XMIDX 2003

Bausteine einer Vertrauens- und Sicherheitsinfrastruktur für das Semantic Web

Dipl.-Kfm. Chris Bizer

Freie Universität Berlin
Fachbereich Wirtschaftswissenschaft
Lehrstuhl für Wirtschaftsinformatik - Prof. Dr. Uwe Suhl

Das Semantic Web

- **For the Web to reach its full potential, it must evolve into the Semantic Web, providing a universally accessible platform that allows data to be shared and processed by automated tools as well as by people.** (Tim Berners-Lee, 2002)
  - Erweiterung des WWW um "maschinen-interpretierbare" Informationen
  - WWW ~ globalen, verteilten Dokumenten-Sammlung
  - Semantic Web ~ globalen, verteilten Datenbank
  - Semantic Web Beispielanfrage: Finde alle Professoren in Norddeutschland!
- **Viele Communities arbeiten an der Realisierung der Vision des Semantic Web:**
  - Information Retrieval: Semantische Suche, Bezug zu Topic Maps
  - Computer Linguisten: Semantische Annotierung von Texten
  - Künstliche Intelligenz: Wissensaustausch, globale Wissensbasis
  - Datenbanken: Datenmodell für semi-strukturierte Daten, Integration von Daten unterschiedlicher Formate und Quellen
  - Verteilte Informationssysteme: Datenbasis für Agenten, standardisierte Agenten-Kommunikation, Beschreibung von Web Services
  - Wirtschaftsinformatik: Anwendungspotentiale in den Bereichen E-Commerce, EDI, Knowledge-Management
  - Industrie: Eher abwartend, da viele Entwicklungen noch am Anfang stehen

Agenda

1. Technische Grundlagen
2. Anwendungsbeispiel
3. Die Rolle von Vertrauen im Semantic Web
4. Lösungsansätze und Vergleich der Verfahren
5. Die Rollen von Sicherheit, Privacy und Digital Rights
6. Ausblick

Uniform Resource Identifiers (URIs)

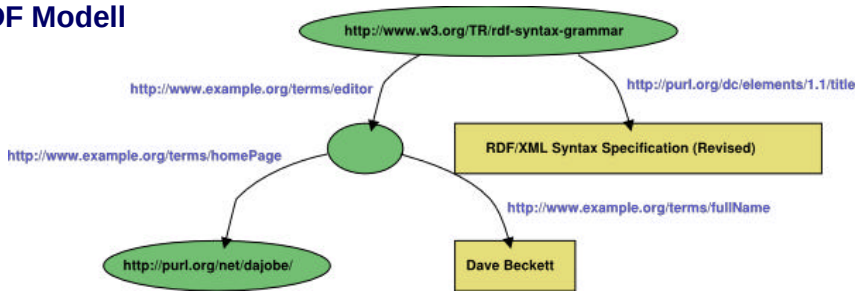
- Alles wird über URIs global eindeutig identifiziert.
  - Dokument: http://www.wiwiss.fu-berlin.de/suhl/index.html
  - Datentyp: http://www.w3.org/TR/xmlschema-2#string (W3C XML Schema Datatypes)
  - Begriff in einem Vokabular: http://purl.org/dc/elements/1.1/title (Dublin Core)
  - Produkte: isbn:0262062321, ean:4-512345-678906
  - Unternehmen: DUNS:80-473-5132 (Dun and Bradstreet)
- Die global eindeutige Identifikation ermöglicht die Integration von Daten aus unterschiedlichen Quellen.
- Es gibt noch keine Übereinkunft über die URI Schemata für Realweltobjekte.

Resource Description Framework (RDF)

- RDF liegt als Datenformat/Datenmodell allen Semantic Web Anwendungen zugrunde.
- RDF Datenmodell (vereinfacht)
  - Modelle bestehen aus eine ungeordneten Menge von Statements
  - Ein Statement besteht aus Subjekt, Prädikat und Objekt
  - Subjekte werden über URIs identifiziert
  - Prädikate werden über URIs identifiziert
  - Objekte sind URIs oder Literals (Text, inkl. Sprache und Datentyp)
- Beispiel Statement
  - Subject: http://www.bizer.de#chris
  - Prädikat: http://www.w3.org/2001/vcard-rdf/3.0#NAME
  - Objekt: Literal: "Chris Bizer"
- Alle RDF Modelle fügen sich zu einem globalen Modell, dem Semantic Web, zusammen.
- Es gibt unterschiedliche Serialisierungs-Syntaxe für RDF Modelle
  - RDF/XML als Austauschformat für das Web
  - N3 als für den Menschen besser lesbare Syntax

Beispiel

- RDF Modell



- RDF/XML Serialisierung (ohne XML und Namespace Deklarationen)

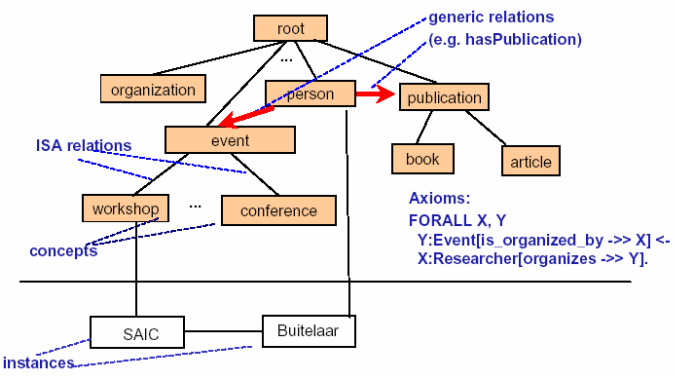
```
<rdf:Description rdf:about="http://www.w3.org/TR/rdf-syntax-grammar">
    <ex:editor>
        <rdf:Description>
            <ex:homePage rdf:resource="http://purl.org/net/dajobe/"/>
            <ex:fullName>Dave Beckett</ex:fullName>
        </rdf:Description>
    </ex:editor>
    <dc:title>RDF/XML Syntax Specification (Revised)</dc:title>
</rdf:Description>
```

RDF Schema (RDF-S) und Web Ontology Language (OWL)

An ontology is a formal, explicit specification of a shared conceptualization.

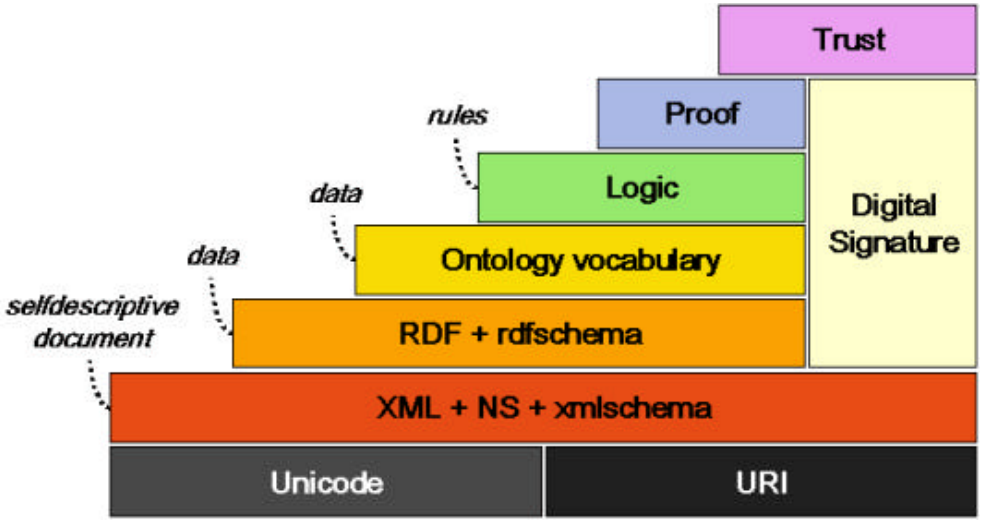
(Gruber, 1993)

- **Ontologien**
  - ermöglichen das gemeinsame Verständnis einer Domain
  - liefern Hintergrundwissen
- **RDF-S**
  - einfache RDF-basierende Sprache zur Definition von Klassen, ihren Eigenschaften und Beziehungen
- **OWL**
  - auf RDF-S aufbauendes zusätzliches Vokabular (cardinality, equivalentClass, sameIndividualAs)

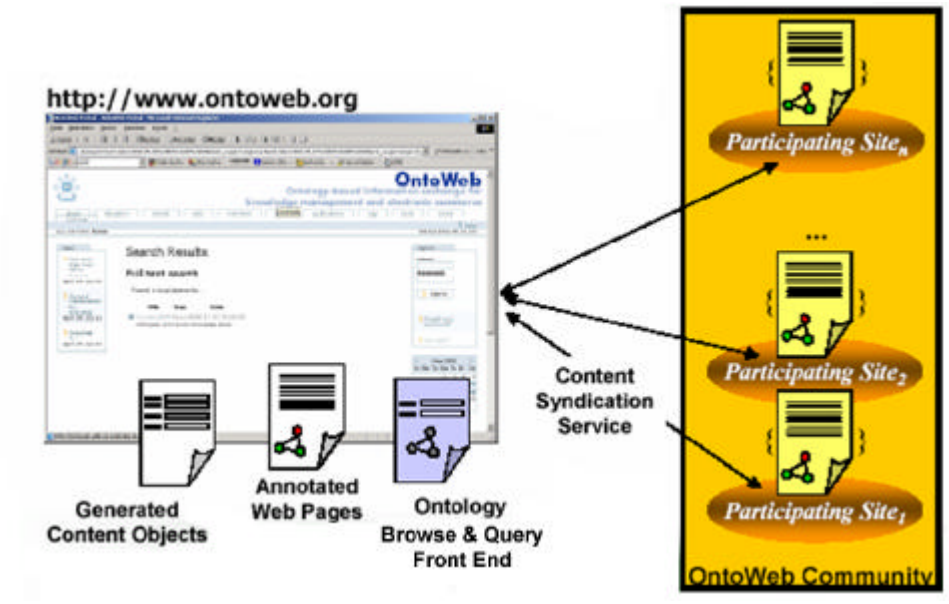


- **Ontologien im Semantic Web**
  - Parallele bzw. gemischte Nutzung verschiedener Ontologien für eine Domain
  - Ontology Mappings ermöglichen die Nutzung von Informationen anderer Ontologien
  - Ontologie Evolution: Längerfristig hofft man, dass sich einzelne Ontologien je Domain durchsetzen werden

The Semantic Web Layer Cake



Anwendungsbeispiel: OntoWeb Community Portal



Vertrauen im Semantic Web

Anyone can say anything about anything!

(Tim Berners-Lee, 2002)

- **Folge 1: Not everything found from the Web is true and the Semantic Web does not change that in any way.**
- **Folge 2: Truth has to be evaluated by each application that processes the information on the Web.**

- **Definition von Vertrauen:**

Readiness of a Trustor to rely to the actions/informations of a Trustee.

(Deutsch, 1973)

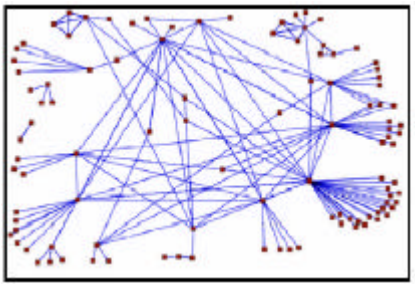
- **Entsteht in einem sozialen Prozess**
- **Beruht auf einer subjektiven Entscheidung des Vertrauenden**
- **Verändert sich im Zeitverlauf**
- **Vertrauensentscheidungen werden von sehr unterschiedlichen Faktoren bestimmt (Unterschiedliche Vertrauensmodelle)**

Ebenen der Vertrauensentscheidung im Semantic Web

- The applications decide what they trust by using the context of the statements; e.g. who said what and when and what credentials they had to say it.
- Ebene 1: Stammen die Informationen wirklich vom Autor?
  - W3C Ansatz: RDF Modelle werden digital signiert
    - XML-DSIG Signaturen und vorhandene Public Key Infrastrukturen verwendbar
    - Problem: Geringe Verbreitung von Schlüsseln und Zertifikaten
  - Alternative: Fundort der RDF Modelle im Web
    - Geringere Sicherheit und ungenauere Zuordnung
    - Aber auch geringerer Aufwand für die Informationsanbieter
- Ebene 2: Ist der Autor vertrauenswürdig?
  - Ansatz 1: Web-Of-Trust mit expliziten Vertrauensaussagen
  - Ansatz 2: Netzwerkanalyse mit expliziten Vertrauensaussagen
  - Ansatz 3: Netzwerkanalyse mit implizieten Vertrauensaussagen
  - Ansatz 4: Pragmatik

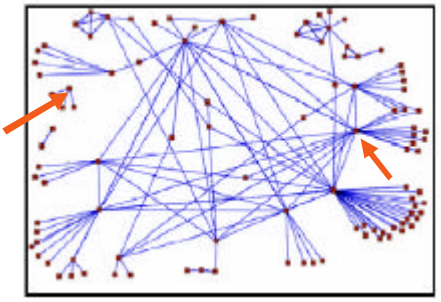
Web of Trust mit expliziten Vertrauensaussagen

- W3C Ansatz: Web of Trust
  - Der Nutzer bestimmt eine Gruppen von Informationsanbietern, denen er vertraut.
  - Nur Aussagen dieser Anbieter oder Anbietern, denen diese Anbieter vertrauen (Grad des Webs), werden berücksichtigt.
  - Es entsteht ein gerichteter, gewichteter Graph.
- MINDSWAP Experiment
  - Friend of a Friend (FOAF) RDF Schema um gewichtete Vertrauensaussagen erweitert.
  - James Hendler et all., WWW 2003
- Vorraussetzungen
  - Relativ enge Community, in der sich die Mitglieder kennen.
    - Bei offenen Systemen meist nicht gegeben.
    - Negative Erfahrungen mit dem Pretty Good Privacy (PGP) Web-of-Trust.
  - Es sind explizite Vertrauensaussagen von jedem Nutzer für jede Anwendungs-Domain erforderlich.
  - Es ist eine hohe Qualität und Aktualität der Vertrauensaussagen erforderlich.
- Fazit: Der Ansatz bedeutet einen sehr hohen Aufwand für den Nutzer und ist für offene Systeme nur bedingt geeignet.



Netzwerkanalyse mit expliziten Vertrauensaussagen

- **Ansatz: Ein außenstehender Knoten nutzt ein vorhandenes Web of Trust.**
- **Einsatzbeispiele**
  - Reputationsdienste auf elektronischen Märkten wie eBay
  - Meinungsportale wie ePinions, Dooyoo oder Ciao
- **TRELLIS Experiment**
  - RDF-basierte Bewertung von Informationsquellen
  - Gil, Ratnakar , ISWC 2002
- **Vorteile**
  - Keine enge Community erforderlich
  - Keine Aussagen von jedem Nutzer zu jeder Anwendungs-Domain erforderlich



- **Nachteile**
  - Erfordert Vertrauen in die Aussagen unbekannter Nutzer
  - Anfälligkeit gegenüber Täuschungsversuchen
  - Motivation zur Abgabe von Vertrauensaussagen

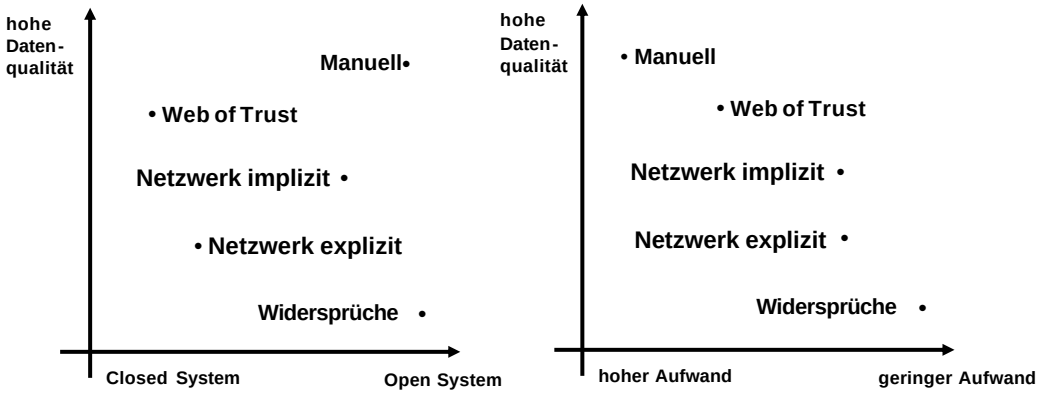
Netzwerkanalyse mit implizieten Vertrauensaussagen

- **Ansatz: Nutzung vorhandener RDF Aussagen.**
- **Beispiel**
  - Vertrauenspolitik: Vertraue den Aussagen aller Professoren, die sich mit dem Thema beschäftigen.
  - Datenbasis: Aussagen zu beruflichen Stellung, Veröffentlichungen, Forschungsinteressen.
- **Verlagerung des Vertrauensproblems, da die benutzten Aussagen über andere Verfahren abgesichert werden müssen.**
- **Vorteile**
  - Geringerer Aufwand, da weniger explizierte Vertauensaussagen erforderlich sind.
  - Eignung zur Übertragung von Realwelt-Vertrauensmodellen auf das Semantic Web.
- **Nachteil**
  - Setzt eine ausreichend große Datenbasis voraus.

Pragmatik

- **Informationen auf Widersprüche gegenüber einer Ontologie oder einer vorhandenen Wissensbasis prüfen.**
  - Ontologie
    - Kardinalitäten
    - Wertebereiche
    - Sonstige Axiome
      z.B. Person kann nicht gleichzeitig Professor und Student sein.
  - Vorhandene Wissensbasis
    - Widersprüche in den Eigenschaften eines Objekts
    - Unterschiedliche URIs bei gleichen Eigenschaften
  - Nachteil: Nur sehr grobe inhaltliche Prüfung möglich
- **Manuelle Prüfung der Informationen**
  - Dieser Ansatz wird in der Ontoweb Beispielanwendung eingesetzt
  - Nachteil: Der hohe Aufwand ist nur für bestimmte Anwendungen und für kleine Informationsmengen vertretbar

Vergleich der Verfahren



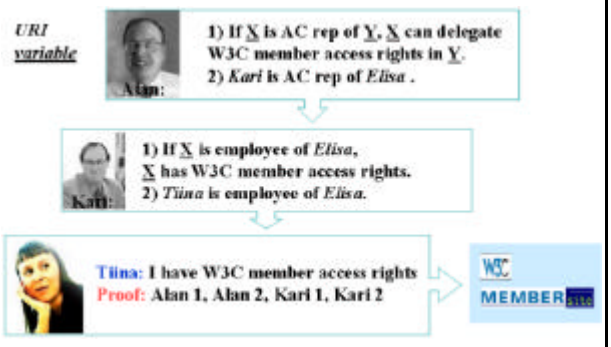
- **Entscheidung für jeden Anwendungsfall**
  - Welche Datenqualität erfordert eine Anwendung?
  - Wie lässt sich diese Qualität mit möglichst geringem Aufwand realisieren?

Beschränkung des Verwendungszwecks

- Das Semantic Web ermöglicht
  - eine einfache Integration von Daten unterschiedlicher Quellen mittels global eindeutiger URIs
  - die Verwendung dieser Daten für vom Autor nicht intendierte Zwecke
- Privacy: Dies ermöglicht die Sammlung personenbezogener Daten
- Marketing: Dies ermöglicht ein verfeinertes Profiling zu Werbe-zwecken
- Ansatz: Auszeichnung zulässiger Verwendungszwecke durch den Informationsanbieter
- Problem: Auszeichnungen ohne rechtliche Verankerung wertlos
- Adaptierbare Standards
  - P3P - Platform for Privacy Preferences
    - W3C Standards zur Formulierung von Datenschutzpolitiken für Webs ites
    - Es existiert bereits eine RDF Version von P3P
  - XrML – eXtensible Rights Markup Language
    - Sprache zur Formulierung von Nutzungsrechten
    - Die Sprache ist für den Medienbereich entwickelt worden

Zugriffsbeschränkung

- Role-Based-Access-Control
  - Vorteil: Erfordert keine zentrale Nutzeradministration
  - Wird als RDF basiertes Verfahren vom W3C eingesetzt
  - Ein Repräsentant jeder Mit-gliedsorganisation benennt die Mitarbeiter dieser Organisation
  - Alle Mitarbeiter von Mitglieds-organisationen können auf den geschützten Bereich der W3C Site zugreifen



- Entwicklung von Security- und Policy-Ontologien
  - Gemeinsame Konzepte unterschiedlicher Verfahren herausarbeiten
  - Verbesserte Interoperrationalität zwischen den Verfahren
  - Beispiele
    - KAoS Access Policy Ontologie: RDF basiertes Framework
    - OASIS eXtensible Access Control Markup Language (XACML) : XML basiertes Framework

Vielen Dank!

Kontakt: Chris Bizer (chris@bizer.de)

Folien online unter: http://www.wiwiss.fu-berlin.de/suhl/bizer/semtrust.ppt