

LinkedSpending

SWP14-LS

- Entwurfsbeschreibung -

letzte Änderung: 31.5.2014

Inhaltsverzeichnis

1. Allgemeines.....3

2. Produktübersicht.....3

3. Struktur des Programmes.....4

 3.1 Der Scheduler und das REST-Interface.....4

 3.2 Der JsonDownloader.....4

 3.3 Der Converter.....4

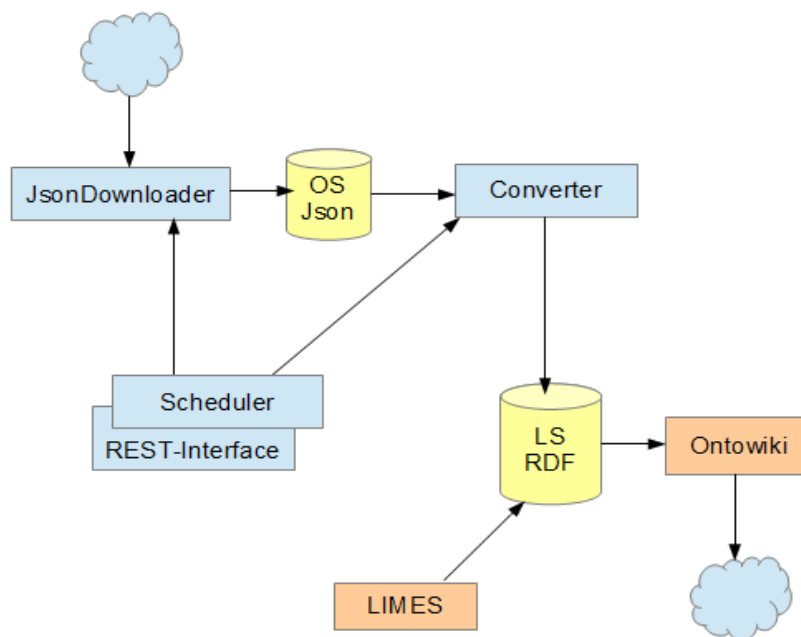
4. Testkonzept.....5

5. Glossar.....6

1. Allgemeines

OpenSpending ist ein Projekt mit dem Ziel Finanzdaten über Regierungs- und Unternehmensausgaben zu sammeln und ansprechend zu visualisieren. LinkedSpending verknüpft diese Datenmenge mit dem SemanticWeb, indem es diese in das RDF-Format überführt, wodurch die Zusammenhänge besser verarbeitet und somit komplexere Visualisierungen automatisch erstellt werden können.

2. Produktübersicht



Das Produkt besteht im wesentlichen aus drei Komponenten. Der JsonDownloader lädt die verfügbaren Datensätze von openspending.org herunter und stellt sie im Json-Format bereit (im Schema OS JSON genannt). Diese Datensätze werden anschließend vom Converter in das RDF-Format überführt und aufbereitet, um sie später mit Hilfe des Ontowiki öffentlich abruf- und nutzbar zu machen (in der Grafik LS RDF genannt). Diesen Gesamtprozess steuert der Scheduler, welcher über ein REST-Interface verfügt. Über diese Schnittstelle wird es außerdem möglich, dass ein Nutzer das Programm selbst steuert.

3. Struktur des Programmes

3.1 Der Scheduler und das REST-Schnittstelle

Nach der Ausführung des Programms wird zunächst der Scheduler gestartet. Dieser initialisiert anschließend einen Http-Server, welcher auf Anfragen reagiert, die an das REST-Interface gestellt werden. Standardmäßig befindet der Scheduler zu diesem Zeitpunkt im automatischen Modus. Das bedeutet, dass ein kompletter Lauf des Programmes bestehend aus einem vollständig Download aller verfügbaren Datensätze und deren Konvertierung, nur zu einem festlegbaren Zeitpunkt gestartet wird. Standardmäßig ist dieser Startzeitpunkt Samstag, 20 Uhr mit einer Wiederholung nach jeweils 2 Wochen. Alternativ zur automatischen Steuerung ist auch ein manueller Betrieb des Programmes möglich. Im manuellen Modus, in den der Scheduler per REST-Anweisung versetzt werden kann, können die Abläufe des Programmes direkt vom Nutzer gesteuert werden. Beispielsweise können komplette Downloads, der Download eines speziellen Datensatzes, die Konvertierung vorhandener Datensätze sowie Pausierung, Fortsetzung und Stoppen des aktuellen Vorganges (also Download oder Konvertierung) ausgelöst werden.

3.2 Der JsonDownloader

Der JsonDownloader ist in der Lage, sowohl einzelne Datensätze als auch das gesamte Angebot an Daten von openspending.org herunterzuladen. In letzterem Fall werden stets mehrere Datensätze parallel zueinander abgearbeitet, wobei größere Dateien in mehrere Teile aufgespalten werden. Nach erfolgreichem Abschluss des Downloadvorganges werden diese dann wieder zu kompletten Datensätzen zusammengefasst. Sollte der Nutzer den Downloader über die REST-Schnittstelle stoppen, so werden unfertige Datensätze und deren Teile wieder gelöscht, um Fehlern vorzubeugen. Wird der JsonDownloader vom Nutzer pausiert, werden auch alle parallelen Datensatzdownloads pausiert.

3.3 Der Converter

Der Converter dient zur Umwandlung der vom JsonDownloader bereit gestellten Json-Dateien in das RDF-Datenformat. Während dieses Vorganges werden die neuen Daten auch aufbereitet, um später komplexere Anfragen über den SPARQL-Endpunkt des Ontowikis zu ermöglichen. Treten Beim Konvertierungsprozess Fehler in größerer Zahl auf, beispielsweise aufgrund fehlerhafter Daten seitens openspending.org oder Fehlern beim Downloadprozess, so bricht der Converter mit einem Hinweis auf die zu große Zahl der Fehler ab. Nach einem erfolgreichen Durchlauf löst der Converter weiterhin eine RSS-Feed-Ausgabe aus, die die Ergebnisse des Konvertierungsprozesses zusammenfasst. Weiterhin kann diese dann auf der Startseite des Ontowiki dargestellt und abgerufen werden.

4. Testkonzept

Da es schwer ist in einem wachsenden Projekt die Korrektheit aller Teile einer Funktion zu überwachen, benutzen wir das Test Framework Junit. Dieses Framework stellt eine Reihe von Funktionen bereit, die es uns erlauben sowohl alle Klassen, als auch Funktionsabläufe zu testen. Dafür werden kleine Funktionen erstellt, welche eine bestimmte Teilkomponente oder Aufgabe testen, bei der das Ergebnis bekannt ist. Das Ergebnis der Funktion wird anschließend mit dem erwarteten Wert verglichen. Stimmen beide Werte nicht überein schlägt der Test fehl.

Die Tests werden von den Entwicklern neben der Implementierung des eigentlichen Programms entwickelt, so dass zu jedem Release die Fehlerfreiheit aller neuen Komponenten gewährleistet ist.

Nach der Fertigstellung zweier oder mehrerer zusammenarbeitenden Komponente werden spezielle Integrationstests erstellt, die prüfen, ob Schnittstellen richtig funktionieren und Funktionsaufrufe den erwarteten Effekt erzielen.

5. Glossar

Dataset:

Ein Dataset bezeichnet eine größere, zusammenhängende Datenmenge und besteht aus Metadaten welche den Inhalt beschreiben und der dazugehörigen Sammlung von Einträgen.

OntoWiki:

Das Ontowiki ist ein semantisches Programm für Wissensmanagement im Semantic Web Kontext.

RDF (Resource Description Framework):

Das Resource Description Framework bezeichnet eine technische Herangehensweise im Internet zur Formulierung logischer Aussagen über. Im RDF-Modell besteht jede Aussage aus den drei Einheiten Subjekt, Prädikat und Objekt.

REST (Representational State Transfer):

REST bezeichnet ein Programmierparadigma für Webanwendungen, welches besagt dass eine URL genau einen Seiteninhalt als Ergebnis einer serverseitigen darstellt.

OpenSpending:

OpenSpending ist eine offene Plattform zum Datenaustausch welche versucht alle Transaktionen zwischen Regierungen und Unternehmen aufzuzeichnen und visuell zu präsentieren.

SemanticWeb:

Das Semantic Web ist eine Instanz von semantischen Netzen und außerdem eine Erweiterung des WWW. Ziel ist es die Bedeutung von Informationen für Computer verwertbar zu machen und damit automatisch für die interessierten Nutzer im Zuge einer Abfrage zu ordnen.

SPARQL (SPARQL Protocol and RDF Query Language):

SPARQL ist eine graphbasierte Abfragesprache für RDF. Mit SPARQL ist es möglich komplexere Anfragen zu stellen als es reine Textsuchen ermöglichen würden.