

A Proposed Framework for Generating Random Objective Exams using Paragraphs of Electronic Courses

Elsaeed E. Abdelrazek

Department of Computer Preparing Teacher
Dumyat University
Dumyat, Egypt

Abstract—Objective exams (OE) plays a major role in educational assessment as well as in electronic learning. The main problem in the traditional system of exams is a low quality of questions caused by some human factors, such as the traditional method for the development of the exam covers a narrow scope of curriculum topics. This does nothing for the separation of teaching process about the examination process. In this study we present a framework that generates three types of Objective exams questions (multiple choice questions (MCQ), true-false question (T/FQ), and complete Questions (CQ) from paragraphs of electronic course. The proposed framework consists of a lot of main stages, it uses both of the natural language processing (NLP) techniques to generate three types of questions (GFQ, T/FQ, and MCQ), and exam maker (EM), it uses the generated questions to produce the objective exams. The proposed system was evaluated by the extent of its ability to generate multiple objective questions. The questions that have been generated from the proposed system was presented to the three of the arbitrators specialists in the field of computer networks to express an opinion on the extent of their relationship to E-course and the accuracy of linguistic and scientific formulation. The results of the study showed an increase in the accuracy and number of the objective exams that were generated through the proposed system compared to the accuracy and number of the exams created by the traditional system this proves the efficiency of the proposed system.

Keywords—Objective exams (OE); Applications Artificial Intelligence (AAI); Random Objective Exams Generation (ROEG)

I. INTRODUCTION

Style currently used in the preparation of the exam is a traditional style, which is dominated by the nature of the challenge it takes to be a professor of time and effort great in search and review content to create a single model of exam with test questions fit in with educational experiences and training activities for students.

Electronic exams (E-exams) of the most prevalent methods of assessment for the purposes of both the evaluation to assess the learners' ability to learn, and to assess the impact of the teaching capabilities of the instructor, and thus Evaluation significantly affect the improvement of learning outcomes.

Exams play an important role in electronic evaluation, and provide an array of benefits for both the learner and the instructor [1].

The random objective exams generation (ROEG) depends on the questions as the main component of any exam. The task of composing exam questions is responsibility of the professor who collects their own exam bank in many forms to help them compose future exams.

ROEG has one of the Applications Artificial Intelligence (AAI) that makes questions Banka electronic courses and extracts data from exams [2].

ROEG can help professors effectively evaluate student's acquisition of essential knowledge and skills thereby enabling professors to focus on more complex educational activities. On the other hand help student focus on the main topics in their study [3].

In this study, intelligent approaches to exam Generation from Paragraphs of electronic courses will be presented which generates three types of Objective exams questions (multiple choice questions (MCQ) which require students to select the correct response from several alternatives, true-false question (T/FQ) and complete Questions (CQ) to supply a word or short phrase).

The following questions should be answered using this study: What stages design of a proposed framework to random objective exams generation (ROEG) from questions bank of electronic courses? How can a proposed framework be developed to help professors in generating objective exam from an E-Course? What are the most important used to evaluate the performance of the proposed Framework?

The objectives of this study are: Reach to stages of preparing a framework proposal to random objective exams generation (ROEG) from paragraphs of electronic courses, Preparing subsystem to query into bank question of E-Course aimed at generating random for three types of questions (MCQ, T/FQ, CQ), Preparing subsystem to generate models of the exam includes random questions which represent the output of the previous subsystem.

The following section will highlight the types of objective exam questions, the flow chart and stages of the Random Objective Exams Generation.

II. THEORETICAL FRAMEWORK

There are many E-Courses producing for students of Educational University, The professor puts a set of questions to create question bank of E-course to assist them in future exam generation for their students.

The exam generation process depends often on questions objectivity pattern, which focus on educational content, where the text is created a lot of objectivity questions related to that specific content [4].

MCQ, T/FQ and CQ generation is the tasks of generating questions from text inputs, having prospective E-course content. Which are used widely as tools for assessing evaluations for most levels of education as framework conceptual understanding of the students can be boosted by posing MCQs on the concepts just taught [5].

The art of formulating questions is one of the fundamental abilities of a good professor .in practicing professorate, an professor must reach the students hidden levels of knowing and awareness in order To help them to reach a high level of thinking in answering questions. Question bank in E-Courses can be classified into five categories they are as follows: Factual Questions (FQ), Inductive Questions (IQ), Analytical Questions (AQ), objective Questions (OQ) and Tag Questions (TQ) [6].

Objective exam questions (OEQ) are those that require a specific answer. An objective Exam usually has only one potential correct answer (there may be some room for answers that are close). OEQ may be constructed so that they contain a list of possible answers, so that the student will be expected to recognize the correct one, Objective Exam items are most often used to assess knowledge of a particular topic, and they typically appear on achievement exams, they are so easy to score, easy to analyze, and so easily tied to learning.

There are many advantages to objective Exams. They can, for example, significantly reduce marking time and analysis of individual questions is more feasible. This allows Professors to quickly identify areas where only a few candidates respond correctly or where most candidates choose the same incorrect option and try to correct any misconceptions.

Fill-in-the-blank questions are a common type of question due to their ease of creation and usefulness in classes across the curriculum. They are considered an Objective Exam Questions because there is only one possible answer that is correct. They are typically used to measure a wide variety of relatively simple skills and specific knowledge.

Also questions can be more easily pre-tested in order to evaluate their effectiveness and level of difficulty. For example, pre-testing may expose questions with design flaws such as good candidates consistently selecting incorrect options.

The types of OEQ as follows: Multiple choice questions (MCQ), true or false questions (T/FQ), Gap fill questions

(GFQ), and matching questions (MQ). Most professors attempt to get a mix of these types of questions in order to best cover the objectives that were part of the lecture plan.

Generating objectivity questions automatically is a relatively new and important research area and potentially useful in computer teacher. Here we first discuss a few systems for objectivity questions generation.

Brown et al. (2005) developed a system for automatic generation of vocabulary assessment questions. They used WordNet for finding definition, synonym, antonym and hyponym in order to generate the Questions focused on attention [7].

Aldabe et al. and Aldabe & Maritxalar developed systems to generate objectivity questions. They have divided the task into six phases: selection sentence, filling blanks, generation of distractors, selection of distractors, evaluation with learners and item analysis [8], [9].

For question selection Agarwal and Mannem used a number of features like: is it first question, contains token that occurs in the title, position of the question in the document, whether it contains abbreviation or superlatives, length, number of nouns and pronouns etc. But they have not clearly reported what should be optimum value of these features or how the features are combined or whether there is any relative weight among the features [10].

Generation of objectivity questions automatically consists of three major steps: 1) selection of sentences from which question can be generated; 2) identification of the keyword which is the correct answer; and 3) generation of distractors that are the wrong answers [11].

III. PROPOSED FRAMEWORK

The proposed framework is capable of generating for objectivity exams on the basis of knowledge and flexibility. Such a system normally establishes a knowledge base to guarantee a high possibility of success and quality of examination.

The proposed framework system of the random objective exams generation (ROEG) goes through several logical subsystems that can be represented by the flowchart shown in Fig. 1.

The proposed system is based on several criteria, the most important of which are the following: Taking into account the relative weight of each educational module within the E-course, taking into account the percentage of representation of each type of the three types of objective Exam Questions, the expense of ease and difficulty of the questions coefficient, avoiding questions repeat within the same exam, to avoid generating questions duplicates leaves.

The proposed framework consists of a lot of main Stages. It uses both of the Natural Language Processing (NLP) techniques to generate three types of questions (GFQ, T/FQ, and MCQ), and Exam Maker (EM), it uses the generated questions to produce the object exam.

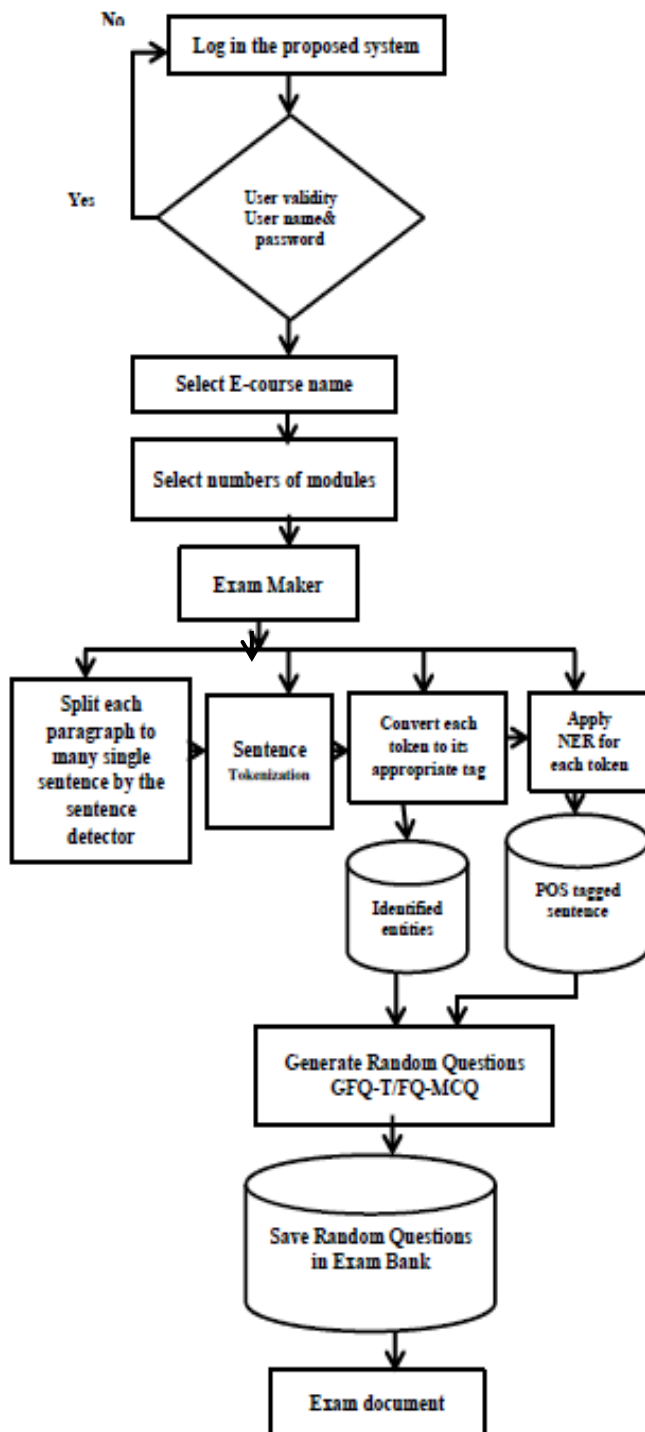


Fig. 1. Flow chart of the Random Objective Exams Generation (ROEG).

The actors in the proposed Framework use case are:

1) Admin, who is responsible for:

a) manage the verbs database use case which is responsible for adding, editing and removing verbs on the verbs database.

b) manage the system dictionary use case which controls the adding or deleting processes on the system dictionary.

c) prepare questions model use case which is used in the development of the generated exam use case.

d) Collected questions randomly for Preparing exam Models which is responsible for providing the professor with exam.

e) Login learners Questions bank Management and get their the powers to deal with the proposed system (user name & password)

2) Professor, who is responsible for:

a) Divide the course into learning modules, which are the foundation stone for generating Questions

b) Prepare questions, which are considered the main input to the generated exam

The three types of objective exams questions (MCQ, T/FQ, CQ) goes through many stages, as follows: data processing of educational module in E-Course, That stage aims to do many processes on the sentences which convey the main part of question; the output of this stage is the part of question tagged sentence and the identified entities on the Educational content of E-course.

The proposed question generation subsystem uses the open natural language processing (NLP) tool at this stage, which is a java library for processing natural text, based on machine learning tools; it supports variety of natural language processing tasks such as the following.

A. Information Extracted Paragraph Stage (IEPS)

The paragraph is a set of interconnected sentences in terms of meaning and relate to only one idea, each sentence containing a verb that expresses the idea of paragraph in the form of a statement, question, instruction, or exclamation and when typing a paragraph should begin with a capital letter.

There are many ways to extract information from the content such as rule learning based method, which use several general rules to extract information from content. The rule-based systems have been mostly used in information extraction from semi-structured web page.

A method is to learn syntactic/semantic constraints with delimiters that bound the text to be extracted, that is to learn rules for boundaries of the target text. Two main rule learning algorithms of these systems are: bottom-up method which learns rules from special cases to general ones, and top-down method which learns rules from general cases to special ones.

The (LP)² algorithms is one of the typical bottom-up methods. It learns two types of rules that respectively identify the start boundary and the end boundary of the text to be extracted from paragraph. The learning is performed from examples in a user-defined corpus (training data set). Training is performed in two steps: initially a set of tagging rules is learned then additional rules are induced to correct mistakes and imprecision in extraction. Three types of rules are defined in (LP)²: tagging rules, contextual rules, and correction rules.

B. Tokenization Stage (TS)

Tokenization is the task of chopping text up into words, phrases, symbols, or other meaningful element called token, perhaps at the same time throwing away certain characters, such as punctuation.

The list of tokens becomes input for further processing such as Grammatical or text mining. The tokenizer will split each sentence to separated tokens. The outputs of this stage are an array of tokens to be used in the following part of speech step.

C. Part of Speech Tagging (POST)

A process of marking up a word in a text as corresponding to a particular part of speech, based on both its definition and its context, which is a very important factor on determining the appropriate tag for each token, its relationship with adjacent and related words in a sentence or paragraph through a set of rules Which can be classified in Table 1.

There are numbers of maximum entropy POST developed in an attempt to further improve the accuracy that can be achieved by the tools which use it; such as open natural language processing, Examples include model Ratnaparkhi [12].

Ratnaparkhi describes Statistical model which trains from a corpus annotated with Part-Of- Speech tags and assigns them to previously unseen text with state-of-the-art accuracy (96.6%). The model can be classified as a Maximum Entropy model and simultaneously uses many contextual “features” to predict the POS tag. Furthermore, and use of specialized features to model difficult tagging decisions, discusses the corpus consistency problems discovered during the implementation of these features, and proposes a training strategy that mitigates these problems [13].

Along with contextual features looking at the surrounding words and tags, there are a number of futures based on the form of the word including the nature of affixes and the inclusion of apostrophes, hyphens, capital letters, and numbers It's also possible to further control the POST by providing it with a POS lookup list which consist of a text file with a word in the sentence and its possible POST on each line.

The question generation subsystem uses the maximum entropy model for its POST, it converts the array of tokens to its POST sentence. When the tag is repeated many times in the sentence, it will be distinguished by a number to talk it easy to recognize it later in the process of generation a question.

TABLE I. CLASSIFICATION RULES OF THE POS TAGS

Tag		Category
NN	Name	Common Noun (Singular or Mass)
NNP		Proper Nouns (Singular)
NNPS		Proper Nouns (Plural)
VB	Verb	Base form
VBD		Past tense
VBG		Present participle
VBN		Past participle
VBP		Non 3rd person singular present
VBZ		3rd person singular present
RP	Adverbs	Prepositions
RBR		Comparative Adverbs
RBS		Superlative Adverbs
CC		Coordinating Conjunctions
UH		Interjections
CD	Number	Cardinal number

D. Named Entity Recognition Stage (NERS)

Named-entity recognition stage (NERS) refers to extraction of data directly from text sentences considering that data extraction tasks are responsible for finding, storing and sorting textual content into categories.

NERS used by Question generation subsystem with its embedded open natural language processing library, which contains a set of pre-trained models for finding entity elements from raw data and can determine the category in which the element belongs; there are English Named Entity Recognition (date - location – organization – percentage - person – time). The system reads the sentence and highlights the important entity elements in the text.

The question Generation subsystem entity finder uses Maximum Entropy model to identify each entity .the Maximum Entropy Named Entity Recognition estimates probabilities based on the principle of making as few assumptions as possible, other than the constraints imposed . Such constraints are derived from training data, expressing some relationship between features and outcome [14].

The question generation adds new constraints to identify more entities, such as the prepositions which are followed by names considers location, and some prefixes such as (Sir., Prof., etc.) which identifies the person entity.

E. Stage of Key Word Answer (KWA) Determination According to the Type of Question Objective

The question generation proposed subsystem generates detect the Answer word required to prepare object question by using the identified entity type contained in the paragraph through some rules which can be explained in Table 2.

TABLE II. RULES FOR KEY WORD ANSWER DETERMINATION

key Word Answer(KWA)	Entity category	Tag	POS Tag paragraph
Who	Person	NN	P
Where	Location	NNP	
	Organization		
When	Period (Date/time/year/hours)	CD	

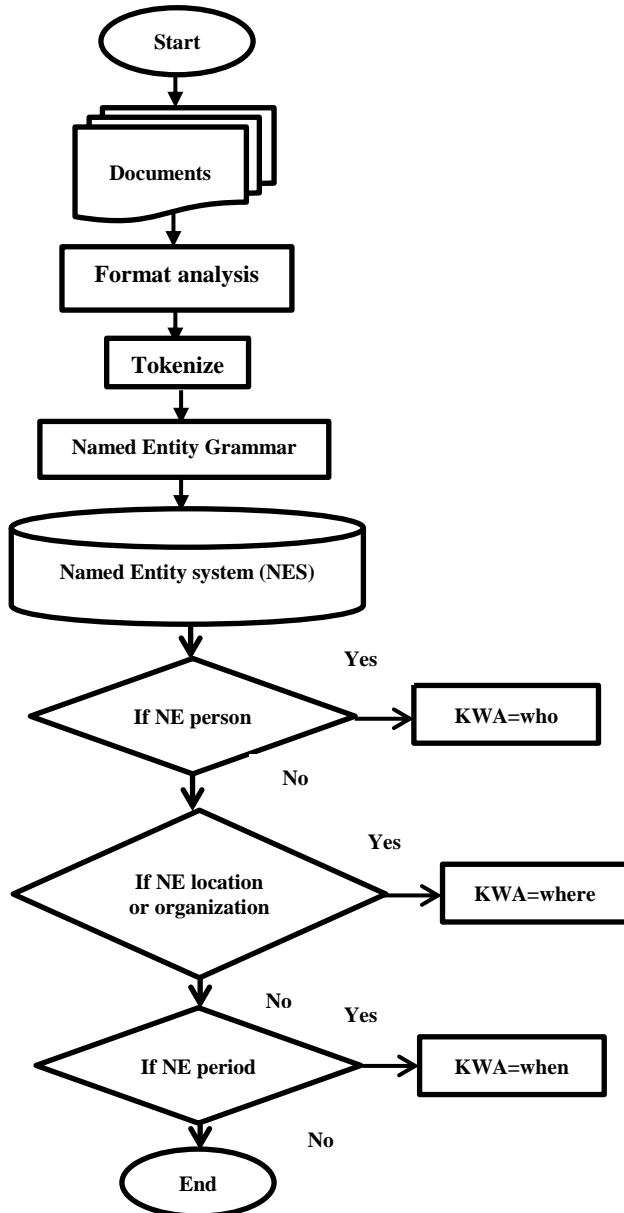


Fig. 2. Stage of key word answer determination according to the type of question objective.

The stage of KWA require analysis of each paragraph of E-Course to determine the entire (person - place - institution - time period) to answer according to the type of question Objective, so that this stage can go through the procedures described in Fig. 2.

F. Objective Question Generating Stage (OQGS)

This stage aims to generate three types of objective exams of each module through the following steps:

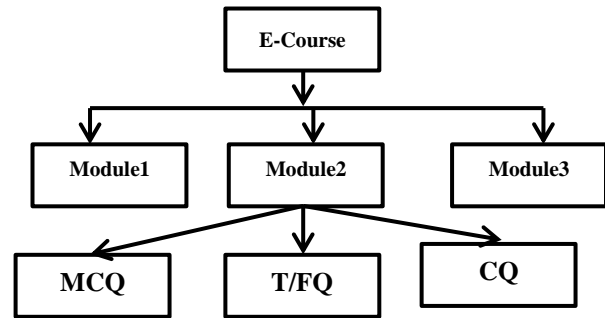


Fig. 3. Division the E-course to many modules.

- Division of the E-course to many modules as shown in Fig. 3.
- Restructuring the E-course manually to get a single shortened version of sentences that conveys the main piece of information.
- Get the tokens of the processed sentence (TS).
- Get the Part of Speech tagged sentence (POST).
- Named entity recognition to identify entities of each token.
- Get the final objective exam questions form.

G. Exam Maker Subsystem (EMS)

A process of preparing the exam maker require to be taken into account the appropriate number of questions in light of the relative weight of each module of E-course through the following equation:

$$NQ = \frac{TM}{TC} * TQ \quad (1)$$

Where:

NQ: number of questions for each Module

TM: total number of pages module

TC: total number of pages course

TQ: total number of exam questions

The exam maker subsystem should determine the number of each type of questions that have been generated for each module. Taking into account the equivalent percentage of representation for every word of the key Words Answer (KWA) such as (who-where-when)within each Module, where the ratio is equal to almost 33.3% .

According to the previous step, the exam maker subsystem presents the exam to the professor for selecting the date and the duration of the exam, and then be the examination ready for printing.

IV. EXPERIMENTAL WORK

The proposed intelligent system consists of a lot of main subsystems, it uses both of (NLP) techniques to generate three types of questions (GFQ, T/FQ, and MCQ), and exam Maker (EM), it has been implemented in the computer networks E-Course, it taught the four year student in the faculty of specific education, computer Teacher preparation department, Damietta university.

The following example illustrates the OEQ process.

The sentence: The best types of network cables are fiber cables and coaxial cables.

- 1) The POS tagged: Illustrated in Table 3.
- 2) The alternatives are:

- A. [The best] of [network] cables are [fiber] cables and [coaxial] cables.
- B. Types of network cables are fiber cables and cables.

TABLE III. THE TOKENS AND ITS CORRESPONDING TAGS

Token	Tag	The Meaning
The	DT	Determiners
best	JJS	Superlative Adjectives
types	NNS	Common Nouns (Plural)
of	IN	Prepositions and Conjunctions
network	NN	Common Nouns (Singular or Mass)
cables	NNS	Common Nouns (Plural)
are	VBP	Verbs (non 3 rd person singular present)
fiber	JJ	Adjectives
cables	NNS	Common Nouns (Plural)
and	CC	Coordinating Conjunctions
coaxial	JJ	Adjectives
cables	NNS	Common Nouns (Plural)

The question generate subsystem is done through Getting the Part Of Speech Tagging (POST) sentence and entities, Selection randomly of alternatives that may be either the first name or adjective, and finally empty the selected alternative.

V. RESULT AND DISCUSSIONS

The proposed system will be evaluated by the extent of its ability to generate multiple objective questions. The questions that have been generated from the proposed system was presented to the three of the arbitrators specialists in the field of computer networks to express an opinion on the extent of their relationship to E-course and the accuracy of linguistic and scientific formulation. Examining what the arbitrators agree upon, the number of valid question and these types that the program generated for the computer networks E-course illustrated in Table 4.

TABLE IV. THE NUMBERS OF THE GENERATED QUESTION FOR THE COMPUTER NETWORKS E-COURSE

Module	Type of questions			Total
	MCQ	T/FQ	CQ	
1	10	6	10	26
2	12	9	8	29
3	8	10	6	24
Total	30	25	24	79

To judge the effectiveness of the proposed system in terms of its ability to query and accuracy in generating questions, so we will use the following [15] :

$$Accuracy = \frac{Correct}{Correct + defective} \tag{2}$$

$$Query = \frac{Correct}{Correct + missed} \tag{3}$$

Where, correct represents the number of questions that have been generated through the proposed system has been evaluated by the arbitrators to correct it linguistically and scientifically. Defective represents the number of questions that have been generated through the proposed system has been evaluated by the arbitrators as incorrect linguistically or scientifically. Missed represents the number of questions that are not being generated by the proposed system, and that should the proposed system is that they generate from the viewpoint of arbitrators. Table 5 and Fig. 4 illustrate that the query rates are greater than accuracy rates because the proposed system passes by each sentence and paragraph in the E-course of computer networks and generates many questions on it. To evaluate the examination generated by the proposed system, a questionnaire was presented to the five of the arbitrator’s specialists in the field of computer networks and teaching methods, to give their opinion on the availability of educational and academic standards in form and content then apply the following equation:

$$Agreement\ Coefficient = \frac{Number\ of\ approvers\ on\ standard\ coefficient * 100}{Total\ number\ of\ arbitrators} \tag{4}$$

This is evident from Table 6 and Fig. 5 that Arbitrator’s approval of the exam to achieve educational and academic standard’s required average coefficient value of the agreement was 90%. These results agreed with the previous studies in the automatic exam generation evaluation, when always the recall is higher than precision because the generated exams from the proposed system are always more than the exams generated by the professors, this prove the effeteness of the proposed system.

TABLE V. THE OBTAINED ACCURACY AND QUERY

Arbitrators no.	Accuracy	Query
1	0.41	0.70
2	0.38	0.61
3	0.44	0.77
The average	0.41	0.69

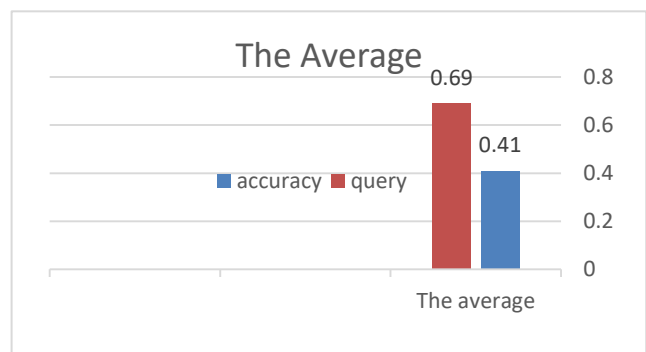


Fig. 4. The average of accuracy and query.

TABLE VI. THE RESULTS OF ARBITRATOR'S OPINION, VALUES OF THE AGREEMENT COEFFICIENT FOR EACH STANDARD AND THE AVERAGE

No.	Standard	The Number of consenting to meet the standard	The agreement Coefficient	The Average
1	Questions fit with the objectives of the module	5	100%	90%
2	the Question covers only one idea	4	80%	
3	Clarity wording of the questions	5	100%	
4	Cover questions per module	4	80%	
5	Questions devoid of grammatical errors	4	80%	
6	Questions devoid of exile	5	100%	
7	Questions measure cognitive aspects	5	100%	
8	Questions measure Analysis capability	5	100%	
9	The answer can be inferred directly from the head of the question	4	80%	
10	Simplicity of Questions	4	80%	
11	Blanks appears at the end of the question	5	100%	
12	Questions fit with individual differences among students	4	80%	
13	Lack of multi-answer of the question	4	80%	
14	Answer of question does not depend on answer of another question	5	100%	

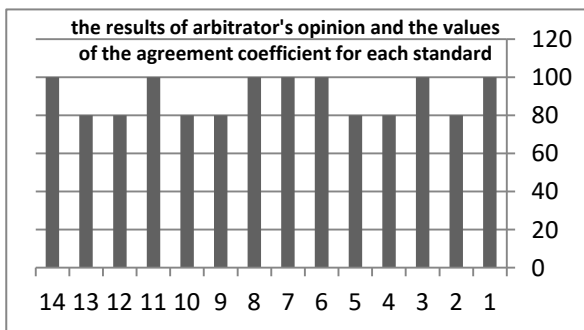


Fig. 5. The results of arbitrator's opinion and the values of the agreement coefficient for each standard.

VI. CONCLUSION

In this study, an approach to automatically generate exam from a paragraphs of E-Course was proposed. This approach is included in a Proposed by the researcher an implemented on the computer networks which is taught to first year student in faculty of specific education, Damietta University, Egypt Automatic exam generation process has gone through many stages will be summarized as follow: Information extracted paragraph stage (IEPS), Tokenization Stage (TS), Part Of Speech Tagging (POST), Named Entity Recognition Stage (NERS), stage of key word Answer (KWA) determination according to the question type Objective, objective question generating Stage (OQGS), and exam maker subsystem (EMS).

REFERENCES

- [1] Van lent, GerbenGLOBAL,2009. Risks and benefits of CBT versus PBT in high-stakes testing. The Transition to Computer-Based Assessment, VOL. (83).
- [2] Fredrik Olsson, 2009. "A literature survey of active machine learning in the context of natural language processing", SICS Technical Report T2009:06, ISSN: 1100-3154.
- [3] Javier Sarsa and Rebeca Soler,2012, E-Learning Quality Relations and Perceptions
- [4] Wolfe,J., Automatic question generation from text –an aid to independent study .In Proceedings of ACM SIGCSE.1976 Available at : <http://dl.acm.org/citation.cfm?id=803459>
- [5] Nicol, D., 2007. E-assessment by design: using multiple choice tests to good effect. Journal of Further and Higher Education, Vol. 31(1), pp. 53-64
- [6] Michael Heilman,2011. Automatic Faculty Question Generation from Text. PH.D. Dissertation, Carnegie, Mellon University, 2011
- [7] Brown, J. C., Frishkoff, G. A., and Eskenazi, M., 2005. Automatic question generation for vocabulary assessment. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Association for Computational Linguistics, pp. 819-826.
- [8] Aldabe, I., Lopez de Lacalle, M., Maritxalar, M., Martinez, E., Uria, L., 2006. ArikIturri: An Automatic Question Generator Based on Corpora and NLP Techniques. In ITS. LNCS 4053, pp. 584594
- [9] Aldabe, I., Maritxalar, M., 2010. Automatic Distractor Generation for Domain Specific Texts. Proceedings of IceTAL, LNAI 6233. pp. 27-38
- [10] Agarwal, M., and Mannem, P., 2011. Automatic gapfill question generation from textbooks. In Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications, pp. 56-64
- [11] Bernhard, D., 2010. Educational Applications of Natural Language Processing. In NATAL. pp. 1-123.
- [12] Kubler,Sandra,et al.2010. "Adding context information to part of speech tagging for dialogues." Ninth International Workshop on Treebanks and Linguistic Theories.
- [13] Ratnaparkhi, Adwait. 1996. A maximum entropy model for part-of-speech tagging. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, University of Pennsylvania, pp. 133–142.
- [14] Chieu,Hai Leong, and Hwee Tou Ng.2002. "Named entity recognition:a maximum entropy approach using global information." Proceeding of the 19th international conference on computation linguistics, Vol.(1).
- [15] Kalady,S.,Elikkottil.A,and Das,R.,Natural language question generation using syntax and keywords. In Kristy Elizabeth Boyer and Paul Piwek, editors, Proceeding of QG2010:The Third Workshop on Question Generation, Pittsburgh,June 2010. Available at: Question generation Org.URL<http://oro.open.au.uk/22343/>.