Mapping Arabic Text Studying Material With The Most Suitable Student Grade

Aisha Basheer, Sara Alawneh, Tala Bostami, Suhad Daraghmeh, Amjad Hawash*
Information & Computer Science Department
An-Najah National University
Nablus, Palestine, P.O. Box:7
aisha.a.basheer87@gmail.com, saraalawneh89@gmail.com, tala.darweesh.boss@gmail.com, suhad@najah.edu, amjad@najah.edu

***Abstract*—Students' poor comprehension is one of the major difficulties that struggle with fully understanding passages. Several factors are related to this problem such as huge material cramming, the absence of creativity, and unsuitable vocabularies. Reading is one of the vital methods to increase students' vocabularies which reflects positively on their comprehension. The suitable mapping between reading materials and students' levels of comprehension is very important in enhancing their reading practice as well as preventing them from feeling frustrated because of the lack of suitable reading materials. In this work, we implemented an Arabic classifier tool based on a Language Modeling Approach used in English and French documents to match reading materials with students' comprehension so that the tool can suggest the lowest level of student grade that fits the tested material. Moreover, the tool can be used to study the comprehension level of students by analyzing their answers to some textual writing assignments as a method to better tune the teaching skills and techniques to increase student's comprehension levels. As an initial effort, 3 subjects were chosen to test the tool (Arabic, Science, and Islamic Studies). A set of students from different schools were chosen for the experimental tests in which good results were achieved.**

***Index Terms*—Additive Method, Arabic Classifier, Maximum Likelihood Estimator, Reading Comprehension, Text Classification.**

I. INTRODUCTION

A learning disability is much related to comprehension problems in such a way that makes learners unable to understand passages and/or materials. Students in different grades are much affected by this problem [1]. Learning disabilities have different signs depending on the level of students. Preschool ages suffer from pronouncing words, trouble learning the alphabet and find it difficult to find the suitable word in some simple conversation [2]. Ages from 5 to 9 have trouble learning the connection between words, misspell some words and show some confusion about basic words when reading. Ages from 10-13 have problems related to difficulty in reading comprehension, trouble following some discussions, and poor handwriting [3].

Reading plays an important role in the process of learning and knowledge acquisition for both children and adults. However, not all texts are accessible to every prospective reader. Reading difficulties can arise when there is a mismatch between a reader's language proficiency and the linguistic complexity of the text they read. In such cases, simplifying the text in its linguistic form while retaining all the content could aid reader comprehension [4].

Dyslexia or reading disorders are one of the results of poor vocabularies that affect students' personality and their ability to involve in society although they don't have problems with vision or comprehension. Enriching students with suitable vocabularies have a direct impact on their reading performance and hence their comprehension [5]. Different pedagogical researches concluded that there are ways to enrich students' vocabularies:

1) **New daily vocabularies:** several minutes a day learning new vocabularies.
2) **Reading:** Reading suitable age materials enables students to gather new vocabularies.
3) **Relate context to vocabularies:** Games, Songs, and music for example are good resources for gaining new vocabularies through fun.
4) **Word association:** is a good way to increase vocabularies in such a way students will be able to match terms with events or for example with dates [6].

Although the Arabic language is considered one of the international languages and the most Semitic languages in terms of the number of speakers and number of vocabularies, research related to the Arabic language face several Natural Language Processing (NLP) problems such as the way the Arabic language is written (connected letters), the absence of diacritics of today's writing that leads to miss-understanding of some words and the ability of some words to have a verb and a name forms that leads to some ambiguity incomprehension are all examples of such problems [7]. As a result, few works are conducted in the field of increasing Arabic text comprehension for both school students and foreigners who would like to learn the Arabic language.

One of the most important issues in increasing students' vocabularies in spoken languages (including Arabic) is relating them to the appropriate materials represented in songs, stories, teaching materials, and others. The suitability of vocabularies is much related to the amount of their comprehension. This amount can be represented by the intersection between what a teacher wants to teach to increase students' comprehension and what they really got and learnt [8].

Corresponding Author: amjad@najah.edu

Our main contribution in this work is to study the impact of mapping the Arabic textual materials with the suitable students' ages on enriching students' comprehension mainly using 3 different studying subjects: Arabic, Science, and Islamic Studies. This contribution can be summarized in the following points:

1) Updated the *Additive Method* by adding a fixed value k to the samples we worked on.
2) Worked on trying several ways to compute the best k value by replacing the addition process with multiplication.
3) Computed the k value depending on the differences in frequencies for words that appears in different consequent levels.
4) Used the *Maximum Likelihood Estimator* that is based on computing the size of a paragraph and the frequencies of its words.
5) Implemented the previous issues in a tool that:
   a) Tests the ability for the user to check the suitability of an entered paragraph with a specific grade.
   b) Suggests the replacement of difficult words with easier ones that are more suitable to student age.
6) The introduction of the "Intra-level" mechanism that checks the suitability of the paragraph with the level of grade suggested by the tool.
7) Enhancing the matching process between tested material and the 12 levels by using synonyms of the words.

Besides the ability to suggest the most suitable materials for students, it is carefully able to help the teachers to choose the most appropriate terms (in the exam preparation process) to be more understandable by students.

The rest of this paper is organized as follows: Previous works are proposed in Section II. Section III discusses methodology of the work while Section IV discusses the Maximum Likelihood Estimator concept. An illustrative example is presented in Section V and the tool architecture is discussed in Section VI. Section VII is related to the conducted experimental tests and collected results and finally Section VIII concludes the work.

II. Related works

Several works and techniques were conducted regards the improvement of students' readability and comprehension. Some of these works emphasize some statistical models to predict reading difficulties, others depend on improving comprehension. Some of the works are related to analyzing the comprehension problem by examining some difficulties that are related (but not limited) to jumping letters, skipping words, the reversal of letters and words, pronunciation, and reading fatigues.

The work presented in [9] is related to developing an Information Retrieval that is not just retrieving the related documents of some query, but also retrieving the suitable documents according to the student's reading level. The authors of the work proposed an applicable statistical model of text for different reading levels. They recast the well-studied problem of readability in terms of text categorization, and use straightforward techniques from a statistical language modeling, in which it is applied for a set of web pages to be classified according to their reading difficulties.

Authors of the work in [10] investigated the relationships between the level of text difficulty, elementary students' reading fluency, and reading comprehension as a way to develop new standards to increase text complexity in elementary classrooms in the USA.

On average, the increased text difficulty level was related to decreased reading fluency, with a small number of exceptions. The authors found that on average, the text difficulty level was negatively related to reading comprehension, although a few studies found no relationship.

Investigating the relationships between the level of text difficulty, elementary students' reading fluency, and reading comprehension as a way to develop new standards to increase text complexity in elementary classrooms in the USA was the work of [10].

The work presented in [11] argues that Good-Turing methods are one means of estimating tasks such as spelling correction, and sense disambiguation and the translation is improved if one can estimate a probability for an object of interest. The paper presented a method that uses the simplest possible smooth, a straight line, together with a rule for switching from Turing estimates which are more accurate at low frequencies. They called this method: Simple Good-Turing (SGT). The accuracy of the SGT is compared to two other methods for estimating the same probabilities, the Expected Likelihood Estimate (ELE), and two-way cross-validation. The SGT method is more complex than ELE but simpler than two-way cross-validation.

Text categorization is the process of grouping documents into categories based on their contents. This process is important to make information retrieval easier, and it became more important due to the enormous textual information available online. The main problem in text categorization is how to improve classification accuracy. Although Arabic text categorization is a new promising field, there are a few kinds of research in this field [12].

The Arabic language is one of the richest languages in the world, where it has many linguistic bases. Because of the few works related to the Arabic language, the classification of Arabic texts is not an easy task. Different algorithms are used to classify Arabic documents. Naïve Bayesian algorithm (NB), K-Nearest Neighbor algorithm (KNN), Support Vector Model (SVM), and Artificial Neural Network (ANN) are all examples of different techniques used [13].

Thanks to the availability of texts on the Web in recent years, increased knowledge and information have been made available to broader audiences. However, how a text is written, its vocabulary, and its syntax—can be challenging to read and understand for many people, especially those with poor literacy, cognitive or linguistic impairment, or that limited knowledge of the language of the text. Texts containing uncommon words or long and complicated sentences can be challenging to read and understand by people as well as challenging to be analyzed by machines [14].

For increasing the student's comprehension of Arabic content, few works were conducted. The authors of the work presented in [15], developed the application of the SQ4-R learning technique applied to a set of students to improve their ability to find words and their meanings from the dictionary. The authors noticed an increase in students' comprehension of searched words. While authors of [16] differentiated between normally achieving students and those at risk. Using a cross-sectional design, the study tested the effect of gender, grade level, and student condition on the variation of Arabic reading skills. Dependent measures of the study included letter-sound identification (LSI), word decoding (WD), phonological awareness (PA) through blending and segmentation, word recognition (WR), reading comprehension (RC), and listening comprehension (LC) in Arabic. Multivariate analysis indicated that gender, grade level, and student condition affected the variation of reading skills.

In this work, and based on the updated Additive Method, we developed a categorization tool that is able to suggest the most suitable students' grade level to a tested Arabic passage or paragraph. Moreover, we introduced the Intra-level feature that adorns the tested material with either a *strong* or *weak* marker within the same grade in order not just to match the tested material with the least suitable grade, but also to able to categorize the material within the same grade since it may contain strong and weak students.

III. Methodology

In the following subsections, we are describing detailed steps for our methodology in this work.

A. Data Gathering

Our work is initially concerned with three school subjects: Arabic Language, Science, and Islamic Studies [1]. We initially collected a set of accurately labeled vocabularies for each single specific grade level related to the three mentioned subjects taken from the Palestinian school curricula for the levels from 1 to 12. The collected data are stored in a special database alongside their frequencies per level[2] and the material they belong to. The number of collected distinct terms for Arabic Language, Science, and Islamic studies is 54733, 29721, and 27042 respectively [3].

B. Data Cleansing

The Arabic language needs special treatment concerning linguistic computations. Because of this, we executed a set of operations on the collected data before they have been used in the work. These operations are done just one time as a preparation step before the involvement of the vocabularies in any kind of computations. These operations can be summarized chronically as follows [4]:

1) **Removing Stop words:** this includes prepositions, adverbs, dialectics, time-related terms, and interrogatives.
2) **Removing punctuation marks:** such as exclamation, questions, quotation, punctuation, commas, and semi-colons.
3) **Removing mathematical signs:** minus, plus, equal, multiplication, divide, and percentage marks (of course these signs are necessary if Mathematics material is included in the study).
4) **Removing low-frequency words:** all words whose total frequencies in all grades are less than 2 and those that are mentioned just in one level.

C. Studying Data

As we mentioned before, the vocabularies with their frequencies according to the level and subject they belong to, are all saved in a special database.

The data has been studied and we found that the amount of frequencies for (nearly) all vocabularies is higher in the upper levels than in the lower ones. This is due to the fact that the books of higher levels are bigger (in terms of the number of pages and the diversity of subjects) than the books of lower levels. This variation in frequencies created a problem because when we search for a famous and well-known word using the tool, we will get a confusing result since the students of all levels can recognize the word. Regardless of this, however, searching for just one word is a straightforward process, the tool just searches for the lowest level able to recognize it and just displays the result for the user. The problem exists in searching for paragraphs, not separate words because paragraphs probably contain several words from several levels. The question is: **what is the suitable minimum level able to recognize the whole paragraph in this case?**

In order to smooth the values of frequencies, we adapted several smooth techniques: we first changed the frequencies for each word per level, then we tried to change the frequencies of words with their first appearance per level, and finally, we changed the frequencies of all words in all levels. Unfortunately non of these tested methods were able to give reasonable results. One of these methods is the *Additive Method* [17], in which a set of specific numbers (threshold) are added to all frequencies of the words in the database according to the levels they belong to in order to adjust these frequencies in a way makes the calculations behind displaying the most relevant level for a given paragraph to be correct.

In order to tune the best numbers to be added to the frequencies of each level, we used to compute the difference in frequencies for the similar words in each consecutive level in order to study the natural evolution and usage of words for students upgrading in study levels. For example, given a

[1]However, the science subject in the secondary levels (from 10 to 12) is divided into 3 different subjects: Physics, Chemistry and Biology. We handled the last subject.

[2]From now and on the terms *level* and *grade* will be exchangeable.

[3]Dealing with phrases like "Life Science" as a whole unit will be one of our future works

[4]We did not use a special segmentation algorithm for the collected texts, rather, we used the splitting techniques that exist in different programming languages.

word W first found in level L, we started to compute the differences of the frequencies of W starting from level L up to the 12^{th} level and record the difference between each subsequent level. This process is done for all words at all levels. However, during our calculations, we found that the words of the preliminary levels have bigger differences in frequencies than the late levels. This makes sense because usually, the late levels use a bunch of words that have smaller frequencies between levels and because of the natural containment of higher levels for the lower ones in terms of used terms. Usually, the students in the elementary levels were given a lot of new words and this makes the difference in frequencies in these levels be big valued. Table I below depicts the number of differences in frequencies computed and used in the *Additive Method* of our work.

TABLE I
LEVELS WITH THEIR COMPUTED NUMBERS.

Level	Number
1	300
2	200
3	50
4	10
5-9	5

Depending on the numbers that appear in the previous table, we multiplied all words frequencies in each level by the number that appears in their corresponding levels (Multiplication is actually a repetitive addition). Relying on this, we preserved the 0 frequencies for those words that do not appear in some levels (usually the words in the higher levels do not appear on the lower ones). We preserved these zeros by using the multiplication of computed numbers with frequencies instead of using the additive method that does not preserve these zeros. Table II contains a sample set of 5 words with their original frequencies in all levels, their frequency differences, and their new computed frequencies, where $O.F.$, $F.D.$ and $N.F.$ are acronyms for *Original Frequencies*, *Frequency Difference* and *New Frequency* respectively. The numbers appear in the $F.D.$ column is calculated by subtraction of consequent grades, and this is why we have —in the last cell because the numbers above them represent grade 12 minus grade 11 frequencies.

IV. MAXIMUM LIKELIHOOD ESTIMATOR

Up to now, we have a database that contains all words that appear on all levels, each of which with its modified frequency per level. Now, given a sample paragraph supplied by some user and in order to compute the amount of its likelihood with the 12 levels, a set of computations take place depending on the following data: 1) Length of supplied paragraph (entered by a user) 2) Paragraph words' occurrences frequencies and 3) Level words' occurrences frequencies. Depending on these values, we computed the *Maximum Likelihood* [18] between the supplied paragraph and the 12 levels using the following formula:

$$L(T|G_i) = \sum_{j=1}^{|V|} c(w_j) * log(P(w_j|G_i)) \tag{1}$$

where:

1) $L(T|G_i)$ is the amount of likelihood L between the set of supplied words in the paragraph T and a given level G_i where $i \in \{1, 2, .., 12\}$.
2) V is the set of unique words that appear in paragraph T.
3) $c(w_j)$ is the amount of word frequency in the supplied paragraph,
4) $P(w_j|G_i)$ is the amount of existence probability between the word w_j and the level G_i.

In order to calculate the value of $P(w_j|G_i)$, we have one of two cases per word w_j: either it exists or not in the database.

1) The word w_j exists in the database:: In this case, the existence of the word w_j has some occurrence at least on one level of the 12 levels. In this case, we applied the following formula:

$$P(w_j|G_i) = \frac{\sum_{k=1}^{12} \alpha_k * P_k}{\sum_{k=1}^{12} \alpha_k} \tag{2}$$

where P_k is the probability of w_j per level k and $\alpha_k = \varphi(i, k, \sigma)$ is a kernel distance function between levels i and k with width parameter α.

The computation of P_k is done as follows:

1) Retrieve the modified frequency value of w_j from the database.
2) Compute the total frequency of w_j in all 12 levels.
3) Compute the probability of the existence of the word w_j in level k by dividing the frequency of the word in that level by the total frequency computed in point 2.

Now, in order to compute the value of α_k, we used a simple regression kernel that depends on the Gaussian centered at mean grade k with a standard deviation σ for the width parameter:

$$\varphi(i, k, \sigma) = \exp\left(\frac{(i-k)^2}{\sigma^2}\right) \tag{3}$$

With training, the optimal value for the width parameter α with respect to the overall root mean square (RMS) error was found to be approximately 6.25-grade levels [9].

2) The word w_j does not exist in the database:: In this case, the length of the word w_j is used in the computations instead of its frequency. The assumption that the length of the word has some relation to the level it may occur is the basis behind this computation. Usually, with the lower levels, the length of words is shorter than those that appear in higher ones. The following equation is used to compute the amount of $P(w_j|G_i)$:

$$P(w_j|G_i) \approx |w_j| * (1 + \frac{(i - |w_j|)}{10}) \tag{4}$$

TABLE II
SAMPLE WORDS FROM DIFFERENT LEVELS AND THEIR FREQUENCIES' COMPUTATION.

Word	معلم Teacher			دبابيس Pins			جار Neighbor			دموع Tears			تلقيح Pollination		
Level #	O. F.	F. D.	N. F.	O. F.	F. D.	N. F.	O. F.	F. D.	N. F.	O. F.	F. D.	N. F.	O. F.	F. D.	N. F.
1	8	3	2400	0	2	0	2	0	600	0	0	0	0	0	0
2	11	29	2200	2	0	400	2	0	400	0	0	0	0	0	0
3	40	24	2000	2	8	100	2	2	100	0	2	0	0	0	0
4	64	24	640	10	0	100	4	2	40	2	0	20	0	0	0
5	127	15	635	10	0	50	6	2	30	2	0	10	0	0	0
6	142	45	710	10	2	50	8	1	40	2	1	10	0	0	0
7	187	14	935	12	0	60	9	2	45	3	5	15	0	0	0
8	201	12	1005	12	0	60	11	2	55	8	0	40	0	0	0
9	213	11	1065	12	0	60	13	0	65	8	3	40	0	0	0
10	224	9	224	12	0	12	13	1	13	11	1	11	0	0	0
11	223	22	233	12	0	12	14	0	14	12	6	12	4	0	4
12	245	—	245	12	—	12	14	—	14	18	—	18	4	—	4

where $|w_j|$ is the length of the word w_j and i is the level being tested.

A. *Levels & Materials Suggestions*

After data collection, cleansing, and computing frequencies, we developed a web-based system connected with the database containing the data and supported by a searching mechanism in order to enable users to feed the system with paragraphs and the latter matches the paragraphs with all words saved in the database to output the user with the most relative class levels suitable with the provided paragraphs. In order to increase the accuracy of the system, we included the synonyms of the words in the searching process (more than 4000 words with their synonyms were taken from an online Arabic dictionary [5]). Now, if users want to search for the most appropriate class levels that match a paragraph of their own, the system executes a matching process and displays the user a list of suitable levels. The user could refine the matching process by using some of the synonyms suggested by the system for all of the words that appear in the paragraph so that the user could use some or all of them and repeat the searching process to get more suitable levels.

In order to have more precise results, we cloned the words for levels 1, 2, and 10 in a separate table and categorized them into two categories: *Strong* and *Weak* according to word length (in terms of characters) and its occurrence frequencies per level. The idea behind this is to provide more accurate results to users in a way they can measure the comprehension strength of students within the same level. This is important because a teacher can assign simple paragraphs to weak students and gradually can assign them harder ones. This enables the teacher to assign students within the same level different homework according to their comprehension ability. Moreover, the work is able to compute the similarity between the entered text and the 3 mentioned subjects, so that a user is able to notice the most related subject to the entered text.

With respect to material suggestion, we accumulated the words that appear in the tested material and compare these words with all words belonging to each studying material under consideration: Arabic, Science, and Islamic Studies. For each material, we counted how many matches between words and computed the percentage of matches by dividing the number of matched words by the total number of words in each material.

[5]https://www.almaany.com/

V. EXAMPLES

In order to illustrate the previous computations and the main idea of mapping a paragraph to the least suitable grade level, an illustrated example is discussed here. Suppose we have the following three phrases:

1) The boy drank the milk -شرب الولد الحليب.
2) Fruits are perfect food الفواكه غذاء مثالي.
3) Diplomatic party حزب دبلوماسي.

We are going to apply the previous computations depending on 3 different grade levels, mainly:1, 6, and 12. The reason why we choose these levels is to apply the computations on levels that represent low, mid, and high students' age grades.

After segmenting (tokenizing) and cleansing the three phrases, a set of keywords with their mapping probabilities with the three mentioned grades are summarized in Table VI where w represents the keyword and $p(w)$ is its corresponding probability with each grade. However, and because of paper space limitation and the inherent repetition of calculations, tables III, IV and V represent the outputs of applying equations 1, 2 and 3 on the words of the first phrase against the 12 levels where $F.A.$ is the Frequency Additive, $Probability$ is the result of dividing the word Frequency Additive in some level by the total frequency summation of the same word in all levels. For example, the $Probability$ value 0.3630 appears in Table IV for the second level is computed by dividing the value 400 by the total 1102, and $P(w_j|G_i)$ represents the probability of the word w_j with respect to the grade being computed for by applying either equation 2 if the word w_j exists in the database or equation 4 if the word does not exist at all.

By applying equation 1 with respect to the values that appear in the mentioned tables:

1) **Phrase #1: شرب الولد الحليب**

TABLE III
COMPUTATIONS RELATED TO THE WORD "DRANK".

Frequency	F. A.	Probability	$P(Drank\|G_i)$
0	0	0.0000	0.1376
2	400	0.3361	0.1397
3	150	0.1260	0.1281
7	70	0.0588	0.1084
16	80	0.0672	0.0907
19	95	0.0798	0.0809
19	95	0.0798	0.0762
19	95	0.0798	0.0708
22	110	0.092	0.0617
26	26	0.0218	0.0507
29	29	0.0263	0.0412
40	40	0.0336	0.0352

TABLE IV
COMPUTATIONS RELATED TO THE WORD "BOY".

Frequency	F. A.	Probability	$P(Boy\|G_i)$
0	0	0.000	0.1345
2	400	0.2495	0.1449
8	400	0.2495	0.1398
10	100	0.0623	0.1206
20	100	0.0623	0.097
23	115	0.0717	0.0819
23	115	0.0717	0.0736
26	130	0.0810	0.0671
28	140	0.0873	0.0578
29	29	0.0180	0.0465
34	34	0.0212	0.0365
40	40	0.0249	0.0298

TABLE V
COMPUTATIONS RELATED TO THE WORD "MILK".

Frequency	F. A.	Probability	$P(Milk\|G_i)$
1	300	0.1098	0.2084
6	1200	0.4392	0.1901
7	350	0.1281	0.1539
13	130	0.0475	0.1097
17	85	0.0311	0.0730
20	100	0.0366	0.0526
23	115	0.0420	0.0452
27	135	0.0494	0.0425
35	175	0.0640	0.0386
42	42	0.0153	0.032
45	45	0.0164	0.0269
55	55	0.0201	0.0228

TABLE VI
THE PROBABILITIES OF THE APPEARANCE OF THE KEYWORD IN GRADES 1, 6, AND 12.

Type w	Grade 1 P(w)	Grade 6 P(w)	Grade 12 P(w)
Drank شرب	0.1376	0.0809	0.0352
Boy الولد	0.1345	0.0819	0.0298
Milk الحليب	0.2084	0.0526	0.0228
Fruits الفواكه	0.0758	0.0956	0.0426
Food غذاء	0.0972	0.0853	0.0474
Perfect مثالي	4.4023	0.0505	0.1704
Party حزب	2.1440	0.0024	0.3410
Diplomatic دبلوماسي	6.0884	0.0681	0.1119

$L(T|Grade_1) = log(0.1376) + log(0.1345) + log(0.2084) = -2.4132$

$L(T|Grade_6) = log(0.0809) + log(0.0819) + log(0.0526) = -3.4569$

$L(T|Grade_{12}) = log(0.0352) + log(0.0298) + log(0.0228) = -4.617$

$Prediction = Max(-2.4132, -3.4569, -4.617) = -2.4132$ and this goes to Grade 1.

2) **Phrase #2:** الفواكه غذاء مثالي

$L(T|Grade_1) = log(0.0758) + log(0.0972) + log(0.00004) = -6.5306$

$L(T|Grade_6) = log(0.0956) + log(0.0853) + log(0.0505) = -3.3843$

$L(T|Grade_{12}) = log(0.0426) + log(0.0427) + log(0.170) = -3.4614$

$Prediction = Max(-6.5306, -3.3843, -3.4614) = -3.3843$ and this goes to Grade 6.

3) **Phrase #3:** حزب دبلوماسي

$L(T|Grade_1) = log(0.000000214) + log(0.00006) = -10.891$

$L(T|Grade_6) = log(0.00242) + log(0.06810) = -3.7813$

$L(T|Grade_{12}) = log(0.3410) + log(0.1119) = -1.4181$

$Prediction = Max(-10.891, -3.7813, -1.4181) = -1.4181$ and this goes to Grade 12.

From the mentioned examples, the implemented algorithms succeeded in selecting the least suitable grade for each of the three phrases, where the maximal amount of $L(T|G)$ is selected.

VI. TOOL IMPLEMENTATION AND ARCHITECTURE

In order to make it easy for users to test the suitability of a given material with the most related grade, all of the previous equations are implemented in a client/server architecture tool. A 3 tiers architecture tool has been developed using HTML, JavaScript, and CSS for the front-end tier (Presentation Layer), PHP for the back-end tier (Computation Layer), and MySQL for the storage layer. Figure 1 below depicts the architectural structure of the tool where the request/reply protocol is used for data transmission.

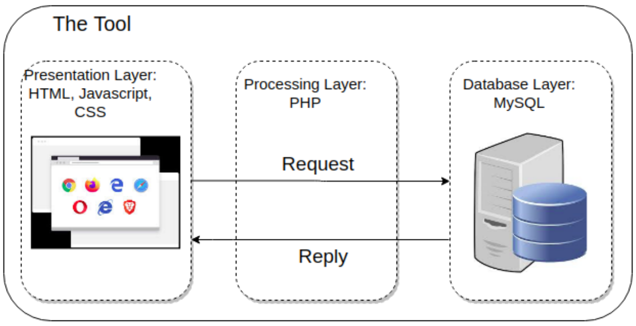

Fig. 1. The tool architecture.

The presentation layer enables users to support the tool with paragraphs in different ways. Users can directly write their paragraphs in a special text box or they can upload a complete text file(s) to be completely mapped by the tool with the suitable grade(s). In either case, the tool and after executing the mapping process by the computation and database layers displays the results to the users in a suitable format. The tool is able to process a set of text files by supporting it with the folder path on the client-side in which the tool evolves the files located there and starts the mapping process per file. Extensions of the files could be ".txt" or ".doc".

Figure 2 represents the result of suitability for some text with the 3 subjects: Arabic, Science, and Islamic Studies represented in percentages.

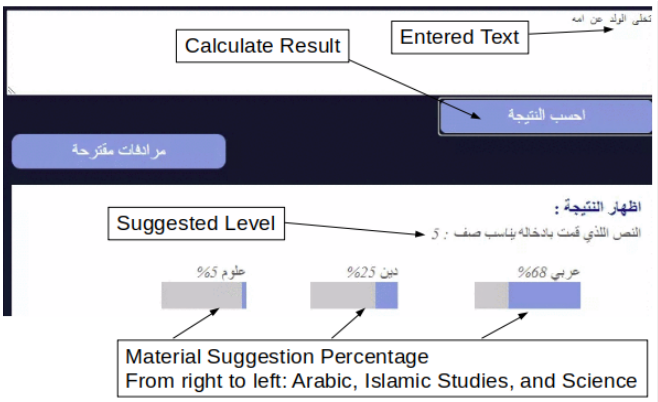

Fig. 2. A sample text with the results of matching with the three subjects: Arabic, Science and Islamic Studies.

As we mentioned before, for all concepts extracted from the different 3 subjects, we accompanied with each concept a set of lexemes to be used in the suitability suggestion activity. The users of the tool are able to ask for alternative meanings of some words to check the suitability of some other grades. Figure 3 below illustrates the ability to ask the tool to display all lexemes of some word that appears in the text. This enables the users to better choose words that are more suitable to grades that appear in Figure 4 [6].

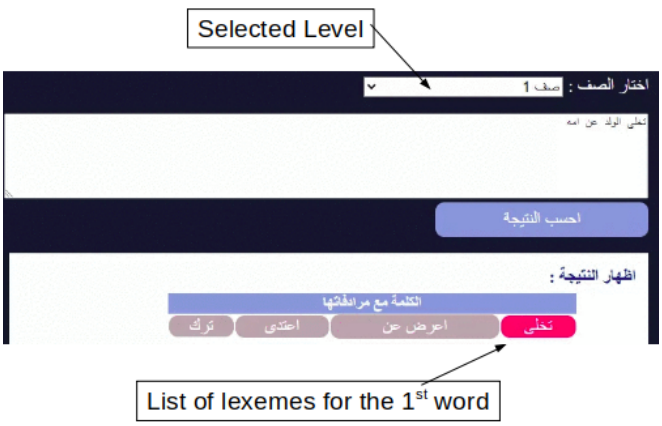

Fig. 3. The set of lexemes for words appear in the text.

[6] Suggested lexemes are listed according to the selected level.

Fig. 4. The result of suitability suggestion after selecting some lexemes.

Selecting a file to be examined against the 12 grades is shown in Figure5 while selecting a folder to check all files that exist in it is shown in Figure 6.

Fig. 5. Measuring the suitability of a complete file.

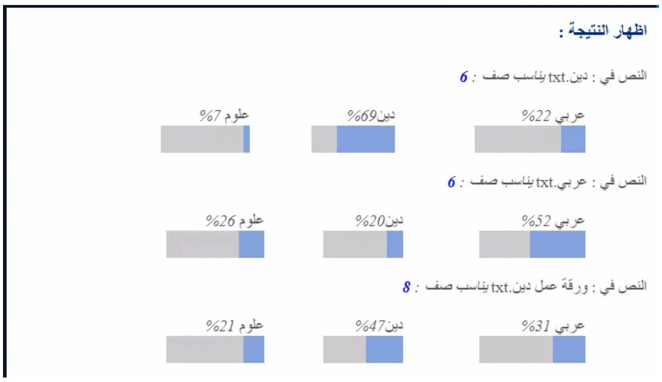

Fig. 6. Measuring the suitability of a list of files.

Because it is usual for some grades to have different students' educational levels within the same class, the tool is able to categorize the tested text into two categories (Intra-level): *strong* and *weak*. Figure 7 depicts this facility.

The sequence diagram in figure 8 depicts the actions performed to make the grades suggestion upon providing the tool with a sample paragraph. The activity begins when the user asks the system to suggest the suitable grades by applying the function *reuwstGrade(Paragraph)* where the sample paragraph is parameterized for the function. Upon this request, the object *Web Browser* takes the paragraph and using a special JavaScript code executed the function *sendText(Paragraph)* to the object *PHP Code* that executes

Fig. 7. Intra-level facility: the text here goes to strong students.

the function *Tokenize(Paragraph)* to extract the important tokens in the paragraph and save them in the array *Tokens[]*. After that the function *getFrequencies(Tokens[])* is executed to get all frequencies in the database for all tokens appear in the array *Tokens[]*. The retrieved frequencies are returned to the object *PHP Code* in the 2-Dimensional array *Frequencies[][]* in which each token is mapped with its frequency. This array is used by the function *computeGrades(Frequencies[][])* returns a list of numbers represent the amount of association between the paragraph and 12 grades all saved in *Grades[]* array that is later used by the function *computeMaximum(Grades[])* to return the most suitable grade to the paragraph saved in the variable *Grade* in which it is returned to *Web Browser* object that notifies the user by the return back function *notifyUser(Grade)*. Of course, these sample activities that appear in the sequence diagram can be generalized in the case the parameterized paragraph becomes a text file.

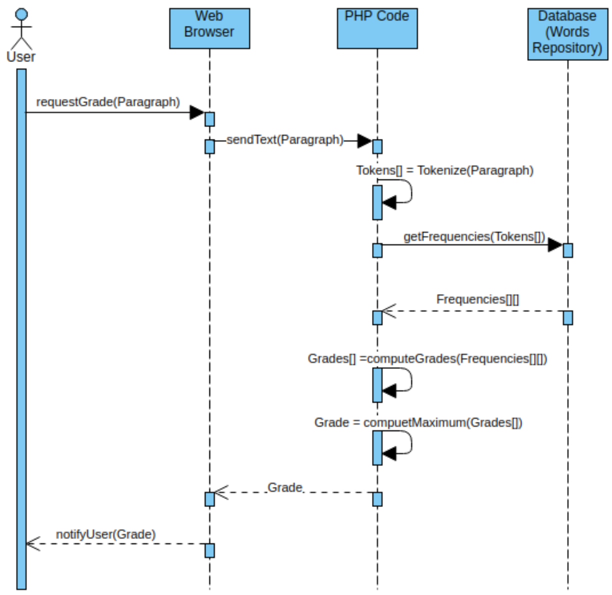

Fig. 8. Requesting The Suitable Grade Sequence Diagram.

The database structure of the tool is very simple. It is composed of a main table named *Subject* with four fields: *Term*, *Islamic*, *Science* and *Arabic*. This table contains unique words (terms) and the frequency of each one in the 3 different subjects. The table contains 83672 different words for all 12 grades.

VII. Experimental Tests

In order to test the implemented work, we conducted a comprehensive test that is composed of two different parts:

1) *Stories Suitability*: Matching some stories with the different grades.
2) *Exams Suitability*: Check the suitability of some exams against the grades.

In the first part of the test, we asked a set of school teachers to choose different stories (25) that were used in teaching (the teachers previously knew their suitability with grades) and ask the tool to check their suitability against the different grades. By this, we are examining the amount of correctness of the tool since we are comparing its results with the results of experts (teachers). Figure 9 depicts the results of this test where the x-axis represents the number of samples used in the test and the y-axis represents the 12 grades. The blue line depicts the suggested grade of the tool for each story while the other line represents the actual grade (from teachers) of each story. From the graph, there are only 5 samples that were exactly correct, 12 samples with a difference of one grade, and only one sample with a difference of three grades. The amount Correlation Coefficient between the expected and actual results is 0.84. The value is close to 1 and positive which means a strong correlation between the expected and the actual grades.

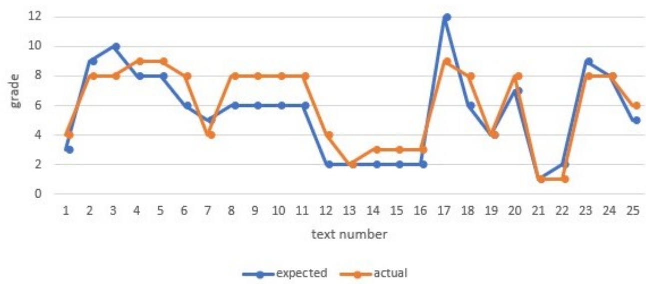

Fig. 9. Expected vs. Actual grades for a set of stories.

The second part of the test is related to examining the suitability of some exams with the 12 grades. Different levels of exams (Arabic 33, Science 18, and Islamic Studies 22) have been selected by some teachers (experts) and the tool is asked to suggest the suitability of each exam with the grades. Figures 10, 11 and 12 depict the result of this test where the figures illustrate the amounts of differences between the levels of the expected and actual exams for the 3 subjects: Arabic, Science, and Islamic Studies. The Correlation Coefficient for the Arabic exams is 0.76, the Science exams 0.08, and the Islamic Studies 0.17.

Fig. 10. Expected vs. Actual grades for a set of Arabic exams.

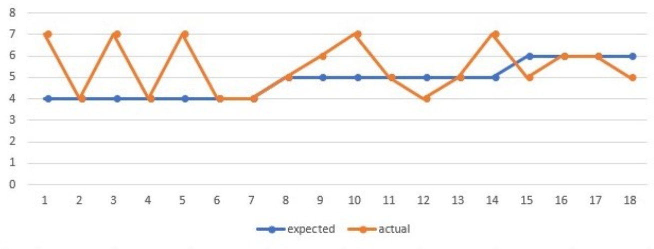

Fig. 11. Expected vs. Actual grades for a set of Science exams.

VIII. Conclusion

This work is related to enhancing the readability comprehension of Arabic texts for students at school levels. In this work, we implemented a tool to match Arabic materials to the least suitable school level using three different subjects: *Arabic*, *Science*, and *Islamic Studies*. We updated the *Additive Method* by multiplying frequencies of words by a special value k that is found by computing the differences of actual frequencies of words between the consecutive school levels where terms are found. Lexemes are also used in the work to enhance the matching process. A comprehensive test is conducted in which promising results were achieved.

For future works, we are planning to include all studying subjects so that we will have a bigger collection of words, for example substituting the science subject with specific ones like Biology, Chemistry, and Physics. Synonyms, antonyms, and hyponyms will also be considered to increase the variety of Arabic text styles. We are planning to include Named Entity Recognition to handle the out of vocabulary problem as well as consider the n-gram language model and phrases manipulation. Scientific text classification for large-scale environments will be one of our future works as well, where SciBERT tool will be involved to deal with Arabic texts.

References

[1] S. Sadatmir, V. Nimehchisalem, and A. N. Abdullah, “Validation of ‘self-perception of esl learners’ listening comprehension problems and metacognitive listening strategy use’questionnaire,” *Advances in Language and Literary Studies*, vol. 9, no. 6, pp. 158–171, 2018.

[2] F. Jeremy, “Educating students with learning disabilities in inclusive classrooms,” 2018.

[3] M. T. Davis and J. Southward, “Practical strategies for conducting and documenting classroom observations in sld evaluation,” *The DiaLog*, vol. 48, p. 4–8, 2019.

[4] S. V., *Analyzing Text Complexity and Text Simplification: Connecting Linguistics, Processing and Educational Applications*. PhD thesis, 01 2015.

[5] C. Elbro, “Dyslexia as disability or handicap: When does vocabulary matter?,” *Journal of Learning Disabilities*, vol. 43, no. 5, pp. 469–478, 2010. PMID: 20445205.

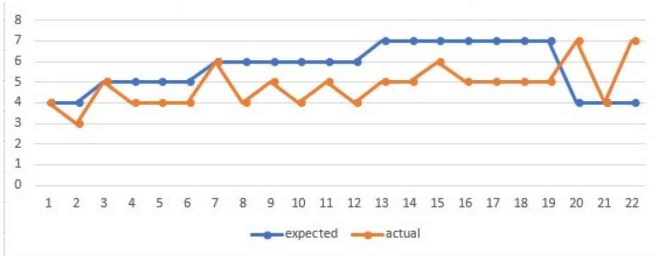

Fig. 12. Expected vs. Actual grades for a set of Islamic Studies exams.

[6] R. v. Nikijuluw, “Enriching students’ vocabulary by implementing semantic mapping strategy at the tenth grade of smk negeri 1 ambon,” 01 2018.

[7] Q. Al-Radaideh, “Applications of mining arabic text: A review,” in *Recent Trends in Computational Intelligence* (A. Sadollah and T. S. Sinha, eds.), ch. 5, Rijeka: IntechOpen, 2020.

[8] K. Aesaert, R. Vanderlinde, J. Tondeur, and J. van Braak, “The content of educational technology curricula: a cross-curricular state of the art,” *Educational Technology Research and Development*, vol. 61, no. 1, pp. 131–151, 2013.

[9] K. Collins-Thompson and J. Callan, “Predicting reading difficulty with statistical language models,” *JASIST*, vol. 56, pp. 1448–1462, 11 2005.

[10] S. J. Amendum, K. Conradi, and E. H. Hiebert, “Does text complexity matter in the elementary grades? a research synthesis of text difficulty and elementary students’ reading fluency and comprehension,” *Educational Psychology Review*, vol. 30, pp. 121–151, 2018.

[11] W. A. Gale and G. Sampson, “Good-turing frequency estimation without tears,” *J. Quant. Linguistics*, vol. 2, pp. 217–237, 1995.

[12] A. Odeh, A. Abu-Errub, Q. Shambour, and N. Turab, “Arabic text categorization algorithm using vector evaluation method,” *International Journal of Computer Science and Information Technology*, vol. 6, p. 83–92, Dec 2014.

[13] A. M. F. A. Sbou, “A survey of arabic text classification models,” *International Journal of Informatics and Communication Technology*, vol. 8, pp. 25–28, 2019.

[14] X. Wan, “Automatic Text Simplification,” *Computational Linguistics*, vol. 44, pp. 659–661, 12 2018.

[15] N. Erlina, “The increase of students’ arabic reading skills through the application of sq4r learning technique,” *International Journal of Language Education and Culture Review*, vol. 4, pp. 37–57, 2018.

[16] M. Emam, A. Kazem, T. Al Said, W. Al-Maamary, and R. Al-Monzery, “Variations in arabic reading skills between normally achieving and at risk for reading disability students in second and fourth grades,” *Journal of European Studies*, vol. 6, pp. 17–30, 08 2014.

[17] D. Balota, A. Aschenbrenner, and M. Yap, “Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations,” *Journal of experimental psychology. Learning, memory, and cognition*, vol. 39, 04 2013.

[18] A. Yuan, N. Jeffries, and G. Zheng, “Maximum likelihood,” in *Brenner’s Encyclopedia of Genetics (Second Edition)* (S. Maloy and K. Hughes, eds.), pp. 328 – 329, San Diego: Academic Press, second edition ed., 2013.