A Concept-Suggestion Engine for Professional Multimedia Archives

Marco A. Palomino[1], Michael P. Oakes[1], and Tom Wuytack[2]

[1] University of Sunderland - Informatics Centre
St Peter's Way, Sunderland SR6 0DD, United Kingdom
{marco.palomino, michael.oakes}@sunderland.ac.uk

[2] Belga News Agency
Rue Frederic Pelletier 8b, 1030 Brussels, Belgium
wut@belga.be

Abstract. Choosing the optimal set of keywords to represent a search engine query is not a trivial task, and may involve an iterative process such as relevance feedback, repeated unaided attempts by the user or the automatic suggestion of additional terms, which the user may select or reject. This is particularly true of a multimedia search engine which searches on concepts as well as user-input terms, since the user is unlikely to be familiar with the full range of system-known concepts in advance. We propose two concept suggestion strategies: suggestion by semantic similarity, where the closest matching concept definitions in a glossary to the initial user input are found using the cosine similarity measure, and suggestion by normalised textual matching, where concepts are suggested if their headwords match the user input. Both methods were evaluated by comparing machine suggestions with concepts suggested by professional annotators, using the measures of micro- and macro- precision and recall. Although normalised text matching was the simpler technique, it performed much better on the recall-based measures, and only slightly less well on the precision-based measures.

Keywords: Information retrieval, multimedia, concepts, ground-truth.

1 Introduction

Popular multimedia search engines, such as *Google* [1], *YouTube* [2] and *Blinkx* [3], provide access to their repositories via text, as this is still the easiest way for their users to express their information needs. The indices of these search engines point to pictures, videos and some other resources on the Web based on their file names, surrounding text or transcripts. Regrettably, this results in disappointing performance when the visual content is not reflected in the associated text. Hence, a current trend in information retrieval consists of learning a lexicon of semantic concepts from multimedia examples and employing them as entry points when querying the Web [4].

As part of the EU-funded *Video and Image Indexing and Retrieval in the Large-Scale (VITALAS)* project [5], which aims to offer a system dedicated to the intelligent access to professional multimedia archives, we have experimented with several techniques for the automatic suggestion of concepts derived from user-specified sample pictures and their textual captions. Currently, VITALAS employs a vocabulary of 525 concepts, whose entries vary from pure video format—like a detected *overlayed text*—settings and scenarios—like an *interview*—objects—like an *elephant*—or events—like a *celebration after scoring*.

The VITALAS concept vocabulary gives users semantic access to picture and video, allowing them to query on the presence or absence of content elements. However, selecting the right topic from a large vocabulary is a lengthy and resource-consuming task that should not be performed manually. Therefore, we have developed a *suggestion engine* that analyses textual captions contained in an archive and automatically derives the most relevant concepts for querying such an archive. The results yielded by our suggestion engine have been compared with observations made by professional annotators, who reviewed the pictures and linked them to some of the concepts that best described them. Thus far, our evaluation indicates that simple textual matches between picture captions and the concept vocabulary produces better suggestions than other more sophisticated approaches.

The remainder of this document is structured as follows: Section 2 describes the multimedia collection that we employed in our study. Section 3 presents the concept vocabulary from which we derive our suggestions and elaborates on the acquisition of ground-truth annotation. Section 4 explains the two different strategies that we propose to implement a suggestion engine. Section 5 reports on the evaluation of our results. Section 6 introduces the related work in this area of research, and Section 7 states our conclusions.

2 Belga's Multimedia Archive

In order to undertake our research, we made use of a large multimedia archive owned by the *Belga News Agency* [6]. Belga's archive covers Belgian and international news and current affairs—politics and economics, finance and social affairs, sports, culture and personalities. Although Belga's content is published in four different formats—text, pictures, audio and video—, this paper concentrates on pictures and their associated textual captions, exclusively.

A *caption* is a free-text field whose content is created by photographers, and offers an explanation or designation accompanying a picture posted on Belga's website. Each group of photographers has its own conventions and styles to present information. As a consequence, certain captions include not only text relative to pictures, but also names of photographers, their initials and acronyms of press agencies, as well as the dates when the pictures were taken or published and some other ancillary information. Since none of these particulars are deleted before posting, we decided to keep them in our analysis too.

To ensure that we had enough material to carry out our work, Belga granted us access to a set of 1,727,159 pictures and captions that were published on its website between 22 June 2007 and 2 October 2007. Figure 1 displays an example of a typical picture and caption posted on Belga's website.

Fig. 1. Example of a Belga Picture and Caption: Italian Prime Minister, Romano Prodi (R), shakes hands with US President George W. Bush prior to their press conference at Chigi Palace in Rome, Italy, in June 2007. President Bush, who is in Rome as part of his trip through Europe, will also meet Silvio Berlusconi.

3 VITALAS Annotated Concept Vocabulary

Acquiring a suitable list of concepts for multimedia purposes is a major research challenge. Naphade *et al.* have built a multimedia ontology based on extensive analysis of video-archive query logs [7], and the *MediaMill Challenge* has employed this ontology as the main source for its 101 concept vocabulary [8].

As opposed to MediaMill, the VITALAS concept vocabulary is largely derived from the automatic extraction of keywords that characterise Belga's archive. Nevertheless, it has been refined manually over time. Originally, the vocabulary was derived from a comparison between Belga's captions and a model of general English language. The words that deviated from the model were very specific to the captions and thus made appropriate keywords to characterise the archive. Professional annotators evaluated the keywords and removed those that they considered unsuitable. The remaining keywords became the first entries in the VITALAS concept vocabulary. Later on, these entries were extended manually to guarantee that the vocabulary comprised as many categories as available in the news domain. Finally, we mined Belga's query logs programmatically to extract keywords that reflected the most important concepts from the users' perspective[1]. Some of these concepts were also added to the vocabulary.

[1] Belga gave us access to its query logs for exactly the same period when the collection of chosen captions was published on its website: 22 June 2007 – 2 October 2007.

Further details related to the VITALAS concept vocabulary have been published by Palomino *et al.* [9], and the entire concept vocabulary for the VITALAS project is available at the authors' website (`http://osiris.sunderland.ac.uk/~cs0mpl/VITALAS/`).

3.1 Ground-Truth Annotation

Recognising the importance of having a picture database containing ground-truth annotations parsed by humans, VITALAS selected 100,000 pictures published on Belga's website, and employed professional annotators to determine some of the concepts that best described them.

For annotation purposes, the presence of a concept was assumed to be binary: it was either visible in a picture or not—the location of the concept in the picture was not taken into account. A total of 1,000 pictures were annotated for each of the 525 concepts. However the set of 1,000 pictures annotated for a particular concept were not necessarily the same as those annotated for any other concept.

Approximately 500 of the pictures annotated for a particular concept were chosen from the results of queries submitted by Belga users who had included that concept name. The rest of the pictures were chosen randomly. Hence, the first half of pictures is likely to contain positive samples, whereas the second one is likely to contain negative samples. Achieving this balance was vital for the evaluation of our experiments.

The ground-truth annotation also included the provision of a textual *definition*—or *description*—for each concept, together with relevant keywords and references to positive images. Table 1 shows an example of a VITALAS concept—*food*—accompanied by its description and reference to positive images. Table 1 also shows a picture that has been annotated positively as an image that does correspond with the concept *food*.

Concept name: food

Concept description: An image showing any substance reasonably expected to be ingested by a human or an animal for nutrition or pleasure.

Relevant keywords: Cooking, meal.

Examples of positive images: A picture of a table showing a served meal; a picture of dishes ready to be consumed; a picture of meat, fish, fruit or vegetables for sale in a market.

Table 1. Example of Disambiguation and Positively Annotated Picture

The VITALAS manual annotation process has yielded an incomplete, but reliable ground-truth for our concept vocabulary. Certainly, we would like to have all the pictures annotated for all of the concepts; yet, despite resource limitations, we have gathered a reasonably large subset of annotated pictures.

4 Concept Suggestion Strategies

We identify two different approaches for suggesting to users the most relevant concepts related to a particular picture caption: suggestion based on the *semantic similarity* between the caption and the textual description of each concept; and suggestion based on *normalised-textual matching* between the caption and the concept vocabulary. In the following subsections, we detail these strategies.

4.1 Suggestion by Semantic Similarity

As explained in subsection 3.1, each concept ω in the VITALAS concept vocabulary is associated with a textual description d_ω. Then, we can measure the semantic similarity between a caption and the textual description of each different concept. Both captions and descriptions are *normalised* before examining their semantic similarity: all text is converted to lower case, punctuation and numbers are removed, and extremely common and semantically non-selective words are deleted—the stop-word list that we are using was built by Salton and Buckley for the experimental *SMART* information retrieval system [10]. In addition, all text is *stemmed*, reducing inflectional and derivationally related forms of a word to a common base—the particular algorithm for stemming English words that we are using is *Porter's algorithm* [11].

We represent each concept description as a *vector*, whose entries correspond to unique normalised words. Since concept descriptions are written in natural language, word distribution corresponds, roughly, with *Zipf's law* [12]. Therefore, the vector space model proposed by Salton *et al.* [13] is appropriate for our semantic analysis. Specifically, with a collection of descriptions D, a concept description d_ω in D, and a caption q containing words t_i, we use the following implementation of the vector space model to compute the *cosine similarity* between caption q and concept description d_ω:

$$sim(q, d_\omega) = \frac{\sum\limits_{k \in (q \cap d_\omega)} tf_{kq} \cdot tf_{kd_\omega}}{\sqrt{\sum\limits_{k \in d_\omega} (tf_{kd_\omega})^2} \sqrt{\sum\limits_{k \in q} (tf_{kq})^2}},$$

where tf_{kq} is the frequency of the $k-th$ word contained in caption q, and tf_{kd_ω} is the frequency of the $k-th$ word contained in description d_ω.

It can be demonstrated that the resulting similarity between q and d_ω ranges from 0 meaning *no match*, to 1 meaning *complete match*, with in-between values indicating intermediate similarity [14]. Hence, we may chose a threshold and suggest concept ω as a possible *match* for q only if the similarity between q and d_ω is above the threshold.

4.2 Suggestion by Normalised Textual Matching

Considering the relatively small size of both captions and concept descriptions, it is computationally inexpensive to calculate their semantic similarity. Using an Intel© Xeon© CPU 5150 processor with 2GB of RAM, running under Microsoft Windows XP 2002 SP2, we can calculate the semantic similarity between a single caption and *all* of the 525 concept descriptions in less than a few hundred milliseconds. However, another strategy that we chose to assess due to its simplicity and extremely fast performance was the straight textual matching between the captions and the concept vocabulary.

As in the case of the semantic similarity, this second strategy begins by normalising the caption. Yet, in this case, we also normalised the concept vocabulary. As a second step, we look for exact matches between the words in the normalised caption and those available in the normalised vocabulary. For illustration purposes, Table 2 displays an example of a caption, its normalised version and the resulting matches with the concept vocabulary.

Original caption:	*Soccer Italy training—Italian forward Alessandro Del Piero of Juventus Turin practices his penalties during training at Wembley Stadium this afternoon, 11 February, before tomorrow's World Cup qualifying match against England.*
Normalised caption:	*soccer itali train italian forward alessandro piero juventu turin practic penalti train wemblei stadium afternoon tomorrow world cup qualifi match england*
Concept vocabulary:	`abbey ... cup ... soccer ... stadium ...`
Normalised vocabulary:	`abbei ... cup ... soccer ... stadium ...`
Matches:	`cup, soccer, stadium`

Table 2. Example of Normalised Textual Match

In the case of concept names made of more than one word—such as, `davis_cup`—different heuristics may be applied. We may look for precise matches of all the words contained in the concept name, which would limit the number of matches considerably, but would ensure that only captions referring explicitly to the concept name are matched.

A more relaxed approach would be to select one single word as the *headword*. For instance, we may say that `davis` is the headword for the concept `davis_cup`, and we will automatically associate all the matches of `davis` with this concept. This is the approach that we have pursued.

Due to space limitations, we cannot list the headwords for all of the concepts in the VITALAS vocabulary in these pages. Readers, however, are welcome to visit the authors' website for further details on this matter [15]. It should be observed that for certain concepts—such as `ac_milan_soccer`—two headwords have been selected—`milan` and `soccer`—though they do not necessarily have to appear together to provide a match—an appearance of either `milan` or `soccer` would provide a match for the concept `ac_milan_soccer`.

In the following section, we report on the use of different thresholds for this strategy. Of course, as we lower the threshold a larger number of false positives is suggested to the users. Nevertheless, *recall* also increases, which is more important than *precision* for concept suggestion, because users will benefit from being able to choose from a variety of possible additional concepts and can easily reject unsuitable ones.

5 Evaluation

Precision and *recall* [16] are two standard measures used in information retrieval to evaluate performance. Precision and recall are defined in terms of a set of retrieved documents and a set of relevant documents. Given the particular characteristics of the multimedia archive that we have used, and the conditions of the ground-truth annotation that we have exploited, we have taken a modified version of the traditional definitions of precision and recall.

For the remainder of this document, we refer to the *recall for caption* q ($\mathbb{R}_q$), and the *precision for caption* q ($\mathbb{P}_q$), as

$$\mathbb{R}_q \equiv \frac{\mid \Omega_A^q \cap \Omega_M^q \mid}{\mid \Omega_A^q \mid},$$

$$\mathbb{P}_q \equiv \frac{\mid \Omega_A^q \cap \Omega_M^q \mid}{\mid \Omega_M^q \mid},$$

where Ω_A^q is the set of concepts that the annotators associated with the picture whose caption is q, and Ω_M^q is the set of concepts that our automatic suggestion strategy proposed for the picture whose caption is q.

Averaging over the total number of pictures in the collection $\mathfrak{C}$, we made use of the following definitions for *micro-recall* ($\mu_{\mathbb{R}}$), *micro-precision* ($\mu_{\mathbb{P}}$), *macro-recall* ($\mathcal{M}_{\mathbb{R}}$) and *macro-precision* ($\mathcal{M}_{\mathbb{P}}$) [17],

$$\mu_{\mathbb{R}} \equiv \frac{\sum\limits_{q \in \mathfrak{C}} |\Omega_A^q \cap \Omega_M^q|}{\sum\limits_{q \in \mathfrak{C}} |\Omega_A^q|} \quad , \quad \mu_{\mathbb{P}} \equiv \frac{\sum\limits_{q \in \mathfrak{C}} |\Omega_A^q \cap \Omega_M^q|}{\sum\limits_{q \in \mathfrak{C}} |\Omega_M^q|}$$

$$\mathcal{M}_{\mathbb{R}} \equiv \frac{\sum\limits_{q \in \mathfrak{C}} \mathbb{R}_q}{|\mathfrak{C}|} \quad , \quad \mathcal{M}_{\mathbb{P}} \equiv \frac{\sum\limits_{q \in \mathfrak{C}} \mathbb{P}_q}{|\mathfrak{C}|}.$$

Table 3 displays the values of each of the measures defined above for the two suggestion strategies described in Section 4—the highest values achieved for each measure are presented in bold font.

Measure	Threshold	Semantic Similarity	Textual Match
$\mu_{\mathbb{R}}$			0.75
$\mu_{\mathbb{P}}$			0.17
$\mathcal{M}_{\mathbb{R}}$			0.77
$\mathcal{M}_{\mathbb{P}}$			0.21
$\mu_{\mathbb{R}}$	0.2	0.34	
$\mu_{\mathbb{P}}$	0.2	**0.26**	
$\mathcal{M}_{\mathbb{R}}$	0.2	0.35	
$\mathcal{M}_{\mathbb{P}}$	0.2	**0.35**	
$\mu_{\mathbb{R}}$	0.15	0.49	
$\mu_{\mathbb{P}}$	0.15	0.17	
$\mathcal{M}_{\mathbb{R}}$	0.15	0.49	
$\mathcal{M}_{\mathbb{P}}$	0.15	0.27	
$\mu_{\mathbb{R}}$	0.1	0.65	
$\mu_{\mathbb{P}}$	0.1	0.08	
$\mathcal{M}_{\mathbb{R}}$	0.1	0.66	
$\mathcal{M}_{\mathbb{P}}$	0.1	0.16	
$\mu_{\mathbb{R}}$	0.05	**0.83**	
$\mu_{\mathbb{P}}$	0.05	0.02	
$\mathcal{M}_{\mathbb{R}}$	0.05	**0.84**	
$\mathcal{M}_{\mathbb{P}}$	0.05	0.06	

Table 3. Evaluation Results

Semantic similarity with a low threshold—0.05—performs better than normalised textual matching on the recall-based measures. High recall is more important than high precision for concept suggestion, since users will benefit from being able to choose from a range of possible additional concepts and can easily reject unsuitable ones.

Normalised textual matching performed better than semantic similarity when the threshold is above 0.05. Indeed, if we computed a *micro F-measure*, defined as $\mu_F \equiv \frac{2\mu_{\mathbb{P}}\mu_{\mathbb{R}}}{\mu_{\mathbb{P}}+\mu_{\mathbb{R}}}$, and a *macro F-measure*, defined as $\mathcal{M}_F \equiv \frac{2\mathcal{M}_{\mathbb{P}}\mathcal{M}_{\mathbb{R}}}{\mathcal{M}_{\mathbb{P}}+\mathcal{M}_{\mathbb{R}}}$, we would realise that the F-based measures for normalised textual matching—namely, $\mu_F = 0.28$ and $\mathcal{M}_F = 0.33$—are better than those of semantic similarity at any threshold.

Apart from its simplicity, one of the advantages of normalised textual matching is the fact that this technique is the fastest one to execute. Therefore, we envisage that an improved version of the normalised textual matching approach, where the annotators pick up the headwords, and possibly extend them manually to better reflect concept relations, may yield very good results.

6 Related Work

The study published by Hoogs *et al.* [18] is among the first ones to add semantics to concept detection, by establishing links with a general-purpose ontology, which connected a limited set of visual attributes to *WordNet* [19]. However, combining low-level visual attributes with concepts in an ontology is a rather difficult task, due to the so-called *semantic gap* between them [20].

To cope with the demand for ground-truth, Lin *et al.* initiated a collaborative annotation effort for the *TRECVID 2003 benchmark* [21]. Using tools from Christel *et al.* [22] and Volkmer *et al.* [23], a common annotation effort was again made for the *TRECVID 2005 benchmark*, yielding a large set of annotated examples for 39 concepts taken from a predefined collection [7]. We have provided a larger compilation, increasing the concept vocabulary to 525 concepts, and getting 1,000 annotated pictures per concept.

The work reported by Snoek *et al.* [24] is closely related to ours. They did implement a concept suggestion strategy based on semantic similarity; yet, they made use of the *Lucene search engine* [25] as part of their implementation, and the goal of their study was different, as they attempted to obtain semantic descriptions and structure from WordNet. The results presented by Snoek *et al.* are not conclusive, but we may consider following their recommendations on *ontology querying* and *Resnik's measure of information content* [26] in future versions of our research.

7 Conclusions

We have described two methods of concept suggestion, with the aim of helping multimedia search engine users enhance their initial keyword queries with additional terms corresponding to system-known concepts, namely suggestion by semantic similarity and normalised textual matching. Although normalised textual matching was the simpler technique, it performed much better on the recall-based measures at most of the thresholds tried for the semantic similarity approach, and only slightly less well on the precision-based measures. As explained in Section 5, high recall is more important than high precision for query term suggestion, since the user will benefit from being able to choose from a range of possible additional concepts, and can easily reject unsuitable ones. However, in approaches such as the one proposed by Palomino *et al.* [27], where discovered additional concepts are automatically added to the query without prior user approval, precision is more important, since the inclusion of non-relevant concepts in the query can severely degrade performance.

Even though we have evaluated the quality of our concept selection using recall- and precision- based measures, we still need to measure the effect of our concept suggestion facilities on the overall search engine performance. Such evaluations are being undertaken by our research partners at the *Institut National de l'Audiovisuel* [28], using recall, precision, and subjective measures of user satisfaction with the overall system.

7.1 Future Work

In future versions of our suggestion engine, we are considering to determine the degree of association between every pair of concepts by means of a concept-to-concept similarity matrix. To produce each entry in this matrix, we plan to represent the relation between each concept and all of the captions in different vectors. The entries in these vectors would contain the frequency of appearance of the concept headwords in each caption of the collection. Then, the similarity of a pair of concepts would be given by the cosine similarity of their corresponding vectors. For instance, the most similar VITALAS concepts to `actress` would be `filming` (0.40), `film_festival_cannes` (0.39), `festival` (0.27), `academy_award` (0.15), `actor` (0.15) and `award` (0.15).

Once the similarity matrix is created, we can suggest concepts relevant to a particular picture by scanning its caption and searching for occurrences of the concept headwords on it. Afterwards, we derive from the similarity matrix the "similarity" between the appearing headwords and all the concepts in the vocabulary. As in the case of the semantic similarity approach, we define a threshold and suggest to the user only those concepts whose corresponding values are above the threshold.

By generating a similarity matrix automatically from Belga's specific documents, we expect to produce concept suggestions adapted to that particular domain much better than if we had used the term relations available from a thesaurus or library classification.

Acknowledgments. This research was supported under the EU-funded VITALAS project—project number FP6-045389. The authors are very grateful to Belga News Agency for providing the data used to carry out their research, and to the following people for participating in the ground-truth annotation: Jonathan Dujardin, Jeroen Van Den Haute, Katrien Vanmechelen, Sini Cheang and Hecham Azarkan.

References

1. Google: *Web, Image and Video Search.* `http://www.google.com/`.
2. YouTube: *YouTube, LLC.* `http://www.youtube.com/`.
3. Blinkx: *Video Search Engine.* `http://www.blinkx.com/`.

4. Worring, M., Snoek, C.G.M., Huurnink, B., van Gemert, J.C., Koelma, D.C., de Rooij, O.: The MediaMill Large-Lexicon Concept Suggestion Engine. In *Proceedings of the 14th ACM International Conference on Multimedia*, Santa Barbara, CA, Association for Computing Machinery (October 2006) 785–786
5. VITALAS: *Video and Image Indexing and Retrieval in the Large Scale.* `http://www.vitalas.org/`.
6. Belga: *Belga News Agency.* `http://www.belga.be/`.
7. Naphade, M., Smith, J.R., Tesic, J., Chang, S.F., Hsu, W., Kennedy, L., Hauptmann, A., Curtis, J.: Large-Scale Concept Ontology for Multimedia. *IEEE MultiMedia* **13**(3) (2006) 86–91
8. Snoek, C.G.M., Worring, M., van Gemert, J.C., Geusebroek, J.M., Smeulders, A.W.M.: The Challenge Problem for Automated Detection of 101 Semantic Concepts in Multimedia. In *Proceedings of the 14th ACM International Conference on Multimedia*, Santa Barbara, CA, Association for Computing Machinery (October 2006) 421–430
9. Palomino, M.A., Oakes, M.P., Wuytack, T.: Automatic Extraction of Keywords for a Multimedia Search Engine Using the Chi-Square Test. In *Proceedings of the 9th Dutch-Belgian Information Retrieval Workshop (DIR 2009)*, Enschede, The Netherlands (February 2009) 3–10
10. Buckley, C.: Implementation of the SMART Information Retrieval System. Technical Report TR85-686, Computer Science Department, Cornell University, Ithaca, New York (May 1985)
11. Porter, M.: An Algorithm for Suffix Stripping. *Progam* **14**(3) (July 1980) 130–137
12. Gelbukh, A., Sidorov, G.: Zipf and Heaps Laws' Coefficients Depend on Language. In *Proceedings of the Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City (February 2001) 332–335
13. Salton, G., Wong, A., Yang, C.: A Vector Space Model for Automatic Indexing. *Communications of the ACM* **18**(11) (November 1975) 613–620
14. Widdows, D.: Measuring Similarity and Distance. In *Geometry and Meaning*, CSLI Publications (November 2004)
15. Palomino, M.A.: *VITALAS Concept-Suggestion Engine.* `http://osiris.sunderland.ac.uk/~cs0mpl/VITALAS/`.
16. Belew, R.K.: *Finding Out About: A Cognitive Perspective on Search Engine Technology and the WWW.* Cambridge University Press, Cambridge, UK (February 2001)
17. Joachims, T.: *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms.* Kluwer Academic Publishers (April 2002)
18. Hoogs, A., Rittscher, J., Stein, G., Schmiederer, J.: Video Content Annotation Using Visual Analysis and a Large Semantic Knowledgebase. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Madison, WI (June 2003) 327–334
19. Fellbaum, C.: *WordNet: An Electronic Lexical Database.* MIT Press, Cambridge, MA (May 1998)
20. Dorai, C.: Bridging the Semantic Gap in Content Management Systems: Computational Media Aesthetics. In *Proceedings of the International Conference on Computational Semiotics for Games and New Media*, Amsterdam, The Netherlands, Kluwer Academic Publishers (September 2001) 94–99
21. Lin, C.Y., Tseng, B.L., Smith, J.R.: Video Collaborative Annotation Forum: Establishing Ground-Truth Labels on Large Multimedia Datasets. In *Proceedings of the TRECVID 2003 Workshop*, Gaithersburg, MD (November 2003)

22. Christel, M., Kanade, T., Mauldin, M., Reddy, R., Sirbu, M., Stevens, S., Wactlar, H.: Informedia Digital Video Library. *Communications of the ACM* **38**(4) (April 1995) 57–58
23. Volkmer, T., Tahaghoghi, S., Thom, J.A.: Modelling Human Judgement of Digital Imagery for Multimedia Retrieval. *IEEE Transactions on Multimedia* **9**(5) (August 2007) 967–974
24. Snoek, C.G., Huurnink, B., Hollink, L., de Rijke, M., Schreiber, G., Worring, M.: Adding Semantics to Detectors for Video Retrieval. *IEEE Transactions on Multimedia* **9**(5) (August 2007) 975–986
25. Lucene: *The Lucene search engine.* `http://lucene.apache.org/`.
26. Resnik, P.: Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal, Canada (1995) 448–453
27. Palomino, M.A., Oakes, M.P., Xu, Y.: An Adaptive Method to Associate Pictures with Indexing Terms. In *Proceedings of the 2nd International Workshop on Adaptive Information Retrieval*, London, UK (October 2008) 38–43
28. INA: *Institut National de l'Audiovisuel.* `http://www.ina.fr/`.