

18<sup>th</sup> International Conference on Knowledge-Based and Intelligent  
Information & Engineering Systems - KES2014

## Topic segmentation for textual document written in Arabic language

Anja Habacha Chaibi\*, Marwa Naili, Samia Sammoud

*RIADI-ENSI, University of Manouba, Manouba 2010, Tunisia*

---

### Abstract

Topic segmentation is important for many natural language processing applications such as information retrieval, text summarization... In our work, we are interested in the topic segmentation of textual document. We present a survey of related works particularly C99 and TextTiling. Then, we propose an adaptation of these topic segmenters for textual document written in Arabic language named as ArabC99 and ArabTextTiling. For experimental results, we construct an Arabic corpus based on newspapers of different Arab countries. Finally, we evaluate the performance of these new segmenters by comparing them together and to related works using the metrics WindowDiff and F-measure.

© 2014 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Peer-review under responsibility of KES International.

**Keywords:** Topic segmentation; Arabic language processing; ArabC99; ArabTextTiling.

---

### 1. Introduction

The aim of topic segmentation is to divide a document into segments, such that each segment is thematically coherent and consecutive segments are about different topics. This technique is used to improve the access to information. For example, in information retrieval, short relevant text segments that directly correspond to the user's query can be returned instead of long documents. For text summarizing, a better summary can be obtained from topically segmented documents. For the last years, several approaches have been proposed for the topic segmentation and they can be classified in endogenous approach and exogenous approach. The first approach exploits the information contained in the text to be segmented such as lexical repetition. In the other hand, the second approach uses external resources like: thesaurus, dictionary and co-occurrence network.

While extensive research has targeted the topic segmentation for the English language, few have studied it in other languages especially for the Arabic language. Indeed, for the last years, many topic segmenters for English language have been presented by several authors. For example, we mention TextTiling which is developed by Hearst<sup>1</sup>. She uses a sliding window and computes similarities between adjacent blocks based on their frequency vectors. Choi<sup>2</sup> presented a new topic segmenter which is C99. This algorithm is based on lexical cohesion and it uses the cosine

---

\* Corresponding author. Tel.: +216-95-238-571.

E-mail address: [anja.habacha@ensi.rnu.tn](mailto:anja.habacha@ensi.rnu.tn)

metric to compute similarity between sentences. Later, Choi<sup>3</sup> improved his algorithm by using Latent Semantic Analysis (LSA) to extract semantic knowledge from corpora. Ferret<sup>4</sup> proposed his own topic segmenter F06 which is based on TextTiling algorithm. Later, he improved it by using the thematic similarities between words and he named it F06T. He used also a co-occurrence network in his third algorithm F06C. Then, he combined F06T and F06C and proposed the F06CT algorithm.

Unlike the English language, there is a lack of research for the Arabic language. In fact, the specific issue dealing with topic segmentation for the Arabic language is raised in the research of Brants et al.<sup>5</sup> in 2002, El-Shayeb et al.<sup>6</sup> in 2007, Tourir et al.<sup>7</sup> in 2008 and Harrag et al.<sup>8</sup> in 2010. In the work of Brants et al.<sup>5</sup>, new topic segmenter has been presented, namely TopSeg, which is based on Probabilistic Latent Semantic Analysis (PLSA). The aim of this segmenter is to identify boundaries between concatenated texts. The limit of this work is that TopSeg ignores the stemming step and uses full forms terms and character n-grams which increase the execution time. For the evaluation, Brants et al.<sup>5</sup> used the error probability which have been criticized for being biased, e.g., it penalizes false negatives (missed boundaries) more than false positives (erroneous additional boundaries). In the work of El-shayeb et al.<sup>6</sup>, a comparative analysis of three different text segmentation algorithms (SeLeCT, LCseg and TextTiling) on Arabic news stories, was presented. Also, a combined system of SeLeCt and LCseg (ModSeleCT) was described. For that, each algorithm was implemented, adopted for Arabic language and evaluated on an Arabic Reuters news story dataset. The objective of this work is to identify boundaries between 1000 concatenated news stories. For the evaluation, four metrics have been used: Recall, Precision, Rseg and WindowDiff. In the work of Tourir et al.<sup>7</sup>, an automatic technique to help segment the Arabic texts while preserving the semantics was presented. This technique is based on an empirical study on the sentences and clauses connectors. In order to evaluate the segmentation process, only ten Arabic essays were segmented and the results were compared to manual segmentations performed by linguistic experts. So, for the comparison they used two factors: correct hit and incorrect hit. The correct hit represents the position marked by the process as a segment boundary and agreed by the judge. Incorrect hit represents the position marked by the process as a segment boundary and the judge disagrees with it. In the work of Harrag et al.<sup>8</sup>, the two topic segmenters C99 and TextTiling was adapted to the Arabic language using the light stemmer. As result, two segmenters have been presented: TopSegArab and ArabTiling. To evaluate the performance of these two systems, only five texts were segmented and the results were compared with the judgments of a group of seven readers. The comparison was accomplished by using three metrics: Recall, Precision and F-measure which is more specific to the information retrieval than the topic segmentation.

In this paper, we dedicated our research to the topic segmentation on Arabic text corpus. Therefore, this paper is organized as follows: Section 2 presents the adaptation of C99<sup>2</sup> and TextTiling<sup>1</sup> to the Arabic language; Section 3 describes our Arabic test corpus; Section 4 deals with the evaluation of the proposed segmenters; and finally, section 5 is dedicated to the conclusion and our future work.

## 2. Adaptation to Arabic language

Currently, the most known algorithms for topic segmentation are C99<sup>2</sup> and TextTiling<sup>1</sup>. These two segmenters are dedicated to the English language. So, we decide to adapt them to the Arabic language. Therefore, in this section, two text segmentation systems are presented, namely ArabC99 and ArabTextTiling and they are based respectively on C99<sup>2</sup> algorithm and TextTiling<sup>2</sup> algorithm.

### 2.1. ArabC99

ArabC99 is a topic segmenter dedicated for the Arabic language. It is based on lexical cohesion and it goes through two important steps: pre-processing and segmentation. The pre-processing step includes the following operations: words extraction, stop words elimination and stemming. Indeed, ArabC99 extracts all the words from the input text. Then, the not useful words are eliminated by using a list of stop word specific to the Arabic language. Next, a stemming program is applied in order to provide the root of each useful word. For that, we use the Shereen Khoja's stemmer<sup>9</sup> which is used to Arabic language. In fact, Khoja's stemmer<sup>9</sup> removes the longest suffix and the longest prefix. Then it matches the remaining word with verbal and nouns patterns, to extract the root. The stemmer makes

use of several linguistic data files such as a list of all diacritic characters, punctuation characters, definite articles and 168 stop words. To better explain the pre-processing step, we present the following example. As shown in Fig. 1, the three operations have been applied to one arabic sentence along with a translation to the English language.

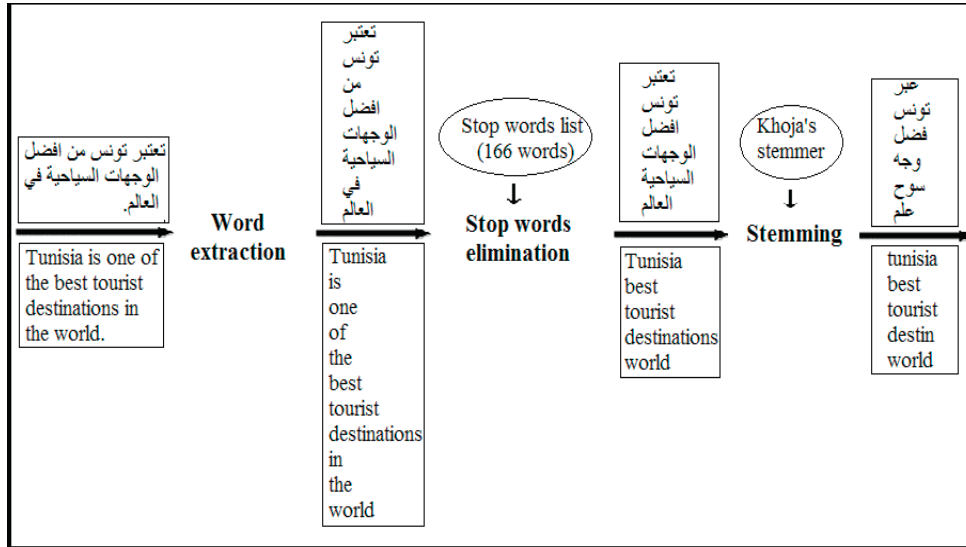


Fig. 1. The pre-processing step.

The second step is the segmentation which is the same as C99 algorithm<sup>2</sup>. As shown in Fig. 2, this step is divided in four phases. The first phase is the construction of the frequency dictionary. In fact, each sentence is represented by a vector which contains the frequency of each word. The second phase is the similarity matrix construction. In fact, the similarity between a pair of sentences is computed using the cosine measure as shown in equation 1. Once the similarity matrix is constructed, the rank matrix will be calculated. Each value in the similarity matrix is replaced by its rank in the local region. As shown in equation 2, the rank is the number of neighboring elements with a lower similarity value. Finally, the fourth phase is dedicated to identify the topic boundaries by using Reynar's maximisation algorithm<sup>10</sup>.

$$Sim(x, y) = \frac{\sum f_{x,j} f_{y,j}}{\sqrt{\sum f_{x,j}^2 \sum f_{y,j}^2}} \quad (1)$$

With  $f_{x,j}$  denote the frequency of word  $j$  in sentence  $x$  and  $f_{y,j}$  denote the frequency of word  $j$  in sentence  $y$ .

$$Rank(x, y) = \frac{\text{Number of elements with a lower similarity value compared to } sim(x, y)}{\text{Number of elements examined}} \quad (2)$$

## 2.2. ArabTextTiling

ArabTextTiling is a topic segmentation algorithm for Arabic language. Like ArabC99, this algorithm contains two important steps: pre-processing and segmentation. The pre-processing step consists of the succession of the following operations: words extraction, stop words elimination and stemming. For the stemming phase, we also use Khoja's stemmer<sup>9</sup>. In fact, the pre-processing step of ArabTextTiling is the same as the pre-processing step of ArabC99 (Fig. 1). The second step is the same as TextTiling<sup>1</sup> and it includes three phases (Fig. 3).

The first one is the blocks construction by using a sliding window. Then, the similarity between blocks is calculated by the cosine measure as in equation 1. This formula yields a score between 0 and 1, inclusive. These scores will be

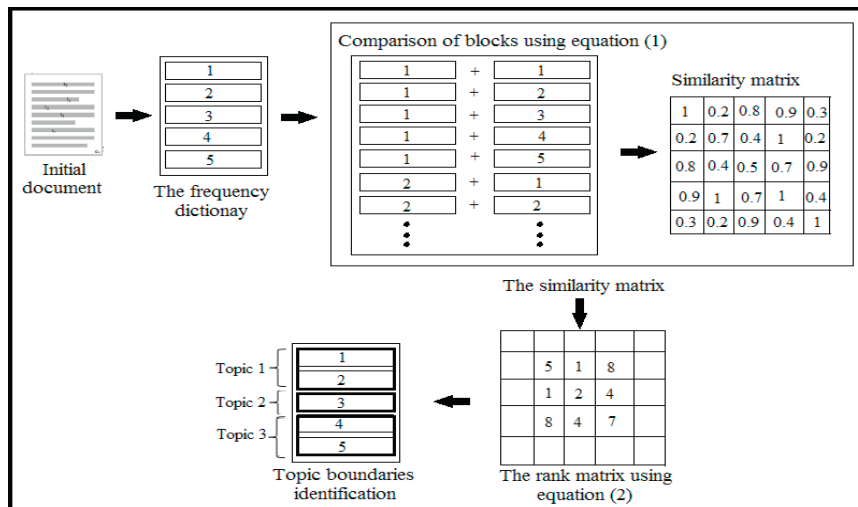


Fig. 2. The second step of ArabC99

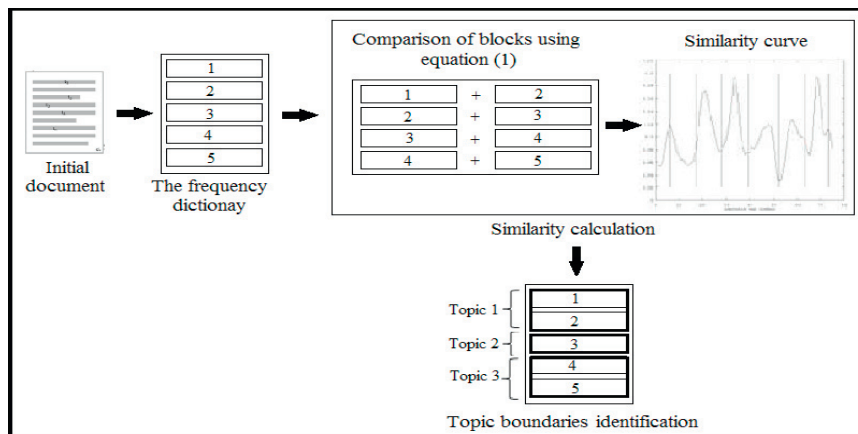


Fig. 3. The second step of ArabTextTiling.

plotted as shown in Fig. 3. Finally, boundaries are determined by changes in the sequence of similarity scores. These changes are detected by recording the higher peaks in the similarity curve.

### 2.3. Example of topic segmentation

To better understand the concept of topic segmentation, we present the following example. In this example, we have performed ArabTextTiling on an Arabic document. This document is constructed from a concatenation of two articles from the Algerian newspaper Al-Chourouk<sup>11</sup>. The first paper is about the protest of higher school teachers to condemn the phenomenon of school violence in Algeria. The second paper is about the organization of the General Assembly of the federal electoral Algerian Football. As shown in Fig. 4, our segmenter identified correctly and automatically the two parts of the document.

نظم، أسس أساتذة الثانويات بالجزائر العاصمة، وقفة احتجاجية لمدة ساعة، تنديدا بظاهرة العنف المدرسي التي تفتشت بشكل جد ملفت للانتباه. بحيث عبروا عن سخطهم واستيائهم تجاه التلاميذ الذين أصبحوا يمارسون العنف ضد أساتذتهم من دون أي خوف ونظرا إلى انتشار ظاهرة العنف الممارس من طرف التلاميذ ضد الأساتذة، قرر أساتذة الثانويات بالجزائر وسط، شرق، غرب، تنظيم وقفة احتجاجية لمدة ساعة بداخل مؤسساتهم التربوية، بغية إسماع أصواتهم للوصاية، لكي تتدخل في الوقت المناسب لإيجاد الحلول التي من شأنها التقليل - على الأقل - من الظاهرة التي تفتشت بشكل رهيب وملفت للانتباه، بحيث أصبح الأستاذ مهددا بشكل يومي من اعتداءات التلاميذ ضدهم وكانت نقابات التربية المستقلة، وعلى رأسها النقابة الوطنية لعمال التربية والتكوين، قد أشارت في الدراسة التي أعدتها، أن الأساتذة صاروا هم ضحية اعتداءات التلاميذ داخل المؤسسات التربوية، خاصة فئة الأساتذات منهن وبالطور الثانوي، وليس العكس، خاصة في ظل غياب مواد قانونية تحصن الأستاذ من الاعتداءات الجسدية التي تواجهه أثناء ممارسة وظيفته بالمقابل، فقد حرصت الوصاية على إصدار مناشير وتعليمات من شأنها منع العقاب البدني واللفظي ضدهم. في الوقت الذي شددت بأن الأساتذات بالطور الثانوي هن أكثر عرضة للاعتداءات الجسدية واللفظية كما بينت نفس الدراسة، أن التلميذ محمي من اعتداءات الأستاذ، لكن بالمقابل لا توجد حماية للعنف الذي يمارسه التلميذ ضد أستاذه بالقسم، وعليه في حال اعتداء التلميذ على أستاذه داخل القسم فإن المعتدي بحال على مجلس التأديب، وغالبا ما تكون الأحكام الصادرة في حق التلميذ "رمزية" لا تتعدى التوبيخ البسيط أو الملاحظة الشفهية وأقصاها النقل إلى مؤسسة أخرى. ومن ثمة فإن هذا النوع من العقوبات يدفع بالتلاميذ إلى التفكير في الانتقام من الأستاذ مرارا إلى غاية تنفيذه، خاصة في ظل استمرار غياب أطر حماية الأستاذ بكل أشكالها.

اجتمعت الإثنين لجنة الترشيحات المكلفة بتنظيم الجمعية العامة الانتخابية للاتحادية الجزائرية لكرة القدم، المقررة يوم 7 مارس المقبل بوهان، بمقر الفاف بدالي إبراهيم تحت إشراف رئيسها أحمد ميراك الذي تم تنصيبه على هامش أشغال الجمعية العامة العادية، المنعقدة نهاية الأسبوع الفارط بوهان وخصصت هذه الجلسة لدراسة ملفات المترشحين للرئاسة، سيما وأن المهلة القانونية التي كانت الفاف قد حددتها قد انقضت الخميس المنصرم، وروراوة كان الوحيد الذي أودع ملف الترشيح رسميا لدى الأمانة العامة للاتحادية، وحسب مصدر عليم فإن اللجنة التي تتشكل من 7 أعضاء، منهم 4 رؤساء رابطات، وممثلي نوادي شباب قسنطينة، أولمبي المنية ووافق سطيف، سترس ملفات الأعضاء الذين أدرجهم روراوة ضمن قائمته للظفر بصفة العضوية في المكتب الفيدرالي، لينتبع بعدها إشعار المترشح روراوة بالقرارات المتخذة بخصوص كل عنصر، علما وأن قرارات لجنة الترشيحات قابلة للطعن، لتعقد جلسة ثانية للنظر في الطعون، وسيتم إثرها الإعلان عن القائمة الرسمية التي سيدخل بها روراوة الجمعية العامة الانتخابية لخلافة نفسه للعهدة الثانية على التوالي والثالثة له، رغم أن الأمر محسوم بصفته المترشح الوحيد وكشف مصدرنا بأن روراوة شطب 3 أسماء رسميا من قائمته، ويتعلق الأمر بكل من الحكم الدولي السابق إبراهيم الجزار، اللاعب الدولي السابق علي فرقاني والدكتور ياسين زرقيني، حيث قرر ضم رئيس اللجنة الطبية الدكتور علي بقدح، في انتظار الكشف عن آخرين لاستكمال القائمة.

Fig. 4. The result of ArabTextTiling

### 3. Arabic test corpus

We have used Arabic newspapers: the Arabic version of The Diplomatic World<sup>12</sup> and Tunisian<sup>13</sup>, Egyptian<sup>14</sup> and Algerian<sup>11</sup> versions of a newspaper named Al-Chourouk. Indeed, we have collected 120 articles dealing with various topics such as politics, sports, culture, history, technology and arts. Then these articles are divided into four collections of documents. The first collection contains 35 articles from the Tunisian newspaper. The second collection includes 35 articles from the Egyptian newspaper. The third collection consists of 30 articles from the Algerian newspaper. Finally, the fourth collection is made up of 20 articles from The Diplomatic World newspaper.

The originality of our research is presented by the fact that we used an important amount of data (120 raw texts) unlike the works of Tourir et al.<sup>7</sup> (10 texts) and Harrag et al.<sup>8</sup> (5 texts). Also, the aim of our work is not only the identification of boundaries between concatenated texts like the most of research which has presented until now. Indeed, we have constructed two sets. For the first set we have combined four documents in series in order to identify boundaries between texts. While for the second set and as shown in Fig. 5, we have combined four documents by section and in turn.

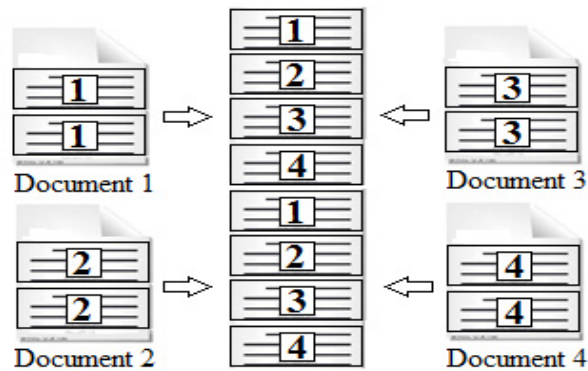


Fig. 5. The construction of the second set.

#### 4. Evaluation

Unlike the inefficient metrics which have been used in the works of Bants et al.<sup>5</sup> and Touri et al.<sup>7</sup>, we used the two metrics F-measure and WindowDiff for the evaluation. Indeed, the WindowDiff metric is a variant of the Pk measure which uses a sliding window. This window allows to take into account the number of boundaries between two sentences remote with a distance  $k$  ( $k$  is the length of the sliding window). In the other hand, the F-measure metric has been seeing as a compromise between the two major standards for the evaluation which are Recall and Precision. So, with the use of F-measure and WindowDiff, we have evaluated our new topic segmenters for the two test sets which have been described in the previous section.

##### 4.1. Evaluation results using the first set:

In this subsection, we present the results of each algorithm (ArabC99 and ArabTextTiling) for the first set with the use of curves and summary tables to properly analyze the results.

As we said in section 3, there are four collections of documents which are well identified in the following curves. In fact, the first collection which is related to the Tunisian newspaper<sup>13</sup> is limited by the origin of the curve and point A. The second collection of article which is from the Egyptian newspaper<sup>14</sup> is limited by the two points A and B. The third collection which contains articles from the Algerian newspaper<sup>11</sup> is limited by B and C. Finally, the fourth collection which is related to the Arabic version of The Diplomatic World<sup>12</sup> is limited by C and D.

Fig. 6 shows the F-measure values as a function of 120 documents. The ArabC99 performance is better than ArabTextTiling performance especially for the second and fourth collection. For the first collection, the performance of the two segmenters gets close.

In table 1, the average F-measure results of ArabC99 and ArabTextTiling is shown. The results show that ArabC99 outperforms ArabTextTiling in all the four collections of documents. Also, for the fourth collection, F-measure values of the two segmenters decreases relative to the other collections.

Table 1. Average F-measure results for the first set.

Segmentation	First collection	Second collection	Third collection	Fourth collection
ArabTextTiling	0.366	0.173	0.422	0.094
ArabC99	0.523	0.559	0.803	0.395

Fig. 7 shows the WindowDiff values as a function of the four collections of documents. Largely, the performance of ArabC99 is better than the performance of ArabTextTiling. Indeed, the error rate of ArabC99 is less than the error



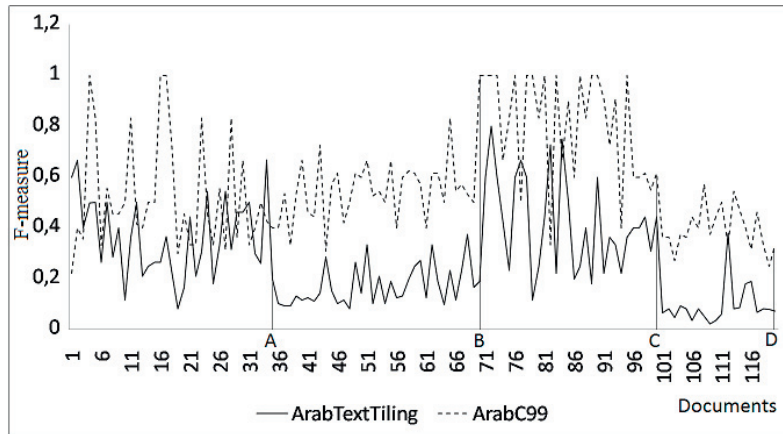


Fig. 6. F-measure results for the first set.

rate of ArabTextTiling especially for the two last collections.

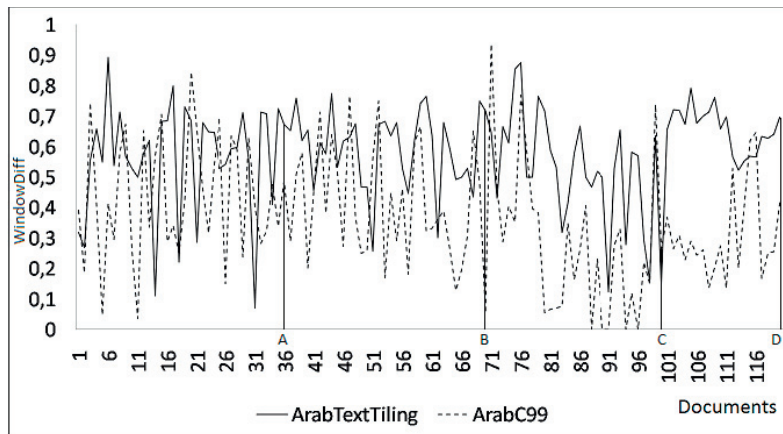


Fig. 7. WindowDiff results for the first set.

The table 2 contains the average WindowDiff results of ArabC99 and ArabTextTiling for the first set. The results show that ArabC99 has the smallest error rate compared to ArabTextTiling especially for the two last collections.

Table 2. Average WindowDiff results for the first set.

Segmentation	First collection	Second collection	Third collection	Fourth collection
ArabTextTiling	0.561	0.596	0.521	0.658
ArabC99	0.434	0.409	0.279	0.311

#### 4.2. Evaluation results using the second set:

In this subsection, we present the results of ArabC99 and ArabTextTiling for the second set using the two metrics F-measure and WindowDiff.

Fig. 8 shows the F-measure results as a function of the 120 documents for our two topic segmenters. For the third and fourth collection, the performance of ArabC99 is always better than the performance of ArabTextTiling.

Yet, for the first and second collection, the performance of the two segmenters gets close and for some documents ArabTextTiling outperforms ArabC99.

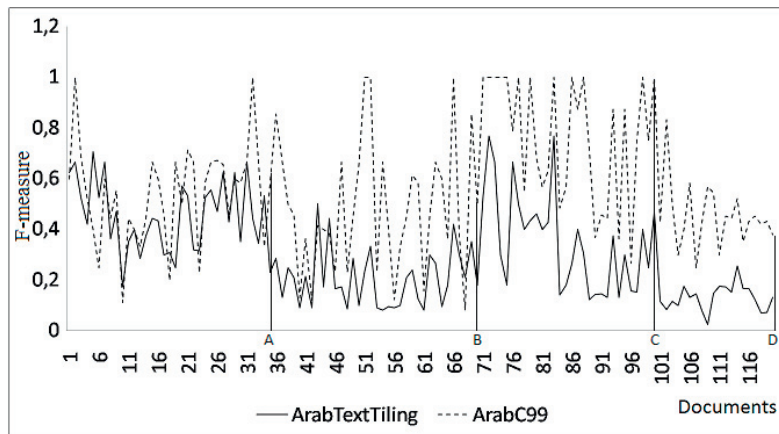


Fig. 8. F-measure results for the second set.

The table 3 contains the average F-measure values. As shown in this table, we notice that the performance ArabC99 is better than the performance of ArabTextTiling. We also notice that the F-measure values of ArabTextTiling decreases for the last collection.

Table 3. Average F-measure results for the second set.

Segmentation	First collection	Second collection	Third collection	Fourth collection
ArabTextTiling	0.452	0.205	0.353	0.131
ArabC99	0.542	0.486	0.767	0.452

Fig. 9 shows the WindowDiff values as a function of the four collections of documents. For the two first collections, the performance of ArabC99 and ArabTextTiling are almost close. In fact, ArabC99 outperforms ArabTextTiling for some documents and the opposite is also true. But, for the last two collections, the performance of ArabC99 is better than the performance of ArabTextTiling especially for the fourth collection.

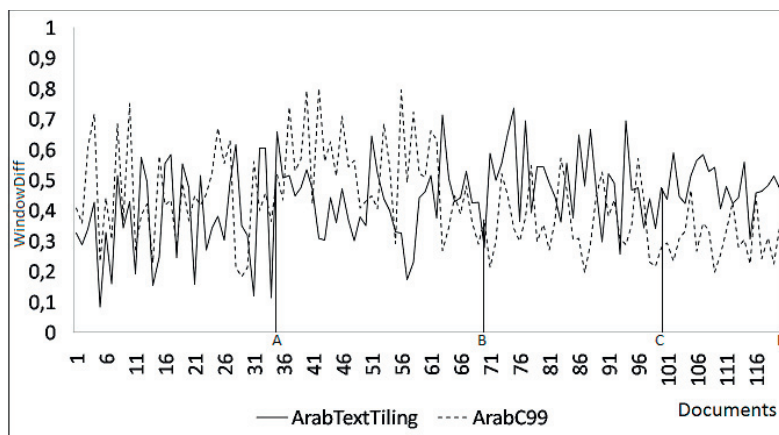


Fig. 9. WindowDiff results for the second set.



The table 4 presents the segmentation performs of ArabC99 and ArabTextTiling using the WindowDiff metric. For the two first collections, ArabTextTiling outperforms ArabC99 with the smallest error rates. While for the last two collection, the performance of ArabC99 is better than the performance of ArabTextTiling.

Table 4. Average WindowDiff results for the second set.

Segmentation	First collection	Second collection	Third collection	Fourth collection
ArabTextTiling	0.451	0.509	0.594	0.578
ArabC99	0.526	0.628	0.438	0.372

### 4.3. Discussion

In general, ArabC99 outperforms ArabTextTiling for the two sets. However, the performance of each segmenter is not the same for the four collections of documents. This difference can be explained by the fact that the four collections are from different sources (Tunisian, Egyptian and Algerian). In fact, these countries use Arabic language but each one of them is characterized by its own style. Also, for the fourth collection that treats only one topic which is politic, the performance of ArabTextTiling decreases and ArabC99 remains the best among the two of them. Furthermore, this collection is also characterized with length of its documents.

Moreover, if we compare our work with the work of Harrag at al.<sup>8</sup>, we can say that our results are different. As shown in table 5, Harrag et al.<sup>8</sup> confirmed that ArabTiling which is the Arabic version of TextTiling is better than TopSegArab which is the Arabic version of C99. Indeed, compared to TopSegArab, ArabTiling has the best values of Precision (0.81) and F-measure (0.65). While, in our work we found that ArabC99 outperforms ArabTextTiling for the two test sets. In fact the evaluation of Harrag at al.<sup>8</sup> is not reliable because they have only used five texts as a data set which is very little for a trustworthy evaluation. Also this difference is caused by the use of a different stemmer which requires a further study and the use of a different data source.

Table 5. Comparison of the algorithms with an Arabic corpus<sup>8</sup>.

Segmentation	Recall	Precision	F-measure
Humain judges 0.81	0.84	0.82	
ArabTiling	0.55	0.81	0.65
TopSegArab	0.54	0.64	0.58

## 5. Conclusion

In this paper, we proposed an adaptation of two topic segmenters (C99 and TextTiling) for textual document written in Arabic language. The Arabic segmenters ArabC99 and ArabTextTiling use Khoja's stemmer. We evaluate these segmenters with two test sets based on newspapers of different Arab countries. The originality of this work is that the evaluation is not limited to the detection of boundary between texts. We notice too that for Arabic language, the difference between the dialects of countries can be important in the topic segmenter process. In addition to that we remark that the adaptation of topic segmenter depends of the choice of the Arabic stemmer and we must study this effect in depth.

From this work several future works can be considered. As a first work, we must evaluate our new segmenters using an Arabic benchmark such as the corpus of Latifa<sup>15</sup> which includes 842684 words and 415 texts. We must also study the integration of Latent Semantic Analysis (LSA) and Probabilistic Latent Semantic Analysis (PLSA) in the topic segmentation. To go further, we should propose a new multilingual hybrid topic segmenter which uses internal and external resources.

## References

1. M. A. Hearst. 1997. TextTiling: "Segmenting text into multi-paragraph subtopic passages". Computational Linguistics, 23 (1) :33-64.

2. F. Y. Y. Choi. "Advances in domain independent linear text segmentation". z. Proceeding of *NAACL-00*, pp 26-33, 2000.
3. F. Y. Y. Choi, P. Wiemer-Hastings, and J. Moore. "Latent Semantic Analysis for Text Segmentation". Proceedings of *EMNLP*, 2001, pp. 109-117.
4. Olivier Ferret. "Improving text segmentation by combining endogenous and exogenous methods". *TAL*. Volume 47 - n2/2006, pp 111-135.
5. T. Brants, F. Chen, and A. Farahat. "Arabic Document Topic Analysis". *TREC 2002*. Gaithersburg: NIST, 2002.
6. M. A. El-Shayeb, S. R. El-Beltagy, and A. Rafea. "Comparative Analysis of Different Text Segmentation Algorithms on Arabic News Stories". In Proc. *IEEE International Conference on Information Reuse and Integration*, 2007, pp. 441-446.
7. A. A. Touir, H. Makhtour, and W. Al-Sanea. "Semantic-Based Segmentation of Arabic Texts". *Inf. Tech. J.*, 7(7)(2008), pp. 1009-1015.
8. F. Harrag, A. H. Cherif, A. S. Al-Salman. "Comparative Study of Topic Segmentation Algorithms Based On Lexical Cohesion: Experimental Results on Arabic Language". *The Arabian Journal for Science and Engineering*, Volume 35, Number 2C, 2010.
9. S. Khoja, "Stemming Arabic Text", <http://zeus.cs.pacificu.edu/shereen/research.htm>. (Accessed April 03, 2013)
10. J. Reynar. "*Topic Segmentation: Algorithms and Application*". Ph.D. thesis. Computer and Information Science. University of Pennsylvania, Pennsylvania, USA, 1998.
11. Echourouk-algerie. <http://www.okbob.net/article-lire-le-journal-echourouk-algerie-75761076.html>. (Accessed April 17, 2013)
12. Le Monde Diplomatique. <http://www.mondiploar.com/index.php3>. (Accessed April 26, 2013)
13. Pressetunisie. <http://www.presetunisie.net/alchourouk.php>. (Accessed April 14, 2013)
14. Shorouknews. <http://shorouknews.com/egypt>. (Accessed April 16, 2013)
15. Latifa Al-Sulaiti's Homepage. <http://www.comp.leeds.ac.uk/eric/latifa/research.htm>. (Accessed March 17, 2014)