# TeachMe: Personal Learning Environment for Children

Rosen Ivanov

***Abstract:*** *The paper presents a mobile environment that enables personalized interactive learning for preschool children and students. The required hardware is a mobile device, that is, a smartphone or tablet computer. The client-side application must support audio capture, audio play, HTTP(S) communication, and JSON parsing. It can be written in HTML5 and JavaScript, Java, Objective C or C#. Server-side software is written in Oracle Java Enterprise Edition. The user interface is specially designed to be easy to use, natural, transparent to technology, and accessible for children. It is based on screen gesture recognition, speech recognition, and speech synthesis. We use inquiry-based learning. Learners can send their requests to the service using their spoken voice. We developed and test two learning scenarios: 1) Preschool learning, starting with numbers, colors, shapes, counting, and getting information about any object in the immediate environment that represents the interests of the learner; and 2) L2 learning: personalized audio-lingual foreign language learning with a focus on word pronunciation and spelling. The learner receives information about the correct intonation and pronunciation of the words and their syntax.*

***Key words:*** *personal learning, inquiry-based learning, children's learning*

## INTRODUCTION

Modern mobile devices such as smartphones and tablets provide opportunities to connect to the Internet through available Wi-Fi wireless networks and radio networks of mobile operators (UMTS, HSPA, or LTE). This connectivity ensures mobility for users of Web- and cloud-based services. Mobile technologies are a key element of systems for mobile and ubiquitous learning. The concept of mobility [5] is defined in three areas: mobility of technologies, mobility of learning, and mobility of learners. Mobile teaching [1, 11] allows the learner to study in a new way that is personalized and learner-centred. When a suitable user interface is selected [12], mobile learning systems can be used successfully by young children.

Innovations in wireless communication and, mobile and Internet technologies such as Semantic and Social Web are responsible for an increase in the number of applications and services for mobile and ubiquitous learning. As shown in [6], the number of mobile and ubiquitous learning articles published from 2001 to 2010 in six major Social Science Citation Index (SSCI) journals is 154. Of these publications, 41 describe systems for teaching elementary school students. Over the last five years, there has been a growing interest in personalized learning tools and technologies. In personal learning services, students play an active part in the learning process, rather than passively consuming content. The teachers can easily create new courseware for any given learning goal.

The teaching of preschool and elementary school children should be an interactive process that takes into account each child's current knowledge level, readiness and skills needed to learn new knowledge, and interests. The training of children must be based on cognitive learning that reinforces their innate curiosity and encourages them to be creative in the learning process. The goal is to make learning accessible from any place and at any time and to make it effective and appropriate for the skill level of each student, allowing for the easy creation and sharing of courseware, to be transparent to the technologies used for its implementation, and to allow realization of a relationship between objects from the real world and their digital description (text, sound, speech, images, and video).

Current trends in learning systems suggest the efficacy of personalized learning [14]. Personal Learning Environments (PLEs) are self-defined collections of services, tools, and devices that help learners to build their own learning resources. Specific PLE features include the following [3]: 1) extracting learning content from social networks, search engines, and feeds; 2) filtering, sorting, and aggregating learning content; 3) exporting results in platform-independent formats such as Extensible Markup Language (XML), Atom, or JavaScript Object Notation (JSON); and 4) sharing learning content with other students. Training content must be delivered to users in platform-independent formats. For example, Sharable Content Object Reference (SCORM) [4] is a collection of standards and specifications for web-based e-learning. SCORM uses XML and JavaScript to realize learning content transfer.

The main objective of this research is to design and implement a mobile environment for interactive and personalized learning of children. The developed service can help children to have more effective and enjoyable learning experiences based on a highly interactive approach for courseware building and personalized teaching. Courseware was constructed from available web resources (news, images, videos, and suggestions).

## METHODOLOGY

In the traditional method of teaching, students are passive participants in the learning process as they receive the knowledge they need from teachers, primarily through oral presentations of training materials. Passive training of children is not effective because it has the following drawbacks: 1) It does not encourage students to work independently. 2) It does not take into account the interests and current knowledge level of each child, while it is known that the amount of time to reach a certain level of knowledge in a given learning area is strictly individual. 3) It does not foster the interests and creativity of students and provides only basic facts that they need to remember. In recent years, the number of mobile and ubiquitous learning systems has increased, as these provide many opportunities for learning to take place beyond the classroom and directly links the real world with its digital representations. Most mobile and ubiquitous learning apps are designed for the education of children older than 6 years [6]. Interactive learning services that are designed especially for younger children (preschool learning) must take into account numerous factors, such as selecting an approach for the development of training materials by teachers or by children's parents, and selecting the format of the learning materials and the method for sharing them among learners. Preschool learning is important for building a foundation for the child's future cognitive, social, and emotional development as well as for his or her creativity and socialization. The training of children 4 through 6 years of age implies the need for teachers [8] because children on their own cannot acquire the necessary basic knowledge and cannot create the training materials, no matter how accessible the user interface is. Teachers should help children when they work with such services until they obtain the necessary confidence and skill to use them independently. Training should not stress the children.

## SYSTEM ARCHITECTURE

In this paper, we present the design and implementation of mobile service for the interactive and personal teaching of children. The design of the user interface and the development of the courseware are methodologically consistent with the characteristics of the intended learners. The service has client-server architecture (see Figure 1). Software on the server side is written in Java Enterprise Edition. Clients of the service must have mobile devices (smartphones or tablet computers) with the following capabilities: implementation of network transfer via protocol HTTP(S), audio capture, audio play, touch screen, and Near Field Communication (NFC) sensor (optional).

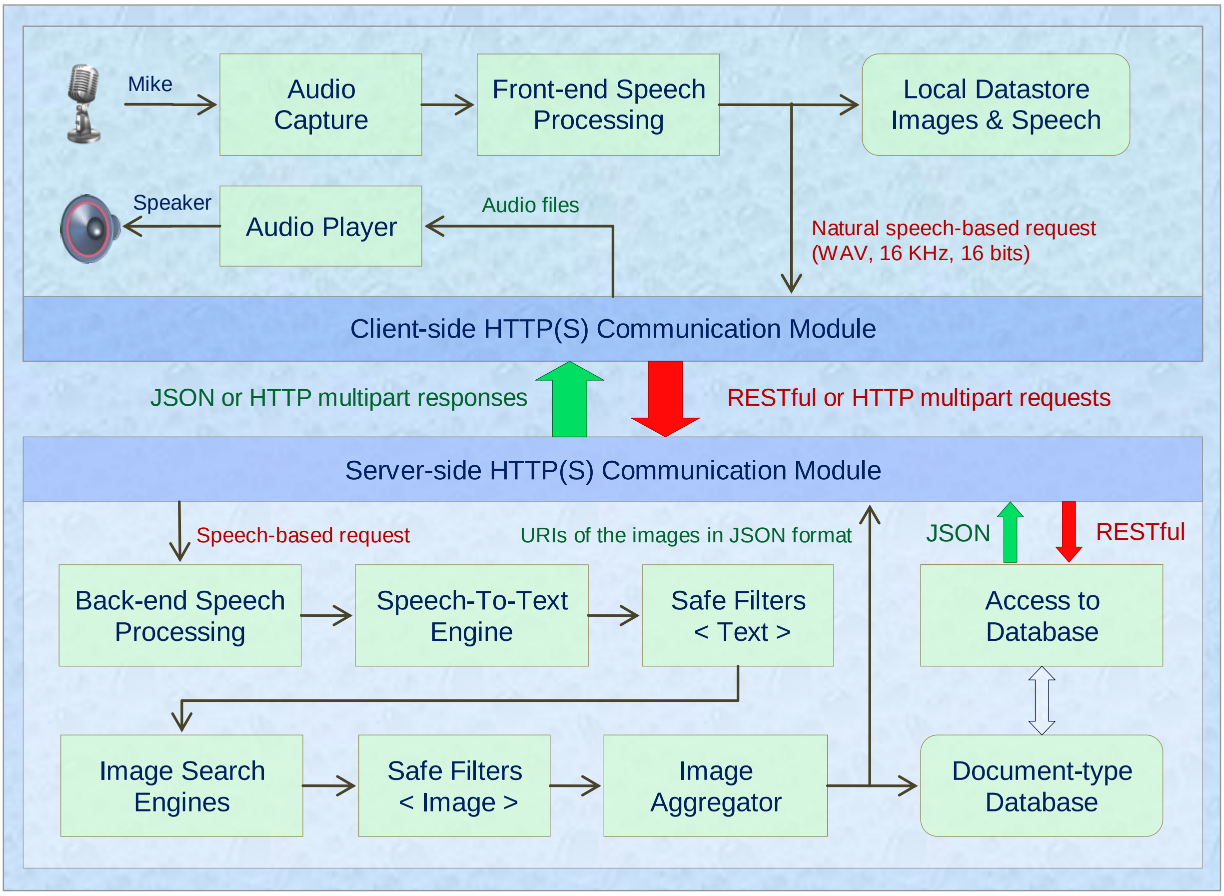



**Fig. 1** Service architecture overview

We use resource-based communication between the clients and the service. In the service proposed, this type of communication is achieved by protocol Representational State Transfer (REST). The responses are in JSON format. Thanks to this platform-independent architecture, access to the service is possible from a wide variety of devices. To create client-side application, developers can use different programming technologies and languages such as HTML5 and JavaScript, the Android Java, Oracle Java Mobile Edition, Apple's Objective-C, and Microsoft C#. We implemented service based on the proposed architecture. The development environment is shown in Table 1.

**Table 1** Development environment

| | |
|---|---|
| Server platform | WEB server: Apache Tomcat Coyote 7.0, RESTfull support: Jersey framework 2.7, Java Development Kit (JDK) 1.7, Database: Apache CouchDB 1.5.0, SSL certificate: Symantec VeriSign |
| Integrated Development Environment (IDE) | NetBeans 7.4 (Java EE) and required plugins |

Access to the service requires user registration. It is implemented in two ways: by entering a user name and password, and automatically through an NFC smartcard (if the mobile device has a built-in NFC sensor). In the latter case, the user name and the password are formed by the ID of the NFC smartcard and the IMEI number of the mobile device. After successful registration, the client receives a user id, an app key, and a secret code. Access to the resources of the service requires OAuth authorization. In this case, the user name and password are not transmitted to the service.

At this stage, the service allows receiving suggestions and image searches. The following describes the process of extracting the resources (images) when a client sends a voice-based request to the service:

1. *Audio capture*. The client selects the language that he or she wants to use, activates an audio recording (16-bit PCM, 16 kHz sampling rate) and pronounces the phrase that contains the query to the service. Languages supported thus far are: English (UK), English (US), French, Spanish, German, Italian, and Bulgarian. The speech signal preprocessing (front-end stage) includes finding the start and stop points of speech signal and normalizing energy in segments between these points. The multi-part request to the service contains the client's voice inquiry in WAV format and parameters that describe this inquiry.

2. *Back-end speech processing*. The aim of this module is to convert the voice-based query to an audio format required by the speech-to-text (STT) engine.

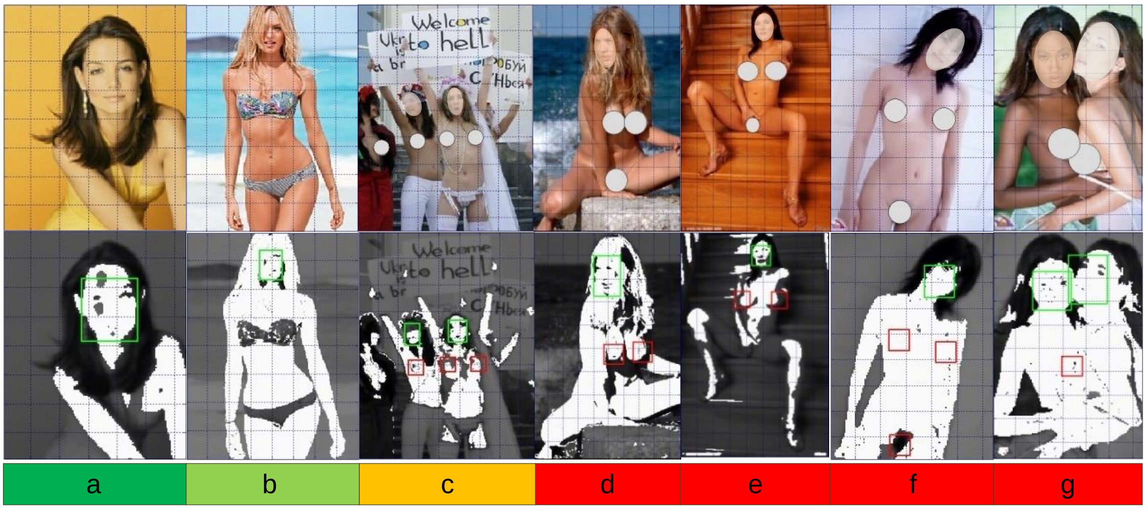
3. *STT Engine*. At this stage, the service uses Google Speech API in order to realize the natural speech recognition and to convert the client's request (16-bit FLAC, 16 kHz sampling rate) to the corresponding text. The response is in JSON format and contains information about the recognized utterance and the probability for correct recognition. To achieve the best recognition accuracy, Google uses state-of-the-art Deep Neural Networks (DNNs) for acoustic modelling of speech [9]. The model is a standard feed-forward neural network with four hidden layers of 2560 nodes, each computing a nonlinear function of the weighted sum of the outputs of the previous layer. The input layer is the concatenation of 26 consecutive frames of 40-dimensional log-filterbank energies calculated on 25 ms windows of speech every 10 ms. The 7969 soft-max outputs estimate the posterior of each acoustic state. This model has 50M parameters, and the overall word-error rate is 12.3%.

4. *Safe search – text filtering*. This module analyses the text returned by STT engine for words or phrases that could be associated with erotic or pornographic content.

5. *Image search*. There are many search engines that allow users to find images hosted on the Web, for example, Google, Bing, Startpage, Flickr, Ask, FAROO, Instagram, etc. Each of the existing image search engines has its advantages and disadvantages. The quality of these services depends on many factors, such as search language, image size, image file format, etc. For this reason, the proposed service uses aggregation of the results of three image search APIs: 1) Google Custom Search API, 2) Microsoft Bing Search API in Windows Azure Marketplace, and 3) Flickr Image Search API. All three services require registration as well as payment if they are used for commercial purposes. Each of the image search APIs used returns the result in XML or JSON format. It contains a URI to the image files that are matched with the phrase to be searched. The users of the service can specify the number of images to return and sorting expressions to apply to the results.

6. *Safe search – image filtering*. Although the client's requests to the service are filtered at different stages of processing (STT engine, text-based filtering, image search engines), it is possible for learners to gain access to images that contain erotic or pornographic content. For this reason, it is necessary that each image be checked to determine whether or not it contains erotic or pornographic parts. Pornographic image recognition is a complex object-recognition task. Most of the current algorithms for the detection of pornographic content in image files are based on skin-color detection [7] and region-of-interest (ROI) detection using AdaBoost or Bag-of-Visual-Words classifiers [13]. The main disadvantage of such algorithms is their continuing high false positive or negative rates; it is difficult to differentiate between human skin coloration and surrounding objects with skin-like color [2], between navel and nipple [15], and between bikini images and some erotic or pornographic images.

We use a fuzzy classifier that fuses three data channels: 1) context-based image features (name, title, and tags), 2) site-safe rating (keywords and content-based analysis of the site from which the image is downloaded), and 3) image content; we combine an adaptive skin-color detection and Speeded Up Robust Features (SURF)-based face and pornographic ROI (PROI) detection. The detection rate for our approach was 95.3%.




| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | katie47<br>**green** | V390913<br>**green** | Corbis-42-27579723<br>**green** | Vika001_large<br>**orange** | anita_mc_08<br>**green** | Beeg<br>**orange** | India02<br>**orange** |
| 2 | www.flickr.com<br>key: **green**<br>desc: **green** | victoriassecret.com<br>key: **green**<br>desc: **green** | corbisimages.com<br>key: **green**<br>desc: **green** | pinkfineart.com<br>key: **red**<br>desc: **orange** | theadultsection.com<br>keys: red[1]<br>desc: red[2]<br>1, 2 – no info | phimx3x.com<br>key: **red**<br>desc: **red** | teenpornstarz.com<br>keys: **red**[1]<br>desc: **red**[2]<br>1, 2 – no info |
| 3 | PROI: 0 | PROI: 0 | PROI: 3 | PROI: 2 | PROI: 2 | PROI: 3 | PROI: 1 |

**Fig. 2** Image classification after filtering: 1) Context, 2) Site rating, 3) Image content (key = keywords, desc = description)

Each image is assigned to one of three basic classes: 1) green (image does not contain erotic or pornographic parts), 2) orange (image is erotic but not pornographic), and 3) red (image is pornographic). The service returns only images with a "green" status.

Figure 2 shows the results returned after the classification of the images that include the human body with various exposed parts. The bikini image from Figure 2b is classified correctly as "green" because it was downloaded from a "green" site and no pornographic ROI was detected. The image shown in Figure 2c is classified as "orange." Although the algorithm has identified three PROIs, which contain nipples, the image is classified as erotic because the other two data channels are showing a "green" status. All other images (d–g) are classified as "red"; they contain pornographic regions, and the text description and sites from which they were downloaded have a "red" or "orange" status.

7. *Return a response*. Images that have a "green" status are recorded in a database. We use CouchDB, a document-oriented NoSQL database server, accessible via a RESTful JSON API. A CouchDB document is an object that consists of named fields and attachments. The response of the service for the client who sent the request is in JSON format. It contains information about the number of found images, links to any of the images, the id of each image, and image MIME file format.

## METHODS

The basic way of acquiring knowledge in the teaching of children ages 3 to 6 is by asking questions. Three-year-olds ask questions that include the words "what," "where," and "who." Children ages 5 and 6 ask questions that require a more complex answer and often include the words "when," "why," and "how." Considering the pedagogical specificities in the teaching of children, the service uses inquiry-based learning; that is, it searches for knowledge by asking questions [10]. Through the process of inquiry, children construct much of their understanding of the natural and human-designed world.

At this stage, we developed and implemented two scenarios for the teaching of children: 1) teaching young children ages 4 through 6, and 2) teaching children ages 7 and older. In Table 2, the parameters describing each of the two scenarios are given.

**Table 2** Learning scenarios description

| Parameter | Scenario 1 | Scenario 2 |
|---|---|---|
| Language skill focus | starting with numbers, colors, shapes, object identification | L2 teaching. Linguistic focus: pronunciation and word spelling |
| Methodology | confirmation inquiry | open inquiry |
| Level | beginners, intermediate | beginners, intermediate |
| Target learners | 4–6 years old | 7 years and older |
| Learning type | supervised | personalized (unsupervised) |

**Scenario 1**

The user interface of the service is based on the recognition of natural speech and voice synthesis, since the majority of children ages 4 through 6 cannot yet read and/or write. The user interface simulates a familiar strategy for children's methods of communication; they ask questions and receive visual answers, i.e., images whose content is related to their queries.

Taking into consideration the level of knowledge of young children, we offer training that begins with recognition of known objects from their surroundings and the world, such as animals, fruits, vegetables, etc. The next level of training is related to the detection of characteristics of objects, such as their shapes and colors. Because most children like to count objects that they recognize, the next level of training is related to recognizing and counting numbers. Several questions that are specific for different levels of training are following: What animal do you see in the picture? Which picture contains a circle? Which picture contains a red square? How many dogs do you see in the picture? Count them!

*The following steps must be taken when creating the training materials:*

1. Select "Create new learning courseware."
2. Specify the name of the courseware via voice.
3. Search for an icon to associate with the teaching course. For example, if the course title is "Wild Animals," the teacher says this phrase aloud and receives in response eight images that are associated with it. He or she selects one of these images for the courseware icon.
4. Select the type of training course. Thus far, the service supports the following types of courses: a) learn what is shown in the picture, b) guess what you see in the picture, c) find a specific image or thematically related images, and d) see the sequence of images and hear their voiced descriptions.
5. The teacher should record all necessary audio files for the implementation of the training course's questions and answers.
6. The trainer should record the expressions that students will hear when they give either a correct or an incorrect answer.
7. Finalize the courseware. All audio recordings, images, and configuration files are stored in the service database. The file that describes the content and type of courseware is in XML format. It is used to share courseware with other learners.

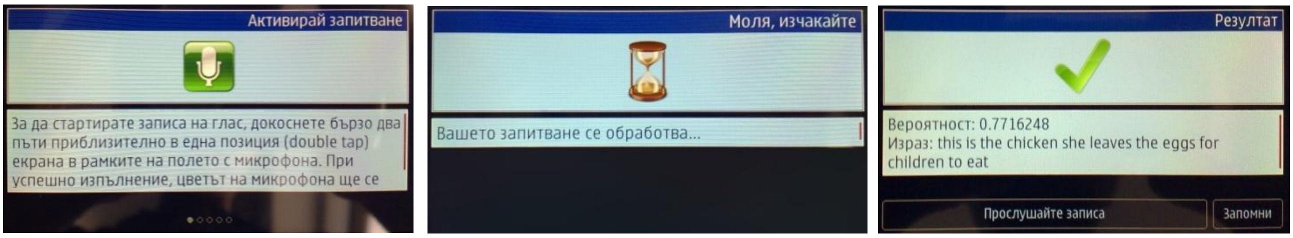
**Scenario 2**

The aim of this scenario is the training of children age 7 and older. An application for learning foreign languages was developed. Mobile applications for language learning are focused on providing knowledge in the following areas: grammar, reading and writing, and listening and speaking. One of the major concerns with children's language learning is the need to teach them to pronounce words correctly. We developed a scenario for mobile L2 learning that provides opportunities for personalized education through reading, listening, grammar, and reviewing the probabilities of correctly pronouncing words and the intonation of phrases. To enable entirely personalized L2 children's learning, the service uses a combination of automatic speech recognition and text-to-speech (TTS) conversion. In this case, it is not necessary to have a teacher qualified to teach L2, as pronouncing words correctly is guaranteed; the speech synthesis engine uses rules for word pronunciation, intonational phrasing, and segmental durations. We use Google Translate API to realize TTS conversion. The relationship between input texts and their acoustic realizations is modelled by a deep neural network (DNN). The use of a DNN can address some limitations of the conventional approach. The DNN-based systems outperformed the Hidden Markov Model (HMM)-based systems with similar numbers of parameters. Deep learning in speech synthesis allows for the modelling of high-dimension, highly correlated features efficiently.

*The following steps must be taken when creating this scenario:*

1. The learner selects the language he or she wants to study.
2. The student says aloud a word or phrase that he or she wants to know how to write and how to pronounce correctly.
3. The speech-based request is converted to an audio file($s$) and sent to the service. Through the speech-to-text engine, the speech-based request is converted to the corresponding text ($T$).
4. By TTS module, the text $T$ is converted to its equivalent synthesized speech $\hat{s}$.

The student can thus see how to write the words and hear how to pronounce them correctly. To achieve this pronunciation, the learner can hear a word or phrase pronounced correctly repeatedly. The whole process can be repeated as many times as necessary, for example, to achieve a probability of correct pronunciation over a preset threshold level. Figure 3 shows the results obtained when testing the mobile app for L2 learning.




a) b) c)

**Fig. 3** Implementation and test of the L2 learning scenario:
a) Double-tap the microphone icon to start audio recording (voice inquiry); b) wait for a response; and c) see how to write and hear how to pronounce the words.

## CONCLUSIONS AND FUTURE WORK

In this paper, we presented a design and a prototype for a mobile environment (tentatively titled TeachMe) that supports children's interactive and personalized learning. The service can provide children with more effective and enjoyable learning experiences, based on its highly interactive approach for courseware building and personalized teaching. We use inquiry-based learning. Learners are able to send their requests to the service via the spoken voice and receive answers as synthesized speech or images. The service supports two types of inquiry-based learning: 1) Confirmation Inquiry, where teachers develop questions and a procedure that guides learners to predefined results; and 2) Open Inquiry, where learners formulate their own research questions, analyze their findings, and share the results.

The primary features of the proposed learning environment are as follows: *First*, the user interface is designed specifically to be easy to use, natural, transparent to technology, and accessible for children. It is based on screen gesture recognition, speech recognition, and speech synthesis. *Second*, to develop the training materials, additional resources and specific technical knowledge are not required, which reduces the time for their implementation. *Finally*, children can ask questions related to all areas of knowledge because of the possibility of aggregation of information from different social networks.

Resources for the service that implements its functionality are available through a specifically designed RESTful JSON API. Future work includes the development of usage samples in a variety of languages (JavaScript, Java Android, Microsoft C#, and Objective-C) to help developers begin using TeachMe API. We plan to perform complete usability tests for the developed learning environment.

## REFERENCES

[1] Ally, M., A. Tsinakos (Editors), Increasing Access through Mobile Learning, Commonwealth of Learning and Athabasca University, 2014.

[2] Chaowei, T., L. Zhenfeng, S. Yanqing, T. Hui, Z. Xu. Joint detection method of pornographic images with skin-like color background, High Technology Letters, 2012, Vol.18(2), pp.132-138.

[3] Conde, M. Á., F.J. García-Peñalvo, M. Alier, J. Piguillem. The Implementation, Deployment and Evaluation of a Mobile Personal Learning Environment, Journal Universal Computer Science, 2013, Vol. 19(7), pp. 854-872.

[4] DigitalThink. SCORM™: The E-Learning Standard, 2003.

[5] El-Hussein, M., J. Cronje, Defining mobile learning in the higher education landscape. Journal of Educational Technology & Society, 2010, Vol. 13(3), pp. 12-21.

[6] Hwang, G.J. and C.C. Tsai, Research trends in mobile and ubiquitous learning: a review of publications in selected journals from 2001 to 2010, British Journal of Educational Technology, 2011, Vol. 42(4), pp. 65-70.

[7] Karavarsamis, S., N. Ntarmos, K. Blekas, I. Pitas. Detecting pornographic images by localizing skin ROIs, Int. Journal of Digital Crime and Forensics, Vol. 5(1), pp. 39-53, 2013.

[8] Lee, J. Preschool Teachers' Shared Believes About Appropriate Pedagogy for 4-Years-Olds', Journal Early Childhood Education, 2006, Vol. 33(6), pp. 433-441.

[9] Lei, X., A. Senior, A. Gruenstein, J. Sorensen. Accurate and Compact Large Vocabulary Speech Recognition on Mobile Devices, INTERSPEECH 2013, pp. 662-665.

[10] Levy, B., E. Thomas, K. Drago, L. Rex. Examining Studies of Inquiry-Based Learning in Three Fields of Education: Sparking Generative Conversation, Journal of Teacher Education, 2013.

[11] Pachler, N., B. Bachmair, J. Cook. Mobile Learning, Springer New York, 2010.

[12] Read, J.C., P. Markopoulos. Child–computer interaction, Int. Journal of Child-Computer Interaction, 2013, Vol. 1, pp. 2–6.

[13] Ries, C.X., R. Lienhart. A survey of visual adult image recognition, Journal Multimedia Tools & Applications, 2014, Vol. 69(3), pp. 661-688.

[14] Wagener, U. Young Children's Self-Regulated Learning: What does it Look Like in the Classroom? Journal of Child and Youth Development, 2013, Vol. 1(1).

[15] Wang, Y., J. Li, H. Wang, Z. Hou. Automatic Nipple Detection Using Shape and Statistical Skin Color Information, Springer-Verlag, LNCS 5916, pp. 644–649, 2010.

## ABOUT THE AUTHOR

Assoc. Prof. Rosen Ivanov, PhD, Department of Computer Systems and Technologies, Technical University - Gabrovo, Phone: +359 887 BULG45, E-mail: rs-soft@ieee.org.