

Recherchedokument

swp-helios

6. Januar 2014

Inhaltsverzeichnis

1	Begriffe/Glossar	2
1.1	Framework	2
1.2	JavaScript	2
1.3	Java	2
1.4	LIMES	2
1.5	Linked Open Data (LOD)	2
1.6	Mapping	2
1.7	Metrischer Raum	3
1.8	Ontologie	3
1.9	RDF-Schema	3
1.10	RDF	3
1.11	SPARQL	3
1.12	URI	4
1.13	Web Application	4
1.14	XML	4
2	Aspekte	4
2.1	Java/JavaScript und Darstellung im Browser	4
2.2	MVC-Architektur	4
2.3	Link Spezifikation	4
2.4	Finden eines effizienten Algorithmus	5
2.5	Umgang mit LIMES	5
2.6	Auswahl des richtigen Ansatzes	5
3	Konzepte	5
3.1	Semantic Web	5
3.2	Linked (Open) Data	5
3.3	Link Discovery	6
4	Quellenangabe	6

1 Begriffe/Glossar

1.1 Framework

Ein Framework ist eine Software, welche eine allgemeine Funktion erfüllt und durch Code des Benutzers an seinen Anwendungszweck angepasst werden kann.

1.2 JavaScript

JavaScript ist eine interpretierte Programmiersprache, welche u.a. von Browsern implementiert wird. Sie wird hauptsächlich benutzt um Code in dem Browser von Besuchern von Internetseiten auszuführen, und damit der Seite dynamische Inhalte hinzuzufügen.

1.3 Java

Java ist eine beliebte, weitläufig benutzte objektorientierte Programmiersprache. Um ein Java-Programm auszuführen, benötigt man die Java Laufzeitumgebung. Dies hat den Vorteil, dass Java Programme unabhängig vom Betriebssystem funktionieren.

1.4 LIMES

Limes ist ein Framework, welches Links im Datennetz auffindet. Es stellt Methoden für eine performante und umfangreiche Linkermittlung innerhalb eines metrischen Raumes bereit. Mittels Web-Interface ist es leicht konfigurierbar. Auch kann es als eigenständiges Tool aus dem Internet heruntergeladen werden, um Links lokal zu ermitteln. Die Ergebnisse werden als RDF oder TSV Daten ausgegeben.

1.5 Linked Open Data (LOD)

Linked Open Data steht für frei verfügbare Daten im Internet, welche mittels Uniform Resource Identifier (URI) identifiziert sind und dadurch per HTTP abgerufen werden können und außerdem per URI auf andere Daten verlinken. Zur Kodierung und Verbindung der Daten wird das Resource Description Framework (RDF) und darauf aufgebaute Standards wie zum Beispiel SPARQL und die Web Ontology Language (OWL) verwendet. Linked Open Data ist gleichzeitig Teil vom Semantic Web.

1.6 Mapping

Mapping (bzw. Data Mapping) beschreibt den Prozess welcher zwischen 2 verschiedenen Datenmodellen eine Beziehung herstellt. Datenmodelle können atomare als auch Metadaten enthalten. Data Mapping dient als erster Schritt

der Informationsintegration. Datentransformation wird verwendet, um die Beziehung zwischen einer Anfangsdatenquelle und einer Zielquelle, in welcher die Daten benutzt werden, zu vermitteln.

1.7 Metrischer Raum

Definiert eine Menge, auf der eine Abstandsfunktion (Metrik) definiert ist, die je zwei Elementen der Menge einen nichtnegativen reellen Wert zuordnet, der als Abstand der Elemente voneinander bezeichnet wird.

1.8 Ontologie

Bezeichnet ein Netzwerk von Informationen und deren logischen Relationen. Eine formale Darstellung einer Menge von Begriffen, ihren Beziehungen untereinander in einem bestimmten Gegenstandsbereich und werden zum Strukturieren von Daten und Beziehungen und zum Wissensaustausch genutzt. Eine Ontologie besitzt eine formale Beschreibung von Daten sowie Regeln über deren Zusammenhang untereinander, dies ermöglicht durch logische Schlussfolgerungen Rückschlüsse zu ziehen, logische Fehler zu erkennen und neues Wissen zu ergänzen. Das RDF-Schema ist ein Beispiel für eine Ontologie.

1.9 RDF-Schema

Das RDF-Schema ist eine Sprache, die dazu verwendet wird Vokabulare zu definieren, welche dann benutzt werden können um Daten in RDF darzustellen.

1.10 RDF

Das Resource Description Framework (RDF) ist eine Sprache. Sie beschreibt Informationen über Ressourcen, welche durch eindeutige Bezeichner (URIs) identifiziert werden. Durch das RDF sind Informationen im Semantic Web maschinell auswertbar (mittels SPARQL), RDF ist damit ein wichtiger Bestandteil des Semantic Webs. Bereits existierende Informationen können mithilfe von z.B RDF-Schema (RDFS) maschinell erkannt werden, obwohl die Informationen nicht direkt vorliegen. Die Daten des RDF sind dabei (wahre) Aussagen über Ressourcen. Die Aussagen liegen dabei als Tripel vor, bei den 3 Elementen handelt es sich um das sogenannte Subjekt (die Quelle über die eine Aussage getroffen wird), das Prädikat (Eigenschaft) und das Objekt (Argument). Eine Menge von Tripeln bildet Graphen. Die Daten des RDF können als XML oder N3 abgespeichert werden, aber auch in Datenbanken (Triplestores).

1.11 SPARQL

SPARQL ist eine graph-basierte Anfragesprache für RDF, die vom W3C als Standard empfohlen wird.

1.12 URI

Ein Uniform Resource Identifier ist eine Zeichenfolge, die abstrakte oder physische Ressourcen identifiziert. Ein bekanntes Beispiel sind die Hypertext Transfer Protocol (kurz HTTP)URLs, die bezogen auf Webseiten als Internetadressen bekannt sind.

1.13 Web Application

Eine Web Application ist ein Anwendungsprogramm, das bei den Benutzern im Webbrowser läuft. Zugegriffen wird über das Internet. Eine Installation der Application ist nicht notwendig, allerdings können Plugins wie Flash oder Java zum Ausführen einer Application benötigt werden.

1.14 XML

(eXtensible Markup Language) ist eine Auszeichnungssprache. XML definiert Regeln um Daten in einer Weise, die von Mensch und Maschine lesbar ist, auszudrücken.

2 Aspekte

2.1 Java/JavaScript und Darstellung im Browser

Zur Darstellung der Verknüpften Datenpunkte im Browser sind Java und JavaScript dank ihrem objektorientierten Paradigmen, der großen Verbreitung und der guten Dokumentation und Kenntnis einzelner Teammitglieder eine gute Wahl, eventuell werden Alternativen in Betracht gezogen, falls diese signifikante Vorteile aufweisen sollten.

2.2 MVC-Architektur

Die MVC-Architektur ist ein Muster für die Modellierung und Implementierung des späteren Programms, dessen Vorteile eine leichte Änderung und Erweiterung des Programmcodes und eine gute Wiederverwendbarkeit sind, weswegen wir dieses Entwurfsmuster in Betracht ziehen, falls keine für unser Projekt bessere Alternative gefunden wird.

2.3 Link Spezifikation

Eine Link Spezifikation ist eine Spezifikation der Bedingungen, unter denen ein Link Discovery Tool (z.B. LIMES) Verbindungen zwischen Datensätzen annimmt.

2.4 Finden eines effizienten Algorithmus

Ein weiter wichtiger Aspekt ist das Finden und Nutzen eines laufzeiteffizienten Algorithmus, der sich gut in der gewählten Programmiersprache implementieren lässt.

2.5 Umgang mit LIMES

Der richtige Umgang mit dem LIMES Framework ist wichtig für ein gutes Projektergebnis und somit ein wichtiger Aspekt für das Projekt. Alle Teammitglieder sollten mit der Funktionsweise von LIMES vertraut sein und effizient mit dessen Tools umgehen können, dies erfordert Einarbeitung und Dokumentation der benutzten Tools und Funktionen.

2.6 Auswahl des richtigen Ansatzes

Das Finden und Nutzen eines nicht zu komplizierten und effektiven Ansatzes ist einer der Hauptaspekte des Projekts, der entscheidend für die spätere Weitergestaltung ist.

3 Konzepte

3.1 Semantic Web

Prinzip, nach dem Informationen im Netz durch Beschreibung ihrer Bedeutung (Semantik) nutzbar und auch für Maschinen verarbeitbar gemacht werden sollen, sodass diese leichter kategorisiert und geordnet werden können - das gesammelte Netz-Wissen soll so für Mensch und auch Maschine leichter zugänglich werden. Die Aufgabe dabei ist es, die gesammelten Informationen durch die Sortierung auf Basis ihres Inhalts automatisch in Beziehung setzen zu können.

3.2 Linked (Open) Data

Linked (Open) Data ist eine Methode, strukturierte Daten (frei und uneingeschränkt) im Web bereitzustellen und sie durch Verknüpfung der Daten untereinander besser nutzbar zu machen. Dabei werden Webstandards wie RDF, URI und HTTP genutzt und so auch automatische Lesbarkeit durch Computern zwischen mehreren Quelltypen ermöglicht, wobei Maschinenlesbarkeit und Verknüpfbarkeit Grundvoraussetzungen darstellen (bei Open Data ist dazu die freie Zugänglichkeit für alle ein weiteres Kriterium). Ein Beispiel für eines der größten Linked Data Projekte ist DBpedia, welches strukturierte Informationen aus Wikipedia generiert und somit komplexere Anfragen ermöglicht. Auch allgemein soll Linked Data dazu verwendet werden, exaktere, treffende

(Such)anfragen zu ermöglichen und Verbindungen innerhalb von einem Thema und verwandten Themen zu erweitern und zu verbessern.

3.3 Link Discovery

Link Discovery bezeichnet den Prozess des Auffindens von potentiellen Verknüpfungen zwischen zwei Datenpunkten aufgrund deren (gemeinsamen) Eigenschaften. Das Ziel ist dabei möglichst strukturierte, logische Verknüpfung von Daten zu finden.

4 Quellenangabe

- [1] XML, <http://en.wikipedia.org/wiki/XML>, 10.57 Uhr, 05.01.2014
- [2] LIMES, <http://aksw.org/Projects/limes>, 13.42 Uhr, 05.01.2014
- [3] SPARQL, <http://de.wikipedia.org/wiki/SPARQL>, 13.42 Uhr, 05.01.2014
- [4] Ontologie, http://de.wikipedia.org/wiki/Ontologie_%28Informatik%29, 13.42 Uhr, 05.01.2014
- [5] Semantic Web, http://de.wikipedia.org/wiki/Semantic_web, 13.48 Uhr, 05.01.2014
- [6] Linked Data,
http://en.wikipedia.org/wiki/Linked_Data, 13.48 Uhr, 05.01.2014,
<http://wiki.dbpedia.org/About>, 13.50 Uhr, 05.01.2014,
<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>,
14.01 Uhr, 05.01.2014
<http://www.semantic-web.at/LOD-TheEssentials.pdf>, 14.01 Uhr,
05.01.2014
- [7] Mapping, <http://www.wisegeek.com/what-is-data-mapping.htm>, 14.30 Uhr, 05.01.2014
- [8] Framework, https://en.wikipedia.org/wiki/Software_framework, 18.36 Uhr, 05.01.2014
- [9] JavaScript, <https://en.wikipedia.org/wiki/JavaScript>, 18.57 Uhr, 05.01.2014
- [10] RDF-Schema, <http://www.w3.org/TR/rdf-schema>, 19.22 Uhr, 05.01.2014
- [11] Metrischer Raum, http://de.wikipedia.org/wiki/Metrischer_Raum, 16.50 Uhr, 05.01.2014