

Konzept für eine neue Generation von Landes-Umweltportalen

R. Weidemann¹, W. Geiger¹, T. Schlachter¹, G. Zilly¹,
R. Ebel², K. Zetzmann³

¹ Forschungszentrum Karlsruhe, IAI
² Landesanstalt für Umwelt, Messungen und Naturschutz
Baden-Württemberg
³ Umweltministerium Baden-Württemberg

EnviroInfo 2008
Workshop „Umweltinformationssysteme
der Öffentlichen Verwaltung“

Übersicht

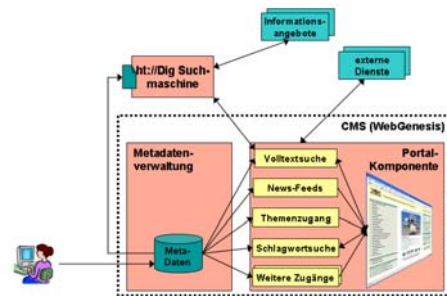
- Motivation
- Struktur von Landes-Umweltportalen
- Anforderungen und Lösungen
- Stand / Erfahrungen



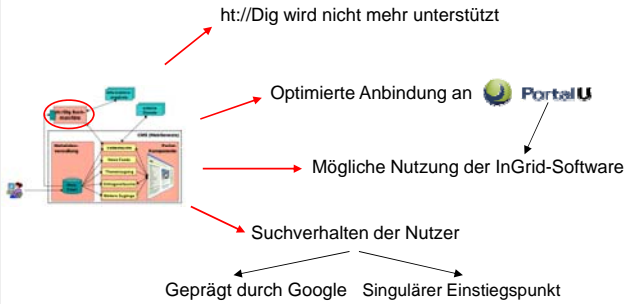
Motivation für Landes-Umweltportale

- Umweltinformationsgesetze verpflichten Behörden, Umweltinformationen aktiv zugänglich zu machen
 - Verschiedene Arten von Informationen (Fachdokumente, Mess- und Sachdaten, Geoinformation, Multimedia, ...)
 - Verteilt auf viele unterschiedliche Web-Angebote
- Ziele der Landes-Umweltportale
 - Singulärer Einstiegspunkt
 - „Einfacher Zugang“ (Zielgruppe: Öffentlichkeit)
 - Umfassender Zugang zu behördlichen, auch kommunalen Informationsangeboten

Bisherige Struktur der Landes-Umweltportale



Gründe für eine Restrukturierung



5 | R. Weidmann | FZKI/IA | 11.09.2008

KIT – die Kooperation von
Forschungszentrum Karlsruhe GmbH
und Universität Karlsruhe (TH)



Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft

Alternativen zur Volltextsuche mit ht://Dig



1. Austausch durch funktional gleichwertige Suchmaschine
 - Lucene/Nutch, dtSearch
 - Kein wesentlicher Mehrwert zu erwarten
2. Einsatz einer kommerziellen Suchmaschine
 - Leistungsfähigere Kategorie (Performanz, DBs, Feeds)
 - Google Search Appliance, Oracle Secure Enterprise Search
3. Einsatz der InGrid-Software
 - Problemlose Integration mit PortalU
 - Mehrfachnutzung von Daten und Diensten
 - Viele weitere Komponenten

- Funktionalität?
- Durchgängiges Konzept?

- Produktreife?
- Komplexität?

6 | R. Weidmann | FZKI/IA | 11.09.2008

KIT – die Kooperation von
Forschungszentrum Karlsruhe GmbH
und Universität Karlsruhe (TH)



Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft

Google Search Appliance (GSA)



- Hard-/Softwarelösung
- Einfaches Lizenzmodell (URLs)
- Skalierbarkeit
- Produktreife, Nutzerakzeptanz
- Funktionalität



Google Search Appliance; © picasa.google.de

7 | R. Weidmann | FZKI/IA | 11.09.2008

KIT – die Kooperation von
Forschungszentrum Karlsruhe GmbH
und Universität Karlsruhe (TH)



Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft

Neues Konzept auf Basis der GSA



- Zwei Alternativen
 1. Beibehaltung der bisherigen Struktur unter Austausch der Volltext-Suchmaschine oder
 2. Rein GSA-basierte Lösung (ev. mit Portal-Framework oder direkte Einbettung in Landes-CMS)
- Entscheidungskriterium
 - Gibt es für Landes-Umweltportale essentielle Funktionen, die mit Alternative 2 nicht realisiert werden können?
- Randbedingung
 - GSA sollte möglichst schnell in den Produktionsbetrieb
- Lösungsweg
 - Stufenweises Vorgehen, ausgehend von Alternative 1 ...

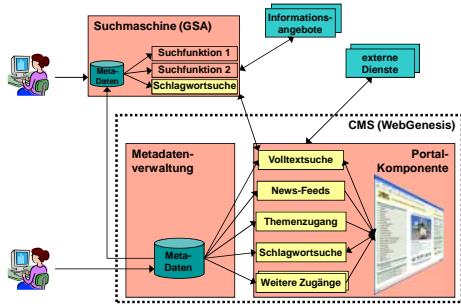
8 | R. Weidmann | FZKI/IA | 11.09.2008

KIT – die Kooperation von
Forschungszentrum Karlsruhe GmbH
und Universität Karlsruhe (TH)



Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft

Sukzessive Nutzung der GSA-Funktionen



Anforderungen und Lösungen



- Typische Probleme bei der Entwicklung von Landes-Umweltportalen und beispielhafte Lösungen mit der GSA

Einbindung strukturierter Daten

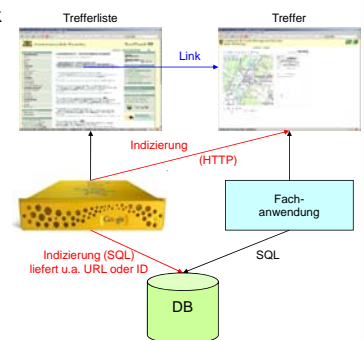


- Anforderung
 - Durchsuchen von Datenbankinhalten
 - Einbeziehung von nicht direkt indizierbaren Objekten (Karten, Bilder, Media-Dateien) durch Anreicherung mit Metadaten
- Mögliche Randbedingungen
 - Reine Datenbankinhalte, nur über SQL-Schnittstelle erreichbar
 - Über Webanwendung verfügbare Inhalte, verborgen hinter Formularen
 - Über Webanwendung verfügbare Inhalte, statisch verlinkt

Indizierung von DB-basierten Webanwendungen



- Indizierung der Datenbank und der Webanwendung
- Treffer verweisen auf Seiten der Webanwendung



Auffinden von Kartendarstellungen

- Beispiel
 - Umweltdatenbanken und -karten online (UDO) auf Basis von disy CadENZA

Trefferliste im Portal Umwelt-BW

Suchergebnisse zu "naturschutzgebiet weingarten"

Treffer 1 bis 10 von insgesamt ca. 295 Ergebnissen

- 1. [Naturschutzgebiete der Gemeinde Weingarten \(Baden\)](#)
- 2. [Naturschutzgebiete der Gemeinde Weingarten \(Baden\), Naturschutzgebiet, NBS, Naturschutzgebietsname, Umgebungsname, Weingarten Moor-Bruchwald Grötzingen, ...](#)
- 3. [http://www13.nrw.baden-wuerttemberg.de/inf/wald/bw/006b/region/bio/-/Schutzgebiete/LandN...](#)
- 4. [Naturschutzgebiet Weingarten Moor - Bruchwald Grötzingen ...](#)
- 5. [Das Naturschutzgebiet Weingarten Moor - Bruchwald Grötzingen liegt im Stadt- und Landkreis Karlsruhe auf dem Gemarkungen Grötzingen und Weingarten ...](#)
- 6. [\(04.10.2009\)](#)
- 7. [http://www13.nrw.baden-wuerttemberg.de/inf/wald/bw/006b/region/bio/-/Schutzgebiete/LandN...](#)



Treffer in UDO

Einbindung des Umweltdatenkatalogs (UDK)

- Anforderung
 - Einbindung von UDK-Objekten in die Suche
- Randbedingung
 - Pflege der UDK-Daten in PortalU oder
 - Pflege der UDK-Daten in eigener InGrid-Installation
- Lösungsmöglichkeiten
 - Direkte Indizierung der Datenbank (s.o.)
 - Indizierung der Objektseiten in PortalU / InGrid
 - Separate Suche in PortalU / InGrid und Zusammenführung der Suchergebnisse

Suchen in Teilräumen

- Anforderung
 - Beschränkung der Suche auf definierte Teilräume, z.B.
 - Angebote bestimmter Informationsanbieter (oder Länder)
 - Einzelne Fachsysteme
 - Themengebiete
- Lösungsmöglichkeiten
 - Zusätzliche, verdeckte Suchbegriffe
 - Separate Indizierung der Teilräume
 - Definition der Teilräume über URL-Muster

Sammlungen / Collections in der GSA

- Definition über
 - URL-Listen (Positiv- und Negativ-Listen)
 - Teilstrings, reguläre Ausdrücke



Semantisch-unterstützte Suche



- Anforderung
 - Seiten sollen auch gefunden werden, wenn nicht die exakten Suchbegriffe verwendet werden:
 - Verschiedene Wortformen
 - Teilworte
 - Umgangssprachliche Begriffe, Synonyme und andere verwandte Begriffe
 - Schreibfehler
- Lösungsmöglichkeiten
 - Reduktion der Inhalte und der Suchbegriffe auf Stammformen
 - Hinterlegung von Wortpaaren (ähnliche Begriffe)
 - Einsatz von Thesauri bei der Indizierung oder Suche
 - Generierung und Prüfung ähnlicher Begriffe über Regeln
 - typische Schreibfehler
 - phonetisch ähnlich

Einbindung des UBA-Thesaurus



- Umweltschlesaurus des Umweltbundesamtes als „Wörterbuch“ für die automatische Erweiterung der eingegebenen Suchbegriffe

{ Trinkwasseruntersuchung, Trinkwasseranalyse,
Trinkwassergütekontrolle, Trinkwasserkontrolle }

Ultrafeinstaub = Nanopartikel

Zersiedlung = Zersiedelung

{ Kohlendioxid, Carborain, CO2,
Kohlendioxidbestimmung, Kohlendioxidemission,
Kohlendioxidhaushalt, Kohlendioxidproblem,
Kohlendioxyd, Kohlenstoffdioxid, Kohlendioxidausstoß }

Zusätzliche Suchergebnisse über Thesaurus



- Automatische Erweiterung der Suchanfrage

Kontextabhängige Informationsangebote



- Anforderung
 - Getriggert durch bestimmte Suchbegriffe sollen vorgegebene Informationen an prominenter Stelle angezeigt werden, z.B.
 - Zentrale Veröffentlichungen der Organisation, z.B. Umweltplan
 - Sicherheitsrichtlinien
 - Aktuelle Werte (Ozon, Hochwasserpegel)
- Lösungsmöglichkeiten
 - Hinterlegung von Schlüsselwort - URL-Paaren
- Lösungen GSA
 - „Keymatch“: Zuordnung von Suchbegriffen zu Internetseiten
 - „One-Box“: Zuordnung von Suchbegriffen und Anwendungen (z.B. separate Suche oder Datenbankabfrage)

Keymatch

Suche nach „Brummtön“

Suchergebnisse zu „Brummtön“

21 | R. Weidmann | FZKJM | 11.09.2008

One-Box (Prototyp)

Suche nach „Nagold“

Nagold

22 | R. Weidmann | FZKJM | 11.09.2008

Schnittstellen zur Einbettung der Suche

- Anforderungen
 - Einbetten eines Suchformulars in Web-Anwendung (CMS)
 - Möglichkeit zur Weiterverarbeitung / Einbettung der Ergebnisliste
 - Administration der Suchmaschine manuell / programmgesteuert
- Lösungsmöglichkeiten
 - Einbetten des Suchformulars
 - Aufruf der Suche über parametrisierte URL
 - Webservice
 - Ergebnisliste
 - (X)HTML, XML, OpenSearch
 - Webservice
 - Administration
 - Web-Oberfläche
 - API, Webservice
 - Parameterdateien

Stand des Projekts September 2008

- Einsatz einer GSA für drei Bundesländer (plus Backup)
- Indizierte Seiten ca. 500.000 ↗
- GSA integriert in die Landes-Umweltportale von Baden-Württemberg, Sachsen-Anhalt und Thüringen
- Landes-Umweltportale von BW und ST sind online
- Verschiedene Collections werden für speziell-zugeschnittene Suchfunktionen eingesetzt
- Fortgeschrittene Funktionalitäten der GSA werden evaluiert und sukzessive in den Produktionsbetrieb überführt

Erfahrungen



- **Positiv**
 - Performanz
 - Stabilität
 - Aufwand zur Nutzung der Grundfunktionalität
 - Akzeptanz bei den Nutzern und Administratoren
 - Leistungsfähige Analysewerkzeuge
 - Synergieeffekt bei der technischen Optimierung des Contents (SEO)
- **Negativ**
 - GSA ist nur sehr eingeschränkt mandantenfähig
 - Keine Programmschnittstellen von Google für die Administration
 - Kompatibilitätsprobleme bei Upgrades



- **Vielen Dank für Ihre Aufmerksamkeit!**
- **Die Landes-Umweltportale im Internet:**
 - <http://www.umwelt.baden-wuerttemberg.de>
 - <http://www.umwelt.sachsen-anhalt.de>
- **Die GSA-Suche ist auch hier integriert:**
 - <http://www.um.baden-wuerttemberg.de>
 - <http://www.lubw.baden-wuerttemberg.de>