

Malte Blattmann
Benedikt Elßmann
Semen Gaidenbrik
Edwin Knese
Dustin Kröger
René Lindenberg
Markus Reinisch
Jonathan Schlue

Betreuer: Robert Schädlich
Abgabedatum: 24.04.2017

Recherchebericht

Begriffe

Abfragesprache

Graphbasierte Abfragesprache

Eine Abfragesprache dient der Informationssuche in Datenbanken. Anders als in einer relationalen Abfragesprache, durchsucht eine graphenbasierte Abfragesprache Graphdatenbanken um Informationen zu erhalten, also eine Datenbank, die Graphen benutzt, um stark vernetzte Informationen darzustellen und abzuspeichern

Relationale Abfragesprache

Eine relationale Abfragesprache agiert auf einer relationalen Datenbank und basiert auf der relationalen Algebra. Das bedeutet, mit einer relationalen Abfragesprache können Daten aus Tabellen abgefragt werden, wobei mehrere Tabellen, die gemeinsame Attribute aufweisen, in den Abfragen kombiniert werden können.

Im Gegensatz zu anderen Datenbankmodellen (z. B. dem hierarchischen) werden dafür teilweise künstliche Schlüsselattribute eingeführt, um die Datensätze leichter unterscheidbar zu machen. In einer relationalen Anfrage können die gesuchten Attribute, die zu durchsuchenden Tabellen und einschränkende Bedingungen vorgegeben werden. [31]

Ajax

Ajax steht für Asynchronous JavaScript and XML und ist ein Konzept zum asynchronen Datenaustausch zwischen einem Browser und dem Server. [19]

CSS

Cascading Style Sheets (kurz: CSS) ist eine formale Sprache zum Festlegen des Erscheinungsbild von Dokumenten (hauptsächlich HTML- und XML-Dokumente) in der Informationstechnik. [17][18]

Graph

Ein Graph ist eine Menge von Knoten und einer Menge von Kanten, die von einem Knoten zu einem anderen Knoten führen.

In einem gerichteten Graphen sind die Kanten beliebig zwischen den Knoten verteilt, in einem ungerichteten Graphen existiert zu jeder Kante eine entgegengesetzt gerichtete Kante. [29]

HTML

Die Hypertext Markup Language ist eine Auszeichnungssprache, die auf Text basiert, um digitale Dokumente wie Texte mit Hyperlinks, Bildern und anderen Inhalten zu strukturieren.

HTTP/S

Das Hypertext Transfer Protocol ist ein Übertragungsprotokoll für Daten über ein Rechnernetz. Mittels HTTPS (S für secure) können die Datenübertragungen verschlüsselt werden. [20] [21]

Java

Java ist eine objektorientierte Programmiersprache. Der Java-Quellcode ist nicht direkt ausführbar; erst der Java-Compiler übersetzt ihn in einen maschinenverständlichen sogenannten Java-Bytecode. Die Maschine, die diesen Bytecode ausführt, ist virtuell – das heißt, der Code wird meist nicht direkt durch Hardware (etwa einen Mikroprozessor) ausgeführt, sondern durch entsprechende Software auf der Zielplattform. Zweck dieser Virtualisierung ist Plattformunabhängigkeit: Das Programm soll ohne weitere

Änderung auf jeder Rechnerarchitektur laufen können, wenn dort eine passende Laufzeitumgebung installiert ist. [10][11]

JS

JS ist eine leichte, interpretierte und objektorientierte Sprache, vorwiegend bekannt als Skriptsprache für Websites. Diese soll die Funktionalität von HTML und CSS erweitern um auch komplexe Nutzerinteraktionen zu ermöglichen. Mit Hilfe von node.js ist es möglich JS auch auf Servern und Computern außerhalb des Webs und Browsern zu nutzen. Zunehmend wird JS auch auf Mikrocontrollern eingesetzt.

Linguistik

Linguistik bedeutet Sprachwissenschaft. Sie beschäftigt sich mit allen Aspekten der Sprache (z. B. der Grammatik). [30]

Markdown

Markdown ist eine Auszeichnungssprache für die Formatierung von Text. Sie wurde 2004 von John Gruber veröffentlicht und wird u.a. von GitLab nativ unterstützt. [13]

Metadaten

Metadaten enthalten Informationen zu anderen Daten. Es handelt sich um Daten, die andere Daten beschreiben.

N3

N3 ist die Abkürzung für Notation3 und ist eine XML-ähnliche, leichter lesbare Art, Daten strukturiert zusammenzufassen. Der N3 Standard findet häufig in Verbindung mit RDF Anwendung.

Ontologie

Ontologien in der Informatik sind meist sprachlich gefasste und formal geordnete Darstellungen einer Menge von Begrifflichkeiten und der zwischen ihnen bestehenden Beziehungen in einem bestimmten Gegenstandsbereich.

OWL

Die Web Ontology Language ist eine Sprache zur Beschreibung von Ontologien, die auf der RDF-Syntax aufbaut und sich auf die Möglichkeit der maschinellen Verarbeitung der Information fokussiert.

RDF

RDF steht für Resource Description Framework und bildet die Basis zur Verarbeitung von Daten innerhalb des semantischen Webs. Eine Resource (Information-(s)quelle) wird über ein Prädikat mit einem Literal (Wert) oder einer anderen Resource verknüpft und so näher beschrieben. RDF baut auf gerichteten Graphen auf und ähnelt UML oder dem ERM als Beschreibungssprache für Daten und deren Beziehungen untereinander.[27]

Semantik

Semantik, auch Bedeutungslehre, nennt man die Theorie oder Wissenschaft von der Bedeutung der Zeichen. Die Semantik als Teildisziplin der Sprachwissenschaft (linguistische Semantik) untersucht die Bedeutung sprachlicher Zeichen.

Die lexikalische Semantik (auch: Wortsemantik) beschäftigt sich mit der Bedeutung von Wörtern und Morphemen wie auch mit der inneren Strukturierung des Wortschatzes insgesamt.

Die Satzsemantik untersucht, wie aus der Bedeutung einzelner Wörter durch ein festes Inventar an Verknüpfungsregeln die Bedeutung von größeren syntaktischen Einheiten Phrasen, Satzgliedern, Teilsätzen und ganzen Sätzen hervorgeht. Die Interpretation eines Satzes muss dabei auf einer Analyse seiner syntaktischen Struktur aufgebaut werden. [9]

SKOS

Das Simple Knowledge Organisation System (SKOS, frei übersetzt „einfaches System zur Organisation von Wissen“) ist eine auf dem Resource Description Framework (RDF) und RDF-Schema (RDFS) basierende formale Sprache zur Kodierung von Dokumentationssprachen wie Thesauri, Klassifikationen oder anderen kontrollierten Vokabularen.

Mit SKOS soll die einfache Veröffentlichung und Kombination kontrollierter, strukturierter und maschinenlesbarer Vokabulare für das Semantische Web ermöglicht werden. SKOS beinhaltet eine Zusammenstellung von Standards und Spezifikationen zur Unterstützung von Wissensorganisationssystemen (Knowledge Organisation Systems) wie

etwa Thesauri, Klassifikationsschemata, Taxonomien und anderen Arten von kontrolliertem Vokabular. [6][7]

Sparqlify

Ist ein Tool mit dessen Hilfe RDF-Schema auf relationale Datenbanken gemappt wird und danach SPARQL-Abfragen auf dieser ausgeführt werden können. [24]

SPARQL

SPARQL ist eine graphenbasierte Abfragesprache für RDF.

SQL

Die Structured Query Language (deutsch "strukturierte Abfragesprache"), kurz SQL, enthält eine relationale Abfragesprache. Sie dient der Abfrage und Bearbeitung von Daten in einer relationalen Datenbank. [28]

Syntax

Unter Syntax versteht man allgemein ein Regelsystem zur Kombination elementarer Zeichen zu zusammengesetzten Zeichen in natürlichen oder künstlichen Zeichensystemen.

Insbesondere versteht man unter Syntax ein Teilgebiet der Grammatik natürlicher Sprachen, das die Zusammenfügung von Wörtern bzw. Wortgruppen zu größeren Einheiten bis hin zu Sätzen behandelt, also die Satzlehre. Die Syntax wird in der Regel unterschieden von der linguistischen Morphologie, die den inneren Aufbau der Wörter behandelt, obwohl die Übergänge zwischen beiden Bereichen fließend sein können. Der Ausdruck Syntax wird für natürliche und formale Sprachen verwendet. [8]

TBX

TermBase eXchange ist eine XML-basierte Auszeichnungssprache, die speziell dem Austausch von Terminologiedaten zwischen Terminologiedatenbanken dient.

Turtle

Turtle, auch Terse RDF Triple Language genannt, ist eine Serialisierung von RDF-Graphen und ist eine Teilmenge von der Notation N3. [22]

URI

Ein Uniform Resource Identifier ist eine Zeichenfolge, die der eindeutigen Identifikation einer Ressource dient.

Die URI gibt auch Aufschluss über die Art der Ressource.

URL

Ein Uniform Resource Locator ist eine URI, die die Ressource zusätzlich im Computernetzwerk lokalisiert.

Plakativ: "WO ist die Ressource?"

URN

Ein Uniform Resource Name ist eine URI im URN-Schema zur genaueren Beschreibung.

Plakativ: "WAS ist die Ressource?"

W3C

Das World Wide Web Consortium (kurz W3C) ist das Gremium zur Standardisierung der Techniken im World Wide Web.

Das W3C entwickelt technische Spezifikationen und Richtlinien in einem ausgereiften, transparenten Prozess, um maximalen Konsens über den Inhalt technischer Protokolle, hohe technische und redaktionelle Qualität und Zustimmung durch das W3C und seine Anhängerschaft zu erzielen. Beispiele für durch das W3C standardisierte Technologien sind HTML, XHTML, XML, RDF, OWL, CSS, SVG und WCAG. [4][5]

Konzepte

CI

Kontinuierliche Integration (auch fortlaufende oder permanente Integration; englisch Continuous Integration (CI)) ist ein Begriff aus der Software-Entwicklung, der den Prozess des fortlaufenden Zusammenfügens von Komponenten zu einer Anwendung beschreibt.

CI führt bei jedem hochgeladenem Commit eines Repositories (bspw. auf Github) festgelegte Skripte aus, die den Zustand des Commits bewerten.

Beispiele: Check von Code-Conventions, Ausführen aller Arten von Tests (Unit, Integration, Interface, ...) Statische Analyse des Codes, Metrikbestimmung und Regeltests, Kompilieren der Software

Bei Fehlschlag einer "Stage" wird sofort gewarnt (bspw. per E-Mail). Damit geht eine Art Instant-Feedback einher, das die Qualität der Softwareentwicklung nachhaltig verbessert. [12]

CD

Continuous Delivery (CD) bezeichnet eine Sammlung von Techniken, Prozessen und Werkzeugen, die den Softwareauslieferungsprozess (englisch: Deployment) verbessern. CD verpackt die Software nach erfolgreicher (!) CI. Nachfolgend wird diese ausgeliefert (Delivery) und ggf. auch gestartet (Deployment). Dabei gibt es keinen Standard und es müssen eigene Skripte geschrieben werden.

Beispiel: Eine Anwendung hat intensive Tests und Metriken durch CI bestanden. Nachfolgend werden von der Anwendung Installer-Pakete für den Windows, Mac und einige Linux Stores/Paketmanager gebaut und in die Verteilzentren (Paket-Repositories) hochgeladen. Das kann durchaus mehrmals pro Tag erfolgen oder per Script auf einmal wöchentlich begrenzt werden. Alternativ wird das Kompilat per SSH auf einen Server übertragen, die alte Version der Software beendet und die neue Version gestartet. [12]

Framework

Ein Programmiergerüst bzw. Framework beschreibt eine oft in der objektorientierten bzw. komponentenbasierten Entwicklung verwendete Programmierunterstützung. Ein Framework dient als Entwurfsmuster und enthält verschiedene Basisbausteine, die in Kombination eine fertige Software ergeben. [16]

Linked Data

Linked Data ist ein Konzept, dessen Grundidee es ist Daten, die miteinander in Beziehung stehen so zu vernetzen, dass ein sinnvolles Netzwerk ("Giant Global Graph"),

bestehend aus verschiedensten Themengebieten, entsteht.

Angestrebt wird hierbei die Beschreibung dieser Beziehung durch standardisierte, maschinenlesbare Beschreibungssprachen wie RDF.

Somit ist Linked Data auch ein Teilgebiet des Semantic Web. [37]

Semantic Web

Unter dem Konzept des Semantic Webs versteht man die Anreicherung von Webdokumenten mit maschinenlesbarem Code. Dadurch wird der Austausch und die Verwertbarkeit dieser vereinfacht und die Daten in einen Kontext (für Maschinen) gesetzt. Dinge von Interesse werden mit einer eindeutigen Adresse versehen und stellen einen Knoten im Graph dar. Die Beziehungen (Kanten) zwischen den Knoten werden ebenfalls mit einer Adresse versehen. [23]

Triplestor

Ein Triplestor ist eine Möglichkeit die auf gerichteten Graphen basierenden RDF Daten zu speichern. Dabei wird die tatsächlich vorhandenen Kanten des Graphen mittels der Subjekt-Prädikat-Objekt Beziehung in einer Textdatei abgespeichert. Ähnlich wie eine relationale Datenbank, speichert man Informationen in einem Triplestore und ruft sie über eine Abfragesprache ab. Im Gegensatz zu einer relationalen Datenbank ist ein Triplestore für die Speicherung und das Abrufen von Triples optimiert. [25][26] Die Komponenten eines Triple, wie die Aussage "Der Himmel hat die Farbe Blau", besteht aus einem Subjekt ("der Himmel"), ein Prädikat ("hat die Farbe") und ein Objekt ("blau"). Dies ähnelt der klassischen Notation eines Entity-Attribut-Wert-Modells innerhalb eines objektorientierten Designs, wobei dieses Beispiel als Entity (sky), ein Attribut (color) und ein Wert (blue) ausgedrückt werden würde.

Aspekte

Datahub

Datahub ist, ähnlich wie Github für Programmcode, eine Plattform für offene Daten, die unter bestimmten Lizenzen veröffentlicht werden. Diese Sammlung von Informationen aus mehreren Quellen ist für die gemeinsame Nutzung und oft die Teilung und den Austausch organisiert.

Für unser Projekt gilt es die vom LiDo umgewandelte RDF-Daten im Datahub zur Verfügung zu stellen.

Gitlab

GitLab ist eine Webanwendung des Unternehmens GitLab Inc., die dem Management von Git-Repositories dient und auf Git basiert.

GitLab bietet hierbei hauptsächlich das kostenlose Hosting von Git Repositories an. Die Anwendung gibt es sowohl gratis, als auch kostenpflichtig mit zusätzlichen Funktionen beispielsweise im Bereich Continuous Integration / Continuous Deployment. GitLab ist bezüglich unseres Projekts ein Aspekt, da es ein grundlegendes Tool zur Koordination unserer Arbeit darstellt.[1][2][3]

Jekyll

Mit einem Werkzeug wie Jekyll können schnell und einfach statische Websites erstellt werden. Dank eines voreingestellten Designs reicht es, Text-Dateien im u.a. Markdown-Format zu hinterlegen. Diese werden folgend durch Jekyll in valides HTML umgewandelt. Jekyll ist durch Plugins und weitere Designs erweiterbar. [15]

LiDo

Die Linguistic Documentation, kurz LiDo, ist eine linguistische Datenbank und Website der Universität Regensburg. Sie enthält diverse linguistische Daten, unter anderem linguistische Begriffe.

Ruby

Ruby ist eine sehr einfach zu erlernende, dynamische Programmiersprache mit weiten Einsatzmöglichkeiten. Von Yukihiro "Matz" Matsumoto im Jahre 1995 entwickelt, kombiniert sie Teile aus den Sprachen Perl, Smalltalk, Eiffel, Lisp und Ada und vereint funktionale und imperative Programmierung. Bekannt ist die Sprache vor allem durch das Framework Ruby on Rails. [14]

SCRUM

Scrum ist eine agile Methode, die verwendet wird um das Management zu fokussieren. Die Methode unterteilt sich in drei Phasen und beginnt mit der Initialphase, in der sich ein genereller Überblick über das Zielsystem verschafft wird. Anschließend beginnt mit den Sprint-Zyklen die Entwicklung einzelner Inkremente des Systems. Abschließend werden die benötigte Dokumentation des Projekts und Erfahrungsberichte zusammengefasst. Scrum orientiert sich an den Prinzipien agiler Methoden, somit ist die Einbeziehung des Kunden, inkrementelle Auslieferung und die flexible Arbeitsweise innerhalb des Teams ein fester Bestandteil der Methode.

Glossar

Abfragesprache . 1

Ajax . 2

CD . 7

CI . 7

CSS . 2

Datahub . 9

Framework . 7

Gitlab . 9

Graph . 2

Graphbasierte Abfragesprache . 1

HTML . 2

HTTP/S . 2

Java . 2

Jekyll . 9

JS . 3

LiDo . 9

Linguistik . 3

Linked Data . 7

Markdown . 3

Metadaten . 3

N3 . 3

Ontologie . 3

OWL . 4

RDF . 4

Relationale Abfragesprache . 1

Ruby . 9

SCRUM . 10

Semantic Web . 8

Semantik . 4

SKOS . 4

SPARQL . 5

Sparqlify . 5

SQL . 5

Syntax . 5

TBX . 5

Triplestor . 8

Turtle . 5

URI . 6

URL . 6

URN . 6

W3C . 6

Literatur

- [1] <https://de.wikipedia.org/wiki/GitLab>
- [2] <https://en.wikipedia.org/wiki/GitLab>
- [3] <https://about.gitlab.com>
- [4] <https://www.w3.org/standards/>
- [5] https://de.wikipedia.org/wiki/World_Wide_Web_Consortium
- [6] <http://www.w3.org/standards/techs/skos>
- [7] https://de.wikipedia.org/wiki/Simple_Knowledge_Organisation_System
- [8] <https://de.wikipedia.org/wiki/Syntax>
- [9] <https://de.wikipedia.org/wiki/Semantik>
- [10] https://www.java.com/de/download/faq/whatis_java.xml
- [11] [https://de.wikipedia.org/wiki/Java_\(Programmiersprache\)](https://de.wikipedia.org/wiki/Java_(Programmiersprache))
- [12] SWT Vorlesung WS 2016 - Professor Dr.-Ing. Klaus-Peter Fähnrich, Dr. Michael Martin, Roy Meissner
- [13] <https://daringfireball.net/projects/markdown/>
- [14] <https://www.ruby-lang.org/de/about/>
- [15] <https://jekyllrb.com/docs/home/>
- [16] <https://de.onpage.org/wiki/Framework>
- [17] https://de.wikipedia.org/wiki/Cascading_Style_Sheets
- [18] <https://de.wikipedia.org/wiki/Stylesheet-Sprache>
- [19] [https://de.wikipedia.org/wiki/Ajax_\(Programmierung\)](https://de.wikipedia.org/wiki/Ajax_(Programmierung))
- [20] https://de.wikipedia.org/wiki/Hypertext_Transfer_Protocol
- [21] https://de.wikipedia.org/wiki/Hypertext_Transfer_Protocol_Secure
- [22] [https://de.wikipedia.org/wiki/Turtle_\(Syntax\)](https://de.wikipedia.org/wiki/Turtle_(Syntax))
- [23] https://de.wikipedia.org/wiki/Semantic_Web
- [24] <http://aksw.org/Projects/Sparqlify.html>
- [25] <https://en.wikipedia.org/wiki/Triplestore>, Zugriff 23.04.2017
- [26] Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, York Sure, Semantic Web Grundlagen, Erste Auflage 2008, Seite 40

- [27] Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, York Sure, Semantic Web Grundlagen, Erste Auflage 2008, Seite 36
- [28] <https://dbs.uni-leipzig.de/file/dbs1-ws1617-kap5-half.pdf>
- [29] http://www.informatik.uni-leipzig.de/alg/lehre/ws12_13/DISK/foalien.pdf Seite 40
- [30] <http://www.duden.de/rechtschreibung/Linguistik>
- [31] <https://de.wikipedia.org/wiki/Abfragesprache>
- [32] <http://www.gruenderszene.de/lexikon/begriffe/html>
- [33] [https://de.wikipedia.org/wiki/Ontologie_\(Informatik\)](https://de.wikipedia.org/wiki/Ontologie_(Informatik))
- [34] <http://www.itwissen.info/Metadaten-meta-data.html>
- [35] <https://www.w3.org/TR/rdf-sparql-query/>
- [36] <http://scrum-master.de/Scrum-Einfuehrung>
- [37] <https://www.w3.org/DesignIssues/LinkedData.html>
- [38] https://de.wikipedia.org/wiki/Web_Ontology_Language
- [39] <http://stackoverflow.com/questions/4913343/what-is-the-difference-between-uri-url-and-urn>
- [40] <https://www.w3.org/OWL/>