

SUI II

Weiterentwicklung der diensteorientierten Infrastruktur des Umweltinformationssystems Baden-Württemberg für die semantische Suche nach Umweltinformationen

U. Bügel; F. Chaves

*Fraunhofer Institut für Optronik, Systemtechnik und Bildauswertung
Fraunhoferstr. 1
76131 Karlsruhe*

C. Döpmeier; T. Schlachter; R. Weidemann

*Karlsruher Institut für Technologie
Institut für Angewandte Informatik
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen*

M. Briesen; G. Lukács

*disy Informationssysteme GmbH
Erbprinzenstr. 4-12
76133 Karlsruhe*

R. Ebel; B. Ellmenreich; M. Tauber

*LUBW Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg
Griesbachstr. 1
76185 Karlsruhe*

A. Sawade

*Ministerium für Umwelt, Naturschutz und Verkehr Baden-Württemberg
Kernerplatz 9
70182 Stuttgart*

1. EINLEITUNG.....	45
2. ANALYSE VON ZIELSYSTEMEN.....	45
3. ÜBERARBEITUNG DER SYSTEMARCHITEKTUR	47
4. NOTWENDIGE ERWEITERUNGEN AM UIS-BERICHTSSYSTEM	48
5. ARBEITEN AM ONTOLOGIE-SYSTEM	48
6. ZUSAMMENFASSUNG.....	49
7. LITERATUR.....	50

1. Einleitung

Im Projekt SUI Phase I („Semantische Suche in Umweltinformationssystemen“) wurden Konzepte für die Verbesserung der Suche nach Informationen in Umweltportalen, insbesondere im Umweltportal Baden-Württemberg, auf Basis von Technologien des Semantic Web erarbeitet /1/, /2/. In der prototypischen Umsetzung konnten nicht alle hierfür notwendigen, sondern nur ausgewählte, unbedingt erforderliche Komponenten implementiert werden. Eine wesentliche Voraussetzung zur vollständigen Umsetzung des SUI-Konzeptes ist die noch fehlende Entwicklung einer generischen Schnittstellenspezifikation, um an das Portal gerichtete Suchanfragen an ausgewählte Zielsysteme (unter notwendiger Transformation z.B. der Suchparameter) weiterzuleiten. Zur Erfüllung dieser Aufgabe sollen zunächst ausgewählte Zielsysteme analysiert, Schnittstellen zu diesen formuliert und dann der Mechanismus zur Definition der Schnittstellen vereinheitlicht werden. Der vorliegende Beitrag reflektiert den aktuellen Stand der Diskussion in der gegenwärtig laufenden Projektphase II.

2. Analyse von Zielsystemen

Als Zielsysteme wurden insbesondere das auf Cadenza Web basierende UIS-Berichtssystem BRSWeb, reine Websysteme wie Themenpark Umwelt und Fachdokumentendienst (FADO) sowie die Volltextsuchmaschine Google Search Appliance (GSA) auf ihre Schnittstellen hin untersucht. Als zentrales Ergebnis der Analyse ausgewählter Selektoren des UIS-Berichtssystems geht hervor, dass die generische Serviceschnittstelle eines Zielsystems die folgenden Funktionalitäten berücksichtigen muss:

- Eine Spezifikation von Eingabeparametern mit der Klassifizierung
 - Konkreter Sachbezug wie Luft-Messstelle, Naturschutzgebiet etc.
 - Raumbezug (Auswahl über Koordinaten, Verwaltungseinheit, evtl. fachliches Objekt)
 - Zeitbezug (Zeitpunkt, Zeitraum)
 - Volltextsuchbegriff oder Schlagwortliste (als einschränkender Filter für die Eingrenzung der Suchergebnisse zu einem konkreten Sachbezug), z.B. über Angabe eines Namens (z.B. Schutzgebietsname)
- Eine Spezifikation des Ausgabeformates sowohl in technischer Form (text/html, application/xml) als auch in inhaltlicher Hinsicht. Die Schnittstellenspezifikation sollte dabei in inhaltlicher Hinsicht sowohl abstrakte als auch sehr konkrete Formate erlauben
 - z.B. RSS, GeoRSS, MediaRSS oder
 - Geodatenformate KML, WMS-Layerspezifikationen, Legato-Layerbeschreibungen

Die Analyse der Suchschnittstellen von Websystemen im Umweltinformationssystem Baden-Württemberg (UIS BW) ergibt, dass solche Systeme in der Regel eine reine Volltextsuche anbieten und diese evtl. auf inhaltliche Bereiche beschränken können. Raum- oder Zeitbezug besitzen diese Suchinterfaces in der Regel nicht.

Vergleicht man diese Anforderungen zur semantischen Beschreibung von Suchschnittstellen der Zielsysteme mit dem Beschreibungsformat für Suchschnittstellen von Websystemen gemäß dem OpenSearch-Standard [3], so fällt auf, dass die *OpenSearch Description Document* Spezifikation bereits den größten Teil dieser Beschreibungsfunktionalität enthält. Deshalb soll in SUI II bei der Beschreibung der Suchschnittstellen von Zielsystemen möglichst auf den OpenSearch-Standard aufgesetzt werden. Ein mögliches Beispiel für die Spezifikation der Such-Schnittstelle eines Zielsystems (hier BRS) gemäß OpenSearch Description zeigt der Sourcecode in Abbildung 1.

```
<?xml version="1.0" encoding="UTF-8"?>
<OpenSearchDescription xmlns="http://a9.com/-/spec/opensearch/1.1/">
  <ShortName>UDO</ShortName>
  <Description>Definition von Suchmöglichkeiten in UDO</Description>
  <Tags>Schutzgebiet Landschaftsschutzgebiet Naturschutzgebiet Biotop Bannwald Schonwald</Tags>
  <Contact>admin@lubw.baden-wuerttemberg.de</Contact>
  <Url type="text/atom+xml"
      template=
"http://brsweb.lubw.baden-wuerttemberg.de/brs-web/alias.cweb?repositoryId=${uis:concept}_nach_${uis:locType}&UDDI=true&value=${uis:locationId}"&format=rss/>
  <Url type="text/html"
      template=
"http://brsweb.lubw.baden-wuerttemberg.de/brs-web/alias.cweb?repositoryId=${uis:concept}_nach_${uis:locType}&UDDI=true&value=${uis:locationId}">
  <LongName>Umweltdaten und Karten Online Suche nach Schutzgebieten</LongName>
  <Image height="64" width="64" type="image/png">http://www.lubw.baden-wuerttemberg.de/websearch.png</Image>
  <Image height="16" width="16" type="image/vnd.microsoft.icon">http://www.lubw.baden-wuerttemberg.de/websearch.ico</Image>
  <Query role="example" uis:concept="naturschutzgebiet" uis:locType="gemeinde" uis:locationId="1234567">
  <Query role="example" uis:concept="landschaftsschutzgebiet" uis:locType="kreis" uis:locationId="80086">
  <Developer>disy</Developer>
  <Attribution>
  Search data Copyright 2010, www.lubw.baden-wuerttemberg.de, All Rights Reserved
  </Attribution>
  <SyndicationRight>open</SyndicationRight>
  <AdultContent>>false</AdultContent>
  <Language>de</Language>
  <OutputEncoding>UTF-8</OutputEncoding>
  <InputEncoding>UTF-8</InputEncoding>
</OpenSearchDescription>
```

Abbildung 1: Beispiel einer OpenSearch Description

Die angegebenen URLs für den Aufruf enthalten Platzhalter der Gestalt $\{searchTerms\}$, $\{uis:concept\}$, $\{uis:locType\}$ und $\{geo:uid\}$, über die Suchparameterwerte mit fest definierter Semantik beim URL-Aufruf an das Zielsystem übergeben werden können. Die Namen der eigentlichen Parameter innerhalb des URL-Aufrufs des Zielsystems sind dabei beliebig. Standardplatzhalter der OpenSearch-Spezifikation wie „searchTerms“ zur Angabe von Volltextsuchbegriffen können ohne Präfix und Angabe eines Namensraums verwendet werden. Die im obigen Beispiel gezeigten Platzhalter mit Präfix „uis“ sind dagegen von ihrer Semantik her SUI- bzw. UIS-spezifisch und müssen daher gemäß der OpenSearch-Spezifikation durch einen eigenen, frei definierbaren Namensraum gekennzeichnet sein. Der Platzhalter „geo:uid“ stammt aus einer OpenSearch-Erweiterung, die vom OGC (Open Geospatial Consortium) im Rahmen des Katalog-Standards definiert wurde, und beschreibt Suchparameter für eine räumliche Suche in Geodaten systemen. Die OGC-Erweiterung deckt alle Anforderungen aus SUI II in Bezug auf räumliche Parameter ab. Eine Erweiterung für die zeitbasierten Parameter, die alle SUI-Anforderungen abdeckt, gibt es ebenfalls.

3. Überarbeitung der Systemarchitektur

Da bei einigen der gewünschten Zielsysteme kurzfristig keine größeren Änderungen vorgenommen werden können, müssen diese so angesprochen werden, wie sie Suchanfragen augenblicklich verstehen. Dies erfordert eine Vorverarbeitung der Anfrage, bei der die im Portal vom Benutzer eingegebenen Suchparameter so aufbereitet werden, dass nur noch Zielsystem-spezifische Parameter an das eigentliche Zielsystem gehen. Die dafür in der Architektur modellierte Vorverarbeitungseinheit „Search Broker“ zwischen der Portalapplikation und dem Zielsystem überführt die eingegebenen Suchparameter zunächst in von den Zielsystemen verstandene Parameter (Abbildung 2). Diese nutzt hierfür sowohl externe Dienste, wie einen Geonamensdienst (Gazetteer Service), der Namen als Ortsnamen interpretiert und genauere Informationen zum Ort zurückgibt, als auch das Ontologiesystem als weitere Quelle, um z.B. den Sachbezug von Suchbegriffen aufzulösen. Das Ontologiesystem muss nicht mehr – wie in SUI I – die gesamte Interpretationsarbeit der Suchparameter leisten, sondern nur noch die semantische Interpretation der Konzept-bezogenen Suchparameter.

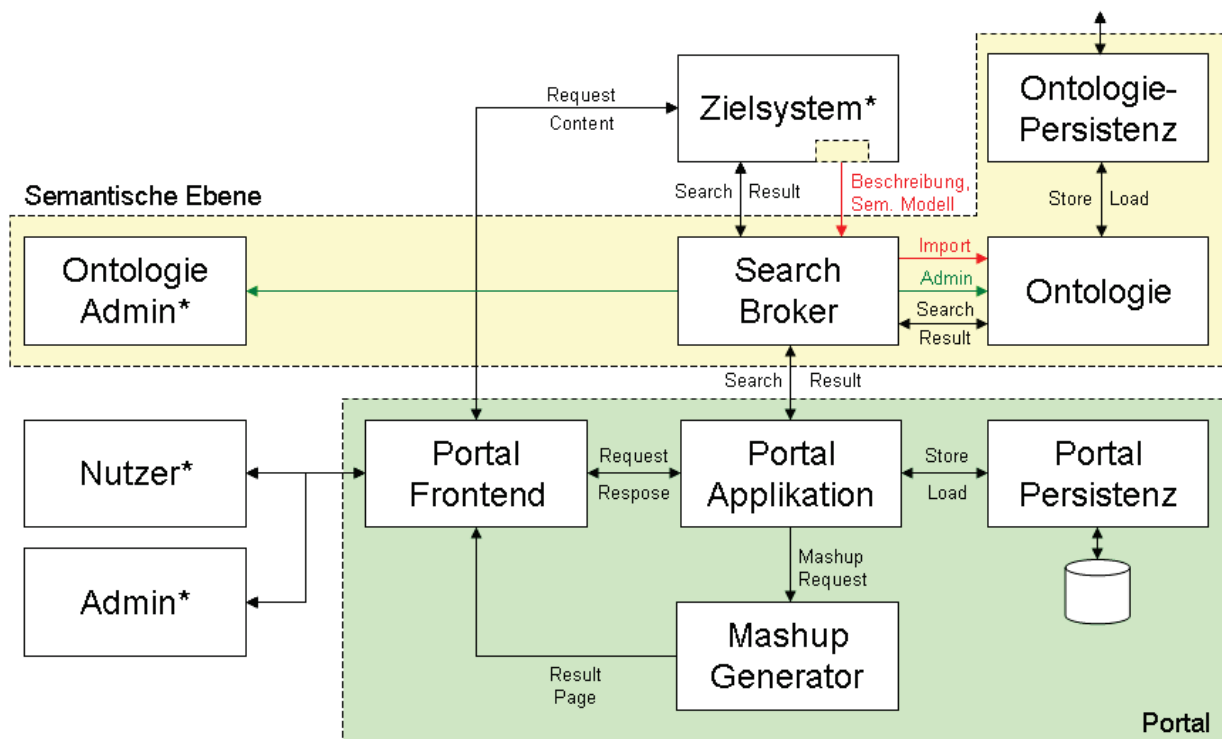


Abbildung 2: Architektur mit Vorverarbeitungseinheit „Search Broker“

4. Notwendige Erweiterungen am UIS-Berichtssystem

Das auf disy Cadenza aufbauende Berichtssystem ermöglicht die Abfrage und die Darstellung von in Datenbanken gehaltenen strukturierten Daten. Für die Integration müssen zwei Arten von Anforderungen bzw. Aufgaben erfüllt werden: solche, die für andere Informationsquellen mit statischen Informationsobjekten – z.B. Dokumente, Bilder, Videos – auch notwendig sind, und solche, die nur für Systeme relevant sind, die diese Informationsobjekte auch dynamisch, z.B. auf Basis einer Datenbank, erstellen.

Als erste Aufgabe bei der Integration jedes Zielsystems müssen geeignete Beschreibungen der Informationsobjekte (z.B. der Cadenza-Selektoren) erstellt werden. Dabei sind drei Arten von Informationen zu berücksichtigen:

1. Informationen für die Recherche (Begriffe, räumliche und zeitliche Abdeckung),
2. Informationen für die Darstellung der Trefferliste (Titel, Beschreibung) und
3. technische Informationen für den Durchgriff auf die Detailergebnisse.

Die zweite allgemeine Aufgabe ist es, eine OpenSearch-konforme Suchmaschine, möglichst auch mit Raum- und Zeitkriterien, über die Beschreibungen zu implementieren.

Eine mächtige Schnittstelle zum Aufruf der Detailergebnisse wird für solche Systeme benötigt, die die Detailergebnisse dynamisch erstellen. Im Fall von disy Cadenza bedeutet dies eine grundlegende Erweiterung der Dienstschnittstelle, mit mächtigen Möglichkeiten zur Parametrisierung der Abfrage und einer flexiblen Darstellung der Bildschirm-Elemente.

Im Rahmen des Projektes werden alle drei Teile für disy Cadenza konzipiert.

5. Arbeiten am Ontologie-System

Das in SUI I entwickelte Ontologie-System wird derzeit nach WebGenesis® Version 8.1 portiert. Die Teil-Ontologien (Lebenslagen, Umweltthemen, Zielsysteminformation, Raumbezug, Zeitbezug) werden gemäß den Anforderungen aus der Zielsystemanalyse für bestimmte Szenarien weiterentwickelt bzw. in Dienste des Search Brokers ausgelagert. Der gewählte Ansatz sieht vor, dass grundsätzlich bereits vorhandene semantische Informationen (z.B. Objektartenkatalog, Semantic Network Services) nativ genutzt und über Import-Schnittstellen in ontologische Elemente überführt werden. Für die Integration der Teil-Ontologien werden zusätzliche Funktionen wie Mapping/Merging von Ontologien entwickelt (Abbildung 3).

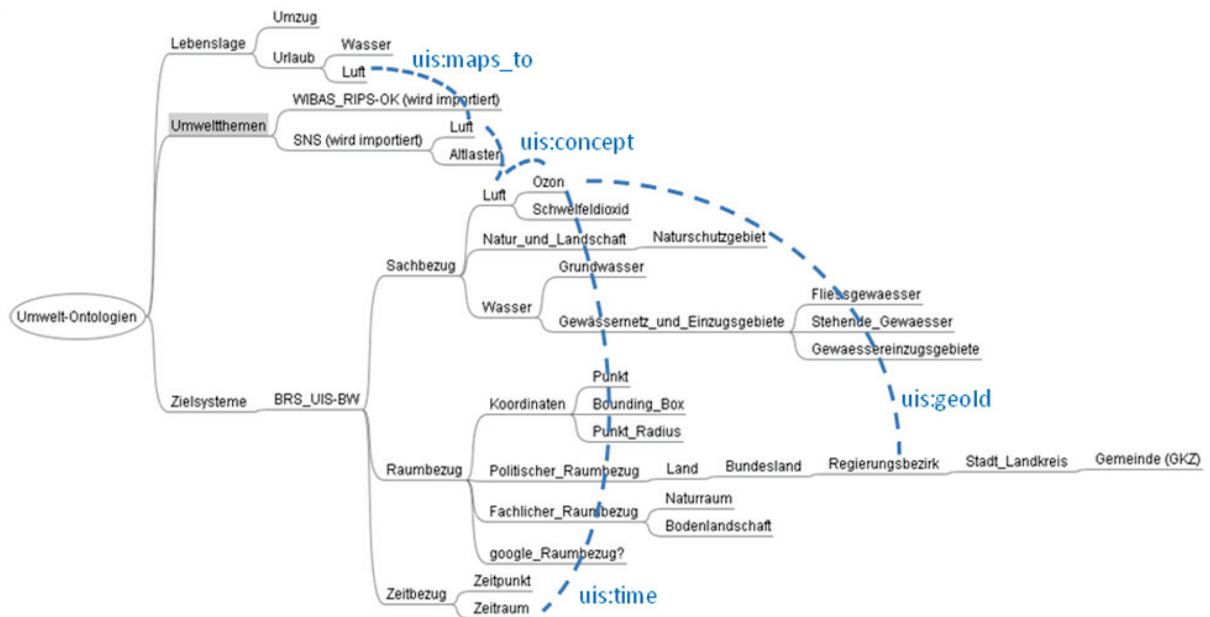


Abbildung 3: Aufbau und Vernetzung der Teilontologien

Die OpenSearch-Beschreibungen der Zielsysteme werden als Annotationen in die Ontologie überführt. Die in den „URL“-Tag-Elementen beschriebenen Abfrage-Schemata werden dabei mit Hilfe der Information aus den „Query“-Tags den Konzepten der Zielsystem-Ontologien zugeordnet und in Relationen der Ontologie gespeichert. Das Beispiel in Abbildung 3 zeigt, wie eine Suche zum Thema Luft mit Ortsangabe auf eine Abfrage des Selektors „Luftmessdaten“ im Zielsystem BRS_UIS-BW mit Hilfe dieser Relationen transformiert werden kann. Aus der Ontologie heraus können somit konkrete Abfragen an Zielsysteme attribuiert und geliefert werden.

6. Zusammenfassung

Im Rahmen eines ersten Schrittes der SUI-Projektphase II wurden ausgewählte Zielsysteme analysiert und aufbauend auf diesen Ergebnissen eine OpenSearch-konforme Beschreibungsschnittstelle zur Zielsystemabfrage definiert. Als nächstes muss nun diese Semantik in Portal, Ontologie und Berichtssystem realisiert werden.

7. Literatur

- /1/ Abecker, A. et al. (2009): SUI – Ein Demonstrator zur semantischen Suche im Umweltportal Baden-Württemberg. In: Mayer-Föll, R., Keitel, A., Geiger, W.; Hrsg.: Kooperative Entwicklung wirtschaftlicher Anwendungen für Umwelt, Verkehr und benachbarte Bereiche in neuen Verwaltungsstrukturen, Phase IV 2008/09, Forschungszentrum Karlsruhe, Wissenschaftliche Berichte, FZKA 7500, S. 157-166.
- /2/ Workshop Umweltgoogle – Landesumweltportale heute und morgen, <http://www.fachdokumente.lubw.baden-wuerttemberg.de/servlet/is/94087/?COMMAND=DisplaySonstigeMedien&FIS=90934&OBJECT=94087&MODE=METADATA>
- /3/ OpenSearch Spezifikation, <http://www.opensearch.org/>
- /4/ Walch, J. et al (2009): OpenGIS® OpenSearch Geospatial Extensions Draft Implementation Standard. Open Geospatial Consortium Inc., Document Reference Number OGC 09-084r1, OpenGIS® Discussion Paper.