JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

USER PROFILE IN PERSONALIZING INFORMATION

H.Ouzif[1], Yassir Aadil[2], I. EL Achkar[3], H. Labriji[4]

[1]Laboratory of Technological Information and Modeling, University Hassan II, Casablanca, Morocco.
E-mail: ouzif.hind@gmail.com
[2]Laboratory of Technological Information and Modeling, University Hassan II, Casablanca, Morocco.
E-mail: aadil.yassir@gmail.com
[3]Laboratory of Technological Information and Modeling, University Hassan II, Casablanca, Morocco.
E-mail: Elachkar.ibtissam@gmail.com
[4]Laboratory of Technological Information and Modeling, University Hassan II, Casablanca, Morocco.
E-mail: labriji@yahoo.fr@gmail.com

Received: 22.04.2020 Revised: 23.05.2020 Accepted: 20.06.2020

ABSTRACT--- In order to meet the users expectations, the information retrieval system classified the users following their profile then affect them to the appropriate chain access of information.
Our approach is founded on the enrichment of users profiles that will be conducted from their social network which represent a rich source of information. However, one of our biggest challenge is when the user profile is newest or not very active, so the user profile may not contain all the interests and information that may be useful for a given mechanism. To bypass these challenges we developed a new mechanism that will allow us to detect similar users to this one and analyze their interest using a similarity techniques then algorithms that will improve its performance in a remarkable way by expanding user's egocentric network which allow adding more nodes from different social networks to cover the entire user's interest. Therefore, our proposal is to combine friends relationships of the user on different types of social network in a single egocentric social graph before the application of the algorithms to get a more complete user profile that will gather the majority of user's interests.

KEYWORDS--- User profile, Egocentric algorithm, Individual algorithm, Social network, Information system, Information retrieval system.

 DOI: http://dx.doi.org/10.31838/jcr.07.14.191

I. INTRODUCTION

The solution to better meet the specific needs of the user is to personalize the search for information, integrating its profile into the access to information chain. In the IT industry, whether in the context of corporate information systems, e-commerce, access to information and knowledge or even leisure, the quality and personalization of information is a major issue, since the success or rejection of these systems depends on the intelligibility, relevance, and adaptation of the information delivered to the user's preferences.

Current access and filtering systems face technological as well as linguistic barriers, caused by the massive production of new resources and information, of a heterogeneous, multiform and multilingual nature, making the production of a relevant quality information a challenge to be overcome. The purpose of customization is to reduce the search space, by facilitating the expression of the user's need and by searching for information on a subject by discarding non-relevant information, knowing that the selected information must be intelligible to the user and usable. The relevance of the information is not, however, an objective measure, generalizable to all users. It is defined by a set of customizable criteria and preferences specific to each user or user community.

In this article we will talk about user profile building using the interactivity between the machine and the user, we will also see the contents of a user profile, as well as its dimensions, its use in the search for information and its personalization, we will simply present two personalization services, which are data filtering and recommendation, taking into consideration the weight of the words in the view page in a session.

II. USER PROFILE CONSTRUCTION

Building and updating a user profile is the most important step in customizing information. The construction techniques are quite a few, but the most natural is to manually fill the boxes of the profile parameters. Except

that sometimes the user even if he knows his preferences well, he cannot express them formally. In addition, this technique remains long and annoying for the user.

As a solution to this problem, there are several methods of constructing user profiles automatically. These methods are linked to the interpretation of client activities on the outcome data. In this article, we will try to explain the method based on interactivity

2.1 Interactivity

User preferences should be the main information specified on the profile. Most of the time, there is a discrepancy between what the user expects and what he really wants, and this fact leads to information systems focusing on a process of discovering the interests of the interactive and incremental user; this approach is based on dialogue between the client who changes depending on the application and the system. The search for information is then carried out on a basis of interactivity.

2.1.1 Interactivity in the Pursuit of Information

The purpose of information-gathering interactivity is to provide the opportunity to capture the interests of a user and also to contribute to the determination of his needs. First, the user implements a simple and vague query that leads to large numbers of results, and in order to enable the user to better target his search, the system proposes criteria in addition to those already in existence, implicitly or explicitly, implicitly when the user is asked to make a choice between the criteria on the basis of a list by checking those that correspond to him most, in case this selection is made in several steps so explicitly, it becomes long and boring. This is why most information systems opt for the implicit deduction of user interests. In the same spirit, the user initially submits a request to the system, which will offer examples of results to choose which ones are of interest to the user. This can then be inferred from an extraction of the most important characteristics on the part of the system on the basis of results deemed relevant by the user.

Apt Decision is an example of an interactive profile building system, it is a real estate search system, which uses a profile initially based on data that the user provided such as the city, the price and the number of rooms, it is on these bases that the system provides a list of ads corresponding to these criteria, a list that does not exceed twelve ads because there is a certain knowledge of the relevance of each criterion subtracted from a study of the operating environment of the agent and the degree of importance of each criterion, such as the fact that we can find an agent who considers the price criterion to be more important than the other criteria.

When a user selects an advertisement, he will discover all the characteristics and criteria, the reaction of the latter will result in the "slipping" of the latter in the boxes of the profile.

The profile boxes are arranged in the order of the most significant to the left and the least to the right, ranked by positive and negative criteria, the user also has the choice to opt for an option that ensures that the system supports the update of the supported profile on the basis of a choice of two ads, however, if the latter prefers not to react directly to the characteristics of the ads, all to lead to an analysis attributes that motivate the customer's choice.

2.1.2 Interactivity in Databases

In the context of databases, interactivity is rarely used because of the static nature of queries. In general, search spaces are bounded and all predicates are expressed at the same time during the query that is executed based on data sources.

The technique that is potentially usable in order to stimulate user interaction is the proposal of criteria that do not appear in the initial request and then launch a new one based on the previous result. This technique can be repeated until it generates customer satisfaction by the results, there is also an approach that is based on memory management by processing information that is deemed relevant by the user, this approach is done by loading into a buffet, tuples on which the user focuses his attention and then send it to the operators.

There is an interactive query execution technique proposed by [Hellerstein 99], it is a technique that is based on the scheduling used for the purpose of online aggregation driven by the user’s need to find immediate accurate estimates of an attribute value in the group by. The idea is characterized by a change in the order in which the processing of data from sources is done and is processed by requests.

The process of executing the request is therefore divided into four phases: production, scheduling, processing and consumption.
Production: This is disk data reading and network loading
Scheduling: The user preferences are taken into account by using a buffer to store the data to be processed first.

III.THE CONTENT OF THE PROFILE AND ITS USE

3.1 Content of Profile Model Dimensions

There are many articles in the personalization literature, these articles offer user profile templates ([Bradley 00], [Mobasher 00], [Shearin 01], [Jung 02], [Crabtree 98] etc.). personalization is also processed through several internet systems such as [Gauch 99] which presents over fifty systems with the techniques used and the information processed, but the reuse of most of these profiles, unfortunately, remains difficult in other applications.

The model described by [Amato 99] provides a generic profile that includes all the parameters used by customization applications. In our approach, these data refer to an eight-dimensional breakdown: the area of interest, ontologies, personal data, quality, customization, security, customer feedback, and various.

These dimensions are all complementary in their content. The only two parts common to the two models are personal data and user behavior data called feedback, as well as security data. other parts of delivery data are included in the subcategory of delivery preferences in customization

Finally, the three subparts of the data collected from the model proposed by [Amato 99] were included in separate categories:

- Content: the subject matter and terms on which the content of the documents relates and which are part of the keywords of the area of interest, the languages and explanations of the latter which correspond to the definitions of the terms of ontologies.
- The data of the structure subpart of the documents that form part of the information of the coding and layout of the customization part
- The dimension of data quality is the explanation of sources.

It can be said that the core of the profile happens to be the area of interest as well as ontologies in our approach. It is these two dimensions that give the user the opportunity to describe precisely the content he is looking for. The presence of a user profile expresses the stable part of his preferences, targeting a specific subject that is part of these centers of interest, however the other dimensions exist to supplement the additional criteria to the client's requests, which alone cannot form a complete request for information. These two dimensions contain the minimum information that gives meaning to the request.

Once the user interest center and ontology phase is completed, the profile content will be supplemented with data from other categories specific to other fields.

3.2 Relationships between Model Parameters

The dependencies that exist between the profile parameters and between the profile and a personalization system will be explained by supporting through a simplified system (figure) of mediation. The purpose of this analysis is to show the links between the profile category parameters and their level of intervention at the mediation system level. It will be assumed that a "Profile Manager" management is an important player in the entire management of the profile, and this on the mediator side, which is linked to each parameter of the profile, so the module has the task of updating the values of the attributes of the profile. It is also considered the execution module that manages data security.

In the next sections, these execution modules, security managers, the query... will appear to explain the relationship between them and the profile data to understand the user's request.

Personal data are not a criterion in the assessment of requests and they vary very little, so the presentation of this category will not be the subject of this section. There is also the different dimension which is only defined according to the content that the user gives and will not be treated as well.

IV.USE OF USER PROFILE IN A PERSONALIZED INFORMATION SEARCH SYSTEM

4.1 Information Search

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

As we have predefined, information is a set of customizable criteria and preferences specific to each user or user community. His search requires "a system that allows to return from a set of documents, those whose content best corresponds to a user's need for information, expressed using a query". The purpose of the information search is to consolidate techniques for the acquisition, organization, storage, search and selection of information.

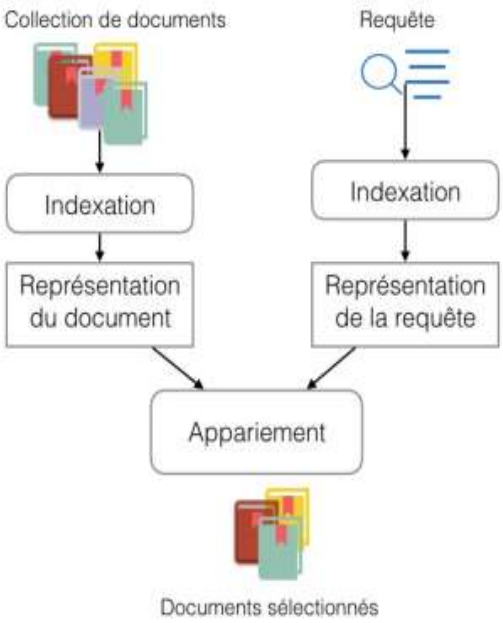

Figure 4.1: Personalization of Information Retrieval

The search for personalized information is intended to address the problem of ambiguities in user-given queries. It uses information about the user and the given query to select specific responses to the user's needs.

To this end, the user profile techniques for the reformulation of requests and the exploitation of the user profile, for the selection of information and the scheduling of search results (Audeh, 2014); Daoud, 2009; Liu, Yu and Meng, 2004; Xu et al., 2007; Zemirli, 2008), as presented above, appears to produce better results than a conventional information search system.

Using the user profile allows the search for personalized information to form the basis for content-based information filtering. With the active user's ratings as the only reference, content filtering must have the active user's ratings on content in the system. Otherwise, the system will lack information and fail to send relevant answers to the user.

Collaborative filtering does not require the provision of specific information on recommendations, it is based on similarity between users to select recommendations, and therefore does not allow the sending of relevant responses to the user.

Hybrid filtering, on the other hand, allows the user to identify the relationship with the recommendations that form the basis of his information filtering.

It should be noted that these 3 types of information filtering are dependent on the relevance of the user profile and therefore on the existence of information that can be exploited by the system, as well as the matching mechanisms used.

V. PERSONALIZATION SERVICES

While the addition of additional information to queries aims to precisely target interesting information but which can be dangerous in the absence of relevance of information, filtering and reordering which are post-processing processes, are intended to adapt the content of the result to a given user.

It has become essential to adapt the way in which information is delivered to users given the development of the Internet, and the enormous amount of information that circulates there, for this reason information systems are starting to offer adapted and personalized services. When a user searches for information, the keywords he uses

in his query are ambiguous and therefore the system does not know the meaning the user wants to give them. in order to solve this problem there are a lot of approaches, in this article we will talk about the results filtering approach, as well as that of the recommendation, the results filtering and the extension of the requests with the interests of the client.

5.1 Filtering the Results

This approach makes it possible to execute the request without taking personalization into account and then apply post-processing on the result in order to eliminate results that are not relevant to the user. Filtering can be done either by applying additional queries to the result, or by treating each element separately in order to study its relevance.

Filtering results has the advantage of its simplicity because it does not require any modification of the functioning of information providers. The disadvantages are the volume of data exchanged between the server and the client which is most often done on the client side, so there is a large transfer of information and in addition there is a risk of deletion of relevant objects by the fact that the elements of the result which are judged to be uninteresting are eliminated definitively.

5.2 The recommendation

To recommend is to offer items to a user according to their preferences or using the experience of other users.
We take the example of [Mobasher 00] as an example of recommendation, we find in this article that each user is associated with a set of transactions T = {t1,…, tm}. Each transaction ti is a multidimensional vector of page views (p = P1, P2 ……, Pi …….. Pn). Then we group these transactions in clusters (groups) in order to maximize the similarity between the transactions of the same cluster and minimize that of the different clusters. Based on one of the clusters, the article defines a user's profile as being a set of couples (pageviews, associated weight):

PRc = {(p, weight (p, PRc)) / p = pageview and weight (p, PRc)> μ}, with μ a given threshold and

weight (p, PRc) = ∑∈ ∗ t C w p t C (,) 1 where t are the transactions of cluster C.

w (p, t) = 1 if p∈t and 0 otherwise

What I propose to detail this formula

Here the weight varies compared to Pi pages and Ti transactions

In this article, we develop even more the formula proposed by [moubacher]

$$\omega_i^c = poids(P_i,c) = \frac{1}{c} * \sum_{t \epsilon c} \omega(P_i,t)$$

$$Rec\ (S, \omega_i^c) = \frac{\sqrt{\frac{1}{c} * \sum_{t \epsilon c}\ \omega_i^c(P_i,T) * \sum\ \omega_i^c * SK}}{\sum\ (SK)^2 * \sum\ (\omega_i^c)^2}$$

$$Rec\ (S, \omega_i^c) = \sqrt{\frac{1}{c} * \sum_{t \epsilon c}\ \omega_i^c(P_i,T) * \frac{\sum\ \omega_i^c * SK}{\sqrt{\sum\ (SK)^2 * \sum\ (\omega_i^c)^2}}}$$

ω_i^c= weight (P_i,c) = 1 / C *, In this formula each pageview Pi corresponds to a value in cluster C and subsequently we will use the two notations to designate a page.

$$\text{Rec (S, } \omega_i^c \text{)} = \sqrt{poids\ (\omega_i^c, C)\ *\ match\ (S, C)}$$

$$\text{match(S,C)} = \sum\ \text{wi * S k/ (s k)}^2 * \sum\ (\omega_i^c)^2$$

$$\text{Rec (S, } \omega_i^c \text{)} = \sqrt{\frac{1}{C}\sum_{t \epsilon c}\ w(P_i,T) * \sum\ w\ i * SK/(SK)^2 * \sum\ (\omega_i^c)}$$

In this formula we indicate that to recommend a page on a REC session (S, Wi) we have two very important factors it is the pageview Pi and the Sk session, according to the demonstration to recommend a page we base ourselves on the user session and the positioning of the View page.

VI. CONCLUSION

We know that each Pi = (M1,... Mi,... Mn) with Mi is the words made up of each Page view so I suggest adding the Word Weight on each Page Pi = (Mi, v) then the Weight of each Pi = 1 / D * SUM (Mi, v) with D = Cluster of Words Mi

And v multidivisional vector of Words v = {v1, v2,.... vi,... vim} we obtain FORMULA (Hind)

rex (s wi) $=\frac{\sqrt{\frac{1}{c}*\sum_{t\in c}\frac{1}{D}*\sum_{v\in D} w(M_i,V)*\sum w\, i*SK}}{\sum (S\,K)^2*\sum (\omega_i^c)^2}$

According to my proposal it is the importance of the Words for each PageView which will highlight the Weight of each Page.

Conclusion

The world is growing, and so are the systems, which makes personalization a crucial approach in the search for information, so price taking into account user preferences is essential, except that it requires reflection at the level of system architecture, languages for querying and building techniques and updating profiles.
Our work is about the construction of the user profile, We propose to take the importance of the Words for each PageView which will highlight the Weight of each Page. To validate the proposed method, our perspective is to carry out tests on a larger number of experiments.

VII. RÉFÉRENCES

[1] [Amato 99] User Profile Modeling and Applications to Digital Libraries, Giuseppe Amato, Umberto Straccia, *Proc. 3rd European Conf. Research and Advanced Technology for Digital Libraries, ECDL,* 1999

[2] [Hellerstein 99] Online dynamic reordering for interactive data processing, V. Raman, B. Raman, and J. Hellerstein, *In VLDB,* 1999

[3] [Bradley 00] Case-Based User Profiling for Content Personalisation, K. Bradley, R Rafter, B. Smyth, Lecture Notes in Computer Science, 2000

[4] [Mobasher 00] Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization, B. Mobasher, H. Dai, T. Luo, M. Nakagawa, Data Mining and Knowledge Discovery,

[5] [Shearin 01] Intelligent Profiling by Example, S. Shearin, H. Lieberman, MIT Media Lab, Cambridge, 2001

[6] [Jung 02] A Formal Model for User Preference, S. Y. Jung, J-H. Hong, T-S. Kim, *IEEE International Conference on Data Mining,* 2002

[7] [Crabtree 98] Knowing Me, Knowing You: Practical Issues in the Personalisation of Agent Technology, B. Crabtree, S Soltysiak, *Proc. Third International Conference on the Application of Intelligent Agents and Multi-Agent Technology (PAAM98), London,* March 23-25, 1998

[8] [Gauch 99] Personalization on the Web, A. Pretschner, S. Gauch, Technical report, In- formation and Telecommunication Technology Center, *Department of Electrical Engineering and Computer Science, The University of Kansas,* 1999.