

# Recommender-gestützte Erweiterung der Discovery in Geoportalen

Bernhard VOCKNER und Manfred MITTLBÖCK

## Zusammenfassung

Geoportale stellen zentrale Zugriffspunkte auf geographische (Metadaten-)Information einer Geodateninfrastruktur (GDI) dar. Die Suche in online verfügbaren Geoportalen ist zumeist auf Ergebnisse einer textbasierten Suche nach Schlüsselwörtern, Titel und Kurzbeschreibungen sowie die Einbeziehung räumlicher und zeitlicher Filter beschränkt. Außen vor bleiben Kontext des Nutzers sowie die Ergebnisse von Suchen anderer Nutzer. Deshalb wird im Rahmen des EU FP7 Projektes EnerGEO der Ansatz verfolgt, durch die Einbeziehung von Empfehler-Systemen ähnlich jener in Online-Versandhäusern zusätzliche, nicht per se aus der Suchanfrage ableitbare Ergebnisse in einer Form wie „*Andere Nutzer die diese Ressource betrachteten interessierten sich auch für...*“ bereitzustellen.

## 1 Rahmenbedingungen

Über die Jahre hinweg haben sich Geoportale als zentrale Zugriffspunkte auf geographische (Metadaten-)Information einer Geodateninfrastruktur (GDI) etabliert (MASSER 2007). Sie ermöglichen mitunter die Eingabe von Metadaten und die Suche nach Inhalten in diversen Metadatenkatalogen. Im Rahmen des EU FP7 Projektes EnerGEO (BLASCHKE et al. 2010) wurde ein Geoportal erstellt, mit dem Zweck, Informationen über bestehende geographische Ressourcen aus der Energie-Domäne einer breiten Nutzergruppe unter Zuhilfenahme effizienter Suchmechanismen zur Verfügung zu stellen. Die eingegebenen Metadaten werden in einer XML-Struktur in einer Datenbank vorgehalten und entsprechen den Vorgaben von ISO 19113, ISO 19114, ISO 19115, ISO 19119, ISO 19139 und der INSPIRE Metadaten Direktive. Die OGC CSW (Catalogue Service Web) Schnittstelle des Open Source Produktes „ESRI Geoportal Server“ ermöglicht technische Interoperabilität und durch so genannte Harvesting-Mechanismen die Vernetzung zwischen unterschiedlichen Katalogen.

## 2 Herausforderungen

Neben technischer Interoperabilität wird in der Umsetzung des EnerGEO Geoportals besonderes Augenmerk auf semantische Interoperabilität gelegt. Das spiegelt sich nicht zuletzt in der erweiterten semantischen Suchmöglichkeit geographischer Ressourcen wider. Semantik spricht die Bedeutung von verwendeten Konzepten zur Metadatenbeschreibung an. Diese Bedeutungen und Wertzuweisungen können je nach fachspezifischer Domäne, Vorwissen, etc. stark variieren und kommen gerade in einer heterogen zusammengesetzten Community wie in EnerGEO zu tragen. Ein damit einhergehender Nachteil ist, dass mögli-

cherweise nicht alle Ergebnisse gefunden werden können, wenn sich die bei der Beschlagwortung verwendeten Konzepte unterscheiden (SMITS & FRIIS-CHRISTENSEN 2007). Dem damit angesprochenen Problem der semantischen Heterogenität kann unter anderem durch die Verwendung von Thesauri und durch die Einbeziehung des Kontexts des Nutzers (z. B. aktueller Standort, verwendete Suchsprache) entgegengewirkt werden (u. a. BUCCELLA et al. 2008).

Als Erweiterung schlagen wir deshalb vor, wie im proof-of-concept für das EnerGEO-Geoportale gezeigt wird, Empfehler-Systeme in den Discovery-Prozess zu integrieren, um automatisiert nicht per se in den Daten vorhandene, sondern von anderen Geoportale-Nutzern erstellte Verknüpfungsmöglichkeiten aufzuzeigen.

### 3 Recommender-gestützte Suchergebnisse

Geoportale stellen standardbasierte Lösungen zur Erfassung von Metadaten und zum Auffinden geographischer Information dar. Die Suche in derartigen Geoportale-Produkten ist dabei zumeist auf Ergebnisse einer textbasierten Suche nach Schlüsselwörtern, Titel und Kurzbeschreibungen sowie die Einbeziehung räumlicher und zeitlicher Filter beschränkt. Außen vor bleiben Kontext des Nutzers sowie die Ergebnisse von Suchen anderer Nutzer. Die Aufgabe eines Empfehler-Systems ist es, herkömmliche Suchergebnisse durch die Einbeziehung dieser Faktoren zu verbessern. Recommender ermöglichen es, Nutzern Empfehlungen basierend auf den Suchanfragen anderer Nutzer in einem System, ähnlich jenem von Online-Versandhäusern zu geben. Das gestattet – neben den semantischen Verknüpfungen über Thesauri und Ontologien – eine weitere, nicht aus den Ressourcen selbst ableitbare Verknüpfung interessanter Datensätze, was vor allem für das domänenübergreifende Auffinden von räumlicher Information einen erheblichen Mehrwert darstellt.

Die Praktik der Einbeziehung von Recommender-Systemen hat sich unter anderem bei Internet-Shops als sehr nützlich erwiesen (z. B. „*Andere Käufer haben auch diese Produkte evaluiert, gekauft ...*“), daher unser Ansatz diese Technologie als zusätzliche Komponente für die Discovery in Geoportale zu integrieren. Eine beispielhafte Implementierung wurde im Rahmen des EU FP7 Projektes EnerGEO durchgeführt. Dieses verwendet als Grundlage ESRI Geoportale Server 1.2, dessen Discovery-Komponente durch die Integration des Open Source Recommender-Systems erweitert wurde.

ESRI Geoportale Server (Open Source) als Vertreter von Geoportalen ist eine Lösung, die einen Editor zur standardkonformen Eintragung von Metadaten bietet, welche als XML-Dokumente in einer Datenbank abgelegt werden. Zum Auffinden geographischer Ressourcen wird standardmäßig eine Suchoberfläche bereitgestellt, die diese XML-Dokumente der Datenbank unter Zuhilfenahme effizienter Indizierungsmechanismen durchsucht. Als Erweiterung dieser Suchfunktion kommt in der vorliegenden Implementierung die Open Source Software „easyrec“ (u. a. GSTREIN 2009) als Recommender-Engine zum Einsatz. Diese aus der Web 2.0 Welt entlehnte Technik lässt Rückschlüsse auf Nutzerverhalten („Andere Nutzer interessierten sich auch für...“) zu und stellt darauf basierend Empfehlungen zur Verfügung.

Die in easyrec implementierte und als „Association Rule Miner (ARM)“ bezeichnete Einkaufswagen-Analyse-Engine basiert auf dem Apriori Algorithmus von R (AGRAWAL &

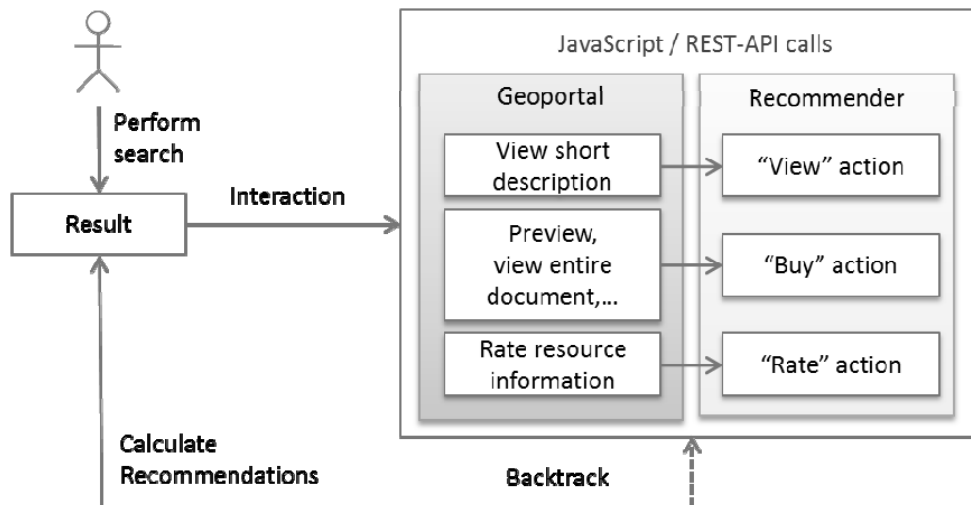
SRIKANT 1994) und „SlopeOne“ (LEMIRE & MACLACHLAN 2005). Apriori ist ein klassischer Lernalgorithmus für Assoziationsregeln eines Recommenders, mit dem Ziel, Zusammenhänge zwischen Artikeln eines Warenkorbes herzustellen. Dadurch lassen sich Aussagen wie „Nutzer die Produkt A und Produkt B in ihren Warenkorb legten, fügten in diesen zu 90 % auch Produkt C“ abbilden. SlopeOne stellt einen Vertreter der artikelbasierten „Kollaborativen Filtermethoden“ dar, die auf der Annahme beruhen, dass sich aufgrund der Verhaltensmuster bestimmter Gruppen auf Interessen Einzelner schließen lässt. In dem von LEMIRE & MACLACHLAN (2005) vorgeschlagenen Algorithmus geht es konkret um die Vorhersage, wie ein Nutzer einen bestimmten Artikel basierend auf den Bewertungen einer Gruppe von Nutzern bewerten würde.

Der in easyrec implementierte Algorithmus betrachtet paarweise Artikel  $\langle X, Y \rangle$ , die mit einer bestimmten signifikanten Wahrscheinlichkeit gemeinsam in verschiedenen Einkaufswagen vorzufinden waren. Übertragen auf das Themengebiet Geoportale lässt sich das Befüllen des (imaginären) Einkaufswagens durch Nutzerinteraktionen wie Anklicken von Ressourcen bewerkstelligen. Eine Feinjustierung der Recommender Engine kann anhand der Angabe der Minimalwerte von zwei Hauptparametern erfolgen: „Support“ ( $\langle X, Y \rangle$ ) und „Konfidenz“ ( $X \rightarrow Y$ ). Support legt fest, wie oft ein Satz an Artikeln  $\langle X, Y \rangle$  zusammen in unterschiedlichen Warenkörben aufscheint, wohingegen Konfidenz die Wahrscheinlichkeit anspricht, dass ein Artikel Y zusammen mit Artikel X auftritt (AGRAWAL & SRIKANT 1994). Die beiden Minimalwerte von Support und Konfidenz legen fest, ob eine Assoziationsregel Berücksichtigung in den Empfehlungen findet oder nicht. Grundsätzlich lässt sich festhalten, dass die Regelsätze bezogen auf ihre Stückzahl weniger und signifikanter werden, wenn die Werte des Supports und der Konfidenz hochgesetzt werden (EASYREC 2011).

## 4 Implementierung

Die Empfehlungen in Recommender-Systemen werden anhand des Betrachtungsverhaltens anderer Nutzer berechnet. Demzufolge wird miteinbezogen, welche Ressourcen von Nutzern näher betrachtet, „gekauft“ oder bewertet wurden. In der vorliegenden Umsetzung werden Klicks auf Suchergebnisse, wie in Markierung 1 in Abb. 2 dargestellt, in Analogie als Betrachtungsinteraktion („view“) verstanden, wohingegen das Ansehen einer Vorschau oder des vollständigen Metadatenatzes als Kaufaktion („buy“) verstanden wird (Abb.1). Werden Ressourcen bewertet („rate“), so wird dies ebenfalls in den Empfehlungsberechnung des Recommenders miteinbezogen. Als weitere Informationen erhält die Recommender Engine Informationen über die ungefähre geographische Position des Nutzers basierend auf dessen IP-Adresse sowie den Nutzernamen, sofern ein Nutzer im Portal angemeldet ist. Das Einloggen eines Nutzers ermöglicht es, nutzerspezifische Empfehlungen basierend auf vorherigen Suchanfragen zu geben. Zusätzlich wird eine eindeutige ID, der Titel der Ressource, der mit der Ressource einhergehende Hyperlink sowie ein Vorschaubild an das Recommender-System gesendet, welche primär der Präsentation der Suchergebnisse des Empfehlens dienen.

Durch eine JavaScript- und REST-API wird die Integration der Recommender-Engine in ESRI Geoportal Server ermöglicht. Sie gestatten es, betrachtete Einträge an das System zu senden, das wiederum automatisiert Empfehlungen berechnet. Diese werden den Nutzern



**Abb. 1:** Workflow des Entstehens von Empfehlungen in der Recommender Engine gesteuert durch Nutzerinteraktionen im Geoportal; Quelle: EIGENE DARSTELLUNG 2012.

nach Klick auf bestimmte Suchresultate im Bereich „Other users also viewed“ zur Verfügung gestellt (Abb. 2) und erlauben das Bereitstellen weiterer potenzieller Resultate, die durch die Suche nach Schlüsselwörtern per se nicht zur Verfügung stünden. Zudem lassen sich auf Wunsch Übersichten gestalten, die die am öftesten betrachteten und am besten bewerteten Ressourcen auflisten.

Die nutzerspezifischen Empfehlungen werden optisch ansprechend in einem so genannten Bildkarussell basierend auf Javascript dargestellt. Dieses enthält einerseits eine Liste der vom Recommender zurückgelieferten Empfehlungen bestehend aus Titel und Vorschau bild der Ressource (sofern vorhanden) bzw. des Typs der Ressource (z. B. Datensatz oder Service). Scroll Buttons ermöglichen es, durch die einzelnen Einträge, die in Fünferpaketen dargestellt werden, zu blättern. Wird auf eine Empfehlung geklickt, so wird dieser Klick durch so genannte „Backtracking“-Mechanismen zurück an die Recommender Engine geschickt (siehe Abb. 2 Markierung 2 bzw. Abb. 1) und fließen als zusätzliche Regelsätze in zukünftige Empfehlungen ein.

## 5 Fazit und Ausblick

Recommender-Systeme stellen eine essenzielle Erweiterung von heutigen Geoportalen dar, da sie die Qualität der Suchergebnisse deutlich verbessern können. Durch die Bereitstellung von zusätzlichen Suchergebnissen, die sich per se nicht aus dem zur Suche verwendeten Schlüsselwort ableiten lassen, können zusätzliche Verknüpfungen zwischen Ressourcen basierend auf den Erfahrungen anderer Nutzer geschaffen und als Resultat präsentiert werden. Die Implementierung im Rahmen des EnerGEO Geoportals hat gezeigt, dass eine Einbeziehung des Kontexts von Nutzern in die Suchresultate diese durch wertvolle Informationen erweitert.

Dennoch existieren Ideen zu einer weiteren Verbesserung der Qualität der Suchergebnisse. Vor allem der Kontext des Nutzers kann durch zusätzliche Elemente (z. B. verwendete Such-Sprache, Fachdomäne) bereichert werden. Des Weiteren ist angedacht, Empfehlungen nicht nur anhand der Klicks anderer Nutzer in die Recommendation Engine einzubeziehen, sondern zusätzliche Information der Metadaten (wie z. B. Kurzbeschreibung, Entstehungshistorie der Ressource) auf semantische Ähnlichkeit zu untersuchen und darauf basierend als weitere Regelsätze in den Recommender zu integrieren um die Qualität der Suchresultatvorschläge weiter zu verbessern.

The screenshot shows the EnerGEO Geoportal interface. At the top, there is a navigation bar with 'HOME', 'SEARCH', 'BROWSE', 'PILOTS', 'MAPPING', and 'MAP VIEWER'. The search bar contains the text 'helioclim' and a 'Search' button. Below the search bar, there are options for 'Records shown from: This Site' and 'Additional Options' including 'WHERE' with radio buttons for 'Anywhere', 'Intersecting', and 'Fully within'. A map of Europe is displayed with a search area highlighted. The search results list includes 'PACA Region (FRANCE) 250 m. resolution Irradiation Map (WMS)', 'Helioclim 3 WPS', and 'Helioclim 3 WMS'. A red box labeled '1)' highlights the 'Helioclim 3 WMS' result. Below the search results, there is a section for 'Other users also viewed' with a red box labeled '2)' containing several recommended items: 'Helioclim-3 Map', 'Helioclim 3 WPS', 'PACA Region (FRANCE) 250 m. resolution Irradiation Map (WMS)', 'Helioclim-3 Free Access', 'Annual BETHY/DLR Above Ground Biomass increase (AGB) for forests (Brandenburg)', and 'Helioclim-1 Surface Solar Radiation'. The footer of the page states: 'This Geoportal was built using the using OpenSource Geoportal Server. Please read the Disclaimer - CSW GetCapabilities'.

**Abb. 2:** Suche im EnerGEO Geoportal, erweitert um Recommender-Empfehlung (Quelle: EIGENE DARSTELLUNG 2012)

## Literatur

- AGRAWAL, R. & SRIKANT, R. (1994), Fast Algorithms for Mining Association Rules in Large Databases. In: VLDB 1994 Proceedings of the 20th International Conference on Very Large Data Bases, 487-499.
- BLASCHKE, T., MITTLBÖCK, M., BIBERACHER, M., GADOCHA, S., VOCKNER, B., HOCHWIMMER, B. & LANG, S. (2010), The GEOSS – EnerGEO portal: towards an interactive platform to calculate, forecast and monitor the environmental impact of energy carriers. In: GREVE, K. & CREMERS, A. B. (Hrsg.) (2010), ENVIRONINFO 2010 – Integration of Environmental information in Europe, Shaker Verlag, Aachen, 2-9.
- BUCCELLA, A., CECHICH, A. & FILLOTRANI, P. (2008), Ontology-driven Geographic Information Integration. A Survey of current Approaches. Computers and Geosciences, DOI: 10.1016/J.Cageo.2008.02.033.
- EASYREC (2011), Association Rule Miner – Online:  
[http://sourceforge.net/apps/mediawiki/easyrec/index.php?title=Association\\_Rule\\_Miner](http://sourceforge.net/apps/mediawiki/easyrec/index.php?title=Association_Rule_Miner)  
(2012-03-30).
- GSTREIN, E. (2009), Adaptive Personalization. A multi view personalization approach incorporating contextual information. Unveröffentlichte PhD Thesis an der TU Wien.
- LEMIRE, D. & MACLACHLAN, A. (2005), Slope One Predictors for Online Rating-Based Collaborative Filtering. SIAM Data Mining (SDM'05), Newport Beach, California, April 21-23, 2005.
- MASSER, I. (2007), Building European Spatial Data Infrastructures. ESRI Press, Redlands, California.
- SMITS, P. & FRIIS-CHRISTENSEN, A. (2007), Resource Discovery in a European Spatial Data Infrastructure. In: IEEE Transactions on Knowledge and Data Engineering, 19 (1), 85-95.