

# Ontologien und Semantische Netze aus Sicht der Datenbank

Karin Patenge, Carsten Czarski  
Oracle Deutschland B.V. & Co. KG  
Potsdam, München

## Schlüsselworte

Datenbank, NCI, Ontologie, OWL, RDF, Semantische Netze, SPARQL

## Einleitung

Was verbindet Personen wie Lukas Cranach d.Ä., Johannes Gensfleisch zum Gutenberg und Erasmus von Rotterdam mit dem Schlachtberg von Bad Frankenhausen? Welcher Grad der Verwandtschaft bestand zwischen Friedrich dem III. und Friedrich der I. aus dem brandenburgisch-preussischem Haus Hohenzollern? Welche Patienten wurden im Jahr 2010 mit Frakturen der oberen Extremitäten eingeliefert, wobei unter „obere Extremitäten“ sowohl Ellenbogen, Unterarm, Handgelenk als auch weitere Bereiche des menschlichen Körpers gemeint sein können? Über wieviel „Ecken“ ist Carsten Czarski mit Larry Ellison bekannt?

Die Auflösung dieser Fragen erfolgt bis zum Ende des Artikels. Dabei stehen die Antworten an sich nicht im Mittelpunkt der Ausführungen. Sie dienen lediglich dazu, auf eine ihnen innewohnende Eigenschaft hinzuweisen: Was hier beispielhaft abgefragt wird, ist domänenspezifisches Wissen aus den Bereichen Kunst, Geschichte, Medizin und Soziale Netze, welches zumeist nur sehr schwer mit den Methoden gängiger relationaler Datenmodellierung (ER<sup>1</sup>-Modelle) abzubilden ist. Vielmehr kommen hier objekt-orientierte Denkansätze und Methoden zur Anwendung, um eine Wissensbasis aufzubauen und diese in geeigneter Weise für anfallende zumeist komplexe Fragestellungen zu verwenden.

Auch im Unternehmenskontext gilt es, immer komplexere Sachverhalte abzubilden, Informationen inhaltlich in Beziehung zu setzen, maschinenlesbar die richtigen Schlüsse daraus zu ziehen oder neues Wissen auf Basis von Regeln daraus abzuleiten.

## Ontologien und semantische Netze

Domänenspezifisches Wissen zu beschreiben und nutzbar zu machen ist in der Informatik eng mit dem Begriff der Ontologien verknüpft. Laut Wikipedia geht es bei einer Ontologie um die Abbildung von „Grundstrukturen der Realität“<sup>2</sup>. Ebenfalls nach Wikiepdia<sup>3</sup> dienen Ontologien dazu, eine Menge von Begriffen (Objekten, Ressourcen) aus einem bestimmten Gegenstandsbereich zu ordnen, formal darzustellen und zwischen den Begriffen bestehende Beziehungen abzubilden. Solche Ontologien enthalten darüberhinaus Regeln, um aus regelbasiert Schlussfolgerungen abzuleiten (Inferenz bzw. *Inferencing*) sowie zur Gewährleistung der Integrität der Informationen. All diese Eigenschaften zusammengenommen, stellt eine Ontologie somit ein Netzwerk (oder auch Graph) von inhaltlich zusammenhängenden Informationen dar. Darüberhinaus dient die Web Ontology Language (kurz: OWL) als Sprache für die Beschreibung von Informationen und ihren inhaltlichen Zusammenhängen.

---

<sup>1</sup> Entity Relationship

<sup>2</sup> <http://de.wikipedia.org/wiki/Ontologie>

<sup>3</sup> [http://de.wikipedia.org/wiki/Ontologie\\_%28Informatik%29](http://de.wikipedia.org/wiki/Ontologie_%28Informatik%29)

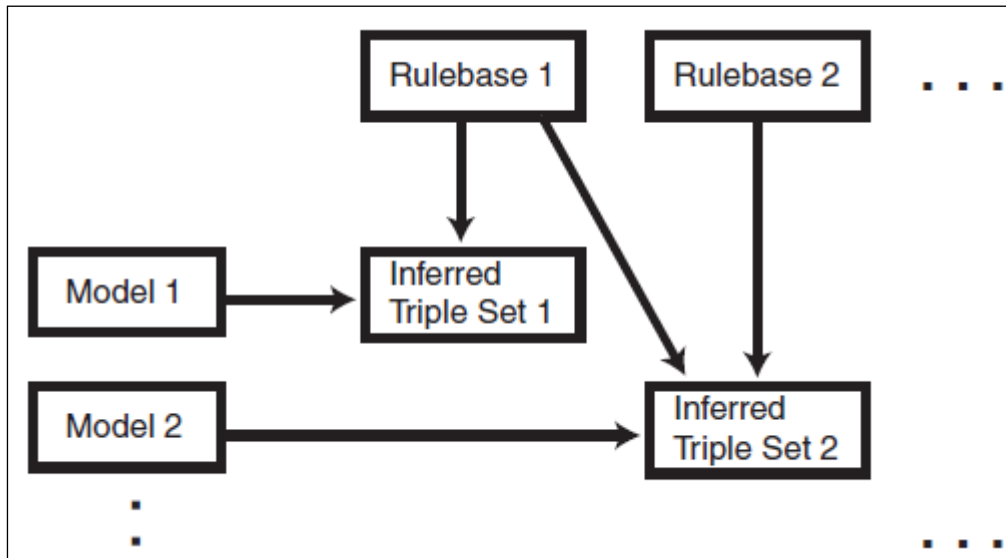


Abb. 1: Inferencing

Beispiele für Ontologien finden sich zahlreich im WWW<sup>4</sup>, für den Bereich der Medizin z.B. die *Gene Ontology*<sup>5</sup> (kurz: GO), die Ontologie zur Anatomie des menschlichen Körpers (engl.: *Foundational Model of Anatomy*, kurz: FMA) oder auch die beispielhaft im Skripting verwendete Ontologie des *National Cancer Institute* (kurz: NCI).

#### Resource Description Framework (kurz: RDF)

Der W3C-Standard RDF definiert ein Datenmodell und ein Framework zum Beschreiben von Ressourcen, unter denen grundsätzlich zunächst einmal um alle Dinge der realen Welt verstanden werden können. Im Unterschied zu einem relationalen Datenmodell verwendet RDF ein sehr allgemeines und flexibles Datenmodell. Alle Informationen über Ressourcen werden in einer Netzstruktur abgebildet. Die kleinste Einheit im RDF bildet hier nicht ein Tupel sondern eine Aussage. Aussagen bestehen immer aus Subjekt (kurz: S), Prädikat (kurz: P) und Objekt (kurz: O). Bezogen auf den Ausschnitt der NCI Ontologie in Abbildung 2 ergeben sich Aussagen wie: Frakturen der unteren Extremitäten (ebenso wie der oberen Extremitäten) sind Unterklassen von Knochenfrakturen. Alternativ läßt sich eine Aussage wie „RDF ist ein W3C-Standard für semantische Netze“ wie in Abb.3 graphisch darstellen.

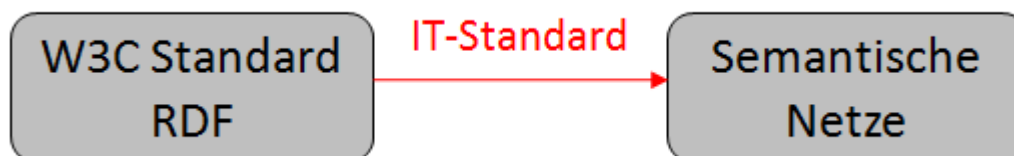


Abb. 2: Eine Aussage als kleinste Einheit des RDF-Standards

RDF-Aussagen werden wegen Ihre S-P-O Struktur auch als *Triples* bezeichnet. Aussagen zusammengefaßt ergeben wie in der NCI Ontologie ein semantisches Netz (siehe Abb. 3).

<sup>4</sup> World Wide Web

<sup>5</sup> <http://www.geneontology.org/>

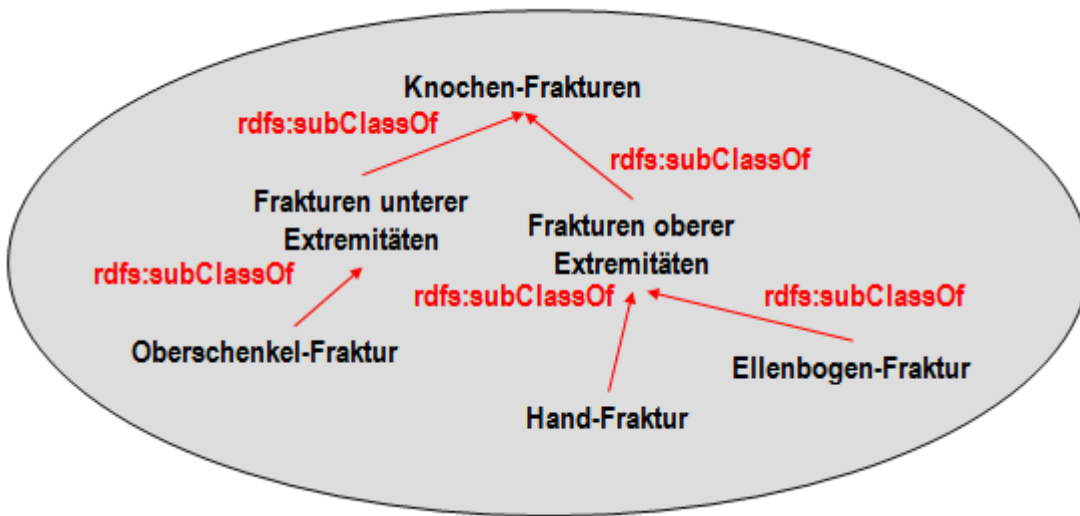


Abb. 3: Ausschnitt aus NCI Ontology (Thesaurus)

Wie arbeitet man nun mit solchen semantischen Netzen? Über das Vorhandensein von Regeln können nun logische Schlussfolgerungen aus der Menge der Aussagen abgeleitet werden. Abbildung 4 zeigt eine solche Regel: „Wenn eine Erkrankung a Unterklasse von Erkrankung b und diese wiederum Unterklasse von Erkrankung c ist, dann ist auch Erkrankung a Unterklasse von Erkrankung c.“

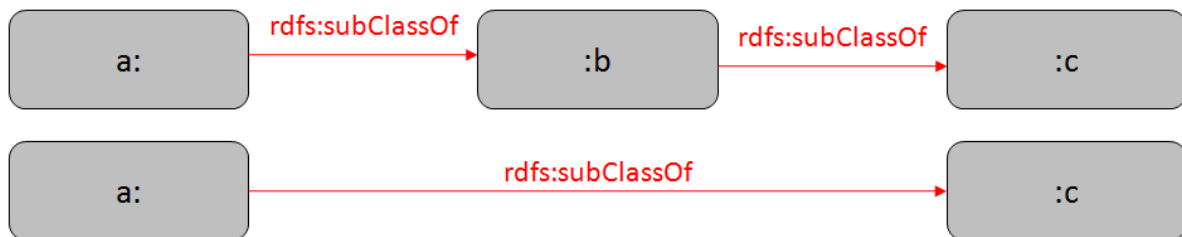


Abb. 4: Regelbasierte logische Schlussfolgerungen (Inferencing)

Für RDF existieren verschiedene Darstellungsformate. Neben der Darstellung als XML Dokument (RDF/XML, *.rdf*) sind N-Triples (*.nt*)<sup>6</sup> oder die Notation 3 (*.n3*) weit verbreitet.

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://de.wikipedia.org/wiki/Resource_Description_Framework">
    <dc:title>Resource Description Framework</dc:title>
    <dc:publisher>Wikipedia - Die freie Enzyklopädie</dc:publisher>
  </rdf:Description>
</rdf:RDF>
```

Abb. 5: Beispiel in XML/RDF Notation<sup>7</sup>

N-Triples stellen semantische Netze als zeilenbasiertes ASCII-Format dar. Jedes Triple repräsentiert eine Zeile und wird durch einen Punkt abgeschlossen.

<sup>6</sup> NCI als *.nt* über Oracle by Example Link (siehe Referenzen) verfügbar

<sup>7</sup> Aus: [http://de.wikipedia.org/wiki/Resource\\_Description\\_Framework#Beispiel](http://de.wikipedia.org/wiki/Resource_Description_Framework#Beispiel)

```

nci_znt
61 <http://www.mindswap.org/2003/nciOncology.owl#Carubicin> <http://www.mindswap.org/2003/nciOncology.owl#Semantic_Type> "Antibiotic" .
62 <http://www.mindswap.org/2003/nciOncology.owl#Carubicin> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#Class> .
63 <http://www.mindswap.org/2003/nciOncology.owl#Carubicin> <http://www.mindswap.org/2003/nciOncology.owl#Synonym> "5,12-Naphthacenedione, 8-acetyl-10-
64 <http://www.mindswap.org/2003/nciOncology.owl#Carubicin> <http://www.mindswap.org/2003/nciOncology.owl#Preferred_Name> "Carubicin" .
65 <http://www.mindswap.org/2003/nciOncology.owl#Carubicin> <http://www.mindswap.org/2003/nciOncology.owl#FULL_SYN> "\n<term-name>Carminomycin</term-n
66 <http://www.mindswap.org/2003/nciOncology.owl#Carubicin> <http://www.mindswap.org/2003/nciOncology.owl#UMLS_CUI> "C0007308" .
67 <http://www.mindswap.org/2003/nciOncology.owl#Carubicin> <http://www.mindswap.org/2003/nciOncology.owl#Synonym> "NSC-180024" .

```

Abb. 6: Beispiel NCI in .nt Notation<sup>8</sup>

Subjekt, Prädikat und Objekt werden durch Leerzeichen voneinander getrennt.

### Anfragesprache für RDF

Mit SPARQL steht eine vom W3C empfohlene graph-basierte Anfragesprache für RDF bereit.

### Semantische Netze und die Oracle Datenbank

Was hat das Ganze nun mit der Oracle Datenbank zu tun?

Diese unterstützt die Speicherung semantischer Netze seit der Version 10g Release 2. Technisch gesehen wird dafür das Netzwerkdatenmodell der Spatial Option zur Datenbank Enterprise Edition (eingeführt mit Version 10g Release 1) genutzt. Daher ist die Lizenzierung dieser Option Voraussetzung für die Nutzung. Voraussetzung ist auch, dass zugehörige Funktionalität, der sogenannte *Semantic Technologies Support* zunächst über ein .sql-Skript freigeschaltet wird, welches sich im \$ORACLE\_HOME/md/admin-Verzeichnis befindet. Eine genaue Beschreibung ist in der *MyOracleSupport* Note 454371.1 (*How To Set Up The Database For Semantic Networks Support*) zu finden.

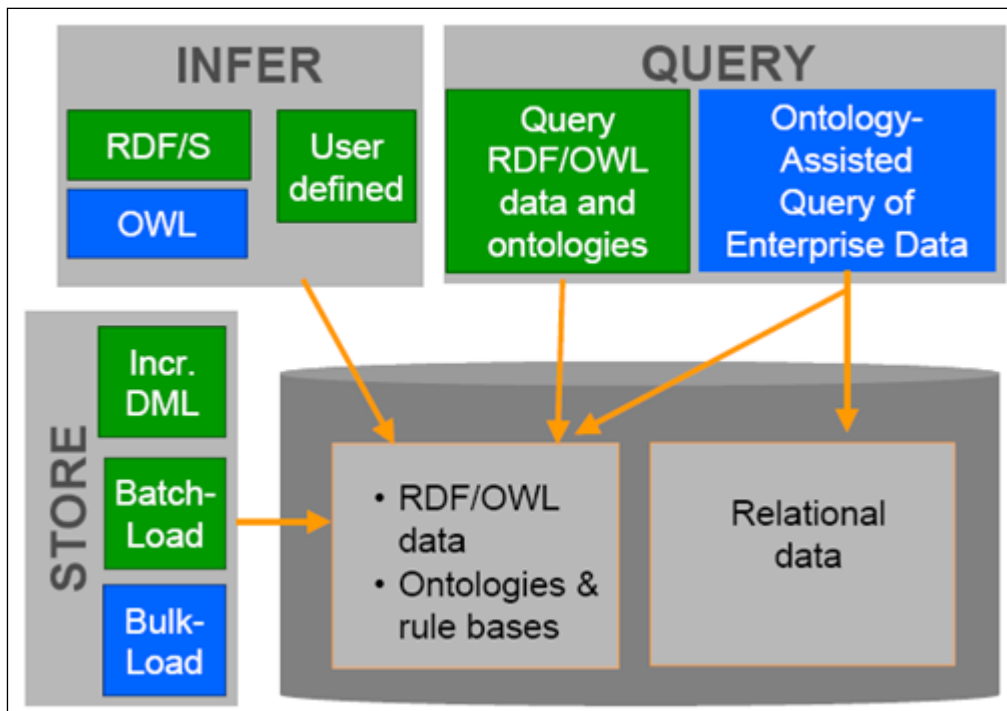


Abb. 7: Semantic Technologies in der Oracle Datenbank

<sup>8</sup> Aus:

[http://www.oracle.com/webfolder/technetwork/tutorials/obe/db/11g/r1/prod/datamgmt/nci\\_semantic\\_network/nci\\_Semantics\\_les01.htm](http://www.oracle.com/webfolder/technetwork/tutorials/obe/db/11g/r1/prod/datamgmt/nci_semantic_network/nci_Semantics_les01.htm)

Wer mit der Oracle Datenbank vertraut ist, weiß, dass auch komplexe Funktionalität über spezielle Datentypen und PL/SQL-Pakete bereitgestellt wird sowie eigene Indizes effiziente Abfragen auf den Informationen ermöglichen.

Für den Umgang mit semantischen Netzen stehen (im Schema MDSYS) u.a. folgende Datentypen, Pakete und Operatoren bereit:

- SDO\_RDF\_TRIPLE\_S  
Datentyp, welcher eine Triple und damit eine RDF-Aussage repräsentiert
- SEM\_APIS  
PL/SQL-Paket zum Einrichten eines semantischen Netzwerkes
- SEM\_MATCH  
Operator, welcher SPARQL-ähnliche Anfragen über SQL in der Oracle Datenbank ermöglicht. Semantisch-indizierte Dokumente können ebenso mit dem SEM\_CONTAINS bzw. SEM\_CONTAINS\_SELECT Man bewegt sich also programmtechnisch weiterhin in der vertrauten Welt von SQL und PL/SQL.
- SEM\_RELATED  
Operator, um die semantischen Beziehungen zwischen Begriffen in einer Ontologie und anderen (z.B. relationalen) Daten festzustellen
- SEM\_DISTANCE  
Operator, welcher aufzeigt, wie weit Begriffen in einer Ontologie voneinander entfernt sind
- SEM\_PERF  
PL/SQL-Paket, welches Werkzeuge zum Überprüfen und Verbessern der Performanz von RDF und OWL beinhaltet
- SEM\_RDFCTX  
PL/SQL-Paket für das Anlegen und Pflegen von semantischen Indizes auf unstrukturierten Inhalten (Dokumenten) in der Datenbank. Externe Datenquellen können hier auch eingebunden werden.

Die Datenbank unterstützt ebenso die bereits weiter oben eingeführte OWL sowie Subsets davon wie (OWL-Lite, OWL-DL und OWL-Full) bzw. neben RDF auch weitestgehend das *Simple Knowledge Organization System* (SKOS) Datenmodell. Dieses kann verwendet werden, um Thesauri, Klassifikationsschemata oder andere Arten von Lexika abzubilden.

### Die ganze Theorie am Beispiel – Ein wenig SQL

Wie wird das Ganze in der Praxis also mittels SQL ausgedrückt? Die nachfolgenden SQL Befehle stellen Beispiele der bisher beschriebenen Funktionalität dar und sind Auszüge bzw. Adaptionen des Tutorials auf *Oracle by Example*.

```
-- Anlegen eines Semantischen Netzes
execute sem_apis.create_sem_network('rdf_users');

-- Tabellen für RDF-Triples anlegen
create table nci_rdf_data (
  id          number,
  triple      sdo_rdf_triple_s
);
-- Semantisches Modell anlegen
execute sem_apis.create_sem_model('nci','nci_rdf_data','triple');

/* Rules index (Entailment) anlegen
 * Objekt, welches vorberechnete Triples enthält, die auf der Basis einer
 * Rulebase ermittelt wurden
 */
```

```

execute sem_apis.create_entailment(
  'nci_idx', sem_models('nci'),
  sem_rulebases('owlprime'),
  0,
  null);

-- Einfache Abfragen
select sem_apis.get_model_id('nci') as model_id from dual;
SELECT a.triple.GET_TRIPLE() AS triple FROM nci_rdf_data a;
SELECT distinct s.triple.GET_SUBJECT() AS subject FROM nci_rdf_data s;
SELECT distinct p.triple.GET_PROPERTY() AS property FROM nci_rdf_data p;
SELECT distinct o.triple.GET_OBJECT() AS obj FROM nci_rdf_data o;

-- Semantischen Index auf einer Patienten-Tabelle anlegen
create index nciIndex on patients_data(diagnosis)
  indextype is mdsys.sem_indextype
  parameters ('ONTOLOGY_MODEL(NCI), RULEBASE(owlprime)');

-- Abfrage auf alle Triples, die 'Finger_fracture' als Subject haben
select p, o
from table(
  sem_match(
    '<http://www.mindswap.org/2003/nciOncology.owl#Finger_Fracture> ?p ?o'
  ),
  sem_models('nci'),
  null,
  null,
  null));

-- Abfrage aller Objekte, die subClassOf 'Finger_fracture'
-- sind unter Berücksichtigung von owlprime
select o
from table(
  sem_match (
    '<http://www.mindswap.org/2003/nciOncology.owl#Finger_Fracture>
<http://www.w3.org/2000/01/rdf-schema#subClassOf> ?o)',
  sem_models('nci'),
  sem_rulebases('owlprime'),
  null,
  null));

-- Tabelle Patients_Data mit Model in Beziehung setzen.
-- Abfrage aller Diagnosen, die subClassOf Upper_Extremity_Fracture sind
select
  id,
  name,
  round(age) as age,
  diagnosis,
  SEM_DISTANCE(123)
from patients_data
where sem_related(
  diagnosis,
  '<http://www.w3.org/2000/01/rdf-schema#subClassOf>',
  '<http://www.mindswap.org/2003/nciOncology.owl#Upper_Extremity_Fracture>',
  sem_models('nci'),
  sem_rulebases('owlprime'), 123) = 1

```

```

ORDER BY SEM_DISTANCE(123) asc;
-- 116 records

-- Im Vergleich dazu:
select * from patients_data
where diagnosis like '%Upper_Extremity_Fracture%';
-- 14 records

ALTER USER scott
IDENTIFIED BY User_Name
REPLACE tiger

```

## Fazit

Mit der Integration von *Semantic Technologies* Funktionalität in die Oracle Datenbank hat sich die Leistungsfähigkeit dieser noch einmal wesentlich gesteigert. Geht es doch nun nicht mehr nur darum, Informationen auch komplexerer Art (wie Geodaten, XML-Dokumente oder Metadaten medizinischer Bilder) mit den geläufigen relationalen Informationen zu "verschneiden". Jetzt eröffnen sich zusätzlich Möglichkeiten, semantische Beziehungen zwischen Informationen abzubilden und auf der Basis von Regelwerken neues Wissen abzuleiten.

## Auflösung der Fragen

Lukas Cranach d.Ä., Johannes Gensfleisch zum Gutenberg und Erasmus von Rotterdam wurden vom Maler Werner Tübke in seinem 123 Meter langem und 14 Meter hohem Geschichtspanorama porträtiert, welches die Schlacht von Bad Frankenhausen im Jahr 1525 im wahrsten Sinne des Wortes „nachzeichnet“?

Friedrich der III., Herzog von Preussen ist der spätere Friedrich der I., erster König von Preussen. Es handelt sich also um ein und dieselbe Person.

Über wieviel „Ecken“ Carsten Czarski mit Larry Ellison bekannt ist, kann er möglicherweise selbst beantworten. Das Prinzip, welches hier Anwendung findet, ist aus Sozialen Netzwerken als *Friend-of-a-Friend* (kurz: FOAF) bekannt

## Referenzen

Oracle Dokumentation: Semantic Technologies Developer's Guide

[http://download.oracle.com/docs/cd/E11882\\_01/appdev.112/e11828.pdf](http://download.oracle.com/docs/cd/E11882_01/appdev.112/e11828.pdf)

Oracle By Example: RDF Semantic Data Management Using the Oracle Spatial 11g Option

[http://www.oracle.com/webfolder/technetwork/tutorials/obe/db/11g/r1/prod/datamgmt/nci\\_semantic\\_network/nci\\_Semantics\\_les01.htm](http://www.oracle.com/webfolder/technetwork/tutorials/obe/db/11g/r1/prod/datamgmt/nci_semantic_network/nci_Semantics_les01.htm)

World Wide Web Consortium: Standards and Drafts (u.a.OWL, RDF, SPARWL)

<http://www.w3.org/TR/>

Jena2 Toolkit

<http://jena.sourceforge.net>

## Kontaktadressen:

Karin Patenge

Oracle Deutschland B.V. & Co. KG

Schiffbauergasse 14

D-14467 Potsdam

Telefon: +49 (0)331-200 7214

Fax: +49 (0)331-200 7559

E-Mail [karin.patenge@oracle.com](mailto:karin.patenge@oracle.com)

Internet: [www.oracle.com](http://www.oracle.com)

Carsten Czarski  
Oracle Deutschland B.V. & Co. KG  
Riesstraße 25  
D-80992 München

Telefon: +49 (0)89-1430 2977  
Fax: +49 (0)89-1430 2977  
E-Mail [carsten.czarski@oracle.com](mailto:carsten.czarski@oracle.com)  
Internet: [www.oracle.com](http://www.oracle.com)