

Projektvertrag

Erstellt am 10.4.2013

Zielbestimmung	1
Produkteinsatz	2
Produktübersicht	2
Produktfunktionen	3
Produktdaten	4
Produktleistungen	5
Qualitätsanforderungen	5
Benutzeroberfläche	6
Nicht funktionale Anforderungen	6
Technische Produktumgebung	7
Spezielle Anforderungen an das Projekt	7
Gliederung in Teilprodukte	7

1. Zielbestimmung

Die Internetseite der Personendatenbank des Projekts "Wissenschaftsbeziehungen im 19. Jahrhundert zwischen Deutschland und Russland auf den Gebieten Chemie, Pharmazie und Medizin" der Sächsischen Akademie der Wissenschaften zu Leipzig soll in eine moderne, kollaborative und semantische Webapplikation überführt werden. Dazu müssen bestehende Daten in eine semantische Datenbank überführt werden, umfangreiche Möglichkeiten geschaffen werden die Daten in einem Backend zu editieren und neue Daten einzufügen. Desweiteren soll ein schlichtes Frontend aufgebaut werden, in dem die Daten übersichtlich dargestellt werden. Dieses soll auch Suchmethoden integrieren, die angemessene Möglichkeiten bietet domänenspezifische Fragestellungen zu beantworten.

Die Personendatenbank enthält bio- und bibliographische Daten über Wissenschaftler, die sowohl in Russland als auch in Deutschland tätig waren.

1.1. Mussbestimmungen

Die Internetseite muss für den Nutzer der Datenbank die gleiche Funktionalität bieten wie die bisherige auf reine HTML-Seiten basierte Lösung. Dies beinhaltet eine Listendarstellung aller Verwalteten Personen und Einzelansichten für diese inklusive Portraitfotos.

Die Internetseite muss eine einfache Editiermöglichkeit für die Editoren der Daten bereitstellen.

Eine einfache Nutzerverwaltung muss in der Lage sein Contenteditoren zu verwalten.

Es muss eine angepasste Suchmöglichkeit geben, die die Suche nach Orten, Jahreszahlen und Institutionen und direkten Verknüpfungen zwischen einzelnen gespeicherten Personen ermöglicht.

1.2. Wunschkriterien

Die verwalteten Daten sollen einen höchstmöglichen Grad an semantischer Aufwertung bieten. Darauf aufbauend sollen unterstützende Werkzeuge implementiert werden, welche die Arbeit der Dateneingabe unterstützt sowie die manuelle semantische Aufwertung einzelner Daten erleichtert.

Wünschenswert ist weiterhin eine Graphendarstellung der Bezüge zwischen den einzelnen Subjekten und Objekten, die in der Datenbank gespeichert sind.

1.3. Abgrenzungskriterien

- Das System bietet keinen definitiven Schutz vor Angriffen

- Das Projekt selbst ist nur mittels moderner Webtechnologien (HTML5, CSS3, JavaScript) benutzbar und kann daher keine große Abwärtskompatibilität bereitstellen¹
- Für eine vollständige inhaltliche Kontrolle der Importierten Daten kann der Auftragnehmer keine Verantwortung übernehmen. Eine Kontrolle der Konvertierten Daten ist vom Auftraggeber zu übernehmen.

2. Produkteinsatz

Die Webapplikation wird auf der einen Seite dazu genutzt bestehende Daten zu bearbeiten und zu erweitern sowie neue Daten hinzuzufügen. Es sollen hierbei bio- und bibliographische Daten, die bei der Arbeit der Wissenschaftlerinnen, die am Projekt beteiligt sind, durch das Einfügen in das Backend semantisch aufgewertet werden. Insbesondere muss die zu modellierende Datenstruktur in der Lage sein alle Daten aufzunehmen bzw. leicht zu erweitern sein um die Datenstruktur mit im Forschungsprozess erwartbaren neu auftauchenden Strukturen erweitern zu können.

2.1. Zielgruppen

Die Webapplikation dem Web User die Möglichkeit geben einfach im Datenbestand zu recherchieren und Suchmöglichkeiten zu nutzen, die dabei helfen fachspezifische Fragestellungen zu klären.

Weiterhin soll es möglich sein den Datenbestand in das Semantic Web einzupflegen. Das bedeutet, dass die Daten nicht nur visuell, sondern auch semantisch im Web zugänglich gemacht werden sollen. Damit soll es Projekten der Fachdomäne ermöglicht werden leicht auf die Daten zuzugreifen und die Daten mit eigenen Beständen abzugleichen, bzw. mit Daten anderer Projekte zu verknüpfen.

2.2. Anwendungsbereiche

Die Arbeit der Medizin- und WissenschaftshistorikerInnen an diesem Projekt ist langfristig ausgelegt. Das bedeutet, dass die Plattform eine stabile Basis bilden sollte und die Persistenz der Daten in angemessenen Rahmen sichergestellt werden muss. Außerdem soll die Plattform einfache Erweiterungsmöglichkeiten bieten um auf technischer Seite leicht anpassbar zu sein und damit der Funktionsumfang leicht auszubauen ist und neuen Zielstellungen angepasst werden kann.

2.3. Betriebsbedingungen

Um das Projekt online stellen zu können, muss die Projektgruppe der Sächsischen Akademie der Wissenschaften zu Leipzig einen Webserver bereitstellen, in dem der Dienst des Projektes angeboten wird. Weiterhin ist ein Datenbanksystem bereitzustellen in dem die Daten gespeichert werden.

Der Client selbst benötigt einen modernen, gängigen Webbrowser, welcher in der Lage ist HTML5, CSS3 und JavaScript zu interpretieren, um die angebotenen Funktionen zu nutzen.

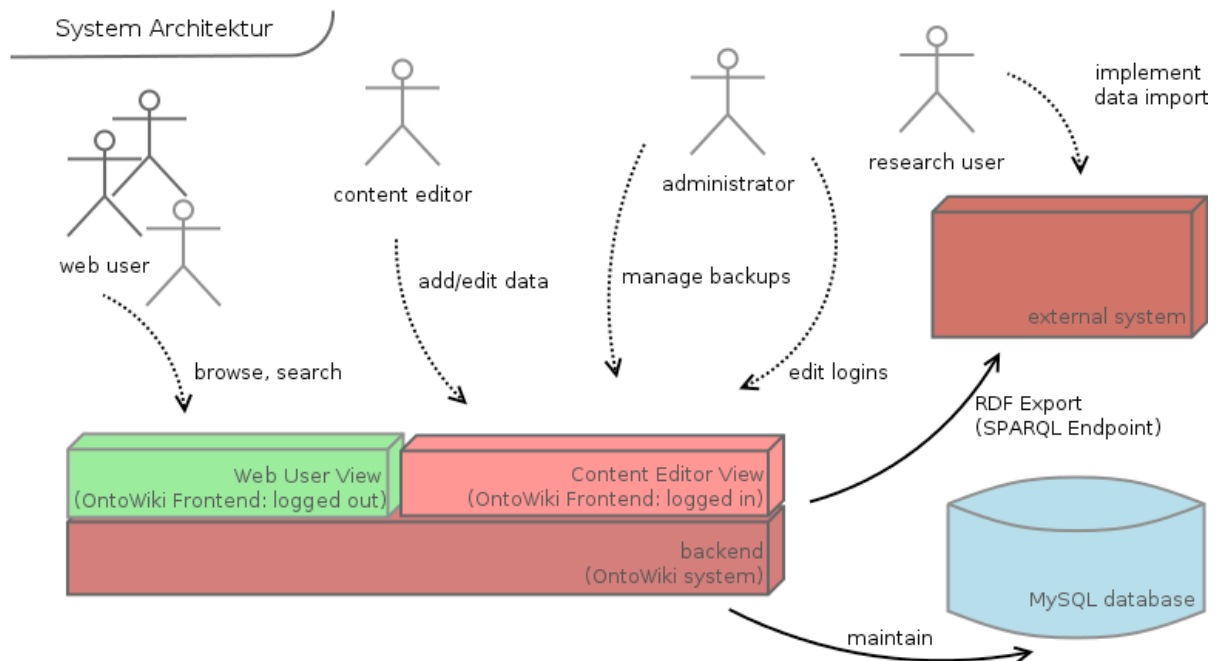
3. Produktübersicht

Auf dem Schema skizziert den Aufbau unseres an die Aufgabenstellung angepassten OntoWiki Systems. Hauptmerkmal ist die Trennung des Zugangs der Web User und der Content Editoren und Administratoren durch den Zustand Logged-in/Logged-out. Alle Daten werden in Modellen in Ontowiki Verwaltet und in einem MySQL Tripelstore über das ERFURT Framework (Bestandteil von

¹ Browser Vorraussetzungen siehe:

<https://github.com/AKSW/OntoWiki/wiki/Browser-Compatibility>

OntoWiki) verwaltet. Der Research User hat die Möglichkeit über einen SPARQL Endpoint Suchanfragen zu stellen.



Web User ist ein Nutzer einer Webapplikation, er kann den Datenbestand nicht ändern, sondern die Funktionen einer Applikation nutzen z.B.: recherchieren und Suchanfragen stellen.

Content Editor hat die Möglichkeit den Datenbestand einer Applikation zu ändern, verfügt also über den Zugang zu dem Backend

Administrator verwaltet die Zugänge der Content User und ist für die Datensicherung verantwortlich.

Research User kann die öffentlichen Daten für weitere professionelle Projekte nutzen. Technisch wird das über einen SPARQL Endpoint realisiert.

4. Produktfunktionen

4.1. Essentielle Produktfunktionen

/LF10/ Geschäftsprozess: Anlegen von Content Editoren

Akteur: Administrator

Beschreibung: Content Editoren werden von dem Administrator angelegt, so dass nur ein begrenzter Kreis die Möglichkeit hat, sich anzumelden und die Datensätze zu editieren.

/LF20/ Geschäftsprozess: Datensicherung

Akteur: Administrator

Beschreibung: Der Administrator kontrolliert die regelmäßigen Datensicherungen.

/LF30/ Geschäftsprozess: Anmeldung/Abmeldung am System

Akteur: Content-Editor

Beschreibung: Der Content Editor soll nach der Rechtevergabe vom Administrator über die An- und Abmeldemöglichkeit verfügen, so dass er die Ressourcen bearbeiten bzw. neu anlegen kann. Dabei kann er auch seine Personendaten editieren und sein Passwort ändern.

/LF40/ Geschäftsprozess: Personen und Daten zu Personen Anlegen

Akteur: Content-Editor

Beschreibung: Als angemeldeter Nutzer kann der Content Editor beliebig viele Personen und mit ihnen verknüpfte Daten anlegen. Die bereits existierenden Personen können aber auch später um die fehlenden Personendaten erweitert werden, es ist also nicht erforderlich alle Daten auf einmal einzugeben.

/LF50/ Geschäftsprozess: Personen und Personendaten Bearbeiten

Akteur: Content-Editor

Beschreibung: Genauso wie beim Daten Anlegen besteht jederzeit die Möglichkeit die schon existierenden Datensätze zu editieren. Der Content-Editor kann die Person bzw. Personen-daten editieren.

/LF60/ Geschäftsprozess: Hinzufügen der Namen in unterschiedlichen Schreibweisen und Sprachen

Akteur: Content Editor

Beschreibung: Der Content Editor kann unterschiedliche Sprachvarianten bzw. Schreib-weisen der Namen beim Eintragen eingeben.

/LF70/ Geschäftsprozess: Namen von Orten und Fachgebieten in verschiedenen Varianten

Akteur: Content Editor

Beschreibung: Sowohl Ortsnamen als auch Fachgebiete können in verschiedenen Varianten hinzugefügt werden, z.B. bei den Ortsnamen ist der Name von den historischen Epochen abhängig, der Content Editor kann also eine damalige äquivalente Bezeichnung dafür verwenden.

/LF80/ Geschäftsprozess: Hinzufügen neuer Kategorien

Akteur: Content Editor

Beschreibung: Neue Kategorie der Lebensstationen kann auf gleiche Weise von dem Content Editor hinzugefügt werden, wie schon vorher die Personen bzw. die Personendaten.

/LF90/ Geschäftsprozess: Veröffentlichungsmodus festlegen

Akteur: Content Editor

Beschreibung: Es können zu jedem Datensatz eine dreistufige Kennzeichnung hinzugefügt werden: Nicht zur Veröffentlichung vorgesehen, In Bearbeitung und Zur Veröffentlichung freigeben.

/LF100/ Geschäftsprozess: Volltextsuche

Akteur: Web User

Beschreibung: Dem Web User soll möglich sein direkt auf der Startseite Suchbegriffe eingeben zu können, die dann automatisch als Volltextsuche zu Ergebnissen führen.

/LF110/ Geschäftsprozess: Erweiterte Suche

Akteur: Web User

Beschreibung: Dem Web User ist es möglich Suchzeiträume einzuschränken sowie gezielt nach Institutionen oder Lebensphasen zu suchen, die mit einzelnen oder mehreren Personen verknüpft sind.

/LF120/ Produktfunktion: Übersichtliche Darstellung

Beschreibung: Im Frontend sollen die Daten schlicht und gut strukturiert zugänglich sein.

/LF130/ Produktfunktion: Überführung der HTML-Daten in eine Datenbank und Erfassung des Vokabulars

Beschreibung: Die bestehenden HTML-Daten werden in eine Datenbank überführt. Personen, Institutionen, Zeitpunkte, Fachgebiete, Titel und Auslandsaufenthalte werden semantisch erfasst

/LF140/ Produktfunktion: Transliterationen von Namen (eng-de, kyrillisch)

Beschreibung: Die Namen werden je nach Eingabeart automatisch (kyrillisch-deutsch und deutsch-kyrillisch sowie kyrillisch-englisch und englisch-kyrillisch) transliteriert.

4.2. Optionale Produktfunktionen

/OF10/ Geschäftsprozess: Kartendarstellung

Akteur: Web User

Beschreibung: Neben den bereits erzeugten Ergebnissen der Suchanfrage werden dem Nutzer die erwähnten Orte auf der Karte (in der damaligen Gliederung Europas) eingeblendet.

/OF20/ Geschäftsprozess: Grafendarstellung von Beziehungen

Akteur: Web User

Beschreibung: Als Ergebnis einer Suchanfrage wird eine Grafendarstellung von Beziehungen verknüpfter Instanzen dargestellt.

/OF30/ Geschäftsprozess : Integration von semantischen Daten aus anderen Quellen

Akteur: Content Editor

Beschreibung: Semantische Daten aus anderen Quellen (Wikipedia, Catalogus Professorum, GND der deutschen Nationalbibliothek) können semantisch integriert werden, möglichst automatisch, d.h. zu den gewählten Datensätzen können semantisch verknüpfte Vorschläge aus anderen Quellen auftauchen.

5. Produktdaten

/LD10/ Grunddaten von Personen (ca.1850)

Beschreibung: Es liegen folgende Daten vor: Nachname, Vorname, Geburtstag, Sterbetag, Geburtsort, Sterbeort, Beruf, interne Identifikationsnummer

/LD20/ Umfangreiche Lebensläufe (ca. 230)

Beschreibung: Es liegen Daten zu Verwandte, Wissenschaftlicher Laufbahn, Veröffentlichungen und umfangreichen Literaturangaben vor.

6. Produktleistungen

/LL10/ Zugriffskontrolle der erstellten Daten

Beschreibung: Noch nicht zur Veröffentlichung bestimmte Daten dürfen dem Web User nicht zugänglich sein

/LL20/ Bessere semantische Aufbereitung von Daten

Beschreibung: Der Server soll eine Funktion bieten, die der besseren semantischen Aufbereitung von Daten dient, so dass die Suchergebnisse über einen höheren Grad der semantischen Bearbeitung verfügen.

7. Qualitätsanforderungen

Um nach der Fertigstellung des Produkts schnell gewährleisten zu können, dass es produktiv eingesetzt werden kann ist es nötig das Hauptaugenmerk auf Funktionalität, Zuverlässigkeit und Benutzbarkeit zu legen.

Es muss gewährleistet sein, dass alle recherchierten Daten des Projekts leicht integriert werden können und die Benutzbarkeit darauf ausgerichtet ist, Nutzern mit normalen Computerkenntnissen einen leichten Einstieg in die Arbeit zu ermöglichen.

Da das Projekt langfristig ausgelegt ist, muss der Zuverlässigkeit ein hoher Stellenwert eingeräumt werden. In dem Zusammenhang ist auf eine ausgereifte Testphase vor dem produktiven Einsatz unabdingbar.

Die Wartbarkeit muss angemessen im normalen Bereich möglich sein. Hierbei sind keine besonderen Anforderungen zu beachten.

Die Übertragbarkeit muss möglich sein, da aber das Produkt eine Grundlage der Arbeit der WissenschaftlerInnen bieten soll ist nicht damit zu rechnen, dass das Produkt schnell durch ein anderes ersetzt wird. Trotzdem spielt die Übertragbarkeit während des gesamten Entwicklungszyklus eine große Rolle. Durch die Exportmöglichkeit des Vokabulars so wie der Daten ist es leicht möglich den Datenbestand in ein Nachfolgeprojekt zu integrieren. Durch diese Technik ist es auch nach durch Vokabular Alignment möglich die Daten in bestehende andere Projekte ganz zu integrieren. Die Produktqualität muss dem Projekt angemessen sein und den produktiven Einsatz ermöglichen.

Produktqualität	Sehr gut	Gut	Normal	Weniger Relevant
Funktionalität				
Angemessenheit		✓		
Richtigkeit		✓		
Interoperabilität	✓			
Ordnungsmäßigkeit		✓		
Sicherheit			✓	
Zuverlässigkeit				
Reife		✓		
Fehlertoleranz		✓		
Wiederherstellbarkeit	✓			
Benutzbarkeit				
Verständlichkeit		✓		
Erlernbarkeit		✓		
Bedienbarkeit	✓			
Effizienz				
Zeitverhalten				✓
Verbrauchsverhalten				✓
Wartbarkeit				
Analysierbarkeit		✓		
Modifizierbarkeit	✓			
Stabilität		✓		
Prüfbarkeit			✓	
Übertragbarkeit				
Anpassbarkeit	✓			
Installierbarkeit		✓		
Konformität		✓		
Austauschbarkeit	✓			

8. Benutzeroberfläche

/BO 10/

Beschreibung: Die Darstellung der Benutzungsoberfläche erfolgt durch einen Web browser, welcher HTML5, CSS3, JavaScript verarbeiten kann.

/BO 20/

Beschreibung: Die Bedienung erfolgt mittels Maus und Tastatur bzw. anderer durch den Browser unterstützte Eingabegeräte.

9. Nicht funktionale Anforderungen

/N 10/

Beschreibung: Die Benutzeroberfläche für der Web User muss einfach zu bedienen und nutzerfreundlich sein.

/N 20/

Beschreibung: Die Benutzeroberfläche für der Content Editor muss einfach zu bedienen und den Nutzer bei der Eingabe neuer Daten unterstützen.

10. Technische Produktumgebung

10.1. Software

- MySQL 5.5.22 (Virtuoso möglich)
- PHP 5.2 oder höher
- Apache Web Server 2.4.1 (Mit mod_rewrite) (NGINX möglich)
- OntoWiki 0.9.7 (Unterpakete stellt OntoWiki bereit)
 - Erfurt (aktuelle Version)
 - ARC2 (aktuelle Version)
 - Zend Framework (aktuelle Version)

10.2. Hardware

- Webserver, auf dem das Projekt online gestellt wird
- MySQL Datenbank
- Client Rechner mit aktuellem Webbrowser²

10.3. Orgware

- Wikiseiten des OntoWiki Projekts (<https://github.com/AKSW/OntoWiki/wiki>)
- Zugang zum Webserver ist notwendig, um etwaige Modifikationen an dem Projekt durchzuführen

10.4. Schnittstellen

- Zum manuellen Im- und Export von semantischen Daten können die Formate RDF/XML, Turtle, RDF/JSON sowie Notation3 genutzt werden
- Weiterhin steht grundsätzlich ein SPARQL-Endpoint zur Verfügung

11. Spezielle Anforderungen an das Projekt

- Es bestehen keine speziellen Anforderungen an das Projekt

12. Gliederung in Teilprodukte

- Tool zum erstellen der semantischen Daten aus der bisherigen HTML Seite und zur Verfügung gestellten Dateien im Python Format .pickle.
- Die eigentliche Webapplication mit der oben beschriebenen Funktionalität

² Browser Vorraussetzungen siehe:

<https://github.com/AKSW/OntoWiki/wiki/Browser-Compatibility>