Recherchenbericht

Inhalt

Literatur/Quellen Angaben
Seite 8

Begriffe

GUI

Die GUI, oder auch Graphical User Interface ist eine grafische Schnittstelle, welche dem Nutzer erlaubt, über Eingabegeräte und grafische Symbole, mit einem Programm zu interagieren. Hierbei wird die Computerlogik dem gemeinen Nutzer mithilfe grafischer Symbole, welche hier verwendet werden, ersichtlich gemacht. Die GUI bildet somit die Schnittstelle zwischen Mensch und Maschine, daher der oft verwendete Name der „MenschMaschineSchnittstelle“.

URI

Ein URI (Uniform Resource Identifier) dient dem Identifizieren von physischen oder abstrakten Ressourcen. Beispiele für URIs sind Internetadressenn wie http://www.google.de, aber auch lokale Pfadangaben wie file:///etc/fstab oder Netzwerkadressen mit spezifischem Benutzer wie ssh://user@host.

URL

URL steht für Uniform Resource Locator und ist die gängigste Adressierungsform für Internetadressen insbesondere für Dokumente innerhalb des World Wide Web. Das URL-Format macht eine eindeutige Bezeichnung aller Dokumente im Internet möglich, es beschreibt die Internetadresse eines Dokuments oder Objekts, das von einem WWW-Browser gelesen werden kann.

Eine URL besteht aus mehreren Komponenten:

- Der Protocol Identifier (PID), Kennzeichnung des Protokolls mit dessen Hilfe der gewünschte Inhalt geholt wird. Beispiele sind http, https, Gopher, FDP... Der PID leitet eine URL ein, an seinem Ende steht <://> um ihn vom folgenden Punkt abzugrenzen.

- Der Recource Name, die vollständige Adresse der gewünschten Inhalte. Der Recource Name ist in mehrere weitere Teile unterteilt: Hostname und dazu optional Pfad, Verzeichnis und Portnummer

- Hostname, bestehend aus Subdomain (z.B. www); Domain und TopLevelDomain(TLD z.B. de; com; info; org) dient zur Lokalisierung des obersten Levels der gewünschten Information

- Pfad; getrennt durch </> dient zur genaueren Lokalisation der Information sollte sich diese in einer beispielsweise verschachtelten Site befinden oder es sich bei der Information um ein Dokument eines anderen Formates handeln (z.B. pdf)

- Verzeichnis, bestehend aus Variable und Wert funktioniert in etwa wie eine Datenbank von Untersites (z.B. die Videosites von www.youtube.com), wird durch ein <?> eingeleitet und Variable und Wert werden durch ein <=> getrennt

- Portnummer;eine Alternative um bestimmte Server oder Hosts ansteuern zu können; notwendig für TCP oder UDP Verbindungen.

Linked Data

Im Semantic Web Terminologie ist Linked Data bezeichnet man ein Verfahren zur Belichtung und Verbindungsdaten im Internet aus verschiedenen Quellen zu beschreiben. So werden Daten mit einem Uniform Resource Identifier (URI) identifiziert, um so auch auf andere Daten verweist, die eventuell weitere Informationen zum selbigen Gegenstand bereithalten. Das wohl bekannteste Beispiel für URIs ist der Uniform Resource Locator (URL), der die Lokation einer Ressource im Internet angibt. Derzeit verwendet die Web Hypertext-Links, die Menschen, um von einem Dokument zu einem anderen bewegen können. Die Idee dahinter ist, dass Linked Data Hyperdata Links Leute lassen oder Maschinen zu finden bezogenen Daten auf dem Web, die zuvor nicht verbunden war.

Abfragesprache

Eine Abfragesprache ist eine formale Sprache, die dem Abfragen von Informationen aus Informationsquellen wie z.B. Datenbanken dient, bzw. die vorliegenden Informationen nach bestimmten Kriterien filtert (vergleichbar mit Suchmaschinen). Verschiedene Abfragesprachen unterscheiden sich dabei in ihrer Komplexität und ihrem Einsatzzweck. Beispiele für Abfragesprachen sind SQL und SPARQL für Datenbanksysteme, oder XQUERY für

XML-Informationssyteme.

Framework

Im SoftwareEngineering ist ein Framework ein modernes Rahmenwerk, das dem Programmierer den Entwicklungsrahmen für seine Anwendungsprogrammierung zur Verfügung stellt und damit die SoftwareArchitektur der Anwendungsprogramme bestimmt. Das Framework wird vorwiegend in der objektorientierten Programmierung eingesetzt und umfasst Bibliotheken und Komponenten wie Laufzeitumgebung und stellt die Designgrundstruktur für die Entwicklung der Bausteine zur Verfügung. Das Framework selbst ist nicht als fertiges Programm zu verstehen. Es kontrolliert weiterhin den Kontrollfluss der Anwendung, sowie dessen Schnittstellen, und gibt letzten Endes auch die Anwendungsarchitektur vor. Eine genauere FrameworkDefinition ist nicht allgemein möglich, da es zu viele Frameworks mit zu vielen zentralen Unterschieden gibt, und somit müssen die Frameworks im Einzelnen beschrieben werden

Link Discovery

LInk Discovery ist eine Technology aus dem Data Mining und dient dem FInden von ähnlichen Mustern oder Inhalten. Dabei werden verschienene Vorgehensweisen unterschieden: die Überwachte und die nicht überwachte Herangehensweise. Bei der überwachten Link Discovery werden bereits gelernte Muster zu den zu überprüfenden Daten hinzugefügt. Der Algorithmus findet dann dem Muster entsprechende Datensätze. Dadurch können geziehlte Informationen gesucht werden. Der Ansatz ist besonders hilfreich, wenn im Voraus bekannt ist wie die Daten aussehen sollten oder über die gewünschten Daten bereits Informationen bekannt sind. Bei nicht überwachten Anätzen werden nur die Daten in den Algorithmus gegeben und Datensätze mit auffälligen Mustern werden dann herausgegeben. Der Vorteil ist, dass wenn die Daten völlig unbekannt sind das Verfahren trotzdem Erfolge bringt und auch neue Muster erkannt werden können,welche zuvor nicht erkannt wurden. Dadurch können Zusammenhänge aller Art gefunden werden.

Cache

Unter einem Cache versteht man einen Zwischenspeicher, welcher dem Anwender in der Regel verborgen ist, der es ermöglicht Inhalte und Daten, die bereits berechnet oder beschafft wurden, schneller wieder zur Verfügung zu stellen. Dies ermöglicht einen schnelleren Zugriff auf Daten, welcher sonst längeren Zeitanspruch benötigen würde. Ein Cache, der noch keine Daten enthält und somit nicht optimal arbeitet wird auch als „kalter Cache“, ein gefüllter als „heißer“ Cache bezeichnet.

Linking

Als Linking wird die Suche nach Links auf einer Website oder in einem Webportal bezeichnet. Eine bekannte Anwendung für diese Suche ist LIMES

Metric Spaces

Ein metrischer Raum wird in der Mathematik als eine Menge verstanden, auf der eine Metrik definiert ist. Diese Metrik ordnet zwei Elementen des Raumes einen nicht negativen reellen Wert zu.
Bei der Web Service Discovery wurde früher der Ansatz über die beschreibenden Sprache gewählt, da die Services aber nicht gleichmäßig verteilt sind liegt es nahe sie über einen metrischen Raum zu modellieren.

Konzepte

Semantic Web

Semantic Web bezeichnet eine Erweiterung für das noch bestehende Web, welches Computern die Möglichkeit verleihen soll, nicht nur nach Daten zu suchen, sondern auch deren Inhalt zu verstehen. Dies ist als Umgang mit der enormen, aber unstrukturierten und dezentralisierten, Datenmenge des Internets angesehen. Um dies zu ermöglichen, müssen diese Daten in einer strukturierten Form aufbereitet werden, um sie für Maschinen interpretierbar zu machen. Dieses Ziel wird durch neue Technologien wie RDF, SPARQL, und OWL erreicht.

Limes

Das Framework LIMES ist konzipiert worden um Links zwischen Entitäten aus Linked Data Sources zu finden. Hierbei kombiniert es mathematische Charakteristika des metrischen Raums, als auch Suffix, Präfix, und Positionsfilterung, um für die Datensätze eine Approximation der Gleichheit zu berechnen. Diese Annäherung wird nun genutzt um eine große Anzahl von Datensatzpaaren auszufiltern, die den Ansprüchen des Mappings nicht genügen.

SPARQL

SQARQL (rekursives Akronym für SPARQL Protocol And RDF Query Language) ist eine graph-basierte Abfragesprache für Daten im RDF-Format. Sie wurde vom W3C entwickelt und 2008 zum Standard ("Recommendation") erklärt und ist mit anderen Abfragesprachen wie SQL vergleichbar. SPARQL unterstützt von Haus aus die Dateiformate XML, JSON, CSV und TSV, um die Interoperabilität mit anderen Programmen / Sprachen zu erleichtern.

RDF

RDF (Ressource Description Framework) beschreibt einen Standard zur Formulierung von Aussagen über beliebige Ressourcen. Diese Aussagen bestehen immer aus einem Subjekt, einem Prädikat und einem Objekt, zum Beispiel: Klaus (Subjekt) studiert (Prädikat) Informatik (Objekt). Aussagen im RDF-Standard lassen sich als gerichteter Graph darstellen, wobei das Prädikat die Kante zwischen Subjekt und Objekt bildet.

Java

Java ist eine objektorientierte, plattformunabhängige Programmiersprache in welcher dieses Projekt verfasst werden soll. Vorteile von Java sind unter Anderem, die Einfachheit gegenüber anderen Progammiersprachen wie C++, sowie die Objektorientierung, die das Programmieren erleichtert. Auch die Plattformunabhängigkeit dieser Sprache ist als Pluspunkt zu betrachten, sowie die Tatsache, das LIMES in Java verfasst ist.

Link Specs

Link Specs sind das Ergebnis eines LIMES Arbeitsvorgangs. Der Nutzer erhält eine .ntDatei, die eine Liste von allen Paaren von Links enthält, von denen vermutet wird, die RDFs seien sich sehr ähnlich. Diese Datei kann der Nutzer privat Speichern, erneut Verändern, und auch ein erneutes Laden ist möglich. Es besteht die Möglichkeit, die Specs öffentlich zu machen, dies lehnt jedoch mit Wahrscheinlichkeit an den Zugriffslevel des Nutzers an. Auch das Anlernen von Link Specs soll über Batch und Active Learning möglich sein, indem ein Nutzer Daten auswählt und hochlädt. Eine Feedbackfunktion soll hierbei integriert sein, um Schwachstellen und stärken über die Erfahrung vieler Nutzer erkennbar zu machen. Großer Wert wird dabei auf eine weitere Funktion gelegt, und zwar auf die Möglichkeit der Bewertung von Link Specs. Zuerst einmal ist es möglich, eine Anzahl zufälliger Link Specs auszuwählen. Der Nutzer kann diese nun völlig frei Bewerten und auch unter negativen sowie positiven Link Specs einordnen. diese Bewertung kann gespeichert werden. Auch diese Funktion kann als Feedback dienlich sein.

Active Learning

Ein Spezialfall des maschinellen Lernens ist das Active Learning. Hierbei ist ein Lernalgorithmus in der Lage den Nutzer, oder eine andere Informationsquelle, anzufragen um gewünschte Ausgaben an neuen Datenpunkten zu erlangen. Active Learning wird in Situationen angewandt, in denen eine riesige Menge an Rohdaten vorhanden ist, aber eine manuelle Kennzeichnung dieser Daten zu teuer wäre. Das Konzept des Active Learning wird von LIMES genutzt.

Batch

Batchverarbeitung, auch Stapelverarbeitung genannt, bezeichnet die sequenzielle Datenverarbeitung am PC. Über BAT-Dateien kann an DOS-Betriebssystemen einen BAtch gestartet werden, welche alle Befehle nacheinander abarbeitet.

Aspekte:

JavaFX

JavaFX ein plattformübergreifendes Framework für Internetanwendungen. Es kann auf diversen Endgeräten zum Einsatz kommen und ist fester Bestandteil der Java Run Time Enviroment.
Zumeist werden JavaFX Anwendungen für Desktop Computer über Java Web Start oder direkt als Java Applet ausgeführt. Die für die JavaFX notwendigen Dateien müssen nicht direkt auf dem Rechner vorhanden sein, da JavaFX mit einem Webserver über die HTTP-GET, REST oder Webservices kommunizieren kann.

RDF/XML

RDF/XML ist ein Standard, um RDF-Daten in Form von XML-Dokumenten abzuspeichern. Hierbei wird die Metasprache XML (Extensible Markup Language) verwendet, um die RDF-Daten in einfachen Textdateien abzuspeichern. Diese Textdateien enthalten keinen Binärcode, sind also menschenlesbar.

Swing

Swing ist eine Java Erweiterung für die Entwicklung von Oberflächen. Dabei können Buttons Labels Fenster und ähnliches verwendet werden. Der Aufbau ist modular und bjektorientiert. Modular da die Elemente zu Gruppen oder verschachtelten Struckturen zusammengefasst werden können und Objektorientiert da jedes Element einzelnd programmiert werden und voneinander unabhängig funktionieren. Der Aufbau ist simpel und ermöglicht dennoch ist das Ergebnis ausreichend für die meisten Anwendungen.

Quellen:

https://de.wikipedia.org/wiki/Abfragesprache

https://de.wikipedia.org/wiki/Resource_Description_Framework

https://en.wikipedia.org/wiki/SPARQL

http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/

https://de.wikipedia.org/wiki/Uniform_Resource_Identifier

https://en.wikipedia.org/wiki/RDF/XML

http://www.w3.org/TR/rdf-syntax-grammar/

http://de.wikipedia.org/wiki/JavaFX

http://www.csie.ntu.edu.tw/~sdlin/publication/linsd_nldrarity.pdf

http://www.sistrix.de

http://www.itwissen.info

http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5198519&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5198519