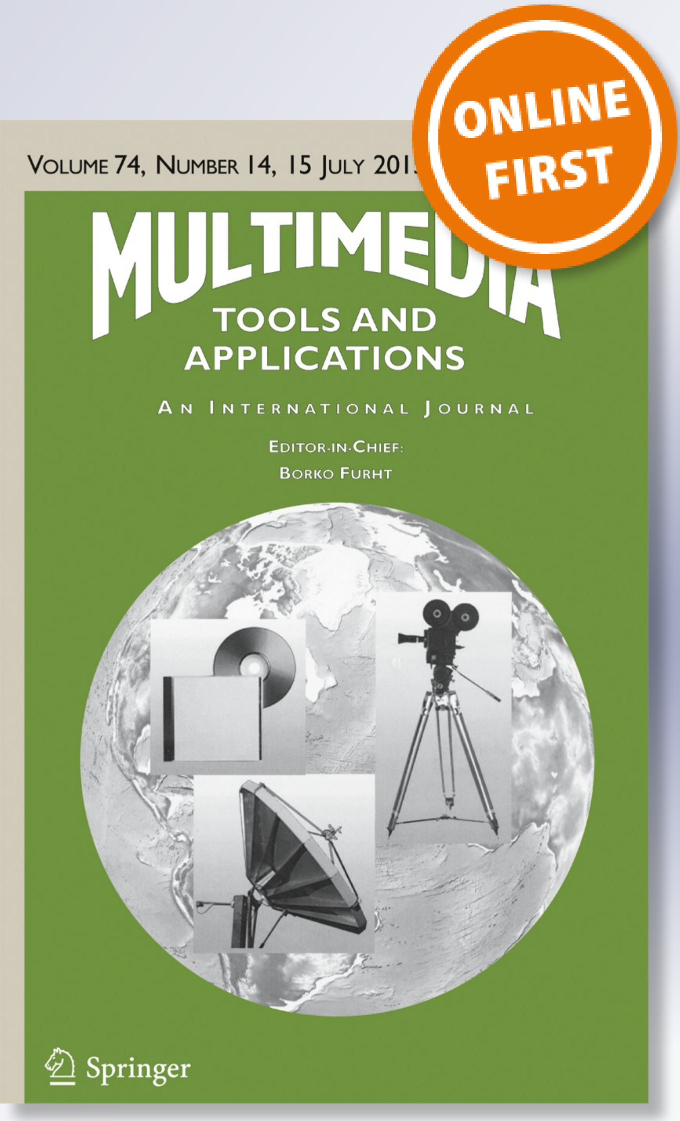
Knowledge based query expansion in complex multimedia event detection

Maaike de Boer, Klamer Schutte & Wessel Kraaij

Multimedia Tools and Applications
An International Journal

ISSN 1380-7501

Multimed Tools Appl
DOI 10.1007/s11042-015-2757-4



Springer

Multimed Tools Appl
DOI 10.1007/s11042-015-2757-4

Knowledge based query expansion in complex multimedia event detection

Maaike de Boer[1,2] · Klamer Schutte[1] ·
Wessel Kraaij[2,3]

Received: 28 November 2014 / Revised: 28 April 2015 / Accepted: 16 June 2015

Abstract A common approach in content based video information retrieval is to perform automatic shot annotation with semantic labels using pre-trained classifiers. The visual vocabulary of state-of-the-art automatic annotation systems is limited to a few thousand concepts, which creates a semantic gap between the semantic labels and the natural language query. One of the methods to bridge this semantic gap is to expand the original user query using knowledge bases. Both common knowledge bases such as Wikipedia and expert knowledge bases such as a manually created ontology can be used to bridge the semantic gap. Expert knowledge bases have highest performance, but are only available in closed domains. Only in closed domains all necessary information, including structure and disambiguation, can be made available in a knowledge base. Common knowledge bases are often used in open domain, because it covers a lot of general information. In this research, query expansion using common knowledge bases ConceptNet and Wikipedia is compared to an expert description of the topic applied to content-based information retrieval of complex events. We run experiments on the Test Set of TRECVID MED 2014. Results show that 1) Query Expansion can improve performance compared to using no query expansion in the case that the main noun of the query could not be matched to a concept detector; 2) Query

✉ Maaike de Boer
maaike.deboer@tno.nl

Klamer Schutte
klamer.schutte@tno.nl

Wessel Kraaij
wessel.kraaij@tno.nl

[1] TNO, Oude Waalsdorperweg 63, 2597 AK The Hague, The Netherlands

[2] Radboud University, Toernooiveld 200, 6525 EC Nijmegen, The Netherlands

[3] TNO, Anna van Buerenplein 1, 2595 DA Den Haag, The Netherlands

Published online: 12 July 2015

expansion using expert knowledge is not necessarily better than query expansion using common knowledge; 3) ConceptNet performs slightly better than Wikipedia; 4) Late fusion can slightly improve performance. To conclude, query expansion has potential in complex event detection.

Keywords Video event classification · Information retrieval · Knowledge bases · Zero-shot retrieval · Semantic analysis

1 Introduction

Retrieving relevant videos for your information need is most often been done by typing a short query in a video search engine such as Youtube [7]. Typically, such visual search engines use metadata information such as tags provided with the video, but the information within the video itself can also be extracted by making use of concept detectors. Concepts that can be detected include objects, scenes and actions [20]. Concept detectors are trained by exploiting the commonality between large amounts of training images. One of the challenges in content-based visual information retrieval is the semantic gap, which is defined as 'the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation' [42]. The importance of bridging the semantic gap is reflected by the emergence of benchmarks such as TRECVID [35] and ImageCLEF [8].

The semantic gap can be split in two sections [15]: the gap between descriptors and object labels and the gap between object labels and full semantics. Descriptors are feature vectors of an image and object labels are the symbolic names for the objects in the image. Full semantics is the meaning of the words in the query or even the intent of the user. The first gap is also referred to as *automatic image annotation* and progress is made rapidly [39, 43]. For the purpose of this paper the second gap is considered.

In the second semantic gap, the challenge is to represent the user intent in terms of the available object labels, which are provided by the concept detectors. State-of-the-art methods used to bridge this second semantic gap include query expansion using knowledge-bases [19] and relevance feedback [36]. Relevance feedback is a method that uses feedback from the user, such as explicit relevance judgments or user clicks, to optimize results. Relevance feedback is a powerful tool, but it requires an iterative result ranking process and dedicated algorithms [36], which is outside the scope of this paper. Another disadvantage of relevance feedback is that the system does not know why a video is not relevant.

Knowledge bases, on the other hand, are interpretable for both systems and humans. Knowledge bases can add more relevant words to the short user query to represent the user intent in a better way. This larger user query contains more words and, thus, more potential to match the object labels. Both common knowledge bases such as WordNet [23] or Wikipedia [16] and expert knowledge bases created by an expert can be used [2, 45]. Common knowledge bases are easy to access and do not require a lot of dedicated effort to construct, but they might not have sufficient specific information and they can be noisy due to disambiguation problems. The lack of sufficient specific information implies that no additional relevant concept detectors can be selected and the noise can cause selection of irrelevant concept detectors. Expert knowledge bases may have sufficient specific information and are less noisy, but it requires a lot of dedicated effort to create them.

Our research focuses on which type of knowledge base is best to use in the domain of complex or high-level events, defined as 'long-term spatially and temporally dynamic object interactions that happen under certain scene settings' [20]. Examples of complex events are *birthday party*, *doing homework* and *doing a bike trick*. In this paper, only textual information is used as input for the system, which it referred to as the zero-example case. In this situation it is unfeasible to create a dedicated detector for each possible word and we, therefore, have to bridge the semantic gap between the predetermined labels assigned to the image and the full semantics of the event. Complex events cannot be captured by a single object, scene or action description and, therefore, complex events have a large semantic gap.

In our experiments, we use the Test Set of TRECVID 2014 Multimedia Event Detection (MED) task [34] to compare retrieval performance on the complex event query, ConceptNet 5 [44] and Wikipedia as common knowledge bases and the textual description provided with the TRECVID task to determine which type of knowledge base is best to use. ConceptNet and Wikipedia are chosen, because both are easy accessible and provide information about complex events. We expect that query expansion has a positive effect on performance, especially if the main noun of the query cannot be detected with the available concept detectors. Because common knowledge bases are not tailored, expert knowledge bases might be able to outperform common knowledge. No difference in performance of ConceptNet and Wikipedia is expected. Fusion, on the other hand, is expected to increase performance, because not all knowledge bases will provide the same information.

In the next section, related work about the query expansion using knowledge bases and complex event detection is reviewed. The third section contains information about the method with the TRECVID MED task and design of the experiment. Section four consists of the results and the last section contains the discussion, conclusions and future work.

2 Related work

2.1 Query expansion using knowledge bases

One of the challenges in keyword search is that the user uses different words in the query than the descriptors used for indexing [4]. Another challenge is that users often provide a short, vague or ill-formed query [4]. In order to find relevant results, the query has to be expanded with relevant, related words, such as synonyms. Computers have no knowledge of our world or language themselves and, therefore, cannot use this information in the way humans do. In order to automatically expand the query without requiring the user to reformulate the query, computer systems should have access to world knowledge and language knowledge. One way to provide this knowledge is to use a knowledge base [4]. Two types of knowledge bases exist: common knowledge bases and expert knowledge bases. In [4] these are called *General World Knowledge Base* and *Domain Specific Knowledge Base*, respectively. Both types of knowledge bases are accessible on the Internet because of the Semantic Web and Linked Open Data (LOD) initiative [1, 40]. The Semantic Web is about exposure of structured information on the Web and the LOD is about linking the structured information. This information is often structured using an ontology, which is a formal way to represent knowledge with descriptions of concepts and relations. An advantage of using ontologies is that they provide a formal framework for supporting explicit, specific and machine-processable knowledge and provide inference and reasoning to infer implicit knowledge [3]. Several standards such as OWL (Web Ontology Language) are easy accessible. A disadvantage of an ontology is that the knowledge has to be inserted in the framework manually.

2.1.1 Common knowledge bases

Many common knowledge bases are available on the Internet and this section can, therefore, not include all available common knowledge bases. Many comparisons between common knowledge bases are available including [26] and [50]. The Linked Open Data initiative gave rise to using existing common knowledge bases in order to expand your own common knowledge base. One example is ConceptNet 5, which is a knowledge representation project in which a semantic graph with general human knowledge is build. This general human knowledge is collected using other knowledge bases, such as Wikipedia and WordNet, and experts and volunteers playing a game called *Verbosity*,[47]. Some of the relations extracted using this game are *RelatedTo*, *IsA*, *PartOf*, *HasA*, *UsedFor*, *CapableOf*, *AtLocation*, *Causes*, *HasSubEvent*, *HasProperty*, *MotivatedByGoal*, *ObstructedBy*, *CreatedBy*, *Synonym* and *DefinedAs*. The strength of the relation is determined by the amount and reliability of the sources asserting the fact. As of April 2012, ConceptNet contains 3.9 million concepts and 12.5 million links between concepts [44]. Experiments on the previous version of ConceptNet, which is ConceptNet 4, indicated that the knowledge base is helpful in expanding difficult queries [21].

Besides factual knowledge, common knowledge base Wikipedia contains encyclopedic information. Wikipedia is a free multi-lingual online encyclopedia edited by a large number of volunteers. Wikipedia contains over 4.8 English million articles. Both information on Wikipedia pages and links between the pages are often used [48]. An open source tool kit for accessing and using Wikipedia is available [29] and many other common knowledge bases include information or links from Wikipedia, such as YAGO2 [18] and ConceptNet [44].

Besides encyclopedic and factual knowledge bases, WordNet is a hierarchical dictionary containing lexical relations between words, such as synonyms, hyponyms, hypernyms and antonyms [28]. It also provides all possible meanings of the word, which are called *synsets*, together with a short definition and usage examples. WordNet contains over 155,000 words and over 206,900 word-sense pairs. WordNet is often used to expand a query with similar words [9] and several similarity measures can be used [37]. Most similarity measures use path-based algorithms.

The common knowledge base sources described above are easy to access, provide enough data for statistical analysis and do not require a lot of human effort to get results, but they might not have sufficient specific information or they might be noisy. Query expansion using these knowledge bases can also suffer from query drift, which means that the focus of the search topic shifts due to a wrong expansion [9]. Query expansion using common knowledge bases most often moves the query to the most popular meaning.

2.1.2 Expert knowledge bases

Besides many common knowledge bases, many expert knowledge bases exist such as in the field of geography [46], medicine [38], multimedia [30], video surveillance [12], bank attacks [13] and soccer [2]. Expert knowledge bases are domain-specific, because disambiguation, jargon and structure of concepts and relations is unfeasible in open domain. Expert knowledge bases are complete and have good performance in information retrieval tasks, but dedicated expert effort in creation of the ontology is a big disadvantage.

2.2 Complex event detection

Complex or high-level events are defined as 'long-term spatially and temporally dynamic object interactions that happen under certain scene settings' [20] or 'something happening at a given time and in a given location' [3]. Research regarding complex event detection and the semantic gap increased with the benchmark TRECVID. Complex events cannot be captured by a single object, scene, movement or action. Research mainly focused on what features and concept detectors to use [14, 30] and how to fuse results of these concept detectors [31]. The standard approach for event detection is a statistical approach to learn a discriminative model from visual examples. This is an effective way, but it is not applicable for cases in which no or few examples are available and the models cannot give interpretation or understanding of the semantics in the event. If few examples are available, the web is a powerful tool to get more examples [24, 27].

On the web, common knowledge bases can be accessed for query expansion in complex event detection. WordNet [28] is for example used to translate the query words in visual concepts [32]. Wikipedia is often successfully used to expand a query in image and video retrieval [19, 22]. A challenge with these methods is that common knowledge sources use text and many words are not 'picturable'. These words cannot be captured in a picture and are often abstract, such as *science*, *knowledge* and *government*. One approach to deal with this challenge is to use Flickr. Both Leong et al. [22] and Chen et al. [10] use Flickr to find 'picturable' words by using the co-occurrance of tags provided with the images resulting from a query. ConceptNet [44] has high potential, but it has not yet shown significant improvement of performance in finding a known item [50].

Expert knowledge bases are not often used in complex event detection. Two examples are the Large-Scale Concept Ontology for Mulitimedia (LSCOM) that contains a lexicon of 1000 concepts describing the broadcast news videos [30] and the multimedia ontology in soccer video domain [2]. The multimedia ontology consists of an ontology defining the soccer domain, an ontology defining the video structure and a visual prototype that links both ontologies. This visual prototype aims to bridge the semantic gap by translating the values of the descriptors in an instance of the video structure ontology to the semantics in the soccer ontology. This ontology is able to detect high-level events such as *scored goal*. Natsev et al. [32] show that in the TRECVID topic domain manual ontologies work on average better than automatic, which uses WordNet and synonymy match, and no query expansion. To our knowledge, the only expert knowledge base for complex events is used in [5] and this knowledge base is not publicly available.

3 Experiments

In our experiments, we compare three types of expansion methods in the field of complex event detection. The first expansion method is considered as our baseline and only uses the complex event query, which has one to four words, to detect the event. The second expansion method uses query expansion with a common knowledge base. We compare two common knowledge bases: ConceptNet 5 and Wikipedia. Both knowledge bases contain information about events, whereas many other knowledge bases only contain information about objects or facts. As opposed to our previous paper [6], WordNet is not used as a common knowledge base, but it is used in another way (see Section 3.2.2). The third expansion method uses query expansion with an expert knowledge base. To our knowledge, no expert knowledge base for our high-level complex events is available and we, therefore, use the textual description provided with the TRECVID Multimedia Event Detection (MED) task as expert knowledge source.

3.1 Task

The open and international TRECVID benchmark aims to promote progress in the field of content-based video retrieval by providing a large video collection and uniform evaluation procedures [41]. Its Multimedia Event Detection (MED) task was introduced in 2010. In the MED task, participants develop an automated system that determines whether an event is present in a video clip by computing the event probability for each video. The goal of the task is to assemble core detection technologies in a system that can search in videos for user-defined events [34].

In this research, two sets of TRECVID MED 2014 are used. The first set is called the Research Set and contains approximately 10.000 videos, which have a text snippet describing the video. The Research Set also has ground truth data for five events. The other set is the Test Set with 23.000 videos and ground truth data for twenty events. For each of the twenty events in the Test Set and the five events in the Research Set a textual description containing the event name, definition, explanation and an evidential description of the scene, objects, activities and audio is used.

The standard performance measure for the MED task is the Mean Average Precision [17]. Performance on the official evaluation of 2013 and 2014 show that complex event detection is still a challenge. In the case with no training examples, which is the representative case for this research, mean average precision is below ten procent.

3.2 Design

This section describes the design of the experiment, which is also shown in Fig. 1.

In the experiments, twenty complex events, such as *dog show*, *felling a tree* and *tailgating*, are evaluated. In this evaluation, a ranked list of all videos is created using the score of a video:

$$S_{e,v,em} = \sum_{c \in CD} \left(\frac{A_{c,e,em} \cdot W_{c,em}}{\sum\limits_{c \in CD} A_{c,e,em} \cdot W_{c,em}} \cdot \max_{k \in v}[CD_k] \right) \tag{1}$$

where $S_{e,v,em}$ is the score of video v for event e in expansion method em, c is a concept, CD is the set of concept detectors, $A_{c,e,em}$ is a binary variable denoting the availability of concept c in event query e in expansion method em, $W_{c,em}$ is the weight of the concept c in expansion method em, and CD_k is the concept detector value in keyframe k.

The name of an event is used as an input for the expansion methods. Each of the expansion methods create a list of weighted words. These weighted words are matched against the available concept detector labels. Our set of concept detectors is limited to less than 2000, so a gap between the words from the expansion methods and the concept detector labels exists. The matching step is, therefore, a filtering step.The value of $A_{c,e,em}$ is one for

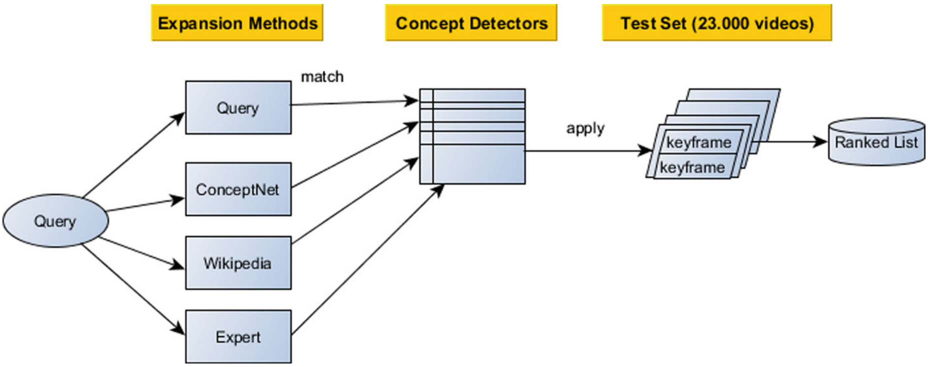


Fig. 1 Design

the selected concept detectors and zero for the concept detector that are not selected. In this way, only the values of the selected concept detectors are considered in the score. Additionally, the sum of the weights of the expansion method is one because of the division. The following sections describe this design in further detail.

3.2.1 Expansion methods

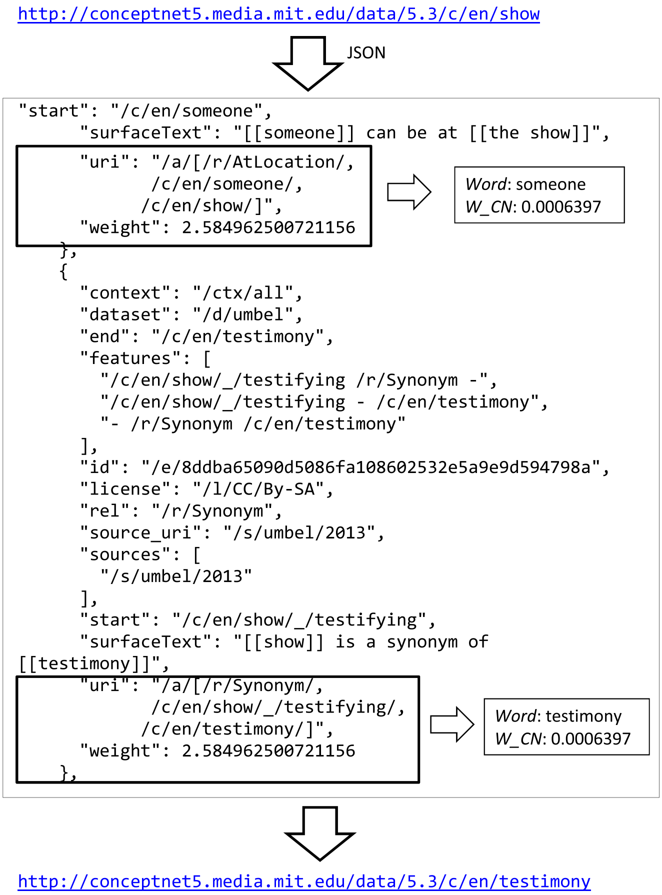
Complex event query The baseline method only uses the complex event query. The query is split into nouns and verbs with equal weights with a total sum of one. In the complex event query, nouns can be compound nouns such as *dog show*. If the compound noun cannot be matched against the available concept detectors (see Section 3.2.2), this noun is split into the separate words, in this case *dog* and *show*. This is also shown in the following formula:

$$W_{c,ceq} = \frac{1}{\sum N_c}, \tag{2}$$

where N_c is the number of concepts in the query

The weight of these words is the previous weight, in this example 1.0, divided by the amount of new words, which is two and, thus, results in a weight of 0.5 for *dog* and and 0.5 for *show*. Negative concepts are not taken into account, which means that the word *vehicle* is not matched against the available concept detectors in the event *winning a race without a vehicle*.

ConceptNet ConceptNet 5 [44] is used to expand the query. Because ConceptNet contains more knowledge about objects and activities compared to events, this expansion method is used to expand the nouns and verbs in the query that have no matching concept detector label. If no label was found in the query, we search for the whole event. For example, in the event *dog show* a concept detector with the label *dog* is present in our collection of concept detectors, but no label is present for *show*. ConceptNet (version 5.3) is automatically accessed through the REST API. All words with the relation *RelatedTo*, *IsA*, *PartOf*, *MemberOf*, *HasA*, *UsedFor*, *CapableOf*, *AtLocation*, *Causes*, *HasSubEvent*, *CreatedBy*, *Synonym* or *DefinedAs* to the searched word are selected. The words with a synonym relation to the searched word are also searched through the REST API. An example is shown in Fig. 2.



http://conceptnet5.media.mit.edu/data/5.3/c/en/show

```
"start": "/c/en/someone",
      "surfaceText": "[[someone]] can be at [[the show]]",
      "uri": "/a/[/r/AtLocation/,
             /c/en/someone/,
            /c/en/show/]",
      "weight": 2.584962500721156
    },
    {
      "context": "/ctx/all",
      "dataset": "/d/umbel",
      "end": "/c/en/testimony",
      "features": [
        "/c/en/show/_/testifying /r/Synonym -",
        "/c/en/show/_/testifying - /c/en/testimony",
        "- /r/Synonym /c/en/testimony"
      ],
      "id": "/e/8ddba65090d5086fa108602532e5a9e9d594798a",
      "license": "/l/CC/By-SA",
      "rel": "/r/Synonym",
      "source_uri": "/s/umbel/2013",
      "sources": [
        "/s/umbel/2013"
      ],
      "start": "/c/en/show/_/testifying",
      "surfaceText": "[[show]] is a synonym of
[[testimony]]",
      "uri": "/a/[/r/Synonym/,
             /c/en/show/_/testifying/,
            /c/en/testimony/]",
      "weight": 2.584962500721156
    },
```

http://conceptnet5.media.mit.edu/data/5.3/c/en/testimony

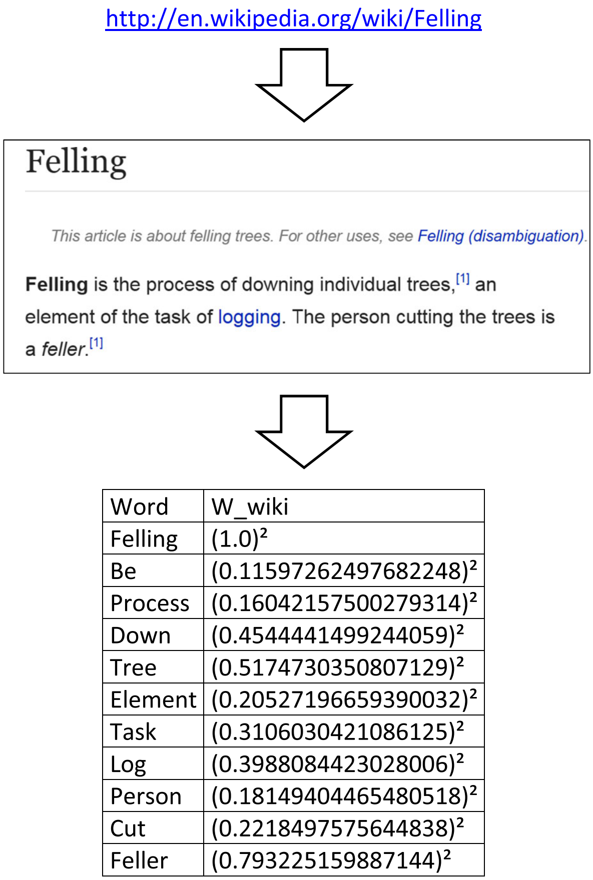
Fig. 2 Example of ConceptNet expansion method

The weight of the words is determined by the weight of the edge ($score_{rel}$) between the found word and the word in the complex event query. The weight is often a value between zero and thirty and is adjusted to a value that is typically between zero and one using:

$$W_{c,cn} = (\frac{score_{rel}}{30})^3 \tag{3}$$

The triple power of the scoring was found by training on the five events in the Research Set. In order to deal with query drift towards the expanded word, the weighted sum of the newly found words is adjusted to the weight of the word searched for. In the event *dog show*, both *dog* and *show* have a weight of 0.5. The sum of the weights of the expanded words of *show* is, thus, 0.5. If the expanded words for *show* are *concert* (0.8), *popcorn* (0.3) and *stage* (0.5), the adjusted weights are 0.25, 0.09375 and 0.15625, respectively.

Wikipedia Wikipedia does have, in contrast to ConceptNet, a lot of information about events. For each event, we automatically search for the corresponding Wikipedia page through the REST API (on October 13, 2014) and manually disambiguate to select the correct Wikipedia page. From this page all text above the table of contents, which we consider



http://en.wikipedia.org/wiki/Felling

Felling

This article is about felling trees. For other uses, see Felling (disambiguation).

Felling is the process of downing individual trees,[1] an element of the task of logging. The person cutting the trees is a *feller*.[1]

Word	W_wiki
Felling	$(1.0)^2$
Be	$(0.11597262497682248)^2$
Process	$(0.16042157500279314)^2$
Down	$(0.454441499244059)^2$
Tree	$(0.5174730350807129)^2$
Element	$(0.20527196659390032)^2$
Task	$(0.3106030421086125)^2$
Log	$(0.3988084423028006)^2$
Person	$(0.18149404465480518)^2$
Cut	$(0.2218497575644838)^2$
Feller	$(0.793225159887144)^2$

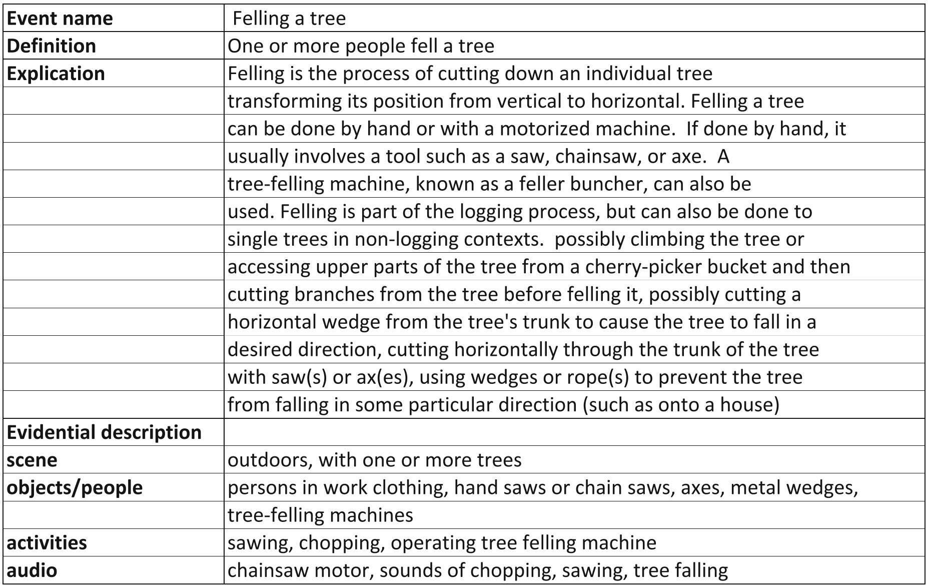
Fig. 3 Example of Wikipedia expansion method

as the general definition part, is scraped. All nouns and verbs are selected using the Stanford Core NLP parser [25]. The weight is calculated using TFIDF. The term frequency (TF) is calculated by counting the amount of times a word is present in the text (f(t,d)). The inverse document frequency (IDF) is calculated by counting the amount of documents the word appears in ($\log \frac{N}{1+|\{d\in D:t\in d\}|}$). The document set is a set of 5798 Wikipedia pages (collected on July 9, 2014). These Wikipedia pages are selected by taking all nouns, verbs, combined nouns and adjective-noun pairs from the text snippets of videos in the Research set. The term frequency is multiplied with the inverse document frequency to obtain the TFIDF. The TFIDF is divided by the highest possible IDF value and squared. This squaring is added because training on the five events in the Research Set increased performance using these steps. This leads to:

$$W_{c,wiki} = \frac{f(t,d)\cdot\log\frac{N}{1+|\{d\in D:t\in d\}|}{}^{2}}{log\frac{N}{1}} \tag{4}$$

where t is the term, d is the current document, $f(t,d)$ is the frequency of t in d, D is the document set, N is the total amount of documents and $|\{d\in D:t\in d\}|$ is the amount of documents in which t appears.

An example can be found in Fig. 3.



Event name	Felling a tree
Definition	One or more people fell a tree
Explication	Felling is the process of cutting down an individual tree
	transforming its position from vertical to horizontal. Felling a tree
	can be done by hand or with a motorized machine. If done by hand, it
	usually involves a tool such as a saw, chainsaw, or axe. A
	tree-felling machine, known as a feller buncher, can also be
	used. Felling is part of the logging process, but can also be done to
	single trees in non-logging contexts. possibly climbing the tree or
	accessing upper parts of the tree from a cherry-picker bucket and then
	cutting branches from the tree before felling it, possibly cutting a
	horizontal wedge from the tree's trunk to cause the tree to fall in a
	desired direction, cutting horizontally through the trunk of the tree
	with saw(s) or ax(es), using wedges or rope(s) to prevent the tree
	from falling in some particular direction (such as onto a house)
Evidential description	
scene	outdoors, with one or more trees
objects/people	persons in work clothing, hand saws or chain saws, axes, metal wedges,
	tree-felling machines
activities	sawing, chopping, operating tree felling machine
audio	chainsaw motor, sounds of chopping, sawing, tree falling

Fig. 4 Example of a textual description in TRECVID MED

Expert The textual description provided with the TRECVID Multimedia Event Detection (MED) task is used as expert knowledge. This description consists of the event name, definition, explanation and an evidential description of the scene, objects, activities and audio. An example of a description is shown in Fig. 4. From all different parts the nouns and verbs are manually extracted. The Stanford Core NLP parser [25] is not used, because the text also contained separate words instead of sentences. This causes problems for the parser. An addition to the selection of these words is that words within brackets or enumerations with an *or* were clustered. This cluster is used to indicate that only one of these concepts has to be available in the video. In texts in which a noun or verb is placed before or after a negation, such as .. *is not considered positive for this event* and *without a vehicle*, are not taken into account. The determination of the weight is equal to the weight determination in the Wikipedia expansion method (W_{wiki}). From the clustered word the compound noun is chosen for determination of term frequency and inverse document frequency. In case no compound word was present the first word is chosen. An example is shown in Fig. 4.

3.2.2 Concept detectors

The list of weighted words from the methods is matched to a set of 1818 concept detector labels, which are (compound) nouns or verbs. This comparison is done in two ways. The first way is to compare the word to the concept detector label. The exact matches are selected. The words without an exact match are compared using WordNet [28]. Both the word and the concept detectors are compared in WordNet using the WUP similarity [49]. Concept detectors with a similarity of 1.0 to a word are selected, which means that both point to the same synset, such as with *foot* and *feet* or *project* and *task*. The only two exceptions are *fight* for *engagement* and *hide* for *fell*. These matches are not taken into account. The selected concept detectors get the weight of the words. If multiple words point to the same concept detector, such as with synonyms, the weights are added. If one word points to multiple concept detectors, such as *dog* from one collection within the set and *dogs* from another collection, the weight is equally divided over the concept detectors. At the end of the matching process the weight of a concept detector is divided by the total amount of weights in order to create a sum of the weights equal to 1.0.

The set of 1818 concept detectors consists of three collections. The first collection consists of 1000 concept detectors and is trained on a subset of the ImageNet dataset with 1.26 million training images as used in ILSVRC-2012 [11]. The second collection has 346 concept detectors, which are trained on the dataset from the TRECVID Semantic Indexing task of 2014. The final collection contains 472 concept detectors and is trained and gathered from the Research Set of TRECVID MED [33]. The last collection originally contained 497 concept detectors, but the detectors directly trained on the high-level events are removed. In this way we can test the elements in the query and query expansion instead of just the (rather good) accuracy of these concept detectors. More details on the concept detectors can be found in [33].

3.2.3 Videos

The test set of TRECVID MED 2014 consists of 23.000 videos. From each video one keyframe per two seconds is extracted. Each concept detector is applied to each keyframe. As a result for each keyframe for each concept detector a value between zero and one is available, which represents the confidence score. The highest confidence score over all keyframes for one video is selected. This score is multiplied by the weight of the concept detector, which was originally coming from the methods. The weighted sum of the concept detector values represents an estimation of the presence of the event in the video. This estimation is used to rank all videos and place the videos in a list in descending order.

4 Results

Results on the Test Set of TRECVID MED 2014 for each of the twenty events are split up in four tables: Tables 1, 2, 3 and 4. Bold digits indicate the highest performance in the row and italic digits indicate random performance. Table 1 shows average performance of the events in which all nouns and verbs in the event query have matching concept detectors. ConceptNet is only used to expand words that could not be matched and, thus, no average performance is available in Table 1 for ConceptNet. Wikipedia has no performance if no page containing the event could be found. Table 2 contains performance of the events in which the main noun of the event query, which is the second noun in compound words, is matched to a concept detector. If no additional words could be found by ConceptNet, performance of ConceptNet is equal to performance of the query. Table 3 shows performance of the events in which at least one word in the event query (not the main noun) could be matched to a concept detector. Table 4 contains information about events in which no word in the event query could be matched to a concept detector. The Mean Average Precision on all twenty events is 0.03865, 0.06143 (0.06220 without same as Query and 0.03047 without *beekeeping*), 0.03024 (0.02042 with random) and 0.03262 for the Query method, ConceptNet, Wikipedia and Expert knowledge, respectively.

Table 1 Average Precision: matching query

Event Name	Query	ConceptNet	Wikipedia	Expert
Cleaning appliance	**0.10228**			0.01055
Town hall meeting	**0.03800**		0.01568	0.00866
Rock climbing	**0.13932**		0.01936	0.01957
Fixing musical instrument	0.04245			**0.04954**
MEAN	**0.08051**		0.01752	0.02208

4.1 Query expansion vs. no query expansion

Comparing average performance of our baseline, which is presented as *Query* in the tables, to each of the other columns shows that query expansion does not always improve performance. Mean Average Precision on all twenty events show the highest value for the method in which no query expansion is used (ConceptNet without *beekeeping*). Table 1 shows that if all nouns and verbs in the query could be matched to a concept detector, average performance is highest for the query. The events *town hall meeting* and *rock climbing* have significantly higher performance for the query compared to the expansion methods. Table 2 shows the same trend as Table 1, but the exception is *tuning musical instrument*. Table 3 shows a mixed performance and in Table 4 performance of the baseline is random and, thus, query expansion methods perform better.

4.2 Expert knowledge vs. common knowledge

The average results regarding common knowledge (ConceptNet without *beekeeping*) and expert knowledge show no clear preference for either method. Comparing the separate results, the performance using expert knowledge is clearly higher in the events *non-motorized vehicle repair*, *tuning musical instrument*, *attempting bike trick* and *working metal craft project*. For the other fourteen events, the common knowledge bases perform equally good or better than expert knowledge.

Table 2 Average Precision: one matching main noun in query

Event Name	Query	ConceptNet	Wikipedia	Expert
Non-motorized vehicle repair	0.02016	0.02016		**0.02915**
Renovating home	**0.01568**	**0.01568**		0.01261
Winning race	**0.04048**	0.01228	0.01181	0.00695
Felling tree	**0.04057**	0.01145	0.01461	0.00656
Parking vehicle	**0.10675**	0.00321	0.00390	0.00404
Tuning musical instrument	0.01496	0.02436	0.02235	**0.05572**
MEAN	**0.03977**	0.01452	0.01317	0.01917

Table 3 Average Precision: one match (not main noun) in query

Event Name	Query	ConceptNet	Wikipedia	Expert
Attempting bike trick	0.07117	0.02361		**0.07486**
Working metal craft project	**0.04621**	0.00336		0.03865
Horse riding competition	0.07655	0.02766	**0.11451**	0.01017
Playing fetch	0.00264	**0.01519**	0.00936	0.00275
Dog show	0.00901	0.05339	0.00943	**0.05362**
MEAN	0.04111	0.02464	**0.04443**	0.03601

4.3 ConceptNet vs. wikipedia

Common knowledge bases ConceptNet (without *beekeeping*) and Wikipedia have comparable Mean Average Precision values. Wikipedia has a higher average precision in Table 3 and ConceptNet has a higher average precision in Table 4. Comparing the different events in Tables 3 and 4, Wikipedia performs better than ConceptNet in *horse riding competition*, *marriage proposal* and *tailgating*.

4.4 Late fusion

In this section, we present the result of late fusion, because we expect that late fusion will help to exploit complementary information provided in the different expansion methods. In late fusion, the scores of the videos ($S_{e,v,EM}$, see 3.2) of the different expansion methods are combined using four different fusion techniques.

The first fusion technique is the arithmetic mean in which the fused score is calculated by:

$$Fa_{e,v} = \frac{1}{EM} \sum_{em \in EM} S_{e,v,em}, \tag{5}$$

where $Fa_{e,v}$ is the fused score for video v and event e, EM is the set of expansion methods and $S_{e,v,em}$ is the score for video v and event e in expansion method em

The geometric mean is used as a second fusion technique:

$$Fg_{e,v} = \prod_{em \in EM} S_{e,v,em}, \tag{6}$$

Table 4 Average Precision: no matching word in query

Event Name	Query	ConceptNet	Wikipedia	Expert
Giving direction location	*0.00095*	**0.00324**		0.00321
Marriage proposal	*0.00219*	0.00203	0.00324	**0.00414**
Beekeeping	*0.00116*	**0.64970**	0.15346	0.23404
Wedding shower	*0.00121*	**0.03929**	0.01301	0.02594
Tailgating	*0.00133*	0.00199	**0.00244**	0.00169
MEAN	0.00137	**0.13925**	0.04304	0.0.05380

Table 5 Mean average precision with late fusion

Fused Part	MAP before fusion	MAP after fusion	MAP increase (%)
Per event	0.08103	0.09198	13.5
Events Table 1	0.08051	0.08051	0.0
Events Table 2	0.03977	0.04030	1.3
Events Table 3	0.05014	0.06130	22.3
Events Table 4	0.13925	0.13925	0.0

where $Fq_{e,v}$ is the fused score for video v and event e, EM is the set of expansion methods and $S_{e,v,em}$ is the score for video v and event e in expansion method em

As a third fusion technique, the highest value for a video is taken:

$$Fm_{e,v} = \max_{em \in EM} S_{e,v,em}, \tag{7}$$

where $Fm_{e,v}$ is the fused score for video v and event e, EM is the set of expansion methods and $S_{e,v,em}$ is the score for video v and event e in expansion method em

The last fusion technique is a weighted mean, in which

$$Fw_{e,v} = \frac{1}{\sum\limits_{em \in EM} W_{em}} \cdot \sum_{em \in EM} (W_{em} \cdot S_{e,v,em}), \tag{8}$$

where $Fw_{e,v}$ is the fused score for video v and event e, EM is the set of methods, W_{em} is the weight for expansion method em and $S_{e,v,em}$ is the score for video v and event e in expansion method em

The fusion score of each combination of two, three and four expansion methods are calculated. In the weighted mean, both values 0.25 and 0.75 are examined as W_{em} for the expansion methods. The results of the fusion optimized per event and optimized per part is shown in Table 5. Results show that Mean Average Precision optimized per event improve from 0.08103 to 0.09198 (+13.5 %) with fusion. Because we are working with the zero-example case, this is our upper boundary. Mean Average Precision optimized per part increases from 0.07538 to 0.07833 (+ 3.9 %) overall. In the column *MAP before fusion* in Table 5, the Mean Average Precision of the query method is used for the events of Tables 1 and 2. Wikipedia is used for the events in Table 3 and if Wikipedia has no result, the query method is used. The results on ConceptNet are used for the events of Table 4. In the fusion of these parts, no single fusion method could outperform the query in complete matched query (events from Table 1) and ConceptNet in the events from Table 4. For the matching main nouns (Table 2) a fusion with the maximum of the query, Wikipedia and the Expert provides highest performance. This fusion method improves 22.7 % on the event *tuning musical instrument* and less than 1.0 % on the other events. In the matching without the main nouns (Table 3) a weighted mean of the query (0.25), Wikipedia (0.75) and the Expert (0.25) provides highest performance. This fusion method improves on the events *attempting bike trick* (7.8 %), *working metal crafts project* (136.5 %) and *dog show* (46.4 %).

5 Discussion, conclusion and future work

In our experiments, the Test Set of TRECVID 2014 Multimedia Event Detection (MED) task [34] was used to compare the effectiveness of our retrieval system for complex event queries. We compared ConceptNet 5 [44] and Wikipedia as common knowledge bases and the textual description provided with the TRECVID task to determine which type of knowledge base is best to use.

Results comparing the baseline with the query expansion methods show that the complex event query not necessarily perform worse than methods using query expansion. These results, however, do not imply that knowledge bases should not be used. It is important to know in which cases a knowledge base can add information and in which cases the complex event query is enough. The results clearly show that if all query terms are found, additional information does not improve performance. This is also the case in most of the events in which the main noun is found. On the other hand, query expansion is beneficial to use in the other events, which confirms our expectations. This brings us to the first conclusion: *1) Query Expansion can improve performance compared to using no query expansion in the case that the main noun of the query could not be matched to a concept detector*.

A result that does not meet our expectations is that query expansion using expert knowledge is not better than query expansion using common knowledge bases. In the events in which no word could be matched to the query, the expert only performs best in *marriage proposal*, whereas the common knowledge bases perform best in the other four events. In the events in which one match in the query is found, expert knowledge and common knowledge both perform best in two of the five events. The second conclusion, therefore, is: *2) Query expansion using expert knowledge is not necessarily better than query expansion using common knowledge*.

Another interesting result is in the comparison of ConceptNet and Wikipedia. The results in Tables 3 and 4 show that Wikipedia only performs better than ConceptNet in *horse riding competition*, *marriage proposal* and *tailgating*. In *horse riding competition*, ConceptNet is used to search for *competition*. This word is general and, therefore, more general words for competitions are found. In Wikipedia, *horse riding competition* is used and one of the key words for the event (*vault*) is found. In *marriage proposal*, ConceptNet has less information than Wikipedia and, therefore, Wikipedia has better performance. In *tailgating*, ConceptNet has other information than Wikipedia. Wikipedia has more information on sports and food, while ConceptNet has more information about the car. Two events in which ConceptNet clearly outperforms all other methods are *beekeeping* and *wedding shower*. Wikipedia and Expert both find *bee* and *apiary*, but other concepts suppress the weight of *apiary*, which decreases performance. In *wedding shower*, the same problem occurs. The concept *party* seems to provide the best information and a low weight of this concept decreases performance. Weighting is, thus, an important part in the expansion methods. In general, we can conclude that, in this configuration, *3) ConceptNet performs slightly better than Wikipedia*.

The last result section shows the results of late fusion. With the twenty events, it is not yet clear which fusion method performs best in which cases. Several events show highest performance using geometric mean, but in the separation of the parts the geometric mean does not have highest performance over a part. Furthermore, some fusion methods improve performance in one event, but decrease performance drastically in other events. For Table 2 the best method per part is a weighted mean. In the events of Table 2, *horse riding competition* has a high performance in the Wikipedia method. In order to not lose this result in the mean, Wikipedia has a weight of 0.75 and the query and expert have a weight of 0.25. ConceptNet, apparently, provides no complementary information and is, therefore, not increasing performance. For Table 3 the best fusion method is an arithmetic mean. The event *working metal crafts projects*, for example, has information in expert about a workshop and kinds of tools and the query has *metal*. Adding this information gives slightly better information than taking a product or taking the maximum. In general, we can conclude that: *4) Late fusion can slightly improve performance*.

To conclude, query expansion is beneficial, especially in events of which the main noun of the query could not be matched to a concept detector. Common knowledge bases do not always perform worse than expert knowledge, which provides options for automatic query expansion from the Web in complex event detection.

The experiments conducted in this paper have some limitations. First, research is only conducted on twenty complex events, which is a very small set. The conclusions on the comparison between the common knowledge bases can, therefore, be different in a larger or different set of complex events. In a larger set of complex events the specific situations in which any of the methods is preferred over the others can be determined in a better way. Second, less than 2000 concept detectors are used. Many words in the query and, especially, the query expansion methods could not be matched to concept detectors. Third, the weight determination ConceptNet, Wikipedia and the expert expansion method is trained on the Research Set with only five events. This amount of events is not enough to train on and the weighting is, therefore, not optimal. Fourth, the fusion methods as well as the weights in the weighted mean are not fully explored.

In the future, we want to compare the kind of information available in the expert knowledge and in common knowledge in order to determine what kind of information provides the increase in performance in complex event detection. This can be combined with the further exploration of fusion methods. Other common knowledge bases, such as YAGO2 and Flickr, are possibly worth integrating in our system. Another interesting option is to examine the use of (pseudo-) relevance feedback. This feedback can also be combined with, for example, common knowledge sources.

Acknowledgments We would like to thank the VIREO team from the City University of Hong Kong for the application of their concept detectors on the TRECVID MED 2014 data sets and both the technology program Adaptive Multi Sensor Networks (AMSN) and the ERP Making Sense of Big Data (MSoBD) for their financial support.

References

1. Baeza-Yates R, Ciaramita M, Mika P, Zaragoza H (2008) Towards semantic search. In: Natural language and information systems. Springer, pp 4–11
2. Bagdanov AD, Bertini M, Del Bimbo A, Serra G, Torniai C (2007) Semantic annotation and retrieval of video events using multimedia ontologies. In: International conference on semantic computing. IEEE, pp 713–720
3. Ballan L, Bertini M, Del Bimbo A, Seidenari L, Serra G (2011) Event detection and recognition for semantic annotation of video. Multimed Tools Appl 51(1):279–302
4. Bodner RC, Song F (1996) Knowledge-based approaches to query expansion in information retrieval. Springer, Berlin, pp 146–158. ISBN:3-540-61291-2
5. de Boer M, Schutte K, Kraaij W (2013) Event classification using concepts In: ICT-Open, pp 39–42
6. Bouma H, Azzopardi G, Spitters M, de Wit J, Versloot C, van der Zon R, Eendebak P, Baan J, ten Hove JM, van Eekeren A, ter Haar F, den Hollander R, van Huis J, de Boer M, van Antwerpen G, Broekhuijsen J, Daniele L, Brandt P, Schavemaker J, Kraaij W, Schutte K (2013) TNO at TRECVID 2013: multimedia event detection and instance search. In: Proceedings of TRECVID 2013
7. Burgess J, Green J (2013) YouTube: online video and participatory culture. Wiley. ISBN-13: 978-0745644790

8. Caputo B, Müller H, Martinez-Gomez J, Villegas M, Acar B, Patricia N, Marvasti N, Üsküdarlı S, Paredes R, Cazorla M et al (2014) ImageCLEF 2014: overview and analysis of the results. In: Information access evaluation. Multilinguality, multimodality, and interaction. Springer, pp 192–211
9. Carpineto C, Romano G (2012) A survey of automatic query expansion in information retrieval. ACM Comput Surv (CSUR) 44(1):1
10. Chen J, Cui Y, Ye G, Liu D, Chang SF (2014) Event-driven semantic concept discovery by exploiting weakly tagged internet images. In: Proceedings of international conference on multimedia retrieval. ACM, p 1
11. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: Conference on computer vision and pattern recognition. IEEE, pp 248–255
12. Francois AR, Nevatia R, Hobbs J, Bolles RC, Smith JR (2005) VERL: an ontology framework for representing and annotating video events. MultiMedia IEEE 12(4):76–86
13. Georis B, Maziere M, Bremond F, Thonnat M (2004) A video interpretation platform applied to bank agency monitoring. In: IEEE Intelligent Surveillance Systems (IDSS-04), pp 46–50
14. Habibian A, van de Sande KE, Snoek CG (2013) Recommendations for video event recognition using concept vocabularies. In: Proceedings of the 3rd ACM conference on International conference on multimedia retrieval. ACM, pp 89–96
15. Hare JS, Lewis PH, Enser PG, Sandom CJ (2006) Mind the gap: another look at the problem of the semantic gap in image retrieval. In: Electronic imaging 2006, pp 607,309–607,309. International society for optics and photonics
16. Hassan S, Mihalcea R (2011) Semantic relatedness using salient semantic analysis. In: Proceedings of AAAI confences on artificial intelligence, pp 884–889
17. Hauptmann AG, Christel MG (2004) Successful approaches in the TREC video retrieval evaluations. In: Proceedings of the 12th annual ACM international conference on multimedia. ACM, pp 668–675
18. Hoffart J, Suchanek FM, Berberich K, Weikum G (2013) YAGO2: a spatially and temporally enhanced knowledge base from wikipedia. Artif Intell 194:28–61
19. Hoque E, Hoeber O, Strong G, Gong M (2013) Combining conceptual query expansion and visual search results exploration for web image retrieval. J Ambient Intell Human Comput 4(3):389–400
20. Jiang YG, Bhattacharya S, Chang SF, Shah M (2012) High-level event recognition in unconstrained videos. In: International journal of multimedia information retrieval, pp 1–29
21. Kotov A, Zhai C (2012) Tapping into knowledge base for concept feedback: leveraging conceptnet to improve search results for difficult queries. In: Proceedings of the fifth ACM international conference on Web search and data mining. ACM, pp 403–412
22. Leong CW, Hassan S, Ruiz ME, Mihalcea R (2011) Improving query expansion for image retrieval via saliency and picturability. In: Multilingual and multimodal information access evaluation. Springer, pp 137–142
23. Liu XH (2002) Semantic understanding and commonsense reasoning in an adaptive photo agent. Ph.D. thesis. Massachusetts Institute of Technology
24. Ma Z, Yang Y, Cai Y, Sebe N, Hauptmann AG (2012) Knowledge adaptation for ad hoc multimedia event detection with few exemplars. In: Proceedings of the 20th ACM international conference on multimedia. ACM, pp 469–478
25. Manning CD, Surdeanu M, Bauer J, Finkel J, Bethard SJ, McClosky D (2014) The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, pp 55–60
26. Mascardi V, Cordì V, Rosso P (2007) A comparison of upper ontologies. In: WOA, pp 55–64
27. Mazloom M, Habibian A, Snoek CG (2013) Querying for video events by semantic signatures from few examples. In: MM'13, pp 609–612
28. Miller GA (1995) Wordnet: a lexical database for english. Commun ACM 38(11):39–41
29. Milne D, Witten IH (2013) An open-source toolkit for mining wikipedia. Artif Intell 194:222–239
30. Naphade M, Smith JR, Tesic J, Chang SF, Hsu W, Kennedy L, Hauptmann A, Curtis J (2006) Large-scale concept ontology for multimedia. MultiMedia. IEEE 13(3):86–91
31. Natarajan P, Natarajan P, Manohar V, Wu S, Tsakalidis S, Vitaladevuni SN, Zhuang X, Prasad R, Ye G, Liu D et al (2011) Bbn viser trecvid 2011 multimedia event detection system. In: NIST TRECVID workshop, vol 62
32. Natsev AP, Haubold A, Tešić J, Xie L, Yan R (2007) Semantic concept-based query expansion and re-ranking for multimedia retrieval. In: Proceedings of the 15th international conference on multimedia. ACM, pp 991–1000
33. Ngo CW, Lu YJ, Zhang H, Yao T, Tan CC, Pang L, de Boer M, Schavemaker J, Schutte K, Kraaij W (2014) VIREO-TNO @ TRECVID 2014: multimedia event detection and recounting (MED and MER). In: Proceedings of TRECVID 2014

34. Over P, Awad G, Michel M, Fiscus J, Sanders G, Kraaij W, Smeaton AF, Quenot G (2013) TRECVID 2013 – an overview of the goals, tasks, data, evaluation mechanisms and metrics. In: Proceedings of TRECVID 2013, NIST, USA
35. Over P, Leung C, Ip H, Grubinger M (2004) Multimedia retrieval benchmarks. Multimedia 11(2):80–84
36. Patil PB, Kokare MB (2011) Relevance feedback in content based image retrieval: a review. J Appl Comput Sci Math 10(10):4047
37. Pedersen T, Patwardhan S, Michelizzi J (2004) Wordnet:: similarity: measuring the relatedness of concepts. In: Demonstration Papers at HLT-NAACL 2004, pp 38–41. Association for Computational Linguistics
38. Pisanelli D (2004) Biodynamic ontology: applying BFO in the biomedical domain. Ontologies Med 102:20
39. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M et al (2014) Imagenet large scale visual recognition challenge. arXiv preprint arXiv:1409.0575
40. Sheth A, Ramakrishnan C, Thomas C (2005) Semantics for the semantic web: the implicit, the formal and the powerful. Int J Semantic Web Inf Syst (IJSWIS) 1(1):1–18
41. Smeaton AF, Over P, Kraaij W (2006) Evaluation campaigns and trecvid. In: MIR '06: Proceedings of the 8th ACM international workshop on multimedia information retrieval. ACM Press, New York, pp 321–330
42. Smeulders AW, Worring M, Santini S, Gupta A, Jain R (2000) Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell 22(12):1349–1380
43. Snoek CG, Smeulders AW (2010) Visual-concept search solved? IEEE Comput 43(6):76–78
44. Speer R, Havasi C (2012) Representing general relational knowledge in conceptnet 5. In: LREC, pp 3679–3686
45. Tu K, Meng M, Lee MW, Choe TE, Zhu SC (2014) Joint video and text parsing for understanding events and answering queries. MultiMedia IEEE 21(2):42–70
46. Vatant B, Wick M (2012) Geonames ontology. http://www.geonames.org/ontology/
47. Von Ahn L, Kedia M, Blum M (2006) Verbosity: a game for collecting common-sense facts. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 75–78
48. Voss J (2005) Measuring wikipedia. In: Proceedings of 10th international conference of the international society for scientometrics and informetrics
49. Wu Z, Palmer M (1994) Verbs semantics and lexical selection. In: Proceedings of the 32nd annual meeting on association for computational linguistics, pp 133–138. Association for Computational Linguistics
50. van der Zon R (2014) A knowledge base approach for semantic interpretation and decomposition in concept based video retrieval. Master's thesis, TU Delft

Maaike de Boer received her BS and MS degrees in Artificial Intelligence from the University of Utrecht, The Netherlands, in 2011 and 2013. Currently, she is a PhD student in the Department of Intelligent Imaging at TNO, The Hague, The Netherlands and the Institute for Computing and Information Sciences at the Radboud University, Nijmegen, The Netherlands. Her present research interests include multimedia event detection, semantic analysis and knowledge bases.

Klamer Schutte received an MS degree in physics from the University of Amsterdam in 1989 and a PhD degree from the University of Twente, Enschede, The Netherlands, in 1994. He had a postdoctoral position with the Delft University of Technologys Pattern Recognition (now Quantitative Imaging) group. Since 1996, he has been employed by TNO, currently as lead research scientist of intelligent imaging. Within TNO he has actively led multiple projects in areas of signal and image processing. Recently, he has led many projects, including super-resolution reconstruction for both international industries and governments, resulting in super-resolution reconstruction based products in active service. His research interests include behavior recognition, pattern recognition, sensor fusion, image analysis, and image restoration.

Wessel Kraaij received his MS degree in Electrical Engineering at the Eindhoven University of Technology, The Netherlands, in 1988. In 1988, he was affiliated as a research assistant at the Eindhoven University of Technology. From 1988 to 1993, he worked on a EU project as a junior researcher at the Tilburg University and in1994 and 1995 he was affiliated as a research assistant at the Utrecht Institute of Linguistics of University of Utrecht. Since 1995, he is a senior scientist in the Department of Media Networking Services at TNO, Delft, The Netherlands. Currently, he is involved in managing the involvement of TNO in a large EU integrated project and he is leading projects. In 1999, he was a visiting researcher at the Université de Montréal, Quebec, Canada. In 2004, he received his PhD in Computer Science from the University of Twente, Enschede, The Netherlands. Since 2003, he is co-coordinator of the NIST TRECVID benchmark workshop on video retrieval. As from 2008, he is an endowed professor at the Radboud University, Nijmegen, The Netherlands. His chair is entitled 'Information Filtering and Aggregation' and he founded the 'Information Foraging Lab' group with a research focus on the user side of search. Wessel is specialized in language technology, information retrieval and multimedia search technology and his main interest is to derive intermediate semantic information from unstructured data sources for indexing, retrieval and knowledge mining purposes. He has published over 150 research papers and has been a member of many technical program committees, such as ACM SIGIR. He has been co-chair or IIIiX 2012, SIGIR 2007 and several DIR workshops.