

Recherchebericht

Thema: SemanticChess
Gruppe: swp13-sc
Verantwortlich: M. Kisilowski
Letzte Änderung: 04.01.13

Inhaltsverzeichnis

1. Begriffe

- 1.1. Framework
- 1.2. Client
- 1.3. Dienst
- 1.4. Server
- 1.5. Apache
- 1.6. Apache Wicked
- 1.7. URI
- 1.8. IRI
- 1.9. Literal
- 1.10. RDF
- 1.11. RDFS
- 1.12. Ontologie
- 1.13. PGN
- 1.14. Konvertierung
- 1.15. Dateiformat
- 1.16. Datenformat
- 1.17. OWL
- 1.18. CGI
- 1.19. JSP
- 1.20. DBpedia
- 1.21. Open Link
- 1.22. Middleware
- 1.23. Universal Data Access Middleware
- 1.24. Abfragesprache
- 1.25. GWT
- 1.26. Spring
- 1.27. Schach-Engine
- 1.28. Elo-Rang

2. Konzepte

- 2.1. Serverkomponenten
 - 2.1.1. PHP
 - 2.1.2. Java-Servlet
 - 2.1.3. IIS
- 2.2. Semantic Web
- 2.3. RDF-Umsetzung
- 2.4. Konvertierung von PGN
- 2.5. Linking mit DBpedia
- 2.6. SPARQL

3. Aspekte

- 3.1. Serverkomponenten
- 3.2. Turtle
- 3.3. RDF/XML
- 3.4. RDF/JSON
- 3.5. Weitere Ontologien
- 3.6. DBpedia
- 3.7. Houdini (Schachengine)

4. Quellen

1. Begriffe

1.1. Framework

Ein Framework ist eine Grundstruktur oder Programmiergerüst, das dem Programmierer einen Entwicklungsrahmen für die Erstellung eines Anwendungsprogramms zur Verfügung stellt, wobei die Wiederverwendbarkeit von architektonischen Mustern genutzt wird, weshalb ein Framework domänenspezifisch oder auf einen einzelnen Anwendungstypen beschränkt ist.

1.2. Client

Ein Client ist eine natürliche Person, die über einem Browser eine Anfrage auf eine Website an einen Server stellt.

1.3. Dienst

Ein Dienst ist ein Programm, das über fest definierte Schnittstellen Eingaben in Ausgaben umwandelt und dabei eine oder mehrere bestimmte Funktionalitäten erfüllt.

1.4. Server

Ein Server, Auch Webserver, ist ein Rechnersystem auf dem ein Programm läuft, ein sogenannter „Dienst“, der Anfragen von Clients beantwortet.

1.5. Apache

Die „Apache Software Foundation“ bietet Unterstützung finanzieller, rechtlicher und organisatorischer Art für viele Open-Source-Projekte, wie dem „Apache http Server“-Projekt und vielen weiteren.

1.6. Apache Wicket

Das Apache Wicket oder kurz Wicket ist eine leichtgewichtete Web-Applikations-Framework für die Programmiersprache Java. Es ähnelt Spring in seiner Absicht und wird hier als Alternative zum Spring-Framework angeführt. Vorteile von Wicket bestehen in einem konsequenten Zwang objektorientiert zu programmieren, welcher im Nachhinein größerer Wiederverwendbarkeit ermöglicht.

1.7. URI((Uniform Resource Identifier)

Ein URI ist ein global eindeutiger Identifikator von Ressourcen (Ressource = Objekt mit einer im Kontext bedeutungsvollen Identität). Gewöhnlich bestehen URIs aus den fünf Teilen scheme (Schema), authority (Anbieter), path (Pfad), query (Abfrage) und fragment (Teil), wobei jede URI mindestens über scheme und path verfügen müssen.

Beispiele: URLs, ISBN,

<http://www.beispiel.org/Bleispiel>

1.8. IRI((Internationalized Resource Identifier)

Ein IRI ist eine Verallgemeinerung der URIs. IRIs können ASCII Sonderzeichen beinhalten. Beispiel: <http://www.beispiel.org/Öäüß>

1.9. Literal

Literale sind Datenwerte, die keine separate Existenz besitzen. Als Literale lassen sich zum Beispiel Wahrheitswerte, Zahlen oder Datumsangaben mit Angabe eines Datentyps spezifizieren und interpretieren (keine Angabe des Datentyps führt zur Interpretation als Zeichenkette).

1.10. RDF(Ressource Description Framework)

RDF ist eine Kernkomponente des Semantic Web und W3C (World Wide Web Consortium, Gremium zur Standardisierung von Webtechnologien) Empfehlung, dient der Repräsentation von Informationen mittels einer Menge von Tripeln (Subjekt, Prädikat, Objekt), wobei Subjekt URI oder leerer Knoten sein darf, Prädikat URI sein muss und Objekt URI, leerer Knoten oder Literal sein kann. Die Tripelmeng kann durch gerichtete Graphen (RDF-Graphen) dargestellt werden.

1.11. RDFS(Ressource Description Framework Schema)

RDFS ist ebenfalls Empfehlung des W3C. RDF legt nur einen Syntax für den gemeinsamen Datenaustausch fest, mit RDFS lassen sich die Ressourcen aus RDF strukturieren (durch Klassen), wodurch einfache Ontologien dargestellt werden können.

- 1.12. Ontologie**
Explizite formale, maschinenlesbare Spezifikation einer Konzeptualisierung (Beziehungen, Ableitungsregeln) zu einem bestimmten Wissensbereich. Ontologien ermöglichen es also ein Netzwerk von Informationen mit logischen Relationen darzustellen, innerhalb dem logische Schlussfolgerungen sowie Gewährleistung der Gültigkeit möglich sind.
- 1.13. PGN**
Portable Game Notation (PGN) ist ein Datenformat zur Speicherung von Schachpartien. Es wurde entwickelt, um den Austausch von Schachdaten zwischen verschiedenen Schachprogrammen zu ermöglichen und zu vereinfachen.
- 1.14. Konvertierung**
Der Begriff Konvertierung bezeichnet die Überführung einer Datei von einem Dateiformat in ein anderes.
- 1.15. Dateiformat**
Ein Dateiformat definiert die Syntax und Semantik von Daten innerhalb einer Datei.
- 1.16. Datenformat**
Ein **Datenformat** ist eine Spezifikation der Datenverarbeitung, die festlegt, wie Daten beim Laden, Speichern oder Verarbeiten programmtechnisch zu interpretieren sind.
- 1.17. OWL (Web Ontology Language)**
Dient der Erstellung und Definition von Ontologien mittels einer formalen Semantik. Baut auf RDF/RDFS-Syntax auf, ist aber mächtiger, da weitere Ausdrucksmittel zur Verfügung gestellt werden.
- 1.18. CGI(Common Gateway Interface)**
CGI ist Standard für Datenaustausch zwischen Webserver und „dritter Software“ (externe Software, die nicht zum Programmpaket des Webserver gehört). Stellt dem Server eine Laufzeitumgebung zur Verfügung, die eine dynamische/interaktive Anfragebeantwortung ermöglicht, z. B. durch CGI-Skripte oder andere Programme.
- 1.19. JSP(JavaServer Pages)**
JavaServer Pages ist eine Skriptsprache, die in HTML und XML eingebettet wird, analog zu PHP, jedoch für Java-Code oder JSP-Funktionen.
- 1.20. DBpedia**
DBpedia ist ein Gemeinschaftsprojekt der *Uni-Leipzig*, der *Freien Universität Berlin* und *OpenLink Software* um strukturierte Informationen aus der Wikipedia zu extrahieren und diese Webanwendungen zugänglich zu machen.
Außerdem ermöglicht DBpedia diese Daten mit Informationen aus anderen Web-Anwendungen zu verbinden.
Kurz: DBpedia ist die RDF-Version der Wikipedia
- 1.21. Open Link**
Ist ein Softwareunternehmen, welches 1992 gegründet wurde und seinen Hauptsitz in den USA hat.
Es ist „Erfinder und Besitzer“ der *Universal Data Access drivers suite*
Open Link entwickelt *middleware* Produkte.
- 1.22. Middleware**
Bezeichnet eine Art von Computersoftware, welche verschiedene Dienste als Teil einer Softwareapplication zur Verfügung stellt.
Wird auch als „Software-glu“ bezeichnet.
- 1.23. UDAM(Universal Data Access Middleware)**
UDAM definiert eine Softwarekategorie, welche transparenten Zugang zu Datenquellen (wie z.B. Datenbanken) für verschiedene Anwendungen bereitstellt.
(z.B. Desktopanwendungen, Entwicklungsumgebungen, Web & Internetpräsenzen)

1.24. Abfragesprache

Eine Abfragesprache ist eine formale Sprache um Informationen (Query) aus einer Informationsquelle zu erhalten. Diese Abfragesprache kann unterschiedlichste Eigenschaften besitzen (z.B. systemnah, benutzerfreundlich). Abfragesprachen sind zum Beispiel SQL, XQuery, Versa, SPARQL. Sie unterscheiden sich u.a. von der Art der Informationsquelle, welche als XML, RDF o. ä. Vorliegen kann.

1.25. GWT(Google Web Toolkit)

Der Namensgebung entsprechend ist das Toolkit zur Erstellung von Webanwendungen konzipiert und wurde von Google als freie Software unter der Apache Lizenz 2.0 veröffentlicht. Eine Besonderheit des GWT ist ein Java-nach-JavaScript-Compiler, welche die vollständige Entwicklung eines Programms in Java unterstützt. Außerdem enthält es auch eine Integration von JUnit und ein kleines Widget-Paket zur GUI-Gestaltung, welches sich an Swing anlehnt.

1.26. Spring

Spring ist ein quelloffenes Framework zur vereinfachten Entwicklung von Java Applikationen und der Förderung guter Programmierpraktiken. Die Entkopplung der Applikationskomponenten ist das Ziel von Spring. Gewährleistet wird dies durch das Zusammenspiel unterschiedlichster Plattformen und Tools. Auch die Jena API besitzt eine Spring-konforme Anbindung. Eine einmal mit Spring erstellte Applikation besitzt danach die vereinfachte Möglichkeit Module zu entfernen, hinzuzufügen oder zu aktualisieren.

1.27. Schach-Engine

Eine Schach-Engine ist ein Bestandteil eines Schachprogramms, die für die Berechnung bzw. Kalkulation bestimmter Züge verantwortlich ist. Eine Engine läuft immer im Hintergrund und ist somit nicht von der Steuerung der Benutzer abhängig.

1.28. Elo-Rang

Der Elo-Rang ist eine Einstufung, mit der man die Spielstärke eines Schachspielers messen kann. Je höher der Elo-Rang, desto besser ist ein Spieler. Schach-Engines besitzen auch einen solchen Elo-Wert, mit dem man die Qualität einer Engine bestimmen kann.

2. Konzepte

2.1. Serverkomponenten

2.1.1. Apache Server

In diesem Kontext der „Apache HTTP Server“ gemeint, ein quelloffenes und kostenloses Programmpaket, das für viele Betriebssysteme verfügbar und durch verschiedene Module (Kryptographie, etc.) erweiterbar ist. „Client“-Anfragen werden statisch (einfache HTML-Seiten) oder dynamisch (generiert durch Skriptsprachen wie PHP) beantwortet.

2.1.2. Tomcat

Tomcat ist ebenfalls ein Apache Produkt. Stellt eine Umgebung zur Ausführung von Java-Code („Servlets“ und „JavaServer Pages“) dar. Wird aus Performance-Gründen meist in den „Apache HTTP Server“ integriert, kann zu Testzwecken aber auch allein ausgeführt werden. Alternativ kann auch „Jetty“ genommen werden, ein kleiner in Java geschriebenes Server zur Ausführung von Servlets.

2.1.3. IIS

„Microsoft Internet Information Services“, ist die Serversoftware von Microsoft, für ASP und .NET-Anwendungen, aber über Erweiterungen auch mit PHP oder JSP nutzbar

2.2. Webtechnologien

2.2.1. PHP

2.2.2. Java Servlet

Java-Klassen, die in einem Webserver ausgeführt werden und „Client“-Anfragen dynamisch beantworten können.

2.2.3. JavaScript

JavaScript ist eine Plattform-unabhängige, dafür jedoch Browser-abhängige Programmiersprache, welche keinesfalls mit Java verwechselt werden darf. Sie findet die meiste Verbreitung zur Bereitstellung interaktiver und dynamischer Inhalte im Web ist dabei allerdings meist auf kleinere Applikationen ausgerichtet.

2.2.4. Ajax

Ajax ist keine Programmiersprache im eigentlichen Sinne. Sie ermöglicht Daten zu senden und zu empfangen, ohne das die gerade betrachtete Website neu geladen wird. Es entsteht eine Interaktivität des Webauftritts, die den meisten Nutzern aus *Google Suggest* bekannt ist. Dabei werden (XML-)Daten im Hintergrund mit JavaScript übertragen und dann dort mithilfe von JavaScript und DOM eingefügt.

2.3. Semantic Web

Das semantische Web ist als Weiterentwicklung des heutigen Web 2.0 angedacht. Aufgrund der hohen (vom Nutzer generierten) Daten- und Informationsfülle, erscheint es sinnvoll Maschinen die Bedeutung (Semantik) der Daten und Informationen beizubringen, damit sie diese automatisiert verarbeiten können. Dies geschieht unter Angabe von zusätzlicher Information und Verlinkung innerhalb von für Maschinen verständlichen Maßstäben. Das Semantic Chess Projekt soll Mechanismen des semantischen Webs nutzen um so der Masse an Informationen über Schachpartien gerecht zu werden.

2.4. RDF-Umsetzung

Ziel des Projektes ist es Informationen über eine Fülle an Schachpartien strukturiert nach bestimmten Kriterien auszuwerten. Das herkömmliche PGN-Format ist dafür nicht sonderlich gut geeignet, weshalb eine Ontologie für Schachpartien entwickelt werden soll, so dass sie in einer RDF-Struktur vorliegen.

Zur Darstellung von RDF-Strukturen und Beschreibung der Schachontologie mit denen die Schachpartien eindeutig beschrieben sein sollen, stehen eine Reihe von Serialisierungen mit unterschiedlichen Vor- und Nachteilen zur Verfügung, die unter dem Gliederungspunkt Aspekte mit ihrer Syntax kurz unter Berücksichtigung der Annahme über die Relevanz für das Projekt vorgestellt werden sollen.

2.5. Konvertierung von PGN

Da PGN nicht proprietär ist, kann man die Daten einer PGN Datei mit fast jedem beliebigen Texteditor lesen. Ich schlage vor, dass man die PGN-Dateien sinnhaft konvertiert.

Bei dieser Konvertierungsart werden Daten „dem Sinne nach“ konvertiert, also unter Beibehaltung des als wesentlich betrachteten Inhaltes. Dies kann mit oder ohne Informationsverlust geschehen, oder es können auch Informationen aus zusätzlichen Quellen durch die Konvertierung angereichert werden.

2.6. Linking mit DBpedia

Linking mit DBpedia bedeutet spezifische Informationen aus der DBpedia für ein Anwendungsprogramm nutzbar zu machen.

Dies funktioniert im allgemeinen über die Abfrage der Daten aus der DBpedia und deren Verknüpfung mit der jeweiligen Anwendung.

DBpedia extrahiert strukturierte Daten/Informationen aus der Wikipedia um Programmen Zugang zu diesen zu gewährleisten. Dieser Zugang wird durch verschiedene Semantic Web Technologien erzeugt.

DBpedia-Daten können über SPARQL oder Linked Data abgefragt werden.

2.7. SPARQL

SPARQL (SPARQL Protocol And RDF Query Language) ist eine Standard Abfragesprache für RDF Formate. SPARQL ist nun von der W3C ein (endgültig) Standardisierter Nachfolger vieler RDF Query Languages.

3. Aspekte

3.1. Serverkomponenten

Es besteht die Möglichkeit eine rein Java-orientierte Lösung anzubieten, da viele Komponenten bereits auf Java setzen und fehlende Teile auch in Java implementierbar wären. Ob dies sinnvoll ist, hängt davon ab, wie gut sich bestimmte Probleme in den jeweiligen Technologien lösen lassen und welche Ziele weiterhin gesetzt werden.

3.2. Turtle

Turtle ist ein vergleichsweise simpel und benutzerfreundliches, für Menschen gut lesbar gehaltenes Format zur maschinenlesbaren Darstellung von RDF-Graphen.

Syntaktisch werden URIs in spitzen Klammern angegeben

(<http://www.beispiel.org/Bleispiel>), Literale in

Anführungszeichen („Musterhausen“@de, „42.132344“^^xsd:float) und je ein Tripel endet mit einem Punkt. Turtle ermöglicht es Präfixe festzulegen, so dass nicht die komplette URI ausgeschrieben werden muss (@prefix bsp:

<http://www.beispiel.org/>. So dass bsp: Bleispiel der obigen URI

entspricht).

Tripel mit gleichem Subjekt, sowie mit gleichem Subjekt und Prädikat, lassen sich unter Verwendung von ; (für Subjekte) und , (für Prädikate) zusammenfassen.

@prefix bsp: <http://www.beispiel.org/>.

bsp: Musterhausen bsp:locatedIn bsp:Musterland,

bsp:Musterkontinent; bsp:hatMueller bsp:Max_Mustermann .

3.3. RDF/XML

Der RDF/XML Syntax ist zwar weniger gut lesbar und wesentlich umständlicher, aber hat sich in vielen Projekten bewährt, weil eine bessere Tool-Unterstützung und eine größere Menge an Programmibliotheken existiert als für die Alternativen. Für unsere Zwecke wird diese Tool-Unterstützung voraussichtlich nicht notwendig sein, trotzdem soll der Syntax hier grob angerissen werden. Wie in XML üblich werden Namensräume genutzt um Tagnamen zu zerlegen, wobei RDF eigene tags den Bezeichner rdf besitzen. Das Subjekt wird als rdf:Description Element kodiert, die URI, wird dabei im about-Attribut angegeben. Prädikate werden durch die im rdf:Description Element enthaltenen Elemente beschrieben. Sie enthalten wiederum das Objekt innerhalb des Attributes rdf:resource.

Literale werden als Freitext innerhalb des Prädikatenelements angegeben.

(Beispiel in der 2. Vorlesung, 17.4.2012, des Semantic Web Moduls – Seite 28 der Uni Leipzig)

3.4. RDF/JSON

JSON stellt ein kompaktes Datenformat zwischen Anwendungen dar und ist

Programmiersprachen unabhängig, da für die gängigsten Sprachen Parser

existieren. RDF/JSON Objekte folgen der Syntax {„S“: {„P“: [O]}}. Das Objekt wird

dabei wie folgt angegeben: Type (URI, Literal, leerer Knoten,

kleingeschrieben), Value (sollte als gesamte URI angegeben werden, beschreibt

Datenwert), Lang (optional: Sprache des Literals, darf – wenn angegeben - nicht

leer sein) und Datatype (optional: URI des Datentyps des Literals). Die Syntax ist

von der Komplexität her zwischen Turtle und RDF/XML einzuordnen. (Beispiel in der 2. Vorlesung, 17.4.2012, des Semantic Web Moduls – Seite 34 der Uni Leipzig)

3.5. Weitere Ontologien

Als optionale zusätzliche Aufgabe wird gefordert die Suche speziell um

Eröffnungen zu erweitern sowie die Möglichkeit sich beim Browsen die Namen

der Eröffnungspartien anzeigen zu lassen. Das Einbinden dieser Funktionalität

kann durch Erzeugen einer zusätzlichen Ontologie bzw. erweitern der bestehenden erfolgen.

3.6. DBpedia

Es ist durchaus angebracht mit der deutschen Version der DBpedia zu arbeiten. Jedoch ist diese bei weitem nicht so ausgebaut wie die Englische. Das arbeiten mit der deutschen Version erfordert möglicherweise in Voraus, neue Mappings zwischen der deutschen Wikipedia und der DBpedia zu erstellen um den Zugriff auf bestimmte Informationen sicherstellen zu können. Im Zuge dessen ist es erforderlich sich mit DBpedia Mapping auseinander zu setzen.

3.7. Houdini

Houdini ist das zur Zeit stärkste und kostenlos erhältliche Programm weltweit. Es war lange umstritten, da der Programmierer des Programms Rybka behauptete, dass der Sourcecode von ihm gestohlen wurde und dadurch Houdini und andere Programme entstanden sind. Später stellte sich heraus, dass das Programm Rybka ein Plagiat von **Fruit**, ebenfalls ein sehr etabliertes Schachprogramm, war. Somit waren alle Zweifel um Houdini beseitigt.

4. Quellen

Server:

http://de.wikipedia.org/wiki/Server_%28Software%29
http://de.wikipedia.org/wiki/Microsoft_Internet_Information_Services

Apache:

<http://httpd.apache.org/>
<http://www.apache.org/foundation/>
http://de.wikipedia.org/wiki/Apache_Software_Foundation
http://de.wikipedia.org/wiki/Apache_HTTP_Server

Tomcat:

<http://tomcat.apache.org/>
<http://www.oracle.com/technetwork/java/overview-137084.html>
<http://www.jsptutorial.org/content/introduction>
<https://www.google.de/search?q=tomcat+und+apache+httpd&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:de:official&client=firefox-a>

RDF-Triple-Stores:

<http://en.wikipedia.org/wiki/Triplestore>

Jena TDB:

<http://jena.apache.org/documentation/tdb/index.html>
<http://jena.apache.org/>

BigOwlin:

<http://www.ontotext.com/owlim/editions>
<http://owlim.ontotext.com/display/OWLIMv53/OWLIM-Lite+Installation>

Sesame:

<http://www.openrdf.org/about.jsp>

OpenLink Virtuoso:

<http://virtuoso.openlinksw.com/>
<http://docs.openlinksw.com/virtuoso/installation.html>

Webtechnologien

http://de.wikipedia.org/wiki/Google_Web_Toolkit
http://en.wikipedia.org/wiki/Spring_Framework
http://en.wikipedia.org/wiki/Apache_Wicket
<https://github.com/AlBaker/SpringJena>
<http://www.alonsoruibal.com/dev/chess/>

Abfragesprachen

http://de.wikipedia.org/wiki/Resource_Description_Framework#Abfragen
<http://de.wikipedia.org/wiki/SPARQL>
http://en.wikipedia.org/wiki/RDF_query_language
<http://jars.de/web/semantische-suche-mit-sparql-und-dbpeda>
<http://seaborne.blogspot.de/2006/04/from-rdql-to-sparql.html>
<http://www.gi.de/nc/service/informatiklexikon/detailansicht/article/anfragesprachen-fuer-das-semantic-web/druckversion.html>

Semantic Web

Vorgängerprojekte
Vorlesung „Semantic Web“
http://de.wikipedia.org/wiki/Semantisches_Web

PGN und Konvertierung

http://de.wikipedia.org/wiki/Portable_Game_Notation
[http://de.wikipedia.org/wiki/Konvertierung_\(Informatik\)](http://de.wikipedia.org/wiki/Konvertierung_(Informatik))