

Recherchebericht

DTP-16: LD Hypstrr - Datadriven Template Publication

7. Januar 2016

INHALTSVERZEICHNIS

1	Glossar	3
1.1	FOAF	3
1.2	Graph	3
1.3	HTML	3
1.4	Linked Data	3
1.5	Literal	3
1.6	Metadaten	3
1.7	Namespaces bzw. Namensräume	3
1.8	Notation3 bzw. N3	3
1.9	Ontologie	4
1.10	OntoWiki	4
1.11	SPARQL	4
1.12	RDF	4
1.13	RDFS	4
1.14	URI	4
1.15	Turtle	5
1.16	Triple-Store	5
1.17	W3C	5
2	Konzepte	5
2.1	Semantic Web	5
2.2	Resource Description Framework	5
2.3	Templating-Systeme	6
2.4	Static Site Generator	6
3	Aspekte	6
3.1	Ziel des Projekts	6
3.2	Static Site Generatoren	7
3.3	Templating-Systeme	7
3.4	RDF-Libraries	8
4	Quellen	8

1 GLOSSAR

1.1 FOAF

FOAF steht für "friend of a friend" und ist ein Projekt zur maschinenlesbaren Modellierung sozialer Netzwerke durch ein RDF-Schema, welches Klassen und Eigenschaften definiert. FOAF ist eines der ersten Anwendungen von Semantic-Web-Technologien.

1.2 GRAPH

Ein Graph ist eine abstrakte Struktur, die eine Menge von Objekten (Knoten) und deren Verbindungen (Kanten) darstellt. Eine Kante existiert immer nur zwischen genau zwei Knoten, ist also eine paarweise Verbindung, und kann gerichtet oder ungerichtet sein.

1.3 HTML

(kurz für: Hypertext Markup Language) ist eine vom W3C entwickelte Markup Sprache, die Texten und Daten Darstellungsformen und Eigenschaften zuweist. HTML-Daten setzen sich aus verschachtelten Tags zusammen und sind Multimedia-kompatibel.

1.4 LINKED DATA

Linked Data ist eine Methode zum Veröffentlichenden strukturierter Daten. Damit werden sie miteinander verknüpft und werden durch semantische Anfragen noch hilfreicher.

1.5 LITERAL

Ein Literal ist im informationstechnischen Sinne eine Folge von Zeichen, die zur direkten Darstellung der Werte von Datentypen zulässig ist. Dies gilt auch im RDF-Modell, wo eine Ressource nur durch eine andere Ressource oder ein Literal beschrieben werden kann.

1.6 METADATEN

Metadaten sind Daten, die Informationen über Eigenschaften anderer Dateien enthalten, jedoch nicht die Datei selbst. So gehört zum Beispiel zu den Metadaten eines Kinofilms der Name des Regisseurs, des Produktionsstudios, etc.

1.7 NAMESPACES BZW. NAMENSRÄUME

Namespaces ermöglichen es, Elemente mit gleichem Namen eindeutig voneinander zu unterscheiden, indem die Elemente und Attribute URIs zugewiesen werden.

1.8 NOTATION3 BZW. N3

Notation3 ist eine formale Sprache zur Beschreibung von semantischen Daten. Das bekannte Turtle bildet eine Untermenge.

1.9 ONTOLOGIE

Eine Ontologie (auch: Vokabular) ist eine sprachlich strukturierte Darstellung von Wissen. Es wird genutzt, um Beziehungen zwischen einzelnen Begriffen darzustellen. Dies ermöglicht mit einem standardisierten Vokabular auch eine Form der digitalen Verwendbarkeit in verschiedenen Systemen und erleichtert das logische Verstehen der Daten.

1.10 ONTOWIKI

OntoWiki ist eine Wiki-artige Software, die das Bearbeiten und Anzeigen von semantischen Daten zulässt. Die Besonderheit besteht darin, Komplexität der Datenspeicherung zu abstrahieren. Die Bedienung erfolgt Wiki-typisch über eine Weboberfläche.

1.11 SPARQL

SPARQL ist eine Abfragesprache für RDF-Daten. Daten können von verschiedenen RDF-Datenquellen in einer Anfrage abgefragt werden. Um die Lesbarkeit zu erhöhen ist es möglich Präfixe zu definieren. Als Ergebnis wird ein RDF-Graph oder ein Ergebnisset geliefert.

1.12 RDF

RDF (kurz für: Resource Description Framework) ist ein System zur Beschreibung von Ressourcen. Es ist eine technische Herangehensweise im Internet zur Formulierung logischer Aussagen über beliebige Ressourcen. Im RDF Modell besteht jede Aussage aus 3 Einheiten: Subjekt, Prädikat und Objekt.

1.13 RDFS

RDFS (kurz für: Resource Description Framework Schema) ist eine von der W3C entwickelte Ontologie-Beschreibungssprache für RDF-Daten. Mit Hilfe von RDFS lassen sich RDF-Ressourcen semantisch durch Eigenschaften und Relationen untereinander beschreiben, wodurch man Ontologien erzeugen kann. RDFS basiert auf der Idee eines mengentheoretischen Klassenmodells. Besitzt man zum Beispiel eine RDF Instanz von Fahrzeug und eine von Auto, so kann man mit RDFS festlegen, dass Fahrzeug eine Klasse und Auto eine Subklasse von Fahrzeug ist.

1.14 URI

URI (kurz für: Uniform Resource Identifier) ist ein Identifikator und besteht aus einer Zeichenfolge, die zur Identifizierung einer abstrakten oder physischen Ressource dient. URIs werden zur Bezeichnung im Internet und dort besonders im World Wide Web genutzt. URIs können als Zeichenfolge in digitalen Dokumente eingebunden werden ("Hyperlink"). Aufbau: scheme, authority, path, query, fragment.

1.15 TURTLE

Turtle steht für "Terse RDF Triple Language" und ist eine Serialisierung für RDF-Graphen. Turtle-Dateien enden auf `.ttl`.

1.16 TRIPLE-STORE

Ein Triple-Store (auch: RDF-Store) ist eine Datenbank zum Speichern und Verwenden von Tripeln durch semantische Anfragen. Triple-Stores speichern wie relationale Datenbanken Informationen in Tripeln, jedoch wurden Triple-Stores dafür optimiert. Tripel können durch RDF oder andere Formate importiert und exportiert werden. Wenn man dem Tripel einen Namen gibt, hat man einen Quad-Store oder einen sogenannten "named graph".

1.17 W3C

W3C (kurz für: World Wide Web Consortium) ist ein internationales Gremium, dessen Ziel es ist, durch zukunftsorientierte Web-Standards in Form von Protokollen und Richtlinien die gesamten Kapazitäten des World Wide Webs zu erschließen und dessen langfristiges Wachstum zu gewährleisten. Es verfolgt genauer die Vorteile des Internets für jedermann zugänglich zu machen, unabhängig von Soft-, Hardware, Sprache, Kultur, etc., sowie den Aufbau eines semantischen Webs.

2 KONZEPTE

2.1 SEMANTIC WEB

Das Web an sich ist weitgehend unstrukturiert und besteht für Maschinen aus bedeutungslosen Zeichenketten. Damit sind diese Informationen schwer für Maschinen verarbeitbar. Die Erweiterung des Webs um das Semantic Web erlaubt eine bessere Maschinenverarbeitbarkeit: Anstatt Aussagen nur in einer natürlichen Sprache auszudrücken, können viele Aussagen, die eine bestimmte Struktur haben, auch maschinenlesbar ausgedrückt werden. Dies erlaubt einer Maschine strukturell zu schlussfolgern.

Ein Beispiel: Gegeben die Aussagen »Alle Katzen sind Tiere.«, »Tom ist eine Katze.« kann die Maschine folgende Aussage schlussfolgern: »Tom ist ein Tier.«. Diese Schlussfolgerung wäre nicht möglich gewesen, wenn die Maschine die Aussagen nur als Zeichenketten, ohne Erkenntnis über ihre innere Struktur, interpretiert hätte. Die Idee ist, dass Aussagen in dieser maschinenlesbaren Form mitgeliefert werden und damit aus der Menge aller Aussagen eine Wissensbasis gebaut werden kann.

2.2 RESOURCE DESCRIPTION FRAMEWORK

RDF (Resource Description Framework) ist eine Sammlung von Standards, die zusammen einen Teil der Idee des Semantic Webs implementieren. Der wesentliche Kern von

RDF ist die Darstellung von Aussagen in der Form (Subjekt, Prädikat, Objekt). Beispiel: »Franz [Subjekt] fährt [Prädikat] nach Berlin [Objekt].« Es ist nur genau ein Objekt möglich. Aussagen wie »Franz [Subjekt] fährt [Prädikat] von Dresden [Objekt] nach Berlin [noch ein Objekt].« sind damit nicht ohne Weiteres in RDF darstellbar. Jede solche Aussage heißt Tripel. Subjekt, Prädikat und Objekt werden jeweils durch eine URI dargestellt. Ausnahme: Ein Objekt darf alternativ anstelle einer URI durch einen String plus einem Datentyp dargestellt werden. Um RDF-Tripel-Mengen zu repräsentieren, gibt es RDF-Serialisierungssprachen. Dazu gehören RDF/XML, Notation 3 und Turtle. Um RDF-Tripel-Mengen zu durchsuchen gibt es RDF-Abfragesprachen. Die Standardabfragesprache heißt SPARQL.

2.3 TEMPLATING-SYSTEME

Ein Template-System ist ein System zur Erzeugung von Dokumenten (beispielsweise HTML) aus Vorlagen (Templates) und einem Datenmodell (z.B. einer Datenbank oder einer Sammlung von Markup-Dokumenten). Die Hauptkomponenten eines Template-Systems sind **Datenmodell, Vorlagen und Template-Engine**.

Das **Datenmodell** kann eine Datenbank sein, eine XML-Datei, eine Datei in einem anderen Markup-Format oder irgendeine andere Art von strukturierten Daten.

Die **Vorlagen** enthalten die Struktur der zu erzeugenden Dokumente. Typischerweise sind Sie in einer Template-Sprache geschrieben, die sich an die syntaktische Struktur der Zieldokumente anlehnt. Simple Vorlagen können Variablen enthalten, die die Template-Engine in einem einfachen Suchen-und-Ersetzen-Verfahren durch Inhalte ersetzt. Komplexere Vorlagen verwenden Template-Sprachen, die Programmier-Konzepte wie Bedingungen und Schleifen unterstützen.

Die **Template-Engine** oder Templating-Sprache erzeugt die Zieldokumente indem sie die Inhalte des Datenmodells entsprechend der Vorlagen strukturiert und gestaltet.

2.4 STATIC SITE GENERATOR

Ein Static Site Generator ist ein Programm, welches mittels eines Template-Systems eine statische Website erzeugt. Der Unterschied zu einem Webframework besteht darin, dass im Betrieb der Webseite keine (dynamische) Verarbeitung von Nutzerdaten stattfindet, sondern alle Inhalte der Webseite statisch bereitgestellt werden. Damit ist der Ressourcenbedarf im Betrieb der Seite geringer und das System typischerweise weniger komplex.

3 ASPEKTE

3.1 ZIEL DES PROJEKTS

Ziel ist die Erstellung einer einfachen, benutzerfreundlichen Anwendung zur Generierung von statischen HTML-Seiten auf Grundlage von gegebenen RDF-Daten und Templating-Sprachen, sodass am Ende W3C-konformer HTML-Code generiert wird und möglichst

alle Daten der RDF-Basis gegenseitig verlinkt werden, sodass man bei einer zusammenhängenden Datenbasis auch am Ende alle Daten gegenseitig erreichen kann. Hierfür eignen sich aktiv supportete, bereits fertige Static Site Generator Tools am besten, welche bereits statische HTML-Seiten aus gegebenen Textbausteinen generieren können. Programmieraufgabe wird also sein, ein Plugin für einen solchen Static Site Generator oder einer Templating-Sprache zu entwickeln, welches RDF-Daten einlesen und zu statischen HTML-Ressourcen verarbeiten kann.

3.2 STATIC SITE GENERATOREN

Es existiert eine Vielzahl von Static Site Generatoren. Im Folgenden werden fünf verschiedene Tools betrachtet.

Jekyll basiert auf der Programmiersprache Ruby. Als Templating-System wird Liquid benutzt. Durch ein sehr mächtiges, integriertes Plugin-System ist es möglich, eigene Plugins ohne Ändern des Jekyll-Sourcecodes zu implementieren. Hinter Jekyll steht eine sehr große Community, die aktiv Support leistet. <http://jekyllrb.com>

Middleman ist ein weiterer Ruby Static Site Generator. Zur Beschreibung der Inhalte wird standardmäßig die Embedded Ruby Template-Engine (ERB) benutzt, es ist jedoch auch möglich Haml oder Liquid einzusetzen. Durch ERB ist es möglich eigene Helper zu programmieren, um beispielsweise mithilfe geeigneter Libraries RDF-Daten zu verarbeiten. Ein ausgereiftes Plugin-System bringt Middleman jedoch im Gegensatz zu Jekyll nicht mit. <https://middlemanapp.com>

Hyde basiert auf der Programmiersprache Python und nutzt die Templating-Sprache Jinja2. Grundlegend unterstützt Hyde auch andere Templating-Sprachen, diese müssten allerdings erst in Hyde implementiert werden. <http://hyde.github.io>

Phrozn basiert auf PHP und unterstützt verschiedene Markup-Formate: Markdown, Textile, Twig, Haml, usw. Es existiert kein offizielles Forum, sondern nur ein IRC-Channel, in dem sich allerdings niemand aufhält. Somit ist es nur schwierig Support zu erhalten. <http://phrozn.info>

Ein Exot stellt das in der Google-Programmiersprache Go programmierte Tool **Hugo** dar. Da Go noch sehr jung und nicht weit verbreitet ist existiert noch keine geeignete RDF-Library, weshalb es folgend nicht weiter betrachtet wird. <http://gohugo.io>

3.3 TEMPLATING-SYSTEME

Basierend auf die oben genannten Anwendungen werden folgend fünf Templating-Systeme betrachtet, die von den Static Site Generatoren unterstützt werden. **Liquid** ist eine in Ruby programmierte Templating-Sprache, welche Tags, Filter und Objekte benutzt, um dynamische Daten zu laden und daraus statische Ressourcen zu generieren („rendern“). Es ist möglich eigene Tags und Filter zu erstellen, um so beispielsweise durch einen einfachen Tag einen QR-Code aus einer beliebigen Zeichenkette zu rendern. Aufgrund einer großen Google-Group und einer aktiven Github-Community wird ein guter Support gewährleistet. <http://liquidmarkup.org>

Die PHP Templating-Sprache **Twig** zeichnet sich durch eine hohe Sicherheit aus, da

ein Sandbox-Mode mitgebracht wird. Zudem ist Twig sehr flexibel, da die Operationen in „lexer“ und „parser“ unterteilt werden. Damit lassen sich eigene Tags und Filter programmieren. Twig wird aktiv gepflegt, besitzt aber keine allzu große Community. <http://twig.sensiolabs.org>

Jinja2 basiert auf Python und ist eine Abwandlung des Web-Frameworks Django. Jinja2 bietet wie Twig einen Sandbox-Mode und ermöglicht die Nutzung eigener Extensions, um so Custom Tags und Filter zu implementieren. Jinja2 wird aktiv entwickelt und aufgrund der Ähnlichkeit zu Django ist die Community entsprechend groß. <http://jinja.pocoo.org/docs/dev/> Die in Ruby programmierte Templating-Sprache **Haml** bringt das Feature mit, Ruby-Code direkt zu evaluieren. Somit ist es relativ einfach Helper zur Verarbeitung von RDF-Daten zu schreiben, sofern es die Ruby-RDF-Library zulässt. <http://haml.info>

3.4 RDF-LIBRARIES

Die vorherige Betrachtung zeigte, dass die Static Site Generatoren bzw. die Templating-Sprachen entweder in Ruby, Python, PHP programmiert wurden.

Für Ruby ist **RDF.rb** die am aktivsten gepflegte RDF-Library. Durch die gute Dokumentation ist ein problemloses Einarbeiten möglich. <http://ruby-rdf.github.io>

Für Python existiert mit **RDFLib** ebenfalls eine sehr ausführlich beschriebene und aktiv entwickelte RDF-Library. <https://rdflib.readthedocs.org/en/stable>

Unter PHP ist **EasyRDF** die bekannteste RDF-Library, welche hinsichtlich aktiver Community, Dokumentation und Funktionalität den anderen beiden Libraries in nichts nachsteht. <http://www.easyrdf.org>

Alle genannten Libraries bieten Support für SPARQL Anfragen.

4 QUELLEN

- <http://www.w3c.de/about/>
- <http://www.w3.org/RDF/>
- <http://stackoverflow.com/questions/9755113/how-are-rdf-and-rdfs-related>
- https://de.wikipedia.org/wiki/Uniform_Resource_Locator
- <https://de.wikipedia.org/wiki/Metadaten>
- <https://de.wikipedia.org/wiki/RDF-Schema>
- <https://de.wikipedia.org/wiki/SPARQL>
- <https://en.wikipedia.org/wiki/OntoWiki>
- <http://aksw.org/Projects/OntoWiki.html>
- <https://web.archive.org/web/20090208105225/>
- <http://www.digitaldivide.net/articles/view.php?ArticleID=20>
- <http://libreas.eu/ausgabe15/texte/001.htm>