CUNI at MediaEval 2012 Search and Hyperlinking Task

Petra Galuščáková and Pavel Pecina
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Charles University in Prague, Czech Republic
{galuscakova,pecina}@ufal.mff.cuni.cz

ABSTRACT

The paper describes the Charles University setup used in the Search and Hyperlinking task of the MediaEval 2012 Multimedia Benchmark. We applied the Terrier retrieval system to the automatic transcriptions of the video recordings segmented into shorter parts and searched for those relevant to given queries. Two strategies were applied for segmentation of the recordings: one based on regular segmentation according to time and the second based on semantic segmentation by the TextTiling algorithm. The best results were achieved by the Hiemstra and TF-IDF models on the LIMSI transcripts and various segmentation.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information search and Retrieval; H.3.4 Systems and Software

1. INTRODUCTION

The Search and Hyperlinking task [1] of the MediaEval 2012 Multimedia Benchmark is defined as finding video segments based on a short query in natural language (Search subtask) and finding links to other relevant segments in the collection (Hyperlinking subtask). The Charles University team participated in the Search subtask only.

The test collection, consisting of video recordings crawled from an Internet video sharing platform, was provided with two automatic speech recognition transcripts of the audio tracks: LIMSI [6] and LIUM [8] and user generated metadata including *titles*, *tags*, and short *descriptions*. Additionally, the data was accompanied by some other features such as concept recognition, shot segmentation and face detection, but this kind of information was not used in our experiments. The collection was split into a development set and a test set, containing 5 288 and 9 550 videos respectively.

This papers reports on the experiments and results carried out by the team of the Charles University in Prague.

2. APPROACH DESCRIPTION

In our approach, we examined two strategies for segmentation of the recordings. First, we split the recordings into sets of overlapping passages of the same length: 45, 60, 90, and 120 seconds. The 45-seconds-long segments start every 15 seconds and the longer ones (60, 90, and 120) start every 30 seconds. At the end, the overlapping passages were removed from the retrieved results and only the better scored passages were kept and submitted.

In the second approach, we attempted to perform semantic segmentation by employing the TextTiling [4] algorithm which divides an input text into semantically coherent segments based on the vocabulary usage, which was also used in the RSR MediaEval Track in 2011 [2]. First, we automatically detect sentences boundaries (based mainly on punctuation) and then run TextTiling with the following settings: average number of words in a sentence set to 27 and average number of sentences in one segment set to 9. These values were optimized on the development set of the test collection and correspond to the 90-seconds-long passages.

Our experiments were carried out with the Terrier[1] information retrieval system – an open source search engine provided with a wide range of indexing and retrieval functionalities. The best results on the development set of the test collection were achieved by employing Hiemstra language model [5] and the TF-IDF vector space model [7].

3. EXPERIMENTS AND RESULTS

We submitted five runs to the Search task: Run 1 and 2 were required by the organizers and exploited only the *title* field of the queries. The remaining three runs used also the *short_title* field. In the baseline we employed the whole recordings with no segmentation. Run 2 was based on the one-best LIUM transcripts, the other runs worked with the LIMSI transcript. In all runs, we added some of the metadata information (*description* and *tags*) to each passage. We retrieved the top 500 results in each run but the overlapping segments in the list of the recordings were removed and therefore the highly ranked list were shorter. The Terrier system was applied with the following settings: Porter Stemmer, stopword list, query expansion, and default parameters for both the TF-IDF and Hiemstra model.

The results obtained on the test set are given in Table 1 in terms of all official evaluation measures: MRR, mGAP, and MASP. Scores are presented for all runs, computed for window sizes of 60, 30, and 10 seconds. The MRR Measure evaluates retrieval quality inside a given window; the mGAP score takes into account the precise starting point of the segment only; and the MASP score considers both the starting and ending points of a relevant segment. The mGAP scores are also calculated with the asymmetrical modified penalty function proposed in [3].

[1] `http://terrier.org`

MediaEval 2012 Workshop, October 4-5, 2012, Pisa, Italy

run	transcripts	model	segmentation	MRR			mGAP				MASP		
				60	30	10	60	30	10	Mod	60	30	10
-	LIMSI	Hiemstra	none	0.339	0.269	0.102	0.207	0.102	0.000	0.399	0.001	0.001	0.000
1	LIMSI	TF-IDF	90 sec	0.422	0.308	0.148	0.258	0.158	0.028	0.448	0.106	0.080	0.037
2	LIUM	Hiemstra	60 sec	0.383	0.344	0.186	0.256	0.167	0.033	0.370	0.110	0.110	0.056
3	LIMSI	TF-IDF	60 sec	0.467	**0.395**	0.193	**0.305**	**0.197**	0.036	0.495	0.156	0.140	0.061
4	LIMSI	Hiemstra	90 sec	**0.470**	0.363	0.193	0.290	0.194	**0.042**	**0.500**	0.123	0.088	0.042
5	LIMSI	Hiemstra	TexTiling	0.278	0.256	**0.199**	0.206	0.161	0.034	0.286	**0.161**	**0.166**	**0.152**

Table 1: Results on the test set for different types of transcripts, segmentation, and retrieval models.

The highest MRR and mGAP scores were achieved by the regular segmentation (60 and 90 seconds). In terms of MASP, the regular segmentation was outperformed by semantic segmentation produced by TextTiling. The mGAP values calculated with the modified penalty function are higher than for the other window sizes. This is caused by the wider window used in the calculation: the window is either 150 or 210 seconds wide, depending on the position of the starting point of the retrieved segment and the starting point of the relevant segment. The differences between Run 3 and Run 4 are small for this score, but the difference between the runs for the LIMSI and LIUM transcripts is substantial. Also, the value achieved by TextTiling is very low, this is even outperformed by the baseline. Interestingly, the MASP score of Run 5 for window size of 30 seconds is higher than the score for window size of 60 seconds.

Figure 1 visualizes the effect of varying the segment length on the retrieval quality of the Hiemstra model (in terms of MRR, mGAP, and MASP measured with 60 second window size). Segmentation to shorter passages gives higher mGAP and MASP scores but the maximum values of MRR are achieved when 90-seconds-long segments are used and both shortening and lengthening decrease the scores.

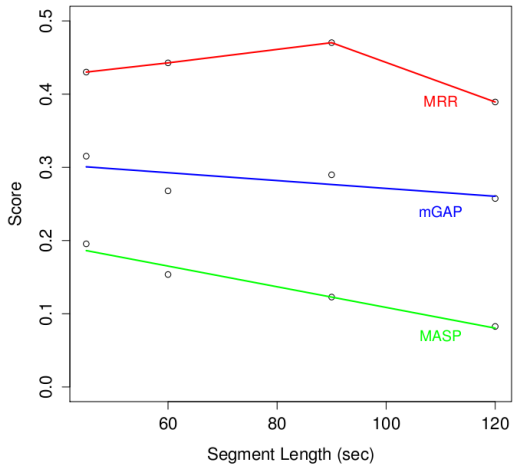

Figure 1: The effect of varying segment length on the retrieval quality (Hiemstra model evaluated with 60 second window size).

4. CONCLUSIONS AND FUTURE WORK

In this paper, we have described our retrieval experiments submitted to the Search task of the MediaEval 2012 Multimedia Benchmark. The Terrier system achieved the best results with the Hiemstra and TF-IDF model on the LIMSI transcripts. The highest MRR and mGAP scores were obtained using regular 60 and 90 second segmentation. In terms of MASP, the regular segmentation was outperformed by semantic segmentation produced by TextTiling. We have also analysed the effect of varying segment length on the retrieval quality. Our future work will focus on improving the segmentation which was shown to have a substantial impact on retrieval quality.

5. ACKNOWLEDGMENTS

This research was supported by the Ministry of Culture of the Czech Republic (project AMalach, program NAKI, grant n. DF12P01OVV022), the Czech Science Foundation (grant n. P103/12/G084) and Ministry of Education of the Czech Republic (project LINDAT-Clarin, grant n. LM2010013).

6. REFERENCES

[1] M. Eskevich, G. J. F. Jones, S. Chen, R. Aly, R. Ordelman, and M. Larson. Search and Hyperlinking Task at MediaEval 2012. In *MediaEval 2012 Workshop*, Pisa, Italy, October 4-5 2012.

[2] M. Eskevich, G. J. F. Jones, C. Wartena, M. Larson, R. Aly, T. Verschoor, and R. Ordelman. Comparing Retrieval Effectiveness of Alternative Content Segmentation Methods for Internet Video Search. In *CBMI 2012*, pages 1–6, June 2012.

[3] P. Galuščáková, P. Pecina, and J. Hajič. Penalty Functions for Evaluation Measures of Unsegmented Speech Retrieval. In *CLEF*, volume 7488 of *Lecture Notes in Computer Science*, pages 100–111. Springer, 2012.

[4] M. A. Hearst. TextTiling: Segmenting Text into Multi-paragraph Subtopic Passages. *Comput. Linguist.*, 23(1):33–64, Mar. 1997.

[5] D. Hiemstra. *Using Language Models for Information Retrieval*. Univ. Twente, 2001.

[6] L. Lamel and J.-L. Gauvain. Speech Processing for Audio Indexing. In *Proc. of the 6th international conference on Advances in NLP*, GoTAL '08, pages 4–15. Springer-Verlag, 2008.

[7] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, 2008.

[8] A. Rousseau, F. Bougares, P. Deléglise, H. Schwenk, and Y. Estève. LIUM's systems for the IWSLT 2011 Speech Translation Tasks. In *IWSLT*, San Francisco (USA), 8-9 Sept 2011.