



TECHNISCHE UNIVERSITÄT
CHEMNITZ

Fakultät für Informatik

CSR-19-08

**Informationsverwertung basierend
auf qualitätsoptimierten
semistrukturierten Datenbeständen
im Wachstumskern “LEDS”**

André Langer · Valentin Siegert · Martin Gaedke

Juli 2019

Chemnitzer Informatik-Berichte

Schlussbericht

nach BNBest-BMBF 98 Nr. 3.1



LINKED ENTERPRISE DATA SERVICES

<https://www.leds-projekt.de>

| | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------|
| Zuwendungsempfänger: Technische Universität Chemnitz Fakultät für Informatik Professur Verteilte und Selbstorganisierende Rechnersysteme | Förderkennzeichen: 03WKCG11D |
| Vorhabenbezeichnung: Wachstumskern LEDS – Linked Enterprise Data Services – Verbundprojekt 01; TP4: Informationsverwertung basierend auf qualitätsoptimierten semistrukturierten Datenbeständen | |
| Laufzeit des Vorhabens: 01.07.2015 – 30.11.2018 | |

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 03WKCG11D gefördert.

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

WACHSTUMSKERNE
UNTERNEHMEN
Die BMBF-Innovationsinitiative
Neue Länder REGION

I. KURZE DARSTELLUNG

1. AUFGABENSTELLUNG

Dies ist der Projektbericht der Technischen Universität Chemnitz zu dem Projekt Linked Enterprise Data Services (LEDS), welches von Juli 2015 bis November 2018 im Förderprogramm Innovative regionale Wachstumskerne des Bundesministeriums für Bildung und Forschung (BMBF) in der Region Westsachsen in Deutschland durchgeführt wurde. Das Bündnis LEDS hatte das Ziel der Evolution von Enterprise IT Infrastrukturen zu semantisch vernetzten Daten-Diensten für elektronische Geschäftsprozesse. Dies umfasste die Entwicklung und Etablierung, einer neuen Generation semantischer, vernetzter Daten-Anwendungen auf Basis des Linked Data Paradigmas. Dafür wurde eine regionale Technologieplattform realisiert, die sowohl Kompetenzen und Methoden, als auch konkrete Technologiebausteine umfasst. Das Bündnis knüpft damit direkt an Entwicklungen zum Web 3.0 und im Bereich E-Government an die Agenda 2020 an.

Die größte Herausforderung für Unternehmensdaten und Anwendungen im E-Commerce und E-Government im nächsten Jahrzehnt ist der Übergang von der Integration statischer und homogen strukturierter Daten hin zur dynamischen Integration heterogener, verteilter und semantisch-strukturierter Daten (z.B. entlang von Wertschöpfungsketten und -netzen). Ziel des Wachstumskerns war es daher, eine Technologieplattform zu entwickeln, zu testen und bei Bündnispartnern und Kunden prototypisch zu implementieren. Unternehmensinterne Daten konnten dadurch mit Hilfe von Referenzdaten aus dem Linked Data Web aufbereitet und als "Corporate Linked Data Cloud" im eigenen Unternehmen und der eigenen E-Commerce-Wertschöpfungskette bereitgestellt werden.

Analog zur Entwicklung von unternehmensinternen Intranets als Ergänzung zum weltweiten Internet steht die Entwicklung hin zu unternehmensinternen Data-Webs unmittelbar bevor. Unternehmen werden damit in die Lage versetzt, ihre internen, strukturierten als auch unstrukturierten Informationen besser zu integrieren und mit dem Wissen des Semantic Webs zu vernetzen. Durch die Verwendung gemeinsamer Ontologien und entsprechender Vokabulare können die Kosten der Datenintegration im Unternehmen insbesondere bei einer großen Anzahl von Datenquellen gegenüber konventionellen Ansätzen wesentlich gesenkt werden. Auf diesem Weg können eine Vielzahl innovativer Anwendungen im Bereich E-Commerce, Suchtechnologie, E-Government u.v.m. implementiert werden. Diese ermöglichen ein neues Nutzererlebnis und den betreibenden Firmen einen Wettbewerbsvorsprung sowie eine aktive Teilhabe an zukünftigen Entwicklungen des Internets. Basierend auf den bereits erworbenen Kompetenzen und bisherigen wirtschaftlichen Erfolgen der Bündnispartner sieht sich das Bündnis in der sehr guten Ausgangslage, ebendiese Entwicklung voranzutreiben sowie einen innovativen Wachstumskern in der Region zu initiieren.

Sechs Arbeitsschwerpunkte (ASP) im Verbundvorhaben leisteten unterschiedliche Beiträge zur Lösung von Problemen potenzieller Kunden. Dies umfasste das Management von Hintergrundwissen (ASP-A), Qualitäts- und Kohärenzaspekte (ASP-B), Wissens-Extraktion und Semantische Inhaltsanalyse (ASP-C), die skalierbare Suche in Linked Data (ASP-D), Vernetzte E-Commerce-Dienste (ASP-E) und Vernetzte E-Gouvernement-Dienste (ASP-F). Mit umfangreicher Expertise im Bereich Web Engineering (insbesondere Mashup, Evolution, Komponententechnologie und Wiederverwendung), Linked Data und Data Quality arbeitete die Professur Verteilte und Selbstorganisierende Rechnersysteme (VSR) der TU Chemnitz im LEDS-Verbundprojekt im Teilprojekt 4 an der Informationsverwertung basierend auf qualitätsoptimierten semistrukturierten Datenbeständen. Dies umfasste die folgenden Teilaufgaben:

ASP-B Qualität und Kohärenz war der Verbesserung der Qualität von Linked Data und der Kohärenz durch entsprechende Linking und Matchingverfahren gewidmet. Die Nutzung von Linked Open Data im Unternehmen ist nur praktikabel, wenn einerseits eine ausreichende Kohärenz zwischen Daten im Daten Web und Unternehmensintern erreicht wird und andererseits die Qualität der Daten in Bezug auf den geplanten Einsatz bewertet werden kann.

Sind diese beiden Aspekte ausreichend berücksichtigt, können Unternehmen durch die Nutzung der Milliarden von frei verfügbaren Wissensseinheiten enorme Potentiale in Bezug auf Flexibilität und Effizienz Ihrer IT-Systeme realisieren. ASP-B löste diese Probleme, indem es leistungsfähigere Software für das Linking und Matching, sowie die Verbesserung der Qualität von Linked Data Wissensbasen lieferte. Während sich die Agile Knowledge Engineering and Semantic Web Group der Universität Leipzig in ihrem Teilprojekt mit der Erprobung von Link Matching Verfahren beschäftigte, konzentrierte sich die Technische Universität Chemnitz auf die Bewertung von Datenqualitätsaspekten von Linked Data. Die Sicherung der Datenqualität in semantischen Wissensbasen ist durch die Erforschung von Verfahren zur Überprüfung von Daten nach den Gesichtspunkten Relevanz, Konsistenz, Aktualität und Richtigkeit sowie der Korrektur von invaliden Datensätzen gekennzeichnet. Die entwickelten Methoden wurden direkt in die Technologieplattform integriert und stellen damit einen entscheidenden Beitrag dar.

Im ASP-C „Extraktion“ ging es darum, unstrukturierte, semi-strukturierte und strukturierte Inhalte (Texte/Dokumente, XML, Datenbanken) zu analysieren und mittels Hintergrundwissen aus dem Daten Web semantisch aufzubereiten. Textanalyseverfahren können aufgrund der extrem aufwändigen Konfiguration und Anpassung im Moment nur sehr eingeschränkt genutzt werden. Durch die geplante Flexibilisierung und automatische Regelgenerierung mittels LOD-Hintergrundwissen erlangen Unternehmen mit den entwickelten Methoden Zugang zur derzeit unter der Oberfläche versteckten Informationsfülle aus Textdokumenten, XML-Dateien und Datenbanken. Im Mittelpunkt stand dabei durch den Verbundpartner AI4BD, Methoden der automatischen Sprachverarbeitung (insb. statistische Ansätze) mit bereits vorhandenen Strukturinformationen sowie Linked Data zu kombiniert, um eine genauere Inhaltsbestimmung von Texten zu ermöglichen. Die Technische Universität Chemnitz trug in diesem ASP zentral zur Weiterentwicklung der LEDS Technologieplattform bei, indem mittels dieser Dienste aus Texten und andere Quellen (insbesondere XML und Datenbanken) extrahierte Informationen für alle anderen Komponenten der Technologieplattform bereitgestellt werden konnten. Damit wurde die Plattform wesentlich flexibilisiert und mit zusätzlichen Inhalten angereichert.

ASP-D „Skalierbare Suche in Linked Data“ widmete sich der Realisierung eines neuen Suchparadigmas, bei dem Nutzer unter Einbeziehung von Hintergrund- und Kontextwissen. Hierzu wurden neue Suchschnittstellen auf Linked Data Wissensbasen erforscht, welche auch Informationen aus sozialen Netzwerken als Linked Data aufbereiten und berücksichtigen. Hauptziel war es., den Nutzern von E-Commerce-Portalen eine neue themenbasierte, kontextualisierte und personalisierte Art der Suche nach Produkten und Angeboten zu ermöglichen. In dem Arbeitspaket wurde eine neue innovative Such-Technik entwickelt, die E-Commerce-Anwendungen mit neuen adaptiven und personalisierten, semantischen Suchfunktionen ausstattet. Methoden des verteilten, skalierbaren RDF-Datenmanagements wurden kombiniert mit Hintergrundwissen, Reasoning- und Question-Answering-Ansätzen und ermöglichten damit ein gänzlich neues Suchparadigma ("Social-aware search"). Die in diesem ASP entwickelte Suchtechnologie wurde als eine zentrale Komponente der Technologieplattform insbesondere in ASP-E und ASP-F genutzt.

2. VORAUSSETZUNGEN, UNTER DENEN DAS VORHABEN DURCHGEFÜHRT WURDE

Das Vorhaben Linked Enterprise Data Services setzt auf dem Linked Data Paradigma auf. Linked Data ist die Idee, verfügbare Daten semantisch über eindeutige Identifikatoren miteinander zu vernetzen und über eine standardisierte generische Schnittstelle verarbeitbar zu machen. In der Vergangenheit wurde dies vor allem im Open Data Umfeld für freie, quelloffene Daten eingesetzt. Doch können von dieser Technologie auch Unternehmen profitieren, welche vor der Aufgabe stehen, eine Vielzahl an unternehmensinternen Daten aus heterogenen Datenquellen verschiedener Systeme miteinander in Einklang zu bringen, mit Informationen aus anderen Quellen zu kombinieren und daraus relevantes Wissen abzuleiten. Im Kontext der digitalen Transformation verschiedener Anwendungsbereiche steht dies im direkten Einklang mit der Digital Single Market Strategie der Europäischen Union und den Big Data Trends im Bereich Industrie 4.0, sowie der Agenda 2020 im Bereich E-Government.

Unter diesem Hintergrund setzt das Verbundprojekt LE DS direkt auf W3C Standardisierungen und weiteren Entwicklungen im Bereich des Semantic Webs auf und führt diese fort. Zur Informationsverarbeitung werden dafür zur Verfügung stehende Programm Bibliotheken eingesetzt, welche etabliert sind und unter einer offenen Softwarelizenz stehen.

Gleichzeitig ist der Umgang mit Linked Data eine Grundvoraussetzung für den Erfolg des Wachstumskernprojekts. Einerseits sollten offene Datenquellen in ausreichendem Maße bereits zur Verfügung stehen, deren vorhandenes Informationspotential in einem industriellen Kontext weiter erschlossen und genutzt werden kann. Andererseits müssen auch Industrievertreter in die Lage versetzt werden, Linked Data als potentiell gewinnbringende Zukunftstechnologie zu sehen und in existierende Abläufe zu integrieren. Die LE DS-Technologieplattform kann Institutionen daraufhin in die Lage versetzen, vorhandene Datenbestände semantisch zu erschließen und als Linked Data zukünftig zu verarbeiten.

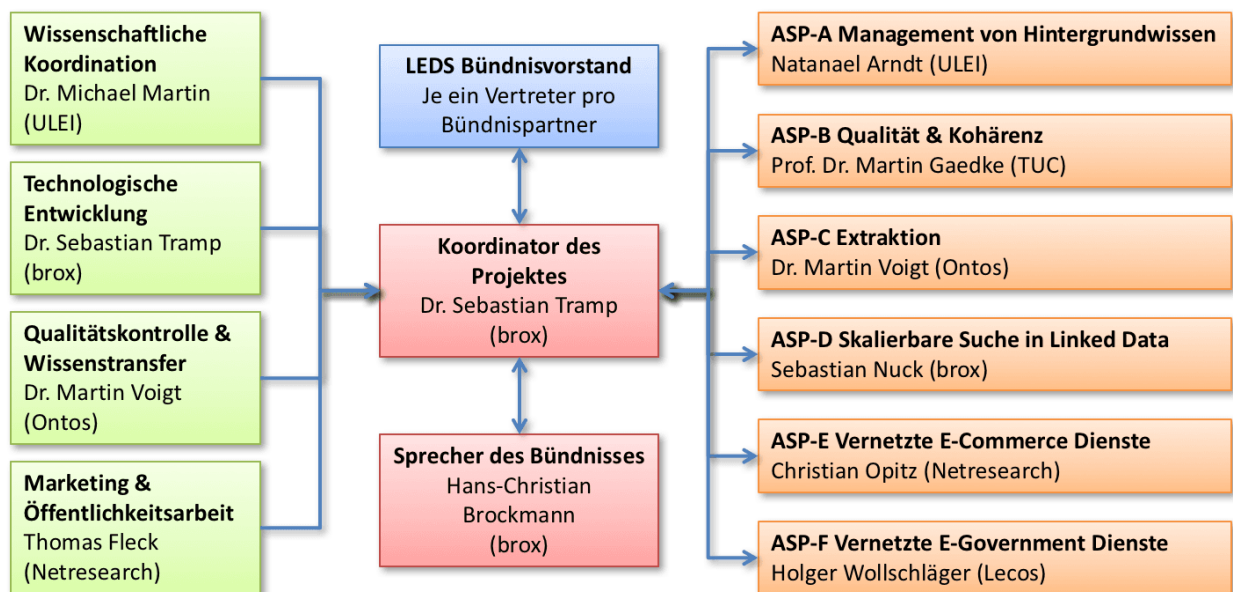
Für dieses Vorhaben wurden die Kernkompetenzen verschiedener Projektpartner benötigt. Dies umfasste im LE DS Bündnis mit der TU Chemnitz und der Universität Leipzig die Forschungskompetenz von zwei universitären Einrichtungen. Dabei konzentrierte sich die Tätigkeit der TU Chemnitz auf den Bereich der Datenqualitätsanalyse, der Wissensextraktion und der Entwicklung neuer, adaptiver, menschenbezogener User Interfaces mit Methoden des Web Engineerings und komponentenorientierten Designs, während sich das Team der Universität Leipzig um semantische Datenmanagementaspekte kümmerte. Die wesentliche Voraussetzung zur Bereitstellung der Forschungskompetenz lag dabei im Personalbereich mit der Anstellung von erfahrenem wissenschaftlichem Personal mit entsprechender Qualifikation in den benötigten Kernbereichen. Diese erforderte praktische Berufserfahrung im Umgang mit industriellen Daten und erweiterte Expertise im Management von Linked Data. Weiterhin wurden mit der brox IT Solutions GmbH Niederlassung Leipzig, der Firma AI4BD Deutschland (früher Ontos) GmbH und der Firma Netresearch DTT GmbH drei industrielle Partner eingebunden sowie mit der Lecos GmbH eine städtische Organisation, bei denen alle entwickelten Verfahren industrienah getestet werden konnten.

Als Folge daraus, war das dargelegte komplexe Teilvorhaben nicht ohne Förderung des LE DS-Gesamtprojekts umsetzbar, da die geplante Forschung und Entwicklung, aufgrund der notwendigen Personal- und Sachaufwendungen, durch andere Stellen nur unzureichend finanziert wäre. Im Kontext von aktuellen EU-Ausschreibungen wurde kein anderes geeignetes Programm gefunden, welches sowohl die beschriebenen Themenschwerpunkte und Zielsetzungen hinreichend abgedeckt als auch die in diesem Bündnis vorhandene räumlich nahe Kooperation mit den Bündnispartnern ermöglicht hätte.

3. PLANUNG UND ABLAUF DES VORHABENS

Basierend auf den bereits erworbenen Kompetenzen und bisherigen wirtschaftlichen Erfolgen der Bündnispartner sah sich das Bündnis in der sehr guten Ausgangslage, die in Abschnitt 1 beschriebene Entwicklung voranzutreiben sowie einen innovativen Wachstumskern in der Region zu initiieren.

Zur Gewährleistung der Kohärenz und Schlagkräftigkeit des Bündnisses wurde das Arbeitsprogramm (Verbundprojekt) in die sechs bereits beschriebenen Arbeitsschwerpunkte ASP-A bis ASP-F untergliedert, in denen die TU Chemnitz im ASP-B die Rolle des Teilprojektleiters einnahm (siehe Abbildung) und in den Arbeitsschwerpunkten ASP-C und ASP-D weitere wesentliche Beiträge lieferte.



Innerhalb eines Zeitraumes von 3 Jahren wurde die Realisierung einer Technologieplattform vereinbart, in deren Anschluss die Technologieplattform in Produkte und Dienstleistungen der Bündnispartner prototypisch integriert wurde. Der Fokus der Arbeiten in diesem Zeitraum war dabei:

- Jahr 1: Erste Iteration der Entwicklung und Integration der Basistechnologien
- Jahr 2: Zweite Iteration der Basistechnologieentwicklung, Beginn der Anwendungsfalladaption
- Jahr 3: Finalisierung der Technologieplattform-Integration, Anpassung und Evaluierung in Anwendungsfällen
- Jahr 4-5 (nach Ende der Förderung): Produktisierung der Technologieplattform durch die industriellen Bündnispartner, Evolution des gemeinsamen Bündnisses, welches Weiterentwicklung, Wartung und Dienstleistungen rund um die LEADS-Technologieplattform anbietet.

Das Vorgehen umfasste dabei für die TU Chemnitz die nachstehenden Arbeitspakete:

| AP | Titel | Zeitraum |
|---------|------------------------------------------------------------------------------------------|-----------|
| 1.4.B.1 | Spezifikation (Anforderungs- und Funktionsdefinition) der Linked Data Quality Komponente | M1 - M9 |
| 1.4.B.2 | Schnittstellen und Architekturprototyp der Linked Data Quality Komponente | M7 - M15 |
| 1.4.B.3 | Entwicklung erster Demonstrator der Linked Data Quality Komponente | M13 - M21 |
| 1.4.B.4 | Test und Optimierung der Linked Data Quality Komponente | M19 - M27 |
| 1.4.B.5 | Erweiterung der Linked Data Quality Komponente | M25 - M40 |
| 1.4.B.6 | Abschlusstests, Integration und Dokumentation der Linked Data Quality Komponente | M28 - M41 |

| AP | Titel | Zeitraum |
|-----------|---------------------------------------------------------------------------------------------------------------------|-----------------|
| 1.4.C.1 | Spezifikation (Anforderungs- und Funktionsdefinition) des Extraktionsmechanismus für semistrukturierte Datenquellen | M1 - M9 |
| 1.4.C.2 | Schnittstellen und Architekturprototyp des Extraktionsmechanismus für semistrukturierte Datenquellen | M7 - M15 |
| 1.4.C.3 | Entwicklung erster Demonstrator des Extraktionsmechanismus für semistrukturierte Datenquellen | M13 - M21 |
| 1.4.C.4 | Test und Optimierung des Extraktionsmechanismus für semistrukturierte Datenquellen | M19 - M27 |
| 1.4.C.5 | Erweiterung des Extraktionsmechanismus für semistrukturierte Datenquellen | M25 - M38 |
| 1.4.C.6 | Abschlusstests, Integration und Dokumentation des Extraktionsmechanismus für semistrukturierte Datenquellen | M28 - M40 |
| 1.4.D.1 | Spezifikation (Anforderungs- und Funktionsdefinition) der Social-aware Suche in Linked Data Quellen | M1 - M9 |
| 1.4.D.2 | Schnittstellen und Architekturprototyp der Social-aware Suche in Linked Data Quellen | M7-M15 |
| 1.4.D.3 | Entwicklung erster Demonstrator der Social-aware Suche in Linked Data Quellen | M13 - M21 |
| 1.4.D.4 | Test und Optimierung des Social-aware Suche in Linked Data Quellen | M19 - M27 |
| 1.4.D.5 | Erweiterung der Social-aware Suche in Linked Data Quellen | M25 - M33 |
| 1.4.D.6 | Abschlusstests, Integration und Dokumentation der Social-aware Suche in Linked-Data-Quellen | M28 - M41 |

4. WISSENSCHAFTLICHEM UND TECHNISCHEM STAND, AN DEN ANGEKNÜPFT WURDE, INSBESONDERE

Linked Data Grundforschung. Linked (Open) Data ist eine auf Web-Standards basierende Methode zur einheitlichen Speicherung, Publikation und Verknüpfung von Daten unterschiedlicher Art und Herkunft im RDF-Format. Seit der Prägung des Begriffs im Jahre 2006 wurden bereits hunderte von Datenquellen bzw. Wissensbasen unterschiedlicher Domänen als Linked Open Data (LOD) veröffentlicht und miteinander vernetzt. Auch viele Unternehmen (u.a. Google [Goel et al., 2009], Yahoo! [Mika, 2009], Oracle [Chong et al., 2005, Sundara et al., 2010]), Organisationen (u.a. BBC [Kobilarov et al., 2009], New York Times) und wissenschaftliche Projekte haben begonnen Linked Data zu veröffentlichen und zu nutzen. Die standardisierte und generische Repräsentation der Informationen ermöglicht eine einfache Zugänglichkeit und Nutzung heterogener Informationen in vielzähligen Anwendungen. Weiterhin verspricht man sich durch eine zunehmende Verlinkung der Datenquellen signifikante Vorteile in Bezug auf deren semantische Integration, die insbesondere für Webdaten mit traditionellen Ansätzen (Data Warehouses, Mediator/Wrapper-Architekturen) nur mit hohem Realisierungsaufwand und geringer Skalierbarkeit realisierbar sind [Rahm et al., 2005, Leser and Naumann, 2007]. Linked Data verspricht effektive Datenintegrationsmöglichkeiten u.a. durch die generische Repräsentation und Verwaltung heterogener Daten, die semantische Beschreibung der Daten mittels Ontologien sowie die explizite semantische Verlinkung von Objekten verschiedener Quellen durch besonders sameAs-Äquivalenzbeziehungen. Solche Links ermöglichen die Fusionierung von sich entsprechenden Informationen, die zur Anreicherung und Verknüpfung mit weiteren Informationen nutzbar ist. Somit könnten etwa zum Autor einer Publikation weitere Informationen z.B. zu seinem Institut, seinen Kollegen etc. gewonnen werden. Weiter können über Produktverlinkungen Preise unterschiedlicher Anbieter sowie Produktbeschreibungen und Nutzerbewertungen aus verschiedenen Quellen aggregiert und ausgewertet werden.

Bisherige Arbeiten im Bereich Linked Data fokussierten zunächst auf die Basisaufgaben der Generierung bzw. Extraktion von RDF-Daten sowie die effiziente Speicherung und Nutzung einzelner Datenquellen.

(a) Extraktion. In vielen Fällen werden RDF-Daten nicht manuell erstellt, sondern aus bestehenden Quellen extrahiert. Extraktions-Ansätze werden in der Regel entsprechend des Strukturierungsgrades der zugrundeliegenden Datenquellen klassifiziert: unstrukturierte, semi-strukturierte und strukturierte Quellen. Die Extraktion von Fakten aus dem Klartext umfassenden unstrukturierten Quellen erfordern Methoden des Natural Language Processing. Semi-strukturierte Quellen, z.B. Wiki-Markup, können mit anwendungsspezifischen Extraktionsmethoden, wie im DBpedia-Extraktionsframework [Lehmann et al., 2009], erschlossen werden. DBpedia [Lehmann et al., 2014] nutzt eine Reihe von semi-strukturierten Elementen in Wikipedia (wie z.B. Infoboxen, Kategorien und Links zwischen verschiedenen Spracheditionen) zur Extraktion und Repräsentation von Faktenwissen in RDF. Durch die Abdeckung einer Vielzahl von Informationsdomänen, die Verfügbarkeit in verschiedenen Sprachen und den Gemeinschafts-Konsens den Wikipedia/DBpedia widerspiegelt, hat sich DBpedia zu einem zentralen Knotenpunkt im Linked Data Web entwickelt. Dieser umfasst im Moment mehr als eine Milliarde Fakten. Eine Vielzahl von Forschungsansätzen und kommerziellen Anwendungen (z.B. Zemanta.com, OpenCalais.com) nutzen DBpedia inzwischen als Basistechnologie. Strukturierte Quellen umfassen insbesondere relationale Datenbanken, für welche bereits mehrere Ansätze zum RDF-Mapping existieren (insbesondere Triplify [Auer et al., 2009], D2RQ [Bizer and Cyganiak, 2006], sparqlmap [Unbehauen et al., 2013] und morph [Priyatna et al., 2014]). Die W3C RDB2RDF Arbeitsgruppe [Wang and Strong, 1996] hat mit R2RML einen Standard zur Umwandlung von relationalen Daten in RDF erarbeitet.

(b) Speicherung / Abfragen. Das starke Wachstum von Linked-Data-Quellen lieferte wichtige Impulse zur Entwicklung von Datenbank-Techniken für die native Verwaltung von RDF-Daten. Wenngleich relationale Datenbanken bei vergleichbaren Datenstrukturen noch eine wesentlich höhere Leistung als native RDF-Datenbanken aufweisen, konnte auch deren Skalierbarkeit durch die Entwicklung effizienter Speicher- und Clustering-Lösungen deutlich verbessert werden. Native RDF-Datenbanken eignen sich vor allem in Situationen mit sehr heterogenen Daten, die nicht leicht in einem relationalen Schema erfasst werden können. [Boncz et al., 2014] beschreibt hier den nach wie vor bestehenden Unterschied in der Performanz zwischen RDF- und relationalen Datenbanken.

(c) Klassifizierung / Anreicherung. Aufgrund der noch oft schwach ausgeprägten oder fehlenden Strukturinformationen bzw. Ontologiebezüge in LOD-Quellen ist deren Anreicherung um solche Strukturen für viele Anwendungen wesentlich, insbesondere zur Datenintegration und für fortgeschrittene Ansätze zur Link Discovery. Da sowohl OWL als auch RDF auf Beschreibungslogiken beruhen, kommen prinzipiell korrekt und vollständig eingestufte Reasoner-Implementierungen (z.B. ElVira, Pellet, FACT++, Hermit, KAON und Racer zur Wissensanreicherung in Frage. Allerdings machen deren Leistungsanforderungen sie für die Anwendung auf Wissensbasen der Größenordnung der LOD-Datenquellen ohne Anpassungen und Preprocessing unbrauchbar. Deshalb wurden mehrere unvollständige Inferenz-Methoden u.a. innerhalb des LarkC-Projekts [Fensel et al., 2008] entwickelt. Anreicherungs-Methoden basieren unter anderem auf induktiver logischer Programmierung und formaler Konzept-Analyse. Vielversprechende Ansätze in diesem Bereich sind DL-Learner [Lehmann, 2009], ORE [Lehmann and Bühmann, 2010] und OntoComp [Sertkaya, 2009].

Link Discovery und Object-Matching. Die semantische Verknüpfung von LOD-Wissensbasen ist für wesentliche Nutzungsformen von zentraler Bedeutung. Besonders wichtig ist dies für Datenintegrationsaufgaben, die Ontologie-übergreifende Anfragebearbeitung oder die Anreicherung von Fakten mit direkt oder indirekt verknüpften Informationen anderer Quellen. Der am häufigsten verwendete Link-Typ ist hierzu owl:sameAs, um auszudrücken, dass sich zwei URI-Referenzen auf dieselbe Entität beziehen [Halpin et al., 2010]. Da, wie erwähnt, der derzeitige Grad der Verlinkung von LOD-Wissensbasen noch äußerst gering ist, besteht ein großer Bedarf an leistungsfähigen Verfahren zur Link Discovery bzw. insbesondere zur Entdeckung von sameAs-Links. Hierzu gibt es erst vergleichsweise wenige und meist recht einfache oder auf einzelne Domänen beschränkte Ansätze [Cudré-Mauroux et al., 2009, Nikolov et al., 2009, Volz et al., 2009a, Hogan et al., 2010, Sleeman and Finin, 2010, Papadakis et al., 2011]. In jüngster Zeit wurde jedoch begonnen generische Frameworks wie SILK [Volz et al., 2009b] und LIMES [Ngonga Ngomo and Auer, 2011] zu entwickeln.

Relativ intensiv untersucht wurde dagegen bereits die Problematik des Ontology Alignment [Rahm and Bernstein, 2001, Leser and Naumann, 2007, Bellahsene et al., 2011], um Korrespondenzen zwischen sich entsprechenden Konzepten verschiedener Ontologien zu ermitteln. Die Verfahren lassen sich daher prinzipiell auch zur Erkennung von Ontologie-Links zwischen LOD-Quellen nutzen. Viele der derzeitigen Ansätze wurden im Rahmen der seit 2005 bestehenden Benchmark-Initiative OAEI (Ontology Alignment Evaluation Initiative) für unterschiedliche Testfälle vergleichend evaluiert [Euzenat et al., 2009]. Trotz erzielter Fortschritte besteht insbesondere für größere Ontologien (z.B. Web Directories) noch Verbesserungsbedarf. Diesem kann eventuell durch verstärkte Nutzung von Ontologie-Instanzen begegnet werden [Rahm, 2011]. Seit 2009 gibt es im Rahmen des OAEI-Wettbewerbs auch dedizierte Tests zum Instanz-Matching (Object-Matching bzw. Link Discovery), welche jedoch nur von wenigen Systemen und mit vielfach unzureichenden Ergebnissen (F-Measure von unter 50%) bearbeitet wurden [Euzenat et al., 2009].

Der relativ junge Entwicklungsstand von Link Discovery wird deutlich, wenn man ihn mit dem Stand der Wissenschaft zur bereits seit längerem untersuchten Problemstellung der Duplikaterkennung bzw. des Object-Matchings (auch Objektidentifizierung, Record Linkage oder Entity Resolution) für strukturierte Daten vergleicht. Der Wissensstand in diesem Gebiet wird z.B. in den Übersichtsartikeln [Naumann and Herschel, 2010, Köpcke and Rahm, 2010, Batini and Scannapieco, 2006] dargelegt. Methoden zum Object Matching und auch zur Link Discovery müssen zwei wesentliche Ziele in Einklang bringen: (1) Die Methoden sollen **effizient** sein (geringe Laufzeit, Skalierbarkeit auf große Datenmengen). (2) Andererseits sollen sie aber auch **effektiv** sein, um zuverlässig und präzise korrespondierende Objekte zu erkennen.

Zur Bestimmung aller korrespondierenden Objekte ist prinzipiell jedes Objekt mit jedem anderen Objekt zu vergleichen. Die quadratische Komplexität dieser Vorgehensweise ist für größere Datenmengen jedoch zu ineffizient. Sogenannte Blocking-Strategien erlauben die Anzahl der notwendigen Vergleiche zu reduzieren (siehe z.B. [Baxter et al., 2003, Bilenko et al., 2006, Draisbach and Naumann, 2009] für einen Überblick). Statt alle Objekte miteinander zu vergleichen, werden bei diesen Strategien die Objekte aufgrund bestimmter Schlüsselwerte in (eventuell überlappende) Partitionen eingeteilt und Objekte jeweils nur innerhalb einer Partition verglichen.

Eine orthogonal nutzbare Optimierung besteht in der parallelen Berechnung von Korrespondenzen bzw. Links auf verteilten Infrastrukturen. Hierzu bieten sich die zunehmend verfügbaren Cloud-Infrastrukturen an, sowie Implementierungen auf Basis von Programmiermodellen wie MapReduce. Bisher gibt es erst sehr wenige Untersuchungen zur effizienten Umsetzung von Object Matching-Verfahren [Vernica et al., 2010, Wang et al., 2010] bzw. Link Discovery [Isele et al., 2010] auf Basis von MapReduce.

Die meisten Object Matching- und Link Discovery-Ansätze basieren auf der Feststellung von Ähnlichkeiten zwischen Attribut- bzw. Property-Werten. Für die Ähnlichkeitsvergleiche werden dabei typischerweise zeichenkettenbasierte Abstandsmaße [Cohen et al., 2003, Elmagarmid et al., 2007] oder gelegentlich auch domänenspezifische Metriken [Chaudhuri et al., 2009] herangezogen. Im Object Matching-Bereich wurde ferner erkannt, dass es zur Verbesserung der Effektivität sinnvoll ist, neben den Attributähnlichkeiten zusätzliches Kontextwissen in Form von semantischen Beziehungen zu berücksichtigen [Chen et al., 2009, Kang et al., 2008, Bhattacharya and Getoor, 2007, Bhattacharya and Getoor, 2006, Chen et al., 2005, Culotta and McCallum, 2005, Dong et al., 2005, Singla and Domingos, 2004]. Somit können etwa für mehrere Personen mit demselben Namen anhand von Kontextinformationen wie die Zugehörigkeit zu einer Organisation oder die Information über Koautoren nützlich sein, um tatsächliche Entsprechungen besser zu identifizieren. Die in LOD-Quellen vorliegenden Beziehungen eignen sich ideal für die Realisierung solcher kontextbasierten oder semantischen Link Discovery-Verfahren.

Eine Erkenntnis aus bisherigen Object Matching-Untersuchungen [Doan et al., 2003] ist, dass einzelne Match-Verfahren meist keine ausreichende Effektivität erreichen, sondern dass hierzu die Kombination unterschiedlicher, sich ergänzender Verfahren notwendig wird. Eine optimale Auswahl, Kombination und Konfiguration von Verfahren zu finden, gestaltet sich jedoch aufgrund der vielfältigen Möglichkeiten selbst für Experten oftmals schwierig und zeitaufwändig. Es werden daher zunehmend Techniken des maschinellen Lernens vorgeschlagen, um die Auswahl, Kombination und Konfiguration von Verfahren zu automatisieren. Hierbei stellt sich die Schwierigkeit wie geeignete Trainingsdaten auszuwählen sind, um den manuellen Aufwand möglichst gering zu halten [Bilenko and Mooney, 2003, Köpcke and Rahm, 2008]. Besonders vielversprechend sind deshalb Ansätze des aktiven Lernens [Sarawagi and Bhamidipaty, 2002, Sarawagi et al., 2002, Arasu et al., 2010, Ngonga Ngomo and Lyko, 2012, Isele and Bizer, 2013], welche automatisch geeignete Objektpaare zur interaktiven Match-Bewertung durch den Nutzer bestimmen. Damit sollen mit minimalem Nutzeraufwand effektive Match-Strategien erlernt werden.

Datenqualität und -kohärenz. Zur Sicherung der Qualität von Daten in klassischen Datenspeichern (wie relationalen Datenbanken) existieren sowohl etablierte Qualitätskriterien [Wang and Strong, 1996] als auch entsprechende Technologien zur Überprüfung und Auflösung von Inkonsistenzen oder Leistungsproblemen [Bergdahl et al., 2007]. Kriterien und Indikatoren für die Beurteilung der Datenqualität von Linked Data Quellen [Guéret et al., 2012] sowie der Implementierung in konkrete Technologien [Fürber and Hepp, 2011] sind eingeschränkt vorhanden. Erste Ansätze um Daten aus sozialen Netzwerken für die Überprüfung der Qualität von bestimmten Wissensbasen heranzuziehen, wurden u.a. in [Lu et al., 2010, Anagnostopoulos et al., 2008, Goyal et al., 2010] betrachtet. Es existiert momentan jedoch keine ganzheitliche Lösung, die semantisch aufbereitete Informationen aus heterogenen sozialen Netzwerken unter Beachtung des Datenschutzes zur Sicherung der Datenqualität in Linked Data Quellen systematisch verwertet.

Semistrukturierte Datenbestände weisen keine explizite Definition und Typisierung aller Datenobjekte auf. Der strukturierte Anteil innerhalb von semistrukturierten Datenbeständen ist durch den jeweiligen Datenverwendungszweck bestimmt. Auf Basis dieser Eigenschaften bieten semistrukturierte Datenbestände eine hohe Erweiterbarkeit sowie Flexibilität beim Datenaustausch bzw. beim Bezug von Daten aus heterogenen Datenquellen. Dies ist durch die weite Verbreitung von XML als Auszeichnungssprache für semistrukturierte Daten belegt [Kim et al., 2009]. Ebenso sind semistrukturierte Daten im Bereich sozialer Netzwerke als Repräsentationsform für soziale Inhalte vertreten und dienen somit als wesentliche Informationsquellen [IBM Studie, Sep 2011, <http://www-03.ibm.com/press/de/de/pressrelease/35352.wss>].

Zur Extraktion von Daten aus semistrukturierten Datenbeständen sind sowohl Ansätze für strukturierte als auch unstrukturierte Daten relevant. Die Anwendbarkeit dieser existierenden Ansätze auf die Domäne der semistrukturierten Daten muss analysiert und geeignete Ansätze ggf. adaptiert werden. Unstrukturierten Daten können gezielt kontextspezifische Strukturen hinzugefügt werden, um deren Interpretation und Einordnung zu erleichtern. Solche Strukturen sind u.a. Schlüsselwörter [Zangerle et al., 2011], welche bestimmten Objekten (Text, Foto, Video etc.) zugeordnet wurden, sowie Kennzeichnungen in Form von Zeitstempeln, Ortsangaben oder Bewertungen.

Die föderierte Abfrage von Informationen und deren Aufbereitung zu einer zentralen geordneten Ergebnismenge ist bisher nur rudimentär erforscht. Ebenso die kalkuliert ungenaue, aber performante Ermittlung von Informationszusammenhängen [Bonatti et al., 2011]. Es existieren bisher nur wenige oder unzureichende Werkzeuge zur Umsetzung. Der Bedarf nach Skalierbarkeit stellte eine weitere Herausforderung dar. Diese ist essentiell für die Umsetzung wirtschaftlicher Szenarien, ist jedoch kaum in Arbeiten adressiert und nur für spezielle Anwendungsszenarien umgesetzt [Thompson and Personick, 2009].

Unsere Patentrecherche mit DEPATISnet ergab weiter zwei relevante Patente: CN000001766871A und DE000010331817A1. Diese stehen jedoch dem angestrebten Verwertungsplan im Rahmen von Lehrveranstaltungen und Publikationen nicht entgegen.

5. ZUSAMMENARBEIT MIT ANDEREN STELLEN

Die Arbeit im LEDS Verbundprojekt konzentrierte sich auf einen kooperativen Austausch zwischen den Bündnispartnern brox IT-Solutions GmbH (Leipzig), AI4BD Deutschland GmbH (Leipzig), Netresearch DTT GmbH (Leipzig), Lecos GmbH (Leipzig), Universität Leipzig und der Technischen Universität Chemnitz. Im Zuge dessen wurde auch eine Zusammenarbeit mit dem Behindertenverband Leipzig zur prototypischen Anwendung der LEDS Technologieplattform initiiert.

Hinzu kamen im wissenschaftlichen Umfeld Kontakte zu anderen Forschungseinrichtungen und Informationsanbietern wie der Technischen Informationsbibliothek (TIB) Hannover. Diese führten während der Projektlaufzeit aber zu keinen gemeinsamen Publikations- oder Ergebnisverwertungsaktivitäten.

Weiterhin wurden mehrere öffentlichkeitswirksame Veranstaltungen durchgeführt. Hierbei sind vor allem die Beteiligung an der SEMANTiCS Konferenz im Jahr 2016 in Leipzig als auch die Organisation der 47. Jahrestagung der Gesellschaft für Informatik im Jahr 2017 zu nennen, bei der mit der Gesellschaft für Informatik e.V. unmittelbar zusammengearbeitet wurde. Daneben beteiligte sich das LEDS Projekt mehrfach am Leipziger Semantic Web Tag (LSWT) und war auf anderen Konferenzen wie der International Conference on Web Engineering (ICWE) präsent, welche durch die International Society for Web Engineering (ISWE) unterstützt wurde.

II. EINGEHENDE DARSTELLUNG

1. DER VERWENDUNG DER ZUWENDUNG UND DES ERZIELTEN ERGEBNISSES IM EINZELNEN, MIT GEGENÜBERSTELLUNG DER VORGEgebenEN ZIELE

In diesem Abschnitt stellen wir zunächst den Hintergrund des Projekts im Detail dar und gehen anschließend auf die konkret erreichten Meilensteine und Ergebnisse ein.

Für viele Geschäftsprozesse werden häufig umfangreiche Wissensbasen durchsucht, um Ergebnisse zu komplexen Anfragen effizient zu liefern. Die Qualität der Ergebnisse hängt hierbei, neben der Anfrage an sich, direkt von der Qualität der Daten in den verwendeten Wissensbasen ab. Kriterien wie Vollständigkeit, Korrektheit, Konsistenz, Genauigkeit und Aktualität bestimmen die Datenqualität. Um Kunden als Dienstleistung hochqualitative Resultate im Zuge der Beantwortung von Suchanfragen anbieten zu können, sollen semistrukturierte Daten aus heterogenen Wissensbasen systematisch verwertet werden. Für den Prozess der Optimierung von Suchergebnissen ist die Nutzbarmachung von Daten aus einem sozialen Kontext unter Berücksichtigung aktueller Datenschutzrichtlinien von besonderem Interesse. Soziale Netzwerke stellen aufgrund ihres großen Nutzerkreises sowie ihrer hohen Frequentierung eine Quelle für aktuelle und umfassende Daten dar, die durch Extraktion, Analyse und Aufbereitung sowohl zum Abgleich, Ergänzung und Korrektur als auch zur Optimierung und Reorganisation von semantischen Wissensbasen erschlossen werden kann.

Semistrukturierte Datenbestände weisen keine Spezifikation sämtlicher Datenobjekte auf, da ihr struktureller Anteil durch den Verwendungszweck definiert ist. Aus diesem charakteristischen Merkmal leitet sich eine hohe Erweiterbarkeit und Flexibilität ab. Semistrukturierte Daten werden daher als Repräsentationsform für soziale Inhalte im Bereich sozialer Netzwerke verwendet. Sie sind somit wesentliche Informationsquellen zu deren Extraktion geeignete Verfahren für strukturierte sowie unstrukturierte Daten evaluiert und adaptiert werden müssen. Methoden um Daten zugeordnete kontextspezifische Strukturen, wie bspw. sowohl Ortsangaben, Schlagwörter und Bewertungen als auch Gruppenreferenzen und -kommentare maschinell zu detektieren und gesondert zu verarbeiten, sind zu untersuchen. Neben der Extraktion an sich ist die semantische Aufbereitung der extrahierten Inhalte – in einer geeigneten Form wie RDF – ein zentraler Gegenstand der Forschung und Entwicklung. Somit können die in einer maschinenverwertbaren Form vorliegenden Daten durch nachfolgende Methoden effektiver prozessiert werden. Weiterführende Informationen sind den Sektionen Motivation und Zielsetzung in der Beschreibung von „Arbeitsschwerpunkt C: Informationsextraktion“ zu entnehmen.

Nach der Extraktion und Transformation von Daten aus semistrukturierten Wissensbasen, ist die Analyse und Optimierung der Datenqualität entscheidend für die nachfolgenden Verarbeitungsschritte. Zur Abschätzung der Qualität einer Wissensbasis ist diese auf Konsistenz d.h. auf Duplikatfreiheit und Korrektheit der Ableitungsbeziehungen zu prüfen. Aufgrund des Datenumfangs und des nachrichtengetriebenen Charakters sozialer Netzwerke können solch aufbereitete „soziale Daten“ 1) einer Konsistenzanalyse zuträglich sein, falls ein bestimmtes Verhältnis von unterstützenden bzw. widersprechenden Aussagen vorgefunden wird, und 2) die aktuellen Informationen als ein Maß zur Evaluierung der zeitlichen Relevanz fungieren. Weiterhin kann die Ermittlung der Genauigkeit einer Information durch eine Analyse aller Informationsbestandteile auf Basis sozialer Daten unterstützt werden. Eine umfassende Beschreibung dieses Themas ist in „Arbeitsschwerpunkt B: Qualität und Kohärenz“ vorhanden. Diese semantisch aufbereiteten und hinsichtlich Qualität optimierten Daten aus semistrukturierten Wissensbasen sollen genutzt werden, um u.a. automatisch die Relevanz dieser Daten zu bestimmen. Hierfür ist geplant ebenso die verarbeiteten kontextspezifischen Strukturen, welche den Daten zugeordnet waren, als Merkmale zur Relevanzbestimmung einzusetzen. Nähere Informationen sind in den Abschnitten zur Themenbasierten Suche in „Arbeitsschwerpunkt D: Skalierbare Suche in Linked Data“ verfügbar. Auf Basis der skizzierten Motivation leiten sich folgende Ziele für dieses Teilvorhaben „Informationsverwertung basierend auf qualitätsoptimierten semistrukturierten Datenbeständen“ ab:

- **Z-1.4.1 Erfassung und Aufbereitung von semistrukturierten Daten:** Evaluation existierender Ansätze, Adaption geeigneter Methoden und Entwicklung von Algorithmen zur Erfassung (Analyse, Extraktion) und Aufbereitung (Transformation, Gruppierung, Referenzierung) semistrukturierter Daten aus heterogenen Datenquellen und nachfolgender Analyse struktureller Datenmerkmale. Das Ziel Z-1.4.1 gilt als erreicht, wenn mindestens 80% der ersichtlich verwertbaren Bestandteile eines durchschnittlichen semistrukturierten Datensatzes (bspw. Top 10 der unter Alexa gelisteten deutsch/englischsprachigen Webseiten) auch maschinell erfasst, aufbereitet und verknüpft werden können. Das Teilvorhabenziel Z-1.4.1 ist "Arbeitsschwerpunkt C: Informationsextraktion" zugeordnet und geht mit folgenden Zielen dieses Arbeitsschwerpunktes einher:
 - Z-C4 Erfassung und Analyse von semistrukturierten Daten
 - Z-C5 Aufbereitung und Aggregation von semistrukturierten Daten
 - Z-C6 Verknüpfung mit existierenden Wissensbeständen
- **Z-1.4.2 Qualitätssicherung und Klassifikation von semantischen Daten:** Erforschung von Verfahren zur Überprüfung der 1) Datenqualität nach den Gesichtspunkten Relevanz, Konsistenz, Aktualität und Richtigkeit sowie der Korrektur von invaliden Datensätzen und 2) automatischen thematischen Einordnung bzw. Reorganisation von semantischen Daten. Das Ziel Z-1.4.2 gilt als erreicht, wenn semistrukturierte Daten zu mindestens 75% korrekt klassifiziert werden können sowie mindestens 90% der in aufbereiteten semistrukturierten Daten enthaltenen Fehler automatisch erkannt werden. Das Teilvorhabenziel Z-1.4.2 ist "Arbeitsschwerpunkt B: Qualität und Kohärenz" zugeordnet und geht mit folgenden Zielen dieses Arbeitsschwerpunktes einher:
 - Z-B1 Link-Generierung basierend auf Link-Eigenschaften
 - Z-B4 Sicherung der Datenqualität in semantischen Wissensbasen
 - Z-B5 Klassifikation zu integrierender Daten
- **Z-1.4.3 Themen- und Relevanzbasierte Anfragebeantwortung:** Forschung und Entwicklung von Verfahren zur Anfragebeantwortung und Relevanzabschätzung unter Berücksichtigung des sozialen Kontextes. Das Ziel Z-1.4.3 gilt als erreicht, wenn basierend auf qualitätsoptimierten Daten ein detektierter sozialer Kontext so verwertet werden kann, dass Ergebnisse für Anfragende relevanter bzw. themenbezogener sind vgl. zu keiner Detektion. Das Teilvorhabenziel Z-1.4.3 ist "Arbeitsschwerpunkt D: Skalierbare Suche in Linked Data" zugeordnet und geht mit folgenden Zielen dieses Arbeitsschwerpunktes einher:
 - Z-D1 Information Retrieval für lokale RDF-Informationen
 - Z-D3 Analyse und Matching von Suchverfahren
 - Z-D6 Heuristisches Reasoning auf RDF-Wissensrepräsentationen

Entsprechend dieser Zielstellung konnten mit der bereitgestellten Zuwendung zwei Projektmitarbeiter während der Projektlaufzeit in Vollzeit forschersich tätig sein. Dabei wurden die in der nachstehenden Tabelle dargestellten Meilensteine erreicht. Details zur Verwendung der Zuwendung und der Publikation der Ergebnisse sind in den Abschnitten 2 und 6 genauer dargestellt.

| Meilenstein | Erreichtes Ergebnis | Aufwand |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| MS-1.4.1 | Schnittstellen-, Technologie- und Architekturspezifikation der Linked Data Quality Komponente, des Extraktionsmechanismus für semistrukturierte Datenquellen und Social-aware Suche in Linked Data Quellen (M6) | 12 PM |
| | Evaluierung existierender Technologien zur Qualitätssicherung ist durchgeführt. | |

| Meilenstein | Erreichtes Ergebnis | Aufwand |
|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| | <p>Verfahren zur Sicherung der Datenqualität in Linked Data Quellen ist spezifiziert.</p> <p>Evaluationsresultate von Technologien für semistrukturierten Datenbestände sind bekannt.</p> <p>Schnittstellen und Methoden des Extraktionsmechanismus sind spezifiziert.</p> <p>Experimentierprozesse für themenbasierte Suche und für Suche unter Betrachtung von sozialen Zusammenhängen sind definiert.</p> | |
| MS-1.4.2 | Fertigstellung der Basis-Implementierung der Linked Data Quality Komponente, des Basis-Extraktionsmechanismus für semistrukturierte Datenquellen und der Basis-Komponente für Social-aware Suche in Linked Data Quellen (M18) | 24 PM |
| | <p>Basissatz von Qualitätssicherungs-Algorithmen als Demonstratorkomponente ist implementiert.</p> <p>Basisversionen von Algorithmen zur Datenaufbereitung, Transformation und Gruppierung sind erarbeitet.</p> <p>Implementierung der ersten Version des Extraktionsmechanismus für semistrukturierte Daten ist fertiggestellt.</p> <p>Mechanismus zur Suche unter Gewichtung der regionalen Eigenschaften eines Nutzers ist fertiggestellt.</p> | |
| MS-1.4.3 | Fertigstellung der finalen Implementierung der Linked Data Quality Komponente, des finalen Extraktionsmechanismus für semistrukturierte Datenquellen und der finalen Komponente für Social-aware Suche in Linked Data Quellen (M30) | 24 PM |

| Meilenstein | Erreichtes Ergebnis | Aufwand |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| | <p>Finaler Satz von Methoden zur Qualitätssicherung von Linked Data Quellen ist erarbeitet.</p> <p>Alle spezifizierten Algorithmen zur Datenaufbereitung, Transformation und Gruppierung sind implementiert.</p> <p>Finale Version des Extraktionsmechanismus für semistrukturierte Daten ist fertiggestellt.</p> <p>Mechanismus zur Suche unter Gewichtung der Freunde- und Freundesfreunde-Beziehung eines Nutzers ist fertiggestellt.</p> <p>Mechanismus zur Suche unter Gewichtung der erkennbaren Expertise (bspw. Umweltschutzexperte) innerhalb eines sozialen Netzwerks ist fertiggestellt.</p> | |
| MS-1.4.4 | <p>Fertigstellung der Stabilisierung der Linked Data Quality Komponente, des Extraktionsmechanismus für semistrukturierte Datenquellen und der Komponente für Social-aware Suche in Linked Data Quellen sowie Integration der Komponenten in die LEDES-Technologieplattform (M41)</p> | 22 PM |
| | <p>Resultate von Funktions-, Integrations- und Leistungstests zur Linked Data Quality Komponente, zum Extraktionsmechanismus für semistrukturierte Daten und zur Social-aware Suche sind bekannt.</p> <p>Implementierung der Linked Data Quality Komponente, des Extraktionsmechanismus und der Social-aware Suche ist stabilisiert und optimiert.</p> <p>Integration der Linked Data Quality Komponente, des Extraktionsmechanismus für semistrukturierte Daten und der Social-aware Suche in LEDES-Technologieplattform ist durchgeführt.</p> <p>Dokumentation zur Linked Data Quality Komponente, zum Extraktionsmechanismus und zur Social-aware Suche ist vervollständigt.</p> | |

2. DER WICHTIGSTEN POSITIONEN DES ZAHLENMÄßIGEN NACHWEISES

Entsprechend des Gesamtfinanzierungsplans wurde das Teilprojekt TP4: Informationsverwertung basierend auf qualitätsoptimierten semistrukturierten Datenbeständen im Wachstumskern LEDES mit einer bewilligten Zuwendung von 412.707,00 EUR durchgeführt. Innerhalb des Projektzeitraums entstanden dabei Ausgaben in Höhe von 400.584,98 EUR. Die wichtigsten Positionen des zahlenmäßigen Nachweises waren dabei Personalkosten für die Tätigkeit wissenschaftlicher Mitarbeiter im Projekt in Höhe von 388.503,94 EUR sowie Reise- und Publikationskosten in Höhe von 8.713,09 EUR. Daneben entstanden sind geringfügige weitere unmittelbare Kosten zur Projektdurchführung in Höhe von 3.367,95 EUR. Der nicht ausgeschöpfte Projektmittelrahmen entsprechend des Kassenbestands zum 30.11.2018 in Höhe von 14.546,42 EUR wurde zurücküberwiesen. Eine detaillierte Darstellung der gesamten Kalkulation ist im Verwendungsnachweis zu finden.

3. DER NOTWENDIGKEIT UND ANGEMESSENHEIT DER GELEISTETEN ARBEIT

Die im Rahmen des Teilvorhabens zu entwickelnden Technologien bildeten wesentliche Bestandteile der Arbeitsschwerpunkte zu den Themengebieten Datenqualität, Informationsextraktion, und Suche. Ohne die technologischen Beiträge des Teilvorhabens wäre der Erfolg der Arbeitsschwerpunkte gefährdet gewesen und somit auch der Erfolg des Gesamtvorhabens. Zur Forschung und Realisierung der benötigten Technologiebausteine wurde wissenschaftlich fundiert vorgegangen. Hierzu wurden etablierte Methoden des Software- und Web Engineerings eingesetzt, insbesondere ein komponentenorientiertes Design, wodurch Applikationen entstanden, welche sich unkompliziert in die LEDS-Technologieplattform integrieren lassen, aber auch unabhängig davon eigenständig über wohl definierte Schnittstellen (REST APIs) nutzbar sind.

Im Arbeitsschwerpunkt ASP-B war dafür die Realisierung einer Datenqualitätssicherungskomponente wichtig. Diese war notwendig, um in einem Unternehmenskontext die syntaktische und semantische Korrektheit der zu verarbeitenden Daten sicherzustellen und umfasste sowohl bereits vorliegende, strukturierte Daten als auch Eingabedaten, welche aus anderen Datenquellen extrahiert und mit bestehenden Datensätzen kombiniert wurde. Hierzu wurde etablierte Datenqualitätsmetriken aus sorgfältig definierten Qualitätsdimensionen eingesetzt und auf interne und externe Linked Data Quellen angewendet. Weiterhin wurden etablierte Standards wie das Data Quality Vocabulary der W3C Working Group für Daten im World Wide Web dazu benutzt, um Messergebnisse interoperabel und maschinell weiter verarbeitbar in einem standardisierten Format bereitzustellen. Der aktuelle Stand der Technik wurde mit Methoden des Quality Engineerings erweitert, um Anforderungen an die Qualität von Datenquellen einheitlich semantisch zu spezifizieren und Datenqualitätsmessergebnisse zwischen verschiedenen Datenquellen zu vergleichen. Alle Ergebnisse wurden auf internationalen Konferenzen innerhalb der zurückliegenden drei Jahre ge-reviewed und als fachlich angemessen und relevant eingestuft und publiziert.

Die Extraktionskomponente aus semistrukturierten Datenquellen im Arbeitsschwerpunkt ASP-C war notwendig, da im industriellen Umfeld vielfach Datenbestände in tabellarischen Dateiformaten wie csv oder xls vorliegen oder auch durch Anwendungen als JSON bereitgestellt werden. Komponenten zur Extraktion von Linked Data aus diesen Dateien gab es zum Beantragungszeitpunkt des Projekts nur eingeschränkt, wodurch dieser Technologiebaustein für den Einsatz im Unternehmensdatenkontext essentiell notwendig war. Zur Realisierung wurden Methoden des Data Engineerings eingesetzt und durch Schema Mapping den vorhandenen Daten korrespondierende Linked Data Konzepte zugeordnet. Hierfür wurden etablierte Ontologien wie FOAF und schema.org eingesetzt. Alle Ergebnisse wurden auf internationalen Konferenzen innerhalb der zurückliegenden drei Jahre ge-reviewed und als fachlich angemessen und relevant eingestuft und publiziert.

Von industriellen Unternehmen im E-Commerce-Umfeld war darüber hinaus ein praktischer Anwendungsfall zum Einsatz von Linked Data Technologien von entscheidender Bedeutung. Besonderes Potential wurde in der Kombination von unternehmensinternen (Produkt-)Daten und Interaktionsdaten mit externen Datenbeständen gesehen. Im Arbeitsschwerpunkt ASP-D wurde hierfür ein Suchinterface auf einem Linked Data Bestand geschaffen, welches eine neue Art der adaptiven Suche einsetzt unter Zugriff auf externe, semi-strukturierte Daten aus sozialen Netzwerken. Als reales Anwendungsbeispiel wurden hierfür Informationen aus der Graph API der zu diesem Zeitpunkt größten sozialen Netzwerk-Plattform Facebook verwendet. Zum Zugriff auf diese Daten war innerhalb der Suchkomponente die aktive Anmeldung und Freigabe durch den jeweiligen Nutzer notwendig, wodurch Aspekte der Privatsphäre im Umgang mit personalisierten Daten beachtet und eingehalten wurden. Zur Authentifizierung und dem Zugriff auf externe Datenbestände wurden standardisierte Protokolle wie OAuth 2.0 benutzt. Für die Anwendung der kontextuellen Informationen auf das Sucherlebnis und die Reihenfolge der Suchergebnisse wurde ein gewichtiges Scoring Model definiert. Alle Ergebnisse wurden auf internationalen Konferenzen innerhalb der zurückliegenden drei Jahre ge-reviewed und als fachlich angemessen und relevant eingestuft und publiziert.

Das zuvor dargelegte komplexe Teilvorhaben "Informationsverwertung basierend auf qualitätsoptimierten semistrukturierten Datenbeständen" war nicht ohne Bewilligung des Gesamtprojektes umsetzbar, da die geplante Forschung und Entwicklung, aufgrund der notwendigen Personal- und Sachaufwendungen, nur unzureichend finanziert gewesen wäre. Die durch die Grundfinanzierung bereitgestellten Mittel sind der Erfüllung des Lehrauftrages zugewiesen und konnten daher nicht zur Bewältigung der Aufgaben im Teilvorhaben herangezogen werden. Folglich waren die Mittel der Grundfinanzierung – über Landesstellen – ungenügend. Ohne Förderung des Teilvorhabens wären die vorgestellten Themen, wenn überhaupt, nur mit großer Zeitverzögerung und deutlichen Qualitätseinbußen realisierbar gewesen.

4. DES VORAUSSICHTLICHEN NUTZENS, INSBESONDERE DER VERWERTBARKEIT DES ERGEBNISSES IM SINNE DES FORTGESCHRIEBENEN VERWERTUNGSPLANS,

Als Universität strebt die TU Chemnitz keine wirtschaftliche Verwertung der Forschungsergebnisse an.

Durch die automatische Extraktion und Aufbereitung von Informationen aus semistrukturierten Datenbeständen wird eine bisher noch unzureichend genutzte Datenquelle erschlossen und somit u.a. ein Zugang zur maschinellen Verwertung von Inhalten aus sozialen Netzwerken etabliert. Soziale Inhalte wie Mitteilungen, Bewertungen, Kritiken sowie Beziehungsgefüge und Gruppenzugehörigkeiten können aktiv zur Optimierung von Dienstleistungen, bspw. für die zugeschnittene Beantwortung von Suchanfragen, verwendet werden und folglich der Verbesserung von Effektivität und Kundenzufriedenheit zuträglich sein.

Die Qualität von existierenden semantischen Wissensbasen kann durch die systematische Nutzung von extrahierten und semantisch aufbereiteten Daten aus semistrukturierten Quellen erhöht werden, indem diese sowohl zur Prüfung der Konsistenz, Vollständigkeit, Korrektheit und Aktualität als auch zu ggf. durchzuführenden Korrekturmaßnahmen eingesetzt werden. Aufgrund des großen Nutzerkreises, der hohen Freqüentierung und des nachrichtengetriebenen Charakters sozialer Netzwerke sind deren aufbereitete Inhalte gut geeignet, um einen Abgleich der Daten durchzuführen. Von der qualitativen Sicherung und Optimierung der Daten in der Wissensbasis profitieren alle angekoppelten Dienste und Prozesse durch u.a. hochwertigere Suchergebnisse im Kontext von Konsistenz, Korrektheit, Vollständigkeit und Aktualität. Unternehmen, welche große Bestände von heterogenen Daten für Kunden oder eigene Absichten verwalten, profitieren von den Forschungs- und Entwicklungsergebnissen, indem sowohl semantische Zusammenhänge zwischen den verteilten Datensätzen automatisch geprüft, erkannt und ggf. Verbindungen hergestellt werden als auch Duplikate sowie fehlerhafte, unvollständige oder nicht mehr aktuelle Datensätze gefunden und optional bereinigt werden. Durch diese Technologien können nicht nur neu zu integrierende Unternehmensdaten korrekt gruppiert und zugeordnet werden, sondern auch existierende Datensätze so reorganisiert werden, dass Abfragen effizienter ablaufen. Unternehmen und Institutionen, welche diese technologische Basis nutzen, können somit auf umfassend geprüfte und optimierte Datenbestände zugreifen, was u.a. in reduzierten Aufwand für Wartung und Veredelung der Daten resultiert und damit einhergehend zu Einsparungen von Personalkosten führt. Dies ermöglicht ebenso Geschäftsmodelle für Datendienstleister, die eine Standardvariante oder veredelte Varianten von Daten anbieten könnten. Neben dieser wirtschaftlichen Bedeutung erzielen Kunden und Partner durch Anwendungen, welche die themen- und relevanzbasierte Suche integrieren, auf spezifische soziale Präferenzen zugeschnittene Resultate. Solche sozialen Präferenzen stellen bspw. Bewertungen von und Kritik an Informationen durch bestimmte Gruppen wie Fachexperten eines Themas, Bekanntenkreis des Fragestellers oder der Fragesteller selbst dar.

Die wissenschaftliche Anschlussfähigkeit ergibt sich durch die geplante breite und fortlaufende Diffusion der Forschungsergebnisse bereits während der Projektlaufzeit. Hierbei wurden die Forschungsergebnisse sowohl intern hinsichtlich ihrer wissenschaftlichen Tragfähigkeit und Praxistauglichkeit zu prüfen als auch öffentlich zur Diskussion gestellt und beurteilt. Um möglichst gezielt die kritische Öffentlichkeit anzusprechen - insbesondere Forscher aus anderen akademischen Einrichtungen und interessierte Vertreter anderer Unternehmen - wurden die Resultate in Form von Publikationen bei international anerkannten Fachkonferenzen eingereicht.

Neben der Präsentation von Forschungsergebnissen hielten geprüfte Inhalte ebenso Einzug in thematisch passende Lehrveranstaltungen wie die Vorlesung XML und Current Trends in Web Engineering, als auch in Seminare unserer Professur an der TU Chemnitz und einem Planspiel mit speziellem Fokus auf Linked Data im Jahr 2018 (siehe Verwertungsplan), die somit über die Projektlaufzeit hinaus nachhaltig wirken. Des Weiteren konnten die erworbenen Projekterfahrungen und -ergebnisse im Anschluss an das Projektvorhaben verwertet werden, indem Anträge auf Folgeprojekte für weiterführende, aufbauende Forschung und Entwicklung im Themenbereich bei dem BMBF, der DFB, dem ESF und dem EU H2020 generiert werden konnten.

Zuletzt arbeiten unsere Studierenden inzwischen kontinuierlich in weiterführenden Forschungsarbeiten an der Fortführung unserer Projektergebnisse im LEDS-Kontext. Beispiele sind ein Service zur Abfrage geeigneter Linked Data URIs für auszudrückende Konzepte in Semantic Web Anwendungen, ein Linked Data gestütztes User Interface zur Annotation von Forschungsdaten und eine SoLiD-gestützte Plattform zur Veröffentlichung von Forschungsdaten. Des Weiterhin arbeiten derzeit zwei wissenschaftliche Mitarbeiter über das LEDS-Projekt hinaus an einem PhD-Projekt mit dem Schwerpunkt auf der Nutzung von Linked Data.

5. DES WÄHREND DER DURCHFÜHRUNG DES VORHABENS DEM ZE (ZUWENDUNGSEMPFÄNGER) BEKANNT GEWORDENEN FORTSCHRITTS AUF DEM GEBIET DES VORHABENS BEI ANDEREN STELLEN,

Nach einer umfangreichen Vorarbeitsphase erfolgte die Beantragung des LEDS-Vorhabens erfolgreich im Jahr 2014. Zu diesem Zeitpunkt waren die Basistechnologien innerhalb der Semantic Web Community bereits grundlegend standardisiert und im Begriff der praktischen Anwendung. Die im LEDS-Projekt fokussierten Technologien waren zum Zeitpunkt der Beantragung in dem angestrebten Ausmaß noch nicht vorhanden. Im Verlauf des Projekts zeigten sich aber ähnliche Bestrebungen auch in Forschungsaktivitäten anderer Institutionen.

So entstand im Rahmen eines EU FP7 Projekts "Diachron" bereits ein systematischer Überblick über Datenqualitätsmetriken für Linked Data Quellen im Jahr 2014 [Zaveri et. al, 2014]. Die Entwicklung der Linked Data Komponente im LEDS Projekt griff diese Publikation auf, wodurch Doppelforschung frühzeitig vermieden werden konnte. Ebenso arbeiteten andere Forschergruppen in der Folge an eigenen Unterstützungstools zur Messung von Datenqualität. Beispiele sind Luzzu [Debattista et. al, 2014], KBMetrics [Ruan et. al, 2015] oder the Roomba OpenData Checker [Ahmad et. al, 2015]. Andere Tools wie RDFAlerts [Hogan et. al, 2010], SWIQA [Fürber et. al, 2011] oder Sieve [Mendes et. al, 2012] waren aus den State of the Art Vorrecherchen bereits bekannt, aber mit den Anforderungen des Projekts nicht innerhalb der LEDS-Technologieplattform nutzbar. Weitere Tools anderer Stellen erschienen für einen Einsatz sinnvoll, jedoch waren diese häufig nicht mehr verfügbar oder wurden nicht mehr weiterentwickelt (End-of-Life). Die im LEDS-Projekt entwickelte "SemQuire" Linked Data Quality Assessment Komponente stellt daher gegenüber anderen inzwischen existierenden Werkzeugen einen Mehrwert da, da sich diese über wohldefinierte Schnittstellen auch in automatisierte Prozessabläufe integrieren lässt und die Bewertung interner und externer Datenquellen sowohl im Push als auch im Pull-Verfahren ermöglicht.

Bei der Extraktion von Linked Data aus unstrukturierten und semistrukturierten Datenquellen veränderte sich ebenso die Verfügbarkeit der angebotenen Werkzeuge. Inzwischen existieren mehrere Lösungen sowohl als Online-Lösung als auch als Offline-Standalone-Tool zur Konvertierung von Nicht-RDF-Dateiinhalten in ein RDF-Serialisierungsformat. Eine gute Übersicht liefert <https://www.w3.org/wiki/ConverterToRdf>. Die im LEDS-Projekt entwickelte "KESeda" Extraktionskomponente hat gegenüber diesen Lösungen jedoch den Vorteil, dass sie nicht nur eine rein syntaktische Konvertierung in ein anderes Datenformat durchführt, sondern zusätzliche Hintergrundinformationen einbeziehen kann und über ein grafisches Nutzerinterface eine vorkonfigurierbare, semi-automatisierte, überwachte Herangehensweise anbietet. Hierdurch werden die Erkennungsraten und semantisch korrekten Zuordnungen der extrahierten Informationen deutlich verbessert.

Die Analyse personalisierter Rankingfaktoren wurde in anderen Arbeiten ebenso bereits diskutiert. Ein Schwerpunkt war hier die Wahrscheinlichkeit der Beeinflussung durch Freundschaftsbeziehungen im sozialen Umfeld [Anagnostopoulos et al., 2008, Goyal et al., 2010]. Beispiele für eine social-aware Top-k spatial keyword search wurden durch [Wu, 2014] dargestellt. [Köster et. al, 2016] diskutierten dann die Rolle von Empfehlungssystemen im M-Commerce. Das LEDS-Projekt mit dem Ergebnis der Social-aware search zeigte praktisch die Interaktion von Webshops mit Daten aus sozialen Netzwerken auf und lieferte dadurch einen realen Proof-of-Concept. Gleichzeitig zeigten wir Möglichkeiten auf, wie semistrukturierte Informationen Daten aus WebAPIs anderer Anbieter für die Weiternutzung in einem Linked Data Kontext aufbereitet werden können. Zum aktuellen Zeitpunkt werden hierfür von sozialen Netzwerken auch neue Zugriffsmöglichkeiten wie beispielsweise über GraphQL bereitgestellt, allerdings ist der Zugriff auf diese Daten nach anhaltenden Data Privacy Debatten stark reglementiert. Das LEDS-Projekt lieferte daher über den aktuellen Stand der Technik hinausgehende Forschungsergebnisse, die wir national und international publizierten und so der gesamten Forschungscommunity zugänglich machen.

6. DER ERFOLGTEN ODER GEPLANTEN VERÖFFENTLICHUNGEN DES ERGEBNISSES NACH NR. 11

Im Rahmen des LEDS-Verbundprojekts entstanden durch das Forschungsteam der Technischen Universität Chemnitz unter anderem die folgenden neun Publikationen:

Langer, André; Göpfert, Christoph; Gaedke, Martin: F.I.E.L.D.S. - Analyzing Form Input interfaces for Explicit Linked Data handling in document Submission systems. Proceedings of 17th International Conference WWW/Internet (ICWI2018), pp. 3 - 10, ISBN/ISSN 978-989-8533-82-1; 2018.

Langer, André; Siegert, Valentin; Göpfert, Christoph; Gaedke, Martin: SemQuire - Assessing the Data Quality of Linked Open Data Sources. Proceedings of 18th International Conference on Web Engineering (ICWE2018), pp. 163 - 175, ISBN/ISSN 978-3-030-03055-1; 2018.

Langer, André; Gaedke, Martin: DaQAR - An ontology for the uniform exchange of comparable Linked Data Quality Assessment Requirements. Proceedings of 18th International Conference on Web Engineering (ICWE2018), pp. 234-242, ISBN/ISSN 978-3-88579-669-5; 2018.

Krug, Michael; Wiedemann, Fabian; Ast, Markus; Gaedke, Martin: The SmartComposition Approach for Creating Environment-Aware Multi-screen Mashups. Rapid Mashup Development Tools. RMC 2016. Communications in Computer and Information Science, vol 696. Springer, Cham, ISBN/ISSN 978-3-319-53174-8; 2017.

Langer, André; Krug, Michael; Moreno, Luis; Gaedke, Martin: Utilizing Linked Data Structures for Social-aware Search Applications. Proceedings of the 47. Jahrestagung der Gesellschaft für Informatik e.V. (GI), pp. 1903-1914, ISBN/ISSN 978-3-88579-669-5; 2017.

Seidel, Martin; Krug, Michael; Langer, André; Gaedke, Martin: KESeDa: Knowledge Extraction from Heterogeneous Semi-Structured Data Sources. Proceedings of the 13th International Conference on Semantic Systems, pp. 129-136, ISBN/ISSN 978-1-4503-4752-5; 2016.

Langer, André; Gaedke, Martin: FAME.Q - A formal approach to master quality in Enterprise Linked Data. Proceedings of the 15th International Conference WWW/Internet (ICWI2016). Mannheim, Germany. October 28 - 30, 2016, pp. 51-58, ISBN/ISSN 978-1-5108-3297-8; 2016.

Burian, Frank: CloudLinkSync: URI-basierte Adressierung lokaler Kopien von Cloud-Ressourcen. Proceedings of Studierendensymposium Informatik 2016 der TU Chemnitz, pp. 141-144, ISBN/ISSN 978-3-944640-85-3; 2016.

Krug, Michael; Gaedke, Martin: AttributeLinking: Exploiting Attributes for Inter-Component Communication. Post-proceedings of the 16th International Conference on Web Engineering (ICWE2016) Workshops, 2016.

Krug, Michael; Gaedke, Martin: SmartComposition: bringing component-based software engineering to the web. Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services (iiWAS '15), 2015. ACM, New York, NY, USA, Article 63, 4 pages. DOI: <http://dx.doi.org/10.1145/2837185.2837247>

Scholtz, Anna; Wild, Stefan; Gaedke, Martin: Systematic composition of web-based applications with focus on security. Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services (iiWAS '15), 2015. ACM, New York, NY, USA, Article 85, 5 pages. DOI: <http://dx.doi.org/10.1145/2837185.2837250>

Heil, Sebastian; Wild, Stefan; Krug, Michael; Gaedke, Martin: Towards Reputation-Aware Expert Finding with Linked Open Data. SEMANTiCS (Posters, Demos, SuCESS), 2016.

III. ERFOLGSKONTROLLBERICHT

1. DEN BEITRAG DES ERGEBNISSES ZU DEN FÖRDERPOLITISCHEN ZIELEN

Das LEDS Verbundprojekt wurde im Rahmen der Programmfamilie "Unternehmen Region" des Bundesforschungsministeriums für Bildung und Forschung (BMBF) gefördert, welches regionale Innovationsbündnisse in Ostdeutschland unterstützt, um zukunftsfähige technologische Profile zu entwickeln. Ziel ist es dabei, Akteure mit hochspezialisierten wirtschaftlichen, wissenschaftlichen und technologischen Kompetenzen mit Innovationspotenzial zu regionalen Clustern zusammenzufassen und die Stärke der Region Westsachsen in diesem Bereich ausnutzen, um in diesem Gebiet neue, langfristig erfolgreiche Wirtschaftsfelder aufzubauen. Mit dem klaren thematischen Fokus auf der Nutzung von Linked Data Technologien in einem wirtschaftsbezogenen Umfeld wurde das Vorhaben LEDS innerhalb des Schwerpunkts innovativer regionaler Wachstumskern gefördert. Als weiteres Ergebnis wurde durch das Projekt talentierten Wissenschaftlern und Fachkräften in der Industrie eine Perspektive verschafft, mit der eccenca GmbH eine innovative Unternehmensgründung forciert und ein Beitrag geleistet zu wettbewerbsfähigen Regionen mit eigenem wirtschaftlichem und wissenschaftlichem Profil.

Als möglicher Zielmarkt für die LEDS-Technologieplattform und die Arbeiten der TU Chemnitz wurde im Speziellen der E-Commerce-Bereich anvisiert. Dieser unterscheidet sich in einigen wesentlichen Faktoren von traditionellen Handelsmärkten. So reagiert dieser sehr agil auf sich ergebende bzw. neue Möglichkeiten und gehört zu den transparentesten Märkten überhaupt. Dies stellt die Marktteilnehmer unter besonderen Druck, schnell auf Entwicklungen der Konkurrenz bzw. auf das Verlangen der Kunden zu reagieren. Daraus ergeben sich wiederum deutlich kürzere Entwicklungszyklen als in traditionellen Unternehmen. Die auf Basis des Linked-Data-Paradigmas repräsentierten Inhalte und semantischen Strukturen unterstützen die erforderliche Agilität in solchen Märkten entscheidend, da LEDS-Systeme das im Web verfügbare Wissen besser interpretieren können, so dass darauf aufbauend z.B. neuartige, themen- und kontext-basierte Integrations-, Such- und Anfragemöglichkeiten entwickelt werden können.

Unternehmen werden damit in die Lage versetzt, ihre internen, strukturierten als auch unstrukturierten Informationen besser zu integrieren und mit dem Wissen des Semantic Web zu vernetzen. Durch die Verwendung gemeinsamer Ontologien und entsprechender Vokabulare können die Kosten der Datenintegration im Unternehmen, insbesondere bei einer großen Anzahl von Datenquellen gegenüber konventionellen Ansätzen, wesentlich gesenkt werden. Auf diesem Weg können eine Vielzahl innovativer Anwendungen im Bereich E-Commerce, Suchtechnologie, E-Government u.v.m. implementiert werden. Diese ermöglichen ein neues Nutzererlebnis und den betreibenden Firmen einen Wettbewerbsvorsprung sowie eine aktive Teilhabe an zukünftigen Entwicklungen des Internets.

I.a.1 Beitrag des Verbundvorhabens zur Lösung der Probleme potenzieller Kunden des Bündnisses

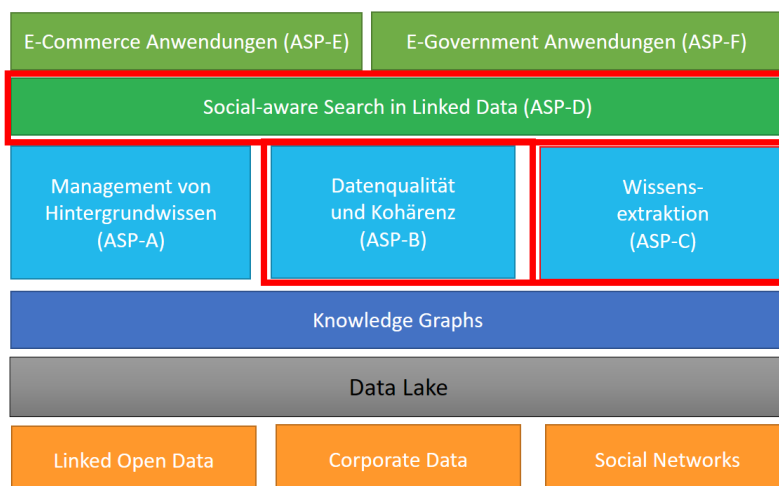
Zunächst werden in der nachstehenden Tabelle die zum Vorhabenbeginn identifizierten Kundenprobleme noch einmal aufgezeigt

| Kundenprobleme | Entwicklungspotential | Neue Geschäftsfelder | Beteiligte Partner |
|-------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|------------------------------------------------------------------------------------------|------------------------------------------------|
| P1: Integration einer großen Anzahl von Informationssystemen im Unternehmen ist trotz Service- | Anwendung und Ausnutzung des leichtgewichtigen Linked Data-Paradigma | Lizensierung der Technologieplattform als Linked Enterprise Data Integrationskomponente; | Lecos, brox, Netresearch, eccenca, (ULEI, TUC) |

| Kundenprobleme | Entwicklungspotential | Neue Geschäftsfelder | Beteiligte Partner |
|----------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|
| orientierten Architekturen ungelöst | für Enterprise Data Integration | Realisierung von Integrationsprojekten | |
| P2: Integration von Daten und Wissen aus dem Daten Web in E-Commerce und E-Government-Anwendungen | Entwicklung von Strategien zur synchronen Co-Evolution von LOD-Hintergrundwissen und Unternehmensdaten | Lizensierung von Software zur Unterstützung der Co-Evolution, kundenspezifische Anpassung, Wartung, Support und Consulting | Netresearch, Lecos, brox, , (ULEI) |
| P3: Analyse von E-Commerce-relevanten Web 2.0 Quellen (Wikis, Blogs, Soziale Netzwerke etc.) | Entwicklung von Methoden zur Inhaltsanalyse mit Nutzung von LOD-Hintergrundwissen | Bereitstellung von Inhaltsanalysediensten u.a. auch für Trendmining, Sentimentanalyse, Konkurrenzmonitoring etc. | Netresearch, AI4BD, (TUC) |
| P4: Vereinfachung der Suche auf Web-Portalen und E-Commerce-Sites | Entwicklung von Themen- und Motiv-basierten Suchfunktionen | Steigerung der Kundenzufriedenheit und des Umsatzes pro Kunde | Netresearch, Lecos, (ULEI) |
| P5: Aufwendige Schulung und Einarbeitung von Mitarbeitern für E-Commerce und Support | Entwicklung von Unterstützung für multimodale, semantische E-Commerce-Prozesse | Erweiterung der Funktionalität und Konkurrenzfähigkeit der E-Commerce-Plattformen der Partner | Netresearch, AI4BD, brox, eccenca, (ULEI) |
| P6: Aufwendige, kosten und ressourcenintensive Integration von E-Government und E-Commerce Daten und Diensten | Entwicklung von Strategien zur dynamischen Vernetzung und Integration von Daten und Diensten im E-Government | Lizensierung von Software und Erarbeitung von Dienstleistungsangeboten für öffentliche Verwaltungen und Unternehmen | brox, eccenca, Lecos, AI4BD, (ULEI) |

Als Lösung wurde innerhalb von drei Jahren eine digitale Technologieplattform, der so genannte LEDS Technologiestack, realisiert und entsprechend des Verwertungsplans in die Produkte und Dienstleister der beteiligten Partner integriert, als auch durch Ausgründung der Firma eccenca GmbH als Kerngeschäftsmodell neuen Kunden zugänglich gemacht.

Zur Umsetzung der Technologieplattform wurde das Projekt in mehrere Arbeitsschwerpunkte gemäß der nachstehenden Abbildung aufgeteilt. Der Arbeitsbereich der TU Chemnitz ist dabei farblich in Rot hervorgehoben.



Die Ergebnisse aus den Arbeitsschwerpunkten ASP-A, ASP-B und ASP-C stellen Basistechnologien für die Technologieplattform bereit, der Beitrag der TU Chemnitz fokussierte dabei die Entwicklung einer Datenqualitätssicherungskomponente ("SemQuire"), die Entwicklung eines Extraktionstools von Wissen aus semistrukturierten Datenbeständen ("KESeDa"), sowie die Implementation und Evaluation eines Prototypen für eine social-aware Search ("SAS"). Alle erarbeiteten Lösungen wurden auf nationalen und internationalen Konferenzen veröffentlicht und damit einem breiten Publikum zugänglich gemacht.

I.a.2 Beitrag des Verbundvorhabens zur technologischen (Weiter-)Entwicklung der Kernkompetenz/der Technologieplattform bzw. zum Ausgleich von Defiziten

Die Arbeiten der TU Chemnitz leisteten einen entscheidenden Beitrag zur Entwicklung der gemeinsamen Technologieplattform.

Als Koordinator des Arbeitsschwerpunkts ASP-B im Leds Projekt stellte die TU Chemnitz mit SemQuire eine Datenqualitätssicherungskomponente für Linked Data bereit, welche in der Lage ist 55 verschiedene Metriken aus unterschiedlichen Datenqualitätsbereichen zu messen und automatisiert strukturiert für die Analyse und Weiterverarbeitung durch andere Komponenten im Leds-Technologiestack bereitzustellen. Zwar beschäftigte sich die Industrie und Forschung in der Vergangenheit bereits umfangreich mit technischen Fragestellungen des Datenmanagements und auch mit der Beschreibung von Datenqualität, jedoch existierte bis zum Beginn des Projekts noch keine Softwarekomponente, welche dieses Vorgehen auf semantisch zugängliche Unternehmensdatenbestände anwenden und automatisiert in bestehende Unternehmensprozesse integrieren konnte. Dabei ist das Management von Wissen aus qualitativ hochwertigen Daten ein entscheidender Wettbewerbsfaktor für den Geschäftserfolg im 21. Jahrhundert. Obwohl in den letzten Jahren vielfach über Big Data Aspekte im Unternehmensumfeld diskutiert wurde, wurde die automatisierte Prüfung von Datenqualitätsmetriken häufig vernachlässigt. Innerhalb der Leds-Technologieplattform erlaubt die SemQuire Datenqualitätssicherungskomponente sowohl die Bewertung unternehmenseigener Datenbestände als auch aus externen Linked Open Data Quellen im Push- und Pullverfahren basierend auf offenen Standards wie W3C RDF, SPARQL und DQV.

Weiterhin ist die Erschließung von Wissen durch Extraktion aus existierenden Datenbeständen für Unternehmen eine zweite wichtige Herausforderung gewesen. Im Arbeitsschwerpunkt ASP-C konzentrierte sich die TU Chemnitz auf die Entwicklung einer Extraktionskomponente zur Bereitstellung von Linked Data aus semistrukturierten XML- oder JSON-Daten unter dem Namen KESeDa. Zum Vorhabenbeginn gab es hierfür noch keine vorhandene Lösung, welche einfach in bestehende Prozesse integrierbar war und in der Lage war, mithilfe existierender Ontologien, Schema Mapping und Hintergrundwissen Daten aus vorhandene Datensätzen semantisch zugänglich zu machen.

Besonders Projektpartner aus dem E-Commerce-Bereich interessierten sich für die Verbindung unternehmensinterner Daten mit nutzerspezifischen Interaktionsdaten und öffentlich verfügbaren Hintergrundinformationen zu dem sozialen Umfeld eines Benutzers. Im Arbeitsschwerpunkt ASP-D nutzte die TU Chemnitz das Linked Data Paradigma, um Unternehmen in diesem Bereich eine neue Art der adaptiven Suche in E-Commerce-Shops bereitzustellen. Hierfür wurde mit der Social-aware Search Komponente ein Prototyp realisiert, welcher die Basisfunktionalitäten der LEDS-Technologieplattform intelligent miteinander verband, um Wissen zum sozialen Umfeld eines Shopbesuchers aus existierenden semistrukturierten Datenquellen zu extrahieren und so zu verknüpfen, dass dem Nutzer bei der Suche nach Produkten ein noch besseres Shopperlebnis geboten werden kann. Die Realisierbarkeit eines derartigen Vorgehens wurde praktisch aufgezeigt und im Rahmen des Projekts evaluiert.

I.a.3 Beitrag des Verbundvorhabens zur Entwicklung neuer Produkte und/oder Dienstleistungen

Als Forschungseinrichtung strebt die TU Chemnitz keine wirtschaftliche Verwertung der Forschungsergebnisse an. Dementsprechend wurde oder wird keine Auftragsentwicklung für neue Projekte durchgeführt. Die industriellen Geschäftspartner der TU Chemnitz im LEDS-Projekt nutzen das erarbeitete Wissen und die geschaffenen Basistechnologien weiter zur Etablierung neuer Geschäftsfelder im Rahmen der gegründeten eccenca GmbH in Form einer Datenplattform für den industriellen Umgang mit Linked Data.

Entsprechend des Verwertungsplans konnten die Projektergebnisse aber zur Dissemination von Forschungsergebnissen genutzt werden. Auch in nachfolgenden Projektanträgen haben die Erkenntnisse aus dem LEDS-Projekt für die TU Chemnitz einen entscheidenden Wert, da das Wissen und die entwickelten Komponenten in zukünftige Vorhaben direkt eingebracht werden kann. Im Rahmen zweier Promotionsprojekte passiert dies derzeit bereits in den Bereichen des Managements von Forschungsdaten und der Forschung an Content Trust.

Zuletzt haben die Ergebnisse aus dem LEDS-Verbundprojekt unmittelbaren Eingang in die Lehrtätigkeiten der Professur Verteilte und Selbstorganisierende Rechnersysteme (VSR) an der TU Chemnitz gehalten und sind fester Bestandteil der Vorlesungen XML und Current Trends in Web Engineering geworden. Zusätzlich arbeiten Studierende im Rahmen eines Planspiels Web Engineering in Projektteams an neuen Forschungs- und Geschäftsideen basierend auf den Entwicklungen in der Linked Data Community und dem LEDS Projekt. Dabei wird auch in Abschlussarbeiten kontinuierlich ein Fokus auf die Ausschreibung von Themen im Bereich von Linked Enterprise Data Services gelegt. Außerdem findet ein Austausch von Studierenden und Nachwuchsforschern im Netzwerk des LEDS-Bündnisses statt.

2. DAS WISSENSCHAFTLICH-TECHNISCHE ERGEBNIS DES VORHABENS, DIE ERREICHTEN NEBENERGEBNISSE UND DIE GESAMMELTEN WESENTLICHEN ERFAHRUNGEN

Die Forschungsarbeit der TU Chemnitz lieferten entsprechend des Abschnitts 1 einen wesentlichen Beitrag zum LEDS Technologiestacks der Bündnispartner.

Dabei wurden mit der Technologieplattform im Gesamtprojekt die folgenden Lösungskonzepte verfolgt:

Z1: Integration und Management von Hintergrundwissen und Unternehmensdaten (siehe P1, P2 aus I.a). Die Technologieplattform realisiert einen generischen Prozess, der existierende Wissensbasen und Vokabulare der Linked Open Data Cloud entsprechend den gewünschten Kriterien aufbereitet, mit unternehmenseigenen Datenquellen integriert und die konsistente, synchrone Co-Evolution interner und externer Datenquellen sicherstellt.

Z2: Unterstützung der verteilten Evolution und Syndikation von öffentlichen und privaten Daten (siehe P1, P2 aus I.a). Datenquellen im Unternehmen, bei Kooperationspartnern innerhalb von Wertschöpfungsnetzwerken und im Linked Data Web entwickeln sich unabhängig voneinander weiter. Ein zentrales LEADS-Ziel ist es diese Evolution durch geeignete Tracking-, Syndikations- und Migrations-Strategien zu unterstützen. Dabei soll der flexible Datenaustausch in Wertschöpfungsnetzwerken mit einer Vielzahl von beteiligten Unternehmen reibungslos unterstützt werden.

Z3: Überwachung der Datenqualität und Zugriffskontrolle (siehe P1, P2 aus I.a). Die Bewertung der Datenqualität von Linked Open Data ist essentiell für unternehmenskritische Anwendungen. In LEADS werden Qualitätsbewertungstechniken für Linked Open Data erforscht und in die Technologieplattform integriert. Die Bewertungskriterien berücksichtigen Metriken entlang den Dimensionen Bereichen Kontext (z.B. Provenance), intrinsische Kriterien (z.B. Korrektheit, Genauigkeit), Repräsentation, Dynamik und Zugriff. Darüber hinaus werden Techniken entwickelt, die eine feingranulare Zugriffskontrolle auf Linked Enterprise Data im Unternehmen für verschiedene Nutzergruppen ermöglichen.

Z4: Semantische Inhaltsanalyse auf Basis von Daten-Web Hintergrundwissen (siehe P3 aus I.a). Methoden der automatischen Sprachverarbeitung (insbesondere statistische Ansätze) sollen mit bereits vorhandenen Strukturinformationen (z.B. RDFa) sowie Informationen aus der Linked Data Cloud kombiniert werden, um eine genauere Inhaltsbestimmung von Webseiten (insbesondere von für E-Commerce-Prozesse relevante Web 2.0 Quellen wie Blogs oder Kommentare aus sozialen Netzwerken) zu ermöglichen und diese damit in den E-Commerce Wertschöpfungskreislauf zu integrieren. Neben der Integration von unstrukturierten Informationsquellen zielen wir mit der LEADS-Technologieplattform auch auf semi-strukturierte (z.B. Web-Daten) und strukturierte (z.B. relationale Datenbanken) Datenquellen.

Z5: Skalierbare themen-/motivbasierte, personalisierte Suche (siehe P4 aus I.a). Die Technologieplattform wird mit neuartigen Suchstrategien ausgestattet, die anders als existierende Suchparadigmen nicht allein auf vorhandene Datenbestände fokussieren, sondern zusätzlich das mittels Z1 und Z2 bereitgestellte Hintergrundwissen ebenso integrieren, wie personalisierte, kontext- und aufgabenbezogene Einschränkungen und Präferenzen. Ein Schwerpunkt bildet in diesem Zusammenhang auch die Realisierung performanter und intuitiver Benutzerschnittstellen.

Z6: Multimodale, semantische E-Commerce-Unterstützung im Onlinehandel (siehe P5 aus I.a). Multichannel-Kommunikation und eine effiziente Unterstützung in der Kundenberatung spielen eine entscheidende Rolle in der Online-Abwicklung von Beratungs- und Bestellvorgängen. Diese Verbindung stellt hohe Anforderungen an den Zugang zu den erforderlichen Informationen. Es werden Werkzeuge zur semantischen Transformation von E-Commerce-Daten sowie eine multimodale Kundenschnittstelle entwickelt, die insbesondere die themenbasierte Suche und Systeme zur Ermittlung von kontextsensitivem Hintergrundwissen integrieren. Ziel ist es Kunden sowohl bei Informations-, Support- und Bestellvorgängen über Internet-Portale mit relevanten Zusatzinformationen zu versorgen und zielgerichteter zu bedienen.

Z7: Vernetzung von E-Government-Daten (siehe P6 aus I.a). In diesem Anwendungsfall wird die Technologieplattform im Zusammenspiel von öffentlichen Verwaltungen und der freien Wirtschaft evaluiert. Durch die Integration verschiedenster Datenquellen (wie. z.B. zu Flächennutzung, Immobilien, Kaufkraft, existierenden wirtschaftlichen Strukturen, Geodaten) wird die Technologieplattform die Vernetzung von öffentlichen (E-Government) und privaten (E-Commerce) Diensten ermöglichen und Impulse für eine neue Generation von kombinierten Public-Private-Services liefern.

Nachfolgend sind die erzielten wissenschaftlich-technischen Ergebnisse der TU Chemnitz in allen Arbeitsschwerpunkten mit den dabei gemachten Erfahrungen näher dargestellt und erläutert:

AP 1.4.B.1: Spezifikation (Anforderungs- und Funktionsdefinition) der Linked Data Quality Komponente (M1-M9)

Zunächst wurde ein Grundverständnis für die Thematik "Qualität und Kohärenz" geschaffen, indem eine Zusammenfassung zu grundlegenden Themen wie Datenverarbeitung und Datenformaten, im Besonderen in Hinblick auf deren Strukturiertheit, erarbeitet wurde. Auch die Entwicklung des Internets, Webs und die damit verbundenen neuen Herausforderungen insbesondere für Unternehmen wurden betrachtet. Bei dieser Thematik hat das Thema "Linked Data" eine zentrale Rolle gespielt.

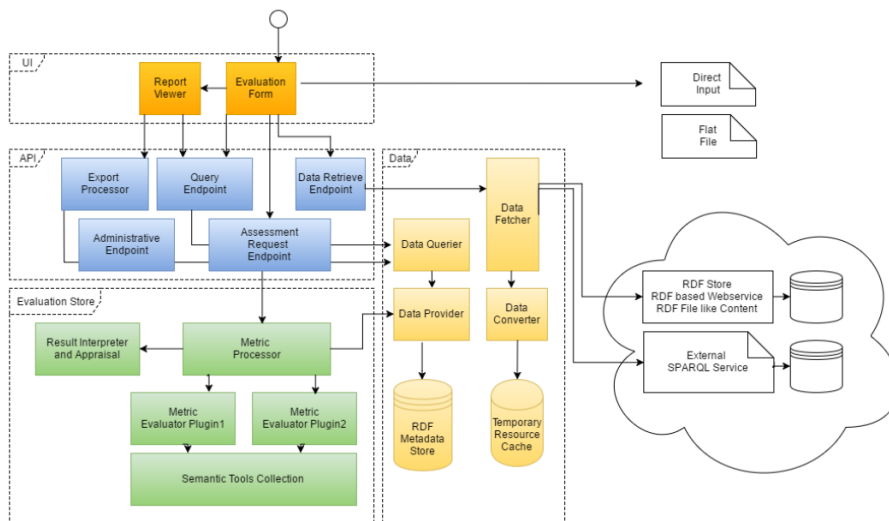
Des Weiteren wurde eine Anforderungsanalyse durchgeführt. Dazu wurde untersucht, welche Qualitätskriterien existieren und inwiefern diese relevant beispielsweise für die Verbesserung von semantischen Suchanfragen sind. Dementsprechend wurde eine umfassende Literaturrecherche und Stand der Technik Untersuchung angeschlossen. Im Einzelnen wurden dabei viele Arbeiten gefunden, welche Informationsqualität allgemein betrachten. Zusätzlich wurden die Recherchen im Hinblick auf die Kontexte Web, Linked Data sowie der Anfragebeantwortung erweitert.

Die gefundenen Qualitätskriterien wurden anhand bekannter Kriterien, beispielsweise nach ihrer Bestimmung, des Informationstyps oder der Messbarkeit, gruppiert. Es wurden Metriken untersucht und Strategien für Qualitätsverbesserung aus der Literatur gewonnen. Weiterhin wurden Qualitätsaspekte, strukturiert nach Qualitätsdimensionen, im Detail untersucht. Dazu wurde zunächst jeder Aspekt definiert und erläutert. Nachfolgend wurden Gebiete herausgearbeitet, in denen der jeweilige Aspekt besonders wichtig ist und gegebenenfalls notwendige Voraussetzungen für die Möglichkeit der Bewertung erörtert, sowie mögliche Metriken genauer untersucht. Eine Liste von existierenden Frameworks zur Untersuchung von Datenqualität in semantischen Wissensbasen ist ebenso Teil des Berichtes. Abschließend wurden Konzeptideen zur Qualitätssicherung erarbeitet.

Der Meilenstein *MS-1.4.B.1 Schnittstellen-, Technologie- und Architekturspezifikation der Linked Data Quality Komponente* wurde abgeschlossen. Das dazugehörige Deliverable gibt als Ergebnis der Evaluation aus dem ersten Projektzeitraum einen umfassenden Überblick über existierende Technologien und Tools, welche für die Entwicklung einer Linked Enterprise Data Service Plattform verwendet werden können. Als Resultat einer vorab durchgeführten Literaturrecherche werden geeignete Qualitätskriterien und dazugehörige Metriken aufgelistet und in vier verschiedenen Gruppen aufgeteilt. Davon ausgehend werden alle funktionalen und technischen Anforderungen an die Qualitätskomponente abgeleitet und genannt, sowie konzeptionell überführt in einer ersten Systemarchitektur dargestellt.

AP 1.4.B.2: Schnittstellen und Architekturprototyp der Linked Data Quality Komponente (M7-M15)

Zur fundierten Umsetzung der Linked Data Quality Komponente wurde zunächst ein Basis-Algorithmus zur Berechnung und Bewertung von Qualität einer Datenquelle definiert und mit existierenden Ansätzen verglichen. Dabei wurde festgestellt, dass bereits vorhandene Technologien in diesem Bereich zumeist informelle, eigene Algorithmen einsetzen und keine fundierte Formalisierung, besonders im Umgang mit Linked Data Sources existiert. Diese Formalisierung wurde vorgenommen und in einer wissenschaftlichen Veröffentlichung mit dem Titel „*FAME.Q – A Formal Approach to Master Quality in Enterprise Linked Data*“ vorbereitet. Darauf aufbauend können nun die bereits im AP 1.4.B.1 identifizierten Qualitätskriterien in einem ersten Basissatz in einer Demonstrator-Komponente implementiert werden. Weiterhin wurden existierende Protokolle und Ontologien zur Anbindung der Qualitätskomponente in die LEDS-Technologieplattform untersucht. Ziel ist es, die Qualitätskomponente auf etablierten Standards aufzubauen und so eine möglichst große Interoperabilität zu erreichen. Im Ergebnis wurde dies technisch durch mehrere unabhängige Systemkomponenten umgesetzt, welche ihre Dienste als Webservice zur Verfügung stellen, die den REST Prinzipien folgen, und darüber Daten maschinenlesbar und automatisierbar austauschen können. Zum systemübergreifenden Datenaustausch war besonders der aktuelle W3C Draft zu einem *Data Quality Vocabulary* interessant, welcher ebenfalls durch die Linked Data Quality Komponente in der LEDS Plattform unterstützt werden sollte.



Nach der Erarbeitung des Architekturkonzepts für die Datenqualitätssicherungskomponente und der Identifikation geeigneter Übertragungsprotokolle und Ontologien konnte dies planmäßig weiter konkretisiert und in eine Referenzarchitektur- und Schnittstellenbeschreibung mit allen benötigten Teilkomponenten überführt werden. Dazu wurden von uns alle Module in ihrer Funktion inklusive der erwarteten Ein- und Ausgabewerte spezifiziert, der erwartete Programmablauf bei einer prototypischen Implementation schematisch dargestellt, sowie das Kommunikationsprotokoll zur Nutzung und Steuerung der Datenqualitätskomponente über spezifische Webservice-Endpunkte konkret beschrieben. Besonderes Augenmerk lag hierbei auf die spätere Integration der Komponente in die gemeinsame LEDS-Technologieplattform, weswegen Daten sowohl im Push- als auch im Pull-Verfahren von verschiedenartigen Datenquellen entgegengenommen und analysiert werden können. Ebenfalls wurden drei grundlegende Testszenarien formuliert, welche mithilfe der Linked Data Quality Komponente durchführbar sein sollen. Die Ergebnisse aus der konzeptionellen Betrachtung im AP 1.4.B.2 wurden anschließend benutzt, um in den Folgemonaten einen ersten Demonstrator einer Qualitätssicherungskomponente zu implementieren.

AP 1.4.B.3: Entwicklung erster Demonstrator der Linked Data Quality Komponente (M13-M21)

Ziel des Arbeitsschwerpunkts 1.4.B.3 war die Realisierung eines ersten Demonstrators, der praktisch in der Lage ist, Datenqualitätsaspekte in mehrere Bewertungsdimensionen auf Linked Enterprise Datenbeständen zu messen. Hierfür wurde zunächst ein Datenmodell definiert, welches es sowohl erlaubt, die gewünschten Qualitätsdimensionen mit Metadaten zu beschreiben, als auch die zu bewertenden Daten zu verwalten und Messungen darauf durchzuführen. Das Datenmodell ist konsistent, kompatibel zu anderen weltweit vorgeschlagenen Datenqualitätsaustauschformaten, und basiert ebenfalls auf einer Formulierung mittels Linked Data. Technisch abgebildet wird das Datenmodell in einer StarDog Graph-Datenbank, welche auch in anderen Arbeitspaketen die favorisierte Plattform bildet. Softwareseitig wurde der Demonstrator in drei Teilen implementiert: Einem Web-Frontend, welches eine grafische Bedienoberfläche zur manuellen Nutzung der Linked Data Quality Component bereitstellt, einem REST-Webservice, welcher die (semi-)automatische Ansteuerung und Nutzung der Qualitätskomponente durch andere Anwendungen der Technologieplattform mithilfe des Kommunikationsprotokolls aus AP 1.4.B.2 erlaubt, sowie einer Prozessor-Komponente, welche die Implementation aller zur Verfügung stehenden Qualitätsmetriken enthält. Diese drei Teilkomponenten wurden bis M18 implementiert und können seitdem wie geplant miteinander agieren und inzwischen erste Messungen von Qualitätsmetriken vornehmen. Ebenso besteht eine maschinenlesbare Exportmöglichkeit der Messergebnisse im DQV-Format. Der Meilenstein *MS-1.4.B.2 Fertigstellung der Basis-Implementierung der Linked Data Quality Komponente* wurde damit planmäßig erreicht und alle Ergebnisse im Deliverable MS-1.4.B.2 dokumentiert.

In dem darauffolgenden Projektabschnitt wurde ein erster praktisch nutzbarer Demonstrator der Linked Data Quality Komponente entwickelt. Dieser umfasst sowohl einen automatisiert ansteuerbaren Webservice zur Prüfung spezifischer Qualitätscharakteristiken vorliegender Daten als auch ein nutzerfreundliches Web-Interface zur manuellen Angabe und Prüfung konkreter Linked Data Quellen. Ebenso besteht eine Exportmöglichkeit der Analyseergebnisse in einem semantischen Datenformat unter Nutzung des aktuellen Drafts der W3C Working Group für ein *Data Quality Vocabulary*. Innerhalb des Demonstrators wurden zunächst vereinzelte Basisalgorithmen zur Prüfung der Linked Data *Accessibility*, *Performance* sowie intrinsische Metriken und statistische Eigenschaften eines Linked Datasets umgesetzt, die im weiteren Verlauf effizient getestet und miteinander in Beziehung gesetzt werden können. Die gesamte Komponente wurde anschließend bereits in ein *Docker*-Image überführt, um sie innerhalb der LEDS-Technologieplattform ohne viel Aufwand mit anderen Projektkomponenten kombinieren zu können. Zusätzlich erstellten wir einen Continuous Integration Workflow, der jegliche Aktualisierungen an der Linked Data Quality Komponente automatisch auch in die LEDS Technologieplattform übernimmt.

AP 1.4.B.4: Test und Optimierung der Linked Data Quality Komponente (M19-M27)

Innerhalb des Entwicklungsplans war in den zurückliegenden Monaten die Implementierung weiterer Basis-Algorithmen zur Analyse von Linked Data Qualitätscharakteristiken, die Evaluation der korrekten Funktionsweise bereits implementierter Metriken sowie eine fortführende Optimierung der implementierten Funktionalitäten vorgesehen. Dafür wurde zunächst mithilfe des *Mocha* Frameworks eine erste Testplattform aufgesetzt, womit implementierte Metriken zur Qualitätssicherung von Linked Data Quellen isoliert getestet werden konnten. Ebenso wurde intensiv an einem Austauschformat für die Spezifikation von Qualitätsanforderungen an Linked Data Quellen gearbeitet, der den State of the Art weiter voranbringt, da hier innerhalb der wissenschaftlichen Community größere Defizite entdeckt worden waren.

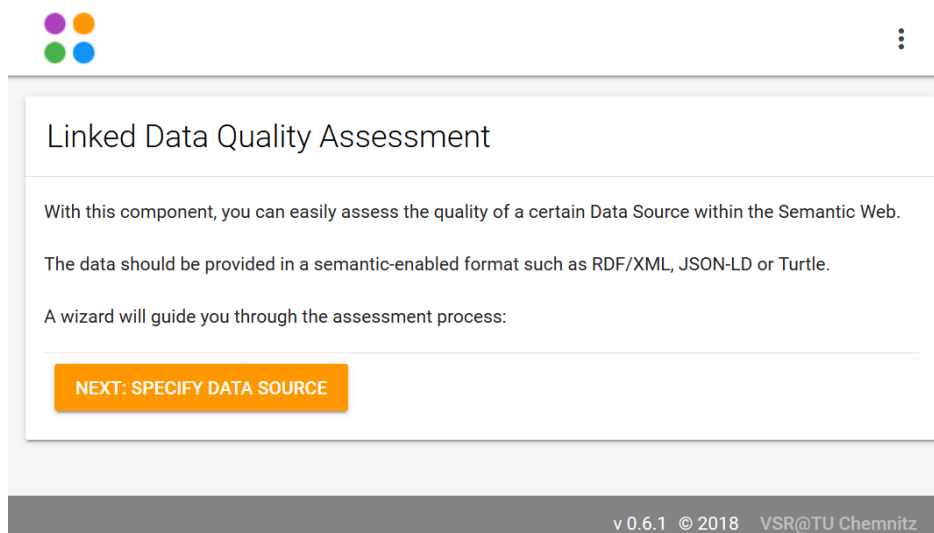
Inzwischen waren 22 verschiedene Basis-Qualitätsmetriken aus den Bereichen Linked Data Accessibility und Representation implementiert, die in der früheren State of the Art - Analyse als geeigneter Qualitätsaspekt zur Messung identifiziert und anwendungsdomänenunabhängig empfohlen worden waren. Zu jeder umgesetzten Qualitätsmetrik wurde ein separater Test geschrieben, um die Korrektheit der Messung an Beispieldaten aus der Realwelt testen zu können. Ebenso umgesetzt und demonstriert wurde die praktische Messung mehrerer oder aller möglichen Qualitätsmetriken innerhalb einer Webanwendung mit graphischer Benutzeroberfläche (WebUI), über welche ein Benutzer systemgeführt durch einen Wizard zunächst die zu testende Datenquelle spezifizieren, die zu prüfenden Qualitätsaspekte auswählen und anschließend die Messoperationen durchführen lassen kann. Die Messergebnisse werden anschließend übersichtlich präsentiert, und sind innerhalb der LEDS-Technologieplattform in einem maschinenlesbaren Format zur Weiterverarbeitung durch andere Softwarekomponenten exportierbar.

Im Anschluss daran wurden wohl-definierte Software-Tests durchgeführt und ausgewertet. Bei der Evaluation der Gruppe *Accessibility*-Metriken traten keine erkennbaren Fehler auf. Bei den implementierten *repräsentationalen* Metriken konnten während der Tests plattformspezifische Spezialfälle beobachtet werden, bei denen angeforderte Datensätze nicht in dem erwarteten Serialisierungsformat korrekt zurückgegeben wurden. Die Erkenntnisse der Tests wurden genutzt, um die Basisimplementierung anzupassen und die Qualität der durchgeführten Messungen zu optimieren. Weiterhin wurden zusätzliche Arbeiten am Web User Interface zur Präsentation der Messergebnisse durchgeführt und ferner die Datenhaltung im Backend der Datensicherungskomponente verbessert.

AP 1.4.B.5: Erweiterung der Linked Data Quality Komponente (M27-M40)

In der Folge lief die Implementierung weiterer relevanter Metriken, welche verschiedene Qualitätseigenschaften auf intrinsischer und kontextueller Ebene analysieren können. Darüber hinaus wurde kontinuierlich an der weiteren Umsetzung der Forschungsleistung zur nutzungsbezogenen Spezifikation von Qualitätsanforderungen an eine konkrete Datenquelle, welche eine überblickshafte Gesamtbewertung des Qualitätszustands einer potentiellen Datenquelle zu einem konkreten Zeitpunkt ermöglicht, gearbeitet. Hierzu wurden die Einzelmessungen der Qualitätsmetriken mit Erwartungswerten abgeglichen und gewichtet zu einer Gesamtschätzung kombiniert. Eventuell gefundene Qualitätsprobleme können dabei dem Benutzer dargestellt werden. Der Meilenstein *MS-1.4.B.3 Fertigstellung der finalen Implementierung der Linked Data Quality Komponente* wurde in der ersten Jahreshälfte 2018 entsprechend des Arbeitsplans erfolgreich erreichen.

Zunächst wurde die bestehende Implementation der Datenqualitätssicherungskomponente entsprechend des ursprünglichen Umsetzungskonzepts dahingehend überprüft und erweitert, dass Daten sowohl im Push-Verfahren als direkte Eingabe oder Upload, als auch im Pull-Verfahren von externen Datenquellen von einer URL oder einem SPARQL-Endpunkt geladen werden können. Zusätzlich wurde ein Caching-Mechanismus implementiert, der bereits geladene Daten nicht noch einmal erneut abrufen, sondern bei Bedarf als lokale Kopie verarbeitet, sofern dies die konkrete Metrik erlaubt. Dadurch verbesserte sich sowohl der Kommunikationsoverhead als auch die Skalierbarkeit der Anwendung bei größeren Datenmengen. Weiterhin wurde daran gearbeitet, mehrere Metriken die nicht voneinander abhängig sind, parallel berechnen zu lassen, um so zu einer schnelleren Gesamtqualitätsbewertung zu kommen. Hierbei konnte in der Praxis sowohl die Implementierung in NodeJS als auch die konzeptionelle Umsetzung einzelner Metriken als eigenständige Module und der Ergebnisabruf über einen eindeutigen Deeplink als vorteilhaft bestätigt werden. Das Management und das Interface zur Anbindung der Datenqualitätssicherungskomponente wurde mit den anderen Projektpartnern besprochen. Die Konfiguration der Komponente wurde so zentralisiert, dass sie über eine eigenständige Konfigurationsdatei ohne weitere Nutzerinteraktion stattfinden kann. Die weitere Spezifikation eines konkreten Messauftrags erfolgt direkt durch WebAPI Calls. Dies umfasst sowohl die zu prüfende Datenquelle, die durchzuführenden Messungen als auch eventuelle Rahmenbedingungen. Letzteres ist weiterhin Gegenstand unserer laufenden Forschung, bei der benötigte Informationen zur Messung *kontextueller* Metriken identifiziert werden. Insgesamt stellte sich über die zurückliegenden Monate heraus, dass die Umsetzung einer Komponente zur Datenqualitätsbewertung sehr komplex ist, und die Implementation einer Vielzahl von Datenqualitätsmetriken aus unterschiedlichen Schwerpunktbereichen erfordert, um ein umfassendes Gesamtbild der zu verarbeitenden Unternehmensdaten betrachten und realistisch einschätzen zu können. Zusätzlich wurden seit Jahresbeginn aus diesem Grund 17 weitere Datenqualitätsmetriken zur Analyse *intrinsischer* Eigenschaften von Datensätzen implementiert.





Result page for 1558258029422

Task description

Data Source <http://dbpedia.org/sparql?query=CONS...>
Media Type application/sparql-results+xml
Source Type <http://res.semquire.net/concepts/SPARQLSource>
Data Dump /data/Distribution1558258025158_1
Start Time 19.5.2019, 11:27:09
End Time 19.5.2019, 11:28:40

Overall Quality Assessment

★★★★☆ Good (0.63)

Measurements

Accessibility dimensions → Availability → *DereferencedForwardLinksMetric* Value: 3

v 0.6.1 © 2018 VSR@TU Chemnitz

```
.myQualityChecker
  rdf:type dqv:SoftwareAgent;
  rdfs:label 'SemQuire Linked Enterprise Data Services Quality Assessment Component';
  .

.myQualityMetadata
  rdf:type dqv:QualityMetadata;
  prov:wasAttributedTo :myQualityChecker;
  prov:generatedAtTime '2019-05-19T09:33:43.024Z';
  prov:wasGeneratedBy :myQualityChecking;
  .

.myQualityChecking
  rdf:type prov:Activity;
  rdfs:label 'Another awesome SemQuire Quality assessment check';
  prov:wasAssociatedWith :myQualityChecker;
  prov:used <http://res.semquire.net/data/Distribution1558258025158\_1>;
  prov:generated :myQualityMetadata;
  prov:endedAtTime '2019-05-19T09:28:40.678Z';
  prov:startedAtTime '2019-05-19T09:27:09.422Z';
  .

<http://res.semquire.net/data/Distribution1558258025158\_1>
  rdf:type <http://www.w3.org/ns/dcat#Distribution>;
  dcat:modified2019-05-19T09:27:05.158Z;
  dcat:accessUrl <http://dbpedia.org/sparql?query=CONSTRUCT%0D%0A%7B%0D%0A%3Fs%20%3Fp%20>;
  dcat:downloadUrl <http://res.semquire.net/data/Distribution1558258025158\_1>;
  dcat:mediaType application/sparql-results+xml;
  dqv:hasQualityMeasurement <http://res.semquire.net/results/Metric1558258029422\_0>;
  .

<http://res.semquire.net/results/Metric1558258029422\_0>
  rdf:type <http://www.w3.org/ns/dqv#QualityMeasurement>;
  dqv:computedOn <http://res.semquire.net/data/Distribution1558258025158\_1>;
  dqv:isMeasurementOf <http://res.semquire.net/concepts/DereferencedForwardLinksMetric>;
  dqv:value '3';
  .

<http://res.semquire.net/results/Metric1558258029422\_1>
  rdf:type <http://www.w3.org/ns/dqv#QualityMeasurement>;
  dqv:computedOn <http://res.semquire.net/data/Distribution1558258025158\_1>;
  dqv:isMeasurementOf <http://res.semquire.net/concepts/DumpDownloadAvailableMetric>;
  dqv:value 'false';
  .

<http://res.semquire.net/results/Metric1558258029422\_10>
  rdf:type <http://www.w3.org/ns/dqv#QualityMeasurement>;
  dqv:computedOn <http://res.semquire.net/data/Distribution1558258025158\_1>;
  dqv:isMeasurementOf <http://res.semquire.net/concepts/LowLatencyMetric>;
  dqv:value '0.427762174';
  .
```

AP 1.4.B.6: Abschlusstests, Integration und Dokumentation der Linked Data Quality Komponente (M28-M41)

Die Linked Data Quality Komponente wurde bereits frühzeitig in Form eines Docker-Containers in die LEDS-Technologieplattform integriert und ist in der Lage, über die in der konzeptionellen Phase vorgesehenen Webservice-Schnittstellen mit anderen Komponenten der Technologieplattform zu kommunizieren. Weiterhin ist bereits der Datenabruf über die projektpartnereigene *DataPlatform* erfolgreich umgesetzt und getestet.

Nach Abschluss des Arbeitspakets 1.4.B.4 und der Implementation intrinsischer Datenqualitätsmetriken als Erweiterung der Linked Data Quality Komponente in AP 1.4.B.5 konnte nun auch die Durchführung von Abschlusstests am finalen Prototypen der Datenqualitätssicherungskomponente voranschreiten. Wir erstellten hierfür basierend auf den bereits spezifizierten Testszenerien für alle Metriken Softwaretests mithilfe des Mocha Testframeworks und Realweltdaten verschiedener Linked Data Provider. Die Testergebnisse wurden evaluiert, auftretende Probleme identifiziert und korrigiert. Mit Erreichen des Meilensteins *MS-1.4.B.4 „Fertigstellung der Stabilisierung der Linked Data Quality Komponente sowie Integration in die LEDS Technologieplattform“* erfolgte dann die abschließende Dokumentation der Datenqualitätssicherungskomponente.

Die Bereitstellung der Linked Data Quality Komponente erfolgte unter <https://docker-registry.eccenca.com/tuc-semquire>.

AP 1.4.C.1: Spezifikation (Anforderungs- und Funktionsdefinition) des Extraktionsmechanismus für semistrukturierte Datenquellen (M1-M9)

Im Rahmen des Arbeitspakets 1.4.C.1 wurde eine Stand-der-Technik-Analyse von Extraktionsmechanismen für semistrukturierte Daten durchgeführt. Dazu wurden zuerst wichtige Begriffe definiert, die für das Grundverständnis der Extraktionsproblematik notwendig sind. So musste beispielsweise eine Abgrenzung von vollstrukturierten, semistrukturierten und unstrukturierten Daten vorgenommen werden – dies gelingt häufig nur sinnvoll im Kontext eines Anwendungsbezugs. Weiterhin wurden Anforderungen an eine Extraktionslösung aufgestellt. Anhand dieser Anforderungen wurden existierende Ansätze im Gebiet der Datenextraktion untersucht. Dazu wurden besonders die folgenden Bereiche beleuchtet: Wrapper Induction Werkzeuge, Sprachen für Wrapper Entwicklung, NLP-basierte Ansätze, HTML-basierende Werkzeuge, sowie Modelling-basierende Werkzeuge. Neben der Untersuchung der allgemeinen, theoretischen Ansätze wurden auch existierende Tools auf ihre Einsetzbarkeit als Teillösung im Rahmen des Projektes überprüft.

Zusätzlich wurde begonnen, ein Architekturkonzept für eine mögliche Extraktionskomponente für semistrukturierte Daten zu erarbeiten. Dieses Konzept beinhaltet unter anderem eine Komponente für die Erkennung von Datentypen, eine weitere Komponente zur Extraktion von nützlichem Wissen aus API-Dokumentationen, sowie der Transformationskomponente für die Überführung in RDF. Als primäre Datenquelle sollen Daten aus sozialen Netzwerken analysiert und weiterverarbeitet werden.

Aufbauend auf der umfangreichen Stand-der-Technik-Analyse, welche im vorherigen Berichtszeitraum durchgeführt wurde, ist das initiale Architekturkonzept für eine mögliche Extraktionskomponente für semistrukturierte Datenquellen weiterentwickelt worden. Weiterhin wurden Technologien zur Umsetzung evaluiert. Entsprechend ist dieses Arbeitspaket mit der Fertigstellung eines Spezifikationsdokuments des Extraktionsmechanismus abgeschlossen und der Meilenstein *MS-1.4.C.1 Schnittstellen-, Technologie- und Architekturspezifikation des Extraktionsmechanismus für semistrukturierte Datenquellen* fristgerecht erreicht worden.

AP 1.4.C.2: Schnittstellen und Architekturprototyp des Extraktionsmechanismus für semistrukturierte Datenquellen (M7-M15)

Anhand des in AP 1.4.C.1 entwickelten Architektur- und Technologiekonzepts wurde ein initialer Prototyp entwickelt. Dabei wurden erste Algorithmen und Komponenten zur Datenverarbeitung und Analyse implementiert. Aktuell ist hier ein erster lauffähiger Demonstrator entstanden, welcher JSON-Daten mittels verschiedener Matching-Algorithmen in RDF-Daten wandelt und dabei versucht, möglichst passende Klassen für Objekte zuzuweisen. Dieser Prototyp befindet sich aktuell in stetiger Weiterentwicklung und fungiert als Basis für die nachfolgenden Arbeiten. Der dazugehörige Meilenstein *MS-1.4.C.2 Fertigstellung Basis-Extraktionsmechanismus für semistrukturierte Datenquellen* ist laut Arbeitsplan in Monat M18 fällig und planmäßig erreichbar.

Aufbauend, auf dem in AP 1.4.C.1 entwickelten Architektur- und Technologiekonzepts wurde in diesem Arbeitspaket ein initialer Prototyp des Extraktionsmechanismus für semistrukturierte Datenquellen entwickelt. In einem mehrstufigen Prozess, welcher aus verschiedenen Komponenten zur Datenanalyse und Datenverarbeitung besteht, können XML- und JSON-Daten verarbeitet werden. Durch die Möglichkeit, verschiedene Konfigurationsoptionen über eine spezifizierte REST-Schnittstelle festlegen zu können, ist der Extraktionsmechanismus an viele Einsatzszenarien anpassbar. Als Ergebnis des Bearbeitungsprozesses entsteht eine semantisch annotierte RDF-Datei, welche in nachfolgenden Systemen, z.B. zum Einsatz in Suchsystemen, eingebunden werden kann.

Eine detaillierte Beschreibung der Architektur, der Komponenten, sowie der Schnittstellenspezifikation wurde fristgerecht im Deliverable für den Meilenstein *MS-1.4.C.2 Fertigstellung Basis-Extraktionsmechanismus für semistrukturierte Datenquellen* im Monat 18 dokumentiert.

AP 1.4.C.3: Entwicklung erster Demonstrator des Extraktionsmechanismus für semistrukturierte Datenquellen (M13-M21)

In diesem Arbeitsschwerpunkt wurden die gewonnenen Ergebnisse in einem ersten Demonstrator des Extraktionsmechanismus für semistrukturierte Datenquellen umgesetzt. Es wurden unterschiedliche Analyse-Module, Matching-Algorithmen und entsprechende Schnittstellen implementiert. Diese sind über eine REST-Schnittstelle als Web-Dienst verfügbar gemacht worden. Ziel des Demonstrators ist es, zu zeigen, wie die entwickelten Ansätze praktisch angewendet werden können und folglich zu einer Zuweisung von Datenobjekten zu semantischen RDF-Klassen führen. Zur besseren Demonstration der Dienste wurden zusätzlich eine visuelle Benutzeroberfläche im Form einer Webanwendung programmiert, welche es erlaubt, den Extraktionsmechanismus zu konfigurieren und zu initiieren. Weiterhin wurde ein Ergebnisevaluationswerkzeug umgesetzt, mit dem grafisch und textuell die extrahierten Daten exploriert, sowie statische Auswertungen analysiert werden können.

Die REST-Schnittstelle wurde unter Verwendung von standardisierten Ansätzen, etwa der RESTful API Modeling Language (RAML, <https://raml.org>), beschrieben und definiert. RAML basiert auf YAML und bietet umfassende Möglichkeiten RESTful-APIs zu spezifizieren, um damit die Schnittstellen für andere Anwendungen einfach anbindbar zu machen. Weiterhin ist die Weboberfläche des Prototyps weiterentwickelt und durch bessere Verlinkung von Extraktion und Evaluation nutzerfreundlicher gestaltet worden. Zum Ende des geplanten Arbeitspakets ist damit ein gut demonstrierbarer Ansatz zur Wissensextraktion aus semistrukturierten Daten geschaffen und das Ziel, zu zeigen, wie die entwickelten Ansätze praktisch angewendet werden können und folglich zu einer Zuweisung von Datenobjekten zu semantischen RDF-Klassen führen, erreicht worden. Durch die Einbindung des Demonstrators in die projektumfassende Orchestrationsumgebung wurde gewährleistet, dass die Ergebnisse aufwandsarm getestet und angewendet werden können. Dies ist durch den Einsatz des Docker-Container-Ansatzes partnerübergreifend umgesetzt.

AP 1.4.C.4: Test und Optimierung des Extraktionsmechanismus für semistrukturierte Datenquellen (M19-M27)

Arbeitspaket 1.4.C.4 beschäftigt sich mit der Optimierung des zuvor fertiggestellten, ersten Demonstrators zur Extraktion von Wissen aus semistrukturierten Datenquellen. Dies beinhaltet sowohl ausführliche Tests mit unterschiedlichsten Eingabedaten, die Evaluation der Extraktionsergebnisse, sowie die anschließende Verbesserung der umgesetzten Algorithmen. Die dabei festgestellten Mängel werden im weiteren Projektverlauf kontinuierlich ausgebessert und, wenn nötig, durch die Umsetzung von zusätzlicher Funktionalität bearbeitet. Als erste Ergebnisse der Optimierung konnte die Weboberfläche des Prototyps durch bessere Verlinkung von Extraktion und Evaluation nutzerfreundlicher gestaltet werden, in dem eine direkte Navigation von der Initialisierung der Extraktion zum Auswertungswerkzeug möglich gemacht wurde.

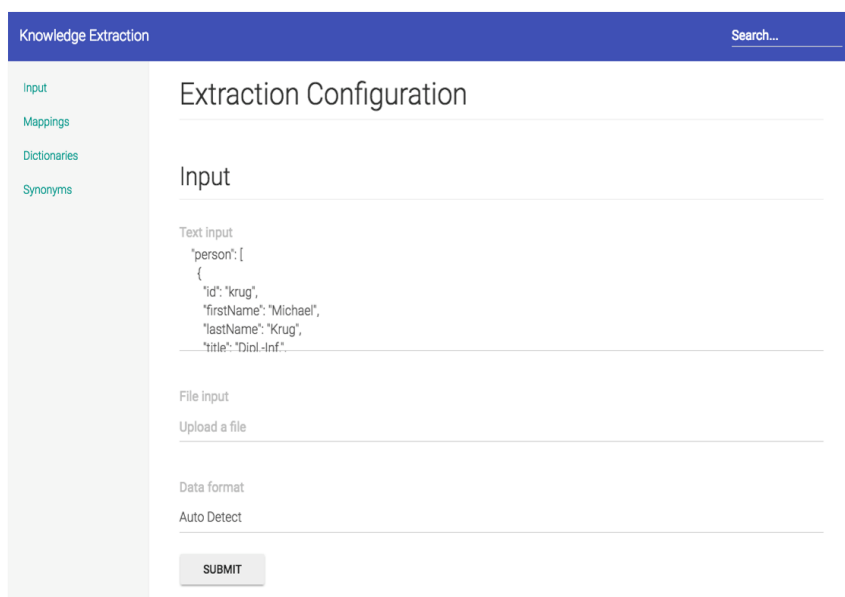
Zum Test der implementierten Funktionen wurde das Framework *Node Tap* eingesetzt und systematisch alle Einheiten validiert. Die dabei erkannten Probleme, v.a. bezüglich der Testbarkeit einzelner Funktionsteile, wurden im Rahmen des AP analysiert und behoben. Arbeitspaket 1.4.C.4 wurde planmäßig abgeschlossen und beinhaltet die Evaluation der umgesetzten Funktionalitäten und entsprechender Analysen und Optimierungen. Die zuvor eingeführte Testsuite wurde um die Abdeckung der neu implementierten Logik aus AP 1.4.C.5 erweitert.

AP 1.4.C.5: Erweiterung des Extraktionsmechanismus für semistrukturierte Datenquellen (M25-M38)

Im Arbeitspaket 1.4.C.5 wurde der stabile Extraktionsmechanismus für semistrukturierte Datenquellen schrittweise erweitert. Dazu wurden weitere Ontologien und Vokabulare eingebunden, um eine noch aussagekräftigere und vielseitiger einsetzbare Lösung zu erhalten. So ist die Integration von domänenspezifischen Ontologien wichtig, um spezielle Anwendungsfälle der Industrie bedienen zu können.

Neben der Einbindung und Erprobung neuer Vokabulare – beispielhaft wäre hier DBpedia zu nennen – wurden einzelne Analyse- und Mappingfunktionen in den Aspekten Performance und Accuracy verbessert. Durch das Zwischenspeichern von Teilergebnissen und Analyseresultaten konnte eine bessere Laufzeit erreicht werden. Die prototypische Plattform zum Management und Konfiguration des Extraktionsmechanismus ist umgesetzt, wird im Rahmen dieses Arbeitspakets noch finalisiert und abschließend vollständig dokumentiert.

Die Implementierung einer prototypischen Plattform zum Management und Konfiguration des Extraktionsmechanismus wurde finalisiert. Damit steht dem Anwender eine grafische Weboberfläche zur Verfügung, welche es erlaubt, zum einen den Extraktionsprozess zu Konfigurieren und zu Initialisieren und zum anderen ein umfassendes Werkzeug zur Analyse und Auswertung der Extraktionsergebnisse bietet. Weiterhin wurde der Extraktionsmechanismus selbst weiter optimiert und hinsichtlich Laufzeit und Genauigkeit verbessert.



The screenshot displays the 'Knowledge Extraction' web application interface. The main heading is 'Extraction Configuration'. On the left, there is a navigation menu with four items: 'Input', 'Mappings', 'Dictionaries', and 'Synonyms'. The 'Input' section is currently active. It contains three input fields: 'Text input' with a JSON example, 'File input' with an 'Upload a file' button, and 'Data format' with a dropdown menu set to 'Auto Detect'. A 'SUBMIT' button is located at the bottom of the configuration area. A search bar is visible in the top right corner of the application header.

Mappings

Select predefined mappings for
people

| Use | Keyword | Property | URI | Class |
|-------------------------------------|-----------|------------|------------------------------|--------------------------|
| <input checked="" type="checkbox"/> | firstName | givenName | http://schema.org/givenName | http://schema.org/Person |
| <input checked="" type="checkbox"/> | lastName | familyName | http://schema.org/familyName | http://schema.org/Person |
| <input type="checkbox"/> | + | Keyword | Property | URI |

Dictionaries

| Use | Type |
|-------------------------------------|----------|
| <input checked="" type="checkbox"/> | city |
| <input checked="" type="checkbox"/> | language |

Synonyms

| Use | Word | Synonyms |
|-------------------------------------|------------|--------------------------------------------------------------------|
| <input checked="" type="checkbox"/> | name | login |
| <input checked="" type="checkbox"/> | givenName | firstname, first_name, first name, given name, given_name, vorname |
| <input checked="" type="checkbox"/> | familyName | lastname, last_name, last name, family name, family_name, nachname |

AP 1.4.C.6: Abschlusstests, Integration und Dokumentation des Extraktionsmechanismus für semistrukturierte Datenquellen (M28-M40)

Zur Umsetzung der Integration der Extraktionskomponente in die LEDS-Technologieplattform wurde mit Docker-Compose ein Multi-Container-Ansatz gewählt. In dieser Orchestrierung sind alle entwickelten Bausteine als einzelne, ausführbare Container bereitgestellt und werden mittels einer speziellen Konfiguration miteinander verbunden. Dabei werden die bereitgestellten Schnittstellen untereinander angesprochen und damit ein ganzheitlicher Workflow gewährleistet.

Anschließend wurde im Rahmen dieses Arbeitspakets die finale Dokumentation bearbeitet und weitere Tests mit unterschiedlichen Eingabedaten vorgenommen. Ebenso wurden die in AP 1.4.C.5 vorgenommenen Erweiterungen wie in AP 1.4.C.4 in die Testumgebung aufgenommen und damit validiert.

Die zuvor umgesetzte Integrationslösung auf Basis von Docker-Compose wurde im Rahmen des Arbeitspakets 1.4.C.6 erweitert. Zusätzlich wurden weitere Tests mit unterschiedlichen Eingabedaten vorgenommen und die im AP 1.4.C.5 umgesetzten Erweiterungen in die Testumgebung aufgenommen und validiert.

Zuletzt wurde die finale Integration des Extraktionsmechanismus auf Basis von Docker-Compose umgesetzt. Die implementierten Erweiterungen wurden in die Testumgebung aufgenommen und validiert. Bei den durchgeführten umfassenden Tests der Lösung konnten letzte Fehler identifiziert und behoben werden. Abschließend wurde die gesamte, in ASP-C entwickelte Extraktionskomponente, inklusive der Plattform zum Management und Konfiguration, detailliert dokumentiert. Dies beinhaltet sowohl interne Arbeitsabläufe, externe Schnittstellen als auch Ergebnisse der Testdurchführungen.

Die Bereitstellung der Extraktionskomponente für semistrukturierte Datenquellen erfolgte unter <https://docker-registry.eccenca.com/tuc-keseda>

AP 1.4.D.1: Spezifikation (Anforderungs- und Funktionsdefinition) der Social-aware Suche in Linked Data Quellen (M1-M9)

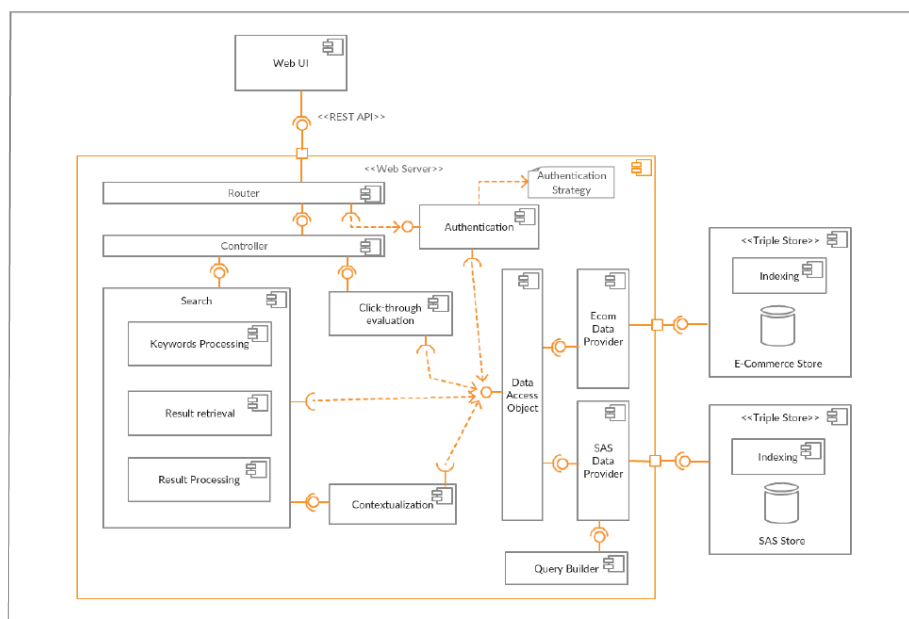
Im Arbeitspaket 1.4.D.1 wurde im ersten Projekthalbjahr an der Definition der Social-aware Suche gearbeitet. Es wurden Funktionsdefinitionen und Anwendungsszenarien erstellt, welche anhand einiger Use-Cases verdeutlicht wurden. Diese Ausarbeitungen sollen die Anwendbarkeit einer solchen Suche, vor allem im E-Commerce Sektor, verdeutlichen. Ferner stellen die Szenarien eine Diskussionsgrundlage zwischen den Projektpartnern dar, um die gemeinsame Zielsetzung zu festigen. Ebenso dienen die Anwendungsfälle dazu, Anforderungen an eine Suche in Wissensbasen mit sozialem Kontext zu definieren. Diese Analyse ist zum Zeitpunkt des Berichtes noch nicht abgeschlossen.

Im zweiten Projekthalbjahr wurde die Spezifikation der Social-aware Suche finalisiert. Der Meilenstein *MS-1.4.D.1* wurde erreicht und hierzu ein Deliverable erstellt, welches alle Erkenntnisse aus unseren bisherigen Untersuchungen enthält. Darin wird zunächst ein Überblick gegeben, welche verschiedenen Sucharten es gibt und welche im LEDES-Anwendungskontext in Frage kommen.

Mit dem Arbeitsschwerpunkt der TU Chemnitz, eine Komponente für eine *Social-aware Suche* umzusetzen, wurde hier im Ergebnis insbesondere eine personalisierte und kontextualisierte Schlüsselwortsuche herausgearbeitet und in ein erstes mögliches Konzept einer Komponentenarchitektur überführt.

AP 1.4.D.2: Schnittstellen und Architekturprototyp der Social-aware Suche in Linked Data Quellen (M7-M15)

Passend zu der im *AP 1.4.D.1* erarbeiteten Anforderungs- und Funktionsdefinition wurden eine Reihe möglicher Einsatzszenarien in einem ersten umfassenden Use Case konkretisiert und mit den Projektpartnern diskutiert. In der Folge wurde eine Recherche nach geeigneten Testdaten aus dem E-Commerce-Umfeld durchgeführt, welche sich jedoch schwierig gestaltete. Zumeist lagen derartige Daten noch nicht in Linked Data-Form vor, weswegen das Nutzerszenario für den Architekturprototypen auf die semantische Suche in einem beliebigen fachspezifischen Datenbestand erweitert wurde. Darauf aufbauend wurden existierende Ontologien untersucht und miteinander verglichen, welche die semantische Ablage von Suchinformationen und personalisierten Kontexten erlauben. Im Ergebnis wurden momentan zwei Vokabulare von schema.org und sindice.com als geeignet identifiziert, die in der Folge näher bewertet werden müssen. Außerdem wurden die Anforderungen an den Architekturprototypen in einem weiteren Iterationsschritt geschärft, sodass dieser nun den zuvor definierten Use Case komplett abbilden kann.



Im Arbeitsschwerpunkt 1.4.D.2 wurde aufbauend auf den in AP 1.4.D.1 erarbeiteten Use Cases, sowie Funktions- und Anforderungsdefinitionen ein initialer Schnittstellen- und Architekturprototyp entwickelt. Konzeptionell soll dieser Architekturprototyp in der Lage sein, Suchoperationen auf einem semantischen Datenbestand zuzulassen und dabei gleichzeitig Hintergrundwissen über den aktuellen Nutzer einzubeziehen, um ihm adaptiv eine Ergebnisliste mit den für ihn passendsten Suchergebnissen bereitzustellen. Um einen Bezug zu klassischen E-Commerce-Anwendungen zu halten, wurde ein Keyword-basierter Suchansatz als Grundlage gewählt und mit Technologien und Ansätzen des Semantic Webs erweitert. Der Architekturprototyp wurde in mehrere Komponenten unterteilt und dokumentiert, welche unter anderem den Datenabruf aus sozialen Netzwerken, die Kontextualisierung der Sucheingabe oder das adaptive Result Ranking regeln. Neben standardisierten Web-Schnittstellen können verschiedene Triple-Stores angebunden werden und Daten zur Suche bereitstellen, als auch persistent Metadaten zu persönlichen und sozialen Attributen der Suchakteure verwalten und zueinander in Beziehung setzen. Alle Daten werden dabei homogen als Linked Data repräsentiert, weswegen der Prototyp unabhängig von einem konkreten Social Web Platform Provider und Suchdatenbestand interagieren kann und ebenso in der Lage ist, uniforme Netzwerkbeziehungen über verschiedenartige soziale Profile herzustellen, sofern diese geeignete Informationen für eine Social-aware Suche bereitstellen.

AP 1.4.D.3: Entwicklung erster Demonstrator der Social-aware Suche in Linked Data Quellen (M13-M21)

Als Vorarbeit für die Entwicklung eines ersten Demonstrators für die *Social-aware Suche* wurde zwischenzeitlich bereits ein initialer Prototyp umgesetzt, um die generelle Möglichkeit des Abrufs personenbezogener Daten aus sozialen Netzwerken am Beispiel von Facebook zu eruieren. Dieser Test war positiv und floss in die Implementierung der *Social-aware Search Component* unmittelbar ein.

Als Beispiel-Datenbestand wurden im ersten Schritt bibliographische Daten wissenschaftlicher Publikationen benutzt, da diese zum einen eine Vielzahl strukturierter Meta-Informationen bereitstellen, als auch Experten- und Domänenabhängig ähnliche Suchverhalten erwarten lassen. Die Referenzarchitektur des Prototyps der Social-aware Suche umfasst eine Implementierung basierend auf NodeJS mit Anbindung einer Stardog Graph-Datenbank sowie Komponenten zur Authentifizierung und zum Zugriff auf die Profil- und Netzwerkdaten des aktuellen Nutzers in sozialen Netzwerken am Beispiel des Facebook OpenGraphs. Zusätzlich wurde zu Demonstrationszwecken synthetisch ein umfangreicheres soziales Freudenetzwerk mit Relationen zwischen verschiedenen gewichteten Freundesklassen generiert und darauf ein erster Algorithmus implementiert, der anhand des aktuellen Suchkontexts, dem bisherigen Klickinteresse auf spezifische Suchergebnisobjekte, sowie demographischer Nutzercluster ein adaptives Search Ranking vornimmt, welches im Vergleich zu einem klassischen Ergebnisranking für den aktuellen Nutzer relevantere Suchergebnisse zuerst aufführt. Eine zusätzliche Visualisierungskomponente erlaubt zu Forschungszwecken die Nachvollziehbarkeit der Suchergebnisreihenfolge und soll in weiterführenden wissenschaftlichen Publikationen die Thematik anschaulicher und diskutierbar machen. Der Meilenstein *MS-1.4.D.2 Fertigstellung Basis-Komponente für Social-aware Suche in Linked Data Quellen* wurde planmäßig erreicht.

Aufbauend auf Beispiel-Datensätzen als Linked Data war es nun möglich zu zeigen, wie soziale Eigenschaften im Rahmen einer Suche zur Verbesserung der Ergebnisqualität angewendet werden können. Anhand eines Suchkontexts, dem bisherigen Klickinteresse auf bestimmte Suchergebnisobjekte, sowie demographischer Informationen wird ein adaptives Search Ranking vorgenommen, welches im Vergleich zu einer klassischen Suchergebnissortierung für den aktuellen Nutzer relevantere Suchergebnisse zuerst aufführt. Mithilfe verschiedener, auch interaktiver Visualisierungsansätze wird durch eine Graphkomponente zu Analysezielen außerdem dargestellt, welche Faktoren die Suche konkret beeinflussen. Abschließend wurde die Suchdienst-Komponente aus dem ASP-D ebenfalls als *Docker* Image der LED-Technologieplattform zur Verfügung gestellt und über einen Continuous Integration Workflow bei Änderungen automatisch aktualisiert.

AP 1.4.D.4: Test und Optimierung der Social-aware Suche in Linked Data Quellen (M19-M27)

Hier wurde die Implementierung des Social-aware Suchdienstes zur Suche in Linked Data Quellen evaluiert und optimiert. Regulatorische Grenzen machten die Weiterentwicklung schwer, welche in der Praxis die Nutzung personenbezogener Daten und sozialer Netzwerkbeziehungen zu Applikationszwecken von Datenanbieterseite reglementieren. Aus konzeptioneller Sicht hat dies auf die Social-aware Search Komponente keinen Einfluss, da diese auf beliebigen sozialen Netzwerk-Anbieterdaten arbeiten kann, allerdings verlängerte dies die Entwicklungsarbeit und war zu Projektbeginn als Trend nicht absehbar. Erkenntnisse aus der Entwicklung des Demonstrators wurden in der Veröffentlichung „*Utilizing Linked Data Structures for Social-aware Search Applications*“ publiziert.

Die Suchdienstkomponente wurde kontinuierlich verbessert. Zunächst wurde für die im weiteren Verlauf geplanten funktionellen Erweiterungen der Prototyp technisch überarbeitet (mittels Refactoring-Methoden) und ein Test-Framework (*Mocha*) zur Sicherstellung der korrekten Funktionsweise der Suchkomponente integriert. Außerdem wurde Suchdienstkomponente an die zentrale Benutzer- und Konfigurationsverwaltung der brox/eccenca angebunden. Nach erstem Nutzerfeedback wurde außerdem die Darstellung der Knoten innerhalb der Visualisierungskomponente verbessert. Zuletzt fokussierten wir uns auf den Einsatz der Social-aware Search Komponente im E-Commerce-Umfeld und befassten uns hierfür mit geeigneten, existierenden Produktdaten. Das Angebot an verfügbaren Shopdaten in einem bereits vorliegenden Linked Data Format war unerwartet wenig.

Die Optimierung und Evaluation des Social-aware Suchdienstes zur Suche in Linked Data Quellen erfolgte zu Großteilen bereits im ersten Projekthalbjahr 2017. Die Ergebnisse der durchgeführten Tests wurden abschließend noch einmal überprüft. Dabei führten fehlgeschlagene Tests zu einer konsequenten Überarbeitung der Softwarekomponente zur Sicherstellung der korrekten Funktionsweise. Besonders die Implementierung der verwendeten Distanzmetriken konnte noch einmal verbessert werden.

AP 1.4.D.5: Erweiterung der Social-aware Suche in Linked Data Quellen (M25-M33)

Im zweiten Projekthalbjahr 2017 wurde eine Reihe an Erweiterungen an der Social-aware Search Komponente durchgeführt. Zunächst erfolgte ein umfassendes Refactoring des Quellcodes der ersten Version der Softwarekomponente, um diesen besser wartbar und weiterentwickelbar zu gestalten. Es wurde die Bereitstellung als Docker-Container optimiert und die Version zugrundeliegender Softwarekomponenten aktualisiert. Die Überprüfung und Bereitstellung von Demodaten für die Social-aware Suche wurden direkt integrierten in den Softwareprototypen integriert. Zusätzlich wurde zu Demonstrationszwecken eine alternative, von der Facebook-Plattform unabhängige Anmeldemethode implementiert. Im Weiteren wurde die Social-aware Such-Komponente um ein Assistenz-System zum Vorschlagen geeigneter Suchbegriffe (Keyword Auto-Suggestion) umgesetzt. Im Gegensatz zur klassischen Herangehensweise von Suchmaschinen, einem Nutzer nach Eingabe der ersten Buchstaben eines Schlüsselworts auf statistischer Basis Suchbegriffe vorzuschlagen, die in der Vergangenheit am meisten von anderen Nutzern gesucht wurden, wendeten wir unseren Social-aware Approach auch auf den Vorschlag potentiell interessanter Schlüsselwörter an und demonstrierten damit einen weiteren sinnvollen Einsatz einer Social-aware Suche in Linked Data Beständen. Die Implementation der Funktion erfolgte zum Jahresende 2017.

SOCIAL AWARE SEARCH

Social Context ON

social

André Langer | Log Out

Search Results

| # | Publication Name | Description |
|---|------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | Ranking Web Services using Centralities and Social Indicators | Nowadays, developers of web application mashups face a sheer overwhelming variety and pluralism of web services. Therefore, choosing appropriate web s... |
| 2 | WebID+ACO: A distributed identification mechanism for social web | This paper defines an approach to managing digital identity requiring special-purpose technology on the browser client. We propose a mechanism using s... |
| 3 | The Social Routing Principle | ... |
| 4 | Scope-Aware Delegations in Distributed Social Networks | Swiftly meeting challenges by dist... |

Search Result Score

Social clicks: 4
Anonymous clicks: 0
Agents: 4
Gender Matches: 2
Total Score: 4.87

Social Graph

Die Arbeiten an der Social-aware Search Komponente im Jahr 2018 konzentrierten sich zu Jahresbeginn darauf, alle bereits in der zweiten Jahreshälfte 2017 vorbereiteten Funktionalitäten fertig zu implementieren. Im Speziellen umfasste dies die Anwendung des Social-aware Search Konzepts auf einen weiteren Use Case in Form einer Keyword Recommendation (Auto Suggestion) noch während der Nutzer eine Sucheingabe tätigt. Die implementierte Funktionalität wurde getestet und funktionierte ordnungsgemäß wie erwartet. Daneben wurden weitere Optimierungen und Fehlerbehebungen an der prototypischen Implementierung der Softwarekomponente durchgeführt (Multi-Session-Support). Das Management und die Konfiguration der Social-aware Suchkomponente wurde im Rahmen des E-Commerce-Use Cases mit anderen Projektpartnern besprochen. Die Konfiguration der Gewichtungsparmeter kann dabei direkt in dem Suchprototypen vorgenommen werden. Für die spätere praktische Weiternutzung stellt zunehmend die veränderte Geschäftspolitik von Social Media Plattformdaten-Anbietern wie Facebook eine Hürde da, nur noch in eingeschränktem Maße den Zugriff auf nutzerspezifische Daten zu erlauben. Hinzu kam ein administrativer Reviewprozess ab dem Sommer 2018, der nur noch eingeschränkt die Weiternutzung des entwickelten Demonstrators erlaubte. Nichtsdestotrotz konnten wir innerhalb des LEDES-Projekts die Machbarkeit und den Mehrwert einer Social-aware Suchlösung in Linked-Data-Datenbeständen demonstrieren und AP 1.4.D.5 planmäßig und erfolgreich abschließen.

AP 1.4.D.6: Abschlusstests, Integration und Dokumentation der Social-aware Suche in Linked-Data-Quellen (M28-M41)

Analog zu den in den anderen Arbeitspaketen entwickelten Softwarekomponenten wurde auch die Social-aware Suchkomponente frühzeitig als Docker-Container für die LEDES-Technologieplattform zur Verfügung gestellt. Parallel dazu erfolgte eine Überarbeitung und Optimierung der Orchestrierung und der Präsentationswebsite. Eine Anbindung des Datenabrufs von Linked Data über die projektspezifische *DataPlatform* erfolgte ähnlich wie bei der Qualitätssicherungskomponente ebenso bereits, so dass ein uniformer Datenzugriff gegeben ist.

In der ersten Jahreshälfte 2018 konzentrierten sich die Aktivitäten auf die Konzeption konkreter Testszenarien sowie die Durchführung und Auswertung der Software-Tests. Probleme traten lediglich bei der Authentifizierung über den zentralen OAuth2-Dienst der LEDES-Technologieplattform zum Zugriff auf semantische Metadaten der Softwarekomponente auf, die in Rücksprache mit den Verbundpartnern rasch identifiziert und behoben werden konnten. Zur Evaluation der korrekten Funktion aller Module wurden 201 verschiedene Tests durchgeführt. Nach Abschluss der Arbeiten im AP 1.4.D.5 erfolgte der Abschlusstest und die finale Dokumentation der Social-aware Suchkomponente entsprechend des Arbeitsplans. Der Meilenstein *MS-1.D.3 "Fertigstellung des finalen Demonstrators mit integrierter Social-Aware- Suche"* wurde erreicht.

Die Bereitstellung der Social-aware Such-Komponente erfolgte unter <https://docker-registry.eccenca.com/tuc-sasearch>.

3. DIE FORTSCHREIBUNG DES VERWERTUNGSPLANS.

Erfindungen / Schutzrechtsanmeldungen und erteilte Schutzrechte

Im Berichtszeitraum wurden im Rahmen des Projektes keine Schutzrechte angemeldet.

Wirtschaftliche Erfolgsaussichten nach Projektende

Im Vergleich zum Verwertungsplan hatten sich keine Änderungen im Hinblick auf die wirtschaftlichen Erfolgsaussichten nach Projektende ergeben.

Wissenschaftliche und/oder technische Erfolgsaussichten nach Projektende

Im Hinblick auf die wissenschaftlichen und technischen Erfolgsaussichten nach Projektende zeigten sich frühzeitig erste Erfolge in Lehre und Forschung. Das Thema Linked Data for Enterprises und Semantic Web Data Management war auf verschiedenen internationalen Konferenzen nachgefragt und es ist zu erwarten, dass dazu auch in den Folgejahren weitere Forschungsaufgaben gelöst werden müssen und publiziert werden. Daher wurde die Thematik *Semantic Web* und *Linked Data* frühzeitig in das universitäre Ausbildungsprogramm der Professur Verteilte und Selbstorganisierende Rechnersysteme an der TU Chemnitz integriert. Weiterhin zeigten Studierende der TU Chemnitz deutliches Interesse an diesem Problemfeld, weswegen in der Folgezeit speziell in diesem Bereich durchgeführte Abschlussarbeiten und für die Wirtschaft spezifisch geschulte Absolventen zunehmen werden.

Mittels studentischer Abschluss- und Projektarbeiten konnte die Thematik auch erfolgreich an den wissenschaftlichen Nachwuchs transportiert werden, um somit auch die zukünftigen Fachkräfte für den Bedarf an Expertenwissen im Bereich der semantischen Technologien in der Wirtschaft vorzubereiten. Thematisch beschäftigten sich die Arbeiten unter anderem mit der Integration klassischer Daten entlang von Wertschöpfungsketten in Industrieumgebungen mittels Linked Data und dem Einfluss von Ranking Faktoren auf Suchverhalten im Semantic Web Umfeld, die Erstellung eines Scoring Models, die Analyse natürlichsprachiger Nachrichtenmeldungen auf Glaubwürdigkeit oder auch die Linked Data gestützte Analyse von Quelltexten. Ebenso wird die Thematik Semantic Web kontinuierlich im Bereich der Lehre nachgefragt und erzeugt permanentes Interesse. Forschungsbezogen waren wir mit den Eigenentwicklungen aus dem LEDS-Projekt in wiederkehrenden Kontakt mit den führenden Arbeitsgruppen im Semantic Web – Umfeld und auf den etablierten Konferenzen der Semantic Web und Web Engineering Community vertreten. Projektseitig profitierten auch andere Projekte der Professur aus den Erfahrungen des LEDS-Teams, bspw. in einem *eHealth Research Lab* oder in neuen Projektanträgen für zukünftige Projekte mit dem BMBF oder in H2020.

Besonders zeigte sich dies im Rahmen unseres Workshops *LEDSPLaY - Linked Enterprise Data Services, Provenance, Linking and Quality* auf der Konferenz *INFORMATIK 2017* Ende September 2017 in Chemnitz. Die *INFORMATIK* ist die Expertenveranstaltung der größten Informatikvereinigung im deutschsprachigen Raum (GI e.V.) und lieferte 13 qualitativ hochwertige wissenschaftliche Beiträge und eine fachlich gewinnbringende Diskussion der bisherigen Ergebnisse und möglicher Zukunftsperspektiven.

Im Forschungsbereich war die Thematik zunehmend nachgefragt. Semantic Web Tracks und Panels finden sich inzwischen auf allen renommierten Konferenzen weltweit. Industriepartner fragten vor allem die Integration von Semantic Web Technologien in ihren Wertschöpfungsketten nach, um Daten aus heterogenen Datenbeständen sinnvoll miteinander verknüpfen zu können. Darüber hinaus sind taxonomische Klassifikationsmöglichkeiten von Daten und Technologien stetig nachgefragt, als auch die Möglichkeit mit standardisierten Vokabularen (Ontologien) Beziehungen auszudrücken und Daten zwischen verschiedenen Systemen auszutauschen. Der Technologie-Stack des LEDS Projekts kann hierfür inzwischen bereits eine geeignete Lösung präsentieren.

Auch hat die LEDS Abschlussveranstaltung die Relevanz und die Notwendigkeit weiterer Forschungs- und Entwicklungsarbeiten im Rahmen des Semantic Webs aufgezeigt.

Wissenschaftliche und wirtschaftliche Anschlussfähigkeit

Im Vergleich zum Verwertungsplan haben sich keine Änderungen im Hinblick auf die wissenschaftliche und wirtschaftliche Anschlussfähigkeit nach Projektende ergeben.

Aus Sicht der TU Chemnitz wird die wissenschaftliche Anschlussfähigkeit als sehr gut eingeschätzt. Mit der Datenqualitätssicherungskomponente SemQuire wurde eine Grundlage geschaffen, von der auch zukünftige Projektaktivitäten stark profitieren können. Die Untersuchung von Datenqualitäts- und Kurationskriterien ist nach wie vor ein hochaktuelles Thema, was Projektausschreibungen wie die BMBF Bekanntmachung 1791 Richtlinie zur Förderung von Forschungsvorhaben zur Entwicklung und Erprobung von Kurationskriterien und Qualitätsstandards von Forschungsdaten im Zuge des digitalen Wandels im deutschen Wissenschaftssystem. Bundesanzeiger vom 13.06.2018 zeigen. Ebenso hat das systematische Management und der Umgang mit projektbezogenen Daten eine stetig zunehmende Bedeutung in Datenmanagementplänen als Anforderung vieler Projektausschreibungen auf nationaler und internationaler Ebene. Aus dem LEDS-Projekt heraus haben sich hierzu bereits zwei unmittelbare Dissertationsprojekte ergeben, welche sich mit dem Linked Data basierten Management von Forschungsdaten und der Forschung an Trust-Aspekten im Umgang mit Informationen beschäftigen. Weiterhin ergaben sich aus dem LEDS-Projekt heraus nachhaltig Kontakte zu anderen Forschern weltweit im Bereich des Datenqualitätsmanagements, in Initiativen zur Etablierung offener Knowledge Graphen für Forschungsaktivitäten, oder auch in Anwendungsbereichen wie dem Web of Things. Hier ist über das LEDS-Projekt hinaus innerhalb der kommenden drei Jahre eine weiterführende, intensive Zusammenarbeit angedacht. Besonders erfreulich sind dabei aktuelle Bestrebungen der TU Chemnitz zur Weiterarbeit in einem Informationsinfrastrukturprojekt in einem Sonderforschungsbereich innerhalb der kommenden 5 Jahre.

Auch von Industriepartnern werden zunehmend neue Methoden und Ansätze im Umgang mit Linked Data angefragt und die Anschlussfähigkeit als sehr gut eingeschätzt. So haben sich im Rahmen des Projektendes des LEDS-Verbundprojekts Anfragen aus der Industrie ergeben, welche Unterstützung im semantischen Datenmanagement für ein ad-hoc business reporting, für neue Mobilitätskonzepte im Nah- und Fernverkehr oder zur Entwicklung einer Sprachapplikation zur Jobsuche auf Linked Data Beständen erbitten. In einem ersten Schritt werden hierzu im kommenden Jahr studentische Abschlussarbeiten betreut werden. Parallel dazu werden weiterführende Projekte innerhalb der nächsten zwei Jahre etabliert.

Auf die entwickelten Methoden und Softwarekomponenten wurden seitens der TU Chemnitz aus wissenschaftlicher Sicht keine Schutzrechte angemeldet.

Zu den Verbundpartnern des LEDS-Projekts besteht auch nach Projektende weiterhin ein sehr guter Kontakt. Die eccenca GmbH setzt kontinuierlich die Etablierung der LEDS-Projektergebnisse in einem tragfähigen Geschäftsmodell um. Eine zukünftige weitergehende Zusammenarbeit innerhalb der kommenden drei Jahre ist angedacht.

Entsprechend den Bedingungen des Wachstumskernprojekts wird die TU Chemnitz den Verwertungsplan kontinuierlich fortschreiben und die Forschungsergebnisse aus dem LEDS Projekt nachhaltig in der Lehrausbildung in Chemnitz und zukünftigen Forschungsprojekten integrieren und weiter ausbauen. Absolventen der Informatikstudiengänge an der TU Chemnitz werden diese Expertise nutzen können, um technisch hochqualifiziert als potentielle Nachwuchskraft in den Arbeitsmarkt der Region Westsachsen einsteigen zu können, entweder industriell in groß- und mittelständigen Unternehmen als Senior Developer, Data Engineer oder Projektmanager, oder auch als Nachwuchswissenschaftler im Fachbereich Web- und Data Engineering.

Für weitere Details verweisen wir auch auf die ausgefüllte Tabelle zur Verwertung.

4. ARBEITEN, DIE ZU KEINER LÖSUNG GEFÜHRT HABEN,

Alle in der Vorhabenbeschreibung beschriebenen Meilensteine und Ergebnisse wurden, wie in III.2 beschrieben, planmäßig erreicht.

Bei der Datenqualitätssicherungskomponente SemQuire wurde eine Auswahl von 55 intrinsischen, kontextuellen, repräsentativen und zugänglichkeitsprüfenden Metriken für Linked Data Quellen realisiert. Die Umsetzung weiterer Metriken ist weiterhin Forschungsgegenstand und über das LEDS-Projekt hinaus im Gange. Dies umfasst besonders Metriken, welche Techniken des Data Minings oder des maschinellen Lernens (AI) einsetzen und auf großen Datenbeständen operieren. Weiterhin wurden Metriken für einen sehr spezifischen Anwendungsfokus von unseren Untersuchungen ausgeklammert, ebenso Implementierungen für die es bereits einen ausreichend guten Toolsupport gibt, die mit überschaubarem Aufwand in den Technologiestack integriert werden könnten.

Auch der Extraktionsmechanismus für semistrukturierte Daten KESeDa lieferte eine prototypische Lösung. Weitergehende Forschungsaktivitäten waren auch hier angedacht, um beispielsweise Linked Data auch als csv-Datenbeständen extrahieren zu können. Jedoch fanden sich hierzu bereits existierende Tools zur Konvertierung von tabellarischen Daten in eine RDF-Serialisierung.

Die Arbeiten an der Social-aware Suchkomponente wurden durch Data Privacy Geschehnissen auf sozialen Netzwerk-Plattformen im Jahr 2018 stark beschränkt. Während die Personalisierung des E-Commerces zum Zeitpunkt der Vorhabenbeantragung ein Thema mit hohem Zukunftspotential war, wurde der Zugriff auf Hintergrundinformationen zu einem Nutzer und dessen sozialen Umfeld in der Folge immer stärker durch Plattformbetreiber wie Facebook eingeschränkt. Dennoch konnten die Arbeit einer prototypischen Umsetzung einer Social-aware Suchkomponente auf Linked Data Beständen und die Integration in den LEDS-Technologiestack wie vorgesehen abgeschlossen werden. Der praktische Einsatz in Webshops realer Kunden unseres Verbundpartners konnte jedoch aufgrund vertraglicher Rahmengenzen nicht erfolgen. Ebenso konnten uns während der LEDS-Projektlaufzeit in nicht ausreichendem Maße E-Commerce-Datenbestände bereitgestellt werden, sodass die Social-aware Suchkomponente nur auf Demodaten evaluiert werden konnte.

5. PRÄSENTATIONSMÖGLICHKEITEN FÜR MÖGLICHE NUTZER

Bereits innerhalb des LEDS-Projekts fanden neben monatlichen Telefonkonferenzen zur Abstimmung unter den Projektpartnern auch mehrere Workshops zu verschiedenen Themen statt. So wurde zu Projektbeginn beispielsweise frühzeitig ein Treffen zwischen der TU Chemnitz und Universität Leipzig organisiert, bei dem die bisherigen und aktuellen Forschungsprojekte, Ansätze und Werkzeuge vorgestellt wurden.

Weiterhin haben wissenschaftliche Mitarbeiter der TU Chemnitz mehrere nationale und internationale Konferenzen besucht und in deren Kontext Ergebnisse aus dem LEDS Projekt diskutiert und publiziert. Vorbereitend begann dies bereits im Dezember 2015 mit der internationalen Konferenz iiWAS'15 (The 17th International Conference on Information Integration and Web-based Applications & Services) in Brüssel. Dort wurden zwei Publikationen präsentiert, die darüber hinaus mit Experten des Fachgebiets intensiv diskutiert werden konnten. Dies waren zum einen "SmartComposition: Bringing Component-Based Software Engineering to the Web" sowie "Systematic Composition of Web-based Applications with Focus on Security". Die Arbeiten befassen sich unter anderem mit der komponentenbasierten Entwicklung von Webanwendungen und der entsprechenden Kommunikationsinfrastruktur, sowie der Unterstützung des Designs von webbasierten Anwendungen mit speziellem Fokus auf Sicherheit. Im Rahmen des LEDS Projekts waren diese Entwicklungen vor allem im Bereich der Komposition von Linked-Data-orientierten Webanwendungen, sowie bei der Absicherung dieser in Bezug auf Autorisierung und Authentifizierung von Bedeutung – etwa mit dem auf Linked-Data-orientierten Identitätsmanagement sowie Autorisierungs- und Authentifikationsansatz WebID vom W3C wichtig.

Zur Vorstellung der ersten erarbeiteten Forschungsergebnisse wurde im Juni 2016 die Konferenz ICWE 2016 (The 16th International Conference on Web Engineering) in Lugano, Schweiz, besucht. Im Rahmen dieser internationalen Konferenz wurden sowohl die Publikationen „*AttributeLinking: Exploiting Attributes for Inter-Component Communication*“ und „*The SmartComposition Approach for Creating Environment-Aware Multi-Screen Mashups*“, als auch weitere Ergebnisse der Forschungsgruppe VSR vorgestellt und mit Experten des Fachgebiets intensiv diskutiert. Hierbei wurde auch an der zweiten internationalen *Mashup Challenge* teilgenommen. In diesem Wettbewerb konnten Teilnehmer aus der ganzen Welt ihre Entwicklungen im Rahmen der Kompositionstechnologien vorstellen und gegenseitig bewerten. Dabei konnten wir mit unserem Beitrag „*The SmartComposition Approach for Creating Environment-Aware Multi-Screen Mashups*“ die besten Bewertungen erreichen und damit den ersten Platz in diesem internationalen Vergleich belegen, welcher mit einer Urkunde am Ende der Konferenz ausgezeichnet wurde.

Daneben entstanden zwei weitere wissenschaftliche Publikationen mit den Titeln „*KESeDa - Knowledge Extraction from Heterogeneous Semi-Structured Data Sources*“ sowie „*Towards Reputation-Aware Expert Finding with Linked Open Data*“, welche öffentlichkeitswirksam auf der SEMANTiCS Konferenz in Leipzig Mitte September vorgestellt werden. Zusätzlich präsentierten wir auf der International Conference for WWW/Internet (ICWI) in Mannheim vom 28.-30. Oktober 2016 unsere ersten Ergebnisse zur Formalisierung von Datenqualität durch die Vorstellung des Fullpapers „*FAME.Q – A Formal Approach to Master Quality in Enterprise Linked Data*“.

Einen zentralen Meilenstein stellte dabei die erste öffentlichkeitswirksame Veranstaltung des LEDS-Projekts im Rahmen der 13. internationalen Konferenz für semantische Systeme (SEMANTiCS 2016) dar. Zwischen dem 12. und 15. September 2016 wurden in Leipzig verschiedene Workshops, Tutorials und Sessions mit zahlreichen Vorträgen rund um das Thema Semantische Systeme durchgeführt, an denen ca. 280 Forscher, Experten und Anwender aus aller Welt teilnahmen. Der Wachstumskern war durch einen betreuten Stand und einen eigenen Track vertreten. Im Zuge der Veranstaltung wurden neben der Vorstellung des Wachstumskerns auch mehrere wissenschaftliche Publikationen präsentiert und diskutiert.

Einen Schwerpunkt im Jahr 2017 nahm die Durchführung der Wissenschaftskonferenz INFORMATIK 2017 ein, in der durch unser Team ein eigener, öffentlichkeitswirksamer Workshop LEDSPLaY ausgerichtet wurde, auf dem wir aktuelle Forschungsergebnisse über „*Utilizing Linked Data Structures for Social-aware Search Applications*“ einem breiten Fachpublikum präsentierten und gleichzeitig den Austausch zu weiteren Themen aus dem Fachbereich Linked Enterprise Data Services, Provenance, Linking und Quality förderten.

Im April 2018 wurde das letzte Projekttreffen mit Strategie-Workshop an der Universität Leipzig durchgeführt, bei dem aktuelle Entwicklungen im Markt sowie im Forschungsumfeld vorgestellt wurden. Zusätzlich wurden mit allen anwesenden Projektpartnern neue Forschungsergebnisse besprochen und der strategische Einfluss des LEDS-Projekts auf unsere zukünftige Lehrausbildung und Forschungstätigkeit skizziert.

Zur optimalen Kommunikation und Präsentation der erreichten Vorhabenziele wurde in Zusammenarbeit mit dem Leipziger Videoproduzenten Uwe Willmann eine Reihe von vier kurzen Filmclips erstellt, welche - unterstützt durch Interviews und Screencasts der entwickelten Tools - die folgenden Schwerpunkte zeigen u: 1. Angewandte Methodiken im LEDS-Projekt, 2. E-Commerce Showcases, 3. (Linked) Open Data zum Anfassen, 4. Open Data & Open E-Government. Die Videos wurden öffentlich auf der Projekt-Webseite sowie auf verschiedenen Medienplattformen zur Verfügung gestellt und als öffentlichkeitswirksames und besonders zugängliches Werbemittel eingesetzt.

Vom 5. bis 8. Juni 2018 fand in Cáceres, Spanien, die 18. International Conference on Web Engineering (ICWE2018) statt, an der Mitarbeiter der TU Chemnitz zwei wissenschaftliche Publikationen vorstellen und diskutieren konnten: zuerst *“DaQAR - An ontology for the uniform exchange of comparable Linked Data Quality Assessment Requirements”* im Semantic Web Research Track der Konferenz, und zusätzlich auf dem International Workshop on Engineering Open Data (WEOD) der Beitrag *“SemQuire - Assessing the Data Quality of Linked Open Data Sources”*. Zusätzlich konnte die Konferenz genutzt werden, um sich mit anderen Forschungsgruppen besser zu vernetzen, deren Forschungsarbeit sich ebenfalls auf Anwendungen des Semantic Webs konzentriert. So entstanden neue Kontakte zur Gruppe von Dr. Javier Izquierdo der Universität Katalonien, zu Dr. Gabriela Bosetti von der Universität La Plata und es wurden weiterführende Gespräche mit Prof. Dr. Jens Lehmann geführt.

Außerdem fand am 18.06.2018 mit dem „*Leipziger Semantic Web Tag*“ in der Hainstraße in Leipzig die offizielle Projektabschlussveranstaltung statt. Im Rahmen eines ganztägigen Vortags- und Workshop-Programms konnten die Projektergebnisse vielen interessierten Besuchern aus Wirtschaft und Wissenschaft vorgestellt werden und weiterführende Arbeiten besprechen (<https://www.leds-projekt.de/lswt2018.html>).

Gegen Ende des Projektes wurde eine weitere Publikation F.I.E.L.D.S. - Analyzing Form Input interfaces for Explicit Linked Data handling in document Submission systems auf der 7th International Conference WWW/Internet (ICWI2018), Budapest, Ungarn publiziert und anschließend in einer erweiterten Version in einem Journal article im IADIS International Journal on Computer Science and Information Systems, Vol. 13, No. 2 publiziert. Seitdem wird die Referenz auf das LEDS-Projekt auch in aktuelle und zukünftige Präsentationen unserer Professuraktivitäten integriert.

Zusätzlich wurden öffentlichkeitswirksam alle LEDS-projektbezogenen Informationen auf der Website <https://www.leds-projekt.de> veröffentlicht.

Aktuelle Meldungen wurden über Social Media Kanäle wie <https://twitter.com/LEDSProjekt> verbreitet. Die produzierten Videobeiträge zum LEDS Projekt wurden unter <https://vimeo.com/ledsprojekt> bereitgestellt. Details zu den entstandenen wissenschaftliche Publikationen wurden außerdem auf gängigen Plattformen wie ResearchGate, Gogole Scholar und Bibsonomy bereitgestellt.

Medienwirksam wurden außerdem mehrere Interviews gegeben. Beispiele hierfür sind:

- [Big Data funktioniert nur mit semantischen Technologien - 5 Nutzungsszenarien](https://www.industry-of-things.de/big-data-funktioniert-nur-mit-semantischen-technologien-5-nutzungsszenarien-a-592187/)
<https://www.industry-of-things.de/big-data-funktioniert-nur-mit-semantischen-technologien-5-nutzungsszenarien-a-592187/>
- [Qualitätssicherung von Daten als Chance und nicht nur Pflicht verstehen](https://www.leds-projekt.de/de/aktuelles/2016/Qualitaetssicherung-von-Daten-als-Chance-und-nicht-nur-Pflicht-verstehen.html)
<https://www.leds-projekt.de/de/aktuelles/2016/Qualitaetssicherung-von-Daten-als-Chance-und-nicht-nur-Pflicht-verstehen.html>
- [Mit praktischer Erfahrung punkten](https://www.tu-chemnitz.de/tu/pressestelle/aktuell/7651)
<https://www.tu-chemnitz.de/tu/pressestelle/aktuell/7651>

Darüber hinaus wurden durch Projektmitarbeiter Workshops wie beispielsweise beim Data Science Open Space oder Developer Open Space in Berlin ausgerichtet.

Präsentationsfolien der Forschungsbeiträge wurden gesammelt unter https://www.slideshare.net/LEDS_Projekt.

6. DIE EINHALTUNG DER AUSGABEN- UND ZEITPLANUNG.

Das Vorhaben lag zum Projektende innerhalb der ursprünglichen Arbeits- und Zeitplanung. Allerdings gab es bedingt durch personelle Fluktuationen innerhalb des VSR-Projektteams der TU Chemnitz eine genehmigte Verlängerung der Projektaktivitäten um 5 Monate und eine Rückzahlung nicht ausgeschöpfter Projektmittel, welche nachstehend kurz beschrieben wird.

Das Projekt wurde mit dem Zuwendungsbescheid vom 28.05.2015 zum 01.07.2015 genehmigt. Die Besetzung der ersten wissenschaftlichen Mitarbeiterstelle erfolgte zum 01.08.2015. Die ursprünglich vorgesehene Besetzung der zweiten Stelle zum 01.08.2015 mit einem Studenten aus dem Masterstudiengang Web Engineering musste zunächst verworfen werden, da der Student krankheitsbedingt seinen Abschluss nicht plangemäß erreichen konnte.

Trotz intensiver Suche nach einer geeigneten Ersatzkandidatin bzw. einem geeigneten Ersatzkandidaten konnte erst keine geeignete Bewerberin bzw. kein geeigneter Bewerber gefunden werden. Drei Monate später konnte die Stelle mit der Einstellung eines Absolventen am 01.11.2015 besetzt werden. Die zeitliche Verzögerung konnte zunächst sehr gut kompensiert werden, da die sonst übliche Einarbeitungszeit weitestgehend entfiel. Die Professur konnte trotz dieser insgesamt angespannten Situation die geplanten Forschungsarbeiten in einer ersten Iteration erreichen.

Aufgrund der verzögerten Einstellung wurden im Jahr 2015 weniger Personalmittel benötigt. Diese wurden bereits in Absprache mit dem Projektträger in das Folgejahr 2016 übertragen. Hierdurch wurde es möglich, die Entwicklung erster Prototypen und vertiefende Ergebnis-Iterationen durch eine weitere Person zu beschleunigen. Abgesehen von der Verschiebung der angesprochenen finanziellen Mittel von 2015 in 2016 konnte der ursprüngliche Arbeitsplan eingehalten werden.

Durch den beschriebenen Mittelübertrag in das Folgejahr 2016 konnte das Forscherteam planmäßig mit einer weiteren Person verstärkt werden. Zum 01.05.2016 wurde ein weiterer wissenschaftlicher Mitarbeiter an der Professur eingestellt, der mit zehnjähriger Berufserfahrung aus der Wirtschaft im Informationsmanagement bereits eine fortgeschrittene Qualifikation mitbrachte und so das bisherige Projektteam unterstützen und bereichern konnte. Parallel dazu verließ einer der bisherigen Mitarbeiter zum 30.06.2016 auf eigenen Wunsch die Professur, wodurch nun planmäßig zwei wissenschaftliche Mitarbeiter innerhalb des LEDES-Wachstums-kerns beschäftigt waren.

Der Mittelabruf lag in dieser Übergangszeit bedingt durch den zusätzlichen Personaleinsatz zur Kompensation der Mitarbeiterfluktuation zum Ende dieses Berichtszeitraums zunächst über der ursprünglich geplanten Kostenkalkulation. Insgesamt werden alle geplanten Mittel in der Kostenplanung wie vorkalkuliert benötigt, abgerufen, und der Finanzierungsplan zum Jahresende wieder ausgeglichen. Der veranschlagte Aufgaben- und Zeitplan wird der Praxis entsprechend bestätigt. Der vorgesehene Arbeitsplan konnte im Berichtszeitraum eingehalten werden.

Es fanden im Rahmen der Veröffentlichung der wissenschaftlichen Ergebnisse inländische Reisetätigkeiten zu Konferenzen im In- und Ausland statt. In Rücksprache wurde weiterhin eine Umwidmung von Mitteln beantragt und bewilligt, die ursprünglich für Verbrauchsmaterialien geplant waren und stattdessen in technisches Arbeitsmaterial investiert werden mussten.

Im Jahr 2017 kam es erneut zu unvorhergesehenen Abweichungen von den ursprünglichen Planungen. Zeitraum vom 04.02.2017 bis zum 03.04.2017 befand sich ein im Projekt angestellter Mitarbeiter in Elternzeit. Zusätzlich fiel unser zweiter wissenschaftlicher Mitarbeiter aufgrund einer akuten Erkrankung langfristig bis Jahresende aus. Entsprechend waren sowohl Arbeits-, Zeit- als auch Kostenplanung davon betroffen.

Als Konsequenz daraus kam es im aktuellen Berichtszeitraum zu ungeplanten Entwicklungspausen, die entweder innerhalb der Entwicklungsgruppe an der Technischen Universität Chemnitz direkt kompensierbar waren oder die zu einer zeitlichen Verschiebung in der Ressourcen- und Ergebnisplanung führten.

Ein nahtloser Ersatz der Mitarbeiter-Tätigkeiten war durch mehrere Gründe nicht möglich. Einerseits war die Dauer der Erkrankung unseres Mitarbeiters nicht direkt absehbar. Die Einarbeitung eines neuen Mitarbeiters in den aktuellen Projektstand hätte ein deutliches Hintergrundwissen vorausgesetzt. Außerdem waren andere Mitarbeiter an unserer Professur parallel bereits in andere Projektaktivitäten eingebunden. Die Genesung unseres Mitarbeiters im Januar 2018 war daher sehr erfreulich, und der Rückstand besonders im Arbeitsschwerpunkt B konnte rasch aufgeholt werden, um alle geplanten Projektergebnisse fristgerecht zum Projektende wie geplant präsentieren zu können. Unterstützend hierfür wurde die Unterstützung eines weiteren Mitarbeiters in der ersten Jahreshälfte 2018 geplant.

Damit verbunden war auch eine Verschiebung in der Kostenplanung, da bedingt durch den Einsatz einer Krankengeldzahlung im Zeitraum der längeren Erkrankung und Wiedereingliederung unseres Mitarbeiters nicht alle Personalmittel wie geplant abgerufen werden konnten. Ebenso führte die Durchführung unserer öffentlichkeitswirksamen Veranstaltung im Rahmen der Konferenz INFORMATIK 2017 in Chemnitz mit der lokalen Präsentation unserer wissenschaftlichen Ergebnisse vor Ort zu einer Einsparung im geplanten Reisebudget. Aufgrund der Personalsituation wurde eine kostenneutrale Projektverlängerung beantragt und bewilligt. Dadurch standen fünf zusätzliche Monate zur Fertigstellung der geplanten Projektergebnisse zur Verfügung. Dafür wurde ein neuer Arbeitsplan erstellt, welcher erwartungsgemäß eingehalten werden konnte.

7. PERSONELLE BESETZUNG

Soll-Ist-Personaleinsatz über die gesamte Projektlaufzeit. Mit (*)-Stern gekennzeichnet sind die Zeiträume zur Einarbeitung neuer Mitarbeiter ins Projekt.

| Monat | PM Soll | PM Ist |
|-------|---------|--------|
| 1 | 2 | 0 |
| 2 | 2 | 1* |
| 3 | 2 | 1 |
| 4 | 2 | 1 |
| 5 | 2 | 2* |
| 6 | 2 | 2 |
| 7 | 2 | 2 |
| 8 | 2 | 2 |
| 9 | 2 | 2 |
| 10 | 2 | 2 |
| 11 | 2 | 3* |
| 12 | 2 | 3 |
| 13 | 2 | 2 |
| 14 | 2 | 2 |
| 15 | 2 | 2 |
| 16 | 2 | 2 |
| 17 | 2 | 2 |
| 18 | 2 | 2 |
| 19 | 2 | 2 |
| 20 | 2 | 1 |
| 21 | 2 | 1 |

| Monat | PM Soll | PM Ist |
|-------|---------|--------|
| 22 | 2 | 2 |
| 23 | 2 | 2 |
| 24 | 2 | 1 |
| 25 | 2 | 1 |
| 26 | 2 | 1 |
| 27 | 2 | 1 |
| 28 | 2 | 1 |
| 29 | 2 | 1 |
| 30 | 2 | 1 |
| 31 | 2 | 1 |
| 32 | 2 | 2* |
| 33 | 2 | 2,5* |
| 34 | 2 | 3 |
| 35 | 2 | 3 |
| 36 | 2 | 3 |
| 37 | 3 | 3 |
| 38 | 2 | 2 |
| 39 | 2 | 2 |
| 40 | 2 | 2 |
| 41 | 1 | 1 |

Chemnitzer Informatik-Berichte

In der Reihe der Chemnitzer Informatik-Berichte sind folgende Berichte erschienen:

- CSR-12-01** Studentensymposium Informatik Chemnitz 2012, Tagungsband zum 1. Studentensymposium Chemnitz vom 4. Juli 2012, Juni 2012, Chemnitz
- CSR-12-02** Tom Kühnert, Stephan Rusdorf, Guido Brunnett, Technischer Bericht zum virtuellen 3D-Stiefeldesign, Juli 2012, Chemnitz
- CSR-12-03** René Bergelt, Matthias Vodel, Wolfram Hardt, Generische Datenerfassung und Aufbereitung im Kontext verteilter, heterogener Sensor-Aktor-Systeme, August 2012, Chemnitz
- CSR-12-04** Arne Berger, Maximilian Eibl, Stephan Heinich, Robert Knauf, Jens Kürsten, Albrecht Kurze, Markus Rickert, Marc Ritter, Schlussbericht zum InnoProfile Forschungsvorhaben sachsMedia - Cooperative Producing, Storage, Retrieval and Distribution of Audiovisual Media (FKZ: 03IP608), September 2012, Chemnitz
- CSR-12-05** Anke Tallig, Grenzgänger - Roboter als Mittler zwischen der virtuellen und realen sozialen Welt, Oktober 2012, Chemnitz
- CSR-13-01** Navchaa Tserendorj, Uranchimeg Tudevtagva, Ariane Heller, Grenzgänger - Integration of Learning Management System into University-level Teaching and Learning, Januar 2013, Chemnitz
- CSR-13-02** Thomas Reichel, Gudula Rüniger, Multi-Criteria Decision Support for Manufacturing Process Chains, März 2013, Chemnitz
- CSR-13-03** Haibin Xu, Thomas Reichel, Gudula Rüniger, Michael Schwind, Softwaretechnische Verknüpfung der interaktiven Softwareplattform Energy Navigator und der Virtual Reality Control Platform, Juli 2013, Chemnitz
- CSR-13-04** International Summerworkshop Computer Science 2013, Proceedings of International Summerworkshop 17.7. - 19.7.2013, Juli 2013, Chemnitz
- CSR-13-05** Jens Lang, Gudula Rüniger, Paul Stöcker, Dynamische Simulationskopplung von Simulink-Modellen durch einen Functional-Mock-up-Interface- Exportfilter, August 2013, Chemnitz
- CSR-14-01** International Summerschool Computer Science 2014, Proceedings of Summerschool 7.7.-13.7.2014, Juni 2014, Chemnitz
- CSR-15-01** Arne Berger, Maximilian Eibl, Stephan Heinich, Robert Herms, Stefan Kahl, Jens Kürsten, Albrecht Kurze, Robert Manthey, Markus Rickert, Marc Ritter, ValidAX - Validierung der Frameworks AMOPA und XTRIEVAL, Januar 2015, Chemnitz

Chemnitzer Informatik-Berichte

- CSR-15-02** Maximilian Speicher, What is Usability? A Characterization based on ISO 9241-11 and ISO/IEC 25010, Januar 2015, Chemnitz
- CSR-16-01** Maxim Bakaev, Martin Gaedke, Sebastian Heil, Kansei Engineering Experimental Research with University Websites, April 2016, Chemnitz
- CSR-18-01** Jan-Philipp Heinrich, Carsten Neise, Andreas Müller, Ähnlichkeitsmessung von ausgewählten Datentypen in Datenbanksystemen zur Berechnung des Grades der Anonymisierung, Februar 2018, Chemnitz
- CSR-18-02** Liang Zhang, Guido Brunnett, Efficient Dynamic Alignment of Motions, Februar 2018, Chemnitz
- CSR-18-03** Guido Brunnett, Maximilian Eibl, Fred Hamker, Peter Ohler, Peter Protzel, StayCentered - Methodenbasis eines Assistenzsystems für Centerlotsen (MACeLot) Schlussbericht, November 2018, Chemnitz
- CSR-19-01** Johannes Dörfelt, Wolfram Hardt, Christian Rosjat, Intelligente Gebäudeklimatisierung auf Basis eines Sensornetzwerks und künstlicher Intelligenz, Februar 2019, Chemnitz
- CSR-19-02** Martin Springwald, Wolfram Hardt, Entwicklung einer RAD-Plattform im Kontext verteilter Systeme, März 2019, Chemnitz
- CSR-19-03** André Böhle, René Schmidt, Wolfram Hardt, Evaluation von Signaleigenschaften zur Lokalisierung von Einschlägen mit Piezokeramischen Sensoren, März 2019, Chemnitz
- CSR-19-04** Johannes Götze, René Schmidt, Wolfram Hardt, Hardwarebeschleunigung von Matrixberechnungen auf Basis von GPU Verarbeitung, März 2019, Chemnitz
- CSR-19-05** Vincent Kühn, Reda Harradi, Wolfram Hardt, Expert System for Adaptive Flight Missions, Juni 2019, Chemnitz
- CSR-19-06** Samer Salamah, Guido Brunnett, Christian Mitschke, Tobias Hess, Synthesizing gait motions from spline-based progression functions of controlled shape, Juni 2019, Chemnitz
- CSR-19-07** Martin Eisoldt, Carsten Neise, Andreas Müller, Analyse verschiedener Distanzmetriken zur Messung des Anonymisierungsgrades θ , Juni 2019, Chemnitz
- CSR-19-08** André Langer, Valentin Siegert, Martin Gaedke, Informationsverwertung basierend auf qualitätsoptimierten semistrukturierten Datenbeständen im Wachstumskern "LEDS", Juli 2019, Chemnitz

Chemnitzer Informatik-Berichte

ISSN 0947-5125

Herausgeber: Fakultät für Informatik, TU Chemnitz
Straße der Nationen 62, D-09111 Chemnitz