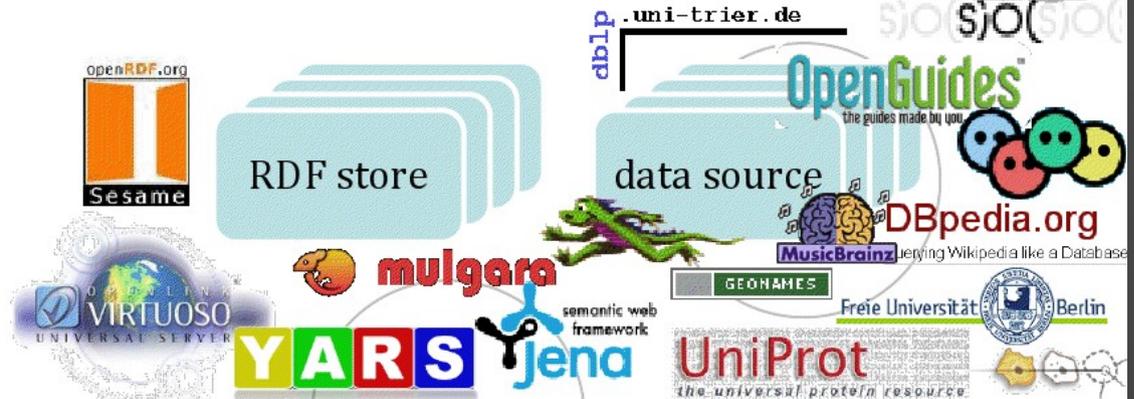
object-relational mapping

database

semantic web applications:

Ruby on Rails

ActiveRDF



ActiveRDF

OO vs. RDF

	OO	RDF
Paradigma	Close World	Open World
Class Membership	Fest, während Instantiierung Objekt an eine Klasse gebunden	Frei (rdf:type, properties) Ressource kann zu mehreren Klassen gehören
Class Hierarchy	Meistens 1 Superklasse	Mehrfachvererbung, sowie Zyklische Vererbung
Attribute vs. Property	Attribute Lokal, nur von innerhalb dieser Klasse Oft fester Datentype	Property Standalone Instance Benutzbar von jeder Ressource Werte mehrerer Datentypen
Flexibility	Keine Veränderung während Laufzeit	RDF ist designed für heterogene Daten von unterschiedlichen Quellen. Schema und Daten verändern sich zu Laufzeit

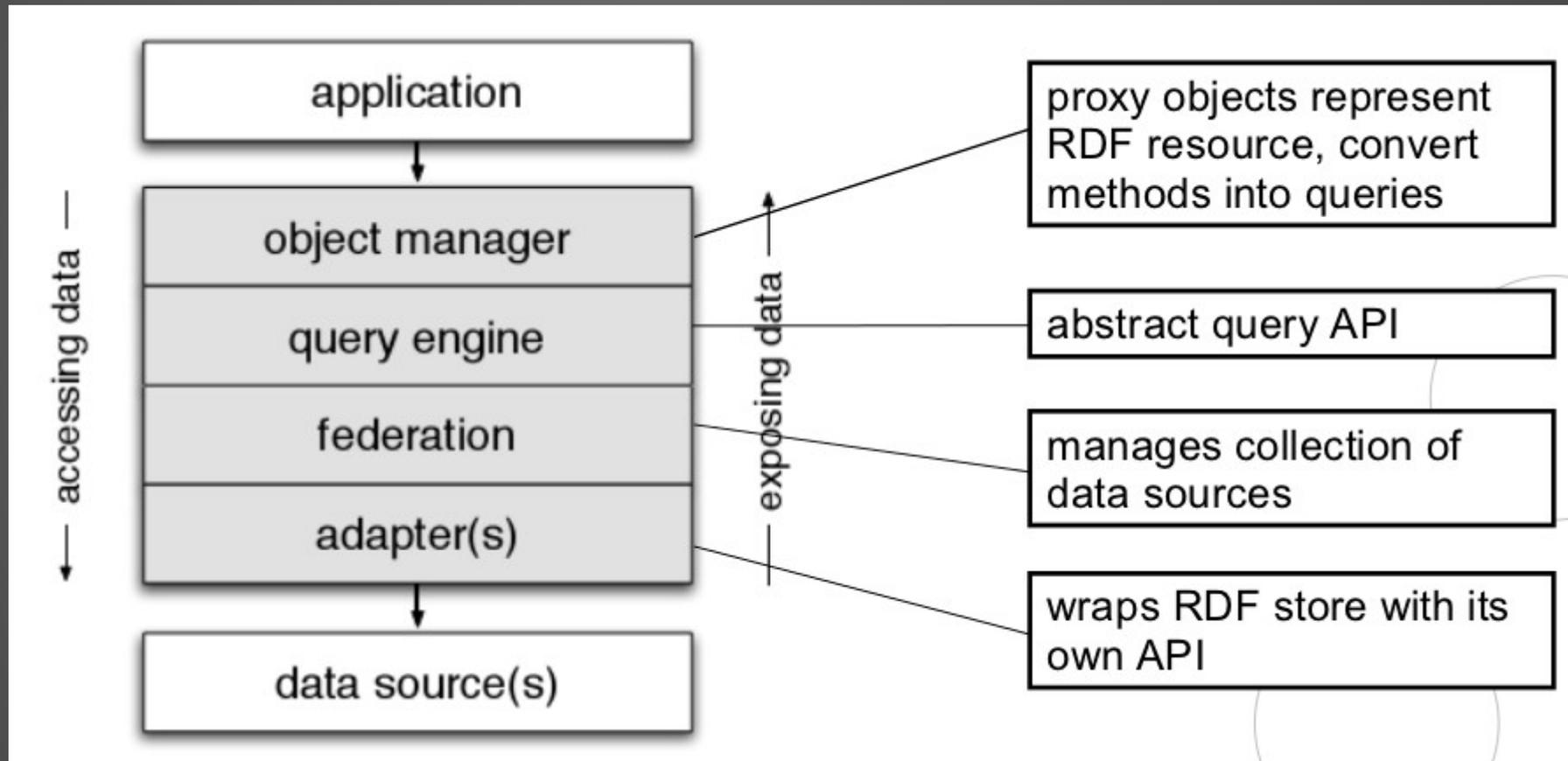
ActiveRDF

Lösungsansatz

- Erzeuge eine DSL
 - Scriptspachen, Metaprogrammierung, dynamische Typisierung
Laufzeitintrospection
- Ruby
 - Instanzspezifisches Verhalten der Objekte möglich
 - Objekte dürfen von ihrer Klassen-Definition abweichen
 - Klassenhierarchie kann überschrieben werden
 - Daten und Sprache verändern sich zur Laufzeit

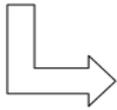
ActiveRDF

Architektur



ActiveRDF

```
foaf:Person isa rdfs:Class.  
john isa foaf:Person.  
john foaf::name "John Breslin".
```



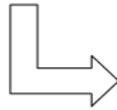
```
Person  
name:string
```

```
john: Person  
name: "John Breslin"
```

- **Intuitives Mapping**
 - RDF Klassen zu OO Klassen
 - RDF Ressourcen zu OO Objekten
 - RDF Tripel zu OO Attribute Werten
- **Multiple Klassenzugehörigkeit**
 - Standartmethoden die sich auf das **class** Attribut beziehen werden überschrieben und verwenden **rdf:type**

ActiveRDF

```
foaf:Person isa rdfs:Class.  
john isa foaf:Person.  
john foaf::name "John Breslin".
```



```
Person  
name:string
```

```
john: Person  
name: "John Breslin"
```

Virtual API

- Methoden wie **john.age** oder **Person.find_by_name('John')** werden nicht generiert
- Zuhilfenahme von **MethodNotFoundError**
- **ObjectManager** Proxy fängt den Aufruf ab
- Aufruf wird in ein Query übersetzt

Faceted Browsing

- **Data Navigation UI**
 - RDF data ist groß, stark verbunden und ohne festes Schema.
 - Anforderung an skalierbare, graphbasierte, generische Datennavigation
- **Bekannte Techniken für RDF Navigation:**
 - Keyword Suche e.g **Swoogle**
 - Simple Informationsanfrage, nicht geeignet für Lernen und Nachforschung
 - Explizite Anfragen e.g **Sesame**
 - Schwer und erfordert Kenntnis der Datenstruktur
 - Graph Visualisierung e.g **IsaViz**
 - Skaliert nicht
 - **Faceted Browsing**
 - Ursprünglich manuell konstruiert und domainabhängig
 - Unterstützt nicht von vornherein Graphnavigation

Faceted Browsing

The screenshot shows the iTunes application window with a progress bar at the top indicating "Copying 13 of 938: Rose Rogue". The main interface is divided into three vertical panes for faceted browsing: Genre, Artist, and Album. The Genre pane shows "All (25 Genres)" with a scrollable list including 2001, Acid, Alternative, Blues, Dance Hall, Drum & Bass, Dub, Electronic, Electronica/Dance, and Fusion. The Artist pane shows "All (54 Artists)" with a scrollable list including Afro Cuban All Stars, Ahpex Twin, Air, Amon Tobin, Anima Sound System, Aphex Twin, Asian Dub Foundation, Beck, Bent, and Bent - Ariels. The Album pane shows "All (59 Albums)" with a scrollable list including A Toda Cuba Le Gusta, Acid Dub Nucleik, Amelie, Ariels, Asnakjálkar í apabúningum, Bass Temperature, Beautiful Garbage, Big Mother Is Watching You, Boogie Woogie Christmas, and Brother's Gonna Work It Out. Below these panes is a table of music tracks with columns for Name, Time, Artist, Album, Genre, My Rating, Play Count, and Last Played. The table contains 20 rows of tracks, including "Habana Del Este", "Los Sitió Asere", "Salsa - Afro Cuban All Stars...", "Tanga, Rumba", "Amor Verdadero", "Fiesta De La Rumba", "Pio Mentiroso", "Maria Caracoles", "Elube Changó", and "Come To Daddy, Pappy mix".

Name	Time	Artist	Album	Genre	My Rating	Play Count	Last Played
✓ Habana Del Este	6:38	Afro Cuban All S...	A Toda Cuba Le ...	Acid			
✓ Habana Del Este	6:38	Afro Cuban All S...	A Toda Cuba Le ...	Acid			
✓ Los Sitió Asere	5:20	Afro Cuban All S...	A Toda Cuba Le ...				
✓ Los Sitió Asere	5:20	Afro Cuban All S...	A Toda Cuba Le ...				
✓ Salsa - Afro Cuban All Stars...	0:58	Afro Cuban All S...	A Toda Cuba Le ...	Salsa			
✓ Salsa - Afro Cuban All Stars...	0:58	Afro Cuban All S...	A Toda Cuba Le ...	Salsa			
✓ Tanga, Rumba	3:30	Afro Cuban All S...	A Toda Cuba Le ...	Latin			
✓ Tanga, Rumba	3:30	Afro Cuban All S...	A Toda Cuba Le ...	Latin			
✓ Amor Verdadero	6:38	Afro Cuban All S...	A Toda Cuba Le ...	Jazz			
✓ Amor Verdadero	6:38	Afro Cuban All S...	A Toda Cuba Le ...	Jazz			
✓ Fiesta De La Rumba	5:53	Afro Cuban All S...	A Toda Cuba Le ...	Jazz			
✓ Fiesta De La Rumba	5:53	Afro Cuban All S...	A Toda Cuba Le ...	Jazz			
✓ Pio Mentiroso	4:37	Afro Cuban All S...	A Toda Cuba Le ...	Other			
✓ Pio Mentiroso	4:37	Afro Cuban All S...	A Toda Cuba Le ...	Other			
✓ Maria Caracoles	4:48	Afro Cuban All S...	A Toda Cuba Le ...	Jazz			
✓ Elube Changó	4:02	Afro Cuban All S...	A Toda Cuba Le ...	Jazz			
✓ Come To Daddy, Pappy mix	4:21	Ahpex Twin	Come To Daddy	Electronic			
✓ Come To Daddy, Pappy mix	4:21	Ahpex Twin	Come To Daddy	Electronic			
✓ Flim	2:57	Ahpex Twin	Come To Daddy	Electronic			
✓ Flim	2:57	Ahpex Twin	Come To Daddy	Electronic			
✓ Come To Daddy, Little Lord ...	3:50	Ahpex Twin	Come To Daddy	Electronic			
✓ Come To Daddy, Little Lord ...	3:50	Ahpex Twin	Come To Daddy	Electronic			

Faceted Browsing

Einschränkungen

- Facet Auswahl ist typischerweise auf eine Domain konfiguriert
 - Musik: Genre, Interpret, Album, Titel
 - Bücher: Titel, Verlag, Autor
 - Rezepte: Zutaten, Art, Kalorien, Zubereitungszeit
 -
- Facets offenbaren vernetzte Daten nicht vollständig
 - Nehmen Homogenität an
 - Fokussieren auf ein Ressourcentyp

Facet Browsing

... on the semantic Web

- Theory

Informationsraum zerteilt in Dimensionen - Facets .

Facets - wichtige charakteristische Eigenschaften

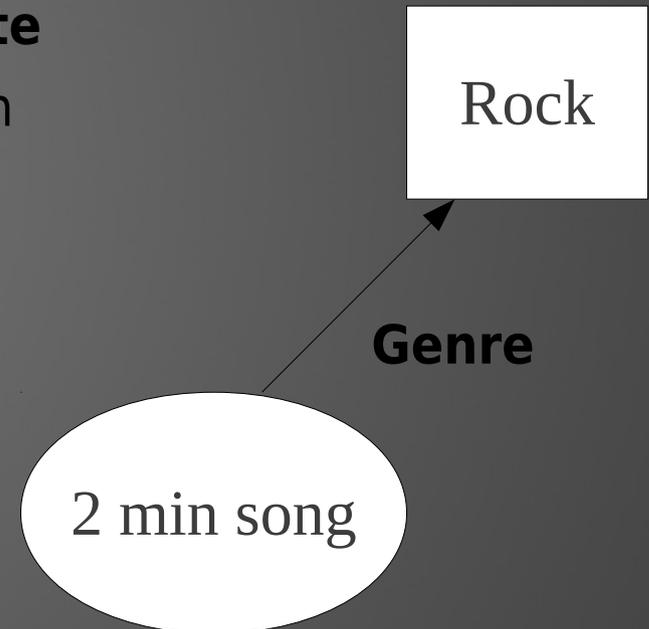
Ein Facet besitzt mehrere **Restriktionselemente**

Benutzer wählt Restriktionselemente und filtert den Informationsraum

–

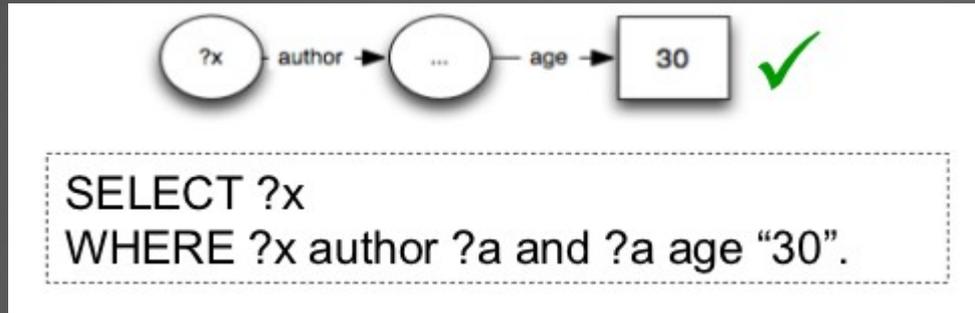
- Facets in Semantic Web

- Informationsraum: Menge aller Tripel
- Facets: Menge der Prädikate (Properties)
- Restriktionselemente: Menge der Objekte

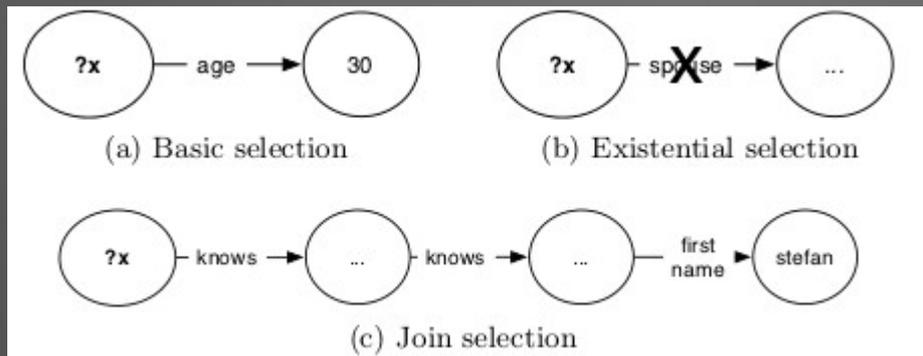


Faceted Browsing

- Faceted Browsing ist Query Construction



- Arten von Queries



Faceted Browsing

- Automated Facet Ranking

- Bestmögliche Navigation im Baum

- Kriterien:

Predicate Balancing

- Alle Restriction Values etwa gleich oft

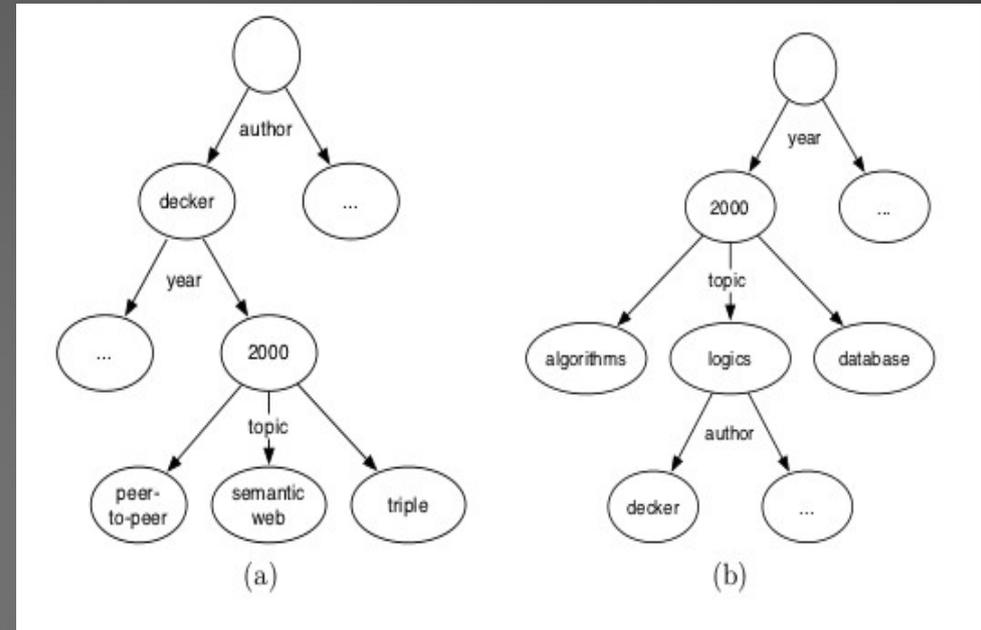
Object Cardinality

- Limitierte Anzahl an Restriction values

Predicate Frequency

- Ein "gutes" Prädikat kommt oft vor

- Die Werte werden bei jedem Schritt neu berechnet



Faceted Browsing

- Evaluation

	solved	unsolved		keyword	query	faceted
			easiest to use	13.33%	0%	86.66%
keyword	15.55%	84.45%	most flexible	13.33%	26.66%	60%
query	15.55%	84.45%	most dead-ends	53.33%	33.33%	13.33%
faceted	74.29%	25.71%	most helpful	6.66%	0%	93.33%
			preference	6.66%	6.66%	86.66%

(a) Task solution rate

(b) Post-test preferences

LarkC

- Large Knowledge Collider
 - Large Scale Semantic Computing, distributed Reasoning
Europäische Union
 - Teilnehmer
 - STI Innsbruck, University of Innsbruck (Austria) (coordinator),
 - AstraZeneca R&D (Sweden), CEFRIEL Società consortile a responsabilità limitata (Italy), High Performance Computing Center, University of Stuttgart (Germany), Max Planck Institute for Human Development (Germany), Saltlux Inc. (Korea), Siemens AG (Germany), University of Sheffield (United Kingdom), Vrije Universiteit Amsterdam (Netherlands), Web Intelligence Institute (China), International Agency for Research on Cancer (France), Information Retrieval Facility (Austria), Ontotext AD (Bulgaria)
 - Budget: 10 Millionen €

- **Towards LarkC: a Platform for Web-scale Reasoning**
 - Dieter Fensel (University of Innsbruck)
 - Frank van Harmelen (Vrije Universiteit Amsterdam)
 - Bo Andersson (Astrazeneca AB)
 - Paul Brennan (International Agency for Research on Cancer)
 - Hamish Cunningham (University of Sheffield)
 - Emanuele Della Valle (CEFRIEL)
 - Florian Fischer (University of Innsbruck)
 - Zhisheng Huang (Vrije Universiteit Amsterdam)
 - Atanas Kiryakov (OntoText Lab (Sirma AI Ltd.))
 - Tony Kyung-il Lee (Saltlux)
 - Lael Schooler (Max Planck Institute for Human Development, Berlin)
 - Volker Tresp (Siemens)
 - u.a

- Plattform für Web-Scale Reasoning
 - “... einige sehr fortgeschrittene Szenarien erfordern Reasoning über 10 Milliarden RDF Tripel in weniger als 100ms”
 - Aktuelle Logik basierte Reasoning Systeme skalieren nicht
 - Plan für eine massiv verteilte incomplete Reasoning Plattform:
 - Bereichern des aktuellen logikbasierten Reasoning durch Methoden des Maschinenlernens, Informationstheorie, Datenbanken und Wahrscheinlichkeitstheoretisch
 - Anwenden kognitiv inspirierter Methoden
 - (spreading activation, reinforcement, u.a.)
 - Entwickeln einer verteilten Reasoning Plattform auf einem high-performance computing cluster
 - Pluginbasierte Architektur (Reasoning Engines austauschbar)

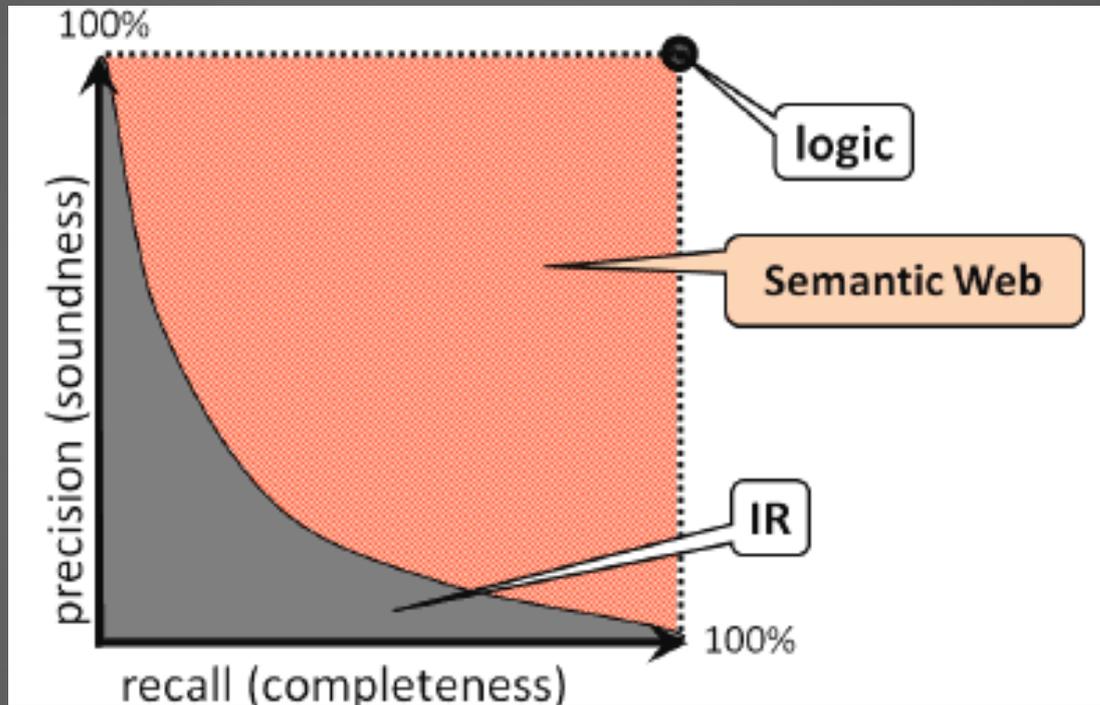
LarkC

- State-of-the-art
 - Reasoning in kleinen, geschlossenen, konsistenten und statischen Domains
 - Kleine Menge von Axiomen und Fakten
 - DL Reasoner können 10^5 Axiome(Regeln) handhaben, skalieren schlecht bei großen Instance Sets
 - Logic Programming Engines – ebenfalls ca. 10^5 Axiome und ca. 10^6 Instance Sets, aber nur simple logische Schlussfolgerungen
 - Es besteht eine tiefe Nichtübereinstimmung zwischen **Reasoning auf Web-Scale Ebene** und effizienten **Reasoning Algorithmen** basierend auf **Prädikatenlogik**
 - Folgt aus den Kernannahmen der Computational Logic

Computational Logic	Semantic Web Reasoning
Kleine Menge von Axiomen	Der Web spiegelt die Gesamtheit des menschlichen Wissens wieder
Kleine Menge von Fakten	30 Milliarden Webseiten * 100 Fakten/Seite = 3 Trillionen Fakten
Vollständigkeit der Inferenzregel	Web hat keine festen Grenzen. Unrealistische Anforderung an die Inferenzmaschine
Korrektheit der Inferenzregel und Konsistenz. Axiome spiegeln die Wahrheit wieder. Erschließe das wahre Wissen aus den Axiomen	Information ist unzuverlässig von Anfang an. Unmöglich die Wahrheit zu erschließen.
Statische Domain	Fakten ändern sich, während des Zugriffsprozesses

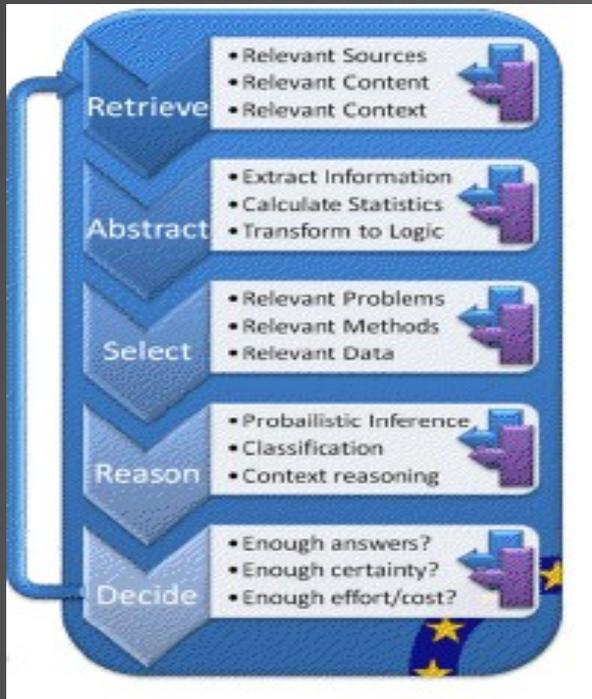
LarkC

- Balance zwischen Korrektheit und Vollständigkeit



Positionierung zwischen traditionellem Suchen und traditionellem Reasoning

LarkC



- 1) Frage rohen Inhalt mir Assertions ab (Vermute einen Beitrag zu Lösung)
- 2) Abstrahiere in die für das heterogene Reasoning benötigte Form
- 3) Wähle den am meist versprechenden Ansatz zuerst
- 4) Reasoning durch mehrere Mittel
- 5) Entscheide ob genügend akkurate und präzise Lösungen gefunden wurden
- 6) Versuche weiter wenn nicht und die Zeit es zulässt

WebPIE

- [Web-scale Parallel Inferenz Engine](#)

Jacopo Urbani, Spyros Kotoulas, Jason Maassen, Niels Drost, Frank Seinstra,

Frank van Harmelen, Henri Bal

LarKC und Prof. Henri Bal's Distributed Computing Group

Department of Computer Science, Vrije Universiteit Amsterdam

WebPIE

- **Web-scale Parallel Inferenz Engine**

Semantic Web Reasoning durch Menge von Map- Reduce Operationen (basiert auf Hadoop)

Map Reduce Framework:

Map Phase verteilt die lokalen Aufgaben

Reduce Phase fügt die Ergebnisse zusammen

```
if (<?p isA transitiveProperty>, <?a ?p ?b>, <?b ?p ?c>)  
then <?a ?p ?c>
```

three-way-join, unzählige rulesets, exponentielle Komplexität

WebPIE

Angegangene Probleme:

- Parallele Joins
 - Verbessertes load balancing durch Umverteilung bestimmter Regeln in separate MapReduce Jobs
- Doppelte Inferenzen
 - Einige Aussagen können auf Basis mehrerer Inputs und Regeln getroffen werden
 - Gruppieren der Statements bei den Inferred Statements sowie spezifische Optimierungen in Sonderfällen
- Rekursives Anwenden der Regeln
 - Optimieren der Reihenfolge der Regelanwendung um die Iterationsanzahl zu reduzieren

iQser Technologies

- Bottom up Approach by IQser Technologies

Prof. Dr. Bela Mutschler
(Projektleiter)

Hochschule Ravensburg-Weingarten

Bernd Michelberger

Hochschule Ravensburg-Weingarten

Prof. Dr. Manfred Reichert

Universität Ulm / Institut DBIS

Richard Oed

Daimler AG / Abteilung GR/PSP

Markus Hipp

Daimler AG / Abteilung GR/PSP

Dr. Jörg Wurzer

iQser Technologies

iQser Technologies

- iQser GIN Plattform

- Semantische Middleware

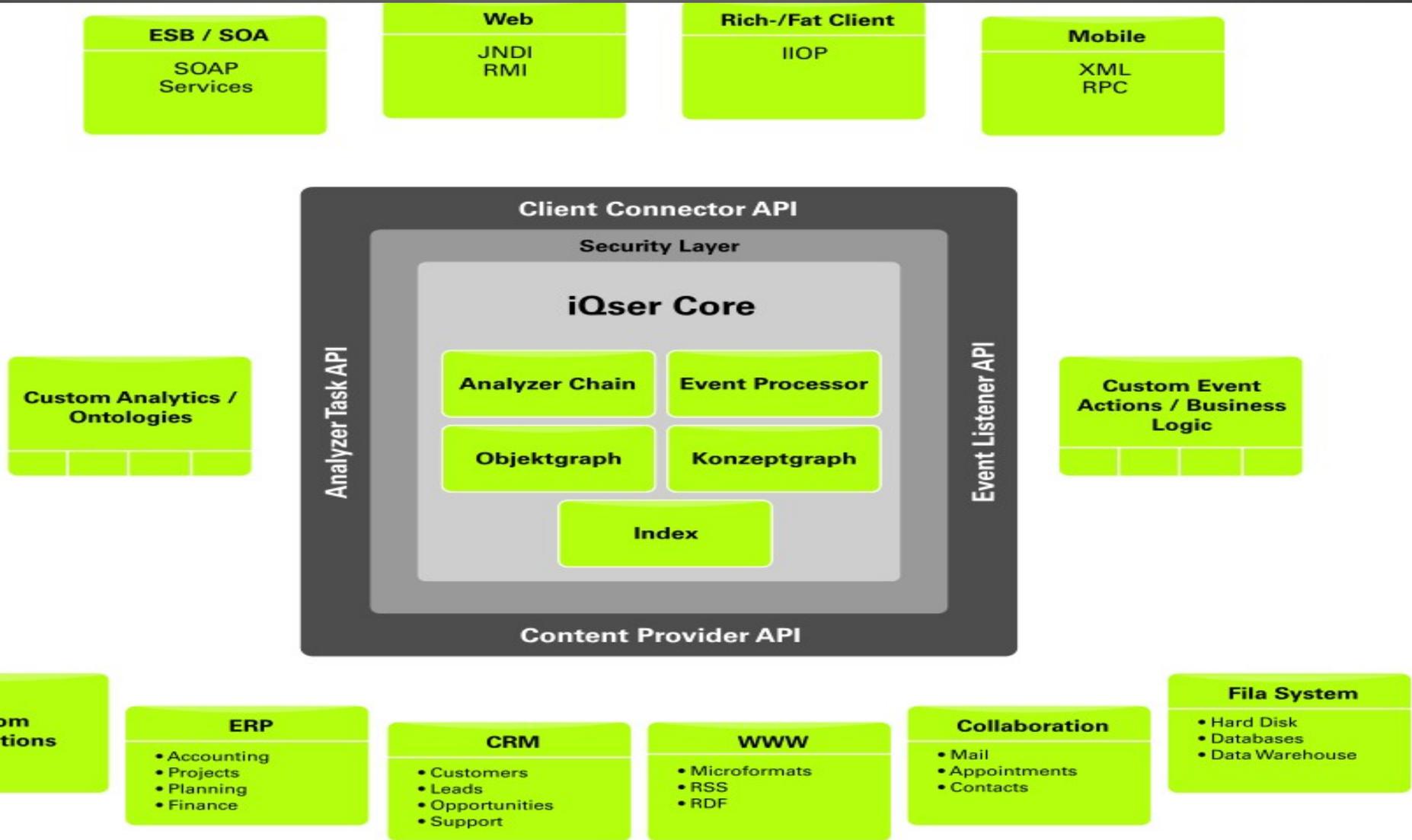
Zentrale semantische Komponente zur Integration in die bestehende IT-Infrastruktur

- Automatisch abgeleitetes, übergreifendes Datenmodell

- Uniform Information Layer

- Zugriffspunkt auf Daten aus verteilten, heterogenen Quellen und Systemen
- Daten nach einer einheitlichen semantischen Beschreibung selektieren und abrufen
- e.g. Report über alle Kunden, die durch die Verknüpfung zu bestimmten technischen Problemfällen charakterisiert sind, die erst nach mehreren Rückfragen gelöst wurden

iQser Technologies



Iqser Technologies

- Bottom up Approach
 - kein Wissen über die eingebundenen Systeme und ihrer Heterogenität für solche Abfragen notwendig
 - Keine Datenmigration, Mapping zwischen den verteilten Systemen oder Modellierung einer übergeordneten Ontologie notwendig
 - Ableitung einer flachen Ontologie aus den Objekten verfügbarer Datenquellen
 - Aufgabe, eine theoretische Ontologie mit den realen Daten mit einem Mapping zu verknüpfen entfällt
 - statistischer Verfahren extrahiert signifikante Konzepte wie Namen, Orte, Fachtermini etc. automatisch und vernetzten diese

Zusammenfassung

- **ActiveRDF**

- Wie modelliert man RDF Daten in OO?

- **Faceted Browsing**

- Welche Anwendungsfälle für Applikationen gibt es?

- **LarkC**

Welche Plattformen ermöglichen eine Umsetzung der Semantic Web Leitideen

- Wie händelt man logisches Reasoning auf riesigen Datenmengen?

- **Bottom up Reasoning**

- Wie annotiert man sinnvoll die Information aus reinen Texten?

Abgrenzung

- Übertragen eines realen Datensatzes (relational) auf ein RDF Modell
- Etablieren eines Semantic Web Ecosystems (Plattform), integrieren der bestehenden Web 2.0 Anwendung
- Auffinden der “semantischen” UseCases
 - Evaluieren von Faceted Browsing an konkretem Anwendungsfall,
 - Tauglichkeit von automatischen Facet Ranking
 - Effizienz der Facetbildung bei großen Datenmengen
 - Szenariofindung für Reasoning am konkretem Datenset

Annotation der Textinhalte

- Entwicklung eines Interfaces
- Bottom up Analyse
- Aufzeigen des Mehrwertes der Semantischen Technologien

Quellen

- Segaran, Evans, Taylor, **Programming the Semantic Web**, O'Reilly, 2008
- Oren, Delbru, **ActiveRDF: Object-Oriented Semantic Web Programming**, 2007
- Oren, Delbru, **ActiveRDF: object-oriented RDF in Ruby**, 2006
- E. Oren, R. Delbru, and S. Decker. **Extending faceted navigation for RDF data**. 2006
- V. Geroimenko and C. Chen, (eds.) **Visualizing the Semantic Web**. Springer-Verlag, 2006.
- Fensel, Harmelen, **Towards LarKC: a Platform for Web-scale Reasoning**, 2008

Quellen

- R. Krummenacher, I. Toma, D. Fensel, R. Brehar, and S. Nedevschi, **Instrumenting and Monitoring the LarKC Research Infrastructure**, 2011.
- Axel Polleres, **OWLIM: A family of scalable semantic repositories**, 2011
- J. Urbani, S. Kotoulas, J. Maassen, N. Drost, F. Seinstra, F. van Harmelen, and H. Bal, **WebPIE: a Web-scale Parallel Inference Engine**, 2010
- Jörg Wurzer, **iQser GIN Plattform - Intelligente Semantische Middleware**, 2010
- Jacopo Urbani, Spyros Kotoulas, Eyal Oren, and Frank van Harmelen, **Scalable Distributed Reasoning using MapReduce**, 2009