



**- Angebots- und Vertragswesen: Angebot**

**Version: 1.0(final)**

<b>Projektbezeichnung</b>	MSP-13 - Integration eines Semantischen Tagging Systems in Microsoft Sharepoint
<b>Projektleiter</b>	Martin John
<b>Verantwortlich</b>	Projektleiter
<b>Erstellt am</b>	11.01.2013
<b>Zuletzt geändert</b>	21.01.2013 21:30
<b>Bearbeitungszustand</b>	<input type="checkbox"/> in Bearbeitung
	<input type="checkbox"/> vorgelegt
	<input checked="" type="checkbox"/> fertig gestellt
<b>Dokumentablage</b>	C:\Aktuell\Uni LE\SWT Praktikum SoSe 13\Aufgaben\Projektangebot_v10final.odt
<b>V-Modell-XT Version</b>	1.4

## Änderungsverzeichnis

Änderungen			Geänderte Kapitel	Beschreibung der Änderungen	Autor	Zustand
Nr.	Datum	Version				
1	11.01.2013	0.1	Alle	Initiale Produkterstellung	JM	
2	18.01.2013	0.1r	5,6	Fachlich überarbeitet	JM	
3	21.01.2013	0.2	9	Gelöscht	RA	

## Prüfverzeichnis

Die folgende Tabelle zeigt einen Überblick über alle Prüfungen – sowohl Eigenprüfungen wie auch Prüfungen durch eigenständige Qualitätssicherung – des vorliegenden Dokumentes.

Datum	Geprüfte Version	Anmerkungen	Prüfer	Neuer Produktzustand
21.1.2013	0.2	Kap9 gelöscht, Glossar hinzugefügt Rechtsschreibprüfung	RA	

**INHALTSVERZEICHNIS**

1 Einleitung.....	4
2 Zielbestimmung.....	4
3 Produkteinsatz.....	4
4 Designübersicht.....	5
5 Funktionalität und Arbeitspakete.....	6
5.1 Zwingende funktionale Anforderungen.....	6
5.2 Optionale funktionale Anforderungen.....	7
5.3 Nicht funktionale Anforderungen.....	7
5.4 Aufwandsbetrachtung.....	7
6 Qualitätssicherung.....	7
6.1 Konzeption .....	7
6.2 Qualitätsanforderungen.....	7
7 Glossar.....	8

## 1 EINLEITUNG

Ziel dieses Projektangebotes ist es, die formellen und informellen Erwartungen des Kunden schriftlich zu erfassen und darzustellen und in Berücksichtigung von Machbarkeiten die entstehenden Anforderungen entsprechend ihres Aufwandes zu kategorisieren.

Als Grundlage für dieses Angebot gilt die Aufgabenstellung:

Mit Hilfe eines existierenden Knowledge Extraktion Frameworks (z.B. [AKSW Fox](#)) sollen Dokumente in einem [Microsoft Sharepoint Server](#) automatisch getagt werden. Dies soll automatisch im Hintergrund passieren und die erzeugten Metadaten sollen gespeichert werden und angezeigt werden können. Die Integration mit dem Sharepoint Server soll auf typische Weise erfolgen, d.h. dies ist eine gute Möglichkeit, sich umfassend in die Plattform einzuarbeiten.

Tagging ist ein wesentlicher Bestandteil des vernetzten Arbeitens. Dabei ist die bisherige Prämisse das manuelle Tagging von Inhalten. Die Nutzbarkeit dieser Tags ist jedoch recht eingeschränkt.

Die Nutzung von semantischem Tagging erhöht die Verwendbarkeit des Tagging, denn damit können informell verfasste Inhalte maschinenlesbar gestaltet werden. Die Inhalte sind damit besser automatisch auswertbar. Eine verbesserte Auswertung von Inhalten kann wesentlich zum Wissensmanagement in Organisationen beitragen, indem große Wissensmengen miteinander in Beziehung gesetzt werden können.

Für das semantische Tagging werden Schlagworte als RDF-Triples abgespeichert. Diese Triples verbinden Ressourcen mit anderen Ressourcen. Für das Tagging bedeutet das, ein Tag-Begriff (Subjekt) wird in Beziehung gesetzt (Prädikat) zu einer anderen Ressource (Objekt) mit dem Sinn, den eigentlichen Tag näher zu beschreiben.

Im Folgenden wird dargestellt, welche Leistungen in Anlehnung an die Anforderungen durch die Projektgruppe erfüllt werden können.

## 2 ZIELBESTIMMUNG

Ziel der Projektarbeit ist die exemplarische Erstellung eines lauffähigen Betriebssystems mit einem Microsoft Sharepoint Server und einer Datenbank als Imagedatei für Oracle VirtualBox. In der Datenbank werden RDF-Triples verwaltet, die während des semantischen Taggings auf den Sharepoint-Inhalten entstehen.

## 3 PRODUKTEINSATZ

Das Ergebnis der Arbeit hat exemplarischen Charakter und soll nachweisen, wie es möglich ist, Daten-Ressourcen einer Arbeitsgruppe mit anderen verfügbaren Datenquellen in Beziehung zu setzen. Das Projekt ist somit eine Grundlage dafür, dass jeglichen Organisationen große Mengen an sich stetig entwickelten Inhalten so verwalten zu können, dass aggregiertes Wissen erhalten und geteilt innerhalb der Organisation geteilt werden kann.

## 4 DESIGNÜBERSICHT

Die folgende Abbildung stellt Entwurfsweise die allgemeine Funktionalität des Projekts dar:

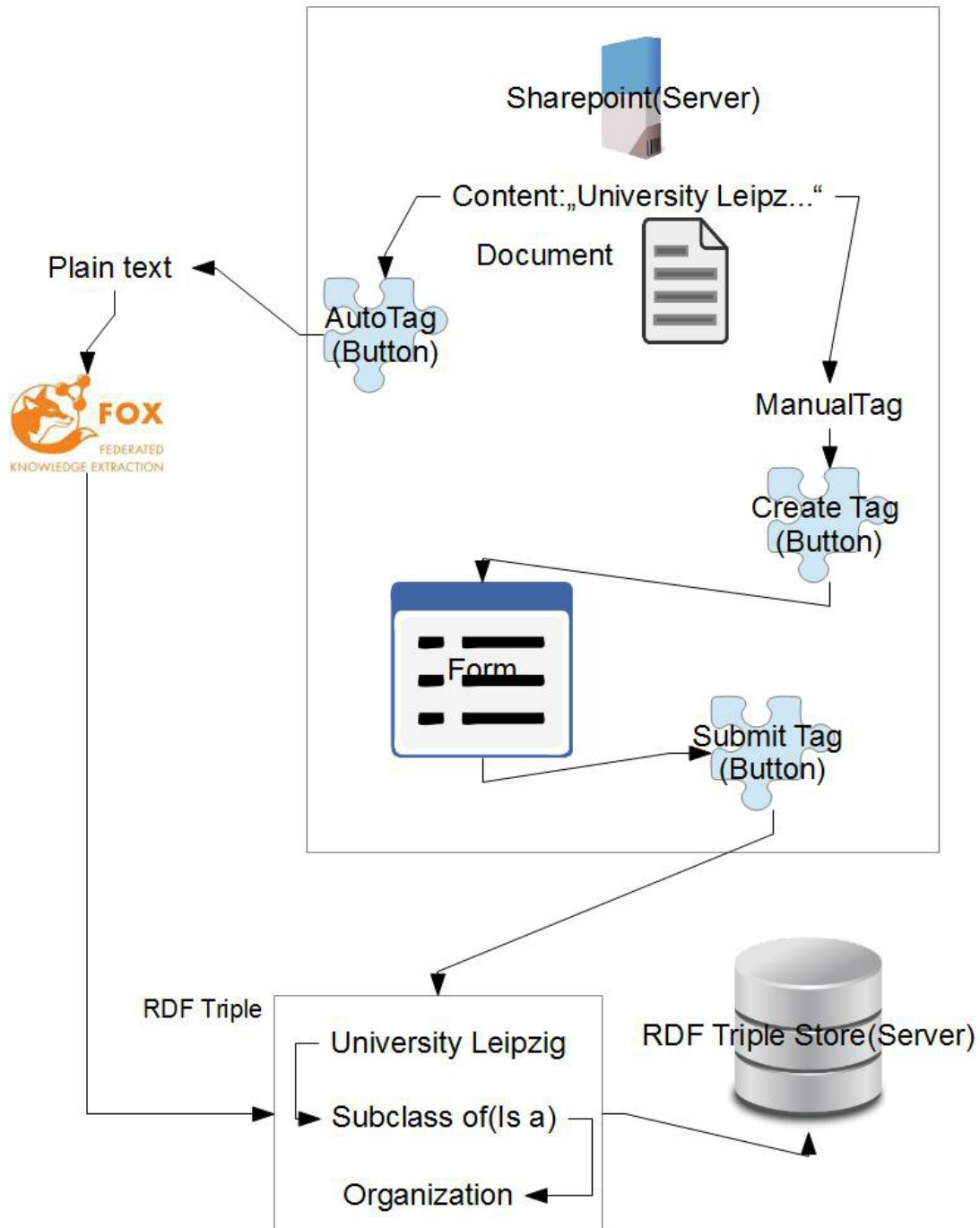


Abbildung 1: Funktionsschema

## 5 FUNKTIONALITÄT UND ARBEITSPAKETE

### 5.1 Zwingende funktionale Anforderungen

- ZF 000 Es muss ein funktionierendes Oracle VirtualBox Image erstellt werden
- 10 Enthält Betriebssystem Microsoft Windows, Version beliebig
  - 20 Enthält Microsoft Sharepoint (siehe ZF 100)
  - 30 Enthält eine Datenbank, die RDF-Triples verwaltet kann (siehe ZF 200)
- ZF 100 Es muss ein exemplarisches nutzbares Sharepoint-Portal erstellt werden
- 10 Das Portal ist im lokalen Netzwerk über einen internetfähigen Browser erreichbar
  - 20 Es gibt eine Seite, auf der Freitexteinträge verfasst werden können
  - 30 Es gibt eine Seite, auf der Dokumente beliebigen Formats hochgeladen und gelöscht werden können(Größe?)
  - 40 Dokumente vom Format .pdf, .doc, .docx, .xls, .xlsx, .ppt, .pptx können im SP-eigenen Viewer angezeigt werden
- ZF 200 Es muss eine nutzbare Datenbank für das Verwalten von RDF-Triples erstellt werden (=Triple Store)
- 10 Der Triple Store muss per SPARQL abfragbar sein
  - 20 Der Triple Store ist im lokalen Netzwerk erreichbar
  - 30 Der Triple Store muss in ihrer Struktur auf Analysierbarkeit angelegt sein
  - 40 RDF Triples müssen in einer Liste angezeigt werden können und müssen editierbar und löscherbar sein
- ZF 300 Manuelles semantisches Tagging muss im Sharepoint Portal möglich sein
- 10 Es ist möglich eine Textstelle in einem Freitexteintrag zu markieren und per Schaltfläche einen Tag hinzuzufügen
  - 20 Freitexteinträgen können per Schaltfläche Tags hinzugefügt werden (Metadaten)
  - 30 Es ist möglich zu einem hochgeladenen Dokument ein Tag per Schaltfläche zu erstellen (Metadaten)
  - 40 Für das Erstellen eines Tags wird ein Formular angezeigt
  - 41 Im Formular wird das Subjekt eines RDF-Triples automatisch befüllt
  - 42 Im Formular erfolgt die Festlegung des RDF-Prädikats per Drop-Down-Menü
  - 43 Die Erstellung eines Tags im Formular kann abgebrochen werden
  - 44 Ein im Formular erstelltes Tag muss erst bestätigt werden, bevor es in der Datenbank gespeichert wird.
  - 45 Bereits bestehende Tags werden bei erneuter Eingabe erkannt und der Nutzer wird entsprechend informiert. Das Tag wird nicht überschrieben.

## 5.2 Optionale funktionale Anforderungen

- OF 000 In Fließtexten erstellte Tags müssen erkennbar sein
- OF 100 Alle Tags zu einem Text oder Dokument müssen unter dem Text oder Dokument per Klick auf eine Schaltfläche aufgelistet werden
- OF 200 Das Hinzufügen von Tags ist per Kontextmenü möglich (markieren, Rechtsklick, ->Tag hinzufügen)
- OF 400 Automatisches Tagging wird auf Knopfdruck durchgeführt mittels eines Knowledge Extraction Frameworks
- OF 500 Windows-Installationsdatei für beliebige Sharepoint Server
- OF 600 Beim manuellen Tagging werden Vorschläge durch AKSW Fox innerhalb der Form unterbreitet

## 5.3 Nicht funktionale Anforderungen

- ZN 000 In der Dokumentation ist die Möglichkeit der Übertragbarkeit auf andere Systeme auszuführen in Hinsicht auf Aufwand und Kompatibilität

## 5.4 Aufwandsbetrachtung

- |        |                                  |     |
|--------|----------------------------------|-----|
| ZF 000 | Funktionierende VBOX-Image       | 5%  |
| ZF 100 | Sharepoint Portal aufsetzen      | 10% |
| ZF 200 | Datenbank erstellen              | 25% |
| ZF 300 | Manuelles Tagging implementieren | 60% |
- Die Datenbank kann hierbei parallel erstellt werden.

# 6 QUALITÄTSSICHERUNG

## 6.1 Konzeption

Für die Qualitätssicherung wird entsprechend den Anforderungen des Auftraggebers ein Qualitätssicherungskonzept erstellt, dass sich in folgende Punkte unterteilt:

Allgemein

Dokumentationskonzept

Testkonzept

Die genauen Ausführungen sind in der Anlage A festgehalten.

## 6.2 Qualitätsanforderungen

Produktqualität	Sehr gut	Gut	Normal	Nicht relevant
Funktion	•			
Zuverlässigkeit			•	
Benutzbarkeit		•		
Effizienz				•
Änderbarkeit			•	
Übertragbarkeit		•		

## 7 GLOSSAR

### Sharepoint

Sharepoint ist ein Produkt von Microsoft welches die Zusammenarbeit von Projektteams koordinieren soll. Die Vorteile sind eine gemeinsame Plattform, die Zusammenarbeit und die schnelle Reaktionsfähigkeit auf die Businessanforderungen. Es ist Dokumentenmanagementsystem und Informationsmanagementsystem (Webseiten mit einem intuitiven CMS). Damit ist die Plattform anpassbar auf die Prozesse und Strukturen von Unternehmen, Gruppen und Projekten. Es kann Daten von verschiedenen Formaten und Herstellern über eine einzige Schnittstelle verfügbar machen. Es gibt die Varianten "Sharepoint Foundation" mit den Grundfunktionalitäten und "Sharepoint Server". Sharepoint Server ist die portalbasierte Version von Microsoft Sharepoint. Eine Sharepoint Server Umgebung verfügt über mindestens einen Sharepoint Foundation Server. Zusätzlich zu den Funktionen der Foundation bietet Server weitere Dienstanwendungen und spezielle Funktionen.

### Knowledge Extraction Framework

Ist die Schaffung von Wissen aus strukturierten (relationalen Datenbanken, XML) und unstrukturierten (Text, Dokumente, Bilder) Quellen.

FOX erstellt RDF Tripel. Es ist ein Java Framework.

### CMS

Ein Content-Management-System ist eine Software zur gemeinschaftlichen Erstellung, Bearbeitung und Organisation von Inhalten (Content) zumeist in Webseiten, aber auch in anderen Medienformen.

### RDF

Resource Description Framework - ermöglicht die Beschreibung von Web-Ressourcen, wobei es auf XML basiert. RDF dient der Darstellung von Internetdaten, und ist dabei auch für Computer verständlich. RDF kann bei der Bildung von Semantic Web angewandt werden. RDF ist ein Tripel, dh.: zur Darstellung von einem Ressource braucht man einen Subjekt, Prädikat und Objekt.

### SPARQL

SPARQL ist eine graph-basierte Anfragesprache für RDF.

### Tagging

Ist eine Kategorisierung von Informationen.

Tagging System

Ein System, das die Art und Weise des Taggings organisiert und die Verwaltung und Nutzung der Tags ermöglicht. Dabei gibt das System vor, ob das Tagging mittels freiem Vokabular oder mittels einer Auswahl an möglichen Tags durchgeführt wird.

### Metadaten

Sind strukturierte Daten, die Informationen über andere Informationsressourcen enthalten.

### Framework

Ein Framework ist ein Programmiergerüst, das in der Softwaretechnik, insbesondere im Rahmen der objektorientierten Softwareentwicklung sowie bei komponentenbasierten Entwicklungsansätzen, benutzt wird.

### NIF

Ist ein RDF/OWL<sup>1</sup>-basiertes Format, das die Arbeit zwischen Natural Language Processing (NLP) Tools, Sprachressourcen und Anmerkungen erreichen will.

### NLP

Natural Language Processing ist ein Bereich der Informatik, Künstliche Intelligenz und Linguistik mit den Wechselwirkungen zwischen Computern und Menschen (natürlich) betreffenden Sprachen.

### C#

C# ist eine von Microsoft im Rahmen von .NET entwickelte Sprache. Das .NET-Framework ist eine Plattform bestehend aus einer Laufzeitumgebung, in der die Programme ausgeführt werden und einer Sammlung von Klassenbibliotheken und Schnittstellen. Hauptsächlich ist C# objektorientiert.

---

1 OWL: Web Ontology Language, Beschreibungssprache nach der Syntax von RDF.



**Server**

Ist ein Software Programm, dass mit anderen Programmen kommuniziert (Bsp. Clients, Server etc.) um bestimmte Dienstleistungen anzubieten.

**RDF-Graph-Modell**

Die grundlegende Struktur des RDF ist eine Ansammlung von 'Triplen'. Jedes Triple besteht aus einem Subjekt, Prädikat (auch Eigenschaft) und Objekt. Der Verbund mehrerer solcher Triple heißt RDF Graph, welcher mittels Knoten (entsprechen Subjekt oder Objekt) und Kanten (repräsentieren Prädikate) als gerichteter Graph dargestellt werden kann. Jedes Triple gibt Auskunft über ein Relation (Prädikat) zwischen zwei Dingen die durch Subjekt und Objekt beschrieben werden. In natürlicher Sprache würde sich ein Triple in etwa so lesen: Subjekt X hat die Eigenschaft (Prädikat) z, deren Wert/Aussage gleich Objekt Y ist. Die Aussage eines RDF Graphen setzt sich aus den Aussagen jedes seiner Triple zusammen (mittels log. AND-Verknüpfung).

Ein Knoten eines RDF Graphen kann eine URI-Reference, eine URI, ein Literal oder 'Blank' sein. Kanten (Eigenschaften) sind URI-Referenzen welche die Relation zwischen Subjekt und Objekt beschreiben. Eine URI-Referenz, bzw. ein Literal welches als Knoten verwendet wird gibt an, was dieser Knoten repräsentiert.

**URI / URL / URN**

Uniform Resource Identifier, kurz URI, sind Zeichenfolgen, welche beliebige Ressourcen (Webseiten, Bilder, etc.) aus dem Web mittels eindeutiger Adressen identifiziert. URIs können als Zeichenfolge in digitale Dokumente, z.B. HTML oder PHP eingebunden werden.

URLs und URNs sind dabei Unterarten von URIs.

Uniform Resource Locators (URLs) benennen eine Ressource über ihren primären Zugriffsmechanismus wie http oder ftp. Unified Resource Names (URNs) identifizieren eine Ressource mittels eines vorhandenen oder frei zu vergebenden Namens, z. B. urn:isbn.

**URI-Referenz**

Eine URI-Referenz ist eine Zeichenfolge, die eine URI und somit die Ressource dieser URI repräsentiert. URI-Referenzen werden zu vollwertigen URIs konvertiert, indem sie anhand einer Basis-URI nach einem bestimmten Algorithmus aufgelöst werden. (wikipedia)

**Literale**

Literale geben die Werte einer Aussage/Eigenschaft (z.B. Name, Zahlen, Datum usw.) als String an. Jeder Wert eines Literals könnte auch als URI angegeben werden, doch ist es oftmals einfacher und anschaulicher ein Literal zu verwenden.

Man unterscheidet:

- 'plain literals' – normaler String mit einer Aussage
- 'typed literals' – String mit angefügter URI-Referenz zu einem Datentyp (Wertangabe)
- Ein Literal kann das Objekt eines RDF-Ausdrucks sein, nie aber ein Subjekt oder Prädikat.

**Entailment**

Bedeutungen und Folgerungen innerhalb eines RDF-Graphen wird durch das formale Konzept des Entailment (Folgebeziehung) gestützt.

Einfach ausgedrückt: Ausdruck A hat einen anderen Ausdruck B zur Folge, falls jede Bedingung die A 'wahr' macht auch B 'wahr' macht.

Auf dieser Grundlage kann man die 'Wahrheit' von B folgern, wenn man von der 'Wahrheit' von A ausgeht oder dies bereits gezeigt hat.

**Blanks**

Ein 'Blank'-Knoten besteht weder aus URI noch aus einem Literal und wird in der RDF- Syntax als UNIQUE-Knoten verwendet, welcher in mehreren RDF-Statements verwendet werden kann, aber keinen eindeutigen Namen hat.

**ABBILDUNGSVERZEICHNIS**

Abbildung 1: Abb. 1: Funktionsschema.....5

Symbole und Grafiken unter Nutzung von openclipart.org erstellt.