Application of the adaptive content concept for an e-learning resource

Dmitry A. Aldunin
Doctoral Student, Department of Innovation and Business in Information Technologies
National Research University Higher School of Economics
Address: 20, Myasnitskaya Street, Moscow, 101000, Russian Federation
E-mail: daldunin@hse.ru

Abstract

Nowadays, the e-learning market is rapidly growing both financially and geographically. More and more often, e-learning resources involve a multicultural audience and are becoming available to people with different educational backgrounds. However, there are cognitive specificity and different approaches to the learning process in different cultures. This paper is devoted to illustrating a possible solution for adaptation of content of an e-learning resource to a multicultural audience. The solution described applies the adaptive content concept based on individual educational trajectories and preparing content according to the individual cultural characteristics of learner and his or her competencies, both obtained and desired.

During the research, the learner-centric model of learning processes was developed. In the article, both high-level and detailed models are presented. Principles of planning the individual learning trajectory based on the learner's obtained and desired competencies, and statistical data about his or her learning style are also described. As an example of the possibility to apply historical data on how learning style affects successful passing through the learning course, the statistical analysis is provided. The analysis relies on person-course de-identified dataset from seven courses on HarvardX and MITx platforms provided during the 2013/14 academic year. This analysis demonstrated the statistical significance of several parameters. A comparison of algorithms for estimating the probability of successfully passing the course depending on the learning style, is also presented.

Key words: e-learning, cross-cultural context, mathematical modeling, web-development, individual learning trajectory, learning analytics.

Citation: Aldunin D.A. (2016) Application of the adaptive content concept for an e-learning resource. *Business Informatics,* no. 4 (38), pp. 27–34. DOI: 10.17323/1998-0663.2016.4.27.34.

Introduction

Nowadays, the cross-cultural problem is one of the basic things to consider when creating an e-resource focused on representatives of different cultures. Since the advent of using information technology for delivery of information, new tools and methods to present it have appeared. Some of them both partly decrease and increase the psychological tension of the process of acquiring knowledge. According to G. Uzilevsky [1], ergonomic semiotics is a scientific practice which studies the problems common to semiotics, linguistics and ergonomics, and is able to resolve a number of problems associated with the need to make information intelligible to the target audience, depending on the cognitive specificity caused by cultural differences. In the era of the worldwide Internet's intromission into most areas of life, the implementation of cross-cultural aspects of ergonomic semiotics is very important.

On the one hand, the content of usual e-resources certainly embodies the national culture of its creators [2]. On the other hand, content must meet the cultural and pragmatic expectations of the learner for maximal efficiency of presenting information. For example, it is possible to observe that Western educational content is characterized by openness, clear logic and predictability, its dosage of information and no hidden content; while Eastern content usually has a complex structure and is very context-related [2].

Pragmatic factors must necessarily be taken into account during the process of creating and monitoring resources aimed at a multicultural audience. There are different cognitive models of perception and processing of information in the Eastern (collectivistic and inclined to dialectic) cultures and the Western (individualistic and inclined to formal logic) cultures. In the West, the information is mostly directly perceived through the prism of the person's individual perception and is superimposed on the existing information background. In the East, users simply collect information without personal critical reflection when consuming information content [2].

1. Basis of cultural classification

The central object of my adaptive content concept is the learner's profile consisting of two parts: cultural and cognitive, plus competency profiles. The cultural and cognitive profile contains information about culture-related features: cognitive specificity, specificity of train of thought, and affective differences. According to M. Kholodnaya's research [3], the cognitive style reflects the way of perceiving, analyzing, structuring and categorizing the world, the style of learning. The competency profile contains a list of the learner's knowledge and skills, and the vector of his or her educational and professional development (what he or she is interested in for further learning).

It makes sense to base the cultural and cognitive profile on the six cultural dimensions invented by Geert Hofstede [4, 5] (Power Distance (PDI), Collectivism vs Individualism (IND), Masculinity vs Femininity (MAS), Uncertainty Avoidance (UAV), Long-term vs. Short-term orientation (LTO), Indulgence versus Restraint (IVR)) because there a test exists which provides an exact value for each dimension, making it very convenient for programming. This basis was used when mapping with the cultural classifications of F. Trompenaars [6], H. Triandis [7] and R. Lewis [8]. While Hofstede's classification was driven by an individual's associates, other classifications are more individual and thus complement the former. The profile may also be extended with the binary classification of learning styles introduced by R. Felder and L. Silverman [9]. In this model, learning styles are described as a value between pairs of extremes: active or reflective, sensing or intuitive, verbal or visual, and sequential or global. There also is the test for getting values for a specific learner [10].

The competency profile should be based on a competency classificatory. Competencies should be divided into systematic competencies (core competencies, such as time-management, statistics analyzing, working in group, etc.) and professional competencies (related to the exact professional field). Professional competencies should themselves be divided in two: instrumental competencies, and social and personal competencies. The latter competency group is related to the cultural and cognitive profile and that is how connection between profiles is carried out.

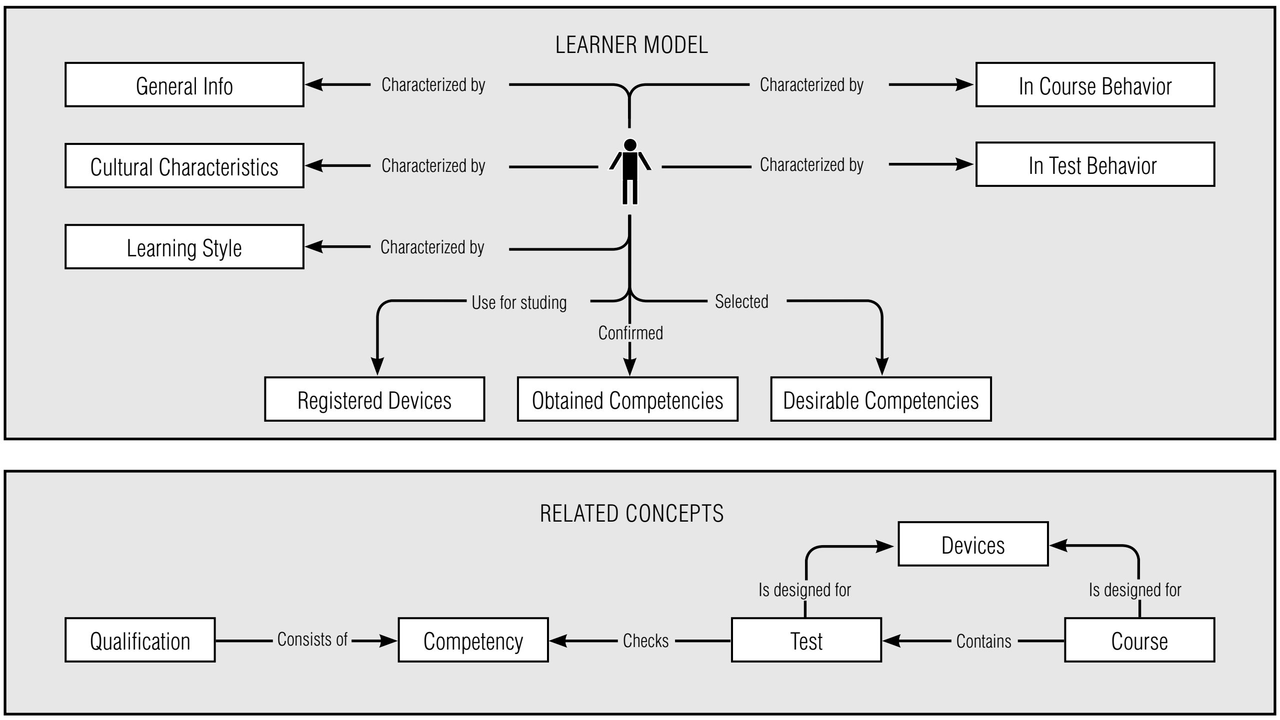

Fig. 1. Learner-centric ontology

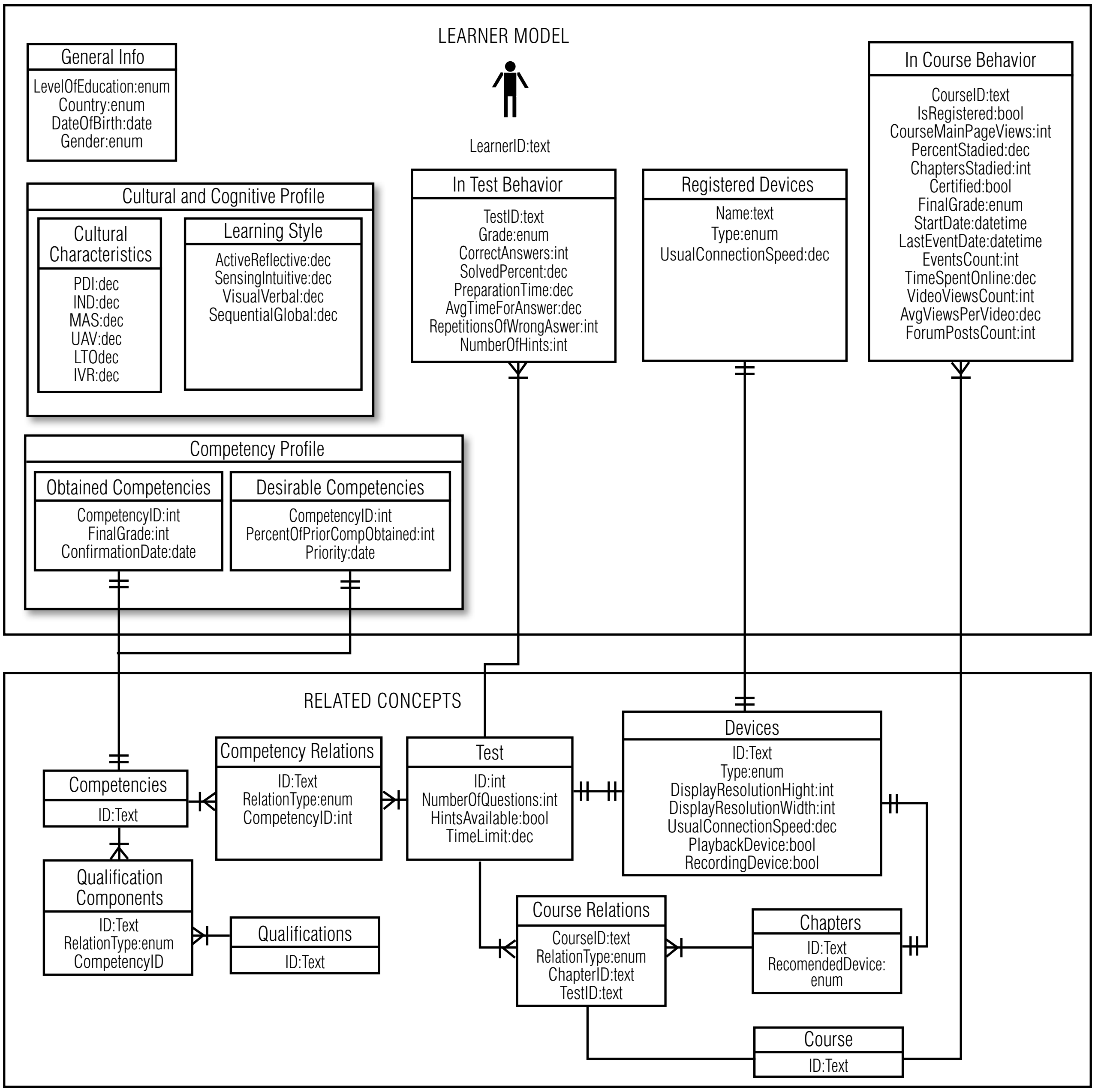

Fig. 2. Detailed model with attributes and relations of entries

2. Learner's model

Lots of models for the learner and learning process were presented by different researchers. For example, domain ontology by P.Q. Dung, A.M. Florea designed for application in a multi-agent e-learning system [11] or the model by A. Garcia-Gabot, taking the learner's device into account [12]. But these models lack a complex view of the learner, specificity of the e-learning process, and domain modeling. The following ontology model solves this problem (*Figure 1*).

The full model describing the relations and attributes of entities is the following (*Figure 2*).

The learner-centricity of this model is strongly associated with the idea of customer centricity in today's commerce, which is the key to business maturity[1]. As e-learning is actually a business process with certain revenues for both sides, it is obvious that learner-centricity is an important step for general e-learning development all over the world.

[1] "The journey toward greater customer centricity", Ernst & Young LLP, 2013

3. Individual learning trajectory

An individual educational trajectory (IET) – the plan of study, taking into account preferences of the learner in the final set of acquired competences, as well as his penchant for specific forms of presentation of cognitive objects and tasks, the different methods of research and forms of expressing ideas depending on the conditions of cultural and personal characteristics. The learner should select learning courses or desired competencies for building the IET. Every course may be characterized by a set of competencies according to the competency classificatory mentioned before, the duration and price. Moreover, the duration of some courses may depend on the specific learner and the time to obtain the same competency also varies in different courses. In addition, qualifications (sets of competencies) may be formed. Therefore, when a learner, for example, selects the qualifications of an IT business analyst it means he should get a number of competencies (business process analysis, system analysis, process modeling, etc.) and a list of required courses may be formed. There are different ways to form this list: minimize the number of courses, minimize the duration, minimize the price, etc. The learner should also have the option to select courses him or herself.

Thus, it is possible to base the mathematical model on the goal of obtaining a set of n competencies in m courses. For each competency *comp* there is a set of courses which provide that competency $\langle c_1,...c_i\rangle, 1 \le i \le m$; and vice versa, for each course there is a list of competencies it provides $\langle postcomp_1,..., postcomp_i\rangle, 1 \le i \le n$. Moreover, for each course there are lists of pre-requisite competencies $\langle precomp_1, \ldots, precomp_i\rangle, 1 \le i \le n$.

For each course there is a value $0 \le h \le 1$, which defines the difficulty of passing this course depending on the share of required pre-requisite competencies for this course having been obtained before the start.

Times required for obtaining competencies may be defined as

$$T = CT \cdot (1 - H), \tag{1}$$

where $T = \begin{bmatrix} T_1 \\ \ldots \\ T_m \end{bmatrix}$ – time for passing courses;

$CT = \begin{bmatrix} CT_1 \\ \ldots \\ CT_m \end{bmatrix}$ – common time for passing courses;

$H = \begin{bmatrix} H_1 \\ \ldots \\ H_m \end{bmatrix}$ – difficulty of courses.

The total time of passing all the required courses depends on the limit of courses to study at once, and the sequence of courses. And as soon as the total time depends on how prepared a learner is for each course, maximization of cross-usage of competencies by learning courses is a perfect way to determine their sequence and minimize the total time. This means that competencies provided by one course should be used by another one and so on whenever possible.

The total price of studying the courses depends on the total time because of the fixed costs per unit time (FC) like payment for an Internet connection, device amortization, etc.:

$$PR_{Total} = \sum_{i=1}^{m} PR + FC \cdot T_{Total}, \tag{2}$$

where $PR = \begin{bmatrix} PR_1 \\ \ldots \\ PR_m \end{bmatrix}$ – prices of courses.

Knowing the cultural specificity of a learner makes it possible to predict probable difficulties during a given course and even the probability of successfully passing an exam. Prediction of probable difficulties is possible because of:

✦ analysis of whether the learner has required competencies;

✦ analysis of correlations between the learner's cultural specificity and required competencies;

✦ analysis of how learners with similar cultural characteristics passed this course exam;

✦ analysis of which learner's behavior correlates with successfully passing courses.

As an example of the last type of analysis, several courses from combined HarvardX and MITx person-course academic year 2013 de-identified dataset[2] were analyzed using linear regression with the behavior parameters listed above. In *Table 1*, the listed regression coefficients define which behavior patterns are more or less important for passing the course. All of them have p-value below 0.05, which means that all coefficients are statistically valuable. The coefficient of determination (R^2) defines which percent of the results may be explained by the model.

Such analysis allows us to directly support learners having probably the wrong learning style or at least to inform them about possible problems and their reasons.

[2] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/26147

Table 1.

Regression analysis of behavior leading to successful passing courses

	R^2	Number of interactions with course	Number of days of activity	Number of video views	Number of course chapters studied	Number of forum posts
Introduction to Solid State Chemistry	0.81	0.479	0.182	–0.237	0.462	–0.007
Introduction to Computer Science and Programming	0.78	0.370	0.406	–0.162	0.240	–0.005
Introduction to Biology	0.75	0.814	0.173	–0.328	0.164	–0.021
Electricity and Magnetism	0.82	0.676	0.341	–0.243	0.104	–0.006
Mechanics Review	0.84	0.467	0.050	–0.043	0.454	0.017
Health in Numbers	0.74	0.624	0.070	–0.295	0.424	–0.017
Human Health and Global Environmental Change	0.67	0.793	0.031	–0.286	0.236	0.015

4. Adaptation of learning content

A learning course may be represented as a number of learning materials (multimedia or text documents). Every learning material should be divided into blocks. The block is an atomic part of learning material which should be marked by the most suitable ranges of Hofstede's cultural dimensions (one, two or even all six dimensions for very specific content) and index within material. Such a concept allows us to construct the most viable variant of learning material for a given learner. For example, if we describe the case of a company's success, the personal impact of a great leader may be focused for individualists, while statistical data may be focused for those who avoid uncertainty, etc. When the learner opens a document, it should be prepared by selecting the most suitable block for each index value or the default if no suitable block is found.

Moreover, the learning material representation interface may be adapted, too. There are five key points for such adaptation [9]:

✧ **metaphors** – typical images used in design;

✧ **mental models** – models of real life trains of thought, the use of which is stimulated when using the site;

✧ **navigation** – features of moving within web pages and between them;

✧ **interaction** – specificity of human-computer interaction;

✧ **appearance** – design features: colors, shapes, layout, sounds, the use of multimedia, etc.

Adaptation on most of these points may be automated. For example, computer vision algorithms (such as CaptionBot powered by Microsoft Cognitive Services) allow us to automatically describe pictures.

5. Mathematical models and algorithms

Application of the adaptive content concept requires some mathematical models and algorithms. The first task that must be solved is how to predict the possibility of successfully completing the course.

For solving this task, it is possible to use historical data on how learners with different values on cultural dimensions passed the given course exams. Logistic regression may be used for analysis of such panel data. The binary logistic model is used to estimate the probability of a binary response based on independent predictor variables:

$$\mathbb{P}\{y=1|x\} = F\left(x_1, x_2, \dots, x_n\right), y \in \{0,1\}. \qquad (3)$$

In this case, the binary response is success or failure in a course. Analyzing the statistics of previous students, it is possible to discover how important different parameters are for success in a particular course and use the measures of this importance (regression coefficients θ) multiplied by values for current student x to predict his result. The logistic function (or sigmoid) is:

$$F = \frac{1}{1+e^{-\Theta^T x}} = \frac{1}{1+\ e^{-(\Theta_1 x_1 + \Theta_2 x_2 + \dots + \Theta_n x_n)}}. \qquad (4)$$

Regression coefficients should be identified for every course using historical data about course completion as a training set. From the machine time cost point of view, it makes no sense to recalculate regression coefficients after getting the result for every new student, since its impact will be very low. For example, the HarvardX and

MITx de-identified dataset of academic year 2013[3] contains more than 640,000 entries (each represents one individual's activity in one edX course) for only 13 learning courses.

The next task is sequencing of courses. As was mentioned before, this algorithm should be based on maximizing cross-usage of competencies by learning courses. The depth-first traversal is the most appropriate algorithm for this task. It may build an optimal sequence from the list of courses and competencies they require and provide. The advantages of this algorithm are simplicity and avoiding the non-optimality of usual search.

The last top-tier task is forming groups of students with similar characteristics. This process may have different goals: from forming the exact number of groups within one course to forming groups according to one or several characteristics. For different goals, it makes sense to use different clustering algorithms such as K-Means, DBSCAN or Mean shift. Full descriptions of the algorithms may be found, for example, in [13].

K-means clustering

This method of vector quantization aims to partition n observations (x_1, x_2, ... x_n) into $k \le n$ clusters $S = \{S_1, S_2, \ldots S_n\}$ based on the nearest mean:

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2, \tag{5}$$

where μ_i is the mean of points in S_i.

Advantages:

1. K-means clustering is NP-hard problem, however, there are efficient heuristic algorithms. That makes K-means faster than most of other clustering algorithms.

Disadvantages:

1. K-means clustering requires pre-defined number of clusters.

2. K-means clustering has strong sensitivity to outliers and noise.

3. K-means clustering doesn't work well with a non-circular cluster shape.

DBSCAN (Density-based spatial clustering of applications with noise)

This algorithm groups points that are packed in space closely (have many neighbors), marks points that lie alone as outliers (low-density, nearest neighbors are too far away). DBSCAN is one of the most cited in scientific literature[4].

Advantages:

1. DBSCAN does not require a pre-defined number of clusters, opposing to k-means.

2. DBSCAN is able to find clusters of any form.

3. DBSCAN is robust to outliers.

4. DBSCAN requires only two parameters (ε and the minimum number of points required to form a dense region) and nearly ignores the order of entries being analyzed.

Disadvantages:

1. DBSCAN is not entirely deterministic

2. The scale and data must be well understood to correctly choose a meaningful distance threshold ε.

Mean shift (mode-seeking algorithm)

This algorithm is a non-parametric feature-space analysis approach for using discrete data for locating the maxima of a density function, from which the data was sampled. The mean shift algorithm is iterative and starts with an initial estimate x:

$$m(x) = \frac{\sum_{x_i \in N(x)} x_i K(x_i - x)}{\sum_{x_i \in N(x)} K(x_i - x)}, \tag{6}$$

where K is the weighted mean of the density in the window;

$N(x)$ is the neighborhood of x (a set of points for which $K(x) \neq 0$).

The difference $m(x) - x$ is actually called a mean shift.

Advantages:

1. The mean shift does not require a pre-defined number of clusters, opposing to k-means.

2. The mean shift is able to find clusters of any form.

3. The mean shift requires only one parameter (the size of the window).

4. The mean shift is well adapted for using a Gaussian kernel.

Disadvantages:

1. Selection of a window size for the mean shift algorithm is not trivial.

[3] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/26147

[4] DBSCAN is on rank 41 within most cited data mining articles according to Microsoft academic search, when accessed on: 21.08.2016

Turning to the technical part, it is important to note that a database is preferable for such service rather than data storage, because the main type of queries will be selection, not insertion of entries. Integration with existing information systems (such as learning management systems of universities, etc.) may be realized via mediating the replication layer and table triggers. Thus, any change in one system may be replicated in a number of tables of the other system. This makes it possible not to adapt the service database's structure for every integration.

Conclusion

Application of the adaptive content concept in the described way may significantly improve an e-learning process:

✦ increase the apprehensibility of learning materials. The process of e-learning does not provide such tutor support as full-time education does; it is mostly an autodidactic process. Thus, it is very important to decrease the psychological tension caused by feeling alone, having an uncertain interface and unintelligible learning materials;

✦ possibility to build an individual educational trajectory;

✦ possibility to form the learner's competency profile. Such a profile is also useful because, in fact, it is a ready-to-use part of the CV.

Further problems to study are the following:

✧ the need to prioritize the cultural dimensions during formation of learning materials in the case of mixed cultures (such as Chinese students studying in the USA for a long time) and, consequently, the development of a system of weighted coefficients;

✧ problems related to differentiation of cultural groups in the target audience (collaboration of representatives of different groups, etc.);

✧ the need to monitor the user's satisfaction and to make a forecast of changes in the user's preferences on the basis of data about users with similar parameters.

This publication is part of a series of works devoted to the analysis of conformity of cultural specificity of target audience and methods of content and design adaptation needed to ensure maximum comfort and ease of the e-learning process for this audience. The result of the entire work will be the development of a software environment that facilitates the preparation and demonstration of content sensitive to cultural specificity. ■

References

1. Uzilevsky G.Y. (2000) *Nachala ergonomicheskoy semiotiki* [Beginnings of ergonomic semiotics]. Orel: ORAGS (in Russian).
2. Huang K.-H., Deng Y.-Sh. (2008) Social interaction design in cultural context: A case study of a traditional social activity. *International Journal of Design*, no. 2, pp. 81–96.
3. Kholodnaya M.A. (2004) *Kognitivnye stili. O prirode individual'nogo uma* [Cognitive styles. On the nature of an individual mind]. St. Petersburg: Piter (in Russian).
4. Hofstede G. (1980) *Culture's consequences: International differences in work related values.* Beverly Hills, CA: Sage Publications.
5. Hofstede G., Hofstede J.G., Minkov M. (2010) *Cultures and organizations: Software of the mind.* N.Y.: McGraw-Hill.
6. Trompenaars F., Hampden-Turner C. (2004) *Managing people across cultures.* London: Capstone Publishing.
7. Triandis H. (2011) Culture and self-deception: A theoretical perspective. *Social Behavior and Personality*, no. 1, pp. 3–14.
8. Lewis R. (2006) *When cultures collide. Managing successfully across cultures.* Boston, London: Nicholas Brealey International.
9. Felder R.M., Silverman L.K. (1988) Learning and teaching styles in engineering education. *Engineering Education*, no. 7 (78), pp. 674–681.
10. Soloman B.A., Felder R.M. (2016) I*ndex of learning styles questionnaire*. North Carolina State University. Available at: http://www.engr.ncsu.edu/learningstyles/ilsweb.html (accessed 01 September 2016).
11. Dung P.Q., Florea A.M. (2013) Adaptation to learners' learning styles in a multi-agent e-learning system. *Internet Learning*, no. 1, pp. 11–20.
12. Garcia-Gabot A. (2013) *A proposal of a multi-agent system for adapting learning contents to user competencies, context and mobile device.* Research paper. Bratislava: Slovak University of Technology.
13. Gan G., Ma C., Wu J. (2007) *Data clustering: Theory, algorithms, and applications.* Philadelphia: SIAM.

Применение концепции адаптивного контента для электронного образовательного ресурса

Д.А. Алдунин
аспирант кафедры инноваций и бизнеса в сфере информационных технологий
Национальный исследовательский университет «Высшая школа экономики»
Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: daldunin@hse.ru

Аннотация

В настоящее время рынок электронного обучения быстро растет как в финансовом, так и в географическом плане. Электронные образовательные ресурсы все чаще привлекают мультикультурную аудиторию, становятся доступны для людей с различным уровнем образования. В то же время в различных культурах существует своя специфика восприятия информации и различные подходы к образовательному процессу. Данная работа посвящена иллюстрации возможного решения для адаптации контента электронного образовательного ресурса для мультикультурной аудитории. Описанное решение основано на применении концепции адаптивного контента для построения индивидуальной образовательной траектории и подготовки контента в соответствии с индивидуальными культурными особенностями обучающихся и его компетенциями (имеющимися и желаемыми).

В ходе исследования была разработана модель образовательного процесса, ориентированная на конкретного обучающегося. В статье представлены как модель верхнего уровня, так и детализированная модель. В статье также описаны основные принципы формирования индивидуальной образовательной траектории на основании имеющихся и желаемых компетенций пользователя и статистических данных о его стиле обучения. В качестве демонстрации возможности применять исторические данные о том, как образовательный стиль влияет на успешное завершение курса, приведен статистический анализ данных об изучении студентами семи курсов на платформах HarvardX и MITx в течении 2013/14 учебного года, показавший статистическую значимость целого ряда параметров. Также в работе представлено сравнение алгоритмов, применимых для оценки вероятности успешного завершения курса обучающимся в зависимости от его стиля обучения.

Ключевые слова: электронное обучение, кросс-культурный контекст, математическое моделирование, веб-разработка, индивидуальная образовательная траектория, образовательная аналитика.

Цитирование: Aldunin D.A. Application of the adaptive content concept for an e-learning resource // Business Informatics. 2016. No. 4 (38). P. 27–34. DOI: 10.17323/1998-0663.2016.4.27.34.

Литература

1. Узилевский Г.Я. Начала эргономической семиотики. Орел: ОРАГС, 2000.
2. Huang K.-H., Deng Y.-Sh. (2008) Social interaction design in cultural context: A case study of a traditional social activity // International Journal of Design. 2008. No. 2. P. 81–96.
3. Холодная М.А. Когнитивные стили. О природе индивидуального ума. СПб: Питер, 2004.
4. Hofstede G. Culture's consequences: International differences in work related values. Beverly Hills, CA: Sage Publications, 1980.
5. Hofstede G., Hofstede J.G., Minkov M. Cultures and organizations: Software of the mind. N.Y.: McGraw-Hill, 2010.
6. Trompenaars F., Hampden-Turner C. Managing people across cultures. London: Capstone Publishing, 2004.
7. Triandis H. Culture and self-deception: A theoretical perspective // Social Behavior and Personality. 2011. No. 1. P. 3–14
8. Lewis R. When cultures collide. Managing successfully across cultures. Boston, London: Nicholas Brealey International, 2006.
9. Felder R.M., Silverman L.K. Learning and teaching styles in engineering education // Engineering Education. 1998. No. 7 (78). P. 674–681.
10. Soloman B.A., Felder R.M. Index of learning styles questionnaire. North Carolina State University. [Электронный ресурс]: http://www.engr.ncsu.edu/learningstyles/ilsweb.html (дата обращения 01.09.2016).
11. Dung P.Q., Florea A.M. Adaptation to learners' learning styles in a multi-agent e-learning system // Internet Learning. 2013. No. 1. P. 11–20.
12. Garcia-Gabot A. A proposal of a multi-agent system for adapting learning contents to user competencies, context and mobile device. Research paper. Bratislava: Slovak University of Technology, 2013.
13. Gan G., Ma C., Wu J. Data clustering: Theory, algorithms, and applications. Philadelphia: SIAM, 2007.