

Projektangebot

Inhaltsverzeichnis:

1. Zielbestimmung
2. Produkteinsatz
3. Designübersicht
4. Funktionalität und Arbeitspakete
5. Produktleistungen
6. Qualitätssicherung
7. Glossar

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.

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.

Die Arbeit der MedizingeschichtlerInnen 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.

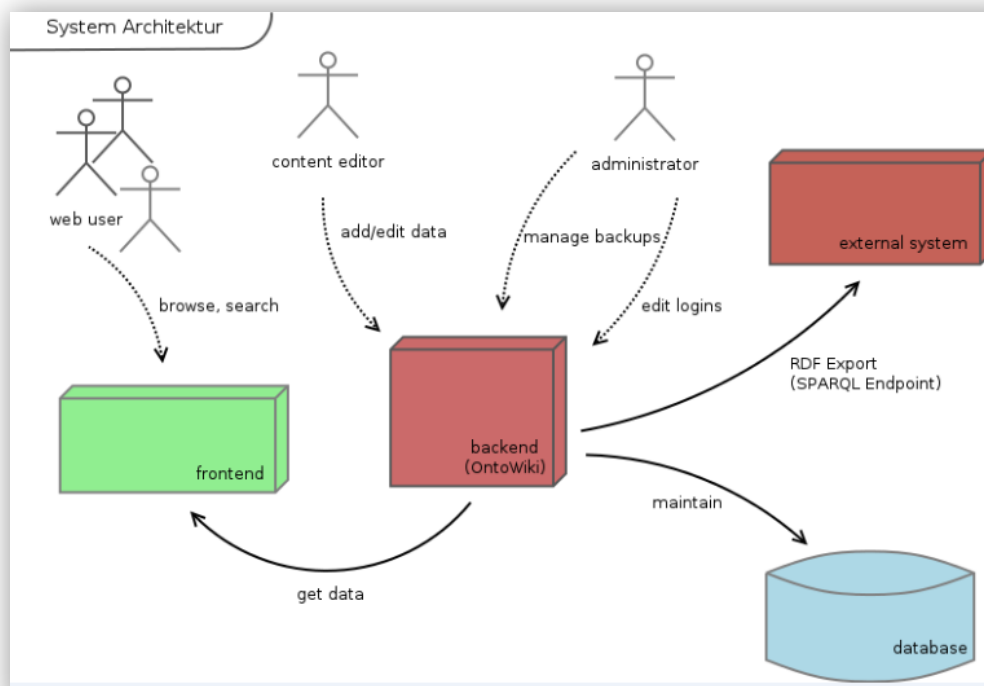
Auf der anderen Seite soll die Webapplikation dem Web User die Möglichkeit geben einfach im Datenbestand zu recherchieren und Suchmöglichkeiten zu nutzen, die dabei helfen fachspezifische Fragestellungen zu klären.

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.

3. Designübersicht

Auf dem Schema der Systemarchitektur (Schema 3.1) sind 4 Schichten zu unterscheiden. Das Frontend ist die Stelle, auf die der Web User zugreifen kann, indem er eine Suchanfrage stellt. Danach werden ihm Suchergebnisse angezeigt. Das Frontend bezieht sich direkt auf das Backend. In dieser Systemarchitektur ist das Backend als Ontowiki zu verstehen, wo der Content Editor unterschiedliche Inhalte editieren kann und wo der Administrator nicht nur die Datensicherung sicherstellt, sondern auch die Logins zum System ermöglicht. Vom Backend erfolgt der Export von RDF-Tripeln, die dank dem SPARQL-Endpoint, wo die Suchanfragen gestellt werden, angezeigt werden können. Alle Backend-Daten werden in einer Datenbank gespeichert.

Schema 3.1



Rollen:

Administrator – ist für das Verwalten der Zugänge zum Backend zuständig und kontrolliert die Datensicherung.

Content Editor – zu seinen Aufgaben gehören sowohl die Erstellung von Personen und Daten zu den Personen als auch die Editiermöglichkeiten der bereits existierenden Datensätzen. (Schema 4.1)

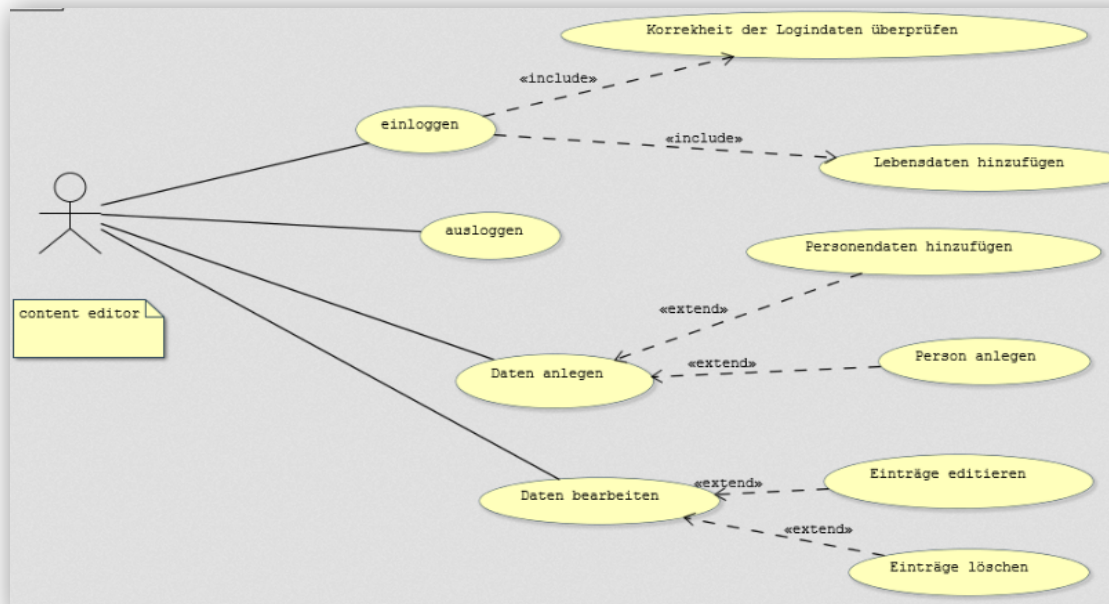
Beispielanwendung: Die Person loggt sich ein, nachdem sie von dem Administrator Login-Daten bekommen hat. Der Content-Editor trägt Lebensdaten einer Person der Personendatenbank ein.

Web User – darf in den Daten freizugänglich recherchieren. Nach der Suchanfrage werden ihm die Suchergebnisse angezeigt. (Schema 4.2)

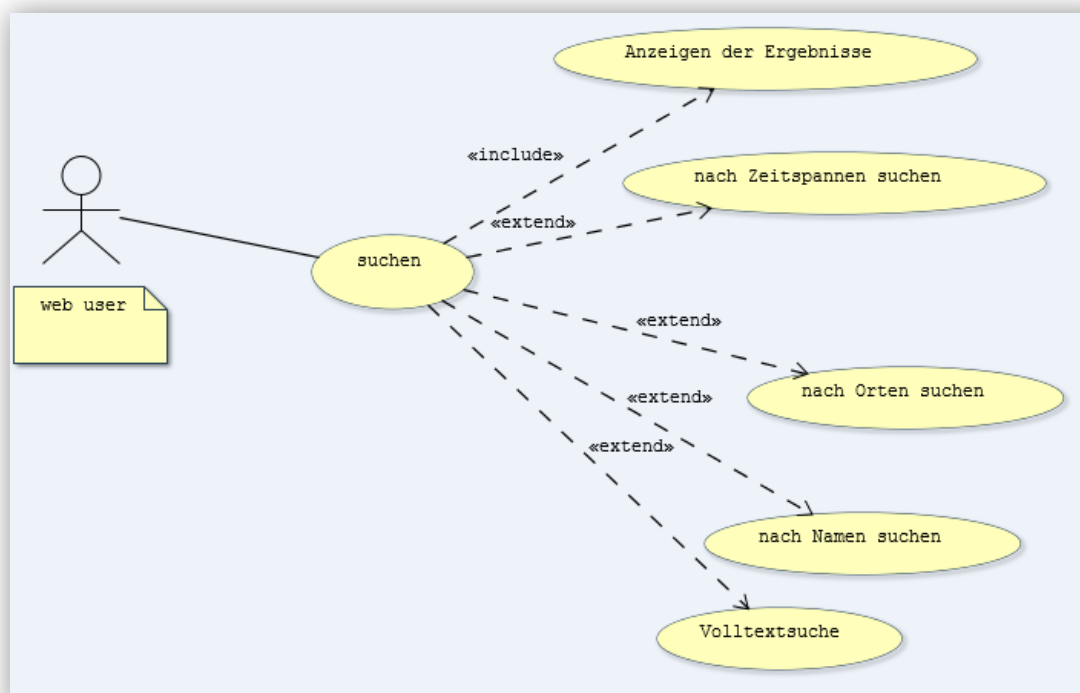
Beispielanwendung: Der Datennutzer sucht nach einem Wissenschaftler und spezifiziert seine Anfrage, indem er seine Daten eingibt bzw. Lebensabschnitte und Orte auswählt. Danach werden die Suchergebnisse angezeigt.

Nutzer des SPARQL-Endpoints – Nutzer mit fortgeschrittenem Wissen, der die semantischen Daten für andere Projekte nutzen möchte.

Beispielanwendung: Semantische Daten im RDF-Format über den SPARQL-Endpoint exportieren.



Schema 4.1: Use-Case Diagramm – Content Editor



Schema 4.2: Use-Case Diagramm – Web User

4. Funktionalität und Arbeitspakete

4.1 Funktionalitäten:

/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: 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.

/LF100/ 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.

/LF110/ Geschäftsprozess: Übersichtliche Darstellung

Akteur: -

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

Serverseitige Funktionen:

/LF120/ Geschäftsprozess: 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

/LF130/ Geschäftsprozess: 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.

Optionale Produktfunktionen:

/LF140/ 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.

/LF150/ Geschäftsprozess: Grafendarstellung von Beziehungen

Akteur: Web User

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

/LF160/ 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.

Serverseitige Funktionen:

/LF170/ Geschäftsprozess: 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.

Mögliche Erweiterungen im Nachfolgeprojekt:

Je nach dem Aufwand der Muss-Funktionen ist es auch möglich, dass die optionalen Anforderungen erst in einem Nachfolgeprojekt realisiert werden können. Das Nachfolgeprojekt kann um automatische Generierung der semantisch verknüpften Datensätzen erweitert werden. Es könnten dann beim Eintragen der semantischen Datensätze sofort unterschiedliche Begriffe die semantische Verbindung zu den bereits eingetragenen Daten aufweisen Anzeigen. Ein Beispiel wären Städtenamen oder Institutionen, bei denen automatisch Vorhandene Daten Vorgeschlagen werden. Dies würde der Kontrolle der richtig gesetzten Verknüpfung dienen.

4.2 Arbeitspakete:

Vorprojekt 10%

Im Rahmen des Vorprojekts sollen die Grundfunktionen des Softwarepaketes implementiert werden. Zielsetzung ist es ein OntoWiki System aufzusetzen und einen Versuchsdatensatz mit geringem Umfang einzufügen um die Funktionsweise von OntoWiki zu testen und dem Auftraggeber Grundsätzliche Funktionsweisen demonstrieren zu können.

Auf der anderen Seite soll ein einfaches Frontend die Grundfunktionalität der Darstellung der Daten für den Web User zeigen und einfache Suchfunktionen demonstriert werden können. Nähere Informationen befinden sich im Dokument Vorprojekt.

Datentransformation/ Haltung 40%

Ein Schwerpunkt der Arbeit wird der Umgang mit den Daten sein, da die Datengrundlage der Schlüssel für das Funktionieren des Softwarepakets darstellt. Je mehr Zeit in diese Grundlagenarbeit investiert wird, desto größer sind die Möglichkeiten später das System zu erweitern.

Teilaspekte sind die Modellierung der Datenstruktur als Ontologie in enger Zusammenarbeit mit dem Auftraggeber. Hierbei soll durch Vokabular-Alignment eine größtmögliche Kompatibilität mit bestehenden semantischen Web Ressourcen ermöglicht werden.

Das Hauptaugenmerk muss hier auf Personen, Institutionen, Zeitpunkt, berufliche Laufbahn sowie Fachgebiete gesetzt werden, da diese die Hauptgrundlage für die Recherchemöglichkeiten liefern.

Optional soll versucht werden einen möglichst großen Teil der weiteren Daten semantisch zu spezifizieren.

Dann muss eine Schnittstelle aufgebaut werden, die es ermöglicht die Altdaten (vorhanden in HTML-Struktur) zu importieren und semantisch aufzuwerten.

Backendentwicklung 30%

Die Entwicklung des Backends umfasst das Aufsetzen des Frameworks, Anpassung und Implementierung von Widgets zur Dateneingabe und Darstellung der Daten im Backend, Spezifizierung und Realisierung von Schnittstellen und Implementierung von Serverfunktionalitäten.

Weiterhin ist hier die Implementierung von optionalen in den Funktionalitäten formulierten Umsetzungen von Anforderungen zu nennen.

Frontendentwicklung 30%

In diesem Arbeitspaket soll die Darstellung für den Web User realisiert werden. Hier liegt das Hauptaugenmerk auf die schlichte, übersichtliche Darstellung der Daten und die Implementierung der spezifizierten Suchmöglichkeiten sowie optionaler Funktionalitäten für das Frontend wie graphische Darstellungen.

Einrichten der Software auf dem Server der SAW und Übergabe 10%

Der letzte Teil der Arbeit wird die Portierung des Softwaresystems auf die Server der SAW und die eventuelle Schulung des Personals.

5. Qualitätssicherung

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.

Insbesondere ist eine Rückkopplung zum Auftraggeber und eine in einem iterativen Prozess stattfindende Anpassung an die Bedürfnisse des Auftraggebers unabdingbar um einen schnellen Übergang in den produktiven Einsatz der Applikation zu ermöglichen. Die Umsetzung des Projekt Catalogus Professorum Lipsiensis hat gezeigt, dass der Umfang dieser Arbeit nicht zu unterschätzen ist (Vgl. Riechert et al. 2010). Dieser Punkt wiegt schwerer als die Effizienz, da ohne Zuverlässigkeit das Produkt nicht einsatzfähig ist.

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. Zusätzlich ist es durch die geplante Architektur ohne viel Aufwand möglich vorhandene Schnittstellen für die Übertragung auf Folgeprojekte zu gewährleisten.

Die Produktqualität muss dem Projekt angemessen sein und den produktiven Einsatz ermöglichen.

Übersicht der Qualitätsanforderungen:

	Sehr gut	Gut	Normal	Weniger Relevant
Produktqualität			✓	
Funktionalität		✓		
Zuverlässigkeit	✓			
Benutzbarkeit		✓		
Effizienz				✓
Wartbarkeit			✓	
Übertragbarkeit				✓

Quellen:

Riechert, Thomas/Morgenstern, Ulf/Auer, Sören/Tramp, Sebastian/Martin, Michael: Knowledge Engineering for Historians on the Example of the *Catalogus Professorum Lipsiensis*. In: P.F. Patel-Schneider et al.(Eds.): ISWC 2010, Part II, LNCS 6497, pp. 225–240, 2010.

6. Glossar

Altdaten

Das sind alle Daten, die bisher erfasst wurden und die in Form der bestehenden HTML-Daten zur Verfügung gestellt werden.

Backend

Das Backend ist der Kern des Systems, das die Datenverwaltung und Datenhaltung realisiert. Dem Administrator und den Content Editoren werden hier alle sie betreffenden Funktionalitäten zur Verfügung gestellt.

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

Frontend

Das Frontend ist die Schnittstelle des Systems zu dem Web User. Hier werden die Daten für die Öffentlichkeit visuell dargestellt.

OntoWiki ist ein Wikisystem, in dem alle Daten als Tripel gespeichert werden, was ihre Verknüpfungen ermöglicht. Es ist auch ein Werkzeug, das der kollaborativen Arbeit mehrerer Nutzer dient. Web User können verschiedene Inhalte anlegen und editieren und auch dazu Annotationen bzw. Kommentare schreiben.

SPARQL Endpoint

Durch diese Schnittstelle wird es externen Projekten ermöglicht direkt auf die Daten zuzugreifen.

Vokabular bildet eine Menge von gewählten Bezeichnungen, welche die Datenstruktur in einer Ontologie beschreiben.

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.