

Research on Fuzzy Recommendation System Based on User Profile

Jingfeng Shao, Xiaoyu Yang, Chuangtao Ma

Xi'an Polytechnic University, Xi'an, China

Email: shawnxpu@163.com

How to cite this paper: Shao, J.F., Yang, X.Y. and Ma, C.T. (2017) Research on Fuzzy Recommendation System Based on User Profile. *Open Journal of Social Sciences*, 5, 1-11.
<https://doi.org/10.4236/jss.2017.510001>

Received: August 5, 2017

Accepted: October 10, 2017

Published: October 13, 2017

Abstract

To provide the personalized services for learners according to user's learning behavior, firstly, through theory of Fuzzy Ontology, the ambiguity and uncertainty of information in E-learning recommendation system was analyzed, and fuzzy information was extended to the ontology. Secondly, the user interest estimation based on behavior was studied in term of user's learning preferences and cognitive status, user profile was described by the learning object, and then a user model updating algorithm was proposed. Finally, a fuzzy recommendation system based on user profile was built. As verified by experiment, the results have shown that the system can provided the personalized learning services according to user's learning behavior, and it is useful in describing user preferences.

Keywords

Recommendation System, User Profile, Interest Degree, Fuzzy Ontology

1. Introduction

With the rapid development of educational information technology, network technology and the various teaching tools were widely used in the teaching process. Under the internet background, E-learning has become an important method of education and teaching, because it took full advantage of the new communication mechanism and abundant learning resources, and realized a new learning way, the goal is to provide modern information technology for user. So, E-learning is an online learning process that the learning and teaching activities is mainly completed through the network [1]. To be compared with traditional teaching methods, E-learning can make user learn much more modern knowledge via internet, and can provide a new learning and teaching model for

teachers and students. Hence, in the education internationalization environment, for each learner, how to use E-learning recommendation system to exchange information with active learners, to analyze individual needs, to provide interest content and so on, all these has become a hot research issue.

User profile is a description file of user's interest, and also is the data and structure of user's preferences [2]. In the meanwhile, *User profile* can reflect the user's preference information, and can be used to store the information of the semantic web. So, its aim is to express user preferences and to promote information storage and retrieval process. Contrarily, the semantic web used the formalized ontology to create machine-readable data, while the ontology is useful in dealing with machine-identifiable descriptive information. However, when, the formal concept of ontology in the general domain is incompetent, uncertain and ambiguous. So, to describe the fuzzy concept and fuzzy relation of imprecise, uncertain and fuzzy information, fuzzy theory was extended to the ontology [3], and then made it be formed fuzzy ontology, because fuzzy ontology can deal with fuzzy knowledge, and can describe and retrieve text and multimedia objects in the fuzzy information.

To this end, a fuzzy ontology based on *user profile* was introduced, and a method to describe *user profile* was proposed. Furthermore, domain ontology was expanded, and user preferences, interests and hobbies were described, and the teaching activities in E-learning recommendation system were elaborated. For learning object, they can use fuzzy logic to define, implement, verify and improve the performance of retrieval, classification and management, and can construct fuzzy ontology process based on *user profile*, and then can apply *user profile* into the recommendation system[4].

2. Construction of Fuzzy Ontology

Ontology is a clear formal specification of the shared conceptual models. The process of ontology construction was believed as the relationship among concepts. In the real world, the ambiguity and uncertainty of information was expressed the fuzzy knowledge of the fuzzy ontology. This meaning was that fuzzy knowledge has two advantages, one was to improve the concept of formal description via the ontology, the other was to help describe and deal with fuzzy information in many application fields. Thus, the fuzzy ontology can extract some useful information from the data source, and this extraction process can automatically construct without human participation. For example, Makokha [5] proposed a conceptual and web database based on the intelligent decision analysis system model, and used the model to accurately retrieve the relevant information. Of course, this model used the conceptual latent semantic indexing technique based on the term similarity function, and automatically constructed the ontology from the text document set. Betty [6] described a method that used fuzzy ontology to improve the retrieval performance of semantic documents, and then introduced an information retrieval algorithm based on fuzzy concept network.

3. Fuzzy Recommendation System Based on User Profile

According to the definition of Anaya [7], the recommended system was system which can output the personalized recommendation content, and can provide the interesting, personalized and useful object in the optional space for users. So, the components of *user profile* based on fuzzy ontology and the automatic construction process was introduced, of which *user profile* was a description file of the user's interest, was an embodiment of the user's personalized needs, and also was a relevant learning object composition [8].

3.1. User Interest Estimation Based on Behavior

The representation of the user model determines the model that reflects the ability and computability of the real information, and also limits the selection of the modeling method and the model updating algorithm in a certain extent. The hierarchical vector space model uses the tree structure to classify the interest topics and makes use of the vector space model to represent the user's interest. So, a three-layer tree structure to represent the user interest model was proposed, because a topic can have multiple topic features. Of which, the first layer node represented the user, the second layer node represented the user's interest topic, and the third layer node represented the characteristics values of user's interests' topic. Thus, the specify model representation method was shown in **Figure 1**.

If the user has m different interest topics, then the user interest model can be expressed as the following eigenvector.

$$Model = \{(V_1, W_1, n_1), (V_2, W_2, n_2), \dots, (V_m, W_m, n_m)\}$$

where V_i is the i th topic feature vector, W_i is the subject weight, and n_i is the number of document instances of the i th topics ($1 \leq i \leq m$). Now, W_i was initialized, and $W_i = U(page_1)U(page_2)\dots U(page_m)$, of which $U(page_k)$ ($k = 1, 2, \dots, m$)

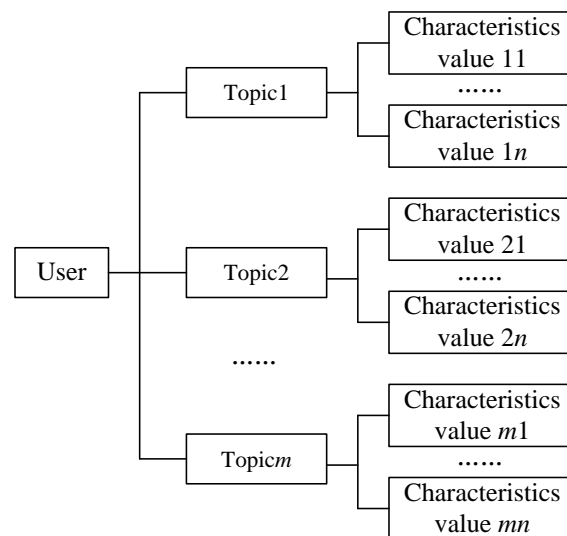


Figure 1. User profile based on vector space model.

represented the page interest $page_k$ of the user's interest. While the document instance abstraction key characteristics in topic S_i was denoted as $\{(k_{i1}, p_{i1}), (k_{i2}, p_{i2}), \dots, (k_{im}, p_{im})\}$, of which (k_{ij}, p_{ij}) was the j th interest key word of in S_i class, k_{ij} was the key word, and p_{ij} was the weight of the keyword k_{ij} .

In general, the user has five typical browsing behaviors, which included saving page, printing page, *Bookmark*, visiting times and staying time. These five behaviors are the best combination to estimate the user's interest through the correlation analysis.

Let $U(p)$ represent the user's interest in the page p , and then $U(p)$ can be expressed as follows.

$$U(p) = (S(p), P(p), B(p), F(p), T(p)) \tag{1}$$

where $S(p)$ was the saving page, $P(p)$ was the printing page, $B(p)$ was *Bookmark*, $F(p)$ was the page visiting frequency, and $T(p)$ was the staying time on the page. For a page, the user had three visiting behaviors that are saving page, printing page and *bookmark*, one of behaviors expressed that the user had a high interest degree in the page. So, the interest degree was defined as follows.

$$U(p) = \begin{cases} 1 & S(p), P(p), B(p) \text{ was happened at least one} \\ f(F(p), T(p)) & \text{otherwise} \end{cases} \tag{2}$$

We can see in Formula (2), the higher the user visited the frequency of the page, the slower the page was scanned, and the more the user was interested the page. This result explained that frequency, browsing speed and interest degree had the inter-relationship, and can be denoted as follows.

$$f(F(p), T(p)) = \alpha * \frac{F(p)}{\max_{v \in P}\{F(v)\}} + (1 - \alpha) * \frac{T(p) / Size(p)}{\max_{v \in P}\{T(v) / Size(v)\}}, \alpha \in [0, 1] \tag{3}$$

where p is the page and $Size(p)$ is the size of the page p . To simplify the experimental process, we made $\alpha = 0$, and the calculation of interest degree mainly considered the size of the text, which was changed from the access page to a text instance. To this end, the inter-relationship can be simplified as follows.

$$f(F(p), T(p)) = \frac{1 / Size(p)}{\max_{v \in P}\{1 / Size(v)\}} \tag{4}$$

3.2. User Model Updating Algorithm

In the *Web* system, the interest learning is an indispensable part of the knowledge acquisition link, because it provided the personalized service for the user. To describe the user's personal interest, the initial user model based on the user registration information or the information was established. Hence, once the current interest of the user was motivated, the dynamic description of user interest was adaptively modified. For the user, they can modify the current model, and in the meanwhile the current interest model can be changed with the user's interest. In this way, an updating mechanism was taken place, and the updating

algorithm based on *User Profile* was shown as follows.

Input: $\{ \langle e_1, U_1 \rangle, \langle e_2, U_2 \rangle, \dots, \langle e_m, U_m \rangle \}$ // m is the window size, feature similarity threshold $s \in (0,1)$ and subject number threshold K ;

Output: Updated interest model

Function: $Size(model)$ is the number of topics in the current model

Steps:

- 1) for each e_j in $\{ \langle e_1, U_1 \rangle, \langle e_2, U_2 \rangle, \dots, \langle e_m, U_m \rangle \}$ do
- 2) Let $MaxSim = \max_k \{ Sim(V_k, e_j) \}$ // Calculate the maximum similarity
- 3) if $MaxSim < s$ and $Size(model) < K$ then
- 4) Create new subject $S_k = \langle V_k, P_k \rangle$ with $V_k = e_j, P_k = U_j$
- 5) else for $\forall m \ Sim(V_m, e_j) > s$ do
- 6) $V_m^{new} = \gamma(V_m^{old}, \langle e_j, U \rangle)$
- 7) $W_m^{new} = \varphi(W_m^{old}, \langle e_j, U \rangle)$
- 8) end for
- 9) end if
- 10) end for

In the algorithm, if the learning document topic was changed very broadly, then the high threshold s would lead to the unlimited growth of interest topic. To overcome this phenomenon, we take use of the threshold K to limit the number of topic in the model. If the number of topic reached K , a K -nearest neighbor algorithm was triggered, and it can update the features and weights of the topic, which the similarity degree was greater than s . Furthermore, the parameter s and K were similar to the *MTDR* algorithm proposed in [5], $s = 0.207$ was obtained by the initial experiment, and K was determined in term of the requirement of the individual service system. In this paper, the user model updating algorithm took full account of the topic and its weight updating, we made $K = 10$, and used the user feedback in the fixed window to update the system. For the topic characteristics vector, each topic instance can get, initialize and update via the weighted average method, which were shown as follows.

$$V_i^{new} = \gamma(V_i^{old}, \langle e, U \rangle) = \frac{\sum_{e_j \in E_j} U(e_j) * e_j}{\sum_{e_j \in E_j} U(e_j)} \tag{5}$$

For the corresponding weight of the topic, the user feedback $\langle e_j, U_j \rangle$ was given and the updating process was shown as follows.

$$\begin{aligned} W_i^{new} &= \varphi(W_i^{old}, \langle e_j, U_j \rangle) = W_i^{old} * U_j \\ &= U(\omega_1)U(\omega_2) \dots U(\omega_{n_i}) * U_j \end{aligned} \tag{6}$$

$\omega_j (j = 1, 2, \dots, n_i)$ was denoted as the page.

3.3. Fuzzy User Profile

In the context of E-learning, *user profile* was described by the learning object. In the meanwhile, the user's preference was described by two elements of *user profile*, and usually, they were expressed as a set $U(\xi, P)$ [9]. Here ξ was the

concept set that described the user's preference, P was the fuzzy ontology of the concept set, and was corresponded to preference. The set of ξ was a set of concepts c_i , ω_i was the weight of a concept, and its goal was to describe the user's preference. As shown in Equation (7).

$$\xi^u = \{(c_1^u, \omega_1^u), (c_2^u, \omega_2^u), \dots, (c_n^u, \omega_n^u)\} \tag{7}$$

This method was quantified the weight of each concept in the set of learning objects, which included the following two step.

Step 1. The weight of each concept c_i in the learning object d was a fuzzy value. Now, according to FIS-CRM18 membrane model (fis-crm (c_i, d)), the basic weight value of each concept was calculated, and then can re-adjusted the weight values based on the synonyms and frequency in the user's document.

Step 2. If a concept was more frequent than other concepts in the user's documentation, it was considered that the concept was highly relevant with *user profile*. Of course, there existed some exceptions. To obtain this correlation, the weight of each concept in *user profile* must consider, and the frequency in the learning objects. So, the weight of the concept c_i in user profile u can be calculated by the Formula (8).

$$\omega_i^u = \sum_{j \in u} w_{ij} * (1 + \frac{docs(c_i, u)}{|T|}) * Ln(\frac{|U|}{U(c_i)} + 1) \tag{8}$$

In the above formula, ω_j was the relevance degree between the concept c_i and learning object d_j , docs (c_i, u) was the document number of the concept c_i in *user profile*, $|T|$ was the total number of learning objects in *user profile*, $|U|$ was the total number of *user profile* in the E-learning environment [10], and $U(c_i)$ was the number of *user profile*, and indicated that the concept c_i had the determining membership.

After the weight of the concept was calculated, the correlation distribution between the concepts can be determined. Thus, the normalized weight would be distributed in the interval [0,1], and only the concept had an enough large weight ($\omega > 0.5$), because it is useful for improving the learning activities in the E-learning environment [11].

The fuzzy ontology was a vector network, and it took the concept from the learning object. Thus, a fuzzy ontology can be seen as a directed graph set [12]. Furthermore, each point represented a concept, and an edge represented the relation between concepts.

4. Experiment and Assessment

The recommendation system can provide some useful and interesting learning object for users. The main task in assessing process was to use the learning objects stored in the learning object library to construct *user profile* and to evaluate it.

4.1. Experiment

The experiment was carried out in the Metadata built by us for Learning, which supported open access in the form of domain ontology to obtain learning re-

sources, and especially for teaching resources [13]. The basic components of Metadata were able to share and reuse objects, which included the storing resource indexes and knowledge base for metadata, and were able to support the transparent retrieval on a heterogeneous repository network, which included the integrating search engine, integrating other resources, searching for network users and restoring visualization detector and so on. Thus, the user interest degree was shown in **Table 1**.

As the learning object, all users were the contributor of the learning object, and users in the experiment must upload more than 10 English-based *pdf* files to the database, because the number of documents was so much as to establish a good *user profile*. For example, 10 users used simple query interface (SQI) to collect about 2400 learning objects, of which SQI provided the standardized communication among users, and be able to realize the joint queries. Thus, the selected learning object can extract the text content from a *pdf* document, the learning object can generate a set of user documents, and the result was shown in **Figure 2**. In the meanwhile, this document set was divided into two parts, which included the production of user features and the test of recommended system.

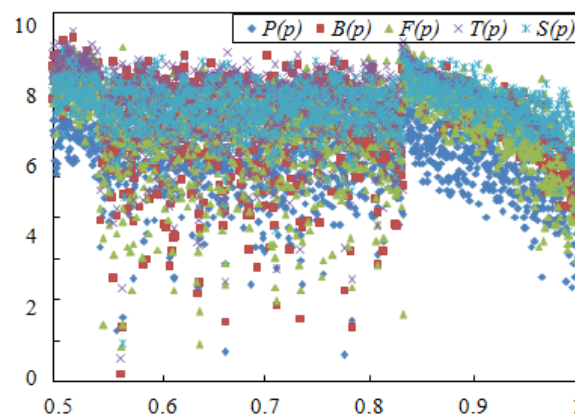


Figure 2. The text content from 2400 learning objects.

Table 1. The user interest degree.

$U(p)$	$S(p)$	$P(p)$	$B(p)$	$F(p)$	$T(p)$
User 01	0.17	1.00	0.50	0.17	0.50
User 02	0.20	0.50	0.50	1.00	0.17
User 03	0.33	0.20	0.25	1.00	0.50
User 04	0.67	0.50	0.67	0.40	0.40
User 05	0.50	0.33	0.17	0.33	0.50
User 06	1.00	0.14	1.00	1.00	0.50
User 07	0.50	0.25	0.33	0.20	0.50
User 08	0.33	0.25	0.14	1.00	1.00
User 09	0.33	0.33	0.50	0.33	0.33
User 10	1.00	0.25	0.33	1.00	0.50

4.2. Ontology Assessment

Using the above method, the term set and the fuzzy ontology of each user was got, these terms set and the ontology set described the user characteristics, and were described the basic semantics of the learning object. Thus, through being compared to produce all the relationships of ontology, these relationships included four semantic relations, and were extracted from ours database, which included synonyms, synonyms, subtitles and upper words. After that, a structured analogy and a similarity relation defined by the conceptual network was used, and *Precision* and *Recall* was evaluated [14]. Here, *Precision* was a selected element correlation probability, *Recall* represented the probability that a related element was selected, and *F test* was the harmonic average between *Precision* and *Recall*. Thus, they were calculated by Formulas (9)-(11) to determine the performance of the system.

$$Precision = (\text{number of retrieved relevant}) / (\text{number of retrieved irrelevant}) \quad (9)$$

$$Recall = (\text{number of retrieved relevant}) / (\text{number of retrieved irrelevant}) \quad (10)$$

$$F = (2 * Precision * Recall) / (Precision + Recall) \quad (11)$$

In this paper, the ontology evaluation standard was adopted, and the experiment was used to verify the *Precision*, *Recall* and *F test* proposed in the recommended system (**Table 2**).

10 users were selected in the test, 620 learning objects with different fields were collected by ours database. The data set was divided into training set and test set. 10 *user profile* was established by 512 learning objects of the training set. The test set contained the remaining 220 learning objects.

Compared the recommended content from the system with the extracted content from metadata (if the user is a contributor of the learning object, this learning object is recommended to the user), the values of *Precision*, *Recall*, and *F test* were obtained. And then, the experimental data were shown in **Table 3**, and the experimental results were shown in **Table 4**.

In **Table 4**, the average values of *Precision*, *Recall*, and *F test* is 86.6%, 83.2% and 88.0%, respectively. These data results also represented that the proposed system had a good data process performance, and the whole data process was shown in **Figure 3**.

We can see from **Figure 3**, the system is flexible and can automatically generate good recommended content based on *user profile*.

Table 2. The meaning of the various symbols.

	Selected	Not selected	Total
Relevant	<i>rt</i>	<i>Nrt</i>	<i>Nr</i>
Irrelevant	<i>it</i>	<i>Nit</i>	<i>Nt</i>
Total	<i>tl</i>	<i>Ntl</i>	<i>N</i>

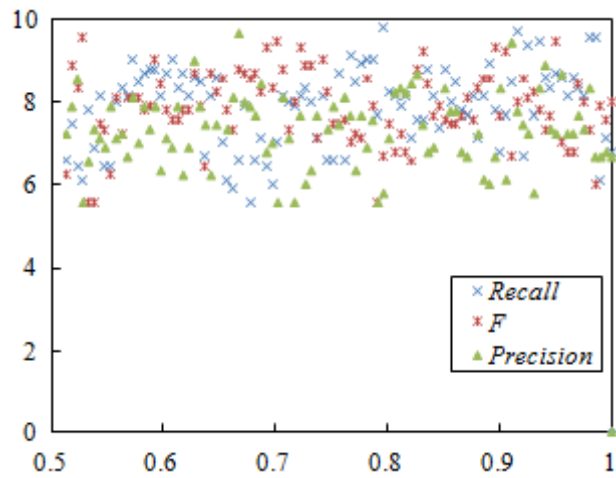


Figure 3. Data process for fuzzy recommendation system.

Table 3. The experimental data.

	<i>rt</i>	<i>Nrt</i>	<i>it</i>	<i>Nr</i>	<i>tl</i>
User 01	20	5	6	20	23
User 02	18	5	6	19	14
User 03	14	0	1	12	17
User 04	22	4	3	23	15
User 05	15	0	2	16	18
User 06	10	0	5	17	12
User 07	11	2	2	11	15
User 08	14	0	1	16	15
User 09	15	6	7	14	18
User 10	6	2	8	6	10

Table 4. The experiment results.

	<i>Precision</i>	<i>Recall</i>	<i>F</i>
User 01	0.78	0.95	0.99
User 02	0.84	0.91	0.85
User 03	1.00	0.72	0.86
User 04	0.76	1.00	0.99
User 05	0.98	0.88	0.94
User 06	0.98	1.00	0.98
User 07	0.87	0.79	0.86
User 08	1.00	0.85	0.89
User 09	0.57	0.65	0.78
User 10	0.88	0.57	0.66
Average	0.866	0.832	0.880

5. Conclusion

The flexibility of fuzzy ontology supported a wide range of information retrieval and filtering problems. A method to describe *user profile* using fuzzy ontology was proposed, and this method can effectively improve the learning behavior of E-learning, and use the recommendation system to provide the interested learning object for users. *User profile* combined a related concept set to define a fuzzy ontology that comprehensively described the user's preferences, and introduced a user-selected or created learning object by generated *user profile*. The experiment was realized on ours database platform, and the results have showed that the proposed system was reasonable and valid in the terms of *Precision* and *Recall*. Further, the future task is to improve the quality of *user profile*, the aim is to use pruning process to avoid the irrelevant concepts to disturb users, to consider the feedback information from the user, and even to use some hybrid screening technology to add more detailed correlation experiments.

Acknowledgments

Supported by Shaanxi Province Education and Teaching Reform Project in China (No.15BY67), Education and Teaching Reform Project of China Textile Industry Federation (No.2017BKJGLX158), Shaanxi Province Education Science "13th Five-Year" Planning Project in China (No.SGH16H061), and Shaanxi Province Science and Technology Project in China (No.2017GY-039).

References

- [1] Su, Y.R. and Xue, C.B. (2014) Explore the Development and Practical Application of English Teaching. *International Journal of Technology Management*, No. 3, 1-3.
- [2] Nan, C.P. and Guo, X. (2014) Analysis of the Application of a Happy Learning Method in College English Classroom Teaching. *International Journal of Technology Management*, No. 4, 11-13.
- [3] Arinto, P.B. (2016) Issues and Challenges in Open and Distance E-Learning: Perspectives from the Philippines. *International Review of Research in Open and Distributed Learning*, **17**, 23-27. <https://doi.org/10.19173/irrodl.v17i2.1913>
- [4] Zhao, K. and Chan, C.K.K. (2014) Fostering Collective and Individual Learning through Knowledge Building. *International Journal of Computer-Supported Collaborative Learning*, **9**, 63-95. <https://doi.org/10.1007/s11412-013-9188-x>
- [5] Makokha, G.L. and Mutisya, D.N. (2016) Status of E-Learning in Public Universities in Kenya. *International Review of Research in Open and Distributed Learning*, **17**, 40-48. <https://doi.org/10.19173/irrodl.v17i3.2235>
- [6] Betty, M.R. (2014) Using a Social Networking Tool for Blended Learning in Staff Training: Sharing Experience from Practice. *Journal of Neonatal Nursing*, **20**, 90-94. <https://doi.org/10.1016/j.jnn.2014.03.005>
- [7] Anaya, A.R., Luque, M. and Garcia-Saiz, T. (2013) Recommender System in Collaborative Learning Environment an Influence Diagram. *Expert Systems with Applications*, **40**, 7193-7202. <https://doi.org/10.1016/j.eswa.2013.07.030>
- [8] Diaby, M., Viennet, E. and Launay, T. (2014) Exploration of Methodologies to Improve Job Recommender Systems on Social Networks. *Social Network Analysis and*

Mining, **4**, 1-17. <https://doi.org/10.1007/s13278-014-0227-z>

- [9] Tsai, W.H., Lin, Y.T. and Lee, K.R. (2016) Development of Social-Aware Recommendation System Using Public Preference Mining and Social Influence Analysis: A Case Study of Landscape Recommendation. *Journal of Internet Technology*, **17**, 561-569.
- [10] Malik, M.W. (2010) Factor Effecting Learner's Satisfaction Towards E-Learning: A Conceptual Framework. *OIDA International Journal of Sustainable Development*, **2**, 77-82.
- [11] Simon, R.J.M. and Mucklow, J. (2012) E-Learning Initiatives to Support Pre-Scribing. *British Journal of Clinical Pharmacology*, **74**, 621-631. <https://doi.org/10.1111/j.1365-2125.2012.04300.x>
- [12] Johnson, L. and Adams, B.S. (2016) NMC Horizon Report: 2016 Higher Education Edition. The New Media Consortium, Austin.
- [13] Ana, I., Molina, M.A., Redondo, C.L., *et al.* (2014) Assessing the Effectiveness of New Devices for Accessing Learning Materials: An Empirical Analysis Based on Eye Tracking and Learner Subjective Perception. *Computers in Human Behavior*, No. 31, 475-490. <https://doi.org/10.1016/j.chb.2013.04.022>
- [14] Soussan, D., Marisa, S., Tom, T. and Generation, Y. (2014) Web Design and Eye Tracking. *International Journal of Human-Computer Studies*, No. 5, 307-323.