User Profile Analysis Using an Online Social Network Integrated Quiz Game

Yusuf YASLAN *1, Halil GÜLAÇAR1, Melih Nazif KOÇ1

1Istanbul Technical University, Faculty of Computer and Informatics Engineering, Computer Engineering Department, 34469, İstanbul

Abstract: User interest profiling is important for personalized web search, recommendation and retrieval systems. In order to develop a good personalized application one needs to have accurate representation of user profiles. Most of the personalized systems generate interest profiles from user declarations or inferred from cookies or visited web pages. But to achieve a certain result that satisfies the user needs, explicit definition of the user interests is needed. In this paper we propose to obtain interest profiles from a quiz game played by the user where at each play he/she is asked 10 questions from different categories with different difficulty levels. The developed quiz game is integrated to Facebook online social network. By doing so, we had the chance to extract each user's both explicit Facebook interest profiles and implicit interest profiles from quiz game answers. These profiles are used to extract different features for each user. Both implicit interest profile and explicit interest profile features are evaluated for clustering and interest ranking tasks separately. The experimental results show that the implicit interest profile features have promising results on personalized systems.

Çevrimiçi Sosyal Ağa Entegre Bilgi Yarışması Oyunu Kullanarak Kullanıcı Profili Analizi

Özet: Kullanıcı ilgi alanlarının profillenmesi, kişiselleştirilmiş web araması, öneri ve bulgetir sistemleri için önem taşımaktadır. İyi kişiselleştirilmiş bir uygulama geliştirmek için kullanıcı profillerinin doğru bir şekilde temsil edilmesine ihtiyaç duyulmaktadır. Kişiselleştirilmiş sistemlerin çoğu kullanıcı bildirimlerinden, çerezlerden veya ziyaret edilen web sayfalarından ilgi alanı profilleri üretmektedir. Ancak, kullanıcı ihtiyaçlarını karşılayan belirli bir sonuca ulaşmak için, kullanıcı ilgi alanlarının açıkça tanımlanması gerekmektedir. Bu çalışmada kullanıcıların ilgi alanlarını içeren profilleri, kullanıcılara farklı zorluk seviyelerinde, değişik kategorilerden 10 sorunun sorulduğu, bir bilgi yarışması oyunu ile çıkarılmıştır. Geliştirilen bilgi yarışması oyunu, Facebook çevrimiçi sosyal ağa entegre edilmiştir. Bu sayede, her kullanıcının hem Facebook sosyal ağ paylaşım sitesinde bulunan açık ilgi alanı profilleri, hem de bilgi yarışması oyunu cevaplarından elde edilen örtülü ilgi alanı profilleri çıkarılmıştır. Bu profiller her kullanıcı için farklı öznitelik çıkarımında kullanılmıştır. Hem örtülü ilgi alanı profilleri hem de açık ilgi alanı profil bilgileri kümeleme ve ilgi alanı sıralama yöntemleri ile incelenmiştir. Deneysel sonuçlar, örtülü ilgi alanı profil özelliklerinin kişiselleştirilmiş sistemlerde kullanımı için umut verici sonuçlar verdiğini göstermiştir.

1. Introduction

Personalization has gained great importance with the abundant data available on the Internet. One of the aims of the personalized applications is to give relevant information that matches a user's personal interests and provide efficient information access.

Therefore, building user profiles for long term and short term interests is crucial for personalization. A user profile, is a structured information that contains user's preferences, behavior and context [1]. Besides user's expertise, know-how can be extracted for social networking and experts for specific topic areas can be found [2, 3]. Most of the personalization systems generate user profiles either declared by users or inferred from browsing habits like visited

pages, click streams, bookmarks and past activities like queries [4, 5]. In the literature there is a vast amount of research on user profiling where most of the approaches are developed for recommendation systems. The majority of the recommendation systems developed in industry and academia employ content-based, collaborative and hybrid filtering approaches. In content-based filtering, the right information to be retrieved is selected by comparing the user profile with the contents of items. However collaborative filtering chooses the items, based on the similarities between users' profiles [6]. In other words, the rating of a user for an item is predicted based on the items that are rated by other users. However, in many problems the number of ratings for training can be very small that makes the rating data insufficient to obtain a prediction. In order to tackle with this problem one can use a hybrid system that combines content-based and collaborative filtering approaches [7]. The advent of personalization systems contributes the performance improvements not only in large businesses but also in small and medium enterprises. They have impacts on how products are sold and how companies approach to user needs [8]. On the other hand, personalized data such as location, likes and interests can be used by advertisers for targeted marketing. Most of the social networks use personalized data for these type of purposes. On the other hand, information extraction about users from their social contents is called as "social correlation" and most of the methods use natural language processing methods to extract keywords for users [9]. However, collecting this information is a difficult task for third parties and is the subject of much research. Nowadays crowdsourcing approaches are considered and Games With a Purpose (GWAP) systems are used to collect data [10]. GWAP games are simple and entertaining applications that are used to solve secondary problems such as object segmentation, protein folding, image indexing and labeling.

In this paper user interest profiles are extracted using log-files of a quiz game integrated in Facebook online social network. Facebook has become popular since people can share their thoughts, photos, videos and other liked media information with each other. Briefly Facebook can be summarized as: "Looking at, looking up or keeping up with people" [11]. The growth of the online social networks has increased the studies in broad areas made on them. Some examples cover customer loyalty [12], social actions [13], user engagements [14], psychological and sociological [15] studies. Detailed descriptions on social media users and usage are briefly given in [16].

On the other hand, there is a massive amount of people who play games that are integrated with Facebook [17]. Recently, in [10] a casual Facebook game has been used to collect data for movie matching and recommendation. Each player is tasked to match a set of movies with their friends. Actual movie likes of friends are obtained from Facebook 'like' information. The evaluation is obtained using the data generated by 27 mutual friends and friends-of-friends. Experimental results show that the crowdsourced game based recommendation is superior to collaborative and content-based filtering approaches. In [2], user interest profiles are extracted without user interaction utilizing WhoKnows? and RISQ! semantic quiz games. In order to extract the interest profiles 14 people were asked to select and order 10 interest categories from a hierarchy. 12 of the participants returned orderings back. The evaluation was obtained for those 12 people and longest common sequences of self-assessment based and quiz game based lists are obtained. The average longest common subsequence was 5.5, meaning that more than half of the orderings were correct. In [18] an automatic extraction and identification of users' responses from Facebook medical quizzes is proposed. The system crawled quiz questions that are regularly posted at The New England Journal of Medicine (NEJM) Facebook page's timeline and related answers given by the users. The proposed approach does not consider any user profiling and is capable of automatically identifying the correct and wrong answers to a quiz. Unlike the previous works, in this paper user interest profiles are obtained for 200 people selected from 2118 players who played the Facebook integrated quiz game. We have selected the users among the players who have answered questions from all categories and have Facebook profile information. The interest categories of each person is extracted explicitly from that person's Facebook profiles. In our experiments we used 7 categories namely literature, social sciences, music, sport, technology & science, politics and general knowledge. Besides, we also constructed feature vectors for each user based on the correctly answered questions and their categories, Facebook profile texts and their categories. These feature vectors are employed to cluster users. User clusters are obtained by the feature vectors extracted from quiz game and Facebook profiles separately. Recently in [19] user profile clustering patterns have been analyzed with the MCA K-Means algorithm using Facebook users' liking pattern. However, in addition to Facebook profile patterns we also consider quiz game patterns and the clustering results obtained with these two different feature sets are compared with each other with Rand index. In our experiments we employed the K-Means and Spectral Clustering algorithms. The experimental results show that Spectral Clustering performs better than K-Means clustering and the clustering obtained with the binary feature sets of questions and Facebook profile words has the higher clustering similarity. The rest of the paper is as follows: In section 2 we introduce the developed Facebook integrated quiz game platform.
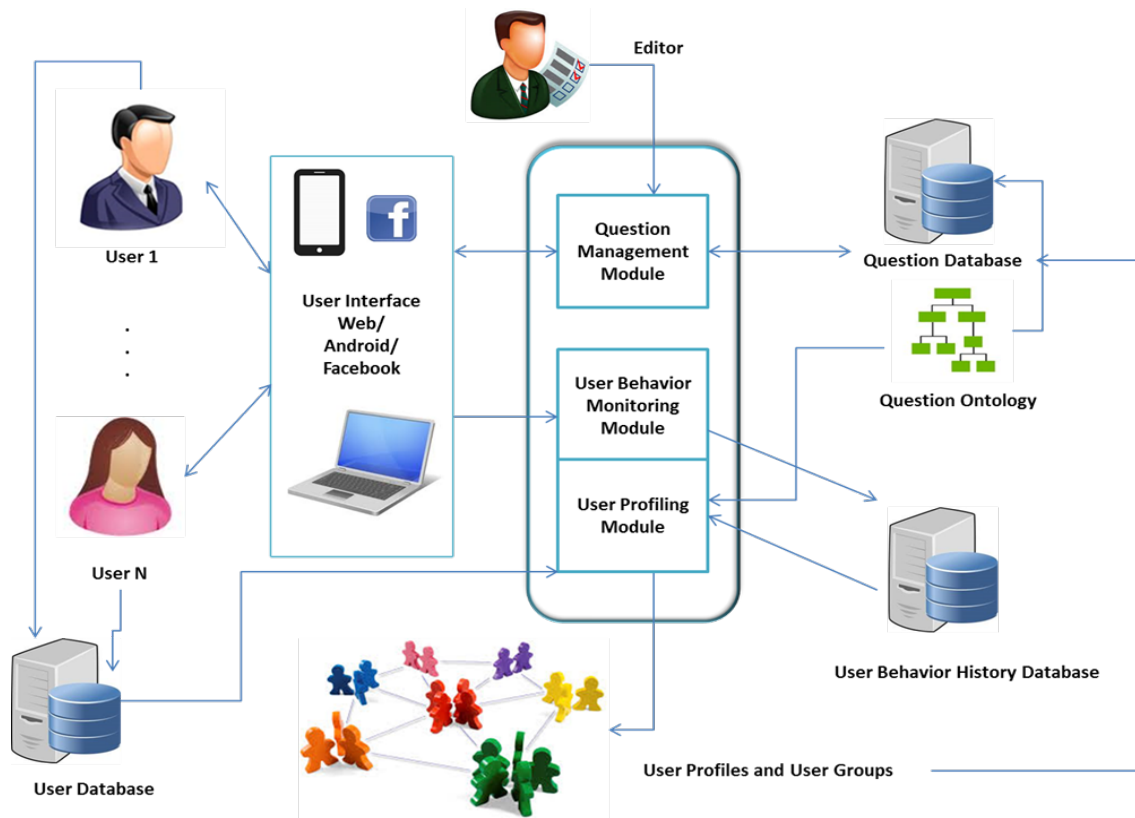
Figure 1. Developed IQ (Intelligent Questions) Quiz Game Platform.

Section 3 summarizes the collected data from Facebook and quiz game for each user. Besides the methods that are used to analyze the data are given in Section 3. Then experimental results for clustering and interest rankings are given. The last section concludes the paper.

2. Quiz Game Platform: Intelligent Questions (IQ) Game Platform

In order to extract the implicit user profiles a quiz game platform namely IQ (Intelligent Questions) platform is developed. The modules of the IQ platform are given in Figure 1 and they are as follows:

a) User Interface
b) Question Database: This database is used to store quiz questions from various categories with different difficulty levels.
c) Question Management Module: This module is used to manage the questions stored in the database. Before publishing the questions, each question is checked by an editor using this module.
d) Question Ontology: Every question has a category and subcategory information which are represented by question ontology which is used for user interest profiling.
e) User Behavior History Database: User answers to questions and their statistics are stored in this database.
f) User Behavior Monitoring Module: This module analyses the users' game play behaviors and usages.

g) User Profiling Module: User profiling is obtained by this module analyzing the user behavior history database.

The IQ Platform is implemented as a Facebook quiz game application, namely 10Soru which is developed using PHP, MySQL, HTML, CSS and JavaScript technologies. In the 10Soru game, a user is asked to answer 10 questions each of which has 4 answers from various categories with different difficulty levels. A sample question page of the game is given in Figure 2. The game has also many information pages such as; leader board and previous game statistics.

Figure 2. 10Soru Facebook Quiz Game Application.

The number of daily games for each user is limited to 5 and users can increase this number by inviting their

Facebook friends. Moreover, users can share their scores in their Facebook wall at the end of the game. For each quiz a set of 10 questions is formed from the questions that haven't been answered by that user. These 10 questions are prepared from various categories namely literature, social sciences, music, sport, technology & science, politics and general knowledge with different difficulty levels (Very Easy, Easy, Normal, Hard and Advertisement). For each user a game score is calculated using the total number of correctly answered questions, game duration and the number of used jokers. The best score of each day becomes the daily score of the user and this score determines the position of the user among all users in the daily leader board. Additionally, weekly and monthly leader boards are also shown.

For each game every user has the following 4 jokers:
• 50%: Eliminates two options which aren't correct.
• Skip: Skips the current question and asks it again after the 10th question.
• Change: Changes the question with a new one.
• Double Answer: Gives 2 answer rights to the user.

When a user uses 50% and Double Answer jokers respectively, he/she certainly answers correctly. Although it is a normal situation, this situation means that user doesn't know the correct answer of this question. Note that this score is accepted as correct during user score calculation for the leader board. But it is accepted as a wrong answer during analysis.

For each user the following explicit Facebook social network data are obtained from the users with their permissions:
Facebook Profile information (name surname, birthdate, e-mail address, gender)
• Facebook groups information (group ID and group name)
• Facebook pages information (page ID, page name and page category)
In conclusion, for each user two types of data sets, namely quiz game and social network datasets, are obtained.

3. Dataset and Methods

Facebook groups and pages give clue about users' interests and profiles. One can find the people who have similar profiles according to whether they are members of the same group or the same page. On the other hand, there may be more than one groups that belong to the same topic. In order to detect it, we separate Facebook group and page names into words. We applied stemming on the words using Zemberek (github.com/ahmetaa/zemberek-nlp) Turkish natural language processing library. Besides Facebook profiles we also obtained Quiz game answers. For each user we constructed different types of feature vectors and we analyzed them by

clustering and obtaining the longest common subsequences of the interest rankings.

3.1. Feature vectors used for clustering

For each user who plays the game we constructed a profile feature vector using the Facebook profile and quiz game answers. Initially, we evaluated these profile feature vectors in clustering task. Two different types of features are considered. The first one is Facebook profile features and the second one is quiz game profile features.

These features and their definitions are given in Table 1 and Table 2 respectively. Similarly, the dimensions of the feature vectors are given in Table 3.

Table 1. Profile feature vectors extracted from Facebook profiles.

Feature Vector	Definition
Facebook page and group IDs (pib)	If a user likes that page or join that group, then it is 1; otherwise 0.
Facebook page categories (pcb)	If a user likes any page about that category, then it is 1; otherwise 0.
Facebook page categories (pcf)	The frequency of the Facebook page category among the all liked categories.
Words obtained from Facebook pages and groups by using Zemberek (wb)	If the user profile contains that word then it is 1; otherwise 0.
Words obtained from Facebook pages and groups by using Zemberek (wf)	The frequency of the word at a user's profile.

Table 2. Profile feature vectors extracted from Quiz game.

Feature Vector	Definition
Question categories (qcb)	If the correct answers of a user for a category is more than the wrong answers, then it is 1; otherwise 0.
Question categories ratio (qccr)	The ratio between the number of correctly answered questions and the number of asked questions for a specific category.
Question IDs (qb)	If a question is asked for multiple times and the user's correct answers are more than the wrong answers it is 1; otherwise 0.
Question IDs ratio (qcr)	If a question is asked for multiple times the ratio between the correct answers and total answers.

3.2. User clustering using Facebook and quiz game based profile feature vectors

Using the proposed quiz game platform, we extracted both implicit quiz profile features and explicit Facebook profile features. By applying clustering on these feature spaces, our aim is to find how similar clustering results can be obtained using implicit and explicit feature spaces. If the overlap is high, this

represents that one feature set can be used as an alternative for the other feature set. For clustering analysis, we used two important and popular clustering methods, K-Means and Spectral Clustering, and compared their results using clustering Rand index.

Table 3. Dimensions of the feature vectors.

Feature Name	Dimension
pib	15247
pcb	189
pcf	189
wb	12848
wf	12848
qcb	7
qccr	7
qb	3128
qcr	3128

K-Means Clustering: K-Means clustering is a commonly used clustering algorithm that partitions a dataset into K clusters. It is an iterative algorithm where at each iteration the cluster centers are refined and based on the distance between instances and cluster centers of the partitions are updated. The algorithm iterates until a convergence criterion is met and applies to steps: (a) Calculates the distances between instances and cluster centers, then assigns each instance to the nearest cluster and (b) Updates the cluster centers.

Spectral Clustering: Spectral Clustering technique has many applications in machine learning, image, speech processing and computer vision domains. It applies eigenvalue decomposition on the Laplacian matrix that is constructed from the similarity information between instances. The clustering problem is defined as graph partitioning problem and eigenvalue decomposition is dependent to the affinity (similarity) matrix. The algorithm partitions data samples into different clusters so that samples with high similarity values reside in the same clusters. Let $X = \{x_1, x_2, ..., x_n\}$ be the data samples. An undirected graph $G = (V, E)$ can be obtained from the similarities between the data samples and the weighted adjacency matrix of the graph can be represented with $W = \{w_{ij}\}$, $i,j = 1,..,n$ [20]. The degree of a vertex v_i is defined as follows:

$$d_i = \sum_{i=1}^{n} w_{ij} \qquad (1)$$

Degree values of each vertex constitutes the diagonal degree matrix, D, as the degree d_1, d_2,...,d_n values are on the diagonal. By using the degree matrix D and affinity matrix W we can define the normalized graph Laplacian matrix, L_{sym}, as follows:

$$L_{sym} = I - D^{-1/2} W D^{-1/2} \qquad (2)$$

The spectral clustering algorithm computes the first k eigenvectors $u_1, u_2, ..., u_k$ of L_{sym} matrix. Then it constructs a U matrix containing u_i vectors as columns. Each row is normalized to 1 and corresponding instances are clustered using K-Means algorithm. A detailed description and other versions of spectral clustering algorithm can be found in [20].

3.3. User interest profile ranking

In order to extract the users' interests, we computed two different interest profiles. One from the Facebook profiles of the users and the other from the quiz game results. We assumed that if a user is interested in a topic category he/she likes many pages from that category at Facebook. Similarly, if a user can correctly answer many questions about a category at the quiz game it can be said that he/she is interested in that domain. For each user we obtained an interest value for each category. The Facebook category interest value for a specific category is the ratio between the number of pages that a person liked from that category and the total number of pages he/she liked. The quiz game category interest value is the ratio between the normalized correctly answered questions from a category and total number of asked questions from that category. The normalized correctly answered questions is the difference between the total number of correctly answered questions and the quarter number of the wrongly answered questions. Note that interest values for each category would be different from the other categories. In order to obtain a ranking between categories we need to have a comparable interest values for each category. Therefore, we applied z-score normalization for categories. After this normalization we rank the categories according to their interest values. This gives the interest profile ranking for each user. In the experimental results we compared the both interest rankings obtained from the Facebook profiles and the quiz game.

4. Experimental Results

Experimental results are first obtained for clustering. We compared different profile feature vectors that are extracted from the Facebook and the quiz game. The pairwise similarities of the clustering results are compared with Rand index. The last experiments are performed on the interest rankings.

4.1. Clustering results

Rand Index: Since we try to find whether a partitioning obtained using Facebook profile features can be also obtained using quiz game features we compared different clustering results of the feature types. In order to compare these clustering results we used Rand index as a measure of agreement. Let $X = \{x_1, x_2, ..., x_n\}$ be n data samples to be clustered and two clusterings of them are $Z = \{z_1, z_2, ..., z_{K1}\}$ and $Z' = \{z'_1, z'_2, ..., z'_{K2}\}$. Then the Rand index is defined as follows [21]:

$$c(Z, Z') = \sum_{i<j}^{n} s_{ij} \qquad (3)$$

Where $s_{ij} = 1$, if there exist k and k' such that both x_i and x_j are in both Z_k and Z'_k. Also $s_{ij} = 1$, if there exist k and k' such that x_i is in both Z_k and Z'_k, while x_j is neither in Z_k and Z'_k. For the other cases $s_{ij} = 0$. The Rand index lies between 0 and 1. If the two partitions agree with each other the Rand index is 1. The Rand index values for K-Means and Spectral Clusterings are given in Table 4 and Table 5 for k = 3 respectively. Similarly, clustering results for k = 5 are given in Table 6 and Table 7. From the tables it can be seen that the spectral clustering gives more overlapping results for both Facebook and quiz game profile features. Besides the best Rand index results are obtained using wb and qb feature sets. This means that binary representations of the Facebook words and quiz questions give similar clustering results.

Table 4. Rand Index for K-Means clustering (k=3).

Facebook/Quiz	qbc	qccr	qb	qcr
pib	0.36	0.52	0.53	**0.63**
pbc	0.55	0.49	0.49	0.46
pcf	0.43	0.49	0.52	0.55
wb	0.35	0.52	0.53	**0.63**
wf	0.42	0.51	0.52	0.58

Table 5. Rand Index for Spectral clustering (k=3).

Facebook/Quiz	qbc	qccr	qb	qcr
pib	0.37	0.35	0.77	0.41
pbc	0.38	0.35	**0.78**	0.42
pcf	0.50	0.50	0.53	0.51
wb	0.37	0.34	0.77	0.41
wf	0.41	0.38	0.75	0.44

Table 6. Rand Index for K-Means clustering (k=5).

Facebook/Quiz	qbc	qccr	qb	qcr
Pib	0.27	0.32	0.38	0.50
Pbc	**0.64**	0.62	0.58	0.51
Pcf	0.59	0.57	0.54	0.49
Wb	0.52	0.53	0.53	0.52
Wf	0.26	0.32	0.38	0.50

Table 7. Rand Index for Spectral clustering (k=5).

Facebook/Quiz	qbc	qccr	qb	qcr
pib	0.24	0.25	0.73	0.26
pbc	0.28	0.29	0.71	0.31
pcf	0.62	0.62	0.39	0.62
wb	0.22	0.24	**0.75**	0.25
wf	0.37	0.37	0.64	0.39

4.2. Ranking results

We computed two interest rankings; one from Facebook profiles and the other from quiz game answers. We accepted the Facebook interest profile as the base profile and we aimed to compute the similarities between the Facebook profiles and the interest profile obtained using the quiz game. We compared two rankings using the Longest Common Subsequence algorithm.

The Longest Common Subsequences (LCS) algorithm is a dynamic programming algorithm and it is generally used to compare strings or word lists. Additionally, some studies used LCS to detect similarities between two rankings [2]. The algorithm gives an output the most repeated subsequence. Therefore, if two rankings continue with a similar order, the output of the algorithm is near the ranking number. In this work, two rankings obtained from users' Facebook profiles and quiz game answers were used for parameters of LCS inputs. In a similar study in [2] the average LSC for 12 users was found 5.5 for 10 categories which means more than half of them were ordered correctly. After the implementation of LCS on the 10Soru database, subsequence length was found as averagely 3.74 over 7 categories. We also conducted hypothesis testing on the experimental results using Monte Carlo simulation method. For $\alpha = 0.05$, we found p = 0.001 meaning that we were successful to reject the H0, where the H0 hypothesis states that LCS are coming from two different random rankings.

5. Conclusion

Personalization systems have gained great importance with the advance of internet technologies. In this paper in order to have better personalized systems we proposed to obtain user interest profiles from a quiz game where at each play the user is asked 10 questions from different categories with different difficulty levels. We have incorporated the developed quiz game with the Facebook online social network. For each user both explicit Facebook profile and implicit quiz game profile features are extracted. These features are compared in clustering and interest profile ranking tasks. Experimental results show that the binary quiz category and Facebook word features give promising clustering results with Spectral Clustering algorithm. The last experiments are conducted on the comparisons of the interest rankings. The average LCS among seven categories is found 3.74 which is also consistent with previous results and show that more than half of the orders are correct. The experimental results performed on real dataset gave us promising results. Integrating new question ontologies and using the friendship information (i.e. collective classification) is our future work.

Acknowledgment

References

[1] Kadima, H., Malek, M. 2010. Toward ontology-based personalization of a recommender system in social network, in Proceedings of the SoCPaR, IEEE, in Trevor P. Martin; Azah Kamilah Muda; Ajith Abraham; Henri Prade; Anne Laurent; Dominique Laurent & Virginie Sans, ed., pp.119-122.

[2] Knuth, M., Ludwig, N., Wolf, L. and Sack, H. 2011. The Generation of User Interest Profiles from Semantic Quiz Games, in Proceedings of the Second International Workshop on Mining Ubiquitous and Social Environments MUSE 2011; Martin Atzmueller & Andreas Hotho, ed.; pp. 43-53.

[3] Bozzon A., Brambilla M., Ceri S., Silvestri M., Vesci G. 2013. Choosing the Right Crowd: Expert Finding in Social Networks, in Giovanna Guerrini and Norman W. Paton ed., Joint 2013 EDBT/ICDT Conferences, Genoa, Italy, March 18-22.

[4] Limam L., Coquil D., Kosch H., Brunie L. 2010. Extracting User Interests from Search Query Logs: A Clustering Approach, in Proceedings of the 21st International Workshop on Database and Expert Systems Applications (DEXA 2010), A Min Tjoa and Roland Wagner ed., pp. 5-9.

[5] White W. R., Bailey P., Chen L. 2009. Predicting User Interests from Contextual Information. in Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval James Allan and Javed A. Aslam and Mark Sanderson and ChengXiang Zhai and Justin Zobel ed.; pp. 363-370.

[6] Trajkova, J., Gauch, S. 2004. Improving Ontology-Based User Profiles, in Proceedings of the RIAO, CID, Christian Fluhr; Gregory Grefenstette & W. Bruce Croft, ed., pp. 380-390.

[7] Wu, M.-L., Chang, C.-H., Liu, R.-Z. 2014. Integrating content-based filtering with collaborative filtering using co-clustering with augmented matrices. Expert Systems with Applications, 41, pp. 2754-2761.

[8] Parra-Arnau, J., Rebollo-Monedero, D., Forné, J. 2014. Measuring the privacy of user profiles in personalized information systems. Future Generation Computer Systems, 33, pp. 53-63.

[9] Ko H. G., Ko I. Y., Kim T., Lee D. 2013 Hyun S. J. Identifying User Interests from Online Social Networks by Using Semantic Clusters Generated from Linked Data, in Proceedings of the ICWE 2013 Workshops, LNCS, Quan Z. Sheng and Jesper Kjeldskov ed.; pp. 302-309.

[10] Banks, S., Rafter, R., Smyth, B. 2015. The Recommendation Game: Using a Game-with-a-Purpose to Generate Recommendation Data, in Hannes Werthner, Markus Zanker, Jennifer Golbeck & Giovanni Semeraro, ed., 'RecSys', ACM, pp. 305-308.

[11] Joinson A. N. 2008. Looking at, Looking up or Keeping up with People?: Motives and Use of Facebook, in Proceedings of the Conference on Human Factors in Computing Systems (CHI) 2008, Mary Czerwinski and Arnold M. Lund and Desney S. Tan ed. Florence, Italy.

[12] Baird C.H., Parasnis G. 2011. From social media to social customer relationship management. Strategy and Leadership, 39, pp. 30-37.

[13] Cheung C.M., Lee M.K. 2010. A theoretical model of intentional social action in online social networks. Decision Support Systems, 49, pp. 24-30.

[14] Khobzi, H., Teimourpour, B. 2015. LCP Segmentation: A framework for Evaluation of User Engagement in Online Social Networks. Computers in Human Behavior, 50, pp. 101-107.

[15] Ortigosa A., Carro R. M., Quiroga J. I. 2014. Predicting User Personality by Mining Social Interactions in Facebook. Journal of Computer and System Sciences, 80, pp. 57-71.

[16] Weller K. 2016. Trying to understand social media users and usage. Online Information Review, 40, pp. 256 – 264.

[17] Hicks K., Gerling K., Kirman B., Linehan C. 2015. Dickinson P. Exploring Twitter as a Game Platform; Strategies and Opportunities for Microblogging-based Games, in Proceedings of The ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play (CHI PLAY).

[18] Rodriguez-Gonzalez A., Ruiz E. M., Pujadas M. A. M. 2016. Automatic Extraction and Identification of Users' Responses in Facebook Medical Quizzes. Computer Methods and Programs in Biomedicine, 127, pp. 197-203.

[19] Van Dam J. W., Van de Velden M. 2015. Online Profiling and Clustering of Facebook Users. Decision Support Systems, 70, pp. 60-72.

[20] Luxburg U. 2007. A Tutorial on Spectral Clustering. Statistics and Computing, pp. 395-416.

[21] Rand W.M. 1971. Objective Criteria for the Evaluation of Clustering Methods. Journal of the American Statistical Association, pp. 846-850.