P- ISSN 1991-8941 E-ISSN 2706-6703 *Journal of University of Anbar for Pure Science (JUAPS)* Open Access
2018,12 (3) :71-83

Textual Analysis Applications: Subject Review

Abeer K. Al-Mashhsdany [(1)] **Safa A. Ahmed** [(2)]

[(1)]Department of Computer Science, College of Science, Al-Nahrain University, Baghdad-Iraq.
[(2)]Chemical Engineering Department, University of Technology

ARTICLE INFO

Received: 26 / 4 /2018
Accepted: 26 / 12 /2018
Available online: 20/12/2018

DOI: 10.37652/juaps.2022.171861

Keywords:
textual analysis,
morphological,
syntactic analysis,
semantic approach.

ABSTRACT

This paper is a literature survey about applications of textual analysis. It aims to provide brief description about the common textual analysis applications. The paper talks about the dictionary which is mostly, one of the main components for textual analysis applications. The paper highlights a number of related examples that were proved in previous published papers. Common features for the related examples are illustrated. And their results are discussed. It will be shown that "morphological and syntactic analysis" is a proved approach. Also, it will be shown that text similarity based on "morphological and syntactic analysis" approach has more accurate results than text similarity based on semantic approach..

1. Introduction

Firstly, the researchers like to explain the word "text". When considering text as a semantic level, it is just sequences of sentences that constitute a certain unity of meaning [1].

When a textual analysis is performed on a text, this means that an educated guess is made in some interpretations which are likely made of that text [2]. Textual analysis provides methods for describing and interpreting the characteristics of a message (may be recorded or visual) [3].

* Corresponding author at: Chemical Engineering Department, University of Technology ,Baghdad,Iraq.E-mail address: 80155@uotechnology.edu.iq

In textual analysis process; it is important to improve the acquisition and selection of texts to be studied, and determine the appropriate approach for analyzing them. There are many approaches; morphological and syntactic analysis, interaction analysis, content and semantic analysis, machine learning techniques, the discourse analysis, rhetorical criticism, narrative analysis, and the performance studies. [4, 5, 6, 7]

Text analysis tools may generate new texts through implementing some computer processes. The original texts are decomposed and then recomposed to be ready for interpretation. The new text can be analyzed too [8]. Performing text analysis needs understanding

the lexicon of the language, so computer –based dictionary is very important in text analysis systems [9].

Common applications in text analysis, such as; document clustering, document filtering, and information retrieval. As a preprocessing for those applications, some operations are needed, such as; extracting features, representing document, and signature creation. At most text analysis applications, the preprocessing operations are the same. [10, 11]

This paper is a literature survey about textual analysis applications. The next section presents a brief description about textual analysis applications. Then, three sections describe about some examples – papers considered as textual analysis applications-. Then conclusion will be discussed.

2. Textual Analysis Applications

This section aims at providing good and brief description about textual analysis applications. Before listing the applications, the researchers shed light on the term dedicated-dictionary, which refers to a database of words that belongs to a specific domain. Mostly, the dedicated-dictionary unit is considered to be an important component of any text analysis system. If he wants to know the reason, he has to remember that all text analysis applications are belong to artificial intelligence branch. Artificial intelligence is the branch of computer science which deals with tasks that require intelligent when performed by humans. All linguistic applications belong to artificial intelligence branch, and textual analysis systems are part of the linguistic applications [12, 4, 13]. Humans have the ability to perform intelligent tasks because they have the brains. Humans are provided with high quality database stored inside their brain. So to computerize any task, this requires intelligent; database component must be the first part to be thought about. Linguistic applications databases store a lot of words. Such databases are called dictionaries. Because of the limitations found in computerized systems, the dictionary usually is dedicated to a specific domain, so it is called "dedicated-dictionary". Sometimes the term "lexicon" is used instead of dictionary. Lexicon is not just list of words; instead, in a lexicon words are classified into classes, and they may be provided with features and rules. [12, 4, 13, 9]

The first concerned application will be "spellchecker". Discovering typos at any input text – written in any language- is the task of spellchecker. Not just discovering error, but also providing list of suggestions to correct wrong words. [14, 4]

"Phrase similarity" is the next application. Its job is to discover matching meaning between phrases; the phrase may be a sentence or part of sentence. Performing phrase similarity with high accuracy is important to achieve other jobs such as "rule consistency". Some intelligent applications acquire knowledge from external environments, acquired knowledge may be inconsistent, and so it must be validated. [4]

"Keyword extraction" is the next application. Its job is to analyze an input document, and then find its keywords. Any document's content can be described by just a little set of terms, which are called keywords. Keywords are useful for a variety of purposes including; summarizing, translating, indexing, labeling,

information retrieval, categorization, clustering, and searching. [15]

"Text summarization" is the next application. Summary is a text that abstracts one or more texts. It expresses important information in original text(s). Size of summary is usually less than half of the original text(s), however it is not accepted to be longer than the half. A summarization method depends on selecting important sentences from original text(s), and then it is called extractive summarization method. In such methods, keyword extraction gives good advice to recognize important sentences. [16]

"Text similarity" is the next application. It is the process of calculating similarity between texts. In such process, text summary gives good advice, it facilitates calculating similarity, especially in case of long texts. That is because text summary achieves size reduction for long texts. [17]

"Text analysis in translation" is the next application. A model is generated for translation-oriented source-text analysis applicable to all text types and genres independent of the language. [18]

"Information retrieval" is the next application. Its job is to retrieve appropriate documents that are chosen according to a query. Also "text mining systems" are used for the same purpose. [19, 10]

"Information extraction" is the next application. Now the job is to analyze text then extract pieces of information from text. The wanted pieces are pre-defined like a company name, or sentences describing some activities -teaching activity - as an example. [10]

"Question answering" is the next application. Its job is providing answers to questions. Answering application is based on retrieving. In order to provide an answer, document(s) which are relevant to a question must be retrieved. Sometimes a single document is not enough to construct complete answer. Seeking parts of answer from number of documents requires providing good combination process. [11]

"Document clustering" is the next application. The job of clustering is organizing documents into groups. This application is based on semantic similarity. [10]

"Classifying" is the next application. Many terms are commonly used for this application; document categorization, document routing, document filtering. Anyway, this application classifies documents into predefined groups. This application gives good advice in another textual analysis application such as word sense disambiguation, information retrieval, and any application requiring document organization. [5]

"Entity linkage" is the next application. Here the job is to find relationships between entities. An example of relationships is "President". So an example of entity linkage is recognizing that "Jon Mano" is President of "Mercedes Company". [10]

"Event detection" is the next application. This application is interested in important events that happen in real time. Tasks of current application are; detecting those events, deliver summary, providing accepted evidence, and supporting to whose interested with. [11]

3. Textual analysis examples

3 .1 Morphological and Syntactic Analysis Examples

This section is interested with morphological and syntactic analysis approach. Here the researchers like to highlight six examples in textual analysis; each one of them is based on morphological and syntax analysis approach. All six examples have several common features, see the list below.

- All examples require dedicated dictionaries to perform their jobs.
- All examples store "stop-noisy-words" in their dictionaries. "Stop-noisy-words" for different examples may have some differences.
- All examples receive text as input to its application (program). Language of text is either Arabic or English.
- All examples perform a linguistic preprocessing as the first step when receiving a text. Such preprocessing includes; front token, filtering (remove stop-noisy-words), replacing synonyms, and detecting negation.
- All examples perform rooting-stemming process. An example may be root-based, stem-based, or root-stem based. This process requires programming the linguistic rules of the specific language.

The following is a summary about each example. For each one; tasks, methods, components and results are illustrated.

3.1.1 [20] Arabic Diagnosing Expert System Shell (ADESS)

This work includes two phases, textual analysis system is found at first phase - automatic acquisition phase-. This textual phase acquires knowledge from external environment (human expert) in an automatic way. The acquired knowledge is a text (phrase) written in Arabic language. ADESS constructs an Arabic Morphological system [AMS], which is root-based. ADESS is provided with medical dictionary. The acquired knowledge is analyzed by AMS in order to achieve roots. Here textual analysis purposes are; spellchecker, phrases similarity, avoiding rule duplication, and avoiding inconsistent rules.

ADESS automatic acquisition phase is decomposed into three components; automatic knowledge acquisition & knowledge engineer module, Arabic morphological module, and refine knowledge module. The Arabic morphological module is decomposed into two components; "dictionary component", and "rules - analysis component". The validated acquired knowledge is stored at the knowledge base, which is one component of ADESS Architecture.

ADESS succeeds in achieving its purposes by merging AMS with the automatic knowledge acquisition. ADESS succeeds in discovering and correct typos. ADESS succeeds in discovering phrases that are matched in meaning, and then preventing duplication and inconsistent rules. But there is a weakness point. ADESS tried to cover all Arabic morphological rules. But because of their wideness, and also because of irregular cases, the rules cannot be covered completely.

3.1.2 [21] Knowledge acquisition & hybrid Inference with a Stem-Based approach (KISB)

KISB is composed from three main components; Arabic Analyzer System [AAS], Knowledge Acquisition Module [KAM], and Hybrid Inference Module [HIM].

KISB's AAS is stem-based. KISB is provided with medical dictionary. KISB's analyzer is merged with the two modules. The two modules receive text (phrases) written in Arabic language from many users.

KISB ASS is composed from dictionary component and analysis component. KISB dictionary has an advantage upon ADESS dictionary. Because, KISB dictionary stored the words, they are common for all domains isolated from other words that are domain-dependent. KISB analysis performs the process of stemming, which is the process of analyzing a word and then extracting its stem. This process needs programming skills less than needed in case of ADESS analysis (rooting).

AAS provides KAM with ability of preventing typos and duplication. AAS provides HIM with; ability of preventing typos, and ability of computing probability of matching between the stored premise and the received one. With other words, KISB succeeds in achieving several textual analysis purposes; spellchecker, knowledge similarity, premise validation, premise similarity, avoiding duplication, and avoiding inconsistent rules. But KISB dictionary needs more size than ADESS because of stem-based approach. In spite of the larger size, KISB dictionary achieves covering irregular cases that cannot be covered by ADESS.

3.1.3 [22] Root-Stem approach in General Analyzer System for Arabic language (RSGAS)

This paper example succeeds in building general Arabic analyzer system. The purpose is to validate any acquired knowledge from external user. So this work constructs a program that is able to be merged with any other acquired-knowledge-program. The job of the current example is to solve problems related to natural language understanding. Such problems occur in acquired knowledge, for example; typos, inconsistent knowledge, duplication, and similarity. This example manipulates texts written in Arabic language. It differs from the two previous examples in merging the two approaches; root-based, and stem-based. So this work gains the two approaches advantages, and avoiding their disadvantages. Another point of difference that enhances RSGAS performance is that RSGAS adds the feature of antonym to its dictionary. So, in addition to the common linguistic preprocessing that mentioned before, RSGAS adds the process of antonym detection.

RSGAS Arabic analyzer system is decomposed into five main components. The first component is the dictionary module, which stores roots for regular cases, and stores stems for irregular cases. This work is provided with medical-dictionary. Secondly is the word-analysis module, which performs three tasks; extracting the suitable stored root, correct typos, and providing ability of learning new words. Thirdly is the morphological-rules module, which performs two tasks; choosing suitable linguistic rule to analyze a specific word, and discover typos. The fourth component is the meaning module, which has ability to compute ratio of meaning matching for phrases and rules. So it provides the following abilities; achieves knowledge validation, prevents duplication, and prevents inconsistency. Then the last component is system-interface module. This module is a control that controls interconnection among the other four components. Also it controls communication with external environment.

RSGAS is succeeded in the manipulation of problems related to natural language understanding that occur at acquired knowledge. It discovers typos with correction. And because of its ability of computing similarity, it achieves jobs of preventing inconsistent knowledge, and preventing duplication. This work achieves advantage of reducing dictionary size compared with KISB dictionary. At the same time, the problem of irregular cases is solved, so it has many advantages over the two previous techniques ADESS and KISB.

3.1.4 [15] Using English Morphological Analyzer to Decrease the Dictionary Size in Keywords Extraction Techniques

This paper example develops automatic keyword extraction system. It was based mainly on English Morphological Analyzer [EMA], which was a stem-based. This work receives text (paragraphs) written in English language, analyzes text and then extracts list of keywords. EMA is decomposed into five components. A word that must be analyzed is received by the component called analyzer. This component is the control unit of EMA that achieves all required connections in order to extract the stem or root of the input word. Analyzer makes connection with lexicon component trying to find the word. If the process is failed, it will make connection with rule component and lexicon component together trying to decompose the word using English morphological rules. If the process is failed, it will make connection with stemmer analyzer component trying to find the word. Stemmer analyzer component makes connection with stemmer dictionary trying to find the word. In stemmer dictionary words are stored as groups. Each group is a set of stems that belongs to one root. The tested word will be searched in those groups and the root is the required result.

This work is provided with a dedicated-dictionary specialized for artificial intelligence domain. This dictionary contains two parts; the first part represents the lexicon component, and second part represents the stemmer dictionary component. This work achieves reducing dictionary size (31%), because EMA dictionary stores only stems and derives other related words using English morphological rules. In addition to the common linguistic processes that are mentioned above, this work performs others such as; word-frequencies, word position, sentence position, and part of speech (POS). This paper proved that merging those linguistics and statistics approaches achieves more accurate results and the threshold value was 50%.

3.1.5 [16] Text Summarization Based on Several Natural Language Techniques

This paper example developed a method for English text summarization, which is an extractive summarization. This work receives text (paragraphs) written in English language, analyzes it, and then constructs the summary paragraph. The developed method implies statistics and linguistic approaches mainly based on morphological rules. This work is provided with a dedicated-dictionary specialized for artificial intelligence domain. This work makes use of the proved approach at the previous example. So the current work uses EMA and other linguistic processes mentioned at the previous example. Idea of current work is to construct a new paragraph, which is decomposed from important sentences, that are includes keywords inside them. This work achieved successfully the

summary for any input text, and the best threshold value was 60%. Accuracy rate at the best threshold vale was 95%.

3.1.6 [17] Text Similarity Based on English Morphological Analyzer Approach

This paper example developed a method for automatic text similarity testing. This method succeeded in computing similarity ratio between texts. The input for this example is two texts (paragraphs) that are written in English language. The developed method implies statistics and linguistic approaches mainly based on morphological rules. This work is provided with a dedicated-dictionary specialized for artificial intelligence domain. This work makes useful of the proved approach at the previous example. Morphological rules, at current example; depended mainly on the proved method EMA. Here, testing text similarity is based on the two proved approaches at the two previous examples; keyword extraction, and text summarization. For short texts, firstly we extract keywords for the two input texts. Then we compute similarity ratio between the two lists of keywords. For long texts, firstly the researchers construct summaries of the two input texts. Then they compute similarity ratio between the two summaries. This paper achieved successfully high accurate method for automatic text similarity, with accuracy rates bounded from 98% to 100%.

3.2 Machine Learning Techniques Example

This section is interested with a textual application example that is based on machine learning techniques. Many textual applications apply machine learning techniques, for example "text categorization". Really, the current technique merge between machine learning techniques and word morphology found at previous set of examples. In "text categorization" application, two main steps must be performed; providing classification rules, and then classifying unknown documents based on the rules. Creating rules manually by experts achieves high accuracy, but it is costly. Machine learning techniques perform analysis on already-categorized texts, and then automatically create classification rules. Machine learning techniques are cost-saving [5].

3.2.1 [5] Hybrid Intelligent Techniques for Text Categorization

Proposed system mainly performs three steps; representing text documents, constructing classifier, and evaluating performance. First step includes collecting big number of pre-classified text documents. Features are extracted from the collected documents, and then weighted. Feature frequency determines feature weight. Some features are eliminated because of non-informatively. Then we apply stem-based approach to reducing features. Each on is returned to its stem. Threshold value determines the list of distinct features that represent the document.

Second step build the classifier. Supervised learning is applied to learn all features belong to any sub-category of main categories. Here, rough-set theory is used. Last step evaluates proposed system performance. The system performs its processes successfully and achieves accurate rate up to 96%.

3.3 Semantic Analysis Example

This section is interested with semantic analysis approach. This approach depends on the internal relations between segments of meaning. Segments may be either lexemes or words. Syntagmatic relation and paradigmatic relation are important. Here researcher likes to highlight an example in textual analysis, which is based on semantic analysis approach [23].

3.3.1 [23] Text Similarity Based on Modified LSA Technique

LSA stands for [Latent Semantic Analysis]. This paper is interested with similarity between texts based on semantic information. It manipulates texts that are written in English language. In semantic similarity the main process is to find the semantic similarity, distance between texts. Here two approaches were constructed. The two approaches used LSA technique to enhance the main process. These approaches were enhanced for both short sentences and long documents. The first approach determined the degree of similarity by using cosine similarity regardless text's length, while the second approach performs the process with regard text's length. The two approaches used Singular Value Decomposition (SVD) algorithm to generate "terms semantic space". They differed only in steps of calculating the final similarity. The first approach creates a semantic vector for each text. Its contents are; weights of matched words max of semantic similarity for unmatched words. It uses the cosine similarity function to compute final similarity. The second approach creates only one semantic vector. The vector contains; value 1 for matched words, max of max of semantic similarity for unmatched words (calculated using the average weights). Then the researchers calculate the sum values of semantic vector, and get its average. Then they calculate average length of the two texts. To compute final similarity, they divide semantic vector average on length texts average.

The two approaches results were compared with human performance. They achieved accuracy rate 92%, while accuracy of human performance is 76%.

4. Discussions

After reviewing the application and some examples of researches that deals with Textual Analysis, it is better to discuss and summarize the results in tables.

Firstly, the applications about Textual Analysis and examples of researches that used this application are discussed in table (1). The first columns contain the name of application; the other columns refer to the examples that are reviewed in this paper according to sequence that is mentioned in this paper. Number 1 to 6 refer to the examples that use Morphological and Syntax Analysis approach, number 7 refers to the example that uses semantic analysis approach, and number 8 refers to the example that uses semantic analysis approach as shown below:

1: is the first example which is Arabic Diagnosing Expert System Shell (ADESS) in section 3.1.1.

2: is the second example which is Knowledge acquisition & hybrid Inference with a Stem-Based approach (KISB) in section 3.1.2.

3: is the third example which is Root-Stem approach in General Analyzer System for Arabic language (RSGAS) in section 3.1.3.

4: is the fourth example which is Using English Morphological Analyzer to Decrease the Dictionary

Size in Keywords Extraction Techniques in section 3.1.4

5: is the fifth example which is Text Summarization Based on Several Natural Language Techniques in section 3.1.5.

6: is the sixth example which is Text Similarity Based on English Morphological Analyzer Approach in section 3.1.6.

7: is the seventh example which is Hybrid Intelligent Techniques for Text Categorization in section3.2.1.

8: is the eighth example which is Text Similarity Based on Modified LSA Technique in section3.3.1.

Table 1: examples of researches that use the applications of Textual Analysis

application	Example number							
	1	2	3	4	5	6	7	8
Spellchecker	√	√	√	√	√	√	√	
Phrase similarity	√	√	√	×	×	√		√
Keyword extraction	×	×	×	√	√	√	×	×
Text summarization	×	×	×	×	√	√	×	×
Text similarity	×	×	×	×	×	√	×	√
Text analysis in translation	×	×	×	×	×	×	×	×
Information retrieval	×	×	×	×	×	×	×	×
Information extraction	×	×	×	×	×	×	×	×
Question answering	×	×	×	×	×	×	×	×
Document clustering	×	×	×	×	×	×	×	×
Classifying	×	×	×	×	×	×	×	×
Entity linkage	×	×	×	×	×	×	×	√
Event detection	×	×	×	×	×	×	×	×

Table 2 summarized all the examples mentions above. This table contains columns; the first column (no.) refers to the examples that are reviewed in this paper according to sequence that is mentioned in this paper and as shown in the last table. The other columns refer to the researchers of that paper and the year published, the language used in this paper, approach, algorithm, strong point ,weakness point and accuracy rate for each paper.

Table 2: summarized all example of Textual Analysis reviewed in this paper

No.	Researchers	language	approach	algorithm	Strongest point	Weakness point	Accuracy rate
1	-Mashhadany A., Bashaga T. and Samawi V.(2010)	Arabic	morphological and syntax analysis	Root –based AMS	-Small dictionary size. - achieve these tasks: Correct typos, discovering phrases that are matched in meaning, preventing duplication and inconsistent rules.	-Cannot cover the wildness Arabic morphological rules and irregular cases. -need high programming skills.	
2	Al-Mashhadany A.(2012)	Arabic		stem –based AAS	-Achieve those tasks: spellchecker, knowledge similarity, premise validation, premise similarity, avoiding duplication, and avoiding inconsistent rules. -covering irregular cases.	Dictionary needs more size than root-based AMS	
3	Abeer K.Al-Mashhsdany, Abdulwadood K. Al-Mashhadany, Waleed K. Al-Mashhadany(2016)	Arabic		Root -based and stem –based system	-Solved irregular cases, -dictionary size more than root- based AMS but still less than stem based AAS. -makes analyzer close to natural language understanding.	need high programming skills	98%-100%
4	Ahmed T. Sadiq Al-Obaidi1 and Abeer Khalid AL-Mashhadany (2012)	English		Stem – base EMA, merging linguistics and statistical	-Reduce dictionary size to 31%. - success in extraction keywords	need high programming skills	86.6%
5	Abeer K. (2014)	English		Stem – base EMA, merging linguistics and statistical approaches.	increase the performance of text summarization method with minimum dictionary size	need high programming skills	78%-95%

6	**Abeer K. Al-Mashhadany, Sura M. Ali and Sawsan K. Thamer (2014)**	**English**		**Stem – base EMA, merging linguistics and statistical approaches.**	**more flexibility, more accuracy and reducing dictionary size.**	**need high programming skills**	**98-100%**
7	**Ahmed T. Sadiq, Sura Mahmood Abdullah (2013)**	**English**	**machine learning**	**Stem – based approach, machine learning**	**Reduce dictionary size**	**time consuming**	**96%**
8	**Ahmed. T. Sadiq, Khudhair J Kadhim 92015)**	**English**	**semantic analysis**	**Singular value decomposition algorithm**	**can apply on both long texts (documents) or short texts (sentences). It dynamic with change the corpus contents. it has suitable complexity time for applications**	**the corpus do not cover irregular verbs and long abbreviations, v**	**92%**

5. Conclusions

This paper studies some textual-analysis-application examples in details. From those examples the following points are concluded.

1- "Morphological and syntactic analysis" approach is succeeded in constructing the root-based Arabic morphological system in the first example. Here, the Arabic morphological system succeeded in achieving all purposes of this example. Such root-based system has advantage of reducing dictionary size. But, in Arabic morphological rules, the irregular cases and wideness caused weakness point in such root-based system.

2- "Morphological and syntactic analysis" approach is succeeded in constructing the stem-based Arabic analyzer system in the second example. Here the Arabic analyzer system succeeded in achieving all purposes of this example. Such stem-based system has advantage of covering irregular cases that could not be covered by root-based systems. But stem-based system dictionary require more size than dictionary of root-based system.

3- "Morphological and syntax analysis" approach is succeeded in constructing the root-stem-based Arabic analyzer system in the third example. Here the Arabic analyzer system succeeded in achieving all purposes of this example. Such root-stem-based system has advantages of reducing dictionary size and covering irregular cases. Another advantage of RSGAS is the antonym words. Adding more features for dictionary-words enhance the t refinement process and makes analyzer close to natural language understanding.

4- "Morphological and syntactic analysis" approach is succeeded in constructing the stem-based English Morphological Analyzer in the fourth example. EMA succeeded in extracting keywords with average accuracy rate 86.6%, which is better than 85.4 for classical approaches. The best threshold value was 50%. EMA reduced dictionary size into 31%.

5- EMA technique successfully achieved text summarization with minimum dictionary size. EMA technique achieves accuracy rate 95% at best threshold value (60%). Results proved that

EMA has ability to achieve good results (accuracy 78%) at low threshold value (20%).

6- EMA technique successfully achieved high accurate method for automatic text similarity computing. It achieves high accuracy rate between 98% to 100%.

7- Machine learning techniques successfully achieved high accurate method for automatic documents categorization. Merging machine learning techniques with word morphology technique gives high accuracy results and reduces size of dictionary. The average accuracy results for this technique reached to 96%.

8- Semantic analysis approach successfully achieved high accuracy method for automatic text similarity computing. It achieves average accuracy results reached to 92%. But the method based on EMA technique achieved better accuracy results (98% to 100%).

References

[1] A. B., P. S., M. N., G. A., H. A. (2015). Concept-level sentiment analysis with dependency-based semantic parsing: A novel approach. *Cognitive Computation*,4:7: 487–499

[2] R. P. (2018). Measuring populist discourse using semantic text analysis: a comment. *Qual Quant* 52:2163–2172

[3] L. M., C. C. B., and J. S. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research* 28:4:587–604.

[4] A. A.(2012). General Arabic Diagnosing Expert System Shell (GADESS). *LAMBERT Academic Publishing*.

[5] A. T. S., S. M. A. (2013). Hybrid Intelligent Techniques for Text Categorization. *International Journal of Advanced Computer Science and Information Technology (IJACSIT)*. 2:2:59-76.

[6] R. S., B. P., and C. L. (2006). Exploiting multilingual nomenclatures and language-independent text features as an interlingua for cross-lingual text analysis applications. *Cornell university library*.2-12

[7] G. G. (2001).Textual analysis and interpreting research, *SSLiMIT. University of Bologna*. 10:69-88

[8] R. G. (2003). What is Text Analysis, Really?. *Literary and Linguistic Computing*. 18:2:209-219

[9] H. R. , G. R. (1999). Text Analysis: Qualitative and Quantitative Methods. *Text Analysis*. 595-645.

[10] A. R. W., E. G. H., L. L. H. (2007). Text analysis technique. *United States Patent, Patent No.: US* 7,158,983 B2. Jan.

[11] A. R. W., E. G. H., L. L. H. (2008). Text analysis techniques. *United States (12) Patent Application Publication.* Patent No.: US 2008/0109454 A1.

[12] G. F. L. (2002). Artificial Intelligence: Structures and Strategies for Complex Problem Solving. 6the edition .Addison Wesley Longman.

[13] M. W. B., A. B. and A. K. S. (2014). Text analysis. Volume 4. London school of Economics and Political science.

[14] R. A. (2017). Spelling Detection Errors Techniques in NLP: A Survey. *International Journal of Computer Applications*. 172:4:1-5..

[15] A. T. S. A. and A. K. A. (2012). Using English Morphological Analyzer to Decrease the Dictionary Size ins Keywords Extraction Techniques. *International Journal of Research and Reviews in Soft and Intelligent Computing (IJRRSIC)*. 2:1:114-118.

[16] A. K., (2014). Text Summarization Based on Several Natural Language Techniques. *Eng. & Tec. Journal /University of Technology*. .32:2:379-385.

[17] A. K. A., S. M. A. and S. K. T. (2014). Text Similarity Based on English Morphological Analyzer Approach. *Journal of Al-Nahrain University Science*. 71:4:203-212.

[18] C. N. (2005). Text Analysis in Translation: Theory, Methodology, and Didactic Application of a Model for Translation-Oriented Text Analysis. 2nd. Rodopi. Amsterdam.

[19] B. Y. (2008). An evaluation of text classification methods for literary study. *Literary and Linguistic Computing*. 23:3:327–343.

[20] A. A., B. T. and S. V. (2010). Arabic Diagnosing Expert System Shell (ADESS*). The Proceedings of 2nd Conference for Information Technology: Applications and Horizons, University of Technology, Baghdad, Iraq.*

[21] A. A. (2012). Knowlwdge Acquisition & Hybrid Inference with A Stem-Based Approach (KISB). *Journal of Al Nahrain University-Science*. 15: 2:169-183.

[22] A. K. A., A. K. A., W. K. A. (2016). Root-Stem Approach in General Analyzer System for Arabic Language (RSGAS). *J. of University of Al-Anbar for pure science*. 10:3:41-49.

[23] A. T. S., K. J. K. (2015). Text Similarity Based on Modified LSA Technique. *Journal of Advanced Computer Science and Technology Research*. 5:2:64-70.

P- ISSN 1991-8941 E-ISSN 2706-6703 *Journal of University of Anbar for Pure Science (JUAPS)* Open Access
2018,12 (3) :71-83

تطبيقات التحليل النصي: مراجعة الموضوع

عبير خالد المشهداني [١] صفا أمين أحمد [٢]

[١] قسم علوم الحاسوب/كلية العلوم/جامعة النهرين/العراق/بغداد
E-mail:aabeeeeraa@yahoo.com.
[٢] قسم الهندسة الكيمياوية /الجامعة التكنولوجية /العراق /بغداد
80155@uotechnology.edu.iq

الخلاصة:

البحث الحالي عبارة عن مسح ادبي حول تطبيقات تحليل النصوص. يهدف البحث الى توفير شرح وافي حول تطبيقات تحليل النصوص المطروقة. تحدث البحث عن اهمية القاموس وهو عادة احد المكونات الرئيسية التي تدخل ضمن هيكلية تطبيقات تحليل النصوص. تم تسليط الضوء على عدد من الامثلة ذات الصلة، وهي مثبتة من خلال بحوث منشورة سابقا. تم وصف الميزات المشتركة للامثلة المذكورة. كما تم مناقشة النتائج. اثبت البحث الحالي فعالية منهج " التحليل المورفولوجي والنحوي ". كما اثبت ان اعتماد طريقة " التحليل المورفولوجي والنحوي " لاختبار تشابه النصوص يعطي نتائج اكثر دقة من اعتماد الطريقة الدلالية.