International Journal of Spatial Data Infrastructures Research, 2012, Vol.7, 441-463

Recommender-based enhancement of discovery in Geoportals*

B. Vockner[1], M. Belgiu[2], M. Mittlboeck[1]

[1]Research Studios Austria, Studio iSPACE;
bernhard.vockner@researchstudio.at; manfred.mittlboeck@researchstudio.at

[2]Austrian Academy of Sciences, Institute for Geographic Information Science;
mariana.belgiu@oeaw.ac.at

Abstract

In many cases web search engines like Google are still used for discovery of geographic base information. This can be explained by the fact that existing approaches for Geo-information retrieval still face significant challenges. Discovery in currently available Geoportals is usually restricted to text-based search based on keywords, title and abstract as well as applying spatial and temporal filters. Furthermore, user context as well as search results of other users are not incorporated. In order to improve the quality of search results we propose to extend the suitable searching matches in Geoportals with user behaviour and to present them as non-directly linked recommendations like in e.g. Amazon's *"Customers Who Bought This Item Also Bought"* approach. As shown in the proof-of-concept EU FP7 EnerGEO Geoportal, it guarantees results that are not in the data itself but rather derived from the context of other users' searches and views.

Keywords: recommendation, search, discovery, geoportal, SDI

DOI: 10.2902/1725-0463.2012.07.art20

1. INTRODUCTION

Spatial Data Infrastructure (SDI) creates the premises for developing sustained “spatial information highways” (Strobl and Nazarkulova, 2009), making data accessible to as many people as possible (Sadeghi-Niaraki et al, 2010).

SDI initiatives such as the INSPIRE Directive 2007/2/EC, GMES (Global Monitoring for Environment and Security), GEOSS (Global Earth Observation System of Systems), SEIS (Shared Environmental Information System) powered by standardization efforts of international organizations such as OGC (Open Geospatial Consortium) or ISO/TC211 (International Standardization for Standardization) have established frameworks assisting the communication mechanisms between resource providers and users (Masser, 2007). Within the SDI framework, Geoportals are seen as user interfaces and main entry points to geographic information (Strobl and Nazarkulova, 2009). Geoportals may be defined as Internet or Intranet access point to geographic information. They provide integrated access to published resources, including “maps, applications, geographic web services, analytical models, reports, as well as related text material, dissemination articles and journal papers” (Athanasis et al, 2009). Geoportals represent the appropriate application for sharing spatial information in order to optimize data usage and to avoid data duplication.

Metadata of geographic resources (both spatial data and services), to which queries can be posed, are kept in central or distributed databases. They are – as structured information – computer readable and a suitable basis for information sharing. However, these metadata pools need user interfaces for direct interaction with various users and actors. Geoportals serve as standard user interfaces for SDIs and act as access points to metadata. The aim of these Geoportals is to enable users to discover, use and publish metadata of resources they want to share. Therefore, the central task of Geoportals from the perspective of end users is at least the provision of search facilities for discovery.

Over the last years, an increasing number of Geoportals have been established as main SDI gateways to find, evaluate and start “using” geographic information (Masser, 2007). Based on various technical frameworks (such as Open Source Geonetwork - geonetwork-opensource.org or Geoportal Server - geoportal.sourceforge.net), national, international (e.g. EC INSPIRE Geoportal - http://inspire-geoportal.ec.europa.eu) and domain-specific (e.g. http://energeo.researchstudio.at or http://www.geoportal.org) geoportals exist. Existing Geoportal implementation solutions are coming together with the key capabilities required for developing standardized and integrated spatial infrastructures: interfaces for communicating with other infrastructure components over standardized web protocols, standardized interfaces for catalogue service harvesting and resource discovery. Geoportal Server (http://geoportal.sourceforge.net), which is used within the scope of this paper, is an open source solution for creating metadata, discovering and exchanging geospatial information. Moreover, it enables the development of INSPIRE-compliant discovery services and metadata. All the catalogues mentioned provide searchable repositories of information descriptions and – of course – they do support GI discovery, but retrieval still lacks sufficiency.

A major disadvantage within discovery is that not all resources may be provided as search results if the tagging of concepts differs (Smits and Friis-Christensen, 2007). As search results based on keywords are inherently restricted by the ambiguities of natural languages, semantic heterogeneity problems arise (Lutz et al, 2009). In addition to language, the intrinsic characteristics of domain specific information pose great challenges to cross-domains information search and exchange. Therefore, semantic heterogeneity occurs because geographic communities have various perspectives and according to that they may use different terminology to describe the same type of information. Keyword-based search can have low recall as the terminology used by the resource requester may be different from the terminology used by the resource provider. Recall identifies the completeness of retrieval, meaning that users cannot discover all relevant information sources to answer the question they are aiming at (Smits and Friis-Christensen, 2007). On the other hand, results may have low precision, meaning that some of the discovered geographic information is not relevant which is in many cases combined with too many search results presented to the user (Athanasis et al, 2009).

In the last years, sustained efforts have been oriented towards improving users discovery experience in Geoportals. One solution to the already mentioned semantic heterogeneity problems is the development and usage of Knowledge Organisation Systems – KOS (Latre et al, 2012) in the form of thesauri, ontologies or taxonomies. An increasing number of KOSs are available now either as generic approach, such as EIONET GEMET Thesaurus and EuroVoc Thesaurus or KOSs developed for a specific domain such as drought vocabulary (Latre et al, 2012), land cover (Belgiu et al, 2012) or nature conservation thesaurus (De Martino et al, 2011). Developed thesauri represent an explicit specification of domain conceptualization. Domain concepts are mapped against related concepts defined in other formalized concepts schemas. SKOS (Simple Knowledge Organization System) model for instance is used to specify both hierarchical (general/specific or broader/narrower relations) and associative relations between concepts defined in different Knowledge Organization Systems (consistent schemata). Developed KOS' proved to be efficient in overcoming semantic heterogeneity problems (Latre et al, 2012), but a user-centric and dynamic discovery-solution must tackle the information search and retrieval process independently from the information domain and language used when creating metadata. An automated and enhanced semantic discovery approach that takes information domain and user context into account is required. Therefore we propose the usage of technologies like recommendation engines to improve metadata discovery.

In this paper, the concept of recommendation systems as well as an implementation as proof-of-concept will be shown. An enhanced user interface for discovery will be presented, transparently leveraging the complexity of underlying discovery algorithms. Specific keyword-based searches are enhanced with recommendations based on what other users viewed or queried, improving the quality of search results with content that would not have been retrieved by utilizing textual search only. Recommendations are given by a recommender engine that – in addition to semantic linkages and ontologies – makes the discovery of interesting linkages (that are not derivable from the resources themselves) possible. This is especially important for finding spatial information across domains.

The incorporation of new methods for discovery goes in line with the evolutionary concept of SDIs, where we are currently in the transition from a data-centric or product oriented approach towards a user-centric SDI or process oriented approach (Rajabifard et al, 2002). The new SDI generation shows an increasing interest in users' needs and their experience while interacting with available spatial information (Fernández and Castellanos 2006). We need solutions to collect user behaviour and preferences in Geoportals, because *"once collected, this information can be used to improve search engine results by introducing data on favourite datasets"* (Maso et al, 2012). Van Oort et al (2009) explained how Geoportals could benefit from insights from internet marketing theory and define recommender systems as cross-selling approach or the way of *"persuading a visitor or customer to buy or access another associated products"* (van Oort et al, 2009). In this paper, the implemented recommendation system is seen as a contribution to the information sharing transparency and geo-products advertising, and a dynamic mechanism for providing users hints on what might interest them.

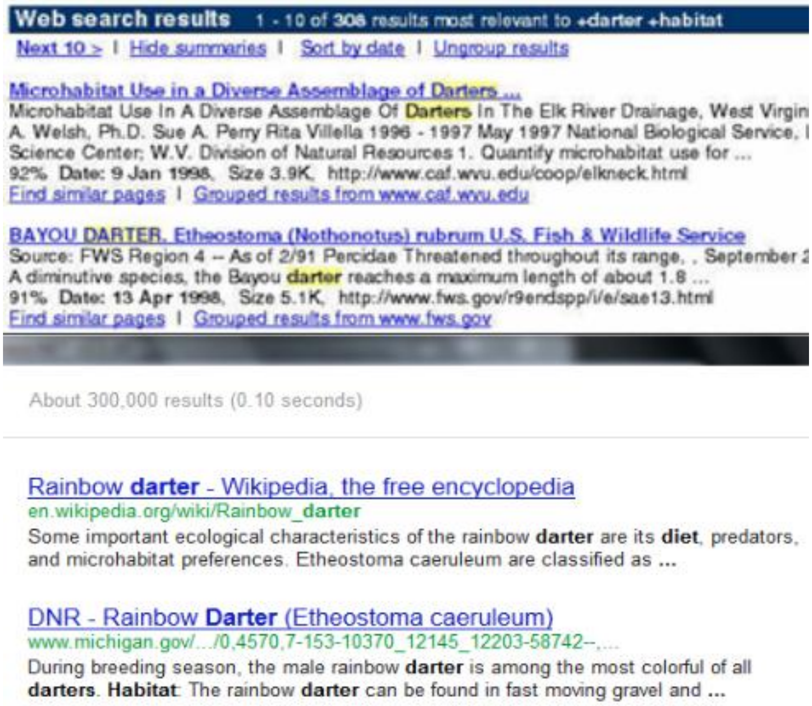
The remainder of this paper is structured as follows. Section 2 provides a short overview of information discovery in the WWW. A short presentation of recommendation search engine functionalities is given in section 3, followed by recommendation strategies within geo-environment introduced in section 4. After a short description of the geoportal implemented within the EU FP7 EnerGEO project together with its improvements in terms of spatial resources discovery experience, section 5 is focusing on the deployed recommendation system and its underlying principles. Section 6 is dedicated to discussions and outlook.

2. INFORMATION DISCOVERY IN THE WWW

Information discovery in the WWW has become a very popular research area as the amount of information to be found is highly increasing. At the beginning of the 1990s catalogues were mainly used to discover information. These catalogues were simply collections of bookmarks maintained by a team of human editors – quite similar to the process of a librarian. The commented bookmark collections were divided into categories and sub-categories. The most prominent examples of these web catalogues include Lycos (founded in 1994) and Yahoo (founded in 1995). Because it was cost-intensive to maintain the resources on the Internet that were growing rapidly, in 1998 Larry Page and Sergey Brin introduced Google where the system itself (the so-called Googlebot) collected and indexed websites instead of human beings. Up to now the basic concept of Google remained unchanged. It uses an algorithm called "PageRank" (Brin and Page, 1998) to rank websites according to their linkages to other websites. It enables full-text search and is based on frequency of occurrence of the searched word(s), position of the search terms (meta-tags, title) as well as user click and rate actions on search results. A single search result (sometimes also called 'search hit') contains a summary of the resource. This concept of result presentation is called 'document surrogate' and has proven successful in its implementation in web search engines.

This simple way of result presentation has not changed much over the years which can be shown by comparing Infoseek in 1997 and Google in 2012 which use nearly the same layout and look (see Figure 1). The reasons why the standard interface is that simple and has remained the same are that users performing a specific task do not want to be distracted by the interface and due to the fact that the WWW is used by a variety of users of all ages, cultures and backgrounds (Hearst, 2009).

Figure 1: Search Results Listings from Infoseek in 1997 (above) and Google in 2012 (below)

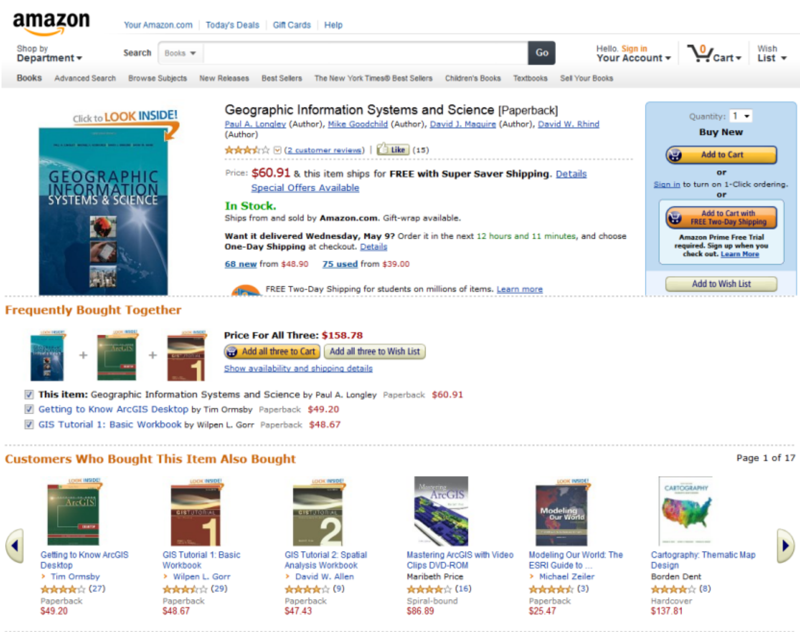


Source: Print Screen of Infoseek Provided by Hearst, 2009

Within the field of information discovery, modern search engines such as Google or Bing mainly use keyword-based algorithms for retrieval of information items as these have shown to be the most efficient and effective in a general-purpose search (Lagoze and Sinhal, 2005). In addition to text-based search, modern search engines usually provide some filters (e.g. type of content, language, temporal extent) and auto suggestion options (such as a dropdown list of elements like “energy”, “energy saving trust” or “energy comparison” when entering the word “ener” for “energy”) while typing in keywords.

Apart from text-based search, online-shops like Amazon have introduced recommendations for their users. This enables the possibility to present their customers additional products they did not think of when searching for an item. For example, if a user is looking for the book “Geographic Information System and Sciences”, the recommender suggests he might also be interested in the book “Getting to know ArcGIS Desktop” as other users often bought these items together and therefore a high likelihood of the items to have something in common exist (Figure 2). Recommendations offer another possibility of information discovery that is thought to be used within Geoportals which from a historical viewpoint are more related to catalogues like Lycos at the beginning of the 1990s than Google at the end of the 1990s. Therefore recommendations can be used to incorporate the benefits of common web technologies to “upgrade” Geoportals to what users experience in their everyday lives when using search engines or online stores. Like recommendations of books are useful for a person looking for specific types of books, we come up with the idea that it might also be useful for a person looking for spatial information.

Figure 2: Recommendations as Used in Amazon.com



3. RECOMMENDATIONS ENHANCING WEB DISCOVERY

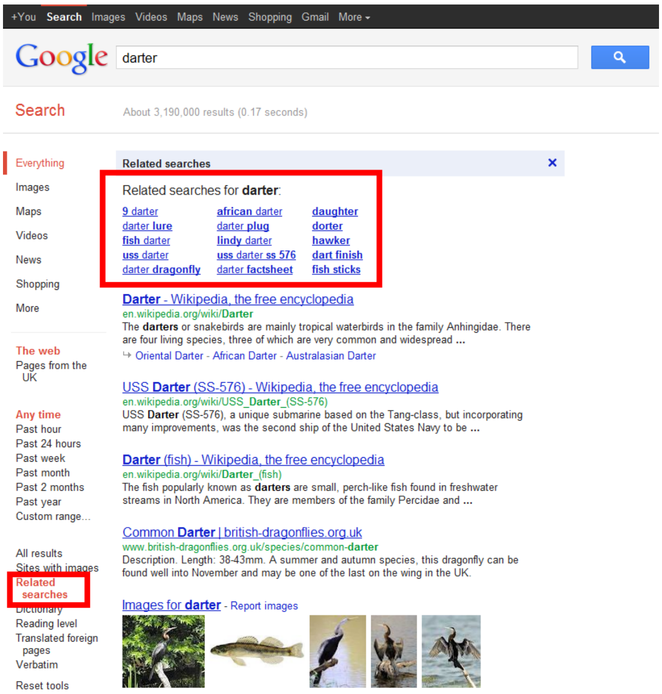
As the name already implies, recommendation engines facilitate the process of discovery giving users meaningful recommendations on what might interest them based on previous users’ behaviour (items previously viewed, bought or rated) and searches of other users. Or in other words: Recommenders provide *“suggestions for items to be of use to a user”* (Ricci et al, 2011; p. 1).

“Items” are all kind of things like books, CDs, PDFs, newspaper articles or in case of Spatial Data Infrastructures and Geoportals they can be thought of spatial resources (e.g. data, services). Recommendations are thought to be provided especially in cases when there is an enormous amount of potential search results. The online-store Amazon is a perfect example for this due to its huge amount of products available.

The most common techniques that are implemented in recommender systems are collaborative filtering (Resnik et al, 1994; Sarwar et al, 2001; Linden et al, 2003) and content-based filtering (Shardanand and Maes, 1995; Balabanović and Shoham, 1997; Pazzani, 1999). The combination of both is referred to as hybrid approaches. Collaborative filtering approaches analyse previous interactions while content-based filtering systems consider attributes of user profiles (Melville and Sindhwani, 2010). Collaborative filtering is dependent on a huge amount of users and tries to predict preferences of users by comparing them with other users. In other words, collaborative filtering can be considered as the machine pendant of asking a friend for recommendations (Farkas, 2007). Content-based filtering methods deal with recommended items based on what a user preferred in previous sessions. Therefore, it requires less information to give recommendations than collaborative filtering techniques. Because interests of users differ, recommendations in an online store are usually personalized. If not logged in, non-personalized recommendations are given as ranked lists of items. Both techniques are implemented in web search engines like Google, but to the knowledge of the authors not in Geoportals.

Although not placed as prominent on their webpage as Amazon, Google also uses Recommendation systems to improve the quality of search results. Google mainly focuses on personalized recommendations. Therefore, it makes use of location information based on IP address and previous searches coming from the IP address or logged in account. Even PageRank depends on recommendations as it takes into account how often a webpage has been cited or referenced. Related items are shown in the “Related searches” section but have to be set in the filter options to be displayed (see Figure 3).

Figure 3: Google Recommendations as “Related Searches”



The practise of integrating recommenders within online stores or web search engines has proven to be very useful for both providers in terms of value for e-commerce (Shani et al, 2005) and customers in form of satisfaction with the results of recommender system based on usability studies (Pu et al, 2012). For example, Felfernig and Gula (2006) state that throughout a survey users utilizing a recommender system were significantly more satisfied with their own decision process of choosing an appropriate item. A survey conducted by ChoiseStream showed that 45% of users are more likely to use an online shop offering recommendations than a shop that does not provide this feature (Pu et al, 2012). Thus, we propose the implementation of recommendation system as an additional component for discovery in metadata geoportals.

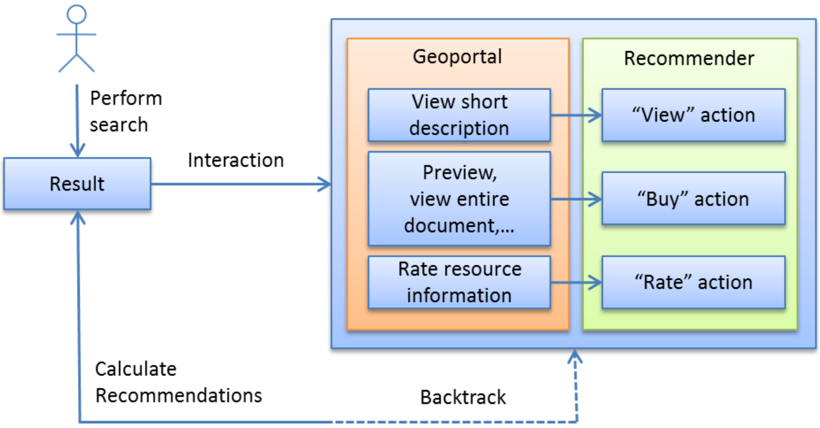
4. RECOMMENDATION STRATEGIES FOR GEO-RESOURCES

As discussed before, integration of recommendations in the EnerGEO Geoportal shall be a powerful method to improve geographic resources discovery and user experience. This can be done by “tracking” the user’s interactions with the search results and the resources itself. User tracking can raise privacy problems. Aggregating information of a specific user can be problematic in legal terms because additional information (i.e. personal interests, etc.) about the user can be inferred and misused or this information could even be sold to advertising companies (Riedl, 2001). Although recommender systems do not directly show “personally identifiable information” by default (Calandrino et al, 2011), they are dependent on a clear understanding of the users’ behaviour and personalized recommendations. Users need to be confident that interactions in the WWW can be logged in order to make systems like recommender engines more efficient regarding user satisfaction with provided results. In geoportal domain there is a reduced number of research activities addressing users’ privacy issues raised by recommendation systems (Canny, 2002; Scipioni, 2011). In addition, privacy is especially related to context. For example a user is more likely to provide information about his favourite places than provide any information about medical treatment he received (Lam and Riedl, 2006).

From a legal point of view, tracking of user interactions in the WWW is dependent on the country the server storing the information located and even leads to different court decisions within one country. Especially saving of IP-addresses over a longer period of time is problematic and therefore not foreseen in the recommender approach. In other cases, it is possible to delete the last digits of the IP-Address, because in that case it is still possible to gather e.g. region information.

An interaction in the context of Geoportals can be regarded as any recorded action that is transmitted to the recommender system or vice versa which can be also referred to as a transaction. An action is simply a recorded click on specific objects of the user interface that is sent to (or received by) the recommender system. The actions themselves can be divided into three categories: “view”, “buy” and “click”. Figure 4 shows the “linkage” of user interactions in Geoportals and “actions” in recommender systems. View actions can be considered as clicks on a specific search result, whereas a click on the details or preview page or invocation of a web service could be thought of a buy action. Further, if the Geoportal offers the possibility to directly download data, this could be a buy action as well. View and buy actions are used as basic rules for the calculation of recommendations. As further input for the recommender system, resources can be rated. Major input to the recommender can be user clicks on specific sections within the Geoportal, mainly regarding the interaction with the search result list. A separate list of recommendations can be used to “backtrack” clicks on recommendations provided additionally to the search result list. Beside others, this offers a means to see which recommendations were of interest to the user. The prototypical online implementation of the recommendation system at the EnerGEO Geoportal website is shown in section 5, where technical implementation is pointed out in more detail.

Figure 4: Recommendation Workflow



To our knowledge, none of the existing Geoportal applications have been extended with recommendation systems. Other systems use ratings, but not in the sense of a recommender system like that used in online stores in the WWW. For instance, the Geoportal APIs implemented on ArcGIS online cloud-based platform is coming together with item ratings capability (www.arcgis.com). Retrieved search results are thus listed based on the number of ratings the web resource has received. Despite its valuable role in helping users to sort out returned search results, rating capability requires active participation of users in rating the items. Recommender-based system on the other hand is tracking users' interaction with spatial items automatically, users participation in items' rating being passive. Beshe (2011) has designed a recommender *"to determine the fitness for use of a spatial dataset and then to use the fitness for use search criteria"*. This approach is focusing mainly on users perspectives on quality dimensions of spatial data and description of spatial resources. In our approach, users' past behaviour in terms of viewed, rated or 'bought item' is used for providing hints on what might interest them.

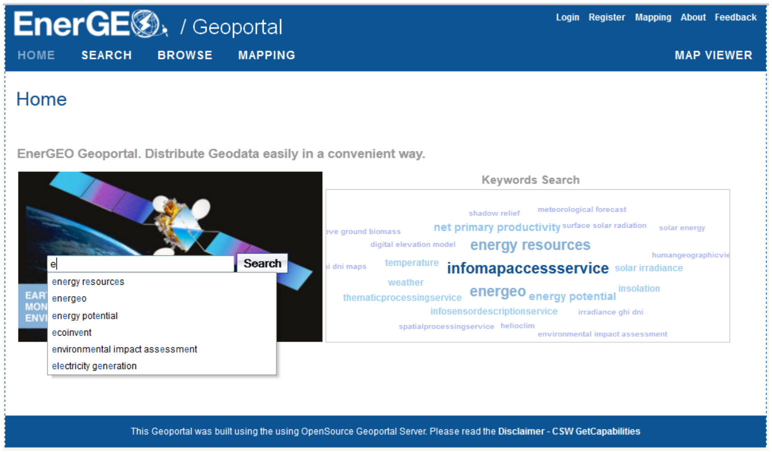
5. RECOMMENDATION SYSTEM IMPLEMENTATION IN THE ENERGEO GEOPORTAL

As part of the EU FP7 project EnerGEO a Geoportal was created fulfilling the purpose of providing information on existing energy resources from the energy domain to a broad audience using standards for technical interoperability as well as efficient discovery mechanisms (Blaschke et al, 2010). The metadata documents registered are compliant with the specifications of ISO Metadata Standards (e.g. ISO 19115) and the INSPIRE Metadata Implementing Rules. The OGC CSW (Catalogue Service Web) interface of the open source product

“Geoportal Server” ensures for technical interoperability and for linkage between different catalogues by so-called harvesting mechanisms.

Searching capabilities of the implemented Geoportal have been extended by incorporating EIONET GEMET (General Multilingual Environmental Thesaurus) Thesaurus. Furthermore, users are provided with keywords in dynamically created tag clouds to see the most popular keyword content of the portal. While typing, the user is assisted with auto suggestion lists of the most popular terms (see Figure 5). Tag clouds technique has been adopted by various Geoportal applications as an approach for improving discovery experience of users. For instance, INSPIRE Geoportal (http://inspire-geoportal.ec.europa.eu/discovery/) supports users in sorting out searched items by using tag cloud where EU Member states and spatial datasets defined in INSPIRE spatial data themes are listed. Another interesting approach using tag clouds in combination with thesauri to guide users through the process of determining broader and narrower terms related to the search term entered was presented by Janowicz et al (2009).

Figure 5: Auto Suggestion Dropdown List while Entering a Search Text



As an extension to general discovery a recommender software product (open source) called “easyrec” (e.g. Gstrein, 2009) is used by the authors for integrating recommendations into the ESRI Geoportal Server framework. Easyrec utilizes a shopping cart analyser called “Association Rule Miner (ARM)” based on the Apriori algorithm R (Agrawal and Srikant, 1994) and “SlopeOne” (Lemire and Maclachlan, 2005). Apriori is a classic learning algorithm for association rules of a recommender system. It enables the manifestation of statements like “users loading their cart with item A and item B also put item C in it with a likelihood of 80%”.

SlopeOne is an item-based collaborative filtering technique which is based on the assumption that the behaviour of certain groups can be used to conclude on particular interests of individuals. The algorithm proposed by Lemire and Maclachlan (2005) aims at predicting how a user would rate an item based on the ratings of a group of users.

The algorithm implemented in easyrec is looking for pairs of items $< X,Y >$ that appeared significantly often together in different baskets. A basket in the case of Geoportals can be referred to as a collection of (click-) actions performed by a specific user. It can be fine-tuned by adjusting two major parameters: support ($< X,Y >$) and confidence ($X \rightarrow Y$). Support determines how often a set of items $< X,Y >$ appears together in different user baskets whereas confidence is about the likelihood that an item Y follows in the presence of item X (Agrawal and Srikant, 1994). The minimum values of support and confidence determine whether an association rule is taken into account for calculation by the recommendation engine or not. Basically, it can be stated that association rules decline in amount and significance if the values of support and confidence are increased (easyrec, 2011).

Figure 6 shows the system architecture of the proposed solution. It is based on two database servers: one for the Geoportal component, one for the Recommendation component. The Geoportal Server solution as well as the open source recommender software easyrec are Java servlets that are deployed on web servers. The Geoportal Server Servlet consists of a web application for user interaction and the metadata catalogue, which forms the basis for entrance to database content. The Ontology Service is directly coupled to the Geoportal Server Servlet and offers the possibility to allow for ontology-based searches within the Geoportal Server search page (see also Figure 8). The Recommender servlet consists of a backend, where recommendations are calculated and an administration website, where containers for items to be considered for recommendation purposes, are created. The recommender engine offers a REST- and JavaScript API to integrate it in websites. In the implementation prototype, a JavaScript extension to Geoportal Server is provided, utilizing the JavaScript API of the Recommender servlet. The Geoportal Server extension is directly integrated into the Geoportal Server Search page (Figure 9). It allows recording of user interactions within the search results page as well as clicks on preview or data download links. The clicks are sent to the recommender component immediately, where they are stored and used for calculation of recommendations based on ARM and SlopeOne algorithms. If the user chooses one of the results in the result list on the right (Figure 8), he utilizes the JavaScript Extension for Geoportal Server to retrieve recommendations for the specific item clicked in the result list.

Figure 6: Proposed System Architecture

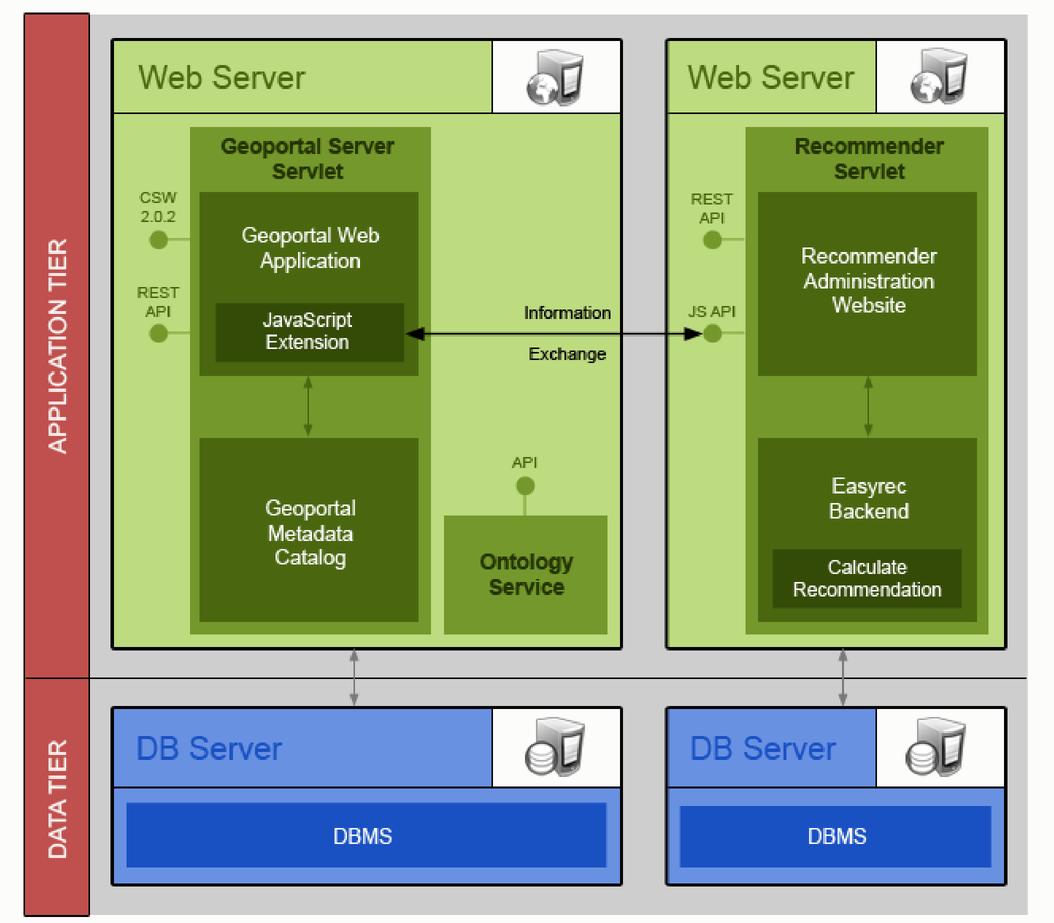
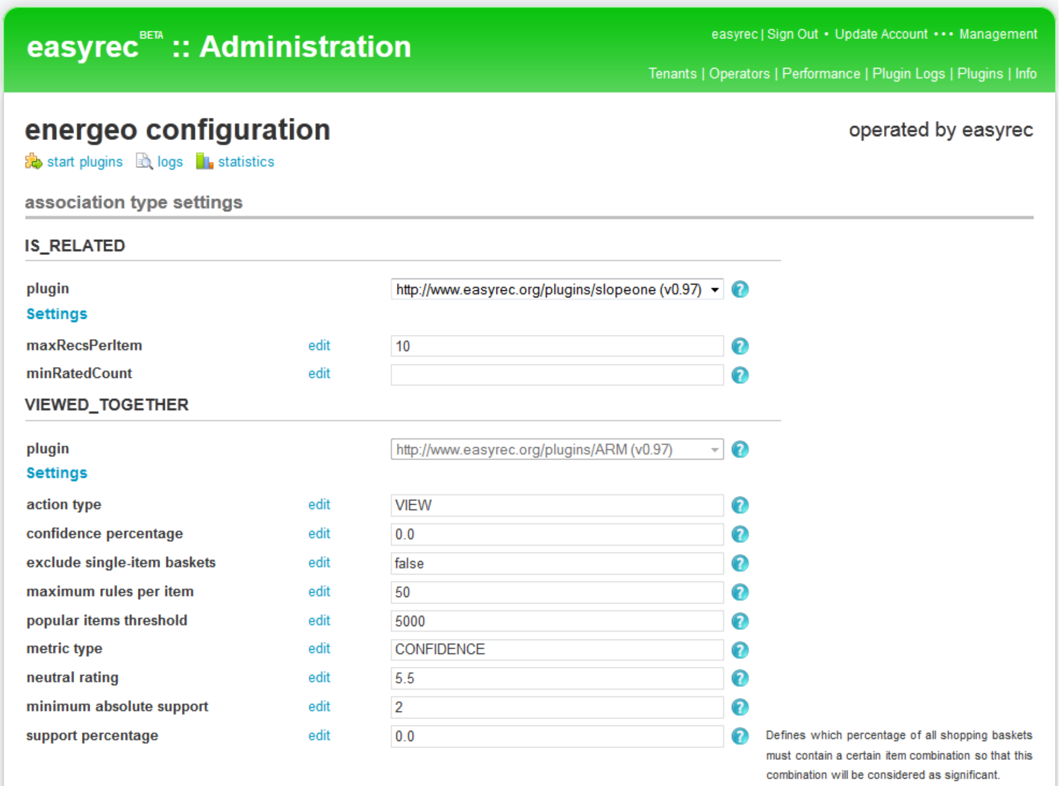


Figure 7 shows some of the parameters of easyrec rule generation. These can be specified for the three action types “view”, “buy” and “rate”. It allows for e.g. specification of the maximum amount of rules being calculated per item, to exclude shopping baskets containing only one item and to edit both the confidence as well as the support percentage. Confidence percentage determines the confidence of item combinations by looking for how often item A occurs together with item B whereas support percentage defines the percentage of the shopping baskets that must contain a certain item combination so this combination is regarded significant (easyrec, 2011).

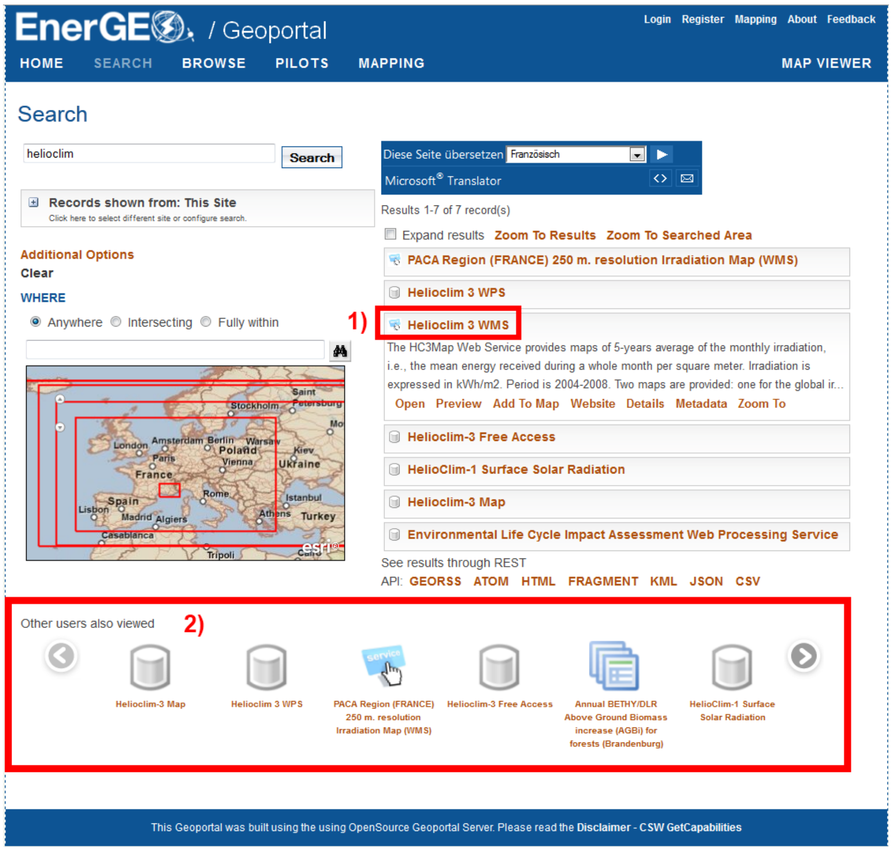
Figure 7: Easyrec Administration User Interface – Adjustable Parameters



Recommendations themselves are calculated based on users' past behaviour regarding items viewed, "bought" or rated. In addition, the approximate location information of the user based on the IP address as well as the username if the user is logged in is incorporated into the recommendation. Logged in users get user-specific recommendation based on former searches instead of general results presented to unidentified users. Further, a unique ID, the title of the resource, a hyperlink pointing to the resource as well as a thumbnail image is sent to the recommendation engine. These parameters are mainly used for presentation purposes of the end result.

Implementation itself requires knowledge of the Geoportal Server Framework, which is mainly based on Java Server Faces and Java as well as JavaScript Libraries. On the other side, easyrec offers a JavaScript and REST-API, both of which can be used to integrate recommendations into Geoportals. From the information automatically gathered, easyrec predicts items the user may be interested in, enhancing search results with elements that are derived from the context. The results of recommendations are provided in a section called "other users also viewed" (see Figure 8, mark 1).

Figure 8: Recommender-Enhanced Discovery in the EnerGEO Geoportal



Beside this form of un-personalized recommendations it is also possible to use personalized recommendations or ranking such as "most viewed" and "best ranked" resources. In that case, clicks on sets of items in former searches are used for further rule-calculation within the recommender system to determine which elements are frequently used together.

The recommendations are presented using a so-called image carousel based on JavaScript (see Figure 8, mark 2). It contains a list of items represented by the title and a thumbnail image representing either a preview image of the resource or at least the data type of the resource. Buttons on the right- and left-hand-side can be used to scroll through the items. When clicking on a resource, its details are shown and the click is "backtracked" by the recommender to be used for further recommender rule generation (Figure 4).

6. OUTLOOK & DISCUSSION

During the last years, consistent research activities have been dedicated to the improvement of users' discovery experience within Geoportals. Semantic heterogeneity proved to be the key challenge of spatial information discovery and integration. To overcome this problem, solutions have been proposed by either incorporating thesauri, ontologies, user's social environment (e.g. current location, query language used) or a combination of them (Buccella et al, 2008). In our approach, we propose discovery enhancement by integrating recommendations based on user interactions into Geoportal solutions. Recommendation systems are an essential extension of today's Geoportals improving the quality of search results significantly. Due to provision of additional search results, which cannot be derived from keyword-based search only, additional links between resources based on the experiences of other users are created and presented as a result. The implementation in the context of the EU FP7 Project EnerGEO has shown that inclusion of the context of users enhances search results with valuable information. The implemented recommendation system is not meant to operate as a standalone solution for spatial information discovery. It is seen as an added value to the existing approaches towards improving search engine results. This system makes the interaction of users with published resources explicit. It is able to trace users' behaviours while searching and interacting with geographic resources and to share registered “traces” with other users and thus providing them hints on what might interest them. Returned resources are further subject to users' exploration to see to what extent discovered resources fulfil their requirements, i.e. fitness of discovered data for their purpose.

The implemented recommendation system is designed to help users “browsing” among huge amount of items and to help them spending less time in finding geographic resources. Therefore, it is seen as a valuable approach towards increasing users' satisfaction while interacting with geographic resources made available by various providers. The functionality of the proposed recommender based system is served out over standardized interfaces supporting integration with other distributed computing infrastructures.

A major benefit of recommender systems in general is that they allow to discover resources similar to what people already liked. As recommendations depend on users' behaviour in the past, the results are not simply "guessed" but calculated by applying mathematic algorithms. As the recommendations are recalculated within a specifiable time, they are always up-to-date.

One major drawback is that recommendations may not coincide with the users' expectations. This is based on the fact that meaningful recommendations can only be provided when large amounts of users are part of a system, as they need user input for generation of rules.

Additional approaches on how to further improve the quality of discovery giving recommendations are foreseen. Especially the context of the user might be enriched by utilizing further elements such as primary search language or domain. Therefore, it is planned to incorporate additional metadata information like abstracts, short descriptions or lineage to the rule sets of the recommendation engine. In addition, the implementation of the location of the user can be improved.

Beside integration of recommender systems, we plan to enhance recommendation results by incorporating contextual similarity within texts registered in the Geoportal as well as matching them with descriptions coming from the INSPIRE data specifications. This information enrichment can be considered a valuable approach towards meta-linking of heterogeneous resources from various domains as well as different types (e.g. spatial and non-spatial) that are of importance for specific user questions. The concept is based on semantic text analysis mechanisms used to extract content from all documents in order to establish “links” between the resources. The results of these “text matchings” can then be used as additional rule sets for calculation of recommendations. In order to validate the approach presented in this paper, a survey will be conducted.

Summing up, in this paper we proposed the integration of recommender systems in the process of metadata discovery within Geoportals. As shown in the proof-of-concept in the EnerGEO Geoportal, it guarantees results that are not in the data itself but rather derived from the context of other users’ searches and views. In addition, integration of contextual similarity of descriptions of INSPIRE data specification elements to further enhance the recommendations may provide additional value for the end user.

ACKNOWLEDGEMENTS

For research purposes, the EU FP7 project EnerGEO project was considered as the perfect environment since project partners come from different European countries having different backgrounds. EnerGEO is the European contribution to the energy theme of GEOSS. The aim of this project is to develop a strategy for a global assessment of the current and future impact of the exploitation of energy resources on the environment and ecosystems and to demonstrate this strategy for a variety of energy resources worldwide. Therefore, the goal is to provide resources in order to generate domain specific knowledge by integrating resources into the common architecture. EnerGEO takes into account

architecture recommendations of GEOSS and other initiatives like GMES and SEIS.

REFERENCES

Athanasis, N., Kalabokidis, K., Vaitis M. and N. Soulakellis (2009). Towards a semantics-based approach in the development of geographic portals, *Computers and Geosciences*, 35(2): 301-308.

Agrawal, R. and R. Srikant (1994). “Fast Algorithms for Mining Association Rules in Large Databases”, *VLDB 1994 Proceedings of the 20th International Conference on Very Large Data Bases, September 12-15, 1994, Santiago de Chile, Chile*, pp. 487-499.

Balabanović, M. and Shoham, Y. (1997). Fab: Content-Based Collaborative Recommendation, *Communications of the ACM*, 40(3): 66-72.

Belgiu, M., Strobl, J. and M. Mittlboeck (2012). “Adding semantics to spatial content: A Land Cover Scenario”, *Proceedings of Global Geospatial Conference 2012, May 14-17, 2012, Québec City, Canada, at* http://www.gsdi.org/gsdiconf/gsdi13/papers/218.pdf

Beshe, T.S. (2011). *Turning spatial data search engine to spatial data recommendation engine*, MSc Research submitted to Faculty of Geo-Information Science and Earth Observation of the University of Twente.

Blaschke, T., Mittlboeck, M., Biberacher, M., Gadocha, S., Vockner, B., Hochwimmer, B. and S. Lang (2010). ”The GEOSS – EnerGEO portal: towards in interactive platform to calculate, forecast and monitor the environmental impact of energy carriers”, in Greve, K. and A.B. Cremers (Eds.), *ENVIRONINFO 2010 – Integration of Environmental information in Europe: Proceedings of the 24th International Conference on Informatics for Environmental Protection*. Aachen: Shaker Verlag, pp. 2-9.

Brin, S. and L. Page (1998). “The Anatomy of a Large-Scale Hypertextual Web Search Engine”, *Seventh International World-Wide Web Conference (WWW 1998), 14-18 April, 1998, Brisbane, Australia*.

Buccella, A., Cechich, A. and P. Fillotrani (2008). Ontology-driven Geographic Information Integration: A Survey of current Approaches, *Computers and Geosciences*, 35(4): 710-723.

Calandrino, J.A., Kilzer, A., Narayanan, A., Felten, E.W. and V. Shmatikov (2011). “"You Might Also Like:" Privacy Risks of Collaborative Filtering”, *Proceedings of the 2011 IEEE Symposium on Security and Privacy (SP)*, 231-246.

Canny, J. (2002). “Collaborative filtering with privacy via factor analysis”, *SIGIR '02 Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, August 11-15, 2002, Tampere, Finland, pp. 238-245.*

De Martino, M. and R. Albertoni (2011). A multilingual/multicultural semantic-based approach to improve Data Sharing in an SDI for Nature Conservation, *International Journal of Spatial Data Infrastructures Research*, 6: 206-233.

easyrec (2011). easyrec Documentation, at http://sourceforge.net/apps/mediawiki/easyrec/ [accessed 6 February 2012].

EuroVoc, at http://eurovoc.europa.eu/drupal/, [accessed 26 September 2011].

ESRI Geoportal Server (2011). At http://geoportal.sourceforge.net/ [accessed 18 October 2012]

Farkas, M.G. (2007). *Social software in libraries: Building Collaboration, Communication, and Community Online*. New Jersey: Information Today.

Felfernig, A. and B. Gula (2006). “An Empirical Study on Consumer Behavior in the Interaction with Knowledge-based Recommender Applications”, *Proceedings of The 8th IEEE International Conference on E-Commerce Technology and The 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services (CEC/EEE’06), June 26-29, 2006, Palo Alto, California, USA.* Washington D.C.: IEEE Computer Society.

Fernández, T.D. and Castellanos, E. (2006). “Towards user-driven Spatial Data Infrastructures. An approach oriented to sustainable development”, *Proceedings of Global Geospatial Conference 2006, Nov. 6-10, 2006, Santiago de Chile, Chile.*

GEMET, at http://www.eionet.europa.eu/gemet, [accessed 26 September 2011].

Gstrein, E. (2009). *Adaptive Personalization: a multi view personalization approach incorporating contextual information*, Unpublished PhD Thesis, Vienna: Vienna University of Technology.

Hearst, M. (2009). *Search User Interfaces*, Cambridge: Cambridge University Press.

Janowicz, K., Schwarz, M. and M. Wilkes (2009). “Implementation and Evaluation of a Semantics-based User Interface for Web Gazetteers”, in Handschuh, S., Heath, T. and V.T. Thai (Eds.), *Proceedings of the IUI'09 Workshop on Visual Interfaces to the Social and the Semantic Web, February 8, 2009,* *Sanibel Island, Florida, USA, 443*, at http://ceur-ws.org/Vol-443/paper2.pdf [accessed 18 October, 2012]

Lagoze, C. and A. Singhal (2005). Information Discovery – Needles and Haystacks, *IEEE Internet Computing,* 9(3): 16-18.

Lam, S.K. and J. Riedl (2006). “Privacy, Shilling and the Value of Information in Recommender Systems”, *Proceeding of User Modeling Workshop on Privacy-Enhanced Personalization, July 25, 2005 Edinburgh, UK*, pp. 85-92.

Latre, A.M., Hofer, B., Lacasta, J. and J. Nogueras-Iso (2012). The development and interlinkage of a drought vocabulary in the EuroGEOSS interoperable catalogue infrastructure, *International Journal of Spatial Data Infrastructures Research,* 7: 225-248.

Lemire, D. and A. Maclachlan (2005). “Slope One Predictors for Online Rating-Based Collaborative Filtering”, *Proceedings of the 2005 SIAM International Conference on Data Mining (SDM'05), April 21-23, 2005, Newport Beach, California.*

Linden, G., Smith, B. and J. York (2003). Amazon.com Recommendations: Item-to-item Collaborative Filtering, *IEEE Internet Computing*, 7(1): 76-80.

Lutz, M., Sprado, J., Klien, E., Schubert, C. and I. Christ (2009). Overcoming semantic heterogeneity in spatial data infrastructures, *Computers and Geosciences,* 35(4): 739-752.

Masser, I. (2007). *Building European Spatial Data Infrastructures*, Redlands, California: ESRI Press.

Maso, J., Pons, X. and A. Zabala, (2012). Tuning the second-generation SDI: theoretical aspects and real use cases, *International Journal of Geographical Information Science*, 26(6): 983-1014..

Melville, P. and V. Sindhwani (2010). “Recommender Systems”, in Sammut, C. and G. Webb (Eds.), *Encyclopedia of Machine Learning.* Berlin/Heidelberg: Springer-Verlag, pp. 829-838.

Pazzani, M.J. (1999). A Framework for Collaborative, Content-Based and Demographic Filtering. *Aritificial Intelligence Review*, 13: 393-408.

Pu, P., Chen, L. and R. Hu (2012). Evaluating Recommender Systems from the User’s Perspective: Survey of the State of the Art, *User Modeling and User-Adapted Interaction*, 22(4-5): 317-355.

Rajabifard, A., Feeney, M.-E.F. and I.P. Williamson (2002). Future directions for SDI development, *International Journal of Applied Earth Observation and Geoinformation*, 4(1): 11-22.

Resnik, P., Iacovou, N., Suchak, M., Bergstrom, P. and J. Riedl (1994). "GroupLens: an open architecture for collaborative filtering of netnews", *Proceedings of ACM 1994 Conference on Computer Supported Cooperative, October 22-26, 1994, Chapel Hill, North Carolina, USA*, pp. 175-186.

Ricci, F., Rokach, L. and Shapira, B. (2011). "Introduction to Recommender Systems Handbook", in Ricci, F., Rokach, L., Shapira, B. and P. Kantor (Eds.), *Recommender Systems Handbook*. Berlin/Heidelberg: Springer-Verlag, pp. 1-35.

Riedl, J. (2001). Personalization and Privacy, *IEEE Internet Computing*, 5(6): 29-31.

Sadeghi-Niaraki, A., Rajabifard, A., Kim, K. and J. Seo (2010). "Ontology based SDI to facilitate spatially enabled society", *Proceedings of GSDI 12 World Conference, October 19-21, 2010, Singapore.*

Sarwar, B. Karypis, G., Konstan, J. and J. Riedl (2001). "Item-Based Collaborative Filtering Recommendation Algorithms", *Proceedings of the 10th international conference on World Wide Web, May 01-05, 2001 Hong Kong, Hong Kong,* pp. 285-295.

Scipioni, M.P. (2011). Towards Privacy-Aware Location-based Recommender Systems, *Proceedings of IFIP Summer School, September 5-9, 2011, Trento, Italy,* at http://uc.inf.usi.ch/sites/all/files/scipioni-trento-paper.pdf [accessed 12 October 2012].

Shani, G., Heckerman, D. and R.I. Brafman (2005). An MDP-Based Recommender System, *Journal of Machine Learning Research*, 6(2): 1265-1295.

Smits, P.C. and A. Friis-Christensen (2007). Resource Discovery in a European Spatial Data Infrastructure, *IEEE Transactions on Knowledge and Data Engineering*, 19(1): 85-95.

Shardanand, U. and P. Maes (1995). "Social Information Filtering: Algorithms for Automating "Word of Mouth"", *Proceeding of CHI '95 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, May 07-11, 1995, Denver, Colorado,* pp. 210-217.

Strobl, J. and A. Nazarkulova (2009). "Places Along the Information Highway: the Long Path towards Spatial Data Infrastructures", *Proceedings of Third Central Asia GIS Conference - GISCA'09-GIScience for Environmental and Emergency Management in Central Asia*, August 27-28, 2009, Bishkek, Kyrgyzstan, pp. 16-21.

Van Oort, P.A.J., Kuyper, M.C., Bregt, A.K. and J. Crompvoets (2009). Geoportals: an internet marketing perspective, *Data Science Journal*, 8(24): 162-181.