Islamic University of Gaza
Deanery of Post Graduate Studies
Faculty of Information Technology

Automatic Arabic Text Summarization System (AATSS) Based on Semantic Feature Extraction

By:

Kathrein Abu Kwaik

Supervised By:

Prof. Nabil M. Hewahi

A Thesis Submitted as Partial Fulfillment of the Requirements
for the Degree of Master in Information Technology

1432 H – August 2011

﴿إنا أنزلناه قرآنا عربيا لعلكم تعقلون﴾ [سورة يوسف 12: 2]

﴿وكذلك أنزلناه قرآنا عربيا..﴾ [سورة طه 20: 113]

﴿قرآنا عربيا غير ذي عوج لعلهم يتقون﴾ [سورة الزمر 39: 28]

﴿كتاب فصلت آياته، قرآنا عربيا لقوم يعلمون﴾" [سورة فصلت 41: 3]

﴿وكذلك أوحينا إليك قرآنا عربيا لتنذر أم القرى ومن حولها ﴾ [سورة الشورى 42: 7]

﴿إنا جعلناه قرآنا عربيا لعلكم تعقلون﴾ [سورة الزخرف 43: 3]

﴿وهذا كتاب مصدق لكتاب موسى لسانا عربيا لينذر الذين ظلموا﴾ [سورة الأحقاف 46: 12]

﴿ نزل به الروح الأمين.. بلسان عربي مبين﴾ [سورة الشعراء 26: 193و195]

﴿ وَهَذَا لِسَانٌ عَرَبِيٌّ مُبِينٌ ﴾ [سورة النحل 16: 103]

Abstract

Recently, one of the problems arisen due to the amount of information and it's availability on the web, is the increased need for effective and powerful tool to automatically summarize text. For English and European languages an intensive works have been done with high performance and nowadays they look forward to multi-document and multi-language summarization. However, Arabic language still suffers from the little attentions and research done in this filed.

In our research we propose a model to automatically summarize Arabic text using text extraction. Various steps are involved in the approach: preprocessing text, extract set of feature from sentences, classify sentence based on scoring method, ranking sentences and finally generate an extract summary. The main difference between our proposed system and other Arabic summarization systems are the consideration of semantics, entity objects such as names and places, and similarity factors in our proposed system. The proposed system has been applied on news domain using a dataset obtained from Falesteen newspaper. Manual evaluation techniques are used to evaluate and test the system. The results obtained by the proposed method achieve 86.5% similarity between the system and human summarization. A comparative study between our proposed system and Sakhr Arabic online summarization system has been conducted. The results show that our proposed system outperforms the Shakr system.

Keywords: *Automatic Text Summarization, Feature Extraction, Manual Evaluation, Natural Language Processing.*

عنوان البحث:

النظام الإلكتروني في تلخيص المستندات العربية اعتماداً على استخلاص الخصائص من حيث المعنى

ملخص:

في الآونة الأخيرة ظهرت العديد من المشاكل نتيجة ازدياد عدد المعلومات المنتشرة على الإنترنت, وأصبح الوصول إلى البيانات المطلوبة والمرادة من أصعب الأمور التي تواجه الباحثين.هذا أدى لظهور الحاجة إلى نظام محوسب آلي يقوم بتلخيص المستندات بشكل الكتروني وعرضها للمستخدم بحيث يستطيع تحديد ما إن كان المستند يفي بالغرض المطلوب أم لا.

إن العديد من أنظمة التلخيص الآلي تدعم اللغات الأوروبية وبالأخص اللغة الإنجليزية وتعطي نتيجة دقيقة جدا في انشاء ملخص عن اي مستند وقد تطور العمل على هذه اللغات لتصل إلى امكانية تلخيص عدة مستندات وبلغات مختلفة .إلا أنه وعلى الرغم من ذلك فإن هنالك قصور كبير في التعامل مع اللغة العربية في هذه المجالات ومازالت الأبحاث مستمرة في تطوير هذا المجال.

في هذا البحث سنقوم بعرض تصميم وانشاء لنظام تلخيص للمستندات العربية على اساس الاستخلاص من النص، هذا المقترح يتكون من عدة مراحل اهمها : مرحلة ماقبل المعالجة واستخراج الخصائص ومرحلة تصنيف الجمل النصية معتمدا على طريقة النقاط ومن ثم مرحلة ترتيب الجمل حسب أوزانها وقيمها وفي النهاية انشاء الملخص فيتم عرضه للمستخدم بطريقة سهلة وسريعة.

إن الإختلاف الرئيسي بين النظام المقترح وباقي الأنظمة العربية المستخدمة في التلخيص الآلي هو الاهتمام بعلم الألفاظ والمعاني سوياً واضافة نظام للتعرف على الاسماء والاماكن وبعض الاحداث الهامة من وجهة نظر الانسان وفحص مدى التقارب والتشابه مابين الجمل في المستند الواحد.

في عملية تقييم النظام وفحصه تم استخدام طريقة التقييم اليدوية وقد توصلنا إلى أن هناك تشابه مابين النظام وبين التلخيص البشري بنسبة 86.5 %. تم عمل دراسة مقارنة ما بين نظامنا وبين نظام صخر للتخليص الآلي. أظهرت النتيجة تفوق النظام المقترح عن نظام صخر.

الكلمات المفتاحية : تلخيص النصوص الآلي, استخلاص الخصائص, التقييم اليدوي, معالجة اللغة الطبيعية

To My Sweety Daughter…

Zain

ACKNOWLEDGEMENT

First, I thank Allah for guiding me and taking care of me all the time. My life is so blessed because of his majesty.

*I would also like to take this opportunity to thank my research supervisor, **Prof. Nabil Hewahi** for giving me the opportunity to work with him and guiding and helping me throughout this research and other courses.*

*I would like to thank the **Arabic language experts** from Islamic university of Gaza (IUG) and the **IT manager from Falsteen newspaper** for their help in collecting and annotating the data and system input data.*

I wish to express my considerable gratitude to many people who, in one way or another, have helped with the process of doing this research.

*I am very grateful to my dear **husband** without his encouragement I can't do this work.*

*Very special thanks to **my Father** for helping me through the research from the beginning until the evaluation step. Without your helping it's very hard to complete this thesis in short time.*

*Thanks, **My mother, Wael and Karo** for all things you do for me, your pray, patience, motivation and continues support.*

*Thanks to all my **friends** in IUG; whom I consider as sisters.*

Thank you all for being always there when I needed you most. Thank you for believing in me and supporting me.

I believe that without your support and your prayers, none of this work would be accomplished.

***Finally**, I hope this thesis be a useful addition to the research activities of Arabic natural language processing.*

Table of Contents

English abstract i

Arabic abstract ii

Dedication iii

Acknowledgment iv

Table of Contents v

List of Tables viii

List of Figures x

List of Abbreviations xi

1. Chapter 1 : Introduction

1.1.Text summarization 1

1.2.Summarization Fields 1

1.3.Taxonomy of Text Summarization 2

1.4.Arabic Natural Language Processing 4

1.5.Examples of Existing Summarization Systems 4

1.5.1. None Arabic Automatic Text Summarization Systems 5

1.5.2. Arabic Text Summarization Systems 6

1.6.Summary Evaluation 7

1.7.Statement of the problem 7

1.8.Objectives 8

1.8.1. Main objective 8

1.8.2. Specific objectives 8

1.9.Significance of the Thesis 9

1.10. Scope and Limitations of the work 10

1.11. Methodology 10

1.12. Thesis Overview 11

2. Chapter 2 : Related works

2.1.Single Document Summarization approaches 12

2.1.1. Arabic Works 16

2.2.Multi-Document Summarization 17

3. Chapter 3 : Automatic Arabic Text Summarization System (AATSS)

3.1.Data Acquisition 19

3.2.Pre-processing and Feature Extraction 19

3.2.1. Document Word Number (tokenization) 21

3.2.2. Sentence Segmentation Features 21

3.2.3. Pre-processing 23

3.2.4. Entity Recognition (ER) 26

3.2.5. Stemming 28

3.2.6. Semantic Feature Extraction 29

3.2.7. Check key phrase 31

3.3.Scoring 32

3.4.Ranking and Summary Generation 33

4. Chapter 4 : Experimental Results and Evaluation

4.1.Implementation 36

4.1.1. System Interface 37

4.1.2. Tools and Programs 38

4.2.Data Set 39

4.3.AATSS examples 39

4.4.System Evaluation 45

4.4.1. AATSS and Human Evaluation 45

4.4.2. Sakhr Summarizer and Human Evaluation 55

4.4.3. Comparison between AATSS and Sakhr online summarizer 62

4.4.4. Discussion 63

5. Chapter 5 : Conclusion and Future works

5.1.Conclusion 65

5.2.Future Work 65

6. References 67

7. Appendix A : System Interface and Code A

8. Appendix B : Summary Examples B

List of Tables

Table (3.1): Examples of sentences and paragraphs 21

Table (3.2): Arabic diacritics 24

Table (3.3): Punctuations 24

Table (3.4): Arabic definite Article 24

Table (3.5): Examples of strange words 24

Table (3.6): Examples of stop words 25

Table (3.7): Examples of pre-processing 25

Table (3.8): Examples of Convert (ـة) 26

Table (3.9): Entity Recognition categories 27

Table (3.10): Entity Recognition example 27

Table (3.11): Stemming example 28

Table (3.12): Example from Al-Taleb Arabic to Arabic dictionary 29

Table (3.13): TF× ISF example 30

Table (3.14): Semantic title similarity calculation 31

Table (3.15): Semantic similarity between two sentence 31

Table (3.16): Key phrase examples 32

Table (3.17): Sentence scoring example 32

Table (3.18): Summary generation example for 5 sentences 33

Table (4.1):.Stem word code 36

Table (4.2):.Examples of summaries generated by AATSS (no summary) 40

Table (4.3): Examples of summaries generated by AATSS (summary 60%) 40

Table (4.4): Examples of summaries generated by AATSS (50%) 42

Table (4.5): Summary Evaluation 1 46

Table (4.6): Summary Evaluation 2 49

Table (4.7): Summary Evaluation 3 52

Table (4.8): Sakhr Evaluation 1 55

Table (4.9): Sakhr Evaluation 2 57

Table (4.10): Sakhr Evaluation 3 59

Table (4.11): F-measure results 62

Table A.1 Calculate Term Frequency A3

Table A.2 Sentence Term Frequency code A5

Table A.3 Remove Diacritics Code A8

Table A.4 Check Strong words Code A9

Table A.5 Compute Score Code A10

Table (B.1): AATSS summary for economic text (1) B1

Table (B.2): AATSS summary for sport text B2

Table (B.3): AATSS summary for economic text (2) B3

Table (B.4): AATSS summary for political text B4

Table (B.5): AATSS summary for technology text B5

List of Figures

Figure (3.1): (A) The proposed AATSS flow-diagram 18

Figure (3.1): (B) The proposed AATSS architecture 18

Figure (3.2): Pre-processing and feature extraction 20

Figure (3.3): Cases of define word in the sentence 22

Figure (3.4): Entity Recognition flow chart 27

Figure (4.1): System interface 38

Figure (4.2): P- measure differences 62

Figure (4.3): R-measure differences 63

Figure (4.4): F-measure differences 63

Figure (A.1): System interface A1

Figure (A.2): Pre-process part A2

Figure (A.3): Variables panel A2

Figure (A.4): Original text area A3

Figure (A.5): Summary text area A3

List of Abbreviations

AATSS	Automatic Arabic Text Summarization System
ANLP	Arabic Natural Language Processing
CR	Compression Rate
DUC	Document Universal Conferences
ER	Entity Recognition
GP	Genetic Programming
GSM	General Statistic Method
HMM	Hidden Markov Model
IDF	Inverse Document Frequency
IR	Information Retrieval
JDK	Java Development Kit
MCS	Multi Classifier System
NLP	Natural Language Processing
P	Precision
PDA	Personal Digital Application
POS	Part of speech
R	Recall
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
RST	Rhetorical Structure Theory
SRL	Semantic Role Labeling
TF	Term Frequency
TF-IDF	Term Frequency × Inverse Document Frequency
TF-ISF	Term Frequency × Inverse Sentence Frequency
TS	Text Summarization

Chapter 1

Introduction

As the amount of information rapidly grows on the internet, there are a lot of difficulties to select the relevant information we need and that satisfy our requirements. Furthermore, publications media varies from specialist journals to newspapers to many other versions.

1.1 Text Summarization

Summarization of texts or Text Summarization (TS) appears as the best solution for users to choose and decide if this document will help her/him or not. Summarization is the process of producing shorter and informative presentation of the most important information from a source or multiple sources of information according to particular needs and specifications. Summarization is not applied only on text documents but also on any multimedia facility. Summarization of text documents became a very important issue due to the very large text sources already available. Users tend to extract the most informative and/or indicative information instead of reading the whole original document(s). [45]

Radev et al.; [40] define a summary as “a text that is produced from one or more texts, that convey important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that".

1.2. Summarization Fields

Nowadays the areas of automatic text summarization are extensive. TS can be useful in many fields such as:

- **Medical area**: A lot of documents are published for medical research in the last two-decades, and in many cases a medical specialist is in a deep need to find relevant information about patient’s conditions timely. So, text summarization here saving time resources and optimizing availability of medical experts.
- **Legal area**: legal resources and documents are sparse and expensive in time and expertise level, and cost, which yields legal experts perform difficult and responsible work. Thus automatic text summarization needed for expert to be able to find compressed and restated content of relevant judicial documents, including laws and their proposals, relevant court decisions or tribunal process summarizations.[41]
- **News area**: thousands of political, sport, economic and other types of news are published every second on the internet. It’s very hard or impossible to browse all of them or either the half. By text summarization user can find which news she/he is concern with before reading the whole text.

On the internet, one can find a lot of examples for automatic text summarization systems as:

- Google[1] News, Microsoft[2] news, Columbia[3] News blaster which return an abstraction summary for news in world.
- Blog summarization tool and aggregation[4] and opinion survey systems
- Sakhr summarization: this can summarize an Arabic text.
- Word summarizer: this also included in Microsoft.Word.
- Personal Digital Application (PDA) which contains summarizer applications.

For any area, automatic text summarization is a very powerful tool to save time and resources, and optimizing availability.

1.3. Taxonomy of Text Summarization

There are several, often related views which can be used to characterized text summarization. The main categories used to classify summarization are listed here:

1. Number of document:

    - Single document summarization: deals with one document and extract informative sentences only from it

    - Multi-document summarization: generates one summary by extracting the most important sentences from multi documents. [19]

2. Number of languages in the document:

    - Mono language summarization: generate summary from a document that contains one language.

    - Multi-languages summarization: here the document or set of documents contain at least two different languages.[31]

3. Form of summary

    - Abstractive summarization: it is the hardest task for computer researchers to solve this type successfully as it is concerned with semantic and language complexity. The summary containing sequence of words not present in the original document.

    - Extractive summarization: it consists of words, sentences and paragraphs that are completely appear in the original document. This approach suffers from inconsistencies, lack of balance and lack of cohesion. Also

[1]http://news.google.com
[2]http://msnbc.msn.com
[3]http://newsblaster.cs.columbia.edu
[4]For instance, http://www.bloghearld.com

some sentence may be extracted out of the context and anaphoric references can be broken. [24]

4. Processing level

    - Surface level approach: in this case the information is represented from the point of shallow features. These include different types of terms, e.g. statistically and positionally salient ones, terms from cue phrases or domain specific and user inputted terms. Usually this approach products extraction based summary as an output.

    - Deeper level approach: this approach may involve sentence generation. The advanced semantic analysis is necessary in order to achieve such task. The output of this approach may be in form of abstracts and extracts. [24]

5. Audiences of the summary (content):

    - Generic summaries: are aimed at a broad community of readers, all major topics are equally important.

    - Query-based summaries: are summaries built on the top of previously submitted user query

    - User or topic focused summaries: are tailored to the interest of the user or represent only particular topic. [24]

6. Details of summary:

    - Indicative summary: they should preserve the most important points in the document. Their aim is to help the user to decide whether the original document is worth reading. Their length ranges between 5 to 10% of the document.

    - Informative summary: it's more large than indicative, which return the important details while reducing information volume. Its length range between 20 to 30% of the complete text.

    - Critical or Evaluative summary: this capture the point of view of the summary author on a given subject (summary of summary). They are little bit out of scope of nowadays automatic summarizers. [24]

7. Genre of document:

    - Depend on the type of document (sport, news, technical, ...)

Any good summarization system should be characterized by the following quantitative features [24]:

- Semantic informativeness: it is viewed as the measure of ability to reconstruct from the summary the original text.
- Coherence: express the way how parts of the summary create together an integrated sequence.
- Compression ratio: the ration between the number of summary sentences and the number of original text sentences.

1.4. Arabic Natural Language Processing

Research in Arabic Natural Language Processing (ANLP) has focused on the manipulation and processing of the structure of the language at morphological, lexical, and syntactic levels. Unfortunately, semantic processing of the Arabic language has not yet received enough attention. [21]

There are some aspects that slow down progress in Arabic Natural Language Processing compared to the accomplishments in English and other European languages including [13]:

- The complex morphology
- the absence of diacritics in written text
- The fact that Arabic does not use capitalization.

In addition to the above linguistic issues, there is also a shortage of Arabic corpora, lexicons and machine readable dictionaries. These tools are essential to advance research in different areas. Despite these difficulties, there has been some success in tackling the problem of Arabic syntax as in [4] [16]. Few attempts have been made to develop Automatic Arabic summarization systems as [1] [14]. These attempts are list in the next section.

1.5. Examples of Existing Summarization Systems

Many systems have been developed in the area of automatic text summarization; some of these are available online while others are commercial software. However, most of these systems have no published documentation explaining their design or method of operation. In this section, we will present the available information about some of these systems.

1.5.1. None Arabic Automatic Text Summarization Systems

- **SweSum:** is the first automatic text summarizer for Swedish[1]. It summarizes Swedish news text in HTML/text format on the WWW. During the summarization 5-10 key words - a mini summary is produced. SweSum is also available for Danish, Norwegian, English, Spanish, French, Italian, Greek, Farsi (Persian) and German texts. SweSum is based on statistical, linguistic and heuristic methods. The system calculates the frequency of the key words in the text, in which sentences they appeared, and the location of these sentences in the text.

- **SUMMARIST:** SUMMARIST is a multilingual text summarization system developed in the Natural Language Group of the Information Sciences Institute of the University of Southern California. It is an attempt to develop robust extraction technology as far as it can go and then continue research and development of techniques to perform abstraction. It produces extract-type summaries from newspaper, newswire, and other texts. It operates on texts in English, Spanish, French, German, Indonesian; older versions have also operated on Japanese. [23]

- **LexRank**: LexRank is a text summarization system developed at University of Michigan by Gunes Erkan and Dragomir Radev, it is available online at http://clair.si.umich.edu/clair/lexrank/

- **Automatic Text Summarizer**: The Automatic Text Summarizer is an online English text summarization tool that is available at http://www.makeuseof.com/dir/automatic-text-summarizer-text-summarization-tool/

- **Kify Online Text Summarizer:** It uses Text Semantic Indexing and some mathematics, the Summarizer decides which parts of a document are important by analyzing the content. It is available online at http://text.kify.com/

- **Intellexer Summarizer 3.1**: It is a commercial tool. Intellexer Summarizer is an innovative program for computers that creates a short summary from any document or a browsed web page. It is available at http://www.fileguru.com/Intellexer-Summarizer/info

- **QuickJist summarizer 1.2:** It is capable of highlighting summaries of web pages directly in browser windows. http://www.filecluster.com/Internet/Browser-Tools/Download-QuickJist-summarizer.html

- **Subject Search Summarizer 2.0:** It is a commercial system available at http://subject-search-summarizer.smartcode.com/info.html

[1]http://swesum.nada.kth.se/index-eng.html

- **Sinope Summarizer**: Sinope summarizers used a semantic analysis to transform the text into a structure for a computer. This structure looks like a network that contains the concepts from the original text and the relationships between them. This structure is called a Semantic Structure and all kinds of mathematical operations can be performed on it that are impossible to perform on the original text. Sinope Summarizer uses advanced mathematical techniques to analyze the Semantic Structure to determine which information elements are important and removes all irrelevant information elements. Sinope is the only system that uses semantic analysis; however, no documentation was to be found. The tool is available at http://www.filecluster.com/Internet/Browser-Tools/Download-Sinope-Summarizer-PE-Trial.html

- **Copernic Summarizer:** It uses statistical and linguistic algorithms. It is an easy to use tool for text extractive single document summarization. It generates summary reports with key concepts (up to 100) and key sentences according to the configured or customized summary length. Also it exports summary reports to various file formats (HTML, XML, Rich Text Format, Text file and URLs). It supports English, French, Spanish and German languages. The tool is available at http://www.copernic.com/en/products/summarizer/

1.5.2. Arabic Text Summarization Systems

- **Lakhas**: An Arabic text summarization system using extraction techniques. It is the first Arabic summarization system to be formally evaluated and compared with English competitors in an evaluation competition [14].

- **AQBTSS**: is a query-based single document summarizer system that takes an Arabic document and a query (in Arabic) and attempts to provide a reasonable summary for the document around this query. [17]

- **ACBTSS:** It integrates Bayesian and Genetic Programming (GP) classification methods in an optimized way to extract the summary sentences. The system is trainable and uses manually labeled corpus. Features for each sentence are extracted based on Arabic morphological analysis and part of speech tags in addition to simple position and counting methods. Initial set of features is examined and reduced to an optimized and discriminative subset of features. Given human generated summaries, the system is evaluated in terms of recall, precision and F-measure. [17]

- **The Summarizer of AramediA**: Based on Sakhr's accumulated research in Natural Language Processing (NLP), and Morphological Analysis over many years. The Summarizer is used for summarizing Arabic and English documents. Based on linguistic analysis of the document, it extracts the main ideas to make it possible for the user to preview these ideas instead of reading the whole document, thus saving time and effort.

- **Sakhr Summarizer:** The Sakhr Arabic summarizer engine identifies the most relevant sentences within a text and displays them in the form of a short text

summary. The summarizer makes it easy to scan just the important sentences within documents, greatly reducing the time needed to read and process items manually. It provides companies with short text summary for each item using a prioritized list of key sentences and gives the ability to select a specific level of summarization. The summarization engine employs the Sakhr Corrector to automatically correct the input Arabic text from common Arabic mistakes, and the keywords extractor to identify a prioritized list of keywords to accurately identify the important sentences. The tool is available at http://textmining.sakhr.com/

1.6. Summary Evaluation

Evaluating a summary is a difficult task because there is no an ideal summary for a given document or set of documents. From any papers or researches concerns with TS we can find that agreement between human summarizers is quite low, both for evaluating the automatic summary and generating manual summaries. Evaluating summary quality is considered to be more difficult than evaluating the summary. Another important problem in summary evaluation is the widespread use of difference metrics for evaluation. There is no standard human or automatic evaluation metric, which makes it very hard to compare different systems and establish a baseline. Besides this, manual evaluation is too expensive: as stated in [29] [22], large scale manual evaluation of summaries as in the Document Universal Conferences (DUC) would require over 3000 hours of human efforts. [12]

Summary evaluation methods attempt to determine how adequate and reliable or how useful a summary is relative to its source. Generally, there are two types of evaluation methods. The first is intrinsic evaluation in which users judge the quality of summarization by directly analyzing the summary. Users judge directly, how well the summary covers main key ideas, or how it compares to an ideal summary written by the author of the source text or a human abstractor. None of these measures are entirely satisfactory. The ideal summary, in particular is hard to construct and rarely unique. In most cases there is no only one correct ideal summary for a given document. The second type of evaluation methods is extrinsic. Users judge a summary‘s quality according to how it affects the completion of some other task, such as how well they can answer certain questions relative to the full source text. Recall-Oriented Understudy for Gisting Evaluation (ROUGE[1])is also used for automatic summary evaluation by counting the number of overlapping units such as word sequences, and word pairs between the computer-generated summary to be evaluated and the ideal summaries created by humans but this system does not support Arabic evaluation. For manual evaluation, extractive approach for summarization enables us to use recall, precession and F-measure[2] to evaluate summaries [44].

1.7. Statement of the problem

The problem of this research is how to develop a model to generate an automatic Arabic text summarization based on extraction that can be valid for various domains

[1](Automatic Evaluation of Summaries Using N-gram Co-Occurrence Statistics)

[2]Precision can be seen as a measure of exactness or fidelity, whereas Recall is a measure of completeness, F-measure balances recall and precision.

with high performance using precision and recall measurements and depending on quantitative features.

The sub problems we face are:

1. What is the proper domain that we shall use to show the power of our proposed model? (Political, technical, sport…)

2. How to collect Arabic vocabularies? Because most of text data corpus do not support Arabic Language.

3. How to pre-process Arabic text?

4. What are the most relevant features to be extracted?

5. How to add semantic information to features?

6. What would be the used methods to score and classify the sentences?

7. How to obtain the dataset for the experimentations?

8. What is the proper approach to produce the summary?

9. How to evaluate the summary?

1.8. Objectives

1.8.1. Main objective

The main objective of this research is the development of an automatic Arabic text summarization model that can be used for various domains to generate a summary for Arabic document with recognizing novelty and ensuring that the final summary is both coherent and complete. We shall try to increase the performance of our proposed model by avoiding the drawbacks of the previous methods as shall be presented in the next chapter such as semantic ignorance and limited number of extracted features.

1.8.2. Specific objectives

The specific objectives of this research are:

- Using a domain that can be widely used and in important area which can give a good evaluation environment.

- Having an Arabic corpus, either by finding out a ready corpus or building a new one with the help of Arabic experts.

- Preprocess the Arabic text using various tools and based on some approaches that we shall develop. .

- Feature extraction using other researchers approaches in addition to certain methods developed by us. We shall have a focus on semantic features and Entity Recognition Features.

- Using special Arabic to Arabic dictionary to add semantic information to feature.

- Developing a simple scoring method that depends on features weight to score each sentence.

- Collecting datasets can be achieved from various domains. These can be obtained from various organizations working in various areas. Good dataset can be collected from news agencies.

- A summary will be produced either by ratio or number of sentences.

- Implement the proposed model.

- Test the capability of the proposed model. We shall use various evaluation techniques (manual and comparing with automatic system) in order to increase the ratio of precision and recall measurements.

1.9. Significance of the thesis

- More support for the Arabic language in the technology area.

- Apply Arabic summarization on large domain areas as an example: these systems can be used by librarians to generate a brief for any documents in libraries, and also can be used in tutoring and learning System.

- Saving efforts and time by helping the user to find which document fits with her/his interest based on the produced summary.

- Increase friendly use of computer to make the computer more Intelligent.

- Applying semantic value on feature extraction which increase the power of the summary

1.10. Scope and Limitations of the work

As presented in the introduction section, there have been two types of summarization systems: single document summarization and multi-document Summarization. In this research we focus on:

- Extraction of automatic summary for single Arabic text document.
- The model is going to be tested on some special domains as (news- sport- economics, technical) despite it is meant to be a general model.
- A manual evaluation technique will be followed because there isn't any automatic evaluation for Arabic summarization.
- The proposed approach accepts only file with .txt format and special Unicode of cp1256.
- The produced summary size is decided by the user as he/she can enter the number of sentence needed in the generated summary or chose the default criteria from section which explain in section 4.3.

1.11. Methodology

Many attempts and achievements have been reached for text summarization in English language and other European and Asian languages. Arabic summarization still suffers from little attention. Most of Arabic summarization systems depend on static features and the syntactic features of the language only.

In our model we shall try to solve an automatic text summarization for Arabic documents based on both syntactic and semantic features of the Arabic language. Four main steps are introduced to build:

1. Data acquisition
2. Preprocessing and feature extraction
3. Scoring
4. Ranking and generate summary.

In the model implementation, we depend mainly on java programming language and using a free source stemmer to stem each Arabic word in the text. For the system interface, the user has an option to assign the number of sentences to be included in the summary and extracted from the text, if this option is not used, the system will generate a summary according to the summary size criteria presented in section 4.3.

1.12. Thesis Overview

Chapter two is a literature survey of the automatic text summarization and current approaches. Chapter three defines in detail the system structure and the proposed approach including the corpus collection and annotation details, features selection and extraction and the used classification approaches (scoring approach). Chapter four includes experiments, the used summary evaluation techniques and system results. Finally chapter five presents the conclusions and future works.

Chapter 2

Related works

Usually, the flow of information in a given document is not uniform, which means that some parts are more important than others. The major challenge in summarization lies in distinguishing the more informative parts of a document from the less ones. Though there have been instances of research describing the automatic creation of abstracts, most work presented in the literature relies on verbatim extraction of sentences to address the problem of single-document summarization.

In this chapter firstly, we look at early work from the 1950s and 60s that kicked off research on summarization. Second, we concentrate on approaches involving machine learning techniques published in the 1990s to today. Finally, we briefly describe some works that concern with automatic Arabic text summarization.

2.1. Single Document Summarization approaches

Father of Information Retrieval (IR), Hans Peter Luhs in 1958 proposed the first approaches of the Automatic text summarization to generate an abstract for document. In his method he used the frequency of particular word as a measure of sentence significant [32]. He derived a significant factor that reflects the number of significant words occurrence within a sentence, and the linear distance between them due to the intervention of none significant words. After that, all of these sentences in the document are ranked and the top ranking sentences are selected to include in abstract summary.

Baxendale also in 1958 [10] proposed a summary that depends on the sentences position in document. He found that in 85% of paragraphs, the topic sentence come in the first and only in 7% it was last. Then one can chose any of these sentences to form the summary.

In 1969 Edmundson [15], developed a protocol for creating manual extracts. He combined the two features from Luhs in [32] and Baxendale in [10] and added another two features: (i) the presence of cue words (like important, hard, significant ...) and (ii) the skeleton of document (title, header, paragraph …), then each sentence has a factor reflect the summation of the four features. The sentences with high factor are chosen for summary. He found in evaluation his protocol that 44% of the auto summaries matched the manual summaries.

Kupiec et al**;** in 1995 [26] proposed a classification technique using Naïve-Bayes method which learning from data and the four features proposed in [15] which are: frequency of particular word in a sentence, sentence position in the document, the presence of cue words and the skeleton of document (title, header, paragraph,..) with other two features (sentence length and the presence of uppercase words), then make each sentence as worthy of extraction or not.

Lin and Hovy in 1997 [30], proposed an approach which is based mainly on sentence position. They say that every text has a predictable discourse structure and the sentences of generator topic centrality tend to occur in certain specific location. They used a newswire corpus and produce a collection of text from TIPSTER program[1] . They measure the yield of each sentence position against the topic keywords. They rank the sentence position to produce optimal position policy for topic position.

Barzilay and Elhadad in 1997 [6], present a new algorithm to compute lexical chains ina text, merging several robust knowledge sources: theWordNet thesaurus, a part-of-speech tagger, shallowparser for the identification of nominal groups, and asegmentation algorithm.They do the following: (i) segment the text, (ii) identify the lexical chain, (iii) using strong lexical chain to identify the sentence worthy of extraction. They describe the term of cohesion in text as a measure of sticking together different parts of the text. Their approach is a notable example when semantically related words are used.

Marcu in 1998 [34]; proposed a unique approach for Automatic Text Summarization. His main assumption is that the sentences in a document do not form flat sequences. He used a discourse theory which is Rhetorical Structure Theory (RST) [33]. In his approach he introduces a text tree to measure distinction between what is more essential to the writer purpose than ordinary text.

Aone et al**;** in 1999 [5] also as [26], proposed a method using Naïve-Bayes classifier but with richer features. They used features like Term Frequency (TF)and Inverse Document Frequency (IDF) to derive signature words.

Lin in 1999 [28] broke away from the assumption that features are independent of each other and tried to model the problem of sentence extraction using Decision Tree instead of Naïve-Bayes classifier. He used some novel features as query signature, IR signature, numerical data, proper name, pronoun or adjective, weekday or moth name and quotations.

Conroy and O'leary in [11] modeled the problem of extracting a sentence from a document using a Hidden Markov Model (HMM). They use only three features: (i) position of sentence in document, (ii) number of terms in sentences and (iii) likeliness of the streams given the document terms. There basic motivation for using Hidden Markov Model is to account local dependencies

Osborne in 2002, [37] claim that existing approaches for summarization have always assumed feature independence. He used a Log-Linear model to obviate this assumption and showed empirically that the system produce better than a Naïve-Bayes classifier. Sex feature are used by this model which are: word pairs, sentence length,

[1]The DARPA TIPSTER Text program was started in 1991 by the Defense Advanced Research Projects Agency (DARPA). It supported research to improve informational retrieval and extraction software and worked to deploy these improved technologies to government users. This technology was of particular interest to defense and intelligence analysts who must review increasingly large amounts of text.

sentence position, and naïve-discourse features like inside introduction or inside conclusion

Svore et al in 2007 [48] proposed an algorithm based on Neural Network and used a set of features stored in database to tackle the summary problem. They produce news extracts system (NetSum) which extracts the most three significant sentences from the news article. He used a Rank-Net algorithm [8] to classify and extract sentences. Training in NetSum is based on modified back-propagation algorithm for two layer networks. The system performance is considered to be better than all the previous systems for news article summarization. The performances of NetSum with external features are statistically significant at 95% confidence. But the main limitation of this system is the generated summary only contains three sentences not more.

Verma et al., in 2007 [49] focus on dynamic summary generation based on user input query. When the user prints some words to find a document, these words may not match the document main idea and retrieving the document abstract will not help the user. Hence the summary needs to be generated in accordance with the user query. In their approach, they couple the summary system with two ontology knowledge sources: WordNet [1] and UMLS[2].

This system consists of three steps:

(i) Evaluation and adjustment of the query in regards to the two previous ontologies.

(ii) Calculation of the distance of document's sentences to the relevant query.

(iii) Calculation of the distance among the candidate summary sentences.

The limitations of this system are: (i) it has designed for specific domain (medical) and (ii) lack of syntax analysis, insufficient query analysis and redundancy reduction problem.

Bawakid and Ovssalah in 2008 [9] proposed a similar approach to Verma et al., in [49] based on semantic analysis of document. In this approach, they used a combination of static features including sentence position and named entity in each sentence- and dynamic feature includes the semantic similarity between sentence and user query.

This system consists of three steps:

[1]WordNet is a large lexical database of English, developed under the direction of George A. Miller. The database consists of linked words, including nouns, verbs, adjectives and adverbs.
[2]UMLS is maintained by US National Library of Medicine and includes three knowledge sources: the Metathesaurus, the Semantic Network and the Specialist lexicon.

(i) Preprocessing each sentence.

(ii) Extracting and analyzing of features.

(iii) Generating summary.

The proposed system can be used for single and multi document summarization. However, this approach suffers from redundant information.

Al-Hashemi in 2010 [3] proposed a technique to produce a summary of an original English text. His model consists of four stages:

(i) Pre-processing stage [stop word removal, Part of speech (POS)]

(ii) Extract important key phrases in the text using special algorithm for ranking the candidate words

(iii) Extract the most ranks sentences

(iv) Filter sentence and assign the document to the related category.

In his work he selects sentence according to many features (sentence position in the document and the paragraph, key phrases existence, existence of indicated words, sentences length and sentence similarity to document class). Then a classical supervised machine learning method is used for document classification. Instance based learning method [2] is the classification method that the proposed system implements. The size of training set is 90 documents and tested by 20 documents. To evaluate the system they used Precision (P) and Recall (R) measurements. The system achieves 70% for overall Precision.

Suanmali et al.; in 2010 [47] proposed Semantic Role Labeling (SRL) approach to improve the quality of the English summary created by the General Statistic Method (GSM). They perform text summarization based on GSM and combine it with the SRL method. For pre-process stage, there are four main activities performed: (i) sentence segmentation, (ii) tokenization, (iii) stop word removal, (iv) word stemming. They extract the important sentences based on both sentence semantic similarity and GSM. To calculate sentence semantic similarity the used: WordNet and PropBank[1] semantic annotation. In feature extraction, they used: (i) syntactic similarity with title, (ii) sentence length, (iii) term frequency, (iv) sentence position, (v) occurrence of proper nouns, numerical data and thematic word. Finally; they combine SRL and GSM to give a score for each sentence using summation

[1]PropBank is a corpus that is annotated with verbal propositions and their arguments—a "proposition bank". Although "PropBank" refers to a specific corpus produced by Martha Palmeret al., the term propbank is also coming to be used as a common noun referring to any corpus that has been annotated with propositions and their arguments.see https://verbs.colorado.edu/propbank/

method. The high score sentences are extracted to form the summary. For evaluation, they used ROUGE with 100 documents and reach F-measure of 0.52431 which is higher than MS-Word summarizer and Baseline summarizer tools.

2.1.1. Arabic Works

It is to be noted based on the available literature and to the extent of the researcher's knowledge, all of the Arabic summarization systems are concerned with single document summarization.

Lakhas, An Arabic summarization system was proposed by Douzidia and Lapalme in [14]. In this system Arabic text is summarized directly into Arabic summary then the summary is translated after that to English text, so it appears as English summarizer not Arabic one because the final summary is in English not in Arabic. Group of features are extracted from each sentence as: (frequency computation, indicative expression, stop word removal) and a score function with weights are calculated. To generate a summary, the top ranked sentences are extracted and apply to summary. After that they use some removal word steps to minimize the summary to 10 words only. They use little features in the score function ignoring very useful features as sentence position, overlapping between sentence and similarity with document title. In their work they also didn't use any classifier to increase the decision for each sentence. The system finally generates English summary so they used ROUGE to evaluate their summary. They also translated the original document to English document and using this one into evaluation with the English summary. ROUGE shows that Lakhas results are ranked about 5^{th} or 6^{th} compared to other system.

Abdullah et al.; in [1] report a proposal of a platform for making the summary of Arabic texts as well as its architecture and its resolution. This platform consists of a group of existing tools from different resources to help human summarizer in the realization of Arabic summaries of text. So they only group and organize a set of tools from the beginning until extracting a summary. They use different modules as: text tokenizer, morphological analyzer, and parser. For extraction they depend on lakhas module in [14] to score sentence. The drawback was the same as lakhas because they use the same version of it without any classifier and didn't add any new relevant features. The evaluation of the platform was carried out on various types of texts length (short text between 30 and 150 words, average texts between 150 and 250 words, long text between 250 and 500 words) according to the execution time of their platform using two type of corpus from news paper articles Dar Al-Hayat and Tunisian education basic year school book. They measure the execution time for each module alone when applying to the summary. In general we can say that they didn't evaluate the summary according to its coherent or completeness or either correctness but only compute the execution time.

Sobh et al.; in [44] introduce an Arabic extractive text summarization system. This system integrates Bayesian and Genetic Programming (GP) classification methods in an optimized way to extract the summary sentences. The system is trainable and use manually labeled corpus. They extract features for each sentence based on Arabic morphological analysis and part of speech tags in addition to simple position and counting methods. After extraction, they use -as we mention before- Bayesian and GP

in different manners to generate some versions of the summary either by integrating the two results or by selecting the max score between them. Using GP method didn't add any powerful value to the model as the result say. Using Bayesian alone increase the precision of the summary and saving the time needed for GP computation. The authors didn't use some useful features as user defined keywords, named entities or indicator phrase which will increase the system controllability and results. Also; if they add some semantic information from lexical resource this will enhance output cohesion. In Evaluation, three important measures are used, precession, recall and F-measure. Precession is a measure of how much of information that the system returned is correct and Recall is a measure of the coverage of the system where F-measure balances recall and precession. They have 4 type of summarization system according to the combination between Bayesian and GP which are: (i) Bayesian, (ii) GP, (iii) Bayesian and GP, (iv) Bayesian or GP. From evolution they found that using Bayesian or GP achieves they highest F-measure between the four approaches which reach to 0.599 when they use only five features (sentence length, sentence paragraph position, sentence similarity, number of infinitives, number of verbs).

Based on the above presented research, most of the works lack to have comprehensive features related to Arabic languages such as using of special words, and various meanings for the same word (semantics and synonyms). Also they lack to have corpus for locations and persons names which can change the sentence weights.

2.2. Multi-Document Summarization:

When a user query about a topic, hundreds of documents are returned to him. If we deal with each document alone and summarized it, then hundred of summaries are generated which are also a problem. In today's community in which time plays important role, multi-document summarizer play essential role in such situations.

The field of multi-document summarization has been pioneered by NLP group at Colombia University. Mckeown and Radev in 1995 from this group develop a summarization system called SUMMONS depend on template-driven message understanding system [36]. Extractive techniques have been applied, making use of similarity measure between pairs of sentences. Different approaches deal with these similarities in different ways: Mckeown et al, in 1999 and Radev et al., in 2000 identify common themes through clustering and then select one sentence to represent each cluster [35] [39]; Barzilay et al., in 1999 generate a composite sentence from each cluster [7].

Some recent work extends multi-document summarization to multi-language environment which proposed by Evans et al**.,** in 2005 [18].

Chapter 3

Automatic Arabic Text Summarization System (AATSS)

In this chapter we shall present our proposed Automatic Arabic Text Summarization System (AATSS). The proposed system will be described using flowcharts, algorithms, figures and tables. Various stages have to be performed to achieve text summarization. The main required steps are shown in Figure 3.1 (A and B) and stated below:

1. Data Acquisition
2. Pre-processing and Feature extraction
3. Scoring
4. Ranking and Generating Summar

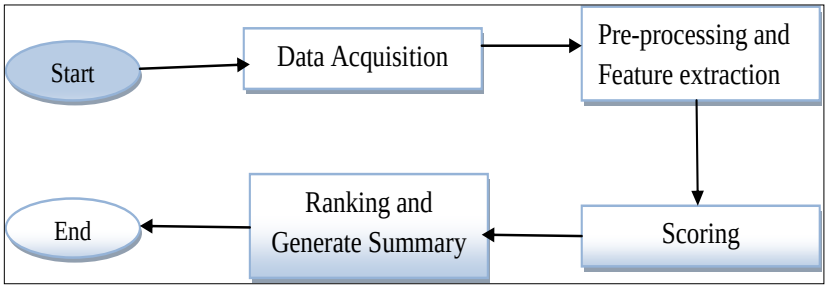

Figure 3.1. (A) The proposed AATSS flow-diagram

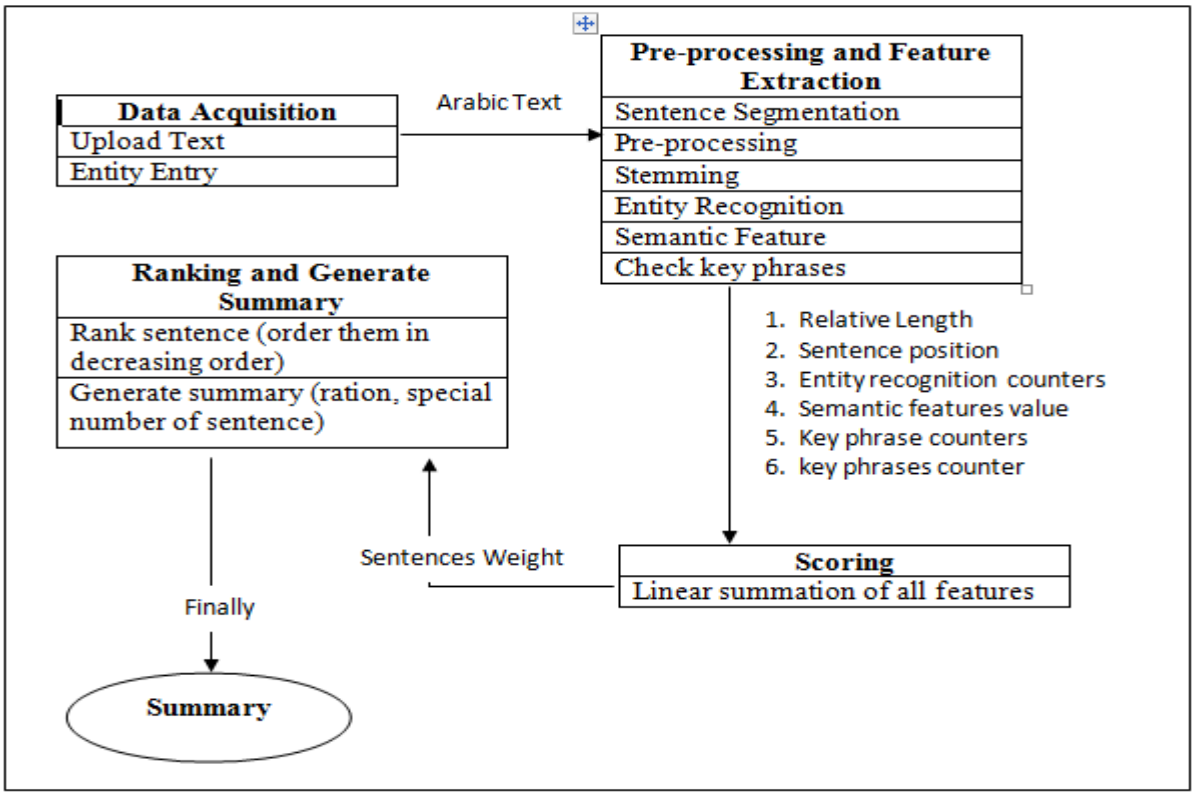

Figure 3.1. (B) The proposed AATSS architecture

3.1. Data Acquisition

This step is considered to be the system inputs. The system inputs can be classified into two types: user text input and support system inputs. In the user text input, the user is asked to upload an Arabic text file with (.txt) format only and a special Unicode of cp1256. This file is one to be summarized. In support system inputs, the user can provide the system with some name entity for persons, places and important terminology related to the problem domain. In the user text input, there is no limitation on the size of the text or number of sentences included in the article. On the other hand user should determine the number of summary sentences he/she needs to include in the final generated summary.

3.2. Pre-processing and Feature Extraction

In contrary of what is used in the most of text summarization systems by separating the stages of pre-processing and feature extraction, in our proposed system we combine the two stages in one. In this system we deal with pre-processing and feature extraction as interrelated stage. We need to extract some features before processing the text as extract number of words for each sentence and the absolute sentence position. Other features need special type of processing as entity recognition feature which requires converting all (ة) to (ه), also semantic feature needs to stem all words before extracting them. Figure 3.2 depicts pre-processing and feature extraction stage and shows how the pre-processing and feature extraction are interrelated.

Figure 3.2 shows the following steps:

1. Firstly we will count the number of words in the document.

2. Apply sentence segmentation to extract relative length feature and sentence absolute position feature.

3. For each sentence in the document, we will apply sentence pre-processing. The resultant sentence after this step is called (pre-process sentence), then we get a (ta'a sentence) by converting (ة) to (ه).

4. Calculate all entity recognition features and remove their occurrences from the input sentence to obtain a (post-pre-sentence). This stage produces five features:

    - Name counter
    - Place counter
    - Special words counter
    - Months and days counter
    - Numerical counter

5. For each post-pre-sentence we apply stemming processing and produce a (stem-sentence) which contains only the stems for each word.

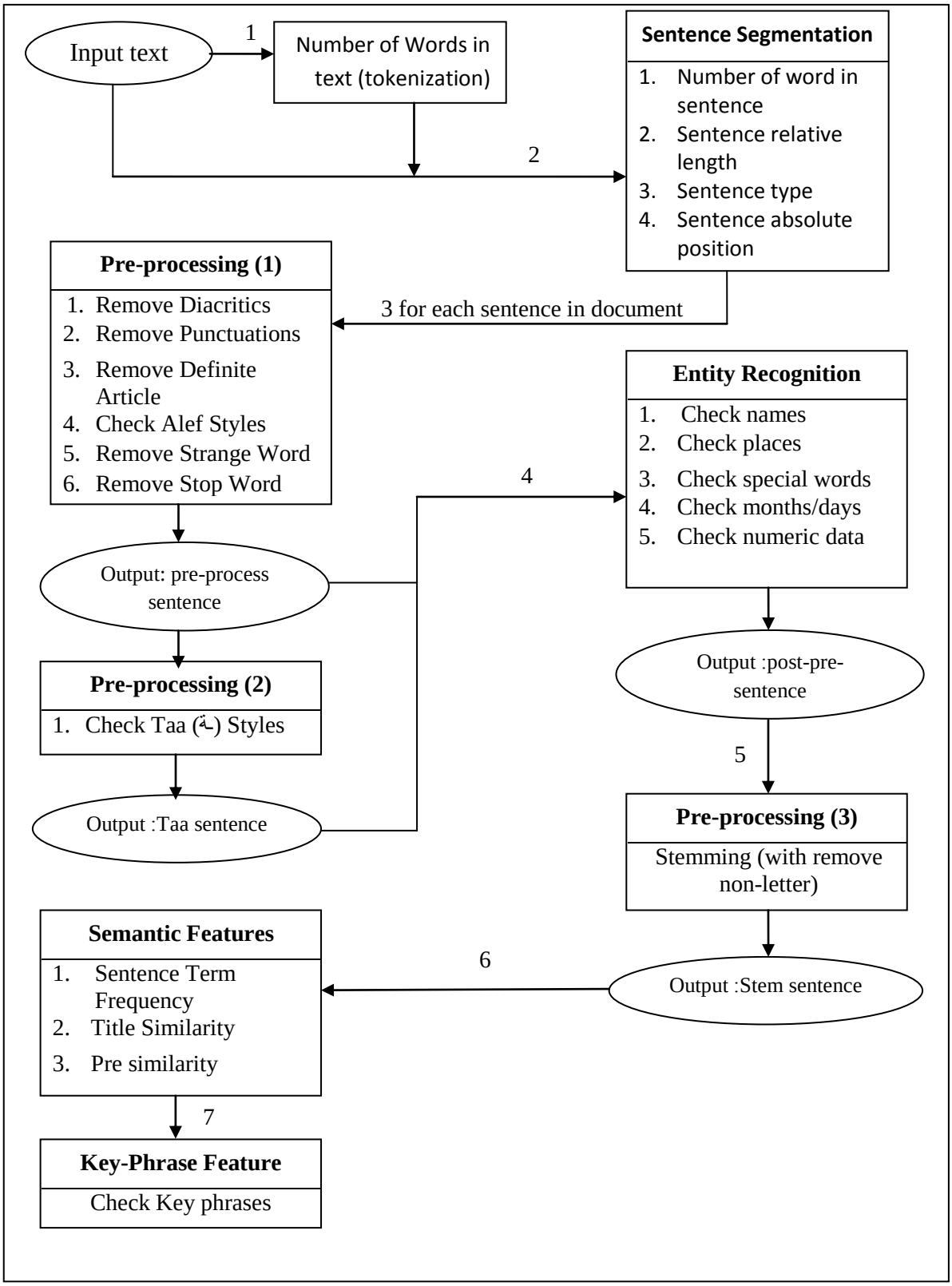

Figure 3.2. Pre-processing and feature extraction

6. Semantic features are extracted in this step from stem-sentence and we have:

    - Term frequency feature.
    - Title similarity feature
    - Similarity with previous sentence feature

7. Also from stem-sentence we check some key phrase and extract the following feature:

    - Strong key phrase feature
    - Middle key phrase feature
    - Weak key phrase feature

Now we will describe briefly each step:

3.2.1. Document Word Number (tokenization)

For each text, we calculate the number of words in it using a word tokenizer based on the following algorithm

1. While (Not End of Text)
    1.1. Tokenize a word
        1.1.1. Increase num_of_word

This is not a system feature but we will use it for another purpose as explained later in sentence length and sentence relative length as in section 3.2.2.

3.2.2. Sentence Segmentation Features

A sentence is that part of the text which ends with full stop (.) whereas a paragraph is ended by new line. So a paragraph might consist of group of sentences. Table 3.1 shows some examples of sentences and paragraphs.

Table 3.1. Examples of sentences and paragraphs

Text	Number of sentence	Number of paragraphs
نفى محافظ شمال سيناء في جمهورية مصر العربية اللواء عبد الوهاب مبروك، ما تناقلته مصادر إسرائيلية بشأن تهريب كميات كبيرة من الأسلحة ومضادات الصواريخ إلى غزة عن طريق ليبيا ومنها إلى شبه جزيرة سيناء ثم عبر الحدود والأنفاق إلى قطاع غزة.	2	2

		وقال مبروك، في تصريحات أوردتها عدد من الصحف المصرية الصادرة السبت 11-6-2011، إن "سيناء مسيطر عليها أمنيا وخاصة في مداخلها ومخارجها في نفق الشهيد أحمد حمدي وكوبري (جسر) السلام وفوق قناة السويس، هذا غير الكمائن الثابتة والمتحركة المدعومة بقوات كبيرة من الجيش المصري".
1	2	أكد مصدر فلسطيني مسؤول، أن الجانب المصري أعاد إغلاق بوابة معبر رفح البري وأبلغ الجانب الفلسطيني على المعبر أن نصف البوابة سيكون مفتوحًا لدخول المسافرين مشيًا على الأقدام، متوقعًا أن تشهد حركة عبور المسافرين لهذا اليوم بطئًا كبيرًا. وأوضح المصدر في تصريح لـ"فلسطين أون لاين"، أن الجانب المصري على المعبر أعاد وضع الألواح الخشبية على بوابة معبر رفح، بحجة الترميم والصيانة.

In sentence segmentation the system's reader reads each sentence and does the following:

A. Extract Number of Words

This is the same as number of words in document but here we calculate it according to sentences. There are three cases to determine a word in the sentence as the following and displays in figure 3.3:

(i) Words in the beginning of sentence: these end with white space as word "تحققت" in the sentence " تحققت التهديدات الإسرائيلية ضد مدينة قلقيلية التي تحولت شوارعها إلى ساحات حرب، دفعت السكان إلى الرحيل عنها إلى القرى والمخيمات المجاورة "طلباً للنجاة"، ليعودوا إليها وقد دمر جزء من بيوت المدينة ونهبت بيوت أخرى من الاحتلال ".

(ii) Words in the middle of sentence: between two white spaces as "التهديدات" or it may be between white space and punctuation like "حرب" in the previous sentence.

(iii) Words in the end of sentence between white space and a full stop (.) like"الاحتلال".

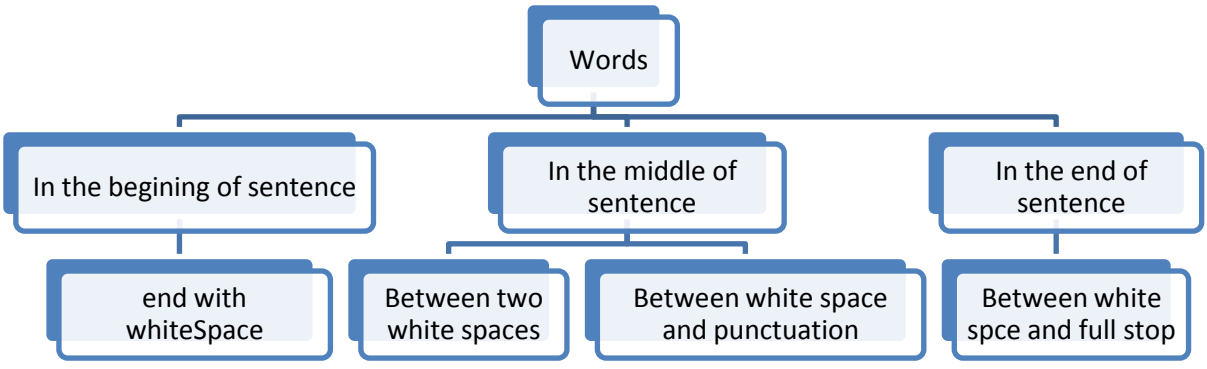

Figure 3.3 Cases of define word in the sentence

B. Sentence Relative Length

This feature is useful to filter out short sentences such as subtitles, author names and date lines commonly found in news articles [47]. The short sentence tends not to be included in summaries [3].

It is used to measure the relative length by counting the number of words in a sentence (as stated in A) and making its length relative to the longest sentence in the current document. We calculate this feature using equation (3.1) due to [47]:

$$\text{Sentence Relative length} = \frac{\text{Number of word in sentence}}{\text{Number of words in longest sentence}} \quad \ldots\ldots\ldots\text{eq (3.1)}$$

C. Sentence Type

This feature helps us to distinguish between titles and sentences. Title is considered to be very important because it affects in computing various features and it will play a main role in feature extraction as a measure of similarity between title and all sentences in the document. Title is the first sentence that appears in a document which is not ended by full stop but breaks with new line to start the article while any other sentence is ended with full stop (.).

D. Sentence Absolute Position

Position of sentence in the document plays a significant factor in finding the sentence that is the most relevant to the topic of the document. Usually the first sentence is the most significant sentence in the document which gives an idea about what the rest of text say; also the last sentence usually contains a conclusion about the text [43]. This feature has presented a good result in several projects as in [9] [15] [44]

To emphasize the significant of different sentence positions, each sentence in the document is given a rank ranging from 1 to max (in our system max is the number of sentences in the document). More weight is given to sentence at the beginning than the rest. We compute sentence absolute position feature using equation (3.2):

$$\text{Sentence Absolute Position} = \frac{\text{Numberofsentences} - \text{sentence posistion} + 1}{\text{Numberofsentences}} \quad \ldots\ldots\ldots\text{eq (3.2)}$$

As example: if we have a document with 6 sentences, then the sentence absolute position for the second sentence in this document = (6-2+1)/6 = 5/6.

3.2.3. Pre-processing

In this stage pre-processing includes filtering each sentence and removing none useful words and characters from it. This stage is summarized as below:

A. Remove Diacritics

In Arabic language there are special notations called diacritics. It used for Arabic grammar and difference in the meaning according to word position in sentence and its POS. Table 3.2 shows the diacritics which are removed from the text:

Table 3.2. Arabic diacritics

◌َ ◌ً ◌ُ ◌ٌ ◌ِ ◌ٍ ◌ّ ◌ْ ◌ٰ

B. Remove Punctuations

As any language, punctuations are used to organize the text and give the sentence a powerful meaning. The punctuation in the text summary does not have any value, so we remove all punctuations which are not full stop. Table 3.3 shows punctuations that should be removed when appear in the text:

Table 3.3. Punctuations

, ¡ : ' ÷ × º >< | \ ¿ ' ! @ # $ % ^ & *) (_ - + = ~ ø ،"

C. Remove Definite Article

In Arabic "ال" character is considered as definite article which converts the word from not definite to definite as word "the" in English. For further processing we need to remove all definite which are shown in table 3.4 to make all words in the same case:

Table 3.4. Arabic definite article

ال, وال, بال, كال, فال, ل

D. Check and Convert "Alef" style

Alef is the first letter in Arabic alphabet, it can be written in different styles as (أ, إ, آ, ا) according to its position in the sentence, diacritics and pronounce as (أحمد,إسلام, آمال, الله). For each occurrence of Alef in the text, the system converts it to (ا)style to make the entire Alef letter in the same style which helps in the semantic features.

E. Remove Strange Words

Strange words are not Arabic words but written with Arabic letters. These words must be removed from the text because they are without any value and do not have Arabic root. Table 3.5 shows some of these words that appear in Arabic text:

Table 3.5. Examples of strange words

خوجة, بلورة, تلفزة, متلفزة, معاريف, هآرتس, جراد, غراد, ديموغرافي, ديمغرافي, الأرنونا, جغرافي, الفيتو, كيبوتس ,فرانس, برس, سوروكا, افريقيا, صهيوني, كوفيه, كوفية, يونايتد, انترناشونال, بواكير, دونم, كيلو, سنتيمتر

F. Remove Stop Words

Stop words are those words which rarely contribute to useful information in terms of document relevance and appear frequently in text but provide less meaning in identifying the important content of document [47].

Those words include prepositions, conjunctions and other high frequency words. Table 3.6 shows some of these words:

Table 3.6. Examples of stop words

ان, بعد, ضد, يلي, الى, ى, في, من, حتى, هو, يكون, به, ليس, أحد, على, وكان, تلك, كذلك, تي, بين, فيها, لم, عليها, إن, على, لكن عن, مساء, لماذا, ليس, ذا, منذ, ذي, أما, حين, من, لا, ليسب, وكانت, أي, ما, عنه, حول, دون, مع, لكنه, لكن, له, هذا, تي, فقط, ثم, هذه, أنه, تكون, قد, بين, جدا, لن, نحو, كان, لهم, لأن, اليوم, لم, هؤلاء, فإن, فيه, ذلك, لو, عند, اللذين, كل, بد, لدى, وثي, أن, مع, فقد, بل ,هو, عنها, منه, بها, في, فهو, تحت, لها, أو, إذ, علي, عليه, كما, كيف, هنا, قد, كانت, لذلك, أمام, هناك, قبل, معه, يوم, منها, إلى, إذا ,هل, حيث, هي, اذا, او, ما, لا, الي, إلي, مازال, لازال, لايزال, مايزال, اصبح, أصبح, أمسى, امسى, أضحى, اضحى, ظل, مابرح, مافتئ, ماانفك, بات, صار, ليس, إن, كأن, ليت, لعل, لاسيما, لايزال, الحالي, حالي, اول, له, ذات, اي, بدلا, اليها, انه, الذين, ذين, فان, ان, الذي, هذا, لهذا, الا, فكان, ستكون, مما, , أبو, بإن, الذي, اليه, يمكن, بهذا, لدي, أن, وهي, أبو, آل, الذي, هن, الذى, ام, ابو, عبد, بروزير, سفير, أستاذ, مهندس, سفير, رئيس, م, ا, ب, ت, ث, ج, ح, خ, د, ر, ز, س, ش, ص, ض, ط, ظ, , ع, غ, ف, ق, ك, ل, ن, ه, و, ي, ى, ء

Table 3.7 shows a sentence from original text and what happen after each pre-processing stage:

Table 3.7. Examples of pre-processing

Original Sentence	يعتبر الفلسطينيونَ خطابَ بنيامين نتنياهو رئيس الوزراء الإسرائيلي أمامَ الكونجرس الأمريكي الشهرِ الماضي بأنهُ خريطة طريق نحو "يهوديةِ الدولة"، وجعل القدس عاصمةً أبديةً لها، ليتكامل خطابه مع الممارسات على أرض الواقع.
1. Remove Diacritics	يعتبر الفلسطينيون خطاب بنيامين نتنياهو رئيس الوزراء الإسرائيلي أمام الكونجرس الأمريكي الشهر الماضي بأنه خريطة طريق نحو "يهودية الدولة"، وجعل القدس عاصمة أبدية لها، ليتكامل خطابه مع الممارسات على أرض الواقع.
2. Remove Punctuation	يعتبر **الفلسطينيون** خطاب بنيامين نتنياهو رئيس **الوزراء الإسرائيلي** أمام **الكونجرس الأمريكي الشهر الماضي** بأنه خريطة طريق نحو يهودية **الدولة** وجعل **القدس** عاصمة أبدية لها **ليتكامل** خطابه مع **الممارسات** على أرض **الواقع**.
3. Remove Definite Article	يعتبر فلسطينيون خطاب بنيامين نتنياهو رئيس وزراء **إسرائيلي أمام** كونجرس **أمريكي** شهر ماضي **بأنه** خريطة طريق نحو يهودية دولة وجعل قدس عاصمة **أبدية** لها يتكامل خطابه مع ممارسات على **أرض** واقع.
4. Check and Convert Alef	يعتبر فلسطينيون خطاب بنيامين نتنياهو رئيس وزراء اسرائيلي امام **كونجرس امريكي** شهر ماضي بانه خريطة طريق نحو يهودية دولة وجعل

	قدس عاصمة ابدية لها يتكامل خطابه مع ممارسات على ارض واقع.
5. Remove Strange Words	يعتبر فلسطينيون خطاب بنيامين نتنياهو **رئيس وزراء** اسرائيلي امام شهر ماضي **بانه** خريطة طريق **نحو** يهودية دولة **وجعل** قدس عاصمة ابدية **لها** يتكامل خطابه **مع** ممارسات **على** ارض واقع.
6. Remove Stop Words	يعتبر فلسطينيون خطاب بنيامين نتنياهو اسرائيلي امام شهر ماضي خريطة طريق يهودية دولة جعل قدس عاصمة ابدية يتكامل خطابه ممارسات ارض واقع.

The final output sentence from this stage in our system is called (pre-process-sentence).

G. Check and Convert (ـة) style

This step is related to previous pre-processing but we need its output separated alone for some extracted features as entity recognition features. For each (ـة) "ta'a marbota" character appears in text we convert it to (ـه) "ha'a style" without dots. This feature helps us to deal with word "مدرسة"and "مدرسه" as the same word. The output sentence here is called (Ta'a sentence). Table 3.8 shows an example of this step applied to the pre-process-sentence obtained from the previous steps.

Table 3.8. Examples of Convert (ـة)

Pre-process-sentence	يعتبر فلسطينيون خطاب بنيامين نتنياهو اسرائيلي امام امريكي شهر ماضي **خريطة** طريق **يهودية دولة** جعل قدس **عاصمة ابدية** يتكامل خطابه ممارسات ارض واقع.
Ta'a-sentence	يعتبر فلسطينيون خطاب بنيامين نتنياهو اسرائيلي امام امريكي شهر ماضي خريطه طريق يهوديه دوله جعل قدس عاصمه ابديه يتكامل خطابه ممارسات ارض واقع.

3.2.4. Entity Recognition (ER)

The motivation for this feature is that the occurrence of proper nouns, referring to people, places and other categories, are clues that a sentence is relevant for the summary as used in [14].

For each (ta'a sentence) we check if there is any entity name occurs in it, if it occurs, we increase a special parameter called –entity counter by 1 and remove this entity name from a (pre-process-sentence). The value of this counter reflects the number of entities in the sentences. If the name counter = 5 then there is five person names in this sentence and if the place counter = 2 then we have only 2 places mentioned in this sentence. Figure 3.4 display the flow chart for this process.

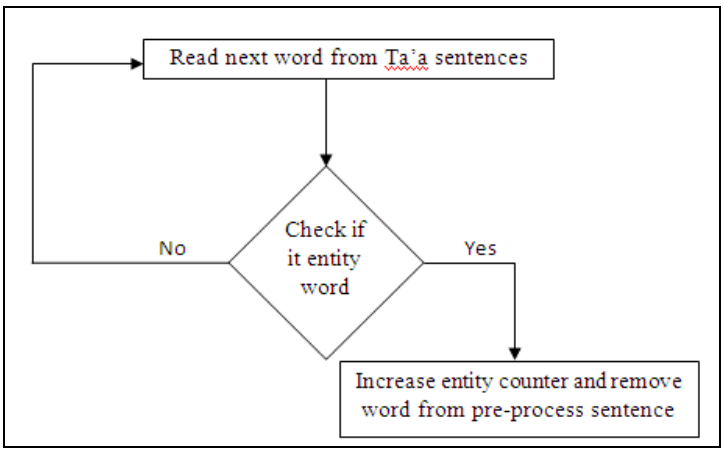

Figure 3.4 Entity Recognition flow chart

Table 3.9 shows the entity type we used in the system with an example for each one.

Table 3.9. Entity Recognition categories

Category mane	Examples	Parameter name
Name entity	اسماعيل هنيه – محمد دحلان – محمود عباس- عبد الله الشامي.	Name recognition counter
Place entity	قاهره – قدس – فلسطين – رام الله – غزة.	Place recognition counter
Months entity and days	يناير – كانون اول – تموز – سبت – ديسمبر.	Months/days recognition counter
Special entity	حركه مقاومه اسلاميه – مجلس امم متحده – جامعه دول عربيه.	Special recognition counter
Numerical entity	20/12/2011 – 123 – 10.5	Numeric recognition counter

Table 3.10 shows an example of ER on "ta'a sentence" and the final output which is called (post-pre-sentence).

Table 3.10. Entity Recognition example

Original Sentence	ويستعيد الثمانيني ذكرى الحرب التي دارت فصولها في شوارع مدينته: "في الخامس من حزيران تعرضت قلقيلية إلى عملية تدمير منظم وهجرها أهالي المدينة إلى القرى المجاورة والى مخيمات نابلس ودمرت بيوتها بنسبة تفوق ال 80 % "، ويضيف: "لقد قامت جرافات الاحتلال الإسرائيلية بتسوية بيوتها الهشة بـ"الأرض".
Pre-sentence	يستعيد ثمانيني ذكرى حرب دارت فصولها شوارع مدينته خامس حزيران

	تعرضت قلقيلية عملية تدمير منظم هجرها اهالي قرى مجاورة مخيمات نابلس دمرت بيوتها بنسبة تفوق 80 يضيف قامت جرافات احتلال اسرائيلية بتسوية بيوتها هشة ارض.
Ta'a sentence	يستعيد ثمانيني ذكرى حرب دارت فصولها شوارع مدينته خامس **حزيران** تعرضت **قلقيليه** عمليه تدمير منظم هجرها اهالي قرى مجاوره مخيمات **نابلس** دمرت بيوتها بنسبةتفوق **80** يضيف قامت جرافات **احتلال اسرائيليه** بتسويه بيوتها هشه ارض.
Post-pre-sentence	يستعيد ثمانيني ذكرى حرب دارت فصولها شوارع مدينته خامس تعرضت عملية تدمير منظم هجرها اهالي قرى مجاورة مخيمات دمرت بيوتها بنسبة تفوق يضيف قامت جرافات احتلال بتسوية بيوتها هشة ارض.

3.2.5. Stemming

Word stemming is the process of reducing inflected or derived words to their stem, base or root form. The purpose of this step is to obtain the stem or root of each word, which emphasize its semantic.

Different approaches for Arabic stemming can be identified, manually constructed dictionaries, algorithmatic light stemmer that remove prefix and suffixes, morphological analyzers that try to find the roots and form of words. Stemmer can be weak, fail to conflate related forms that should be grouped together, or strong, where unrelated forms are conflated [44][47]. In this study we perform stemming using khoja's[1] stemmer [25]. For example a stemming algorithm for Arabic should stem the words (درس, يدرس, مدرس, دارس, دراسة, مدرسة, تدريس) to its word stem (درس).

Using the stem word increases the semantic similarity between sentences and increases the value of Term Frequency through the document. This will consider all different styles of the word which have the same stem as one term and not different words. This step increases the weight of the semantic feature in the sentence.

Before we apply stemming on the text we should remove all non- letters such as numbers and dates from text. Table 3.11 shows a sentence as an example before stemming and after it. We called the output sentence from stemmer (stem-sentence).

Table 3.11. Stemming example

Original Sentence	قال وزير الخارجية المصري نبيل العربي إن مصر ملتزمة برفع الحصار الإسرائيلي عن قطاع غزة، وتقديم التسهيلات للشعب الفلسطيني
Post-pre-sentence	قال خارجية مصري ملتزمة برفع حصار اسرائيلي وتقديم تسهيلات لشعب فلسطيني.

[1]http://zeus.cs.pacificu.edu/shereen/research.htm

Stem-sentence	قول خرج صري لزم رفع حصر اسرائيلي قدم سهل شعب فلسطيني

3.2.6. Semantic Feature Extraction

All the previous Arabic summarization systems extract features from syntactic level of words [1][14][17][44]. In this proposed system we deal with each word in the sentence and with its semantic meaning. To deal with semantic meaning –Arabic synonyms- we use "Al-Taleb Arabic to Arabic dictionary"[1], which give us 100 words with its alternate meaning in Arabic also. Semantic features used in this section and in the next one are one of our main contributions for features extraction in this system. Table 3.12 shows examples from this dictionary.

It is worth noting that we use Arabic synonyms as the only semantic feature. This is true for several approaches such as Arabic WorldNet[2].

Table 3.12. Examples from Al-Taleb Arabic to Arabic dictionary

Word	First meaning	Second meaning
أزعج	أبرم	أسأم
ثمين	غال	نفيس
ذريعة	حجة	عذر
العتمة	الظلمة	الليل
هبوب	ثورة	هيجان
وعر	صعب	خطر
يسير	هين	سهل

Semantic feature extraction stage consists of:

A. Term Frequency

In the context of IR, some very important measures are Term Frequency (TF) and Term Frequency × Inverse Document Frequency (TF-IDF) [42]. In text summarization we can employ the same idea, in this case we have a single document d, and we have to select a set of relevant sentences to be included in the extractive summary out of all sentences in d. Hence, the notion of a collection of document in IR can be replaced by the notion of a single document in text summarization. This new measure will be called Term Frequency × Inverse Sentence Frequency (TF-ISF) [27].

[1] http://www.alrouqy.com/edic.html
[2] http://www.globalwordnet.org/AWN/

We use equation (3.3) due to [27] to calculate semantic TF-ISF:

$$TF_i \times ISF_i = TF_i \times \log \frac{N}{ni} \quad \ldots\ldots\ldots eq\ (3.3)$$

Where TF_i is the term frequency of word i in the document, N is the total number of sentences and n_i is the number of sentences in which word i occurs. The formula (3.3) can be formed in formula (3.4) which used by [27].

$$TF \times ISF = \frac{\sum_{i=1}^{k} Wi\,(S)}{Max\ (\sum_{i=1}^{k} Wi\,(Si^{N}))} \quad \ldots\ldots\ldots eq\ (3.4)$$

Where k is the number of words in sentence (S), W is the term frequency weight for each word i, N is the total number of sentences.

Table 3.13 shows an example of this feature:

Table 3.13. TF× ISF example

Text	<u>الاحتلال يرصد ميزانية لدعم بلدات محيط غزة</u> قررت حكومة الاحتلال الإسرائيلي, أمس، رصد ميزانية بقيمة 184 مليون شيكل، تُضم إلى ميزانية بلدة سديروت وعدة بلدات وقرى إسرائيلية مجاورة للحدود مع قطاع غزة وذلك في مسعى لحجبها عن أعين المقاومة الفلسطينية. وبحسب الموقع الإلكتروني لصحيفة "يديعوت أحرونوت" العبرية، جاء القرار خلال جلسة حكومة الاحتلال الأسبوعية، التي بحثت الأوضاع الاقتصادية في البلدات المحاذية للقطاع. وتأتي الميزانية الجديدة لتطوير السلطات الإسرائيلية في تلك البلدات، إلى جانب دعم المؤسسات المدنية. وقال رئيس الوزراء الإسرائيلي بنيامين نتنياهو تعقيبًا على الموضوع :" لا نريد تصعيدًا على حدود غزة، ومع ذلك فنحن لا نخشاه، ومن يحاول أن يجرب (إسرائيل) لا بد أن يعرف قدرتنا وقوتنا".
Sentence (23 words)	وبحسب الموقع الإلكتروني لصحيفة "يديعوت أحرونوت" العبرية، جاء القرار خلال جلسة حكومة الاحتلال الأسبوعية، التي بحثت الأوضاع الاقتصادية في البلدات المحاذية للقطاع.
Semantic TF× ISF (by computing TF for each word in sentence then apply eq. 3.4	$1\times \log\frac{4}{1} + 1\times \log\frac{4}{1} + 1 \times \log\frac{4}{1} + 1 \times \log\frac{4}{1} + 1 \times \log\frac{4}{1} + 1 \times \log\frac{4}{1} + 2 \times \log\frac{4}{2} + 3 \times \log\frac{4}{2} + 1 \times 2og\frac{4}{2} + 1\times \log\frac{4}{1} + 2 \times \log\frac{4}{2} + 1 \times \log\frac{4}{1} + 1 \times \log\frac{4}{1} + 2 \times \log\frac{4}{2} + 1 \times \log\frac{4}{1} + 3 \times \log 43 + 4 x \log 43 + 3 x \log 43 + 3 x \log 43 =$ **10.2065** TF× ISF = **10.2065 / 16.0227 = 0.637** where **16.0227** is the max TF

B. Semantic similarity with title

The title of the document used to characterize the document is without doubt of paramount importance to quantify the relevance of each sentence/phrase with respect to overall meaning conveyed by the document. Therefore, the evaluated semantic similarity of each sentence and title is explicitly taken into account [9].

This feature is obtained by using the title of the document as "a query" against all sentences of the document, and then the semantic similarity between the document title and each sentence (S) is computed by equation (3.5) due to [47]:

$$\text{Title similarity} = \frac{\text{Number of title word in S}}{\text{Number of words in title}} \quad \ldots\ldots\ldots\text{eq (3.5)}$$

Table 3.14 show how we calculate semantic similarity between title and a sentence

Table 3.14. Semantic title similarity calculation

Title (6 words)	**ويكيليكس**: (**إسرائيل**) **تستعد لحرب واسعة** النطاق
Sentence (15 words)	كشفت وثيقة سربها موقع **"ويكيليكس"** النقاب عن أن (**إسرائيل**) **تتحضر لحرب كبيرة** في الشرق الأوسط
Number of similarity word	**5**
Title similarity	**5 /6 = 0.83**

C. Semantic similarity between sentences

This feature is calculated as previous one but it measures the semantic similarity between each sentence and the previous one so we called it pre-similarity. We used equation (3.6):

$$\text{Pre similarity} = \frac{\text{Number of previous sentence word in S}}{\text{Number of words in previous sentence}} \quad \ldots\ldots\ldots\text{eq (3.6)}$$

Table 3.15 show how we calculate semantic similarity between two sentences.

Table 3.15. Semantic similarity between two sentence

First sentence (15 words)	كشفت **وثيقة** سربها موقع "ويكيليكس" النقاب عن أن (**إسرائيل**) **تتحضر لحرب كبيرة في الشرق الأوسط.**
Second Sentence (23 words)	ونقلت **الوثيقة** عن رئيس هيئة أركان جيش **الاحتلال** جابي أشكنازي خلال لقائه مع وفد من الكونغرس الأمريكي نهاية عام 2009 بأنه كان يقوم **بتجهيز حرب واسعة** النطاق **في الشرق الأوسط** والتي كانت على الأرجح أن تشن ضد حركة حماس أو حزب الله اللبناني
Number of similarity word	**8**
Semantic similarity	**8 /15 = 0.533**

3.2.7. Check key phrase

By key phrase we mean a list of key words that give a candidate sentences more important than other sentences. A group of three types of keywords are extracted from 200 Arabic news articles by a help of an Arabic experts[1]. Hence, we have a list for

[1]Arabic expert from faculty of arts in Islamic university of Gaza (IUG)

strong key words, weak words and keywords between strong and weak. Table 3.16 gives a set of examples for each category.

Table 3.16. Key phrase examples

Strong Word	Middle Word	Weak Word
أكد	أشار	سمع
نفى	أصدر	تزداد
قرر	أوصى	أمضى
زعم	فشل	تبادل

For each key phrase, we obtain its stem and then for each (stem-sentence) we check if it has any word which belongs to one of the three categories. If there is strong word then increase the strong word counter by 1, if there is a middle word then the middle word counter increased by 0.5 and for weak word we increase the weak word counter by 0.25.

3.3. Scoring

The score for each sentence (score (s)) is generated based on the linear combination of the weighted features. Equation (3.7) is used for scoring each sentence.

$$\textbf{Score (s)} = \text{sentence relative length} + \text{sentence absolute position} + \text{name recognition counter} + \text{place recognition counter} + \text{special recognition counter} + \text{months/days recognition counter} + \text{numeric recognition counter} + (\text{TF} \times \text{ISF}) + \text{title similarity} + \text{pre} - \text{similarity} + \text{strong word counter} + \text{middle word counter} + \text{weak word counter} \quad \ldots\ldots\ldots \text{eq (3.7)}$$

Sentence relative length and sentence absolute position are features from sentence segmentation phase in pre-processing and feature extraction step. Name, place, special, months/days and numeric recognition counters, all are features outputted from ER phase. (TF × ISF), title similarity and pre-similarity are semantic features output. Finally; strong, middle and weak word counters extracted from check key phrase step.

Table 3.17 shows example of scoring sentence from text:

Table 3.17. Sentence scoring example

Text	محاولة لاختطاف سفينة إسرائيلية نجحت سفينة إسرائيلية ضخمة, أمس، بالهروب من قوارب تابعة لقراصنة لم تعرف هويتهم بعد، والتي كانت تلاحقها بغرض اختطافها، خلال تواجدها في

	المحيط الهندي. وبحسب صحيفة "معاريف" العبرية، فإن قوارب القراصنة حاولت الاقتراب من السفينة الإسرائيلية التي تحمل على متنها أطنانا من مواد البناء والأجهزة الإلكترونية والكهربائية، بيد أن ربان السفينة زاد من سرعتها، حيث نجح بالهروب بعد ساعتين من المُطاردة في عرض البحر.
Sentence (40 words)	وبحسب صحيفة "معاريف" العبرية، فإن قوارب القراصنة حاولت الاقتراب من السفينة الإسرائيلية التي تحمل على متنها أطنانا من مواد البناء والأجهزة الإلكترونية والكهربائية، بيد أن ربان السفينة زاد من سرعتها، حيث نجح بالهروب بعد ساعتين من المُطاردة في عرض البحر.
Sentence relative length	**1**
Sentence absolute position	**0.5**
Name recognition counter	**1**
Place recognition counter	**0**
Month/days recognition counter	**0**
Numeric recognition counter	**0**
Special recognition counter	**0**
TF× ISF	**1**
Title semantic similarity	**1**
Pre-sentence similarity	**0.4**
Strong word counter	**2**
Middle word counter	**1**
Weak word counter	**0.25**
Total Scoring	**8.15**

3.4. Ranking and Summary Generation

All document sentences are then ranked in descending order according to their score. A set of highest score sentences are extracted as document summary based on the desired number of summary sentences given by the user as input to the system.

Finally the summary sentences are arranged in the original order to insure the readability of generated summary. Table 3.18 shows a summary generation example if user need only 5 sentence in the generated summary

Table 3.18. Summary generation example for 5 sentences

Original Text
مصادر: لقاء بين عباس ومشعل لحسم حكومة التوافق

وقعت مصادر فلسطينية، الاثنين 13-6-2011، إمكانية عقد اجتماع بين رئيس السلطة الفلسطينية محمود عباس ورئيس المكتب السياسي لحركة حماس خالد مشعل لحسم الخلاف بشأن تركيبة حكومة التوافق المقبلة .

وقالت المصادر لوكالة ا لأنباء الألمانية إنه قد تكون هناك حاجة لعقد لقاء ثنائي قريب بين عباس ومشعل في حال استمرار الخلاف الحاصل بين الحركتين على تركيبة حكومة التوافق خاصة منصب رئيس الوزراء خلال اجتماع المصالحة المقرر عقده في القاهرة غدا الثلاثاء .

ورشحت اللجنة المركزية لحركة فتح خلال اجتماع لها برئاسة عباس أول أمس السبت، رئيس حكومة تصريف الأعمال سلام فياض لرئاسة حكومة التوافق الأمر الذي رفضته حركة حماس بشدة . وكان عباس ومشعل عقدا اجتماعا هو الأول من نوعه بينهما عقب توقيع اتفاق المصالحة في القاهرة بعد قطيعة دامت أربعة أعوام .

ومن المقرر أن يعقد وفدان من الحركتين اجتماعا في القاهرة غدا الثلاثاء في مسعى للتوصل إلى اتفاق بشأن حكومة الكفاءات المستقلة التي تم الاتفاق على تشكيلها خلال توقيع الاتفاق في الرابع من الشهر الماضي .

وقال جمال محيسن عضو اللجنة المركزية لحركة فتح إن الحركة تريد رئيس وزراء يمكنه جذب الدعم الدولي و"مهمته ستكون رفع الحصار عن غزة، وليس جلب الحصار إلى الضفة الغربية أيضا ."

في المقابل قال القيادي في حركة حماس صلاح البردويل إن "الحركة لن تقبل بسلام فياض رئيسا للحكومة المقبلة، ولا حتى وزيرا فيها ."

وأعلنت فتح وحماس عقب اجتماع لهما في القاهرة منتصف الشهر الماضي أن تشكيل حكومة التوافق سيتم خلال مهلة شهر

Score	Sentence
14.004	وقعت مصادر فلسطينية، الاثنين 13-6-2011، إمكانية عقد اجتماع بين رئيس السلطة الفلسطينية محمود عباس ورئيس المكتب السياسي لحركة حماس خالد مشعل لحسم الخلاف بشأن تركيبة حكومة التوافق المقبلة
11.310	وقالت المصادر لوكالة الأنباء الألمانية إنه قد تكون هناك حاجة لعقد لقاء ثنائي قريب بين عباس ومشعل في حال استمرار الخلاف الحاصل بين الحركتين على تركيبة حكومة التوافق خاصة منصب رئيس الوزراء خلال اجتماع المصالحة المقرر عقده في القاهرة غدا الثلاثاء
12.011	ورشحت اللجنة المركزية لحركة فتح خلال اجتماع لها برئاسة عباس أول أمس السبت، رئيس حكومة تصريف الأعمال سلام فياض لرئاسة حكومة التوافق الأمر الذي رفضته حركة حماس بشدة
7.310	وكان عباس ومشعل عقدا اجتماعا هو الأول من نوع ه بينهما عقب توقيع اتفاق المصالحة في القاهرة بعد قطيعة دامت أربعة أعوام
6.992	ومن المقرر أن يعقد وفدان من الحركتين اجتماعا في القاهرة غدا الثلاثاء في مسعى للتوصل إلى اتفاق بشأن حكومة الكفاءات المستقلة التي تم الاتفاق على تشكيلها خلال توقيع الاتفاق في الرابع من الشهر الماضي
7.184	وقال جمال محيسن عضو اللجنة المركزية لحركة فتح إن الحركة تريد رئيس وزراء يمكنه جذب الدعم الدولي و "مهمته ستكون رفع الحصار عن غزة، وليس جلب الحصار إلى الضفة الغربية أيضا"
6.342	في المقابل قال القيادي في حركة حماس صلاح البردويل إن "الحركة لن تقبل بسلام فياض رئيسا للحكومة المقبلة، ولا حتى وزيرا فيها"

وأعلنت فتح وحماس عقب اجتماع لهما في القاهرة منتصف الشهر الماضي أن تشكيل حكومة التوافق سيتم خلال مهلة شهر	**7.475**

Summary

مصادر: لقاء بين عباس ومشعل لحسم حكومة التوافق

وقعت مصادر فلسطينية، الاثنين 13-6-2011، إمكانية عقد اجتماع بين رئيس السلطة الفلسطينية محمود عباس ورئيس المكتب السياسي لحركة حماس خالد مشعل لحسم الخلاف بشأن تركيبة حكومة التوافق المقبلة.

وقالت المصادر لوكالة الأنباء الألمانية إنه قد تكون هناك حاجة لعقد لقاء ثنائي قريب بين عباس ومشعل في حال استمرار الخلاف الحاصل بين الحركتين على تركيبة حكومة التوافق خاصة منصب رئيس الوزراء خلال اجتماع المصالحة المقرر عقده في القاهرة غدا الثلاثاء.

ورشحت اللجنة المركزية لحركة فتح خلال اجتماع لها برئاسة عباس أول أمس السبت، رئيس حكومة تصريف الأعمال سلام فياض لرئاسة حكومة التوافق الأمر الذي رفضته حركة حماس بشدة.

وكان عباس ومشعل عقدا اجتماعا هو الأول من نوعه بينهما عقب توقيع اتفاق المصالحة في القاهرة بعد قطيعة دامت أربعة أعوام.

وأعلنت فتح وحماس عقب اجتماع لهما في القاهرة منتصف الشهر الماضي أن تشكيل حكوم ة التوافق سيتم خلال مهلة شهر.

In this chapter we talk about the proposed model used to automatic summarize an Arabic text.

Four stages are implemented in the model which are:

1. Data acquisition: This consists of user input data like an Arabic text and system input data like entity name input.

2. Pre-processing and feature extraction: as inter-related stages and depend mainly in semantic feature extraction like term frequency and title similarity with some syntactic feature as relative length and sentence position.

3. Scoring: by using linear summation for all features and generate a final score weights for each sentence.

4. Ranking and generate summary: rank sentences in decreasing order according to their weights and finally generate summary based on CR or number of sentence entered by user.

Chapter 4

Experimental Results and Evaluation

In this chapter we shall present our experiments on AATSS. Firstly we will state the programming language and tools used to develop the proposed system. A description for the used data set to test the system and evaluate its performance is explained next. Later, Summaries generated by our proposed system, human begin, and Sakhr system will be presented and results will be evaluated by doing a comparison among the three approaches. At the end of this chapter, we shall discuss our results.

4.1. Implementation

To implement the AATSS, we used java programming language because most of pre-processing tools are found in java language beside special tools.

Table 4.1 displays part of the code with some explanation; we shall introduce other code examples in Appendix A.

Table 4.1 Stem word Code

```
// stem the word
public String stemWord ( String word )
{
   // check if the word consists of two letters
   // and find it's root
   if ( word.length ( ) == 2 )
      word = isTwoLetters ( word );

   // if the word consists of three letters
   if( word.length ( ) == 3 && !rootFound )
      // check if it's a root
      word = isThreeLetters ( word );

   // if the word consists of four letters
   if( word.length ( ) == 4 )
      // check if it's a root
      isFourLetters ( word );

   // if the root hasn't yet been found
   if( !rootFound )
   {
      // check if the word is a pattern
      word = checkPatterns ( word );
   }

   // if the root still hasn't been found
   if ( !rootFound )
```

```
        {
            // check for a definite article, and remove it
            word = checkDefiniteArticle ( word );
        }

        // if the root still hasn't been found
        if ( !rootFound && !stopwordFound )
        {
            // check for the prefix waw
            word = checkPrefixWaw ( word );
        }

        // if the root STILL hasnt' been found
        if ( !rootFound && !stopwordFound )
        {
            // check for suffixes
            word = checkForSuffixes ( word );
        }

        // if the root STILL hasn't been found
        if ( !rootFound && !stopwordFound )
        {
            // check for prefixes
            word = checkForPrefixes ( word );
        }
        return word;
    }
```

Explanation	This function is responsible for checking all possibilities of the word and then finally returns its stem if the original word has one, else it returns the same original word without any stem.

4.1.1. System Interface

Our system interface is very simple which consists of two parts. The first part contains the following:

- File browser button: to allow user to choose the Arabic.txt file which need to be summarized
- Text area : view the chosen file before summary
- Data panel: In this part, the user can enter the system variables as: names, places, special words which are used in ER stage. Also the user can enter what so called user variable which is the number of the needed sentences to be in the generated summary. If this number is not specified,

the generated number of sentences will be according to the criteria stated in section 4.3.

In the second part of the system we have a text area which views the generated summary. Figure 4.1 shows the system interface while Appendix A explains how to use it.

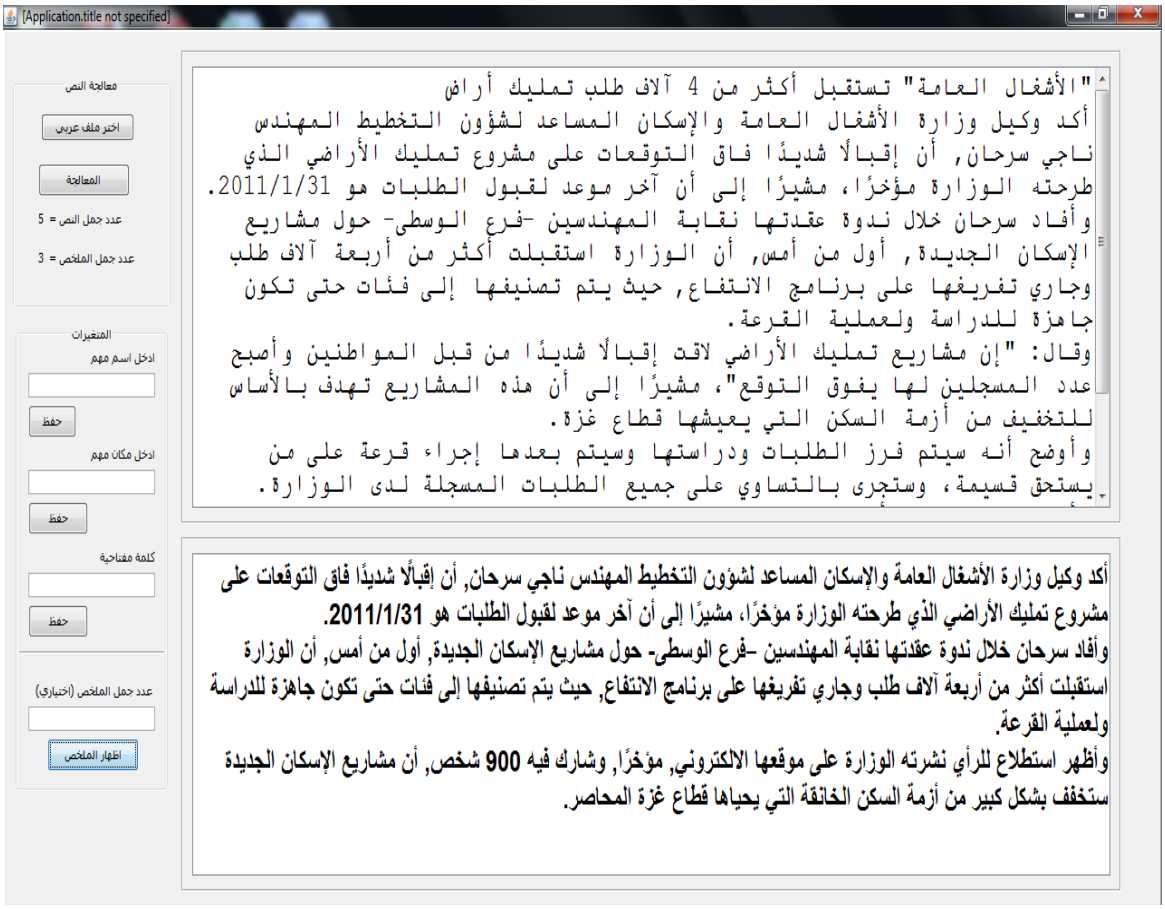

Figure 4.1 System interface

4.1.2. Tools and Programs:

Special tools and programs are used to complete the implementation of AATSS and documentation of the thesis:

- Shreen Khoja Stemmer [25]: this is a free Arabic stemmer. We use it to stem each Arabic word in the document, also it removes all strange words and non- letters from the text.

- Microsoft Access 2007: we used it to access the al-Taleb dictionary words.

- Microsoft SQL server Management Studio: we stored all the data we need to generate summary in SQL tables.

- Microsoft Excel 2007: is used to calculate the test result and compute the final P, R, F measures.

- Net Beans 6.5: this is the program which helps us to build interface and finish the system implementation using java language.

- Java Development Kit (JDK) 1.6: A software development package from Sun Microsystems that implements the basic set of tools needed to write, test and debug Java applications.

- Microsoft Word 2007: the main program used to write the documentation of the system.

4.2. Data Set

For our AATSS we collect a political news articles from "Falsteen newspaper". A randomly 200 news documents are selected between January 2011 and June 2011.

By a help of Arabic expert[1] , we extracted a list of three group keywords which represent the strong, middle and weak keywords, these keywords groups have been illustrated in chapter 3. These keyword classifications are domain independent. In other hand, two human references helped us to extract entity names from these 200 documents which contain names of persons, places, events, special political terminologies and others. These entities are domain dependent. Still these entities can be ignored in general, but very useful if used for scaling the sentence weight.

All data extracted from the documents are stored into SQL server data base to use them later in summary generation. We select only 55 political documents to test and evaluate the system.

4.3. AATSS examples

In this section we will introduce some examples of summary which we generated from the system. For number of sentence included in the summary, we have a default Compression Rate (CR) - the rate between numbers of sentences in the summary to the number of the sentences in document- for each range of sentences in the document as follow:

- If the document sentences < 5 then no summary will be generated.

[1]from Faculty of arts,IUG

- If the document sentence between 5 and 20 sentences, then we use CR = 60%

- Otherwise for all document contain more than 20 sentence then apply CR = 50%.

Table 4.2 shows an example for a text consisting of 2 sentences. This text will not be summarized according to our model.

Table 4.2. Examples of summaries generated by AATSS (no summary)

Original text
محاولة لاختطاف سفينة إسرائيلية نجحت سفينة إسرائيلية ضخمة , أمس، بالهروب من قوارب تابعة لقراصنة لم تعرف هويتهم بعد، والتي كانت تلاحقها بغرض اختطافها، خلال تواجدها في المحيط الهندي. وبحسب صحيفة "معاريف" العبرية، فإن قوارب القراصنة حاولت الا قتراب من السفينة الإسرائيلية التي تحمل على متنها أطنانا من مواد البناء والأجهزة الإلكترونية والكهربائية، بيد أن ربان السفينة زاد من سرعتها، حيث نجح بالهروب بعد ساعتين من المُطاردة في عرض البحر.
Number of sentence = 2 → number of sentence in summary = 2
Summary
محاولة لاختطاف سفينة إسرائيلية نجحت سفينة إسرائيلية ضخمة , أمس، بالهروب من قوارب تابعة لقراصنة لم تعرف هويتهم بعد، والتي كانت تلاحقها بغرض اختطافها، خلال تواجدها في المحيط الهندي. وبحسب صحيفة "معاريف" العبرية، فإن قوارب القراصنة حاولت الاقتراب من السفينة الإسرائيلية التي تحمل على متنها أطنانا من مواد البناء والأجهزة الإلكترونية والكهربائية، بيد أن ربان السفينة زاد من سرعتها، حيث نجح بالهروب بعد ساعتين من المُطاردة في عرض البحر.

Table 4.3 Have a text with 10 sentences, so according to our proposed summary size criteria. The generated summary will have only 6 sentences.

Table 4.3. Examples of summaries generated by AATSS(summary 60%)

Original text
الأوقاف" تنظم ورشة عمل لتقييم أداء الموظفين" نظمت وزارة الأوقاف والشئون الدينية , أمس، ورشة عمل حول "تقييم أداء الموظفين " بالتنسيق والتعاون مع ديوان الموظفين العام , بحضور ومشاركة وكيل الوزارة د .عبد الله أبو جربوع، ومدير عام ديوان الموظفين م.سمير مطير، ومدير عام السياسات والتخطيط بالديوان فدوان أبو شريعة، إلى جانب لفيف من مدراء ورؤساء أقسام الوزارة. وقال أبو جربوع : " إننا محظوظون بطاقم من الموظفين هم الأسبق في الخير والعطاء، ولديهم

الكثير من حيث الأداء والروح الايجابية التي استطاعوا من خلالها أن يزيدوا من عطائهم لتحقيق الأهداف التي نسعى إليها."

وأكد على ضرورة التجديد في العمل لأنه يمثل النجاح من خلال الإ بداع والبعد عن الروتينية , لافتاً إلى أن الإدارة الناجحة لابد لها من تخطيط ومتابعة وتقييم للأداء لضمان نجاحها.

بدورها، أكدت أبو شريعة على أهمية تطوير برنامج نظام تقييم الأداء في الوزارات ومؤسسات السلطة الفلسطينية, منوهة إلى أن الهدف من هذه الورشة هو اطلاع شريحة المقيمين في الوزارات على الإصدار الأول من دليل نظام تطوير تقييم الأداء الحكومي الذي تم إعداده من قبل ديوان الموظفين بغية تحقيق الموضوعية في عملية التقييم.

أوقاف خان يونس

من جهة ثانية، تولى عبد الهادي الأغا مهام مسئولية مديرية أوقاف خان يونس بعد انته اء خدمة مديرها السابق الشيخ سليمان الفرا خلال فترة حافلة بالانجازات والأنشطة المختلفة.
جاء ذلك خلال لقاء جمع كلاً من وزير الأوقاف الدكتور طالب أبو شعر ووكيل وزارة الأوقاف د . عبد الله أبو جربوع، ومدير عام الشئون الإدارية محمد أبو عسكر، وعبد الهادي الأغا وال شيخ سليمان الفرا.

بدوره، ثمن وزير الأوقاف جهود الفرا في خدمة أبنا الشعب الفلسطيني خاصة بمحافظة خان يونس , مشيداً بأدائه المميز والمنشود خلال الفترة التي قضاها.

وأعرب عن أمله في أن تتواصل وتستمر جهوده التي تنم عن مدى حرصه الكبير في إنجاح المسيرة الدعوية التي عهدناها على مدار الأعوام الماضية.

ودعا أبو شعر الأغا إلى ضرورة العمل على تقييم الوضع الحالي ورسم الخطط المستقبلية التي تضمن نجاح العمل والاستمرار به , مشدداً على أهمية وضع السياسيات العامة لاستكمال المشوار الدعوي والعمل على تطوير العمل الإداري للارتقاء بالمستوى المهني والوظيفي لدى الموظفين. من جانبه، أبدى الأغا استعداده لهذه المهمة الموكلة إليه, معرباً عن أمله في أن يكون له الدور المميز من أجل إنجاح الأعمال اليومية وتحقيق الأهداف التي تسعى إليها الوزارة.

Number of sentence = 10 → number of sentence in summary = 6

Summary

الأوقاف" تنظم ورشة عمل لتقييم أداء الموظفين"

نظمت وزارة الأوقاف والشئون الدينية , أمس، ورشة عمل حول "تقييم أداء الموظفين " بالتنسيق والتعاون مع ديوان الموظفين العام, بحضور ومشاركة وكيل الوزارة د .عبد الله أبو جربوع، ومدير عام ديوان الموظفين م.سمير مطير، ومدير عام السياسات والتخطيط بالديوان فدوان أبو شريعة، إلى جانب لفيف من مدراء ورؤساء أقسام الوزارة.

وقال أبو جربوع : " إننا محظوظون بطاقم من الموظفين هم الأسبق في الخير والعطاء، ولديهم الكثير من حيث الأداء والروح الايجابية التي استطاعوا من خلالها أن يزيدوا من عطائهم لتحقيق الأهداف التي نسعى إليها."

بدورها، أكدت أبو شريعة على أهمية تطوير برنامج نظام تقييم الأداء في الوزارات ومؤسسات السلطة الفلسطينية, منوهة إلى أن الهدف من هذه الورشة هو اطلاع شريحة المقيمين في الوزارات على الإصدار الأول من دل يل نظام تطوير تقييم الأداء الحكومي الذي تم إعداده من قبل ديوان الموظفين بغية تحقيق الموضوعية في عملية التقييم.

من جهة ثانية، تولى عبد الهادي الأغا مهام مسئولية مديرية أوقاف خان يونس بعد انتهاء خدمة مديرها السابق الشيخ سليمان الفرا خلال فترة حافلة بالانجازات والأنشطة المختلفة.

بدوره، ثمن وزير الأوقاف جهود الفرا في خدمة أبنا الشعب الفلسطيني خاصة بمحافظة خان يونس , مشيداً بأدائه المميز والمنشود خلال الفترة التي قضاها.

ودعا أبو شعر الأغا إلى ضرورة العمل على تقييم الوضع الحالي ورسم الخطط المستقبلية التي تضمن نجاح العمل والاستمرار به , مشدداً على أهمية وضع السياسيات العامة لاستكمال المشوار الدعوي والعمل على تطوير العمل الإداري للارتقاء بالمستوى المهني والوظيفي لدى الموظفين.

In table 4.4 we have a text with 22 sentences and the generated summary consists of 11 sentences which is equal to 50%.

Table 4.4. Examples of summaries generated by AATSS (50%)

Original text
37مستوطنة تخنق مدينة القدس وتتهدد الأقصى أكد رئيس قسم المخطوطات في المسجد الأقصى د . ناجح بكيرات، أن سلطات الاحتلال الإسرائيلية صادرت حوالي (750) ألف دونم، وأقامت (37) مستوطنة على أراضي الفلسطينيين في القدس المحتلة، ويستوطن فيها منذ العام 1967 (700) ألف يهودي . وقال بكيرات في حوار خاص مع "فلسطين": "لم يتبق للمقدسيين في المدينة سوى 10% من مساحة أراضيهم الإجمالية، بعدما سلبت حوالي (22) مليون دونم (22 ألف كيلو متر مربع) من مساحة فلسطين التاريخية، ولم يتبق للفلسطينيين منها سوى نحو (4-5) مليون دونم . وذكر أن مخططات التهويد في المسجد الأقصى تسارعت خطاها مع سقوط مدينة القدس في قبضة الاحتلال في تلك الفترة، مشيراً إلى أن وتيرة الحفريات بدأت في تلك الفترة بشكل ب سيط ثم أخذت تتطور وتتوسع من خلال مصادرة أراضي الوقف والمقدسيين والسطو على ممتلكاتهم . مشاريع الاستيطان وتابع رئيس قسم المخطوطات بالمسجد الأقصى : "ثم بعد التوسع وضعت خطة إسرائيلية مرسومة لتهويد المسجد الأقصى بالتعاون مع وزارة الآثار ورصدت أماكن بناء المستوط نات بمحيط البلدة القديمة"، مبيناً أن الاستيطان بدأ يتسع علناً بعد يومين من نكسة 1967 في المسجد الأقصى ، وبالقرب من باب المغاربة ، وتم على اثرها هدم حارة المغاربة وإقامة الحي اليهودي مكانه . وأردف بكيرات : " أخذ بعدها طابع الاستيطان يختلف لا سيما بعد إنشاء الأحياء اليهودية على الجبال المحيطة بالقدس مثل جبال خان الأحمر " وتم إقامة مستوطنة معاليه أدوميم هناك، وجبال النبي صموئيل، والتي لا تبعد عن المسجد الأقصى كثيراً . وأشار إلى أن تلك العملية أخذت بالتقدم تجاه المسجد الأقصى، "لتتوقف قليلاً عند التلة الفرنسية، وحي الشيخ جراح لبناء البئر الاستيطانية والوصول للهدف الرئيسي لهم ." وأوضح بكيرات أن الخطوات التي اتبعتها سلطات الاحتلال الإسرائيلية، بدءاً من العام 1970 أدت إلى محاصرة المدينة المقدسة والاستيلاء على أكثر من 450 ألف دونم لصالح مشاريع الاستيطان . وأكد رئيس قسم المخطوطات ، أن الانتهاكات بحق المسجد الأقصى والمدينة المقدسة بدأت تتصاعد لا سيما بعد تولي أرئيل شارون رئاسة الحكومة الإسرائيلية ، وأخذت الأحزاب المعارضة تتسابق في بناء الوحدات السكنية لتبرهن للإسرائيليين أنهم الأجدر لتحقيق أمن(إسرائيل .(

وبين أنه تم بناء ما يقارب من 75 ألف وحدة سكنية عام 2000 ، وبعدها بني ما يقارب من 85 ألف وحدة سكنية، مشيراً إلى أن طابع الاستيطان بدأ يأخذ اتجاهاً مغايراً عن ذي قبل، وبدأت ملاحقة الأحياء العربية، وسلب بيوت المقدسيين في حي الشيخ جراح وسلوان وجبل المكبر إلى أن توقف داخل المسجد الأقصى .

واعتبر بكيرات ، أن اتفاقية أوسلو عززت من الاستيطان في المدينة المقدسة، وأعطت له الضوء الأخضر لزيادة الممارسات الإسرائيلية بحقها، وجعلتنا نطالب بتجميد الاستيطان وإيقاف البناء في الضفة الغربية، وذلك بهدف لفت الأنظار عن المسجد الأقصى ."

وأشار إلى أن سلطات الاحتلال لم تتوقف عن مسلسلها التهويدي في المدينة المقدسة حيث حفر الأنفاق وهدمت البيوت، مضيفاً : " إن التدخلات العربية الرامية إلى وقف التهويد التي تمارسه (إسرائيل) في المسجد الأقصى هو موقف استنكار وشجب ولا يرتقي للمستوى المطلوب بأي شكل من الأشكال ."

واتهم رئيس قسم المخطوطات، العرب بالمسئولية على كل ما يحاك ضد المسجد الأقصى من خلال سكوتهم عما يحدث بشكل يومي ضد القدس المحتلة، وعدم إجبار الاحتلال على التوقف عن جرائمه، والدليل على ذلك إقدام المتطرفين الصهاينة على إحراق المسجد الأقصى في 21 أغسطس عام 1969، وارتكاب المجازر بحق المقدسيين فاكتفت الدول العربية بموقف الشجب والاستنكار ."

ومضى يقول: "إن فتح سفارات للاحتلال الإسرائيلي في مختلف الدول العربية أدى للمضي قدماً في مسلسل التهويد ."

وأكد بكيرات أن الاستيطان في المدينة المقدسة غير الجغرافيا والمعالم المقدسة وفرض واقعًا سكانيًا جديدًا وغير أيضاً الديمغرافية السكانية في المدينة، وفرض ثقافة يهودية جديدة، مضيفاً : "لم يعد الاستيطان بناءً على قدر مما هو ثقافة ومشروع صهيوني خطير خصصت له الدولة العبرية كل السبل والاحتياجات ."

وأشار إلى أن سلطات الاحتلال الإسرائيلية تريد تهويد القدس بشكل كامل وشطب كل المعالم العربية منها، من خلال فرض المشاريع والميزانيات لتطبيع المقدسيين بالطابع اليهودي من خلال الاقتحامات المتكررة الرامية لتغيير الواقع، وقلب المعادلة لصالحهم .

تزوير الحقائق

وبين بكيرات أن الاحتلال الإسرائيل ي عمل جاهداً منذ بداية احتلال المدينة المقدسة على تزوير الحقائق التاريخية وطمس معالمها الأثرية بالكامل للتأكيد على أن مدينة القدس يهودية .

وأضاف:" إن الاحتلال يحاول بكل السبل والوسائل تهويد المدينة المقدسة وإقامة الهيكل المزعوم على أنقاضها، فالمقدسات تنته ك والمساجد تحرق ويتم الاعتداء على الكنائس دون تدخل دولي لوقفهم ."

وبين وجود أكثر من 20 قراراً لترحيل السكان من بيوتهم وإجبار المقدسيين على ترك المدينة كإبعاد النواب المقدسيين كالنائب "أحمد عطون، ومحمد طوطح، ومحمد أبوطير، والوزير خالد أبوعرفة ."

وأشار بكيرات الى أن ذلك يهدف إلى "التضييق على المقدسيين والتفرد في المدينة المقدسة والمسجد الأقصى وإقامة كل ما يحلو لهم من هياكل ومساكن وغيرها دون عرضة من أحد ."

وشدد على ضرورة وضع ملف دولي وعربي مشترك لمواجهة الانتهاكات الإسرائيلية بحق المسجد الأقصى، بالإضافة إلى وضع خطة زمنية محددة لا تزيد عن خمس سنوات لإيقاف تلك الاعتداءات .

وأكد أن القدس "بحاجة إلى قوة فاعلة توقف جميع الاعتداءات التي تحاك ضد المسجد الأقصى، وتفعيل منظمة المؤتمر الإسلامي وجامعة الدول العربية، والشارع العربي، وجعلها قضية على سلم أولوياتهم والدفاع عنها ."

ودعا رئيس قسم المخطوطات إلى استغلال المصالحة الفلسطينية والعمل للفت أنظار العالم الدولي والعربي بالانتهاكات التي ترتكبها قوات الاحتلال الإسرائيلية بحق المسجد الأقصى والمقدسيين، وتفعيل قضية مقاومة المحتل والتوقف عن المفاوضات.

Number of sentence = 22 → number of sentence in summary = 11

Summary

37مستوطنة تخنق مدينة القدس وتتهدد الأقصى

أكد رئيس قسم المخطوطات في المسجد الأقصى د .ناجح بكيرات، أن سلطات الاحتلال الإسرائيلية صادرت حوالي (750) ألف دونم، وأقامت (37) مستوطنة على أراضي الفلسطينيين في القدس المحتلة، ويستوطن فيها منذ العام 1967 (700) ألف يهودي.

وتابع رئيس قسم المخطوطات بالمسجد الأقصى : "ثم بعد التوسع وضعت خطة إسرائيلية مرسومة لتهويد المسجد الأقصى بالتعاون مع وزارة الآثار ورصدت أماكن بناء المستوطنات بمحيط البلدة القديمة"، مبيناً أن الاستيطان ب دأ يتسع علناً بعد يومين من نكسة 1967 في المسجد الأقصى ، وبالقرب من باب المغاربة ، وتم على اثرها هدم حارة المغاربة وإقامة الحي اليهودي مكانه.

وأردف بكيرات : " أخذ بعدها طابع الاستيطان يختلف لا سيما بعد إنشاء الأحياء اليهودية على الجبال المحيطة بالقدس مثل جبال خان الأحمر " وتم إقامة مستوطنة معاليه أدوميم هناك، وجبال النبي صموئيل، والتي لا تبعد عن المسجد الأقصى كثيراً.

وأكد رئيس قسم المخطوطات ، أن الانتهاكات بحق المسجد الأقصى والمدينة المقدسة بدأت تتصاعد لا سيما بعد تولي أرئيل شارون رئاسة الحكومة الإسرائيلية ، وأخذت الأحزاب المعارضة تتسابق في بناء الوحدات السكنية لتبرهن للإسرائيليين أنهم الأجدر لتحقيق أمن(إسرائيل.(

وبين أنه تم بناء ما يقارب من 75 ألف وحدة سكنية عام 2000 ، وبعدها بني ما يقارب من 85 ألف وحدة سكنية، مشيراً إلى أن طابع الاستيطان بدأ يأخذ اتجاهاً م غايراً عن ذي قبل، وبدأت ملاحقة الأحياء العربية، وسلب بيوت المقدسيين في حي الشيخ جراح وسلوان وجبل المكبر إلى أن توقف داخل المسجد الأقصى.

واعتبر بكيرات ، أن اتفاقية أوسلو عززت من الاستيطان في المدينة المقدسة، وأعطت له الضوء الأخضر لزيادة الممارسات الإسرائيلية بحقها، وجعلتنا نطالب بتجميد الاستيطان وإيقاف البناء في الضفة الغربية، وذلك بهدف لفت الأنظار عن المسجد الأقصى."

وأشار إلى أن سلطات الاحتلال لم تتوقف عن مسلسلها التهويدي في المدينة المقدسة حيث حفر الأنفاق وهدمت البيوت، مضيفاً : " إن التدخلات العربية الرامي ة إلى وقف التهويد التي تمارسه (إسرائيل) في المسجد الأقصى هو موقف استنكار وشجب ولا يرتقي للمستوى المطلوب بأي شكل من الأشكال."

واتهم رئيس قسم المخطوطات، العرب بالمسئولية على كل ما يحاك ضد المسجد الأقصى من خلال سكوتهم عما يحدث بشكل يومي ضد القدس المحتلة، وعدم إجبار الاحتلال على التوقف عن جرائمه، والدليل على ذلك إقدام المتطرفين الصهاينة على إحراق المسجد الأقصى في 21 أغسطس عام 1969، وارتكاب المجازر بحق المقدسيين فاكتفت الدول العربية بموقف الشجب والاستنكار."

أكد بكيرات أن الاستيطان في المدينة المقدسة غير الجغرا فيا والمعالم المقدسة وفرض واقعًا سكانيًا

جديدًا وغير أيضاً الديمغرافية السكانية في المدينة، وفرض ثقافة يهودية جديدة، مضيفاً : "لم يعد الاستيطان بناءً على قدر مما هو ثقافة ومشروع صهيوني خطير خصصت له الدولة العبرية كل السبل والاحتياجات."

وأشار إلى أن سلطات الاحتلال الإسرائيلية تريد تهويد القدس بشكل كامل وشطب كل المعالم العربية منها، من خلال فرض المشاريع والميزانيات لتطبيع المقدسيين بالطابع اليهودي من خلال الاقتحامات المتكررة الرامية لتغيير الواقع، وقلب المعادلة لصالحهم.

وأكد أن القدس "بحاجة إلى قوة فاعلة توقف جميع الاعتداءات التي تحاك ضد المسجد الأقصى، وتفعيل منظمة المؤتمر الإسلامي وجامعة الدول العربية، والشارع العربي، وجعلها قضية على سلم أولوياتهم والدفاع عنها".

4.4. System Evaluation

System evaluation is a hard task as mentioned in section 1.6 because it is very difficult to find an ideal text summary for a given document or a set of documents.

Classification approach for generating automatic summaries makes it easier for evaluating extractive summaries. Three important measures are commonly used, precision, recall and F-measure [20] [46]. Precision is a measure of how much of the returned information by the system is correct.

Precision = Number of system correct summary sentences / Number of system summary sentences … eq. (4.1)

Recall is a measure of the coverage of the system.

Recall = Number of system correct summary sentences / Number of human summary sentences … eq. (4.2)

Usually Recall and Precision are antagonistic to one another. A system strives for coverage will get lower precision and a system strives for precision will get lower recall. F-measure balances recall and precision using a parameter β. The F-measure is defined in equation (4.3) as follows:

$$F = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \quad \text{.....eq. (4.3)}$$

When β is one, Precision P and Recall R are given equal weight. When β is greater than one, Precision is favored, when β is less than one, recall is favored. In the following experiments β equals one.

4.4.1. AATSS and Human Evaluation

To evaluate our AATSS we used two human references to summarize 55 political news documents. For each reference and document we compute the three measurements P, R and F. Table 4.5 to Table 4.7 show some examples of this evaluation. In each of the Tables, a document has been summarized by AATSS system, Human reference-1 and human reference-2, then P, R and F measures are computed.

For each document the average of F measure has been computed. The average of all averages of F measure for all the chosen documents is considered the system performance. The more average of F measure (for all the documents) is approaching to 1; the better is the system's performance.

Table 4.5 show an evaluation for Arabic text by comparing AATSS with two human summaries and for each human we compute P,R and f- measures. In final we compute the average f-measure between the two summaries.

Table 4.5. Summary evaluation 1(the underlined sentences in the summary of AATSS are existing in human summary)

Original text
العبادسة: "حرب غزة" والممارسات الإسرائيلية أثبتتا همجية الاحتلال أكد النائب عن كتلة التغيير والإصلاح البرلمانية الدكتور يحيى العبادسة، أن الحرب الإسرائيلية على قطاع غزة، والممارسات الإسرائيلية العدوانية والعنصرية أثبتتا همجية الاحتلال وعدم تورعه عن استهداف كافة شرائح المجتمع الفلسطيني دون أي اكتراث. وأوضح العبادسة، خلال اعتصام نظمه "مجمع ناصر الطبي" بمدينة خان يونس، إحياءً لذكرى مرور عامين على "حرب غزة"، أمس، أن الشعب الفلسطيني وقواه الحية "مصممون على ملاحقة الاحتلال في كافة المحافل وصولا إلى تقرير المصير وتحرير الأرض الفلسطينية." مأزق حقيقي وقال: " إن (إسرائيل) تعيش في مأزق حقيقي، بعد الحرب التي شنتها على قطاع غزة، والتي أثبتت عدوانيتها وهمجيتها"، مبينًا أن كافة الأصوات الداعية إلى إشعال فتيل حرب جديدة على القطاع مهددة بالملاحقة في العديد من دول العالم. وثمن القيادي في حركة المقاومة الإسلامية "حماس" جهود الشرائح المختلفة في فضح ممارسات الاحتلال في دولة العالم . وأضاف: "إن نموذج قطاع غزة الذي يرفع راية المقاومة أنتج صمودًا وتحديًا أمام قوة وهمجية الاحتلال، يحتذى به أمام دول العالم وشعوبها"، منتقدًا المشاريع القي تصب في مصلحة الاحتلال وتنفيذ أهدافه من خلال التسوية السلمية وملاحقة المقاومة. وشارك في الاعتصام الذي عرض خلاله العديد من القذائف الإسرائيلية التي استهدفت منازل المواطنين والمنشآت المدنية، نواب المجلس التشريعي الفلسطيني وممثلون عن الفصائل الفلسطينية وقيا دة الشرطة والعديد من الأطباء والممرضين العاملين في المجمع. بدوره؛ أثنى رئيس بلدية خان يونس محمد الفرا، على جهود الطواقم الطبية العاملة في القطاع الصحي بالمدينة، على دورها وجهدها أثناء الحرب الإسرائيلية على قطاع غزة، في إنقاذ المصابين، وانتشال جثث الشهداء م ن أنقاض المباني المدمرة. وقال الفرا: "نقف اليوم على أعتاب ذكرى حرب إسرائيلية نشتم منها رائحة الغدر والتي سقط على إثرها أكثر من 1400 شهيد، وآلاف الجرحى في اثنين وعشرين يوماً "، مبينًا أن أعداد الشهداء وطريقة استشهادهم تفند كافة الادعاءات الإسرائيلية بحماية المدنيين . ادعاء الديمقراطية وأضاف: "لقد سقط القناع عن وجه (إسرائيل) في ادعائها للديمقراطية وحماية حقوق الإنسان، من خلال استهداف

الأبرياء، وعدم الاكتراث لحياتهم خلال الحرب."

من جهته؛ أكد مدير مجمع ناصر الطبي، د . يوسف أبو الريش، استمرار كافة الهيئات الطبية ا لعاملة على مواصلة مسيرتها وعطائها في سبيل خدمة المواطنين الفلسطينيين، في ظل الظروف والأوضاع المختلفة التي قد تحدث.

وأوضح أبو الريش أن الطواقم الطبية العاملة في المجمع عملت بجهد وتفاني في إنقاذ الشهداء، وعلاج الجرحى المصابين خلال الحرب والذين وصلوا المستشفى من كافة المناطق والأحياء، مثمناً دور المؤسسات الأخرى التي تعاونت مع المستشفى.

وشكر الطواقم الطبية في المجمع التي واصلت الليل بالنهار من أجل خدمة شعبهم الفلسطيني، موجهًا رسالة إلى العالم والمؤسسات الحقوقية بضرورة التحرك لوقف جرائم دولة الاحتلال الإسرائيلي بحق الشعب الفلسطيني.

Human reference 1 summary	System summary
العبادسة: "حرب غزة" والممارسات الإسرائيلية أثبتتا همجية الاحتلال	العبادسة: "حرب غزة" والممارسات الإسرائيلية أثبتتا همجية الاحتلال
1.أكد النائب عن كتلة التغيير والإصلاح البرلمانية الدكتور يحيى العبادسة، أن ا لحرب الإسرائيلية على قطاع غزة، والممارسات الإسرائيلية العدوانية والعنصرية أثبتتا همجية الاحتلال وعدم تورعه عن استهداف كافة شرائح المجتمع الفلسطيني دون أي اكتراث.	1.أكد النائب عن كتلة التغيير والإصلاح البرلمانية الدكتور يحيى العبادسة، أن الحرب الإسرائيلية على قطاع غزة، والممارسات الإسرائيلية العدوانية والعنصرية أثبتتا همجية الاحتلال وعدم تورعه عن استهداف كافة شرائح المجتمع الفلسطيني دون أي اكتراث.
2.وأوضح العبادسة، خلال اعتصام نظمه "مجمع ناصر الطبي" بمدينة خان يونس، إحياءً لذكرى مرور عامين عل ى "حرب غزة "، أمس، أن الشعب الفلسطيني وقواه الحية "مصممون على ملاحقة الاحتلال في كافة المحافل وصولا إلى تقرير المصير وتحرير الأرض الفلسطينية."	2.وقال: " إن (إسرائيل) تعيش في مأزق حقيقي، بعد الحرب التي شنتها على قطاع غزة، والتي أثبتت عدوانيتها وهمجيتها "، مبينًا أن كافة الأصوات الداعية إلى إشعا ل فتيل حرب جديدة على القطاع مهددة بالملاحقة في العديد من دول العالم.
3.وقال: " إن (إسرائيل) تعيش في مأزق حقيقي، بعد الحرب التي شنتها على قطاع غزة، والتي أثبتت عدوانيتها وهمجيتها"، مبينًا أن كافة الأصوات الداعية إلى إشعال فتيل حرب جديدة على القطاع مهددة بالملاحقة في العديد من دول العالم . وثمن القيادي في حركة المقاومة الإسلامية "حماس" جهود الشرائح المختلفة في فضح ممارسات الاحتلال في دولة العالم.	3.وأضاف: " إن نموذج قطاع غزة الذي يرفع راية المقاومة أنتج صمودًا وتحديًا أمام قوة وهمجية الاحتلال، يحتذى به أمام دول العالم وشعوبها "، منتقدًا المشاريع التي تصب في مصلحة الاحتلال وتنفيذ أهدافه من خلال التسوية السلمية وملاحقة المقاومة.
4.وشارك في الاعتصام الذي عرض خلاله العديد من القذائف الإسرائيلية التي استهدفت منازل المواطنين والمنشآت المدنية، نواب المجلس التشريعي الفلسطيني وممثلون عن الفصائل الفلسطينية وقيادة الشرطة والعديد من الأطباء والممرضين العاملين في المجمع.	4.وشارك في الاعتصام الذي عرض خلاله العديد من القذائف الإسرائيلية التي استهدفت منازل المواطنين والمنشآت المدنية، نواب المجلس التشريعي الفلسطيني وممثلون عن الفصائل الفلسطينية وقيادة الشرطة والعديد من الأطباء والممرضي ن العاملين في المجمع.
5.بدوره؛ أثنى رئيس بلدية خان يونس محمد الفرا، على جهود الطواقم الطبي ة العاملة في القطاع الصحي بالمدينة، على دورها وجهدها أثناء الحرب	5.بدوره؛ أثنى رئيس بلدية خان يونس محمد الفرا، على جهود الطواقم الطبية العاملة في القطاع الصحي بالمدينة، على دورها وجهدها أثناء الحرب الإسرائيلية على قطاع غزة، في إنقاذ المصابين،

الإسرائيلية على قطاع غزة، في إنقاذ المصابين، وانتشال جثث الشهداء من أنقاض المباني المدمرة. 6.وأضاف: " لقد سقط القناع عن وجه (إسرائيل) في ادعائها للديمقراطية وحماية حقوق الإنسان، من خلال استهداف الأبرياء، وعدم الاكتراث لحياتهم خلال الحرب".	وانتشال جثث الشهداء من أنقاض المباني المدمرة. 6.وقال الفرا : "نقف اليوم على أعتاب ذكرى حرب إسرائيلية نشتم منها رائحة الغدر والتي سقط على إثرها أكثر من 1400 شهيد، وآلاف الجرحى في اثنين وعشرين يوماً "، مبينًا أن أعداد الشهداء وطريقة استشهادهم تفند كافة الادعاءات الإسرائيلية بحماية المدنيين. 7.وأضاف: "لقد سقط القناع عن وج ه (إسرائيل) في ادعائها للديمقراطية وحماية حقوق الإنسان، من خلال استهداف الأبرياء، وعدم الاكتراث لحياتهم خلال الحرب".

F	R	P
[(1+1)*(5/7)*(5/6)]/[(5/7)+(5/6)] = 0.769	5/6	5/7

Human reference 2 summary	System summary
العبادسة: "حرب غزة" والممارسات الإسرائيلية أثبتتا همجية الاحتلال 1.أكد النائب عن كتلة التغيير والإصلاح البرلمانية الدكتور يحيى العبادسة، أن الحرب الإسرائيلية على قطاع غزة، والممارسات الإسرائيلية العدوانية والعنصرية أثبتتا همجية الاحتلال وعدم تورعه عن استهداف كافة شرائح المجتمع الفلسطيني دون أي اكتراث. 2.وقال: " إن (إسرائيل) تعيش في مأزق حقيقي، بعد الحرب التي شنتها على قطاع غزة، والتي أثبتت عدوانيتها وهمجيتها"، مبينًا أن كافة الأصوات الداعية إلى إشعال فتيل حرب جديدة على القطاع مهددة بالملاحقة في العديد من دول العالم . وثمن القيادي في حركة المقاومة الإسلامية "حماس" جهود الشرائح المختلفة في فضح ممارسات الاحتلال في دولة العالم. 3.وأضاف: " إن نموذج قطاع غزة الذي يرفع راية المقاومة أنتج صمودًا وتحديًا أمام قوة وهمجية الاحتلال، يحتذى به أمام دول العالم وشعوبها"، منتقدًا المشاريع التي تصب في مصلحة الاحتلال وتنفيذ أهدافه من خلال التسوية السلمية وملاحقة المقاومة. 4.وشارك في الاعتصام الذي عرض خلاله العديد من القذائف الإسرائيلية التي استهدفت منازل المواطنين والمنشآت المدنية، نواب المجلس التشريعي	العبادسة: "حرب غزة" والممارسات الإسرائيلية أثبتتا همجية الاحتلال 1.أكد النائب عن كتلة التغيير والإصلاح البرلما نية الدكتور يحيى العبادسة، أن الحرب الإسرائيلية على قطاع غزة، والممارسات الإسرائيلية العدوانية والعنصرية أثبتتا همجية الاحتلال وعدم تورعه عن استهداف كافة شرائح المجتمع الفلسطيني دون أي اكتراث. 2.وقال: " إن (إسرائيل) تعيش في مأزق حقيقي، بعد الحرب التي شنتها عل ى قطاع غزة، والتي أثبتت عدوانيتها وهمجيتها "، مبينًا أن كافة الأصوات الداعية إلى إشعال فتيل حرب جديدة على القطاع مهددة بالملاحقة في العديد من دول العالم. 3.وأضاف: " إن نموذج قطاع غزة الذي يرفع راية المقاومة أنتج صمودًا وتحديًا أمام قوة وهمجية الاحتلال، يحتذى بـه أمام دول العالم وشعوبها "، منتقدًا المشاريع التي تصب في مصلحة الاحتلال وتنفيذ أهدافه من خلال التسوية السلمية وملاحقة المقاومة. 4.وشارك في الاعتصام الذي عرض خلاله العديد من القذائف الإسرائيلية التي استهدفت منازل المواطنين والمنشآت المدنية، نواب المجلس التشري عي الفلسطيني وممثلون عن الفصائل الفلسطينية وقيادة

الفلسطيني وممثلون عن الفصائل الفلسطينية وقيادة الشرطة والعديد من الأطباء والممرضين العاملين في المجمع. 5.بدوره؛ أثنى رئيس بلدية خان يونس محمد الفرا، على جهود الطواقم الطبية العاملة في القطاع الصحي بالمدينة، على دورها وجهدها أثناء الحرب الإسرائيلية على قطاع غزة، في إنقاذ المصابين، وانتشال جثث الشهداء من أنقاض المباني المدمرة. 6.وقال الفرا : " نقف اليوم على أعتاب ذكرى حرب إسرائيلية نشتم منها رائحة الغدر والتي سقط على إثرها أكثر من 1400 شهيد، وآلاف الجرحى في اثنين وعشرين يوماً"، مبينًا أن أعداد الشهداء وطريقة استشهادهم تفند كافة الادعاءات الإسرائيلية بحماية المدنيين. 7.وأضاف: "لقد سقط القناع عن وجه (إسرائيل) في ادعائها للديمقراطية وحماية حقوق الإنسان، من خلال استهداف الأبرياء، وعدم الاكتراث لحياتهم خلال الحرب". 8.وشكر الطواقم الطبية في المجمع التي واصلت الليل بالنهار من أجل خدمة شعبهم الفلسطيني، موجهًا رسالة إلى العالم والمؤسسات الحقوقية بضرورة التحرك لوقف جرائم دولة الاحتلال الإسرائيلي بحق الشعب الفلسطيني	الشرطة والعديد من الأطباء والممرضين العاملين في المجمع. 5.بدوره؛ أثنى رئيس بلدية خان يونس محمد الفرا، على جهود الطواقم الطبية العاملة في القطاع الصحي بالمدينة، على دورها وجهدها أثناء الحرب الإسرائيلية على قطا ع غزة، في إنقاذ المصابين، وانتشال جثث الشهداء من أنقاض المباني المدمرة. 6.وقال الفرا : " نقف اليوم على أعتاب ذكرى حرب إسرائيلية نشتم منها رائحة الغدر والتي سقط على إثرها أكثر من 1400 شهيد، وآلاف الجرحى في اثنين وعشرين يوماً "، مبينًا أن أعداد الشهداء وطريقة استشهادهم تفند كافة الادعاءات الإسرائيلية بحماية المدنيين. 7.وأضاف: "لقد سقط القناع عن وجه (إسرائيل) في ادعائها للديمقراطية وحماية حقوق الإنسان، من خلال استهداف الأبرياء، وعدم الاكتراث لحياتهم خلال الحرب".

P	R	F
1	0.875	0.933
Average F		
0.851		

Table 4.6 show an evaluation for Arabic text by comparing AATSS with two human summaries and for each human we compute P,R and f- measures. In final we compute the average f-measure between the two summaries.

Table 4.6. Summary evaluation 2 (the underlined sentences in the summary of AATSS are existing in human summary)

Original text
متابعة أردنية لاتهام (إسرائيل) لتاجر أردني بالاتصال بمنظمة "إرهابية" قالت وزارة الخارجية الأردنية إنها تتابع باهتمام مع الجانب الإسرائيلي، تداعيات اعتقال السلطات للمواطن الأردني أمجد عبد المجيد سدر (46 عاما) بعد توجيه اتهامات له تنطوي على اتصاله بجهاز وجسم خارجي, وتقديم خدمات

لمنظمة إرهابية واستعماله مواد غير قانونية لدعم الإرهاب، بحسب محاميه.

وأفاد مسئول في الخارجية، فضَّل عدم الكشف عن اسمه أن الخارجية ومنذ ورود معلومات إليها عما قامت به السلطات الإسرائيلي، بحق المواطن سدر "تحركنا لاستجلاء الموقف، وأجرينا اتصالات مع الخارجية الإسرائيلية عبر سفارتنا في تل أبيب".

وكانت السلطات الأمنية الإسرائيلية وجهت لسدر الذي يعمل تاجراً ، بعد أن اعتقلته الجمعة الماضية (12/31) في جسر الشيخ حسين أثناء عودته من زيارة إلى القدس، ثلاث تهم تشمل "الاتصال بجهاز وجسم خارجي، وتقديم خدمات لمنظمة إرهابية، واستعمال مواد غير قانونية لدعم الإرهاب"، وفق محاميه محمد عابدين.

وقال عابدين في تصريحات لصحيفة "الغد" الأردنية إن موكله سدر، الذي يقبع الآن في سجن "بيتح تكفا" سيئ الصيت، نفى كليا صحة التهم الثلاث.

وبين عابدين أن موكله سدر أكد خلال التحقيق، الذي أجرته السلطات الأمنية الإسرائيلية معه، براءته كليا من هذه التهم، نافيا جملة وتفصيلا صحتها، ومبدياً في الوقت نفسه استغرابه من أن توجه له مثل هذه التهم، التي قال إنه يسمع عنها لأول مرة "وهو التاجر المنشغل بتجارته وعمله".

وأوضح المحامي عابدين أن هذه التهم "غير قانونية"، واصفا إياها بـ "التعسفية"، لأن سدر مواطن أردني، وأي ملاحقة قضائية يجب أن تكون في بلده، ومن قبل القضاء الأردني، وليس من قبل السلطات الأمنية الإسرائيلية . وقال إن عقوبة هذه التهم، في حال إدانة موكله، تصل إلى السجن ما بين سبع إلى عشر سنوات.

يشار إلى أن سدر تمكن من دخول فلسطين المحتلة، بعد حصوله على تأشيرة دخول صرفرت عن السفارة الإسرائيلية في عمان، وتمنح تلك السفارة مئات المواطنين الأردنيين سنويا، تصريحاً لدخول فلسطين المحتلة، فيما يعتبر معارضون أردنيون ذلك بأنه اعتراف بالاحتلال، حيث أعدوا "قائمة سوداء" بأسماء من يحصل على تأشيرة إسرائيلية.

Human reference 1 summary	System summary
متابعة أردنية لاتهام (إسرائيل) لتاجر أردني بالاتصال بمنظمة "إرهابية"	متابعة أردنية لاتهام (إسرائيل) لتاجر أردني بالاتصال بمنظمة "إرهابية"
1.قالت وزارة الخارجية الأردنية إ نها تتابع باهتمام مع الجانب الإسرائيلي، تداعيات اعتقال السلطات للمواطن الأردني أمجد عبد المجيد سدر (46 عاما) بعد توجيه اتهامات له تنطوي على اتصاله بجهاز وجسم خارجي , وتقديم خدمات لمنظمة إرهابية واستعماله مواد غير قانونية لدعم الإرهاب،بحسب محاميه.	1.قالت وزارة الخارجية الأردنية إنها تتابع باهتمام مع الجانب الإسرائيلي، تداعيات اعتقال السلطات للمواطن الأردني أمجد عبد المجيد سدر (46 عاما) بعد توجيه اتهامات له تنطوي ع لى اتصاله بجهاز وجسم خارجي , وتقديم خدمات لمنظمة إرهابية واستعماله مواد غير قانونية لدعم الإرهاب،بحسب محاميه.
2.وكانت السلطات الأمنية الإسرائيلية وجهت لسدر الذي يعمل تاجراً ، بعد أن اعتقلته الجمعة الماضية (12/31) في جسر الشيخ حسين أثناء عودته من زيارة إلى القدس، ثلاث تهم تشمل "الاتصال بجهاز وجسم خارجي، وتقديم خدمات لمنظمة إرهابية، واستعمال مواد غير قانونية لدعم الإرهاب "، وفق محاميه محمد عابدين.	2.وأفاد مسئول في الخارجية، فضَّل عدم الكشف عن اسمه أن الخارجية ومنذ ورود معلومات إليها عما قامت به السلطات الإسرائيلي، بحق المواطن سدر "تحركنا لاستجلاء الموقف، وأجرينا اتصالات مع الخارجية الإسرائيلية عبر سفارتنا في تل أبيب".
3.وبين عابدين أن موكله سدر أكد خلال التحقيق، الذي أجرته السلطات الأمنية الإسرائيلية معه، براءته كليا من هذه التهم، نافيا جملة وتفصيلا صحتها، ومبدياً في	3.وكانت السلطات الأمنية الإسرائيلية وجهت لسدر الذي يعمل تاجراً ، بعد أن اعتقلته الجمعة الماضية (12/31) في جسر الشيخ حسين أثناء عودته من زيارة إلى القدس، ثلاث تهم تشمل "الاتصال بجهاز وجسم خارجي، وتقديم خدمات لمنظمة إرهابية،

الوقت نفسه استغرابه من أن توجه له مثل هذه التهم، التي قال إنه يسمع عنها لأول مرة "وهو التاجر المنشغل بتجارته وعمله". 4.يشار إلى أن سدر تمكن من دخول فلسطين المحتلة، بعد حصوله على تأشيرة دخول صدرت عن السفارة الإسرائيلية في عمان، وتمنح تلك السفارة مئات المواطنين الأردنيين سنويا، تصريحاً لدخول فلسطين المحتلة، فيما يعتبر معارضون أردنيون ذلك ب أنه اعتراف بالاحتلال، حيث أعدوا "قائمة سوداء " بأسماء من يحصل على تأشيرة إسرائيلية.	واستعمال مواد غير قانونية لدعم الإرهاب "، وفق محاميه محمد عابدين. 4.يشار إلى أن سدر تمكن من دخول فلسطين المحتلة، بعد حصوله على تأشيرة دخول صدرت عن السفارة الإسرائيلية في عمان، وتمنح تلك السفارة مئات المواطنين الأردنيين سنويا، تصريحاً لدخول فلسطين المحتلة، فيما يعتبر معارضون أردنيون ذلك بأنه اعتراف بالاحتلال، حيث أعدوا "قائمة سوداء" بأسماء من يحصل على تأشيرة إسرائيلية.

P	R	F
0.75	0.75	0.75

Human reference 2 summary	System summary
متابعة أردنية لاتهام (إسرائيل) لتاجر أردني بالاتصال بمنظمة "إرهابية"	متابعة أردنية لاتهام (إسرائيل) لتاجر أردني بالاتصال بمنظمة "إرهابية"
1.قالت وزارة الخارجية الأردنية إنها تتابع باهتمام مع الجانب الإسرائيلي، تداعيات اعتقال السلطات للمواطن الأردني أمجد عبد المجيد سدر (46 عاما) بعد توجيه اتها مات له تنطوي على اتصاله بجهاز وجسم خارجي , وتقديم خدمات لمنظمة إرهابية واستعماله مواد غير قانونية لدعم الإرهاب،بحسب محاميه.	1.قالت وزارة الخارجية الأردنية إنها تتابع باهتمام مع الجانب الإسرائيلي، تداعيات اعتقال السلطات للمواطن الأردني أمجد عبد المجيد سدر (46 عاما) بعد توجيه اتهامات له تنطوي على اتصاله بجهاز وجسم خارجي , وتقديم خدمات لمنظمة إرهابية واستعماله مواد غير قانونية لدعم الإرهاب،بحسب محاميه.
2.وأفاد مسئول في الخارجية، فضَّل عدم الكشف عن اسمه أن الخارجية ومنذ ورود معلومات إليها عما قامت به السلطات الإسرائيلي، بحق المواطن سدر "تحركنا لاستجلاء الموقف، وأجرينا اتصالات مع الخارجية الإسرائيلية عبر سفارتنا في تل أبيب".	2.وأفاد مسئول في ا لخارجية، فضَّل عدم الكشف عن اسمه أن الخارجية ومنذ ورود معلومات إليها عما قامت به السلطات الإسرائيلي، بحق المواطن سدر "تحركنا لاستجلاء الموقف، وأجرينا اتصالات مع الخارجية الإسرائيلية عبر سفارتنا في تل أبيب".
3.وبين عابدين أن موكله سدر أكد خلال التحقيق، الذي أجرته السلطات الأمنية الإسرائيلية معه، براءته كليا من هذه التهم، نافيا جملة وتفصيلا صحتها، ومبدياً في الوقت نفسه استغرابه من أن توجه له مثل هذه التهم، التي قال إنه يسمع عنها لأول مرة "وهو التاجر المنشغل بتجارته وعمله".	3.وكانت السلطات الأمنية الإسرائيلية وجهت لسدر الذي يعمل تاجراً ، بعد أن اعتقلته الجمعة الماضية (12/31) في جسر الشيخ حسين أثناء عودته من زيارة إلى القدس، ثلاث تهم تشمل "الاتصال بجهاز وجسم خارجي، وتقديم خدمات لمنظمة إرهابية، واستعمال مواد غير قانونية لدعم الإرهاب "، وفق محاميه محمد عابدين.
4.وأوضح المحامي عابدين أن هذه التهم "غير قانونية"، واصفا إياها بـ"التعسفية"، لأن سدر مواطن أردني، وأي ملاحقة قضائية يجب أن تكون في بلده، ومن قبل القضاء الأ ردني، وليس من قبل السلطات الأمنية الإسرائيلية . وقال إن عقوبة هذه التهم، في حال إدانة موكله، تصل إلى السجن ما بين سبع إلى	4. يشار إلى أن سدر تمكن من دخول فلسطين المحتلة، بعد حصوله على تأشيرة دخول صدرت عن السفارة الإسرائيلية في عمان، وتمنح تلك السفارة مئات المواطنين الأردنيين سنويا، تصريحاً لدخول فلسطين المحتلة، فيما يعتبر معارضون أردنيون ذلك بأنه اعتراف بالاحتلال، حيث أعدوا "قائمة

عشر سنوات. 5.يشار إلى أن سدر تمكن من دخول فلسطين المحتلة، بعد حصوله على تأشيرة دخول صدرت عن السفارة الإسرائيلية في عمان، وتمنح تلك السفا رة مئات المواطنين الأردنيين سنويا، تصريحاً لدخول فلسطين المحتلة، فيما يعتبر معارضون أردنيون ذلك بأنه اعتراف بالاحتلال، حيث أعدوا "قائمة سوداء " بأسماء من يحصل على تأشيرة إسرائيلية.	سوداء" بأسماء من يحصل على تأشيرة إسرائيلية.

P	R	F
0.75	0.6	0.667
Average F		
0.708		

Table 4.7 show an evaluation for Arabic text by comparing AATSS with two human summaries and for each human we compute P,R and f- measures. In final we compute the average f-measure between the two summaries.

Table 4.7. Summary evaluation 3 (the underlined sentences in the summary of AATSS are existing in human summary)

Original text
"الأوقاف" تنظم ورشة عمل لتقييم أداء الموظفين نظمت وزارة الأوقاف والشئون الدينية, أمس، ورشة عمل حول "تقييم أداء الموظفين" بالتنسيق والتعاون مع ديوان الموظفين العام, بحضور ومشاركة وكيل الوزارة د.عبد الله أبو جربوع، ومدير عام ديوان الموظفين م.سمير مطير، ومدير عام السياسات والتخطيط بالديوان فدوان أبو شريعة، إلى جانب لفيف من مدراء ورؤساء أقسام الوزارة. وقال أبو جربوع : " إننا محظوظون بطاقم من الموظفين هم الأسبق في الخير والعطاء، ولديهم الكثير من حيث الأداء والروح الايجابية التي استطاعوا من خلالها أن يزيدوا من عطائهم لتحقيق الأهداف التي نسعى إليها". وأكد على ضرورة التجديد في العمل لأنه يمثل النجاح من خلال الإبداع والبعد عن الروتينية , لافتاً إلى أن الإدارة الناجحة لابد لها من تخطيط ومتابعة وتقييم للأداء لضمان نجاحها. بدورها، أكدت أبو شريعة على أهمية تطوير برنامج نظ ام تقييم الأداء في الوزارات ومؤسسات السلطة الفلسطينية , منوهة إلى أن الهدف من هذه الورشة هو اطلاع شريحة المقيمين في الوزارات على الإصدار الأول من دليل نظام تطوير تقييم الأداء الحكومي الذي تم إعداده من قبل ديوان الموظفين بغية تحقيق الموضوعية في عملية التقييم. أوقاف خان يونس من جهة ثانية، تولى عبد الهادي الأغا مهام مسئولية مديرية أوقاف خان يونس بعد انتهاء خدمة مديرها السابق الشيخ سليمان الفرا خلال فترة حافلة بالانجازات والأنشطة المختلفة. جاء ذلك خلال لقاء جمع كلاً من وزير الأوقاف الدكتور طالب أبو شعر ووكيل وزا رة الأوقاف د . عبد الله أبو

جربوع، ومدير عام الشئون الإدارية محمد أبو عسكر، وعبد الهادي الأغا والشيخ سليمان الفرا.

بدوره، ثمن وزير الأوقاف جهود الفرا في خدمة أبنا الشعب الفلسطيني خاصة بمحافظة خان يونس , مشيداً بأدائه المميز والمنشود خلال الفترة التي قضاها.

وأعرب عن أمله في أن تتواصل وتستمر جهوده التي تنم عن مدى حرصه الكبير في إنجاح المسيرة الدعوية التي عهدناها على مدار الأعوام الماضية.

ودعا أبو شعر الأغا إلى ضرورة العمل على تقييم الوضع الحالي ورسم الخطط المستقبلية التي تضمن نجاح العمل والاستمرار به, مشدداً على أهمية وضع السياسيات العامة لاستكمال المشوار الدعوي والعمل على تطوير العمل الإداري للارتقاء بالمستوى المهني والوظيفي لدى الموظفين . من جانبه، أبدى الأغا استعداده لهذه المهمة الموكلة إليه, معرباً عن أمله في أن يكون له الدور المميز من أجل إنجاح الأعمال اليوم ية وتحقيق الأهداف التي تسعى إليها الوزارة.

Human reference 1 summary	System summary
"الأوقاف" تنظم ورشة عمل لتقييم أداء الموظفين	"الأوقاف" تنظم ورشة عمل لتقييم أداء الموظفين
1.نظمت وزارة الأوقاف والشئون الدينية , أمس، ورشة عمل حول "تقييم أداء الموظفين " بالتنسيق والتعاون مع ديوان الموظفين العام , بحضور ومشاركة وكيل الوزارة د .عبد الله أبو جربوع، ومدير عام ديوان الموظفين م .سمير مطير، ومدير عام السياسات والتخطيط بالديوان فدوان أبو شريعة، إلى جانب لفيف من مدراء ورؤساء أقسام الوزارة.	1.نظمت وزارة الأوقاف والشئون الدينية, أمس، ورشة عمل حول "تقييم أداء الموظفين" بالتنسيق والتعاون مع ديوان الموظفين العام , بحضور ومشاركة وكيل الوزارة د .عبد الله أبو جربوع، ومدير عام ديوان الموظفين م .سمير مطير، ومدير عام السياسات والتخطيط بالديوان فدوان أبو شريعة، إلى جانب لفيف من مدراء ورؤساء أقسام الوزارة.
2.وقال أبو جربوع : " إننا محظوظون بطاقم من الموظفين هم الأسبق في الخير والعطاء، ولديهم الكثير من حيث الأداء والر وح الايجابية التي استطاعوا من خلالها أن يزيدوا من عطائهم لتحقيق الأهداف التي نسعى إليها".	2.وقال أبو جربوع : " إننا محظوظون بطاقم من الموظفين هم الأسبق في الخير والعطاء ، ولديهم الكثير من حيث الأداء والروح الايجابية التي استطاعوا من خلالها أن يزيدوا من عطائهم لتحقيق الأهداف التي نسعى إليها".
3.بدورها، أكدت أبو شريعة على أهمية تطوير برنامج نظام تقييم الأداء في الوزارات ومؤسسات السلطة الفلسطينية, منوهة إلى أن الهدف من هذه الورشة هو اطلاع شريحة المقيمين في الو زارات على الإصدار الأول من دليل نظام تطوير تقييم الأداء الحكومي الذي تم إعداده من قبل ديوان الموظفين بغية تحقيق الموضوعية في عملية التقييم.	3.بدورها، أكدت أبو شريعة على أهمية تطوير برنامج نظام تقييم الأداء في الوزارات ومؤسسات السلطة الفلسطينية, منوهة إلى أن الهدف من هذه الورشة هو اطلاع شريحة المقيمين في الوزارات على الإصدار الأول من دليل نظام تطوير تقييم الأداء الحكومي الذي تم إعداده من قبل ديوان الموظفين بغية تحقيق الموضوعية في عملية التقييم.
4.من جهة ثانية، تولى عبد الهادي الأغا مهام مسئولية مديرية أوقاف خان يونس بعد انتهاء خدمة مديرها السابق الشيخ سليمان ال فرا خلال فترة حافلة بالانجازات والأنشطة المختلفة.	4.من جهة ثانية، تولى عبد الهادي الأغا مهام مسئولية مديرية أوقاف خان يونس بعد انتهاء خدمة مديرها السابق الشيخ سليمان الفرا خلال فترة حافلة بالانجازات والأنشطة المختلفة.
5.ودعا أبو شعر الأغا إلى ضرورة العمل على تقييم الوضع الحالي ورسم الخطط المستقبلية التي تضمن نجاح العمل والاستمرار به , مشدداً على أهمية وضع السياسيات العامة لاستكمال المشوار الدعوي والعمل	5.بدوره، ثمن وزير الأوقاف جهود الفرا في خدمة أبنا الشعب الفلسطيني خاصة بمحافظة خان يونس , مشيداً بأدائه المميز والمنشود خلال الفترة التي

على تطوير العمل الإداري للارتقاء بالمستوى المهني والوظيفي لدى الموظفين.	قضاها. 6.ودعا أبو شعر الأغا إلى ضرورة العمل على تقييم الوضع الحالي ورسم الخطط المستقبلية التي تضمن نجاح العمل والاستمرار به , مشدداً على أهمية وضع السياسيات العامة لاستكمال المشوار الدعوي والعمل على تطوير العمل الإد اري للارتقاء بالمستوى المهني والوظيفي لدى الموظفين.

P	R	F
0.833	1	0.909

Human reference 2 summary	System summary
"الأوقاف" تنظم ورشة عمل لتقييم أداء الموظفين	"الأوقاف" تنظم ورشة عمل لتقييم أداء الموظفين
1.نظمت وزارة الأوقاف والشئون الدينية , أمس، ورشة عمل حول "تقييم أداء الموظفين " بالتنسيق والتعاون مع ديوان الموظفين العام , بحضور ومشاركة وكيل الوزارة د .عبد الله أبو جربوع، ومدير عام ديوان الموظفين م .سمير مطير، ومدير عام السياسات والتخطيط بالديوان فدوان أبو شريعة، إلى جانب لفيف من مدراء ورؤساء أقسام الوزارة.	1. نظمت وزارة الأوقاف والشئون الدينية, أمس، ورشة عمل حول "تقييم أداء الموظفين" بالتنسيق والتعاون مع ديوان الموظفين العام , بحضور ومشاركة وكيل الوزارة د .عبد الله أبو جربوع، ومدير عام ديوان الموظفين م .سمير مطير، ومدير عام السياسات والتخطيط بالديوان فدوان أبو شريعة، إلى جانب لفيف من مدراء ورؤساء أقسام الوزارة.
2.وقال أبو جربوع : " إننا محظوظون بطاقم من الموظفين هم الأسبق في الخير والعطاء، ولديهم الكثير من حيث الأداء والروح الايجابية التي استطاعوا من خلالها أن يزيدوا من عط ائهم لتحقيق الأهداف التي نسعى إليها".	2.وقال أبو جربوع : " إننا محظوظون بطاقم من الموظفين هم الأسبق في الخير والعطاء، ولديهم الكثير من حيث الأداء والروح الايجابية التي استطاعوا من خلالها أن يزيدوا من عطائهم لتحقيق الأهداف التي نسعى إليها".
3.بدورها، أكدت أبو شريعة على أهمية تطوير برنامج نظام تقييم الأداء في الوزارات ومؤسسات السلطة الفلسطينية, منوهة إلى أن الهدف من هذه الورشة هو اطلاع شريحة المقيمين في الوزارات على الإصدار الأول من دليل نظام تطوير تقييم الأ داء الحكومي الذي تم إعداده من قبل ديوان الموظفين بغية تحقيق الموضوعية في عملية التقييم.	3.بدورها، أكدت أبو شريعة على أهمية تطوير برنامج نظام تقييم الأداء في الوزارات ومؤسسات السلطة الفلسطينية, منوهة إلى أن الهدف من هذه الورشة هو اطلاع شريحة المقيمين في الوزارات على الإصدار الأول من دليل نظام تطوير تقييم الأداء الحكومي الذي تم إعداده من قبل ديوان الموظفين بغية تحقيق الموضوعية في عملية التقييم.
4.من جهة ثانية، تولى عبد الهادي الأغا مهام مسئولية مديرية أوقاف خان يونس بعد انتهاء خدمة مديرها السابق الشيخ سليمان الفرا خلال فترة حافلة بالانجازات والأنشطة المختلفة.	4.من جهة ثانية، تولى عبد الهادي الأغا مهام مسئولية مديرية أوقاف خان يونس بعد انتهاء خدمة مديرها السابق الشيخ سليمان الفرا خلال فترة حافلة بالانجازات والأنشطة المختلفة.
5.ودعا أبو شعر الأغا إلى ضرورة العمل على تقييم الوضع الحالي ورسم الخطط المستقبلية التي تضمن نجاح العمل والاستمرار به , مشدداً على أهمية وضع السياسيات العامة لاستكمال المشوار الدعوي والعمل	5.بدوره، ثمن وزير الأوقاف جهود الفرا في خدمة أبنا الشعب الفلسطيني خاصة بمحافظة خان يونس , مشيداً بأدائه المميز والمنشود خلال الفترة التي

على تطوير العمل الإداري للارتقاء بالمستوى المهني والوظيفي لدى الموظفين.	قضاها. 6.ودعا أبو شعر الأغا إلى ضرورة العمل على تقييم الوضع الحالي ورسم الخطط المستقبلية التي تضمن نجاح العمل والاستمرار به, مشدداً على أهمية وضع السياسيات العامة لاستكمال المشوار الدعوي والعمل على تطوير العمل الإداري للارتقاء بالمستوى المهني والوظيفي لدى الموظفين.

P	R	F
0.833	1	0.909
Average F		
0.909		

4.4.2. Sakhr Summarizer and Human Evaluation

In chapter 1, we talk about Sakhr as an Arabic summarization tool which identifies the most relevant sentences within a text and displays them in the form of a short text summary.

In order to compare between AATSS and Sakhr online summarizer[1] we should firstly compare each other with human summary to obtain average F-measure. Table 4.8 to table 4.10 shows some texts but summarized with Sakhr on line summarizer rather than AATSS.

Table 4.8 show an evaluation for Arabic text by comparing Sakhr with two human summaries and for each human we compute P,R and f- measures. In final we compute the average f-measure between the two summaries.

Table 4.8. Sakhr evaluation 1 (the underlined sentences in the Sakhr summary are existing in human summary)

Original text
"الأشغال العامة" تستقبل أكثر من 4 آلاف طلب تمليك أراض أكد وكيل وزارة الأشغال العامة والإسكان المساعد لشؤون التخطيط المهندس ناجي سرحان , أن إقبالًا شديدًا فاق التوقعات على مشروع تم ليك الأراضي الذي طرحته الوزارة مؤخرًا، مشيرًا إلى أن آخر موعد لقبول الطلبات هو 2011/1/31. وأفاد سرحان خلال ندوة عقدتها نقابة المهندسين – فرع الوسطى- حول مشاريع الإسكان الجديدة , أول من أمس, أن الوزارة استقبلت أكثر من أربعة آلاف طلب وجاري تفريغها على برنامج الانتفاع, حيث يتم تصنيفها إلى فئات حتى تكون جاهزة للدراسة ولعملية القرعة. وقال : "إن مشاريع تمليك الأراضي لاقت إقبالًا شديدًا من قبل المواطنين وأصبح عدد المسجلين لها يفوق

[1]http://textmining.sakhr.com/

التوقع"، مشيرًا إلى أن هذه المشاريع تهدف بالأساس للتخفيف من أزمة السكن التي يعيشها قطاع غزة.

وأوضح أنه سيتم فرز الطلبات ودراستها وسيتم بعدها إجراء قرعة على من يستحق قسيمة، وستجرى بالتساوي على جميع الطلبات المسجلة لدى الوزارة.

وأظهر استطلاع للرأي نشرته الوزارة على موقعها الالكتروني, مؤخرًا, وشارك فيه 900 شخص, أن مشاريع الإسكان الجديدة ستخفف بشكل كبير من أزمة السكن الخانقة التي يحياها قطاع غزة المحاصر.

Human reference 1 summary	Sakhr summary
"الأشغال العامة" تستقبل أكثر من 4 آلاف طلب تمليك أراض 1.أكد وكيل وزارة الأشغال العامة والإسكان المساعد لشؤون التخطيط المهندس ناجي سرحان , أن إقبالًا شديدًا فاق التوقعات على مشروع تمليك الأراضي الذي طرحته الوزارة مؤخرًا، مشيرًا إ لى أن آخر موعد لقبول الطلبات هو 2011/1/31. 2.وأوضح أنه سيتم فرز الطلبات ودراستها وسيتم بعدها إجراء قرعة على من يستحق قسيمة، وستجرى بالتساوي على جميع الطلبات المسجلة لدى الوزارة. 3.وأظهر استطلاع للرأي نشرته الوزارة على موقعها الالكتروني, مؤخرًا, وشارك فيه 900 شخص, أن مشاريع الإسكان الجديدة ستخفف بشكل كبير من أزمة السكن الخانقة التي يحياها قطاع غزة المحاصر	"الأشغال العامة" تستقبل أكثر من 4 آلاف طلب تمليك أراض 1.أكد وكيل وزارة الأشغال العامة والإسكان المساعد لشؤون التخطيط المهندس ناجي سرحان , أن إقبالًا شديدًا فاق التوقعات على مشروع تمليك الأراضي الذي طرحته الوزارة مؤخرًا، مشيرًا إلى أن آخر موعد لقبول الطلبات هو 2011/1/31. 2.وأفاد سرحان خلال ندوة عقدتها نقابة المهندسين – فرع الوسطى- حول مشاريع الإسكان الجديدة , أول من أمس, أن الوزارة استقبلت أكثر من أربعة آلاف طلب وجاري تفريغها على برنامج الانتفاع , حيث يتم تصنيفها إلى فئات حتى تكون جاهزة للدراسة ولعملية القرعة. 3.وقال : "إن مشاريع تمليك الأراضي لاقت إقبالًا شديدًا من قبل المواطنين وأصبح عدد المسجلين لها يفوق التوقع"، مشيرًا إلى أن هذه المشاريع تهدف بالأساس للتخفيف من أز مة السكن التي يعيشها قطاع غزة.

F	R	P
0.33	0.33	0.33

Human reference 2 summary	Sakhr summary
"الأشغال العامة" تستقبل أكثر من 4 آلاف طلب تمليك أراض 1.أكد وكيل وزارة الأشغال العامة والإسكان المساعد لشؤون التخطيط المهندس ناجي سرحان , أن إقبالًا شديدًا فاق التوقعات على مشروع تمليك الأراضي الذي طرحته الوزارة مؤخرًا، مشيرًا إلى أن آخر موعد لقبول الطلبات هو 2011/1/31. 2.وأفاد سرحان خلال ندوة عقدتها نقابة المهندسين – فرع الوسطى- حول مشاريع الإسكان الجديدة , أول من أمس, أن الوزارة استقبلت أكثر من أربعة آلاف طلب وجاري تفريغها على برنامج الانتفاع , حيث	"الأشغال العامة" تستقبل أكثر من 4 آلاف طلب تمليك أراض 1.أكد وكيل وزارة الأشغال العامة والإسكان ا لمساعد لشؤون التخطيط المهندس ناجي سرحان , أن إقبالًا شديدًا فاق التوقعات على مشروع تمليك الأراضي الذي طرحته الوزارة مؤخرًا، مشيرًا إلى أن آخر موعد لقبول الطلبات هو 2011/1/31. 2.وأفاد سرحان خلال ندوة عقدتها نقابة المهندسين – فرع الوسطى- حول مشاريع الإسكان الجديدة, أول من أمس, أن الوزارة استقبلت أكثر من أربعة آلاف طلب وجاري تفريغها على برنامج الانتفاع , حيث

يتم تصنيفها إلى فئات حتى تكون جاهزة للدراسة ولعملية القرعة. 3.وأظهر استطلاع للرأي نشرته الوزارة على موقعها الالكتروني, مؤخرًا, وشارك فيه 900 شخص, أن مشاريع الإسكان الجديدة ستخفف بشكل كبير من أزمة السكن الخانقة التي يحياها قطاع غزة المحاصر	يتم تصنيفها إلى فئات حتى تكون جاهزة للدراسة ولعملية القرعة. 3.وقال: " إن مشاريع تمليك الأراضي لاقت إقبالًا شديدًا من قبل المواطنين وأصبح عدد المسجلين لها يفوق التوقع"، مشيرًا إلى أن هذه المشاريع تهدف بالأساس للتخفيف من أزمة السكن التي يعيشها قطاع غزة.

P	R	F
0.667	0.667	0.667
Average F		
0.5		

Table 4.9 show an evaluation for Arabic text by comparing Sakhr with two human summaries and for each human we compute P,R and f- measures. In final we compute the average f-measure between the two summaries.

Table 4.9. Sakhr evaluation 2 (the underlined sentences in the Sakhr summary are existing in human summary)

Original text	
الكتلة الإسلامية تنظم حفلاً فنياً وسط القطاع نظمت الكتلة الإسلامية في مدينة دير البلح، وسط قطاع غزة، أمس، احتفالاً فنياً بمناسبة الذكرى الثانية للحرب على قطاع غزة، بمشاركة قادة الكتلة وحركة "حماس"، ورئيس بلدية دير البلح سعيد نصار، ولفيف من الطلاب. وفي كلمة الكتلة الإسلامية التي ألقاها طارق شاهين، تحدث فيها عن الحرب على غزة مستذكراً الشهداء القادة الذين ارتقوا خلال الحرب وعلى رأسهم ال دكتور نزار ريان، ووزير الداخلية سعيد صيام، وشهداء الكتلة الإسلامية الذين استشهدوا في مدارسهم وجامعاتهم. وأكد شاهين على دور الكتلة في تجسيد الذكرى الثانية في قلوب الطلاب "لتتناقلها الأجيال جيلاً بعد جيل "، شاكراً الحضور على تلبيتهم لدعوة الكتلة للمشاركة في الاحتفال. وتخللت الاحتفال فقرات إنشادية وعرض مرئي خاص بالحرب الأخيرة، إضافة إلى عرض مسرحي لفرقة النشامى الفنية يُعرض لأول مرة للفنان نبيل الخطيب، والتي كانت بعنوان "مهمة في غزة " حيث جسدت واقع جيش الاحتلال خلال الحرب. من جهة أخرى، نظمت الكتلة في مخيم البريج، وسط القطاع، أسبوعاً دعوياً بالتعاون مع جمعية التبيان لرعاية حفظة القرآن تحت عنوان "علي خُطى الحبيب... إنا ماضون."	
Sakhr summary	Human reference 1 summary
الكتلة الإسلامية تنظم حفلاً فنياً وسط القطاع	الكتلة الإسلامية تنظم حفلاً فنياً وسط القطاع

1. نظمت الكتلة الإسلامية في مدينة دير البلح، وسط قطاع غزة، أمس، احتفالاً فنياً بمناسبة الذكرى الثانية للحرب على قطاع غزة، بمشاركة قادة الكتلة وحركة "حماس "، ورئيس بلدية دير البلح سعيد نصار، ولفيف من الطلاب. 2. وفي كلمة الكتلة الإسلامية التي ألقاها طارق شاهين، تحدث فيها عن الحرب على غزة مستذكراً الشهداء القادة الذين ارتقوا خلال الحرب وعلى رأسهم الدكتور نزار ريان، ووزير الداخلية سعيد صيام، وشهداء الكتلة الإسلامية الذين استشهدوا في مدارسهم وجامعاتهم. 3. وتخللت الاحتفال فقرات إنشادية و عرض مرئي خاص بالحرب الأخيرة، إضافة إلى عرض مسرحي لفرقة النشامى الفنية يُعرض لأول مرة للفنان نبيل الخطيب، والتي كانت بعنوان "مهمة في غزة" حيث جسدت واقع جيش الاحتلال خلال الحرب.	1. نظمت الكتلة الإسلامية في مدينة دير البلح، وسط قطاع غزة ، أمس، احتفالاً فنياً بمناسبة الذكرى الثانية للحرب على قطاع غزة، بمشاركة قادة الكتلة وحركة "حماس "، ورئيس بلدية دير البلح سعيد نصار، ولفيف من الطلاب. 2. وفي كلمة الكتلة الإسلامية التي ألقاها طارق شاهين، تحدث فيها عن الحرب على غزة مستذكراً الشهداء القادة الذين ارتقوا خلال الحرب وعلى رأسهم الدكتور نزار ريان، ووزير الداخلية سعيد صيام، وشهداء الكتلة الإسلامية الذين استشهدوا في مدارسهم وجامعاتهم. 3. وأكد شاهين على دور الكتلة في تجسيد الذكرى الثانية في قلوب الطلاب "لتتناقلها الأجيال جيلاً بعد جيل "، شاكراً الحضور على تلبيتهم لدعوة الكتلة للمشاركة في الاحتفال. 4. وتخللت الاحتفال فقرات إنشادية وعرض مرئي خاص بالحرب الأخيرة، إضافة إلى عرض مسرحي لفرقة النشامى الفنية يُعرض لأول مرة للفنان نبيل الخطيب، والتي كانت بعنوان "مهمة في غزة" حيث جسدت واقع جيش الاحتلال خلال الحرب.

F	R	P
0.857	1	0.75

Human reference 2 summary	Sakhr summary
الكتلة الإسلامية تنظم حفلاً فنياً وسط القطاع 1. نظمت الكتلة الإسلامية في مدينة دير البلح، وسط قطاع غزة، أمس، احتفالاً فنياً بمناسبة الذكرى الثانية للحرب على قطاع غزة، بمشاركة قادة الكتلة وحركة "حماس "، ورئيس بلدية دير البلح سعيد نصار، ولفيف من الطلاب. 2. وفي كلمة الكتلة الإسلامية التي ألقاها طارق شاهين، تحدث فيها عن الحرب على غزة مستذكراً الشهداء القادة الذين ارتقوا خلال الحرب وعلى رأسهم الدكتور نزار ريان، ووزير الداخلية سعيد صيام، وشهداء الكتلة الإسلامية الذين استشهدوا في مدارسهم وجامعاتهم. 3. وتخللت الاحتفال فقرات إنشادية وعرض مرئي خاص بالحرب الأخيرة، إضافة إلى عرض مسرحي لفرقة النشامى الفنية يُعرض لأول مرة للفنان نبيل الخطيب، والتي كانت بعنوان "مهمة في غزة" حيث	الكتلة الإسلامية تنظم حفلاً فنياً وسط القطاع 1. نظمت الكتلة الإسلامية في مدينة دير البلح، وسط قطاع غزة، أمس، احتفالاً فنياً بمناسبة الذكرى الثانية للحرب على قطاع غزة، بمشاركة قادة الكتلة وحركة "حماس "، ورئيس بلدية دير البلح سعيد نصار، ولفيف من الطلاب. 2. وفي كلمة الكتلة الإسلامية التي ألقاها طارق شاهين، تحدث فيها عن الحرب على غزة مستذكراً الشهداء القادة الذين ارتقوا خلال الحرب وعلى رأسهم الدكتور نزار ريان، ووزير الداخلية سعيد صيام، وشهداء الكتلة الإسلامية الذين استشهدوا في مدارسهم وجامعاتهم. 3. وأكد شاهين على دور الكتلة في تجسيد الذك رى الثانية في قلوب الطلاب "لتتناقلها الأجيال جيلاً بعد جيل "، شاكراً الحضور على تلبيتهم لدعوة الكتلة للمشاركة في الاحتفال. 4. وتخللت الاحتفال فقرات إنشادية وعرض مرئي

جسدت واقع جيش الاحتلال خلال الحرب.	خاص بالحرب الأخيرة، إضافة إلى عرض مسرحي لفرقة النشامى الفنية يُعرض لأول مرة للفنان نبيل الخطيب، والتي كانت بعنوان "مهمة في غزة" حيث جسدت واقع جيش الاحتلال خلال الحرب.

P	R	F
0.75	1	0.857
Average F		
0.857		

Table 4.10 show an evaluation for Arabic text by comparing Sakhr with two human summaries and for each human we compute P,R and f- measures. In final we compute the average f-measure between the two summaries.

Table 4.10. Sakhr evaluation 3 (the underlined sentences in the Sakhr summary are existing in human summary) this text summarized by AATSS in table 3.6

Original text
"الأوقاف" تنظم ورشة عمل لتقييم أداء الموظفين نظمت وزارة الأوقاف والشئون الدينية, أمس، ورشة عمل حول "تقييم أداء الموظفين" بالتنسيق والتعاون مع ديوان الموظفين العام, بحضور ومشاركة وكيل الوزارة د.عبد الله أبو جربوع، ومدير عام ديوان الموظفين م.سمير مطير، ومدير عام السياسات والتخطيط بالديوان فدوان أبو شريعة، إلى جانب لفيف من مدراء ورؤساء أقسام الوزارة. وقال أبو جربوع : " إننا محظوظون بطاقم من الموظفين هم الأسبق في الخير والعطاء، ولديهم الكثير من حيث الأداء والروح الايجابية التي استطاعوا من خلالها أن يزيدوا من عطائهم لتحقيق الأهداف التي نسعى إليها". وأكد على ضرورة التجديد في العمل لأنه يمثل النجاح من خلا ل الإبداع والبعد عن الروتينية , لافتاً إلى أن الإدارة الناجحة لابد لها من تخطيط ومتابعة وتقييم للأداء لضمان نجاحها. بدورها، أكدت أبو شريعة على أهمية تطوير برنامج نظام تقييم الأداء في الوزارات ومؤسسات السلطة الفلسطينية , منوهة إلى أن الهدف من هذه الورشة هو ا طلاع شريحة المقيمين في الوزارات على الإصدار الأول من دليل نظام تطوير تقييم الأداء الحكومي الذي تم إعداده من قبل ديوان الموظفين بغية تحقيق الموضوعية في عملية التقييم. أوقاف خان يونس من جهة ثانية، تولى عبد الهادي الأغا مهام مسئولية مديرية أوقاف خان يونس بعد انتهاء خدمة مديرها السابق الشيخ سليمان الفرا خلال فترة حافلة بالانجازات والأنشطة المختلفة. جاء ذلك خلال لقاء جمع كلاً من وزير الأوقاف الدكتور طالب أبو شعر ووكيل وزارة الأوقاف د . عبد الله أبو جربوع، ومدير عام الشئون الإدارية محمد أبو عسكر، وعبد الهادي الأغا والشيخ سليمان الفرا. بدوره، ثمن وزير الأوقاف جهود الفرا في خدمة أبنا الشعب الفلسطيني خاصة بمحافظة خان يونس , مشيداً بأدائه المميز والمنشود خلال الفترة التي قضاها.

وأعرب عن أمله في أن تتواصل وتستمر جهوده التي تنم عن مدى حرصه الكبير في إنجاح المسيرة الدعوية التي عهدناها على مدار الأعوام الماضية.

ودعا أبو شعر الأغا إلى ضرورة العمل على تقييم الوضع الحالي ورسم الخطط المستقبلية التي تضمن نجاح العمل والاستمرار به, مشدداً على أهمية وضع السياسيات العامة لاستكمال المشوار الدعوي والعمل على تطوير العمل الإداري للارتقاء بالمستوى المهني والوظيفي لدى الموظفين . من جانبه، أبدى الأغا استعداده لهذه المهمة الموكلة إليه, معرباً عن أمله في أن يكون له الدور المميز من أجل إنجاح الأعمال اليومية وتحقيق الأهداف التي تسعى إليها الوزارة.

Human reference 1 summary	Sakhr summary
"الأوقاف" تنظم ورشة عمل لتقييم أداء الموظفين	"الأوقاف" تنظم ورشة عمل لتقييم أداء الموظفين
1.نظمت وزارة الأوقاف والشؤون الدينية, أمس، ورشة عمل حول "تقييم أداء الموظفين " بالتنسيق والتعاون مع ديوان الموظفين العام , بحضور ومشاركة وكيل الوزارة د .عبد الله أبو جربوع، ومدير عام ديوان الموظفين م .سمير مطير، ومدير عام السياسات والتخطيط بالديوان فدوان أبو شريعة، إلى جانب لفيف من مدراء ورؤساء أقسام الوزارة.	1.نظمت وزارة الأوقاف والشئون الدينية, أمس، ورشة عمل حول "تقييم أداء الموظفين" بالتنسيق والتعاون مع ديوان الموظفين العام , بحضور ومشاركة وكيل الوزارة د .عبد الله أبو جربوع، ومدير عام ديوان الموظفين م .سمير مطير، ومدير عام ال سياسات والتخطيط بالديوان فدوان أبو شريعة، إلى جانب لفيف من مدراء ورؤساء أقسام الوزارة.
2.وقال أبو جربوع : " إننا محظوظون بطاقم من الموظفين هم الأسبق في الخير والعطاء، ولديهم الكثير من حيث الأداء والروح الايجابية التي استطاعوا من خلالها أن يزيدوا من عطائهم لتحقيق الأهداف التي نسعى إليها".	2.وأكد على ضرورة التجديد في العمل لأنه يمثل النجاح من خلال الإبداع والبعد عن الروتينية , لافتاً إلى أن الإدارة الناجحة لابد لها من تخطيط ومتابعة وتقييم للأداء لضمان نجاحها.
3.بدورها، أكدت أبو شريعة على أهمية تطوير برنامج نظام تقييم الأداء في الوزارات ومؤسسات السلطة الفلسطينية, منوهة إلى أن الهدف من هذه الورشة هو اطلاع شريحة المقيمين في الوزارات على الإصدار الأول من دليل نظام تطوير تقييم الأداء الحكومي الذي تم إعداده من قبل ديوان الموظفين بغية تحقيق الموضوعية في عملية التقييم.	3.جاء ذلك خلال لقاء جمع كلاً من وزير الأوقاف الدكتور طالب أبو شعر ووكيل وزارة الأوقاف د . عبد الله أبو جربوع، ومدير عام الشئون الإدارية محمد أبو عسكر، وعبد الهادي الأغا والشيخ سليمان الفرا
4.من جهة ثانية، تولى عبد الهادي الأغا مهام مسئولية مديرية أوقاف خان يونس بعد انتهاء خدمة مديرها السابق الشيخ سليمان الفرا خلال فترة حافلة بالانجازات والأنشطة المختلفة.	
5.ودعا أبو شعر الأغا إلى ضرورة العمل على تقييم الوضع الحالي ورسم الخطط ال مستقبلية التي تضمن نجاح العمل والاستمرار به , مشدداً على أهمية وضع السياسيات العامة لاستكمال المشوار الدعوي والعمل على تطوير العمل الإداري للارتقاء بالمستوى المهني والوظيفي لدى الموظفين.	

P	R	F
0.333	0.2	0.25

Human reference 2 summary	Sakhr summary
"الأوقاف" تنظم ورشة عمل لتقييم أداء الموظفين 1. نظمت وزارة الأوقاف والشئون الدينية, أمس، ورشة عمل حول "تقييم أداء الموظفين " بالتنسيق والتعاون مع ديوان الموظفين العام , بحضور ومشاركة وكيل الوزارة د .عبد الله أبو جربوع، ومدير عام ديوان الموظفين م .سمير مطير، ومدير عام السياسات والتخطيط بالديوان فدوان أبو شريعة، إلى جانب لفيف من مدراء ورؤساء أقسام الوزارة. 2. وقال أبو جربوع : " إننا محظوظون بطاقم من الموظفين هم الأسبق في الخير والعطاء، ولديهم الكثير من حيث الأداء والروح الايجابية التي استطاعوا من خلالها أن يزيدوا من عطائهم لتحقيق الأهداف التي نسعى إليها". 3. بدورها، أكدت أبو شريعة على أه مية تطوير برنامج نظام تقييم الأداء في الوزارات ومؤسسات السلطة الفلسطينية, منوهة إلى أن الهدف من هذه الورشة هو اطلاع شريحة المقيمين في الوزارات على الإصدار الأول من دليل نظام تطوير تقييم الأداء الحكومي الذي تم إعداده من قبل ديوان الموظفين بغية تحقيق الموضوعية في عملية التقييم. 4. من جهة ثانية، تولى عبد الهادي الأغا مهام مسئولية مديرية أوقاف خان يونس بعد انتهاء خدمة مديرها السابق الشيخ سليمان الفرا خلال فترة حافلة بالانجازات والأنشطة المختلفة. 5. ودعا أبو شعر الأغا إلى ضرورة العمل على تقييم الوضع الحالي ورسم الخطط المستقبلية التي تضمن نجاح العمل والاستمرار به , مشدداً على أهمية وضع السياسيات العامة لاستكمال المشوار الدعوي والعمل على تطوير العمل الإداري للارتقاء بالمستوى المهني والوظيفي لدى الموظفين.	"الأوقاف" تنظم ورشة عمل لتقييم أداء الموظفين 1. نظمت وزارة الأوقاف والشئون الدينية, أمس، ورشة عمل حول "تقييم أداء الموظفين" بالتنسيق والتعاون مع ديوان الموظفين العام , بحضور ومشاركة وكيل الوزارة د .عبد الله أبو جربوع، ومدير عام ديوان الموظفين م .سمير مطير، ومدير عام السياسات والتخطيط بالديوان فدوان أبو شريعة، إلى جانب لفيف من مدراء ورؤساء أقسام الوزارة. 2. وأكد على ضرورة التجديد في العمل لأنه يمثل النجاح من خلال الإبداع والبعد عن الروتينية , لافتاً إلى أن الإدارة الناجحة لابد لها من تخطيط ومتابعة وتقييم للأداء لضمان نجاحها. 3. جاء ذلك خلال لقاء جمع كلاً من وزير الأوقاف الدكتور طالب أبو شعر ووكيل وزارة الأوقاف د . عبد الله أبو جربوع، ومدير عام الشئون الإدارية محمد أبو عسكر، وعبد الهادي الأغا والشيخ سليمان الفرا

P	R	F
0.333	0.2	0.25

Average F
0.25

4.4.3. Comparison between AATSS and Sakhr online summarizer

A comparison between 55 political document summaries is done using both AATSS and Sakhr. For each human we compute the f-measure for each 55 document he summaries as we done in table 4.5 to table 410. Then we compute the average f-measure for this human by dividing the total f-measure by 55. Finally we compute the average f-measure between two humans.

Table 4.11 shows the F-measures for the two systems and two human references.

Table 4.11. F-measure results

	Human reference 1	Human reference 2	Average F-measure
AATSS	87.4%	84.7%	**86.5%**
Sakhr	76.57%	74.09%	**75.33%**

Figure 4.2 to Figure 4.4 represent the difference between P, R and F-measures

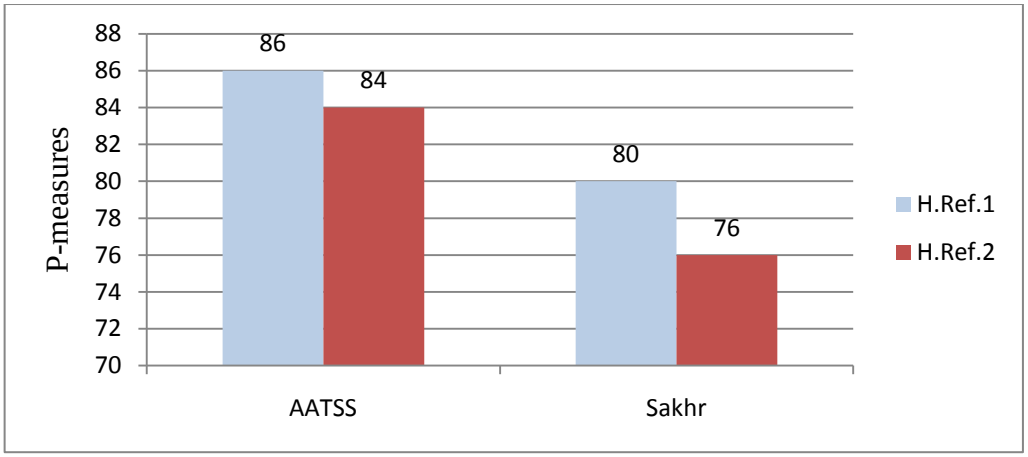

Figure 4.2. P- measure differences

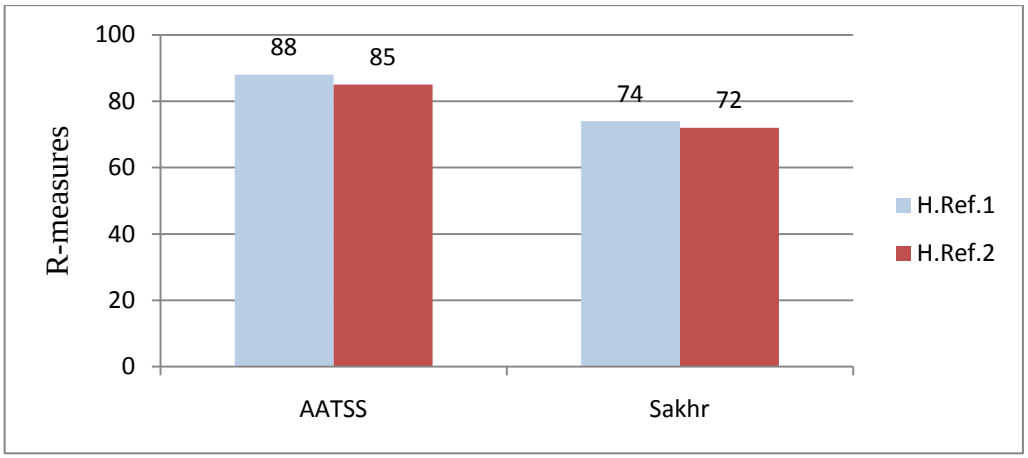

Figure 4.3. R-measure differences

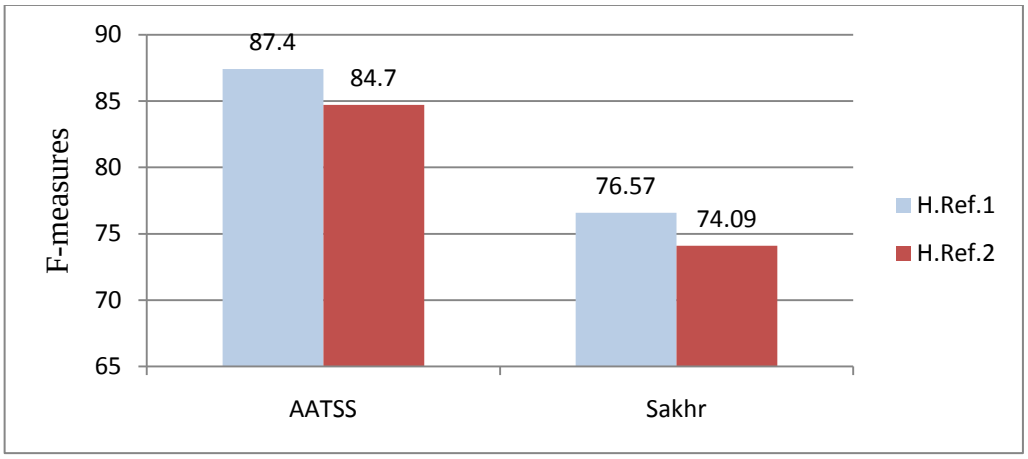

Figure 4.4. F-measure differences

We compare AATSS with Sakhr free online summarizer which is not the best version of Sakhr summarizer. In this free summary there is not a special entity recognition system and it does not support semantic Arabic synonyms. According to default criteria Sakhr free summarizer does well for the short text not long one. Comparing AATSS with the commercial version of Sakhr will have another f-measures value and another comparison value.

4.4.4. Discussion

From the previous experiments and comparisons we can find that:

1. AATSS is outperforming Sakhr summarizer based on F- measure, this is due to :

    a. Entity Recognition in AATSS is more specialized than Sakhr

b. Check for strong, middle and weak keywords in the sentences adding a big value for AATSS.

c. In AATSS we deal with semantic TF and semantic similarity for title and sentences.

2. Sakhr deals good with short documents which contain less than 6 sentences whereas in long documents that might exceed 15 sentences, Sakhr‘s performance become worse. This could be because Sakhr does not extract more than 5 sentences from the original text which might decrease the value of the summary.

3. AATSS is very similar to human summarization as it achieves similarity of 86.5% (F-measure).

4. AATSS have some drawbacks as :

a. There is no automatic Arabic entity recognition system for Arabic so user should enter them manually to obtain good summary.

b. As in all Arabic summarization systems, AATSS lacks to Arabic pronoun resolution system which increase the semantic cohesion between sentences. As example, this sentence may be extracted to summary ‘‘وأضاف أنه يجب علينا الإعتناء بنظام التكافل الإجتماعي لتقوية الروابط في المجتمع الفلسطيني’’ while we don’t really know who is the person mentioned by word ‘‘أضاف’’.

Chapter 5

Conclusion and Future Works

5.1 Conclusion

In this thesis, an automatic system for Arabic extractive text summarization has been introduced. The proposed method is called AATSS and is based mainly on sentence weighting and scoring. The system summarizes the text in three sizes based on the original text, if the text size is less than 5 sentences, no summary will be generated, if the size of the text is 5 or more sentences and less than 20 sentences, a 60% of the original text will be generated as a summary, otherwise, 50% of the text will be generated as a summary.

Four stages are introduced in AATSS which are sometimes interrelated with each other, these stages are: data acquisition, pre-processing and feature extraction, scoring sentence and finally generates summary.

Our AATSS deals with extracted semantic information which increase the capability by giving weights to sentences that have words with the same meaning. On the other hand this process will increase the cohesion between sentences in the extracted summary. Also, an entity recognition module is applied on the documents to identify the most importance sentences.

To evaluate AATSS we used two experiments. The first experiment is a comparison between our system and two human beings. A manual technique which computes precision, recall and F measures between the system and two human summarizers has been conducted. The overall F measure was 86.5% similarity between them. In the second experiment, we compare AATSS with online Sakhr summarizer. We found that AATSS is more powerful than Sakhr as the last only achieve 75.3% similarity with humans.

Finally, our system is optimized, easy to use, general to any domain area and able to produce summaries comparable to human generated summaries. We expect the system to be used for a wide range of applications.

5.2 Future work

We should apply a number of suggested techniques to enhance the system:

[1] Selecting more features like adding semantic information from comprehensive lexical resource such as WordNet, but for Arabic language, may enhance output cohesion and help in feature selection.

[2] One problem with extracted sentences, they may contain anaphora links to the rest of the text. This has been investigated by [38]. Several heuristics have been proposed to solve this problem such as including

the sentence just before the extracted one. Anaphora solving seems to be interesting point of research.

[3] Integration between scoring method and classification algorithms like Bayesian classifiers [44] has the advantage of being fast and simple and the results were good enough. We may try using a multi-classifier system (MCS); this may increase system complexity and may enhance the results.

[4] Adopting alternative techniques for evaluation will help better understanding the nature of the summarization problem. For example: testing the system performance for accomplishing another task such as question answering or document classification. A major research area is developing an automatic evaluation for Arabic summaries.

[5] Developing an automatic Arabic entity recognition system. English language work in this context can be useful to develop a similar system for Arabic language.

[6] Developing an Arabic pronoun resolution system which increases the semantic cohesion. This process can be done using other developed systems in other languages.

References

1. Abdallah, M., Aloulou, C. and Belguith, L., "Toward a Platform for Arabic Automatic Summarization", in the proceeding of the International Arab Conference on Information Technology (ACIT'2008), Jordan, (2008).
2. Aha, D., Kilber, D. and Albert, M., "Instance-Based Learning Algorithms", Kluwer Academic Publishers, Vol. 6, pp. 37-66, (1991)
3. Al-Hashemi, R., "Text Summarization Extraction System (TSES) Using Extracted Keywords", International Arab Journal of e-Technology, Vol. 1, No. 4,pp. 164-168,(2010)
4. Al-Shammari, E and Lin, J., "Towards an Error-free Arabic stemming", in the proceedings of the 2nd ACM workshop on Improving non English Web Searching, California, USA, pp. 9 – 16, (2008).
5. Aone, C.; Okurowski, M.; Gorlinsky, J. and Larsen, B., "A Trainable Summarizer with Knowledge Acquired from Robust NLP Techniques.", in I. Mani and M. Maybury (editors) in Advances in Automatic Text Summarization, MIT Press, pp. 17 – 80, (1999).
6. Barzilay, R. and Elhadad, M., "Using Lexical Chains for Text Summarization", in the proceedings of the ACL workshop on Intelligent Scalable Text Summarization, Madrid, pp. 10 – 17, (1997).
7. Barzilay, R. and McKeown, K. and Elhadad, M., "Information Fusion in The Context of Multi-document Summarization", in the Proceedings of the 37th annual meeting of the association for computational Lingustics on Computational Lingustics, College Park, Maryland, Association for Computational Linguistics, pp. 550 – 557, (1999).
8. Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., and Hullender, G., "Learning to Rank using Gradient Descent", in the Proceedings of the 22nd international conference on Machine learning, NewYourk, USA, pp. 89–96, (2005)
9. Bawakid, A. and Oussalah, M., "A Semantic Summarization System: University of Birmingham at TAC 2008", in the Proceedings of the Text Analysis Conference, Gaithersburg, Maryland, USA , (2008).
10. Baxendale, P., "Machine-made Index for Technical Literature: an Experiment", IBM journal of Research and Develoment, vol. 2, no. 4, pp. 354 – 361, (1958).
11. Conroy, J. and O'leary, D., "Text Summarization Via Hidden Markov", in the proceedings of the 24th annual International ACM SIGIR conference on Research and Development in Information Retrievel, Louisiana; USA, pp. 406 – 407, (2001).
12. Das, D. and Martins, A., "A Survey on Automatic Text Summarization Literature Survey for the Language and Statistics", II Course at CMU, (2007).
13. Diab, M., Jurafsky, D. and Hacioglu, K., "Automatic Processing of Modern Standard Arabic Text", Arabic Computational Morphology, Springer Netherlands, vol. 38, pp. 159 – 179, (2007).
14. Douzidia, F.S. and Lapalme, G., "Lakhas, an Arabic Summarization System",in the Proceeding.of 2004 Document Understanding Conference, Boston, MA (2004).

15. Edmundson, H., “New methods in automatic extracting”, Journal of ACM (JACM), vol. 16, no. 2, pp. 264 – 285, (1969).
16. Elabbas, B., “Perspectives on Arabic Linguistics XIX; Papers from the Nineteenth Annual Symposium on Arabic Linguistics”, John Benjamin's Publishing Company”, Urbana, (2005)
17. El-Haj, M., Kruschwitzc, U. and Fox, CH., “Experimenting with Automatic Text Summarization for Arabic”, in the Proceedings of the 4th conference on Human language technology, Berlin, (2009)
18. Evans, D., “Similarity-based Multilingual Multi-Document Summarization”, Technical Report CUCS-014-05, Columbia University, New York, USA, (2005).
19. Goldstein, J., Mittal, V., Carbonell, J. and Kantrowitz, M., “Multi-document Summarization by Sentence Extraction”, in the Proceedings of the 2000 NAACL-ANLP Workshop on Automatic Summarization, Seattle, Washington, pp. 40-48, (2000).
20. Gong, Y. and Liu, X., “Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis” , In the proceedings of Special Interest Group on Information retrieval, SIGIR, ACM, pp. 19–25, (2001)
21. Haddad, B. and Yassen, M., “A Compositional Approach towards Semantic Representation and Construction of ARABIC”, Logical Aspects of Computational Linguistics, Berlin / Heidelberg: Springer, pp. 147 – 161, (2005).
22. Hassel, M. and kth, N., “Automatic Text Summarization Evaluation. A Survey of Methods and Tools”, GSLT Information Access Course.
23. Hovy, E. and Lin, CH., “Automated Text Summarization and the SUMMRIST System”, in the proceeding of TIPSTER 98 workshop, Baltimore, Maryland, (1998)
24. Ježek, K. and Steinberger, J., “Automatic Text Summarization (The State of the Art 2007 and New Challenges).”, in the proceeding of Document Understanding Confernce (DUC), Rochester, New York USA ,(2007).
25. Khoja, Sh. and Garside, R., “Stemming Arabic text”, Computer Science Department, Lancaster Uni- versity, Lancaster, UK, http://www.comp.lancs.ac.uk/computing/users/khoja/stemmer.ps, (1999).
26. Kupiec, J., Pedersen, J. and Chen, F., “A Trainable Document Summarizer”, in the proceedings of the 18th annual international ACM SIGIR conference on Research and Development in Information Retreivel, New York; USA, pp. 68 – 73, (1995).
27. Larocca Neto, J., Santos, A. D., Kaestner, C. and Freitas, A., “Document Clustering and Text Summarization”, in the Proceeding of 4th International Conference. Practical Applications of Knowledge Discovery and Data Mining (PADD-2000), London, pp. 41-55, (2000).
28. Lin, Ch., “Training a Selection Function for Extraction”, in the proceedings of the 8th international conference on information and knowledge management, Kansas City; USA, pp. 55 – 62, (1999)
29. Lin, C., “Rouge: A Package for Automatic Evaluation of Summaries”. In the Proceedings of the ACL-04 Workshop, Spain, (2004).

30. Lin, C. and Hovy, E., “Identifying Topics by Position”, in the proceedings of the 15th conference on Applied Natural Language Processing, Morristown, NJ, USA, pp. 283 – 290, (1997).
31. Litvak, M., Lipman, H., Ben Gur, A., Last, M., Kisilevich, S. and Keim, D., “Towards Multi-lingual Summarization: A Comparative Analysis of Sentence Extraction Methods on English and Hebrew Corpora”, in the Proceedings of the CLIA/COLING 2010. Beijing, China, (2010).
32. Luhn, H., “The Automatic Creation of Literature Abstracts”, IBM journal of Research and Develoment, vol. 2, no. 2, pp. 159 – 165, (1958).
33. Mann, C. and Sandra, A., “Rhetorical Structure Theory: A Theory of Text Organization”, pp. 87-190, (1987).
34. Marcu, D., “Improving Summarization Through Rhetorical Parsing Tuning”, in the proceedings of the 6th workshop on very large corpa, Montreal; Canada, pp. 206 – 215, (1998).
35. McKeown, K., Klavans, J., Hatzivassiloglou, V., Barzilay, R. and Eskin, E., “Towards Multidocument Summarization by Reformulation: Progress and Prospects”, in the Proceedings of the AAAI/IAAI, Orlando, Florida, United States, American Association for Artificial Intelligence, pp. 453 – 460, (1999).
36. McKeown, K. and Radev, D., “Generating Summaries of Multiple News Articles”, Washington, United States , ACM, pp. 74 – 82, (1995).
37. Osborne, M., “Using Maximum Entropy for Sentence Extraction”, Association for Computational Linguistics, vol. 4, pp. 1 – 8, (2002).
38. Paice, C., “Constructing Literature Abstracts by Computer: Techniques and Prospects”, Information processing and management, Vol. 26, pp. 171-186, (1990)
39. Radev, D. R., Jing, H. and Budzikowska, M., “Centroid-based Summarization of Multiple Documents: Sentence Extraction, Utility-based Evaluation, and User Studies”, in the Proceedings of the NAACL-ANLP 2000 workshop on Autmatic Summarization, Association for Computational Linguistics, pp. 21 – 30, (2002).
40. Radev, D. and Mckeown, K., “Introduction to the Special Issue on Summarization”, Computational Liguistics, vol. 28, no. 4, pp. 339 – 408, (2002).
41. Saravanan, M., Ravindran, B. and Raman, S., “Improving Legal Document Summarization Using Graphical Models”, in Proceeding of the 2006 conference on Legal Knowledge and Information Systems, JURIX, pp. 51-60, (2006)
42. Salton, G. and Buckley, C. “Term-weighting Approaches in Automatic Text Retrieval”, Information Processing and Management, Vol. 24, pp. 513-523. (1988).
43. Sekine, S. and Nobata, C., “Sentence Extraction with Information Extraction Techniques”, In the Proceeding of ACM SIGIR'01 Workshop on Text Summarization. New Orleans, pp.1115-1129, (2001).
44. Sobh, I., Darwish, N. and Fayek. M. "An Optimized Dual Classification System for Arabic Extractive Generic Text Summarization", in proceedings of the Seventh Conference on Language Engineering, ESLEC. (2007).

45. Sobh, I., Darwish, N. and Fayek. M. "A Trainable Arabic Bayesian Extractive Generic Text Summarizer", in the proceedings of the Sixth Conference on Language Engineering, ESLEC, pp. 49-154, (2006).
46. Steve, J., Stephen, L. and Gordon, W., "Interactive Document Summarization Using Automatically Extracted Key phrases", in the proceedings of the 35th Annual Hawaii International Conference on System Sciences, (2002).
47. Suanmali, L., Salim, N. and Binwahlan, M., "SRL-GSM: A Hybrid Approach on Semantic Role Labeling and General Statistic Method for Text Summarzation", Journal of Applied Science, Vol. 10, N. 3, pp. 166-173, (2010).
48. Svore, K., Vanderwende, L. and Burges, C., "Enhancing Single-document Summarization by Combining RankNet and Third-party Sources", in the proceedings of the EMNLP-CoNLL, Association for Computational Linguistics, (2007).
49. Verma, R. and Chen, P., "A Semantic Free-text Summarization System Using Ontology Knowledge", in the proceeding of the Document Understanding Conference (DUC), Rochester, New York; USA, (2007).

Appendix A

AATSS Interface and Code

A.1. AATSS Interface

The system interface from Figure A.1 to Figure A.5 is shown in this appendix with some explanations, Figure A.1 show the whole system interface with original text and its summary:

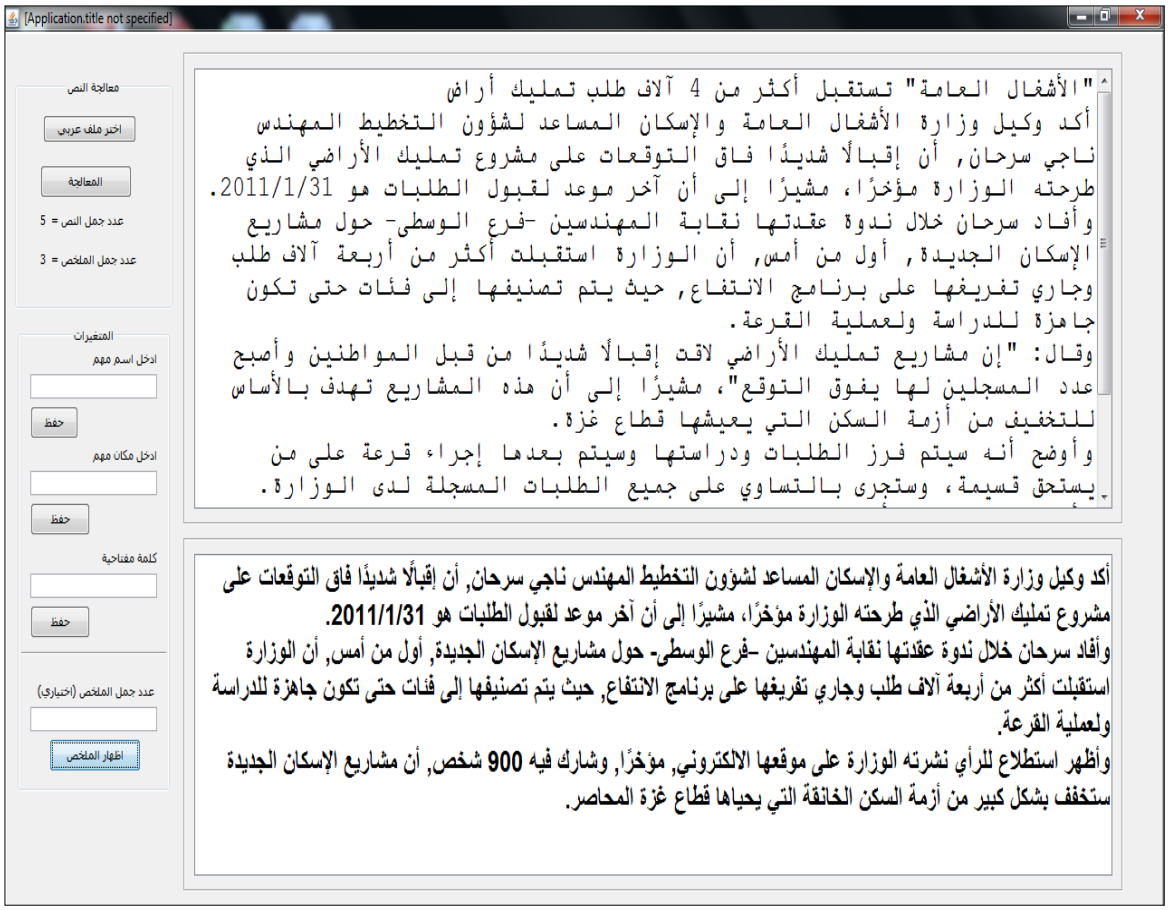

Figure A.1 System interface

Figure A.2 shows the pre-processing part which consists of two buttons. The first button (اختار ملف عربي) used to choose a ".txt" file from the local disk where the second (المعالجة) used to pre-process the chosen file. Two labels are activated in pre-process part after the (المعالجة) button finished its functions. The first displays the number of sentences in the original document and the second label displays the number of the sentences in the summary.

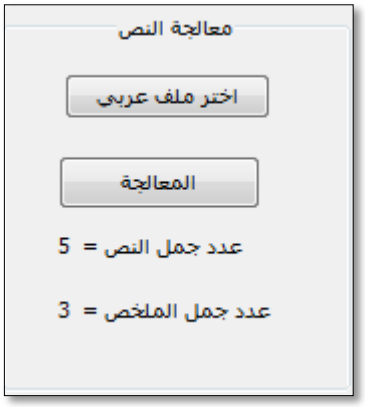

Figure A.2 Pre-process part

Figure A.3 shows the variables panel. All variables here is optional for user, but it's more recommended to enter the first three variables – which used to enter name entity, place entity and special keywords- to obtain a good summary. The final parameter is used to determine the number of sentences in the summary; if the user ignores it then the system will use the default numbers of sentences mentioned in section 4.3.

Figure A.3 Variables panel

Figure A.4 displays the text area for the original text and Figure A.5 displays the summary text area.

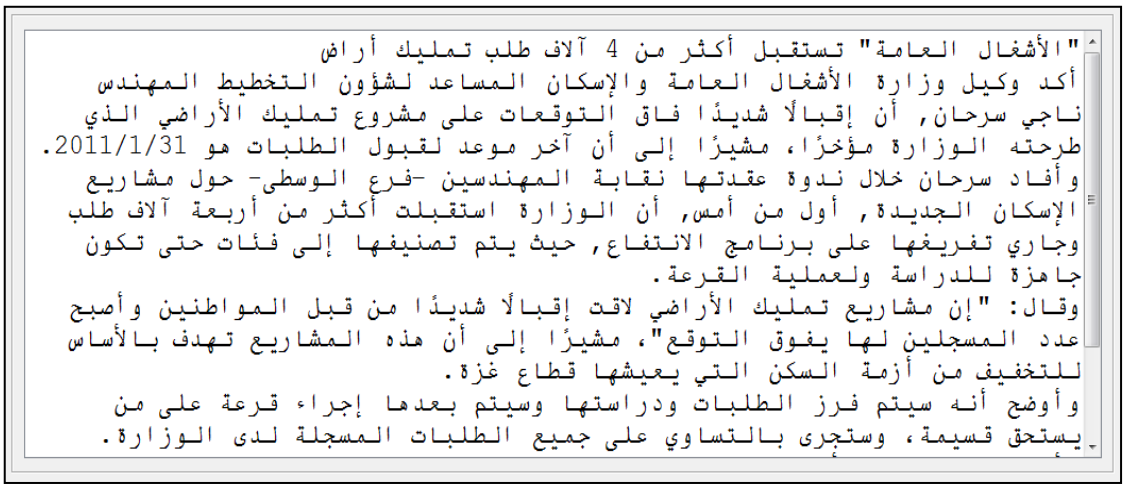

Figure A.4 Original text area

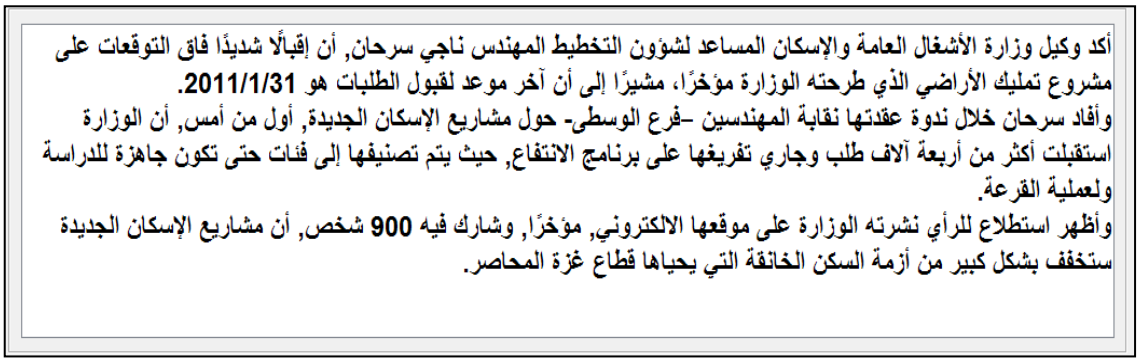

Figure A.5 Summary text area

A.2. AATSS Code

Table A.1 to A.5 display parts of code using to implement the model

Table A.1 Calculate Term Frequency

```
private void TRF(){
   StringBuffer word = new StringBuffer();
   String onelines="";
   Statement stmt = null;
   Statement stmtup = null;
   ResultSet rs = null;
   ResultSet rs2 = null;
   String SQl ="";
   String SQlup ="";
   int sent_id=0;
```

```
try
{
   stmt = SummarizationApp.con.createStatement();
   stmtup = SummarizationApp.con.createStatement();
   SQl = "SELECT sent_id,STEM_SENT FROM FEATURES WHERE
            NEWS_ID = "+SummarizationView.Doc_ID;
   rs = stmt.executeQuery(SQl);
   SQlup = "SELECT max(sent_id) FROM features where news_id =
              "+SummarizationView.Doc_ID;
   rs2 = stmtup.executeQuery(SQlup);
   rs2.next();
   int max = rs2.getInt(1);
   while(rs.next())
   {
      sent_id++;
      onelines = rs.getString(2);

      for(int i=0;i<onelines.length();i++)
      {
         if(!Character.isWhitespace(onelines.charAt(i)))
         {
            word.append ( onelines.charAt ( i ) );
         }
         else
         {
           if ( word.length ( ) != 0 )
               {
                  SQlup = "SELECT word,freq,num_sent,last_id FROM TermF
                             where word = '"+word.toString()+"'";
                  rs2 = stmtup.executeQuery(SQlup);
                  double tf=0;
                  if(rs2.next())
                  {
                     int freq = rs2.getInt(2);
                     int num_sent = rs2.getInt(3);
                     int last_id = rs2.getInt(4);
                     if(freq ==3)
                        System.out.print(word.toString());
                     if(sent_id == last_id)
                     {
                        tf = ++freq *(Math.log1p(max/num_sent));
                        SQlup= "UPDATE TermF SET freq = "+(freq)+", tf =
                                  "+tf+" where word = '"+word.toString()+"'";
                     }
                     else
                     {
                        tf = ++freq *(Math.log1p(max/++num_sent));
                        SQlup= "UPDATE TermF SET freq = "+(freq)+",
```

```
                                    num_sent = "+(num_sent)+", last_id =
                                    "+sent_id+", tf = "+tf+" where word =
                                    '"+word.toString()+"'";
                            }


                          }
                          else
                          {
                            tf =Math.log1p(max/1);
                            tf= 1 *(tf);
                            SQlup = "insert into termf (word,freq,num_sent,last_id,tf)
                                     values('"+word.toString()+"',1,1,"+sent_id+","+tf
                                     +")";
                          }

                          stmtup.executeUpdate(SQlup);
                         word.setLength ( 0 );
                       }
                    }
                  }
                }

              }
              catch(SQLException ex)
              {
                 System.out.println("SQl Exception : "+ex.getMessage());
              }


            }
```

Explanation	This function is responsible for computing how many times each word in the document are appeared also with its semantic vocabularies. Each word Frequency counter is saved into SQL table to use this value in the sentence term frequency as in table A.2.

Table A.2 Sentence Term Frequency code

```
private void SemTRF(){
   StringBuffer word = new StringBuffer();
   String onelines="";
   Statement stmt = null;
   Statement stmtup = null;
   Statement stmtvoc = null;
   ResultSet rs = null;
   ResultSet rs2 = null;
```

```
            String SQl ="";
            String SQlup ="";
            int sent_id=0;
            double total =0;
            double max = 0;
            Vector sent = new Vector();

           try
            {
               stmt = SummarizationApp.con.createStatement();
               stmtup = SummarizationApp.con.createStatement();
               stmtvoc = SummarizationApp.con.createStatement();
               SQl = "SELECT sent_id,STEM_SENT FROM FEATURES WHERE
NEWS_ID = "+SummarizationView.Doc_ID;
               rs = stmt.executeQuery(SQl);
              while(rs.next())
               {
                  sent_id++;
                  onelines = rs.getString(2);

                  for(int i=0;i<onelines.length();i++)
                  {
                     if(!Character.isWhitespace(onelines.charAt(i)))
                     {
                        word.append ( onelines.charAt ( i ) );
                     }
                     else
                     {
                       if ( word.length ( ) != 0 )
                           {
                              sent.addElement(word.toString());

                              int founded =0;
                              String SQLvoc = "SELECT F_VOC,S_VOC,TH_VOC FROM
VOCABULARY WHERE F_VOC = '"+word.toString ( )+"' OR S_VOC =
'"+word.toString ( )+"' OR TH_VOC = '"+word.toString ( )+"'";
                              ResultSet rsvoc = stmtvoc.executeQuery(SQLvoc);
                              while(rsvoc.next())
                              {
                                 if((!rsvoc.getString(1).equals(word.toString())))
                                    {
                                       founded =0;
                                       for(int cc =0;cc<sent.size();cc++)

if(sent.elementAt(cc).toString().equals(rsvoc.getString(1)))
                                              founded ++;
                                       if (founded == 0)
                                        sent.addElement ( rsvoc.getString(1) );
```

```
                    founded =0;

                }
                if((!rsvoc.getString(2).equals(word.toString())))
                {
                    founded =0;
                    for(int cc =0;cc<sent.size();cc++)

            if(sent.elementAt(cc).toString().equals(rsvoc.getString(2)))
                        founded ++;
                    if (founded == 0)
                     sent.addElement ( rsvoc.getString(2) );

                    founded =0;
                }

if((!rsvoc.getString(3).equals(word.toString()))&&(!rsvoc.getString(3).equal
s("")))
            {
                founded =0;
                for(int cc =0;cc<sent.size();cc++)

                    if(sent.elementAt(cc).toString().equals(rsvoc.getString(3)))
                        founded ++;
                    if (founded == 0)
                     sent.addElement ( rsvoc.getString(3) );

                    founded =0;

                }
            }
            word.setLength ( 0 );
          }
        }

      }

      for(int cc =0;cc<sent.size();cc++)
      {
          SQlup = "SELECT TF FROM TermF where word =
                '"+sent.elementAt(cc).toString()+"'";
        rs2 = stmtup.executeQuery(SQlup);
         if(rs2.next())
           {
            total = total + rs2.getDouble(1);

           }

      }
```

```
                if(max<total)
                   {
                      max = total;
                   }

                SQlup = "UPDATE FEATURES SET Sem_TRF = "+total+" WHERE
                 SENT_ID = "+sent_id+" AND NEWS_ID =
                 "+SummarizationView.Doc_ID;
                stmtup.executeUpdate(SQlup);
                total =0;
                sent.removeAllElements();
              }

              SQl = "SELECT Sem_TRF,sent_id FROM FEATURES WHERE
                        NEWS_ID = "+SummarizationView.Doc_ID;
              rs = stmt.executeQuery(SQl);

             while(rs.next())
             {
                 double synt_trf=(rs.getDouble(1)/max);
                SQlup = "UPDATE FEATURES SET Sem_TRF = "+synt_trf+" WHERE
                          SENT_ID = "+rs.getString(2)+" AND NEWS_ID =
                          "+SummarizationView.Doc_ID;
                stmtup.executeUpdate(SQlup);
               // total =0;
             }


           }
           catch(SQLException ex)
           {
              System.out.println("SQl Exception : "+ex.getMessage());
           }
        }
```

Explanation	This function is responsible for computing sentence Term frequency by using each word frequency value stored before in table A.1.

Table A.3 Remove Diacritics Code

```
        // remove diacritics from the word
        public boolean removeDiacritics ( String currentWord, StringBuffer
modifiedWord ) {
              boolean diacriticFound = false;
              modifiedWord.setLength ( 0 );
              Vector diacritics = ( Vector ) PreProcessFiles.elementAt ( 3 );
```

```
        for ( int i = 0; i < currentWord.length ( ); i++ )
           // if the character is not a diacritic, append it to modified word
           if ( ! ( diacritics.contains ( currentWord.substring ( i, i+1 ) ) ) )
              modifiedWord.append ( currentWord.substring ( i, i+1 ) );
           else
           {
              diacriticFound = true;
           }
        return diacriticFound;
     }
```

Explanation	This function is responsible for removing all diacritics from each word in the Arabic text.

Table A.4 Check Strong words Code

```
private void ChkStrongWords() {
      try
      {
         Statement st = SummarizationApp.con.createStatement();
         Statement stmt = SummarizationApp.con.createStatement();
         String stem="";
         String SQLs= "SELECT STEM_SENT,SENT_ID FROM FEATURES
                       WHERE NEWS_ID = "+ Doc_ID;
         ResultSet rss = stmt.executeQuery(SQLs);
         while (rss.next())
         {
            stem = rss.getString(1);
            Scanner scanner = new Scanner(stem);
            String SQL = "SELECT ID,STRONG FROM WORDS where strong
                          != '0'";
            ResultSet rs = st.executeQuery(SQL);
            String strongfind ="";
            int strongfinds =0;
            while(rs.next())
            {
              strongfind = scanner.findInLine(rs.getString(2));
              if(strongfind != null)
              {
                 strongfinds++;
              }

            }

            SQL = "UPDATE FEATURES SET CHKSTRONG =
                  "+strongfinds+" WHERE SENT_ID = "+rss.getInt(2)+"
                  and news_id = "+ Doc_ID;
```

```
                    st.executeUpdate(SQL);
                    strongfinds =0;

                }
        }
            catch(SQLException ex)
            {
               System.out.println(ex.getMessage());
            }
          }
```

Explanation	This function is responsible for checking all occurs of strong key-phrase which we described in section 3.2.7. after counting all occurs of strong word in the sentence. An update query is applied on feature SQL table to store the value of the strong word counter as a new features.

Table A.5 Compute Score Code

```
        public void score()
        {
           try
           {
              Statement stmt = SummarizationApp.con.createStatement();
              Statement stmtup = SummarizationApp.con.createStatement();
              String SQL = "select * from features where news_id =
"+SummarizationView.Doc_ID;
              String SQLup="";
              ResultSet rs = stmt.executeQuery(SQL);
              //ResultSet rsup =null;
              while(rs.next())
              {
                 double score  = 0;
                 score                                                    =
rs.getDouble(7)+rs.getDouble(8)+rs.getDouble(9)+rs.getDouble(10)+rs.getDouble(11)+
rs.getDouble(12)+rs.getDouble(13);
                 score = score + rs.getDouble(24)+rs.getDouble(26)+rs.getDouble(22);
                 score = score + rs.getDouble(14)+rs.getDouble(15)+rs.getDouble(16);
                 SQLup = "UPDATE FEATURES SET SCORE = "+score+" WHERE
NEWS_ID = "+SummarizationView.Doc_ID+" AND SENT_ID = "+rs.getString(3);

stmtup.executeUpdate(SQLup);//rs.getDouble(25)+rs.getDouble(21)++rs.getDouble(23)

              }
           }
           catch(SQLException ex)
           {
              System.out.println(ex.getMessage());
```

} }	
Explanation	This function is responsible for linear summation of all features stored in SQL table to generate a sentence final weights for each sentence in the Arabic document.

Appendix B

AATSS Examples

In this appendix, some examples for AATSS with various domains

Table B.1. AATSS summary for economic text (1)

Original text	
إدخال دفعة سيارات حديثة وبضائع لغزة قررت سلطات الاحتلال الإسرائيلي، الخميس 4-8-2011، فتح معبر كرم أبو سالم جزئياً لإدخال شاحنات محملة بالضائع والمساعدات، إضافة إلى دفعة جديدة من المركبات الحديثة إلى قطاع غزة . وأوضح رئيس لجنة تنسيق إدخال البضائع للقطاع رائد فتوح أن سلطات الاحتلال سمحت بإدخال (290-300 شاحنة) عبر معبر كرم أبو سالم، محملة بالمساعدات والبضائع للقطاعين التجاري والزراعي وقطاع المواصلات . وأضاف أن من ضمن الشاحنات 21 مركبة حديثة، وعشر شاحنات أسمنت وحديد بناء و64 شاحنة حصمة لمشاريع وكالة الغوث وتشغيل اللاجئين الفلسطينيين "الأونروا"، إضافة إلى ضخ كميات محدودة من غاز الطهي . وكانت سلطات الاحتلال سمحت، الأربعاء، بإدخال 274 شاحنات من ضمنهم 83 تجارة، 30 قمح ، 55 زراعة، 32 مساعدات . فيما تم أيضاً إدخال (45 شاحنة) حصمة لمشاريع وكالة الغوث، و14 شاحنة حصمة و6 أسمنت وحديد بناء لمشاريع"UNDP" ، وتسع شاحنات من حمولة أسطول الحرية محملة بزوايا حديد وبلاستيك وألواح جبس، كما تم ضخ 183060 كيلو غاز طهي.	
Name entity	رائد فتوح
Place entity	قطاع غزه – كرم سالم
Key words	وكاله غوث – لجنه تنسيق ادخال بضائع
Number of summary sentences	4
Summary text	
إدخال دفعة سيارات حديثة وبضائع لغزة قررت سلطات الاحتلال الإسرائيلي، الخميس 4-8-2011، فتح معبر كرم أبو سالم جزئياً لإدخال شاحنات محملة بالضائع والمساعدات، إضافة إلى دفعة جديدة من المركبات الحديثة إلى قطاع غزة. وأوضح رئيس لجنة تنسيق إدخال البضائع للقطاع رائد فتوح أن سلطات الاحتلال سمحت بإدخال (290-300 شاحنة) عبر معبر كرم أبو سالم، محملة بالمساعدات والبضائع للقطاعين التجاري والزراعي وقطاع المواصلات. وأضاف أن من ضمن الشاحنات 21 مركبة حديثة، وعشر شاحنات أسمنت وحديد بناء و64 شاحنة حصمة لمشاريع وكالة الغوث وتشغيل اللاجئين الفلسطينيين "الأونروا"، إضافة إلى ضخ كميات محدودة من غاز الطهي. فيما تم أيضاً إدخال (45 شاحنة) حصمة لمشاريع وكالة الغوث، و14 شاحنة حصمة و6 أسمنت وحديد بناء لمشاريع"UNDP" ، وتسع شاحنات من حمولة أسطول الحرية محملة بزوايا حديد وبلاستيك وألواح جبس، كما تم ضخ 183060 كيلو غاز طهي.	

Table B.2. AATSS summary for sport text

Original text	
فلسطين تودع تصفيات مونديال 2014 ودّع المنتخب الوطني الفلسطيني لكرة القدم تصفيات آسيا المؤهلة لمونديال البرازيل 2014، عندما خرج من لقاء الإياب على استاد الشهيد فيصل الحسيني بالرام، الخميس 28-7-2011، متعادلاً بهدفين لمثلهما أمام منتخب تايلاند الذي نجح في التأهل لدور المجموعات بنتيجة اللقاءين كونه فاز ذهاباً على أرضه (1/0 .(وخيب المنتخب الفلسطيني آمال الجماهير الكبيرة التي ملأت المدرجات، والتي كانت تعول على تحقيق إنجاز للرياضة الفلسطينية بالصعود إلى الدور الثالث من تصفيات المونديال لأول مرة في تاريخ مشاركات فلسطين، ولكن الرياح لم تجر بما تشتهي السفن، وغاب التوفيق عن مهاجمي الفريق رغم الأفضلية التي رسمها اللاعبون مع الربع ساعة الأولى عندما تقدم المنتخب بهدف مراد عليان (6)، بعد فاصل من التمريرات المتقنة بين إسماعيل العمور وفهد عتال ليختتمها عليان في الشباك، ومع اعتماد المنتخب على انطلاقات عليان وعرضيات العمور بقي الهجوم الفلسطيني أخطر، ولكن الأمور تحولت مع الثلث ساعة الأخيرة للشوط الأول فامتلك التايلنديون المبادرة وهددوا مرمى محمد شبير، ليحرز داتساكورن تونقلاو هدف التعادل في الدقيقة (33) أنهى بها نتيجة الشوط الأول . الحكم البحريني صالح العباسي أشهر البطاقة الصفراء لكل من مراد عليان وإسماعيل العمور، وأصبحت الأفضلية لتايلاند بعد الهدف وتراجع أداء الوطني، بداية الشوط الثاني شهدت خروج فهد عتال ودخول علي الخطيب، ومن تمريرة العمور العرضية استقبلها مراد عليان وسددها بقوة ليبعدها الدفاع إلى خارج الملعب، وتواصل الضغط الفلسطيني في محاولة للبحث عن هدف الذي لم يكن حينها كافياً للتأهل في المقابل حافظ دفع تايلاند على شباكه، مما اضطر المدير الفني موسى بزاز لاستبدال أشرف نعمان بالمهاجم إياد أبو غرقود، وسليمان العبيد مكان إسماعيل العمور . وفي الوقت الذي غادرت فيه معظم الجماهير الفلسطينية من أرض الملعب نجح مراد عليان في الدقيقة (90) من إحراز الهدف الثاني بعد تسديدة قوية سكنت الشباك، مما منح الوطني بصيصاً من الأمل، وفي اللحظات الحرجة من عمر اللقاء أشهر الحكم البطاقة الحمراء لأحمد محاجنة بعد تدخل قوي على الكرة، واحتسب ركلة حرة مباشرة سددها داتساكورن تونقلاو بقوة لتسكن الشباك هدف التعادل لتايلاند، لتتبخر حينها آمال فلسطين وتنطلق أفراح تايلاند بالصعود لدور المجموعات . تشكيلة المنتخب الفلسطيني:" محمد شبير (حارس مرمى)، أحمد محاجنة، عمر جعرون، عبد اللطيف البهداري، خالد مهدي، إسماعيل العمور، حسام وادي، أشرف نعمان، خضر يوسف، مراد عليان، فهد عتال.	
Name entity	محمد شبير - أحمد محاجنة- عمر جعرون- عبد اللطيف البهداري- خالد مهدي- إسماعيل العمور- حسام وادي- أشرف نعمان- خضر يوسف- مراد عليان- فهد عتال – فيصل الحسيني - داتساكورن تونقلاو – صالح عباسي - موسى بزاز – اياد غرقود – سليمان عبيد
Place entity	اسيا – برازيل- تايلاند- فلسطين - رام
Key words	مونديال
Number of summary sentences	3
Summary text	
فلسطين تودع تصفيات مونديال 2014 ودّع المنتخب الوطني الفلسطيني لكرة القدم تصفيات آسيا المؤهلة لمونديال البرازيل 2014، عندما خرج من لقاء الإياب على استاد الشهيد فيصل الحسيني بالرام، الخميس 28-7-2011، متعادلاً بهدفين لمثلهما أمام منتخب تايلاند	

الذي نجح في التأهل لدور المجموعات بنتيجة اللقاءين كونه فاز ذهاباً على أرضه (1/0.(

وخيب المنتخب الفلسطيني آمال الجماهير الكبيرة التي ملأت المدرجات، والتي كانت تعول على تحقيق إنجاز للرياضة الفلسطينية بالصعود إلى الدور الثالث من تصفيات المونديال لأول مرة في تاريخ مشاركات فلسطين، ولكن الرياح لم تجر بما تشتهي السفن، وغاب التوفيق عن مهاجمي الفريق رغم الأفضلية التي رسمها اللاعبون مع الربع ساعة الأولى عندما تقدم المنتخب بهدف مراد عليان (6)، بعد فاصل من التمريرات المتقنة بين إسماعيل العمور وفهد عتال ليختتمها عليان في الشباك، ومع اعتماد المنتخب على انطلاقات عليان وعرضيات العمور بقي الهجوم الفلسطيني أخطر، ولكن الأمور تحولت مع الثلث ساعة الأخيرة للشوط الأول فامتلك التايلنديون المبادرة وهددوا مرمى محمد شبير، ليجرز داتساكورن تونقلاو هدف التعادل في الدقيقة (33) أنهى بها نتيجة الشوط الأول.

الحكم البحريني صالح العباسي أشهر البطاقة الصفراء لكل من مراد عليان وإسماعيل العمور، وأصبحت الأفضلية لتايلاند بعد الهدف وتراجع أداء الوطني، بداية الشوط الثاني شهدت خروج فهد عتال ودخول علي الخطيب، ومن تمريرة العمور العرضية استقبلها مراد عليان وسددها بقوة ليبعدها الدفاع إلى خارج الملعب، وتواصل الضغط الفلسطيني في محاولة للبحث عن هدف الذي لم يكن حينها كافياً للتأهل في المقابل حافظ دفع تايلاند على شباكه، مما اضطر المدير الفني موسى بزاز لاستبدال أشرف نعمان بالمهاجم إياد أبو غرقود، وسليمان العبيد مكان إسماعيل العمور.

Table B.3. AATSS summary for economic text (2)

Original text	
طوكيو الأغلى وتونس ثاني أرخص مدينة عالمية مازالت طوكيو المدينة الأكثر غلا ء في العالم للعيش فيها، وجاءت بالإضافة إلى أوسلو النرويجية وأوساكا اليابانية ضمن المدن الخمسة الأولى الأكثر غلاء، الأمر المثير أن تونس احتلت المركز الثاني من أسفل السلم، أي أنها جاءت بين ارخص مدن العالم . غير أن المسح الأخير لأكثر الأماكن غلاء في العالم، والصادر عن مركز المعلومات لمجلة الإيكونوميست، كشف عن تغييرات جذرية خلال العام الماضي . وأصبحت أستراليا من بين أكثر الدول الصاعدة، جراء قوة عملتها الدولار الأسترالي أمام العملات الأخرى، ما أدى إلى رفع مستويات المعيشة في سيدني، التي احتلت المركز السادس عالمياً من حيث الغلاء، ثم ملبورن التي تلتها وحلت سابعة، ثم بعدها بست مراكز جاءت مدينة بيرث واحتلت المركز 13 وجاءت بعدها مباشرة مدينة بريسبين . أوروبياً، احتلت العاصمة الفرنسية باريس المركز الرابع، ثم زيوريخ السويسرية في المركز الخامس ففرانكفورت الألمانية في المركز الثامن وجنيف في المركز التاسع، واحتلت سنغافورة المركز العاشر بين أغلى مدن العالم . وبيّن المسح حجم التحولات خلال العقد الأخير، حيث كانت هونغ كونغ تحتل المركز الثالث عالمياً، ولكنها باتت في المركز 22 حالياً، أما شنغهاي فكانت في المركز 16 ولكنها تراجعت إلى المركز 48 أما بكين فكانت في المركز 11 لتتراجع إلى المركز 64 . وتحتل نيويورك الأمريكية المركز 49 بين أغلى المدن . من بين أكثر المدن صعوداً على سلم الغلاء جاءت مدينة بودابست البلغارية وقفزت 17 مركزاً حيث تحتل حالياً المركز 76 عالمياً . أما أكثر المدن تراجعاً من حيث الغلاء فهي اسطنبول التركية التي تراجع من المركز 24 إلى المركز 52 . أما أرخص خمس مدن في العالم فهي، كراتشي وتحتل المركز 133، وتونس وتحتل المركز 132 ومومباي وتحتل المركز 131 وطهران وتحتل المركز 130 ونيودلهي وتحتل المركز 129.	
Name entity	
Place entity	طوكيو – تونس – اوسلو – اوساكا - استراليا – سيدني – ملبورون – هونغ كونغ – شنغهاي – نيويورك-

	بوادبست – اسطنبول – كراتشي – مومباي – طهران – نيودلهي – بريسبين – بريث – باريس- زيوريخ – فرانكفورت – جنيف – سنغافورة.
Key words	مركز المعلومات لمجلة الإيكونوميست
Number of summary sentences	5

Summary text

طوكيو الأغلى وتونس ثاني أرخص مدينة عالمية

مازالت طوكيو المدينة الأكثر غلا ء في العالم للعيش فيها، وجاءت بالإضافة إلى أوسلو النرويجية وأوساكا اليابانية ضمن المدن الخمسة الأولى الأكثر غلاء، الأمر المثير أن تونس احتلت المركز الثاني من أسفل السلم، أي أنها جاءت بين ارخص مدن العالم.

وأصبحت أستراليا من بين أكثر الدول الصاعدة، جراء قوة عملتها الدولار الأسترالي أمام العملات الأخرى، ما أدى إلى رفع مستويات المعيشة في سيدني، التي احتلت المركز السادس عالمياً من حيث الغلاء، ثم ملبورن التي تلتها وحلت سابعة، ثم بعدها بست مراكز جاءت مدينة بيرث واحتلت المركز 13 وجاءت بعدها مباشرة مدينة بريسبين.

أوروبياً، احتلت العاصمة الفرنسية باريس المركز الرابع، ثم زيوريخ السويسرية في المركز الخامس ففرانكفورت الألمانية في المركز الثامن وجنيف في المركز التاسع، واحتلت سنغافورة المركز العاشر بين أغلى مدن العالم.

وبيّن المسح حجم التحولات خلال العقد الأخير، حيث كانت هونغ كونغ تحتل المركز الثالث عالمياً، ولكنها باتت في المركز 22 حالياً، أما شنغهاي فكانت في المركز 16 ولكنها تراجعت إلى المركز 48 أما بكين فكانت في المركز 11 لتتراجع إلى المركز 64.

أما أرخص خمس مدن في العالم فهي، كراتشي وتحتل المركز 133، وتونس وتحتل المركز 132 ومومباي وتحتل المركز 131 وطهران وتحتل المركز 130 ونيودلهي وتحتل المركز 129.

Table B.4. AATSS summary for political text

Original text

واشنطن تضغط على نتنياهو لاستئناف المفاوضات

ذكرت صحيفة "هآرتس" العبرية، الاثنين 13-6-2011، أن واشنطن تمارس ضغطا على رئيس الوزراء الإسرائيلي بنيامين نتنياهو لقبول المقترح الأمريكي بشأن استئناف المفاوضات الفلسطينية الإسرائيلية، وفقا لما جاء فى خطاب الرئيس الأمريكي باراك أوباما فى التسع عشر من الشهر الماضي .

ونقلت الصحيفة في تقرير أوردته فى موقعها الإلكتروني عن إسحاق مولهو المبعوث الشخصي لنتنياهو بعد لقائه كبار المسئولين الأمريكيين الأسبوع الماضي في واشنطن قوله "هناك خيبة أمل أمريكية إزاء موقف رئيس الحكومة الإسرائيلية بأنه يعرقل مساعي الولايات المتحدة للحيلولة دون صدور قرار أممي يعترف بدولة فلسطينية فى سبتمبر المقبل ."

وقالت الصحيفة إن الأمريكيين أبلغوا مولهو بأنه لعرقلة المبادرات الأوروبية ولاسيما الاقتراح الفرنسي بعقد مؤتمر دولي للسلام في باريس يتعين على الإسرائيليين التقدم بشيء ملموس مثل موافقة نتنياهو على التفاوض على أساس المقترح الأمريكي الذي جاء به خطاب أوباما .

ونقل مصدر إسرئيلى عن أحد المسئولين الأمريكيين قوله "إن نتنياهو يطلب دعما أمريكيا له فى سبتمبر.. ولكنه لا يعطينا أي أدوات نساعده بها فى هذا الصدد وبدلا من أن يساعدنا يجعل الأمر أكثر صعوبة ."

وتابعت الصحيفة أنه نتيجة لذلك اشتكى مسئولون أمريكيون من أن أوباما لم يستطع الحصول على تعهد بريطاني

وفرنسي بمعارضة الطلب الفلسطيني بشأن دولة فلسطينية خلال جولته الأوروبية الشهر الماضى .

وأشارت "هاآرتس" إلى أن وزير الخارجية الألماني جيدو فيستر فيله سيلتقي مع نتنياهو غدا الثلاثاء لتحذيره من أنه إذا ما استمر هذا المأزق الدبلوماسى، ستضطر ألمانيا إلى جانب كل من بريطانيا وفرنسا لدعم مشروع قرار في c يقر بالمقترح الأمريكي كأساس لاستئناف المحادثات بين الفلسطينيين والإسرائيليين.

Name entity	جيدو فيستر فيله – نتنياهو - اوباما – بنيامين نتنياهو - إسحاق مولهو
Place entity	ألمانيا – بريطانيا – فرنسا – باريس – واشنطن – ولايات متحده
Key words	هاآرتس – مجلس امن
Number of summary sentences	4

Summary text

واشنطن تضغط على نتنياهو لاستئناف المفاوضات

ذكرت صحيفة "هاآرتس" العبرية، الاثنين 13-6-2011، أن واشنطن تمارس ضغطا على رئيس الوزراء الإسرائيلي بنيامين نتنياهو لقبول المقترح الأمريكي بشأن استئناف المفاوضات الفلسطينية الإسرائيلية، وفقا لما جاء فى خطاب الرئيس الأمريكي باراك أوباما فى التاسع عشر من الشهر الماضي.

ونقلت الصحيفة في تقرير أوردته فى موقعها الإلكتروني عن إسحاق مولهو المبعوث الشخصي لنتنياهو بعد لقائه كبار المسئولين الأمريكيين الأسبوع الماضي في واشنطن قوله "هناك خيبة أمل أمريكية إزاء موقف رئيس الحكومة الإسرائيلية بأنه يعرقل مساعي الولايات المتحدة للحيلولة دون صدور قرار أممي يعترف بدولة فلسطينية فى سبتمبر المقبل."

وقالت الصحيفة إن الأمريكيين أبلغوا مولهو بأنه لعرقلة المبادرات الأوروبية ولاسيما الاقتراح الفرنسي بعقد مؤتمر دولي للسلام في باريس يتعين على الإسرائيليين التقدم بشىء ملموس مثل موافقة نتنياهو على التفاوض على أساس المقترح الأمريكي الذي جاء به خطاب أوباما.

وأشارت "هاآرتس" إلى أن وزير الخارجية الألماني جيدو فيستر فيله سيلتقي مع نتنياهو غدا الثلاثاء لتحذيره من أنه إذا ما استمر هذا المأزق الدبلوماسى، ستضطر ألمانيا إلى جانب كل من بريطانيا وفرنسا لدعم مشروع قرار في مجلس الأمن يقر بالمقترح الأمريكي كأساس لاستئناف المحادثات بين الفلسطينيين والإسرائيليين.

Table B.5. AATSS summary for technology text

Original text

النت السريع بغزة قريباً

اعتمد وزير الاتصالات وتكنولوجيا المعلومات الدكتور يوسف المنسي، 13-3-2011، المقترح المعدل الذي تقدمت به شركة الاتصالات الفلسطينية "بال تل" لتقديم خدمة الانترنت عبر النفاذ المباشر "Bitstreem" اعتبارا من مطلع حزيران المقبل بعد الوصول إلى أفضل الشروط التي تحقق مصلحة المواطنين .

وأوضح المنسي في بيان صحفي وصل "فلسطين أون لاين نسخة عنه، أنه تم الإيعاز لشركة الاتصالات بالبدء بتنفيذ المشروع فوراً وذلك بعد انتهاء اللجنة المكلفة من دراسة المشروع والتشاور مع كافة القطاعات المجتمعية وتقديم توصياتها بهذا الشأن .

وأشار المنسي إلى أن الوزارة حرصت خلال فترة الدراسة والتشاور التي استغرقت أربعة أشهر على تحقيق أكبر قدر من الشروط الايجابية لصالح المواطنين وتلافي أية عيوب وسلبيات قد تنجم عن التسرع في تطبيق هذه الخدمة .

ونبه المنسي إلى أن الوزارة تُمارس دورها الفاعل في الرقابة على شركة الاتصالات بشقيها الخلوي والثابت

والانترنت، مضيفاً "أن الوزارة تسعى بشكل حثيث لضمان أفضل الخدمات للمواطنين وبما يضمن عدم تعرضهم للاستغلال والاحتكار أو لرداءة الخدمات ."

وأشاد المنسي بجهود اللجنة المكلفة بدراسة المشروع وعملها الدؤوب لانجاز المشاورات في أقل وقت ممكن والوصول إلى أفضل المواصفات لتطبيق خدمة الانترنت عبر النفاذ المباشر "B.S.A" دون أية مشاكل أو سلبيات قد يُعاني منها المواطن مستقبلاً .

يذكر أن الوزير المنسي كلف لجنة برئاسة المهندس سهيل مدوخ مدير عام الحاسوب الحكومي، وعضوية كلاً من م. جلال إسماعيل مدير عام ديوان الوزير، م. مروان حماد مدير عام التراخيص وآخرين من الوزارة .

وتشاورت اللجنة مع ممثلي مؤسسات المجتمع المدني ومزودي خدمة الانترنت والشركات لتقديم المقترحات بشأن تطبيق الخدمة بجودة عالية وتخفيض الأسعار، وخلصت إلى عدد من التوصيات التي التزمت بها الشركة وبناء عليه تقرر البدء بتطبيق الخدمة اعتباراً من مطلع حزيران يونيو القادم .

وأشار المنسي إلى أن مُستخدمي الانترنت سيتمكنون من التمتع بسرعات عالية وتكلفة محدودة وصولاً بالخدمة إلى مستويات عالية مميزة.

Name entity	يوسف منسي – سهيل مدوخ – مروان حماد – جلال اسماعيل
Place entity	
Key words	شركة اتصالات – تكنولوجيا معلومات - انترنت
Number of summary sentences	5

Summary text

<u>النت السريع بغزة قريباً</u>

اعتمد وزير الاتصالات وتكنولوجيا المعلومات الدكتور يوسف المنسي، 13-3-2011، المقترح المعدل الذي تقدمت به شركة الاتصالات الفلسطينية "بال تل" لتقديم خدمة الانترنت عبر النفاذ المباشر "Bitstreem" اعتبارا من مطلع حزيران المقبل بعد الوصول إلى أفضل الشروط التي تحقق مصلحة المواطنين.

وأوضح المنسي في بيان صحفي وصل "فلسطين أون لاين نسخة عنه، أنه تم الإيعاز لشركة الاتصالات بالبدء بتنفيذ المشروع فوراً وذلك بعد انتهاء اللجنة المكلفة من دراسة المشروع والتشاور مع كافة القطاعات المجتمعية وتقديم توصياتها بهذا الشأن.

ونبه المنسي إلى أن الوزارة تُمارس دورها الفاعل في الرقابة على شركة الاتصالات بشقيها الخلوي والثابت والانترنت، مضيفاً "أن الوزارة تسعى بشكل حثيث لضمان أفضل الخدمات للمواطنين وبما يضمن عدم تعرضهم للاستغلال والاحتكار أو لرداءة الخدمات."

وتشاورت اللجنة مع ممثلي مؤسسات المجتمع المدني ومزودي خدمة الانترنت والشركات لتقديم المقترحات بشأن تطبيق الخدمة بجودة عالية وتخفيض الأسعار، وخلصت إلى عدد من التوصيات التي التزمت بها الشركة وبناء عليه تقرر البدء بتطبيق الخدمة اعتباراً من مطلع حزيران يونيو القادم.

وأشار المنسي إلى أن مُستخدمي الانترنت سيتمكنون من التمتع بسرعات عالية وتكلفة محدودة وصولاً بالخدمة إلى مستويات عالية مميزة.