Beyond Shot Retrieval: Searching for Broadcast News Items Using Language Models of Concepts

Robin Aly[1], Aiden Doherty[2], Djoerd Hiemstra[1], and Alan Smeaton[2]

[1] Datbase Systems, University Twente, 7522AE Enschede, The Netherlands
{r.aly,hiemstra}@ewi.utwente.nl
[2] CLARITY: Center for Sensor Web Technology, Dublin City University, Ireland
{aiden.doherty,alan.smeaton}@dcu.ie

Abstract. Current video search systems commonly return video shots as results. We believe that users may better relate to longer, semantic video units and propose a retrieval framework for news story items, which consist of multiple shots. The framework is divided into two parts: (1) A concept based language model which ranks news items with known occurrences of semantic concepts by the probability that an important concept is produced from the concept distribution of the news item and (2) a probabilistic model of the uncertain presence, or risk, of these concepts. In this paper we use a method to evaluate the performance of story retrieval, based on the TRECVID shot-based retrieval groundtruth. Our experiments on the TRECVID 2005 collection show a significant performance improvement against four standard methods.

1 Introduction

Video search systems have usually concentrated on retrieval at the shot level, with a shot being the smallest unit of a video which still contains temporal information [16]. However not as much attention has been focused on searching for larger/semantic units of retrieval, generally referred to as stories. We believe that users may relate better to these retrieval units. However, current retrieval models for video search are difficult to adapt to these story retrieval units, since they are tailored to find shots, which are most often represented by a single keyframe. Therefore, the main contribution of this paper is a retrieval framework (based on language modelling of semantic concepts, for example a "Person", "Outdoor" or "Grass"), which can be applied to longer video segments, such as a news item.

Throughout this paper we assume that the user's information need is specified by a textual query. Given the growing prominence and attention afforded to lifelog data from wearable cameras such as the SenseCam, where audio data isn't recorded [3], we want our model to be also applicable to search in video data without considering the audio stream. As a result we focus on working with concepts extracted from the content of images. Current concept based video retrieval systems normally operate on a fixed-number of features per retrieval unit, for example the confidence scores of detectors for a number of concepts [6,

18]. Therefore, it is difficult to extend these models to search for news items of varying length.

To solve this, our approach uses an analogy to text retrieval and considers the frequency of a concept, in parallel to the frequency of a term, for ranking. If we knew the occurrence or absence of a concept in each shot of a news item we can determine its frequency simply by counting. However, because of the varying number of shots, the absolute concept frequencies are difficult to compare among items. Instead, we use them indirectly by calculating the probability that a concept is produced by a news item, following the language modelling approaches in text retrieval [8, 15].

After the definition of the above ranking function, we have to cope with two additional problems: (1) We have to identify which concepts to use for retrieval, because they are not necessarily named in the query text and (2) the occurrences of the concepts are uncertain. For (1) we adapt an approach from prior work which uses a labelled training collection to find useful concepts for shot search [1].

To handle the uncertainty (2), we propose the following procedure: Given the probabilistic output of the concept detectors, we calculate the expected language model score for an item with this distribution. That is, the score we expect for this item considering all possible combinations of concept frequencies. Furthermore, we include the standard deviation around the expected score, which expresses the associated risk of using the expected score as the correct score i.e. given that not all image concept detectors are entirely accurate, there is merit in boosting certain "risky" items in the ranked list. This is similar to the Mean-Variance Analysis framework from Wang [22], which considers uncertainty of scores in text retrieval.

The rest of this paper is structured as follows: In Section 2 we describe related work to this paper. Section 3 describes our ranking framework for news items. The experiments which show the effectiveness of our framework are described in Section 4. We finish this paper with conclusions and proposals for future work in Section 5.

2 Related Work

Firstly in Section 2.1 we present some background on how concepts are detected and on existing concept selection strategies for shot retrieval. Section 2.2 describes existing methods of multimedia retrieval for ranking general documents. Finally, in Section 2.3 we describe the Mean-Variance Analysis Framework, which was recently proposed by Wang [22], the principles of which we adopt in our way of treating the uncertainty of concepts.

2.1 Concept Detection and Selection

Concept Detection Snoek *et al.* [18] give a general overview of techniques in concept detection and retrieval. The features of each shot are mostly extracted from a single key frame. The predominant technique for concept detectors are

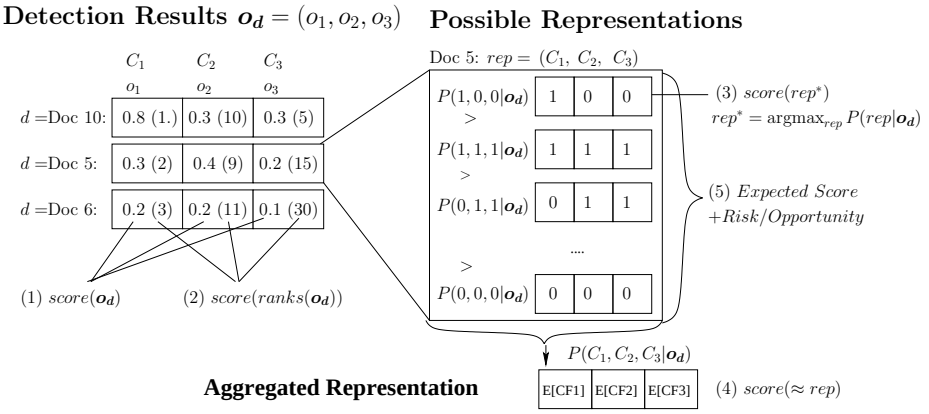


Fig. 1. Classification of multimedia retrieval methods (story confidence scores and and rank to the left, possible concept occurrences of shots within the story to the right): (1) Rank based, (2) Score based, (3) Top-1, (4) Expected Representation and (5) A combination of the expected score and the associated risk, which is proposed in this method.

support vector machines, which emit a confidence score o, describing their belief, whether a concept occurred in a certain shot. Furthermore, Platt presents in [14] a method to transform this confidence score into a probability.

Concept Selection As concepts are not necessarily named in textual queries, a retrieval system has to have a method which selects important concepts and assigns weights to these concepts. Hauff et al. use in [7] text retrieval on a collection of textual concept descriptions to select good concepts. Wikipedia articles and concatenated WordNet Glosses are investigated as description sources. The score of a text retrieval system is then used to measure the importance of the concept. Natsev *et al.* [12] propose a method called "statistical corpus analysis" which learns the useful concepts from the term-concept co-occurrences in the search collection. Our recently proposed concept selection is similar to the latter but uses the development collection, which has been used for detector training [1]. Here, the presence of the concepts is known and their importance to a query is determined by executing the text query on the textual representations of the development collection and by assuming the top-N documents to be relevant.

2.2 Retrieval Models in Multimedia IR

Because our model uses concept frequencies under uncertainty, it is related to two disciplines in multimedia retrieval: (1) concept based shot retrieval, which operates on the occurrence of concepts, and (2) spoken document retrieval, which considers the frequency of terms. Figure 1 demonstrates the relationships between the two disciplines, based on a retrieval scenario where concepts are used to return a ranked list of either shots or news items. To the left ("Detection

Results") we see the rank and confidence scores $\boldsymbol{o_d} = (o_1, o_2, o_3)$ for each concept occurrence. On the right ("Possible Representations"), we see the possible concept occurrences $C_1 - C_3$ in document 5.

We identify four different classes of how systems rank these documents, two for unit ranking and two for unit representation. Class (5) represents our proposed approach which combines the possible representations to the expected score and additionally considers the risk of using this score.

Confidence Score Based (1) Many approaches from this class are variations of the score functions $CombMNZ$ and $CombSUM$, which originate from meta search, see [2]. In parallel to class (1) score based approaches are only applicable to a fixed number of concepts. Another problem is the normalization and weighting of the confidence scores.

Component Rank Based (2) Systems of this class use methods such as the Borda Count method, which originates from election theory in politics. See Donald *et al.* [6] for the application in concept based video retrieval. However, this method is not directly applicable to longer video segments since it relies on the fixed number of concepts to rank.

Top-1 Representation (3) Systems in this class consider the most probable representation. Therefore, the system completely ignores other possible representations. While this technique was found suitable for retrieval tasks with high detection quality [21], Momou *et al.* report in [10] that under high word error rate, the performance deteriorates quickly.

Expected Component Value (4) More recent approaches in spoken document retrieval use the lattice output of the speech recognizer to obtain more variability [4, 5]. For example, Chia [5] calculates the expected term frequencies of the top-N documents, weighted by their probability. For high N and a linear score function this value approximates the expected score, which is one part of our ranking framework. However, there is no notion of the risk a system takes, if it uses this score.

2.3 Uncertainty in Text IR - Mean-Variance Analysis

Markowitz proposes in [11] the Portfolio Selection theory in economics. Wang [22] successfully transferred this theory into the Mean-Variance Analysis framework for uncertain scores in text information retrieval. According to Wang, a retrieval system should rank a news item at position m which optimizes the following expression, which considers "risk" of the item:

$$d^* = \underset{d}{\operatorname{argmax}}\, E[r_d] - b\, w_m\, \operatorname{var}[r_d] - 2b \sum_{i=1}^{m-1} w_i\, \operatorname{cov}[r_d, r_{d_i}] \qquad (1)$$

Here, r_d is the uncertain score of the news item d and the three ingredients of the framework are: (1) The expected uncertain score $E[r_d]$; (2) the variance

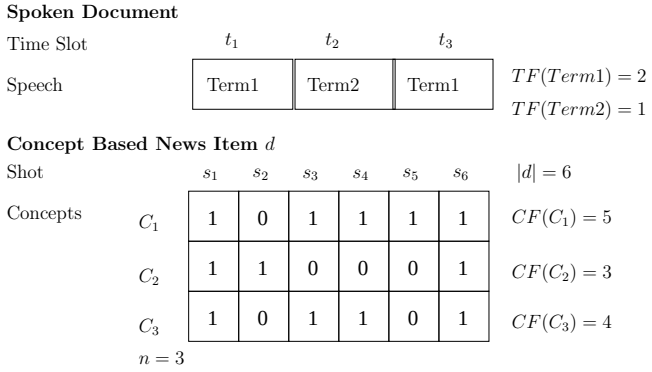


Fig. 2. A concept based news item representation and its analogy to a spoken document

var$[r_d]$ of the possible scores nearby $E[r_d]$; and (3) the covariance cov$[r_d, r_{d_i}]$, which specifies whether the scores of news item d and d_i are correlated. The parameter b represents the different risk affinities of different users. For $b > 0$ is the risk averse user who prefers to have a stable ranking, while for $b < 0$ he is willing to take more risk to have the chance of getting relevant news items to the top of the ranking. For $b = 0$ the system ranks by the expected score of each news item, which is the same as ranking by the original score function.

Essentially given the problem of the semantic gap in image retrieval, and hence the difficulty in producing highly accurate semantic concept detectors, we believe there may be merit in boosting the rank of more risky items to the top of the ranked list (given that the safe items may be proposed by mildly accurate concept detectors). Indeed this approach builds upon that proposed by Varian [20].

3 News Item Search

3.1 Concept Based News Item Representation

A news broadcast video can naturally be segmented into news items. Furthermore, these items can be subdivided into shots. Until now, this unit was used to present results to the user. Figure 2 shows how concept-based and spoken document-based representations would approach the problem of representing a news story. The spoken document consists of three spoken words at time position $t_1 - t_3$ and the news item of six shots $s_1 - s_6$. The concept lexicon consists of three concepts $C_1 - C_3$. We denote the presence of concept i in shot j as $C_{ij} \in \{0, 1\}$ where 1 stands for the presence of the concept. On the right, we see the term and concept frequencies as the count of the values on the left, as the analogy to spoken documents. We can express the frequency as a sum: $CF(C_i) = \sum_j C_{ij}$. For a given information need, we then select n important concepts $C_1, ..., C_n$ according to our prior work [1] and the vector of the concept frequencies $rep = (CF(C_1), ..., CF(C_n))$ is our document representation.

3.2 Concept Language Models

We now describe our ranking function for concept based news item retrieval. To our knowledge, this is the first proposal of a concept retrieval function for this domain. The basic idea behind our approach is to consider the occurrence and absence of a concept as two words of the language of this concept. Therefore a word is either "present" or "absent" and instead of a single stream of terms we have multiple "concept streams". As mentioned before, by simply counting we can get the concept frequency in a news item.

Because the concept frequencies between news items are difficult to compare we consider, in parallel to language modelling [8, 15], the probability that a concept is present in a news item. However, the extracted concepts will not fully reflect the content of the video. For example, since they are normally extracted at discrete points in time a concept detector may miss the occurrence of a concept in the news story as a whole. To solve this, we apply Dirichlet smoothing [23] to the probability and obtain the language model score for concept C_i in news item d :

$$P(C_i|d) = \frac{CF(C_i) + \mu P(C_i)}{|d| + \mu} \tag{2}$$

where $P(C_i)$ is the prior of encountering a concept C_i in the collection, $|d|$ is the length (in shots), finally μ is the scale parameter of the Dirichlet prior distribution. We now can rank news items by the probability of drawing a list of concepts independently from their "concept stream":

$$score(rep) = P(C_1, ..., C_n|d) = \prod_i^n P(C_i|d) \tag{3}$$

Here, the right part calculates the probability of sampling these concepts independently from the news item d.

3.3 Uncertain Concept Occurrences

Until now we have considered concept based news item search for the case of known concept occurrences. However in reality we will only have uncertain knowledge about their presence through the output of detectors. Let o_{ij} be the detector's confidence score that concept C_i occurs in shot s_j and $\boldsymbol{o_d}$ is now the combination of all confidence scores of an news item. For each concept in each shot of the news item this output can be transformed into a probability: $P(C_{ij} = 1|o_{ij})$. This probability can for example be estimated by a method described by Platt [14], which considers each concept occurrence independently. Work in the somewhat related domain of lifelogging has shown that the occurrences of many concepts within a shot and within adjacent shots is statistically dependent on each other [3]. However in this work we concentrate on more generic representations and leave the investigations of these dependencies for future work.

With this knowledge, we can determine the probability distribution over the possible frequency values $CF(C_i)$. For example, the probabiltiy that concept C_i has a frequency of 1 in a news item with $|d| = 3$ is

$$P(CF(C_i) = 1|\boldsymbol{o_d}) = P(\boldsymbol{C_i} = 1, 0, 0|\boldsymbol{o_d}) + P(\boldsymbol{C_i} = 0, 1, 0|\boldsymbol{o_d}) + P(\boldsymbol{C_i} = 0, 0, 1|\boldsymbol{o_d})$$

where $\boldsymbol{C_i}$ is a short form for (C_{i1}, C_{i2}, C_{i3}) and the first probability is calculated as follows:

$$P(\boldsymbol{C_i} = 1, 0, 0|\boldsymbol{o_d}) = P(C_{i1} = 1|o_{i1})(1 - P(C_{i2} = 1|o_{i2}))(1 - P(C_{i2} = 1|o_{i3}))$$

The probability of a whole representation is calculated as follows $P(rep|\boldsymbol{o_d}) = \prod_i^n P(CF(C_i)|\boldsymbol{o_d})$. Furthermore, the expected concept frequency of a concept, which is for example used by Chia*et al.* [5], can be determined as $E[CF(C_i)|\boldsymbol{o_d}] = \sum_{j=1}^n P(C_{ij}|o_{ij})$.

3.4 Retrieval under Uncertainty

We now describe how we combine the concept language score and the uncertainty of the concept occurrences into one ranking function. Because of the representation uncertainty, the document score is a random variable S_d.

Similar to the Mean-Variance Analysis framework from Wang [22], our framework now ranks by a combination of the expected score and the standard deviation of the score:

$$RSV(d) = E[S_d|\boldsymbol{o_d}] - b\sqrt{\text{var}[S_d|\boldsymbol{o_d}]} \tag{4}$$

Here, RSV(d) is the final score under which a document is ranked, $E[S_d|\boldsymbol{o_d}]$ is the score we expect from the distribution of concept occurrences and $\text{var}[S_d|\boldsymbol{o_d}]$ is the variance of the score around the expected value and specifies how dispersed the scores are. the risk factor was easier to control when considering the standard deviation[3] rather than the variance which was used in the Mean-Variance Analysis framework. The constant b specifies the risk perception of a system in the same way as in the Mean-Variance Analysis framework. We now define the expected score, the first component of our ranking framework:

$$E[S_d|\boldsymbol{o_d}] = \sum_{rep} score(rep) P(rep|\boldsymbol{o}_d) \tag{5}$$

That is, we iterate over all possible concept frequency combinations, calculate the resulting score and merge these scores according to their probability of being the right representation. Additionally the variance of the score, which represents the risk can be defined as:

$$\text{var}[S_d|\boldsymbol{o_d}] = E[S_d^2|\boldsymbol{o_d}] - E[S_d|\boldsymbol{o_d}]^2 \tag{6}$$

[3] standard deviation = square root of variance

$$\text{with } E[S_d^2|\boldsymbol{o_d}] = \sum_{rep} score(rep)^2 P(rep|\boldsymbol{o}_d) \tag{7}$$

While there is a very large number of possible representations ($2^{n|d|}$) in fully calculating Equations 5 and 7, we apply the Monte Carlo estimation method which samples from the given distribution. The method is defined as follows: Let $rep_{d_1}, ..., rep_{d_N}$ be random samples from $P(Rep|\boldsymbol{o_d})$. The expectations from Equation 5 and Equation 7 can then be approximated by:

$$E[S_d|\boldsymbol{o_d}] \simeq \frac{1}{N}\sum_{l=1}^{N} score(rep_{d_l}) \qquad E[S_d^2|\boldsymbol{o_d}] \simeq \frac{1}{N}\sum_{l=1}^{N} score(rep_{d_l})^2$$

To attain a random sample rep_{d_1} of a news item we iterate over each shot j and flip for each concept i a coin with the probability $P(C_{ij}|o_{ij})$ for head, the output of the concept detector. If we observe head (i.e. the probability is greater than a random number from the interval $[0:1]$), we add one to the concept frequency $CF(C_i)$ of this concept. After processing all concepts for all shots we calculate the score of the sample according to Equation 3. Because the standard error of the Monte Carlo estimate is in the order of $\sqrt{N}$ we achieve a relatively good estimate already with few samples.

4 Experiments

4.1 Experiment Setup

Our experiments are based on the TRECVID 2005 dataset which comprises 180 hours of Chinese, Arabic and English broadcast news [17]. NIST announced the automatic shot segmentation from Peterson [13] as the official shot boundary reference, defining a total of $45,765$ shots. For the segmentation of the videos into news items, we used the results from [9], which essentially looked for the anchor person in the video, to determine an item change. This segmentation resulted in $2,451$ news items of an average length of 118 seconds. We associated a shot with a news item, if it began within the time interval of the aforementioned news item. This resulted on average in 17.7 shots per news item.

Because of the novelty of our approach no standard set of queries with relevance judgments existed for this search task. Therefore, we decided on using the 24 official, existing queries from TRECVID 2005, replacing the *"Find shot of ..."* with *"Find news items about ..."*. Furthermore, we assumed that a news item is relevant to a given query, if it contained a relevant shot (which can be determined from the standard TRECVID groundtruth set). We argue that for most topics this is realistic since the user is likely searching for the news item as a whole, rather than shot segments within it.

We used the lexicon of 101 concepts and the corresponding detector set from the MediaMill challenge experiment for our experiments [19]. The reason for this is that it is an often referenced stable detector set with good performance on

the mentioned data set. As detailed above, we use a concept selection method to select important concepts for a query [1]. Therefore we executed the query with a standard text retrieval engine on a textual representation of the development set and assume the top-N documents to be relevant. We then used the first n most frequent concepts for this query, as we have used before [1].

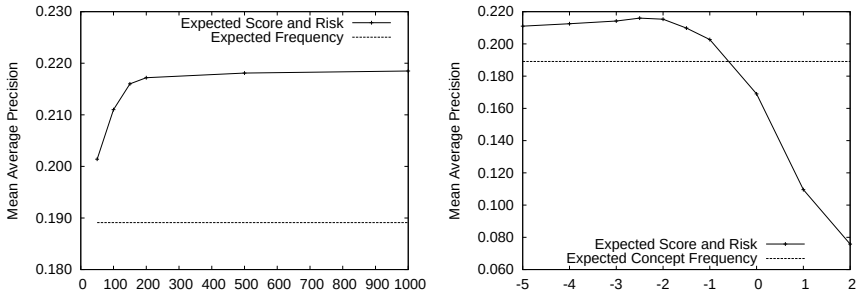
We compared our approach to four other approaches (classes 1-4 discussed in Section 2). As the approaches from concept based shot retrieval only work on a fixed number of features we used the average probability of each considered concept as the score for this concept $s(C_i) = \sum_j P(C_{ij}|o))/|d|$. To quickly recap, the considered approaches are: (1) Borda Count which considers the rank of the average concept occurrence probability [6], (2) CombMNZ which multiplies the scores as long as they are not zero [2]. (3) Top-1, which ranks the news items by the concept language model score of the most probable representation. To be more concrete, a concept occurrence was counted if the probability of the concept was above 0.5. The resulting concept frequencies were then used to calculate the concept language model score described in Equation 3. Finally, (4) we used an approach similar to that of Chia [5], which uses the expected concept frequency as the concept frequency in Equation 3.

4.2 Comparison to other Methods

Table 1 shows the result of the comparison of the described methods with our expected score method. The first row, n, beneath the class names indicates the number of concepts under which this class performed the best. We see that classes (1)-(3) perform much worse than the two methods which include multiple possible concept frequencies. Among them, there is only a small difference. For our method we used $N = 200$ samples, a Dirichlet prior of $\mu = 60$, and a risk factor $b = -2$. Since these parameters returned the best results while using few samples. To rule out random effects, we repeated the run ten times and report the average. Our method is significantly better that the expected frequency method and has a mean average precision of 0.214.

Table 1. Results of the comparison of our method (5) against four other methods described in related work. The actual methods are (1) Borda Count fusion, (2) $CombMNZ$, (3) Top-1, (4) Considering the expected frequency and (5) Our method of taking the expected score plus a risk expression. The MAP of our method has been successfully tested for significant improvement with a paired t-test with significance level 0.05, marked by $*$.

	(5) Expected Score + Risk	(1) Rank	(2) Score	(3) Top-1	(4) Expected Frequency
n	10	1	10	5	10
MAP	0.214$*$	0.090	0.105	0.094	0.192
P10	0.291	0.000	0.045	0.245	0.287



(a) Number of Samples N ($b = -2$) (b) Risk Factor b ($N = 200$)

Fig. 3. Robustness study of parameter settings.

4.3 Study of Parameter Values

In Figure 3 we summarise the result of a study over the two most important parameters in our model. Here we also set $\mu = 60$ and repeated each run ten times, to rule out random effects. In Figure 3 (a) the sensitivity of our method over the number of samples is shown. We see that already with few samples ($N = 50$) the performance is better than the expected concept frequency. As usual for a Monte Carlo estimator, the precision increases in line with the square root of the number of samples. After $N = 250$ samples we barely see any improvement.

Figure 3 (b) shows the behaviour of our model for changes of the risk parameter b. We see, with values of $b > -1$ our method performs worse than the expected concept frequency. The reason for this is that the concept detectors still show a low performance and therefore the variance of the concept frequencies can be quite high. A risk averse system will, in extreme cases, value the item with a certain score of practically zero higher than an item with still slightly lower score but a higher variance. This also makes clear why a positive risk perception increases performance.

5 Conclusions

In this work we proposed a ranking method for longer video segments than the commonly assumed retrieval unit of a video shot. Because of the novelty of the task we focused on the search for news items, a particular segment type. After identifying four major classes of ranking methods for general multimedia data, we found that current shot based methods are hard to adapt to longer video segments. Therefore, we proposed a new ranking function, a concept based language model, which ranks a news item with known concept occurrences by the probability that important concepts are produced by the item. However, since we only have probabilistic knowledge about the concept occurrence we included

this uncertainty in our ranking framework by considering the expected language model score plus the associated risk, represented by the standard deviation of the score.

We also proposed a means to creating a groundtruth for future story retrieval tasks, whereby we infer relevance from existing TRECVID judgements on shot-based retrieval. In the experiment which we performed on the TRECVID 2005 test collection, our proposed concept based language modelling retrieval method was able to show significant improvements over all representative methods from the identified classes.

We have shown that models which consider all possible concept frequencies perform better than systems that take only one of the scores, ranks, or most probable representations of documents into account. We have also shown that our method, which considers the expected score of a concept based language model, performs significantly better than an adapted method from spoken document retrieval (which takes the expected concept frequency and only then applies the concept based language model).

References

1. R. Aly, D. Hiemstra, and A. P. de Vries. Reusing annotation labor for concept selection. In *CIVR '09: Proceedings of the International Conference on Content-Based Image and Video Retrieval 2009*. ACM, 2009.
2. J. A. Aslam and M. Montague. Models for metasearch. In *SIGIR '01: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 276–284, New York, NY, USA, 2001. ACM.
3. D. Byrne, A. Doherty, S. C.G.M, G. Jones, and A. F. Smeaton. Everyday concept detection in visual lifelogs: Validation, relationships and trends. *Multimedia Tools and Applications Journal*, 2009.
4. C. Chelba and A. Acero. Position specific posterior lattices for indexing speech. In *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 443–450, Morristown, NJ, USA, 2005. Association for Computational Linguistics.
5. T. K. Chia, K. C. Sim, H. Li, and H. T. Ng. A lattice-based approach to query-by-example spoken document retrieval. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 363–370, New York, NY, USA, 2008. ACM.
6. K. M. Donald and A. F. Smeaton. A comparison of score, rank and probability-based fusion methods for video shot retrieval. In *Image and Video Retrieval*, volume Volume 3568/2005, pages 61–70. Springer Berlin / Heidelberg, 2005.
7. C. Hauff, R. Aly, and D. Hiemstra. The effectiveness of concept based search for video retrieval. In *Workshop Information Retrieval (FGIR 2007), Halle, Germany*, volume 2007 of *LWA 2007: Lernen - Wissen - Adaption*, pages 205–212, Halle-Wittenberg, 2007. Gesellschaft fuer Informatik.
8. D. Hiemstra. *Using Language Models for Information Retrieval*. PhD thesis, University of Twente, Enschede, January 2001.
9. W. H. Hsu, L. S. Kennedy, and S.-F. Chang. Video search reranking via information bottleneck principle. In *MULTIMEDIA '06: Proceedings of the 14th annual ACM*

international conference on Multimedia, pages 35–44, New York, NY, USA, 2006. ACM.

10. J. Mamou, D. Carmel, and R. Hoory. Spoken document retrieval from call-center conversations. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 51–58, New York, NY, USA, 2006. ACM.
11. H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.
12. A. P. Natsev, A. Haubold, J. Tešić, L. Xie, and R. Yan. Semantic concept-based query expansion and re-ranking for multimedia retrieval. In *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, pages 991–1000, New York, NY, USA, 2007. ACM.
13. C. Petersohn. Fraunhofer hhi at trecvid 2004: Shot boundary detection system. In *TREC Video Retrieval Evaluation Online Proceedings, TRECVID*, 2004.
14. J. Platt. *Advances in Large Margin Classifiers*, chapter Probabilistic outputs for support vector machines and comparison to regularized likelihood methods, pages 61–74. MIT Press, Cambridge, MA, 2000.
15. J. M. Ponte. *A language modeling approach to information retrieval*. PhD thesis, University of Massachusetts Amherst, 1998.
16. A. F. Smeaton, P. Over, and A. Doherty. Video shot boundary detection: Seven years of trecvid activity. *Computer Vision and Image Understanding*, 2009.
17. A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and trecvid. In *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, New York, NY, USA, 2006. ACM Press.
18. C. G. M. Snoek and M. Worring. Concept-based video retrieval. *Foundations and Trends in Information Retrieval*, 4(2):215–322, 2009.
19. C. G. M. Snoek, M. Worring, J. C. van Gemert, J.-M. Geusebroek, and A. W. M. Smeulders. The challenge problem for automated detection of 101 semantic concepts in multimedia. In *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*, pages 421–430, New York, NY, USA, 2006. ACM Press.
20. H. R. Varian. Economics and search. *SIGIR Forum*, 33(1):1–5, 1999.
21. E. M. Voorhees and D. Harman. Overview of the ninth text retrieval conference (trec-9). In *In Proceedings of the Ninth Text REtrieval Conference (TREC-9*, pages 1–14, 2000.
22. J. Wang. Mean-variance analysis: A new document ranking theory in information retrieval. In *ECIR '09: Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, pages 4–16, Berlin, Heidelberg, 2009. Springer-Verlag.
23. C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Trans. Inf. Syst.*, 22(2):179–214, 2004.