# CUNI at MediaEval 2014 Search and Hyperlinking Task: Search Task Experiments

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

## ABSTRACT

In this paper, we describe our participation in the Search part of the Search and Hyperlinking Task in MediaEval Benchmark 2014. In our experiments, we compare two types of segmentation: fixed-length segmentation and segmentation employing Decision Trees on a set of various features. We also show usefulness of exploiting metadata and explore removal of overlapping retrieved segments.

## 1. INTRODUCTION

The main aim of the Search sub-task is to find video segments relevant to a given textual query. This problem is an important part of the Spoken Content Retrieval [8, 10] research area, which has been emerging in recent years.

All experiments presented in this paper were conducted on the BBC Broadcast data. A total of 1335 hours of video was available for training and 2686 hours for testing. We exploited subtitles, automatic speech recognition (ASR) transcripts by LIMSI [6], LIUM [9], and NST-Sheffield [7], all available for the task. Detailed information about the task and data can be found in the task description [2].

## 2. SYSTEM DESCRIPTION

Based on the results of our previous experiments [3], we employed the Terrier IR system[1] and its implementation of the Hiemstra language model [5] with stemming and stopwords removal.

Two strategies were used for segmentation of the recordings: 1) we divided the video recordings into segments of fixed length and 2) we used segmentation system which employed Decision Trees (DT) [3]. This system makes use of several features including cue word n-grams (word n-grams frequently occurring at the segment boundary, e.g. "if", "I'm", "especially", "the") and cue tag n-grams (tag n-grams frequently occurring at the segment boundary, e.g. "VBP PRP VBG"), silence between words, division given in transcripts, and the output of the TextTiling algorithm [4]. For each word in the transcript, it decides whether the segment ends after this word or not. The created segments may overlap. The system was trained on the data from Similar Segments in Social Speech Task in MediaEval 2013 [11].

## 3. SYSTEM TUNING

Based on our previous experiments, we set the segment length in the fixed-length segmentation to 60 seconds and the shift between the overlapping segments to 10 seconds. The segment length applied in the segmentation system was tuned on the training data and set to 50 seconds and 120 seconds for the Search sub-task. We also experimented with post-filtering of the retrieved segments – we either used all the retrieved segments or we removed segments which partially overlapped with another higher ranked segment.

We also employed the metadata provided for the task. For each recording we extracted the title, episode title, description, short episode synopsis, service name and program variant and appended the text to each segment from that recording.

## 4. RESULTS

The results for the Search sub-task are given in Table 1. We present scores of six evaluation measures: Mean Average Precision (MAP), Precision at 5 (P5), Precision at 10 (P10), Precision at 20 (P20), Binned Relevance (MAP-bin), and Tolerance to Irrelevance (MAP-tol) [1].

Unsurprisingly, the best results are achieved in experiments using subtitles. Generally, most of the results obtained with the LIMSI transcripts are higher than the corresponding results with the LIUM and NST-Sheffield transcripts. The only exception are the experiments employing overlapping segments. The results with the NST-Sheffield transcripts are higher than the corresponding results with the LIUM transcripts.

In most of the cases, the concatenation of the segment with metadata improved the results, despite the drop in the P5 score for all types of transcripts. Apart from several values of P and MAP-bin for the LIUM transcript, the fixed-length segmentation outperforms the Decision Trees-based segmentation with 120-seconds-long segments. Though the 50-seconds-long segments created using Decision Trees notably outperform the fixed-length segments measured by MAP and precision-based measures, they are outperformed by the fixed-length segmentation using the MAP-bin and MAP-tol measures.

All measures, except the MAP-tol measure, are notably higher in the experiments in which we did not remove partially overlapping segments from the list of the retrieved segments. Due to the nature of these measures, it is not possible to distinguish, whether a user had already seen the retrieved segment or not. Therefore, all the relevant segments,

[1] http://terrier.org

*MediaEval 2014 Workshop,* October 16-17, 2014, Barcelona, Spain

| Transcripts | Segment. | Seg. Len. | Metadata | Overlap | MAP | P5 | P10 | P20 | MAP-bin | MAP-tol |
|---|---|---|---|---|---|---|---|---|---|---|
| Subtitles | Fixed | 60s | No | No | 0.4209 | 0.7933 | 0.7433 | 0.5950 | 0.3192 | **0.3155** |
| Subtitles | Fixed | 60s | Yes | No | 0.5127 | 0.7467 | 0.7267 | 0.6100 | 0.3433 | 0.3023 |
| Subtitles | Fixed | 60s | Yes | Yes | 4.3527 | 0.7867 | 0.7733 | 0.7683 | **0.4150** | 0.1459 |
| Subtitles | DT | 120s | Yes | No | 0.3692 | 0.7467 | 0.7133 | 0.6050 | 0.2606 | 0.2157 |
| Subtitles | DT | 120s | Yes | Yes | **16.3486** | **0.8400** | **0.8367** | **0.8433** | 0.3172 | 0.0515 |
| Subtitles | DT | 50s | Yes | No | 0.8028 | 0.7867 | 0.7667 | 0.6933 | 0.3199 | 0.2350 |
| LIMSI | Fixed | 60s | No | No | 0.3534 | **0.7133** | 0.6600 | 0.5317 | 0.2916 | 0.2633 |
| LIMSI | Fixed | 60s | Yes | No | 0.4725 | 0.6667 | 0.6633 | 0.5467 | 0.3160 | **0.2696** |
| LIMSI | Fixed | 60s | Yes | Yes | 4.3000 | 0.6733 | 0.7133 | 0.7400 | **0.3822** | 0.1344 |
| LIMSI | DT | 120s | Yes | No | 0.3750 | 0.6933 | 0.6600 | 0.5383 | 0.2759 | 0.2054 |
| LIMSI | DT | 120s | Yes | Yes | **4.6366** | **0.7133** | **0.7300** | **0.7617** | 0.3706 | 0.1007 |
| LIUM | Fixed | 60s | No | No | 0.2836 | **0.6667** | 0.6067 | 0.4800 | 0.2227 | 0.2080 |
| LIUM | Fixed | 60s | Yes | No | 0.4371 | 0.6333 | 0.6400 | 0.5367 | 0.2651 | **0.2327** |
| LIUM | Fixed | 60s | Yes | Yes | 3.8328 | 0.6333 | 0.6767 | 0.6817 | 0.3180 | 0.1118 |
| LIUM | DT | 120s | Yes | No | 0.3538 | 0.6533 | 0.6300 | 0.5450 | 0.2659 | 0.2009 |
| LIUM | DT | 120s | Yes | Yes | **4.0709** | 0.6533 | **0.6800** | **0.6900** | **0.3345** | 0.0990 |
| NST-Sheffield | Fixed | 60s | No | No | 0.3279 | 0.6867 | 0.6467 | 0.5050 | 0.2646 | 0.2405 |
| NST-Sheffield | Fixed | 60s | Yes | No | 0.4645 | 0.6667 | 0.6600 | 0.5667 | 0.2974 | **0.2598** |
| NST-Sheffield | Fixed | 60s | Yes | Yes | 4.1241 | 0.6933 | 0.7000 | 0.7300 | **0.3560** | 0.1209 |
| NST-Sheffield | DT | 120s | Yes | No | 0.3627 | 0.6733 | 0.6567 | 0.5633 | 0.2624 | 0.2133 |
| NST-Sheffield | DT | 120s | Yes | Yes | **10.0198** | **0.7267** | **0.7533** | **0.7650** | 0.3342 | 0.0675 |

**Table 1: Results of the Search sub-task for different transcripts, segmentation types, segment lengths, metadata, and removal of overlapping segments. The best results for each transcript are highlighted.**

which frequently overlap each other, increase the score. The MAP-tol measure is not influenced by this behavior as it takes into account only the relevant content which had not been already seen by a user. Therefore, the highest MAP-tol scores are achieved for the fixed-length segmentation when the overlapping retrieved segments are removed.

## 5. CONCLUSION

In our experiments in the Search sub-task, we have experimented with subtitles and three ASR transcripts. The subtitles outperformed all used ASR transcripts. However, the LIMSI transcripts also generally scored well and they slightly outperformed the NST-Sheffield transcripts. The LIUM transcripts achieved the lowest scores in most of the cases. Moreover, we have confirmed usefulness of the metadata and effectiveness of simple segmentation into fixed-length segments.

We have also pointed out the problems with partially overlapping segments occurring in the results. Such segments can greatly increase MAP scores, however they could not be expected to be helpful for the users. Therefore, the MAP-tol measure could be preferred in such cases.

## 6. ACKNOWLEDGMENTS

This research is supported by the Czech Science Foundation, grant number P103/12/G084, Charles University Grant Agency GA UK, grant number 920913, and by SVV project number 260 104.

## 7. REFERENCES

[1] R. Aly, M. Eskevich, R. Ordelman, and G. J. F. Jones. Adapting Binary Information Retrieval Evaluation Metrics for Segment-based Retrieval Tasks. *CoRR*, abs/1312.1913, 2013.

[2] M. Eskevich, R. Aly, D. N. Racca, R. Ordelman, S. Chen, and G. J. F. Jones. The Search and Hyperlinking Task at MediaEval 2014. In *Proc. of MediaEval*, Barcelona, Spain, 2014.

[3] P. Galuščáková and P. Pecina. Experiments with Segmentation Strategies for Passage Retrieval in Audio-Visual Documents. In *Proc. of ICMR*, pages 217–224, Glasgow, UK, 2014.

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

[5] D. Hiemstra. *Using Language Models for Information Retrieval.* PhD thesis, University of Twente, Enschede, Netherlands, 2001.

[6] L. Lamel and J.-L. Gauvain. Speech Processing for Audio Indexing. In *Proc. of GoTAL*, pages 4–15, Gothenburg, Sweden, 2008.

[7] P. Lanchantin, P.-J. Bell, M.-J.-F. Gales, T. Hain, X. Liu, Y. Long, J. Quinnell, S. Renals, O. Saz, M.-S. Seigel, P. Swietojanski, and P.-C. Woodland. Automatic Transcription of Multi-genre Media Archives. In *Proceedings of SLAM Workshop*, pages 26–31, Marseille, France, 2013.

[8] M. A. Larson and G. J. F. Jones. *Spoken Content Retrieval: A Survey of Techniques and Technologies*, volume 5 of *Found. Trends Inf. Retr.* Now Publishers Inc., Hanover, MA, USA, 2012.

[9] A. Rousseau, P. Deléglise, and Y. Estève. Enhancing the TED-LIUM Corpus with Selected Data for Language Modeling and More TED Talks. In *Proc. of LREC*, pages 3935–3939, Reykjavik, Iceland, 2014.

[10] S. Rüger. *Multimedia Information Retrieval.* Synthesis Lectures on Information Concepts, Retrieval and Services. Morgan & Claypool Publishers, San Rafael, CA, USA, 2010.

[11] N. G. Ward, S. D. Werner, D. G. Novick, E. E. Shriberg, C. Oertel, L.-P. Morency, and T. Kawahara. The Similar Segments in Social Speech Task. In *Proc. of MediaEval*, Barcelona, Spain, 2013.