International Journal for Traffic and Transport Engineering, 2023, 13(1): 79 - 95

DOI: http://dx.doi.org/10.7708/ijtte2023.13(1).07 *UDC: 004.455.1.055:004.738.5*

THE USER PROFILING PROCESS

Valentina Radojičić[1], Slobodan Mitrović[2]

[1,2] University of Belgrade –Faculty of Transport and Traffic Engineering

Received 16 January 2023; accepted 20 February 2023

Abstract: The growth of available information on the Internet and enormous diversity in user's behaviour indicate the exceptional importance of service personalization. Creating of the individual user profile is a key activity that precedes service personalization. Forming an automated user profile is the main challenge in development of the personalized applications that fully meet needs of the user. This research includes the information that needs to be modelled to represent different user profiles, how the information is collected, how to construct the user profile and finally how the user profile is used to deliver a personalized service. The proposed user profiling process includes three basic activities: data collection, user profile construction and personalization. Paper gives the comparison of the user profile methods and techniques. The findings showed that proposed user profiling process improves construction of the accurate user profile for effective service personalization.

Keywords: constructing, data collecting, user profile, modelling, personalization, profiling process.

1. Introduction

Development of information and communication technology has brought an understandable require to have personalized information systems with purpose to adjust information functionality to the specific interest of users. Today, there are numerous online services available to the users through various electronic platforms (e.g. smart phones and televisions, personal computers, etc.) (Radojičić & Mitrović, 2022). User requirements are very heterogeneous, so the creation of user profiles becomes very important for service providers in order to achieve successful service personalization and to adjust service offers to user needs. Personalized services are aimed at meeting the user's requirements and needs, in order to provide increasing user satisfaction. The success of service depends on how well the service provider identifies the user's behaviour, bearing in mind that the user profile is a virtual representation of the user himself. Research works regarding user personalisation are considered as very actual and they involve different approaches related to data science including artificial intelligence. One of the notable applications of the user personalization is the recommender system (Gauch *et al.,* 2007).

A user profile is a set of information representing a user via user related rules, settings, needs, interests, behaviours and preferences (Cufoglu, 2014), (Araniti *et al.,* 2003), (Henczel, 2004). This collection of personal information can either be represented as static data (e.g. native country) or dynamic data (e.g. needs). The content

[1] Corresponding author: valentin@sf.bg.ac.rs

and amount of the information within a user profile can vary depending on the application area such as healthcare sectors, banking sectors, social media, e-commerce, security, access control and social networking (Middleton *et al.,* 2004), (Webb *et al.,* 2001). However, regardless of the information, the accuracy of the user profile is based on how the user information is gathered and organized, and how accurately this information reflects the user. In other words, it depends on the user profiling process in which the information is gathered, organized and interpreted to create the summarization and the description of the user (Henczel, 2004). Various methods, techniques and algorithms within the user profiling process have been proposed in the literature. This paper investigates some aspects of the user profiling process and proposed systematic classification schema for the user profile construction and personalization. The proposed classification schema includes three main activities: data collecting, profiles construction and personalization. This paper gives comparison of the user profile methods through a discussion of techniques used for the user profiling, advantages and disadvantages of available methods for future service personalization and tries to explore all aspects of user profiles modelling. The accuracy of the user profile created depends on how user information is collected and organized. Moreover, how accurately this information about the user is updated during the time. Earlier techniques required collecting data directly from users, when the system explicitly asking necessary data from users. But this method is not considered effective because the user is more often than not interested to providing input data directly. Recent research has focused more to user profiling implicitly based on user actions and behaviour.

This paper investigates some of the most popular techniques for collecting information about users, representing, and constructing the user profiles. It includes both kind of these techniques, implicit and explicit, investigating their contrast through the review of related research including the profile construction difference. This paper is organized as follows: Section 2 presents the literature review. Section 3 presents the classification schema for the user profiling research work. Subsections conduct surveys the research efforts for data collection, building and modelling a user profile, and personalization respectively. Finally, the conducted study is summarized in section 4.

2. Literature Review

User profiling and service personalization is one of the topics that received a special attention in the previous decade, considering wide range of applications. Along with the fact that more than 3000 papers have been published so far, this literature review will present those researches that could serve as a theoretical basis, as well as those articles that refer to the particular researches, as guidelines for solving certain types of problems. Earlier review papers (Gauch *et al.,* 2007), (Cufoglu, 2014) aim (1) to give a general overview of works dealing with user profiling and related concepts, (2) to consider the advantages and disadvantages of current methods for future personalization of services, and (3) to show some details about simulations based on classification and clustering algorithms with a real-world user profile dataset. Research survey (Gauch *et al.,* 2007) presents some early user profiling approaches. Although most of presented solutions have been replaced by newer ones, this work is considered as important because of the theoretical foundations of user profiling. A brief overview of user profiling (Cufoglu, 2014) is considered as significant because it adopts the current classification of profile types and data collection methods. Additionally, this is one of the first surveys that recognizes the classification of personalization methods into collaborative, content-based and hybrid. Also, this paper presents theoretical foundations of the service personalization, since it comprises a review of corresponding works on that topic. Their classification is also confirmed by (Kanoje *et al.,* 2014).

The review paper (Farid *et al.,* 2018) was written with the aim of (1) reviewing the types of information that are necessary for modelling and displaying different user models through the presented research, (2) identifying ways to collect information and maintain the model, as well as identifying models to enable service personalization. Having this in mind, the authors propose a new classification scheme, which includes three user profiling phases: data collection, presentation and construction of profiles, as well as personalization. This scheme is based on the classification of user profiling research studies. A more detailed and comprehensive review of approaches and techniques related to user profiling is presented in the review paper (Eke *et al.,* 2019). This article includes more detailed theoretical statements and an overview of works on the user profile, which were used to adopt the description, characteristics and taxonomy of the user profile. Also, they provided the study of existing modelling of user profiling in terms of data collection, feature extraction, profiling techniques, as well as profiling approaches (with identification of their strengths and weaknesses) and performance metrics. Special emphasis is given to user profile modelling techniques. The paper also discusses research challenges with a focus on privacy, data sets, cold start issues, user trust issues, and computational complexity. Additionally, authors identify a research direction that would offer solutions to the challenges identified in the work, with the aim of further improving mechanisms of user profiling. The theoretical importance of the review is supported later by (Costanzo *et al.,* 2019), which investigates the differences between user preferences in explicit and implicit user profiles, which are obtained by applying advanced user profile modelling techniques.

Considering the problem of constructing user profiles and personalization, there is a significant number of papers, which are focused on a certain type of problem and which can be technologically solved by user profiling / personalization. One of the most important topics is recommender systems, which are applied in almost every segment of the modern life of online users and it is widely covered by papers like (Abdollahpouri *et al.,* 2019), (Jannach *et al.,* 2021), (Lee *et al.,* 2019), (Da'u & Salim, 2019), (Kulkarni *et al.,* 2020), (Milano *et al.,* 2020), (Quijano-Sánchez *et al.,* 2020), (Bai *et al.,* 2019) and (Fkih, 2022). In the papers (Bai *et al.,* 2019) and (Kulkarni *et al.,* 2020) a useful theoretical presentation of recommender system is given. In addition to the above, Kulkarni *et al.* (2020) provides an overview of the main paradigms of recommender systems using explicit and implicit information. On such a basis, several methods were applied with the aim of effectively completing the design of the system recommended for e-learning improvement.

In (Dong *et al.,* 2021), the authors conduct a systematic mapping study on Review-based user profiling (RBUP). They gave an emphasis on those papers, which present research results related to the generic analysis of RBUP process with identification of potential research directions. Analysis included the observation of processes, through which user profiling could be performed. The analysis identified two research challenges. The first one is related to the generation of the initial user (explicit) profile, while the second one is related to the permanent updating of profile information, in order to synchronize that state with the user's current preferences. These problems could be overcome with the application of machine learning techniques. It must be noted that other techniques could be engaged, as well, and this stream is supported with a number of papers related with other techniques of user profiling. Hawashin, *et al.* (2021) propose the use of intelligent agents for extracting the interests of users who can be grouped according to gender or professional orientation type. Since the lack of classification is related to the so-called conflict of identified interests, the authors suggest ranking agents according to their ability to extract the interests that best reflect the users. Therefore, this paper takes user feedback about the most representative agent, and then classifies it in order to predict the most representative agent for new users. El Houda *et al.* (2019) propose the use of genetic algorithms along with user queries, in order to extract key user interests and/or correct their weighting coefficients. Then, the identified key interests would be used to discover new interests needed to update the user profile, as well as to remove those ones, which are no longer needed. Bearing in mind that user profiles refer to groups of individuals with random characteristics, profiling techniques can be used to create fake profiles (so-called bots), which can be further used in different kinds of campaigns. In order to develop mechanisms for efficient fake profiles recognition, Hayawi *et al.* (2022) propose the use of deep neural networks as the engine that can classify Twitter accounts as "human" or "bot". Such an engine would use metadata obtained from the Twitter account user profile elements such as description, number of followers and number of tweets. Also, it would handle mixed feature data types including numeric, binary and textual, transforming the observed model into hybrid. The model uses long-term memory units (LSTM) and dense layers to accept and process mixed types of input data. Its efficiency is evaluated using a collection of publicly available datasets. A similar study was conducted by Heidari *et al.* (2022), who developed several machine learning models in order to detect bots based on extracted user profile from Tweeter message. Bearing in mind that a user's profile is technically rely on the user's online postings, the authors set development goals, which include improved bot detection through machine learning models based on users' personal information. It is based on a similarity of personal information making it quite difficult to distinguish a bot from a human user. However, this similarity is actually recognized as the advance in the new machine learning model that provides a way to detect social bots with high predictive accuracy based on personal information.

Hu *et al.* (2017) proposed a vector space model for forming an accurate user profile, which comprises a word vector and a weight vector, which make it possible to infer basic information about the user by analysing the user's search logs in a certain period of time. One of the recent papers (Mamun *et al.,* 2021) further analyses the development of models for online users profiling. This analysis includes existing approaches, methods, and research challenges. The paper also shows ways to apply profiling of online users, through the identification of patterns appearance, as well as online interest patterns, regarding the data collection, feature extraction and technical profiling. In their work on service personalization, Abri *et al.* (2020) propose a model for assessing personalization using a thematic user profile, which combines personalized and non-personalized thematic models. This model introduces a new metric related to the thematic distribution of user documents in the thematic user profile, on which the potential for personalization could be assessed for all user queries. There is a significant number of authors who observe from different aspects the personalization of services throughout different recommender systems. Kulkarni *et al.* (2020) reviews the main paradigms of recommender systems, which use explicit and implicit feedback and different methodologies that have been applied to design recommender systems in e-learning systems. The paper also summarizes the e-learning concepts, viewed from the aspect of the recommender system. Lops *et al.* (2019) provide an overview of works related to the modelling of recommender systems based on collaborative filtering as well as content-based filtering. Nilashi *et al.* (2018) propose a recommender system development method that is using a collaborative filtration approach. By applying dimensionality reduction and ontology techniques, the authors try to solve two recognized shortcomings - sparsity and scalability. Fkih (2022) evaluates recommendation systems using similarity measures in collaborative learning systems, which are presented in an experimental study. Valcarce *et al.* (2019) present an embedded model for user representation in memory-based recommender systems that rely solely on user preferences through ratings. The authors use the dropout variant as profile regularization to improve performance and prevent overfitting. Da'u & Salim (2019) published a systematic review of deep learning-based recommender systems that summarizes and analyses existing studies based on relevant research publications.

Jannach *et al.* (2021) also present a detailed review of Conventional Recommender Systems considering existing approaches to conversational recommendations, which could be categorized by different dimensions/contexts, such as supported user intentions or background knowledge.

3. The User Profiling Process

A major challenge within the user profiling process is how the user profile can be built to accurately reflect users' preferences (Farid *et al.,* 2018). Figure 1 presents the user profiling process that generally consists of three main activities: data collecting, profiles construction and personalization. The first activity considers gathering of data about the users and how it can be extracted. Second, building and modelling of the user profile investigated the way for representing user profile groups and techniques of its construction. Third, how to identify the individual user and how to explore the applications in order to provide personalized services to him?

3.1. Data Collecting

The main activity of the user profiling process is collecting information about the users: gathering, obtaining and extracting data. The level of generated data related to user profile is highly correlated to profile accuracy. There are static and dynamic types of data that could be able to define a user profile. Static data assumes the user personal data or demographic information, knowledge and skills, user needs and goals (age, sex, income level, race, employment, location, homeownership, level of education, e-mail address, screen size or other device info, etc.). Otherwise, dynamic data are user's behaviour, interests and preferences (behaviours of clicking, watching, browsing, etc.). In contrast to the static profile, the dynamic profile is auto-generated by the system and consequently, the user attribute and contents go through changes over time. In dynamic profiling, the profile information about user's behaviour seeks to determine future information about the user more than the present information (Araniti *et al.,* 2003). In other words, it is referred to as a behavioural or adaptive profile. The dynamic profile is always accurate in a situation where there is a high velocity of delivery data. It may differentiate between short-term and long-term interests. Short-term profiles represent the user's current interests whereas long-term profiles indicate interests that are not subject to frequent changes over time (Eke *et al.,* 2019). Likewise, dynamic data could be less-dynamic which does not change very often (name, age, level of education, roaming, profile picture on Facebook, Instagram etc.) and more-dynamic (Calls and SMS messages, likes etc.) which changes on at least a daily basis and real-time, which needs to be tracked frequently in real-time, in order to achieve satisfactory observations (location, orientation, light level and gyroscope, etc.) (Smailović, 2016).

The user profiling methods could be explicit, implicit, or hybrid. Explicit user profiling methods of gathering data are used static and predictable characteristics of the user (Poo *et al.,* 2003). These are manual techniques such as registration forms, questionnaires, or user classified training sites, or by asking users to rate items, or by tracking users' query words. Disadvantage of this method is that the explicit profiles have a static nature and are valid only until the user changes his interest and preferences parameters. The explicitly created user profile is called explicit or static user profile.

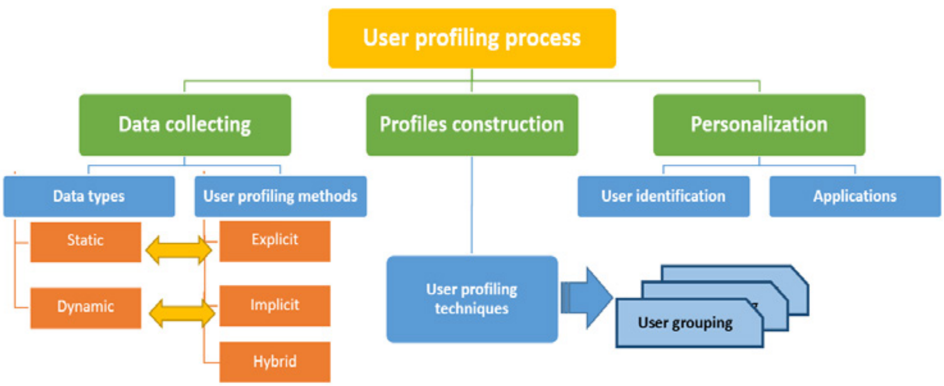

Fig 1.
The User Profiling Process

On the other hand, implicit information is gathered dynamically by continuous monitoring the user's interest and preferences through interactions by the system automatically. Unlike static profiling method, dynamic profiling uses the implicit method and analyses user's behaviour pattern or usage history to determine user's interests (Henczel, 2004). Hence, the user profile created by implicit method could be referred as implicit/dynamic user profile. The implicit method is also called as Behavioural profiling, Adaptive Profiling or Ontological Profiling of the user (Poo *et al.,* 2003). It attempts to infer the user's interests or context from the processed logs, browsing history from web, proxy servers, purchased items, examined products, bookmarked pages, preferred brands, restaurants rated, followers on social media, GPS data logged, links sent to friends, or the information gathered automatically without any effort from the user. Some of the main operations that are being pointed out specifically are: text tracing, link pointing, but not clicking the link, link clicking, text selection, scrolling a window at a certain speed, registering a page as a bookmark, saving an HTML document, printing a page, moving a window of the Web browser, etc. A lot of research literature can be found which discussed some filtering techniques, some of which could be rule based filtering, collaborative filtering and content based filtering techniques (Kanoje *et al.,* 2014).

It is also possible to produce a hybrid user profile which can be achieved in two ways (Henczel, 2004). The first way starts by using the explicit techniques to collect the initial data, followed by the implicit techniques to update the user profile. In the second way, on the other hand, implicit techniques are followed by the explicit techniques. In general, it has been cited that the hybrid methods are more efficient than both of the fundamental methods. This approach helps profiling more efficient and maintains the accuracy of temporal information as information gets updated temporally (Kanoje *et al.,* 2014). Table 1 compares the aforementioned user profile methods (Cufoglu, 2014).

Different kind of user's data could be extracted depending on collection method and its nature determines the way and the recommending items that the system can execute personalization. Extracting data is nothing else but obtaining the useful information about a user from different sources. For this cause many methodologies and models have been used by different researches like data extraction from different sources such as web and social media websites. It also comprises some techniques related to user behaviour analysis which help user profiling system gather interesting information about users (Kanoje *et al.,* 2014).

3.2. Profiles Construction

User profiles are constructed from information sources using a variety of construction techniques mostly based on machine learning or information retrieval. Depending on the user profile representation desired, different techniques may be appropriate. The profiles may be constructed manually by the users or experts, however, this is difficult and time consuming for most users and would be a barrier to widespread adoption of a personalized service. Techniques which automatically construct the profiles from user feedback are much more popular. Although some approaches use genetic algorithms or neural networks to learn the profiles, there is a whole spectrum of simpler, more efficient approaches based on probabilities or the vector space model, which are used. All of them could be chosen with the same goal: to keep the accuracy of user preferences in up-to-date state. Profile updating can be done automatically and/or manually. Automatic methods are preferred because it is less intrusive to the end user (Gauch, et al., 2007).

Table 1

Comparison of the User Profile Methods

User profiling methods	Descriptions	Techniques	Advantages	Disadvantages
Explicit	User manually creates user profile;	Questionnaires, registration forms, user classified training sites, rating items, tracking users' query words;	High quality of information;	The explicit profiles are valid only until the user changes his interest and preferences parameters; Lot of efforts from user to update the profile information;
Implicit	System monitor users interactions between user and content;	Filtering techniques, Machine learning algorithms, artificial intelligence, etc.;	The information gathered automatically without any effort from the user; Easily update by automatic methods;	Initially requires a large amount of interaction between user and content before an accurate user profile is created; Typically less accurate than explicit data;
Hybrid	Combination of explicit and implicit user profiles;	Both explicit and implicit techniques;	Combines explicit and implicit methods to leverage the benefits of both methods; accuracy is maintained by temporal updating;	N/A

Profile Integration assumes a problem of data cleaning. Its might happen that some of data that has been collected might be a duplicate or it may look like duplicate. So it is necessary to identified duplicate data. After the profile integration is done it is necessary to put the users into the different groups. This can be done by grouping the users based on their behaviours into the same group by various techniques. Usually, clustering algorithms are deployed to group user data objects depending on the information contained in the data that specifies the objects and their association. The group behaviour determines the grouping of users into separate classes using some clustering algorithm. It also deals with the assignment of a set of observations into the clusters in such a way that the observation in a similar cluster looks alike in some sense. Furthermore, it uses a classifier called K-means to classify objects based on attributes into k numbers of the groups (Eke *et al.,* 2019).

3.2.1. User Profiling Techniques and User Grouping

In recent years, several modelling techniques have been employed in the construction of the user profile: neighbourhood-based, machine-learning, ontology-based, statistical modelling and filtering approach (Eke *et al.,* 2019). A neighbourhood with adequate knowledge can assist each user to build the neighbourhood user profile in order to address the inherent shortage of information in personal interest representation. A machine-learning algorithm is used for training and testing of data. Computers learn from the algorithm that consists of two broad forms, namely supervised learning (input mapped to desired output) and unsupervised learning (auto-detection of data disregarding pattern to class assignment). In a supervised learning approach, the systems learn how to perform a task of new observation classification from the input data. The algorithm learns from the available training data and uses its application on real data (Kotsiantis, 2007). The most useful supervised learning for profiling is K-Nearest Neighbour, Naive Bayes and Support Vector Machine (Bradley *et al.,* 2000), (Eke *et al.,* 2019). Ontology-based techniques could improve user profiling because of evolution of the semantic web. These techniques could efficiently share relevant user's information in other systems, without common difficulties (Maria *et al.,* 2007). A statistical model is a technique that uses keywords as a dataset for building a user profile. In web system, this technique may consist of a highly frequent word obtained from the visited web page by the user (Tang *et al.,* 2010). On the other hand, the filtering recommendation approach to profiling is a method of filtering information that meets the user's specific need in different situations and removes the irrelevant information about the user. This approach consists of rule-based, content-based, collaborative-based and hybrid methods.

Rule-based approach are specified by the information system based on the demographic similarity or a static profile of users obtained via the registration process by asking users a set of questions (Choi & Han, 2008). Demographic recommendations consist of: (1) Collecting demographic information about the users; (2) Aggregating the users into the clusters; (3) Using a similarity measure and data correlation; (4) Generating cluster-based recommendation; It`s effectiveness usually relies on knowledge quality of the rules. Demographic recommendations are efficient but require domain engineering by human experts and involve expensive collection of demographic data, do not track the changes in the population. Demographic similarity does not necessarily imply preference similarity. However, it has poor maintenance issues and is prone to bias since the input is the subject of the user's self-description or their interests (Eke *et al.,* 2019).

Content-based filtering methods are the popular techniques of recommendation systems. Those methods also called as Cognitive filtering methods. The recommendations depends on users former choices. Item description and a profile of the user`s orientation play an important role. Content-based filtering algorithms usually identify and count similarities, as well as items with specific keywords. By using user's features and likes they create a dataset used for recommending with similar things, which could be probably liked. It uses the online information and what it is able to collect, and then makes recommendations accordingly. This type of recommender system is hugely dependent on the inputs provided by users (examples included Google, Wikipedia, etc.). For example, when a user searches for a group of keywords, then Google displays all the items consisting of those keywords. This technique mostly depends on the explicit ratings or preferences given by the users for particular item and tries to find users with similar ratings for the item. But in practice users never tend to give explicit ratings to the system. Therefore, a mechanism is needed that will implicitly identify the rating or preference of a user for the particular item. However, when the information is not sufficient, it will lead to a cold-start problem. The cold start problem is common in learning and adapting dynamic user profiles for personalization, where the system is not capable of providing an effective personalization service in order to learn the user profile (Eke *et al.,* 2019). The content-based recommendation system usually works on two methods, both of them using different models and algorithms. One uses the vector spacing method, while the other uses the classification method. The vector spacing method implies that user interest's keywords are extracted from visited documents during browsing. They are represented either by a single vector that includes all the interest or with multiple vectors, which reflects interest in several domains. In this method the effectiveness of the user profiles depends on the vectors' degree of generalization. The classification method creates a decision tree and find out what the user really wants. The other knowing content-based filtering methods are Latent Semantic Indexing, Learning Information Agents, Neural Network Agents etc. (Tang *et al.,* 2010).

Collaborative filtering is most extensively used approach to design recommender system. User's interest in an item is established based on the user's previous interests on the same item. These methods are established on gathering and examining a large amount of information which based on users behaviour, activities or preferences and anticipating taste of that particular user by using their similarity with other users. Collaborative filtering uses large number of users' evaluations as a source for its recommender dataset. This information is usually recorded as a matrix, with the rows representing users and the columns representing items (Kanoje *et al.,* 2014). The basic premise of such systems is that the historical data should be sufficient to generate a prediction. It does not depend on machine decomposable message. This type of filtering works with an algorithm that aggregates the feedback provided by different users and recommends items for users by considering the similarities between users in order to offer recommendations to the target users. There are two groups of methods regarding collaborative filtering: Memory-based and Model-based methods. Memory-based methods are the most basic because they use no model at all (Kanoje *et al.,* 2014). They assume that predictions could be based solely on "memory" of past data and typically use a simple distance measurement approach, such as the nearest neighbour. Model-based approaches, on the other hand, usually suppose some form of the underlying model and attempt to ensure that any predictions made fit the model properly. The model-based approach assumes that users of the same group (e.g. age, sex, social group) based on their similar behaviour and have the same user profile group as a result. This technique is heavily dependent on clustering of profiles and efficiency to associate them to users. Collaborative filtering algorithm is highly scalable and produces high - quality recommendations with large datasets. These filtering approaches cannot help in a cold-start situation with the absence of the user's initial ratings (Gauch *et al.,* 2007), (Farid *et al.,* 2018).

A hybrid method, also referred as hybrid filtering method, usually uses content-based and collaborative methods to combine the advantages and overcome the limitations of both methods (Gauch *et al.,* 2007), (Farid *et al.,* 2018), (Kelly & Teevan, 2003). In this way the immediate users' profile availability is guaranteed, since this method supplies a more accurate description of user preferences and interests, which are further permanently updated with results of monitored user-system interaction (Radojičić & Mitrović, 2022). Generally, the hybrid method assigns to new user a default profile with the use of the collaborative method and further enhances the profile using the content-based method (Radojičić & Mitrović, 2022). In the literature four hybrid user profiling techniques have been introduced Smailović (2016). These are called: 'static content profiling', 'dynamic content profiling', 'static collaborative profiling', and 'dynamic collaborative profiling'. The static content profiling is the combination of static profiling and content-based method. Here, the information about user's interests is gathered during registration. Consequently, in dynamic content profiling, information about user's interests are retrieved via monitoring user's behaviour. Moreover, in static collaborative profiling, information relating to user's interests is collected based on user's explicit requests.

On the other hand, dynamic feedback from the users initiates possibility to gather information from users' grouped by similar behaviours. This method is called dynamic collaborative profiling. This approach has been proven effective in several application areas such as web search, electronic commerce, sensing, monitoring, and financial-based systems. The hybrid filtering approach is introduced to overcome some common problem that are associated with above filtering approaches such as cold start problem, over specialization problem and sparsity problem (Thorat & Barve, 2015). Another motive behind the implementation of hybrid filtering is to improve the accuracy and efficiency of the recommendation process. Table 2 compares the aforementioned three main filtering methods with respect to their techniques, advantages and disadvantages (Cufoglu, 2014).

Table 2

Comparison of Three Main Filtering Methods

Filtering methods	Description	Techniques	Advantages	Disadvantages
Content-based methods	Depends on users former choices; Try to recommend items based on similarity count;	Vector Space model; Latent semantic indexing; Learning information agents; Neural network agents; Classification method;	Objective analysis of large multimedia sources without involvement of the users;	Depends on the explicit ratings or preferences given by the users; When the information is not sufficient, it will lead to a cold-start problem; Hard to introduce recommendations;
Collaborative-based methods	Established on gathering and examining a large amount of information which based on users behaviour, activities or preferences;	Memory-based and Model-based	Content independent; Historical data should be sufficient to generate a prediction; More accurate than content-based filtering; Highly scalable and produces high - quality recommendations with large datasets;	Poor predictive capabilities when a new item is introduced into the database due to lack of ratings; Cannot help in a cold-start situation with the absence of the user's initial ratings;
Hybrid methods	Uses content-based and collaborative methods to combine the advantages and overcome the limitations of both methods;	Static content based profiling; Dynamic content based profiling; Static collaborative profiling; Dynamic collaborative profiling;	Overcome cold start problem, specialization problem and sparsity problem; Improve the accuracy and efficiency of the recommendation process;	The method weaknesses can outweigh the strengths if the hybrid method is not well designed;

3.3. Personalization

The third and final activity in the development of the user profiling process refers to personalization. Personalization is the ability to provide content and services tailored to individuals based on knowledge about their preferences and behaviour. Moreover, personalization is defined as adjusting the service to suit the interests, preferences and needs of the user (Kelly & Teevan, 2003). Personalization exploits user profile to filter information and provide personalized services in various areas, such as personalized recommendation systems, mobile services, social networks, multimedia systems, web search and browsing, etc.

Service personalization is functional change of the information content in order to increase its ability to recognize individuality. User as an individual could be modelled through knowledge, interests, goals and motivation, background, personality and traits, interactions with system etc. In accordance of data types, there are two service personalization methods: implicit and explicit personalization. In implicit personalization, information about the user is collected implicitly (through intuitive interaction with the observed system). Therefore, the user should not be aware of the information collection process. On the other hand, in explicit personalization, user profile information is collected through direct engagement with the user. In that case, a user is aware of the information gathering process. In implicit personalization, accuracy improves with the user's continued use of the system. However, in explicit personalization, accuracy of personalized information heavily depends on the involved user (Radojičić & Mitrović, 2022). Personalization requires identification of the individual user at the first place. In other words, activity of personalization begins in an explicit form, because information related to the user's interests and preferences have to be explicitly provided by the user to the system. This could be considered as the serious challenge, considering the fact that users rarely provide all their information accurately, since they often feel that their privacy is compromised. Hence, this is considered the most important shortcoming in the personalization activity. However, the accuracy of the entered information does not have to be considered as critical issue if the context of personalization is related to the group user profile development. Conversely, this could be serious issue for any system that places explicit personalisation at the function of the individual user (Gauch *et al.,* 2007).

Personalized search/content browsing is the most common way to use personalized systems. It could be considered as an activity of searching for content that might include documents, web content, social network interactive content, audio and video material, gaming session and software content and appearance customisations within operating systems different platforms.

In all cases, preference is given to the user's interest identified from the created user profile, while the accuracy of the system's performance is improved by predicting the user's interest better compared to conventional search. In this case, the important role is played by recommender systems, which have already been described in this text. It is important to notice that the search results and content navigation are dependent on the dynamics of changing preferences during the search. By this way, personalization can influence the time component within the level of prediction accuracy have to be reached.

3.3.1. Areas of Application

Farid *et al.* (2018) in their review listed the areas of application of personalized systems:

- Personalized search and web browsing: this type of online search occurs in accordance with the identified interests of the user. It operates by activating a prediction model based on the available data gathered from the profile (e.g. on the results of the analysis of website visits, by collected cookies, session data, etc.).
- Recommender systems: this is one of the most common examples of the application of personalization in practice. A recommender system is actually a subclass of an information filtering system that operates with the intention of predicting an item's rating or a user's preferences. The personalized recommendation system works with two goals: (1) to recognize the user as efficiently as possible and (2) to recommend items from the domain of his interest as accurately and efficiently as possible (Farid *et al.,* 2018).
- Adaptive learning systems: adaptive learning is one of the forms of e-learning, in which students direct it towards the set course goals based on their abilities

and preferences. The functioning mechanisms of these systems are referenced in works such as (Farid *et al.,* 2018), (Kulkarni *et al.,* 2020).

- Visualization: profile visualization of users' semantic models on social networks refers to the possibility of displaying and organizing user interest models based on ontology principles (Wang & Chang, 2014).
- Personalized online social network services: the representation of data on the user interface of the page or application of the observed social network is formed based on the application of profiling techniques. On this basis, groupings of topics and the order of presentation are classified in accordance with the identified belonging to user groups, as well as according to the calculated relevance, based on the various mechanisms mentioned in the literature review.

4. Conclusion

This paper provides a very clear and simplified block diagram of the user profiling process. We discuss about the characteristics, limitation and relationship between the static and dynamic data types which require application of the existing methods to user profile construction. Identifying reliable data sources is the main task which needs to be work on. This task of identifying the sources needs to be automated so that existing systems can make use of various techniques to create a novel user profile that would help in various applications. Beside this, we compared the explicit, implicit and hybrid user profiles methods by studying the advantages and disadvantages between them. Many operations need to be considered in constructing an adequate and consistent user profile group with similar needs, behaviour and preferences of the users. Such facts include choosing appropriate techniques. In this paper, filtering techniques are described in detail and their comparative characteristics are given. The hybrid filtering approach is introduced to overcome some common problem that are associated with Content-based and collaborative-based methods such as cold start problem, over specialization problem and sparsity problem. Another motive behind the implementation of hybrid filtering is to improve the accuracy and efficiency of the recommendation process. Personalization in context of user profiling modelling is the last activity, but a very important one. It implies separating a user as some individual from the users group, and directing the selected activities towards him. For future work, it is necessary to introduce the framework for evaluating recommender systems, which is based on recommender systems evaluation. By this system, the evaluation space of recommender systems has to be evaluated.

Acknowledgements

This paper was partly supported by The Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

Abdollahpouri, H.; Mansoury, M.; Burke, R.;Mobasher, B. 2019. The Unfairness of Popularity Bias in Recommendation. Available from Internet: <https://arxiv.org/abs/1907.13286#>.

Abri, S.; Abri, R.; Cetin, S. 2020. Estimating Personalization using Topical User Profile. In *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - KDIR,* 145–152.

Araniti, G.; De Meo, P.; Iera, A.; Ursino, D. 2003. Adaptively controlling the qos of multimedia wireless applications through "user profiling" techniques, *IEEE Journal on Selected Areas in Communications* 21(10): 1546-1556.

Bai, X. et al. 2019. Scientific Paper Recommendation: A Survey, *IEEE Access* 7: 9324-9339.

Bradley, K., Rafter, R.; Smyth, B. 2000. Case-Based User Profiling for Content Personalisation. In *Lecture Notes in Computer Science.* Springer Berlin Heidelberg, 62-72.

Choi, O.; Han, S. 2008. Personalization of Rule-based Web Services, *Sensors* 8(4): 2424-2435.

Costanzo, L. L.; Deldjoo, Y.; Dacrema, M. F.; Schedl, M.; Cremonesi, P. 2019. *Towards Evaluating User Profiling Methods Based on Explicit Ratings on Item Features.* Available from Internet: <https://arxiv.org/abs/1908.11055>.

Cufoglu, A. 2014. User Profiling - A Short Review, *International Journal of Computer Applications* 108(3): 1-9.

Da'u, A.; Salim, N. 2019. Recommendation system based on deep learning methods: a systematic review and new directions, *Artificial Intelligence Review* 53(4): 2709-2748.

Dong, X.; Li, T.; Song, R.; Ding, Z. 2021. Profiling users via their reviews: an extended systematic mapping study, *Software and Systems Modeling* 20(1): 49-69.

Eke, C.; Norman, A.; Shuib, L.; Nweke, H. 2019. A Survey of User Profiling: State-of-the-Art, Challenges, and Solutions, *IEEE* Access 7: 144907-144924.

El Houda, B.; Nadjia, B.; Abdelkrim, M.; 2019. Queries-Based Profile Evolution using Genetic Algorithm. In *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA).* IEEE. Available from Internet: <https://ieeexplore.ieee.org/document/9035351>.

Farid, M.; Elgohary, R.; Moawad, I.; Roushdy, M. 2018. User Profiling Approaches, Modeling, and Personalization. In *Proceedings of the 11th International Conference on Informatics & Systems (INFOS 2018).* Available from Internet: <https://ssrn.com/abstract=3389811>.

Fkih, F. 2022. Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison, *Journal of King Saud University - Computer and Information Sciences* 34(9): 7645-7669.

Gauch, S.; Speretta, M.; Chandramouli, A.; Micarelli, A. 2007. User Profiles for Personalized Information Access, *The Adaptive Web. Lecture Notes in Computer Science* 4321: 54-89.

Hawashin, B.; Mansour, A.; Fotouhi, F.; Alzu'bi, S.; Kanan, T. 2021. A Novel Recommender System Using Interest Extracting Agents and User Feedback. In *2021 International Conference on Information Technology (ICIT)*, 674-678.

Hayawi, K.; Mathew, S.; Venugopal, N.; Masud, M.; Ho, P. H. 2022. DeeProBot: a hybrid deep neural network model for social bot detection based on user profile data, *Social Network Analysis and Mining* 12(43). Available from Internet: <https://link.springer.com/article/10.1007/s13278-022-00869-w>.

Heidari, M.; Jones, J. H. J.; Uzuner, O. 2022. Online User Profiling to Detect Social Bots on Twitter. Available from Internet: < https://arxiv.org/abs/2203.05966>.

Henczel, S. 2004. Creating user profiles to improve information quality. Available from Internet: <https://www.researchgate.net/publication/262729945_Creating_User_Profiles_to_Improve_Information_Quality>.

Hu, J.; Jin, F.; Zhang, G.; Wang, J.; Yang, Y. 2017. A User Profile Modeling Method Based on Word2Vec. In *2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, 410-414.

Jannach, D.; Manzoor, A.; Cai, W.; Chen, L. 2021. A Survey on Conversational Recommender Systems, *ACM Computing Surveys* 54(5): 1-36.

Kanoje, S.; Girase, S.; Mukhopadhyay, D. 2014. User profiling trends, techniques and applications. Available from Internet: <https://doi.org/10.48550/arXiv.1503.07474>.

Kelly, D.; Teevan, J. 2003. Implicit feedback for inferring user preference, *ACM SIGIR Forum* 37(2): 18-28.

Kotsiantis, S. B.; Zaharakis, I.; Pintelas, P. 2007. Supervised machine learning: a review of classification techniques, *Emerging artificial intelligence applications in computer engineering* 160(1): 3-24.

Kulkarni, P.; Rai, S.; Kale, R. 2020. Recommender System in eLearning: A Survey. In *Proceedings of International Conference on Computational Science and Applications,* 119-126.

Lee, H.; Im, J.; Jang, S.; Cho, H.; Chung, S. 2019, MeLU: Meta-Learned User Preference Estimator for Cold-Start Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19*, 1073–1082.

Lops, P.; Jannach, D.; Musto, C.; Bogers, T.; Koolen, M. 2019. Trends in content-based recommendation, *User Modeling and User-Adapted Interaction* 29(2): 239-249.

Mamun, M.; Al-Digeil, M.; Ahmed, S. 2021. Profiling Online Users: Emerging Approaches and Challenges. In *Securing Social Networks in Cyberspace*, 221-240.

Maria, G.; Akrivi, K.; Costas, V.; Constantin, H. 2007. Creating an ontology for the user profile: Method and applications. In *Proceedings AI*AI Workshop RCIS*, 407-412.

Middleton, S.; Shadbolt, N.; De Roure, D. 2004. Ontological user profiling in recommender systems, *ACM Transactions on Information Systems* 22(1): 54-88.

Milano, S.; Taddeo, M.; Floridi, L. 2020. Recommender systems and their ethical challenges, *AI & Society* 35(4): 957-967.

Nilashi, M.; Ibrahim, O.; Bagherifard, K. 2018. A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques, *Expert Systems with Applications* 92: 507-520.

Poo, D.; Chng, B.; Goh, J.M. 2003. A hybrid approach for user profiling. In *36th Annual Hawaii International Conference on System Sciences*, 9p.

Quijano-Sánchez, L.; Cantador, I.; Cortés-Cediel, M.; Gil, O. 2020. Recommender systems for smart cities, *Information Systems* 92: 101545.

Radojičić, V.; Mitrović, S. 2022. Modelling of Telecommunication Users Profiles [In Serbian: Modeliranje korisničkih profila telekomunikacionih servisa]. In *Proceedings the XL Symposium on new postal and telecommunications technologies - PosTel 2022*, 213-222.

Smailović, V. 2016. *User profile for provisioning information and communication services based on user influence.* PhD thesis. University of Zagreb, Faculty of Electrical Engineering and Computing.

Tang, J.; Yao, L.; Zhang, D.; Zhang, J. 2010. A Combination Approach to Web User Profiling, *ACM Transactions on Knowledge Discovery from Data* 5(1): 1-44.

Thorat, P. B.; Barve, S. 2015. Survey on collaborative filtering, content-based filtering and hybrid recommendation system, *International Journal of Computer Applications* 110(4): 31–36.

Valcarce, D.; Landin, A.; Parapar, J.; Barreiro, Á. 2019. Collaborative filtering embeddings for memory-based recommender systems, *Engineering Applications of Artificial Intelligence* 85: 347-356.

Wang, Y.; Chang, S. K. 2014. User Profile Visualization to Facilitate MSLIM-Model-Based Social Influence Analysis Based on Slow Intelligence Approach, *International Journal of Software Engineering and Knowledge Engineering* 24(10): 1573-1587.

Webb, G.; Pazzani, M.; Billsus, D. 2001. Machine Learning for User Modeling, *User Modeling and User-Adapted Interaction* 11(1/2): 19-29.