Defining and Evaluating Video Hyperlinking for Navigating Multimedia Archives

Roeland J.F. Ordelman
University of Twente &
Netherlands Institute for
Sound and Vision
The Netherlands

Maria Eskevich
EURECOM
Sophia Antipolis
France

Robin Aly
Data Management
University of Twente
The Netherlands

Benoit Huet
EURECOM
Sophia Antipolis
France

Gareth J.F. Jones
ADAPT Centre / CNGL
School of Computing
Dublin City University, Ireland

ABSTRACT

Multimedia hyperlinking is an emerging research topic in the context of digital libraries and (cultural heritage) archives. We have been studying the concept of video-to-video hyperlinking from a video search perspective in the context of the MediaEval evaluation benchmark for several years. Our task considers a use case of exploring large quantities of video content via an automatically created hyperlink structure at the media fragment level. In this paper we report on our findings, examine the features of the definition of video hyperlinking based on results, and discuss lessons learned with respect to evaluation of hyperlinking in real-life use scenarios.

Keywords

Video hyperlinking, multimedia archives, digital libraries, media fragment, benchmark evaluation, crowdsourcing

1. INTRODUCTION

Users of the World Wide Web are very familiar with using hyperlinks to navigate from one information source to another. Links are associated with 'anchors points' in the text, and may be placed for a number of reasons. For example, to direct the reader to explanatory material, as in the case of Wikipedia links to concept definitions, or to related materials, such as related news stories. While anchors and links have traditionally been placed into text documents manually, work on automated Wikification [11] has initiated a line of work on creating link structures automatically. This paper adds to this line of work.

Besides the linking of text documents, there is growing interest in automated hyperlinking between elements of multimedia content. Hyperlinking research and development initiatives are emerging in the broadcast domain, digital libraries and multimedia (cultural heritage) archiving [4, 14]. Use scenarios include the exploration of additional information sources while accessing content in a linear fashion [8, 9, 13] or scenarios in which links to other relevant videos in a collection could help an end-user (e.g., general public, media researchers) to explore an audiovisual archive [12, 5, 20], or even one step further, to link video fragments for the purpose of creating narratives, e.g., in a journalist type of use scenario.

We have been studying video hyperlinking for a number of years within the context of a series of search and hyperlinking tasks within the MediaEval benchmark evaluation campaigns[1]. Our task specifically addresses automated video-to-video hyperlinking. Instead of considering video hyperlinking from a multimedia ontology perspective, we approach it as an audio-visual source data perspective. Taken in this way, a video archive consists of an infinite set of media fragments[2] that can be related to other media fragments based on multiple (semantic) representations of the multimodal information in the data that are not necessarily based upon ontologies. A typical use case we consider is the navigation through large quantities of locally archived or distributed video content via a link structure at the level of media fragments. This navigation can have an exploratory nature or be part of a storytelling framework that assembles related media fragments on the basis of a topic. We distinguish the video hyperlinking use case from others (e.g., recommendation, near-duplicate detection) by focusing on 'give me more information *about* this anchor' instead of 'give me more *based on* this anchor or entity'.

Taking this perspective, we envisage video hyperlinking systems to as being akin to video search systems: they extract a query representation from an anchor media fragment, apply this to a video search system, and identify potentially relevant fragments to present to the user. However, there

---

[1] http://www.multimediaeval.org/
[2] See http://www.w3.org/TR/media-frags/

are also notable differences between our conceptualization of video hyperlinking and user-based videos search. We distinguish four areas of comparison: (1) the creation of anchors in video hyperlinking – in relation to the creation of queries in video search, (2) anchor representation in the video hyperlinking framework – compared to query representation in the video search framework [19], (3) the practical use of the anchor representation in systems that provide relevant link targets – compared to the use of a query in video search, and (4) dealing with the results in search and video hyperlinking. These comparisons motivate the agenda of our search and hyperlinking evaluations within MediaEval and, in 2015, in the context of the TREC Video Retrieval Evaluation[3] (TRECVID) [15].

This paper is structured as follows: we first examine these four areas of the video hyperlinking framework in more detail and discuss related challenges in Section 2. In Section 3, we describe our strategy for involving end-users in the evaluation framework and focus on our experiences in (manual) anchor formulation and representation, addressing the specific nature of multimodal anchors created for video hyperlinking. In Section 4, we wrap up with some lessons learned from our work so far, and we conclude with a discussion on future steps in a benchmark evaluation context in Section 5.

2. DEFINITION AND CHALLENGES

2.1 Definition

In order to clarify the features of hyperlinking in the context of our work, we formally define the video hyperlinking process to be composed of the following steps:

1. Anchor Identification. FOr a given video (v), the anchor identification process (ai) determines a set of segments (identified by their start time (s) and end time (e)) for which users might benefit from a link to related information, formally:

$$ai(v) = [(s, e)*]$$

We define an anchor to be the triple of video (v), start time (s) and end time (e). Note that alternative definitions could model anchors as entities or words being said or use a spatial dimension (particular areas of the video in order to allow clicking), or some multimodal combination of these. We choose this simpler definition here since our work is still at an exploratory stage in terms of the defintion of video hyperlinking.

2. Anchor Representation. The anchor representation process (ar) takes an anchor (v, s, e) and generates a query q input for a video search system. This seeks to identify relevant information for the content of the anchor:

$$ar(v, s, e) = q$$

The anchor representation sstage may be composed of multiple steps. For example, an algorithm may first extract entities which appear within the video segment as a representation of its semantics and form a query out of the appearing entities. Note that q can, and often will be, multimodal. It can therefore contain parts that

___
[3]http://trecvid.nist.gov/

address the audio, the visual channel and the manual metadata associated with the video. Furthermore, in a more general definition, the query generation process could yield multiple queries if the anchor contains multiple facets that should be linked. This is similar to ambiguous queries, currently investigated in text retrieval (see e.g., [1]), where text queries represent different information needs. For the purposes of our current studies we focus on anchors that are unambiguous.

3. Target Search. The target search process (S) takes the query generated by the query generation process (stage 2) and produces a ranked list of video segments, defined by video v, start time s and end time e, as well as a score Sc:

$$S(q) = [(v, s, e, Sc)*]$$

4. Target Presentation. The target presentation process selects parts of the generated ranked list of search results for presentation to the user. We model target presentation as a separate stage since we envisage that this component of a video hyperlinking system may be more complex than merely presenting a list of highest ranked results to the user.

2.2 Anchor Identification

When linking text documents, the key challenges are first, to identify anchors for potential links, and second, to select relevant content. In an automatic approach to *textual* hyperlinking, the anchor text acts as a search query for textual content to be linked to. Both query and content are represented in the same modality - text. One of the main issues when switching to the audio-visual domain, is that both the anchor and the link targets are *multimodal* in nature.

This mulitmodality means that the process of the anchor definition in video is more complex for audio-visual hyperlinking. This complexity arises both due to the nature of the media, e.g. it is not clear what visual features should be extracted from a video fragment for use in the query, and in the combination of visual, audio and text features within the query and the subsequent search process (e.g., [17]). Ideally anchors should be automatically identified and analyzed, but this is itself currently a research challenge, and within the framework of the MediaEval Search and Hyperlinking task, we have so far experimented with a completely manual scenario in which users define anchor segments within a video by providing a start time and an end time. This is not a task with which current users are generally familiar with and selecting 'interesting' segments in a video to form anchors is new to them, and they will probably not be fully aware of the concepts of multimodality as applied in video search. Our task for MediaEval 2015 will include an exploratory sub-task on automatic anchor identifcation within broadcast video.

2.3 Anchor Representation

After having created an anchor the multimodal information incorporated within it has to be translated into a query representation including relevant textual and visual features. This is then fed into a video search system as a (structured) query. In addition to being defined as a media fragment with specific start and end time, an anchor may incorporate the anchor's context such as the video from which it is taken as

a whole or a window of a certain length before and after the anchor. In addition to anchor content, the query representation can makes use of any available archival metadata[4], including subtitles and/or features extracted automatically from the audio and visual channels, e.g., speech transcripts and visual features. It is important to note that automatic analysis may result in errors or noise in the extracted features, Of course, if sufficient resources are available the audio and visual content of the anchor can be described manually. The form and detail of such manual annotation could be enforced via a suitable design of interface.

2.4 Target Search

Once a query has been extracted from an anchor, it can be used to search for video segments that are suitable as targets for a hyperlink. The key question is what 'suitable' means, how should relevance be interpreted in the context of video hyperlinking, and related to this, what should a video hyperlinking system actually be doing while 'searching' for relevant link targets. Although the relevance should ultimately be assessed by users, it is important to have a clear understanding of what a novel and, from a user perspective, unknown scenario such as video hyperlinking is aiming at in order to set-up an evaluation framework that enables advancing our understanding of it.

Our current working hypothesis is that given a multimodal anchor representation, the goal of searching in a video hyperlinking context is to find content that is *about* what is represented in the anchor – we sometimes refer to this as 'topically related' – and not content that is *based upon* it, which is *similar* to it, or has identical semantic labels. One important implication of this is that the context of the anchor in the 'about' case is of significant importance, whereas in the 'similarity' case, one could argue that it is the other way around.

In current state-of-the-art video search systems, such as the one we use in the search task that precedes the anchor creation process in our evaluation framework ([18], see section 3.2 below), multimodal queries are (potentially) either broken down and taken up by individual system components (e.g., face recognition, visual concept detection, speech/speaker recognition, etc.) by means of parsing a full text query, (e.g., extracting named entities) or by asking the user to manually split a multimodal query into parts via an advanced search type of interface with different fields (for e.g., archival metadata (text), spoken content, and various visual search options).

In a video hyperlinking system it is the anchor representation that needs to be broken down into parts that can be taken up by the appropriate system components. In addition, the outputs of these individual components need to be combined in a ranking function in such a way that it reflects the goals of the video hyperlinking scenario.

2.5 Target Presentation

The next challenge in the process is how the list of potential hyperlink targets which emerges from the search process will be used in a real-life hyperlinking scenario. Probably the presentation of a ranked, but potentially still extensive, list of 'search results', as in the video search scenario, will not be the most appropriate format to serve as an instrument to navigate through a linked video structure. Instead, one would expect that a hyperlinking system should optimize its precision on a relatively small top-level part of the list. Also, the multimodality of the anchor may provide a clue for the organization of search results. For example by clustering results on the basis of visual or audio cues, or by taking other characteristics of the anchor into account, such as its type, which we currently define as being based on the whole scene, the speech, a moving object, a static object, or music.

The design of an evaluation framework for the video hyperlinking scenario, and especially the expression of the tasks of anchor creation and anchor target relation assessment by target end users, demands a deep understanding of multiple dimensions of the content and content relations. In the following sections we describe how we engaged with target end users and relate our findings to our ongoing efforts to improve the task design.

3. STUDYING END-USER BEHAVIOUR

3.1 Overall task design

Participants in the MediaEval 2013 Search and Hyperlinking benchmark evaluation task benchmark evaluation task were asked to automatically generate potential link targets on the basis of anchors created manually by potential end users. The anchors and link targets were taken from a collection of 1, 260 hours of BBC broadcast video. The average length of a video was roughly 30 minutes and most videos were in the English language [6]. Along with the video, participants were provided with subtitles and automatic speech recognition (ASR) transcripts, archival metadata and automatically identified visual shot boundaries with a single visual key-frame extracted for each shot. For each anchor participants were required to return a ranked list of the most likely target video segments for given the anchor from within the collection of broadcast videos. Participants prioritized one of their submitted runs to be used in the second stage of the user study in which we created a pool of the top 10 ranked results submitted by each participant for manual relevance assessment. This resulted in 2081 anchor video target video fragments to be evaluated, reduced to 2078 after duplicate removal.

3.2 User Studies

The creation of the MediaEval 2013 Search and Hyperlinking task included two types of end user study: 1) *anchor identification* (as input for the automatic hyperlinking search task) and 2) *relevance assessment* of the results of the search task submitted by the participants. Our underlying use case envisaged a scenario with 'public' users engaging with the exploration of large audio-visual (broadcast) archives using a video hyperlinking approach, To identify a cohort of suitable potential users, we employed a recruitment bureau to select users (30 in total) of varying age and background to participate in the anchor selection session. We asked the same users to participate again for the assessment part of the study. We required these users to be representative of the general population of Internet users: the recruitment bureau selected participants by age group (16 to 30), with a high computer familiarity level (as reported

---

[4]archival metadata here refers to the metadata manually created in the production and archive workflow of the content, such as title, summary details, keywords, and other feature, e.g. people appearing in the video, location, etc. In practice archival metadata is often sparse.

Figure 1: Amazon MTurk HIT example that contains the anchor and target videos (1 and 3), anchor explanation given by the participants (2), the field to express the decision on the hyperlink relevance (4), and the field for the explanation of this decision (5). The transcript questions (6) and notes are used to assess general quality of the submitted work.

by themselves), and as having an active online presence on a regular basis. In addition, they were all UK citizens speaking English as their native language, which was a requirement given the BBC content used in our evaluations.

In parallel to the local team of target users in the relevance assessment phase, we used 'workers' from the Amazon Mechanical Turk (AMT) crowdsourcing platform [5] to increase the scale of the assessment of relevance assessments. By having both assessments from a local, controlled group of target users, and users of the crowdsourcing platform, potential discrepancies between judgments, could help us to validate assessments, and to gain insights into potential differences in judgments between groups of assessors. For example, contradictions between the judgements of the anchor creator and an AMT worker on the relevance of an identified target might suggest that the anchor was only relevant personally to the specific user and not general enough. On the AMT platform we do not have reliable access to information about the workers, but, based on our worker selection process, we assume that the users that who come to work on such platforms reliably represent the target audience of our task scenario. To improve the intrinsic involvement of users in the study, we gave them a monetary compensation: at a per hour rate to the recruited users performing the task at our premises, and per task rate to the AMT workers.

3.2.1 Anchor Identification by End-Users

Previous work suggests that it may be difficult for users to identify hyperlinks in material that they are not genuinely interested in [2]. So instead of providing participants with a video for the anchor selection task, we set it within the context of a search scenario in which anchor identification took place as a second step of a two stage activity. Participants watched a video clip corresponding to a relevant search result (see [3] for a more elaborate description). For the search part we provided them with a prototype audio-

visual search system[6] that permits both visual and textual based search. Although the participants expressed their interest in visual search facilities verbally, we observed that most of them actually preferred search options based on text features (speech transcripts (subtitles or ASR output) and archival metadata). However, since they were able locate interesting videos to start the anchor creation task, we see no grounds to believe that this behaviour has an impact on the outcome of our experiment. It is interesting to note though that in the query formulation process in a standard search task, users appear not to take intuitively to the use of unfamiliar search options for different modalities, which may influence the way they approached the anchor creation task. On the other hand, it could well be that the visual search facilities just did not work that well in this task given the relatively small data set (successfully detected visual concepts are sparse) for the queries of users that are not familiar with state-of-the-art visual search options.

For a given relevant clip found in the search process, we asked the user to identify anchors that thwy would want to link to related video content using a virtual cutter tool which enabled them to define a starting point and an end point for a fragment that s/he would like to have linked to additional information. For these anchors, we also asked the user to give a textual description of what was contained in the anchor (e.g., "about a world famous singer and his relation with organized crime"), main characteristics of the anchor as starting point for linking –the whole scene, the speech, a moving object, a static object, or the category 'other' that could be used for music as the main characteristic for an anchor–, and a description of what they expected to see in the link targets (e.g. "mafia clips; connections between mafia and other singers/famous people").

We found that participants created anchors that referred primarily to spoken content and whole scenes, anchors referring to visual objects were under-represented. This finding seems to be inline with the user behaviour during the search

---

stage, favouring textual searches over visual searches. In total, the session resulted in 98 anchors. Since we had only a comparatively small video collection available from the BBC archives, this could have resulted in sparseness in the linking results in terms of the number of relevant links for each anchor. In order to avoid this possible effect, we filtered the anchors to produce an evaluation test set. To do this, we used the same prototype audio-visual search system as the user study participants to check whether it was possible to find more than one relevant segment within the collection for the given target descriptions. We also checked how these descriptions were as queries, on the scale from 1 to 5, from being an easily found audiovisual concept to a topic that requires elaborated search. This filtering resulted in a set of 30 anchors for the evaluation task. It is interesting to note that these filtered anchors consisted only of the two types most common type in the evaluation set, i.e. "whole scene" and "speech".

3.2.2 Search Assessment by End-Users

Using the submissions of the task participants, we gathered two sets of ground truth relevance assessments of proposed hyperlinks. Firstly, we formulated an Human Intelligence Task (HIT) on the platform to judge the relevance of a link target, see Figure 1. Secondly, we carried out a locally controlled user session, where participants of our original user trial in which the anchors were created (with a few exceptions as some were not able to return) assessed the relevance of the links create for their anchors. To allow comparisons between the two groups of judgments, we provided AMT workers and local users with the same task, except that the AMT workers were required to answer an additional question specifically designed for the detection and filtering of improper work submissions. This question is shown as question 6 in Figure 1. Given an anchor, all participants were provided with target video excerpts and the textual description that participants provided in the first session about potentially relevant target content. We edited the text of the anchor descriptions by creating sentences in order to have coherent natural language sentences across all the anchors, without changing the original keywords.

In an earlier set-up of the crowdsourcing task (the Search and Hyperlinking Task at MediaEval 2012 [7]) that used video with a creative common license, the workers had access to the video context of both anchor and link target for evaluation. The interface had pointers to the anchor/target segments start and end times, and assessors were potentially able to check the context of both videos. As the license for the BBC video collection only allowed us to show excerpts of videos on the AMT platform, the workers could not access the context of both videos, and their decision on the relevance was based only on the target segments they were exposed to. The local participants at our premises had no such restrictions.

4. FINDINGS AND REFLECTIONS

On the basis of the analysis of the data from the evaluation set-up (anchor identification and search assessment) and from the participants in the benchmark evaluations, a number of observations can be made. These can be taken as lessons learned during the design of the video hyperlinking benchmark evaluations, specifically to enhance our approach towards the creation of 'gold-standard' anchors given a video
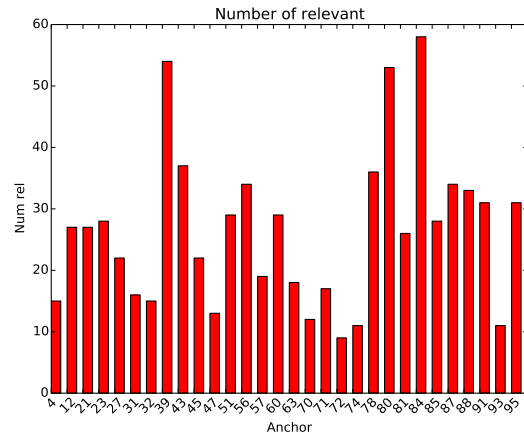

Figure 2: Total number of found relevant target links for 30 anchors using AMT relevance assessment.

and a certain use scenario, the development of mechanisms for anchor creation, and our strategy with respect to the assessment of link target relevance.

One important observation is that the sparseness of potential link targets for the user generated anchors in a limited video collection seems not to be a significant issue. Figure 2, based on further crowdsourcing relevance assessment of 34 submitted runs. This shows that participants find between 9-58 relevant targets per anchor. This indicates that the properties of the test collection we are using in terms of size and variety is sufficient for our goals.

Another observation concerns the agreement between end-users and AMT workers in their assessment of the link targets. We calculated the average number of (non)relevant target links that both end users and AMT workers agreed upon, and contrasted these values with the rest of the judgments that mismatched between two user studies. Figures 3 and 4 show that for both types of anchors even though the amount of (non)relevant links may vary, the level of disagreement stays around 21.5-21.6%. Although, a difference in agreement for relevant versus non-relevant could be expected (it is easier to judge about irrelevant results than results that are 'close'), these findings hint at design artifacts. For example, it is possible that the descriptions that go with the anchors as a reference for the AMT workers are not formulated clearly enough, a problem that the anchor creators themselves obviously do not have. This indicates that we have to re-design the tool that we are using in the anchor creation process in order to ensure that the mental process of creating anchors is annotated more richly, e.g., by forcing creators to label their actions in more detail. In this context, we also need to be aware of differences in the underlying intention of anchor creators: either from the perspective of *a content producer* generating anchors s/he thinks an end-user would be interested in, or from the perspective of *an end-user* wanting a certain anchor in the context of watching a video clip. As we have a pool of anchor creators from different categories – in additon to the general public, we also have access to content creators, journalists and researchers– we need to be sure that we separate and/or log different perspectives appropriately.
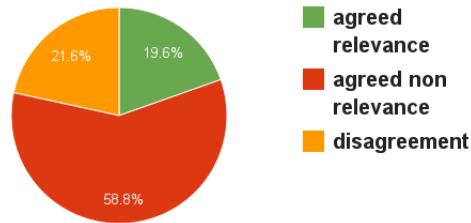
Figure 3: Average number of positive and negative relevance judgments that end-users and AMT workers agree upon, and the average level of disagreement, for the anchor type "whole scene"
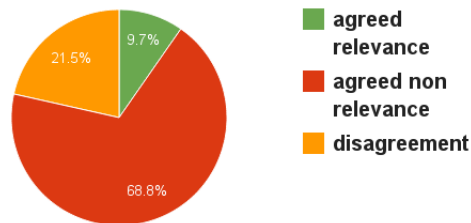

Figure 4: Average number of positive and negative relevance judgments that end-users and AMT workers agree upon, and the average level of disagreement, for the anchor type "speech"

5. CONCLUSION AND FUTURE WORK

In this paper we examined the definition and evaluation of video hyperlinking in the context of navigating multimedia archives. We provided a formal definition of the task and reported on the challenges that we have identified within four task component stages: anchor identification, anchor representation, target search, and target presentation. We described our strategy to involve end users in an evaluation framework and lessons learned from our work so far that will be taken up during our ongoing efforts to improve our evaluation set-up with respect to test collection creation and relevance assessment.

There are two main topics that we are planning to focus on in the near future. Firstly, we start with the design of an automatic anchor identification task as the first step towards automating both anchoring and target selection. Secondly, we will work on the elaboration of the theoretical framework of video hyperlinking. One important theme that is not yet adequately addressed is how our information retrieval perspective on video hyperlinking aligns with related work in the semantic web community and vice versa.

In the longer term, we are planning to set-up an evaluation framework that focuses on the fourth step in the video hyperlinking process, Target Presentation. Here, the focus will be on the evaluation of strategies that deal with issues in real-life video hyperlinking application scenarios with respect to the practical use of along list potentially of relevant targets given an anchor. Although the specifics of the application scenario are expected to play an important role here, examples one could think of are the clustering of targets, the elimination of near-duplicates [10], using narrative mod-

els, or thread-based visualizations such as the Fork Browser in [16].

6. ACKNOWLEDGMENTS

References

[1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In *WSDM '09: Proceedings of the Second ACM International Conference on Web Search and*, pages 5–14, New York, NY, USA, 2009. ACM.

[2] R. Aly, K. McGuinness, M. Kleppe, R. Ordelman, N. E. O'Connor, and F. de Jong. Link anchors in images: Is there truth? In *Proceedings of the 12th Dutch Belgian Information Retrieval Workshop (DIR 2012)*, pages 1–4, Ghent, 2012. University Ghent.

[3] R. Aly, R. Ordelman, M. Eskevich, G. F. Jones, and S. Chen. Linking inside a video collection - what and how to measure? In *Proceedings of the 22nd International Conference on World Wide Web Companion, IW3C2 2013, Rio de Janeiro, Brazil*, pages 457–460, Brazil, May 2013. ACM.

[4] M. Bron, B. Huurnink, and M. de Rijke. Linking Archives Using Document Enrichment and Term Selection. In *Research and Advanced Technology for Digital Libraries*, volume 6966, pages 360–371. 2011.

[5] M. Bron, J. van Gorp, F. Nack, and M. de Rijke. Exploratory Search in an Audio-Visual Archive: Evaluating a Professional Search Tool for Non-Professional Users. In *1st European Workshop on Human-Computer Interaction and Information Retrieval (EuroHCIR 2011)*, 2011.

[6] M. Eskevich, R. Aly, R. Ordelman, S. Chen, and G. J. Jones. The Search and Hyperlinking Task at MediaEval 2013. In *MediaEval 2013 Workshop*, Barcelona, Spain, October 18-19 2013.

[7] M. Eskevich, G. J. Jones, R. Aly, R. J. Ordelman, S. Chen, D. Nadeem, C. Guinaudeau, G. Gravier, P. Sébillot, T. de Nies, P. Debevere, R. Van de Walle, P. Galuscakova, P. Pecina, and M. Larson. Multimedia information seeking through search and hyperlinking. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, ICMR '13, pages 287–294, New York, NY, USA, 2013. ACM.

[8] A. Girgensohn, L. Wilcox, F. Shipman, and S. Bly. Designing affordances for the navigation of detail-on-demand hypervideo. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '04, pages 290–297, New York, NY, USA, 2004. ACM.

[9] P. Hoffmann, T. Kochems, and M. Herczeg. HyLive: Hypervideo-Authoring for Live Television. In M. Tscheligi, M. Obrist, and A. Lugmayr, editors, *Changing Television Environments*, volume 5066 of *Lecture Notes in Computer Science*, pages 51–60. Springer Berlin Heidelberg, 2008.

[10] J. Liu, Z. Huang, H. Cai, H. T. Shen, C. W. Ngo, and W. Wang. Near-duplicate video retrieval: Current research and future trends. *ACM Comput. Surv.*, 45(4):44:1–44:23, Aug. 2013.

[11] R. Mihalcea and A. Csomai. Wikify!: Linking Documents to Encyclopedic Knowledge. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (CIKM '07)*, pages 233–242, 2007.

[12] J. Morang, R. J. F. Ordelman, F. M. G. de Jong, and A. J. van Hessen. InfoLink: analysis of Dutch broadcast news and cross-media linking. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2005)*, pages 1582–1585, Los Alamitos, 2005. IEEE Computer Society.

[13] L. Nixon, M. Bauer, and C. Bara. Connected media experiences: web based interactive video using linked data. In *Proceedings of the 22nd international conference on World Wide Web companion*, pages 309–312. International World Wide Web Conferences Steering Committee, 2013.

[14] D. W. Oard, A. S. Levi, R. L. Punzalan, and R. Warren. Bridging communities of practice: Emerging technologies for content-centered linking. In *Museums and the Web*, 2014.

[15] P. Over, G. Awad, M. Michel, J. Fiscus, G. Sanders, W. Kraaij, A. F. Smeaton, and G. QuÃl'enot. Trecvid 2014 – an overview of the goals, tasks, data, evaluation mechanisms and metrics. In *Proceedings of TRECVID 2014*. NIST, USA, 2014.

[16] C. G. M. Snoek, K. E. A. van de Sande, O. de Rooij, B. Huurnink, J. R. R. Uijlings, M. van Liempt, M. Bugalho, I. Trancoso, F. Yan, M. A. Tahir, K. Mikolajczyk, J. Kittler, M. de Rijke, J.-M. Geusebroek, T. Gevers, M. Worring, D. C. Koelma, and A. W. M. Smeulders. The MediaMill TRECVID 2009 semantic video search engine. In *Proceedings of the 7th TRECVID Workshop*, Gaithersburg, USA, November 2009.

[17] D. Stein, E. Apostolidis, V. Mezaris, N. de Abreu Pereira, J. Müller, M. Sahuguet, B. Huet, and I. Lašek. Enrichment of news show videos with multimodal semi-automatic analysis. *Proceeding of the NEM-Summit, Istanbul, Turkey*, 2012.

[18] T. T. T. Tommasi, R. Aly, K. McGuinness, K. Chatfield, R. Arandjelovic, O. Parkhi, R. Ordelman, A. Zisserman. Beyond metadata: searching your archive based on its audio-visual content. In *IBC 2014*, Amsterdam, The Netherlands, 2014.

[19] H.-K. Tan, C.-W. Ngo, and X. Wu. Modeling video hyperlinks with hypergraph for web video reranking. In *Proceedings of the 16th ACM international conference on Multimedia*, MM '08, pages 659–662, New York, NY, USA, 2008. ACM.

[20] S. Tan, C.-W. Ngo, H.-K. Tan, and L. Pang. Cross media hyperlinking for search topic browsing. In *Proceedings of the 19th ACM international conference on Multimedia*, MM '11, pages 243–252, New York, NY, USA, 2011. ACM.