

Recherchebericht

Personalisierte Stundenpläne

Autor: Ken Kretschmar

07.01.2015

Inhaltsverzeichnis

Begriffe	2
Konzepte	5
iCalendar	5
RDF	5
SPARQL	5
Triplestore	6
Aspekte	7
Endpunktgebundenheit und deren Komplexität	7
Kompatibilität des Kalenders	7
Benutzeroberfläche	7
Quellen	8

Begriffe

Aggregatfunktion

Aggregatfunktionen in Abfragesprachen fassen mehrere Ergebnisse zusammen. Beispielsweise ermittelt die Funktion AVG() den Durchschnittswert über alle ausgegebenen Werte.

CSV

Eine CSV-Datei (Comma-separated values) speichert Tabellen-Daten in Klartext. Eine CSV-Datei besteht aus einer Anzahl an Einträgen, die durch ein bestimmtes Trennzeichen und Zeilenumbruch getrennt werden.

Datenbankmanagementsystem

Software, die für ein Datenbanksystem installiert und konfiguriert wird. Sie ermöglicht den Zugriff auf die in der Datenbank enthaltenen Daten und entscheidet damit maßgeblich über die Funktionalität des Systems.

Event

Ein Event ist eine Veranstaltung oder ein Termin, welches eine feste Zeit in Anspruch nimmt. Es hat einen festen Anfangs- und Endpunkt, und für diese Zeit ist man bei einem Kalender dann ausgeplant/beschäftigt.

FMI Model

Eine Ontologie. Sie wird benutzt, um Informationen über die Mathematik- und Informatikfakultät der Universität Leipzig zu strukturieren.

foaf

ein Projekt, welches die maschinenlesbare Modellierung sozialer Netzwerke als Ziel hat. Es basiert auf RDFS.

GROUP BY

Befehl in SQL. Gruppiert die Ergebnisse nach den übergebenen Attributen.

iCalendar

Ein Kalender-Standardformat, welches Internetnutzern erlaubt, Terminwünsche und Aufgaben an andere zu übermitteln. (Dateiendung .ical; .ics; .ifb)

Journal

Ein Journal ist eine Beschreibung, welche an ein bestimmtes Datum gebunden ist. Es wird z.B. genutzt, um tägliche Ereignisse oder gemeisterte Herausforderungen festzuhalten. Es nimmt keine Zeit im Kalender in Anspruch.

Linked Data

Methode zum veröffentlichen strukturierter Daten, so dass diese durch semantische Anfragen mehr Nutzen bekommen. Dabei baut es auf Standard Web Technologien, wie HTTP, RDF, URIs, auf.

N3

Notation 3. Eine Formale Sprache, welche eine Syntax für RDF ist.

Namespaces

Namensräume. Sie dienen der Organisation von Identifiern, so dass die lokale und dezentrale URI-Vergabe nicht zu Konflikten führt.

Ontologie

Eine Ontologie ist eine sprachlich gefasste, formal geordnete Darstellung einer Menge von Begrifflichkeiten sowie der zwischen diesen bestehenden Beziehungen. Sie wird genutzt, um einfaches Austauschen von Wissen in digitalisierter und formaler Form zwischen Anwendungsprogrammen und Diensten zu ermöglichen.

Open Data

Beschreibt alle Datenbestände, die im Interesse der Gesellschaft ohne jedwede Einschränkung zur freien Nutzung, zur Weiterverbreitung und zur freien Weiterverwendung frei zugänglich gemacht werden. Diese Daten sind in einer strukturierten Form abgelegt, wodurch sie sowohl maschinell als auch von Menschen verarbeitet werden können.

Bsp.: Lehrmaterial, Geodaten, Statistiken etc

Open Data Uni Leipzig:

Auf der Basis bekannter semantischer Technologien (RDF, Sparql, Linked Data) wird eine Datenbasis aufgebaut und weiterentwickelt. Diese Datenbasis enthält Informationen zu Lehrveranstaltungen an der Universität Leipzig. Zum Verständnis dieser werden Kenntnisse über RDF, Sparql, Linked Data, Namespaces und URI empfohlen.

ORDER BY

Wird in SQL verwendet, um die Ergebnisse nach den mitgegebenen Attributen zu sortieren.

OWL

Web Ontology Language. Eine Spezifikation des W3C, um Ontologien anhand formaler Beschreibungssprache zu erstellen, publizieren und zu verteilen. Durch OWL werden Termini einer Domäne und deren Beziehungen formal so beschrieben, dass Software deren Bedeutung verarbeiten kann. OWL basiert auf der RDF-Syntax.

RDF

Resource Description Framework. Ein Standard für die Beschreibung von Metadaten. Besteht aus einem Triple der Form Subjekt, Prädikat, Objekt.

RDFS

Resource Description Framework Schema. Ontologie-Beschreibungssprache für RDF.

Ressource

Ein eindeutig bezeichneter Gegenstand, worüber man Aussagen treffen möchte.

Semantic Web

Semantic Web bezeichnet eine „Bewegung“ im Internet, geführt von W3C, welche die Masse an Daten im Internet, welche größtenteils ungeordnet auftreten, zu ordnen und in eine normierte Form zu überführen, wodurch diese Daten leichter zugänglich und vor allem maschinell einfach zu verarbeiten sind. Das Endziel ist ein „web of data“.

SPARQL

SPARQL (rekursives Akronym für SPARQL Protocol And RDF Query Language) ist eine graphenbasierte Abfragesprache für RDF. Dadurch können Benutzer Abfragen gegen Daten, die nach der RDF-Spezifikation des W3C aufgebaut sind, durchführen. SPARQL macht sich dabei den Aufbau der Daten nach Triplet „Subjekt-Prädikat-Objekt“ (siehe RDF) zunutze.

URI

Unique Resource Identifier. Besteht aus einer Zeichenfolge, die zur Identifizierung einer abstrakten oder physischen Ressource dient. URIs werden für Webseiten und Webservices im Internet als Bezeichnung verwendet

Variable

Ein formalsprachlicher Platzhalter, in das beliebige Ausdrücke eingesetzt werden können.

W3C

World Wide Web Consortium. Internationale Organisation zur Standardisierung von Techniken des World Wide Web.

XML

Extensible Markup Language. Eine Auszeichnungssprache, welche verwendet wird, um hierarchisch, meist als Baum, strukturierte Daten in Form von Textdateien darzustellen. Sie wird für den plattform- und implementationsunabhängigen Austausch von Daten zwischen Computersystemen eingesetzt.

Konzepte

iCalendar

iCalendar wurde von der Internet Engineering Taskforce entwickelt, um ein Standardformat für den offenen Austausch von kalenderbezogenen Daten über das Internet zu ermöglichen. Es basiert dabei auf dem früher entwickelten vCalendar des Internet Mail Consortium. Mit ihm ist es möglich kalenderspezifische Daten zu erzeugen, wie Events, To-Do-Listen, Journal und anderes.

iCalendar wird von vielen Anwendungen wie Google Calendar, Apple Calendar, IBM Lotus Notes, Yahoo! Calendar, Mozilla Thunderbird, Microsoft Outlook und vielen Anderen genutzt, um einen digitalen Kalender zu erstellen, organisieren und sich mit anderen Nutzern über Termine abzusprechen.

Die Dateien sind in einem einfachen Text geschrieben und können daher leicht erzeugt werden. Im Normalfall nutzt iCalendar die UTF-8-Kodierung und jede Zeile ist dabei maximal 75 Oktetts lang und endet mit `\r\n`, also DOS-Zeileneendungen. Jede Datei beginnt mit und endet mit „BEGIN:VCALENDAR“ bzw. „END:VCALENDAR“ .

RDF

Ursprünglich wurde das Resource Description Framework zur Beschreibung von Metadaten konzipiert. Jedoch entwickelte es sich zum Grundbaustein des so genannten SemanticWebs, da es zum Beispiel Programmen ermöglicht Daten sinnvoll darzustellen.

Desweiteren ist es durch RDF möglich Datensätze unterschiedlicher Struktur auf gleiche Art zu verarbeiten oder zu vereinen.

RDF wird als Tripel formalisiert welches aus Subjekt, Prädikat und Objekt besteht, wobei jedoch eine weitere Ressource als Subjekt verwendet werden kann. Damit spannt RDF im mathematischen Sinne einen gerichteten Graphen auf bei dem die Kanten die Relation zwischen zweier Knoten (repräsentativ für Ressourcen) sind.

RDF ist unabhängig einer texturellen Repräsentation und wird durch N3 formalisiert.

Der Austausch der in RDF formulierten Aussagen findet mit dem RDF-Schema (RDFS) statt, bei dem ein einheitliches Vokabular zur Verfügung gestellt wird. RDFS liegt die Idee eines mengentheoretischen Klassenmodells zugrunde.

Dadurch ist es möglich, Ontologien einheitlich zu formalisieren und weiter zu verarbeiten, wobei es wichtig ist, Klassen und Eigenschaften getrennt zu modellieren.

SPARQL

Die genaue Funktionsweise lässt sich mit folgendem Beispiel leichter erläutern:

```
PREFIX foaf: http://xmlns.com/foaf/0.1/
SELECT ?name ?email
WHERE {
  ?person a foaf:Person.
  ?person foaf:name ?name.
  ?person foaf:mbox ?email.
}
```

Die obige Abfrage-Beispiel nutzt die Ontologie-Definition „foaf“, oft auch „friend-of-a-friend“-Ontologie genannt. Die Abfrage soll von allen Personen in der Datenbank die Namen und die Email ausgeben. Dies geschieht zunächst durch die Angabe des Prefix. Dies erlaubt es die URIs verkürzt darzustellen.

Die eigentliche Abfrage beginnt mit dem SELECT. Hier werden Variablen angegeben, die durch die Query ausgegeben werden sollen, also der Name und die Email. Variablen fangen immer mit „?“ angeführt.

In der WHERE-Klausel wird genauer definiert, was gesucht ist. Hier wird in der ersten Zeile festgelegt, dass die Variable ?person (nach der Definition in „foaf“) eine Person sein soll. Alle weiteren Zeilen nutzen die oben genannte Triplet-Struktur:

Für die eben als Person definierte Variable ?person soll nun ?name das Objekt des Prädikates „foaf:name“ bilden. Zuletzt soll für jedes Subjekt „Person“ das Objekt zum Prädikat „foaf:mbox“ in ?email ausgegeben werden.

Außerdem sind noch weitere Abfrage-Formen und Variationen möglich:

Die oben genutzte SELECT-Klausel sorgt dafür, dass die Abfrage in Tabellenform angegeben wird.

CONSTRUCT würde die obigen Informationen im RDF-Format zurückgeben.

Die ASK-Klausel wird genutzt um eine simple Wahr/Falsch-Abfrage über eine RDF-Datenbank zu erhalten.

Schließlich sind mit SPARQL noch die von SQL bekannten Operationen wie GROUP BY, ORDER BY und die Aggregatfunktionen möglich.

Triplestore

Ein Triplestore ist ein Datenbankmanagementsystem, welches RDF als Datenmodell benutzen. Sie benutzen somit die bereits unter RDF beschriebenen Tripel und die Abfragesprache SPARQL. Für das Projekt existiert ein bereits vorhandener Triplestore¹. An diesen können über einem SPARQL-Endpunkt Abfragen erstellt und die Resultate in verschiedenen Formaten (u.A. RDF, CSV, XML) gespeichert werden. Zusätzlich wurde für die Datenbank eigens ein RDF-Schema (<http://od.fmi.uni-leipzig.de/model/>) erstellt wurde um die Daten in der Datenbank in Beziehung zu stellen.

¹ Siehe <http://od.fmi.uni-leipzig.de/> (06.01.15)

Aspekte

Endpunktgebundenheit und deren Komplexität

Die bereits erwähnte SPARQL-Schnittstelle der Universität Leipzig (od.fmi.uni-leipzig.de) ist durch die Projektdefinitionen vorgegeben. Durch die Komplexität der Abfragesprache wird unter Umständen die Möglichkeit der Optionsbereitstellung in der Benutzeroberfläche eingeschränkt. Außerdem muss geklärt werden, wie von der zu erstellenden Software auf die SPARQL-Schnittstelle zugegriffen werden kann. Für die meisten Programmiersprachen stehen allerdings vollständige APIs zur Verfügung.²³⁴ Weiterhin könnte die Bereitstellung des eigenen RDF-Schemas (<http://od.fmi.uni-leipzig.de/model/>) die Einarbeitung in die Datenbank erschweren, da eigene Schemata oftmals nicht gut dokumentiert und eventuell schlecht designed sind. Dennoch ist die Einarbeitung für das Projekt unerlässlich.

Dennoch gibt es einige Funktionen, die die Entwicklung des Programms vereinfachen. So ermöglichen die unter „od:RegularEvent“ gespeicherten Informationen, dank der deutlichen strukturellen Ähnlichkeit zu dem iCalendar vorhandenen Event-Format „VEVENT“, eine gute Datengrundlage zur Erstellung der exportierbaren Kalender-Daten.

Außerdem können über bestimmte Prädikate („od:containsLV“, „od:containsKurs“, „od:recommendedFor“, „od:toStudiengangSemester“) für eine angegebene Studiengang-Semester-Kombination eine Vorauswahl an Lehrveranstaltungen getroffen werden, die anschließend in den Kalender eingetragen werden können. Somit muss der Benutzer nicht jeden belegten Kurs einzeln wählen.

Zudem können diese genutzt werden, um das Lernveranstaltungsverzeichnis zu strukturieren.

Kompatibilität des Kalenders

Durch die Wahl des iCalendar-Standards entsteht eine sehr hohe Kompatibilität gegenüber verschiedensten Kalender-Applikationen, da der Standard vom Großteil moderner Kalender-Applikationen verwendet wird.

Benutzeroberfläche

Bezüglich der grafischen Benutzeroberfläche ist die Wahl zunächst nicht durch Vorgaben eingeschränkt. Dadurch bieten sich die Möglichkeiten, entweder die das Projekt als Web-Applikation über eine Webseite zur Verfügung zu stellen, dies hat den Vorteil, dass sie ohne zusätzliche Installationen im Browser ausgeführt werden kann. Nachteilig ist hier lediglich, dass sichergestellt werden muss, dass die Webseite in allen gängigen Browsern darstellbar und benutzbar ist. Mit dieser Option könnte zusätzlich die Erstellung einer Mobil-Variante der Webseite in Betracht gezogen werden.

Alternativ kann die Applikation als eigenständiges Programm realisiert werden, welches auf dem Benutzergerät zunächst installiert/heruntergeladen werden muss, um es auszuführen. Dies bietet den Vorteil der Unabhängigkeit gegenüber des Projektserverns. Auch hier muss jedoch sichergestellt werden, dass die Applikation auf allen möglichen Endgeräten funktioniert.

² sparqlwrapper mit python (<https://rdflib.github.io/sparqlwrapper/>)

³ jena-sparql-api mit java (<http://aksw.org/Projects/jena-sparql-api.html>)

⁴ Erfurt mit php (<http://wiki.aksw.org/Projects/Erfurt>)

Quellen

- <http://www.leipzig-data.de/open-data/> (03.01.2015)
- http://en.wikipedia.org/wiki/Open_data (03.01.2015)
- <http://bis.informatik.uni-leipzig.de/OLAT/LVPlanung/ODQuelle> (06.01.2015)
- http://en.wikipedia.org/wiki/Semantic_Web (01.01.2015)
- <http://de.wikipedia.org/wiki/RDF-Schema> (29.12.2014)
- <http://de.wikipedia.org/wiki/RDF> (29.12.2014)
- <http://www.w3.org/RDF/> (29.12.2014)
- <http://www.linkeddatatools.com/introducing-rdf> (30.12.2014)
- <http://en.wikipedia.org/wiki/SPARQL> (19.12.2014)
- <http://www.linkeddatatools.com/querying-semantic-data> (20.12.2014)
- <http://en.wikipedia.org/wiki/ICalendar> (19.12.2014)
- <http://www.informatik.uni-leipzig.de/~graebe/Texte/delfi-11.pdf> (29.12.2014)
- <http://de.wikipedia.org/wiki/Datenbank#Datenbankmanagementsystem> (05.01.2015)