A First Summarization System of a Video in a Target Language

Kamel Smaïli, Dominique Fohr, Carlos González-Gallardo, Michal Grega, Lucjan Janowski, Denis Jouvet, Artur Komorowski, Arian Kozbial, David Langlois, Mikolaj Leszczuk, et al.

A First Summarization System of a Video in a Target Language *

K. Smaïli[1], D. Fohr[1], C.E. González-Gallardo[2], M. Grega[3], L. Janowski[3], D. Jouvet[1], A. Komorowski[3], A. Koźbiał[3], D. Langlois[1], M. Leszczuk[3], O. Mella[1], M.A. Menacer[1], A. Mendez[4], E. Linhares Pontes[2], E. SanJuan[2], D. Świst[3], J.M. Torres-Moreno[2,5], and B.Garcia-Zapirain[4]

[1] Loria University of Lorraine - France
[2] LIA Université d'Avignon et des Pays de Vaucluse - France
[3] AGH University of Science and Technology Kraków - Poland
[4] University of DEUSTO Bilbao - Spain
[5] Ecole Polytechnique de Montréal - Canada.

Abstract. In this paper, we present the first results of the project AMIS (Access Multilingual Information opinionS) funded by Chist-Era. The main goal of this project is to understand the content of a video in a foreign language. In this work, we consider the understanding process, such as the aptitude to capture the most important ideas contained in a media expressed in a foreign language. In other words, the understanding will be approached by the global meaning of the content of a support and not by the meaning of each fragment of a video.
Several stumbling points remain before reaching the fixed goal. They concern the following aspects: Video summarization, Speech recognition, Machine translation and Speech segmentation. All these issues will be discussed and the methods used to develop each of these components will be presented. A first implementation is achieved and each component of this system is evaluated on a representative test data. We propose also a protocol for a global subjective evaluation of AMIS.

Keywords: Video Summarization · Speech Recognition · Machine Translation · Text Boundary Segmentation · Text Summarization · Sentence Compression.

Introduction

Nowadays, the information is widely available in different medias: TV, social networks, newspapers, etc. The main difference in comparison to what we had one or two decades before is that people can access to the videos of social networks in foreign languages. When, the video does not necessitate any understanding, there is no main problem. In the opposite, when the information necessitates to understand the language, a human being is limited in terms of mastering foreign language, even if YouTube proposes a rough translation of some contents. In

* Supported by Chist-Era (AMIS project).

AMIS, a Chist-Era project[6], the main objective is to make available a system, helping people to understand the content of a source video by presenting its main ideas in a target understandable language. The understanding process is considered here to be the comprehension of the main ideas of a video. We think that the best way to do that, is to summarize the video for having access to the essential information. Henceforth, AMIS focuses on the most relevant information by summarizing it and by translating it to the user if necessary. As a result, AMIS will permit to have another side of story of an event since we can, for instance, have the Russian version of the war in Syria.

Several skills are necessary to achieve this objective: video summarization, automatic speech recognition, machine translation, text summarization, etc. Each output of a a sub-system of AMIS can enrich in upstream or downstream the other modules. That makes AMIS working such as a workflow where the flow refers to the information necessary for a component. In this article, we will present the first result that works such as a pipeline system connecting the output of each component to the input of the next one.

Different components of AMIS

Video Summarization

We designed and developed an operational framework for summarization of newscasts and reports [11]. The framework is designed in such a way, that it allows for easy experimentation with different approaches to video summarization. The framework hosts several high- and low-level meta-data extraction algorithms (referring to our former research conducted within the scope of e.g. IMCOP project [1]) that include detection of the anchor-person, recognition of day and night shots and extraction of low-level video quality indicators. The main summarization processes start with Shot Boundary Detection (SBD). This algorithm helps in prediction whether video is static or dynamic. Also, through SBD we can calculate and compare data per shots instead of frames which is a way more efficient while analyzing video clips over longer durations. The video quality indicators mentioned above are used for calculating coefficient of activity which is a product of two indicators – Spatial Activity and Temporal Activity. These indicators show amount of details appearing on the frame and how dynamic the frame is in comparison to the previous frame, respectively. The coefficient of activity is calculated in two steps – firstly per each frame and then as an average per each shot.

The final summarization is built from shots with higher or equal value of coefficient of activity than average value for entire video.

Speech recognition system

Automatic speech recognition (ASR) and machine translation (MT) are among the AMIS project key technologies for understanding videos. Although these

[6] http://deustotechlife.deusto.es/amis/

items were developed as separate modules, the goal is to include them more tightly with other modalities and other processes. AMIS project deals with videos in French, English and Arabic languages. Arabic is considered as one of the foreign input language of the videos. That is why an Arabic automatic speech recognition system has been developed in Loria. This system is based on the state-of-the-art methods and is trained on large acoustic and text corpora. For that we developed *ALASR: Arabic Loria Automatic Speech Recognition system* [16]. A speech recognition system needs at least two components: an acoustic model and a language model. In the following, each component is described by presenting the different steps to train such models. Training necessitates two kinds of data: acoustic and textual, which are presented in the following.

Acoustic model The development of the acoustic model is based on Kaldi [18] recipe, which is a state-of-the-art toolkit for speech recognition based on Weighted Finite State Transducers [17]. The ASR system uses 13-dimensional Mel-Frequency Cesptral Coefficients (MFCC) features with their first and second order temporal derivatives, which leads to 39-dimensional acoustic features. For Arabic, 35 acoustic models (28 consonants, 6 vowels and silence) are trained. The emission probabilities of the HMM models are estimated by DNN (namely DNN-HMM). The DNN-HMM are trained by applying sMBR criterion [24] and using 40 dimensional features vector (fMLLR) [6] for speaker adaptation. The topology of the neural network is as follows: a 440-dimensional input layer (40 $\times$ 11 fMLLR vectors), 6 hidden layers composed by 2048 nodes and a 4264-dimensional output layer, which represents the number of HMM states. And finally, the total number of weights to estimate is about 30.6 million.

Language model In Arabic, even in newspapers,several words could be simplified especially at the beginning or at the end by replacing a specific letter by another one or by omitting the *hamza* symbol. Unfortunately, both writing of a word may exist in a same document. This leads to share a probability over two forms that correspond to the same word. For instance, the word إستعمَال (*uses*) is the right way to write it, but people could omit the *hamza* and write it such as: استعمَال. Obviously, this is not the only case, several other points have to be treated. Some of them are specific to Arabic and others are used in the majority of other natural languages. Several preprocessing tasks have been done on the corpora before calculating the language model [16]. The training corpus is composed by GigaWord and the speech transcripts. As this data set is unbalanced, a 4-gram language model, for each part of this corpus, has been developed and combined linearly. The optimal weights are determined in order to maximize the likelihood of the development corpus. Due to the memory constraints, the full 4-gram language model has been pruned by minimizing the relative entropy between the full and the pruned model [21].

Machine Translation

A statistical machine translation has been developed in the direction Arabic – English, since Arabic is considered such as the foreign language of the video to translate to a summarized video in English. For training, we use a parallel (Arabic — English) corpus of 9.7 million sentences extracted from United Nation corpus concerning the period of January 2000 – September 2009. A 4-gram language model has been trained on the target language of the mentioned parallel corpus. The vocabulary contains 224,000 words. The development and the test corpus are composed by 3,000 parallel sentences. This component is a statistical machine translation, but in the next version of AMIS, we will use a neural network machine translation [15].

Text Summarization

Automatic Text Summarization (ATS) is a Natural Language Processing (NLP) task [23]. The main objective of ATS is to find the most important information from a text source in order to produce an abridged and informative version. In the AMIS project, the ATS component aims to summarize a newscast or report video in French, English or Arabic languages based only on the textual information provided by the ASR or by the MT modules.

ATS is divided in three different families depending on the followed approach to generate a summary: summarization by extraction, summarization by abstraction and summarization by sentence compression [23].

We implemented mainly the extractive summarization paradigm in the project because it is more robust to external noise like speech disfluencies and ASR errors [2, 4]. Also, we performed some exploratory experiments using the sentence compression paradigm.

Extractive Text Summarization Extractive Text Summarization (ETS) aims to select the most pertinent segments of the transcribed video based on different criteria like information content, novelty factor and relative position. ARTEX (Autre Résumer de TEXtes), originally developed for French, English and Spanish is an ETS system described in [22] by Torres-Moreno *et al.*. Within the AMIS project, a Modern Standard Arabic (MSA) extension has been developed and added to ARTEX, making possible the generation of summaries in all the languages involved in AMIS.

Multi-Sentence Compression Multi-Sentence Compression (MSC) combines the information of a cluster of similar sentences to generate a new sentence, hopefully grammatically correct, which compresses the most relevant data of this cluster. Among several state-of-the-art MSC methods, Linhares Pontes *et al.* [12] used an Integer Linear Programming (ILP) formulation to guide the MSC using a list of keywords. Our system incorporates this approach to use the

keywords of a transcribed video to guide the compression of similar sentences and to improve the informativeness of summaries

Independently of the summarization approach ATS relies on the existence of sentences either to select, reformulate or compress the source text. One main issue to take into account is that in the AMIS project, the source from which a text summary is created is the transcript of a ASR system or its translation; this transcript does not contain any punctuation mark, hence sentences are nonexistent. In order to solve this problem, a specialized module of sentence boundary detection has been developed.

Sentence Boundary Detection Deep research has been done concerning Sentence Boundary Detection (SeBD) of the ASR transcripts covering the three languages of the project: English, French and Arabic. Written text differs from spoken language in the way the writer/speaker expresses its ideas. In spoken language, sentences are not as well defined as in written text, in this context a segment is defined by a sentence-like unit (SU). Well-formed sentences, phrases and single words can be considered SUs [13].

The developed SeBD system uses mainly textual features and convolutional neural networks (CNN) to segment the transcripts and generate SUs [7]. Arabic SeBD module was trained with a subset of 50 million words of the Arabic Gigaword Corpus.

Besides ATS, SeBD has shown to be of vital importance for other NLP tasks in the AMIS project like NMT. For these reason a stand-alone version of the module has been deployed to be used independently by other AMIS modules.

Audio summarization

Audio summarization task is a new approach to generate speech-to-speech summaries. This is a difficult task, because in the audio signal there is not any linguistic information (words, sentences, etc.). Although low level features such as signal intensity, starting and ending time could be used for extracting and selecting informative segments, reliable methods for automatic extraction of high level features (prosodic and linguistic features) from spontaneous speech have not yet been established [5, 25]. We explore several neural architectures using deep learning in order to find the best features in the hidden layer. These features will allow us to capture the most important abstract structure (linguistic level) from a low level resource (signal).

Global architecture

Several architectures are proposed to summarize a video in a target language. These architectures are presented in Figure 1. The four ($SC1$, $SC2$, $SC3$ and $SC4$) architectures are different. In $SC1$, a summary video is achieved directly

without using any knowledge on the audio content of the video. The content of the result of this summary is then transcribed thanks to ALASR (our speech recognition system) then translated to English. The result of the translation is integrated such as subtitles to the summarized video.

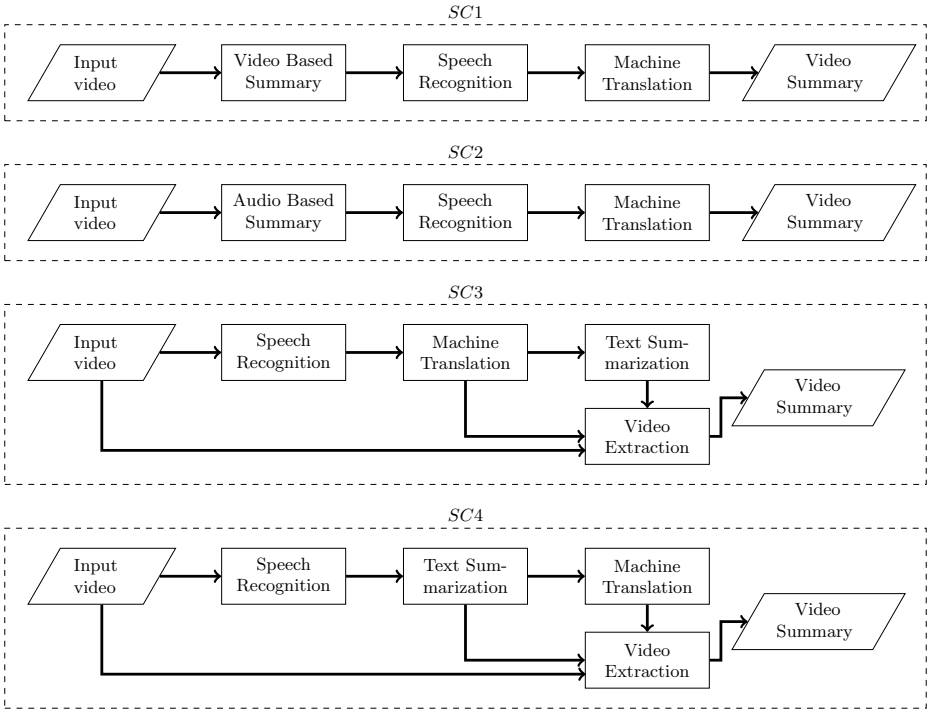

Fig. 1. Different architectures for summarizing a video to a target language

$SC2$ is an original architecture in which an audio summary is proposed on the audio part of the original video. The result is then transcribed and translated such as in $SC1$.

$SC3$ and $SC4$ are similar architectures, they take benefit from the result of ALASR. The result of this step is then a text in Arabic, then the blocks (Machine Translation + Text Summarization) and (Text Summarization + Machine Translation) are respectively performed on $SC3$ and $SC4$.

In this paper, only the first architecture ($SC1$) is presented.

Objective Evaluation

In order to appreciate the relevance of the whole system developed, it is necessary to evaluate it globally, but also each component should be evaluated. In the following, we will evaluate each component since the consequences of the weakness of a system could be propagated to the other components.

Video Summarization The framework is evaluated using annotated video sequences. A pool of experts decides which frames are key frames (meaning: very important frames, core of the video) and which have to be in the summary. We use VLC media player to extract frames from single shots. Obviously, this evaluation process is time-consuming and subjective when the key frames are chosen. In order to describe the obtained results, we are calculating Precision, Recall and F1 score for each sequence and algorithm.

We considered video summarization based only on the visual evaluation. Both human and algorithm were provided with video without any additional audio description. Of course a person creating the evaluation could understand some written text appearing on the screen. Our goal was to validate if the summary created by human is similar to the summary created by algorithm, focusing on the visual part only. Such evaluation can be found in literature [8, 9, 19, 20] just to name a few. Nevertheless the summary of news was not considered. In our research we noted that making a reasonable summary for human observer is very difficult. As a consequence comparing to an algorithm is not as precise as for other cases considered in literature.

The first problem we found is the length of summary. In Table 1 the length of summaries provided by a human observer are presented. We can see that the shortest is just 15% of the original video while the longest is 61%. Comparing such different solutions is difficult. The difference comes from very different natures of the video news, which can span from a talking head to a report from a field where there is an action. In order to help with comparing human and automatic summarization, the automatic algorithm has the information about the length of the summary provided by humans.

The automatic and human summaries were compared by precision and recall metrics. We calculated how many frames marked by human were also marked by an algorithm. So true positive means that both human and algorithm marked the same frame. True negative means that both human and algorithm did not mark specific frame. The results obtained for 50 evaluated sequences are presented in Table 2. The obtained results are not very good but even comparing summaries provided by two humans are not much better. The problem is the content, which is already a summary.

Arabic Automatic Speech Recognition : The ASR system for Arabic has been trained on an acoustic corpus of 63 hours (Nemlar [14] and NetDC [3]). The Language Model has been trained on the GigaWord. The vocabulary is composed of 95k words with an average of 5.07 pronunciations for each entry. After several

Video ID	Summary length	Source length	Percentage
1	205	473	43%
2	86	187	46%
3	76	277	27%
4	69	200	35%
5	85	186	36%
6	119	194	**61%**
7	41	281	**15%**
8	41	233	18%

Table 1. Summary length comparison. The summaries in this table are created by human.

Recall	Precision	F1	Accuracy
0.13	0.36	0.19	0.36

Table 2. Performance of video summarization

tests, tuning and improvements ALASR achieves the performance presented in Table 3. This performance is achieved on a tuning and a test corpora of 31,000

	Dev WER	Test WER
ALASR	13.07	14.02

Table 3. Performance of ALASR in terms of WER

sentences for each of them. This test is done on data not extracted from our video database. The issue is that we do not have any reference transcription corpus for these videos. The evaluation is then impossible. To overcome this problem, we decided to build a pseudo-reference by aligning, for each YouTube video from Euronews channel, the automatic transcription and textual data from the corresponding Youtube and Euronews webpages. The transcription is considered as a reference if the WER is under a chosen threshold [10].

Experiments have been done on a corpus of 1,300 sentences (a mixture of transcriptions from YouTube and Euronews). We have to notice, that this transcription does not correspond exactly to what has been pronounced. Consequently, the performance we provide below is under-estimated. Under these conditions, ALASR achieves a WER of 36.5.

Machine Translation The machine translation system has been evaluated on separated corpora: on a general one and on data extracted from our video database. In Table 4, we give the result of Arabic-English machine translation

system evaluated on 3,000 sentences extracted from a corpus of the United Nations [26]. As for our ASR system, the evaluation on our database is difficult

	Test (3k sentences)
BLEU	39

Table 4. The evaluation of the Arabic–English MT system.

because we do not have any reference corpus for Machine Translation. To overcome this limit, we decided to create two reference corpora. The corpora are composed respectively by the translation of 197 videos of Euronews that correspond to 1,253 sentences achieved by Google and SYSTRAN. It means that we consider the result of the translation of Google respectively SYSTRAN as the references. The results are given in Table 5. The achieved BLEU for our

System	**AMIS**	**Google**	**Systran**
AMIS		26.7	9.9
Google	26.7		12.8
Systran	10	12.9	

Table 5. MT evaluation on AMIS data.

system when SYSTRAN is considered such as the reference is 9.9, while the results when the translation of Google is considered as the reference is equal to 26.7. To understand the weak performance we get with SYSTRAN as reference, we used Google as a translator. The achieved BLEU for this experience is 12.8. This shows that both Google and our system fails to get good results with a translation produced by SYSTRAN.

Sentence Boundary Detection The SeBD module was approached as a classification task, where the system should decide if a target word corresponds or not to a boundary between two SUs. Table 6 shows the results of a strict evaluation for the Arabic SeBD module over 12M samples of the Arabic Gigaword corpus. The method seems to perform really well concerning the "no boundary" class (NO_BOUND); both Precision and Recall achieve a value over 92%. By contrast the performance related to the "boundary" class (BOUND) drops almost 15% for Precision and 32.5% for Recall. The unbalanced nature of the data influences this drop in performance. The "no boundary" class represents the 84% of the samples, against the 16% from the "boundary" class. Further work is in develop to reduce the gap between both classes.

Class	Precision	Recall	F1
NO_BOUND	0.928	0.963	0.945
BOUND	0.782	0.638	0.700

Table 6. Performance of the AMIS SeBD

Subjective evaluation

In the previous section, we evaluated each component independently from the others, in this section we propose a method to evaluate the system as a whole. The method is based on a questionnaire that involves the final users, which is considered as the best indicator of the quality of this type of systems. The evaluators of the summarized videos will be 12 participants in total, 6 in Arabic (3 men and 3 women per language). The inclusion criteria are being 18 years old, with at least high school level, while the exclusion criteria is having understanding problems, reading or writing impairment. The material to be evaluated will consist of 25 videos tackling mixed topics: Politics, Soccer, War, Homosexuality. The number of summarized videos per user participating as tester is 3. The aim of the designed questionnaires is analyzing the quality of the video summarization with the best precision, taking into account the proposed resources. So, we will include 2 generic questions for all the videos, that can be evaluated from 0 ("Not done") to 4 ("Excellent"), and a set of 3/5 specific questions (depending of the length of the video) with 3 possible answers. In that way, we will know if the summary is understandable and if there is any part out of context, using the generic questions, and if the main ideas of the original video have been gathered by the summarization, using the video specific questions. The assessment data analysis will consist on statistical analysis of questionnaires and the application of some machine learning techniques, if possible for clustering and comparison purposes between genders, languages, etc.

Conclusion

In this article, several research aspects have been investigated through AMIS project. An understanding system of a foreign video has been developed by summarizing the source video. The objective was to capture the main idea of a video and to restore it into English. A whole system has been carried out by implementing several sub-systems, where each of them represents a real scientific challenge. The different sub-systems were assembled to give rise to the first version of AMIS. The system is operational and the results are certainly improvable, but are better than what we expected. We thought that the serialization of these systems had to produce very bad results, but this is not the case. We are currently working on developing more relevant architectures that would probably yield better results. In parallel, each sub-system is subject to regular improvements making the global system better.

Acknowledgment

We would like to acknowledge the support of Chist-Era for funding this work through the AMIS (Access Multilingual Information opinionS) project. Research work funded by the National Science Center, Poland, conferred on the basis of the decision number DEC-2015/16/Z/ST7/00559.

References

1. Baran, R., Zeja, A.: The imcop system for data enrichment and content discovery and delivery. In: 2015 International Conference on Computational Science and Computational Intelligence (CSCI). pp. 143–146 (Dec 2015). https://doi.org/10.1109/CSCI.2015.137
2. Bell, P., Lai, C., Llewellyn, C., Birch, A., Sinclair, M.: A system for automatic broadcast news summarisation, geolocation and translation. In: INTERSPEECH. pp. 730–731 (2015)
3. Choukri, K., Nikkhou, M., Paulsson, N.: Network of data centres (netdc): Bnsc-an arabic broadcast news speech corpus. In: LREC (2004)
4. Christensen, H., Kolluru, B., Gotoh, Y., Renals, S.: From text summarisation to style-specific summarisation for broadcast news. In: European Conference on Information Retrieval. pp. 223–237. Springer (2004)
5. Furui, S., Kikuchi, T., Shinnaka, Y., Hori, C.: Speech-to-text and speech-to-speech summarization of spontaneous speech. IEEE Trans. Speech and Audio Processing **12**(4), 401–408 (2004)
6. Gales, M.J.: Maximum likelihood linear transformations for hmm-based speech recognition. Computer speech & language **12**(2), 75–98 (1998)
7. González-Gallardo, C.E., Torres-Moreno, J.M.: Sentence boundary detection for french with subword-level information vectors and convolutional neural networks. arXiv preprint arXiv:1802.04559 (2018)
8. Gygli, M., Grabner, H., Gool, L.V.: Video summarization by learning submodular mixtures of objectives. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3090–3098 (June 2015). https://doi.org/10.1109/CVPR.2015.7298928
9. Huang, M., Mahajan, A.B., Dementhon, D.F.: Automatic performance evaluation for video summarization. Tech. rep.
10. Jouvet, D., Langlois, D., Menacer, M.A., Fohr, D., Mella, O., Smaïli, K.: Adaptation of speech recognition vocabularies for improved transcription of youtube videos. In: Proceedings of the ICNLSSP Conference (2017)
11. Leszczuk, M., Grega, M., Koźbiał, A., Gliwski, J., Wasieczko, K., Smaïli, K.: Video summarization framework for newscasts and reports – work in progress. In: Dziech, A., Czyżewski, A. (eds.) Multimedia Communications, Services and Security. pp. 86–97. Springer International Publishing, Cham (2017)
12. Linhares Pontes, E., Huet, S., Linhares, A.C., Torres-Moreno, J.M.: Multi-sentence compression with word vertex-labeled graphs and integer linear programming. In: Proceedings of TextGraphs-12: the Workshop on Graph-based Methods for Natural Language Processing. Association for Computational Linguistics (2018)
13. Liu, Y., Chawla, N.V., Harper, M.P., Shriberg, E., Stolcke, A.: A study in machine learning from imbalanced data for sentence boundary detection in speech. Computer Speech & Language **20**(4), 468–494 (2006)

14. Maegaard, B., Choukri, K., Jørgensen, L.D., Krauwer, S.: Nemlar: Arabic language resources and tools. In: Arabic Language Resources and Tools Conference. pp. 42–54 (2004)
15. Menacer, M.A., Langlois, D., Mella, O., Fohr, D., Jouvet, D., Smaïli, K.: Is statistical machine translation approach dead? In: ICNLSSP 2017 - International Conference on Natural Language, Signal and Speech Processing. pp. 1–5. ISGA, Casablanca, Morocco (Dec 2017), https://hal.inria.fr/hal-01660016
16. Menacer, M.A., Mella, O., Fohr, D., Jouvet, D., Langlois, D., Smaïli, K.: Development of the Arabic Loria Automatic Speech Recognition system (ALASR) and its evaluation for Algerian dialect. In: ACLing 2017 - 3rd International Conference on Arabic Computational Linguistics. pp. 1–8. Dubai, United Arab Emirates (Nov 2017), https://hal.archives-ouvertes.fr/hal-01583842
17. Mohri, M., Pereira, F., Riley, M.: Speech recognition with weighted finite-state transducers. In: Springer Handbook of Speech Processing, pp. 559–584. Springer (2008)
18. Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K.: The kaldi speech recognition toolkit. In: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding. IEEE Signal Processing Society (Dec 2011), iEEE Catalog No.: CFP11SRW-USB
19. Quemy, A., Jamrog, K., Janiszewski, M.: Unsupervised video semantic partitioning using ibm watson and topic modelling. In: Proceedings of the Workshops of the EDBT/ICDT 2018 Joint Conference (EDBT/ICDT 2018). pp. 44–49 (March 2018)
20. Sharghi, A., Laurel, J.S., Gong, B.: Query-focused video summarization: Dataset, evaluation, and a memory network based approach. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017. pp. 2127–2136. IEEE Computer Society (2017). https://doi.org/10.1109/CVPR.2017.229, https://doi.org/10.1109/CVPR.2017.229
21. Stolcke, A.: Entropy-based pruning of backoff language models. arXiv preprint cs/0006025 (2000)
22. Torres-Moreno, J.M.: Artex is another text summarizer. arXiv preprint arXiv:1210.3312 (2012)
23. Torres-Moreno, J.M.: Automatic Text Summarization. Wiley and Sons, London (2014)
24. Veselỳ, K., Ghoshal, A., Burget, L., Povey, D.: Sequence-discriminative training of deep neural networks. In: Interspeech 2013 (2013)
25. Zhang, J.J., Fung, P.: Active learning with semi-automatic annotation for extractive speech summarization. ACM Transactions on Speech and Language Processing (TSLP) **8**(4), 6 (2012)
26. Ziemski, M., Junczys-Dowmunt, M., Pouliquen, B.: The united nations parallel corpus v1. 0. In: LREC (2016)