SumSAT: Hybrid Arabic Text Summarization Based on Symbolic and Numerical Approaches

Said Moulay Lakhdar and Mohamed Amine Cheragui
Mathematics and Computer Science Department
Ahmed Draia University
Adrar, Algeria
moulaylakhdarsaid@yahoo.fr, m_cheragui@univ-adrar.dz

Abstract

The increase in number and volume of electronic documents makes the development of applications such as text summarization crucial, in order to facilitate the task for persons who want to consult their documents. The purpose of an electronic document summary is the same as that of a book abstract; it informs the reader about the subject matter. The usefulness of the summary is distinguished by the limited time devoted to its reading to synthesize all the ideas that the author wants to spend.

The objective of this paper is to present our SumSAT tool, which is an Arabic text summarization system, adopting an extraction approach. The originality of our work lies in the use of a hybrid methodology that combines three methods: contextual exploration, indicative expression, and graph method. The proposed strategy is evaluated by comparing the obtained results with human summaries using recall and precision metrics.

1 Introduction

Considered for a long time as one of the main topic of natural language processing (Luhn, 1958), Text summarization has only grown in importance since the late 90s with the proliferation of Internet use and the emergence of large amounts of information (Maâloul, 2012), which has forced researchers to make more effort to make the text summarization process more efficient. This effectiveness is linked to two (02) essential factors, on the one side reducing the size of the text and on the other side keeping the basic idea (or ideas) that are conveyed by the text.

The purpose of this paper is to present SumSAT which is a text summarization system developed for the Arabic texts. The originality of our work lies in making a contribution not only in the pre-processing phase which consists in preparing the text for the summarization process but also in the processing phase where we have chosen a hybrid strategy that showcases several techniques from different approaches.

The rest of the paper is organized as follows: Section 2 will focus on the principle of Text summarization. Section 3, briefly describes works in the literature that are related to Arabic text summarization. Section 4 presents our hybrid approach based on contextual exploration, Indicative expression and graph method. Section 5 introduces the SumSAT tool. The results of experiments on the dataset of Arabic are discussed in Section 6. Finally, a conclusion that presents the assessment of our work associated with perspectives and future work.

2 Text Summarization Between Abstraction and Extraction

There are two very divergent approaches to automatically generate summaries (Pai, 2014; Munot and Govilkar 2014; Allahyari et al., 2017). Summarization based on Abstraction and Summarization based on Extraction.; the first one (Abstraction approach) comes from the field of artificial intelligence and aims to use natural language processing techniques (such as semantic representation and modification, text understanding) to generate a new summary (with new words) that covers the main ideas found in the original text. This production process remains relatively difficult to compute, and text generation is still very imperfect (Pal and Saha 2014; Zhu et al.,2009; D'Avanzo et al., 2004).

However, in the extraction-based approach, the main purpose is to extract the most important or significant phrases in the original text and combining them to make a summary. Its objective is to produce the summary without going through deeper analysis, so the main task is to determine the relevance of these phrases according to one or more criteria (generally a statistical features) (Mohamed, 2016; Oufaidaa et al.,2014).

3 Related Work

Compared to other languages such as English, works on the Arabic language are very few due mainly to its morphological and syntactic complexity. The table below gives an indication of some tools and works done on Arabic text summarization (Douzidia and Lapalme, 2004; Sobh et al., 2006; Schlesinger et al., 2008; Mahmoud et al., 2009; Alotaiby et al., 2012; Belguith, 2014; AL-Khawaldeh and Samawi , 2015; Belkebir and Guessoum, 2015; Lagrini et al., 2017).

Tool and Work	Methodology	
LAKHAS (Douzidia and Lapalme)	Numerical	Sentence position terms frequency title words cue words
Al Sanie	Symbolic	RST (Rhetorical Structure Theory)
Sobh, Ibrahim, Nevin Darwish, and Magda Fayek	Numerical	Bayesian Genetic Programming classification
CLASSY (Schlesinger, Judith D., Dianne P. O'leary, and John M. Conroy)	Numerical	Log-likelihood
AQBTSS and ACBTSS (Mahmoud O.EI-Haj and Bassam H. Hammo)	Numerical	TF-IDF

Table 1: Summarizing reviewed Works and tools (A).

Tool and Work	Methodology	
Alotaiby, Fahad, Salah Foda, and Ibrahim Alkharashi.	Numerical	Frequency of non-stop words Machine Learning
Belghuith	Hybrid	RST Machine Learning
LCEAS (AL-Khawaldeh and Samawi)	Hybrid	Based on semantic relations Roots extraction
Belkebir and Guessoum	Numerical	Machine Learning
Samira Lagrini, Mohammed Redjimi and NabihaAzizi	Hybrid	RST machine Learning

Table 2: Summarizing reviewed Works and tools (B).

4 SumSAT's general architecture

SumSAT is a text summarization system by extraction. To generate a summary, our system operates in three main steps, which are:

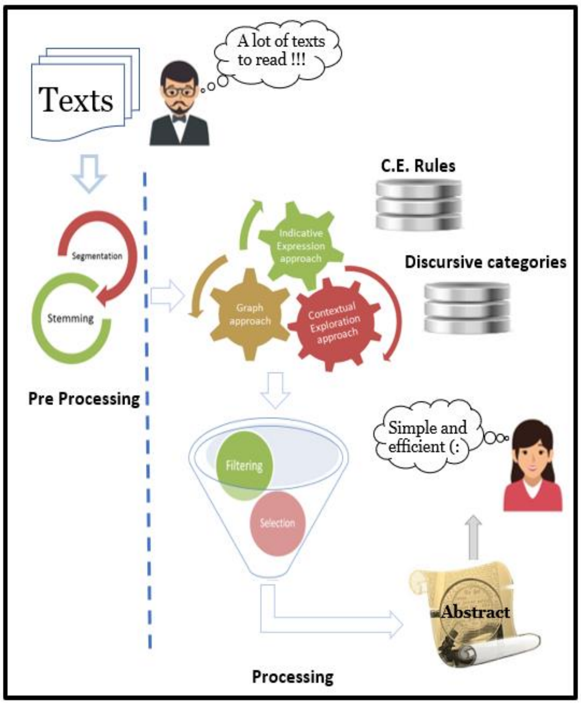


Figure 1: General architecture of SumSAT.

4.1 Step 1: Pre-processing

This phase is divided into two sub-phases:

Segmentation: Since the text summarization operation consists in selecting relevant phrases, the first task to be performed is the segmentation of the source text into phrases. The method used to divide a text is based on the contextual exploration method, where the input is a plain text in the form of a single text segment. The segmentation starts with detecting the presence of indicators, which are punctuation marks (« . », « ; », « : », « ! », « ? ».). If there is an indicator, segmentation rules will be applied to explore the contexts (before and after) to ensure that additional indicators are present and that certain conditions are met. In the case of an end of a phrase, this decision is converted into the action of segmentation of the text into two textual segments. Thus, and by repeating this operation on the resultant segments, we obtain a set of textual segments which, placed next to each other, which form the input plain text.

It is important to specify that in our segmentation the dot « . » cannot be always considered as an indicator of a sentence end; i.e., cases like : abbreviation, acronym or a number in decimal, where particular rules can be added.

Stemming: This operation consists of transforming, eventually agglutinated or inflected word into its canonical form (stem or root) (Roubia et al., 2017). In our case, we need the results of the Stemming in the graph method in order to define the most important phrases. To generate these roots, we use the Full-Text Search technique, which allows us to generate the roots of words composing the phrases and eliminate the stopwords. This technique also generates other features such as ranking (rank value) to classify the found phrases, in order to filter the relevant ones according to their scores.

4.2 Step 2: Processing

Since we adopt an extraction approach, the main task is to evaluate the phrases, select the most relevant ones, then build the summary. We adopted a hybrid approach combining three methods: the contextual exploration (main methods), the indicative expression and graph model (secondary methods). The secondary methods will scramble on the result of the principal method to give better results or provide a solution in the case that contextual exploration is not efficient.

Contextual Exploration method: This method has been chosen in order to produce a consistent summary and to offer users the possibility to choose the summary by point of view, where the information to be summarized is classified into discursive categories. The contextual exploration module receives a segmented text as input (the result of the segmentation module). The first task is to detect the presence of some linguistic indicators in each segment. Once an indicator is found, all contextual exploration rules related to that indicator will be set to find additional clues and to verify the conditions required by that rule. If all conditions are verified, an annotation action, determined by the exploration rule, is performed on the segment exactly where the linguistic indicator is placed.

For our SumSAT System, we have defined 13 discursive categories, each category has its own complementary clues (See figure 2).

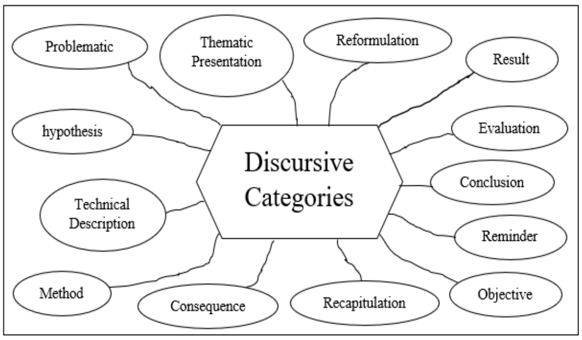
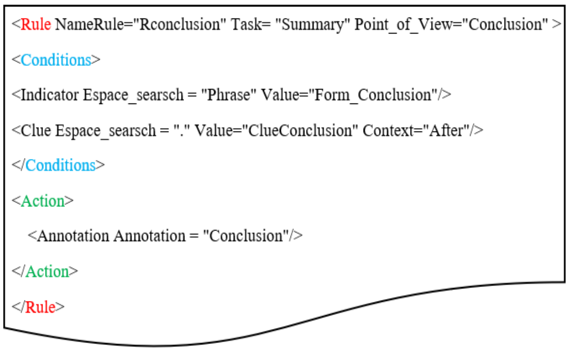


Figure 2: The discursive categories defined for SumSAT.

Example: The following example illustrates an application of our method to select sentences that contains information about the discursive category "conclusions and results". One of the rules associated with this category is as follows:



```
<Rule NameRule="Rconclusion" Task= "Summary" Point_of_View="Conclusion" >
<Conditions>
<Indicator Espace_searsch = "Phrase" Value="Form_Conclusion"/>
<Clue Espace_searsch = "." Value="ClueConclusion" Context="After"/>
</Conditions>
<Action>
   <Annotation Annotation = "Conclusion"/>
</Action>
</Rule>
```

Figure 3: Example of discursive rule.

The rule, delimited by the tag (<Rule> and </Rule>), consists of two parts :

- Condition part: delimited by (<Conditions> and </Conditions>): It groups together information about the indicator (delimited by <Indicator and />) associated with an information category, and information about the additional clues (<clue and />) that are associated with it.

- Actions part: delimited by (<Actions> and </Actions>): Action to be done, after verifying the existence of additional clues and the required conditions.

  Where:

  ✓ NameRule: the name that identifies the rule.

  ✓ Task: The task this rule performs since contextual exploration can be used for annotation and summary generation, as it can be used for segmentation.

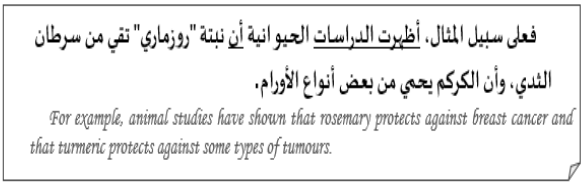
  ✓ Point of View: Represents the category name of the information retrieved.

  ✓ Search_space: Space or context, where the additional clue is located; whether the search is done in the phrase itself or in the paragraph.

  ✓ Value: It is the name of the file where the indicators are stored, or the name of the file where the clues are stored, associated with this category of information.

  ✓ Context: Specifies whether the search for additional clues should be done before or after the indicator.

Consider the following phrase to be annotated (applying the rule mentioned above):



فعلى سبيل المثال، أظهرت الدراسات الحيوانية أن نبتة "روزماري" تقي من سرطان الثدي، وأن الكركم يحمي من بعض أنواع الأورام.

For example, animal studies have shown that rosemary protects against breast cancer and that turmeric protects against some types of tumours.

Figure 4: Example of a contextual exploration rule.

In this phrase, it can be said that the complementary clue (أن) is present after the indicator (أظهرت الدراسات). Therefore, the action to be taken is indicated in the actions part (delimited by <Actions> and </Actions>); so, this phrase assigned the value 'Conclusion' to indicate that it contains information concerning a result or conclusion.

Indicative expression: This method is selected to offer the possibility of generating a summary of a general order, or a specific field; sport, culture, economy, etc. This method consists of identifying phrases that contain indicators. These indicators are determined according to the field of the text to be analyzed, and its main task is to identify indicators in phrases, neglecting the additional clues. Using the following formula:

$$Score_{cue}(S) = \begin{cases} 1 & if \quad S \text{ is an indicator} \\ 0 & else \end{cases} \quad (01)$$

Graph method: In order to reduce the deficiencies of SumSAT's, we have used a hybrid approach that integrates a symbolic method (E.C.) and numerical methods (graph and indicative expression methods). The use of this hybrid approach allowed us to offer the user the possibility to choose a summary by point of view through contextual exploration, as well as the possibility to choose a default summary, to cover cases where the information is not present in the form of a discursive category.

The generation of the summary, using the graph method, consists of selecting the most representative phrases of the source text, since it attributes to the sentences a relevance score or similarity measure by calculating the number of intersection terms. These terms are the result of the stemming process performed in the pre-processing process.

Suppose that we have a text composed of six sentences (P1, P2, P2, ..., P6). After applying stemming for each sentence, the total number of terms shared with all the others are given in the table below:

Phrases	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**
Total number of Stems (Roots) shared with all other phrases	9	8	7	3	6	5

Table 2: Phrases Weight.

Modelling this problem for the summary is like considering: The document as a graph, the phrases as nodes of this graph, the intersections of the phrases as edges of this graph, the total number of intersecting terms (stems or roots), of a phrase with all the others, as a weight of the node representing this phrase. Finally, to generate the summary we use the Greed algorithm.

	P1	P2	P3	P4	P5	P6
P1	0	0	1	1	1	1
P2	0	0	1	1	0	1
P3	1	1	0	0	1	0
P4	1	1	0	0	0	1
P5	1	0	1	0	0	1
P6	1	1	0	1	1	0

Table 3: Phrase intersection matrix

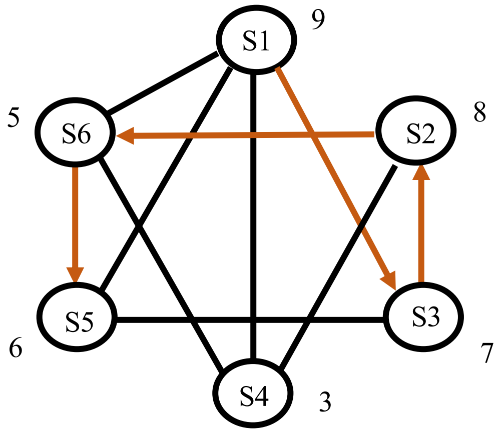


Figure 5: Pathway followed using the Greedy algorithm.

4.3 Step 3: Filtering and selection

The generation of the summary must take into consideration the user's requirements, and the compression ratio to determine the relevant phrases to be selected. The final summary is made up of all phrases that fulfill the following conditions:

- Phrases that belong to the discursive categories, or to the selected domains (chosen by the user) ;
- And/or the phrases that appear in the list of nodes visited by the graph method (the case of the default summary) ;
- The number of phrases is limited by the summary rate, introduced by the user ;
- The appearance order of the phrases in the summary must respect the order of these phrases in the source text.

In order to generate a dynamic summary, a link is established between the summary phrases and their corresponding phrases in the source text.

5 Presentation of SumSAT

SumSAT (Acronym of Summarization System for Arabic Text) is a web application system that runs at web browsers. Its execution is local to the IIS server (Internet Information Server), of Windows. The interaction between our system and Microsoft SQL Server is done by queries (T-SQL transactions). SumSAT is introduced to the user through a GUI, based on HTML5, ASP, C#, and Silverlight.

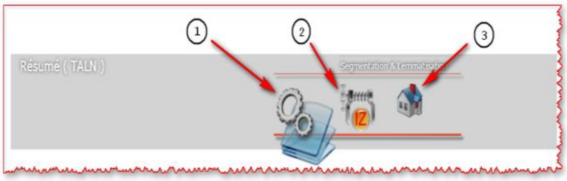
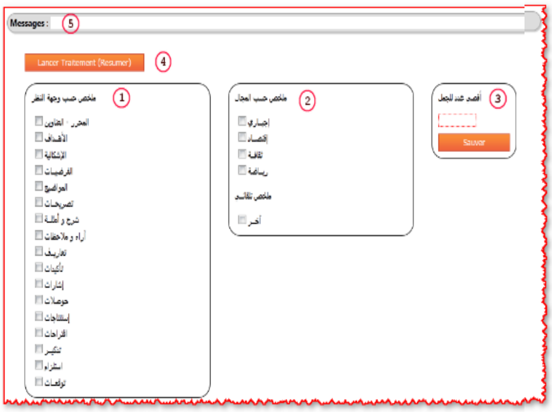


Figure 6: GUI Main Menu.



Messages :

Lancer Traitement (Resumer)

ملخص حسب وجهة النظر

□ المحرر - العناوين
□ الأهداف
□ الإشكالية
□ الفرضيات
□ المواضيع
□ تصريحات
□ شرح و أمثلة
□ آراء و ملاحظات
□ تعاريف
□ تأكيدات
□ إشارات
□ حوصلات
□ إستنتاجات
□ اقتراحات
□ تنكير
□ استلزام
□ توقعات

ملخص حسب المجال

□ إخباري
□ إقتصاد
□ ثقافة
□ رياضة

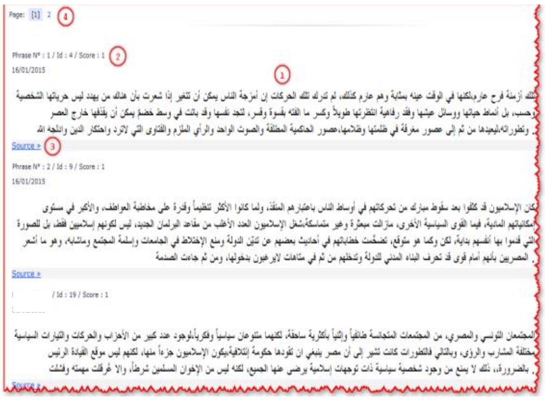
ملخص تلقائي

□ آخر

أقصى عدد للجمل

Sauver

Figure 7: GUI Generation of Summary.



Page: [1] 2

Phrase N° : 1 / Id : 4 / Score : 1
16/01/2015

تلك أزمنة فرح عارم،لكنها في الوقت عينه بمثابة وهم عارم كذلك، لم تدرك تلك الحركات إن أمزجة الناس يمكن أن تتغير إذا شعرت بأن هناك من يهدد ليس حرياتها الشخصية وحسب، بل أنماط حياتها ووسائل عيشها وفقد رفاهية انتظرتها طويلاً وكسر ما ألفته بقسوة وقسر، لتجد نفسها وقد باتت في وسط خضمّ يمكن أن يقذفها خارج العصر وتطوراته،ليعيدها من ثم إلى عصور مغرقة في ظلمتها وظلامها،عصور الحاكمية المطلقة والصوت الواحد والرأي الملزم والفتاوى التي لاترد واحتكار الدين والذات الله .

Source »

Phrase N° : 2 / Id : 9 / Score : 1
16/01/2015

كان الإسلاميون قد كافؤوا بعد سقوط مبارك من تحركاتهم في أوساط الناس باعتبارهم المنقذ، ولما كانوا الأكثر تنظيماً وقدرة على مخاطبة العواطف، والأكثر في مستوى إمكانياتهم المادية، فيما القوى السياسية الأخرى، مازالت مبعثرة وغير متماسكة،شغل الإسلاميون العدد الأغلب من مقاعد البرلمان الجديد، ليس لكونهم إسلاميين فقط، بل للصورة التي قدموا بها أنفسهم بداية، لكن وكما هو متوقع، تضمّنت خطاباتهم في أحاديث بعضهم عن تديّن الدولة ومنع الإختلاط في الجامعات وإسلمة المجتمع وماشابه، وهو ما أشعر المصريين بأنهم أمام قوى قد تحرف البناء المدني للدولة وتدخلهم من ثم في متاهات لايرغبون بدخولها، ومن ثم جاءت الصدمة .

Source »

/ Id : 19 / Score : 1

المجتمعان التونسي والمصري، من المجتمعات المتجانسة طائفياً وإثنياً بأكثرية ساحقة، لكنهما متنوعان سياسياً وفكرياً،لوجود عدد كبير من الأحزاب والحركات والتيارات السياسية مختلفة المشارب والرؤى، وبالتالي فالتطورات كانت تشير إلى أن مصر ينبغي ان تقودها حكومة إئتلافية،يكون الإسلاميون جزءاً منها، لكنهم ليس موقع القيادة الرئيس بالضرورة،، ذلك لا يمنع من وجود شخصية سياسية ذات توجهات إسلامية يرضى عنها الجميع، لكنه ليس من الإخوان المسلمين شرطاً، والا غُرقت مهمته وفشلت .

Source »

Figure 8: GUI of the Result (summary)

6 Experimentation and Results

SumSAT's summary generation is based on a hybrid approach where the discursive annotation constitutes its main task: the generated summary is based on the concept of point of view. Therefore, the relevance of a phrase depends on the presence of surface linguistic markers characterizing (referring to) a discursive category. The evaluation of the summary generation process by point of view consists of the evaluation of the discursive annotation task made by SumSAT.

The objective of this evaluation is to know the percentage of phrases correctly annotated by the system, compared to the total number of annotated phrases, and compared to the total number of manually annotated phrases (reference summary). This can be expressed by measuring:

6.1 The precision rate

The number of correct discursive categories, detected by the system, compared to the total number of discursive categories detected by the system.

6.2 The Recall Rate

The number of correct discursive categories, detected by the system, compared to the total number of discursive categories presented in the reference summary

The precision and recall rates are calculated as follows:

$$\text{Precision } (\%) = ({}^{a}/_{b}) * 100 \qquad (02)$$

$$\text{Recall } (\%) = ({}^{a}/_{c}) * 100 \qquad (03)$$

Where :

- a : Number of automatically assigned correct annotations.
- b: Number of automatically assigned annotations.
- c: Number of manually assigned correct annotations.

For this purpose, we have set up corpora composed of twenty-five texts, and their corresponding summaries (The reference summaries are manually compiled by two experts). For each of the selected texts, we have proceeded to the generation of summaries, by discursive categories one by one. The evaluation consists of applying the metrics, in order to criticize and conclude based on the results obtained.

The results of the calculated rates, as well as the precision and recall results, are illustrated in Table 5, 6 and 7 and by representative graphs (Figure 9, 10 and 11). These results are calculated for all the selected texts in the corpora, and for each of the discursive categories adopted by SumSAT. For all categories, the precision rate is higher than 66%, except for four of them (hypothesis, Recapitulation, Reminder, Prediction), which have a precision rate between 40% and 50%. Similarly, the recall rate is higher than 66%, except for three categories that have a recall rate between 30% and 50% (Prediction, Definition and Reminder) . This shows that SumSAT has promising results which can be improved, despite the difficulties of generating coherent summaries.

- Precision rate: These results show that much more work needs to be done on refining surface markers to maximize this rate. In technical terms, it is necessary to work on two parameters. The first parameter, related to regular expressions, detects discursive markers (indicators and additional clues). The second parameter, linguistic (the good choice of these discursive markers).
- Recall rate: The results show that the work which can contribute to improving these results will be linguistic, especially the collection of discursive markers in order to enrich linguistic resources.

Note that the obtained results are influenced by the divergence of the texts from the point of view of style, discursive and argumentative strategies, and the covered topic. This means that the surface markers, for some categories, are rarely the same from one text to another. Similarly, the indicators are sometimes weak and cannot refer to a discursive category. Moreover, the additional clues are sometimes equivocal.

Category	Precision (%)	Recall (%)
Objective	73,68	82,35
hypothesis	42,03	70
Conclusion	77,78	70
Explanation	88,57	95,38
Consequence	77,27	70,83

Table 3: SumSAT evaluation using P/R (01).

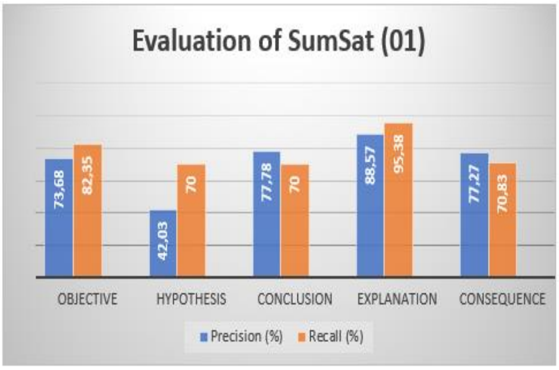


Figure 9: Graphical representation of SumSAT's evaluation results (01).

Category	Precision (%)	Recall (%)
Definition	66,67	32,67
Confirmation	97,5	82,98
Problematic	66,67	66,67
Reminder	50	44,02
Recapitulation	50	88,24

Table 4: SumSAT evaluation using P/R (02).

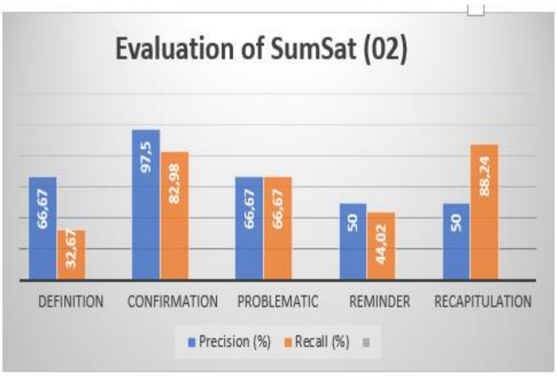


Figure 10: Graphical representation of SumSAT's evaluation results (02).

Category	Precision (%)	Recall (%)
Author, Title & Subtitle	91,94	91,94
Thematic	85,71	66,67
Prediction	50	50
Finding & opinion	90	69,26
Enunciation	94,94	91,85

Table 5: SumSAT evaluation using P/R (03).

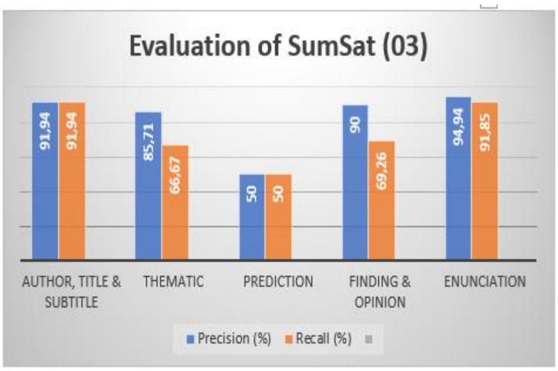


Figure 11: Graphical representation of SumSAT's evaluation results (03).

7 Conclusion and Continuing Efforts

In this paper, we have presented SumSAT which is an Arabic text summarization system that adopts a hybrid approach (i.e: contextual exploration method, indicative expression method, and graph model method) to build summary. The work we have done has given us an overview of the difficulties that we have encountered in the field of Arabic text summarization. In pre-processing, the incorrect use of punctuation marks (author's style) induces segmentation errors, and as a result, the relevance of phrases is incorrect, which gives an incoherent summary. On the processing phase, one of the difficulties met, and which influences the performance of the system, is the manual search for linguistic markers, to enrich the list of discursive categories. This task costs time and resources, which has reduced the list of the information offered by SumSAT. In addition, we found that the representative phrases with a high weight may not be selected because of the restrictions on the incrementation of the list of visited summits when the transition is made only between the adjacent ones (Graph model method).

Based on the obtained results, we propose an amelioration of the methods used to generate the summary by making a modification, such that the glutton algorithm (graph model method) gives the advantage to the representative nodes, without being limited by the transitions between the adjacent summits. Also, the integration of a tool for identifying surface linguistic markers in documents is a good way to enrich the system's linguistic resources.

References

Allahyari Mehdi , Pouriyeh Seyedamin , Assefi Mehdi, Safaei Saeid , Trippe D. Elizabeth, Gutierrez B.Juan and Kochut Krys. 2017. Text Summarization Techniques: A Brief Survey. *In Proceedings of arXiv. rXiv:1707.02268.*

Alkhawaldeh Fatima Taha and Samawi W. Venus. 2015. Lexical cohesion and entailment based segmentation for arabic text summarization (lceas). *The World of Computer Science and Information Technology Journal*. (WSCIT)5 (3): 51-60.

Alotaiby Fahad, Foda Salah and Alkharashi Ibrahim. 2012. New approaches to automatic headline generation for Arabic documents. *Journal of Engineering and Computer Innovations*. Vol. 3(1), pp. 11-25.

Belguith Lamia Hadrich. 2014. Automatic summarization. Natural Language Processing of Semitic Languages. Springer Berlin Heidelberg.371-408.

Belkebir Riadh. and Guessoum Ahmed. 2015. A supervised approach to arabic text summarization using adaboost. New Contributions in Information Systems and Technologies. Springer International Publishing. 227-236.

D'Avanzo Ernesto, Magnini Bernardo and Vallin Alessandro. 2004. Keyphrase Extraction for Summarization Purposes: The LAKE System at DUC-2004. *In Proceedings of the 2004 Document Understanding Conference (DUC2004),* Boston, MA.

Douzidia Fouad Sofiane and Lapalme Guy. 2004. Lakhas, an Arabic Summarization System. *In Proceedings of the 2004 Document Understanding Conference (DUC2004),* Boston, MA.

Lagrini Samira, Redjimi Mohamme and Azizi Nabiha. 2017. Extractive Arabic Text Summarization Approaches. *In Proceeding of the 6th International Conference, ICALP 2017,* Fez, Morocco.

Luhn Hans Peter. 1958. The automatic creation of literature abstracts. *IBM Journal of Research and Development, Volume 2 Issue 2.*

Maâloul Mohame Hedi. 2012. Approche hybride pour le résumé automatique de textes. Application à la langue arabe. *thesis doctorat*, University Aix-Marseille.

Mahmoud El hadj, Kruschwitz Udo and Fox Chris. 2009. Experimenting with automatic text summarisation for Arabic. *Language and Technology Conference.* Springer Berlin Heidelberg. 490-499.

Mohamed Ashraf Ali. 2016. Automatic summarization of the Arabic documents using NMF: A preliminary study. *In Proceedings of the 11th International Conference on Computer Engineering & Systems (ICCES).* Egypt

Munot Nikita and Govilkar S. Sharvari. 2014 . Comparative Study of Text Summarization Methods. *International Journal of Computer Applications*. Volume 102–No.12.

Oufaida Houda., Noualib Omar and Blache Philippe. 2014. Minimum redundancy and maximum relevance for single and multi-document Arabic text summarization. *Journal of King Saud University - Computer and Information Sciences,* Volume 26, Issue 4.

Pal Alok Ranjan and Saha Diganta. 2014. An Approach to Automatic Text Summarization using WordNet. *In Proceedings of the 4th Advance Computing Conference (IACC)*, Page(s): 1169 – 1173. India

Pai Anusha. 2014. Text Summarizer Using Abstractive and Extractive Method. *International Journal of Engineering Research & Technology*, Vol. 3 Issue 5.

Rouibia Rima., Belhadj Imane. & Cheragui Mohamed Amine. 2017. JIDR: Towards building hybrid Arabic stemmer. *In Proceeding of the 1st IEEE International Conference on Mathematics and Information Technology (ICMIT)*. Adrar. Algeria.

Schlesinger D. Judith , O'Leary P. Dianne and Conroy M. John 2008. Arabic/English multi-document summarization with CLASSY—the past and the future. *In Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*. Springer Berlin Heidelberg. 568-581,

Sobh Ibrahim , Darwish Nevin Mahmoud , Fayek Magda B. 2006. An Optimized Dual Classification System for Arabic Extractive Generic Text Summarization. Available at: http://www.rdi-eg.com/rdi/technologies/papers.htm.

Zhu Junyan, Wang Can, He Xiaofei, Bu Jiajun, Chen Cun, Shang Shujie, Qu Mingcheng and Lu Gang. 2009. Tagoriented document summarization. *In Proceedings of the 18th International Conference on World Wide Web*, WWW '09, ACM, New York, NY, USA.