# DCU ADAPT @ TRECVid 2015: Video Hyperlinking Task

Shu Chen[1,2], Keith Curtis[1], David N. Racca[1], Liting Zhou[1,2], Gareth J.F. Jones[1], and Noel E. O'Connor[1]

[1] ADAPT Centre, School of Computing, Dublin City University
Dublin 9, Ireland
{kcurtis, dracca, gjones}@computing.dcu.ie
[2] INSIGHT Centre for Data Analytics, Dublin City University
Dublin 9, Ireland
shu.chen4@mail.dcu.ie, tintin.joo@gmail.com, noel.oconnor@dcu.ie,

**Abstract.** Automatic construction of inter-video hyperlinks between items of video content has the potential to support user navigation and browsing within video archives. We describe DCU ADAPT's participation in the TRECVid 2015 Video Hyperlinking task. This task combines both multimedia analysis and hyperlink construction. We provide a detailed description of our approach to the task and report results of our experimental investigation in terms of various hyperlinking evaluation metrics.

**Keywords:** Multimedia Hyperlinking, Data Fusion, Query Expansion, Information Retrieval

## 1 Introduction

Information retrieval (IR) systems provide tools to enable users to satisfy their information needs [3]. In the standard approach to IR, the user enters a search query expressing their information need, in response to which the IR system returns a number of items which may satisfy this information need. An alternative approach to query drive search is navigation in which users are able to follow links from a current item to other items of potential interest.

The TRECVid 2015 Video Hyperlinking Task is a benchmark to explore the dynamic creation of video-to-video hyperlinks from an anchor point in one video to regions of other videos which may be of interest to the viewer of the first video, enabling users to navigate easily between videos. The TRECVid 2015 task builds on the activities of the MediaEval workshop series which first proposed a video Search and Hyperlinking task in 2012 [8]. The TRECVid 2015 Video Hyperlinking task focuses on the automatic construction of effective hyperlinks across TV collections. The task simulates a scenario where a user is searching for a known segment in a video collection, and on occasion may find that this segment is not sufficient to address their information need or they may wish to explore other related video segments [8]. The distinguishing feature between video hyperlinking and other IR systems according to [10], is that video hyperlinking explains the source anchor by constructing hyperlinks to other resources.

Dublin City University (DCU) participated in the Search and Hyperlinking task at the MediaEval workshop from 2012 to 2014 [4–6]. In 2015 the ADAPT Centre @ DCU participated in the Video Hyperlinking task at TRECVid [11]. Our experimental investigation aimed to integrate knowledge gained from our existing experience in video hyperlinking, with further data fusion and IR methods to improve the hyperlinking system proposed in [5].

This paper is structured as follows: Section 2 describes our hyperlinking model. Section 3 describes our experimental investigation, and Section 4 concludes the paper.

## 2 Design of Hyperlinking Model

DCU's method for video hyperlinking at TRECVId 2015 follows on from our previous work carried out for the earlier MediaEval tasks [4–6]. It consists of three components: query anchors, target segments and hyperlinks. A query anchor is used as the input for our hyperlinking system. A target segment is a potentially relevant video clip for the query anchor identified using the hyperlinking system. A hyperlink indicates the relevance relationship between a query anchor and a target segment. The section introduces our strategies for each of these components.

### 2.1 Query Anchor Analysis

The TRECVid 2015 Video Hyperlinking task provides query anchors as the input query for the hyperlinking system. A query anchor contains the "start time", "end time" and "video name", without a clear description of the query content representing what it is about this content that users are interested in finding related information for. Thus, it is necessary to extract the query content to predict potentially interesting information. Our previous research [4–6] involved using the spoken transcripts to represent the query content.

In this paper, we apply a classic query expansion strategy proposed in [15] to enrich the query description beyond using the words in the ASR transcript between its start and end times. This procedure adds spoken terms from the regions of the transcript surrounding the query anchor. Our solution uses two fixed length segments before and after the query anchor as the query context. The spoken terms in the query context can be numerous, with only some of them having the potential to be beneficial to the hyperlinking process. Thus, we apply a filtering strategy to select terms which should be representative of this region of the anchor transcript. This strategy calculates a weight for each term in the query context, referred to as the "offer weight" [14]. The use of the offer weight for question expansion was introduced in [15], where the offer weight for term $i$ is calculated as:

$$\text{OfferWeight}(t_i) = r_i \cdot \text{Score}_{\text{relevance}}(t_i), \tag{1}$$

where $r_i$ indicates the number of relevant documents containing the term $t_i$. The $\text{Score}_{\text{relevance}}(t_i)$ is defined as:

$$\text{Score}_{\text{relevance}}(t_i) = \log \frac{(r_i + 0.5) \cdot (N - n_i - R + r_i + 0.5)}{(n_i - r_i + 0.5) \cdot (R - r_i + 0.5)}, \tag{2}$$

where $R$ is the number of documents relevant to the current query, $n_i$ is the number of documents in the data collection containing $t_i$, and $N$ is the size of the data collection [15]. This equation uses the value 0.5 to prevent potential issues of division by 0. A high offer weight means that a term has a lower document frequency in the whole collection and higher document frequency in relevant items. In our hyperlinking system, $n_i$ is the number of target segments containing a term in the query context. and $N$ the total number of target segments. We define $R = 1$, assuming that the query context is the single relevant document for the query anchor, and the value of $r_i$ is set as 1 due to $R = 1$. Thus, our methodology to create the hyperlinking query is: 1) extract the spoken terms in the query anchor as $T$; 2) extract the spoken terms in the query context, as $T_c$; 3) calculate the offer weight of each term in $T_c$; 4) select the top $K$ terms in $T_c$ according to their offer weights; and 5) combine $T$ with the top $k$ selected terms.

### 2.2 Target Segment Detection

The spoken transcripts provided by LIMSI as part of the task data [2] were used to represent spoken information. These transcripts were divided into sentences whose boundaries were determined using a combination of audio features and semantic information [2]. The start time of each sentence is used as the start point of a target segment. The sentences located from this start point for a duration of 90 seconds were combined to create the textual description of the target segments.

### 2.3 Hyperlinking Construction

For the MediaEval 2013 Hyperlinking task, our approach integrated video-level and segment-level features in an attempt to improve hyperlinking quality [5]. We proposed an assumption that if a complete video is relevant to a query anchor, a target segment extracted from this video may be relevant to this query anchor. Thus, based on this hypothesis, we created a hierarchical hyperlink model which divided the hyperlinking process into two steps: video-level and segment-level hyperlinking. The former searches for videos which are relevant to the query video and ranks each video in the collection by a similarity score. The second step identifies potential target segments relevant to the query anchor within each video. The final retrieval list is created by fusing the results from both video-level and segment-level hyperlinking. Mathematically, merging these two lists uses a standard data fusion process. In our previous work [5], we adopted a simple linear late fusion scheme as follows:

$$\text{Score}_{\text{fuse}} = w_v \cdot R_{\text{videolevel}} + w_s \cdot R_{\text{segmentlevel}}, \tag{3}$$

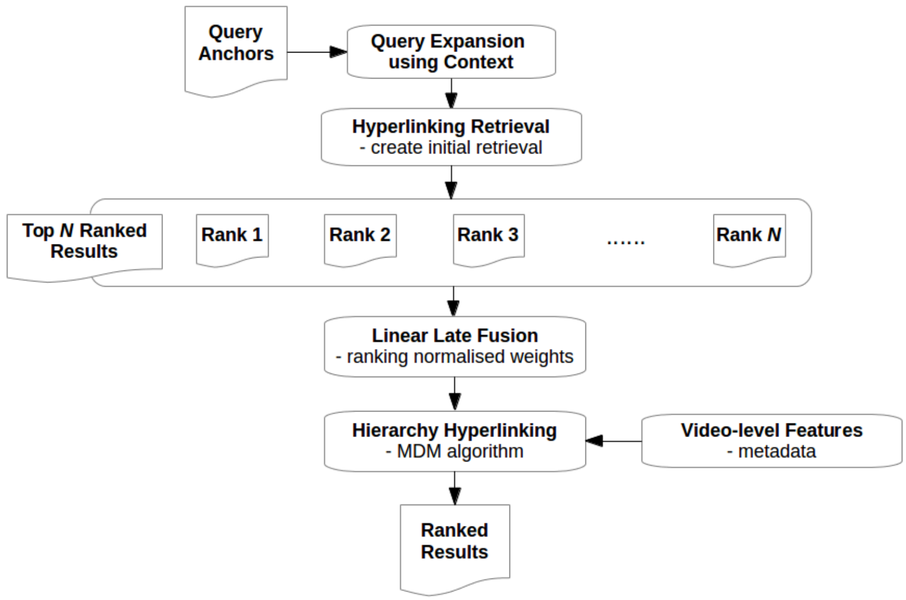


Fig. 1: DCU ADAPT hyperlinking model.

where $R_{\text{videolevel}}$ and $R_{\text{segmentlevel}}$ represent the ranked lists constructed by the video-level and segment-level hyperlinking respectively, and $w_v$ and $w_s$ are the corresponding scalar fusion weights.

Estimating the fusion weights is an open issue in the data fusion process. Our previous work [6] showed that using supervised learning to determine the fusion weights is unreliable,. This was illustrated by showing that this strategy can perform worse than using a simple solution of equal fusion weights. For our TRECVid 2015 runs, our hyperlinking investigation utilised an alternative Maximum Deviation Method (MDM) solution proposed in [16] to determine the fusion weights.

Figure 1 illustrates the hyperlinking system for our TRECVid 2015 submission. We follow our strategy, proposed in [4], of using Apache Lucene (4.9.0) to index and search both video-level and segment-level features. During the indexing stage, we use the Porter Stemming algorithm [12] to implement word stemming. Stop words are then removed based on the stopword list described in [4]. Next, we apply the strategy proposed in [7] to rerank the initial retrieval list. The system firstly uses query expansion to create the anchor query. An initial retrieval list is then created searching the set of target segments based on the LIMSI transcripts. This method assumes that the top $M$ retrieval results are relevant to the query. Taking the top $M$ results as $M$ new queries, the retrieval system creates $M$ retrieval lists. Linear fusion is applied to merge all $M + 1$ retrieved lists (the initial list and $M$ new retrieval lists). We assume that the merged ranked list will place better potential hyperlink target segments at a higher rank. This fusion process is defined according to:

$$\text{Score}_{\text{initial}} = \sum w_i \cdot \text{Score}(d, R_i), \tag{4}$$

where $R_i$ is the retrieved list using the $i$th result in the initial retrieve list, $\text{Score}(d, R_i)$ is the score of the document $d$ in $R_i$ and $w_i$ is the corresponding fusion weight.

We propose that the fusion weight should reflect the diversity of retrieved document at different ranks. Thus, a rank dependent weight for document $i$ is calculated as follows:

$$w_i = \frac{1}{r_d + 1} \tag{5}$$

where $r_d$ is the rank position of $d$ in $R_i$. The merged ranked list is then the final segment-level hyperlinking result. In our submission, we use MediaEval 2013 and 2014 data collections as the training set, and assign the value of $M$ as 15.

We use the metadata supplied with the videos to perform video-level hyperlinking. This metadata was created manually by BBC for each video. The metadata for each video includes multiple features, e.g. upload data, authors, description, etc. In our investigation, we only use the "description" feature, which constitutes one or two sentences which describe the primary focus of the video. Indexing and searching of the metadata again uses the same approach to search using Lucene as the the LIMSI transcripts.

The final output of our hyperlinking system is created by merging the video-level and segment-level hyperlinking results. The fusion weights are determined by applying the MDM algorithm.

## 3 Experimental Evaluation

The TRECVid 2015 Video Hyperlinking task used a dataset which was comprised of 3,520 BBC TV videos. The content collection was originally broadcast between 12th May, 2008 and 31st July, 2008. The average length of a video is roughly 45 minutes, and all videos are in the English language. The evaluation metrics include Precision@N (P@N), MAP [1], and MAiSP [13]. DCU submitted two runs: "tv15lnk-DCU-L-3-SsF-I-M-qexpmdm" and "tv15lnk-DCU-L-1-SsF-I-M-qexplatemdm".

- tv15lnk-DCU-L-3-SsF-I-M-qexpmdm (DCU-L-M-MDM): This method created the segment-level hyperlinking results directly using the expanded query without applying [7]'s method.
- tv15lnk-DCU-L-1-SsF-I-M-qexplatemdm (DCU-Llate-M-MDM): The method used the expanded query to create an initial retrieval list and then applied the fusion methods described in [7] to create the segment-level hyperlinking results.

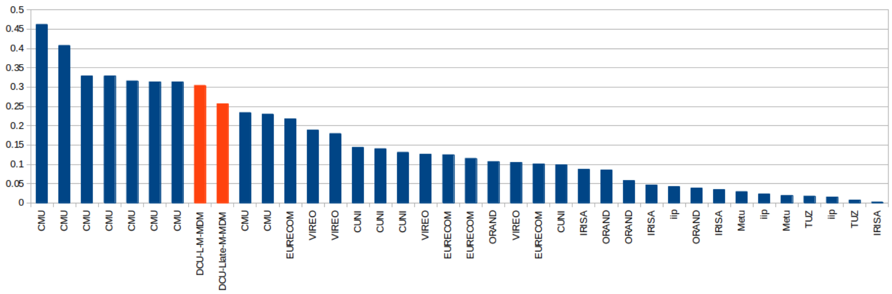


Fig. 2: Hyperlinking experimental evaluation results for all participants (MAP).

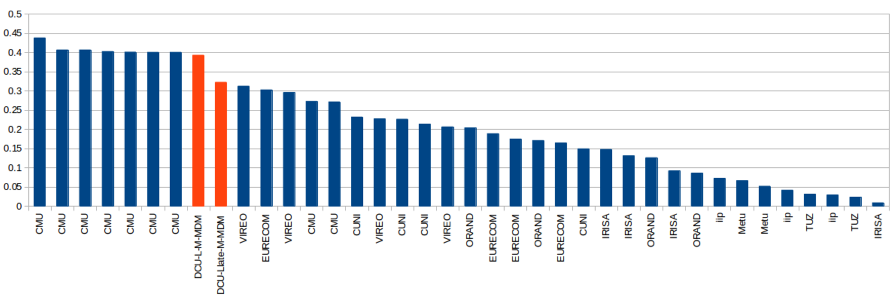


Fig. 3: Hyperlinking experimental evaluation results for all participants (P@20).

Figure 2 shows the MAP values for all participants submissions. Our best run, DCU-L-M-MDM, achieved MAP of 0.3044 and is ranked 8th of the 38 submitted runs. This run ranked 2nd of the 10 participating research groups. Compared with the run at position 1 submitted by CMU, the MAP of DCU-L-M-MDM was 0.1579 lower.

The results in Figure 2 show that DCU-L-M-MDM achieves relatively high precision compared with DCU-Llate-M-MDM. This means that re-ranking the initial ranked list retrieved by using the segment-level feature actually decreased the hyperlinking performance.

### 3.1 Analysis of Results

Figure 3 shows the P@20 values of all submitted runs. These results confirm our findings that reranking the initial retrieval list has a negative impact on hyperlinking creation, since DCU-Llate-M-MDM receives a relatively low value compared to DCU-L-M-MDM. The difference between the best run and DCU's best run is only 0.045 (from 0.4380 to 0.3930). Our results shown in Figures 2 and 3 demonstrate that the query expansion strategy produces a relatively high precision at rank 20.

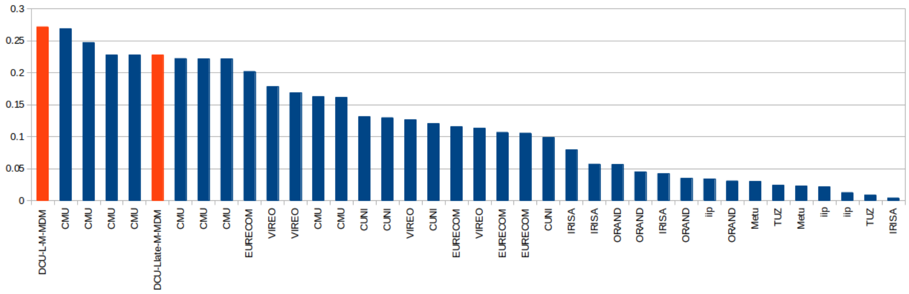


Fig. 4: Hyperlinking experimental evaluation among all participants (MAiSP). DCU's runs are marked as red.

Figure 4 shows that our DCU-L-M-MDM submission achieves the best MAiSP compared with all other runs. MAiSP is determined by the length of segments and the length of relevant period in the corresponding segment [9,13]. [9] indicates that low MAiSP means that a low percentage of relevant content in the retrieved segments. This metric takes the overlap between rank of retrieved results and relevant content in the retrieved segments into consideration, while the MAP metric only considers the rank of retrieved segments. Thus, the achievement of DCU-L-M-MDM in MAiSP demonstrates its advantage in efficiency of retrieving segments which cover the potentially interesting points in the relevant content marked in the ground truth.

## 4 Conclusions and Future Work

This paper has described DCU's submissions to the TRECVid 2015 Hyperlinking task. These continue our previous strategy of using a linear fusion scheme to integrate segment-level and video-level hyperlinking results in video search. We apply a query expansion strategy using the spoken terms in the query context to enrich the hyperlinking query. Moreover, we applied the reranking method described in [7]. According to the experimental evaluation, we conclude that this reranking method is not effective for this task, since DCU-L-M-MDM outperforms DCU-Llate-M-MDM in terms of all evaluating metrics. Compared with other participants, our run performs best in terms of MAiSP. These experimental results confirm that using query expansion and appropriate data fusion approach is a reasonable researching topic in multimedia hyperlinking. Thus, our future work will investigate how to analyse the hyperlinking query content. Since our experimental investigation involves only LIMSI transcripts and metadata, we also plan to extend our work to investigation of other multimodal features in the linking process.

**Acknowledgements** This research was supported by Science Foundation Ireland (SFI) as a part of the CNGL Centre for Global Intelligent Content (Grant No: 12/CE/I2267) at DCU within the ADAPT centre (Grant No. 13/RC/2106).

## References


1. R. Aly, M. Eskevich, R. Ordelman, and G. J. F. Jones. “Adapting Binary Information Retrieval Evaluation Metrics for Segment-based Retrieval Tasks”. *Computing Research Repository*, 2013.
2. S. Ayache and G. Queenot. “Video Corpus Annotation using Active Learning”. In *European Conference on Information Retrieval (ECIR)*, pages 187–198, Glasgow, Scotland, 2008.
3. R. Baeza-Yates and B. Ribeiro-Neto. *“Modern Information Retrieval”*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.
4. S. Chen, G. J. F. Jones, and N. E. O’connor. “DCU Linking Runs at Mediaeval 2012: Search and Hyperlinking Task”. In *Proceedings of the MediaEval 2012 Multimedia Benchmark Workshop*, Pisa, Italy, 2012.
5. S. Chen, G. J. F. Jones, and N. E. O’connor. “DCU Linking Runs at Mediaeval 2013: Search and Hyperlinking Task”. In *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop*, Barcelona, Spain, 2013.
6. S. Chen, G. J. F. Jones, and N. E. O’connor. “DCU Linking Runs at Mediaeval 2014: Search and Hyperlinking Task”. In *Proceedings of the MediaEval 2014 Multimedia Benchmark Workshop*, Barcelona, Spain, 2014.
7. S. Chen, K. McGuinness, R. Aly, N. E. OConnor, and T. Tuytelaars. “AXES @ TRECVid 2014: Instance Search”. In *Proceedings of TREC Video Retrieval Evaluation*, 2014.
8. M. Eskevich, G. J. F.Jones, S. Chen, R. Aly, R. Ordelman, and M. Larson. “The Search and Hyperlinking Task at Mediaeval 2012”. In *Proceedings of the MediaEval 2012 Multimedia Benchmark Workshop*, Pisa, Italy, 2012.
9. M. Eskevich, W. Magdy, and G. J. F. Jones. “New Metrics for Meaningful Evaluation of Informally Structured Speech Retrieval”. In *Proceedings of the 34th European Conference on Information Retrieval (ECIR 2012)*, pages 170–181, Barcelona, Spain, 2012.
10. R. J. F. Ordelman, M. Eskevich, R. Aly, B. Huet, and G. J. F. Jones. “Defining and Evaluating Video Hyperlinking for Navigating Multimedia Archives”. In *Proceedings of the 24th International Conference on World Wide Web*, pages 727–732, Florence, Italy, 2015.
11. P. Over, G. Awad, M. Michel, J. Fiscus, W. Kraaij, A. F. Smeaton, G. Queenot, and R. Ordelman. “TRECVid 2015 - An Overview of the Goals, Tasks, Data, Evaluation Mechanisms and Metrics”. In *Proceedings of TREC Video Retrieval Evaluation 2015*, 2015.
12. M. F. Porter. “An Algorithm for Suffix Stripping”. *Program*, 14:130–137, 1980.
13. D. N. Racca and G. J. F. Jones. “Evaluating Search and Hyperlinking: An Example of the Design, Test, Refine Cycle for Metric Development”. In *Proceedings of the MediaEval 2015 Multimedia Benchmark Workshop*, Wurzen, Germany, 2015.
14. S. E. Robertson. “On Term Selection for Query Expansion”. *Journal of Documentation*, 46:359–364, 1990.
15. S. E. Robertson and K. S. Jones. *“Simple, Proven Approaches to Text Retrieval”*. Computer Laboratory, University of Cambridge, 1994.
16. P. Wilkins. *“An Investigation into Weighted Data Fusion for Content-based Multimedia Information Retrieval”*. PhD thesis, Dublin City University, 2009.