Environmental monitoring: personalized access approach to the remote sensing information resources

To cite this article: I V Kovalev *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **315** 052011

IOP Conf. Series: Earth and Environmental Science **315** (2019) 052011 doi:10.1088/1755-1315/315/5/052011

Environmental monitoring: personalized access approach to the remote sensing information resources

I V Kovalev[1, 2, 3], D I Kovalev[1], M V Karaseva[1, 2], V V Brezitskaya[1] and A A Voroshilova[3]

[1]Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., Krasnoyarsk, Russia, 660037
[2]Siberian Federal University, 79, Svobodny Pr., Krasnoyarsk, Russia, 660041
[3]Krasnoyarsk City Hall of Science and Technology, 61 Uritskogo Str., Krasnoyarsk, Russia, 660049

E-mail: karaseva-margarita@rambler.ru

Abstract. The article considers the necessity to apply a user profile operating with the earth remote sensing information used in the environmental monitoring. The stages of forming a user profile are presented; a dynamic profile model is given. It allows to take into account both long-term and short-term profiles. The main attention is paid to the following issues in describing the stages of user personalization: monitoring the user's activity; creating user profiles; evaluating the effectiveness of personification. The proposed approach has been successfully implemented for the environmental monitoring of the power plants.

1. Introduction

The satellite technologies based on the use of remote sensing (RS) information are becoming increasingly widespread in solving a variety of problems from the environmental monitoring and the search for mineral resources to monitoring territories and defense problems. A significant part of remote sensing information is used for the scientific purposes. Here, informational resources of American organizations and the U.S. services such as Geological Survey (USGS), National Oceanic and Atmospheric Administration (NOAA), National Aeronautics and Space Administration (NASA) are in demand [1, 2]. Operational and archival Earth remote sensing information is offered by a large number of satellite operators, both commercial and public. The procedures for evaluating space images are sufficiently developed and implemented through the Internet.

However, satellite technologies suppose the application of not only space images, but also parametric information related to the Earth, ocean and atmosphere (geophysical, geological, meteorological, etc.). Various information centers, information banks and repositories, including thematic data centers that are part of the NOAA structure, provide such information in large volumes, e.g. National Climatic Data Center, National Geophysical Data Center, National Oceanographic Data Center, National Coastal Development Data Center; repository Ameri Flux (http://public.ornl.gov/ameriflux/), network resource FLUXNET (http://fluxnet.ornl.gov/fluxnetdb), Goddard Earth Sciences Dataand Information Services Center (GES DISC) NASA, multi-scale multicomponent system Integrated Ocean Observation System (IOOS) (http://www.ocean.us) and etc.

The modern Russian ground complex for receiving, processing and distributing (GCRPD) space information from a remote sensing satellite includes heterogeneous and dispersed information systems (IS) that belong to various ministries, departments and separate organizations. A lot of centers have poor technical instrumentation and they are equipped with small receiving antennas. It does not ensure the possibility of receiving the full flow of space information (SI) from the advanced Russian remote sensing satellites.

The existing methods and forms of customer service have a low efficiency in the execution of applications for space imagery and they do not provide the required reliability of orders fulfillment for remote sensing satellites and space products for its processing. The access to archives of the stored space information is difficult due to the large number of such archives and the low level of interaction among them on account of departmental disunity. All this complicates sharply the possibilities of the efficient application of the aggregate of the available remote sensing information and reduces the interest of home and, moreover, foreign potential consumers to Russian space information.

In this situation, the effective organization of access to information resources of the remote sensing is extremely important for consumers working with a geographically distributed information system (GDIS) of the RS. The departmental identity of the centers (information systems) and stations makes no difference to the user. Their functional unification is used taking into account the elaborated uniform and the agreed operating rules.

2. Methods of personalized access to RS information resources

It is possible to improve the quality of information services provided to users of the GDIS through the application of personalization, i.e., the provision of content and services to individual users in accordance with their requirements and taking into account data on their preferences and behavior.

Often when searching for information in GDIS, users do not know what exactly they need to find. Often the search is complicated by the fact that the information sought is not clearly displayed in the navigation elements of the system or specific information systems. In this case, users need help finding. The system monitors its activities and identifies its goals and preferences to help a user. The determination of the user's goals and the formation of his model is an important problem in online systems [3, 4].

Many users customize their IS workspace to their needs with built-in configuration tools. The disadvantage of this approach is that after it has been set up, users do it again rarely, although over the time since the last setting, functions needed by the user could be added to the IS. They could be unknown and, therefore, they could be unusable.

It is clear from the information above that online IS users need to organize the convenient personalized access to their operating areas, and it is desirable to carry out personalization at least partially automatically based on the information collected while the monitoring the user activity.

The personalized support for navigation in the online IS is designed to enhance the functionality for an individual user based on the characteristics by providing links to information of potential interest. Personalized navigation support is needed for both corporate IS and web sites, since their users face the same type of problems.

2.1. Personalized navigation support method

The personalized navigation support in online IS and websites is implemented as a direct guidance or adaptive sorting. The first method is suitable for providing the user with a link to the best page for the next move in accordance with the information about his links. The IS can highlight the link to the best page or suggest an additional link “Next” to move to it.

The second method is more flexible and it helps to recommend a page that is not directly related to the current link but not a link represented on it. The adaptive sorting method means the arranging the page links in accordance with the user's links model and some criteria of the link relevance. The lack of adaptive sorting is in constant changing the order of the links in the following requests for the page.

At the same time, studies have shown that a stable order of menu items is very important for users [5]. However, experimental studies gave the results that adaptive sorting significantly reduces the time spent on navigation in online information systems, where each page contains a large number of context-independent references.

The personalization introduction in the IS will make them more convenient for users mainly by reducing the time spent on obtaining the necessary information. Modern personification methods require further elaboration of all its stages: monitoring user activity, creating user profiles, and providing personalized information and services.

2.2. Stages of the personification process
It is required to decide what and how should be personalized realizing personalization in the online IS. Also it necessary to take into account the information about users, their actions in IS, what actions of other users they resemble, how one can extract this information from the available information and create users' profiles.

Modern approaches to personification are based on monitoring the users' online IS activity, forming his profile based on statistics collected while monitoring and personalizing content or services in accordance with the user profile.

Consider the user's access to the online IS to view the page. Processing a user's query begins with its authorization, for example, by name and password, cookie, IP-address of his computer or personal certificate, resulting in a unique user in the system.

At the next stage, the user profile is modified based on the information collected about it by the system. This information may contain information about its movements through the pages in the online IS, the text of search queries, actions related to viewing information, demographic and psychographic data, databases of IS information, as well as journals of the IS and the web server. This information determines the interests, occupation and preferences of the user depending on both a type of space images queried and a type of parametric information associated with the Earth, the ocean and the atmosphere. The information may be geophysical, geological, and meteorological.

At the final stage, in accordance with a user's profile, the personalized information is selected from the IS database and the code of the page sent to the user is formed.

2.3. Monitoring of the user activity
The methods for monitoring the user activity are divided into passive and active. In the passive method, a user independently specifies his preferences by filling out questionnaires or participating in surveys. The information entered by him is brought into the database and used to form his profile.

The main drawback of passive information gathering is the reluctance of users to provide personal information because of laziness, or (for websites) fear that information will fall into the hands of intruders. Users provide personal information if it is to their advantage, and they trust the website owner.

In active methods, the server software collects information. Nowadays, a lot of active methods of collecting and analyzing information on various aspects of user interaction with IShave been developed [6]. There exist methods that analyze the keywords of the pages queried by the user, the texts of his search queries, routes for navigating through the pages of the system, information about queries of various kinds. Naturally, these methods have specific advantages and disadvantages.

For example, highlighting keywords that are frequently found on the user-queried pages is a very resource-intensive method for determining its preferences. Each word is characterized by the frequency of emergency on the queried pages. Keywords form a multidimensional space where a visitor preference is determined by a dot. A number of keywords reaches hundreds, so a problem of determining the preferences of each user is of a high dimension, and even methods for reducing the dimension of the feature space cannot improve the situation significantly. The advantage of this method is its high accuracy of determining the interests of the user. Also, this method allows users to be combined according to similar interests into clusters [7].

Some experimental methods for each page indicate meta-information or keywords. And it is too time-consuming, requiring the involvement of experts. In real conditions, this method is not applicable, as it will require too high labor costs.

The analysis of the users' routes of the online IS reveals popular routes to navigate the site. This method is suitable for personalizing navigation to an individual user or a group with similar preferences. The method can help to find out from which pages and where users are moving. The disadvantage of this method is its computational complexity.

Only a small number of the currently proposed methods for monitoring the activities of users of online IS are applicable in practice. A good monitoring method should be resource-insensitive, provide acceptable accuracy in determining preferences, goals, or objectives, and it shouldn't require the interference of experts in their work.

3. Formation of users' profiles

The users' profiles are created on the basis of the information collected during the monitoring their activities. The users' profiles can be mathematical models of goals or user's interests. Most systems apply models with a unique structure. For example, the user's profile may contain relative frequency of occurrence if the keywords of the pages queried by a user were collected during the monitoring process [8].

In some systems users' profiles consist of two parts. The first one is the actual part of the profile. It contains information about a user indicated by him at registration and participation in surveys. Also it contains statistical information obtained during the monitoring of his activities, for example, the average number of IS pages viewed per day. The second part of the profile is behavioral. It simulates the usual actions of the user; it can be formed on the basis of the available information using the data mining methods.

Consider one of the personification methods in more detail [9]. One can define that documents can be represented as vectors of a latent semantic space. In order to monitor and continuously analyze possible changes in the user's interests, it is proposed to introduce the concept of temporal measurement in the latent semantic space, thereby considering not the latent semantic space itself, but its modification, the time latent semantic space. Each measurement (with the exception of a temporary one) of such a vector space represents conditional probabilities for a given class $P(\bullet| z)$. Documents are vectors with weight coefficients (coordinates) $P(z|d)$, the temporal measurement is assumed to be zero.

The query, as well as the documents themselves, can be represented as vectors in a temporary latent semantic space. In addition to $P(z| Q)$ weights, they have an additional (temporary) measurement (current weight), initially equal to a certain positive value decreasing over time. It assumes that a user's interest in a certain field decreases in the absence of its emergence in queries for a long time. If a user initiates a query related to a specific category from his current profile, then the weight of this category can either be stabilized for a certain time or increased.

According to the geometry of the latent semantic space, a query consisting of terms is projected into the latent semantic space. Thus, the hypersurface S_i formed by the query Q_i is the intersection of probabilistic surfaces of all classes introduced on the latent semantic space where this term appears with a certain probability:

$$S_i = \bigcap_k H_{ki}.$$

The adaptive user profile correction algorithm is based on implicit user feedback, which is implemented based on the history of its queries. The input of the algorithm is a user's query, the output is one or more triples of the form (C_i, W_i, α_i), where C_i is a category of interests, W_i is a current weight, $\alpha_i$is a level of variability (the meaning of this value is to reflect how much the interests of the user are changed within the current query in relation to the previous queries).

So, the user profile is a set of triples. At the same time, it is organized in such a way that the interests of the user are divided into two types: short-term (*short-term profile*) and long-term (*long-term profile*).

As a rule, the capacity of a long-term profile is larger than that of a short-term one. The profile structure can be represented by a table, as it is given in Table 1. In this case, it is considered that triples, in which the value of the current weight is positive, belong to the short-term profile, and if the weight is negative, then it belongs to the long-term profile. At the same time, for triples in the short-term profile, the current weight decreases linearly, whereas for triples in the long-term profile, the decrease in weights is exponential.

Table 1. Short-term user profile in relation to problems of the Ministry of Emergency Situations in operating with remote sensing information.

Category	Fires	Flood	Seismic activity	Other emergencies
Current weight	95	85	35	70
Level of variation	0.60	0.45	0.20	0.15

Formally, the current profile of i is described as follows:

$$Pr_i = \{(C_j, W_j, \alpha_j)_i, \ j=1, k \}.$$

Wherein

$$Pr_i = PrR_i \cup PrL_i,$$

where $PrR_i = \{(C_j, W_j, \alpha_j)_i | \ \forall W_j \geq 0, \ j=1, k\}$ is short term profile,
$PrL_i = \{(C_j, W_j, \alpha_j)_i | \ \forall W_j < 0, \ j=1, k\}$ is long term profile.

The level of variability (α_i) is calculated as the proximity of two consecutive queries Q_i and Q_{i-1}represented in the frequency space of their terms:

$$\alpha_i = \frac{\sum_w \tilde{n}(Q_i, w)\tilde{n}(Q_{i-1}, w)}{\sqrt{\sum_{w'} \tilde{n}(Q_i, w)^2 \sum_d \tilde{n}(Q_{i-1}, w)^2}},$$

where $\tilde{n}(Q_i, w)$areweighted frequency terms.

As one can see, the proposed method is the most progressive: it allows a user to enter two user profiles (long-term and short-term ones), as well as carry out a dynamic recalculation of profile parameters depending on the user behavior.

The efficiency of personification can be evaluated in various ways, depending on the specifics of the problem being solved [10]. The evaluation of the efficiency of personification is considered as a return on its implementation in IS. In particular, system developers should decide whether to collect additional data (multi-version approach [11]), develop more advanced analysis algorithms, create better user profiles, and improve the personification mechanism. The properly organized iterative process will allow understanding user needs better; to make focus the work of IS on their needs more precisely and to offer them better personalization [12].

4. Conclusion

The article defines the characteristics of the user profile for information systems with support for personalization. In conclusion, we will present a procedure that allows us to form a model of user queries, as well as a user profile, which is essential when considering consumer factors influencing the decision-making process, analyzing the consequences that this or that situation (this or that decision) and etc.

Obviously, the formed user profile is reflected in the choice of a rational strategy of behavior in problem situations; it depends not only on quantitative ratios, but on qualitative formulations. The choice of a category or user profile is often determined by a set of typical situations that can happen in the system only in the form it was fixed in practice.

When forming a user profile, a number of attributes are taken into account. They characterize various situations that can happen in the process of a user operating with remote sensing information:

- name of the situation;
- description of its causes;
- indication of the consequences to which it may lead;
- temporal characteristics that determine the time of occurrence and end of the situation;
- name of the structural units of the enterprise where the situation happened;
- description of the measures necessary for its successful resolution;
- indication of factors influencing the decision-making process;
- list of documents regulating the user's behavior in the current situation;
- list of data from various information and measurement systems used by the user in the process of developing a solution.

It should be noted that during the operation of GDIS, some of the characteristics mentioned above may not be used by the consumer of remote sensing information in the decision-making process. In addition, a number of characteristics may remain undetermined by experts for a considerable period of time. In these cases, only known (basic) characteristics will be taken into account and the rest will not be considered in the information system when forming a user profile.

Acknowledgements
The reported study was funded by Russian Foundation for Basic Research, Government of Krasnoyarsk Territory, Krasnoyarsk Regional Fund of Science, to the research project: «Multiversion method for improving information reliability of environmental monitoring of thermal power plants», 18-48-240007.

References
[1] Durbha S S and King R L 2005 Semantics-enabled framework for knowledge discovery from earth observation data archives *IEEE Transactions on Geoscience and Remote Sensing* **43(11)** 2563-72
[2] Bingham A W, McCleese S, Stough T, Deen R G, Hussey K and Toole N 2009 Earth Science Datacasting: Informed Pull and Information Integration *IEEE Transactions on Geoscience and Remote Sensing* **47(10)** 3570-80
[3] Resnick M L and Vaughn W V 2006 Best practices and future visions for search user interfaces *Journal of the American Society for Information Science* **57(6)** 781–7
[4] Tenopir C 2008 Online Systems for Information Access and Retrieval *School of Information Sciences - Faculty Publications and Other Works*
[5] Maloney J J 1983 *Online searching technique and management* (Chicago: ALA)
[6] Walker G and Janes J 1999 *Online retrieval: Dialogue of theory and practice* (Englewood: Libraries Unlimited)
[7] Wang Y D and Forgionne G 2006 A decision-theoretic approach to the evaluation of information retrieval systems *Information Processing and Management* **42(4)** 863–74
[8] Tenopir C and Ennes L 2001 Reference services in the new millennium *Online* **25(4)** 40–5
[9] Evangelopoulos N, Zhang X and Prybutok V 2012 “Latent Semantic Analysis: Five Methodological Recommendations *European Journal of Information Systems* **21(1)**
[10] Petrosyan M O *et al.* 2016 *IOP Conf. Ser.: Mater. Sci. Eng.* **155** 012032
[11] Saramud M V *et al.* 2019 *E3S Web of Conferences* **75** 01005
[12] Kovalev I V *et al.* 2016 *IOP Conf. Ser.: Mater. Sci. Eng.* **122** 012018