Proceedings of the Computational
Linguistics-Applications Conference pp. 53–60 ISBN 978-83-60810-47-7

MUSE – A Multilingual Sentence Extractor

Marina Litvak†, Mark Last*, Menahem Friedman* and Slava Kisilevich‡
†Sami Shamoon Academic College of Engineering
Beer-Sheva, Israel
Email: marinal@sce.ac.il
*Ben Gurion University of Negev
Beer Sheva, Israel
Email: {litvakm, mlast, fmenahem}@bgu.ac.il
‡University of Konstantz
Konstantz, Germany
Email: slaks@dbvis.inf.uni-konstanz.de

***Abstract*—MUltilingual Sentence Extractor (MUSE) is aimed at multilingual single-document summarization. MUSE implements the supervised language-independent summarization approach based on optimization of multiple statistical sentence ranking methods. The MUSE tool consists of two main modules: the training module activated in the offline mode, and the on-line summarization module. The training module can be provided with a corpus of summarized texts in any language. Then, it learns the best linear combination of user specified sentence ranking measures applying a Genetic Algorithm to the given training data. The summarization module performs real-time sentence extraction by computing sentence rankings according to the weighted model induced in the training phase. The main advantage of MUSE is its language-independency – it can be applied to any language given a gold standard summaries in that language. The performance of MUSE in our previous works was found to be significantly better than the best known state-of-the-art extractive summarization approaches and tools in the three different languages: English, Hebrew, and Arabic. Moreover, our experimental results in a cross-lingual domain suggest that MUSE does not need to be retrained on each new language, and the same weighting model can be used across several languages.**

***Index Terms*—automated summarization, multi-lingual summarization, language-independent summarization, genetic algorithm, optimization, linear combination**

I. Introduction

DOCUMENT tools should identify a minimum number of words and/or sentences to express a document's main ideas. In this way, high quality summarization tools can significantly reduce the information overload faced daily by many professionals in a variety of fields.

The publication of information on the Internet in an ever-increasing variety of languages amplifies the importance of developing multilingual summarization tools. There is a particular need for language-independent statistical tools that can readily be applied to text in any language without depending on language-specific linguistic analysis. In the absence of such tools, the only alternative to language-independent summarization is the labor-intensive translation of the entire document into a common language.

Since a pure statistical method usually characterizes only one sentence feature, various attempts were made to use a combination of several methods as a ranking function [1], [2]. MUSE continues this effort by learning the best linear combination of 31 statistical language-independent sentence ranking features using a Genetic Algorithm (GA). With this approach, MUSE can be easily applied to multilingual extractive summarization. All sentence features comprising the linear combination are based on either a vector or a graph representation using a mere word and sentence segmentation of a document.

MUSE implements the multilingual summarization[1] approach introduced in [4].[2] Evaluation of MUSE on three monolingual (English, Hebrew and Arabic) and one bilingual corpora of English and Hebrew documents has shown the following:
- MUSE performance is significantly better than TextRank [5] and Microsoft Word's Autosummarize tool in all tested languages, as demonstrated in Table II.
- In English, MUSE outperforms such known summarization tools as MEAD [1] and SUMMA [2].
- MUSE does not need to be retrained on each language and the same model can be used across at least three–English, Hebrew, and Arabic–different languages.

Table I demonstrates the results of training and testing comprising the average ROUGE values obtained for English[3], Hebrew[4] and bilingual corpora using 10-fold cross validation and reported in [4].

Table II shows the comparative results, also reported in [4], (ROUGE mean values) for three corpora, with the best summarizers on top. Results contain comparisons between: (1) a multilingual version of TextRank (denoted by

[1]Multilingual summarization is defined by [3] as "processing several languages, with summary in the same language as input"

[2]A web application of the MUSE-based summarizer will soon be made available at http://www.cs.bgu.ac.il/∼litvakm/

[3]We used the corpus of summarized documents available at the Document Understanding Conference, 2002 [6] for English. This benchmark dataset contains 533 news articles, each accompanied by two to three human-generated abstracts of approximately 100 words each.

[4]For the Hebrew language we generated a corpus of 50 summarized news articles of 250 to 830 words each from the Website of the *Haaretz* newspaper (http://www.haaretz.co.il)

ML_TR) (Mihalcea, 2005), (2) Microsoft Word's Autosummarize function (denoted by MS_SUM), and (3) the best single scoring method in each corpus. As a baseline, we compiled summaries created from the initial sentences (denoted by POS_F). MUSE performed significantly better than TextRank in all three corpora and better than the best single methods COV_DEG in English and D_COV_J in Hebrew corpora respectively.

TABLE I
RESULTS OF 10-FOLD CROSS VALIDATION (LITVAK ET AL., 2010B)

	ENG	HEB	MULT
Train	0.4483	0.5993	0.5205
Test	0.4461	0.5936	0.5027

TABLE II
SUMMARIZATION PERFORMANCE. MEAN ROUGE-1 (LITVAK ET AL., 2010B)

Metric	ENG	HEB	MULT
MUSE	**0.4461**	**0.5921**	**0.4633**
COV_DEG	0.4363	0.5679	0.4588
D_COV_J	0.4251	0.5748	0.4512
POS_F	0.4190	0.5678	0.4440
ML_TR	0.4138	0.5190	0.4288
MS_SUM	0.3097	0.4114	0.3184

II. MULTILINGUAL SENTENCE EXTRACTOR (MUSE): OVERVIEW

A. Methodology

MUSE implements a *supervised* learning approach to language-independent extractive summarization where the best set of weights for a linear combination of sentence scoring methods is found by a genetic algorithm trained on a collection of document summaries. The weighting vector thus obtained is used for sentence scoring in future summarizations. Since most sentence scoring methods have a linear computational complexity, only the training phase of our approach is time-consuming.

Using MUSE, the user can choose the subset of totally 31 sentence metrics that will be included in the linear combination. All metrics are based on different text representation models and are language-independent since they do not rely on any language-specific knowledge. Figure 1 demonstrates the taxonomy of all 31 metrics. We divided them into three main categories—*structure*-, *vector*-, and *graph*-based—according to their text representation model, where each sub-category contains group of metrics using the same scoring method.

A detailed description of sentence metrics used by MUSE can be found in (Litvak et al., 2010b).

We found the best linear combination of the metrics depicted in Figure 1 using a Genetic Algorithm (GA). GAs are categorized as global search heuristics. Figure 2 shows a simplified GA flowchart. A typical genetic algorithm requires (1) a genetic representation of the solution domain, (2) a fitness function to evaluate the solution domain, and (3) some basic parameter settings like selection and reproduction rules.

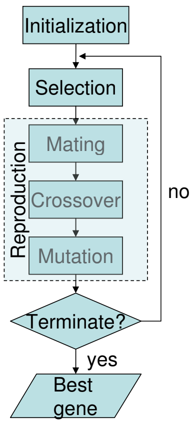


Fig. 2. Simplified flowchart of a Genetic Algorithm (Litvak et al., 2010b)

We represent each solution as a vector of weights for a linear combination of sentence scoring metrics—real-valued numbers in the unlimited range normalized in such a way that they sum up to 1. The vector size is fixed and it equals the number of metrics used in the combination.

Defined over the genetic representation, the fitness function measures the quality of the represented solution. We use ROUGE-1 and ROUGE-2, Recall (Lin & Hovy, 2003) as a fitness functions for measuring summarization quality—similarity with gold standard summaries, which should be *maximized* during the training (optimization procedure). We used annotated corpus of summarized documents where each document is accompanied by several human-generated summaries—abstracts or extracts as a training set.

The reader is referred to (Litvak et al., 2010b) for a detailed description of the optimization procedure implemented by MUSE.

Algorithms 1 and 2 contain the pseudo-code for two independent phases of MUSE: training and summarization, respectively. Assuming efficient implementation, all metrics have a linear computational complexity relative to the total number of words in a document - $O(n)$. As a result, summarization computation time, given a trained model, is also linear (at factor of the number of metrics in a combination). The training time is proportional to the number of GA iterations multiplied by the number of individuals in a population times the fitness evaluation (ROUGE) time. On average, in our experiments the GA performed $5-6$ iterations—selection and reproduction—before reaching convergence.

B. Architecture

The current version of MUSE tool can be applied only to text documents or textual content of the HTML pages. It consists of two main modules: the *training module* activated in offline, and the real-time *summarization module*. Both modules utilize two different representations of documents described in (Litvak et al., 2010b): vector- and graph-based.

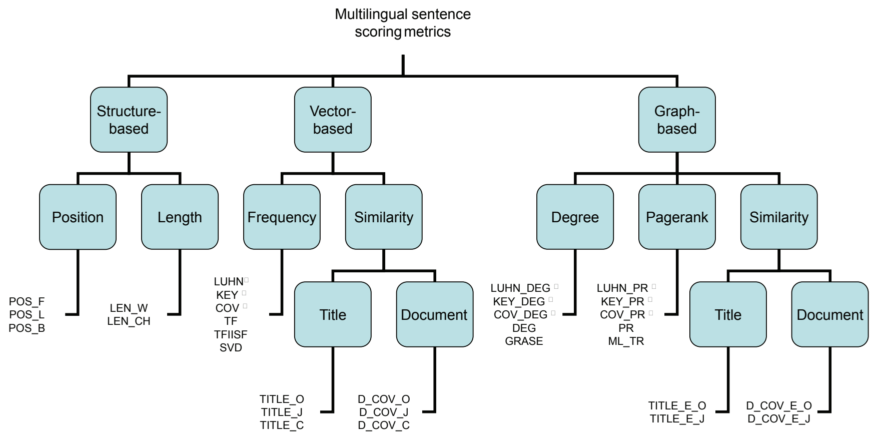


Fig. 1. Taxonomy of language-independent sentence scoring metrics (Litvak et al., 2010b)

Algorithm 1 Step 1: Training

Input: Gold Standard - a corpus of summarized documents D, N chosen metrics
Output: A weighted model W - vector of weights for each of N metrics
  Step 1.1: Compute M **- sentence-score matrix**
  for all $d \in D$ **do**
    Let R_1, R_2, and R_3 are d representations
    for all sentences $s \in d$ **do**
      Calculate N metrics using R_1, R_2, and R_3
      Add metrics row for s into M
    end for
  end for
  Step 1.2: Compute a vector W **of metrics weights**
  Run a Genetic Algorithm on M, given D:
  Initialize a population P
  repeat
    for all solution $g \in P$ **do**
      Generate a summary a
      Evaluate a by ROUGE on summaries of D
    end for
    Select the best solutions G
    P - a new population generated by G
  until convergence - no better solutions are found
  return a vector W of weights - output of a GA

Algorithm 2 Step 2: Summarizing a new document

Input: A document d, maximal summary length L, a trained weighted model W
Output: A set of n sentences, which were top-ranked by the algorithm as the most important.
  Step 2.1: Compute a score of each sentence
  Let R_1, R_2, and R_3 are d representations
  for all sentense $s \in d$ **do**
    Calculate N metrics using R_1, R_2, and R_3
    Calculate a score as a linear combination according to W
  end for
  Step 2.2: Compile the document summary
  Let $S = \emptyset$ be a summary of d
  repeat
    get the top ranked sentence s_i
    $S = S \bigcup s_i$
  until S exceeds max length L
  return S

The *preprocessing module* is responsible for constructing each representation, and it is embedded in both modules.

The *training module* receives as input a corpus of documents, each accompanied by one or several gold-standard summaries—abstracts or extracts—compiled by human assessors. The set of documents may be either monolingual or multilingual and their summaries have to be in the same language as the original text. The *training module* applies a genetic algorithm to a document-feature matrix of precomputed sentence scores with the purpose of finding the best linear combination of features using any ROUGE metric (ROUGE-1 Recall as a default or specified by end-user) as a fitness function. The output/model of the training module is a vector of weights for user-specified sentence ranking features. In the current version of the tool, the user can choose from 31 vector-based and graph-based features. The recommendation for the best 10 features one can find in (Litvak et al., 2010a).

The *summarization module* performs summarization of input text/texts in real time. Each sentence of an input text obtains a relevance score according to the trained model, and the top ranked sentences are extracted to the summary in their original order. The length of resulting summaries is limited by a user-specified value (maximum number of words in the text extract or a ratio). Being activated in real-time, the *summarization module* is expected to use the model trained on the same language as input texts. However, if such model is not available (no annotated corpus in the text language), the user can choose the following: (1) the model trained on some other language/corpus can be used (in (Litvak et al., 2010b) we show that the same model can be efficiently used across different languages), or (2) user-specified weights for each sentence feature (from 31 provided in the system) in the linear combination can be used for summarization.

The *preprocessing module* performs the following tasks: (1) sentence segmentation, (2) word segmentation, (3) vector space model construction using *tf* and/or *tf-idf* weights, (4) a word-based graph representation construction, (5) a sentence-based graph representation construction, and (6) document metadata construction, including such information like frequency (tf and tf-idf) for each unique term, its location inside the document, etc. The outputs of this submodule are: sentence segmented text (SST), vector space model (VSM), the document graphs, and the metadata stored in the xml files. Steps (1) and (2) are performed by the *text processor submodule*, which is implemented using *Strategy Design Pattern* (Freeman et al., 2004) and consists of three elements: *filter, reader* and *sentence segmenter*. The *filter* works on the Unicode character level and performs such operations like identification of characters, digits, punctuations and normalization (optional for some languages). The *reader* invokes the *filter*, constructs word chunks from the input stream and identifies the following states: *words, special characters, white spaces, numbers, URL links* and *punctuation marks*. The *sentence segmenter* invokes *reader* and divides the input space into sentences. By implementing different filters, the reader can work either with a specific language (taking into account its intricacies) or with documents written in arbitrary language. Figure 4 presents a small text example and its word-based graph representation.

Figure 3 shows the general architecture of the MUSE system.

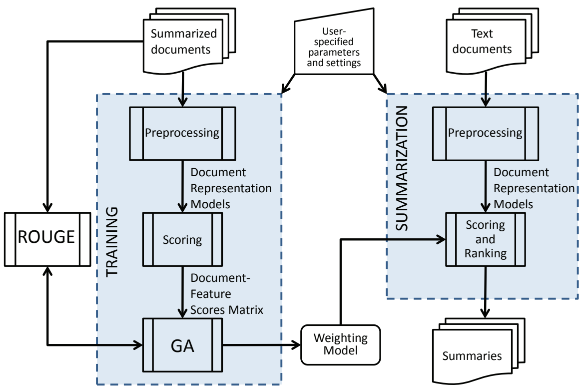


Fig. 3. MUSE architecture

C. Use Case

MUSE has five possible use cases demonstrated in Figure 9 and briefly described in Table III: Configure, Train, Summarize, Rouge, and Evaluate.

In the **Configure** use case, the user is required to specify the following parameters: folder paths to input documents being summarized, gold standard, output summaries, maximal length of a summary, preprocessing settings. The advanced user can change the default settings for a GA: population size, crossover and mutation probabilities, etc., training settings: splitting to the training and test data, preprocessing settings: maximal size of a graph representation, etc.. All specified settings can be used in the next user actions aka training and/or summarization or stored for the later use.

0	Hurricane Gilbert Heads Toward Dominican Coast.
1	Hurricane Gilbert swept toward the Dominican Republic Sunday, and the Civil Defense alerted its heavily populated south coast to prepare for high winds, heavy rains and high seas.
2	The storm was approaching from the southeast with sustained winds of 75 mph gusting to 92 mph.

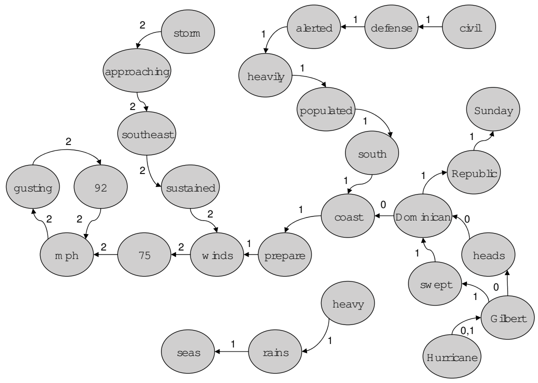


Fig. 4. Text document (top) and its graph representation (bottom)

In the **Train** use case, the user trains genetic algorithm on the training corpus of annotated documents, where the optimal weights for the linear combination of sentence scoring metrics are determined. The weighted model can be used in the subsequent **Summarize** use case or stored for the later use.

The MUSE system can be evaluated on a new annotated dataset in the **Evaluate** use case.

In the **Summarize** use case, the user can get a summary for a single input document or set of summaries for a set of input documents. The pre-specified, new or modified settings can be used. The pre-trained or a new trained weighted model can be used.

In the fifth, **Rouge** use case, the user can apply ROUGE evaluation toolkit on existing—just produced or stored in advance—summaries. The ROUGE-1 Recall metric is applied by default, but it can be changed in the configuration use case.

D. Features

MUSE software has the following unique features:

- **Multilingual summarization.** The way in which MUSE processes a text is fully *multilingual*. All statistical metrics for sentence ranking used by MUSE do not require any language-specific analysis or knowledge, that allow MUSE to process texts in any language. Figures 6, 7, and 8 demonstrate the documents and their summaries—in a source language and translated to English—generated

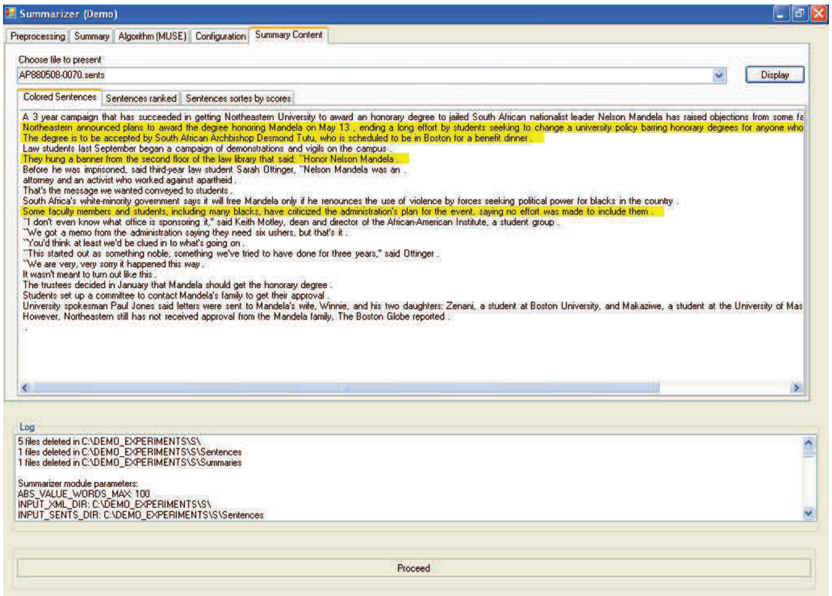


Fig. 5. Input file (AP880228-0097 from DUC 2002 collection) and its extract by MUSE.

أميركا: خطوة غير مسبوقة في□الدولي»

واشنطن- رويترز ـ اتخذت الولايات المتحدة خطوات غير مسبوقة في صندوق النقد الدولي في مسعى لإجبار أوروبا على التخلي عن بعض نفوذها في المجلس التنفيذي للصندوق لصالح الاقتصادات الناشئة. وقال دبلوماسيون من عدة دول أعضاء في الصندوق إن الولايات المتحدة التي شعرت بخيبة الأمل لرفض أوروبا اقتسام المزيد من الصلاحيات، رفضت هذا الشهر مساندة قرار يبقي على الهيمنة الاوروبية على المجلس التنفيذي للصندوق الذي يضم 24 عضوا. وفشلت واشنطن في جهود تبذلها منذ فترة طويلة لخفض عدد مقاعد المجلس التنفيذي للصندوق من 24 إلى 20 في اطار صلاحيات أوسع من شأنها أن تمنح القوى الاقتصادية الصاعدة قول أكبر في قرارات الصندوق بصورة تعكس تنامي نفوذها الاقتصادي. ورفضت أوروبا فكرة التنازل عن تسعة مقاعد تشغلهم حاليا في المجلس وعبرت اقتصادات ناشئة مثل تركيا عن اهتمامها بالحصول على مقعد في المجلس التنفيذي للصندوق. وتعكس هيمنة الدول الاوروبية والولايات المتحدة على الصندوق النظام العالمي بعد الحرب العالمية الثانية والذي يواجه تحديات مع صعود دول مثل الصين.

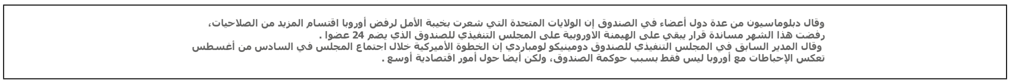
والمجلس التنفيذي أحد أهم أجهزة اتخاذ القرارات في الصندوق فقد وافق على قروض طارئة بمليارات الدولارات لدول هزتها الازمة الاقتصادية العالمية ويشرف على طريقة ادارة الصندوق.

وقال المدير السابق في المجلس التنفيذي للصندوق دومينيكو لومباردي إن الخطوة الأميركية خلال اجتماع المجلس في السادس من أغسطس تعكس الإحباطات مع أوروبا ليس فقط بسبب حوكمة الصندوق، ولكن أيضا حول أمور اقتصادية أوسع. ومن بين هذه القضايا الخلافات حول قواعد السيولة الجديدة للبنوك العالمية واصرار أوروبا على التقشف المالي فيما تؤكد واشنطن على الحاجة الى تأمين الانتعاش الاقتصادي قبل تطبيق سياسات متقشفة. ولم يسبق أن استعرضت الولايات المتحدة عضلاتها بهذه الطريقة العلنية.

وقال رئيس معهد أوكسفورد للسياسات الاقتصادية وزميل معهد بروكينغز في واشنطن لومباردي «انها خطوة عدوانية تولدت عن احساس قوي بالاحباط حيال ما تعتبره الولايات المتحدة عجزا أوروبيا عن تعزيز عملية اصلاح صندوق النقد الدولي». وقال مسؤول كبير في وزارة الخزانة الأميركية إن انتخابات المجلس التنفيذي للصندوق فرصة لبلورة سبل لتعديل تركيبة المجلس وجعله أكثر تمثيلا. وأضاف أن «الوزير تيموثي جايتنر يساند اصلاح المجلس التنفيذي للصندوق كي يعكس بصورة أفضل حقائق الاقتصاد العالمي اليوم ولضمان زيادة تمثيل الاسواق الناشئة والبلدان النامية». ورأى المسؤول انه بعد الخطوة الأميركية الكرة الان في الملعب الاوروبي خلال المناقشات التي يرجح أن يجريها وزراء المالية الاوروبيون خلال اجتماعهم الدوري المقبل.

وأشار مسؤولون أوروبيون إلى رغبتهم في بحث ادخال تغيرات على تمثيل أوروبا في المجلس التنفيذي للصندوق لكن ليس هناك توافق في الآراء حول سبل تنفيذ ذلك. وقال متحدث باسم وزير المالية الالماني فولفجانج شيوبله ان المجلس يجب أن يضم 24 دولة

(a) Summarized document



وقال دبلوماسيون من عدة دول أعضاء في الصندوق إن الولايات المتحدة التي شعرت بخيبة الأمل لرفض أوروبا اقتسام المزيد من الصلاحيات،
رفضت هذا الشهر مساندة قرار يبقي على الهيمنة الاوروبية على المجلس التنفيذي للصندوق الذي يضم 24 عضوا .
وقال المدير السابق في المجلس التنفيذي للصندوق دومينيكو لومباردي إن الخطوة الأميركية خلال اجتماع المجلس في السادس من أغسطس
تعكس الإحباطات مع أوروبا ليس فقط بسبب حوكمة الصندوق، ولكن أيضا حول أمور اقتصادية أوسع .

(b) Original summary

Diplomats from several countries that are members of the IMF said that the United States felt disappointed by the refusal of Europe to share more powers, and this month it refused to support a resolution to maintain the European domination of the IMF Executive Board, which includes 24 members.
The former director of the Executive Board of the Fund Domenico Lombardi said that the American move during the Board meeting on the sixth of August reflected frustrations with Europe, not only over the Fund governance, but on wider economic issues.

(c) Translated summary

Fig. 6. Arabic document titled *"America: an unprecedented step in the "International (Monetary Fund)""* and its summary.

by MUSE for Arabic, Hebrew and English languages, respectively. Figures 6, 7 and 8 demonstrate extracts produced by MUSE for Arabic, Hebrew and English texts, respectively.

- **Flexible pre-processing.** User is allowed to add the following language-specific analysis to the MUSE processing: stopwords removal, stemming, sentence segmentation and POS tagging, by configuring the system and providing necessary tools or data. For example, the user can decide that he/she wants to remove stopwords by providing the stopwords list in the processed language and turning on the "remove stopwords" parameter. Also, many parameters for constructing a document representation are configurable.

נתניהו ואבו מאזן הסכימו: לסיים מו"מ תוך שנה
לפי דיווח ב"ניו יורק טיימס", מזכירת המדינה האמריקאית קלינטון תודיע היום על חידוש השיחות הישירות בין ישראל לפלסטינים.
עוד נאמר בדיווח כי ראש הממשלה ויו"ר הרשות הפלסטינית הסכימו לסיים את השיחות בתוך שנה.
על הפרק: כל סוגיות הסדר הקבע
תגיות: הקוורטט,ברק אובמה,אבו מאזן,משא ומתן ישיר בקרוב שוב בבית הלבן? אבו מאזן עם אובמה ייענה להזמנה?
נתניהו עם אובמה
מזכירת המדינה האמריקאית הילרי קלינטון צפויה להודיע היום (שישי) על כך שישראל והפלסטינים יחדשו בתחילת החודש הבא את השיחות הישירות בין הצדדים, לאחר הפסקה של כשנה וחצי. לפי הדיווח ב"ניו יורק טיימס", נתניהו ויו"ר הרשות הפלסטינית אבו מאזן, הסכימו לסיים את השיחות בתוך שנה. עוד נמסר כי במשא ומתן הישיר יידונו כל סוגיות הסדר הקבע ובהן מעמדה של ירושלים, גבולות המדינה הפלסטינית החדשה, ערבויות ביטחוניות לישראל וזכות השיבה של הפליטים הפלסטינים .
מוקדם יותר אמר גורם דיפלומטי כי הקוורטט יקרא היום לישראל ופלסטינים להתחיל במשא ומתן ישיר בוושינגטון בשניים בספטמבר. הגורם ציין כי ההערכה היא ששני הצדדים ייענו להזמנה, והוסיף שנשיא ארה"ב ברק אובמה ישתתף אף הוא בשיחות.
אמש דווח כי בטיוטת ההודעה שצפוי לפרסם הקוורטט□ המורכב מארה"ב, האו"ם, האיחוד האירופי ורוסיה - לא יהיה מוזכר באופן מפורש הצורך בהקפאת הבנייה בהתנחלויות, כפי שהופיע בהצהרות הקוורטט הקודמות. עוד דווח כי בהודעה יצוין ששיחות השלום צפויות להגיע לסיומן בתוך כשנה.
עם זאת, המקורות דיווחו כי בטיוטת ההודעה נכתב כי "הקוורטט מאשרר את מחויבותו המלאה להצהרותיו הקודמות", בהן נקראה ישראל לעצור את הבנייה בהתנחלויות. על פי המקורות, בטיוטה נכתב עוד כי "משא ומתן ישיר ודו צדדי שיפתור את כל נושאי הליבה יוביל להסדר שנידון בין הצדדים, שיסיים את הכיבוש, ותוצאותיו יהיו מדינה פלסטינית שחיה בשלום לצד ישראל". עוד דווח, שבטיוטה נכתב כי "המשא ומתן יכול להסתיים בהצלחה תוך שנה", וכן כי "הצלחתו תדרוש את תמיכתם של מדינות ערב."
ביום ראשון החליטו שרי השביעייה להתעלם מהצהרת הקוורטט הבינ"ל, בנושא פתיחת המשא ומתן הישיר בין ישראל והפלסטינים. את הצהרת הקוורטט, הגדירו שרי השביעייה כ"עלה תאנה" של הפלסטינים לדחות את המהלך. בנוסף, הוחלט להמתין לזימון של האמריקאים, שאמור להתקבל אצל הצדדים בימים הקרובים, לפתוח בשיחות ישירות במצרים או בוושינגטון.

(a) Summarized document

לפי הדיווח ב"ניו יורק טיימס", נתניהו ויו"ר הרשות הפלסטינית אבו מאזן, הסכימו לסיים את השיחות בתוך שנה.
אמש דווח כי בטיוטת ההודעה שצפוי לפרסם הקוורטט□ המורכב מארה"ב, האו"ם, האיחוד האירופי ורוסיה - לא יהיה מוזכר באופן מפורש הצורך בהקפאת הבנייה בהתנחלויות, כפי שהופיע בהצהרות הקוורטט הקודמות.
על פי המקורות, בטיוטה נכתב עוד כי "משא ומתן ישיר ודו צדדי שיפתור את כל נושאי הליבה יוביל להסדר שנידון בין הצדדים, שיסיים את הכיבוש, ותוצאותיו יהיו מדינה פלסטינית שחיה בשלום לצד ישראל. "

(b) Original summary

According to a report in The New York Times, Netanyahu and Palestinian Authority Chairman Mahmoud Abbas, agreed to complete the talks within a year. Last night it was reported that the draft notification that is expected to be published by the Quartet - consisting of U.S., UN, EU and Russia - will not explicitly mention the need to freeze settlement construction, as appeared in the previous Quartet statements.
According to the sources, the draft stated that "direct and bilateral negotiations that will solve the core issues will lead to an agreement between the parties, and will end the occupation, and its results will be a Palestinian state living in peace alongside Israel."

(c) Translated summary

Fig. 7. Hebrew document titled *"Netanyahu and Abbas agreed to complete negotiations within a year"* and its summary.

BBC News - Images reveal Indonesian tsunami destruction

Aerial images from the tsunami-hit Mentawai Islands in Indonesia have revealed the extent of destruction, as officials raised the death toll to 311.
Flattened villages are plainly visible on the images, taken from helicopters circling the islands.

Rescuers have finally reached the area where 13 villages were washed away by the 3m (10ft) wave but 11 more settlements have not yet been reached. President Susilo Bambang Yudhoyono has arrived in the region.
He cut short a trip to Vietnam to oversee the rescue effort and has been briefed by officials in the port city of Padang on Sumatra.
He then began the journey to the remote and inaccessible Mentawai Islands, where he will also meet the governor of the area.
A 7.7-magnitude undersea earthquake triggered the tsunami two days ago.
But the BBC's Karishma Vaswani, in Jakarta, says rescue teams have still not arrived at the worst-affected communities, where the scale of the damage is still unclear.
More than 300 people are still missing, authorities say, and there are growing fears that many or most of those were swept out to sea by the tsunami.
Communication down

The first cargo plane loaded with tents, medicine, food and clothes landed on the islands on Wednesday.
But officials have had less luck transporting goods by boat some 175km (110 miles) across choppy seas from Padang.

"We're still looking for a means of transportation to be able to carry relief goods and personnel," local official Hidayatul Irham told the BBC's Indonesian service.
He said rescue teams dispatched to the island were unable to send back adequate reports because lines of communication with the remote islands were so bad.
Local disaster official Ade Edward said more than 400 people were still missing and 16,000 refugees had been moved to higher ground from the coastal areas.
The first images emerging from the islands, taken on mobile phones, showed bodies being collected from empty clearings where homes and buildings once stood.
Later, Vice-President Boediono and his entourage took helicopters to the island and released aerial images showing widespread destruction of buildings.
District chief Edison Salelo Baja said corpses were strewn along beaches and roads.

Government helicopters were able to survey the damage on Wednesday Locals were given no indication of the coming wave because an early-warning system put in place after the devastating 2004 tsunami had stopped working.
Ridwan Jamaluddin, of the Indonesian Agency for the Assessment and Application of Technology, told the BBC's Indonesian service that two buoys off the Mentawai islands were vandalised and out of service.
"We don't say they are broken down but they were vandalised and the equipment is very expensive. It cost us five billion rupiah each (£353,000; $560,000).
However, even a functioning warning system may have been too late for people in the Mentawai Islands.
The vast Indonesian archipelago sits on the Pacific Ring of Fire, one of the world's most active areas for earthquakes and volcanoes.
More than 1,000 people were killed by an earthquake off Sumatra in September 2009.
In December 2004, a 9.1-magnitude quake off the coast of Aceh triggered a tsunami in the Indian Ocean that killed a quarter of a million people in 13 countries including Indonesia, Sri Lanka, India and Thailand.

(a) Summarized document

More than 300 people are still missing, authorities say, and there are growing fears that many or most of those were swept out to sea by the tsunami .
Ridwan Jamaluddin, of the Indonesian Agency for the Assessment and Application of Technology, told the BBC's Indonesian service that two buoys off the Mentawai islands were vandalised and out of service .
However, even a functioning warning system may have been too late for people in the Mentawai Islands .

(b) Summary

Fig. 8. English document titled *"Images reveal Indonesian tsunami destruction"* and its summary.

- **A rich choice of statistical sentence features.** User can choose from 31 sentence features provided by MUSE for summarization by configuring their weights in a linear combination. Our recommendation for 10 best metrics identified by cluster analysis of their performance on English and Hebrew corpora can be found in (Litvak et al., 2010a).
- **Easy to use.** Text and HTML documents can be summarized with just one click. Figure 5 shows extract produced by MUSE for one of the DUC 2002 (DUC, 2002) documents.
- **Cross-lingual use of multiple ranking models.** MUSE allows to use the same trained model across different languages. As result, no need in retraining on a new

TABLE III
USE CASE DESCRIPTION

Use Case	Goal	Precondition	Postcondition	Brief
Configure	Specify preprocessing and summarizer settings	None	Stored configuration file for later use	User specifies all necessary parameters as: path to the input document/s, summary length, gold standard folder, etc. User can store his settings for the later use.
Train	Train a model	Parameters settings, train and test data	Trained model - weights for linear combination	Train the genetic algorithm on the training document set. The trained model can be stored for the later use.
Evaluate	Evaluate the system	Parameters settings, gold standard summaries	Average train and test scores (ROUGE)	Evaluate the system on the given document set using 10-fold cross-validation. The average ROUGE score is presented to the user.
Summarize	Summarize the input document/s	Settings and a weighted model	Summary for each input document	User can get a summary for each input document and store it in the file system. Also, different output representations are provided to the user: sentence scores, highlighted and sorted sentences.
Rouge	Calculate ROUGE score for the input summaries	Settings, input and gold standard summaries	ROUGE score	User can get a ROUGE score for the input summaries given gold standard summaries for the document set.

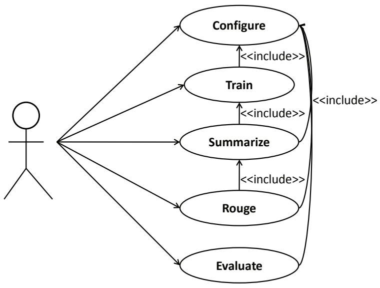


Fig. 9. Use case diagram of MUSE

corpus, language or genre. User is allowed to store the trained model in the file system for later use.

- **Storing summaries for later use.** The final summary can be exported to a file according to the user's settings.
- **Storing statistics for later analysis.** The results for the future statistical analysis (like summaries per individual metric, ROUGE score per document, etc.) can be stored in the file system for later use.
- **Summary options.** The size of the summary can be assigned according to specific goals: a cursory examination or a detailed survey. Specify the summary length by either the number of words/sentences in the summary or a percentage of the original text.
- **Output interpretability.** User can obtain various output representations: input document with colored extracted sentences, scores per sentence, and sentences sorted by their score.
- **High-level configuration.** MUSE has flexible pre-processing and optimization options. The user can work in the default mode as well as in advanced one. Advanced users can configure the settings of the genetic algorithm, pre-processing, etc.

III. CONCLUSIONS AND FUTURE WORK

In this article we described MUSE—a *supervised* language-independent summarizer for the text documents based on sentence extraction.

We described and detailed the MUSE architecture, approach and use cases.

MUSE does not require any language-specific knowledge and can be applied to any language with a minimum amount of text pre-processing. Moreover, our experiments show that the same weighting model is applicable across multiple languages.

In general, we can conclude that combination of as many independent statistic features as possible can compensate the lack of linguistic analysis and knowledge for selecting the most informative sentences to a summary. More features can be added to our system, and/or another supervised model can be used for the learning and optimization of a linear combination. We believe that such an approach generally works when retrained on different genres and languages.

We can recommend the following: If a corpus in the target language exists, the best approach is to train MUSE on the target-languge corpus, while periodically updating the trained model when new annotated data becomes available. If there is a corpus in any source language, but no high-quality target-language corpus is available, we would recommend to use the model trained on the source language corpus for summarizing documents in the target language.

In future work, MUSE may be evaluated on additional languages and language families, incorporate threshold values for threshold-based metrics into the GA-based optimization procedure, improve performance of similarity-based metrics in the multilingual domain, apply additional optimization techniques like Evolution Strategy (Beyer & Schwefel, 2002), which is known to perform well in a real-valued search space, and extend the search for the best summary to the problem of multi-objective optimization, combining several summary quality metrics.

ACKNOWLEDGMENTS

We are grateful to Nataly Alalouf for implementing the stand-alone and the web versions of MUSE.

REFERENCES

[1] D. Radev, S. Blair-Goldensohn, and Z. Zhang, "Experiments in single and multidocument summarization using mead," *First Document Understanding Conference*, 2001.

[2] H. Saggion, K. Bontcheva, and H. Cunningham, "Robust generic and query-based summarisation," in *EACL '03: Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics*, 2003.

[3] I. Mani, *Automatic Summarization*. Natural Language Processing, John Benjamins Publishing Company, 2001.

[4] M. Litvak, M. Last, and M. Friedman, "A new approach to improving multilingual summarization using a genetic algorithm," in *Proceedings of the Association for Computational Linguistics (ACL) 2010*, 2010, pp. 927–936.

[5] R. Mihalcea, "Language independent extractive summarization," in *AAAI'05: Proceedings of the 20th national conference on Artificial intelligence*, 2005, pp. 1688–1689.

[6] DUC, "Document understanding conference," 2002, http://duc.nist.gov.

[7] C.-Y. Lin and E. Hovy, "Automatic evaluation of summaries using n-gram co-occurrence statistics," in *NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, 2003, pp. 71–78.

[8] M. Litvak, S. Kisilevich, D. Keim, H. Lipman, A. Ben-Gur, and M. Last, "Towards language-independent summarization: A comparative analysis of sentence extraction methods on english and hebrew corpora," in *Proceedings of the CLIA/COLING 2010*, 2010.

[9] E. Freeman, E. Freeman, K. Sierra, and B. Bates, *Head First Design Patterns, First Edition, Chapter 1*. OŔeilly Media, Inc, 2004, p. 24.

[10] H.-G. Beyer and H.-P. Schwefel, "Evolution strategies: A comprehensive introduction," *Journal Natural Computing*, vol. 1, no. 1, pp. 3–52, 2002.