

Aufgabenblatt 2

Inhaltsverzeichnis

1	Begriffe	2
1.1	Ontologie	2
1.2	RDF	2
1.3	RDF-S	2
1.4	OWL	2
1.5	SPARQL	3
1.6	Wiki	3
1.7	PHP	3
1.8	Zend	4
1.9	XML	4
1.10	Ajax	5
2	Konzepte	5
2.1	Semantic Web	5
2.2	MVC	6
2.3	Mengen- („Eimer“) Konzept	6
3	Applikationsstudie	7
3.1	Ontowiki	7
3.2	Erfurt	8
3.3	Plugins, Komponenten, Module	8

1 Begriffe

1.1 Ontologie

Eine Ontologie ist eine formale Repräsentation von einer Menge von Begriffen (Vokabular) und Beziehungen zwischen ihnen. Ontologien enthalten Inferenz- und Integritätsregeln, also Regeln der Schlussfolgerung und zur Gewährleistung ihrer Gültigkeit. Ontologien dienen als Möglichkeit Wissen strukturiert aufzubereiten und zugänglich zu machen - dies macht sie zum zentralen Bestandteil des „Semantic Web“. Viele Wissenschaften nutzen Ontologien in dem sie ihr Spezialwissen modellieren und Schlussfolgerungsprozesse erstellen damit deklaratives Wissen, Problemlösungstechniken und Schlussfolgerungsmechanismen von mehreren Systemen geteilt werden können. Zum Beispiel in der Bioinformatik oder der künstlichen Intelligenz.

1.2 RDF

Das Akronym RDF steht für Resource Description Framework und bezeichnet eine formale Sprache zur Beschreibung von Metainformationen. RDF ist ein Datenmodell für Ressourcen und Relationen zwischen diesen. Es stellt eine einfache Semantik für das Datenmodell zur Verfügung. Diese Datenmodelle können in XML repräsentiert werden. Das eigentliche RDF-Modell besteht aus Tripeln der Form: Ressource, Eigenschaftselemente und Objekt. Ressourcen können zum Beispiel Web-Seiten, Bücher oder Gemälde sein, auch wenn sie teilweise nicht über das Internet erreichbar sind. Wichtig dabei ist aber, das man die Ressource in irgendeiner Form eindeutig hinterlegt. Die Eigenschaftselemente dagegen haben die Aufgabe, eine Aussage über die Ressource zu treffen, und stellen einen Bezug zum Objekt her. Die Objekte beschreiben nun den Wert einer Ressource. Durch die Kombination zu einem RDF-Tripel wird so eine Aussage über ein bestimmtes Objekt gemacht. Wenn es nun mehrere Aussagen über mehrerer Objekte gibt, entsteht ein RDF-Modell, das Aussagen über verschiedene Objekte einer Domäne und Beziehungen untereinander zulässt.

1.3 RDF-S

Das Resource Description Framework Schema liefert das Vokabular zur Beschreibung der Eigenschaften und Klassen von RDF Ressourcen, mit einer Semantik für verallgemeinernde Hierarchien solcher Eigenschaften und Klassen (z.B. Vererbung). Dabei liegt RDFS die Idee eines mengentheoretischen Klassenmodells zugrunde, indem Klassen und Eigenschaften separat modelliert werden. Mit diesem Klassenmodell ist es möglich eine formale Beschreibung der Semantik, der verwendeten RDF-Elemente, zu erstellen.

1.4 OWL

OWL ist die dritte W3C-Spezifikation zur Erstellung von Semantic Web Applikationen. Aufbauend auf RDF und RDFS definiert OWL die Arten von Beziehungen, die in RDF mittels eines XML-Vokabulars ausgedrückt werden können, um die Hierarchien und Beziehungen zwischen verschiedenen Ressourcen zu veranschaulichen. Semantic Web-Ontologien bestehen aus einer Taxonomie und einer Reihe von Inferenzregeln, anhand derer Maschinen logische Schlüsse ziehen können.

Eine Taxonomie ist in diesem Zusammenhang ein Klassifikationssystem (Hierarchien von Klassen).

OWL wird verwendet um Ressourcenklassen Eigenschaften zuzuweisen und dieselben Eigenschaften auch an deren Unterklassen zu vererben. OWL verwendet auch die XML-Schemadatentypen und unterstützt Klassenaxiome wie z.B. `subClassOf`, `disjointWith`, usw. und Klassenbeschreibungen wie `unionOf`, `intersectionOf`, usw. Auch viele andere Konzepte wurden in OWL integriert, wodurch OWL zur umfangreichsten heute existierenden Standard-Ontologiebeschreibungssprache wurde. Zusätzliche Sprachkonstrukte ermöglichen es, prädikatenlogische Ausdrücke zu formulieren. Je nach Einsatz (bzw. geforderter Mächtigkeit) wurden die Sprachebenen OWL Lite, OWL DL und OWL Full definiert. Für den Einsatz von OWL Lite/DL wurden Einschränkungen definiert, welche die Entwicklung von Tools erleichtern bzw. vollständige Inferenz ermöglichen sollen.

1.5 SPARQL

SPARQL ist eine graph-basierte Anfragesprache für RDF. Der Name (gesprochen: „sparkle“) steht für „SPARQL Protocol and RDF Query Language“. Die Sparql-Engine, die den Query verarbeitet, sucht dann in der gegebenen RDF-Wissensbasis nach dem Teilgraphen, der in der WHERE-Klausel angegeben ist.

Das folgende Beispiel findet die Namen aller afrikanischen Hauptstädte und das Land in dem sich jedes befindet.

```
PREFIX abc: <http://example.com/exampleOntology#>
SELECT ?capital ?country
WHERE {
  ?x abc:cityname ?capital.
  ?y abc:countryname ?country.
  ?x abc:isCapitalOf ?y.
  ?y abc:isInContinent abc:africa.
}
```

Variablen werden mit vorangestelltem „?“ gekennzeichnet (möglich ist auch „\$“). Dabei werden alle Variablenbelegungen für „?capital“ und „?country“ zurückgegeben, die auf die Muster dieser vier RDF-Tripel passen.

Weil das Ausschreiben der URIs die Leserlichkeit einer Abfrage mindert, können Präfixe verwendet werden. Hier steht ein „abc:“ für `http://example.com/exampleOntology#` external

1.6 Wiki

Wikis ermöglichen es verschiedenen Autoren, gemeinschaftlich an Texten zu arbeiten. Wikis sind eine Unterkategorie von Content-Management-Systemen (CMS), setzen aber auf die Philosophie des offenen Zugriffs, im Unterschied zu teils genau geregelten Arbeitsabläufen (workflow) von Redaktionssystemen. Die Änderbarkeit der Seiten durch jedermann setzt eine ursprüngliche Idee des World Wide Web erstmals konsequent um und erfüllt weiterhin eine wesentliche Anforderung an soziale Software. Jedoch ist nicht jedes Wiki für alle lesbar oder schreibbar, es gibt auch Systeme, die Zugriffssteuerung für bestimmte Seiten und Benutzergruppen erlauben. In der Wirtschaft dienen Wikis häufig zum Wissensmanagement, mit den Zielen, erhöhte Transparenz des vorhandenen Wissens herzustellen, Prozesse zu optimieren und Fehler zu vermeiden.

Im Falle von OntoWiki geht es darum Ontologien nach dem oben genannten Wiki-Prinzip zu bearbeiten.

1.7 PHP

PHP (rekursives Akronym für „PHP: Hypertext Preprocessor“) ist eine Skriptsprache mit einer an C angelehnten Syntax, die hauptsächlich zur Erstellung von dynamischen Webseiten oder Webanwendungen verwendet wird. PHP ist Open-Source-Software. PHP zeichnet sich besonders durch die leichte Erlernbarkeit, die breite Datenbankunterstützung und Internet-Protokolleinbindung sowie die Verfügbarkeit zahlreicher, zusätzlicher Funktionsbibliotheken aus. Viele interaktive Webseiten und web-basierte Dienste sind in PHP umgesetzt (zum Beispiel auch OntoWiki). Bei einem HTTP-Request an ein PHP Skript wird dieses vom Interpreter verarbeitet und nicht direkt an den Client gesendet. Erst die Ausgabe des PHP-Interpreters wird an den Browser geschickt. In den meisten Fällen ist dies ein HTML-Dokument es sind jedoch viele Dokumenttypen möglich (XML, PDF, Bilder etc.)

1.8 Zend

Die Zend Engine ist ein Compiler für die Skriptsprache PHP und eine virtuelle Maschine.

Das Zend Framework ist ein Komponenten-orientiertes Framework für PHP 5, das heißt die Klassen und Pakete können unabhängig voneinander und auch in Kombination mit denen anderer Hersteller genutzt werden. Die Funktionen sind komplett objektorientiert realisiert. Wichtige Komponenten sind:

- MVC-Verwaltung
- Authentifizierung, Benutzerverwaltung
- Datenbank-Zugriff
- Lokalisierung: Übersetzung, Währungen, Zeit/Datumsformate
- E-Mail-Versand

1.9 XML

Das XML steht für eXtensible Markup Language und ist eine Auszeichnungssprache zur Beschreibung strukturierter Daten in Form von Textdaten. XML ist wie andere Verfahren für den Austausch von Daten zwischen Computersystemen besonders über Internet einsetzbar. Es gibt verschiedene Beispiele für XML-Sprachen z.B: RSS, MathML, GraphML, XHTML, XML-Schema. XML ist eine Teilmenge von SGML (Standard Generalized Markup Language) und wurde entwickelt, um eine einfache Implementierung und Zusammenarbeit sowohl mit SGML als auch mit HTML (Hypertext Markup Language) zu sichern. Jedes XML-Dokument besteht aus einer Folge von Textzeichen, im einfachsten Fall ASCII und hat eine logische und eine physische Struktur. Die physische Struktur besteht aus:

- XML-Deklaration
- Entitäten

aber Die logische Struktur besteht aus:

- Verarbeitungsanweisungen
- Elemente(werden wie in HTML mit Klammern (Tags))
- Attribute(jedes Attribut besteht aus dem Attribut-Name und dem Attribut-Wert).
- Kommentare
- Text

1.10 Ajax

Ajax steht für „Asynchronous JavaScript and XML“. Es bezeichnet ein Konzept der asynchronen Datenübertragung zwischen einem Server und dem Browser, das es ermöglicht, innerhalb einer HTML-Seite (durch clientseitige Skriptsprachen) selbstständig HTTP-Requests durchzuführen, ohne die Seite komplett neu laden zu müssen. Dabei können gewisse Teile einer HTML-Seite oder auch reine Nutzdaten (XML, JSON) sukzessiv bei Bedarf nachgeladen werden, womit Ajax eine Schlüsseltechnik zur Realisierung des Web 2.0 darstellt, denn es lassen sich Anwendungen erzeugen, die in ihrer Bedienbarkeit intuitiv sind und weniger an das seitenorientierte Browsen erinnern als eher an Desktop-Applikationen.

Es gibt zahlreiche Ajax-Frameworks die umfangreiche Funktionalität wie Auto-Vervollständigung, Layoutelemente wie Tabs, Buttons, Pop-Ups, Rich-Text-Editors, Tabellen zur Verfügung stellen. Z.B. JQuery, Yahoo YUI, Sajax, Xajax.

2 Konzepte

2.1 Semantic Web

Semantic Web wird als eine Erweiterung des Internets verstanden, und hat zum Ziel Daten zusätzlich mit maschinenlesbaren Attributen zu beschreiben. Dies wird durch sogenannte Metainformationen erreicht, welche Webdokumenten zusätzlich angefügt werden, um den Inhalt des Dokumentes semantisch zusammenzufassen. Dadurch wird es für Maschinen möglich, diese Dokumente zu interpretieren. So kann der Inhalt eines Dokumentes maschinell erfasst und ausgewertet werden. Informationen können somit automatisiert miteinander in Beziehung gesetzt werden. Zusammenhänge werden sichtbar, die vorher nicht intuitiv erkennbar waren.

Das Semantic Web wird durch W3C Recommendations spezifiziert und besteht aus vielen Teil-Technologien wie URI, XML, RDF, RDFS, OWL, SPARQL und anderen. Wie diese Schichten aufeinander aufbauen ist in der Grafik dargestellt.

Im Unterschied zum Information-Retrieval mit Informationsextraktion (IR/IE), das auf unstrukturierten Daten operiert, setzt das Semantische Web Annotationen (Metadaten) voraus. Die Bedeutung der dargebotenen Inhalte wird also mit Hilfe einer Auszeichnungssprache explizit dazugeschrieben und nicht erst später heuristisch interpretiert wie in der Computerlinguistik. Die Annotation geschieht unter Einsatz von festgelegten Vokabularien und Ontologien, beispielsweise mittels RDF oder OWL.

Eine Annotation von HTML/XML-Seiten im Web geschieht z. B. durch die Wissens-/Ontologie-Repräsentationssprachen (RDF) oder der darauf aufbauenden Web Ontology Language (OWL).

Zum einen geht es darum, bessere Kategorisierungsmöglichkeiten zur Verfügung zu stellen. Andererseits lassen sich mittels Annotation Schlussfolgerungen zu treffen. Beispielsweise besagt die Annotation einer Webseite, dass sie sich mit „Fußball“ beschäftigt. Und Aus der verwendeten

Ontologie geht hervor, dass es sich bei „Fußball“ um eine bestimmte „Sportart“ handelt. Man käme also zu dem Schluss, dass die Website das allgemeinere Thema „Sport“ behandelt, obwohl dies nicht ausdrücklich in den Metadaten hinterlegt wurde.

Bei einer entsprechenden Begriffswahl in der Annotation ließe sich somit ein hoher Automatisierungsgrad bei der Verarbeitung von Websites erzielen. So wäre es sehr wünschenswert, wenn in naher Zukunft Semantische Suchmaschinen durch die Implementierung Semantischer Netze auch komplexere Anfragen direkt beantworten könnten. Das Ergebnis der Suchanfrage „Wie viele Tore hat Diego Maradona bei der Fußball-WM im Jahre 1982 geschossen?“ würde dann lediglich diese eine benötigte Information enthalten. Und nicht wie bisher eine Menge von Dokumenten, die mehr oder weniger diese Begriffe enthalten - denn die Informationsextraktion ist dann wieder vom Menschen abhängig.

2.2 MVC

Model-View-Controller (MVC, „Modell/Präsentation/Steuerung“) bezeichnet ein Architekturmuster zur Strukturierung von Software-Entwicklung in die drei Einheiten Datenmodell (engl. Model), Präsentation (engl. View) und Programmsteuerung (engl. Controller). Ziel des Musters ist es, einen flexiblen Programmentwurf zu machen, der u. A. eine spätere Änderung oder Erweiterung erleichtert und eine Wiederverwendbarkeit der einzelnen Komponenten ermöglicht.

Modell (model): Das Modell enthält die darzustellenden Daten und gegebenenfalls (abhängig von der Implementation des MVC-Patterns) auch die Geschäftslogik. Es ist von Präsentation und Steuerung unabhängig. Die Bekanntgabe von Änderungen an relevanten Daten im Modell geschieht nach dem Entwurfsmuster „Beobachter“. Das Modell ist das zu beobachtende Subjekt, auch Publisher, also „Veröffentlicher“, genannt.

Präsentation (view): Die Präsentationsschicht ist für die Darstellung der benötigten Daten aus dem Modell und die Entgegennahme von Benutzerinteraktionen zuständig. Sie kennt sowohl ihre Steuerung als auch das Modell, dessen Daten sie präsentiert, ist aber nicht für die Weiterverarbeitung der vom Benutzer übergebenen Daten zuständig. Im Regelfall wird die Präsentation über Änderungen von Daten im Modell mithilfe des Entwurfsmusters „Beobachter“ unterrichtet und kann sich daraufhin die aktualisierten Daten besorgen. Die Präsentation verwendet das Entwurfsmuster „Kompositum“.

Steuerung (controller): Die Steuerung verwaltet eine oder mehrere Präsentationen, nimmt von ihnen Benutzeraktionen entgegen, wertet diese aus und agiert entsprechend. Zu jeder Präsentation existiert ein Modell. Es ist nicht die Aufgabe der Steuerung, Daten zu manipulieren. Die Steuerung entscheidet aufgrund der Benutzeraktion in der Präsentation, welche Daten im Modell geändert werden müssen. Sie enthält weiterhin Mechanismen, um die Benutzerinteraktionen der Präsentation einzuschränken. Präsentation und Steuerung verwenden zusammen das Entwurfsmuster „Strategie“, wobei die Steuerung der Strategie entspricht. Der Controller kann in manchen Implementierungen ebenfalls zu einem „Beobachter“ des Modells werden, um bei Änderungen der Daten die View direkt zu manipulieren.

2.3 Mengen- („Eimer“) Konzept

Das Eimerkonzept (unsere Wortschöpfung) stellt ein Layout- und Bedienkonzept für den Umgang mit und die Bearbeitung von Mengen von gleichartigen Objekten und deren Beziehungen dar. Hierbei wird zur Abstraktion das aus der Realwelt bekannte Bild der Eimer wieder verwendet, welche Objekte enthalten können. Mittels bekannter intuitiver Nutzungstechniken wie Drag and Drop (Objekte mit dem Mauszeiger ziehen und über dem Zielort loslassen) lassen sich diese Eimer einfach verwalten.

Die Eimer repräsentieren hierbei Ressourcen sowie deren Eigenschaften und Beziehungen.

Erster Entwurf

Anhand einer ersten Grobskizze des möglichen Aufbaus der Benutzeroberfläche (siehe Anhang) soll dieses Konzept näher erläutert werden:

Die Notation orientiert sich dabei an UML, speziell die Verwendung von Rechtecken zur Zusammenfassung ähnlicher Elemente. Diese Rechtecke repräsentieren hier jedoch die oben genannten Objekte. Diese Objektmengen können eingegrenzt werden (z.B. nur alle mit der Eigenschaft x).

Unterhalb besteht zudem die Möglichkeit, von den oberhalb ausgewählten Objekten zu jenen zu navigieren, die zu diesen in einer Beziehung stehen.

Desweiteren soll es in einem weiteren Modus möglich sein, zwei Mengen zu konstruieren und dann alle Verbindungen zwischen diesen anzeigen zu lassen. Im Bild fehlt noch eine Werkzeugleiste (Toolbox), auf der ebenfalls alle Bedienelemente frei angeordnet werden können.

Während der Nutzer mit diesen Eimern hantiert (die ja eigentlich ein SPARQL Query repräsentieren) wird der Querystring im mittleren Teil dargestellt und verändert sich entsprechend der graphischen Darstellung.

Gleichzeitig wird darunter in einer tabellenartigen Struktur die Ergebnismenge als Auflistung der Elemente und deren Eigenschaften dargestellt.

3 Applikationsstudie

3.1 Ontowiki

OntoWiki ist ein semantisches Wiki und Open-Source Software. Es soll als Ontologie Editor und als Wissensmanagementsystem dienen. Es ist in PHP geschrieben und nutzt eine SQL- oder Virtuoso-Datenbank. Es setzt auf das Semantic Web Framework Erfurt und das Zend Framework auf. Für das User Interface werden Ajax Technologien benutzt, dabei wird auf jQuery gesetzt.

Features:

- Wiki-Prinzip (jeder kann editieren)
- Versioning
- Inline Editing
- Autoadaptives User Interface

Es besitzt zahlreiche Interfaces wie:

- SPARQL Endpoint

- Linked Data Endpoint
- WebDAV
- REST API
- Command Line Interface
- LDAP

Ist erweiterbar durch

- Plugins
- Views/Templates
- Themes
- Lokalisierungen

Das in OntoWiki umgesetzte MVC-Prinzip orientiert sich an der in Zend vorgegebenen Realisierung:

- OntoWiki_View leitet sich von Zend_View und OntoWiki_Controller_Action von Zend_Controller_Action ab.
- OntoWiki_Controller_Action stellt die generelle Controller-Superklasse von OntoWiki dar und beinhaltet dabei unter anderem die grundlegenden Controller-Methoden wie init, postDispatch, getModel, getEventDispatcher und renderView.
- Eine eigene Model-Klasse, wie bei MVC eigentlich üblich, gibt es nicht. Die zu bearbeitenden Daten liegen bei OntoWiki im Wissensbasis-Format vor

Links

<http://www.informatik.uni-leipzig.de/~auer/publication/ontowiki.pdf> external (englisch)

<http://www.informatik.uni-leipzig.de/~auer/publication/WissensarbeitMitOntoWiki.pdf> external

<http://pcai042.informatik.uni-leipzig.de/~swp08-4/download/loesung5.pdf> external

3.2 Erfurt

Die Manipulation von Wissensbasen ist eine zentrale Funktionalität von OntoWiki, zu deren Realisierung das Erfurt Semantic Web Framework verwendet wird.

OntoWiki nutzt Erfurt als zentralen Bestandteil, z.B. auch um Authentifikation und Zugangskontrollmechanismen zu realisieren. Die Daten dafür werden natürlich auch in RDF Stores gehalten und es wird ein Vokabular dafür zur Verfügung gestellt. Erfurt basiert auf dem Zend Framework und ist in PHP5 geschrieben. Die Architektur ist MVC-orientiert und das Framework stellt verschiedene Event-Mechanismen zur Ablaufsteuerung zur Verfügung. Die RDF-Stores können in verschiedenen Backends realisiert sein: als Zend_Db (also z.B. MySQL, Oracle, MS SQL, SQLite, ...) oder mittels OpenLink Virtuoso (XML, Triple Store, ...).

3.3 Plugins, Komponenten, Module

OntoWiki unterscheidet zwischen drei Arten von Erweiterungen:

- Komponenten
- Module
- Plugins

Komponenten: Komponenten sind einsetzbare MVC-Controller, an die Requests weitergeleitet werden. Gewöhnlich aber nicht notwendiger Weise stellen Komponenten Inhalte für das Hauptfenster zur Verfügung und können in diesem Fall über die Navigation für Nutzer zugänglich sein. In anderen Fällen können Komponenten als Controller dienen, die asynchrone Requests bearbeiten. Komponenten werden statisch konfiguriert über die component.ini Datei im Ordner der Komponente.

Typische Anwendungen für Komponenten:

- ein neues Tab im Hauptfenster (z.B. für eigene Views auf Daten)
- eine neue Webservice-Schnittstelle

Module: Module werden als kleine Fenster dargestellt, die zusätzliche User Interface-Elemente zur Verfügung stellen, mit denen der Nutzer Einfluss auf den Hauptfensterinhalt hat. Weil Module sehr dynamische Erweiterungen sind, können sie sowohl statisch als auch dynamisch konfiguriert werden. Die statische Konfiguration funktioniert genauso wie bei anderen Erweiterungen; eine module.ini Datei befindet sich im Hauptverzeichnis des Moduls. Desweiteren muss ein Modul von OntoWiki_Module abgeleitet sein und kann viele Methoden überladen, um dynamisch anpassbar zu sein. Wenn vorhanden, überschreiben Rückgabewerte von Funktionen die Einstellungen aus der module.ini.

Typische Anwendungen für ein Modul:

- ein zusätzliches Fenster oben rechts mit eigener Navigation
- ein zusätzliches Fenster im Hauptfenster für spezifische Ressourcen

Plugins: Plugins sind die einfachsten aber auch flexibelsten Erweiterungen. Sie bestehen aus beliebigem Code, der bei bestimmten Events ausgeführt wird. Plugins müssen ihre Events in der plugin.ini registrieren, die im Hauptverzeichnis des Plugins liegen muss.

Typische Anwendungen für ein Plugin:

- eine Aktion (Action) wird jedesmal ausgeführt, wenn ein Tripel hinzugefügt wird (z.B. für eigene Regeln)
- einfach alles wofür man ein Event benutzen kann :-)

Link: <http://code.google.com/p/ontowiki/wiki/ExtensionCookbook> external