



## A Word & Character N-Gram based Arabic OCR Error Simulation model

<sup>1</sup>Mostafa Ezzat,<sup>2</sup>Tarek El Ghazaly,<sup>3</sup>Mervat Gheith

Computer Sciences Department, Institute of Statistical Studies & Research, Cairo University, EGYPT

<sup>1</sup>Mostafa.Ezzat@cu.edu.eg

<sup>2</sup>Tarek.Elghazaly@cu.edu.eg

<sup>3</sup>MGheith@issr.cu.edu.eg

### Abstract

This paper provides a new model aimed to enhance Arabic OCR degraded text retrieval effectiveness. The proposed model based on simulating the Arabic OCR recognition mistakes based on both, word based and Character N-Gram approaches. Then we expand the user search query using the expected OCR errors. The resulting search query expanded gives high precision and recall values in searching Arabic OCR-Degraded text rather than the original query. The proposed model showed a significant increase in the degraded text retrieval effectiveness over the previous models. The retrieval effectiveness of the new model is %93, while the best effectiveness published for word based approach was %84 and the best effectiveness for character based approach was %56.

### Indexing terms/Keywords

Arabic OCR Degraded Text Retrieval; Arabic OCR-Degrade Text; Orthographic Query Expansion; Synthesize OCR-Degraded Text.



## Council for Innovative Research

Peer Review Research Publishing System

**Journal:** INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol 12, No. 8

[editor@cirworld.com](mailto:editor@cirworld.com)

[www.cirworld.com](http://www.cirworld.com), [www.ijctonline.com](http://www.ijctonline.com)



## INTRODUCTION

Printed documents has never lost its importance, although the number of documents available as character-coded text is now increasing as a result of electronic publishing, but it's not the same case for Arabic. Actually, it's still uncommon to find character-coded Arabic documents instead of Arabic document images on the Web. This makes searching documents by user query very hard.

Without automated process, the only way the user can find the desired document would be only manually or consulting a person who is familiar with the searched documents. Since searching character coded documents is the easiest way, and automating the process means generating the character-coded representations of the documents. We can generate the character-coded representation of the documents by rekeying the documents' text or creating metadata about the documents such as titles, summaries, or keywords. But these approaches would be labor intensive and impractical for large numbers of documents. But we can produce the character-coded representation of the document by scanning the documents and then use Optical Character Recognition (OCR), which is an automated process that converts document images into character-coded text. However, the results often contain errors.

Further, Arabic has special properties like orthography, which is how words are written and complex morphology, which is how words are constructed, adversely affect the accuracy of OCR. On the other hand, the OCR process is inexpensive and thus well suited for large document collections.

### 1.1 Orthographic Properties of Arabic

Arabic is a right-to-left connected language that uses 28 letters, and its shapes change depending on their positions in words. Fifteen letters contain dots to differentiate them from other letters. And depending on the discretion of the document producer, these letters may or may not have diacritics (short vowels). Ligatures, which are special forms for some character sequences, and kashidas, which are symbols that extend the length of words, are often employed in printed text. Having a connected script that uses ligatures and kashidas complicates isolation of individual letters in OCR, and dots and diacritics make OCR sensitive to document noise and speckle. Most Arabic OCR systems segment characters [4, 5, 6, 7], while a few opted to recognize words without segmenting characters [8, 9], and another system developed by BBN avoids character segmentation by dividing lines into slender vertical frames (and frames into cells) and uses an HMM recognizer to recognize character sequences [10].

This paper presents OCR-Degradation synthesizing model that is trained on different shapes of the words generated from the OCR recognition process and the model is independent of the technique used in the OCR process itself. Then using the training information, we reformat the user search query by inserting all the possible shapes of each query word to generate the search query that will return all the documents related to the original search query.

## PREVIOUS WORK

Building an OCR-Degraded text synthesization model is the most important part of information retrieval when the retrieved text is generated from the OCR process. In this case there are many factors that affect the OCR process, like the Paper quality which is affected by the time, human errors done during the scanning process, the accuracy of the OCR system used and the font which the document was initially printed with and many others.

All these factors and others affects the information retrieval process done on the degraded text returned from the OCR process. so modeling the OCR-Degraded text is a major part in any research that deals with information retrieval. In this section we will present two models were developed by ElGhazaly and Darwish as parts of their research in the information retrieval in the same language or in different languages. To the best of our knowledge, they are the only models built specially for Arabic OCR-Degraded Text retrieval.

### 2.1 ElGhazaly OCR-Degraded Synthesizing Model

The model is a Word Based model that is trained and tested on complete words. ElGhazaly generated documents of single word per line, and then print, scan and OCR these documents using Sakhr OCR, then he manually align the OCR recognition results with the clean text document. During this manual alignment he checked if there is a deformation of the OCR-Degraded word, then he stored the deformed OCR-Degraded Word and its original word as a training pair in the OCR Errors Database [2, 22].

Then ElGhazaly used this OCR Error model to expand the query used in the information retrieval. This Orthographic Query Expansion approach attempts to find different misrecognized versions of a query word in the collection being searched. But ElGhazaly didn't proof the retrieval effectiveness of his model.

### 2.2 Darwish OCR-Degraded Synthesizing Model

Darwish model is based on the character level not the word level as ElGhazaly. The main idea of the model was based on the fact that the context affects the character shape in a word, because Arabic letters are connected and change shape depending on their position in the word and some special shapes are formed when special characters are sequenced in the same word, so his assumption was that the position of a letter being recognized and the letters surrounding it ("context") would be important in developing a good model.

The model was simply based on aligning the OCR recognition results from the print (300x300), 200x200, and 200x100 versions of the Zad collection (the training and testing data set as we will mention later in this section) with the associated clean text version of the same collection. The alignment was done using SCLITE which is an application from the National



Institute of Standards and Technology (NIST), which employs a dynamic programming string alignment algorithm that attempts to minimize the edit distance between two strings. Basically, the algorithm uses identical matches to anchor alignment, and then uses word position with respect to those anchors to estimate an optimal alignment on the remainder of the words. Two factors affected his alignment process. The printed and clean text versions in the Zad collection were obtained from different sources that exhibited minor differences (mostly substitution or deletion of particles such as in, from, or, and then), and secondly some areas in the scanned images of the printed page exhibited image distortions that resulted in relatively long runs of OCR errors.

Based on the alignment done using the SCLITE and the other algorithm Darwish implemented, he built a "garbler" tool which reads in the clean word  $C_1..C_i..C_n$  and synthesizes OCR degradation to produce a garbled word  $D'_1..D'_j..D'_m$ . This "garbler" chooses and performs a random edit operation (insertion, substitution or deletion) based on the probability distribution for the possible edit operations. The garbler was the tool that simply takes any word and generates the different error shaped that would be produced if this word was printed and OCR-ed.

### 2.3 The Accuracy of the Models

It was important for us to see the accuracy of both models, because this will show us the direction of our research. Because each model has a different direction, one of them is based on word based approach and the other one is based on character based approach.

ElGhazaly model produced accuracy 84.74%, He trained model on dataset of 53,787 words, and then he tested his model on a test set contained 51,658 words [21]. ElGhazaly introduced his own accuracy measure, which is the number of accurate replacements (with respect to the training set size) divided by the total number of OCR-Degraded words. In other words, if the mistaken OCR-degraded word is available in the training set with the correct original word, then this will be considered as accurate replacement. Otherwise, it will be considered as not accurate one. But ElGhazaly didn't actually index and search his test data set calculate the precision, recall and mean average precision of his model [21].

On the other hand Darwish model produced accuracy for 3- gram or 4 gram character indexing was 87% [1, 24], but here we must illustrate the difference in both accuracy measurements, because Darwish was measuring according to the character level which means, For example, consider a page of 20 lines, each line has on average 10 words, and each word has on average 5 characters. This means the page has 1000 characters in 200 words. If we consider the OCR output of this page to have only 20 character errors each in a separate word, this means character accuracy of 98% where it means word accuracy of 90%. Darwish Also illustrated that the best mean average precision of his model which was "0.56" [1, 25]

## THE PROPOSED OCR-DEGRADED ARABIC TEXT RETRIEVAL MODEL

The OCR-Degraded Arabic Text Retrieval model doesn't depend on the OCR software used and is not aimed to enhance the OCR recognition accuracy, but the model improves the retrieval effectiveness of the degraded text returned from the OCR operation. The new model combined from two steps, the first step is synthesizing the OCR-Degraded text, and the second step is then expanding the user search query using the expected OCR error generated from the Arabic OCR error simulation. The main goal of the model is to enhance the retrieval effectiveness of large degraded text produced from automatic operation, OCR, such as electronic library information retrieval. Unfortunately, despite the claims of OCR commercial vendors; OCR recognition accuracy rates are far from perfect, especially for languages like Arabic [14]. This directly affects the accuracy of text retrieval process regarding these documents.

As we mentioned before, in this model we are not concerned with enhancing the OCR recognition accuracy. Also, we are not concerned with correcting the OCR recognition outputs. Our aim of the proposed model is to synthesize the text degradation produced by the OCR application.

Many models simulate the OCR degradation errors have been developed. But these models are either word based models or character based models, and each one of them has its limitations. The word based models, as an example, are dependent on the size of the training data set, in the other hand; the character based models are in-dependent of the training data set. But the character based models covers only the words with the number of characters it was trained on it. If the model is based on five-character gram, so it will be trained on the words with length five characters. However, in Arabic, as character shape defers up to its position in the words (begin, middle, end, Isolated). So, it is too difficult to include all these variables (7-gram character for example) plus the character position in one model. However, even if the 7-gram is reached, the 8-grams will not be covered and so on. To cover these limitations our proposed model is based on both character n-gram and word based approaches to Synthesize the OCR degradation error. The next section (3.1) describes the OCR-Degradation synthesizing model, which is a character based and word based model trained and tested on complete words. This model supports the maximum n-grams in the training set with respect to the words' positions. Then in section (2.2) we will show how we used this model to enhance Arabic OCR Degraded Text retrieval.

### 3.1 The OCR-Degradation Synthesizing Model

As we mentioned, the proposed OCR-Degradation synthesizing model is both a character based and word based model. This model supports the maximum n-grams in the training set with respect to the characters' positions. The following figure illustrates the main idea of the proposed model.

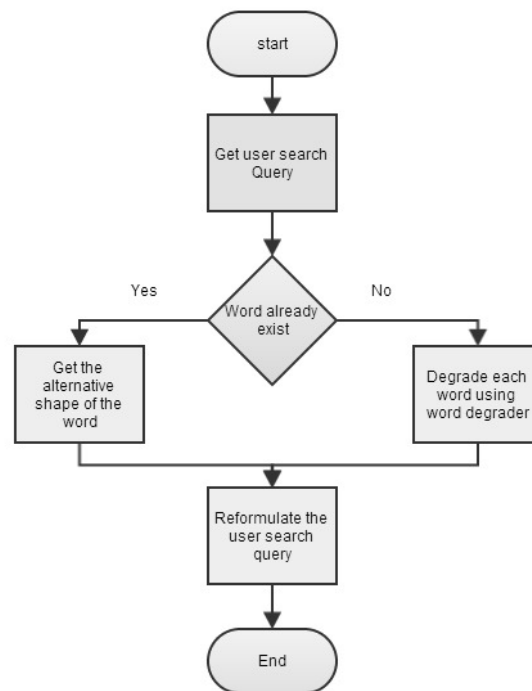


Fig.1.OCR-Degradation Synthesizing Model

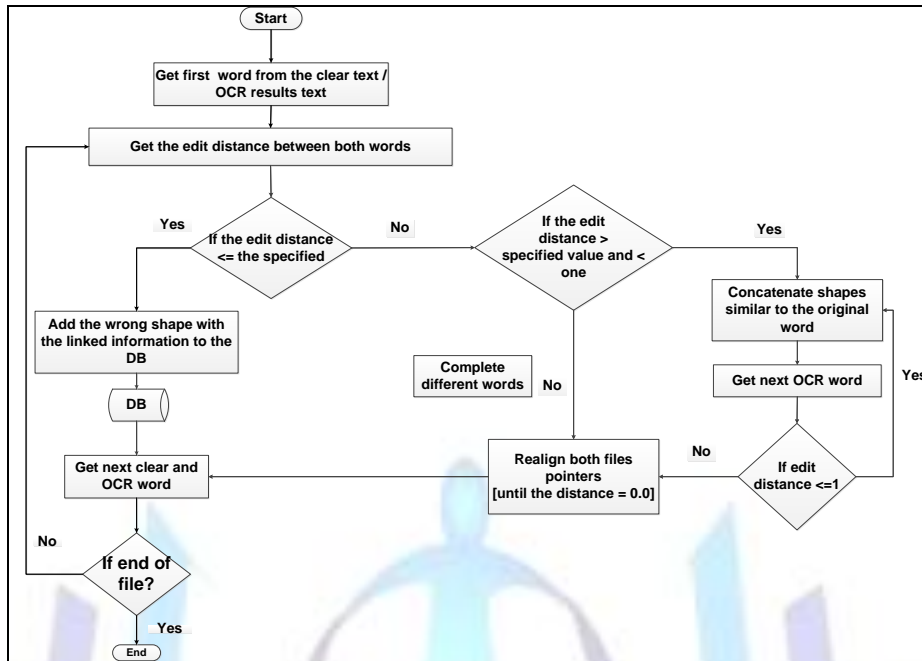
### 3.2 The Word based OCR Degradation Synthesizing Model

Getting the alternative degraded shapes of the word is done by the word based model, which is based on aligning the OCR degraded words and the clean text words, the alignment operation is done by calculating the edit distance between both words. usually the edit distance is calculated by calculating the number of operations that must be done on the word to convert it to the other, these operations are insertions, deletions, and substitutions operations. Calculating the edit distance between two Arabic words requires caution, because Arabic words may be formatted with prefix letters which sometimes forms half the word length, which means, if we calculated the edit distance between two words contains the same prefixes but the remaining of the two words is completely different, our model will consider them as the same word but some degradation happened while the OCR. To solve this problem our model ignores the Arabic prefixes letters while computing the edit distance between any words.

The word based model is built to be trained on any recognition accuracy returned from the OCR system; which means specifying the deformation level the user of the model wants to cover in his training data set; and this is controlled using the edit distance. The model checks each word in each document of the OCR-Degraded documents' set against the corresponding original word, if the edit distance between the two words less than or equal a selected value, the model stores the deformed shape linked with the original word as a training pair in the OCR errors database. Otherwise the model ignores these words and fetches the next words from the OCR-Degraded and cleans text files. Figure 2 illustrates the word based OCR-degradation synthesizing model.

And to complete our proposed model we built a software application we called The "Aligner", which performs all the actions illustrated in the model and constructs the training database. The "Aligner" takes both Original and OCR-ed text files and tokenize them into words, then based on the edit distance between them, the tool updates our training database that stores the original word shape and the corresponding shapes of the words appeared in the OCR version.

The word based model starts by fetching the first word from the clean the OCR-Degrade text files, and then the model calculates the edit distance between both words. Here we have three cases we shall handle, the first case, the edit distance value is less than or equal the accepted value, the second case, the value is larger than the accepted value but less than or equal one, and the third and final case, the value is larger than one. The first case means that the edit distance is within the accepted value specified by the model user in the application interface; so the model adds both words to the training database.



**Fig.2.TheWord based OCR Degradation Synthesizing Model**

The second case means that either the degraded word is part of the original word, but the OCR application split it while the recognition process, which happens sometimes for long words, or the recognition process was bad for this word and the system recognized only few characters of the word. This point formed a limitation in the word based models we checked, they didn't consider if the word was recognized as two and sometimes three parts [2,22].

We decided to cover this limitation in our model. This means, if the OCR system recognized one word into two or three words, our system will consider these parts as one word and store them in the training database as one shape corresponding to the original shape. Finally, the third case, if the edit distance between both words is larger than one, which means that both words are completely different, this means that the model has lost the correct position of anchors corresponding to each file, in this case our model realigns the anchors until both files' anchors are pointing to the same word. The "Aligner" updates the training database with degraded shapes with their corresponding original shapes; the database structure was designed to keep only one entry of the original word and any number, but unique, of the deformed shapes of this word. The database also stores the context of the deformed shapes, like the preceding word and the succeeding word. This information will be used in our future work which is checking how the OCR-Degraded word context affects the retrieval effectiveness.

### 3.3 Character based Model

Because getting the alternative degraded shapes of the word is done using the word based model is limited by the training set size. We decided to build another sub-model that handles the other cases where the user search query words don't exist in the training dataset. The model uses the same training dataset we used in the word based degradation model.

Using the modified alignment tool, the model compares every degraded word produced from the OCR operation with the corresponding clean word. And if the edit distance between both words is less than the specified value in the alignment tool, the model checks the word characters character by character. The model then stores the deformed character in the database. Figure 3 illustrates the character based synthesizing model. The tool also stores the preceding and succeeding character of the deformed character.

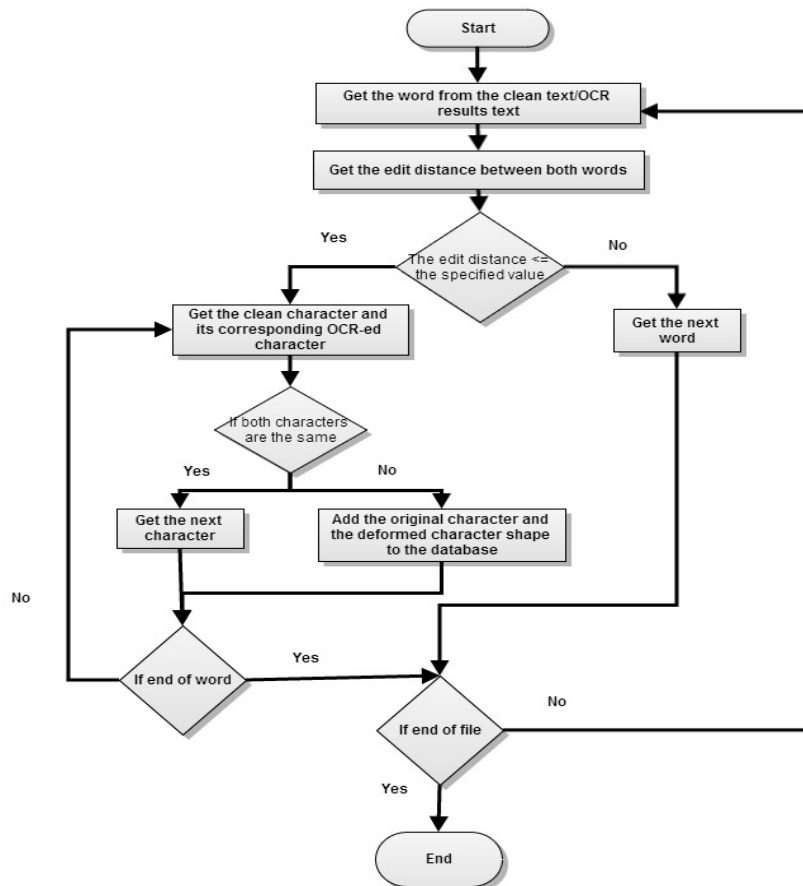


Fig.3. Character based synthesizing model

### 3.4 Orthographic Query Expansion

The second step in the Orthographic OCR degradation model after synthesizing the OCR-Degraded text is the Orthographic query expansion using the word based OCR error simulation model and the Character n-gram model is to expand the user search. And to expand the search query we built a tool named “Query Builder” which takes the original user search query and generates the deformed OCR- Degraded query based on the training database built using the word based model and the character based model. The query builder gets each word in the user search query and checks if there are related degraded shapes in the training database generated from the word based model, and then adds these shapes to the search query, otherwise it uses the “Query Degrader” which generates the word degraded shapes based on the character based model.

### 3.5 Training the OCR Degradation Synthesizing Model

Both word based model and character based models were trained on the electronic version of the Arabic three volume book “Abgad El Eloum” or “Alphabet of science” and for short naming we will call it “ABGAD”. For the experiment we printed the first two volumes of the book, which contained 205,761 words, then we scanned and OCR-ed the first two volumes using the best Arabic OCR application, Sakhr Automatic reader version 10. This application produces 99.8% accuracy for high-quality documents 96% accuracy for low-quality documents[14].

We, in this phase, are working on clean text electronic version ABGAD, and the corresponding OCR produced degraded text. And to make the alignment between both texts we built an Arabic word alignment tool. We named it the “Aligner” this tool takes the clean text and the OCR generated text and generates the training database. The generated database constructed from 1750 documents (305,586 words), consists of 397, 50 unique words, 193,756 words OCR-ed correctly and 230, 49 word OCR-ed wrong.

### 3.6 Testing the Orthographic Query Expansion Model

To test and verify that OCR degradation search model we built increases the accuracy of the degraded text information retrieval, we have to make a group of tests. Evaluating retrieval effectiveness requires the availability of a test document



collection with an associated set of topics and relevance judgments. Relevance judgments are the mappings between topics and documents in the collection that are relevant to them. The cost of producing relevance judgments for a large collection is very high and dominates the cost of developing test collections [15].

There are three ways to produce relevance judgment, the first and the oldest one is “Manual search”, where a relevance judge manually searches and assesses documents for a topic until the judge is convinced that all relevant documents are found [17]. The second way is “Pooling”, which is assessing manually the relevance of the union of the top n documents from multiple retrieval systems for every topic [16]. The third way is exhaustively searching the documents for relevant documents [18].

All these ways miss some of the relevant documents, and our main objective is the assessment of relevance. But studies shown that relevance judgments can be reliably used to correctly differentiate between retrieval systems provided that a sufficient number of queries are used [4, 19, 16]. Voorhees estimated the number of sufficient queries to be about 25 [16]. But here in our test we extended the search query to be 35 queries to collect more data about the retrieval accuracy.

**Orthographic Query Expansion Model Test Set Statistics.** We based our test on two test sets have been selected from the Training and test pool. The first one includes 20 documents (66,985 words), which is the third volume of “ABGAD”. And to memorize the training set was the first two volumes only of the book. And the second data test set contains 2,730 long documents containing 621,763 words. There is no intersection between the Training and Test sets.

The second test data is “ZAD” data collection, which is a 14th century religious book called Zad Al-Me’ad, which is free of copyright restrictions and for which an accurately character-coded electronic version exists [20].

### 3.7 Testing the Orthographic Query Expansion Model Accuracy

This section describes the development of a test collection that can be used to evaluate alternative techniques for searching scanned Arabic text and a set of experiments that were designed to identify the effect of the proposed model on retrieval effectiveness. IR evaluation measures are concerned with precision and recall given:

		Relevant	Not Relevant
Retrieved		A	B
Not Retrieved	Re-	C	D

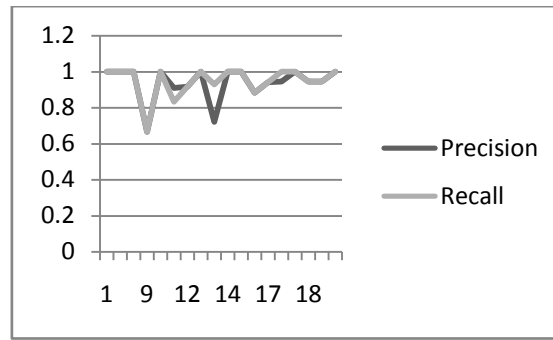
$$\text{Precision} = \frac{A}{(A+B)} \quad \text{Recall} = \frac{A}{(A+C)}$$

Precision measures the fraction of retrieved documents that are relevant. Recall measures the fraction of all relevant documents that are actually retrieved. IR-effectiveness measures use precision and recall in different ways. For example, precision at n measures the precision after a fixed number of documents have been retrieved. Another is precision at specific recall levels, which is the precision after a fraction of relevant documents are retrieved.

For our experiment we indexed our clean and OCR-Degraded documents on the best Arabic search engine, IDRISI 6.0[14]. One of IDRISI features is creating separate search collection for different group of documents depending on the user requirements. So we created a separate collection for the clean text documents and another collection for the OCR-Degraded documents. The author of the paper, a native speaker of Arabic, developed 35 topics and exhaustively searched the collection for relevant documents.

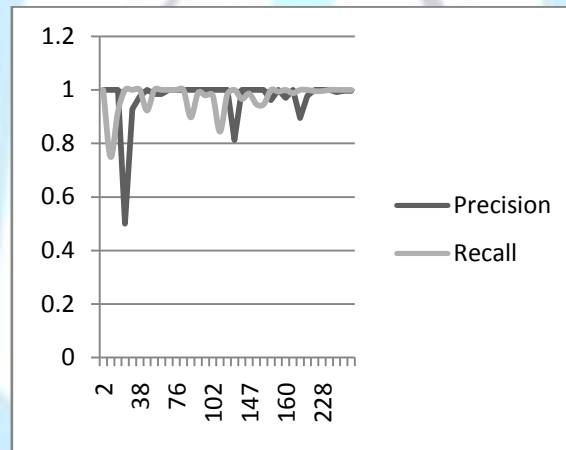
Then using the “Query Degraded” tool that takes the user clean text query and based on the training database generates the relevant OCR-Degraded text query, we passed both the original search query and the OCR degraded query to IDRISI to search the clean text collection and the OCR-Degraded text collection separately. Using this way we can compare the results of both collections and get the precision and recall to check our model accuracy.

After completing the experiment on the thirty five queries we analyzed the results and found that. For test data set 1, the number of relevant documents per topic ranged from one (for one topic) to eighteen, averaging 14. For test data set 2 the number of relevant documents per topic ranged from two (for one topic) to 224, averaging 121. The average query length used for the test data set 1 is 4.1 words and for test data set 2 is 5.2 words. The following figure displays the search precision and recall returned from the search engine relevant to the number of retrieved document for test data set 1.



**Fig.4. Test Data Set 1 Precision and Recall relevant to no. of documents retrieved**

And the following figure displays the search precision and recall returned from the search engine relevant to the number of retrieved document for test data set 2. The figure shows that the recall and precision values returned from searching OCR-Degraded text relative to the number of returned document in the test data set 2. Unfortunately Darwish didn't mention the precision and recall measurements for his model. He was most concerned with the mean average precision value of his information revival system, the mean average precision is the most commonly used evaluation metric which eases comparison between systems [16]. The best mean average precision resulted by Darwish was "0.56" [1] for 3-Gram and 4-gram steamed words, in the other hand the mean average precision resulted by the proposed model for test data set 1 is "0.96" and for test data set 2 is "0.93".which means a significant improvement in the information retrieval effectiveness.



**Fig.5. Test Data 2 Precision and Recall relevant to no. of documents retrieved**

## CONCLUSION

In the second section we provided a comprehensive study of the previous models designed to simulate OCR-Degraded text, to the best of our knowledge Darwish and Elghazaly models are the only models designed especially for Arabic recognition, then we provided our model which we consider advanced the state of the art for Arabic OCR-Degraded text retrieval in specific and IR in general. Perhaps the most challenging contribution was the development of new OCR degraded synthesizing model that improves information retrieval effectiveness. The proposed new model has shown high effectiveness than the previously developed n-gram or either word level models.

The model was trained on 1500 documents containing 205,696 words, consists of 38,840 unique words, 182,647 words read correctly and 230,49 word read wrong. And the model was tested on two different data sets, on the first data set, which consisted only of 20 documents containing 66,985 words.

And the second test data set were 2,730 separate documents containing 621,763 words. There were no intersection between the training data sets and test data sets. When the model was applied on the first data set, The mean average precision produced was "0.95", and when the model applied on the second data set the mean average precision produced was "0.93". In the other hand Darwish shown that the best mean average precision produced by his model for 3-gram indexed data set was "0.56" [1]. We here must illustrate that ElGhazaly didn't measure his model performance against indexed data set [2] and he depended only on calculating the number of OCR-Degraded words in the training set as an accuracy measurement.

We must illustrate also that we built all the required applications to proof the model accuracy and effectiveness, starting from the word aligner to the query expansion. On the other hand we showed that other researchers used external applications like SCLITE an application from the National Institute of Standards and Technology (NIST) or manual alignment to align the degraded text with the clean text.





The main limitation of the proposed OCR-Degraded synthesizing model is its dependency on the training set size, as it's a word based model. On the other hand the character based model proposed by Darwish doesn't have this limitation. Also we must illustrate that our model handles the word splitting problem while recognition and this limitation exists in both models we compared with. Table 1 illustrates the main comparison points between the three models mentioned in this paper.

**Table 1. COMPARISON BETWEEN THE THREE MODELS**

Comparison Point	Word based Model	Character Based Model	Pro-posed Model
The model considered word context error that may affect the recognition operation.	✓	✗	✓
The model is word based and covers any word with any length once it was trained on it	✓	✗	✓
The accuracy measurement used based on the word level not on character level	✓	✗	✓
The model solved the word concatenation problem	✗	✓	✓
The model tried to simulate different degradation levels	✗	✓	✓
The model idea can potentially be useful in widespread applications	✓	✓	✓
The model retrieval effectiveness is tested on indexed data set	✗	✓	✓
The model considered enhancing the scanned images before OCR	✗	✗	✓
The model accuracy is not affected by the training set size	✗	✓	✓

## References

- [1] Darwish, K. "Probabilistic Methods for Searching OCR-Degraded Arabic Text," A PhD Dissertation, University of Maryland, College Park, 2003.
- [2] Elghazaly, T. "Cross Language Information Retrieval (CLIR) for digital libraries with Arabic OCR-Degraded Text," A PhD Dissertation, Cairo University, Faculty of Computers and Information, 2009.
- [3] Chen, A. and F. Gey. Building an Arabic Stemmer for Information Retrieval. In TREC. 2002. Gaithersburg, MD.
- [4] Burgin, B., Variations in Relevance Judgments and the Evaluation of Retrieval Performance. Information Processing and Management, 1992. 28(5): p. 619-627.
- [5] Callan, P., Z. Lu, and B. Croft. Searching distributed collections with inference networks. in SIGIR. 1995.
- [6] Blando, L., Evaluation of Page Quality Using Simple Features. 1995, University of Nevada, Las Vegas: Las Vegas, NV.
- [7] Chen, S., S. Subramaniam, R., R. Haralick, and I. Phillips. Performance Evaluation of Two OCR Systems. in Annual Symp. on Document Analysis and Information Retrieval. 1994.
- [8] Darwish, K. and D. Oard. CLIR Experiments at Maryland for TREC 2002: Evidence Combination for Arabic-English Retrieval. in TREC. 2002. Gaithersburg, MD.
- [9] Cole, A., D. Graff, and K. Walker, Arabic Newswire Part 1 Corpus (1-58563-190-6), Linguistic Data Consortium (LDC).
- [10] Darwish, K. Building a Shallow Morphological Analyzer in One Day. in ACL Workshop on Computational Approaches to Semitic Languages. 2002.
- [11] Rice, S., F. Jenkins, and T. Nartker, The fifth annual test of OCR accuracy. 1996, Information Science Research Institute, University of Nevada, Las Vegas.



- [12] D. K. Harman. Overview of the first Text REtrieval Conference (TREC-1). In Proceedings of the First Text REtrieval Conference (TREC-1), pages 1–20. NIST Special Publication 500-207, March 1993
- [13] Tapas Kanungo, Gregory A. Marton, Osama Bulbul. "OmniPage vs. Sakhr: Paired model evaluation of two Arabic OCR products," in Proc. of SPIE Conf. on Document Recognition and Retrieval, 1999
- [14] WWW.SAKHR.COM, last visited on June 2013.
- [15] Soboroff, I., C. Nicholas, and P. Cahan. Ranking retrieval systems without relevance judgments. in SIGIR. 2001.
- [16] Voorhees, E. Variations in Relevance Judgments and the Measurement of Retrieval Effectiveness. in SIGIR. 1998. Melbourne, Australia.
- [17] Wayne, C. Detection & Tracking: A Case Study in Corpus Creation & Evaluation Methodologies. in Language Resources and Evaluation Conference. 1998. Granada, Spain.
- [18] Tseng, Y. and D. Oard. Document Image Retrieval Techniques for Chinese. In Symposium on Document Image Understanding Technology. 2001. Columbia, MD.
- [19] Salton, G. and M. Lesk, Relevance Assessments and Retrieval System Evaluation. Information Storage and Retrieval, 1969. 4: p. 343-359.
- [20] Publishers, Al-Areeb Electronic.
- [21] Elghazaly, T. , Fahmy A., "Query Translation and Expansion for Searching Normal and OCR-Degraded Arabic Text," Lecture Notes in Computer Science 5449- Computational Linguistics and Intelligent Text Processing , Springer-Verlag Berlin Heidelberg, Germany, pp. 481-497, ISSN 0302-9743, 2009.
- [22] Tarek A. Elghazaly and Aly A. Fahmy, "English/Arabic Cross Language Information Retrieval (CLIR) for Arabic OCR-Degraded Text", Communications of the IBIMA, Volume 9, number 25, pp. 208-218, ISSN 19437765, 2009.
- [23] Darwish, K. and D. Oard. CLIR Experiments at Maryland for TREC 2002: Evidence Combination for Arabic-English Retrieval. in TREC. 2002. Gaithersburg, MD.
- [24] Darwish, K. and D. Oard. Term Selection for Searching Printed Arabic. in SIGIR.2002.
- [25] Darwish, K. and D. Oard. Probabilistic Structured Query Methods. To appear in SIGIR. 2003.