إقـــــرار

أنا الموقع أدناه مقدم الرسالة التي تحمل العنوان:

تلخيص تلقائي للنصوص العربية

Automatic Arabic Text Summarization

أقر بأن ما اشتملت عليه هذه الرسالة إنما هو نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه حيثما ورد، وإن هذه الرسالة ككل أو أي جزء منها لم يقدم من قبل لنيل درجة أو لقب علمي أو بحثي لدى أي مؤسسة تعليمية أو بحثية أخرى.

DECLARATION

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualification

Student's name:	اسم الطالب/ة: اياد جهاد الفرا
Signature:	التوقيع:
Date:	التاريخ: 30 / 01 / 2016

The Islamic University Gaza
Higher Education Deanship
Faculty of Engineering
Computer Engineering
Master of Computer Engineering

غزة – الإسلامية الجامعة
العليا الدراسات عمادة
الهندسة كلية
هندسة الحاسوب
ماجستير هندسة الحاسوب

تلخيص تلقائي للنصوص العربية

Automatic Arabic Text Summarization

Submitted by:

Eng. Eyad Elfarra

Supervised by:

Prof. Mohammed Mikki

A Thesis Submitted in Partial Fulfillment of Requirements for the Degree of Master in Computer Engineering Department in Islamic University of Gaza-IUG

1437 هـ -2015 م

بسم الله الرحمن الرحيم

الجامعة الإسلامية – غزة
The Islamic University - Gaza

مكتب نائب الرئيس للبحث العلمي والدراسات العليا
هاتف داخلي 1850

الرقم Ref ...../35/غ

التاريخ Date ...../2016/01/18

نتيجة الحكم على أطروحة ماجستير

بناءً على موافقة شئون البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة على تشكيل لجنة الحكم على أطروحة الباحث/ **اياد جهاد رفيق الفرا** لنيل درجة الماجستير في **كلية *الهندسة*** قسم **هندسة الحاسوب** وموضوعها:

تلخيص تلقائي للنصوص العربية

Automatic Arabic Text Summariztion

وبعد المناقشة التي تمت اليوم الاثنين 08 ربيع الآخر 1437هـ، الموافق2016/01/18م الساعة الواحدة ظهراً، اجتمعت لجنة الحكم على الأطروحة والمكونة من:

أ.د. محمد أمين مكي	مشرفاً و رئيساً	.....
د. علاء مصطفى الهليس	مناقشاً داخلياً	.....
أ.د. سامي سليم أبو ناصر	مناقشاً خارجياً	.....

وبعد المداولة أوصت اللجنة بمنح الباحث درجة الماجستير في **كلية *الهندسة*/** قسم وموضوعها: **هندسة الحاسوب.**

واللجنة إذ تمنحه هذه الدرجة فإنها توصيه بتقوى الله ولزوم طاعته وأن يسخر علمه في خدمة دينه ووطنه.

والله ولي التوفيق،،،

نائب الرئيس لشئون البحث العلمي والدراسات العليا

أ.د. عبد الرؤوف علي المناعمة

P.O. Box 108, Rimal, Gaza, Palestine fax: +970 (8) 286 0800 فاكس Tel: +970 (8) 286 0700 ص.ب. 108 الرمال، غزة، فلسطين هاتف
public@iugaza.edu.ps www.iugaza.edu.ps

DEDICATION

To my great father and my great mother

To my wife and my son Mohannad

To my borthers and sisters

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. Mohammad Mikki*** *for giving me the opportunity to work with him and guiding and helping me throughout this research and other courses.*

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.

Very special thanks to ***my Father*** *&* ***My mother*** *for all things you do for me, your pray, patience, motivation and continues support.*

Thanks to ***my wife,*** *for here support and help to complete this thesis.*

Thanks to my ***brothers and sisters*** *for your motivation and support.*

Thanks to all my friends; whom I consider as brothers.
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.

ملخص الدراسة

اللغة العربية تعتبر من اللغات الأكثر شهرة على مستوى العالم، تنبع أهميتها من كونها اللغة الخامسة حول العالم من حيث عدد المتحدثين بها.

إنشاء تلخيص جيد للمستند النصي يعتبر واحد من أهم الفروع في علم اللغويات. التلخيص الجيد للنص يعطي القارئ الأجزاء المهمة مما يوفر عليه الوقت والجهد المبذول في عملية القراءة.

هناك بعض التقنيات المستخدمة في عملية تلخيص النصوص العربية ولكنها قليلة العدد وتحتاج لبعض التحسينات. أحد الطرق المستخدمة هي الطريقة المبنية على القاعدة الرسومية، ولكنها ما زالت بحاجة لبعض التحسينات.

في هذه الأطروحة تم بناء خوارزمية جديدة مبنية على أساس رسومي بياني حيث يتم تمثيل كل جملة في النص بعد معالجتها بنقطة ثم توصيل النقاط مع بعضها وحساب وزن العلاقات بين جميع النقاط الموجودة في الرسم. معالجة النصوص تتم بناءا على قواعد معالجة اللغات الطبيعية ومن ثم يتم ترتيب الجمل طبقا لخوارزمية ترتيب الصفحات التابعة لشركة جوجل.

الوحدات الأساسية المستخدمة في المعالجة في هذه الرسالة هي ثلاث وحدات: وحدة التجذير الكلي، التجذير الخفيف، وعدم استخدام التجذير. بعد المعالجة تم اختيار 40 % من جمل النص الأساسي كجمل في الملخص.

عملية التلخيص تتم هنا عبر 12 خطوة تبدأ بعملية جمع البيانات، المعالجة القبلية، تقسيم النص لوحدات أساسية، التجذير، إزالة الكلمات الشائعة، بناء النص على شكل رسم، حساب قيمة العلاقات بين الجمل، تطبيق خوارزمية الترتيب وأخيرا عملية استخلاص وانشاء الملخص.

تم فحص النظام باستخدام مجموعة من البيانات المجمعة تسمى (EASC) ومن ثم تم استخدام معايير التقييم التالية (Recall , Precision , F-measure) في عملية تقييم النظام.

النتائج أظهرت أن استخدام الوحدة الناتجة عن التجذير الكلي أعطت نتائج أفضل من باقي الوحدات تلتها الوحدة الناتجة عن استخدام التجذير وفي النهاية الوحدة المبنية على استخدام التجذير الخفيف.

ABSTRACT

Arabic language is one of the most famous languages in the world; its importance comes from being the fifth language that has native speakers in the world.

Creating a good summary of the text is one of the most important things in the linguistics because it gives the user the most important paragraphs in the text that he wants to read.

There are some techniques to summarize Arabic language, but they are still little and need to be improved. One approach that are used in text summarization is graph based but it still need enhacment.

This thesis builds a new algorithm called GBATSS (Graph Based Arabic Text Summarizer) to summarize Arabic text depending on NLP and Google page rank algorithm. The system works on three basic units. These units are rooted stem, light stem, and finally no-stem. The system depends on compression ratio of 40 %. The process of summarization is done in 12 stages start from data collection, text preprocessing, text normalization, text tokenization, stemming, stop words removal, building graph, calculating edge weighting, applying page rank, and finally extracting the summary.

Finally, we tested the system using EASC data set and using the recall, precision and f-measure for evaluation process.

The results show that the using of rooted-stem as a basic unit gives the best results then no-stem and finally light-stem

Keywords: *Automatic Text Summarization, Feature Extraction, Summary Evaluation, Natural Language Processing, Google Page Rank, Graph –based Summarization, Arabic language Processing.*

LIST OF CONTENTS

LIST OF ABBREVIATIONS

ACBTSS	Arabic Concept Based single Text Summarization System
ANLP	Arabic Natural Language Processing
AQBTSS	Arabic Query Based single Text Summarization System
ATS	Arabic Text Summarization
CR	Compression Rate
DAG	Directed Acyclic Graph
DUC	Document Universal Conferences
EASC	Essex Arabic Summaries Corpus
EIM	Engineering Information Management
FUF	Functional Unification Formalism
GA	Genetic Algorithm
GP	Genetic Programming
HMM	Hidden Markov Model
IBM	International Business Machines Corporation
IDF	Inverse Document Frequency
IE	Information Extraction
IR	Information Retrieval
ISI	Information Science Institute
JDK	Java Development Kit
ML	Machine Learning

MR	Mathematical Regression
Mturk	Mechanical Turk
MUC	Message Understanding Conferences
NLG	Natural Language Generation
NLP	Natural Language Processing
P	Precision
PF	Path Finder
POS	Part of speech
PR	Page Rank
QA	Question Answering
R	Recall
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
RST	Rhetorical Structure Theory
SBD	Sentence Boundary Disambiguation
SOM	Self Organizing Map
SURGE	Systemic Unification Realization Grammar of English
SVM	Support Vector Machines
TF	Term Frequency
TF-IDF	Term Frequency × Inverse Document Frequency
TF-ISF	Term Frequency × Inverse Sentence Frequency
TS	Text Summarization

URI	Universal Resource Identifier
WF	Word Frequency
WWW	World Wide Web

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

In digital world; there are a huge amount of data, the demand for automatic text summarizer is increasing rapidly. Text summarization is the process of creating short or compressed version of a given text automatically. This version of text provides the user useful data about the original text.

Text summarization main goal is reducing the number of paragraphs and statements in the document as possible as we can to give the readers a useful meaning to decide if the document is useful or not. Text summarization is apart from information retrievals so here we want to talk about information retrievals, text summarization and Arabic language[32].

1.1 Information Retrievals (IR)

Information retrieval is the process of retrieving information resources relevant to an information need from a collection of information resources.

1.2 Natural Language Processing

Natural language processing (NLP) is a field of computer science, related to the area of human–computer interaction. Many challenges in NLP involve natural language understanding, that is, enabling computers to derive meaning from human or natural language input, and others involve natural language generation [1].

Preprocessing is the process of preparing data for the core text-mining task. This process converts the documents from original data source into a format, which is suitable for applying various types of feature extraction methods against these documents to create a new collection of documents fully represented by concepts. The preprocessing phase includes all those routines, processes and methods required to prepare data for a text mining system, which is the core of knowledge discovery operations.

There are many usages of NLP in the technical world. Recently, it has been used with conventional Information Retrieval (IR) search engines to help users navigate quickly through retrieved documents without the need (sometimes) to open the retrieved text [1]. Other NLP applications could also benefit from automatic text summarization (TS), such as: Information Extraction (IE) [2], Text Classification (TS)[3], Question Answering (QA) [4], Natural Language Generation (NLG) [5] and Engineering Information Management (EIM) [6].

1.3 Text Summarization(TS)

The NLP community has explored the subfield of summarization since nearly 1950s [7]. Luhn in[32] Define a summary as text produced from one or more texts that transfers important information in the original text, and that is no longer than half of the original text. Yielded text will make users to decide if this document is important to them or not and making the decision according to this information. Edward Hovy et al. [8] defines the summary as “a text that is based on one or more texts; it has the most important information of the main texts and its content is less than half of the main texts”. Mani [9], describes the text summarization as “a process of finding the main source of information, finding the main important contents and presenting them as a concise text in the predefined template”.

Radev[10] 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". From Radev definition for text summarization we can deduce that text summarization must have the following characteristics:

- Summaries generated from one or multiple documents.
- Only important sentences should be retrieved.
- Summaries should be short as possible as you can.

The process of text summarization can be decomposed into three phases:

1. The analysis phase: analyzes the input text and selects a few salient features.

2. The transformation phase: transforms the results of analysis into a summary representation.

3. The synthesis phase: takes the summary representation, and produces an appropriate summary corresponding to user’s needs.

There are many ways to do automatic text summarization that are classified under one of the following categories:

1. Natural Language processing (NLP) or in another word linguistics: which include lexical chain, Graph theory and WorldNet [53].

2. The statistical techniques: depends on the number of times where some words are repeated in the document then sentences that appear greater than some threshold take it in our summary. Which includes Aggregation Similarity Method [11], Location Method [12],Frequency Method [13], TF-Based Query Method [14].
3. Machine learning (ML) algorithms: like using neural network, genetic algorithm, SOM etc.

Some researchers used a hybrid technique by using one or more technique from these techniques.

These techniques of text summarization are done depending on some features ,which are used in process of weighting the sentences importance in the summary and to decide if this sentence can be included in the generated summary or not. These features are of some kinds, like statistical based on the frequency of some elements in the text, linguistic extracted from a simplified argumentative structure of the text; or heuristic based on sentence length or position and some other features.

1. **Statistic:**
   Based on the frequency of some unites in the text; which give different information about the relevance of sentences for the summary. These features are sufficiently relevant for the single document summarization task [15], some of those features are listed below:

   a) **Numerical data:**
   Numerical data in sentences may indicate an important data like statics، annual budgets, expenses and other number. User needs to see it, so sentences containing numerical data are scored higher than ones without numerical values.

   b) **Word Frequency (WF) or Term Frequency:**
   Words that are repeated in text more than other words mean that these words are important. In sentences that contains words that are repeated more than other this will give them more importance than others. But we must keep in mind that removing stop words that are repeated more than other words in the text and give no meaning.

2. **Linguistic:**
   Extract summary done by depending on the structure of the text, which assume semantics level of representation of the original text and involve linguistic processing at some level . Some type of linguistics is rhetorical structure.

3. Heuristic:

It is based on sentence length or position and some other features as follows:

a) Position Score: The assumption is that certain genres put important sentences in fixed positions. For example, newspaper articles have the most important terms in the first four paragraphs.

b) Title: Words in the title and in following sentences are important and get high score.

c) Similarities between sentences in an article.

d) Indicative Phrases: Sentences containing key phrases like "we conclude that ……"

e) Sentence Length: The score assigned to a sentence reflects the length of the sentence, normalized by the length of the longest sentence in the text [15].

1.4 Categories of Text Summarization

Text summarization can be categorized into many categories depending on the factor we want to take in our accounts. So summaries can be categorize according to :

- **Type of returned summary :**

  - **Extractive summarization methods:**
    The main objective of the extraction-based summarization: is simply selecting sentences with special characteristics and put them together in a summary. The summarized text is extracted from the original text on a statistical basis or by using heuristic methods or a combination of both. Extraction of sentences is done by extracting important sentences by weighting the sentences statistically or by using heuristics properties such as position information are also used for summarization. For example, extracting sentences that follow the key phrase “in conclusion”. This means that the extracted sentences are not changed [16].

  - **Abstractive summarization:**
    In abstract summary, summarizer attempts to understand the text before generating the summary. After understanding the document; summarizer

expresses the summary in new formulation using new terms and new sentences that are not listed in the original text. For example, the phrase “He ate banana, orange and pear” can be summarized as “He ate fruit” [16].

- **Input factor:**

  - **Single document :**
    Extract the most important information from a single document that is supplied to the system as a single summary.

  - **Multiple documents:**
    Extract a single summary from multiple documents with the same topic that will be supplied to the system.

- **Output form of the summary(details or brief)**

  - **Indicative summary:**
    Give the reader a brief idea of the text which should contain the most important points in the document. Goal of this summary is to help the user to decide whether the original document is helpful for reading or not. This is useful in generating the summary of URI that are retrieved from the search engine. It uses compression ratio between 5-10%.

  - **Informative summary:**
    It is larger than indicative, which return more detailed information from the original text. This type of summarization is useful in the process of generating the summary of news feed. It uses compression ratio between 20-30%.

  - **Critical or evaluative summary:**
    Summary of summary that returns author's point of view on a given subject[17].

- **Summary contents :**

  - **Generic summaries:**
    Give general facts of the text with general information that make all major topics in the test in the same importance level.

- **Query-based summaries:**
  Summaries content are generated depending on user needs**.** User enters some topics or terms he wants to know about; then the summarizer generates the summary.

- **User focused:**
  User interest in Math so the content of generated summary will be around this user focus need.

- **Update Summary:**
  Answer the question "what is new?". It take input stream of document and return sub stream of documents. This process is done by tracing the new information that are flaw in the system. The system in this type assuming that the user has read the previous documents (not the previous summaries of those documents).

- **Input/output languages :**

  - **Mono-lingual:**
    The input and output language are the same and there is only one language.

  - **Multilingual:**
    The summarization system can deal with multiple languages. Therefore, the input languages and the output languages are the same in the two documents.

  - **Cross lingual:**
    Input language is differ from the output language in the summarization system.

Extractive methods are usually performed in three steps[18] :

a. **Create an intermediate representation of the original text:**
   In this step, we create a representation of the document by dividing the text into paragraphs, sentences, and tokens. Sometimes we need to perform some preprocessing techniques, such as stop words removal, diacritics removals etc.

b. **Sentence scoring:**
   In this step, we give every sentence a score that describes importance in the document .This score is giving depending on some measure and on the relevancy between sentences in the given documents.

c. **Selecting a summary consisting of several sentences:**
This step combines the scores provided by the previous steps and generates a summary.

1.5 Text Summarization Fields

In our daily life, many fields need text summarization such as:

- **Commercial and advertisement field:**
There are millions of products in the market. Every product has its own description. When the advertiser wants to advertise his products, they need a few number of words to describe his products. Here ATS is needed.

- **News area**:
Every second people publish thousands of political, sport, economic and other types of news. It's very hard or impossible to browse all of them or either the quarter. By text summarization user can find which news he is concerned with reading the whole text.

- **Legal area**:
We have little amount of legal documents. Legal experts time is very expensive. To make legal experts do well they must provided with a summarized document. Therefore automatic text summarization systems will help the legal experts to find compressed and restated content of relevant judicial documents, including laws and their proposals, relevant court decisions or tribunal process summarizations [7].

- **Medical area**:
Medicine progressed every day in the undiscovered diseases and surgical tools and way to heal the patients from their disease. So researcher published every year hundreds of document to describe their works. Doctors and medical specialist need to find relevant information about patient‘s conditions timely. So, text summarization here saves time resources and optimize availability of medical experts.

- **Technical and work reports area:**
In technical world, there are thousands of reports generated every day.
Workers in technical branch do not have enough time to read all these reports

and take decision according to it. Therefore, summarization is a good process for helping the technical people to decide if this report is important or not.

1.6 Arabic Language Processing

A huge number of people around the world speak the Arabic Language. It is used as a communication language between Arabs and non-Arab Muslims as well. So Arabic became an important language on the Internet due to the increasing number of Arabic speaking online users seeking Arabic content and applications. Recent figures from the Internet World Statistics show that there are 112 million Internet users from the Arab World.

Because of the flexibility in structure and writing, Arabic language became a complex language, so when we want to make stemming to it to get its root it will be difficult.

The Arabic language has some in its history, internal structure, strong relationship with Islam, culture and identity. Any Arabic NLP (Natural Language Processing) system that does not take into its account the features of the Arabic language will be inapplicable [19] [20]. Arabic NLP applications must deal with several complex problems relevant to the nature and structure of the Arabic language. Here are some characteristics of the Arabic language:

- Arabic is written from right to left. Like Hebrew, Persian, and Korean.
- There are no capital letters or small letters in Arabic.
- Its letters change shape according to the position of the letter in the word (beginning, middle, or the end of the word).
- The complex morphology.
- The absence of diacritics in written text.
- Modern Standard Arabic does not have an Orthographic representation of short letters, which requires a high degree of homograph resolution and word sense clarification.
- Arabic is a pro-drop language, that is allows subject pronouns to drop [21] subject to retrieval of deletion [22].

Another limitation of Arabic NLP is a shortage of Arabic corpora, lexicons and machine-readable dictionaries. In other side, there has been some success in tackling the problem of Arabic syntax as in [23] [24]. There is some research attempted to develop Automatic Arabic summarization systems as [25] . These attempts are listed in the next section.

1.7 Arabic Text Summarization System

In this section, we will present some Arabic text summarization that are implemented before. These examples are as follows:

- **Lakhas**:
  An extractive Arabic text summarization system. It is the first Arabic summarization system to be formally evaluated and compared with English competitors in an evaluation competition [23].

- **AQBTSS**:
  Query-based single document summarizer system. Takes an Arabic document and a query (in Arabic) and attempts to provide a reasonable summary for the document around this query[24].

- **ACBTSS:**
  Integrates Bayesian and Genetic Programming (GP) classification methods. 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. Evaluation of this system is done by comparing generated summaries with human given summaries in terms of recall, precision and F-measure[24].

1.8 Summary Evaluation

There is no single summary that may be the golden standard in text summarization. In other words; there are many summaries that can be generated for every text documents that depends on the human who generates this summary and the educational and technical background of him. By surfing research papers that are concerned in text summarization we find that the human evaluator does not agree on

one summary for any paragraph. Therefore, the evaluation process of text summarization is a difficult process.

Many metrics are used in generating text summarization. Metrics used differ from paper to paper and from project to project. This variousness makes the evaluation of summarization system is quite fair.
Another problem arises that manual evaluation is too expensive: as stated in [26] [27], large scale manual evaluation of summaries as in the Document Universal Conferences (DUC) would require over 3000 hours of human efforts[28].

Summary evaluation methods attempt to determine how the summary is relative to its source. Generally, evaluation methods can be divided depending on many factors.

- **Intrinsic or extrinsic method:**

  - Intrinsic evaluation methods:
    In intrinsic method users judge if the summary is well or not by itself without comparing if it completes some tasks or not. This is done by deciding if the summary covers main idea or not and if the summary is informative or not. Intrinsic evaluations have assessed mainly the coherence and informativeness of summaries.

  - Extrinsic Evaluation methods:
    Users judge a summary 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.

- **Inter-textual and Intra-textual:**

  - Inter textual evaluation methods :
    Inter textual evaluation methods focus on contrastive analysis of outputs of several summarization systems.

  - Intra-textual evaluation methods:
    Intra-textual methods assess the output of a specific summarization system.

There are many measures for calculating text summarization:

- **Recall-Oriented Understudy for Gisting Evaluation (ROUGE):**
  Evaluate summary by counting the number of overlapping units such as word sequences between the computer-generated summary to be evaluated and the

ideal summaries created by humans but this system does not support Arabic evaluation.

- **Recall:**
  Recall is the number of correct sentences divided by the number of results that should have been returned. In binary classification, recall is called sensitivity. So it can be looked at as the probability that a relevant document is retrieved by the query [38].

  Recall alone is not accurate in measuring the performance of the summary. Because if you returned all sentences in the document then recall value will be 100% so we need to use precision to measure the number of non-relevant sentences.

$$Recall = \frac{Retrieved \cap Relevant}{Relevant} \dots\dots\dots\dots\dots\dots\dots\dots\dots eq(1.1)$$

- **Precision:**
  Precision is the number of correct sentences divided by the number of all returned results. In binary classification, precision is analogous to positive predictive value. Precision considers all retrieved documents[29].

$$Precision = \frac{Retrieved \cap Relevant}{Retrieved} \dots\dots\dots\dots\dots\dots\dots\dots\dots eq(1.2)$$

- **F-measure:**
  Using precision only is not accurate. In addition, using recall only is accurate so we use f-measure to make balancing between these two measures.

$$F = \frac{(\beta^2 + 1)\,PR}{\beta^2(P + R)} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots eq(1.3)$$

  Where:
  B: equal one when β is greater than one, Precision is favored, when β is less than one, recall is favored[29].

P: precision.
R: Recall.

1.9 Statement of Problem

The problem of this research is how to develop a model to generate an automatic Arabic text summarization based on extraction methods that can be valid for various domains with high performance.
This main problem can divided into the following sub problems:

1. What is the proper list of stop words we use?
2. What is the proper data corpus for this system?
3. What is the proper field of documents that will show the power of our summary?
4. How many steps of preprocessing are good for us?
5. What is the better stemming technique for our system?
6. What are the best basic units to use in the system?
7. What are most relevant features to be extracted?
8. How to obtain the proper dataset for system testing?
9. How to reconstruct the summary?
10. What is the proper approach for evaluating the summary?

1.10 Objectives

1.10.1 Main Objectives

The main objective of this research is the development of an automatic Arabic text summarization model depending on Google page rank algorithm that deals directly with graph 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 enhancing preprocessing techniques.

1.10.2 Specific Objectives

The specific objectives of this thesis are:

- Using different domains in the experimental process to find the appropriate one.
- Preprocessing Arabic text using various techniques to improve the quality of resulted summary.
- Feature extraction using other research approaches to extract the specific features in processing.
- Develop simple scoring method that will depend on feature weight and page rank algorithm results for giving scores for each sentence.
- Using EASC (Essex Arabic Summaries Corpus) dataset [64] that contains various domains in the evaluation techniques.
- Summary will be extracted depending on the compression ratio that will be 40% in this experimental thesis.
- Implement the proposed summarization model.
- Evaluate the performance of the proposed model by using different measures in the evaluation stage like precision and recall.

1.11 Significance Of The Thesis:

Arabic text summarization is very important technique in the Arabic world for many reasons:

- Support Arabic contents on the Internet.
- Applying Arabic summarization into multiple domain areas.
- Saving time, cost, and efforts by helping Arabic readers.

- Make computer more intelligent, which is useful for spreading computer technology in Arabic world.

1.12 Scope and Limitations

As introduced before, we have many categories of text summarization. In this research, we focus on:

- Extractive text summarization.
- Arabic single document.
- The document tested in some special domain defined by the EASC corpus that is used.
- The compression ratio used here is 40% to make it fear in comparison phase.

1.13 Methodology

In this research we build Arabic text summarization technique depending on graph based algorithm. To do this process we do the following process to get the results:

1. Enter the following values (dumping factor , number of iterations, Compression ratio).
2. Load single document to be summarized.
3. Text Preprocessing.
4. Text Normalization (remove duplication in spaces, extra commas, " ,' , etc….
5. Tokenization by splitting text to lines.
6. Stop Words removals.
7. Stemming, which can be done in three ways (Root, light, no-stem).
8. Calculating the relations between sentences by using TF-ISF (Term Frequency – Inverse Sentence Frequency) with cosine similarity.
9. Building the Text graph.

10. Representation of sentences by a vertices in the graph

11. Calculating the similarity between sentences that represents the weight of edges between graph vertices.

12. Applying Google PageRank algorithm to the graph.

13. According to compression ratio start the process of finding the candidate sentences to be chosen for the summary (Extract the summary).

In the model implementation, we depend mainly on java programming language and using two types of a free source stemmer to stem each Arabic word in the text. For the system interface, the user has an option to assign the compression ratio that will define the number of sentences to be included in the summary and extracted from the text. Also user can decide the type of basic unit of the summarization system.

1.14 Thesis Organization

In the rest of this documentation, is organized as follows. Chapter 2 presents the related work that doing in the text summarization. Chapter 3 presents the methodology that, doing in the text summarization that contains the processes of collecting data, preprocessing, tokenizing and stemming the text. In addition, it contains the process of building the graph, scoring the nodes then extracting the summary. Chapter 4 presents the results and evaluation of GATSS (Graph-based Arabic Text Summarization System) approach. Chapter 5 presents conclusion and future works.

CHAPTER TWO: RELATED WORK

Many types of text summaries have been discussed in the introduction chapter. Depending on the required application and the user's needs, the target summary can be one of the following types: indicative, informative, topic-oriented, generic, an extract, an abstract, a single-document summary or a multiple-document summary.

Many approaches have been done in TS since 1950s [32]. Advances in NLP tools encouraged researchers to process the TS problem using many approaches such as sentence selection and reduction [33], machine learning techniques [15] [34], using an ontology and lexical chains [35] [36]. Another approach is graph-based methods that have been proposed for the single and multi-document summarization for the English documents. This chapter talks about related work that's done previously in text summarization.

2.1 Related Works in Text Summarization

Mani and Bloedorn [37] proposed an automatic procedure to generate extractive text summary. This approach done by using a machine learning on a training corpus of documents and their abstracts to describe the function which finds the combination of features that is optimal of the summarization task. The resulted summary from this approach will be generic or user specified summary. Using this approach; it is easy to obtain reference summaries, even for big document collections.

Luhn, in [32] proposed a text summarization system that depends mainly on the number of occurrence of specific word if this word appears frequently on the text then it will be a significant word. This approach done in the following process . first stemming words to their root forms. Then deleting stop words from the text. After that compiled a list of content words sorted by decreasing frequency, the index providing a significance measure of the word. The significance of the sentence done finding the occurrence of the significant words in the sentence. All sentences are ranked in order of their significance factor, and the top ranking sentences are finally selected to form the summary. This approach considered the first research in the text summarization.

Baxendale [38], at his research that's done at International Business Machines Corporation (IBM) the researcher used the sentence position feature to find important part in the document. Baxendale uses this feature after examining 200 paragraphs to find that in 85% of the paragraphs the topic sentence came as the first one and in 7% of the time it was the last sentence. According to Baxendale it will be fair enough to select a topic sentence into the summary.

Edmundson [39] describes an extractive text summarization. The developed system using four feature in the summarization two features of word frequency and positional importance were introduced in the previous two works. And another two features that are the presence of cue words (presence of words like significant, or hardly), and the skeleton of the document (whether the sentence is a title or heading). Weights were attached to each of these features manually to score each sentence.

Conroy and O'leary in [40] use Hidden Markov Model (HMM) for solving the problem of extracting sentences from the text. 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.

Marcu [41]; propose a unique approach for Automatic Text Summarization. This approach done by using a discourse theory which is Rhetorical Structure Theory (RST) [42]. In his approach he introduces a text tree to measure distinction between what is more essential to the writer purpose than ordinary text.

Svore et al in [43] propose summarization system by using a modified back-propagation two layer networks Neural Network and a set of features that stored in database. The researchers produce news extracts system (NetSum) which extracts the most three significant sentences from the news article. They used a Rank-Net algorithm [44] to classify and extract sentences. Training in NetSum. The performance of NetSum with external features are statistically significant at 95% confidence. The main drawback of this system is the shorten generated summary that is contain only three sentences not more.

Al-Hashemi [45] propose 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 [46] 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.

2.2 Graph-based Text Summarization

Graph based text summarization is one of techniques that used in the text summarization we talk about researches that's done in this way.

Mihalcea [12] proposes a range of graph-based ranking algorithms, and evaluate their application to automatic unsupervised sentence extraction in the context of a text summarization task. The results obtained from this new unsupervised method are competitive with previously developed state-of-the-art systems.

Yeh *et al*. [47] propose an extractive graph-based summarization method called iSpreadRank. iSpreadRank exploits the concept of spreading activation theory to formulate a general concept from social network analysis by taking into consideration, the importance of its connected nodes also. The algorithm recursively reweighs the importance of sentences by spreading their sentence-specific feature scores throughout the network and adjusts the importance of other sentences.

Litvak and Last [48] propose an extractive graph based summarization system using supervised and unsupervised methods, for identifying the keywords to be used in extractive summarization of text documents. Both these approaches are repsresting document text in a syntactic representation, which enhances the traditional vector-space model by taking into account some structural document features like word co-occurrence, size of the co-occurrence window are considered .
In supervised approach, the training phase was done with the help of classification algorithm by using a summarized collection of documents.
In unsupervised approach, HITS algorithm was run on the document graphs under the assumption that the top-ranked nodes should represent the document keywords.

TextRank demonstrated [49] unsupervised extractive summarization system that relies on the application of iterative graph based ranking algorithms to graphs encoding the cohesive structure of a text. The main characteristics of this system is that it does not rely on any language-specific knowledge resources or any manually constructed training data, so it is portable for a new languages or domains. The author shows that

the iterative graph-based ranking algorithms work well on the task of extractive summarization since they do not only rely on the local context of a text unit (vertex), however it takes the information recursively drawn from the entire text (graph) into account.

Li and Cheng [50] propose a novel algorithm, called TriangleSum for single document summarization based on graph theory. The algorithm builds a dependency graph for the document based on syntactic dependency relation analysis. The nodes represent words or phrases of high frequency, and edges represent dependency relations between them. Then, a modified version of clustering coefficient is used to measure the strength of connection between nodes in a graph. By identifying triangles of nodes, a part of the dependency graph can be extracted. At last, a set of key sentences that represent the main document information can be extracted.

Patil and Brazdil [13] propose a single document graph based extractive summary presented by using a theoretic graph technique called SumGraph. The authors have adopted the concept of Pathfinder Network Scaling (PFnet) technique to compute importance of a sentence in the text. Each text is represented as a graph with sentences as nodes while weights on the links represent intra-sentence dissimilarity.

Wan [51] also proposed agraph based ranking algorithm for multiple document summarization he called this algorithm TimedTextRank. The proposed algorithm overcomes the problems in earlier approaches by introducing temporal dimension. From the preliminary study carried out to measure the effectiveness of the proposed TimedTextRank algorithm, it is seen that use of temporal information of documents based on the graph-ranking for dynamic multi-document summarization leads to results that are promising.

Liu *et al*. [52] proposed a multi-document graph based summarization approach. In this system the proposed algorithm work as following :

i. Trains each sentences by making use of the global features provided by the corresponding sentence using Naïve Bayes Model.
ii. Generate a relevance model for each corpus utilizing the query.
iii. calculating the probability for each sentence in the corpus utilizing the salience model.
iv. Based on the probability value it obtains Personalized PageRank ranking process is performed depending on the relationships among all the other sentences.

v. The redundancy penalty is imposed on each sentence. Finally summary sentences are chosen based on information richness with high information novelty.

Eakan and Radev [53] introduce a stochastic graph-based method for computing relative importance of textual units for Natural Language Processing. they test the technique on the problem of Text Summarization (TS). Extractive TS relies on the concept of sentence salience to identify the most important sentences in a document or set of documents. they consider a new approach, LexRank, for computing sentence importance based on the concept of eigenvector centrality in a graph representation of sentences. In this model, a connectivity matrix based on intra-sentence cosine similarity is used as the adjacency matrix of the graph representation of sentences.

Wan [54] create a graph-based summarization algorithm for multi-document summarization under the assumption that all the sentences in the graph model are indistinguishable. This algorithm take into accounts two different aspects. This aspects are the relationship between sentences with each others in the documents as well the document information to globally reflect the importance the theme of the multi-document cluster.

All of previous summarization system make on graph-based theory but non of them support Arabic text summarization.

2.3 Non Arabic Text Summarization Systems

Summ-It applet :

Research project in Surrey University works by extracting sentences using Lexical Cohesion. For more details about this project you can visit this site :
http://www.mcs.surrey.ac.uk/SystemQ/summary

SweSum :

Research project in Royal Institute of Technology (Sweden). This project extracts sentences to produce an extract type summary. It is closely related to the work at Information Science Institute (ISI). Summaries are created from Swedish or English texts in either the newspaper or academic domains. Sentences are extracted by ranking sentences according to weighted word level features and was trained on a tagged Swedish news corpus. The summarization tool can be hooked up to search engine results.
For more details about this project you can visit this site :

http://www.nada.kth.se/~xmartin/swesum/index-eng.html

The Text summarization Project :

Research project in University of Ottawa . Project founders proposed to use machine learning techniques to identify keywords. Keyword identification can then be used to select sentences for extraction. They planned to use surface level statistics such as frequency analysis and surface level linguistic featurs such as sentence position.
For more information of this project you can visit this site:
http://www.site.uottawa.ca/tanka/ts.html

Summarist :

Research project at University of Southern California . Summarist produces summaries of web documents. It has been hooked up to the Systran translation system to provide a gisting tool for news articles in any language. Summarist first identifies the main topics of the document using statistical techniques on features such as position, and word counts. Current reseach is underway to use cue phrases and discourse structure. These concepts must be interpreted so that of a chain of lexically connected sentences, the sentence with the most general concept is selected and extracted. Subsequent work will take these extracted sentences to construct a more coherent summary.For more details :
http://www.isi.edu/natural-language/projects/SUMMARIST.html

SUMMONS:

Research project in Columbia University . Summons is a multi-document summary system in the news domain. It begins with the results of a MUC-style information extraction process, namely a template with instantiated slots of pre-defined semantics. From this, it can generate a summary by using a sophisticated natural language generation stage. This stage was previously developed under other projects and includes a content selection substage, a sentence planning substage and a surface generation stage. Because the templates have well-defined semantics, the type of summary produced approaches that of human abstracts. That is they are more coherent and readable. However, this approach is domain specific, relying on the layout of news articles for the information extraction stage.for more details visit :
http://www.cs.columbia.edu/~hjing/sumDemo

MultiGen:

Rsearch project also in Columbia University.
MultiGen is a multi-document system in the news domain. It extracts sentence fragments that represent key pieces of information in the set of related documents. This is done by using machine learning to group together paragraph sized chunks of text into clusters of related topics. Sentences from these clusters are parsed and the resulting trees are merged together to form, building logical representations of propositions containing the commonly occuring concepts. This logical representation is turned into a sentence using the FUF/SURGE grammar. Matching concepts uses linguistic knowledge such as stemming, part-of-speech, synonymity and verb classes. Merging trees makes use of identified paraphrase rules.
For more details visit :
http://www.cs.columbia.edu/~regina/demo4/

TRESTLE:

Research project in The Sheffield University. This project produces summaries in the news domain. It uses MUC to extract the main concepts of the text which then presumably is used to generated summaries. Unfortunately, not much information is available on the official website regarding the system architecture. For more details visit :
http://www.dcs.shef.ac.uk/research/groups/nlp/trestle/

2.4 Arabic Text Summarization

In [24] the researcher developed two Arabic summarization systems; the first one is Arabic Query-Based Single Text Summarizer System (AQBTSS) that involves an Arabic document and an Arabic query attempting to provide an acceptable summary for the query of this document. The second one is Arabic Concept-Based Text Summarization System (ACBTSS) that takes a set of words representing a certain concept to be the input to the system instead of a user's query. The two systems share first two phases, which are document selection; where the user selects a document that match his/her query from the document collection, and splitting document into sentences.

In [55] the researcher suggested a platform for summarizing Arabic texts, which consists of set of modules: tokenization module, morphological analyzer module, parser module, relevant sentences extraction module, and extract revision module. The evaluation of this platform is carried out on various types of texts (short, average, long) according to execution time, where it noticed that the run time of the modules of

the platform for a given text, depends on the size of this text, i.e. the more the text is short the more its run time is weak.

Sakhr Summarizer is an Arabic Summarizer engine that finds the most relevant sentences of the source text and displays them as a short summary. The Summarization engine employs the Sakhr Corrector to correct the input Arabic text from common Arabic mistakes automatically, and the Keywords Extractor to identify a prioritized list of keywords to identify the important sentences accurately[24] .

Another system for summarization called Arabic Intelligent Summarizer has been proposed in [56]. This system is mainly based on machine supervised learning technique. The system consists of two phases. The first is the learning phase which informs the system how to extract the summary sentences; SVMs are used for the learning process. The second phase is use phase, which allows the users to summarize a new document.

Sobh et al.; in [28] 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).

Hammo in [57] presents a hybrid technique based on text structure and topic identification. This approach focuses on segment extraction and ranking using heuristic methods that assign weighted scores to segments of text. Also, he use a text categorization system and the Arabic WordNet to identify the thematic structure of the input text in order to select the most relevant sentences. then a tokenizer , a

stemmer and other statistical tools are used to identify relevant segments in the text. The source document is segmented into its major units (title, paragraphs and lines) and then, text-lines are interpreted to extract relevant segments for inclusion in the summary.

Abdel Fattah in [58] proposes an text summarization approach based on several features, including sentence position, positive keyword, negative keyword, sentence centrality, sentence resemblance to the title, sentence inclusion of name entity, sentence inclusion of numerical data, sentence relative length, Bushy path of the sentence and aggregated similarity for each sentence to generate summaries. The researcher investigate the effect of each sentence feature on the summarization task. Then he use all features score function to train genetic algorithm (GA) and mathematical regression (MR) models to obtain a suitable combination of feature weights.

Khalifa in [59] present a technique to segment Arabic discourse into complete sentences based on RST. RST is a linguistically useful method for the summarization purpose, by extracting semantics behind the text. As we said this technique is derived from Arabic Rhetorical system by exploiting the main crucial connector "و" , as defined by Arabic linguists almost one thousand years ago. This approach categorizes the six known rhetorical types of "و" into two classes: segment and unsegment, known as, "Fasl" and "Wasl". Segmentation places are decided according to the type of connector "و". A set of twenty-two syntactic and semantic features devised from "Fasl and Wasl" rhetorical methods, are chosen to categorize each type of "و". The system undergoes the learning and testing stages, using SVM machine learning technique to identify the types of the connector "و".

Al-Sanie in [60] attempts to develop an infrastructure for Arabic text summarization based on RST. Also, RST is used in [61] for the summarization purpose by identifying the rhetorical relationship between the paragraphs and extract the most significant paragraphs as a summary. He suggests different techniques, algorithms, and design patterns to be considered when dealing with summarization application.

In [62], the researcher proposed an Arabic text summarization approach based on extractive graph-based approaches. The researcher uses several basic units such as stem, word, and n-gram are applied in the summarization process. The Arabic document is represented as a graph. To extract the summary the researchers used the shortest path algorithm to extract the summary.
The similarity between any two sentences is determined by ranking the sentences according to some statistical features like TF-IDF. The final score is determined for each sentence using PageRank scoring, and finally, the sentences with high scores are

included in the summary considering the compression ratio. The proposed approach is evaluated using EASC corpus, and intrinsic evaluation method.

All of previous systems are extractive text summarization. Some of these are single document but the others are multiple document. In graph based extractive text summarization , we find little works that summarize Arabic language. These systems have low performance in extracting Arabic text summary.

GBATSS tries to make a new approach that depends on graph based with page-rank algorithm and with some modification preprocessing techniques to improve the performance in text summarization.

CHAPTER THREE: THEORETICAL BACKGROUND

In this research, we build summarization system depending on Google Page-Rank algorithm.

3.1 Google Page Rank Algorithm

In such complex networks as World Wide Web, an important attribute of a node is the in-degree (out-degree); namely the number of inbound (outbound) links on the node[30]. The in-degree of a given page could be considered as an approximation of a page's importance or quality [31]. The PageRank algorithm [31] has extended this idea by not counting the inbound links from all pages equally, but by normalizing via both the importance and the number of outbound links of the neighboring pages. In this respect, the Page-Rank value could serve as a better measure of importance, as it incorporates the paper's visibility and authority at the same time by taking both the number of citations and prestige of the citing papers into account [31]. Defined the PageRank of a Web page A, denoted by PR(A), using equation (3.1).

$$PR(A) = (1-d) + d \cdot \sum_{i} \frac{PR(T_i)}{C(T_i)} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots eq(3.1)$$

Where PR (Ti) denotes the PageRank of page Ti which has connection with page A; C (Ti) denotes the number of outbound links on page Ti; and d is a damping factor which can be set between zero and one.

In previous equation, we see that the PageRank of A is recursively defined by the PageRank of those pages that link to page A. Within the algorithm, the PageRank of pages Ti is always weighted by the number of outbound links C(Ti), leading thereby to a smaller PageRank value transferred from pages Ti to the recipient page A. It is also assumed that any additional inbound link to a recipient page A will always increase A's PageRank.
There is a second version of PageRank algorithm shown in equation (3.2).

$$PR(A) = \frac{(1-d)}{N} + d \cdot \sum_{i} \frac{PR(T_i)}{C(T_i)} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots eq(3.2)$$

Where N is the total number of pages on the web. Actually, the second version of the algorithm does not differ largely from the first one. However, it can better explain the

metaphor of the original Random Surfing model suggested by [31], in which the PageRank of a page is conceived as being the probability for a surfer visiting the page after clicking on many links. Thus, the probability for a surfer keeping clicking on links is given by the damping factor d, which is, depending on the degree of probability, set between zero and one. Since the surfer jumps to another page at random after he stops licking links, the probability therefore is implemented as the complementary part (1 - d) into the algorithm. Due to the huge size of actual web, an approximate iterative computation is usually applied to calculate the PageRank. This means that each page is assigned an initial starting value and the Page Ranks of all pages are then calculated in several computation circles based on one the previous two equations[31].

CHAPTER FOUR: METHODOLOGY

This chapter presents the methodology of GBATSS, Which described by using flowcharts, algorithms, Pseudo codes, figures and tables. We perform various stages for achieving text summarization.

In this study, we want to combine Google's page rank algorithm [63] with weighted graph that represents a single document, where the nodes of the graph represent sentences, and the weight of the edge between each two nodes represents the similarity between these two sentences. Weighted graph representation offers powerful and effective features offered by graph theory. When applying page rank we iterate many times according to the number that defined first, so after each iteration will give the node (vertex) in the graph (document) a rank (value), these values are used to tell us the importance of the node in the document. According to compression ratio, we extract the summary, we select the high-ranking nodes in the graph, and the selected nodes are the extracted summary.

The connection between sentences (edges) can be composed based on similarity between sentences. To calculate the similarity measure we depend on many parameters like: contents overlap. In our system, we use some statistical features to determine the rank of the sentences, such as term frequency and inverse sentence frequency. PageRank formula is used to combine both ranking sentences and calculating similarity between sentences.

To calculate the similarity measure and sentence score, first, we determine the basic unit on which these calculations are based. In this study, two basic units are applied; stem and word. Two types of stemming are used; light and rooted stemmer. Differences in calculating similarity measure, sentence ranking and the quality of the extracted summary are examined according to each unit.
This system processing is done as the following:

1. Enter the following values (dumping factor , number of iterations, Compression ratio).

2. Load single document to be summarized.

3. Text Preprocessing.

4. Text Normalization (remove duplication in spaces, extra commas, " ,' , etc….

5. Tokenization by splitting text to lines.

6. Stop Words removals.

7. Stemming, stemming can be done in three ways (Root, light, no-stem).

8. Building the graph.

9. Representation of sentences by vertexes.

10. Calculating the similarity between sentences that represents the weight of edges between graph vertices.

11. Applying Google PageRank algorithm to the graph.

12. According to compression ratio start the process of finding the candidate sentences to be chosen for the summary (Extract the summary).

Algorithm 4.1(a) shows the pseudo code of GBATSS. This algorithm shows how the process of summarization is done. The process of summarization as shown in the algorithm is done by loading the input document then applying the preprocessing phase by doing normalization stop words .removal and stemming. Then the processes that related with graph by creating, weighting ,and finally indexing the graph.Algorithm 4.1(b) shows edge weighting procedure.

Figure 4.1 shows the flaw of system process in GBATSS that is start by entering the limitation values then loading the source document, then doing the processes of normalization , tokenization ,stop words removal , stemming , and graph processes respectively.

Figure 4.2 shows the proposed architecture of the system and show the details of the sub process in the system. The content of this figure will be discussed in the next paragraph.

Algorithm 4.1(a): Pseudo Code of GBATSS

```
SUMMARIZER ALGORITHM:

Feed the input document
MASTER = entire document.
OUTPUT = output document.
Configure/Set the maximum sentences in the summary <= Total sentences in
document.
for each sentence
        Normalize()
        StopWordsRemoval()
        Stemming()
        createGraphNode( )
        weightGraphEdges()
        indexGraphNode( )
        findWordCounts( )
        findInverseSentenceCounts( )
end for
InvokePageRankAlgorithm(Dumpting_factor , number_of_iteration);
OUTPUT ← Extract_summary(Compresion_Ration)
DISPLAY OUTPUT.
```

Algorithm 4.1(b) : Edge weighting pseudo code.

```
INPUT = sentences
OUTPUT= weights
for each sentence [ call it source node]
        for every other sentence [call it sink node]
                find prod ← termfreq X invSentenceFreq
                if (sentence contains more words than other)
                        then
                                set prod ← 0 [ for the corresponding indexes]
                Save (source prods) //Array of prods calculated for each word in src
                sentence
                Save (sink prods)//Array of prods calculated for each word in sink
                sentence
                weight ← (getDotProduct ( source prods, sink prods )) / (RootSumSquares( source prods) X
                RootSumSquares(sink prods ))
        end for
end for
```

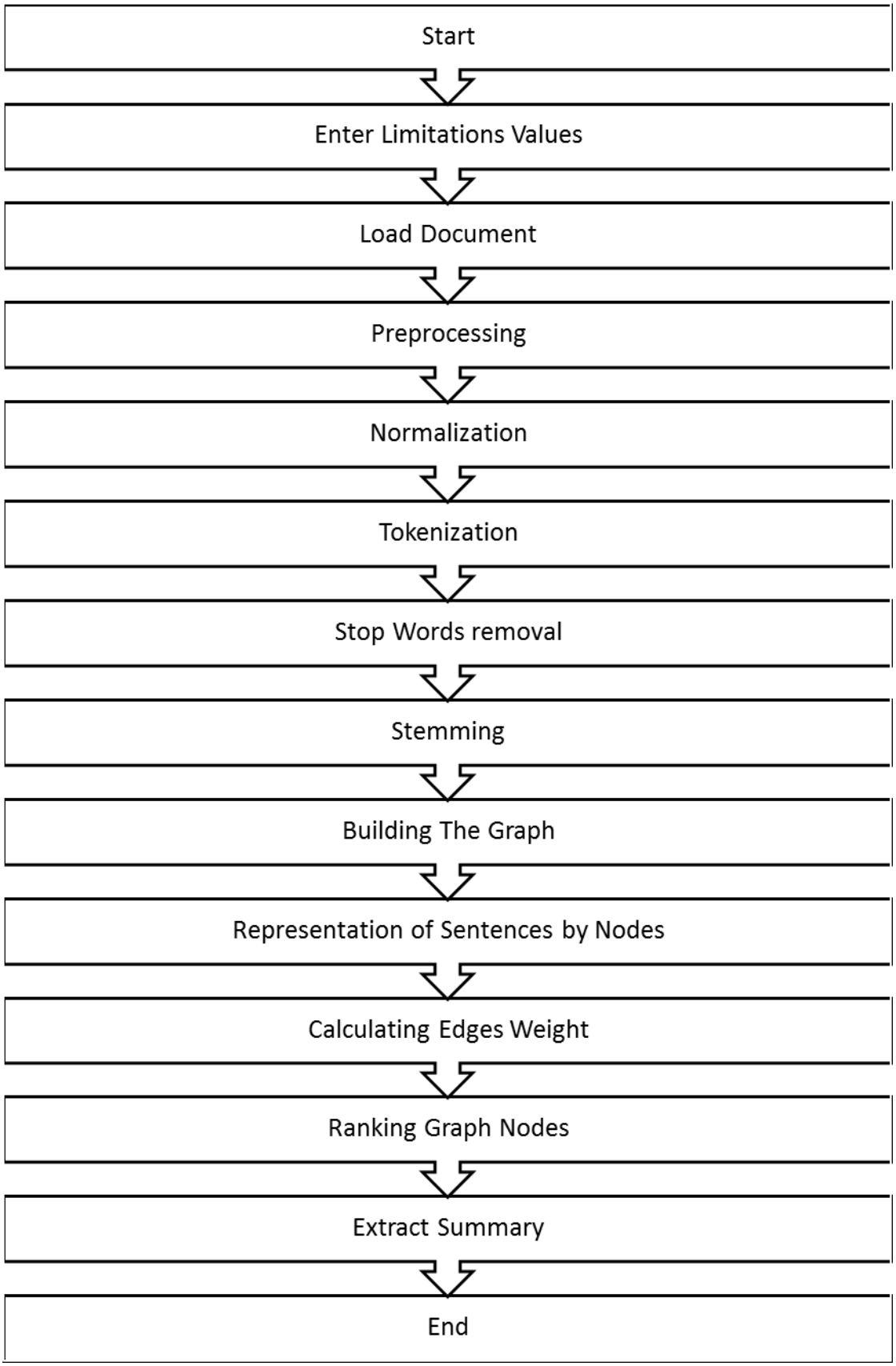

Figure 4.1: Flaw Chart Process

4.1 Enter Limitations values:

In this step, the input factors or limitation values are determined. These factors are the system limitations that will be considered in the summarization process. GBATSS needs the following limitation values:

1. Dumping factor :

   This factor is used to control the weight between sentences in the system, and is important to be supplied to get page rank to work. We use the dumping factor to manage the incoming edges and outgoing edges from the node. After making many studies on the dumping factor Google Page Rank founder finds that the best value to dumping factor is 0.85[63].

2. Compression ratio(CR) :
   Another input you must enter is the value of the compression ratio. This ratio determines the number of sentences that the system will retrieve.

3. Number of iterations:
   Determine the maximum number of iterations you want your program to iterate in the summarization process. When you iterate more you will get a better summary.

4. Determine the type of basic units in the system.

   Another input you enter is the type of basic unit in your summarization system. In this system we have three types of basic units that are as the following:

   a. Word or no-stemmed word.
   b. Rooted Stemmed word.
   c. Light stemmed word.

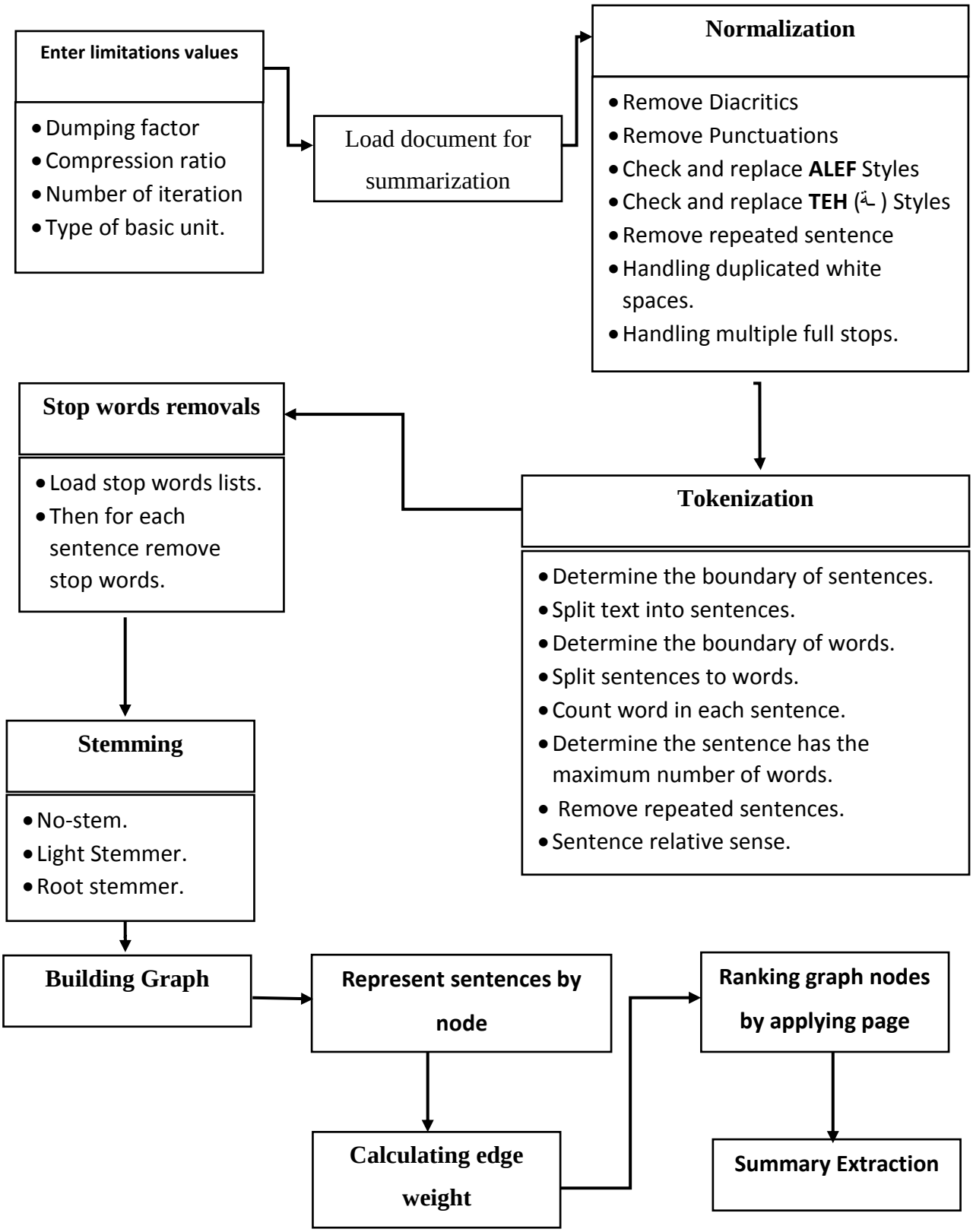

Figure 4.2: Proposed Architecture

4.2 Enter the Document to be summarized:

In this step we load the Arabic document that we want to be summarized; the document should be formatted in the utf-8 charset format, the documents and data set we used in system testing here is the Essex Arabic Summaries Corpus (EASC) [64].

4.3 Preprocessing:

Arabic language categorized as one of languages that has rich and complex morphological and syntactic flexibility [65]. So that dealing with Arabic directly in information retrieval without making any preprocessing steps will make dealing with the text difficult and giving us wrong results, so some language's processing needs to take place before summarization, such as tokenization, stemming and normalization as we will see in the following steps.

4.3.1 Normalization:

Normalization is the action of transforming the text to a new form to make it more consistent using some processing techniques. Normalization has great effects on the quality of the extracted summary, because of removing repeated sentences duplicated white spaces etc… To do normalization we do the following steps:

1. Remove Diacritics.
2. Remove Punctuations.
3. Check and replace **ALEF** Styles.
4. Check and replace **TEH** (ة) Styles.
5. Remove repeated sentences, this step done after stemming.
6. Handling duplicated white spaces.
7. Handling multiple full stops.

Previous Steps of normalization can be divided into two groups first group contains the steps that done before normalization contains steps 1,3,4,6 and 7 as following:

1. Remove Diacritics.

2. Check and replace **ALEF** Styles.
3. Check and replace **TEH** (ة) Styles.
4. Handling duplicated white spaces.
5. Handling multiple full stops.

The second group done after normalization and contains the steps 2 and 5.

1. Remove Punctuations.
2. Remove repeated sentences, this step done after stemming.

4.3.1.1 Remove Diacritics :

There are special notations in Arabic language called diacritics. It is used for making Arabic reader to get the correct pronunciation of the Arabic words .Diacritics are determined according to the Arabic grammar roles. Difference position of the word in the sentence giving us different diacritics for the word and different meaning.

Figure 4.3 shows the diacritics available in Arabic language that will be removed from the text. This figure listed 8 Arabic diacritics contains FATHAH , DAMMA, etc…

Table 4.1 shows an example for removing diacritics from the sentences. In this example original text contains some words that have some diacritics and after applying diacritics removal the result is done as in the example.

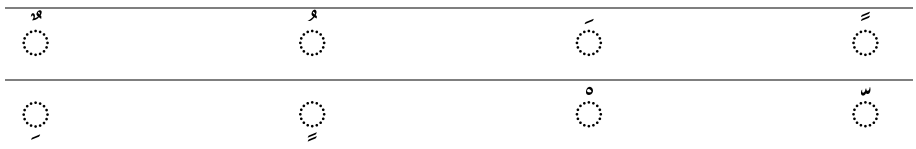

Figure 4.3: Arabic Diacritics

Table 4.1: Diacritics example

Original Text	**النّايْ آلةٌ نفخيةٌ تعدُّ بحق أقدم آلة موسيقية في التاريخ (إذا استثنينا الآلات الإيقاعية) وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة . والناي كلمة فارسية تعني المزمار .**
Remove Diacritics	**الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ (إذا استثنينا الآلات الإيقاعية) وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة . والناي كلمة فارسية تعني المزمار .**

4.3.1.2 Remove Punctuation

Arabic likes any other language need some marks to organize the text to make it more readable and to give the reader the correct meaning of the sentence. These marks called punctuations. Punctuations in the text summary do not have any value, so we remove all punctuations that are not a full stop.

Figure 4.4 shows punctuations that should be removed when they are appearing in the text. These punctuations contains full stop, comma, brackets, etc….

Table 4.2 shows example for removing punctuations from the sentence.

:	**"**	**'**	**.**	**>**	**<**	**(**	**)**	**{**	**}**
@	**#**	**$**	**%**	**^**	**&**	**!**	**/**	**[**	**]**
*****	**~**	**`**	**،**	**؟**	**-**	**×**	**؛**	**_**	**=**
			****	**|**	**-**	**--**			

Figure 4.4: Punctuations

Table 4.2: Punctuation processing example

Original Text	**الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ (إذا استثنينا الآلات الإيقاعية)وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة .**
Remove Punctuations	الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ إذا استثنينا الآلات الإيقاعية وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة

4.3.1.3 Check and Convert ALEF Style:

ALEF is the first letter in Arabic alphabet, we can write ALEF letter in different shapes as (إ، ا، أ، آ) according to its position in the sentence. 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 that helps in the stemming process. The change of ALEF style done only for the letter that at the beginning of the word.
Table 4.2 shows an example of dealing with ALEF style in sentence.

Table 4.3: ALEF style processing example

Original Text	**الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ إذا استثنينا الآلات الإيقاعية وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة**
Check and convert ALEF style	الناي الة نفخية تعد بحق اقدم الة موسيقية في التاريخ اذا استثنينا الالات الايقاعية وللناي عدة اسماء تعرف بها منها الناي القصبة الشبابة المنجيرة

4.3.1.4 Check and Replace TEH (ـة) Styles:

We use this step to the same purpose of the previous one, that there are many mistakes that appear in writing Arabic language, and it helps in stemming process. Dealing with TEH style here is done only at the end of the word.

Table 4.4 shows an example for handling THE style in the sentence.

Table 4.4: TEH style processing

Original Text	**الناي الة نفخية تعد بحق اقدم الة موسيقية في التاريخ اذا استثنينا الالات الايقاعية وللناي عدة اسماء تعرف بها منها الناي القصبة الشبابة المنجيرة**
Check and convert TEH style	الناي اله نفخيه تعد بحق اقدم اله موسيقيه في التاريخ اذا استثنينا الالات الايقاعيه وللناي عدة اسماء تعرف بها منها الناي القصبه الشبابه المنجيره

4.3.1.5 Handling duplicated white spaces :

Duplication in white spaces makes problem in tokenization process because it gives us an extra number of words because of we tell the tokenizer that the separator between every two words is a white space. Therefore, in this step we remove all duplicated white spaces to get a better tokenization process.

Table 4.5 shows example of dealing with duplicated white spaces in the sentence.

Table 4.5 Duplication in white spaces Processing Example

Original Text	**الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ إذا استثنينا الآلات الإيقاعية وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة**
Handling duplicated white spaces	الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ إذا استثنينا الآلات الإيقاعية وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة

4.3.1.6 Handling duplicated full stops:

In tokenization process, we define the sentence that ended with a full stop (.). If there are duplicated full stops in the text, this will make many problems like empty sentences that will take more processing to handle. Therefor here we will remove duplicated full stops to optimize and improve the process of summarization. Handling duplication on full stops done by replacing duplicated full stop by one full stops.

Table 4.6 shows example of handling duplication of full stops in the given sentence.

Table 4.6:Handling Duplication in Full stops Example

Original Text	**الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ إذا استثنينا الآلات الإيقاعية وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة..**
Handling duplicated full stops.	الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ إذا استثنينا الآلات الإيقاعية وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة.

4.3.2 Tokenization

Tokenization or in another word segmentation is the process of dividing your text to a minimum units. This units can be word, sentence etc...

Tokenization is a big challenge in text summarization because it deals closely with the morphological structure of the text in the documents, particularly those that are written in languages of rich and complex morphology, such as Arabic. Identifying sentences in Arabic is not an easy task. The morphological challenge comes from many reasons like:

1. Missing punctuation marks (i.e. “.”, “،”),
2. Arabic sentences do not start with capital letters (as in English, for example).

Information about line numbers within the paragraphs is stored and used in the final phase to preserve the order of the relevant text fragments to be included in the summary.

The function of a tokenizer responsible for two main tasks :

1. Splitting a running text into tokens.

2. Determination of words and sentences boundaries and the separation borders of the token units, multiword expressions, abbreviations and numbers [66].

Process of tokenization is done many steps. These steps are listed in figure 4.5 and will discussed below.

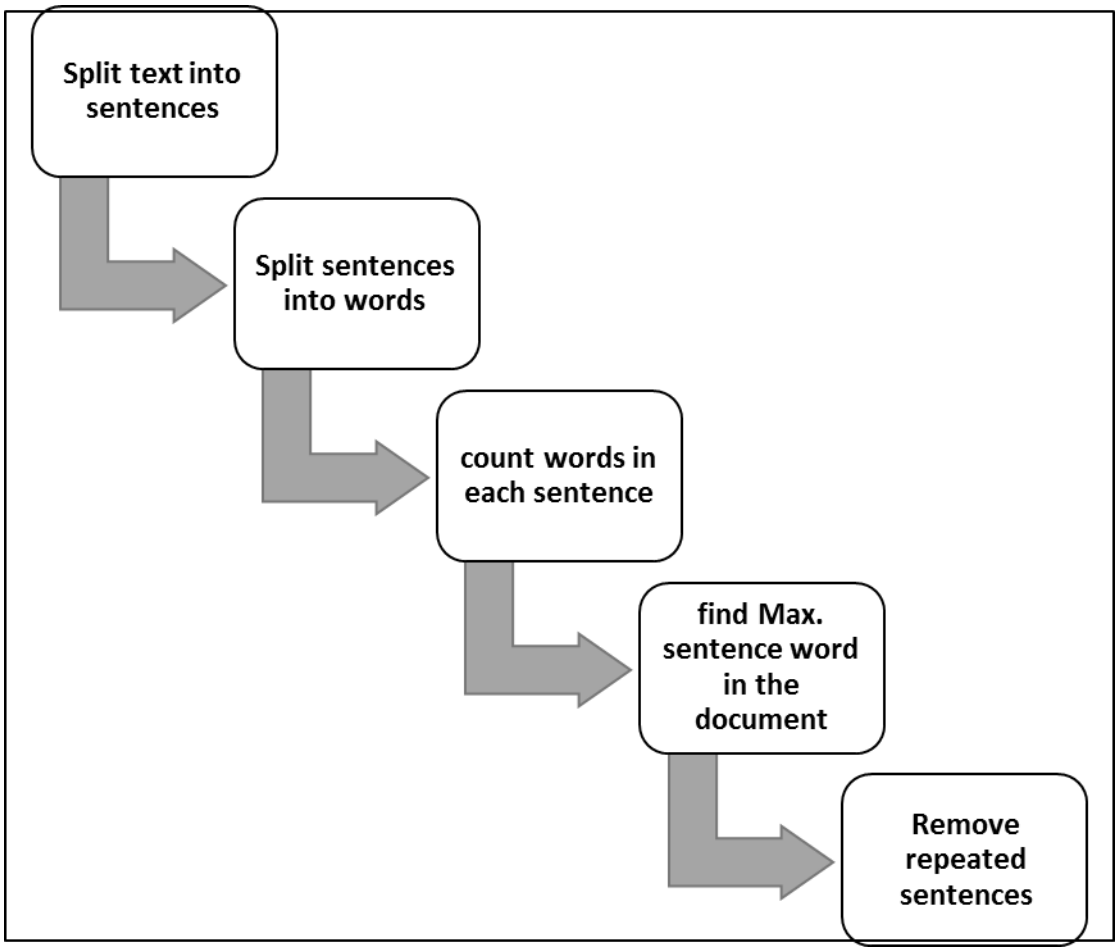

Figure 4.5: Tokenization Process Steps

Tokenization steps can be presents as following:

1. Split text into sentences:

   Sentence Boundary Disambiguation (SBD), or sentence breaking, is the problem in natural language processing of deciding where sentences begin and end. Often; natural language processing tools require their input to be divided into sentences for a number of reasons. Identification of sentence boundary is challenging because punctuation marks are often ambiguous. For example, a period may denote an abbreviation, decimal point, an ellipsis, or an email address ,not the end of a sentence. About 47% of the periods in the Wall Street Journal corpus denote abbreviations. As well, question marks and exclamation marks may appear in embedded quotations, computer code, and slang [66].

   Here we define Arabic sentence that ends with a full stop or with new line breaker.

2. Split sentences into words:

   After splitting our text into sentences, we then divide every sentence into words. Figure 4.6 shows how we can define the word in the sentence. The word can be determined as the following:

   1. In the beginning of the sentence: word ends with white space.

   2. In the middle of the sentence we have two cases :
      i. Between two white spaces.
      ii. Between white space and punctuation.

   3. At the end of the sentence between white space and full stop.

3. Count word in each sentence :

   In this step we count and store the word in each sentence tell we get all the words in the document this step helps us to find the weight of the edges between sentences in the later steps.

4. Find the maximum sentence word in the document:

   This is used for finding maximum term in every sentence and the maximum occurrence term in the document. This step will help us in a later step.

5. Remove repeated sentences:

In this step, we remove repeated sentence because repeated sentences will give the same sentence high rank than other sentences. In addition, it will take a place in the extracted summary and this will yield repeated sentences in the summary, which reduces the quality of the summary. Sentence compare with other sentences in the system, then if they have the same words or basic unit then one of them removed.

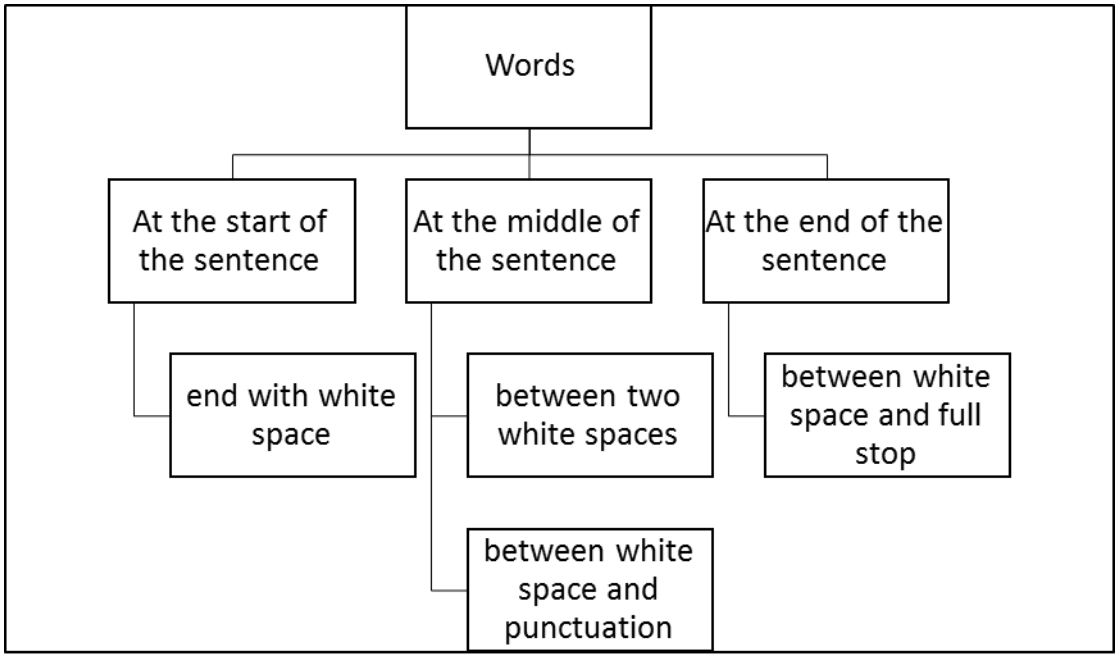

Figure 4.6: How to define words in sentences

4.3.3 Stop Words Removal

This step is important because we want to remove all stop words in the document to be summarized. So what is stop words? Stop words is a set of words that are used commonly in language to do many tasks like a connector or any other tasks that give your sentence a good meaning. They are repeated in the text like(في ، من ، إلى ،على ،…..) . In general stop words removal gives us two major benefits:

1. Improving the efficiency of information retrieval because frequent words have a high tendency to reduce the differences in frequency.

2. Shortens the length of the document and, as a result, affects the weighting process.[67].

Figure 4.7 and algorithm 4.2 show the process of removing stop words.

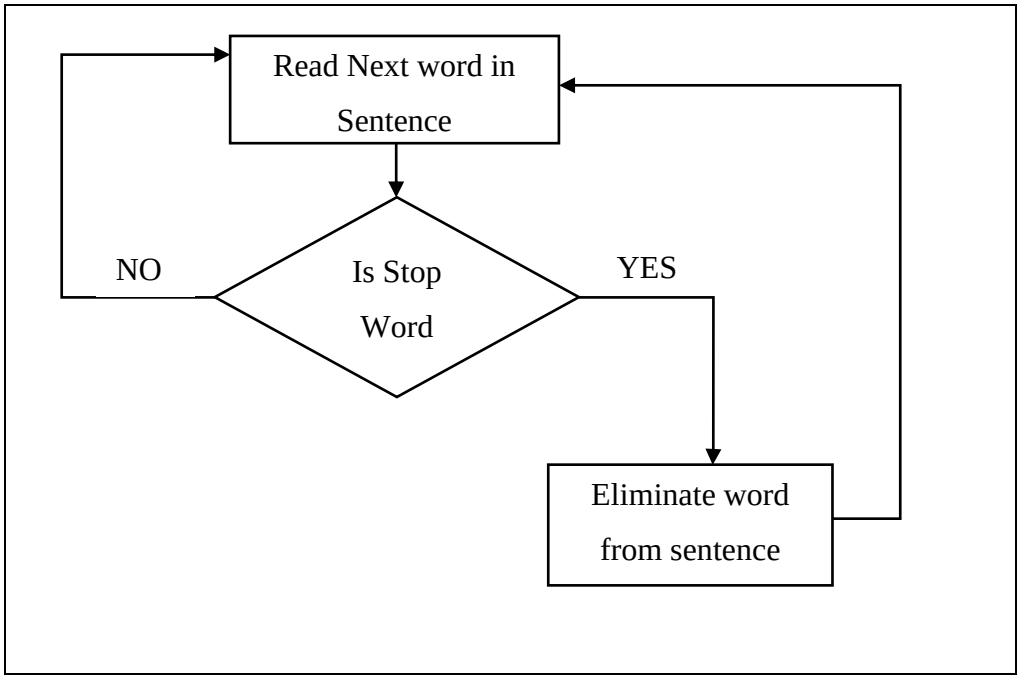

Figure 4.7: Stop words removal

Many researchers collect stop words manually, and then generate all possible forms of the words then include them in the list, which increases the length of the list. However, there is a difference in consideration to decide if the current word is stop word or not such as numbers in some topics, it will be considered as a stop word but in economic topics it will be an important word. another example is dates which are important in historical topics. So the programmer can later determine what words he wants, or we can make the user input the category of the summarized document then according to this we decide which stop words list to use.

Algorithm 4.2: stop words removal Pseudo code

```
for each sentence in document
        for each word in sentence
                if (word is stop word )
                        continue;
                else
                AddToText();
                end if
        end for
end for
```

Stop words can be categorized into many categories[67]:

1. Adverbs.
2. Measurement units.
3. Coins names.
4. Conditional Pronouns.
5. Interrogative Pronouns.
6. Prepositions.
7. Pronouns.
8. Referral Names/ Determiners.
9. Relative Pronouns.
10. Transformers (verbs, letters).
11. Verbal Pronouns and others.

In Arabic language stop words can be categorized to two categories according to that it can extended to prefixes or suffixes:

1. Words can take suffixes or prefixes (كان ، أول).
2. Words cannot take suffixes or prefixes (ثم، أو).

Table 4.7 shows stop words and the affixes that can concatenate with them.

Table 4.7: Stop words affixes

Type	**Genitive (الاضافة)**	**Preposition (الجر)**	Conjugation (التصريف)	**Article (التعريف)**	**Conjunction (العطف)**
Example	كونهم	في ثاني	يكون	الثالث	وكان

Therefore, when you create your stop words list; you must take into your account the stop word and all its situations.

As we said above; there is no specific list for stop words in Arabic. it may vary from researcher to another one and from topic to another topic, in this project we use stop words list that are created in code.google.com. [1]

This project has two lists of stop words. Every list contains 162 words. so here we have 324 words in our list.

Table 4.8 shows an example of stop words removals for a sentence.

[1] https://code.google.com/p/stop-words

Table 4.8: Stop words removal example

Original Text	**الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ إذا استثنينا الآلات الإيقاعية وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة.**
Stop words Removals	**الناي آلة نفخية تعد بحق أقدم آلة موسيقية التاريخ استثنينا الآلات الإيقاعية للناي أسماء تعرف الناي القصبة الشبابة المنجيرة.**

4.3.4 Stemming

Stemming is the process of reducing words to their roots or basic forms through the removal of any affixes attached to them. Reducing words to its root has many benefits like:

- **Compression**: To reduce the size of documents, large words could be stored in their root form. A small program would then be used to return the document to its original form when opened. It would do this by using context and grammar to determine the original form of the word.

- **Spell checking**: Instead of searching for a complete word in a dictionary, only the root would be searched for. This reduces the size of the dictionary.

- **Text searching**: the best example of this is web search engines. Searching for the root of a word gives a wider search than trying to find an exact match.

- **Text Analysis**: For example in statistical text analysis, stemming helps in mapping grammatical variations of a word to instances of the same term[68].

Algorithms for stemming have been studied in computer science since the 1960s [69]. Many search engines treat words with the same stem as synonyms as a kind of query expansion, a process called conflation. Stemming programs are commonly referred to as stemming algorithms or stemmers.

The purpose of this step get the root of the word that will improve the process of weighting the relation between sentences that yield improving the quality of the summarization system.

For example, stemming the Arabic word "قراءة" produces the root "قرأ". This root can also be generated from the word "قارئ". After reducing words to their roots, these generated roots can be used for many applications like compression, spell checking, and text searching. Here we adopted the stemmer of Khoja [68] and light stemmer of Lucene that done by apache[10].

In arabic languages not all words can be stemmed to their original root because of it may be not have original root like preposition or it may be a strange or foreign words. In our system, we use three type of basic units these basic types is as the following:

1. Rooted Stemmer:
   In this stemmer we use khoja[68] stemmer that stem the word to it's root form that has three or four letters.

2. Light Stemmer:
   Light stemmer only remove suffixes and prefixes from the word then return it to the user.

3. No-stem:
   Is the basic unit we want to do in this stage we use the word without any changes.

Figure 4.8 shows categories of basic units, which used in the GBATSS. Table 4.9 shows example of stemmer.

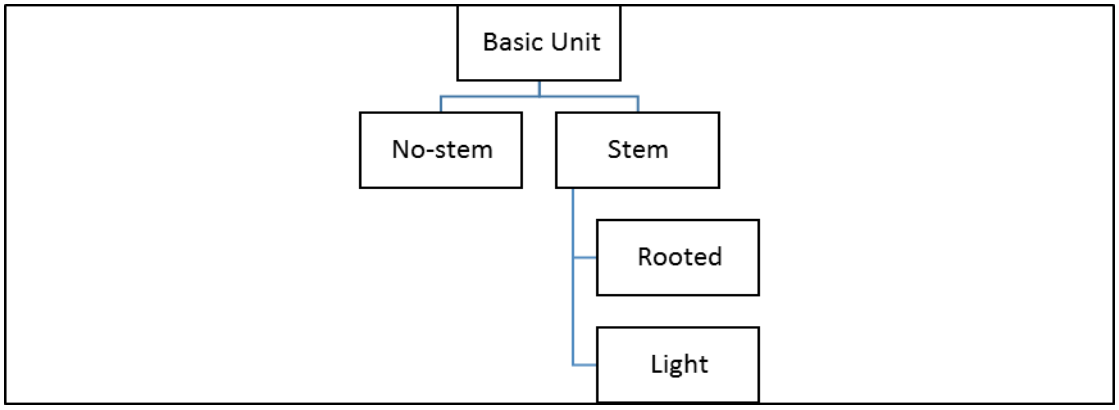

Figure 4.8: Basic units categorization

Table 4.9: Basic units example

Original Text	**الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ (إذا استثنينا الآلات الإيقاعية (وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة**
No-stem	**الناي آلة نفخية تعد بحق أقدم آلة موسيقية التاريخ استثنينا الآلات الإيقاعية للناي عدة أسماء تعرف الناي القصبة الشبابة المنجيرة**
Root Stemmer	**نوي نفخ عدا حقق قدم وسق أرخ ثني الا قعي نوي عدا سمي عرف بها نوي قصب شبب جور**
Light Stemmer	**نا ال نفخ تعد بحق اقدم ال موسيق تاريخ استثنينا ال ايقاع للنا عد اسماء تعرف قصب شباب منجير**

4.4 Building the Graph

The input Arabic text is represented by a graph G. Graph G of a document D is a directed graph G = (V, E), where V is a set of nodes and E is a set of edges [14].

In other words, G is a weighted directed graph whose nodes represent sentences of D and edges weights represent similarity between sentences. Figure 4.9 shows an example of directed weighted graph. In this figure, Symbol S_i represents i^{th} sentence in the text. $W_{i,j}$ represents the weight of the relation between i^{th} sentence and j^{th} sentence in the text.

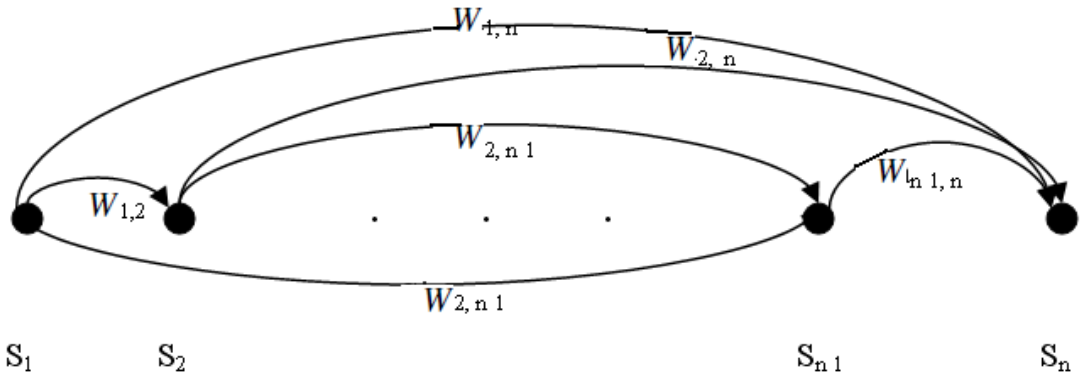

Figure 4.9: Weighted graph

4.5 Representation of Sentences by Nodes

After preprocessing phase, each sentence is provided with an ID, where each ID is represented by a node as dealing with IDs, in some steps, is much easier than the entire sentence. These IDs will be used in the process of finding the summary. Figure 4.10 shows an example of weighted graph that's replace each sentence with ID.

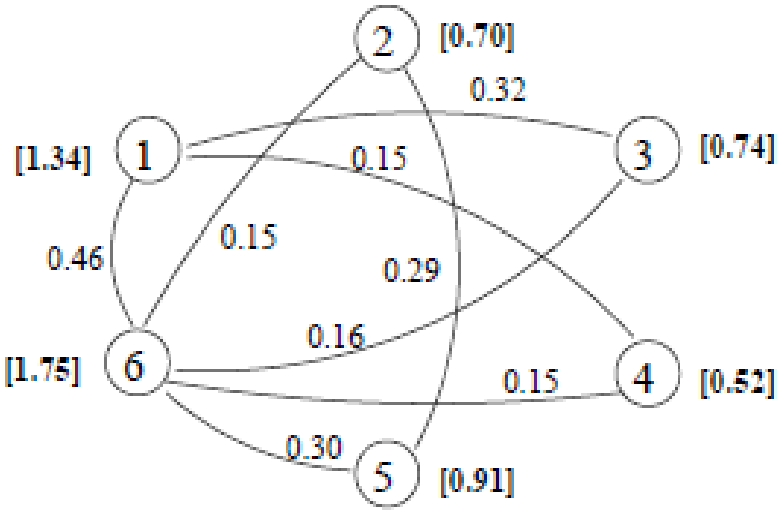

Figure 4.10: Weighted Graph Replaced Sentences with ID

4.6 Calculating Edge Weight

The connection between sentences can be composed on the basis of similarity between the relevant sentences.

The similarity measure is calculated by many parameters, such as content overlap and cosine similarity measures. In this study, the cosine similarity measure is chosen on the basis of term weighting scheme which is the TF-IDF (Term Frequency-Inverse Document Frequency). The **TF-IDF** weighting scheme is a commonly used information retrieval technique for assigning weights to individual terms appearing in the document. This scheme aims at balancing the local and the global term occurrences in the documents [70] [71].

Here **IDF** will be **ISF** (Inverse Sentence Frequency) . **TF-ISF** weights are computed for each sentence, where $\mathbf{s_j}$ shows the $\mathbf{j^{th}}$ sentence and $\mathbf{k_i}$ is $\mathbf{i^{th}}$ index term, $tf_{i,j}$ is said to be 'term frequency' of i^{th} index term in the j^{th} sentence, and isf_i is 'inverse sentence frequency' of i^{th} index term, where N is the number of all sentences and n_i is the number of sentences which contain k_i . The corresponding weight is therefore computed as, $w_{i,j} = tf_{i,j} \times isf_i$. Equation 4.1 shows the equation of ISF calculation. Equation 4.2 shows the formula of TF calculation.

$$ISF_i = \log\frac{N}{n_i} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots eq(4.1)$$

Where:

ISF_i : inverse sentence frequency of i^{th} term.

N: Number of sentences in the text.

n_i: Number of sentences contains i^{th} term.

$$tf_{i,j} = \frac{freq_{i,j}}{max_i\, freq_{i,j}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots eq(4.2)$$

The similarity /edge weight between two sentences s_m and s_n is easily calculated based on cosine measure as in equation 4.3.

$$W(s_m, s_n) = \frac{\sum_{i=1}^{t} w_{i,m} \times w_{i,n}}{\sqrt{\sum_{i=1}^{t} w_{i,m}^2} \times \sqrt{\sum_{i=1}^{t} w_{i,n}^t}} \dots\dots\dots\dots\dots\dots\dots\dots\dots \text{eq}(4.3)$$

Where :t is the number of terms in the sentence.

4.7 Ranking Graph nodes by applying page rank algorithm

The sentences are sorted based on ranks of nodes. The original page rank combines the effect of both incoming and outgoing links. Equation 4.4 shows the original page rank formula.

$$PR(V_i) = (1-d) + d * \sum_{V_j \in In(V_i)} \frac{PR(V_j)}{|Out(V_j)|} \dots\dots\dots\dots\dots\dots\dots\dots\dots . eq(4.4)$$

Where d is a parameter set between 0 and 1.

Equation 4.4 has been adapted to include the notion of edge weights in the graph. Equation 4.4 shows the new formula.

$$PR^w(V_i) = (1-d) + d * \sum_{V_j \in In(V_i)} w_{ji} \frac{PR^w(V_j)}{\sum_{V_k \in Out(V_j)} w_{kj}} \dots\dots\dots\dots . eq(4.5)$$

Where:

$PR^W(V_i)$: is Page rank of vertex V_i

$In(V_i)$: is all the predecessor vertices to node V_i

$Out(V_i)$: is set of vertices that V_i points to .

Graph Implementation possibilities :

- Forward directed : The edges only go out from a sentence to one or more sentences following it.
- Backward directed : The edges only go out from a sentence to one or more sentences preceding it.

The forward DAG(Directed Acyclic Graph) representation has been implemented ,tested and found that the algorithm seems to be biased and has been consistently ranking the sentences in the latter part of the document better than the starting

portions. Hence going by the fact that two sentences are similar if one's contents are similar and one follows the other or vice versa. we can use an undirected graph. This implementation seems to be giving positive and impressive results than its forward directed counter part. The following rules govern the graph structure would be joined to:

1)There is no chronological differences between the sentences , only the contents carry importance.

2)There is also no self-edge, the similarity of every sentence to itself is considered to be 0.

3)There is only one link that connect between two sentences.

This assumption is stated as:

- $i < N : W(s_i, s_i) = 0$

Algorithm 4.3 shows how Page Rank algorithm works. As listed in algorithm 4.3 Page Rank works as follow:

i. Initialize ranks for all sentences in the text equal 1.
ii. Enter while loop until you get converged. Convergence here is the number of iterations GBATSS will iterate.
iii. For every sentence in the document :
    - Initialize master sum to 0.0.
    - Calculate the similarity between current sentence according to equation 4.3.
    - Calculate the rank of the current sentence according to equation 4.5.
    - Save the rank of current sentence.
iv. After calculating rank for every sentence update saved ranks.
v. If converge exit else go to step (iii) and continue.

Algorithm 4.3: Page Rank Algorithm

PAGE RANK ALGORITHM:
Initialize all Ranks = 1 (# of ranks = # of sentences in the document)

while !Converged **iterate**

for i between [1 and numSentences]
sum ← 0.0 //////This is master sum.

for j between [1 and numSentences]
if (j equals i) //////Do not evaluate any sentence with itself.
continue
Wji ← (j < i) ?getSimilarity(j and i):getSimilarity(i and j)
PRVj ← getRank(j)
denSum ← 0.0 /////This is denominator partial sum
for k between [1 and numSentences]
if (k equals j)
continue
Wjk ← (j < k)?getSimilarity(j and k):getSimilarity(k and j)
denSum ← denSum + Wjk

end for
sum ← sum + (Wji * pageRankVj / denSum)

end for
rank ← (1- *DAMPING_FACTOR*) + *DAMPING_FACTOR* **X** sum
tmpranks.save(i, rank)

end for

updateRanks(tmpranks) //This will clear / erase the older ranks and replace them with newer rank, which is used in next iteration and so on.

end while

4.8 Summary Extraction

The 'n' best sentences are chosen based on maximum cut off words/sentences in the summary. The value of n decided depending on the Compression Ratio (CR). This ratio represents a particular proportion of the number of sentences that compose the original text that the user has a choice to choose it.

After the construction of the graph and the calculation of edge weight, and graph nodes are ranked the extraction of the summary is done. Extraction done be selecting the IDs of sentences to be extracted. After selecting IDs then we return to the basic array of sentences we have. then we find the original sentences to be extract. After that we combine the sentences with each other then display the summary. Algorithm 4.4 shows how summary extraction done.

Summary algorithm works as the following:

- ☒ Set the number of maximum sentence you want to include in your summary. This number defined according to compression ratio.
- ☒ Sentences in the summary are sorted from according to its ranks.
- ☒ Check if the total number of selected sentences selected in the summary is equal or greater than the maximum number of the summary stop else continue.
- ☒ After getting the maximum number of the summary, display the summary.

Algorithm 4.4: Summary Algorithm

```
Configure/Set the maximum sentences in the summary <= Total sentences in document.
While ( ! done)
        if (maxSentences <= totalsSentences)
                for each sentence in list
                        if (counter <= maxSentences)
                                OUTPUT.put(currentsentence)
                        else
                                done = true
                        break
                end for
                if (done)
                        break
        else
                OUTPUT “Too many sentences”
        Break
end While

DISPLAY OUTPUT.
```

4.9 Implementation

Java programming language is an open source and platform independent programming language. Many application and programming tools helping libraries that are important in our design, so we use java in programming process of the system. Algorithm 4.5 shows the implementation of the main function in GBATSS. In this function, GBATSS do the following process:

- ☒ load text.
- ☒ Load limitation values.
- ☒ Define basic unit.
- ☒ Preprocessing the document.
- ☒ Build the graph.
- ☒ Invoke page rank.
- ☒ Generate and display the summary.

Algorithm 4.5: Main Function of GBATSS.

```
String sstring = "";//output string to be displayed
      Stemx s = new Stemx();///intialize stem class
      s.load_stop_words();
      String input_text = jTextArea1.getText();//read input string from strign class
      String[] nlines = s.lineSep(input_text);//split the  input string to lines
      String[] st_removed = new String[nlines.length];
      for (int i = 0; i < nlines.length; i++) {
         st_removed[i] = s.stopWordRemoval(nlines[i]);
      }
      String[] out = new String[nlines.length]; //lines after doing modifications.

      double dumping_factor = d_factor.getValue() / 100.0;
      int sentence_number = (int) Math.ceil(out.length * CR.getValue() / 100.0);
      int iteration_number =
Integer.parseInt(jS_iteration_number.getValue().toString());
      int selected_choice = 0;
```

```
        if (rb_all.isSelected()) {
            selected_choice = 1;
        } else {
            if (rb_root_stemmer.isSelected()) {
                selected_choice = 2;
                for (int i = 0; i < st_removed.length; i++) {
                    out[i] = s.stem(st_removed[i]);
                    out[i] = s.preProccessing(out[i]);
                }
            } else {
                if (rb_lightStemmer.isSelected()) {
                    selected_choice = 3;
                    LightStemmer ss = new LightStemmer();
                    for (int i = 0; i < st_removed.length; i++) {
                        out[i] = ss.stem(st_removed[i]);
                        out[i] = ss.preProccessing(out[i]);
                    }
                } else {
                    //word
                    selected_choice = 4;
                    out = st_removed;
                }
            }
        }
        Summarizer summarizer = new Summarizer();
        summarizer.processCmdLine(out, dumping_factor, sentence_number ,
iteration_number);


        summarizer.preProcessDocument();
        summarizer.buildGraph();
```

```
summarizer.invokePageRanking();
ArrayList<String> aa = summarizer.displaySummary();
if (selected_choice == 4) {
   for (int i = 0; i < aa.size(); i++) {
      sstring += aa.get(i) + "\n";
   }
} else {
   for (int i = 0; i < aa.size(); i++) {
      for (int j = 0; j < out.length; j++) {
         String t1 = out[j];
         String t2 = aa.get(i);
         //System.out.println(t1 +"   :   " +t2 + "   :   " + t1.equals(t2));
         if (t1.equals(t2)) {
            sstring += nlines[j] + " . \n";
         }
      }
   }
}
jTextArea2.setText(sstring);
```

Explanation	**This function is responsible for handling all the summarization process by importing normalizing , tokenizing , stemming and then starting the summarization process by creating nodes , building graph , applying ranking ,then extract the summary.**

4.9.1 System interface :

System Interface is very simple as figure 4.11 shows. The interface contains the following elements:

1. Two text area :
   One for the original text or the text to be summarized and the other contains the summarized text.
2. Two sliders one for determining the value of dumping factor and the other for determining the compression ratio.
3. Text field to enter the number of iteration you want the system to run after displaying the resulted summary.
4. Radio Button group that contains the options of stemming type you want to select in the summarization process.
5. Summarize button: when you click this button then the application will run the summarization process.

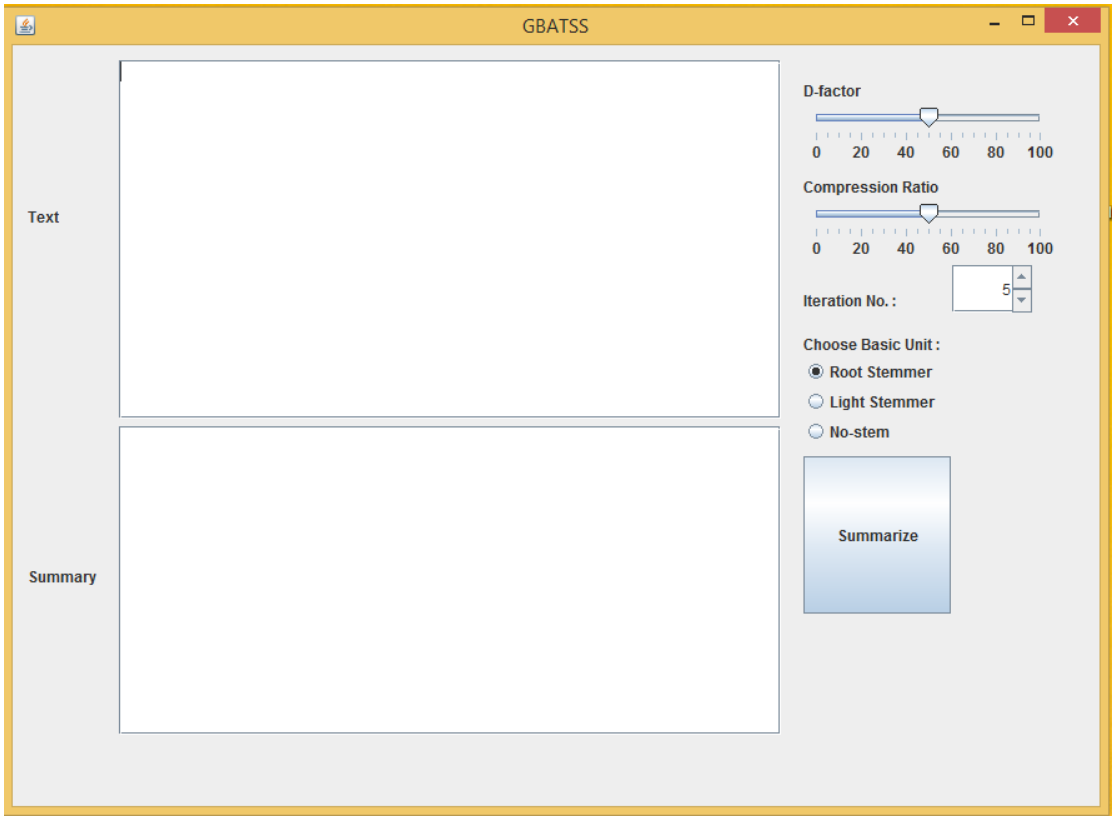

Figure 4.11: Application Interface

4.9.2 Tools and Program:

Special tools and programs are used to complete the implementation of automatic Arabic text summarization and documentation of the thesis:

- Shreen Khoja Stemmer [68]:

  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.

- Lucene light stemmer[10] :

  This stemmer from apache is used for making light stemmer for Arabic word.

- Microsoft Excel 2013:

  It is used to calculate the test result and compute the final P, R, F measures.

- Net Beans 8.0.2:

  Free IDE (Integrated Development Environment) from Oracle. We use it here to help us in developing our java application for summarization. This IDE is very helpful and easy to use.

- Java Development Kit (JDK) 1.8:

  A software development package from Oracle that implements the basic set of tools needed to write, test and debug Java applications.

- Microsoft Word 2013:

  Use this program mainly in writing the documentation of the application and thesis final report.

- Notpad++ :

  We use this application in the process of editing and viewing the code.

- Lucene Normalization package :

  Another free package from Apatche that helps in the normalization process.

CHAPTER FIVE: RESULTS, EVALUATION AND DISCUSSION

In this chapter, we talk about GBATSS (Graph Based Arabic Text Summarization). We take about implemented code, programming language, developing tools, external libraries that are used to develop the proposed system. also we will talk about data set that are used to test the system. Also we will talk about the evaluation of the system and how to measure its performance. Later, Summaries generated by our proposed system, EASC summary and summaries that reported in [62] will be presented and results will be evaluated by doing a comparison among the four approaches. At the end of this chapter, we shall discuss our results.

5.1 Data set:

We use The Essex Arabic Summaries Corpus (EASC) as data set in our system[64]. The EASC is an Arabic natural language resources. It contains 153 Arabic articles each article has 5 human generated extractive summaries with total of 765 summaries for those articles. These summaries were generated using Mechanical Turk [64]. EASC data set contains 10 subjects that are as follow:

- Art.
- Music.
- Environment.
- Politics.
- Sports.
- Health.
- Finance.
- Science and technology.
- Tourism.
- Religion.
- Education.

The articles of this corpus are collected from 3 sources as follows:

- Wikipedia : with 106 articles.
- Alwatan newspaper: with 34 articles.
- Alrai newspaper: with 13 articles.

In this system, we use a compression ratio of 40% and we use three types of basic words for the summary so for every document we have three summaries. We select 10 samples of the data set available so we have 60 summaries that will be compared with the results that are generated in EASC and then evaluate the system.

5.2 System examples :

In this section, we present some examples of the results that that are generated from our summarization system. As we discussed before; we use a compression ratio for 40% and default number of iteration 5 and dumping factor of 0.85.
In addition, we did not select all summaries that are listed in EASC because there are repeated summaries and some outlier summaries that are not valid for testing.

Also we use three types of basic units: no-stem, rooted stem and light stem.
Tables 5.1(a), 5.1(b), 5.1(c) and 5.1(d) show an example of generated summary for each type of basic units.

Table 5.1(a): Basic paragraph

Basic paragraph
الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ (إذا استثنينا الآلات الإيقاعية (وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة . **والناي كلمة فارسية تعني المزمار .** **هي قصبة مفتوحة الطرفين يعزف عليها بواسطة وضع الفم على أحد طرفيها مع إمالته قليلا بزاوية مما يجعل الهواء يصطدم بجدارها الداخلي مصدرا بالنتيجة صوتا شجيا هو أقرب الأصوات وأجملها بالنسبة للإنسان .** **وللناي ستة ثقوب (وأحيانا سبعه) وثقب في منتصف القصبة من الأسفل.** **وتسد هذه الثقوب وتفتح حسب درجة الصوت ولإخراج العلامات بتسلسل يستطيع معه العازف إخراج العلامات الموسيقية لإخراج اللحن المطلوب .** **والثقب الخلفي يسد بالإبهام ويستخدم لإظهار جواب العلامة الدنيا التي تظهر في البداية .** **وتحتاج هذه الآلة إلى براعة شديدة حيث لها 3 تقنيات.** **التقنية الأولى : هي طريقة النفخ حيث أن إخراج الصوت الطبيعي منها هي أول صعوبة يجب التغلب عليها لمن أراد التعلم عليها . لذلك ينصح عادة بأن يتمرن من يريد التعلم بالتدريب على إخراج الصوت أولا ومن ثم عندما يستطيع ذلك يبدأ بالتعلم على إخراج الدرجات الصوتية (تمرين الأصابع) .** **والعازف الخبير يستطيع بتغيير طريقة النفخ التلاعب بهذه الآلة الخطيرة حيث يستطيع العازف المتمكن أن يخرج أكثر من سبع علامات صحيحة أو حتى أكتر من أوكتاف (ديوان) فالآلة تنتج 7 أصوات صحيحة تماما وبالتالي**

يستطيع العازف الخبير أن ينتج 7 أصوات أخفض و 7 أصوات أعلى .

التقنية الثانية : هي إمالة الشبابة بأكثر من زاوية لإخراج أصوات معينة أو ربع التون .

التقنية الثالثة : وهي طريقة سد الثقوب بحيث ينتج الأصوات المطلوبة أو سد الثقوب بطريقة معينة لإنتاج ربع التون (العلامات الشرقية) .

وآلة الناي آلة أساسية في التخت الشرقي التقليدي حيث أن صوتها قريب جدا إلى الأذن البشرية .

وقد اعتاد إخواننا المصريين إضافة عدة آلات تشبهها بالطريقة (مع الاختلاف بطبيعة الصوت) ومنها آلة الكولة ولكنهم لم يلغوا هذه الآلة العظيمة .

عيوب هذه الآلة كثيرة أبرزها أن صوتها ليس ثابتا لذلك تعتمد كثيرا على أذن العازف الذي يجب ان يكون بارعا وذو أذن صافية ممتازة حتى تؤدي دورها الصحيح .

والعيب الثاني أن لكل ناي درجة صوتية يبدأ منها ولذلك تجد أن العازف عادة ما يملك أكثر من ناي وعادة ما يكون عددها سبع وإن كان بعض العازفين المهرة جدا يكتفون بأربع أو ثلاث قصبات ويتحايلون على بقية المقامات.

Table 5.1(b): Generated summary using root stem as basic unit.

Summary generated for root stem as a basic unit
الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ (إذا استثنينا الآلات الإيقاعية (وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة. وتحتاج هذه الآلة إلى براعة شديدة حيث لها 3 تقنيات.التقنية الأولى : هي طريقة النفخ حيث أن إخراج الصوت الطبيعي منها هي أول صعوبة يجب التغلب عليها لمن أراد التعلم عليها . لذلك ينصح عادة بأن يتمرن من يريد التعلم بالتدريب على إخراج الصوت أولا ومن ثم عندما يستطيع ذلك يبدأ بالتعلم على إخراج الدرجات الصوتية (تمرين الأصابع) . التقنية الثالثة : وهي طريقة سد الثقوب بحيث ينتج الأصوات المطلوبة أو سد الثقوب بطريقة معينة لإنتاج ربع التون (العلامات الشرقية) . عيوب هذه الآلة كثيرة أبرزها أن صوتها ليس ثابتا لذلك تعتمد كثيرا على أذن العازف الذي يجب ان يكون بارعا وذو أذن صافية ممتازة حتى تؤدي دورها الصحيح . والعيب الثاني أن لكل ناي درجة صوتية يبدأ منها ولذلك تجد أن العازف عادة ما يملك أكثر من ناي وعادة ما يكون عددها سبع وإن كان بعض العازفين المهرة جدا يكتفون بأربع أو ثلاث قصبات ويتحايلون على بقية المقامات.

Table 5.1(c): Generated summary using light stem as basic unit.

Summary generated from light stem as a basic unit
وتسد هذه الثقوب وتفتح حسب درجة الصوت ولإخراج العلامات بتسلسل يستطيع معه العازف إخراج العلامات الموسيقية لإخراج اللحن المطلوب . لذلك ينصح عادة بأن يتمرن من يريد التعلم بالتدريب على إخراج الصوت أولا ومن ثم عندما يستطيع ذلك يبدأ بالتعلم على إخراج الدرجات الصوتية (تمرين الأصابع) . التقنية الثانية : هي إمالة الشبابة بأكثر من زاوية لإخراج أصوات معينة أو ربع التون . التقنية الثالثة : وهي طريقة سد الثقوب بحيث ينتج الأصوات المطلوبة أو سد الثقوب بطريقة معينة لإنتاج ربع التون (العلامات الشرقية) . عيوب هذه الآلة كثيرة أبرزها أن صوتها ليس ثابتا لذلك تعتمد كثيرا على أذن العازف الذي يجب ان يكون بارعا وذو أذن صافية ممتازة حتى تؤدي دورها الصحيح . والعيب الثاني أن لكل ناي درجة صوتية يبدأ منها ولذلك تجد أن العازف عادة ما يملك أكثر من ناي وعادة ما يكون عددها سبع وإن كان بعض العازفين المهرة جدا يكتفون بأربع أو ثلاث قصبات ويتحايلون على بقية المقامات.

Table 5.1(d): Generated summary using no-stem as basic unit.

Summary generated when using no-stem as a basic unit
الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ (إذا استثنينا الآلات الإيقاعية (وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة. وتحتاج هذه الآلة إلى براعة شديدة حيث لها 3 تقنيات.التقنية الأولى : هي طريقة النفخ حيث أن إخراج الصوت الطبيعي منها هي أول صعوبة يجب التغلب عليها لمن أراد التعلم عليها . لذلك ينصح عادة بأن يتمرن من يريد التعلم بالتدريب على إخراج الصوت أولا ومن ثم عندما يستطيع ذلك يبدأ بالتعلم على إخراج الدرجات الصوتية (تمرين الأصابع) . التقنية الثالثة : وهي طريقة سد الثقوب بحيث ينتج الأصوات المطلوبة أو سد الثقوب بطريقة معينة لإنتاج ربع التون (العلامات الشرقية) . عيوب هذه الآلة كثيرة أبرزها أن صوتها ليس ثابتا لذلك تعتمد كثيرا على أذن العازف الذي يجب ان يكون بارعا وذو أذن صافية ممتازة حتى تؤدي دورها الصحيح . والعيب الثاني أن لكل ناي درجة صوتية يبدأ منها ولذلك تجد أن العازف عادة ما يملك أكثر من ناي وعادة ما يكون عددها سبع وإن كان بعض العازفين المهرة جدا يكتفون بأربع أو ثلاث قصبات ويتحايلون على بقية المقامات.

5.3 System Evaluation :

Summaries evaluation is a difficult process because of that there is more than one summary for each text that differ from person to person. Here we will use the EASC data set that offers five summaries for each text, so we want to evaluate the system for every summary compared with five extracted in the EASC ,then after that we will calculate the average for every summary.

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 as in equation (1.3).

According to equation (1.3), 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 is equal to one.

Figure 5.1 shows the difference between relevant and retrieved sentences and show the difference between precision and recall. In the figure we have the following terms:

- False positive: that means sentences selected in the retrieved summary but it does not exist in the relevant summary.
- False negative: sentences do not exist in the retrieved but exist in the relevant.
- True positive: sentences exist both in the retrieved and relevant summary.
- True negative: sentences do not exist both in the retrieved and relevant summary.
- All: list all sentences in the text.
- Relevant: sentences that are listed in the pre generated summary.
- Retrieved: sentences that result from the summarization system.
- Not relevant: sentences that are not listed in the pre generated summary.

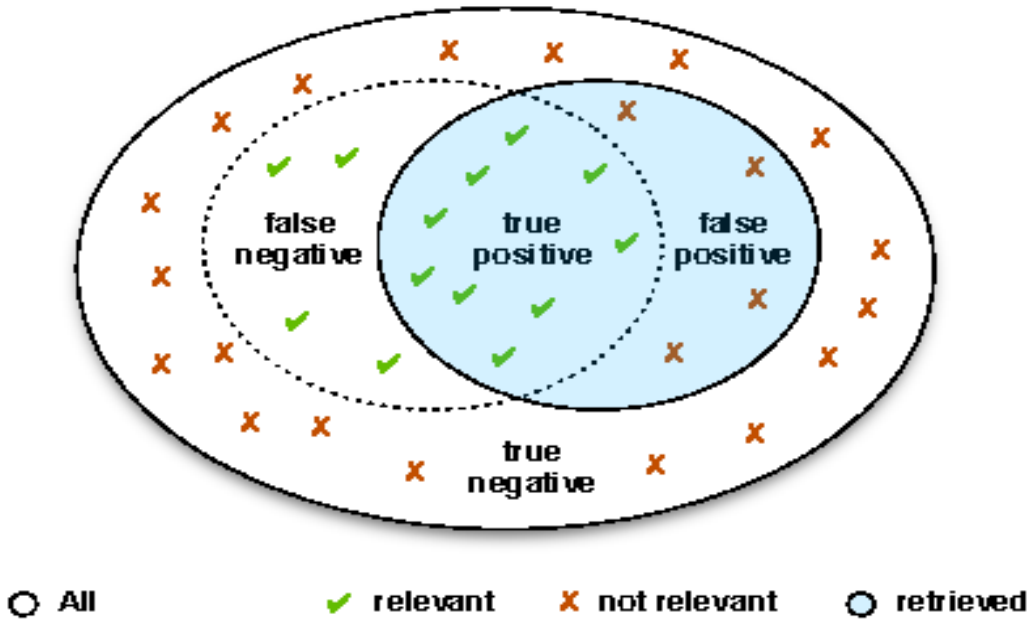

Figure 5.1: classification of resulted values

5.4 System Evaluation with EASC:

In this section, we use EASC generated summarization to compare with GBATSS results. Here we generate three summaries for each text according to the basic units that are used. Then we compare it with the summaries that are listed in EASC. Summaries in EASC have some redundancy and weak summary so in this research we choose best three summaries from them and then use them in the evaluation process of our system. In this evaluation we choose an example of each type of original text like(art, sport, health ,environment , education etc…). The evaluation parameters that are used here are recall, precision and f-measure according to equations (1.1),(1.2) and (1.3) respectively.

Tables 5.2 shows an example of art and music article and the generated summary. Table 5.2(a): shows the original text of the article. Table 5.2(b) shows the first summary listed in EASC corpus. 5.2(c) shows the second summary listed in EASC corpus. 5.2(d) shows the third summary listed in EASC corpus.

Tables 5.3 shows the retrieved summaries of art and music article and there evaluation results. Table 5.3(a): shows the retrieved summary for art and music using rooted-stem as a basic unit and its evaluation results. text of the article. Table 5.3(b) shows

the retrieved summary for art and music using light-stem as a basic unit and its evaluation results. 5.3(c) shows the retrieved summary for art and music using no-stem as a basic unit and its evaluation results.

Tables 5.4 shows the retrieved summaries of sports article and there evaluation results. Table 5.4(a): shows the retrieved summary for sports using rooted-stem as a basic unit and its evaluation results. text of the article. Table 5.4(b) shows the retrieved summary for sports using light-stem as a basic unit and its evaluation results. 5.4(c) shows the retrieved summary for sports using no-stem as a basic unit and its evaluation results**.**

Tables 5.5 shows the retrieved summaries of environment article and there evaluation results. Table 5.5(a): shows the retrieved summary for environment using rooted-stem as a basic unit and its evaluation results. text of the article. Table 5.5(b) shows the retrieved summary for environment using light-stem as a basic unit and its evaluation results. 5.5(c) shows the retrieved summary for environment using no-stem as a basic unit and its evaluation results**.**

Tables 5.6 shows the retrieved summaries of health article and there evaluation results. Table 5.6(a): shows the retrieved summary for health using rooted-stem as a basic unit and its evaluation results. text of the article. Table 5.6(b) shows the retrieved summary for health using light-stem as a basic unit and its evaluation results. 5.6(c) shows the retrieved summary for health using no-stem as a basic unit and its evaluation results.

Table 5.2 (a): Art and music example original text.

Original Text
الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ (إذا استثنينا الآلات الإيقاعية (وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة . **والناي كلمة فارسية تعني المزمار .** **هي قصبة مفتوحة الطرفين يعزف عليها بواسطة وضع الفم على أحد طرفيها مع إمالته قليلا بزاوية مما يجعل الهواء يصطدم بجدارها الداخلي مصدرا بالنتيجة صوتا شجيا هو أقرب الأصوات وأجملها بالنسبة للإنسان .** **وللناي ستة ثقوب (وأحيانا سبعه) وثقب في منتصف القصبة من الأسفل.** **وتسد هذه الثقوب وتفتح حسب درجة الصوت ولإخراج العلامات بتسلسل يستطيع معه العازف إخراج العلامات**

الموسيقية لإخراج اللحن المطلوب .

والثقب الخلفي يسد بالإبهام ويستخدم لإظهار جواب العلامة الدنيا التي تظهر في البداية .

وتحتاج هذه الآلة إلى براعة شديدة حيث لها 3 تقنيات.

التقنية الأولى : هي طريقة النفخ حيث أن إخراج الصوت الطبيعي منها هي أول صعوبة يجب التغلب عليها لمن أراد التعلم عليها . لذلك ينصح عادة بأن يتمرن من يريد التعلم بالتدريب على إخراج الصوت أولا ومن ثم عندما يستطيع ذلك يبدأ بالتعلم على إخراج الدرجات الصوتية (تمرين الأصابع) .

والعازف الخبير يستطيع بتغيير طريقة النفخ التلاعب بهذه الآلة الخطيرة حيث يستطيع العازف المتمكن أن يخرج أكثر من سبع علامات صحيحة أو حتى أكتر من أوكتاف (ديوان) فالآلة تنتج 7 أصوات صحيحة تماما وبالتالي يستطيع العازف الخبير أن ينتج 7 أصوات أخفض و 7 أصوات أعلى .

التقنية الثانية : هي إمالة الشبابة بأكثر من زاوية لإخراج أصوات معينة أو ربع التون .

التقنية الثالثة : وهي طريقة سد الثقوب بحيث ينتج الأصوات المطلوبة أو سد الثقوب بطريقة معينة لإنتاج ربع التون (العلامات الشرقية) .

وآلة الناي آلة أساسية في التخت الشرقي التقليدي حيث أن صوتها قريب جدا إلى الأذن البشرية .

وقد اعتاد إخواننا المصريين إضافة عدة آلات تشبهها بالطريقة (مع الاختلاف بطبيعة الصوت) ومنها آلة الكولة ولكنهم لم يلغوا هذه الآلة العظيمة .

عيوب هذه الآلة كثيرة أبرزها أن صوتها ليس ثابتا لذلك تعتمد كثيرا على أذن العازف الذي يجب ان يكون بارعا وذو أذن صافية ممتازة حتى تؤدي دورها الصحيح .

والعيب الثاني أن لكل ناي درجة صوتية يبدأ منها ولذلك تجد أن العازف عادة ما يملك أكثر من ناي وعادة ما يكون عددها سبع وإن كان بعض العازفين المهرة جدا يكتفون بأربع أو ثلاث قصبات ويتحايلون على بقية المقامات.

Table 5.2(b): First summary listed in EASC corpus.

Summary A
الناي آلة نفخية تعد بحق اقدم آلة موسيقية في التاريخ (اذا استثنينا الآلات الايقاعية (وللناي عدة اسماء تعرف بها منها الناي القصبة الشبابة المنجيرة . التقنية الاولى : هي طريقة النفخ حيث ان اخراج الصوت الطبيعي منها هي اول صعوبة يجب التغلب عليها لمن اراد التعلم عليها . والعازف الخبير يستطيع بتغيير طريقة النفخ التلاعب بهذه الآلة الخطيرة حيث يستطيع العازف المتمكن ان يخرج اكثر من سبع علامات صحيحة او حتى اكتر من اوكتاف (ديوان) فالآلة تنتج 7 اصوات صحيحة تماما وبالتالي يستطيع العازف الخبير ان ينتج 7 اصوات اخفض و 7 اصوات. التقنية الثانية : هي امالة الشبابة باكثر من زاوية لاخراج اصوات معينة او ربع التون . التقنية الثالثة : وهي طريقة سد الثقوب بحيث ينتج الاصوات المطلوبة او سد الثقوب بطريقة معينة لانتاج ربع التون (العلامات الشرقية) . وقد اعتاد اخواننا المصريين اضافة عدة آلات تشبهها بالطريقة (مع الاختلاف بطبيعة الصوت) ومنها آلة الكولة ولكنهم لم يلغوا هذه الآلة العظيمة . عيوب هذه الآلة كثيرة ابرزها ان صوتها ليس ثابتا لذلك تعتمد كثيرا على اذن العازف الذي يجب ان يكون بارعا وذو اذن صافية ممتازة حتى تؤدي دورها الصحيح . والعيب الثاني ان لكل ناي درجة صوتية يبدا منها ولذلك تجد ان العازف عادة ما يملك اكثر من ناي وعادة ما يكون عددها سبع وان كان بعض العازفين المهرة جدا يكتفون باربع او ثلاث قصبات ويتحايلون على بقية المقامات.

Table 5.2(c): second summary listed in EASC corpus.

Summary B
الناي آلة نفخية تعد بحق اقدم آلة موسيقية في التاريخ (اذا استثنينا الآلات الايقاعية (وللناي عدة اسماء تعرف بها منها الناي القصبة الشبابة المنجيرة . وللناي ستة ثقوب (واحيانا سبعه) وثقب في منتصف القصبة من الاسفل. والثقب الخلفي يسد بالابهام ويستخدم لاظهار جواب العلامة الدنيا التي تظهر في البداية . لذلك ينصح عادة بان يتمرن من يريد التعلم بالتدريب على اخراج الصوت اولا ومن ثم عندما يستطيع ذلك يبدا بالتعلم على اخراج الدرجات الصوتية (تمرين الاصابع) . التقنية الثالثة : وهي طريقة سد الثقوب بحيث ينتج الاصوات المطلوبة او سد الثقوب بطريقة معينة لانتاج ربع التون (العلامات الشرقية) . والعيب الثاني ان لكل ناي درجة صوتية يبدا منها ولذلك تجد ان العازف عادة ما يملك اكثر من ناي وعادة ما يكون عددها سبع وان كان بعض العازفين المهرة جدا يكتفون باربع او ثلاث قصبات ويتحايلون على بقية المقامات.

Table 5.2(d): third summary listed in EASC corpus.

Summary C
الناي آلة نفخية تعد بحق اقدم آلة موسيقية في التاريخ (اذا استثنينا الآلات الايقاعية (وللناي عدة اسماء تعرف بها منها الناي القصبة الشبابة المنجيرة . وللناي ستة ثقوب (واحيانا سبعه) وثقب في منتصف القصبة من الاسفل. والثقب الخلفي يسد بالابهام ويستخدم لاظهار جواب العلامة الدنيا التي تظهر في البداية . لذلك ينصح عادة بان يتمرن من يريد التعلم بالتدريب على اخراج الصوت اولا ومن ثم عندما يستطيع ذلك يبدا بالتعلم على اخراج الدرجات الصوتية (تمرين الاصابع) . التقنية الثالثة : وهي طريقة سد الثقوب بحيث ينتج الاصوات المطلوبة او سد الثقوب بطريقة معينة لانتاج ربع التون (العلامات الشرقية) . والعيب الثاني ان لكل ناي درجة صوتية يبدا منها ولذلك تجد ان العازف عادة ما يملك اكثر من ناي وعادة ما يكون عددها سبع وان كان بعض العازفين المهرة جدا يكتفون باربع او ثلاث قصبات ويتحايلون على بقية المقامات.

Table 5.3(a): retrieved summary for art and music article when using rooted-stem as a basic unit.

Summary for Rooted stem as a basic unit
الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ (إذا استثنينا الآلات الإيقاعية (وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة. لذلك ينصح عادة بأن يتمرن من يريد التعلم بالتدريب على إخراج الصوت أولا ومن ثم عندما يستطيع ذلك يبدأ بالتعلم على إخراج الدرجات الصوتية (تمرين الأصابع). التقنية الثالثة : وهي طريقة سد الثقوب بحيث ينتج الأصوات المطلوبة أو سد الثقوب بطريقة معينة لإنتاج ربع التون (العلامات الشرقية). وآلة الناي آلة أساسية في التخت الشرقي التقليدي حيث أن صوتها قريب جدا إلى الأذن البشرية . وقد اعتاد إخواننا المصريين إضافة عدة آلات تشبهها بالطريقة (مع الاختلاف بطبيعة الصوت) ومنها آلة الكولة ولكنهم لم يلغوا هذه الآلة العظيمة . عيوب هذه الآلة كثيرة أبرزها أن صوتها ليس ثابتا لذلك تعتمد كثيرا على أذن العازف الذي يجب ان يكون بارعا وذو أذن صافية ممتازة حتى تؤدي دورها الصحيح . والعيب الثاني أن لكل ناي درجة صوتية يبدأ منها ولذلك تجد أن العازف عادة ما يملك أكثر من ناي وعادة ما يكون عددها سبع وإن كان بعض العازفين المهرة جدا يكتفون بأربع أو ثلاث قصبات ويتحايلون على بقية المقامات.

	Recall	Precision	F-measure
Summary A	**5/8**	**5/7**	**0.667**
Summary B	**4/6**	**4/7**	**0.615**
Summary C	**4/5**	**4/7**	**0.667**
Average	**0.697**	**0.619**	**0.649**

Table 5.3(b): retrieved summary for art and music article when using light-stem as a basic unit.

Summary for light stem as a basic unit
وتسد هذه الثقوب وتفتح حسب درجة الصوت ولإخراج العلامات بتسلسل يستطيع معه العازف إخراج العلامات الموسيقية لإخراج اللحن المطلوب. لذلك ينصح عادة بأن يتمرن من يريد التعلم بالتدريب على إخراج الصوت أولا ومن ثم عندما يستطيع ذلك يبدأ بالتعلم على إخراج الدرجات الصوتية (تمرين الأصابع) . التقنية الثانية : هي إمالة الشبابة بأكثر من زاوية لإخراج أصوات معينة أو ربع التون . التقنية الثالثة : وهي طريقة سد الثقوب بحيث ينتج الأصوات المطلوبة أو سد الثقوب بطريقة معينة لإنتاج ربع التون (العلامات الشرقية) . عيوب هذه الآلة كثيرة أبرزها أن صوتها ليس ثابتا لذلك تعتمد كثيرا على أذن العازف الذي يجب ان يكون بارعا وذو أذن صافية ممتازة حتى تؤدي دورها الصحيح. والعيب الثاني أن لكل ناي درجة صوتية يبدأ منها ولذلك تجد أن العازف عادة ما يملك أكثر من ناي وعادة ما يكون عددها سبع وإن كان بعض العازفين المهرة جدا يكتفون بأربع أو ثلاث قصبات ويتحايلون على بقية المقامات.

	Recall	Precision	F-measure
Summary A	**4/8**	**4/6**	**0.571**
Summary B	**3/6**	**3/6**	**0.5**
SummaryC	**2/5**	**2/6**	**0.364**
Average	**0.467**	**0.5**	**0.478**

Table 5.3(c): retrieved summary for art and music article when using no-stem as a basic unit.

Summary for a word as a basic unit
الناي آلة نفخية تعد بحق أقدم آلة موسيقية في التاريخ (إذا استثنينا الآلات الإيقاعية (وللناي عدة أسماء تعرف بها منها الناي القصبة الشبابة المنجيرة. وتحتاج هذه الآلة إلى براعة شديدة حيث لها 3 تقنيات. التقنية الأولى : هي طريقة النفخ حيث أن إخراج الصوت الطبيعي منها هي أول صعوبة يجب التغلب عليها لمن أراد التعلم عليها. لذلك ينصح عادة بأن يتمرن من يريد التعلم بالتدريب على إخراج الصوت أولا ومن ثم عندما يستطيع ذلك يبدأ بالتعلم على إخراج الدرجات الصوتية (تمرين الأصابع). التقنية الثالثة : وهي طريقة سد الثقوب بحيث ينتج الأصوات المطلوبة أو سد الثقوب بطريقة معينة لإنتاج ربع التون (العلامات الشرقية). عيوب هذه الآلة كثيرة أبرزها أن صوتها ليس ثابتا لذلك تعتمد كثيرا على أذن العازف الذي يجب ان يكون بارعا وذو أذن صافية ممتازة حتى تؤدي دورها الصحيح. والعيب الثاني أن لكل ناي درجة صوتية يبدأ منها ولذلك تجد أن العازف عادة ما يملك أكثر من ناي وعادة ما يكون عددها سبع وإن كان بعض العازفين المهرة جدا يكتفون بأربع أو ثلاث قصبات ويتحايلون على بقية المقامات.

	Recall	Precision	F-measure
Summary A	**5/8**	**5/7**	**0.667**
Summary B	**4/6**	**4/7**	**0.615**
Summary C	**4/5**	**4/7**	**0.667**
Average	**0.697**	**0.619**	**0.65**

Table 5.4(a): Retrieved summary for sports article when using rooted-stem as a basic unit and its evaluation results.

Summary for Rooted stem as a basic unit			
أن أصل نشأة الكاراتيه في اليابان على جزيرة اوكيناوا وسبب موقعها الجغرافى الممتاز حيث قربها على الصين ادى إلى تفاعل تجارة ناجحة مع جيرانها الصينيين مما جعل بعضا من ثقافتة متأثر قليلا في ثقافة الصين وفنونها القتالية . كان الاوكيناويين يسافرون إلى الصين لتعلم فنون القتال وتحديدا الوشوو حيث طورو الكاراتيه من فنونهم القتالية المحلية وبعض من فنون قتالية صينة وهي الوشوو بتحديد . وعندما ارادت اليابان التوحد في فترة ميجى واسقاط الحكم الاقطاعى في كانت هناك بعض المناطق في اليابان لا بد وان توحد بالقوة ومن هذه المناطق ارخبيل ريوكو الذي يوجد فيه ولاية اوكيناوا التي ظهر فيها فن الكاراتيه . فقد هاجمتها قوات هائلة من الساموراى عام 1600م واسقطت النظام الاقطاعى الحاكم للارخبيل وعند ذلك اجرت قوات الساموراى عملية حظر السلاح على سكان الجزيرة خوفا من الثورات ضدهم فاراد الاوكيناويين ان يستعملو اعضاء جسدهم كسلاح للدفاع عن النفس على جزيرة اوكيناوا.			
	Recall	**Precision**	**F-measure**
Summary A	**3/4**	**3/4**	**0.75**
Summary B	**2/3**	**2/4**	**0.57143**
Summary C	**2/4**	**2/4**	**0.5**
Average	**0.639**	**0.583**	**0.607**

Table 5.4(b): Retrieved summary for sports article when using light-stem as a basic unit and its evaluation results.

Summary for light stem as a basic unit			
كان الاوكيناويين يسافرون إلى الصين لتعلم فنون القتال وتحديدا الوشوو حيث طورو الكاراتيه من فنونهم القتالية المحلية وبعض من فنون قتالية صينة وهي الوشوو بتحديد . وعندما ارادت اليابان التوحد في فترة ميجى واسقاط الحكم الاقطاعى في كانت هناك بعض المناطق في اليابان لا بد وان توحد بالقوة ومن هذه المناطق ارخبيل ريوكو الذي يوجد فيه ولاية اوكيناوا التي ظهر فيها فن الكاراتيه . فقد هاجمتها قوات هائلة من الساموراى عام 1600م واسقطت النظام الاقطاعى الحاكم للارخبيل وعند ذلك اجرت قوات الساموراى عملية حظر السلاح على سكان الجزيرة خوفا من الثورات ضدهم فاراد الاوكيناويين ان يستعملو اعضاء جسدهم كسلاح للدفاع عن النفس على جزيرة اوكيناوا . بعد ذلك اسس الخبراء اليابانيون للكاراتيه وفوناكشي جيتشين جمعية اليابان للكاراتيه (Japan Karate Association) ومن هذا المنطلق اشاعو اليابانيون الكاراتيه فى جميع اليابان على الجزر الرئيسية والعالم الخارجي .			
	Recall	**Precision**	**F-measure**
Summary A	**2/4**	**2/4**	**0.5**
Summary B	**1/4**	**1/3**	**0.28571**
Summary C	**2/4**	**2/4**	**0.5**
Average	**0.417**	**0.44**	**0.429**

Table 5.4(c): Retrieved summary for sports article when using no-stem as a basic unit and its evaluation results.

Summary for a word as a basic unit			
ان اصل نشاه الكاراتيه في اليابان علي جزيره اوكيناوا وسبب موقعها الجغرافي الممتاز حيث قربها علي الصين ادي الي تفاعل تجاره ناجحه مع جيرانها الصينيين مما جعل بعضا من ثقافته متاثر قليلا في ثقافه الصين وفنونها القتاليه كان الاوكيناويين يسافرون الي الصين لتعلم فنون القتال وتحديدا الوشوو حيث طورو الكاراتيه من فنونهم القتاليه المحليه وبعض من فنون قتاليه صينه وهي الوشوو بتحديد وعندما ارادت اليابان التوحد في فتره ميجي واسقاط الحكم الاقطاعي في كانت هناك بعض المناطق في اليابان لا بد وان توحد بالقوه ومن هذه المناطق ارخبيل ريوكو الذي يوجد فيه ولايه اوكيناوا التي ظهر فيها فن الكاراتيه بعد ذلك اسس الخبراء اليابانيون للكاراتيه وفوناكشي جيتشين جمعيه اليابان للكاراتيه (Japan Karate Association) ومن هذا المنطلق اشاعو اليابانيون الكاراتيه في جميع اليابان علي الجزر الرئيسيه والعالم الخارجي			
	Recall	**Precision**	**F-measure**
Summary A	**3/4**	**3/4**	**0.75**
Summary B	**2/3**	**2/4**	**0.57143**
Summary C	**1/4**	**1/4**	**0.25**
Average	**0.556**	**0.5**	**0.5238**

Table 5.5(a): Retrieved summary for environment article when using root-stem as a basic unit and its evaluation results.

Summary for Rooted stem as a basic unit			
أطلق علية ابتداء من القرن الأول الميلادي لقب ملك الغابة، ومن أسماء الأسد في اللغة العربية السبع والليث والهزبر والورد والضرغام وأسامة ويسمى بيته عرين . تعيش الآن معظم الجمهرات في إفريقيا الوسطى حيث يظهر أن اعدادها تتناقص بإستمرار، فقد اظهرت إحدى البحوث تراجع اعدادها من حوالي 100،000 في أوائل التسعينات من القرن العشرين إلى حوالي 16،000 إلى 30،000 أسد برّي حالياً . كانت الأسود الآسيويّة تنتشر من تركيا إلى الهند عبر إيران ، ومن القوقاز حتى اليمن . أما الآن فإن ما تبقى منها يعيش في غابة شمال غربي الهند الواقعة في ولاية غوجارات، حيث يعيش 300 أسد في المنطقة المحميّة البالغة مساحتها 1412 كم2 .			
	Recall	**Precision**	**F-measure**
Summary A	**1/1**	**1/4**	**0.4**
Summary B	**4/8**	**4/4**	**0.6667**
Summary C	**4/8**	**4/4**	**0.6667**
Average	**0.6667**	**0.75**	**0.5778**

Table 5.5(b): Retrieved summary for environment article when using light-stem as a basic unit and its evaluation results.

Summary for light stem as a basic unit			
أطلق علية ابتداء من القرن الأول الميلادي لقب ملك الغابة، ومن أسماء الأسد في اللغة العربية السبع والليث والهزبر والورد والضرغام وأسامة ويسمى بيته عرين . تعيش الآن معظم الجمهرات في إفريقيا الوسطى حيث يظهر أن اعدادها تتناقص بإستمرار، فقد اظهرت إحدى البحوث تراجع اعدادها من حوالي 100،000 في أوائل التسعينات من القرن العشرين إلى حوالي 16،000 إلى 30،000 أسد برّي حالياً . كانت الأسود الآسيويّة تنتشر من تركيا إلى الهند عبر إيران ، ومن القوقاز حتى اليمن . أما الآن فإن ما تبقى منها يعيش في غابة شمال غربي الهند الواقعة في ولاية غوجارات، حيث يعيش 300 أسد في المنطقة المحميّة البالغة مساحتها 1412 كم2 .			
	Recall	**Precision**	**F-measure**
Summary A	**1/1**	**1/4**	**0.4**
Summary B	**4/8**	**4/4**	**0.6667**
Summary C	**4/8**	**4/4**	**0.6667**
Average	**0.6667**	**0.75**	**0.5778**

Table 5.5(c): Retrieved summary for environment article when using no-stem as a basic unit and its evaluation results.

Summary for a word as a basic unit			
اطلق عليه ابتداء من القرن الاول الميلادي لقب ملك الغابه، ومن اسماء الاسد في اللغه العربيه السبع والليث والهزبر والورد والضرغام واسامه ويسمي بيته عرين تعيش الان معظم الجمهرات في افريقيا الوسطي حيث يظهر ان اعدادها تتناقص باستمرار، فقد اظهرت احدي البحوث تراجع اعدادها من حوالي 100،000 في اوائل التسعينات من القرن العشرين الي حوالي 16،000 الي 30،000 اسد بري حاليا كانت الاسود الاسيويه تنتشر من تركيا الي الهند عبر ايران ، ومن القوقاز حتي اليمن اما الان فان ما تبقي منها يعيش في غابه شمال غربي الهند الواقعه في ولايه غوجارات، حيث يعيش 300 اسد في المنطقه المحميه البالغه مساحتها 1412 كم2			
	Recall	**Precision**	**F-measure**
Summary A	**1/1**	**1/4**	**0.4**
Summary B	**4/8**	**4/4**	**0.6667**
Summary C	**4/8**	**4/4**	**0.6667**
Average	**0.6667**	**0.75**	**0.5778**

Table 5.6(a): Retrieved summary for health article when using rooted-stem as a basic unit and its evaluation results.

Summary for Rooted stem as a basic unit			
وبعد أن يمر الطعام من الاثنى عشر يصبح صالحاً للامتصاص حيث تتم هذه العملية داخل تلافيف الأمعاء الرفيعة وبنسبة ضئيلة في الأمعاء الغليظة . وتسمى نهاية القولون بالمستقيم الذي يبلغ طوله حوالي 6 بوصات ويقع في تجويف الجزء العجزي من العمود الفقري . وينتهي المستقيم بقناة الشرج التي تكون مقفلة عادة بواسطة عضلة مستديرة قوية تسمى عضلة فتحة الشرج . ويستغرق الطعام مدة 24 ساعة قبل المرور خارج القناة الهضمية. فالمعدة لاتقبل أي طعام بارد جدا فهي تتضرر من دالك وكدالك الطعام الساخن فهو يصيبها بقرحة المعدة .			
	Recall	**Precision**	**F-measure**
Summary A	**2/5**	**2/5**	**0.4**
Summary B	**2/4**	**2/5**	**0.4444**
Average	**0.45**	**0.4**	**0.4222**

Table 5.6(b): Retrieved summary for health article when using light-stem as a basic unit and its evaluation results.

Summary for light stem as a basic unit			
وبعد أن يمر الطعام من الاثنى عشر يصبح صالحاً للامتصاص حيث تتم هذه العملية داخل تلافيف الأمعاء الرفيعة وبنسبة ضئيلة في الأمعاء الغليظة . وتسمى نهاية القولون بالمستقيم الذي يبلغ طوله حوالي 6 بوصات ويقع في تجويف الجزء العجزي من العمود الفقري . وينتهي المستقيم بقناة الشرج التي تكون مقفلة عادة بواسطة عضلة مستديرة قوية تسمى عضلة فتحة الشرج . وتصل نفايات الطعام إلى القولون على هيئة نصف سائل حيث لا يسمح الجسم بخروجها على هذه الهيئة فيقوم القولون بامتصاص معظم السائل من هذه الفضلات ثم يخرج الباقي على هيئة براز .			
	Recall	**Precision**	**F-measure**
Summary A	**3/5**	**3/4**	**0.6667**
Summary B	**2/4**	**2/4**	**0.5**
Average	**0.55**	**0.625**	**0.5834**

Table 5.6(c): Retrieved summary for health article when using no-stem as a basic unit and its evaluation results.

Summary for a word as a basic unit			
وبعد ان يمر الطعام من الاثني عشر يصبح صالحا للامتصاص حيث تتم هذه العمليه داخل تلافيف الامعاء الرفيعه وبنسبه ضئيله في الامعاء الغليظه وتسمي نهايه القولون بالمستقيم الذي يبلغ طوله حوالي 6 بوصات ويقع في تجويف الجزء العجزي من العمود الفقري وينتهي المستقيم بقناه الشرج التي تكون مقفله عاده بواسطه عضله مستديره قويه تسمي عضله فتحه الشرج وتصل نفايات الطعام الي القولون علي هيئه نصف سائل حيث لا يسمح الجسم بخروجها علي هذه الهيئه فيقوم القولون بامتصاص معظم السائل من هذه الفضلات ثم يخرج الباقي علي هيئه براز			
	Recall	**Precision**	**F-measure**
Summary A	**3/5**	**3/4**	**0.6667**
Summary B	**2/4**	**2/4**	**0.5**
Average	**0.55**	**0.625**	**0.5834**

We going to discuss these results. Now we present the summary tables for our works. Table 5.7 shows the detailed results for our work. Then table 5.8 shows the summarized result table then we show the graphs that that result from GBATSS.

Table 5.7 shows the evaluation of detailed results for summary that uses rooted stemmer as a basic unit.
Table 5.8 shows the evaluation of detailed results for summary that uses light stemmer as a basic unit.
Table 5.9 shows the evaluation of detailed results for summary that uses word without any type of stemming as basic unit.

Table 5.7 The detailed evaluation results of Rooted stemmer basic units

		Recall	Precision	F-measure
1	art 1	0.697	0.619	0.649
2	education1	1	0.5	0.6667
3	environment1	0.528	0.8333	0.644
4	environment2	0.667	0.75	0.5778
5	financial1	0.9	0.6	0.7026
6	financial2	0.833	0.5625	0.6667
7	financial3	0.833	0.6667	0.7222
8	health1	0.5	0.6667	0.57143
9	health2	0.45	0.4	0.4222
10	politics1	0.558	0.4375	0.4361
11	sports1	0.639	0.583	0.607
	average	0.691	0.6017	0.6059755

Table 5.8 : The detailed evaluation results of light stemmer basic unit.

		Recall	Precision	F-measure
1	art 1	0.467	0.5	0.478
2	education1	1	0.5	0.6667
3	environment1	0.3889	0.6667	0.4889

4	**environment2**	**0.6667**	**0.75**	**0.5778**
5	**financial1**	**0.7542**	**0.4167**	**0.5224**
6	**financial2**	**0.8333**	**0.5625**	**0.6667**
7	**financial3**	**0.6667**	**0.5**	**0.5556**
8	**health1**	**0.5**	**0.6667**	**0.57143**
9	**health2**	**0.55**	**0.625**	**0.5834**
10	**politics1**	**0.5583**	**0.4375**	**0.4361**
12	**sports1**	**0.417**	**0.44**	**0.429**
	average	**0.618373**	**0.551373**	**0.543275**

Table 5.9: The detailed evaluation results of word as a basic unit.

		Recall	**Precision**	**F-measure**
1	**art1**	**0.697**	**0.619**	**0.65**
2	**education1**	**1**	**0.5**	**0.6667**
3	**env1**	**0.3889**	**0.6667**	**0.4889**
4	**env2**	**0.6667**	**0.75**	**0.5778**
5	**fin1**	**0.9**	**0.6**	**0.7026**
6	**finance2**	**0.75**	**0.5**	**0.5952**
7	**finance3**	**0.667**	**0.5**	**0.5556**
8	**health1**	**0.5**	**0.6667**	**0.57143**
9	**health2**	**0.55**	**0.625**	**0.5834**
10	**politics1**	**0.5583**	**0.4375**	**0.4361**
11	**sports1**	**0.556**	**0.5**	**0.5238**
	average	**0.657627**	**0.578627**	**0.577412**

Table 5.10 shows the final results of evaluation methods that are used in evaluating the system.

Table 5.10: The final results of evaluation methods to all application's results.

	Recall	Precision	F-measure
Rooted	**0.691**	**0.602**	**0.606**
Light	**0.618**	**0.551**	**0.543**
Word	**0.658**	**0.579**	**0.577**

Now we present the charts that describe our works. After that in the next section we present the comparison between GBATSS and another graph based summarization system.

Figure 5.2 shows the chart of the recall evaluation parameter. This chart views the difference in recall parameter among the summaries that resulted from the different types of basic units. According to the chart, we find that when we use rooted-stem as a basic unit, we get the highest recall and when we use the light-stem as a basic unit, we get the lowest recall.

Figure 5.3 shows the chart of the precision evaluation parameter. This chart views the difference in recall parameter among the summaries that resulted from the different types of basic units. According to the chart, we find that when we use rooted-stem as a basic unit, we get the highest precision and when we use the light-stem as a basic unit, we get the lowest precision. In addition we find that the value of precision is greater that recall because the number of sentences differ between relevant and relative summaries.

Figure 5.4 shows the chart of the f-measure evaluation parameter. This chart views the difference in f-measure parameter among the summaries that resulted from the different types of basic units. According to the chart, we find that when we use rooted-stem as a basic unit, we get the highest f-measure and when we use the light-stem as a basic unit, we get the lowest f-measure.

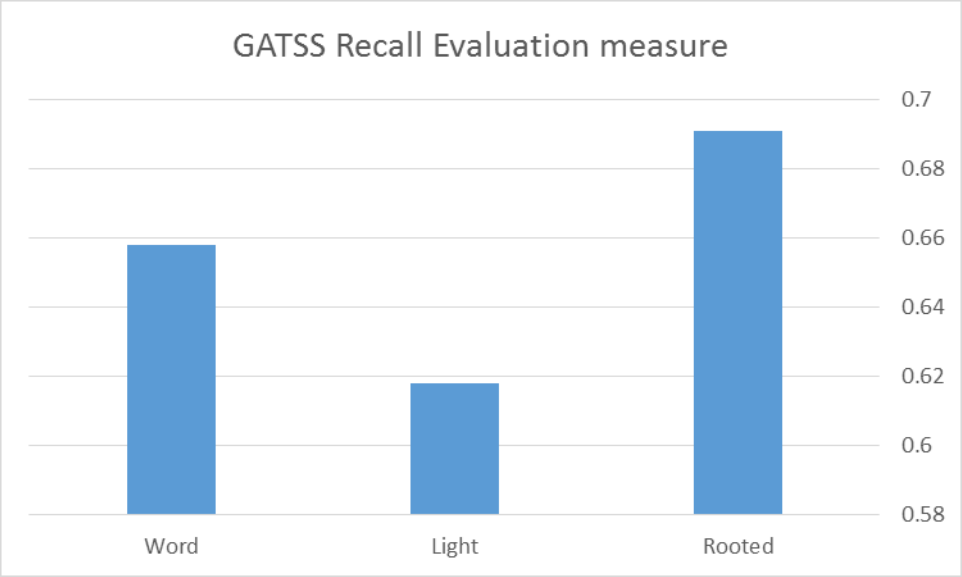

Figure 5.2: GBATSS Recall Evaluation measure.

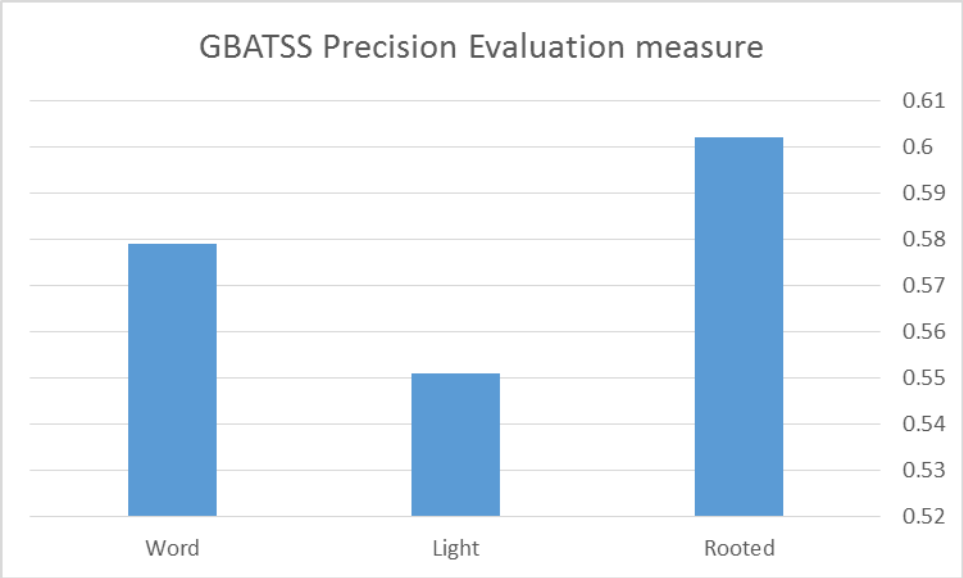

Figure 5.3: GBATSS Precision Evaluation measure.

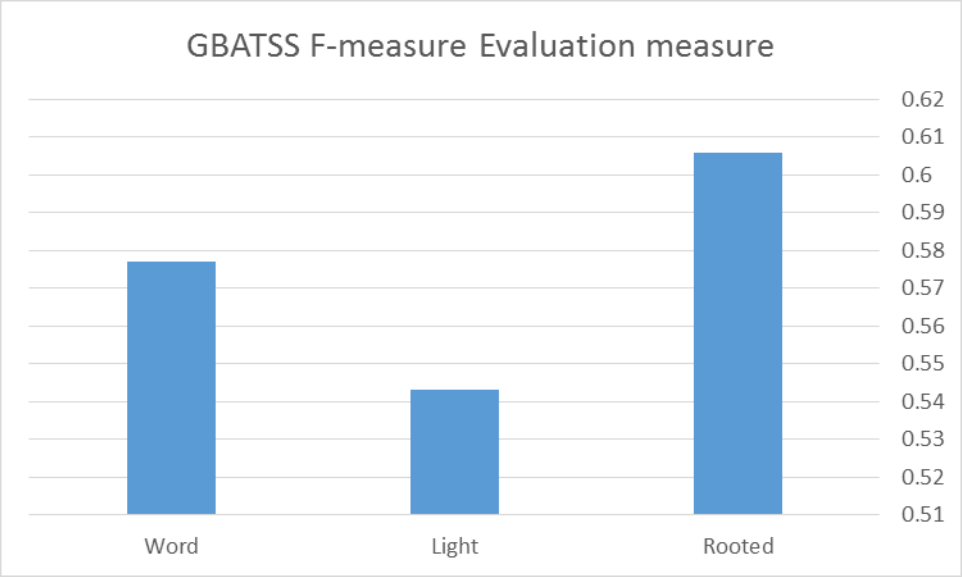

Figure 5.4: GBATSS F-measure evaluation measure.

Figure 5.5 shows a summary for the three evaluation measures that are used in evaluate GBATSS. This chart shows that summary extracted depending on the root-stem basic unit gives us results better than other basic units.

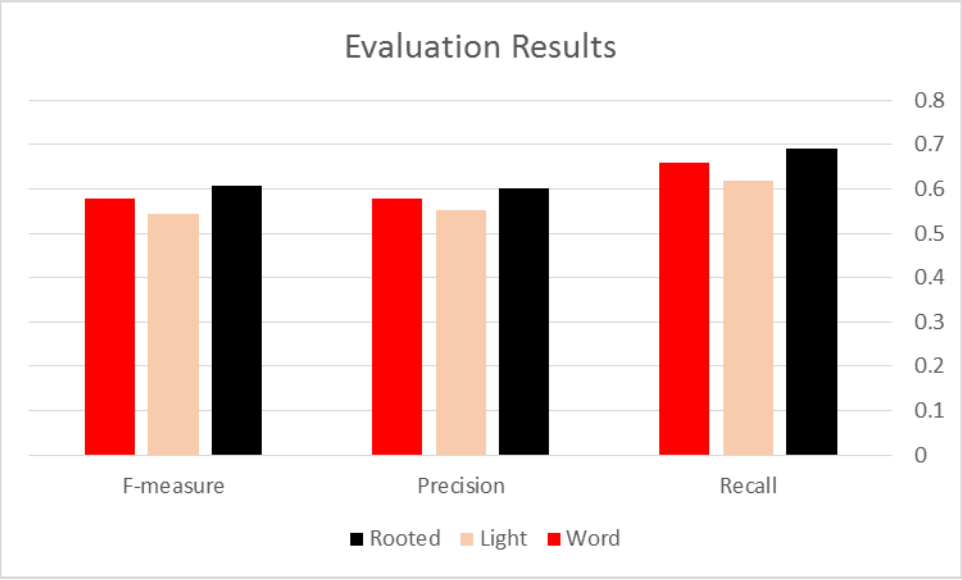

Figure 5.5: Evaluation results

5.4.1 Comparison

By comparing between our application and application done in [1] with same compression rate 40% we find that in f-measure GBATSS have improvement about 15.8%.
In the following figures, we present comparison in recall, precision, and f-measure.

Figure 5.6 shows a comparison chart for recall evaluation measure for two basic units. The first basic unit is rooted stem basic unit and word basic unit. For rooted stem basic unit we find that we have improved about 27%. In word basic unit we have improved about 25%.

Figure 5.7 shows a comparison chart for precision evaluation measure for two basic units the first basic unit is rooted stem basic unit and word basic unit. For rooted stem basic unit we find that we have improve about 13%. In word basic unit we have improve about 11%.

Figure 5.8 shows a comparison chart for evaluation measure for two basic units. The first basic unit is rooted stem basic unit and word basic unit. For rooted stem basic unit we find that we have improved about 16%. In word basic unit we have improved about 15%.

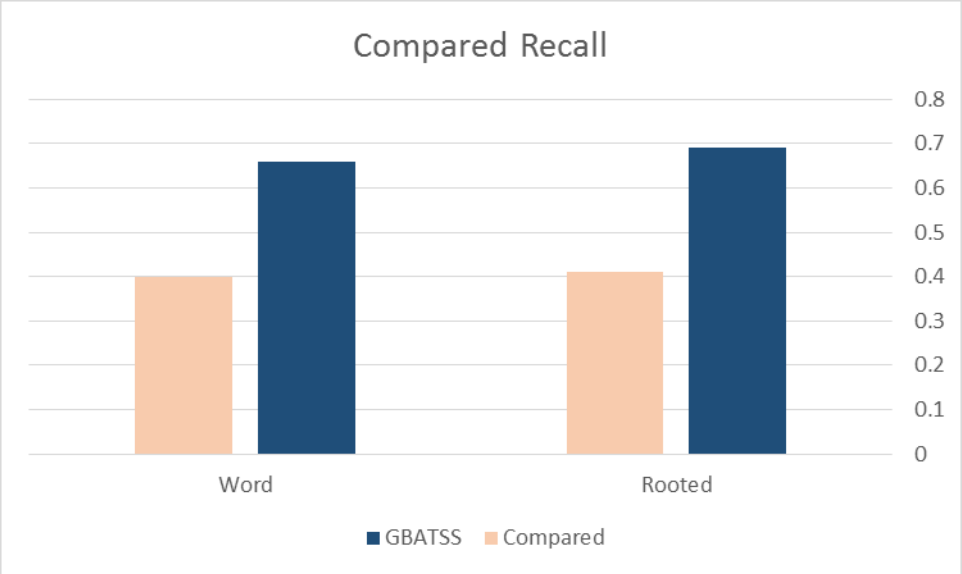

Figure 5.6 compared recall chart

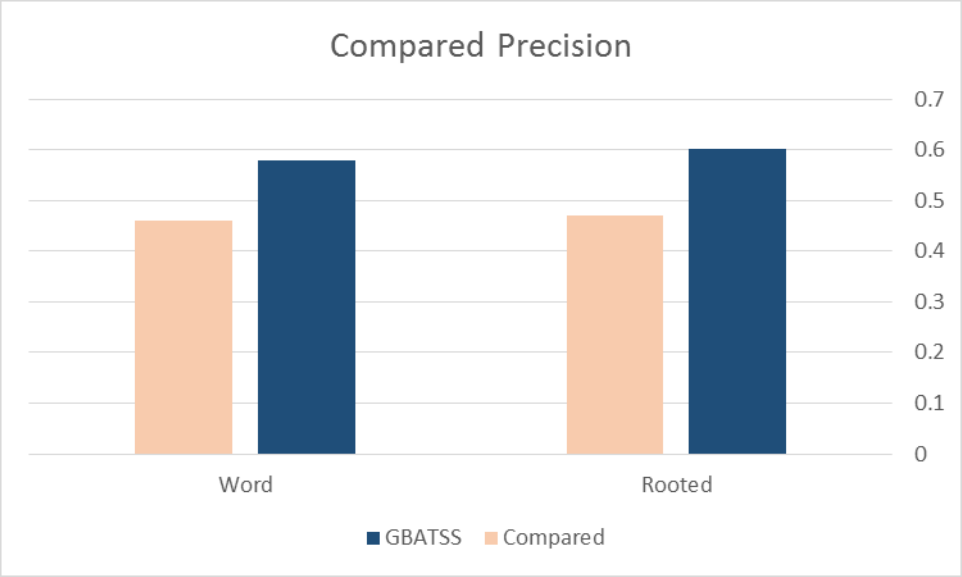

Figure 5.7: Compared precision chart

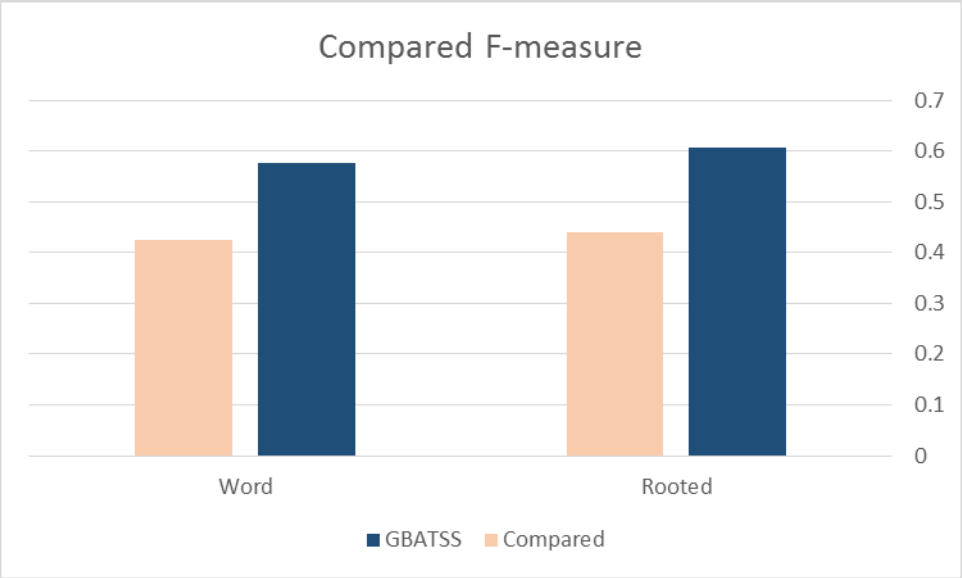

Figure 5.8: Compared F-measure chart

5.4.2 Discussion

From previous discussion and comparison, we found the following:

1. We find that rooted stem basic unit is better than other basic units.

2. GBATSS is outperforming [1] based on F- measure, this is due to :
    a. The improvement in normalization process.
    b. The difference in the stop words list used between the two applications.

3. In some summaries we notice that the resulted summaries have similarity because the content of the basic document vocabulary is near the basic form and has little amount of affixes so the resulted summaries are close enough to each other.

4. In some categories of text articles, we get low evaluation results that's return to the stop words list that are used in the preprocessing stage. Because it was not appropriate to this type of articles.

5. We get an improvement in the summary results because we make some improvements in the preprocessing stage, especially in normalization and stop words list.

CHAPTER Six: CONCLUSION AND FUTURE WORK

6.1 Conclusion

In this thesis, we build an extractive graph based text summarization. The proposed method is based mainly on graph theory with features of terms weighting ,relation between sentences. Then finally we apply Page-Rank algorithm to sort the sentences.

The system works in three basic units. these units are rooted stem , light stem ,and finally word. The system depends on compression ratio of 40 %. The process of summarization is done in many stages starting from data collection , text preprocessing , text normalization, text tokenization , stemming, stop word removals , building graph ,calculating edge weighting ,applying page rank , and finally extracting the summary.

To test and evaluate this system we use EASC data set. After that we use the following evaluation measure recall, precision, and f-measure to evaluate the system.

Then we compare the results with another system .to evaluate the system.
After comparing data with another application we find that, we have improvement about 16% than the other graph based system.

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.

6.2 Future work

Here we propose some techniques to improve our work in future:

1. Using more features like sentence topic relevance, or using another morphological properties.

2. Applying some language morphological techniques to solve the Arabic morphological complex in Arabic language.

3. Using special stop words for every category of text, to enhance the output of the system.

4. Appling modification with page rank to improve the results.

5. Using another basic units like n-grams to find which one is best.

6. Improving stemming techniques in pre-processing stage or using lemmatization technique.

7. Trying to make a hybrid method that uses graph based and rhetorical processing for the text.

8. Improve the pre-processing and normalization process of the system.

9. Try to improve the tokenization process by using other tokenization techniques.

10. Adopting alternative techniques for evaluation that will help better understanding the nature of the summarization problem.

References:

1. Sakai, T. and K. Sparck-Jones. *Generic summaries for indexing in information retrieval*. in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. 2001. ACM.

2. White, M., et al. *Multidocument summarization via information extraction*. in *Proceedings of the first international conference on Human language technology research*. 2001. Association for Computational Linguistics.

3. Ker, S.J. and J.-N. Chen. *A text categorization based on summarization technique*. in *Proceedings of the ACL-2000 workshop on Recent advances in natural language processing and information retrieval: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 11*. 2000. Association for Computational Linguistics.

4. Balage Filho, P.P., et al., *Experiments on applying a text summarization system for question answering*, in *Evaluation of Multilingual and Multi-modal Information Retrieval*. 2007, Springer. p. 372-376.

5. Kan, M.-Y., K.R. McKeown, and J.L. Klavans. *Applying natural language generation to indicative summarization*. in *Proceedings of the 8th European workshop on Natural Language Generation-Volume 8*. 2001. Association for Computational Linguistics.

6. Zhan, J., et al., *Automatic text summarization in engineering information management*, in *Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers*. 2007, Springer. p. 347-350.

7. Saravanan, M., B. Ravindran, and S. Raman, *Improving legal document summarization using graphical models.* Frontiers in Artificial Intelligence and Applications, 2006. **152**: p. 51.

8. Hovy, E. and C.-Y. Lin. *Automated text summarization and the SUMMARIST system*. in *Proceedings of a workshop on held at Baltimore, Maryland: October 13-15, 1998*. 1998. Association for Computational Linguistics.

9. Mani, I., *Automatic summarization*. Vol. 3. 2001: John Benjamins Publishing.

10. Radev, D.R., E. Hovy, and K. McKeown, *Introduction to the special issue on summarization.* Computational linguistics, 2002. **28**(4): p. 399-408.

11. Kupiec, J.M. and H. Schuetze, *System for genre-specific summarization of documents*. 2004, Google Patents.

12. Mihalcea, R. *Graph-based ranking algorithms for sentence extraction, applied to text summarization*. in *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*. 2004. Association for Computational Linguistics.

13. Patil, K. and P. Brazdil, *Sumgraph: Text summarization using centrality in the pathfinder network.* International Journal on Computer Science and Information Systems, 2007. **2**(1): p. 18-32.

14. Wills, R.S., *Google's pagerank.* The Mathematical Intelligencer, 2006. **28**(4): p. 6-11.

15. Amini, M.R., N. Usunier, and P. Gallinari, *Automatic text summarization based on word-clusters and ranking algorithms*, in *Advances in Information Retrieval*. 2005, Springer. p. 142-156.

16. Delort, J.-Y., B. Bouchon-Meunier, and M. Rifqi. *Enhanced web document summarization using hyperlinks*. in *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*. 2003. ACM.

17. Ježek, K. and J. Steinberger. *Automatic text summarization (the state of the art 2007 and new challenges)*. in *Proceedings of Znalosti*. 2008.

18. Nenkova, A. and K. McKeown, *A survey of text summarization techniques*, in *Mining Text Data*. 2012, Springer. p. 43-76.

19. Shaalan 1, K.F., *An intelligent computer assisted language learning system for Arabic learners.* Computer Assisted Language Learning, 2005. **18**(1-2): p. 81-109.

20. Shaalan, K.F., *Arabic GramCheck: A grammar checker for Arabic.* Software: Practice and Experience, 2005. **35**(7): p. 643-665.

21. Farghaly, A. *Subject pronoun deletion rule*. in *Proceedings of the 2nd English Language Symposium on Discourse Analysis (LSDA'82)*. 1982.

22. Frankel, D.S., *Model Driven Architecture Applying Mda*. 2003: John Wiley & Sons.

23. Douzidia, F.S. and G. Lapalme, *Lakhas, an Arabic summarization system.* Proceedings of DUC2004, 2004.

24. El-Haj, M., U. Kruschwitz, and C. Fox, *Experimenting with automatic text summarisation for arabic*, in *Human Language Technology. Challenges for Computer Science and Linguistics*. 2011, Springer. p. 490-499.

25. Al-Shammari, E.T. and J. Lin. *Towards an error-free Arabic stemming*. in *Proceedings of the 2nd ACM workshop on Improving non english web searching*. 2008. ACM.

26. Hassel, M., 2004. Evaluation of automatic text summarization. *Licentiate Thesis, Stockholm, Sweden*, pp.1-75.

27. Das, D. and A.F. Martins, *A survey on automatic text summarization.* Literature Survey for the Language and Statistics II course at CMU, 2007. **4**: p. 192-195.

28. Sobh, I.M.A.H., *An optimized dual classification system for Arabic extractive generic text summarization*. 2009, Faculty of Engineering, Cairo University Giza, Egypt.

29. Powers, D.M., *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.* 2011.

30. Cohen, R., S. Havlin, and D. Ben-Avraham, *Structural properties of scale free networks.* Handbook of graphs and networks, 2003.

31. Brin, S. and L. Page. *The anatomy of a large-scale hypertextual web search engine, 1998*. in *Proceedings of the Seventh World Wide Web Conference*. 2007.

32. Luhn, H.P., *The automatic creation of literature abstracts.* IBM Journal of research and development, 1958. **2**(2): p. 159-165.

33. Jing, H. *Sentence reduction for automatic text summarization*. in *Proceedings of the sixth conference on Applied natural language processing*. 2000. Association for Computational Linguistics.

34. Chuang, W.T. and J. Yang, *Text summarization by sentence segment extraction using machine learning algorithms*, in *Knowledge Discovery and Data Mining. Current Issues and New Applications*. 2000, Springer. p. 454-457.

35. Hirst, G. and D. St-Onge, *Lexical chains as representations of context for the detection and correction of malapropisms.* WordNet: An electronic lexical database, 1998. **305**: p. 305-332.

36. Barzilay, R. and M. Elhadad, *Using lexical chains for text summarization.* Advances in automatic text summarization, 1999: p. 111-121.

37. Mani, I. and E. Bloedorn. *Machine learning of generic and user-focused summarization*. in *AAAI/IAAI*. 1998.

38. Baxendale, P.B., *Machine-made index for technical literature: an experiment.* IBM Journal of Research and Development, 1958. **2**(4): p. 354-361.

39. Edmundson, H.P., *New methods in automatic extracting.* Journal of the ACM (JACM), 1969. **16**(2): p. 264-285.

40. Conroy, J.M. and D.P. O'leary. *Text summarization via hidden markov models*. in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. 2001. ACM.

41. Marcu, D. *Improving summarization through rhetorical parsing tuning*. in *The 6th Workshop on Very Large Corpora*. 1998.

42. Mann, W.C. and S.A. Thompson, *Rhetorical structure theory: A theory of text organization*. 1987: University of Southern California, Information Sciences Institute.

43. Svore, K.M., L. Vanderwende, and C.J. Burges. *Enhancing Single-Document Summarization by Combining RankNet and Third-Party Sources*. in *EMNLP-CoNLL*. 2007.

44. Matveeva, I., et al. *High accuracy retrieval with multiple nested ranker*. in *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. 2006. ACM.

45. Al-Hashemi, R., *Text Summarization Extraction System (TSES) Using Extracted Keywords.* Int. Arab J. e-Technol., 2010. **1**(4): p. 164-168.

46. Aha, D.W., D. Kibler, and M.K. Albert, *Instance-based learning algorithms.* Machine learning, 1991. **6**(1): p. 37-66.

47. Yeh, J.-Y., H.-R. Ke, and W.-P. Yang, *iSpreadRank: Ranking sentences for extraction-based summarization using feature weight propagation in the sentence similarity network.* Expert Systems with Applications, 2008. **35**(3): p. 1451-1462.

48. Litvak, M. and M. Last. *Graph-based keyword extraction for single-document summarization*. in *Proceedings of the workshop on Multi-source Multilingual Information Extraction and Summarization*. 2008. Association for Computational Linguistics.

49. Mihalcea, R. and P. Tarau. *TextRank: Bringing order into texts*. 2004. Association for Computational Linguistics.

50. Li, Y. and K. Cheng. *Single document Summarization based on Clustering Coefficient and Transitivity Analysis*. in *Proceedings of the 10th International Conference on Accomplishments in Electrical and Mechanical Engineering and Information Technology, May*. 2011.

51. Wan, X. *TimedTextRank: adding the temporal dimension to multi-document summarization*. in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. 2007. ACM.

52. Liu, Y., et al. *Personalized PageRank based multi-document summarization*. in *Semantic Computing and Systems, 2008. WSCS'08. IEEE International Workshop on*. 2008. IEEE.

53. Erkan, G. and D.R. Radev, *LexRank: Graph-based lexical centrality as salience in text summarization.* Journal of Artificial Intelligence Research, 2004: p. 457-479.

54. Wan, X. *An exploration of document impact on graph-based multi-document summarization*. in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2008. Association for Computational Linguistics.

55. Abdallah, M., C. ALOULOU, and L. Belguith. *Toward a platform for arabic automatic summarization*. in *Proceedings of the International Arab Conference on Information Technology ACIT*. 2008.

56. Boudabous, M.M., M.H. Maaloul, and L.H. Belguith, *Digital learning for summarizing Arabic documents*, in *Advances in Natural Language Processing*. 2010, Springer. p. 79-84.

57. Hammo, B.H., H. Abu-Salem, and M.W. Evens, *A Hybrid Arabic Text Summarization Technique Based on Text Structure and Topic Identification.* International Journal of Computer Processing Of Languages, 2011. **23**(01): p. 39-65.

58. Fattah, M.A. and F. Ren, *Automatic text summarization.* World Academy of Science, Engineering and Technology, 2008. **37**: p. 2008.

59. Khalifa, I., Z. Feki, and A. Farawila, *Arabic discourse segmentation based on rhetorical methods.* Int. J. Electric Comput. Sci, 2011. **11**(1).

60. AlSanie, W., A. Touir, and H. Mathkour, *Towards an infrastructure for Arabic text summarization using rhetorical structure theory*. 2005, M. Sc. Thesis, King Saud University, Riyadh, Saudi Arabia.

61. Ibrahim, A., T. Elghazaly, and M. Gheith, *A Novel Arabic Text Summarization Model Based on Rhetorical Structure Theory and Vector Space Model.* 2013.

62. Al-Taani, A.T. and M.M. Al-Omour, *An Extractive Graph-based Arabic Text Summarization Approach.*

63. Chakrabarti, S., et al., *Automatic resource compilation by analyzing hyperlink structure and associated text.* Computer Networks and ISDN Systems, 1998. **30**(1): p. 65-74.

64. El-Haj, M., U. Kruschwitz, and C. Fox, *Using Mechanical Turk to create a corpus of Arabic summaries.* 2010.

65. Attia, M.A., *Handling Arabic morphological and syntactic ambiguity within the LFG framework with a view to machine translation*. 2008, University of Manchester.

66. Attia, M.A. *Arabic tokenization system*. in *Proceedings of the 2007 workshop on computational approaches to semitic languages: Common issues and resources*. 2007. Association for Computational Linguistics.

67. El-Khair, I.A., *Effects of stop words elimination for Arabic information retrieval: a comparative study*. International Journal of Computing & Information Sciences, 2006. **4**(3): p. 119-133.

68. Khoja, S. and R. Garside, *Stemming arabic text.* Lancaster, UK, Computing Department, Lancaster University, 1999.

69. Lovins, J.B., *Development of a stemming algorithm*. 1968: MIT Information Processing Group, Electronic Systems Laboratory.

70. Alguliev, R. and R. Aliguliyev, *Experimental investigating the F-measure as similarity measure for automatic text summarization.* Applied and Computational Mathematics, 2007. **6**(2): p. 278-287.

71. Alguliev, R.M., et al., *MCMR: Maximum coverage and minimum redundant text summarization model.* Expert Systems with Applications, 2011. **38**(12): p. 14514-14522.

Appendix A

In this appendix we present articles and summaries that used in GBATSS. Here we present the original summary then the relevant summaries of the article.

Table A.1(a) : Article 1 oiriginal text

Original Text
تبدأ الغرفة التجارية الصناعية في محافظة الأحساء تنفيذ برامجها التدريبية النسائية اعتباراً من يوم الأحد المقبل وذلك من خلال برنامجين نسائيين بعنوان" فنون التعامل مع الزوج" وذلك في الفترة من 7 - 8 من شهر ربيع الثاني الحالي، والبرنامج الآخر "الحب وحده لا يكفي للأبناء" وذلك في الفترة 9 -10 من شهر ربيع الثاني الحالي، وتقدمهما الدكتورة وداد العيسى من دولة الكويت في مقر الغرفة بمدينة المبرز التابعة للأحساء. وأوضح مدير التدريب والتطوير بالغرفة بدر الحليبي أن المحاضرة ستتناول في البرنامج الأول دراسة لأفضل توافق زواجي في العالم، ومظاهر الأسرة السعيدة، ومكونات الزواج السعيد وغير السعيد، وكيف تقيمين علاقاتك الزوجية، والحاجات العاطفية المطلوبة في الزواج، وقواعد تحسين العلاقة الزوجية، بالإضافة إلى مهارات الاتصال بين الزوجين، وفنون التحدث والاستماع، ومهارات حل النزاعات بين الزوجين. كما تتناول عده محاور في البرنامج الآخر ومن أبرزها التواصل الإيجابي مع الأبناء، وعناصر الاتصال مع الأبناء، وتصنيف الدراسات لأنماط شخصية الأبناء، وكيفية الحوار مع الأبناء، وكيف نثق في أنفسنا وفي الأبناء، وكيف نربي الثقة عند الأبناء، وكيف تكسب الأم صداقة ابنها، بالإضافة إلى أسباب المشاكل مع الأبناء، ومهارات حل مشكلات الأبناء.

Table A.1(b) : Article 1 first Summary

Summary A
تبدا الغرفة التجارية الصناعية في محافظة الاحساء تنفيذ برامجها التدريبيةالنسائية اعتباراً من يوم الاحد المقبل وذلك من خلال برنامجين نسائيين بعنوان"فنون التعامل مع الزوج" وذلك في الفترة من 7 - 8 من شهر ربيع الثاني الحالي،والبرنامج الآخر "الحب وحده لا يكفي ل. واوضح مدير التدريب والتطوير بالغرفة بدر الحليبي ان المحاضرة ستتناول فيالبرنامج الاول دراسة لافضل توافق زواجي في العالم، ومظاهر الاسرة السعيدة،ومكونات الزواج السعيد وغير السعيد، وكيف تقيمين علاقاتك الزوجية، والحاجاتالعاطفية المطلوبة في الزواج، وقواعد تحسي.

Table A.1(c) : Article 1 second Summary

Summary B
تبدا الغرفة التجارية الصناعية في محافظة الاحساء تنفيذ برامجها التدريبيةالنسائية اعتباراً من يوم الاحد المقبل وذلك من خلال برنامجين نسائيين بعنوان"فنون التعامل مع الزوج" وذلك في الفترة من 7 - 8 من شهر ربيع الثاني الحالي،والبرنامج الآخر "الحب وحده لا يكفي ل. كما تتناول عده محاور في البرنامج الآخر ومن ابرزها التواصل الايجابي معالابناء، وعناصر الاتصال مع الابناء، وتصنيف الدراسات لانماط شخصية الابناء،وكيفية الحوار مع الابناء، وكيف نثق في انفسنا وفي الابناء، وكيف نربي الثقة عندالابناء، وكيف تكسب الام صداقة ابنها.

Table A.1(b) : Article 1 third Summary

Summary C
واوضح مدير التدريب والتطوير بالغرفة بدر الحليبي ان المحاضرة ستتناول فيالبرنامج الاول دراسة لافضل توافق زواجي في العالم، ومظاهر الاسرة السعيدة،ومكونات الزواج السعيد وغير السعيد، وكيف تقيمين علاقاتك الزوجية، والحاجاتالعاطفية المطلوبة في الزواج، وقواعد تحسي.

Table A.2(a) : Article 2 original text

Original Text
_طرح مدير مدرسة حبيب بن زيد الأنصاري الابتدائية برنامجا هو الأول من نوعه على مستوى منطقة المدينة المنورة التعليمية، إن لم يكن الأول من نوعه على مستوى السعودية على حسب تصريحه، تمثل في إقامة البرنامج التأهيلي للطلاب المستجدين في المرحلة الابتدائية هذه الأيام ولمدة ثلاثة أسابيع، بدلا من إقامته في بداية العام الدراسي . وقال مدير المدرسة أحمد الجمعان عن التجربة التي لا تزال في بدايتها " لا يمكن خوض غمار هذه التجربة في ظل عدم مساندة وتعاون أولياء أمور التلاميذ، فـ90% من التلاميذ الجدد المسجلين انخرطوا في هذا البرنامج الذي نسعى من خلاله إلى كسر حاجز الخوف، وترغيب الطالب في المدرسة، كما نهدف إلى استثمار الوقت في الإجازة من خلال تفعيل دور الأسرة، إذ نوفر لكل طالب قرصا مدمجا (سي دي) لمنهج القراءة والكتابة، ومذكرة تحتوي على أنشطة تعليمية ." وحضور التلميذ إلى المدرسة للمشاركة في هذا البرنامج غير مقيد بمدة زمنية معينة ,ومن دون حقائب مدرسية، كما يترك فيه للطالب حرية التعرف على مرافق المدرسة ومعلمي الصفوف الأولية، كما يتضمن رحلة أسبوعية إلى مدينة ألعاب، وإجراء مسابقات يومية ترفيهية. يذكر أن عدد الطلاب المسجلين في البرنامج 75 طالبا.

Table A.2(b) : Article 2 first Summary

Summary A
_طرح مدير مدرسة حبيب بن زيد الانصاري الابتدائية برنامجا هو الاول من نوعه علىمستوى منطقة المدينة المنورة التعليمية، ان لم يكن الاول من نوعه على مستوىالسعودية على حسب تصريحه، تمثل في اقامة البرنامج التاهيلي للطلاب المستجدين فيالمرحلة الابتدائية هذه الايام.

Table A.2(c) : Article 2 second Summary

Summary B
_طرح مدير مدرسة حبيب بن زيد الانصاري الابتدائية برنامجا هو الاول من نوعه علىمستوى منطقة المدينة المنورة التعليمية، ان لم يكن الاول من نوعه على مستوىالسعودية على حسب تصريحه، تمثل في اقامة البرنامج التاهيلي للطلاب المستجدين فيالمرحلة الابتدائية هذه الايام.

Table A.2(d) : Article 2 third Summary

Summary C
_طرح مدير مدرسة حبيب بن زيد الانصاري الابتدائية برنامجا هو الاول من نوعه على مستوى منطقة المدينة المنورة التعليمية، ان لم يكن الاول من نوعه على مستوى السعودية على حسب تصريحه، تمثل في اقامة البرنامج التأهيلي للطلاب المستجدين في المرحلة الابتدائية هذه الايام.

Table A.3(a) : Article 3 original text

Original Text
الثلج نوع من الهطولات على شكل بلورات دقيقة للجليد تحدث في الفصل البارد لكنها لا تحدث في كل دول العالم. و تزداد غزارة الثلوج وكثافتها كلما اتجهنا قريباً من القطبين الجنوبي و الشمالي. تتم هذه الظاهرة التقاء تيارات هوائية رطبة ودافئة مع تيارات باردة تكون درجة حرارتها 12.5 تحت الصفر و يجب لتكون

الثلج توفر نويات التكاثف التي يتكون عليها الثلج وهي عبارة عن جسيمات صلبة صغيرة جدا عالقة في الجو العلوي مثل ذرات الغبار أو الرماد وعند وجودها تتوفر الحالات الثلاث التي تمكن بخار الماء ليتحول من الحالة الغازية إلى الحالة الصلبة مكونا الثلج و يتم ذلك بتكاثف جزيئات الماء على النواة و في أثناء التصاقها مع بعضها تتم في العادة بناء بللورة الثلج و تكون في هذه المرحلة عبارة عن صفيحة رقيقة ذات ستة جوانب و عندما تسقط هذه البلورة ترتفع درجة الحرارة يتكاثف عليها قدر أكبر من جزيئات الماء و هكذا تنمو حيث يتفرع من الصفيحة البلورية ست أذرع و في درجات الحرارة الأكثر ارتفاعا تذوب حواف البلورة قليلا و ذلك لإتاحة فرصة الالتصاق مع البلورات الأخرى و بهذا تتكون الصفيحة الثلجية.لطالما اتصفت صورة الثلج في ذهن الإنسان بالنقاء و العفة و الطهارة و ربما جاء هذا من لون الثلج الأبيض الناصع، فمثلا قصة بياض الثلج و الأقزام السبعة كان اسم بطلة القصة و هي الفتاة الطيبة بياض الثلج snow white . و في أحيان أخرى يكون الثلج رمزا للموت البطئ و القاسي و الذي جاء بطبيعة الحال بسبب ترافق الثلج مع تيارات باردة و متجمدة و مع غياب للشمس و هبوط حاد للحرارة.

Table A.3(b) : Article 3 first Summary

Summary A
و تزداد غزارة الثلوج وكثافتها كلما اتجهنا قريباً من القطبين الجنوبي و الشمالي. لطالما اتصفت صورة الثلج في ذهن الإنسان بالنقاء و العفة و الطهارة و ربما جاء هذا من لون الثلج الأبيض الناصع، فمثلا قصة بياض الثلج و الأقزام السبعة كان اسم بطلة القصة و هي الفتاة الطيبة بياض الثلج. snow white و في أحيان أخرى يكون الثلج رمزا للموت البطئ و القاسي و الذي جاء بطبيعة الحال بسبب ترافق الثلج مع تيارات باردة و متجمدة و مع غياب للشمس و هبوط حاد للحرارة.

Table A.3(c) : Article 3 second Summary

Summary B
الثلج نوع من الهطولات على شكل بلورات دقيقة للجليد تحدث في الفصل البارد لكنها لا تحدث في كل دول العالم. تتم هذه الظاهرة التقاء تيارات هوائية رطبة ودافئة مع تيارات باردة تكون درجة حرارتها 12.5 تحت الصفر و يجب لتكون الثلج توفر نويات التكاثف التي يتكون عليها الثلج وهي عبارة عن جسيمات صلبة صغيرة جدا عالقة في الجو العلوي مثل ذرات الغبار أو الرماد وعند وجودها تتوفر الحالات الثلاث التي تمكن بخار الماء ليتحول من الحالة الغازية إلى الحالة الصلبة مكونا الثلج. وهكذا تنمو حيث يتفرع من الصفيحة البلورية ست أذرع و في درجات الحرارة الأكثر ارتفاعا تذوب حواف البلورة قليلا و ذلك لإتاحة فرصة الالتصاق مع البلورات الأخرى و بهذا تتكون الصفيحة الثلجية. و في أحيان أخرى يكون الثلج رمزا للموت البطئ و القاسي و الذي جاء بطبيعة الحال بسبب ترافق الثلج مع تيارات باردة و متجمدة و مع غياب للشمس و هبوط حاد للحرارة.

Table A.3(c) : Article 3 third Summary

Summary C
و تزداد غزارة الثلوج وكثافتها كلما اتجهنا قريباً من القطبين الجنوبي و الشمالي. تتم هذه الظاهرة التقاء تيارات هوائية رطبة ودافئة مع تيارات باردة تكون درجة حرارتها 12.5 تحت الصفر و يجب لتكون الثلج توفر نويات التكاثف التي يتكون عليها الثلج وهي عبارة عن جسيمات صلبة صغيرة جدا عالقة في الجو العلوي مثل ذرات الغبار أو الرماد وعند وجودها تتوفر الحالات الثلاث التي تمكن بخار الماء ليتحول من الحالة الغازية إلى الحالة الصلبة مكونا الثلج . و في أحيان أخرى يكون الثلج رمزا للموت البطئ و القاسي و الذي جاء بطبيعة الحال بسبب ترافق الثلج مع تيارات باردة و متجمدة و مع غياب للشمس و هبوط حاد للحرارة.

Table A.4(a) : Article 4 original text

Original Text
الأسد حيوان ضخم من فصيلة السنوريات. تسمى أنثاه لبؤة ويطلق على أطفاله اسم أشبال. أطلق علية ابتداء من القرن الأول الميلادي لقب ملك الغابة، ومن أسماء الأسد في اللغة العربية السبع والليث والهزبر والورد والضرغام وأسامة ويسمى بيته عرين. كان موطن الأسود يشمل عبر التاريخ معظم أوراسيا، من البرتغال إلى الهند، بالإضافة إلى إفريقيا بأكملها.ولكن منذ حوالي 10،000 سنة مضت، إنقرضت الأسود من اوروبة الغربيّة ثم مالبثت أن إنقرضت من باقي اوروبة بحلول القرن الثاني للميلاد، كما إنقرضت الأسود من شمالي إفريقيا والشرق الأوسط في الفترة مابين أواخر القرن التاسع عشر و أوائل القرن العشرين. تعيش الآن معظم الجمهرات في إفريقيا الوسطى حيث يظهر أن اعدادها تتناقص بإستمرار، فقد اظهرت إحدى البحوث تراجع اعدادها من حوالي 100،000 في أوائل التسعينات من القرن العشرين إلى حوالي 16،000 إلى 30،000 أسد برّي حالياً. بالإضافة إلى ذلك فإن جمهرة الأسود الحالية تواجه خطراً اخر يتمثل في عزلة المجموعات عن بعضها جغرافيّاً، مما يزيد من إحتمال التناسل الداخلي بين الأقارب مما يتسبب بمشاكل وراثيّة، وقد أظهرت المؤسسة الكينيّة للحياة البريّة أن المجموعات التي حصل بداخلها تناسل داخلي قد إزداد فيها متوسط عدد الأشبال لكل أنثى، كما تتوقع المؤسسة ازدياد عدد المجموعة بثلاثة اضعاف خلال السنوات العشر المقبلة بسبب إرتفاع نسبة الخصوبة عندها. كانت الأسود الآسيويّة تنتشر من تركيا إلى الهند عبر إيران ، ومن القوقاز حتى اليمن. أما الآن فإن ما تبقى منها يعيش في غابة شمال غربي الهند الواقعة في ولاية غوجارات، حيث يعيش 300 أسد في المنطقة المحميّة البالغة مساحتها 1412 كم2. إنقرض أخر الأسود الأوروبيّة في اليونان بحلول العام 100 للميلاد، ومن السلالات المنقرضة الأخرى: سلالة رأس الرجاء الصالح أسد رأس الرجاء الصالح، سلالة الكهوف أسد الكهوف الأوروبي الذي تعايش مع الإنسان خلال العصر الجليدي الأخير، والسلالة الأميركيّة الأسد الأميركي التي تعتبر قريبة لسلالة

الكهوف.يعتقد أن ذكور الأسود الأوائل كانت عديمة اللبدة وهو الشعر حول العنق، ويبدو أن الذكور الأوروبيّة و ذكور العالم الجديد إستمرّت عديمة الشعر حتى حوالي 10،000 سنة مضت.يعتقد بأن الذكور ذات اللبدة ظهرت منذ 32000 سنة، ويبدو أن الشكل الجديد ذي اللبدة كان له أفضليّة ما جعلته يوسّع موطنه ويستبدل الشكل الأخر في إفريقيا وغربي أوراسيا.يعتقد العلماء بأن اللبدة قد تطوّرت لدى الأسود بسبب ضغط الإنتقاء الجنسي، حيث اصبحت وحدها الأسود ذات اللبدة هي التي تتناسل وهذا ما جعل اللبدة اليوم لاتخدم غاية سوى هذه تقريباً.كان العلماء يعتقدون سابقاً أن حجم اللبدة وكثافتها ولونها دليل على سلالة الأسود المعيّنة، حيث كان يستند إلى هذا في تعريف بعض السلالات مثل أسد رأس الرجاء الصالح و أسد المغرب، أما الآن فقد أصبح يعرف أن العوامل الخارجيّة تؤثر على حجم و لون اللبدة، فقد ظهر أن الأسود في حدائق الحيوان الأوروبيّة و الأميركيّة تنمو لديها لبدة أكبر و أدكن لوناً مما كان سيحصل في موطنها الأصلي بغض النظر عن سلالتها.تمضي ذكور الأسود معظم حياتها خاملة .الأسود حيوانات لاحمة تعيش في مجموعات تسمّى زمراً مفردها زمرة، وتتألّف الزمرة من الإناث ذوات القربى وأشبالها بالإضافة إلى ذكر أو ذكرين أخوين في الغالب والتي تقتضي مهمتهما بإخصاب الإناث و حماية حوز الزمرة.كان يعتقد أن الإناث هي وحدها التي تقوم بعمليّة الصيد، أما الآن فأصبح يعرف أن الذكور تشارك في الصيد أيضاً، فجميع الذكور العازبة التي لم تسيطر على زمرة خاصة بها تصطاد بوتيرةٍ منتظمة، وحتى الذكور المسيطرة تبقى تشارك في الصيد أحياناً إلا أن نسبة مشاركتها تختلف حسب شكل الأرض التي تقطنها وحسب نوعيّة الطرائد المتوافرة.

Table A.4(b) : Article 4 first Summary

Summary A
كانت الاسود الآسيويّة تنتشر من تركيا الى الهند عبر ايران ، ومن القوقاز حتى اليمن.

Table A.4(c) : Article 4 second Summary

Summary B
اطلق علية ابتداء من القرن الاول الميلادي لقب ملك الغابة، ومن اسماء الاسد في اللغة العربية السبع والليث والهزبر والورد والضرغام واسامة ويسمى بيته عرين. ولكن منذ حوالي 10،000 سنة مضت، انقرضت الاسود من اوروبة الغربيّة ثم مالبثت ان انقرضت من باقي اوروبة بحلول القرن الثاني للميلاد،

كما انقرضت الاسود من شمالي افريقيا والشرق الاوسط في الفترة مابين اواخر القرن التاسع عشر و اوائل القرن العشرين. تعيش الآن معظم الجمهرات في افريقيا الوسطى حيث يظهر ان اعدادها تتناقص باستمرار، فقد اظهرت احدى البحوث تراجع اعدادها من حوالي 100،000 في اوائل التسعينات من القرن العشرين الى حوالي 16،000 الى 30،000 اسد برّي حالياً. كانت الاسود الآسيويّة تنتشر من تركيا الى الهند عبر ايران ، ومن القوقاز حتى اليمن. اما الآن فان ما تبقى منها يعيش في غابة شمال غربي الهند الواقعة في ولاية غوجارات، حيث يعيش 300 اسد في المنطقة المحميّة البالغة مساحتها 1412 كم2. يعتقد بان الذكور ذات اللبدة ظهرت منذ 32000 سنة، ويبدو ان الشكل الجديد ذي اللبدة كان له افضليّة ما جعلته يوسّع موطنه ويستبدل الشكل الاخر في افريقيا وغربي اوراسيا. تمضي ذكور الاسود معظم حياتها خاملة . كان يعتقد ان الاناث هي وحدها التي تقوم بعمليّة الصيد، اما الآن فاصبح يعرف ان الذكور تشارك في الصيد ايضاً، فجميع الذكور العازبة التي لم تسيطر على زمرة خاصة بها تصطاد بوتيرةٍ منتظمة، وحتى الذكور المسيطرة تبقى تشارك في الصيد احياناً الا ان نسبة مشاركتها تختل.

Table A.4(d) : Article 4 third Summary

Summary C
الاسد حيوان ضخم من فصيلة السنوريات. تعيش الآن معظم الجمهرات في افريقيا الوسطى حيث يظهر ان اعدادها تتناقص باستمرار، فقد اظهرت احدى البحوث تراجع اعدادها من حوالي 100،000 في اوائل التسعينات من القرن العشرين الى حوالي 16،000 الى 30،000 اسد برّي حالياً. تمضي ذكور الاسود معظم حياتها خاملة.

Table A.5(a) : Article 5 original

Original Text
كشف تقرير لمصلحة الإحصاءات العامة عن استقرار الرقم القياسي لتكاليف المعيشة في السعودية خلال شهر مارس الماضي مقارنة مع فبراير مسجلا ارتفاعا طفيفا نسبته 0.6 %عن الفترة نفسها من العام الماضي. وتضمن التقرير أهم تحركات الأسعار والتغيرات النسبية في الرقم القياسي لتكاليف المعيشة في السعودية، مؤكدا أن شهري فبراير ويناير سجلا تراجعا في الرقم بنسبة 1 %لكل منهما. وبالنسبة للتغيرات على مستوى المجموعات الرئيسة في مارس أظهر التقرير ارتفاع الرقم القياسي لمجموعتي الأثاث المنزلي والسلع والخدمات بنسب 0.5% و 0.6% على التوالي فيما سجلت 4 مجموعات رئيسة أخرى انخفاضا و شملت الأطعمة والمشروبات بتراجع 0.3% والأقمشة والملابس 0.2% والنقل والاتصالات 0.3% والتعليم 0.2 . %أما على مستوى المجموعات الفرعية فكان أكثر المجموعات ارتفاعا الفواكه الطازجة 2.9% تلاها السلع الشخصية 2.4% ثم الحبوب 0.2% والمشروبات 0.2% وأخيرا المفروشات المنزلية بنسبة 0.1%. وتصدرت الخضراوات الطازجة قائمة

الأكثر انخفاضا بنسبة 5.5 %ثم مجموعة الأسماك 3% والبقوليات 1.6 .%على صعيد آخر أظهرت دراسة أعدتها إحدى الشركات العالمية المتخصصة أن متوسط إنفاق الأسرة السعودية الشهري على الخدمات العامة يبلغ 275 ريالا. ويتوزع الإنفاق بنسبة 50% على الهاتف و30%على الكهرباء و12% على الصيانة و8% فقط على المياه المحلاة الحكومية. وشملت الدراسة 1000 أسرة من 8 مناطق إدارية في السعودية. وكان متوسط حجم الأسرة التي شاركت في استبانات الدراسة 5.7 أشخاص ومتوسط مساحة الوحدة السكنية 294 متراً مربعاً وتتألف الوحدة من 5.1 غرف في المتوسط. وقال 57% من المجيبين إنهم يسكنون في منازل مستأجرة ويسكن 70% منهم في شقق.

Table A.5(b) : Article 5 first Summary

Summary A
كشف تقرير لمصلحة الاحصاءات العامة عن استقرار الرقم القياسي لتكاليف المعيشة فيالسعودية خلال شهر مارس الماضي مقارنة مع فبراير مسجلا ارتفاعا طفيفا نسبته 0. 6%عن الفترة نفسها من العام الماضي. وتضمن التقرير اهم تحركات الاسعار والتغيرات النسبية في الرقم القياسي لتكاليفالمعيشة في السعودية، مؤكدا ان شهري فبراير ويناير سجلا تراجعا في الرقم بنسبة 1%لكل منهما. 2% والنقل والاتصالات 0. 3% والتعليم 0.

Table A.5(c) : Article 5 second Summary

Summary B
كشف تقرير لمصلحة الاحصاءات العامة عن استقرار الرقم القياسي لتكاليف المعيشة فيالسعودية خلال شهر مارس الماضي مقارنة مع فبراير مسجلا ارتفاعا طفيفا نسبته 0. وتضمن التقرير اهم تحركات الاسعار والتغيرات النسبية في الرقم القياسي لتكاليفالمعيشة في السعودية، مؤكدا ان شهري فبراير ويناير سجلا تراجعا في الرقم بنسبة 1%لكل منهما. 5% و 0. 3% والتعليم 0. اما على مستوى المجموعات الفرعية فكان اكثر المجموعات ارتفاعا الفواكه الطازجة2.

Table A.5(d) : Article 5 third Summary

Summary C
كشف تقرير لمصلحة الاحصاءات العامة عن استقرار الرقم القياسي لتكاليف المعيشة فيالسعودية خلال شهر مارس الماضي مقارنة مع فبراير مسجلا ارتفاعا طفيفا نسبته 0. وتضمن التقرير اهم تحركات الاسعار والتغيرات النسبية في الرقم القياسي لتكاليفالمعيشة في السعودية، مؤكدا ان شهري فبراير ويناير سجلا تراجعا في الرقم بنسبة 1%لكل منهما. وبالنسبة للتغيرات على مستوى المجموعات الرئيسة في مارس اظهر التقرير ارتفاعالرقم القياسي لمجموعتي الاثاث المنزلي والسلع والخدمات بنسب 0. 3% والاقمشة والملابس 0. 3% والتعليم 0.

Table A.5(e) : Article 5 fourth Summary

Summary D
كشف تقرير لمصلحة الاحصاءات العامة عن استقرار الرقم القياسي لتكاليف المعيشة فيالسعودية خلال شهر مارس الماضي مقارنة مع فبراير مسجلا ارتفاعا طفيفا نسبته 0. وبالنسبة للتغيرات على مستوى المجموعات الرئيسة في مارس اظهر التقرير ارتفاعالرقم القياسي لمجموعتي الاثاث المنزلي والسلع والخدمات بنسب 0. 2% والنقل والاتصالات 0. 3% والتعليم 0. 2. %

Table A.6(a) : Article 6 original

Original Text
تستعد مؤسسة البريد السعودي للترخيص لشركات وطنية للمساهمة والاستثمار في مشروع العنونة الجديدة والترميز الحديث وتأسيس الخدمة البريدية للمواقع التجارية والسكنية حسب الخرائط الرقمية وإنشاء الأنظمة الإلكترونية وإدارة وتشغيل كامل المشروع بأسلوب استثماري. و قدرت مصادر اقتصادية مستقلة أن تصل تكلفة المشروع في مدينة الرياض وحدها 600 مليون ريال بعد أن تم اعتماد البدء في تطبيق المرحلة الأولى للعنونة الجديدة فيها الشهر المقبل0 ويعتمد نظام العنونة على التقسيم المتوالي لكامل مساحة السعودية إلى مناطق بريدية ثم تقسم كل منطقة إلى قطاعات ثم يقسم كل قطاع إلى فروع ثم يقسم كل فرع إلى أقسام ثم يقسم كل قسم إلى مربعات بريدية ، لتكون هذه المتوالية ما سيعرف بالرمز البريدي المكون من 5 خانات تدل على رقم المربع والقسم والفرع والقطاع والمنطقة ويعتبر الرمز البريدي أحد الأجزاء المكونة للعنوان البريدي الذي يتكون بالإضافة إلى الرمز البريدي من رقم الوحدة البريدية واسم الشارع إن وجد0 ويتم الترقيم للمناطق والقطاعات والفروع والأقسام وفقا لأهمية كل محافظة وحسب أولوية التصنيفات الإدارية الخاصة بوزارة الداخلية "إمارة ومحافظة أ. ب ومراكز أ. ب " ويتم ترك الأرقام "صفر" و "واحد" لاستخدامات مؤسسة البريد الخاصة بالصناديق البريدية والدوائر الحكومية وكذلك الاستخدامات غير المتوقعة حالياً.

ونظراً لأهمية مدينة الرياض ولاحتوائها على كثافة سكانية عالية وامتداد عمراني سريع فقد تمت المباشرة فيها كمرحلة أولى لتطبيق النظام البريدي الحديث ،حيث تم الانتهاء من الترميز والعنونة إلكترونيا على أن تبدأ عملية التوزيع خلال الفترة القادمة وسيتم تطبيق النظام على المدن الأخرى على مراحل لاحقة في المستقبل القريب. كما تم تقسيم كامل المملكة إلى 8 مناطق بحيث تشتمل المنطقة البريدية على منطقة إدارية واحدة أو أكثر وقد تم اعتماد التقسيم للمناطق المعمول به حاليا حتى لا يسبب الانتقال إلى النظام الجديد للعنونة إرباكاً في الخدمات البريدية في المرحلة الحالية كما تم تقسيم كل منطقة من المناطق والتي تحمل الخانة الأولى من الرمز البريدي إلى عدة قطاعات بأسلوب علمي تم فيه مراعاة عوامل حدود نطاق الإشراف الإداري للمحافظات ، العوامل الطبيعية والجغرافية، التكلفة الاقتصادية لتوزيع الخدمات البريدية، الكثافة السكانية، بحيث أصبح القطاع يحمل الخانة الثانية من الرمز البريدي.

و تم ترقيم القطاعات البريدية ابتداء بعاصمة المنطقة حيث يتم ترقيم القطاعات الواقعة جنوبا بالنسبة لموقع عاصمة المنطقة بالأرقام الزوجية والقطاعات الواقعة شمالا بالنسبة لموقع عاصمة المنطقة بالأرقام الفردية لتسهيل عملية الاستدلال .

كما تم تقسيم كل قطاع من القطاعات الثمانية المذكورة إلى عدة فروع تم من خلالها مراعاة سهولة الاستدلال بحيث يتم ترقيم الفروع بتحديد "محور الرقم"للقطاع "كطريق الملك فهد بمدينة الرياض"بحيث تكون الأرقام الفردية تصاعديا في الجهة الغربية من المحور والأرقام الزوجية تصاعديا في الجهة الشرقية من المحور وهذا يعني أنه كلما اتجهنا شرقا من المحور تتصاعد الأرقام الزوجية "2 ،4 ،6، 8 " وكلما اتجهنا غربا تتصاعد الأرقام الفردية "3 ، 5، 7، 9" ويتم تقسيم الفروع إلى أقسام متقاربة المساحة ويبدأ اتجاه الترقيم للأقسام من الجهة الأقرب إلى محور الترقيم بحيث يتصاعد الترقيم كلما ابتعدت عن محور الترقيم ، حيث يأخذ القسم الخانة الرابعة من الرمز البريدي.

كما يتم تقسيم الأقسام البريدية إلى مربعات لا تتجاوز مساحتها 1كم مربع للمناطق المزدحمة كما هو الحال في وسط مدينة الرياض ولا تزيد عن 4 كيلو مترات مربعة في المناطق الأقل ازدحاما ، ويمكن أن تزيد مساحة المربع خارج المدن في المناطق قليلة الكثافة السكانية أو الخالية تماما من السكان أي إ حجم المربع يعتمد على الكثافة السكانية، ويمثل المربع الخانة الخامسة من الرمز البريدي مما يجعله جوهر الرمز البريدي والأساس لترقيم الوحدات البريدية.

أما بالنسبة للمناطق المكتظة (الشعبية) يتم تقسيمها إلى (1× 1كم) أو أقل ، وبالتالي يعطى لها الرقم الرابع والخامس من الرمز البريدي.

وتهدف مؤسسة البريد السعودي من خلال المشروع إلى الارتقاء بالخدمات البريدية المقدمة للمواطن والمقيم وتسهيل عملية استلام وتسليم الرسائل والارتقاء بشكل ومضمون الخدمة المقدمة ومواكبة التطورات الحديثة في صناعة البريد وتنفيذ الاستراتيجية التشغيلية الحديثة للبريد ، إضافة إلى تأسيس ثقافة البريد وتكريسها في السعودية.

Table A.6(b) : Article 6 first Summary

Summary A
تستعد مؤسسة البريد السعودي للترخيص لشركات وطنية للمساهمة والاستثمار في مشروعالعنونة الجديدة والترميز الحديث وتاسيس الخدمة البريدية للمواقع التجاريةوالسكنية حسب الخرائط الرقمية وانشاء الانظمة الالكترونية وادارة وتشغيل كاملالمشروع باسلوب استثماري. ونظراً لاهمية مدينة الرياض ولاحتوائها على كثافة سكانية عالية وامتداد عمرانيسريع فقد تمت المباشرة فيها كمرحلة اولى لتطبيق النظام البريدي الحديث ،حيث تمالانتهاء من الترميز والعنونة الكترونيا على ان تبدا عملية التوزيع خلال الفترةالقادمة وسيتم تطبيق النظام ع. كما يتم تقسيم الاقسام البريدية الى مربعات لا تتجاوز مساحتها 1كم مربع للمناطقالمزدحمة كما هو الحال في وسط مدينة الرياض ولا تزيد عن 4 كيلو مترات مربعة فيالمناطق

الاقل ازدحاما ، ويمكن ان تزيد مساحة المربع خارج المدن في المناطق قليلةالكثافة السكانية او الخال.

Table A.6(c) : Article 6 second Summary

Summary B
تستعد مؤسسة البريد السعودي للترخيص لشركات وطنية للمساهمة والاستثمار في مشروعالعنونة الجديدة والترميز الحديث وتاسيس الخدمة البريدية للمواقع التجاريةوالسكنية حسب الخرائط الرقمية وانشاء الانظمة الالكترونية وادارة وتشغيل كاملالمشروع باسلوب استثماري. ونظراً لاهمية مدينة الرياض ولاحتوائها على كثافة سكانية عالية وامتداد عمرانيسريع فقد تمت المباشرة فيها كمرحلة اولى لتطبيق النظام البريدي الحديث ،حيث تمالانتهاء من الترميز والعنونة الكترونيا على ان تبدا عملية التوزيع خلال الفترةالقادمة وسيتم تطبيق النظام ع. ب " ويتم ترك الارقام "صفر" و "واحد" لاستخدامات مؤسسة البريد الخاصة بالصناديقالبريدية والدوائر الحكومية وكذلك الاستخدامات غير المتوقعة حالياً . كما تم تقسيم كامل المملكة الى 8 مناطق بحيث تشتمل المنطقة البريدية علىمنطقة ادارية واحدة او اكثر وقد تم اعتماد التقسيم للمناطق المعمول به حاليا حتىلا يسبب الانتقال الى النظام الجديد للعنونة ارباكاً في الخدمات البريدية فيالمرحلة الحالية كما تم تقسيم كل منط. كما تم تقسيم كل قطاع من القطاعات الثمانية المذكورة الى عدة فروع تم من خلالهامراعاة سهولة الاستدلال بحيث يتم ترقيم الفروع بتحديد "محور الرقم"للقطاع "كطريقالملك فهد بمدينة الرياض"بحيث تكون الارقام الفردية تصاعديا في الجهة الغربية منالمحور والارقام الزوجية . كما يتم تقسيم الاقسام البريدية الى مربعات لا تتجاوز مساحتها 1كم مربع للمناطقالمزدحمة كما هو الحال في وسط مدينة الرياض ولا تزيد عن 4 كيلو مترات مربعة فيالمناطق الاقل ازدحاما ، ويمكن ان تزيد مساحة المربع خارج المدن في المناطق قليلةالكثافة السكانية او الخال.

Table A.6(d) : Article 6 third Summary

Summary C
تستعد مؤسسة البريد السعودي للترخيص لشركات وطنية للمساهمة والاستثمار في مشروعالعنونة الجديدة والترميز الحديث وتاسيس الخدمة البريدية للمواقع التجاريةوالسكنية حسب الخرائط الرقمية وانشاء الانظمة الالكترونية وادارة وتشغيل كاملالمشروع باسلوب استثماري. ونظراً لاهمية مدينة الرياض ولاحتوائها على كثافة سكانية عالية وامتداد عمرانيسريع فقد تمت المباشرة فيها كمرحلة اولى لتطبيق

النظام البريدي الحديث ،حيث تمالانتهاء من الترميز والعنونة الكترونيا على ان تبدا عملية التوزيع خلال الفترةالقادمة وسيتم تطبيق النظام ع.

Table A.6(e) : Article 6 fourth Summary

Summary D
تستعد مؤسسة البريد السعودي للترخيص لشركات وطنية للمساهمة والاستثمار في مشروعالعنونة الجديدة والترميز الحديث وتاسيس الخدمة البريدية للمواقع التجاريةوالسكنية حسب الخرائط الرقمية وانشاء الانظمة الالكترونية وادارة وتشغيل كاملالمشروع باسلوب استثماري. ب ومراكز ا. ونظراً لاهمية مدينة الرياض ولاحتوائها على كثافة سكانية عالية وامتداد عمرانيسريع فقد تمت المباشرة فيها كمرحلة اولى لتطبيق النظام البريدي الحديث ،حيث تمالانتهاء من الترميز والعنونة الكترونيا على ان تبدا عملية التوزيع خلال الفترةالقادمة وسيتم تطبيق النظام ع. و تم ترقيم القطاعات البريدية ابتداء بعاصمة المنطقة حيث يتم ترقيم القطاعاتالواقعة جنوبا بالنسبة لموقع عاصمة المنطقة بالارقام الزوجية والقطاعات الواقعةشمالا بالنسبة لموقع عاصمة المنطقة بالارقام الفردية لتسهيل عملية الاستدلال. كما يتم تقسيم الاقسام البريدية الى مربعات لا تتجاوز مساحتها 1كم مربع للمناطقالمزدحمة كما هو الحال في وسط مدينة الرياض ولا تزيد عن 4 كيلو مترات مربعة فيالمناطق الاقل ازدحاما ، ويمكن ان تزيد مساحة المربع خارج المدن في المناطق قليلةالكثافة السكانية او الخال.

Table A.7(a) : Article 7 original text

Original Text
الإثنا عشري)بالإنجليزية (Duodenum :أو العفج هو أحد أجزاء القناة الهضمية وأول جزء من الأمعاء الدقيقة يصل بين المعدة و الصائم. طوله 25-30 سم وله أربعة قطع تشبه المضلع تنفتح في القطعة الثانية منه كلا من القناتين الصفراوية والمعثكلية ويفصله عن المعدة البواب.يقوم الاثنا عشر بوظيفة أساسية في عملية الهضم.المواد النشوية التي لم يتم تفكيكها في الفم و المعدة يقوم الثنى عشر بإكمال عملية تفكيكها بمساعدة أنزيمات يفرزها البنكرياس كذلك مقاطع البروتينات المسماه ببتونات التي تفككت من البروتينات في المعدة تستمر عملية هضمها في الاثنى عشر بمساعدة أنزيمات أساسية. كذلك يتم تحليل وهضم الدهون وكذا تتحلل وتهضم في الاثنى عشر بواسطة أنزيمات تصل من البنكرياس الفولماريات مثل المواد المضافة للغذاء كاللون والمنكهات وما شابه التي لم تحلل في المعدة والفم.المواد التي تصل من البنكرياس والكبد و المرارة تساعد على تحليل بعض مركبات الغذاء كيميائيا والمواد التي يتم تحليلها كيميائيا هي:بروتينات كروهيدرات ودهون بالاضافة لذلك يتم امتصاص مياة وموادمعدنية الى الدم علما بان المواد المعدنية لا يحل

بها تغيير في العملية الهضمية(مواد غير عضوية). الحامض الملحي الذي تفرزه المعدة يتم معادلته في الاثنى عشر بواسطة (NaHCO3) الذي يفرز من البنكرياس.كذلك يستقبل الاثنى عشر العصارة الصفراء التي تفرزها المرارة.

Table A.7(b) : Article 7 first Summary

Summary A
الإثنا عشري بالإنجليزية Duodenum :أو العفج هو أحد أجزاء القناة الهضمية وأول جزء من الأمعاء الدقيقة يصل بين المعدة و الصائم. المواد النشوية التي لم يتم تفكيكها في الفم و المعدة يقوم الثنى عشر بإكمال عملية تفكيكها بمساعدة أنزيمات يفرزها البنكرياس كذلك مقاطع البروتينات المسماه ببتونات التي تفككت من البروتينات في المعدة تستمر عملية هضمها في الاثنى عشر بمساعدة أنزيمات أساسية. كذلك يتم تحليل وهضم الدهون وكذا تتحلل وتهضم في الاثنى عشر بواسطة أنزيمات تصل من البنكرياس الفولماريات مثل المواد المضافة للغذاء كاللون والمنكهات وما شابه التي لم تحلل في المعدة والفم. كذلك يستقبل الاثنى عشر العصارة الصفراء التي تفرزها المرارة.

Table A.7(c) : Article 7 second Summary

Summary B
الإثنا عشري بالإنجليزية Duodenum :أو العفج هو أحد أجزاء القناة الهضمية وأول جزء من الأمعاء الدقيقة يصل بين المعدة و الصائم. يقوم الاثنا عشر بوظيفة أساسية في عملية الهضم. كذلك يتم تحليل وهضم الدهون وكذا تتحلل وتهضم في الاثنى عشر بواسطة أنزيمات تصل من البنكرياس الفولماريات مثل المواد المضافة للغذاء كاللون والمنكهات وما شابه التي لم تحلل في المعدة والفم. الحامض الملحي الذي تفرزه المعدة يتم معادلته في الاثنى عشر بواسطة (NaHCO3) الذي يفرز من البنكرياس.

Table A.8(a) : Article 8 original text

Original Text
كانت إيطاليا آخر الدول الأوروبية التي دخلت مجال التوسع الاستعماري. وكانت ليبيا عند نهاية القرن التاسع عشر، هي الجزء الوحيد من الوطن العربي في شمال أفريقيا الذي لم يتمكن الصليبيون الجدد من الاستيلاء عليه، ولقرب ليبيا من إيطاليا جعلها هدفا رئيسا من أهداف السياسة الاستعمارية الإيطالية.بدات إيطاليا العزم على احتلال ليبيا فقامت بفتح المدارس في كل من بنغازي و طرابلس لتعلم اللغة الايطالية

وارسلت الارساليات التبشيريه للدين المسيحي افتتحت فروعا لبنك روما واصبحت القنصليه في مدينتى بنغازى و طرابلس مركزا للنشاط السياسى والدعاية الايطاليه والتجسس على اهل البلاد، ولم يصعب على إيطاليا اختلاق الذرائع الواهية لاحتلال ليبيا .فى 27 سبتمبر 1911 م وجهت إيطاليا انذاراً إلى الدولة العثمانية تأخذ عليها فيه انها أهملت شان ليبيا واتهمتها بانها تحرض الليبيين على الرعايا الايطاليين وتضطهدهم. وفى يوم 28 سبتمبر 1911 اقبل الموعد المحدد لانتهاء اجل الانذار كانت السفن الحربية الايطاليه في مياه طرابلس . واعلنت الحرب على تركيا في 29 سبتمبر سنة 1911 م، وبدأت الحرب العثمانية الإيطالية واستطاعت الاستيلاء على طرابلس في 3 أكتوبر من السنة نفسها.وكانت القوات الايطاليه مؤلفة من39 الف جندي و6 الاف حصان وألف سياره و نحو خمسين مدفع ميدان بدأ قصف مدينة طرابلس في الساعة الثالثة والنصف مساء يوم 3 أكتوبر 1911 وأنزلت قوة من البحر عددها يقدر بنحو ألفين جندي وتم إحتلال مدينة طرابلس.ومن هنا بدأ المجاهدون حركتهم يحدوهم الايمان بالحق والدفاع عن العرض والارض وكان عدد المتطوعين نحو 15000 ليبي وتحرك نواب البلاد وزعمائها في ضواحى مدينة طرابلس نحو معسكرات الجهاد ومنهم الشيخ سليمان الباروني نائب الجبل الغربى والشيخ أحمد سيف النصر من زعماء الجنوب والوسط. و السيد احمد الشريف السنوسي في الشرق. وبرزت شخصية شيخ الشهداء عمر المختار في برقة ورفاقه المجاهدين منهم الشهيد عبد القادر يوسف بورحيل والشهيد الفضيل بوعمر بوحو الاوجلى . قاومت القوات الليبية و العثمانية الإيطاليين لفترة قصيرة، ولكن تركيا تنازلت عن ليبيا لإيطاليا بمقتضى المعاهدة التي أبرمت بين الدولتين في 18 أكتوبر 1912م معاهدة أوشي ، وأدرك الليبيون الآن أن عليهم أن ينظموا صفوفهم ويتولوا بأنفسهم أمر المقاومة والجهاد ضد المستعمر، وقد اشتدت مقاومة الليبيين للقوات الإيطالية مما حال دون تجاوز سيطرة الإيطاليين المدن الساحلية، ولما دخلت إيطاليا الحرب العالمية الأولى 1915 م حاولت تركيا استغلال الحركة السنوسية التى كانت بقيادة السيد احمد الشريف السنوسى فمارست تركيا عليه بعض الضغوط لما عرف عليه حبه للمسلمين واحترامه لدولة الخلافه فدخلت قوات الحركه في قتال مع القوات الانجليزيه التى كانت في مصر فابتليت قوات الحركة بخسائر كبيره ، و بعد هزيمة قواته تنازل عن الزعامة لإدريس السنوسي وقاد الجهاد نيابة عنه في المنطقة الشرقية المجاهد عمر المختار. وفي المنطقةالغربية قاد الجهاد سليمان باشا الباروني ومجموعم من المجاهدين في منطقة طرابلس منهم السويحلي والمريض وسوف المحمودي و اعلان الجمهورية الطرابلسية ثم حكومة الاصلاح.

Table A.8(b) : Article 8 first Summary

Summary A
كانت ايطاليا آخر الدول الاوروبية التي دخلت مجال التوسع الاستعماري. بدات ايطاليا العزم على احتلال ليبيا فقامت بفتح المدارس في كل من بنغازي و طرابلس لتعلم اللغة الايطالية وارسلت الارساليات التبشيريه للدين المسيحي افتتحت فروعا لبنك روما واصبحت القنصليه في مدينتى بنغازى و طرابلس مركزا للنشاط السياسى والدعاية الايطاليه والتجس. وفى يوم 28 سبتمبر 1911 اقبل الموعد المحدد لانتهاء اجل الانذار كانت السفن الحربية الايطاليه في مياه طرابلس . وكانت القوات الايطاليه مؤلفة من39 الف جندي و6 الاف حصان والف سياره و نحو خمسين مدفع ميدان بدا قصف مدينة طرابلس في الساعة الثالثة والنصف مساء يوم 3 اكتوبر 1911 وانزلت قوة من البحر عددها يقدر بنحو الفين جندي وتم احتلال مدينة طرابلس. و السيد احمد الشريف السنوسي في الشرق.

Table A.8(c) : Article 8 second Summary

Summary B
كانت ايطاليا آخر الدول الاوروبية التي دخلت مجال التوسع الاستعماري. بدات ايطاليا العزم على احتلال ليبيا فقامت بفتح المدارس في كل من بنغازي و طرابلس لتعلم اللغة الايطالية وارسلت الارساليات التبشيريه للدين المسيحي افتتحت فروعا لبنك روما واصبحت القنصليه في مدينتى بنغازى و طرابلس مركزا للنشاط السياسى والدعاية الايطاليه والتجس. وفى يوم 28 سبتمبر 1911 اقبل الموعد المحدد لانتهاء اجل الانذار كانت السفن الحربية الايطاليه في مياه طرابلس . وكانت القوات الايطاليه مؤلفة من39 الف جندي و6 الاف حصان والف سياره و نحو خمسين مدفع ميدان بدا قصف مدينة طرابلس في الساعة الثالثة والنصف مساء يوم 3 اكتوبر 1911 وانزلت قوة من البحر عددها يقدر بنحو الفين جندي وتم احتلال مدينة طرابلس. و السيد احمد الشريف السنوسي في الشرق. قاومت القوات الليبية و العثمانية الايطاليين لفترة قصيرة، ولكن تركيا تنازلت عن ليبيا لايطاليا بمقتضى المعاهدة التي ابرمت بين الدولتين في 18 اكتوبر 1912م معاهدة اوشي ، وادرك الليبيون الآن ان عليهم ان ينظموا صفوفهم ويتولوا بانفسهم امر المقاومة والجهاد ضد ال.

Table A.8(d) : Article 8 third Summary

Summary C
وفى يوم 28 سبتمبر 1911 اقبل الموعد المحدد لانتهاء اجل الانذار كانت السفن الحربية الايطاليه في مياه طرابلس.

Table A.8(e) : Article 8 fourth Summary

Summary D
كانت ايطاليا آخر الدول الاوروبية التي دخلت مجال التوسع الاستعماري. وكانت ليبيا عند نهاية القرن التاسع عشر، هي الجزء الوحيد من الوطن العربي في شمال افريقيا الذي لم يتمكن الصليبيون الجدد من الاستيلاء عليه، ولقرب ليبيا من ايطاليا جعلها هدفا رئيسا من اهداف السياسة الاستعمارية الايطالية. بدات ايطاليا العزم على احتلال ليبيا فقامت بفتح المدارس في كل من بنغازي و طرابلس لتعلم اللغة الايطالية وارسلت الارساليات التبشيريه للدين المسيحي افتتحت فروعا لبنك روما واصبحت القنصليه في مدينتى بنغازى و طرابلس مركزا للنشاط السياسى والدعاية الايطاليه والتجس. وبرزت شخصية شيخ الشهداء عمر المختار في برقة ورفاقه المجاهدين منهم الشهيد عبد القادر يوسف بورحيل والشهيد الفضيل بوعمر بوحو الاوجلى . قاومت القوات الليبية و العثمانية الايطاليين لفترة قصيرة، ولكن تركيا تنازلت عن ليبيا لايطاليا بمقتضى المعاهدة التي ابرمت بين الدولتين في 18 اكتوبر 1912م معاهدة اوشي ، وادرك الليبيون الآن ان عليهم ان ينظموا صفوفهم ويتولوا بانفسهم امر المقاومة والجهاد ضد ال.