



UNIVERSITÄT LEIPZIG

SOFTWARETECHNIKPRAKTIKUM 2015

Link Discovery

Verantwortlicher:
Sascha Hahne

Betreuer:
Felix Conrads

19. Januar 2015

Inhaltsverzeichnis

1	Projektvision	1
2	Voraussetzungen	1
3	Designübersicht und Funktionalität	1
3.1	Nutzergruppen	1
3.2	Funktionalität	2
4	Arbeitspakete	3
5	Qualitätssicherung	5
	Glossar	6

1 Projektvision

Ziel des Projektes ist es, eine Software zu entwickeln, welche als grafisches Interface für das LIMES-Framework dient. Die GUI soll möglichst einfach zu bedienen sein und einen guten Überblick über die Berechnungen geben. Um auch auf zukünftige Anforderungen an LIMES und neuere Versionen des Frameworks vorbereitet zu sein, ist ein modulares Interface von hoher Priorität, damit Wartbarkeit und Erweiterbarkeit der GUI einfach umzusetzen sind. Das Projekt wird mithilfe von JavaFX umgesetzt werden, damit es auf möglichst vielen Plattformen benutzbar ist.

2 Voraussetzungen

Die grundlegendste Voraussetzung zum Gelingen des Projektes ist natürlich die Bereitstellung und Nutzung von LIMES im Rahmen des Projektes und ein gründliches Verständnis der Funktionalitäten dieses Frameworks. Dafür ist es ebenfalls notwendig, sich die fundamentalen Konzepte des Semantic Web und der Link Discovery anzueignen, um die zu implementierende Oberfläche möglichst nutzerfreundlich, aber dennoch funktional zu gestalten. Als Orientierung bieten sich hierbei bereits realisierte Oberflächen für LIMES an.

3 Designübersicht und Funktionalität

3.1 Nutzergruppen

Endanwender Hat das Ziel, Ähnlichkeiten und Übereinstimmungen in Datensätzen zu finden. Dieser legt dabei Wert auf Übersichtlichkeit und einfache Handhabung der Oberfläche. Häufig verwendete Funktionen sollen schnell erreichbar und die Benutzeroberfläche interaktiv und intuitiv sein. Zur Erleichterung der Handhabung steht dem Nutzer eine Kurzanleitung zur Verfügung, die über die GUI eingesehen werden kann.

Administrator Die Rolle des Administrators unterscheidet sich nur gering von der des Benutzers, da die Endpoints nur eingesehen werden und keine Editierung über die GUI stattfindet. Hauptaufgabe des Administrators ist hierbei die Wartung, Anpassung und mögliche Erweiterung des Quelltextes um weitere Funktionen. Dies wird durch das Qualitätssicherungskonzept sichergestellt.

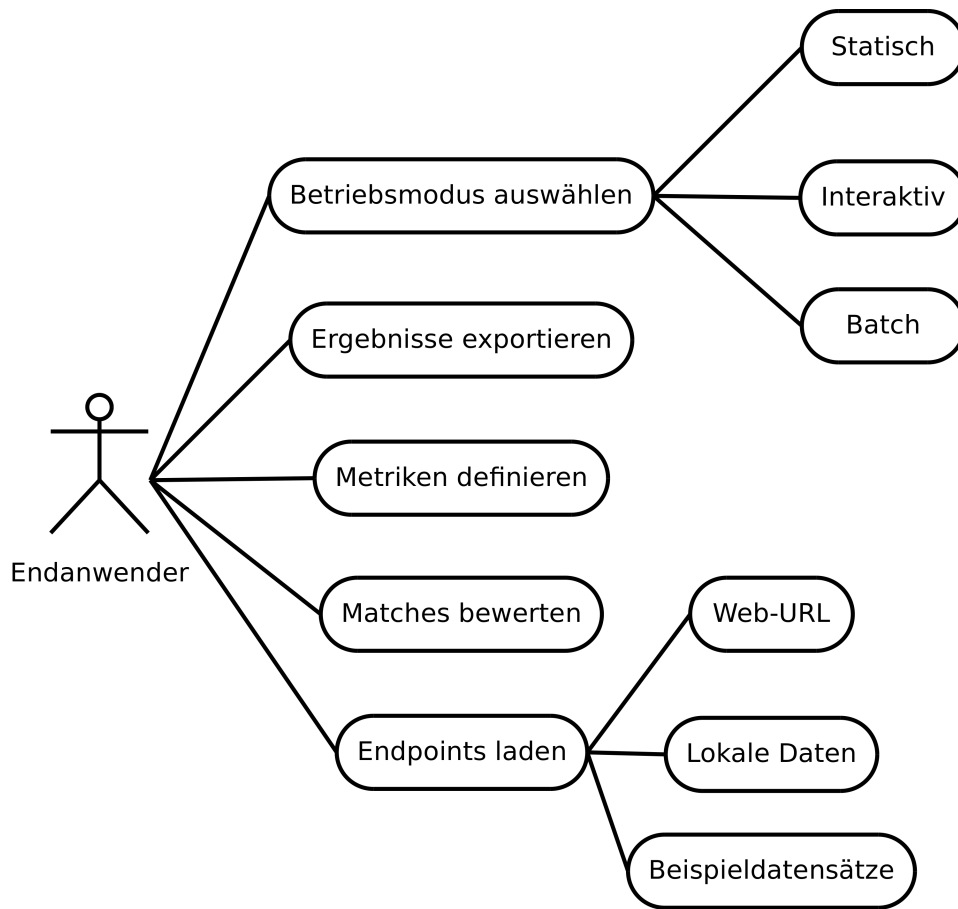


Abbildung 1: Funktionen aus Endanwendersicht.

3.2 Funktionalität

/LF10/ Geschäftsprozess: Auswahl des Betriebsmodus

Akteur: Endanwender

Beschreibung: Der Nutzer kann zu Beginn seiner Anfrage auswählen, welchen Betriebsmodus er verwenden möchte. Er hat dabei die Auswahl zwischen statisch, interaktiv und Batchbetrieb.

/LF20/ Geschäftsprozess: Laden einer Steuerdatei für den Batchbetrieb

Akteur: Endanwender

Beschreibung: Es besteht die Möglichkeit, eine Datei einzulesen, in der die zu verwendenden Metriken und Endpoints definiert sind. Diese werden hintereinander abgearbeitet.

/LF30/ Geschäftsprozess: Definition der Metrik

Akteur: Endanwender

Beschreibung: Der Nutzer kann die zu vergleichenden Attribute, die zugehörigen Metriken, deren Attribute und Wichtung und die Verknüpfungen der Metriken untereinander festlegen.

/LF40/ Geschäftsprozess: Endpoints laden

Akteur: Endanwender

Beschreibung: Die zu verwendenden Datensätze können angegeben werden als Web-URL oder als lokaler Datensatz. Zusätzlich besteht die Möglichkeit, Beispieldatensätze auszuwählen.

/LF50/ Geschäftsprozess: Matches bewerten

Akteur: Endanwender

Beschreibung: Die von LIMES ermittelten Matches werden tabellarisch dargestellt, zusammen mit deren Bewertung. Der Nutzer kann entscheiden, ob es sich dabei tatsächlich um einen Match handelt.

/LF60/ Geschäftsprozess: Exportieren der Ergebnisse

Akteur: Endanwender

Beschreibung: Die fertigen Ergebnisse können als LinkSpecs in Form einer .nt-Datei exportiert werden.

4 Arbeitspakete

1. Benchmark erstellen und GUI an LIMES anbinden

- Zeitlicher Aufwand: 15%
- Muss-Ziele:
 - **/M10/10** Durchführung und Auswertung eines Benchmarks mit 20 verschiedene Konfigurationen
 - **/M10/20** Implementierung Import und Ausführung von Link-Spec-Dateien durch die GUI

2. Konzeption der modularen Oberfläche

- Zeitlicher Aufwand: 15%
- Muss-Ziele:
 - **/M20/10** Erstellung des Konzeptes der modularen Oberfläche

3. Implementierung des Import-Assistenten

- Zeitlicher Aufwand: 10%
- Muss-Ziele:
 - /M30/10 Implementierung eines Assistenten zum Import der lokalen Endpoint-Daten
 - /M30/20 Implementierung eines Assistenten zum Import von Endpoint-Daten mit beliebigen URLs

4. Implementierung der Modellierung der Abfrage

- Zeitlicher Aufwand: 30%
- Muss-Ziele:
 - /M40/10 Implementierung der grafischen Modellierung der Abfragegraphen
- Kann-Ziele:
 - /K40/10 Implementierung einer textbasierten Modellierung der Graphen

5. Implementierung des Exports der LinkSpec

- Zeitlicher Aufwand: 10%
- Muss-Ziele:
 - /M50/10 Speichern der Abfragen als LinkSpec

6. Implementierung der Anzeige der Ausgaben

- Zeitlicher Aufwand: 15%
- Muss-Ziele:
 - /M60/10 Implementierung einer Visualisierung der LIMES-Anfrage

7. Implementierung der Self Configuration und Active Learning

- Zeitlicher Aufwand: 20%
- Muss-Ziele:
 - /M70/10 Implementierung des Self-Configuration-Modus
 - /M70/20 Implementierung des Active-Learning-Modus

5 Qualitätssicherung

Produktqualität	Sehr gut	Gut	Normal	Nicht relevant
Funktionalität			x	
Zuverlässigkeit			x	
Benutzbarkeit	x			
Anpassbarkeit	x			
Übertragbarkeit				x

Bei diesem Softwareprojekt wird das Hauptaugenmerk auf die Konzeption einer intuitiven und erweiterbaren GUI gelegt. Deswegen ist es besonders wichtig im Verlauf des Projekts auf die Qualität der Benutz- und Anpassbarkeit zu legen. Die Benutzbarkeit ist in zweierlei Hinsicht zu gewährleisten: Zu einem soll die Oberfläche bei Nutzern, die möglicherweise noch nicht mit LIMES vertraut sind einen einfachen Einstieg bieten, zum anderen sollen geübte Anwender von LIMES schnell und effizient Berechnungen modellieren speichern können. Da die hier entwickelte Software sich komplett auf das LIMES-Framework stützt, wird auch die Funktionalität fast ausschließlich durch LIMES gewährt, dennoch sollten die grafischen Methoden die hier eigenständig implementiert werden ausreichend betrachtet werden. Durch die einseitige Ausrichtung auf das LIMES-Framework ist die Übertragbarkeit auf andere Frameworks nicht relevant. Zuverlässigkeit spielt in der Entwicklung nur insofern eine Rolle, als dass die GUI die Abfragen an LIMES zuverlässig ausführen muss, da die wichtigen Berechnungen vom Framework übernommen werden.

Glossar

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. 4

Benchmark Die Bewertungs- und Messverfahren zur Beurteilung der Performance einer Software werden als Benchmarks bezeichnet. Diese dienen zum Vergleich der Leistungsfähigkeit verschiedener Systeme.. 3

Endpoint Ein EndPoint ermöglicht es einem User (menschlich oder nicht-menschlich) eine Anfrage an eine Wissensdatenbank mithilfe einer Anfragesprache zu stellen. Das Resultat wird üblicherweise in einem oder mehreren maschinenverarbeitbaren Formaten ausgegeben. Somit stellen EndPoints ein maschinenfreundliches Interface für eine Wissensdatenbank dar. 1–4

Framework Im Software-Engineering 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 Framework-Definition ist nicht allgemein möglich da es zu viele Frameworks mit zu vielen zentralen Unterschieden gibt, und somit müssen die Frameworks im Einzelnen beschrieben werden. 1, 5

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 "Mensch-Maschine-Schnittstelle". 1, 3, 5

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. 1

LIMES Das Framework LIMES ist konzipiert worden, um Links zwischen Entitäten aus Linked Data Sources zu finden. Hierbei kombiniert es mathematische Charakteristika des metrischen Raumes, als auch Suffix, Präfix und Positionsfilerung, 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. 1, 3-5

LinkSpec LinkSpecs sind das Ergebnis eines LIMES-Arbeitsvorgangs. Der Nutzer erhält eine .nt-Datei, 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 LinkSpecs einordnen. diese Bewertung kann gespeichert werden. Auch diese Funktion kann als Feedback dienlich sein.. 3, 4

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.. 3, 4