

# LUPO

## Fortgeschrittene Suchfunktionen in den Landesumweltportalen von Baden-Württemberg, Sachsen-Anhalt und Thüringen

*T. Schlachter; W. Geiger; R. Weidemann; G. Zilly  
Forschungszentrum Karlsruhe GmbH  
Institut für Angewandte Informatik  
Hermann-von-Helmholtz-Platz 1  
76344 Eggenstein-Leopoldshafen*

*R. Ebel; M. Tauber  
Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg  
Griesbachstr. 1  
76185 Karlsruhe*

*A. Sawade; R. Mayer-Föll; K. Zetzmann  
Umweltministerium Baden-Württemberg  
Kernerplatz 9  
70182 Stuttgart*

*V. Bachmann; B. Köther  
Ministerium für Landwirtschaft und Umwelt des Landes Sachsen-Anhalt  
Olvenstedter Str. 4  
39108 Magdeburg*

*D. Keil  
Thüringer Ministerium für Landwirtschaft, Naturschutz und Umwelt  
Beethovenstr. 3  
99096 Erfurt*

*T. Sattler; M. Trumpler  
DECON-network Systemhaus & EDV-Vertriebs GmbH  
Bannwaldallee 24  
76185 Karlsruhe*

<b>1. LANDESUMWELTPORTALE .....</b>	<b>151</b>
<b>2. VOLLTEXTSUCHE MIT DER GOOGLE SEARCH APPLIANCE .....</b>	<b>151</b>
<b>3. FORTGESCHRITTENE SUCHFUNKTIONEN .....</b>	<b>152</b>
3.1 SUCHE IN DATENBANKEN .....	152
3.2 KEYMATCHES .....	152
3.3 ONEBOXES.....	153
3.4 UMWELTTHESAURUS ZUR ERWEITERUNG DER SUCHANFRAGE.....	154
3.5 SCHLAGWORTE AUS DEM SNS.....	154
3.6 VERWENDUNG VON METADATEN .....	155
<b>4. ERFAHRUNGEN BEIM BETRIEB.....</b>	<b>155</b>
<b>5. AUSBLICK.....</b>	<b>156</b>
<b>6. LITERATUR.....</b>	<b>156</b>

# 1. Landesumweltportale (LUPO)

Die Landesumweltportale der Länder Baden-Württemberg, Sachsen-Anhalt und Thüringen bieten sowohl Fachleuten als auch der interessierten Öffentlichkeit einen zentralen Einstiegspunkt inklusive Suchfunktionen zu Umweltinformationen dieser Länder. Damit leisten sie einen wesentlichen Beitrag zu der in den Umweltinformationsgesetzen geforderten aktiven Verbreitung von Umweltinformationen.

Im Jahr 2008 war die bis dahin in den Landesumweltportalen verwendete Suchmaschine ht://Dig durch die leistungsfähigere Google Search Appliance (GSA) abgelöst worden /1/. Bis zur Jahresmitte 2008 wurden im Rahmen des Kooperations-Projekts „LUPO“ zunächst die Kernfunktionalitäten (Volltextsuche) umgesetzt. Während der Laufzeit von KEWA IV sind darüber hinausgehende, fortgeschrittene Suchfunktionen entwickelt worden. Dabei wurde jeweils die Prämisse berücksichtigt, dass auch unerfahrene Benutzer Suchanfragen so einfach wie möglich formulieren können sollen.

# 2. Volltextsuche mit der Google Search Appliance

Mit der GSA steht in den Umweltportalen eine Suchmaschine zur Verfügung, die in ihrer Funktion und Leistungsfähigkeit der bekannten Google-Suchmaschine entspricht. Dies bedeutet, dass zur Formulierung einer Volltext-Suchanfrage lediglich eines oder mehrere Suchwörter in ein Formularfeld („Suchschlitz“) einzutragen sind, wobei auch die Realisierung komplexerer Suchformulare möglich ist.



Abbildung 1: Volltext-Trefferliste im Umweltinformationsnetz Sachsen-Anhalt

Die Suchanfrage produziert eine Trefferliste, die aus Verweisen zu den zuvor indizierten Dokumenten besteht. Jeder einzelne Treffer wird in der Regel durch den Titel des Dokuments, einen Volltext-Ausschnitt der Trefferstelle(n) (Snippet) oder eine Kurzbeschreibung dargestellt, teilweise auch durch weitergehende Informationen zum Format des Dokuments (Icon) oder zum Stand des Dokuments (s. Abbildung 1).

### **3. Fortgeschrittene Suchfunktionen**

Nachdem am Ende der Projektphase KEWA III die von ht://Dig bekannte Funktionalität abgelöst war, ging es in der aktuellen KEWA-Phase darum, die weitergehenden Möglichkeiten der GSA auszuloten und zu implementieren. Dabei sind wesentliche neue Funktionen entstanden. Zudem konnten viele weitere Datenquellen in die Volltextsuche einbezogen werden, deren Indizierung bisher aus technischen Gründen nicht möglich war.

#### **3.1 Suche in Datenbanken**

Ein wesentliches Manko der bisherigen Suchmaschine war es, dass Daten ausschließlich per Volltext-Crawler erschlossen werden konnten, d.h., es war nur möglich, Dokumente zu indizieren, die per Link auf einem statischen Webserver oder in einem Web-Content-Management-System (WCMS) erreichbar waren. Von der Indizierung ausgeschlossen waren dagegen Datenbankinhalte und solche Webangebote, deren Inhalte dynamisch nach einer Formularanfrage erzeugt werden.

Mit der GSA ist es nun möglich, auf verschiedene Datenbanksysteme (z.B. Oracle, MySQL, DB2, Sybase) zuzugreifen und die Inhalte von DB-Queries direkt zu indizieren. Zusätzlich ist es möglich, in den DB-Queries enthaltene URL als Eingabe für die Volltextindizierung zu verwenden. Dadurch können Webanwendungen, die sich hinter komplexen Auswahllisten oder Formularanfragen verbergen, mit der Volltextsuche erschlossen werden.

Beispielhaft sei hier der „Daten- und Kartendienst“ der LUBW Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg /2/ genannt: Über einen eigenen Datenbank-View werden z.B. Informationen zu verschiedenen Schutzgebiets-Typen, Geotopen und Biotopen indiziert. Die Treffer der Ergebnisliste verweisen jeweils auf Seiten der entsprechenden Webanwendung. Neben dem Vorteil, diese Seiten nun in die Volltextsuche integriert zu haben, werden durch die parallele Indizierung von Datensätzen und Seiteninhalten auch passende Treffer gefunden, die den Suchbegriff gar nicht in der HTML-Seite enthalten, z.B. Kartendarstellungen von Schutzgebieten.

#### **3.2 KeyMatches**

KeyMatches sind eine Möglichkeit der GSA, bestimmten Suchbegriffen besonders relevante Webangebote zuzuordnen. Die von Redakteuren in der GSA gepflegte Liste dieser KeyMatches sorgt dafür, dass – getriggert durch den Suchbegriff – ein entsprechender Hinweis neben der eigentlichen Volltext-Trefferliste erscheint (Abbildung 2). Damit ist es zum Beispiel möglich, neue oder besonders häufig nachgefragte Angebote an prominenter Stelle zu präsentieren.

## Suchergebnisse zu "feinstaub "

Versuchen Sie es einmal hier:  
[Feinstaub-Messwerte in Baden-Württemberg](#)

Treffer 1 bis 10 von insgesamt ca. 11300 Ergebnissen.

[Feinstaub \( Landesanstalt für Umwelt, Messungen und Naturschutz > ...](#)  
**Feinstaub ... Feinstaub. Feinstaub** ist in den zurückliegenden Jahren zu einem umweltpolitischen Dauerthema geworden. Viele Bürgerinnen ...  
<http://www.lubw.baden-wuerttemberg.de/servlet/is/20250/>

Abbildung 2: KeyMatch (nach „Versuchen Sie es einmal hier:“) zum Suchbegriff „Feinstaub“

### 3.3 OneBoxes

Bei den OneBoxes handelt es sich um eine GSA-spezifische Funktionalität: Parallel zur Suche im Volltextindex der GSA wird die Suchanfrage (mit den entsprechenden Suchbegriffen) an weitere Systeme weitergegeben. Die GSA erwartet die Antworten dieser Systeme innerhalb eines konfigurierbaren Zeitlimits in einem definierten OneBox-XML-Format, das generisch genug ist, um verschiedenste Inhalte zu transportieren.

Alle rechtzeitig eingegangenen Informationen zu maximal fünf OneBoxes werden zusammen mit der Volltext-Trefferliste ausgeliefert und können durch die Darstellungskomponente (im Falle von LUPO das WCMS WebGenesis®) verwendet und visualisiert werden. Hierdurch ist es möglich, die Suchergebnisse mit vielfältigen passenden Inhalten anzureichern.

The image shows a grid of search results for the term 'feinstaub'. Each result is presented in a distinct box with a title and a list of items. The boxes are: 'AKTUELLE PEGELSTÄNDE' (river levels), 'NEU AUS DEM BILDARCHIV' (bird photo), 'LUFTMESSSTATIONEN' (air quality), 'FEINSTAUB-MESSWERTE' (dust measurements), 'OZON-MESSWERTE' (ozone measurements), 'AKTUELLE PEGELSTÄNDE' (Elbe river levels), 'NEU SCHUTZGEBIET-BILDER' (landscape photo), and 'TREFFER AUS DATENBANKEN' (nature protection areas).

Abbildung 3: Beispiel-OneBoxes aus den LUPO-Umweltportalen

Beispielhaft seien hier einige OneBoxes genannt (s. Abbildung 3): Aktuelle Pegelstände in Baden-Württemberg und Sachsen-Anhalt, Bildarchiv (Arten und Schutzgebiete), Luftmessstationen in Baden-Württemberg, Ozon- und Feinstaubmesswerte in Sachsen-Anhalt, Datenbanktreffer in Baden-Württemberg.

### 3.4 Umweltthesaurus zur Erweiterung der Suchanfrage

Die GSA enthält Standard-Wörterbücher (z.B. für Deutsch, Englisch und Französisch), anhand derer Suchanfragen automatisch erweitert werden können. Der Suchbegriff „gehen“ wird beispielsweise nach dem deutschen Wörterbuch auf die Flexionen „gehe“, „gehst“, „geht“, „ging“, „gegangen“ etc. ausgedehnt, d.h., auch Dokumente, die ausschließlich eine oder mehrere dieser Beugungen enthalten, werden gefunden.

Neben den Standard-Wörterbüchern kann die GSA auch weitere, frei definierbare Wörterbücher enthalten.

Der Umweltthesaurus des Umweltbundesamtes (UmThes) bietet für den Umweltbereich ein definiertes Wortgut mit einer Vielzahl von Verknüpfungen zwischen den verschiedenen Begriffen. Dazu gehören zum Beispiel Ober- und Unterbegriffe sowie Synonyme.

Durch Konvertierung des Umweltthesaurus in das GSA-Wörterbuch-Format konnte eine Untermenge des UmThes für die GSA verfügbar gemacht werden, so dass neben dem Suchbegriff nun automatisch die entsprechenden verwandten Suchbegriffe gefunden werden. Zum Beispiel liefert die Suche nach dem umgangssprachlichen „Müll“ auch Treffer in Dokumenten, welche die Begriffe „Abfall“ oder „Schrott“ enthalten.

### 3.5 Schlagworte aus dem SNS

Bereits vor Einführung der GSA war in den Landesumweltportalen eine Funktion zum Hinweis auf verwandte Suchbegriffe vorgesehen. Hierfür wurde innerhalb der Semantic Network Services (SNS) /3/ des Umweltbundesamtes eine Funktion „SimilarTerms“ entwickelt, welche zu einer Reihe von Suchbegriffen möglicherweise interessante weitere Suchbegriffe liefert, die der Nutzer als Eingabe für die Volltextsuche verwenden kann.

Zur Nutzung in den Landesumweltportalen wurde die Funktion „SimilarTerms“ um das OneBox-Ausgabeformat erweitert. Zusammen mit manuell in der GSA zu pflegenden „Passenden Suchanfragen“ kann auf diese Art und Weise dem Nutzer ergänzend zur automatischen Erweiterung der Suchanfrage aus dem Umweltthesaurus (s.o.) eine qualifizierte Liste von möglichen weiteren Suchanfragen geboten werden (s. Abbildung 4). Der Klick auf einen der dargestellten Suchbegriffe löst eine weitere Suchanfrage unter Verwendung dieses Begriffes aus.

#### Zu "feinstaub" passende Suchbegriffe:

⊕ Abrieb, ⊕ Asbestfeinstaub, ⊕ Asbestfeinstaubemission, ⊕ Dieselruß, ⊕ Feinpartikel, ⊕ Feinstaubabscheider, ⊕ Feinstaubabscheidung, ⊕ Feinstaubausstoß, ⊕ Feinstaubbelastung, ⊕ Feinstaubemission, ⊕ Feinstaubgrenzwert, ⊕ Feinstaubpartikel, ⊕ Feinstaubplakette, ⊕ Feinste Staubpartikeln, ⊕ Feinststaub, ⊕ Flugstaub, ⊕ Flugstaubmessgerät, ⊕ Nanopartikel, ⊕ Nanoskalierter Staub, ⊕ Nanoskalierter platinhaltiger Staub, ⊕ Nanoskaliger Staub, ⊕ PM0,1, ⊕ PM1, ⊕ PM10, ⊕ PM2,5, ⊕ PMx, ⊕ Partikelemission, ⊕ Platin-Feinstaub, ⊕ Schwebstaub, ⊕ Ultrafeine Feststoffpartikel, ⊕ Ultrafeinstaub, ⊕ Zweiundzwanzigste BImSchV

Abbildung 4: Passende Suchbegriffe aus dem SNS und der GSA zum Suchbegriff „Feinstaub“

## 3.6 Verwendung von Metadaten

Neben dem eigentlichen Inhalt enthalten viele Dokumente zusätzliche Metadaten, die z.B. Informationen zu den Autoren oder zum Stand eines Dokuments, dessen Titel, eine Kurzbeschreibung oder eine Liste von Schlagworten liefern können. Die GSA wertet insbesondere die in HTML-Dateien enthaltenen Metadaten aus. Diese können grundsätzlich bei der Suche verwendet werden. In den Landesumweltportalen wird von dieser Funktionalität derzeit nur spärlich Gebrauch gemacht, z.B. um in der Trefferliste – falls vorhanden – statt eines Snippets die Kurzbeschreibung des Dokuments anzuzeigen.

Im Themenpark Umwelt /4/, der ebenfalls den GSA-Index für die Volltextsuche verwendet, werden solche Metaangaben zur Erzeugung einer erweiterten GSA-Trefferliste verwendet, in der z.B. zu jedem Treffer auch passende Icons, Grafiken oder Fotos eingeblendet werden.

## 4. Erfahrungen beim Betrieb

Die Erweiterung der Volltextsuchfunktionen in den Landesumweltportalen konnte nach Inbetriebnahme der GSA Schritt für Schritt erfolgen. Im Falle der KeyMatches und OneBoxes musste zuerst der Zugriff auf diese Daten ermöglicht werden. Hierzu wurde eine Java-Schnittstelle implementiert, die aus den WebGenesis-Templates, welche die Anzeige der Suchergebnisse steuern, angesprochen und abgefragt werden kann. Dieser Mechanismus ist so generisch, dass beliebige weitere OneBoxes mit verschiedensten Inhalten ohne Änderung dieser Schnittstelle angeschlossen werden können.

Die Konstellation, neben der Betriebsversion der GSA noch eine Backup-Maschine als Entwicklungsserver zu betreiben, hat sich in mehrfacher Hinsicht bewährt: Tests von Neuentwicklungen sind so risikofrei möglich und beim bisher einzigen Ausfall der Live-Maschine durch einen Hardware-Defekt konnte der Betrieb dank der Backup-Maschine nach einer Umkonfiguration aufrecht erhalten werden.

Eine Anforderung zur Erweiterung der GSA-Schnittstelle für WebGenesis ergab sich aus dem laufenden Betrieb: Die Möglichkeit zur (Vor-)Konfiguration aller möglicher Parameter und Einstellungen rund um das Suchformular. Hier wurde ebenfalls ein generischer Mechanismus geschaffen, um in einer Properties-Datei beliebige Konfigurationen ablegen und innerhalb von WebGenesis verwenden zu können, so dass selbst innerhalb derselben WebGenesis-Instanz verschiedene Formulare mit den entsprechenden Vorbelegungen konfigurierbar sind. Die GSA-Suche für WebGenesis wird derzeit in Baden-Württemberg auch auf den Websites des Umweltministeriums, der LUBW und der Gewerbeaufsicht verwendet sowie im Themenpark Umwelt und im Intranet der LUBW. Die Konfiguration der GSA beim Anlegen neuer Einträge in den Landesumweltportalen wird über den halbautomatischen Versand von Emails mit den entsprechenden Konfigurationsdaten erleichtert.

Im Sinne der Wiederverwendbarkeit des GSA-Index in anderen Portalen, speziell dem Umweltportal Deutschland PortalU<sup>®</sup>, wurde eine OpenSearch-Schnittstelle für die GSA entwickelt, die eine kaskadierende Suche im PortalU ermöglicht. Somit entfällt die Notwendigkeit einer Mehrfachindizierung der Inhalte durch PortalU und die Landesumweltportale, wodurch sich die Last auf den einzelnen Servern reduzieren wird.

## 5. Ausblick

Neben den Entwicklungen, die direkt in und an der LUPO-Software stattgefunden haben, gibt es weitere Pläne zum Ausbau der Umweltportale: Das Suchen und Finden von Daten mit Geobezug und die semantische Unterstützung der Volltextsuche.

Eine ganz wesentliche Anforderung ist die Möglichkeit zur Suche nach und zur Darstellung von Umweltinformationen mit Geobezug. Viele Umweltdaten haben einen solchen Geobezug, der explizit oder implizit in diesen Daten vorhanden ist. Explizit kann sich dieser Geobezug z.B. durch das Vorhandensein von Metadaten zur geografischen Lage ausdrücken oder ist per se, z.B. bei Kartenlayern, vorhanden. Ein impliziter Bezug kann sich zum Beispiel durch das Vorhandensein von geografischen Bezeichnern in den Metadaten (Gemeindenamen, Gewässernamen, Namen von Schutzgebieten) ausdrücken. Zur Darstellung von Geodaten in den Landesumweltportalen steht ein von der Fa. disy Informationssysteme GmbH entwickelter Kartenclient „Legato“ auf Basis von OpenLayers zur Verfügung, der auf dem Entwicklungsserver bereits prototypisch eingebunden ist.

Mit einem Workshop zum Szenario „What’s in my backyard?“ startete das Projekt SUI zur semantischen Suche in Umweltinformationen. Ziel dieses Projekts ist u.a. die Einbindung einer semantisch unterstützten Volltextsuche in die Landesumweltportale. Die erreichten Ergebnisse sind in einem eigenen Beitrag beschrieben /5/.

All diese neuen Techniken erschließen weitere mögliche Datenquellen, die den derzeitigen Stand von rund 1.200.000 verfügbaren Dokumenten in den drei Landesumweltportalen nicht nur in quantitativer Hinsicht ergänzen sollen. Hierdurch sollen die Attraktivität des Angebots und die Anzahl von derzeit monatlich ca. 50.000 Nutzern (gegenüber ca. 8.000 Nutzern vor Einführung der GSA) weiter gesteigert werden.

## 6. Literatur

- /1/ Schlachter, T. et al. (2008): Landes-Umweltportale – Vernetzung von Informationen in den Umweltportalen von Baden-Württemberg, Sachsen-Anhalt und Thüringen unter Einsatz einer kommerziellen Suchmaschine. In: Mayer-Föll, R., Keitel, A., Geiger, W.; Hrsg.: F+E-Vorhaben KEWA – Kooperative Entwicklung wirtschaftlicher Anwendungen für Umwelt, Verkehr und benachbarte Bereiche in neuen Verwaltungsstrukturen, Phase III 2007/08, Forschungszentrum Karlsruhe, Wissenschaftliche Berichte, FZKA 7420, S. 63-76.
- /2/ Daten- und Kartendienst der LUBW,  
<http://brsweb.lubw.baden-wuerttemberg.de/brs-web/>
- /3/ Rüther, M., Bandholtz, T., Menger, M. (2006): SNS Environmental Vocabulary – from Terms to Ontology. Conf. Semantics 2006, From Visions to Applications - Semantics: The New Paradigm Shift in IT, Wien, 28.-30. November 2006.
- /4/ Döpmeier, C. et al. (2009): Themenpark Umwelt – Optimierung der Volltextsuche und Integration von Panoramabildern und Live-Videos. In diesem Bericht.
- /5/ Abecker, A. et al. (2009): SUI – Ein Demonstrator zur semantischen Suche im Umweltinformationssystem Baden-Württemberg. In diesem Bericht.