Use Case of Cognitive and HCI Analysis for an E-Learning Tool

Marian Cristian Mihăescu [1], Mihaela Gabriela Ţacu [2], Dumitru Dan Burdescu [1]
[1] University of Craiova, Faculty of Automation Computers and Electronics, Romania
[2] University of Craiova, Faculty of Economics and Business Administration, Romania
E-mail: mihaescu@software.ucv.ro, mihatacu@yahoo.com, burdescu@software.ucv.ro

This paper presents the development of a new tool for an e-Learning platform, with implications and analysis from human computer interaction (HCI) and cognitonics perspectives. The goal is to improve the educational process by helping the professor form a correct mental model of each student's performance (capability). Besides this, the developed application is also analyzed from HCI and cognitive perspective with an attempt to offer an effective and highly usable tool. The interplay between cognitive psychology and HCI is emphasized as a fundamental prerequisite for a constructive and argumentative design. The main functionalities offered by the developed application are: evaluation of the level of understanding of the course material by the student and analysis of the difficulty level of the questions proposed by the professor for an exam. The prerequisites for accomplishing the task are a good structure of the on-line educational environment and information about students' activities on the platform. An important part of the process of obtaining the desired result is performing a text analysis and concept extraction on the professor's uploaded courses. The languages supported by this module are both English and Romanian. Using this module, the professor will have the ability to control the difficulty of test and exam questions and make changes accordingly: add, delete or modify test/exam questions. The developed application is thereafter discussed from HCI and cognitive psychology perspectives, such that further analysis and improvements are ready to be achieved. Our approach creates a context in which continuous design, implementation and evaluation has as output a high quality user interface suitable for an Intelligent Tutoring System.

Povzetek: Opisana je nova platforma za učenje, ki omogoča uporabo kognitivnih in HCI analiz.

1 Introduction

During the last years, the interaction between professors and students in on-line educational environments has been considerably improved, especially by developing new tools and implementing different functionalities that integrate intelligent data analysis techniques. An area that still needs further work is the cognitive area, particularly towards helping the professors build more accurate mental models of each student's capabilities.

In regular educational environment, a professor can achieve that mental model by continuously interacting with students and observing their learning skills and capabilities. Online, it is harder to accomplish that, because of the lack of constant and valuable analysis of feedback that is offered by students. That is why the approach to building a professor's mental model of student's activity becomes a tool that can improve an on-line educational environment.

The main purpose of the tool is to help the professor understand and analyze student's activity without having any face-to-face activity.

Besides the design and development of the tool that actually implements the needed functionality, this paper also presents a detailed analysis from human computer interaction (HCI) and cognitive perspective.

The research of HCI and cognitive psychology issues are cornerstone in shifting gears from "technology that solves problems" towards "design that emphasizes the user's needs". These general research areas have a great impact on the field of e-learning due to the wide range media that can produce cognitive affection at various industries. Among the most common options there are simple text, voice, picture, video, or virtual reality.

The final goal of the tool is to extend its usability with respect to the particularities of e-learning environments. That is why the general fundamental usability evaluation formulated by Nielsen [25] needs a proper specific adjustment for e-learning environments. The most important characteristics of a usable e-learning environment should be usefulness, effectiveness, learnability, flexibility and satisfiability [1].

On the other hand, discussing web-based instructional strategies from cognitive point of view requires a different approach that should mainly be constructive (demands for synthesis in the process of building the tool) and argumentative (justify design decisions of the developed tool, critically asses the tradeoffs in alternative designs and conduct usability studies to evaluate prototypes).

2   Related works

Let's regard related works mainly in the following research areas: educational data mining, HCI and cognitonics. Within the educational data mining (EDM) area this paper is closely related to Intelligent Data Analysis performed on educational data.

Among many research problems from EDM, this paper relates to the works that have attempted to represent and analyze educational data with the goal to improve their knowledge level through custom designed data analysis systems and applications. Thus, implementing data mining tasks on educational data (e.g. performed activities, educational resources, messages, etc.) provides the environment in which progress may be achieved.

Many research hours have been allocated to the purpose of extracting key concepts from course materials, messages, questions and finding ways of using them for enhancing the teaching and learning processes [2, 3].

Also, a considerable amount of work has been put into discovering the similarity between concepts. Relevant in this area is the paper that Tara Madhyastha wrote in 2009 called „Mining Diagnostic Assessment Data for Concept Similarity" [4], it presented a method for mining multiple-choice assessment data for similarity of the concepts represented by the multiple choice responses. The result obtained was a similarity matrix that can be used to visualize the distance between concepts in a lower-dimensional space.

The NLP (Natural Language Processing) is another major research area, with a strong focus on documents (text and diagrams). Particularly interesting from our perspective is the research conducted in the domain of linguistics [11, 12]. Important is also the work put into constructing treebanks, both monolingual [13] and parallel. In [9] M. Colhon presents the construction of an English-Romanian treebank, a bilingual parallel corpora with syntactic tree-based annotation on both sides, also called a parallel treebank. Treebanks can be used to train or test parsers, syntax-based machine translation systems, and other statistically based natural language applications.

Agathe Merceron and Kalina Yacef published a case study about how educational data mining algorithms can be used for extracting relevant information from web-based educational systems and how this information can be used for helping teachers and learners [5]. A comprehensive report on the state of the educational data mining was published in 2009 [6] and presented a general view of the EDM field, including the methods used, the papers with the most influence in the field and the key areas of application.

One of the main goals of the educational research is identifying students' current level of understanding. For this purpose, a series of estimates have been used, including DINA model, sum-scores and capability matrix. A comparison between these estimates was presented in „A Comparison of Student Skill Knowledge Estimates" [7].

This paper presents an approach to using concept extraction along with activity monitoring and concept weighting towards constructing accurate models of students' present knowledge and level of understanding of the courses, as well as detecting the difficulty level of each course, course chapter, test question or exam question.

Another scientific discipline whose contribution is necessary is cognitonics [18, 19, 20]. Cognitive aspects are playing an important role in the information society and also in the particular case of e-learning applications. From cognitonics point of view, the developed applications for sustaining on-line courses (or other related activities) should develop creativity, support cognitive-emotional sphere and appreciate the roots of the national cultures. One of the main goals of this quite new research discipline it is the development of a new generation of tools for on-line learning that compensate the broadly observed negative distortions [18].

Among many application domains where cognitonics (cognitive psychology for information society and advanced educational methods) finds a suitable place is e-learning. From this point of view, e-learning tends to need progress from various research domains (e.g. EDM, cognitonics, HCI, etc.) in order to improve its effectiveness and control the main negative side effects regarding linguistic ability, phonological ability, social relationships, etc.

Another research domain that is highly connected with the discussed issues is HCI. Currently, there are numerous research efforts that deal with user interface adaptation in e-Learning Systems, adaptable interfaces featuring multiple views and finally integration of usability assessment frameworks that are designed and refined for the context in which they are applied. HCI issues related to e-learning are user-centered design [21] and user sensitive design [22]. From this perspective, adaptation of knowledge presentation, of interaction style regard specific issues like domain knowledge base generation, user/system interaction modeling, interface evaluation. Poor interaction in various on-line educational activities (e.g. evaluation of exercises after class, quiz games, intelligence analysis, etc.) may find proper solution by employing specific HCI research methodologies related to usefulness, effectiveness, learnability, flexibility and satisfiability.

3   Tools and technologies

The context for which the tool is developed is related to on-line education. From this point of view, the developed tool is actually a web application that gathers various technologies in order to achieve its business goal.

The first step in accomplishing this module's purpose is retrieving the text from documents. For reading .pdf files, we used Apache PDFBox, which is is an open source Java tool for working with PDF documents [10]. For manipulating .doc and .docx files, our choice was Apache POI [14], a powerful Java API designed for handling different Microsoft Office file formats.

Stemming [16] is the process for reducing inflected (or sometimes derived) words to their stem, base or root. The documents written in English were stemmed using the snowball stemmer [15]; as for the Romanian stemmer, we used as a base the PTStemmer implementation [17] and adapted it for the Romanian language by building the corresponding set of grammatical reduction rules: plural reduction, genre reduction, article reduction. The PTStemmer is a toolkit that provides Java, Python, and .NET C# implementations of several Portuguese language stemming algorithms (Orengo, Porter, and Savoy).

The XML processing was done using the Java DOM parser.

4 System architecture and usability assessment

The most important prerequisite for the development of such a tool is an online educational platform that has a proper structure for the educational assets and the ability to integrate proper intelligent data analysis techniques.

The online educational system we have chosen is Tesys Web [8], an e-learning platform used in several faculties from University of Craiova. Tesys has been designed and implemented to offer users a collaborative environment in which they can perform educational activities.

4.1 General architecture

Figure 1 presents the general architecture of the system.
In the left part of the figure we can see only persistent data, basically found on the server, and on the right side the core business logic is presented, it includes the concept extraction, activity monitoring and recommender modules.

Starting from the course documents that were previously uploaded by the professor on the platform, the system extracts the concepts, using a custom concept

extraction module, which incorporates a stemming algorithm and TF-IDF formulas. The obtained data is then transferred into the XML files. The five most relevant concepts are also inserted into the Tesys database, for further use.

As soon as the professor uploads the test questions and specifies each concept's weight for every question, the student's activity monitoring process can begin.

Afterwards, using the concept-weight association, student's responses to the test questions and taking into consideration the performances of student's colleagues, the system will be able to show relevant statistics to the professor, so he can understand each student's learning difficulties as well as the general level of the class.

The recommender module is designed to review the difficulty of the proposed exam questions and advise the professor on lowering or increasing the exam difficulty. All this process is supervised by the professor, who takes the final decision.

4.2 Concept extraction tool

A key feature of this module is the extraction of the most important concepts from every chapter that belongs to a course.

This part of the module is divided into two steps: stemming and computing TF-IDF values.

Several tools and algorithms have been developed for English word stemming, but for the Romanian language this research area is still at the beginning, therefore we developed our own tool and set of rules to accomplish this task.

After the stemming process we use the TF-IDF formulas for every word in the document and then we store the obtained data into an xml file, which has the following structure:

```
<?xml version="1.0" encoding="UTF-8"
standalone="no"?>
<words>
<regul originalForm="reguli"
tf="1.00" idf="0.47"/>
```
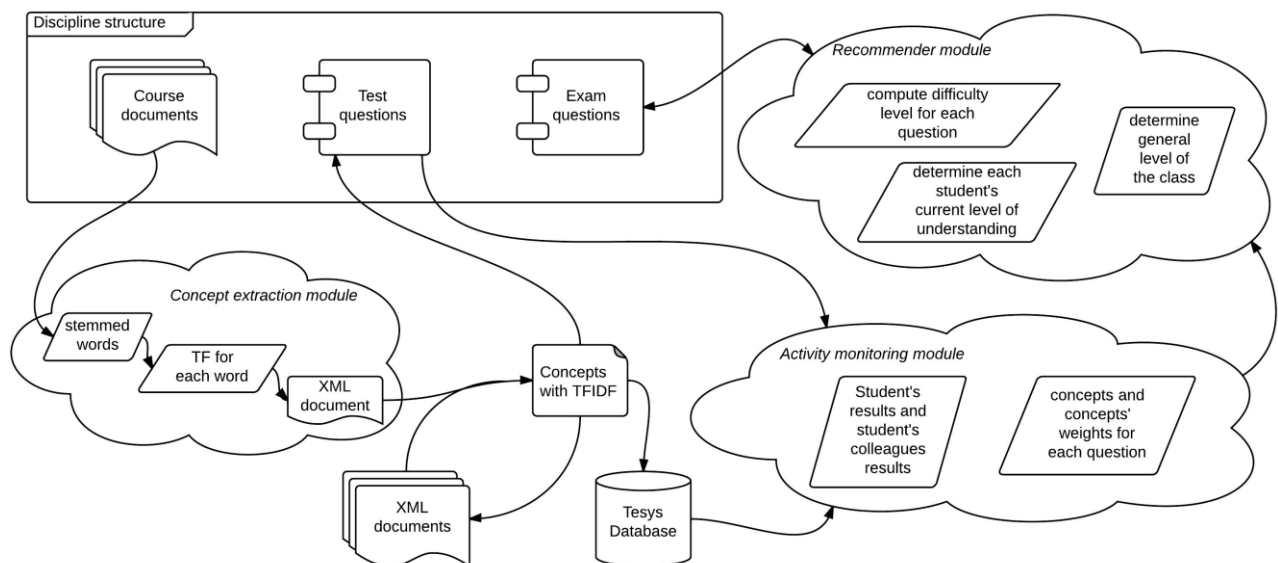

Figure 1: System architecture.

```
<fapt originalForm="fapt" tf="0.74"
idf="0.47"/>
 <atribut originalForm="atribut"
tf="0.45" idf="0.47"/>
 <sistem originalForm="sistem"
tf="0.33" idf="0.47"/>
 <inferent originalForm="inferenta"
tf="0.32" idf="0.55"/>
<algoritm originalForm="algoritm"
tf="0.32" idf="0.55"/>
         . . . . . . .
 <absolut originalForm="absolut"
tf="0.01" idf="1.25"/>
</words>
```

Each concept extracted from the course chapter's document is represented as an element in the xml file. The stemmed form of the word is stored as the element name, and its original form, TF value and IDF value are stored in the element attributes.

The first five concepts that have the highest TF-IDF value are inserted in the database, for further use on the platform.

Figure 2 illustrates part of the interface available to the professor for managing the concepts. It is very straight forward, providing the professor with the list of extracted concepts and some additional options for managing them. These options include: the possibility to add new concepts, modify the existing ones in case they were not correctly extracted and delete the irrelevant concepts, if any.
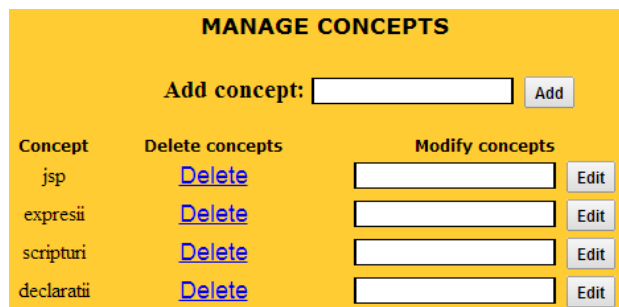

Figure 2: Concepts management.

4.3   Discipline structure

Within the online platform a discipline has the following structure: Chapters, Test Questions, Exam Questions and Concepts.

The chapters are documents uploaded by the professors, which can have one of the following extensions: *.pdf*, *.doc*, *.docx*. These documents are parsed and stemmed, resulting in a list of concepts.

Test questions are the questions used by the students throughout the semester for evaluating their current knowledge. A feature that allows the professor to choose from a list the concepts that are related to the test question and assign them weights was added to the e-learning platform.

The exam questions are the ones from which the students will take the final exam and obtain their final grades.

The concepts are the ones extracted from the chapters' documents which were previously reviewed by the professor.

As presented in Figure 3, for each question it is available the list of concepts extracted from the chapter to which the question belongs. Here is where the professor has the ability to assign the corresponding weights, representing the level of relevance that the concept has to the question.
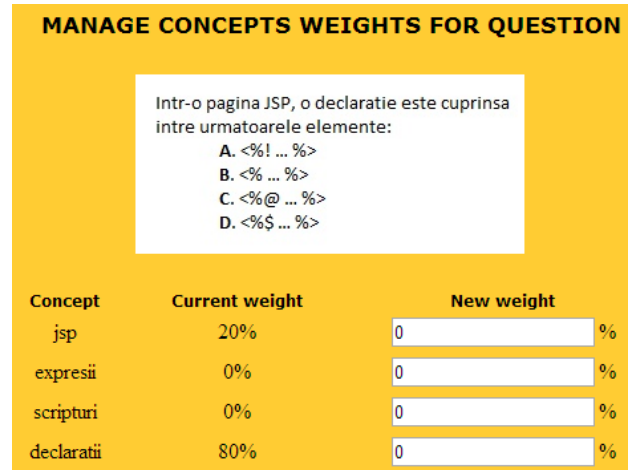

Figure 3: Weights management.

4.4   Activity monitoring

This step is very important because it is decisive for the determination of the student's ability to understand the course material and to highlight his/her progress. The most relevant information can be obtained by evaluating the correctness of the answers from the test questions, taking into consideration the concept associated with those questions and their given weights. This will help the professor figure out which are the concepts that the student has difficulties understanding and how he/she can be helped.

One of the monitoring tools provided to the professor is the graphic presented in Figure 4. By regularly checking this table, the professor can watch the progress of his/her students, be informed of their level of interest in the course material and discover which are the concepts that pose them problems.

4.5   Usability assessment and cognitonics

Usability evaluation is the final and most critical step within the lifecycle of the application, since it may have tremendous implications on the redesign of the user interface and underlying business logic. Applying general heuristics (i.e. with no special tuning to educational context) may be a reasonable option but using approaches that are adapted to e-Learning may offer greater progress [23].

From cognitive perspective, the visual stimuli refer to the following processes: visual search, find, identify,

	First/Last name	Overall understanding level ▾	No. answered questions	Concept 1: jsp	Concept 2: expresii	Concept 3: scripturi	Concept 4: declaratii
1	Student18 Student18	78.37	5	100	73.33	100	100
2	Student17 Student17	77.17	10	100	80	100	100
3	Student14 Student14	56.13	10	100	63.33	100	66.67
4	Student1 Student1	53.7	5	0	60	0	66.67

Figure 4: Activity monitoring.

recognize and memory search [24]. Based on this kind of analysis, specific issues may be obtained specific issues within our e-learning tool. From this perspective, presented educational assets (e.g. quizzes, concepts, etc.) can be characterized as confusing, not findable, etc. and thus reduce the unwanted side effects regarding the distortions in perception of the world caused by information society and globalization.

5   Experimental results

In order to better explain how the system works, we will consider a sample usage scenario. The main steps of the scenario are:

Professor – Concept setup

Let's assume that professor P is a professor on the e-learning platform, and has a course with two chapters. He first has to upload the documents of the chapters. Immediately after he does that, the system parses the documents, applies the stemming algorithm and the TF-IDF formulas and ultimately extracts the most important five key concepts from each file: C11-C15, C21-C25.

This list can be accessed from the course page and is differentiated by the chapter to which they belong. In this moment the professor can review the concepts, he/she can delete the ones that he/she considers irrelevant, or maybe add new ones.

Professor – Test questions setup

This step is performed when professor P loads test questions for the students to answer, and for each question assigns weights of the extracted concepts, denoting the relevance level of every concept for each particular question. The weights have values in the range of [0.0,1.0].

Table 1 presents a possible weight distribution for concepts among questions. The cells corresponding to the concepts that have absolutely no relevance to a question and therefore have a weight of 0.0 are left blank.

	C11	C12	...	C21	C22	C23	C24	C25
Q1		0.1		0.7			0.2	
Q2			0.4		0.6			
Q3		0.7			0.1	0.1		0.1
...								

Table 1: Sample weights distribution.

On the platform, the professor is able to assign the weights as percentages, as previously presented in Figure 3. These weights can be updated at any time, and the progress of the students will be modified accordingly.

Student – Take tests

Let us consider student S1. The first test the student takes contains questions 1, 2, 3, 5, 7. Table 2 presents possible values for the correctness of the answers given by the students that answered these questions.

S1 is the analyzed student, S2 to Sn are the other students that answered the questions.

It is assumed that the tests contain only single choice questions, so the answer can be only evaluated as CORRECT or INCORRECT.

	Q1	Q2	Q3	Q5	Q7
S1	CORRECT	CORRECT	INCORRECT	CORRECT	INCORRECT
S2	INCORRECT	CORRECT	INCORRECT	INCORRECT	CORRECT
S3	CORRECT	CORRECT	INCORRECT	CORRECT	CORRECT
			...		
Sn	CORRECT	CORRECT	CORRECT	CORRECT	INCORRECT

Table 2: Sample Answer Data.

After computing the weights and results, the system will provide the professor with the following statistics:

* student's level of understanding of each concept:

$$LU_c = \frac{\sum_{i=1}^{n-p} w(c, q_i)}{\sum_{j=1}^{n} w(c, q_j)}$$

In this formula, LU_c represents the level of understanding of concept c and $w(c, q_i)$ is the weight associated to the concept for question q_i. The numerator of the fraction is therefore the sum of the weights of the concept for the correctly resolved questions, and the denominator is the maximum amount that could be obtained if the responses to all questions were correct.

- student's performance relative to his colleagues;

$$Performance = \sum_{i=1}^{k} \frac{LU c_i}{Avg\,(LU c_i)}$$

- the difficulty level of every test question

$$Difficulty = \frac{No.\,correct\;answers}{Total\;no.\,answers}$$

- the difficulty level of each concept

S1. Professor – Visualize results : build mental model

After analyzing the presented data, the professor will be able to start creating a mental model regarding the student's current level of understanding of the material and his/her place among the other students. Also, if necessary, the professor might decide to modify the course material, for example add some extra information on a particular concept that the students have trouble understanding.

Another action the professor might choose to take, given the reported level of the class, is increase or decrease the general difficulty level for the test questions, as well as deciding which will be the best exam questions.

6   Conclusions and future works

This paper presents a use case of building a tool for Tesys e-Learning platform and analyzing cognitonics and HCI related issues in an attempt to offer a high quality interaction design that minimizes the cognitive side effects.

The developed tool is presented in detail from architectural and technical point of view, with an emphasis on the design of the user interface and on the data processing issues.

The technical challenges that are addressed in this paper regard building a Romanian stemmer, obtaining concepts, designing mathematical formulas for determining concept and quiz weights and overall students' knowledge levels. From this point of view, the

future works regard validation of these mathematical formulas and possibly inferring better ones. As a general approach, continuous usage of the tool will provide data evidence for our approach.

Another important issue, discussed in this paper, regards the HCI and cognitonics aspects of the user interface designed for this software tool. From this perspective, usability evaluation (general or e-learning related) using HCI specific methodologies represents the final step in obtaining a high quality user interface. From cognitive perspective, the goal is to minimize (or ideally eliminate) the distortions in perception of the world cost by the developed tool. The cognitive aspects validate the student model that is mentally built by the professor while using the tool.

As future works, there are two main directions. One regards properly the analysis of the underlying data (e.g. concepts, weights, formulas) and the other one regards further analysis from HCI and cognitonics perspective of the developed tool. Once progress is made in these directions, similar e-learning tools may also be analyzed providing a framework for progress in this domain.

References

[1]   B. Mehlenbacher, L. Bennett, T. Bird, M. Ivey, J. Lucas, J. Morton, L. Whitman. Usable E-Learning: A Conceptual Model for Evaluation and Design. *Proceedings of HCI International 2005: 11th International Conference on Human-Computer Interaction*. Las Vegas, NV: Mira Digital P, 1-10.

[2]   K. Nakata, A. Voss, M. Juhnke, T. Kreifelts. Collaborative Concept Extraction from Documents (1998). *Proceedings of the 2nd Int. Conf. on Practical Aspects of Knowledge management (PAKM 98)*. 1998. Switzerland.

[3]   J. Villalon, R. A. Calvo. Concept Extraction from Student Essays, Towards Concept Map Mining. *Proceedings of the 2009 Ninth IEEE International Conference on Advanced Learning Technologies*, Riga, Latvia, 2009, Volume 0: 221-225.

[4]   T. Madhyastha, E. Hunt. Mining Diagnostic Assessment Data for Concept Similarity. Journal of Educational Data Mining (JEDM). pp. 72-91. 2010.

[5]   A. Merceron, K. Yacef. Educational Data Mining: a Case Study. *Proceedings of the 12th international Conference on Artificial Intelligence in Education AIED*. pp. 467-474. Amsterdam, The Netherlands. 2005.

[6]   R.S.J.D. Baker, K. Yacef. The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*. pp. 3-17. 2009.

[7]   E. Ayers, R. Nugent, N. Dean. A Comparison of Student Skill Knowledge Estimates. *EDM2009: 2nd International Conference on Educational Data Mining*. Cordoba, Spain. 2009.

[8]   D. D. Burdescu, M. C. Mihaescu, TESYS: e-Learning Application Built on a Web Platform. *ICE-B*. pp. 315-318. 2006.

[9] Colhon, M.: Language Engineering for Syntactic Knowledge Transfer. Computer Science and Information Systems, Vol. 9, No. 3, 1231-1248. (2012)

[10] B. Litchfield. Making PDFs Portable: Integrating PDF and Java Technology. *Java Developer's Journal*. 2005.

[11] D. Cristea, C. Forăscu, "Linguistic Resources and Technologies for Romanian Language", Computer Science Journal of Moldova, vol. 14, no. 1(40). (2006)

[12] D. Klein, C. D. Manning, "Accurate Unlexicalized Parsing", In: Proceedings of the 41st Meeting of the Association for Computational Linguistics, pp. 423-430. (2003)

[13] M. P. Marcus, B. Santorini and M. A. Marcinkiewicz, "Building a Large Annotated Corpus of English: The Penn Treebank". In COMPUTATIONAL LINGUISTICS, vol. 19(2), 313-330. (1993)

[14] Apache POI - the Java API for Microsoft Documents. *http://poi.apache.org/*

[15] Snowball. *http://snowball.tartarus.org/*

[16] Deepika Sharma. Stemming Algorithms: A Comparative Study and their Analysis. *International Journal of Applied Information Systems*. pp. 7-12. 2012.

[17] PTStemmer - A Stemming toolkit for the Portuguese language. *https://code.google.com/p/ptstemmer/*

[18] Fomichov, V. A., Fomichova, O. S. Cognitonics as a New Science and Its Significance for Informatics and Information Society. *Informatica. An International Journal of Computing and Informatics (Slovenia)*, 2006, Vol. 30, No. 4, pp. 387–398; http://www.informatica.si/vol30.htm#No4; retrieved 14.06.2014.

[19] Bohanec, M., Gams, M., Mladenic, D. et al, Eds. (2011). Proceedings of the 14th International Multiconference Information Society – IS 2011, Vol. A, Slovenia, Ljubljana, 10 – 14 October 2011. The Conference Kognitonika/Cognitonics. Jozef Stefan Institute; , http://is.ijs.si/is/is2011/zborniki.asp?lang=eng; pp. 347-430; retrieved 15.05.2014

[20] Gams, M., Piltaver, R., Mladenic, D. et al., Eds. (2013). Proceedings of the 16th International Multiconference Information Society – IS 2013, Slovenia, Ljubljana, 7 – 11 October 2013. The Conference Kognitonika/Cognitonics. Jozef Stefan Institute; http://is.ijs.si/is/is2013/zborniki.asp?lang=eng; pp. 403-482; retrieved 15.05.2014.

[21] D. Norman, S.W. Draper (eds.): User Centered System Design. Earlbaum, Hillsdale (1986)

[22] A. Granić, V. Glavinić. Automatic Adaptation of User Interfaces for Computerized Educational Systems. In: Zabalawi, I. (ed.) Proceedings of the 10th IEEE International Conference on Electronics, Circuits and Systems (ICECS 2003), Sharjah, Dubai, pp. 1232–1235 (2003)

[23] D. Squires, J. Preece. Predicting quality in educational software: Evaluating for learning, usability and the synergy between them. Interacting with Computers 11, 467–483 (1999)

[24] Shu-mei Zhang, Qin-chuan Zhan, He-min Du. Research on the Human Computer Interaction of E-learning. International Conference on Artificial Intelligence and Education (ICAIE), 2010.

[25] Jakob Nielsen and Rolf Molich, Heuristic evaluation of user interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90), Jane Carrasco Chew and John Whiteside (Eds.). ACM, New York, NY, USA, 249-256, 1990.