Journal of Theoretical and Applied Information Technology
31st January 2021. Vol.99. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SPELLING CHECKER FOR DYSLEXIC SECOND LANGUAGE ARAB LEARNERS

SAFAA. M. EL ATAWY[1], HOSNIA.M.M.AHMED[2]

[1] Faculty of Specific Education, Computer Teacher Preparation Dept., Damietta University, Damietta, Egypt.
[2] Faculty of Specific Education, Computer Teacher Preparation Dept., Mansoura University, Mansoura, Egypt.

E-mail: zizoabdo1210@gmail.com

ABSTRACT

People with dyslexia have a lifetime of a poor spelling problem. Several researchers have tried to solve this problem through spell checkers. However, most of these spell checkers are provided to native language owners and do not take into account second-language learners with dyslexia, especially since most of the mistakes they make real mistakes lead to unintended but correct words. So far, there is no root solution to this problem. This research is proposed a spelling checker (DYS-EnSC) based on n-gram technique, look up dictionary, and Damerau-Levenshtein, To create a list of candidates and choose the most suitable candidate for each misspelled word, to detect and correct misspellings of Arab second language learners with dyslexia. The results of this study are included two parts. In the first part, we focus on comparing the performance of GSC (Microsoft Word (MW), A spell, and Language Tool) with the performance of the proposed spelling checker DYS-EnSC. Standard measures (Recall, Precision, and Accuracy) were used. Results suggest that the proposed DYS-EnSC spelling checker was useful in correcting misspellings of second language learners with dyslexia, achieving an accuracy of 93% in detecting misspellings and corrected about 86% outperforming to MW, A spell and Language Tool. In Section 2 of the results of this study, we focus on how successful students with dyslexia were in correcting their misspellings with and without the proposed DYS-EnSC system. Findings suggest that students could correct 9.3% of their misspellings without DYS-EnSC and 86% with it.

Keywords: *Spell-checker, computer assisted language learning, L2 spelling acquisition, Text processing, Dyslexia, N-gram, Damerau-Levenshtein distance, Spelling detection and correction, Spelling Errors.*

1. INTRODUCTION

According to the National Institute of Neurological Disorders and Stroke (NINDS), learning disabilities are disorders that influence a one's ability to understand or utilize spoken or written language and do mathematical calculations, and hence make one unable to learn [1]. Dyslexia is considered the most common learning disabilities among learners [2] and the language-related learning disability that is most common and widely spread [3]. Dyslexia is a health condition with effect on neurobiological functions and is distinguished by one's inability to read and/or spell accurately and / or quickly [4]. The skill of spelling words correctly may be one of the highly appreciated and difficult skills in written communication [5].

A student with dyslexia undoubtedly has a spelling difficulty [6, 7]. As English has become a global language in recent years, it turned to be an essential part of the school curriculum and is being taught as a second language in all learning levels [8]. Non-native English students with dyslexia suffer from the difficulty of learning how to write in English. The essential challenge these students face is that they do not consciously discover spelling mistakes. Therefore, what students with dyslexia write has more mistakes than normal students [3]. Hence emerged the necessity of using spell checkers to help these students to reduce their spelling mistakes and increase the quality of their writing quality.

Computer spell checkers turned out to be a very useful tool for writers and students. These checkers make them able to write texts containing lower error rates, but the case is not like this with students with dyslexia because conventional spell checkers will not be able to discover their spelling mistakes, so, these checkers suggest wrong or misleading corrections [9]. This made devising sophisticated algorithms and methods for automatic text

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

correction a challenge that researchers face constantly. Continuous research in this area is well justified, to upgrade quality, improve performance and widen the prospects of possible applications [10]. Automatic spelling correction is the process of discovering and correcting spelling errors in a text. It is one of the most essential tools in the processing of natural language [11,12] and there are many applications related to it, such as web search query, writing systems, recommend systems, document mining and typos checking before printing and they do a function very similar to spelling correction [13]. Several researches asserted automatic correction system can raise the motivation of students with dyslexia to write, which upgrades the quality of their life and the quality of their writing too. So, this study is motivate by the quest for developing an automatic spelling checker system of texts written in English to help dyslexic students who are Arab non-native students of English to reduce their spelling mistakes and upgrade the quality of their writing. The remainder of this paper is organize as follows: Section 2 discusses relevant work; Section 3 discusses the study problem. Section 4 shows the theoretical background. Section 5 discusses materials and methods; Section 6 analyzes the performance and results of the proposed system; finally conclusion and future work.

2. RELATED WORKS

Research on automatic spelling errors discusses either error detection, correction suggestion, or context-sensitive methods. There are many approaches can be used to detected or corrected misspellings in a text such as looking up the word in a specific dictionary supposed to reflect textual language [14] , edit-distance, n-grams, and noisy channel. Several studies try to correct misspellings in a certain language through various methods. For instance, (Farag et al, 2009) proposed a language independent spellchecker built on enhancing an n-gram based model. The spellchecker suggests corrections for misspellings by picking the most possible words from a list of ranked correction candidates. The suggested correction is derived based on n-gram statistics and lexical sources. The researchers also outlined their evaluations on English and Portuguese referential datasets of misspellings. The results of implementing this method outperformed other latest methods [10]. (Aadil and Bipul, 2016) discussed the complete design and implementation of a spellchecker for Kashmir. The proposed Kashmir spellchecker is a Standalone application and not a part of any word processor. This system have taken care of only non-real word errors. The spellchecker detects about 80% of the misspellings and offers 85% of the suggested corrections [15]. (Nejja and Yousfi, 2018) developed a computerized dictionary prototype for Arabic that restricts access to dictionary information closest to a wrong word using the modified Levenshtein distance to Arabic to detect those closest words. This study using adapting Levenshtein algorithm to meet their requirements and presenting a dictionary management method to minimize the number of comparisons by studied organization and morphology. For organization, suggested a vocabulary management that allows the number of accesses to the lexicon file to be optimize. For morphology, used the derivation concept to minimize the number of comparable words avoiding redundant comparisons [11].(Mridha, et al 2019) proposed an approach to discover missing words and suggest a list of corrections with 82.82% accuracy using the n-gram model to identify any word missing between two words in a sentence. Using probability scoring, they ranked the suggestion list, having identified possible words for the missing word. They utilized a corpus to decide the bigram collection and another corpus to identify the preferred term for missed word, which is a trigram collection. Finally, they used six other corpora to assess their proposed technique. They created all corpora with online collected data [16]. (Maha and William, 2019) proposed an automatic-correction system to discover and correct dyslexic Arabic misspellings using a linguistic model built on the Prediction-by-Partial-Matching (PPM) text compression scheme to generate probable substitutes for every misspelling. The correct substitute for each misspelling is selected based on the length of the trigram compression code. The system is compare with the Farasa tool. The system achieved better results than other tools (recall of 43%, precision 89%, F1 58% and accuracy 81%) [4]. (Miguel and Catherine 2019) assessed the efficacy of spellchecking software developed for L2 learners by randomly selecting 30 texts written by learners of Spanish as a Foreign Language (SFL) from a corpus and analyzing misspellings in them and classified and inputted them into 3 specifically designed spellcheckers to compare their efficacy in discovering misspellings and presenting appropriate feedback with a GSC. Although all three spellcheckers discovered over 85% of the misspellings, they all failed to propose the appropriate substitute for one third of the misspellings. The GSC provided the correct alternative in the rate of (67%). Moreover, the feedback produced by the specialized spellcheckers is restricted to a list of possible alternatives [17]. (Rello,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

L., et al, 2015) presented a spell Real Check that used a statistical dependency parser, a probabilistic language model, and Google n-grams to detect real-world mistack. In the experiment, 17 with dyslexia corrected the sentences in less time with Real Check and more accurately [3].

This study differs from previous work in many regards. First, the proposed correction system was developed for non-native writers, whereas other related studies show that most automatic correction systems were developed for native writers. So, they assume that most misspellings are typographic mistakes in nature, rather than mistakes resulting from the lack of spelling knowledge. Accordingly, when these systems detect a misspelling, they either correct it automatically or suggest a list of alternatives for the misspelt word or a string of characters similar to that of the misspelt word. A native writer will easily select the correctly spelt word of the said menu, if it contains one. However, the task will not be that easy for a second language learner. Hence, our study deals with this issue. Second, most mistakes dealt with in previous studies were related to punctuation and syntax, whereas our study specifically deals with spelling mistakes committed by dyslexic learners of English as a second language, as dyslexic learners have a higher incidence and intensity of misspellings compared with normal learners. Unique English misspellings committed by non-native speakers: the spelling of certain words may contain capital or small letters. Some dyslexic learners have difficulty in distinguishing between the shape of the capital and small cases of the same letter such as ‘a’ and ‘A’. They also cannot distinguish between b and d. It is difficult for them to link a certain sound with its orthography, such as the sound of z in a word like “rose” (pronounced as ROZ). A certain letter may be used in several phonetically syllables. There are also the vowels (a, o, u, I, e), which are used in about 20 different sounds. Certain letters are pronounced differently based on their place in a word (such as the ‘a’ in cat and tail).

The contribution this paper makes is that it proposes an new system called DYS-EnSC for the automatic detection and correction of misspellings made by Arab dyslexic learners of English, based on the statistical approach, n-gram similarity coefficient, look up dictionary Damerau-Levenshtein distance approach to make a list of possible alternatives of the misspelt word, n-gram probability to select The correct alternative for each misspelled word. In addition, this system analyzed the efficacy of using real writings by dyslexic Arab learners, rather than an artificial set of writings, and compared the performance of this system with the performance of three well-known spell checkers.

3. THEORETICAL BACKGROUND

3.1 Error detection and error correction.

Usually students with learning disabilities (LD) misspell 10-20% of their words [18], which limits their writing vocabulary and may make them terminate their writing activity early. So, students with learning disabilities are advised to use word-processing software that contain a spelling checker [19]. A spelling checker is an application that detects misspelt words in a text [20]. Spelling checkers are generally used to detect and correct writing mistakes. We should differentiate the correction and detection of errors. Zamel [21] asserts that irrespective of their knowledge words, when students are unable to correct words, this may be because they cannot detect errors in words to correct them. Plumb [22] found that the most difficult obstacle regarding error correction was not the knowledge of the right way to correct them, but rather an inability to discover these errors. Seen from the perspective of the knowledge-deficit hypothesis they posed, in studies [23, 24, and 25] they indicated that the knowledge of the errors does not necessarily result in adequate automatic correction. Although spell checkers are supposed to be variedly helpful for most writers, they look to be most helpful for students suffering the most in spelling. For them, spell checkers might not only make their spelling better, but also motivate them to utilize a larger vocabulary and write lengthier pieces. The most intuitive question regarding the use of spell checkers is how effective they are in aiding students with varying spelling skills make up for their spelling problems and correcting their misspellings.

3.2 Generic Spell Checkers.

Although students who suffer learning disabilities, especially dyslexic ones, can make use of generic spell checker, there are some restrictions. These include that checkers are made for native writers, and therefore suppose that errors are typing mistakes such as (deleting, changing or adding a letter to a word), rather than because students do not have adequate knowledge of spelling [17]. In this case, the checker shows a list of alternatives with a sequence similar to the one containing the misspelling [26]. It will be easy for native speaker to select the right word if it is provided among these alternatives. However, the case is different with a second-language speaker [27], especially dyslexic persons. Lee [28] thinks that the most important difficulty learner's face when correcting errors is the ability to discover errors. The reason behind this is that GSC are made to correct mistakes made by adults who are somewhat specialists and native speakers of the language, rather than for students who learn it as a second language, especially if they are dyslexic [19]. The objective of this study is to design and implement a system to check spelling and automatically correct any misspelt words for dyslexic learners and compare the performance of the system with the performance of GSC in terms of suggesting correct alternatives for misspelt words dyslexic students and in terms of student ability to use the system to correct their own errors.

4. MATERIALS AND METHODS

A system creating called (Dys_EnSC) to propose spelling correction of misspellings of non-native English students with dyslexia. In the beginning we collected a group of texts written by Dyslexic Second Language Arab Learners to find patterns of errors for Dyslexic, so we can develop the system, it was found that the mistakes they make are considered forgetting, adding or replacing a letters. The system work begins by entering the text to be checked, that works in three steps: The first step is text pre-processing. The second step is Use a statistical model to discover dictionary errors, finally step is Use an edit operation approach to correct errors, Build a list of candidates, selecting the correct proposal from the list of candidates using N-gram probability. Figure 1 explains the flowchart of the proposed system.

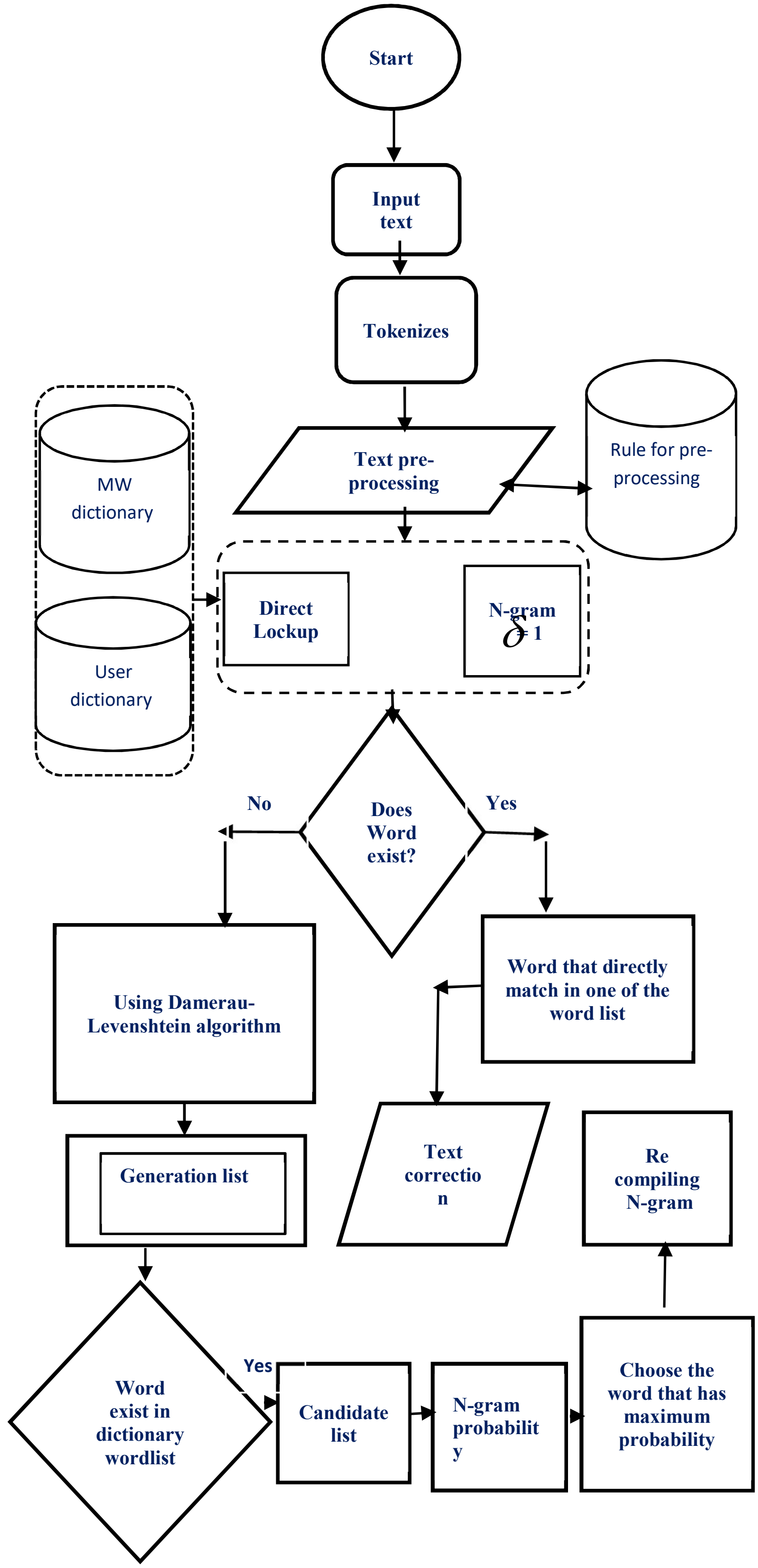

Fig.1 Work Flow Of The Proposed System

4.1 Text pre- processing

After analyzing the dyslexic errors was found, and before implementing the process of discovering and correcting errors, it was found that there are a number of problems that can increase the complexity of the correction process, including repeated letters and the division of words that contain prefixes and suffixes. In the text preprocessing step the excrescence characters in the word were corrected. In English, there are words in which a letter can be repeated twice in a Consecutive, such as ' bigger' a letter 'g' that is repeated twice, and it was also found that words do not contain more than two consecutive letters that are repeated, so the system deletes the consecutive repeated letters that exceed two letters. It also showed through the analysis of errors that students with dyslexia often tend to put a space before the suffixes and after prefixes, which affects the learning of non-native English language dyslexia students such as ‘ drinkable’, When students write it, they put between ' drink ' and 'able' a space. To reduce complexity, collected prefixes and suffixes with words. The table 1 shows the most common prefixes and suffixes of dyslexic corpus analysis.

Table .1 the most common prefixes and suffixes of dyslexic corpus analysis.

prefixes with verbs	Suffixes with verbs	prefixes with noun	Suffixes with noun
re - out	ise	co - mini	Tion- ship
Dis - sub	en	Sub - mis	Ity - age
Over - co	ate	Anti-super	er - ery
un - pre	Fy	auto - tele	Nes-sm
Mis- inter		Bi - tri	ment - ant

also, Students' writings contain special signs, such as (□□, □□, □□, □□, □□, $, @ , ? , : , " , # , …………). To reduce the processing time, this system deleted these signs in the pre-processing step.

4.2 Detection spelling errors

Detection of spelling error is a process that determines if the word is the word or is considered an error [29]. Several techniques are used to detect errors such as n-gram [30], morphological analysis [31], dictionary lookup [32], latent semantic analysis method [33], co-occurrence method [34], and context-vector method [35].

This study based on the n-gram using with dictionary lookup method. The reason for this is that the dictionary lookup method enables us to deal with the most frequently used words in the language, fast technique, can be built a dictionary using any language, in addition to the ability to adapt the lookup as needed. In this suggested method, we used the MW dictionary as part of the spell checker and served as a database for the proposed spelling checker because it is an important dictionary that contains most common words and covers verbs, names, conditions and attributes. However, it may not include specific names, technical terms, acronyms or specialized initials. Therefore, we created a special vocabulary dictionary for high school students, which include names entities that common using in the language, and some special terms. This system is computed the coefficient similarity between two strings, In general, the similarity coefficient δ is calculated by Equation (1).

$$\delta(w_q, w_t) = \frac{| w_q \cap w_t |}{| w_q \cup w_t |} \quad (1)$$

Where w_t and w_q are the n-gram sets for two words w_t and w_q which they compare. $| w_t \cap w_q |$ indicates the number of similar n-gram in w_t and w_q , and

$| w_t \cup w_q |$ indicates the number of unique n grams in the union of w_t and w_q

Dice’s coefficient is used to measure the similarity between two strings based on The N-gram. Consider the word START whose bigrams are: *S,ST,TA,AR,RT,T* To measure the similarity between the words START and STARTING, we can use Dice’s coefficient in the following way. Find all the bi-grams from the word starting

*S, ST, TA, AR, RT, TI, IN, NG, G *

The number of unique bi-grams in the word START is 6 and in the word STARTING is 9. There are 5 common bi-grams in both the words. Similarity measured by Dice’s coefficient is calculated as 2A /

(B+ C), where B and C are the number of unique bigrams in the pair of words; A is the Parts of the common bigrams between the two words.

The proposed detection errors approach is described in Algorithm (1).

Algorithm 1 detection errors approach

For each word w in text t do :

Calculate δ(w) by equ1

Look up w in dictionary

If w not found & δ(w)<1

Then

w is misspelling

End if

End for

4.3 Generation Suggestion Candidate list

By analyzing the list of linguistic errors of students who suffer from dyslexia, it was found that word errors may be substituting one letter for another, deleting a letter, inserting a letter, or transposing two letters. So, the method of edit distance was used, because it is the common method concerned with transforming one word to another. we used the Damerau-Levenshtein algorithm [36] because the Levenshtein distance is the most popular modification distance in the operations of deleting, inserting, and replacing letter , and Damerau-Levenshtein algorithm increases that the operation of transposing two adjacent letters is considered one operation. Based on [41] the equation for Damerau-Levenshtein can be seen in (2).

$$d_{a,b}(i,j)=\left\{\min\left\{\begin{array}{l} d_{a,b}(i-1,j)+1,^{\max(i,j),if\ \min(i,j)=0} \\ d_{a,b}(i,j-1)+1 \\ d_{a,b}(i-1,j-1)+1_{(a_i\neq b_j)},if,i,j>1\&a_i=b_j-1\&a_i-1=b_j \\ d_{a,b}(i-2),(j-2)+1 \quad \min\left\{\begin{array}{l} d_{a,b}(i-1,j)+1 \\ d_{a,b}(i,j-1)+1,else \\ d_{a,b}(i-1,j-1)+1_{(a_i\neq b_j)} \end{array}\right\} \end{array}\right\}^2\right\}$$

By using Damerau-Levenshtein this system creating all possible candidate words for every error word based on if the length of correction candidate word more or less than 1, compared with the word' error then is dismissing edit distance.

4.4 Correction spelling error

The most studied spelling correction algorithms are: edit distance, similarity keys, rule-based techniques, probabilistic techniques, neural networks and n-gram-based techniques [37, 38].

Essentially, the n-gram model is a probabilistic model based on Hidden Markov Model (HMM), originally devised by the Russian mathematician Andrey Markov in the early 20th century and later extensively experimented by Shannon and Chomsky for predicting the next item in a sequence of items. The items can be letters, words, phrases, or any linguistic entity according to the application [39].

A character-based N-gram is a set of N consecutive characters extracted from a word. Typical values for N are 2 or 3 [40]; which correspond to the use of bigrams or trigrams, respectively. Once the candidate list is created, the system uses the bigram model to calculate the probability of word correction. Based on [41] the general N-gram probability can computed by equation 3.

$$p(c_n \mid c_{n-N+1}^{n-1})=\frac{C(c_{n-N+1}^{n-1}c_n)}{C(c_{n-N+1}^{n-1})} \qquad 3$$

Where c_n is a given character & c_{n-N+1}^{n-1} is a fixed-size character of a prefix in corpus C. If the word is more probability, based on n-gram matching, it is specified as a proposed correction. This method depends on the relative frequencies in estimating the probability.

A maximum of five words as a correction are chosen for each misspelling word. Then select the highest

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

probability to be auto correction of misspelling. The proposed system accepts a paragraph and divides it into words and each word divides it into parts bigram and corrects each word and re compiling it again using the chain rule.

4.5 The System GUI.

The system was implemented using MATLAB 2014b. The proposed DYS-EnSC spelling checker. When it is running, the window shown in Figure (2) will appear. Here, the user can input the text to be spell-checked in the space dedicated for this. When "check words" is clicked, the system will begin processing the word and present a list of suggested corrections. The system will present the correct word along with the full correction of the input text in the spaces dedicated for this on the window, which also contains a "new" button to delete the input text and input another one. It also contains an "Exit" button to shut down the program.

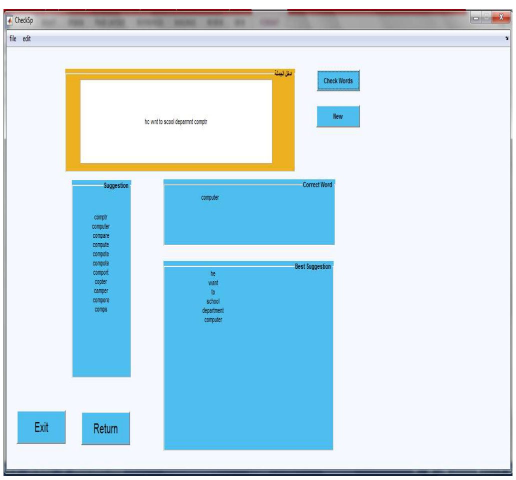

Fig.2 Proposed Spelling Checker (Dys_Ensc) Screen.

4.6 Participants

The Participants consisted of 60 male and female secondary students, 16-19 years old (mean= 17.35, SD= 1.16), in Damietta (32 males and 28 females) with diagnosed dyslexia. All students were Arab speakers and English was their second language. Dyslexia is defined as a disability to learn to acquire and process language, usually manifested in a lack of reading, spelling and writing skills [42]. The IQ of a person suffering from this disorder will be normal or extraordinary. In students aged 16-19, dyslexia manifests itself in forgetting or replacing certain letters of written words. All students got an IQ between 80 and 125 and the reading test results were at least two years before grade average.

5. RESULTS AND DISCUSSION

The findings of this study consist of 2 major parts:

First: those related to the performance of the proposed spelling checker (Dys_EnSC) in terms of automatic correction of misspellings of dyslexia students and comparing its performance with other spelling checkers (MW, A spell, Language Tool).

Second: The findings related to how successful are dyslexia students in using the spelling checker to correct their misspellings.

5.1 Part 1

The first part of the findings focuses on answering the following questions:

Q1: How successful are spelling checkers in identifying and correcting common spelling mistakes made by dyslexic learners of English?

Q2: Are there performance differences between spelling checkers?

Q3: How accurate are the alternatives provided by the proposed spelling checker and at which place does the correct alternative appear in the suggestions list, compared to the alternatives provided by other spelling checkers?

To answer these questions, we evaluated the performance of our proposed spelling checker, DYS_EnSC, specifically designed for dyslexia students, and compared it with free spelling checkers (MW, A spell, Language tool). These checkers were selected, as they are most common and widely used in our Arab region. To evaluate our approach, we took sample writings of participating students as our datasets. All writing samples were administered in students' classrooms, after we required the participants to write a paragraph of 30 lines to express "the importance of reading" and to comment on a photo in at least ten other lines. After students finished writing and commenting, the researchers read students' writing and asked each student to read out the words that are illegible because of the handwriting or spelling issues. The total words of the writing samples were 500.

Misspelt words were marked and counted by 2 assessors. They amounted to 190 words with containing a single misspelling or up to three misspellings per word. These misspellings included addition, deletion and replacement errors. We deleted 20 misspellings for other correctly written words from the dataset (such as “close”, which was replaced by “clothe”), as the spelling checker would not be able to detect these errors. These words were input to different spelling checkers (MW, A spell, Language tool) and to our proposed DYS-EnSC to evaluate our approach.

The evaluation approach used in this study depends on common natural language processing standards (NLP) (Accuracy, Recall, Precision), which we measured both for the detection and correction of misspellings.

A) Error detection evaluation

To evaluate the ability of our DYS-ESC to detect misspellings and compare it with other spelling checkers, we used the standard measures of Accuracy, Recall and Precision. Precision and recall are defined in terms of true positive (TP), false positive (FP) and false negative (FN) as shown in the following equations [43]:

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Where:

True Positive (TP): This means that a spelling error has successfully been detected.

False Negative (FN): This means that word is a misspelling and has not been detected.

False Positive (FP): This means that a correctly spelled word was detected as being a misspelled word.

True Negative (TN): This means that a misspelled word was detected as being a correct word.

Figure 3 shows the findings. As we can see, most misspellings were detected by all 4 spelling-checkers.

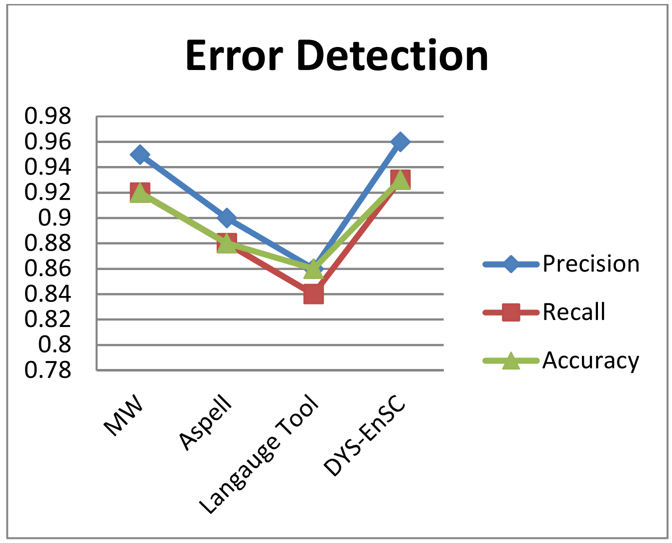

Fig. 3 Percentage Of Errors Detected By MW, A Spell, And Language Tool And DYS-Ensc.

From Figure 3, it is clear that Language tool was the least accurate in detecting misspellings (86%) whereas DYS-EnSC was the most accurate (93%) followed by MW (92%). A spell’s accuracy was 88%. It should be noted that in many cases the performance of our proposed system is close to, or even better than, the checkers with which we compared it, although MW outperformed our system in some cases.

A) Error correction evaluation

Error correction was evaluated to identify whether the system successfully corrected the detected misspelling or not. Precision, Recall and accuracy can then also be calculated using the same equations shown above.

Where

True Positive (TP): This implies that a spelling error was successfully corrected.

False Negative (FN): This means that a misspelling word and was not corrected.

False Positive (FP): This implies that a correctly spelled word was changed.

True Negative (TN): This implies that a correctly spelled word was not changed.

Results shown in Figure 4 were achieved.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

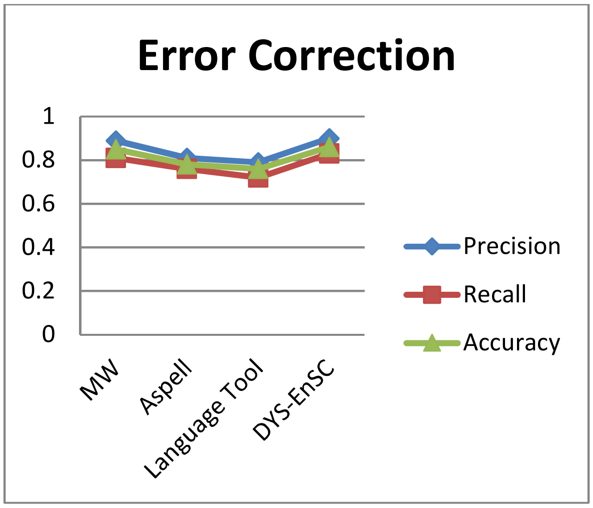

Fig. 4: Percentage Of Errors Corrected By MW, Aspell, And Language Tool And DYS-ESC.

From Figure 4 it is clear that DYS-EnSC achieved the highest accuracy (86%) in correcting misspellings, because it depends on MW dictionary, along with a dictionary developed by us containing words appropriate to the subjects' curriculum. In addition, it provided a list of suggestions and the advantage of automatic correction of the words on the text. MW came second in terms of correction (85%). It provides a list of suggested words that vary in length, along with auto correction. Aspell checks certain misspellings and suggests alternatives. Often, it corrects less word automatically (78%), whereas Language Tool is the least checker in terms of correction (76%). Language Tool corrects grammar and syntax errors in different languages. However, it only shows the unclear message "Possible misspelling" with a list of possible alternatives, same as with MW (see Figure 5). Although Aspell provides comments on the misspelt word, it is unlikely to be useful for students learning English as a second language, especially those with dyslexia. In addition, it does not provide auto correction.

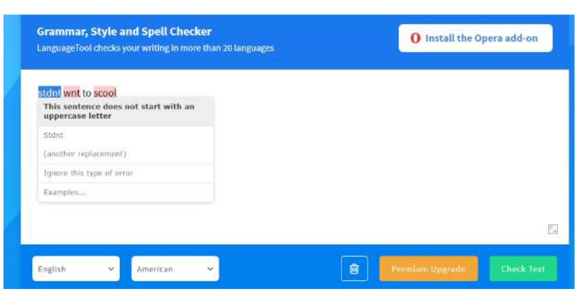

Fig 5: Example Of The Feedback Offered By Language Tool.

It is noteworthy that the only spelling checker that provides competitive result, compared with our proposed system, is MW. However, Word corrected less misspelling than our proposed system, as it achieves highly accurate targeted results. However, as students with dyslexia cannot detect errors consciously, we suppose that a much-targeted system, such as ours, will be useful for this target group. Thus, we answered Q1 and Q2.

To answer Q3, we analyzed the accuracy of the list of alternatives provided by the four spelling checkers to identify the order and position of the correct word on the list of corrections. Table 1 shows the percentage of the order of the correct word on the list of suggested alternatives.

Table 1: The Percentage Of Position Of The Correct Spelling In The List Of Alternatives.

	MW	Aspell	Language Tool	DYS-EnSC
Never	15.06%	22.8%	24.1%	14.01%
First Position.	74.2%	0.01%	68.3%	74.9%
second P.	4.09%	0.4%	5.26%	4.91%
Third P.	3.73%	71.9%	2.03%	2.37%
Fourth P.	1.64%	3.83%	0.31%	0.76%
Fifth P.	0.28%	1,06%	0.00%	3.05%

From the above table it is clear that MW failed to provide the correct alternative in only 15.06% of the cases, with the correct alternative appearing on top of the list in 74.2% of the cases. However, Language Tool failed to provide the correct alternative in only 24.1% of the cases, with the correct alternative appearing on top of the list in 68.3% of the cases. Aspell rarely provided more than 2 alternatives, which made it rank third. Our proposed spelling checker (DYS_EnSC) provided more correct alternatives than Aspell and Language Tool. It also provided several (usually more than eight) alternatives. DYS_EnSC had the highest percentage (3.05%) of times in which the correct word was found in the fifth position.

Although there is, a greater chance the correct word will appear in the list of alternatives provided by DYS_EnSC, this might confuse ESL learners who will find themselves with several alternatives, as their realization for the mistake they made may not be very intuitive. Therefore, the spelling checker provides the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

option of suggesting best alternatives and replacing them automatically in the text. MW did not provide auto correction for many words.

5.2 Part 2

The overall effect of the spelling checker not only depends on its performance, but also on the student's ability to spell and detect the misspelt word. In this section of the study, we will try to answer the following questions:

Q1: How able are Dyslexia students to correct misspellings using the spelling checker?

Q2: How able are Dyslexia students to take alternative action when the correct spelling is not suggested?

Q3: How able are Dyslexia students to look for misspellings not identified by the spelling checker?

To answer these questions, we selected a random sample of the participants with experience in using word processors and spelling checkers. The sample consisted of 25 students (14 males and 11 females) divided into 2 Computer Labs each with 12-13 computers. Students were informed that the objective of the study was to identify the effect of using the spelling checker to help students write better.

These students were trained in our proposed spelling checker. For three months on a regular basis, (average 5 lessens per week, for 12 weeks). Students were able to write in several self-selected topics under the direction of the teacher. Students were required to write about a trip to certain tourist sites in Egypt. On the first day, with the help of the classroom teacher, the researcher made sure that students had the skills required to deal with the spelling checker to correct a document. A set of misspellings was identified to explain the uses of the checker. Correct suggestions were provided for certain misspellings, no correct suggestions were provided for other misspellings. No suggestions or correct words were provided for some misspellings, which were provided as mistakes, such as names.

The researchers and the teachers stated that all students were qualified to the technical work of the spelling checker. Each student worked on a computer under the supervision of the researchers and the classroom teacher to correct words incorrectly identified in the document by the spelling checker. The teacher would read the document sentence by sentence and identify the misspellings not detected by the proposed spelling checker. On the next day, each student worked alone. No formal data were recorded. On the next day, students were required to write on the topic in a document using the proposed spelling checker. After they finished writing the topic and reviewed their writing for 45 minutes in 2 days, the researchers read the documents to make sure that the words were not unclear. When an unclear word was encountered, the student was required to read it. On the following day, students received a copy of their documents and were required to underline any word they though incorrect and suggest corrections. A week later, students were required to use the spelling checker to correct their misspellings. The researcher recorded the spelling of the words identified by the checker, the intended word, the correct suggestion and its order on the list of suggestions.

The percentage of misspellings were calculated by dividing the number of misspelt words on the total number of words. These writings contained an average of 150 words (SD= 95.3) with 17.4% of the words misspelt (SD= 9.3%). The correlation between student writing samples was 4.9. These results suggest that the group as a whole suffer severe spelling issues.

To explore the relationship between the ability to spell words correctly and successful detection of misspellings,, the relationship between the percentage of misspelt words and the percentage of misspellings detected by the student was calculated, as well as the percentage of misspellings detected by the proposed (DYS-EnSC) system and the percentage of corrected misspellings. The results indicated that the students who committed increased errors 37.4 are able to correct less mistakes 8.9, those who have made fewer mistakes are able to correct more mistakes 9.3, and that the spelling checker is able to discover a greater percentage of errors. The Figure.6 shows the relationship between detection and correction spelling errors among students with dyslexia and spelling checker.

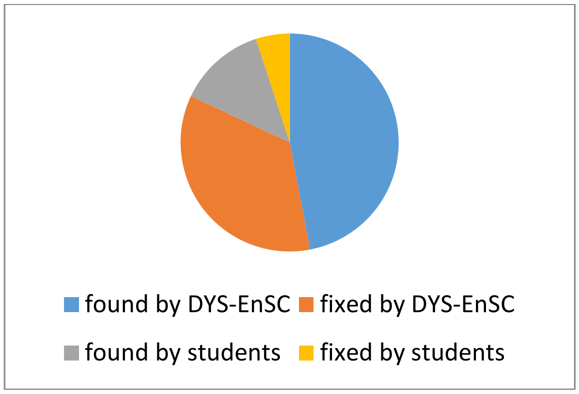

Fig .6 the relationship between detection and correction spelling errors among students with dyslexia and proposed spelling checker.

5.3 DISCUSSION

We compare the results with other similar approaches, and discuss the limitations of our system.

5.3.1 System Evaluation

Our system offers competitive results compared to other similar studies [3, 10, 15, and 17]. However, it should be noted that most of these systems are directed at original writers of the language and not second language learners. This makes the discovery process more difficult. In addition to the fact that most of these studies do not use a test group from the real world, while the use of sentences from the real world makes correction more difficult because the sentences taken from The real world is often similar in spelling, and linguistically as well. This means that the method proposed in our system was effective.

5.3.2 Spellcheckers comparison

Through the previous presentation in 5.1, it is clear that the only auditor who provides competitive results with our proposed system is spellchecker in Microsoft Office 2013, which provides the best results in detection and correction errors; however, the number of errors that cannot be corrected by MS spellchecker is less than what is in our proposed system. Nevertheless, the dyslexic second learner's cannot consciously spot errors [44]. Therefore, we assume that a spellchecker geared toward discovering and correcting their mistakes would be better than the general spelling checkers for this category.

5.3.3 Limitations

Although the system (Dys_EnSC) presented results with high accuracy and high recall, it could not discover some errors and was unable to correct some errors.

These errors can be explained as follows:

1- the system (Dys_EnSC) sometimes cannot detect some errors of the real- word because it is semantically correct such as (close) instead of (clothe) , one solutions is to use anaphora resolution as it in NLP tasks for word processing at the sentence level. It should be noted that there is no software, which used widely in text processing that can detect this kind of errors.

2- Another type of errors is the incorrect synonyms of the word such as (home) instead of (house). This problem can be overcome by using synonym generation systems.

3- in some cases, adding a character or deleting a character may lead to an identical word for a word in the dictionary, therefore the system (Dys_EnSC) cannot detect it as an error, such as (buy) when deleting a character 'u' that becomes 'by' .Also word boundary errors such as 'manmade' and ' man made'. To solve this problem, it can be used pairs instead of tokens through NLP to find more an effective way to represent texts.

6. CONCLUSION

Word processing is one of the most important areas of natural language processing, and spell checkers have become an indispensable tool in modern writing environments where all computerized and smart devices are equipped with spell-checking tools. This study suggested a spelling checker with a high frequency of detection and correction of errors, as it detects about 93% of the spelling errors written by 60 students who are dyslexic and correct about 86% of this error. However, detection and correction can be improved by conducting detailed studies on the errors that make by dyslexic second language arab learners and enriching the database with correct words. The ability to detect can also be improved by expanding the system to detect context errors, which are errors that are correct in spelling form but are wrong in the current context. This might make doing spell checking less complicated for dyslexic second language arab learners. The results confirm the learners' ability to use a spelling aid tools and confirmed that there is a difference between the ability to detect and correct errors. We intend to implement the system in future for Arabic Language.

REFERENCES

[1] Serge A. Afeli, Academic accommodation strategies for pharmacy students with learning disabilities: What else can be done?, Currents in Pharmacy Teaching and Learning 11 (2019) 751–756.

[2] Pavel Zikl, Iva Košek Bartošová , Kateřina Josefová Víšková , Klára Havlíčková , Alice Kučírková , Jolana Navrátilová , Barbora Zetková, The possibilities of ICT use for compensation of difficulties with reading in pupils with dyslexia, Procedia - Social and Behavioral Sciences 176 (2015) 915 – 922.

[3] Luz Rello, Miguel Ballesteros, Jeffrey P. Bigham, A Spellchecker for Dyslexia, Proceedings of the 17th International ACM SIGACCESS Conference on Computer & Accessibility, October 2015, pp. 39-47.

[4] Maha M. Alamri and William J. Teahan, Automatic Correction of Arabic Dyslexic Text, Received: 31 December 2018; Accepted: 1 February 2019; Published: 21 February 2019 , Computers 2019, 8, 19.pp.

[5] Jeanne wanzek, sharon vaughn, jade wexler, elizabeth a.swanson, meghan edmonds, and ae-hwa kim, a synthesis of spelling and reading interventions and their effects on the spelling outcomes of students with ld, journal of learning disabilities vol•ue 39, number 6, november/december 2006, pages 528-54

[6] Claire Gosse, Marie Van Reybroeck, Do children with dyslexia present a handwriting deficit? Impact of word orthographic and graphic complexity on handwriting and spelling performance, Research in Development Disabilities 97 (2020) 103553.

[7] Michael Hebert, Devin M. Kearns, Joanne Baker Hayes, Pamela Bazis, and Samantha Cooper, Why Children With Dyslexia Struggle With Writing and How to Help Them , Lang Speech Hear Serv Sch. 2018 Oct; 49(4): 843–863.

[8] Maja ivancevic otanjac, students with language learning disabilities and difficulties in a foreign language classroom, specijalna edukacija i rehabilitacija (beograd), vol. 15, br. 4. 461-474, 2016.

[9] Pedler, J. (2001). Computer spellcheckers and dyslexics—A performance survey. British Journal of Educational Technology, 32(1), 23-37.

[10] Farag Ahmed, Ernesto William De Luca, and Andreas Nürnberger, Revised N-Gram based Automatic Spelling Correction Tool to Improve Retrieval Effectiveness, Computer Science , no.40, 2009.

[11] Nejja Mohammed, Yousfi Abdellah, The vocabulary and the morphology in spell checker, Procedia Computer Science 127 (2018) 76–81, The First International Conference on Intelligent Computing in Data Sciences.

[12] Pravallika Etoori, Manoj Chinnakotla¸ Radhika Mamidi, Automatic Spelling Correction for Resource-Scarce Languages using Deep Learning, Proceedings of ACL 2018, Student Research Workshop, pages 146–152 Melbourne, Australia, July 15 - 20, 2018. c 2018 Association for Computational Linguistics.

[13] Jui-Feng Yeh, Sheng-Feng Li, Mei-Rong Wu, Wen-Yi Chen, Mao-Chuan Su, Chinese Word Spelling Correction Based on N-gram Ranked Inverted Index List, Proceedings of the Seventh SIGHAN Workshop on Chinese Language Processing (SIGHAN-7), pages 43–48, Nagoya, Japan, 14 October 2013.

[14] G. Andradea, F. Teixeiraa, C. R. Xaviera,b, R. S.Oliveiraa, L. C. Rochaa, A. G. Evsukoffb, HASCH: High Performance Automatic Spell Checker for Portuguese Texts from the Web, Procedia Computer Science 9 (2012) 403 – 411.

[15] Aadil Ahmad Lawaye, Bipul Syam Purkayastha, Design and Implementation of Spell Checker for Kashmiri, Volume: 5 | Issue: 7 | July 2016, ISSN No 2277 - 8179 | IF: 3.508 | IC Value: 69.48.

[16] Mridha,.Mashod Rana, Mohamed Tipu Sultan, An Approach for Detection and Correction of Missing Word in Bengail Sentence, IEE,Copmuter Science, International Conference on Electrical , Computer and Comunication (ECCE) 2019.

[17] Miguel Blazquez and Catherine Fan, The Efficacy of Spell Check Packages Specifically Designed for Second Language Learners of Spanish, Pertanika J. Soc. Sci. & Hum. 27 (2): 847 - 861 (2019).

[18] Graham, S. (1999). Handwriting and spelling instruction for students with learning disabilities: A review. Learning Disability Quarterly, 22(2), 78-98.

[19] MacArthur, C. A., Graham, S., Haynes, J. B., & DeLaPaz, S. (1996). Spelling checkers and students with learning disabilities: Performance comparisons and impact on spelling. The Journal of Special Education, 30(1), 35-57.
[20] Mishra, R., & Kaur, N. (2013). A survey of spelling error detection and correction techniques. International Journal of Computer Trends and Technology, 4(3), 372-374.
[21] Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. TESOL quarterly, 17(2), 165-188.
[22] Plumb, C., Butterfield, E. C., Hacker, D. J., & Dunlosky, J. (1994). Error correction in text. Reading and Writing, 6(4), 347-360.
[23] Heift, T., & Rimrott, A. (2008). Learner responses to corrective feedback for spelling errors in CALL. System, 36(2), 196-213.
[24] Saigh, K., & Schmitt, N. (2012). Difficulties with vocabulary word form: The case of Arabic ESL learners. System, 40(1), 24-36.
[25] Shafto, M. A. (2015). Proofreading in young and older adults: The effect of error category and comprehension difficulty. International journal of environmental research and public health, 12(11), 14445-14460.
[26] Helfrich, A., & Music, B. (2000). Design and evaluation of grammar checkers in multiple languages. In COLING 2000 Volume 2: The 18th International Conference on Computational Linguistics.
[27] Lawley, J. (2016). Spelling: computerised feedback for self-correction. Computer Assisted Language Learning, 29(5), 868-880.
[28] Lee, I. (1997). ESL learners' performance in error correction in writing: Some implications for teaching. System, 25(4), 465-477.
[29] Alkanhal, M. I., Al-Badrashiny, M. A., Alghamdi, M. M., & Al-Qabbany, A. O. (2012). Automatic stochastic arabic spelling correction with emphasis on space insertions and deletions. IEEE Transactions on Audio, Speech, and Language Processing, 20(7), 2111-2122.
[30] Shaalan, K., Aref, R., & Fahmy, A. (2010, March). An approach for analyzing and correcting spelling errors for non-native Arabic learners. In 2010 The 7th International Conference on Informatics and Systems (INFOS) (pp. 1-7). IEEE.
[31] Xie, F., & Jiang, X. M. (2007). Error analysis and the EFL classroom teaching. Online Submission, 4(9), 10-14.
[32] Reynaert, M. W. (2011). Character confusion versus focus word-based correction of spelling and OCR variants in corpora. International Journal on Document Analysis and Recognition (IJDAR), 14(2), 173-187.
[33] Hofmann, T. (2013). Probabilistic latent semantic analysis. arXiv preprint arXiv:1301.6705.
[34] Latif-Amet, A., Ertüzün, A., & Erçil, A. (2000). An efficient method for texture defect detection: sub-band domain co-occurrence matrices. Image and Vision computing, 18(6-7), 543-553.
[35] Billhardt, H., Borrajo, D., & Maojo, V. (2002). A context vector model for information retrieval. Journal of the American Society for Information Science and Technology, 53(3), 236-249.
[36] Damerau, F. (1964). “A technique for computer detection and correction of spelling errors.” Communications of the ACM 7(3): 659-664.
[37] Ali Selamat, Nicholas AKosu,Word- length algorithm for language identification of under-resourced languages , Journal of king saud university - computer and information sciences (2016)28, 457-469.
[38] Rakesh Kumar, Minu Bala, Kumar Sourabh: A study of spell checking techniques for Indian Languages, JK Research Journal in Mathematics and Computer Sciences, Vol. (1) No. (1) March 2018.
[39] Youssef Bassil, Parallel Spell-Checking Algorithm Based on Yahoo! N-Grams Dataset, International Journal of Research and Reviews in Computer Science (IJRRCS), ISSN: 2079-2557, Vol. 3, No. 1, February 2012.
[40] Majumder, P., Mitra, M., & Chaudhuri, B. B. (2002, November). N-gram: a language independent approach to IR and NLP. In International conference on universal knowledge and language.
[41] Daniel Jurafsky, James H Martin. Speech and Language Processing. 2nd Edition. Upper Saddle River: Prentice Hall. 2008.
[42] Adubasim, I. C. J., & Nganji, J. (2017). Dyslexia-A Learning Difference. Autism Open Access, 7, 203.
[43] Engy Yehia, Hussein Boshnak, Sayed AbdelGaber , Amany Abdo, Doaa S.Elzanfaly, Ontology -based clinical information extraction from physician 's free-text Notes, Journal of Biomedical Informatics 98 (2019) 10327.
[44] Rello, L. (2014). DysWebxia: a text accessibility model for people with dyslexia (Doctoral dissertation, Universitat Pompeu Fabra).