Research Article

Arabic Text Summarization from Reader's Perspective

[1]Mohammad Hjouj Btoush, [1]Nasim F. Diab and [2]Khalid N. Alzubi
[1]Department of Computer Science, Prince Abdulah Bin Gazi Faculty of Information Technology,
[2]Department of Management Information System, Faculty of Business,
Al-Balqa' Applied University, Jordan

Abstract: The aim is to build a system that takes the source text as an input and the result will be a summery which contains sentences preserve the main theme of the source. We use a statistical approach to solve the problem of Arabic text summarization. Many researches which study Arabic text summarization problem focus on how writers of articles write, here we make a system take the advantage of how readers of an article read and make comments on article. Our approach assign scores for each segment of an article depending on the location of the sentences, the size of each sentence and term frequency on the article and comments of readers. Our system tested by 100 human evaluators, We gave each evaluator a copy of the summary produced by our system, with a question about the connectivity of ideas and sentences in <article, summary> and asked them to evaluate the summary. We asked them to answer by rejected, not-related, satisfactory, good, accepted, to test our approach which focus on readers' comments to summarize articles.

Keywords: Comment relatedness, extract, term frequency, text summarization

INTRODUCTION

Because of the large amount of information existing recently, many readers do not have enough time to spend on reading, they prefer to read abstracts and extracts, but it is not easy to summarize large documents manually (Tofigh et al., 2013).

The first beginning for text summarization was started in the last century since forty five years ago (Edmundson, 1969).

Automatic text summarization is used to minimize the size of the original text but still including the main theme of the source document. Arabic language has been one of the languages which use automatic text summarization techniques. Statistical approach segment text and assign scores to each part of text. First, the source text is divided into paragraph and sentences, then sentences are selected to be included in the summarization depends on sentence rank (Nobata et al., 2002).

The important sentences in a document is the goal of extraction summarization method to include in the final summary (Tofigh et al., 2013).

One element for text summarization is to choose sentences depending on the location of the sentence. Also we can choose important sentences by the size of the sentence, another element is to choose weighted sentences which contain high terms frequencies and title relatedness sentences (Hammo et al., 2011).

There are many types of text summarization such as informative, indicative, topic oriented, generic, an abstract and an extract (Mani et al., 2002).

Text summarization overlaps with many NLP fields such as web search engines, text classification, question answering.

Text summarization is defined as most difficult problem in Natural Language Processing (NLP), especially non-Roman alphabet languages like Arabic and Chinese (Filho et al., 2007).

NLP applications for Arabic language such as information retrieval, question answering and text summarization need complex processing, which contain: Tokenization, stop words removal and Part of Speech Tagging (POST). There are limited researches in Arabic Natural Language Processing because of the lack of open-source tools and resources as compared with other languages which use the Roman alphabet (Hammo et al., 2011).

Text summarization has many development and approach in recent years because of the high appreciation of its application in learning and education (Spärck Jones, 2007).

Summarization techniques have many features, despite they focus on text, there are wide range of data types which could be included in summarization techniques such as multimedia contents like images and videos (Fan et al., 2008).

Corresponding Author: Mohammad Hjouj Btoush, Department of Computer Science, Prince Abdulah Bin Gazi Faculty of Information Technology, Al-Balqa' Applied University, Jordan

We can find the need to summarize the description if pictures in text summarization techniques. They take profit of the immediate context of the image to extract such information, for instance, text in HTML tags. Their main purpose is to correctly detect and classify entities appearing in images and then, calculate the salience of such entity with the final goal to produce a short annotation for the image (Deschacht and Moens, 2007).

The objective of this study is to contribute to the existing literature body of text summarization by build a system takes the source text as an input and the result will be a summery which contains sentences preserve the main theme of the source.

LITERATURE REVIEW

English language have huge number of solutions for automatic text summarization, other languages have limited research to solve TS problem especially non-Roman alphabet.

Summary for new technologies such as plogs, articles recently enter the world of sTS, all depend on different term frequency types (Balahur-Dobrescu *et al*., 2009).

Few attempts to solve Arabic text summarization problem were established. The first attempt to solve Arabic TS problem was in 2004 by Douzidia and Lapalme (2004) they implement an application called "Lakhas" using machine learning algorithms and sentence extraction. In Sobh *et al*. (2007) designed an Arabic TS system which involved Bayesian and Genetic programming classification. CLASSY, another important system design was published by Schlesinger *et al*. (2008). CLASSY used for multi-document Arabic/English. El-Haj and Hammo (2008) described a system to solve Arabic TS problem depending on queries submitted by users and extract sentences by using a cosine similarity measure. Another important work for Hammo *et al*. (2011) designed a hybrid system which used a filtering technique mixed with statistical method.

System overview: The system takes the source text as an input and the result will be a summery which contains sentences preserves the main theme of the source. The system depends on the following assumptions:

- Terms which are very low or very high is neglected (stop words).
- Thematic terms which is most frequent terms, repeated in the text and in the comment section of the html source.
- If a sentence consists of thematically related terms then this sentence may be included in text summarization.
- Some sentences are included in other sentences, so we can delete it from text summarization (Hammo *et al*., 2011).

- Each comment is represented by one paragraph.

The modules of the system are:

- **Statistical analyzer:** Used to do stop-word removal, term-weight calculations.
- **Segmenter:** Implement a tokenizer to break the original document into words, sentences and paragraphs.
- **Selector:** Determine which sentences will appear in the summary depending in sentences scores. It neglect redundant sentences.
- **Arranger:** To make the summary readable, it order the chosen sentences as it appeared in the original text (Hammo *et al*., 2011; Douzidia and Lapalme, 2004; Sobh *et al*., 2007).

The steps of our Arabic text summarization:

- The source text is segmented into words, sentences, paragraphs.
- Stop-words are identified and removed from original text and comments section.
- Rank sentences of the text by assigning scores to them using location and length of sentences.
- Determine thematic terms of the source text and the comment section of the html file.
- Removing redundant sentences.
- Select high scores sentences.
- Make the selected sentences at the original order of the text (Hammo *et al*., 2011).

METHODOLOGY

Segmenting the original text into words, sentences and paragraphs: Our code implements a tokenizer to segment the original text into paragraphs. Paragraphs are also extracted into sentences and terms.

Identifying sentences in Arabic is so hard because of the fact, sentences are not begin with capital letters and missing punctuation marks.

Rank sentences of the text by assigning scores to them: We use four procedures to pick the suitable sentences and include them in the summary, these procedures are the sentence location, the size of the sentence, term frequency and inverse document frequency and comment relatedness. We use these procedures to simulate how humans try to write and read articles.

Summarization feature such as sentence position and word frequency is important in Text classification.

Sentence location: Humans usually start their articles by using rich opening paragraphs and end them by strong end paragraphs (Edmundson, 1969).

From the previous idea, we make the first and last sentences take high scores, on the contrary of sentences located inside the article.

We used the next equation which give the first and the last sentence the highest location score which equal to 1:

$$S_{location}(L_i) = \text{maximum} \left(\frac{1}{i}, \frac{1}{n-i+1}\right) \qquad (1)$$

where,
i = The sentence location
n = The number of sentences in the article
Li is i^{th} line $S_{location}$ = The score of the sentence location

Sentence size and lengths: The other scoring procedure is to use two types of lengths; Maximum length and Minimum length then sum the two lengths to form the size of the sentence.

The first length, Maximum length assumes that if the sentence is longer than others then it likely be included in the summary.

Many researches tend to assume the Maximum length of Arabic article in the web is equal to 10 tokens after removing stop words (Hammo *et al.*, 2011).

The Maximum length of each sentence is calculated by the equation:

$$S_{\max\_length}(L_i) = \begin{cases} \frac{L_i}{L_{maximum}} \ (\text{if } L_i \leq L_{maximum}) \\ 1 \ (\text{if } L_i > L_{maximum}) \end{cases} \qquad (2)$$

The second length, Minimum length assumes that if the sentence has shorter length than others, then it is consider to be not important sentence.

Many researches tend to assume the minimum length of Arabic article in the web is equal to 5 tokens after removing stop words (Hammo *et al.*, 2011).

The minimum length of each sentence is calculated by the equation:

$$S_{\min\_length}(L_i) = \begin{cases} 0 \ (\text{if } L_i \geq L_{min}) \\ \frac{L_i - L_{min}}{L_{min}} \ (\text{if } L_i < L_{min}) \end{cases} \qquad (3)$$

$$S_{size} = S_{\min\_length} + S_{\max\_length} \qquad (4)$$

It is important to notice that if the sentence is shorter than 5 words then it has negative value (Hammo *et al.*, 2011; Edmundson, 1969).

Term weighting: Term frequency uses counts to produce summaries from scientific documents with the aim to determine the relevance of a sentence in a document. The underlying assumption is that the most frequent words are indicative of the main topic of an article. However, not all the words are taken into consideration. On the contrary, stop words, i.e., words

without carrying any semantic information, such as "إلى" or "من", are not used for computing the term frequency. Under the same assumption, a number of techniques based on term frequency counts have been employed in TS (Luhn, 1958).

Several statistical approaches, such as term frequency or inverse document frequency (tf*idf), are briefly analyzed, as well as the potential problems this kind of features may have. The concept of tf*idf is that frequent terms in a paragraph are important only if they are not very frequent in the whole article (Luhn, 1958).

Another important measure for ranking sentences based on term frequency (*tf*) and inverse document frequency (*idf*). We use the weight tf.idf to calculate how this term associates to the meaning of the original text.

When a sentence contains a high score terms frequency then this sentence is more likely to be included in the summary because of its contribution in the meaning of the document.

After dropping the stop-words, we applied a method to calculate the (*tf.idf*) score for each term (t).

We calculate (*tf*) by determining the term frequency for each token in the text, then calculate the summation of each term in the sentence.

Also, we determine the inverse document frequency by calculating the paragraph frequency, then the weight equation is determined by:

$$\omega(t_i) = tf * Idf = tf\left(\frac{t_i}{p}\right) * log \frac{PN}{pfrequency(t_i)} \qquad (5)$$

where, *pfrequency* is the number of paragraphs including the specified term, PN stand for the whole count of paragraphs the document consist of.

After calculating all terms weight in the document, we calculate the Euclidean normalization for *w* (t) for all words in the paragraphs.

The value of each paragraph Euclidean is calculating by the equation:

$$\vec{p} = \sqrt{\sum_{t \in pi} \omega(ti)^2} \qquad (6)$$

Then we perform the weight normalization by the equation:

$$\omega norm = \frac{\omega(ti)}{\vec{p}} \qquad (7)$$

At last, we can calculate the ith line score in each paragraph by the equation:

$$S_{weight}(l_i) = \sum_{t \in li} \omega norm(ti) \qquad (8)$$

Relatedness to comments: The changes of the society and the new information technology has affected the fast growing of data; Web 2.0 (social web) is one of the most important technology recently, Which leads to

Table 1: Example: Arabic article about studying French

خطوات ''مميزة'' لتعلم اللغة الفرنسية بإتقان
لعل مشاريع تعلم اللغة هي أكثر المشاريع التي نبدأ ولا ننجز نضع المخططات لها ثم تُضيع. فكم من مرة عقدت العزم جازماً وحازماً أنك متم تعلم اللغة الفرنسية نهاية هذه السنة، وأنك بعدها
بادئ قراءة البؤساء وبلغتها الأم لتثبت لنفسك أن هذا المشوار في نهايته
الخطوات لا تخطو وحدها بل تحتاج لتخطوها مركب حسن إدارة الوقت ومحفز الهدف؛ السبب الواضح الذي من أجله قررت خوض هذه التجربة. بدونهما لن تحرز من تقدم مهما تفننت في
الأدوات واستزدت من الخطط.
بناء قاعدة من المفردات والجمل الأساسية والقواعد فما تقرؤه أو تسمعه في أي كتاب أو نص باللغة الفرنسية إنما هو مجموعة من الجمل، فإن أردت فهمها لزمك فهم كلماتها. وما تكتبه أو
تقوله إنما هو مجموعة من الجمل كذلك، فإن أردت تركيبها لزمك فهم قواعدها. لا بد من المفردات والقواعد. لكن السؤال هو : كيف؟ والجواب بسيط: الحفظ
لا بد من مساق متكامل مع مدرس ، ليس من فائدة أن نخدع أنفسنا. فأنت بحاجة إلى من يشرح لك القواعد ويدلك على خفايا اللغة وخباياها ويجنبك الأخطاء. قد تتعلم من مدرس جيد في ساعة
ما تحتاج ساعات لتستخرجه من المواقع أو الكتب
فإن لم يكن ذلك متوفراً لك فابحث جيداً فإن تعذر تماماً فاتجه إلى البدائل عبر الإنترنت . تأتي أهمية هذه الأدوات والتطبيقات والمواقع من أهمية استخدام اللغة يومياً
لا تخدع نفسك هذه التطبيقات مفيدة في سياقها، لكنها لن تغنيك عن الحفظ والالتزام بمساق محدد. لا تقل إنني أتعلم اللغة الفرنسية في حين إن كل ما فعلته هو تحميل خمسة تطبيقات على هاتفك لتستخدمها في كل شهر مرة
اختر مقطعاً خفيفاً باللغة الفرنسية واستمع إليه كل يوم. ثم استمع إليه مع ترجمة. ثم استمع إليه مع متابعة النص الفرنسي. ثم حاول أن تقلد ما يُقال. ثم اختر مقطعاً آخر
لكن تذكر أن الحفظ والتقدم في مساقات اللغة لا ينبغي أن نتوقف في هذه المرحلة. فما تزال تحتاج أن تزيد من رصيدك من المفردات والجمل (وربما الأمثال والتشبيهات الشائعة). وما تزال تحتاج أن تأخذ المساق التالي. لكن الفارق الآن هو أنك تستمتع بتعلم اللغة

new applications such as blogs, forums or social networks. These applications let readers express their opinion toward any topic, product, or service. These applications lead to a new type of summarization called sentiment-based to describe readers opinions (Nenkova, 2006).

The last factor used in scoring the source text in our system is comments relatedness, usually, comments which readers of the web articles include their notes about the original text to emphasize some ideas appeared in the source text from readers perspectives.

All researches try to simulate how writers of articles write, but in our methodology we try also to simulate how readers read an article and think about it.

Most readers in their comments at the original text focus on easy sentences and really they tend to focus on important ideas not detailed sentences.

We faced many difficulties because of the lack in Arabic contents on the web. Also Arabic readers do not tend to comment on articles. Another note is that Arabic readers tend to comment on comments not on the original text especially religious and political comments.

The purpose of this scoring is to provide a summary containing the specific information a reader is interested in. This means that different readers may have different needs, so that summarization systems have to determine the readers profile before they select the relevant information that will be included in the finally summary.

Another important note is that many articles use a title which attract readers and this title is far away from the main idea of the article. The title is كيف تتعلم الفرنسية but the content of the article is speaking about how to study all languages not just French.

Relatedness to the comment of each sentence is calculated by the equation:

$$S_{coment}(L_i) = \sum_{t \in (C \cap L_i)} tf * idf(t) \qquad (9)$$

Total line scoring: At the end of our methodology, we calculate all sentences scores by the summation of all scores methods by the equation:

$$S_{total}(L_i) = \alpha\, S_{location}(L_i) + \alpha\, S_{size}(L_i) + \alpha\, S_{weight}(L_i) + \alpha\, S_{comment}(L_i) \qquad (10)$$

Here, we assume α is equal to 1 (Wang and Yang, 2006).

Example: We have applied the previous statistical measures on the following example to summarize an Arabic article about how to study French and get the advantages of using comments existing by Arabic readers of the article. The article is given in Table 1 which is capture from the web site http://www.arageek.com /2014/10/30/4-steps-to-learn-french-perfectly.html and comments on the article is given in Table 2, the summary of this articles was obtained from our system in Table 3 and 4.

Note that, the original article is not full punctuation, the article have many spellings errors. Another notes in the comments table; comments are written in Colloquial Dialect, they contain many errors in spellings and grammars.

First, we applied our code to segment the article into words, sentences and paragraphs. We use the segmented paragraphs in our method to calculate the various statistical measures including the sentence location score, sentence size, term frequency and inverse document frequency in the article and relatedness comments in that study. Then in Table 3 we show the statistical scores for each element in our method. Note that the negative scores are penalty score and the shaded sentences are the highest score sentences and they probably appeared in the summary.

Eliminating of redundant sentences: The next step in our method for summarized sentences is to reduce the repeated sentences which one sentence is included in another sentence. Here we form our summary without

Table 2: Comments on the article

رائع محمد , الله يجزيك الخير .
نحن في المغرب ندرس اللغة الفرنسية من سن مبكر و نفس الشيء للجزائر و تونس
أو ربما الدخول في حديث مجانين مع نفسك على المرآة ههههههههههههههههههههههههههههه مأساااااااااه تتكرر
لو لم يتوفر مدرس او مركز اريد موقع او كورس ع اليوتيوب يبدأ معى من الصفر واتابعه واحفظ ما فيه ؟؟
مقال رائع لتعلم اللغة الفرنسية
الخطوة الثانية هي الأهم في أن تجد مدر مدرساً

Table 3: Scores of summary sentences for studying French

Line #	Location	Sentence size	Weight	Weight with comment relatedness	Total
1	1.000000	0.6	1.084825	1.08482524	2.684825
3	0.333333	1.0	2.153796	2.56369449	3.897028
4	0.250000	1.0	1.704725	1.70472537	2.954725
5	0.200000	0.9	1.983678	1.98367759	3.083678
6	0.166667	0.5	1.447399	1.44739922	2.114066
11	0.090909	0.8	1.449910	1.44990981	2.340819
15	0.066667	1.0	2.292792	2.54365031	3.610317
22	0.045455	0.9	1.861965	1.86196541	2.807420
24	0.041667	1.0	1.660572	1.66057215	2.702239
28	0.050000	1.0	2.021636	2.18067658	3.230677
29	0.052632	0.9	1.877682	2.00737423	2.960006
30	0.055556	1.0	2.173890	2.43332091	3.488876
33	0.066667	0.7	1.395243	1.39524311	2.161910
34	0.071429	0.9	2.312528	2.47156817	3.442997
38	0.100000	1.0	1.841496	2.50225278	3.602253
41	0.142857	0.6	1.057153	1.05715263	1.800010
44	0.250000	0.7	1.088240	1.24728003	2.197280
45	0.333333	0.8	1.767627	1.76762702	2.900960
47	1.000000	0.2	0.709208	0.86824798	2.068248

Table 4: The summary of the article studying French

لعل مشاريع تعلم اللغات هي أكثر المشاريع التي تُبدأ و لا تُنجز
فكم من مرة عقدت العزم جازماً و حازماً أنك متم تعلم اللغة الفرنسية نهاية هذه السنة
كتبها الخمسة و بلغتها الأم لتثبت لنفسك أن هذا المشوار في نهايته. Les Misérables و أنك بعدها بادئ قراءة البؤساء في نهايته
وها أنت ذا في المكان نفسه تضع الخطة كرة أخرى بفارق بسيط الآن
فقد أضفت بنداً مراجعة سريعة في البداية
بل تحتاج لتخطوها مركب حسن إدارة الوقت و محفز الهدف
بناء قاعدة من المفردات و الجمل الأساسية و القواعد فما تقرؤه أو تسمعه في أي كتاب أو نص باللغة الفرنسية إنما هو مجموعة من الجمل
أنشئ قائمة من المفردات (أو جمل قصيرة) بمعانيها ثم احرص على مراجعتها دورياً
إن أعطيت مساحة من تفكيرك لهذه المهمة (أو اهتمام) ثم فرغت جزءاً من وقتك اليومي لها (عادة)
ليس من فائدة أن نخدع أنفسنا فأنت بحاجة إلى من يشرح لك القواعد و يدلك على خفايا اللغة و خباياها و يجنبك الأخطاء
قد تتعلم من مدرس جيد في ساعة ما تحتاج ساعات لتستخرجه من المواقع أو الكتب
الوضع الأمثل في تقديري هو أن تجد معهداً بأسعار منطقية و مدرس جيد و تلتزم فيه
فإن لم يكن ذلك متوفراً لك فابحث جيداً فإن تعذر تماماً فاتجه إلى البدائل عبر الإنترنت
تأتي أهمية هذه الأدوات والتطبيقات و المواقع من أهمية استخدام اللغة يومياً
ثم استمع إليه مع متابعة النص الفرنسي ثم استمع إليه مع ترجمة
أن الحفظ والتقدم في مساقات اللغة لا ينبغي أن تتوقف في هذه المرحلة
فما تزال تحتاج أن تزيد من رصيدك من المفردات و الجمل (و ربما الأمثال والتشبيهات الشائعة)
لكن الفارق الآن هو أنك تستمع بتعلم اللغة

Table 5: The answers of the human evaluators for each summary

	Rejected		Not-related		Satisfactory		Good		Accepted		
Doc#	H	S	H	S	H	S	H	S	H	S	Sum
1	1		1		1	3	3	3	2	6	20
2					2	2	5	6	3	2	20
3			1			2	7	4	4	2	20
4							8	7	2	3	20
5			1			2	8	6	1	2	20
Sum	1	0	3	0	3	9	31	26	12	15	

H: Humanities teachers; S: Science teacher

repeating any idea. For this purpose we form a procedure to count similar words in sentences which form the summary, then omit any redundant sentence.

In this step, if the summary included a redundant sentence then we omit it and push the next highest sentence score to be included in the summary (Table 5).

Experiments and results: Here, we try to solve Arabic text summarization problems using text extraction techniques, similar to many research applied to the English single document summarization.

We try to produce Arabic text summarizations from html documents converted to text file. Html

Table 6: Overall performance of each group as obtained from the human evaluators

| | | Testers answers | | | | | | | | | | |
| | | Rejected | | Not-related | | Satisfactory | | Good | | Accepted | | Overall |
Group name	Group size	n	(%)	N	(%)	n	(%)	n	(%)	n	(%)	performance
Humanities teachers	50	1	1	3	3	3	3	31	31	12	12	43
Sciences teachers	50	0	0	0	0	9	9	26	26	15	15	41

N: Number of evaluators

Table 7: Overall performance of the summarization system

Evaluator answer	Number of evaluators	(%)	Overall performance of the system (%)
Rejected	1	1	
Not-related	3	3	
Satisfactory	12	12	
Good	57	57	84
Accepted	27	27	
Total	100		

documents is easier to identify comments, paragraphs. The size of the summary produced is 40% of the original text.

An evaluating group was participated in this project. The group of evaluators was asked to read the original text carefully and to perform an evaluation to describe the performance of our approach in Arabic text summarization.

In this section, we illustrate the results of our statistical measures on the evaluation of the summaries produced by the methods we followed in our research.

The experiment environment: The test set is formed by using 5 Arabic html web articles in various topics including education, history, entertainment. Then we apply the statistical methods of our research to produce the summaries, then each summary is printed 20 times, after that we had a total of 100 summaries this gave us one summary for each participant included in the test.

Human evaluators: First, we distribute each pair<article, summary>for a group of evaluators who are teachers working in schools of southern shouna decorate, we asked them to read a summary and answer the evaluation to determine the effectiveness of our summary. Teachers of the evaluators group were of various majors (humanities, science) and have good reading skills. Table 6 shows the groups of evaluators who participate in evaluating our system.

Running the experiments: The evaluation process was performed during the first hour of the work, the 100 pairs<article, summary>were distributed to evaluators, one pair for each teacher and we asked them to read the summary carefully and answer associated evaluation, answers may be one of (0-rejected, 1-not-related, 2-satisfactory, 3-good, 4-accepted). We distributed the pairs for schools in the Southern Shawneh Directorate of Education which is a directorate in the Jordanian ministry of education. Statistics were computed for the analysis of data and results. Table 5 shows the answers of the human evaluators for each summary.

RESULTS AND DISCUSSION

The results obtained from the groups of evaluators are contained in Table 5. Table 6 shows the overall performance of our statistical system then we can set the following conclusions:

- The system works despite the long words per article (750 words/article).
- If the human evaluators evaluate the question as Good or Accepted, so the overall performance of the system is 84% then the summary is counted as working successfully.

Data analysis and interpretation: Table 6 and 7 summarize the results obtained from our system. The 100 evaluators were divided into two groups (humanities and science teachers).

The first group is humanities teachers which consist of 50 testers. One tester of this group chooses to reject some of the summaries (DOC1: 1 tester). Also 3 of testers decided that some summaries are not-related, (DOC1: 1 tester, DOC3: 1 tester, DOC5: 1 tester). In addition 3 of testers decided that some summaries are satisfactory, (DOC1: 1 tester, DOC2: 2 testers). Thirty one of testers decided that some summaries are good, (DOC1: 3 testers, DOC2: 5 testers, DOC3: 7 testers, DOC4: 8 testers, DOC5: 8 testers). Twelve of testers decided that some summaries are accepted, (DOC1: 2 testers, DOC2: 3 testers, DOC3: 4 testers, DOC4: 2 testers, DOC5: 1 tester).

The second group science teachers which consist of 50 testers, 9 of testers decided that some summaries are satisfactory, (DOC1: 3 testers, DOC2: 2 testers, DOC3: 2 testers, DOC5: 2 testers). Twenty six of testers decided that some summaries are good, (DOC1: 3 testers, DOC2: 6 testers, DOC3: 4 testers, DOC4: 7 testers, DOC5: 6 testers). Fifteen of testers decided that some that some summaries are accepted, (DOC1: 6 testers, DOC2: 2 testers, DOC3: 2 testers, DOC4: 3 testers, DOC5: 2 testers).

We suggest to calculate overall performance of a group:

Overall performance of a group = %
good + %accepted   (11)

Now if we count a summary as successful, if the human evaluators marked it as good or as accepted then the overall judgments by humanities teachers group is (43%) and the overall judgments by Science teachers group is (41%).

If we focus on rejected, not-related and satisfied as not successful, we note that a humanities teachers group has rejected one <article, summary> and the summary is DOC1. And marked 3 <article, summary> as not-related and the summaries are DOC1, DOC3, DOC5, Both groups have marked 11 <article, summary> as satisfactory.

We note that both groups (6 testers) have marked DOC1 as not successfully summarized from (16 testers); We note that DOC1 is a scientific article which contain many scientific terms which need in-depth understanding, also the article's sentences is connected and explain each other. Also we found the article itself is poor in its organization and purpose and the readers who wrote the comments were not sure about the subject of the article.

CONCLUSION

Text summarization techniques are now the most popular approaches in Natural Language Processing. It's importance come from the various applications in information industry.

In this project which was implemented using Perl language. The project's code segments an article (5 articles) into paragraphs, sentences and words, then remove stop words from segmented article, after that the code rank sentences using many factors such that sentence position and location, sentence length, terms weighting. At last we used simple excel sheet to select high score sentences to be included in the summary and arrange sentences as they appear in the article to preserve readability of the summary.

The articles which were used in our experiment had a wide range of words count from short length to long article, "How To Study French" article was about 520 words, at the other side "Ten advices to use time efficiently" article consisted of 730 words.

We encountered many problems with Arabic articles, such that missed and incorrect punctuations, another problem was the poor construction of some articles.

Another type of problems were about comments on articles; Arabic readers avoid writing comments on

articles, comments on article are written lately after long time from writing the articles, also many comments were written to replay on another comments not on the original article.

We have designed a system for obtaining Arabic text summarization using segment extraction combined with reader comments relatedness to an article to determine what readers focus on sentences and give it higher scores in selecting the suitable sentences included in the summary. We design a stop-word removal, a tokenizer to get rid of the lack of tools in NLP for Arabic.

We design a text summarization system which take the original article as an input and produce a summary which contained 40% of the original articles, this system is planned to be available online.

At last we have good evaluations feedback for our system from human evaluators. The experiment produced efficient summaries; about 84% of evaluators decided that the performance of the system was good or accepted we forced some problems using comment relatedness ranking such as lack of Arabic comments and comments on comments not on article, also most Arabic readers prefer to use Colloquial Dialect in writing comments.

RECOMMENDATIONS

The feedback from our experiment helped us to determine the effectiveness of our system and determine our next step by combining our approach with entity-level approach such as graph topology which represent the relations of each text unit of the original document to determine connections between short sentences. We tend to increase the number of words per article and include multiple title in articles to test our system performance and compare it with comment relatedness ranking.

We are planning to increase the number of pairs <article, summary>, we aim to include different topics such that political and religious articles. We are also interesting in increasing the number of human evaluators by 1000 evaluators instead of 100 evaluators.

ABBREVIATIONS

The following terms are abbreviations used in this study:

ATS   : Arabic Text Summarization
IDF   : Inverse Document Frequency
NLP   : Natural Language Processing
POST : Part Of Speech Tagging
TF   : Term Frequency
TS   : Text Summarization
IR   : Information Retrieval

REFERENCES

Balahur-Dobrescu, A., M. Kabadjov, J. Steinberger, R. Steinberger and A. Montoyo, 2009. Summarizing opinions in blog threads. Proceeding of the Pacific Asia Conference on Language, Information and Computation Conference, pp: 606-613.

Deschacht, K. and M.F. Moens, 2007. Text analysis for automatic image annotation. Proceeding of the 45th Annual Meeting of the Association of Computational Linguistics, pp: 1000-1007.

Douzidia, F. and G. Lapalme, 2004. Lakhas, an Arabic summarization system. Proceeding of DUC'04, pp: 128-135.

Edmundson, H., 1969. New methods in automatic extracting. J. ACM, 16(2): 264-285.

El-Haj, M. and B. Hammo, 2008. Evaluation of query-based Arabic text summarization system. Proceeding of the International Conference on Natural Language Processing-knowledge Engineering. IEEE Computer Society Press, Los Alamitos, CA, pp: 1-7.

Fan, J., Y. Gao, H. Luo, D.A. Keim and Z., Li, 2008. A novel approach to enable semantic and visual image summarization for exploratory image search. Proceeding of the 1st ACM International Conference on Multimedia Information Retrieval (MIR '08), pp: 358-365.

Filho, P.P.B., V.R. de Uzêda, T.A.S. Pardo and M.D.G.V. Nunes, 2007. Experiments on applying a text summarization system for question answering. In: Peters, C. *et al.* (Eds.), CLEF 2006. LNCS 4730, Springer-Verlag, Berlin, Heidelberg, pp: 372-376.

Hammo, B.H., H. Abu-Salem and M.W. Evens, 2011. A hybrid arabic text summarization technique based on text structure and topic identification. Int. J. Comput. Process. Lang., 23(1): 39-65.

Luhn, H.P., 1958. The automatic creation of literature abstracts. IBM J. Res. Dev., 2(2): 159-165.

Mani, I., G. Klein, D. House and L. Hirschman, 2002. SUMMAC: A text summarization evaluation. Nat. Lang. Eng., 8(1): 43-68.

Nenkova, A., 2006. Summarization evaluation for text and speech: Issues and approaches. INTERSPEECH-2006, Paper No. 2079-Wed1WeS.1.

Nobata, C., S. Sekine, K. Uchimoto and H. Isahara, 2002. A summarization system with categorization of document sets. Proceeding of the 3rd NTCIR Workshop Meeting, pp: 33-38.

Schlesinger, J.D., D.P. O'Leary and J.M. Conroy, 2008. Arabic/English multi-document summarization with CLASSY- the past and the future. In: Gelbukh, A. (Ed.), CICLING 2008. LNCS 4919, Springer-Verlag, Berlin, Heidelberg, pp: 568-581.

Sobh, I., N. Darwish and M. Fayek, 2007. An optimized dual classification system for Arabic extractive generic text summarization. Proceeding of the 7th Conference on Language Engineering (ESLEC). Cairo, Egypt, pp: 149-154. Retrieved from: http://www.RDI-eg.com/ RDI/ technologies/ paper.htm.

Spärck Jones, K., 2007. Automatic summarising: The state of the art. Inform. Process. Manag., 43(6): 1449-1481.

Tofigh, S.M., R. Gopal Ra and H.H. Javadi, 2013. AHP techniques for Persian text summarization. Malays. J. Comput. Sci., 26(1): 2-8.

Wang, F.L. and C.C. Yang, 2006. The impact analysis of language differences on an automatic multilingual text summarization system. J. Am. Soc. Inf. Sci. Tec., 57(5): 684-696.