Collaborative Semantic Annotation of Images: Ontology-Based Model

Damien Zomahoun

COLLABORATIVE SEMANTIC ANNOTATION OF IMAGES: ONTOLOGY-BASED MODEL

Damien E. ZOMAHOUN

Le2i – Laboratory, University of Burgundy, Dijon, FRANCE

ABSTRACT

In the quest for models that could help to represent the meaning of images, some approaches have used contextual knowledge by building semantic hierarchies. Others have resorted to the integration of images analysis improvement knowledge and images interpretation using ontologies. The images are often annotated with a set of keywords (or ontologies), whose relevance remains highly subjective and related to only one interpretation (one annotator). However, an image can get many associated semantics because annotators can interpret it differently. The purpose of this paper is to propose a collaborative annotation system that brings out the meaning of images from the different interpretations of annotators. The different works carried out in this paper lead to a semantic model of an image, i.e. the different means that a picture may have. This method relies on the different tools of the Semantic Web, especially ontologies.

KEYWORDS

Ontology, *Semantic Annotation, Retrieval & images.*

1. INTRODUCTION

Today, images are used in many circumstances, either professional or private. According to the context of use, users have different needs, requirements and constraints. The images were often annotated by a set of keywords, whose relevance related only to one description. The nuance is that an image receives many interpretations. To solve these problems, image annotation systems by the contents have been introduced [7], [8], [9]. The goal of new researches is to annotate images using only their visual content. There are two kinds of techniques: techniques based on the symbolic content and those based on the semantic content [10], [11], [12], [13], [23].

Symbolic content-based techniques use descriptors extracted automatically from images such as color, texture, shape, etc. Despite the different methods and approaches proposed, the meaning of the images remains a highly argued topic. Extensive experiments on image retrieval system by content showed that low-level contents (color, texture and shape) often fail to describe the images [10], [11]. Then, the advent of techniques based on the semantic content has come up to improve the results of image searching. The extraction of images semantics is a very sensitive process that deserves special attention. To compose this semantic, this includes the processes of organization and a selection of the most relevant information.

In this paper, we propose a collaborative approach for image annotation. As in some more recent approaches, it relies on images semantic. Annotators, whose goal is to annotate images, will use it to propose keywords. These proposals will be weighted using the weighting model based on the frequency of occurrence and terms rank. The choice of this technique is justified in the Section IV. Finally, the most representative instances of concepts will be included in the ontology. The rest of the paper is organized as follows. After presenting existing approaches in Section 2, Section 3 and 4 detail our approach and in Section 5 we discuss on the implementation of our model. Finally, section 6 provides the conclusion and outlines our future work.

2. EXISTING APPROACHES

Several image annotation systems have been presented in the literature [1], [2], [3]. The first systems appeared in the 90s, and are text-based [4], [5], [6]. They adopt an approach that consists in describing the visual content by keywords. These keywords are used as indexes to access to the related visual information. The advantage of this approach is that it allows providing access to databases using standard query languages such as SQL. However, this approach requires a lot of manual processing. In addition, the reliability of descriptive data is not assured: they are subjective and may not accurately describe image content.

The use of semantic annotations of images has become a highly explored in information retrieval field. Von-Wun Soo et al. [15] proposed a method of annotation and indexing based on RDF graphs. In this approach, an analyzer (Case-Based Learning (CBL)) is used on one hand to allow converting the concepts proposed in RDF graph, and on the other hand to convert the requests written in a natural language in SPARQL requests. N. Maguesh and al. [18] also present an approach which describes the visual characteristics of an image based on ontology and is similar to the approach proposed by Soo and al. The difference between the two approaches is that the technique from Maguesh et al. does not use keywords, but ontology.

Other techniques, translate keyword requests into formal SPARQL requests. The approach proposed by A. Latreche and al. in [19] involves three phases: the mapping of the keywords to RDF concept, the building of new requests during the mapping and the classification of the requests which consist in identifying the request which responds better to the user's needs.

Some researchers have looked towards automatic annotation. They propose a model that assigns automatically the terms to the images. The goal of these approaches is to estimate if the underlying information contained in an ontology created from a vocabulary of terms can be effectively used together with the visual information extracted from images to produce more accurate annotations. Among these approaches, we can mention the proposal given in [20] and [21]. Authors of [20] propose methods to use a hierarchy derived from ontology to improve automatic image annotation and retrieval. Authors of [21] present M-OntoMat-Annotizer, a tool covering the step of knowledge acquisition for automatic annotation of multimedia content. The tool allows to extract MPEG-7 visual descriptors from both images and videos and to store these descriptors as the so-called visual prototypes of ontology classes.

To improve image retrieval, other approaches have tried to make a combination of techniques such as segmentation of the images into regions and annotation of each region. This is the case for the approaches from Wang et al. in [19], Ning Ruan in [18] and H. Jair and al. in [22]. The approaches proposed in [18] and [22] consist in annotating image regions by using ontologies. The difference between [18] and [22] approaches is that in [22], the images are firstly segmented into regions, whereas in [18], homogeneous regions are recovered thanks to an unsupervised algorithm. In addition, the ontology proposed in [19] consists of three sub-ontologies: domain ontology of animals containing animals' taxonomy derived from WordNet, a description ontology encapsulating descriptions of animals and visual description ontology containing concepts derived from the properties of the images such as contours and color histograms.

3. EMERGSEM APPROACH

Semantic content annotation is the basis for semantic content retrieval. Annotating is the process of adding content-descriptive keywords to image [23]. The techniques we have just cited above aim improving image annotation and perform images retrieval process. However, they don't take into account the authentic meaning of images. As annotators have their own understandings for similar images (each annotator has an interpretation of the images), they often propose the meanings among different data sources. Thus resorting to data sources does not always work well. Figure 1 shows the annotations provided by three annotators.

		
Annotator 1: Persons Dog Sea Tree	**Annotator 2**: Man Woman Beach Dog Sea Forest	**Annotator 3**: Humans Dog Sea Tree Sky House

Figure 1. An image annotated by different annotators

We propose ontology into our system because in ontology, the data are structured. Ontology provides a formal context that may contain explicit definitions of semantics. It is used for concepts modeling (objects) and allows to represent the different types of relationships between image features such as regions. Ontology can be directly processed by a machine, and at the same time can help to extract implicit knowledge through automatic inference [29], [30]. The main contribution of this paper is to propose a collaborative annotation approach based on the ontologies. The collaborative annotation consists of proposing an image to an annotators group. Although there is a tradeoff between rapidity (automatic annotation poses the problem of disambiguation) and precision (manual annotation requires enormous personal and financial means), semi-automatic annotation is still viable choice.

The architecture of the suggested system is shown in Figure 2. The goal of collaborative process is to define the emergent semantics of the image thanks to ontology. The collaborative annotation process is further divided in the following steps:

1.) Ontology model and lexical dictionary are proposed to annotators;
2.) Annotators propose instances indicating the represented image content by ontology concepts;
3.) Once the instances are attributed and store in triples as xml file, image meanings is obtained;
4.) Computation of meanings similarity (detailed in Section 4.3);
5.) Finally, semantic resulting is displayed.

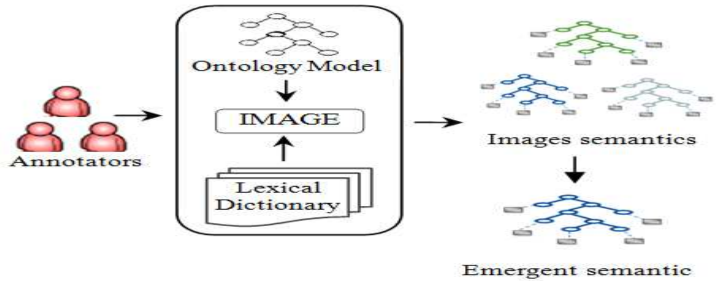

Figure 2. EMERGSEM approach schema

4. OVERVIEW OF THE SYSTEM

4.1. Keywords Usage

Collaborative annotation of image consists submitting the images to a group of annotators. It requires that each annotator assigns/associates one or more keywords to the image by the ontology. To carry out the collaborative annotation, we provide a lexical dictionary to annotators. The use of this dictionary to annotate image highlights a question on effect it products on presently used annotation systems. The answer to this question orients us towards the work which has previously been carried out to resolve the many issues and interactions among variables as subjects, annotators need relevant terms, precision level desired, automatic and intellectual control, and annotators need or desire for special control. At the end of the instantiation, the concepts of ontology are instantiated with the keywords provided by the annotators.

4.2. Image Meanings

Image meanings are obtained through the ontology after keywords attribution to image objects by instantiation process of ontology concepts. Using these semantics, images can be annotated and indexed. The annotation consists in mapping the proposed meanings to images.

The advantage of EMERGSEM approach is that an image can have several meanings. The figure 4 gives an overview of the annotations we can obtain at the end of instantiation process of the image represented in Figure 1.

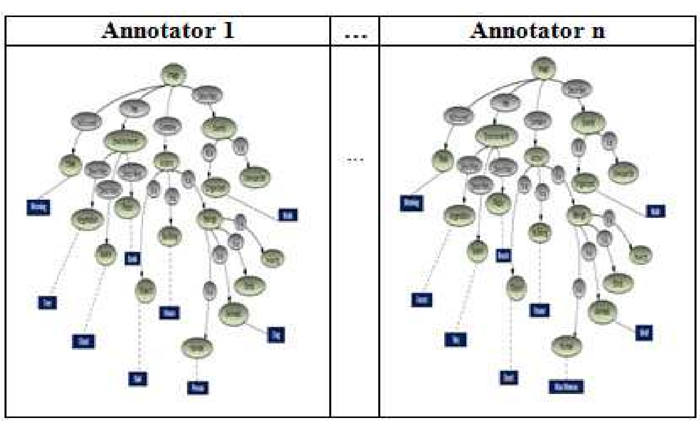

Figure 4. Example of meanings proposed by annotators

EMERGSEM system determines the most relevant semantic(s) by meanings similarity measure. A similarity measure usually stands in the core of an ontology matching procedure. Our approach focuses on the work which introduces instance-based similarity measures in [26]. It uses variable selection in order to represent concepts as sets of characteristic features. The method allows to solve the problem of meanings matching. The goal is to determinate the resulting meaning by taking into account each meaning proposed by annotator.

4.3. Similarity Measure

Instances selection techniques in [27] are used to classify the instances of each ontology concepts, according to the weighting. The goal is to determinate the most representative semantic assigned to an image. Let us recall that the number of proposed meanings to an image is equal to the number of annotators who instantiated the ontology concepts.
Take the example of k meanings represented by Mk (derived from the provided ontology model O) with c_i, their corresponding sets of concepts and $\gamma_{k,i}$, the instances of each concept. The instances of a proposed meaning Mk could be represented by Mk = $\{\gamma_{k,1}, \gamma_{k,2}, \ldots, \gamma_{k,i}\}$. Our approach consists in weighting instances from identical concepts of each provided meanings.

The weighting of semantics depends on the instances weighting representing each concept of an ontology. It is usually based on the frequency of occurrence of the instances [25]. Unfortunately, while a (medium or low frequency) keyword that appears frequently in an image is likely to be more important than one that occurs rarely, experiences confound this effect. To improve instances evaluation, the measure of the average weight corresponding to the position of the concerned instances will be added to the frequency of occurrence. The weight is given to the instances following the order of choice. The advantage of this method is that it considers that the order of appearance of the terms plays a very important role, because the instance selected first should have a more significant evaluation with respect to an instance at the end of the list.

$$\textbf{Eval term}_i = (\textbf{Frq}_i) * (\sum \textbf{Weight}_{ij} / \textbf{N}_i) \quad (1),$$

with Frq_i: frequency of terms i.
$Weight_{ij}$: the weight of $Term_i$ for ranking position j
N_i: the number of occurrences of $Term_i$.
The frequency is given by: $Frq_i = i / \sum i$.

Table 1 presents the instances weighting. The frequency, weight average and evaluation value of each instance is shown. The assignments are ordered from rank or position 1 to n. A weight is assigned to instances is calculated by:

$$\textbf{Weight} = (\textbf{Nb} - (\textbf{Rg i} - 1)) / \textbf{Nb} \quad (2),$$

where Nb is the maximum number of instances assigned by an annotator and Rg_i is the instances position.

For the same concepts of interest, $c_1 \in (M_1, M_2)$, we carry out an instance procedure independently on each of their corresponding sets and evaluate the instances by their importance. In consequence, the concepts c_1 can be represented by the relevant instance of the list of their corresponding instances weightings.

A comparing procedure will be defined as a procedure ƛ ; which takes many meanings (M_1, M_2, M_3, …, M_i) of the same image and produces another one meaning M_E that we call "emergent semantic", by comparing the instances of each concept (of the meanings) proposed by annotators. In fact, the resulting merged meaning M_E is all meanings M_1, M_2, M_3, …, M_n, enriched with links to their concepts [28].

To indicate that the procedure described above is applied on the meaning M_1, M_2, M_3,…, M_n resulting in an output meanings M_E, we will use the denotation ⅄ (M_1, M_2, M_3, …, M_n) = M_E. The semantics are the resultant of all the meanings proposed by the annotators as shown in Figure 5.

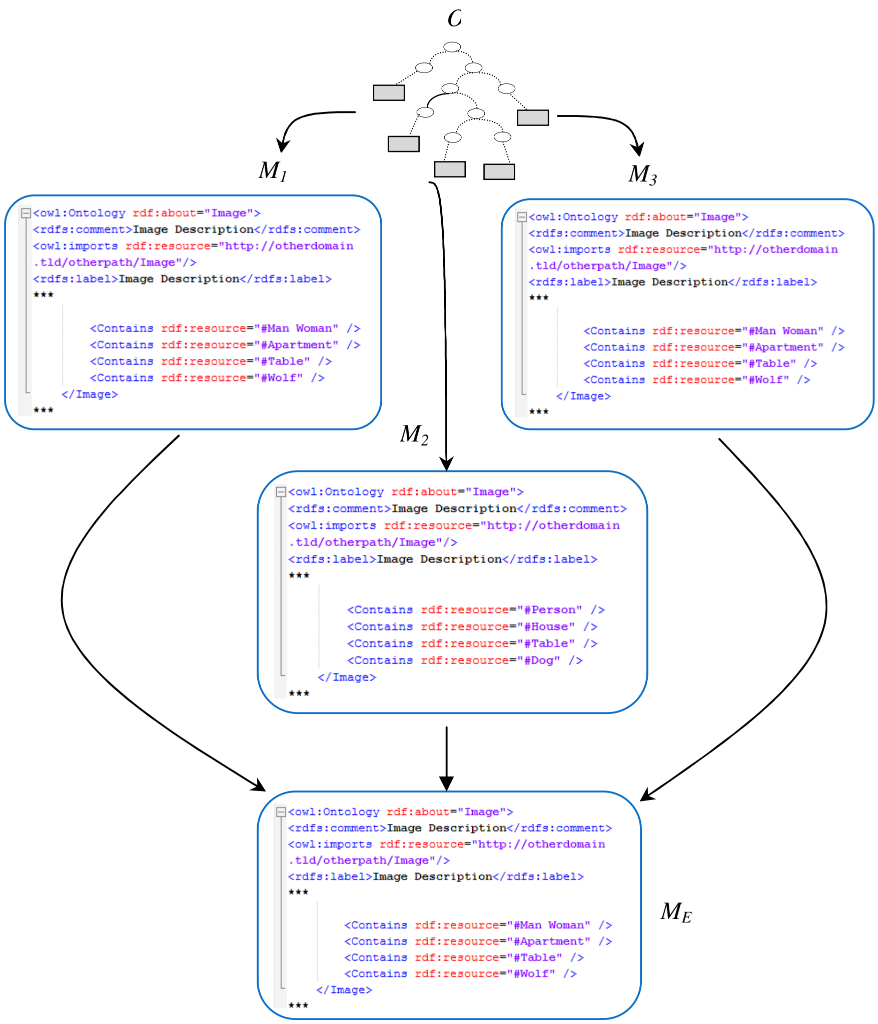

Figure 5. Resulting meaning determination

5. DISCUSSION

To properly annotate an unknown image, its semantic should be obtained. In this work the resulting semantic is extracted based on further analysis of the different meanings proposed by annotators.

The model proposed in this paper is simple and offers many advantages. It allows to annotate any kind of image in any field. It focuses to extract the shared semantic of image. Semantics proposed by annotators representing image semantics are matched. We provide an evaluation of the suggested matching approach. In our experimentation, we consider the concepts from the ontology, respecting several criteria as we show in Table 1, Table 2 and Figure 6.

The process of annotating an unknown image is implemented in steps. Instances of concepts are provided by the annotators (according to the ontology model) to obtain the possible semantics of images (Table 1). The similarity measure that we used, allows to extract a relevant instance of each concept (Table 1 and Figure 6).

Table 1. Example of semantics proposal

	Instances of concepts				
Meaning M_1	Woman.Man	Dog	Tree	House	Ball
Meaning M_2	Persons	Wolf	Forest	Apartment	Ball
Meaning M_3	Persons	Dog	Forest	Apartment	Bowl
Meaning M_4	Child	Wolf	Plant	Apartment	Ball
Meaning M_5			***		
Ontology M_n	Woman.Man	Dog	Forest	House	Apple

The weightings used in similarity measure are computed by evaluation method based on frequency and weight (Section 4.3). The evaluation relies on the training data set provided.

Table 2. Instances weighting

Concept	Instances	Frequency	Weight	Evaluation
Animal	Dog	0,1022	0,8100	0,0828
	Wolf	0,0114	0,6700	0,0076
	Cat	0,0114	1,0000	0,0114
Human	Woman.Man	0,1022	0,6211	0,0634
	Persons	0,0114	0,3300	0,0037
	Child	0,0114	1,0000	0,0023
Object	Ball	0,1022	0,7311	0,0747
	Bowl	0,0114	0,7500	0,0085
	Apple	0,0114	1,0000	0,0114
Building	Apartment	0,0795	0,6814	0,0542
	House	0,0227	0,5800	0,0132
	Pavilion	0,0114	1,0000	0,0036
Vegetation	Forest	0,0682	0,6783	0,0462
	Tree	0,0341	0,8300	0,0282
	Plant	0,0114	0,9200	0,0105

Figure 6 shows the evaluations of instances. As mentioned above, this evaluation takes into account the frequency and weight of instances. Instances that have high weighting are extracted to represent the ontology concepts.

The combination of these methods calls for the development of more improvement evaluation that allow to extract efficiently the relevant instances that are of interest to users rather than being "hard-wired" into the annotation engines provided by most of the current vendors that, primarily, focus automatic annotation. The second requirement of interactivity also calls for the development of tools allowing users to provide inputs into the annotation process in an interactive manner, preferably via some well-defined user interface.

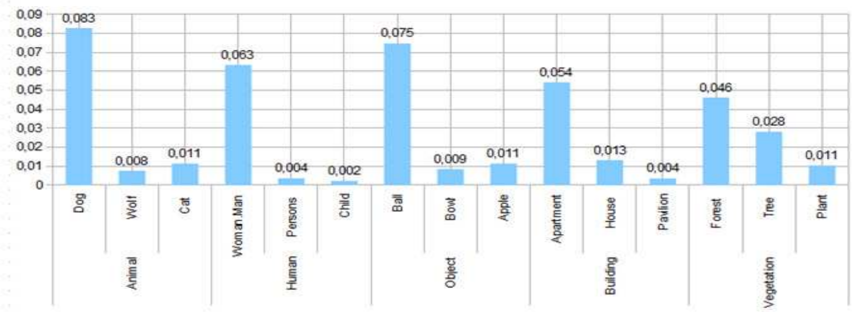

Figure 6. Instances Evaluation

The Figure 7 is an example of relevant instances from the Figure 6.

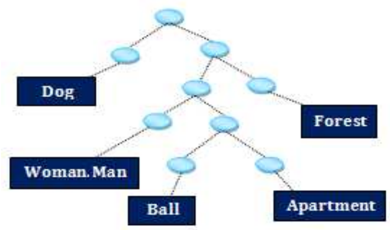

Figure 7. Relevant instances determination

Our training dataset consists of the training instances proposed by annotators. The relevant $\gamma_{k,i}$ represent the instances that describe the concepts of ontology and the obtained semantic becomes the emergent semantic of image.

The collaborative approach that we propose in this paper focuses on the performance. Image annotation will not be the work of one person, but the contribution of annotators group. Each annotator will therefore propose a meaning of a given image. To these meanings derives the resulting semantic. The objective of this approach is not just to get the meanings of an image, but its shared semantics.

Many important problems still need to be addressed like, for instance, populating each ontology with existing or built image datasets, deciding on an appropriate representation of the instances, solving various complexity issues in the matching process related to the number of concepts and instances in the targeted ontologies. Although our preliminary experimental results are encouraging, the work of implementing and evaluating the propositions of this paper on a larger scale is still in progress.

We have implemented an annotation system. We integrate in the system, classical and collaborative annotations with ontology to evaluate the relevancy of our approach (Figure 9 and 10). When we compare the two systems, we note that the collaborative system produces the shared semantic of image and facilitates images retrieval. The Table 3 shows the comparing results of our experiment.

Table 3. Comparing data of semantic annotations

	Classic Annotation	Collaborative Annotation
Complexity (second)	12,25	37,61
Performance (second)	08,17	04,97
Error reduction (%)	0,418	0,012

For the complexity, we take into account annotation time average, and retrieval time for the performance. The graph of comparative parameters is represented (Figure 8).

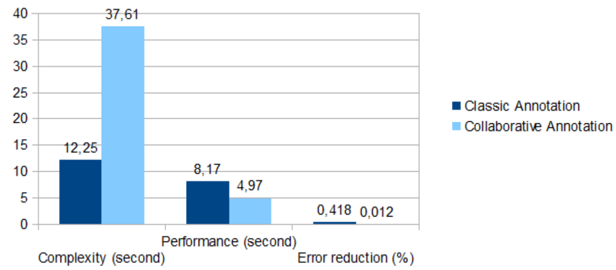

Figure 8. Comparing graph of Table 3

We note that the collaborative annotation require much more time than the other annotations. By comparing the performance and error curves, we observe that they are similar, indicating that more errors numerous, more the retrieval is difficult.

6. Conclusion And Future Works

The different images annotation approaches we studied, have their advantages and disadvantages. Approaches using the keywords are easy to apply with an acceptable accuracy of image retrieval, whereas those using ontologies are semantically rich. They meet the needs of full descriptions of images retrieval and improve the accuracy of the retrieval. Although there is a tradeoff between complexity and performance, collaborative annotations are still viable choices when better performance is considered.

Our system has been tested on a database of images and the results showed that, compared to existing annotation systems, collaborative annotation ensures accuracy regarding the real meaning of the images and also facilitates their retrieval.

Our next work will base on semantic propagation. As the images we annotated are classified, we thought it would be indispensable to proceed to the propagation of semantic annotations once an image is annotated. So, we will need objects recognition algorithm and another for objects position recognition. Then, we will propose a semantic indexing method. The goal is to exploit the advantages that ontology offers to us for images retrieval. We made and assume this choice because this technique relies on the fact that the semantic of the images must take into account different proposal of semantics suggested by the annotators. To achieve this, we will study the importance of the users and their preferences during indexing images. The Figure 9 and 10 represent one part of collaborative annotation process.

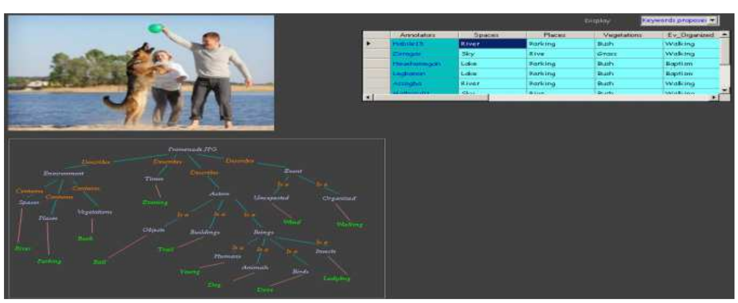

Figure 9. Image semantic determination

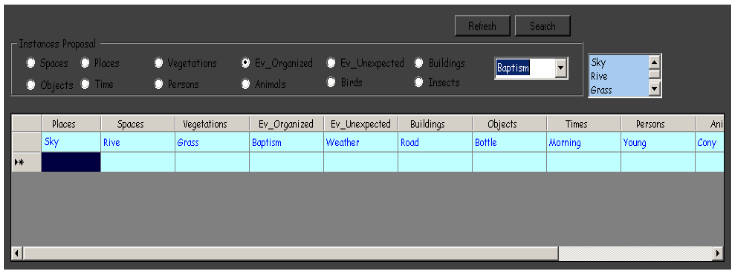

Figure 10. Image retrieval

ACKNOWLEDGMENTS

This work has been funded by Electronic, Computer Science and Image Laboratory (LE2I) and by Doctoral School SPIM, FRANCE.

REFERENCES

[1] P. Aigrain et al. Content-based representation and retrieval of visual media: a-state-of- the-art review. Multimedia Tools and Applications, 3(3) :179–202, septembre 1996.

[2] J.J. Fan and K.Y. Su. An efficient algorithm for matching multiple patterns. IEEE Trans. Knowl. Data Eng., 5(2) :339–351, avril 1993.

[3] G. Salton and M. J. McGill. An introduction to modern information retrieval. McGraw-Hill, New York, 1983.

[4] A. D. Bagdanov, M. Bertini, A. D. Bimbo, G. Serra, and C. Torniai. Semantic annotation and retrieval of video events using multimedia ontologies. In ICSC, pages 713–720, 2007.

[5] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan. Matching words and pictures. JMLR, 3:1107–1135, 2003.

[6] S. Dasiopoulou and I. Kompatsiaris. Trends and issues in description logics frameworks for image interpretation. In Artificial Intelligence: Theories, Models and Applications, volume 6040, pages 61–70. 2010.

[7] S. Bres, J.M. Jolion, F. Lebourgeois « Traitement et analyse des images numériques» Hermes Lavoisier 2003.

[8] I.T. Young, J.J. Gerbrands et L.J. Van Vliet « Fundamentals of Image Processing » Université de Delft. (sur internet http:// www.ph.tn.tudelft.nl/ ~lucas /publications /1995/FIP95TYJGLV/FIP2.2.pdf)

[9] D. Lingrand « Introduction au Traitement d'images » Vuibert , 2004.

[10] Atanas Kiryakov, Borislav Popov, Ivan Terziev, Dimitar Manov, Damyan Ognyanoff, Semantic Annotation, Indexing, and Retrieval.

[11] Saša Nešic ,Mehdi Jazayeri, Fabio Crestani, Dragan Gaševic, Concept-Based Semantic Annotation, Indexing and Retrieval of Office-Like Document Units, 2010.

[12] Alain Boucher, Thi-Lan Le, Comment extraire la sémantique d'une image, 2005.

[13] Alain PUJOL Contributions à la Classification Sémantique d'Images, 2009.

[14] Svenonius, Elaine. Unanswered Questions in the Design of Controlled Vocabularies, J. Am. Soc. Inf. Sci., 37(5):331-340, Sept. 1986.

[15] Von-Wun Soo, Chen-Yu Lee, Chung-Cheng Li, Shu Lei Chen, Ching-chih Chen, "Automatic metadata creation: Automated semantic annotation and retrieval based on sharable ontology and case-based learning techniques". Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries, May 2003.

[16] N. Magesh, Dr. P. Thangaraj, Semantic Image Retrieval Based on Ontology and SPARQL, Query, International Conference on Advanced Computer Technology (ICACT), 2011.

[17] A. Latreche, A. Lehireche, and K. Benyahia, "Interrogation à base d'annotation sémantique", in proc. International Conference on Computer science and its applications (ciia), 2009.

[18] Ning Nuan, Ning Huang, Wen Hong, Geoscience And Remote Sensing Symposium, 2006. Igarss 2006. IEEE International Conference on July 31 2006-AUG. 4 2006.

[19] Wang H., Chia L. T., Liu S., « Semantic Retrieval with Enhanced Matchmaking and Multi-Modality Ontology ». IEEE International Conference on Multimedia and Expo. p. 516 – 519, 2007.

[20] Srikanth M., Varner J., Bowden M., Moldovan D.: Exploiting ontologies for automatic image annotation. In: Proceedings of the 28th International ACM Conference on Research and Development in Information Retrieval (SIGIR), New York, NY, USA, ACM Press (2005) 552-558.

[21] Saathoff, C., Timmermann, N., Staab, S., Petridis, K., Anastasopoulos, D., Kompatsiaris, Y.: M-ontomat-annotizer: Linking ontologies with multimedia low-level features for automatic image annotation. In: Demos and Posters of the 3rd European SemanticWeb Conference (ESWC). (2006)

[22] H. Jair Escalante, Carlos Hernández-Gracidas, Jesús A. González, Aurelio López, Manuel Montes, Eduardo Morales, Enrique Sucar, Luis Villaseñor,

[23] Dominic Mainz, Katrin Weller, Jürgen Mainz, Semantic Image Annotation and Retrieval with IKen, International Semantic Web Conference Posters & Demos, volume 401 of CEUR Workshop Proceedings, CEUR-WS.org, (2008).

[24] Haav, H. M., Lubi, T.-L.: A Survey of Concept-based Information Retrieval Tools on the Web. In Proc. of 5th East-European Conference ADBIS*2001, Vol 2., Vilnius "Technika", pp 29-41.

[25] Rana EL CHARIF, Analyse des paramètres de pondération dans le cadre de collections volumineuses, rapport de DEA d'Informatique, p. 23-31, 2006.

[26] K. Todorov, P. Geibel, and K.-U. K¨uhnberger. Extensional ontology matching with variable selection for support vector machines. In CISIS, pages 962–968. IEEE Computer Society Press, 2010.

[27] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. JMLR, 3(1):1157–1182, 2003.

[28] Nicholas James and Konstantin Todorov and Celine Hudelot, Ontology matching for the semantic annotation of images, Fuzzy Systems (FUZZ), IEEE International Conference, 2010.

[29] J. Hunter. Adding multimedia to the semantic web - building an mpeg-7 ontology. In SWWS, pages 261–281, 2001.

[30] B. Yao, X. Yang, L. Lin, M. W. Lee, and S. C. Zhu. I2th: Image parsing to text description. In Proc. of IEEE, 2009.

Born 13.04.1983. Ph.D. student at the University of Burgundy, Laboratory of Electronics, Informatics and Images (Le2i) in Dijon, France.

Damien. E. Zomahoun develops the annotation applications of images. He is an author of the publication concerning Exploiting Semantic Indexing Images for Emergence Recommendation Semantics System and Contributions of Semantic Web Tools to Relational Models.