

**A. YEROKHIN,**  
*Doctor of Sciences, Professor,  
Chief of Informatics Department of  
Kharkiv National University of Internal Affairs*  
**A. KOLCHENKO,**  
*teacher of  
Kharkiv National University of Internal Affairs*  
**M. KOLCHENKO,**  
*student of  
Kharkiv National University of Radioelectronics*

---

### THE DEVELOPMENT OF INFORMATION SYSTEM BASED ON THE USER INTERESTS

---

In former services (Web 1.0), user, for the most part, played a passive role of consumer services.

The concept of Web 2.0 implies the active involvement of users to Internet resources – quite often a large value (if not the principal) is assigned to the content created by users. The most striking examples of this are multimedia content services – Wikipedia, Last.fm, Flickr, YouTube.

Information, necessary to the user, is usually given by different services.

However, most services provide only part of the information. User is obliged to work with a set of such services and «glue» different information mentally to draw the whole picture of some problem domain.

Fortunately, there are software solutions that allow to «glue» the separate pieces of information to obtain a single picture. This is mashup technology. There are different solutions for mashups development. The most known of them are mashup editors (for example, Yahoo! Pipes, Dapper and Microsoft Popfly).

The next step for improving the quality of interaction is the personalization of Web applications. Personalization allows adapt a web application for the needs of given user, based on his interests and preferences. Personalization is the some kind of the information filter, because it allows to display only relevant content to the user. The core of personalization is the User model. The User profile is the widely accepted kind of the such model.

User Profile keeps characteristics of the user, essential for given task. Thus, for the music

service, it's necessary to store user's age, location, history of visits to the resource, the history of navigation and search, the list of favorite artists and tags, and other information in the User Profile.

This profile does not allow to receive relevant musical information only. Not obvious at first sight example is: the system offers the user to purchase a ticket to a concert of favorite artist in a nearby town.

How to generate user profile? It is necessary to collect some data about the user. There are two approaches to create a profile. When building the user's profile a distinction is made between explicit and implicit forms of data collection.

Examples of explicit data collection include the following:

- asking a user to rate an item on a sliding scale;
- asking a user to rank a collection of items from favourite to least favorite;
- presenting two items to a user and asking him/her to choose the best one;
- asking a user to create a list of items that he/she likes.

Examples of implicit data collection include the following:

- observing the items that a user views in an online store;
- analyzing item/user viewing times;
- keeping a record of the items that a user purchases online;
- obtaining a list of items that a user has listened to or watched on his/her computer;
- analyzing the user's social network and

discovering similar likes and dislikes.

The class of systems that use content filtering algorithms for the issue of user-relevant information is called Recommender systems.

Recommender systems forms a specific type of information filtering (IF) technique that attempts to present information items (movies, music, books, news, images, web pages, etc.) that are likely of interest to the user. Typically, a recommender system compares the user's profile to some reference characteristics, and seeks to predict the 'rating' that a user would give to an item they had not yet considered. These characteristics may be from the information item (the content-based approach) or the user's social environment (the collaborative filtering approach).

It is important to note, that the profile should reflect the current interests of the user. This implies the need for continuous updating of data about the user.

Thus, the advantages of personalization are:

1. Improve the user experience;
2. Save time;
3. Save money.

Disadvantages of personalization:

1. Need to register user;
2. Violation of privacy of user data.

**Conceptual design.** Web-based application

that combines different services, such as Last.fm, Flickr, YouTube and Flickr was developed.

The application allows user to search for the artists, then to view information about them, including images (from Flickr), videos (from YouTube) and incoming events (from Last.fm, displayed Google.Maps). Application does not require user registration.

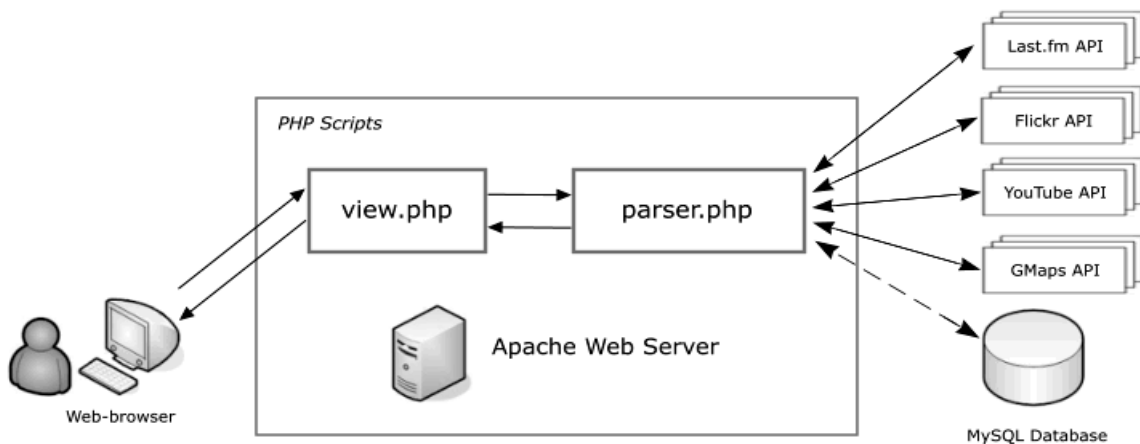
This work is based on the previous one, but significantly extends its functionality.

The main feature of the current project is detection of users and resources that fall into given user interests.

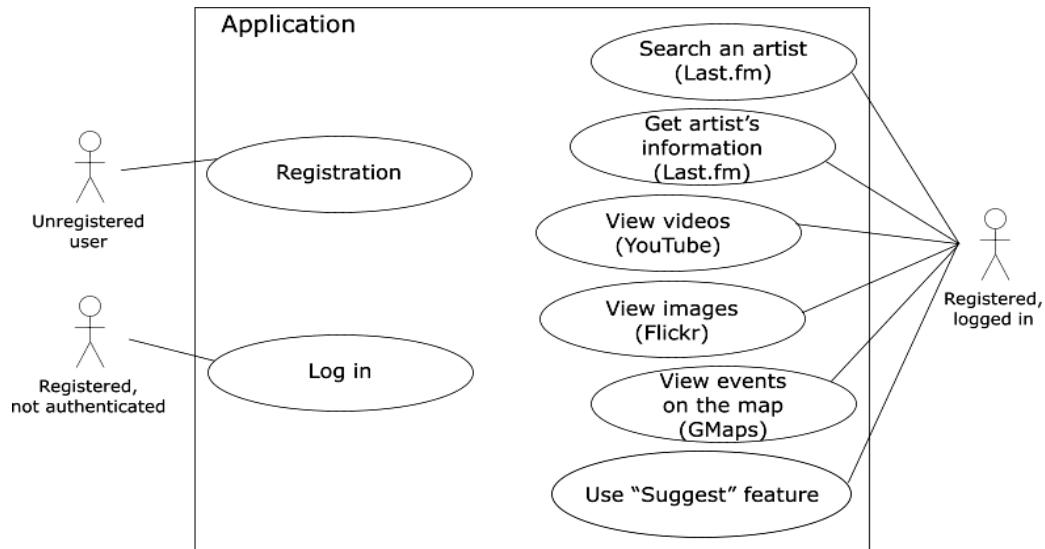
System automatically generates user profile while he interacts with system. This profile is not static, there is an adaptation mechanism that updates profile when user changes his interests – i.e. the profile reflects user interests and preferences.

Based on the user profile, system proposes a list of resources which can be interesting for the user. Besides, system displays a list of users with similar interests. System provides an ability to add these users as friends for subsequent communication.

**Technical approach.** The Figure 1 describes the architecture of the system. Use case diagram is shown on the Figure 2.



**Figure 1. Architecture of the system**



**Figure 2. Use case diagram**

*Defining user preferences.* The profile of each user contains such data sets:

$\langle \text{tagi} \rangle \langle \text{counti} \rangle \langle \text{timei} \rangle \langle \text{indexi} \rangle$ ,

where  $\langle \text{tagi} \rangle$  – is resource tag, the user works with. It is added to profile if there is not there yet, so the every tag in profile is unique,

$\langle \text{counti} \rangle$  – is simply counter for the every tag. When the user works (views or searches for definite resource) with appropriate resource, the counter is increased,

$\langle \text{timei} \rangle$  – is time in seconds, the user spent at the resource with given tag,

$\langle \text{indexi} \rangle$  – is integral value of the user interest to appropriate tag. It is estimated by the expression:

$\langle \text{indexi} \rangle = (\langle \text{counti} \rangle / \text{total\_count}) * (\langle \text{timei} \rangle / \text{total\_time})$ ,

where  $\text{total\_count} = \text{Sum}(\langle \text{counti} \rangle)$ ,

$\text{total\_time} = \text{Sum}(\langle \text{timei} \rangle)$ .

The first  $k$  sets  $\langle \text{tagi} \rangle \langle \text{counti} \rangle \langle \text{timei} \rangle$

$\langle \text{indexi} \rangle$  in user profile which have biggest values of the field  $\langle \text{indexi} \rangle$  define the user preferences. The more value takes the  $k$ , the wider circle of user interest the system takes into account. So, the  $k$  is important system parameter.

*Detection of users and artists that fall into user interests.* For the friends search it is necessary to select the fields  $\langle \text{tagi} \rangle \langle \text{indexi} \rangle$  from the other user profiles and order results by descendant. The founded topmost users are potentially could became the friends of given user.

The search of resources that fall into given user interests executes the same way. We need to select only resources with tags that have the biggest values in user profile.

In future project can be extended by such services, as:

- advising of users about the events (for example, about concerts) of artists, which most correspond to tastes of user;

- complete discography of an artist, get lyrics, using API of the lyricwiki.org service.

*Надійшла до редколегії 29.10.2009*

#### **Annotation**

The modern approach in the development of complex data-driven web applications orientated on data from different sources is described. The role of the user's profile as a component of such application is shown. Then the unobtrusive generation of user's profile is discussed. The user's profile reflects his interests. Based on presume of similarity, the method for finding resources that fall into user's interests is developed.

Описано сучасний підхід до побудови складних веб-додатків, що орієнтовані на дані з різних джерел. Показано роль профілю користувача як складової частини цих додатків. Описано новий метод генерації профілю користувача, що відображає інтереси користувача. Розроблено метод пошуку ресурсів, що відповідають інтересам користувача.

Описан современный подход к построению сложных веб-приложений, ориентированных на данные из разных источников. Показана роль профиля пользователя как составляющей части этих приложений. Описан новый метод генерации профиля пользователя, отражающего его интересы. Разработан метод поиска ресурсов, соответствующих интересам пользователя.