

Enhanced Approach of Automatic Creation of Test Items to Foster Modern Learning Setting

Christian Gütl^{1, 2}, Klaus Lankmayr¹ and Joachim Weinhofer¹

¹Institute for Information Systems and New Media (IICM), TU-Graz, Austria

²Curtin University, Perth, Australia

Christian.Guetl@iicm.tu-graz.ac.at

lanki@sbox.tugraz.at

jowein@sbox.tugraz.at

Abstract: Assessment has to be seen as an integrated and important activity in the learning process. In particular modern educational approaches - such as self-directed or exemplary learning - and personalized learning activities cause a tremendous effort or make it even impossible to prepare appropriate and individualized test items, assess them and provide feedback. This situation has motivated the Advanced Educational Media Technologies (AEMT) Group at Graz University of Technology to initiate a research program on e-assessment to cover the entire life cycle of the assessment process by semi-automated and automated approaches. In this paper we focus on the automated process to create different types of test items out of textual learning content, more precisely to create single choice, multiple-choice, completion exercises and open ended questions. Started in previous research activities by applying statistic approaches to summarize learning content and to identify most important words, our most recent approach applies a combination of statistic and natural language processing as well as semantic methods to identify most important concepts on an abstracted level. In our enhanced approach, identified concepts and differently related concepts represent the underpinning input for the test item creation. The implemented prototype can process learning content stored in various file formats, extracts most important content and related concepts, creates different types of test items and reference answers, and supports online assessments as well as exports the test items in QTI format. In this paper we cover in particular the following aspects: the motivation for automated test item creation, related work, requirement and design of the enhanced tool, implementation and usage viewpoints. Furthermore, we outline a study on the evaluation of the prototype, which suggests that the quality of extracted concepts and created test items is comparable with such ones provided by humans.

Keywords: e-assessment, automated test item creation, distance learning, self-directed learning, natural language processing, computer-based assessment

1. Introduction

Our knowledge-driven and globalized world requires continuous adaptation of knowledge and skills from members of the society. Life-long learning is the key in such an environment and new pedagogical approaches such as exemplary-based learning and self-directed learning are becoming increasingly popular. (Gütl, 2010) It is also commonly agreed that assessment must be an integrative task within educational processes in order to provide feedback to students and teachers as well and to gain information to adapt the learning task (Bransford, Brown, & Cocking 2000). However, the creation of appropriate test items is a time consuming task, in particular to assess content alternatives and different knowledge levels in adaptive eLearning environments. Moreover, in self-directed learning settings or more general in life-long learning settings it is almost impossible to provide prepared test items but it is also challenging to assess student assignments and provide feedback. (Gütl, 2008) This situation has motivated Advanced Educational Media Technologies (AEMT) Group at Graz University of Technology to initiate a research program on e-assessment to cover the entire life cycle of the assessment process by semi-automated and automated approaches.

In this paper we want to focus on the automatic and semiautomatic test item creation from text-based learning content. To this end, the paper is structured as followed: first we will give background information and related work on both the concept extraction and automatic test item creation. This is followed by requirements, design and development of the introduced tool, the *Automatic Question Creator (AQC)*. A discussion from the users' viewpoint as well as an initial qualitative study of the quality of the extracted concepts and created test items give first insights of the practical usage.

2. Background and related work

In the processes of the automated creation of assessment items, one of the important tasks is to identify and extract the main concepts of natural language texts of the learning content. *Concept extraction* is a major task in natural language processing and is thoroughly examined in past and present, such as in (Moens & Angheluta, 2003; Villalon & Calvo, 2009). The fundamental ideas referring to

concept extraction are based on the findings of Luhn in the late 1950s who detected a relatedness of word frequencies to the main contents of a text. In the late 1970s Edmundson improved this method by combining cue phrases, word frequencies, title words and the position of words in a paragraph. Kupiec, Pederson and Chen (1995) extended this method by considering acronyms and proper nouns additionally. Gelfand, Wulfekuhler and Punch (1998) used semantic relationship graphs which depict semantic dependencies of words recursively. Frank et al. (1999) created a domain-specific key phrase extraction (KEA) that uses a Naive Bayes classification depending on word frequency and the position of the first occurrence of the word. KEA was extended by Turney (2003) who enhanced the algorithm by co-occurrences which consider the customariness of two words together in the WWW. Song, Han and Rim (2004) generated lexical chains and a concept score depending on word association, the depth in WordNet hierarchy and a semantic relation weight. Hassan, Mihalcea and Banea (2007) use a text rank algorithm that takes account of the context of a word by transforming the document into a graph and calculating node weights. Ledeneva, Gelbukh and Garcia-Hernández (2008) evaluate n-grams, consisting of n words, instead of single words to determine the importance of concepts. A more detailed discussion of methods and approaches can be found elsewhere, such as at (Liu & Yang, 2009; Hovy, Kozareva & Riloff, 2009).

By further focusing on research of *automated test item creation*, an extensive literature review has shown just few pre-existing approaches and tools (Gütl, 2008; Lankmayr, 2010). Coniam (1997) developed a method for automatic multiple choice item creation for filling-in blank test items. There are two distinctive ways to identify the fill-in blank area: (a) a user defined n-th word deletion depending on a predefined entry point, and (b) the part of speech tag. Moreover several distractors are extracted from a list derived from the Bank of England Corpus whereby these words have similar word frequencies in that corpus as the selected word. Fischer and Steinmetz (2000) are using special meta information, like relations between key phrases that are contained in the Multibook system to generate test items with the help of predefined templates. The approach from Mitkov and Ha (2003) to create multiple choice questions uses WordNet to calculate distractors for key terms. The questions are built by a rule based transforming of sentences into interrogative clauses. Brown, Frishkoff and Eskenazi (2005) developed the REAP system that is able to provide users texts suitable for their reading levels and to generate appropriate multiple choice and assignment items. By help of WordNet using definitions, synonyms, antonyms, hypernyms and hyponyms question items are formed to improve word knowledge by evaluating user statistics. An approach based on machine learning was introduced by Hoshino and Nakagawa (2005). Thereby k-Nearest neighborhood, naïve Bayes classification and a suitable training set are utilized to identify the positions of the blanks in news articles for creating multiple choice items. Chen, Liou and Chang (2006) have built grammar tests by transforming sentences extracted from the WWW. The transformation is done by applying manually generated patterns and is used for creating multiple choice items and error detection tests. Rus, Cai and Graesser (2007) introduced methods for generating questions with the help of patterns, templates and a special markup language named QG-ML. The patterns are characterized by semantic, lexical and syntactical structures whereas the templates describe methods to implement these structures to generate questions. Gütl (2008) described a system that uses the summary of a document to identify key concepts (named entities) and that generates completion tests as well as limited choice items.

The evaluation of the current state of research suggests that approaches using machine learning are strongly depending on the training set and the knowledge domain. Most of the illustrated systems are applying either statistical or semantic methods and are not able to fulfill the requests given by the variety of assessment item types. Moreover, pre-existing approaches and tools are not sufficiently flexible and extendable to support the above mentioned variety of application scenarios and learning settings. For this reason we developed a system, the *Automatic Question Creator (AQC)*, which builds on a combination of statistical, semantic and structural analysis to accomplish a step-by-step extraction of relevant concepts from natural language texts.

3. Requirements, design and development

3.1 Objectives and high level requirements

The goal of the Automatic Question Creator (AQC) is to provide a tool which supports the creation of text items or even generates them automatically from the learning content. A flexible design should enable various groups to use the tool stand-alone or to integrate it in a learning platform as well as adjust the tool according to the specific learning setting. This has led us to specify to following requirements on an abstract level:

- Support of various input file formats from local file systems and from internet resources
- Domain knowledge and document structure independency
- Identification of most important concepts
- Creation of test items and reference answers based on identified concepts
- Support of open ended, single choice, multiple-choice and completion exercises
- Variability, configurability, modularity, extensibility and performance
- interoperability with existing eLearning systems

3.2 Conceptual architecture and tools

The high-level conceptual design of the AQC which is outlined in Fig. 1; it illustrates core conceptual units and pre-existing tools as well. The system can be unfold into three main modules: (1) The *Pre-processing* module deals with format conversion, text cleaning methods and transformation into an internal XML schema which contains all necessary data for further processing. (2) The *Concept Extraction* module performs structural, statistical and semantic analysis, runs term weighting and finally extracts the most suitable phrases; a detailed description is given in Section 3.3. (3) The *Assessment Creation* module determines the most appropriate sentence for each phrase and adds the previous and the following sentences to provide sufficient context information. Moreover the module identifies distractors and antonyms, creates question items and reference answers, and finally transforms those items in QTI standard.

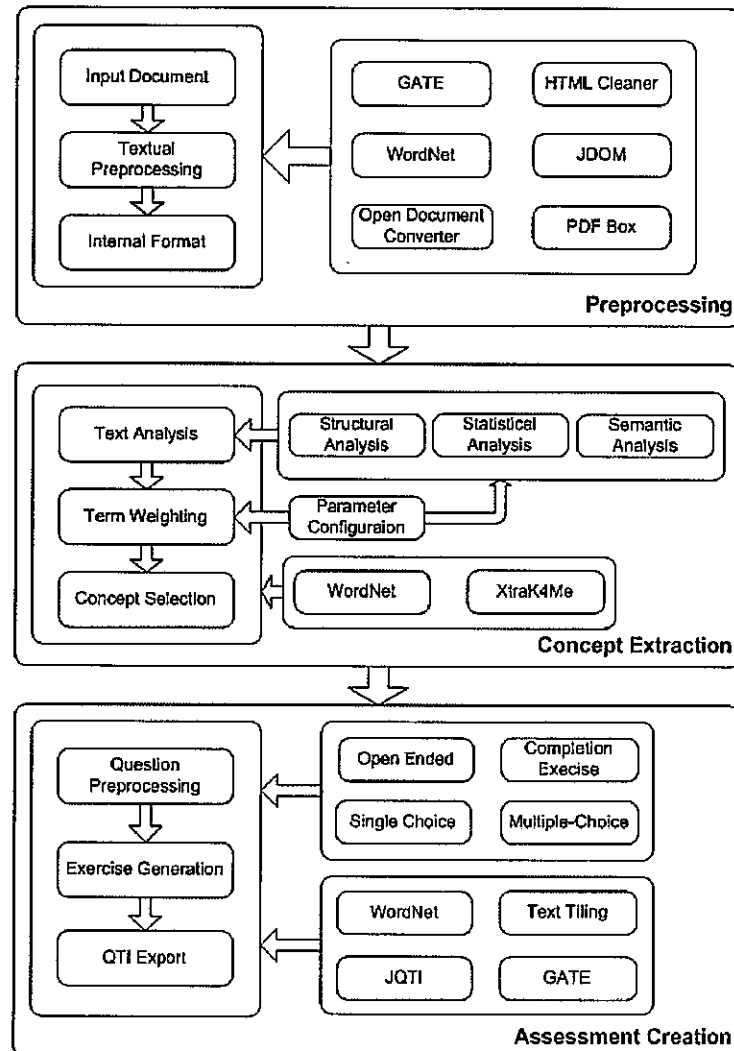


Figure 1: Conceptual design of automatic question creator

The two main components integrated in the implemented system are GATE and WordNet. The GATE framework, especially the ANNIE plug-in, is used for basic text processing and annotation. Thereby the text is split up into tokens and sentences, the part of speech classification as well as name entity recognition, noun chunking and co-reference resolution of each token is performed. (GATE, 2010) The semantic analysis is processed with WordNet whereby semantic and lexical relations between words are calculated as well as distractors and antonyms are selected. (WordNet, 2010) Format conversion for Word, Open Document Text and HTML is utilized to transform the input files into a HTML format by using JODConverter (2010). PDF files are transformed with the help of PDFBox (2010) that is able to extract the textual information from such files, structural information is added manually by applying predefined patterns. Content of the WWW, such as Wikipedia, is also supported as input source by the Automatic Question Creator. Afterwards the generated HTML file is cleaned up using HTML Cleaner (2006) to ensure a conversion to XML with JDOM (2010). The concept extraction done by the AQC is assisted by XtraK4Me of Schutz (2008). QTI exportation and rendering is done with JQTI (2008).

3.3 Data structure and applied methods

The main idea of our enhanced approach is to combine statistical, semantic and structural analysis to find most relevant words in learning content or respectively concepts suitable for creating test and exercises. Based on general word frequencies of the stemmed text the AQC transforms those frequencies into weights for each word. In a second step in the process chain, these weights are adapted by a configurable set of algorithms that evaluate dependencies of the words according to the appearance in the text, such as in title, abstract, keywords, headlines. Also structure and formatting style as well as word types are considered in the process. Depending on the set of the highest weights and further configurable parameters the AQC generates single choice, multiple-choice, completion exercises and open ended questions. Moreover the system is capable of exporting the test items including reference answers into the QTI format to allow integration into other learning and assessment systems.

In order to support the process chain, an internal data structure is applied which is organized into three main elements as illustrated in Figure 2. A *Word Element* contains all necessary textual and structural information of each token retrieved from GATE, WordNet and format conversion as well as from statistical and semantic analysis. Each token is also associated with a *Weight Element* that stores a weight of each algorithm performed for concept extraction. The *Sentence Element* is calculated for each sentence in the text and contains the sentence boundaries, the related concepts and the sentence weight.

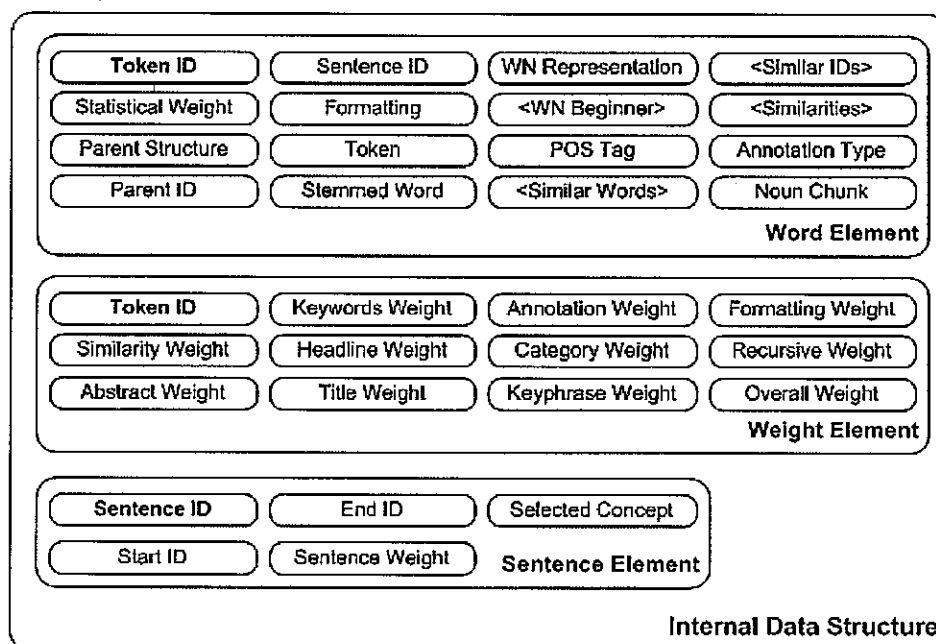


Figure 2: Internal data representation

The overall weight of words is composed of its statistical weight based on word occurrence w_1 (see Table 1, line 1) and several other weights w_i (see Table 1, line 2 - 11) that are retrieved by applying statistic, semantic and structural analysis. Most of these methods are subject to the distance of words in WordNet hierarchy. The influence of each those weights on the overall weight can be adjusted by a set of independent parameters $k_{m,i}$. Our first approach for the calculation of the overall weight $w(i)$ for a word i is shown in equation (1), further experiments and improvement are subject to future work. To ensure stop word elimination only nouns and verbs are considered. A more detailed description of the weighting process and the applied methods can be found at Weinhofer (2010).

Table 1: Algorithms, weights and configurable parameters

Module I_n	Weight	# Adjustable Parameters k_m	Description
1	$w_{stat}(i)$	1	statistical weight, normalized number of occurrences of a stemmed word in a section
2	$w_{sim}(i)$	2	weight derived from statistical weights of similar words, depends on similarity measures retrieved from WordNet
3	$w_{title}(i)$	1	semantic relation to the words in the title
4	$w_{headline}(i)$	6	semantic relation to the words in the corresponding headline depending on the headline layer (up to 6)
5	$w_{abstract}(i)$	1	semantic relation to the words in the abstract
6	$w_{keywords}(i)$	1	semantic relation to keywords
7	$w_{annotation}(i)$	17	weight for the special annotations retrieved by GATE, the 17 annotation types can be handled individually
8	$w_{category}(i)$	25	weight according to the 25 unique beginners retrieved from WordNet
9	$w_{formatting}(i)$	1	Weight depending on the text formatting
10	$w_{keyphrase}(i)$	1	weight for phrases supplied form XtraK4Me algorithm
11	$w_{recursive}(i)$	2	recursive similarity weight calculation, consideration of lexical chains

In a further step, for each noun which is above a predefined threshold, a set of phrases that contain this word is built for each of the sections. Then all phrases of each set are weighted by summing up the overall weights of all words contained in a phrase. The highest weighted phrase of each set is chosen as potential concept. Finally the concept extraction is accomplished by building a collection of the best of these concepts for each section of the text.

$$w(i) = w_{stat}(i) * \left(k_1 + \sum_{l=2}^{11} [(w_l + 1) \sum_m k_{m,l}] \right) \quad (1)$$

For *Completion Exercises* the previous and following sentences are added to the selected sentence to offer additional context information to the user. In all of those sentences the selected concepts get replaced with fill-in blank areas to avoid unnecessary hints. *Multiple-choice item* also requires distractor calculation. Basically the distractors are determined by searching coordinate terms for the whole question phrase in WordNet. If this calculation fails, the phrase gets split in all possible coherent n-grams and the coordinate terms for the longest sequence are randomly selected. In the worst case only a single word of a concept delivers suitable results. Due to the circumstance that there are very few proper nouns and no dates included in WordNet, a special case appears if the concept is assigned to a special annotation type. In this case three random phrases sharing the same annotation type are chosen as distractors from the underlying document. *Single choice items* can be generated by searching antonyms for single words in a concept and replacing the original word. Since the result of this procedure is seldom satisfying, the same method is repeated with all adjectives, verbs and nouns of the whole sentence. *Open ended exercises* are generated using several patterns depending on the special annotation type in the selected concept. Due to the fact of implementing a fully automatic assessment system the difficulty according to open ended questions is to compute a reference answer automatically. To meet this challenge the AQC uses the text tiling algorithm to find the most proper text block containing the extracted concept.

The created test items are finally transformed into the QTI standard as single XML file for each question item. The reason for that exportation is to afford an opportunity of integrating the generated test

items in learning management systems or other assessment tools. Moreover an implementation of a web service to access the extracted phrases and the created test items is planned.

4. Usage viewpoint and first findings

This section outlines AQC from the user's point of view which is restricted in this first version to semi-automatic test item creation. The process steps are as follows: First, an input file in one of the supported formats has to be selected either from the local file system or from an Internet resource. The text is converted and filtered as well as a control output is generated. In the next step the user can induce the annotation process and the internal data structure is built. The result of the annotation is shown and the user can initiate the weighting process for concept identification. Figure 3 illustrates an example of a weighted text and the calculated weighting factors of a token which results from selected methods. In this step the user can initially set or change the weighting factors of the methods or even select and unselect methods to be applied (see Figure 4).

Figure 3: Annotated and weighted text

Figure 4: AQC configuration panel

The next step in the process chain is the selection of the most important concepts to finally create the test items. Figure 5 illustrates a sample of concept extraction whereby the highest weighted phrase for each section of the content is listed on the screen. The user is enabled to deselect unwanted phrases as well as add unconsidered phrases or single words in a chapter of the text. Based on the final settings the test item creation is initiated. Different types of test items can be selected and instances of created items can be viewed. An example of a generated multiple-choice test item is outlined in Figure 6 that shows the representation of the QTI item in HTML.

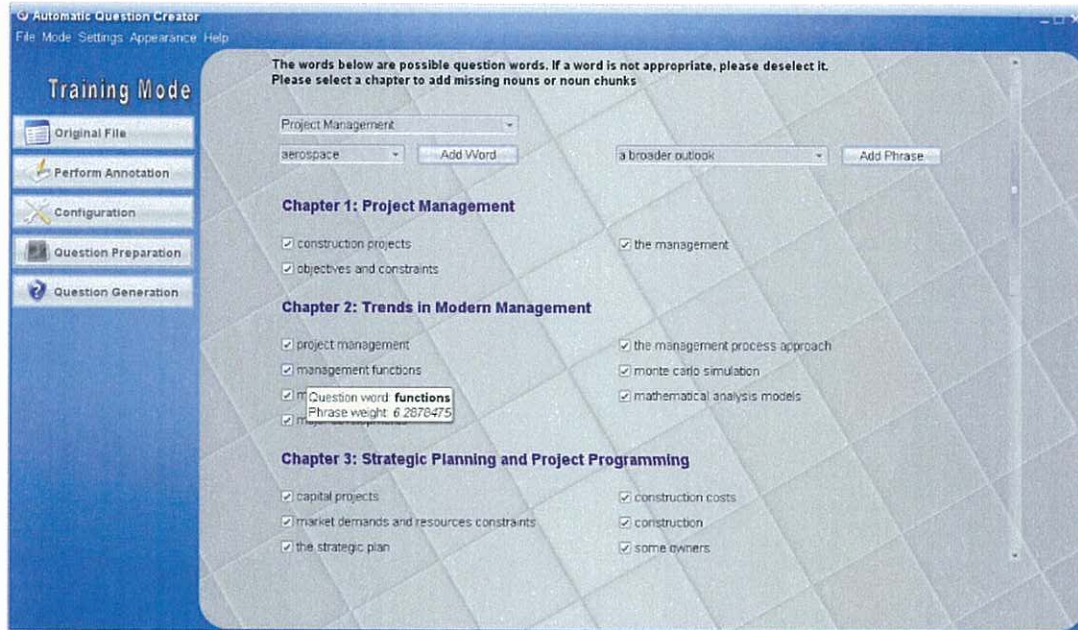


Figure 5: Example of extracted concepts

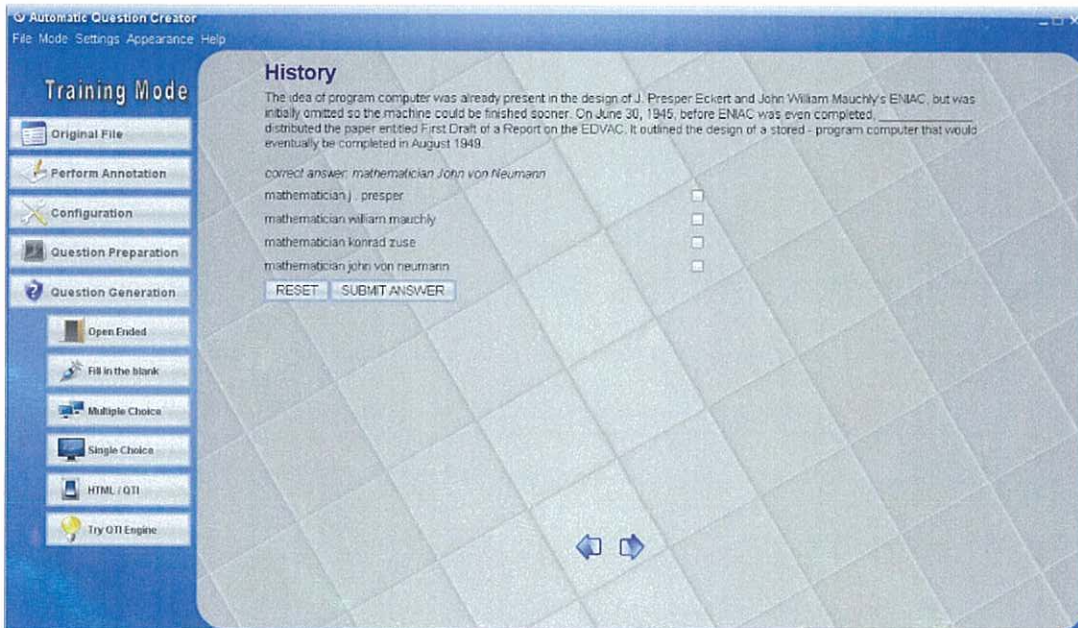


Figure 6: Example of a multiple-choice item

To verify the implemented system, especially the quality of the extracted concepts and the created test items, an initial qualitative user study has been administered. Therefore learning content of 'project management in construction' from MIT OpenCourseWare has been selected (Hendrickson, 2000). Five subjects (one college student and four students at university level) were asked to follow the evaluation procedure, see also (Lankmayr, 2010):

- Phase 1: Read carefully the text

- Phase 2: Identify the five most expressive concepts of each section
- Phase 3: Construct two open ended, single choice, multiple-choice and completion exercises for each section
- Phase 4: Evaluate concepts and text items for each of the four item types
Subjects were presented two manually and two automatically created test items (in arbitrary order) for each question type and sections.

The quality measure for assessing test items was derived from the observation matrix of Canella, Ciancimino and Campos (2010). The observation matrix originally consists of the pertinence, the level, the terminology and the interdisciplinarity criteria for evaluating test items created by students. In our context the terminology and the interdisciplinarity are not appropriate due to the usage of patterns and the focus on specific topics. Therefore we adapted the procedure towards the following criteria to evaluate the quality of the AQC:

- Concept: relevance of a concept in a section (for all question types)
- Pertinence: relevancy of a question in the given context (for all question types)
- Level: level of difficulty of a question (for all question types)
- Distractor: quality of the listed distractors (for multiple-choice items)
- Answer: quality of the reference answer (for open ended items)

Figure 7 shows the comparison between the percentage of relevant and irrelevant concepts extracted by the AQC for each chapter and the whole document. To gather chapters the document is divided by headlines of the highest level. A concept was counted as relevant if the average value of subjects' relevance appraisal was above 50 percent. As illustrated by Figure 7 the performance of concept extraction is satisfying with 70 % for the whole documents and between 60 and 80 % on the chapter level.

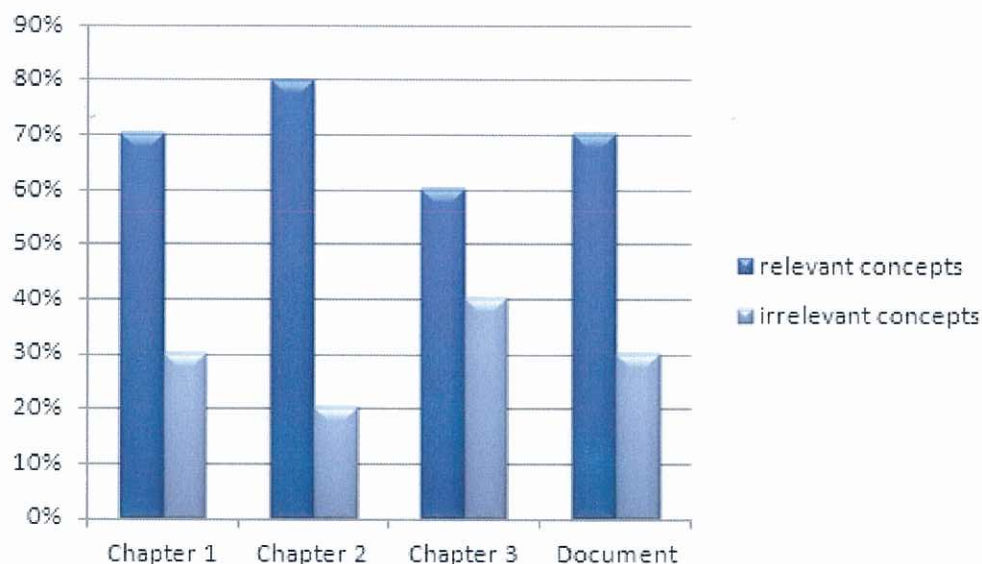


Figure 7: Evaluation of extracted concepts

Figure 8 shows the comparison of manual and automated test item creation and depicts the mean values and standard deviations of four performance indicators as described above. Subjects rate both manual and automated created test items quite good in terms of the relevance (pertinence) and the appropriateness of the provided reference answer as well as the quality of answer options of multiple-choice items (distractors). The degree of difficulty (level) was rated significantly lower. The average time consumption for accomplishing the evaluation was about 150 minutes. In contrast the test item creation process by the AQC has finished in about 5 minutes. Our preliminary results suggest that subjects could not distinguish between manually and automatically created test items as well as the effort to create test items can be reduced significantly.

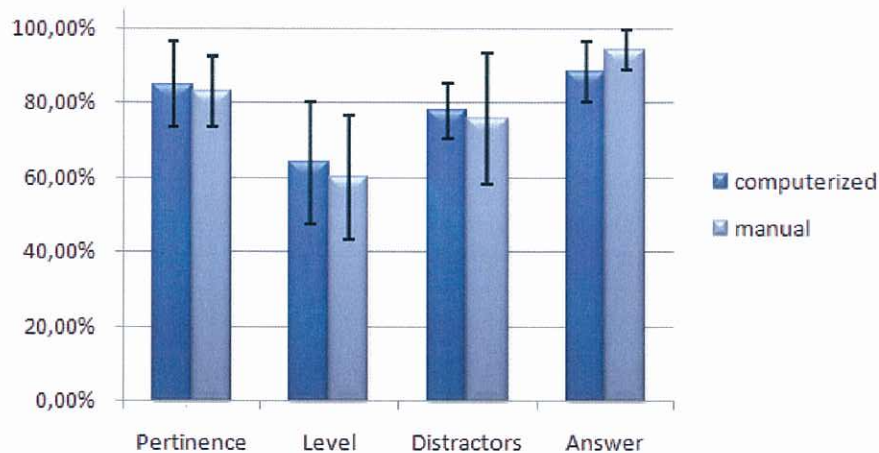


Figure 8: Comparison of manual and automatic created questions

5. Conclusions and future work

Assessment has to be seen as an integrated and important activity in the learning process. In particular modern educational approaches - such as self-directed or exemplary learning - and personalized learning activities cause a tremendous effort or make it even impossible to prepare appropriate and individualized test items, assess them and provide feedback. To overcome this problem, we advocate a tool which automatically creates test items from learning content, administer knowledge assessment and provide feedback.

We have introduced a concept and prototype implementation, that is capable of handling various text formats and WWW resources, that annotates the corpus using GATE, that applies statistical, semantic and structural methods for identifying key concepts. Based on these concepts the Automatic Question Creator (AQC) generates open ended, single choice, multiple-choice and completion exercises and exports those into QTI items. The initial qualitative evaluation confirmed first promising results and showed the advantages of the system. Encountered problems include the high time complexity for text annotation and WordNet-based operations but also the selection of not relevant concepts due to lack of common sense knowledge and domain knowledge.

First findings suggest that the support to create or even the automated creation of test items offers major advantages for tutors as well as for learners in various learning settings, from face-to-face to blended to remote learning. The main benefit for tutors demonstrates economy of time even if only a certain percentage of the created questions can be used without further processing. For learners the AQC enables (semi-)automated self assessment. Also AQC can be incorporated in (adaptive) eLearning systems and create on the fly test items to keep track on the knowledge state of the learners.

Although the first results are promising, there is room for improvements, such as enhancing and expanding the annotation and the extraction process in the natural language processing chain. Moreover we would prefer using natural language generation for question creation and reference answer construction. To enrich the functionality and the usability of the AQC output we would also like to enable tutors to take more control during the process chain, such as deselect inappropriate concepts or annotate others from the text to be considered. In the next step we will enable the fully automatic creation of test items, develop a Web-based interface, implement interfaces for integration in learning management systems, support multiple language, and administer experiments on different approaches to weight calculation for words and concepts.

References

- Bransford, J.D., Brown, A.L., & Cocking, R.R. (Eds.) (2000) *"How People Learn: Brain, Mind, Experience, and School. Expanded Edition"*, Washington DC: National Academies Press.
- Brown, J.C., Frishkoff, G.A. and Eskenazi, M. (2005) "Automatic Question Generation for Vocabulary Assessment", *Proceedings of the Human Language Technology Conference on Empirical Methods in Natural Language Processing*: 6 – 8 October 2005, Vancouver, British Columbia, Canada, pp 819- 826.

- Canella, S., Ciancimino, E. and Campos, M.L. (2010) "Mixed e-Assessment: an application of the student-generated question technique", Paper read at IEEE International Conference EDUCON 2010, Madrid, Spain, April.
- Chen, C.Y., Liou, H.C. and Chang, J.S. (2006) "FAST: An Automatic Generation System for Grammar Tests", *Proceedings of the COLING/ACL on Interactive presentation sessions*, pp 1-4.
- Coniam, D. (1997) "A Preliminary Inquiry Into Using Corpus Word Frequency Data in the Automatic Generation of English Language Cloze Tests", *Calico Journal*, Vol 14, No. 2-4, pp 15-34.
- Fischer, S. and Steinmetz, R. (2000) "Automatic Creation of Exercises in Adaptive Hypermedia Learning Systems", *Proceedings of the eleventh ACM on Hyper-text and hypermedia*, pp 49 -55.
- Frank, E., Paynter, G.W., Witten, I.H., Gutwin, C. and Nevill-Manning, C.G. (1999) "Domain-specific Keyphrase Extraction", *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence IJCAI*, pp 668-673.
- GATE (2010). "GATE: general architecture for text engineering", [Online], The University of Sheffield, <http://gate.ac.uk>
- Gütl, C. (2008) "Automatic Limited-Choice and Completion Test Creation, Assessment and Feedback in modern Learning Processes", Paper read at LRN Conference 2008, Guatemala, February 12th – 16th.
- Gütl, C. (2010) "The Support of Virtual 3D Worlds for enhancing Collaboration in Learning Settings", in Francesca Pozzi and Donatella Persico (Eds.) *Techniques for Fostering Collaboration in Online Learning Communities: Theoretical and Practical Perspectives*, IGI Global, 2010 (accepted)
- Hassan, S., Mihalcea, R. and Banea, C. (2007) "Random-Walk Term Weighting for Improved Text Classification", *Semantic Computing, ICSC 2007*, pp 242–249.
- Hendrickson, C. (2000) "Project Management in Construction", [online], Carnegie Mellon University, <http://pmbbook.ce.cmu.edu/index.html>
- Hoshino, A. and Nakagawa, H. (2005) "A real-time multiple-choice question generation for language testing: a preliminary study", *Proceedings of the second workshop on Building Educational Applications Using NLP*, pp 17-20.
- Hovy, E., Kozareva, Z. and Riloff, E. (2009) "Towards Completeness in Concept Extraction and Classification", *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Vol. 2, pp 948-857.
- HTML Cleaner (2006) "HTML Cleaner Project", Sourceforge.net, <http://htmlcleaner.sourceforge.net>
- JDOM (2010) "JDOM", <http://www.jdom.org>
- JODConverter (2010) "JODConverter", Art of Solving, www.artofsolving.com/opensource/jodconverter
- JQTI (2008) "JQTI", University of Southampton, <http://jqti.qttools.org>
- Kupiec, J., Pederson, J. and Chen, F. (1995) "A trainable document summarizer", *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, pp 68-73.
- Lankmayr, K. (2010) "Design and Development of Concepts for Automatic Exercise Generation", [online], Graz University of Technology, <http://www.iicm.tu-graz.ac.at/thesis/MA%20Lankmayr.pdf>
- Ledeneva, Y., Gelbukh, A. and García-Hernández, R. A. (2008) "Terms Derived from Frequent Sequences for Extractive Text Summarization", *Computational Linguistics and Intelligent Text Processing, 9th International Conference, CICLing 2008, Haifa, Israel, February 2008, Proceedings*, pp 593-604, Springer Verlag, Heidelberg.
- Liu, N. and Yang, C.C. (2009) "Keyphrase Extraction for Labeling a Website Topic Hierarchy", *Proceedings of the 11th International Conference on Electronic Commerce, Taipei, Taiwan, 2009*, pp 81-88.
- Mitkov, R. and Ha, A. L. (2003) "Computer-Aided Generation of Multiple-Choice Tests", *Proceedings of the HLT-NAACL 2003 workshop on Building educational applications using natural language processing*, pp 17-22.
- Moens, F.M. and Angheluta, R. (2003) "Concept extraction from legal cases: the use of a statistic of a coincidence", *Proceedings of the 9th international conference of Artificial intelligence and law, Scotland, United Kingdom, 2003*, pp 142–146.
- PDFBox (2010) "Apache PDF Box, Java- PDF Library", The Apache Software Foundation, <http://pdfbox.apache.org/>
- Rus, V., Cai, Z. and Graesser, A. C. (2007) "Experiments on Generating Questions About Facts", *Computational Linguistics and Intelligent Text Processing, 8th International Conference, CICLing 2007, Mexiko City, Mexiko, February 18-24, 2007, Proceedings*, pp 444–455.
- Schutz, A. (2008) "SMiLE: XtraK4Me - Extraction of Keyphrases for Metadata Creation", SmlE: Semantic Information Systems and Language Engineering Group, <http://smile.deri.ie/projects/keyphrase-extraction>
- Song, Y.-I., Han, K.-S. and Rim, H.-C. (2004) "A Term Weighting Method based on Lexical Chain for Automatic Summarization", *Computational Linguistics and Intelligent Text Processing, 5th International Conference, CICLing 2004, Seoul, Korea, February 2004, Proceedings*, pp 636–639, Springer Verlag, Heidelberg.
- Turney, P.D. (2003) "Coherent Keyphrase Extraction via Web Mining", *Proceedings of IJCAI'03*, pp 434–439.
- Villalon, J. and Calvo, R.A. (2009) "Concept Extraction from student essays, towards Concept Map Mining", *Proceedings of the 9th IEEE International Conference on Advanced Learning Technologies, ICALT 2009*, pp 221-225.
- Weinhofer, J. (2010) "Extraction of relevant semantic data from natural language texts in the view of automatic question generation", [online], Graz University of Technology, <http://www.iicm.tu-graz.ac.at/thesis/MA%20Weinhofer.pdf>
- WordNet (2010). "WordNet: A lexical database for English", Princeton University, <http://wordnet.princeton.edu>