

Sentic Album: Content-, Concept-, and Context-Based Online Personal Photo Management System

Erik Cambria · Amir Hussain

Received: 15 August 2011 / Accepted: 16 April 2012 / Published online: 3 May 2012
© Springer Science+Business Media, LLC 2012

Abstract The world of online personal photo management has come a long way in the past few years, but today, there are still huge gaps in annotating, organizing, and retrieving online pictures in such a way that they can be easily queried and visualized. Existing content-based image retrieval systems apply statistics, pattern recognition, signal processing, and computer vision techniques but these are still too weak to ‘bridge the semantic gap’ between the low-level data representation and the high-level concepts the user associates with images. Image meta search engines, on the other hand, rely on tags associated with online pictures but results are often too inaccurate since they mainly depend on keyword-based rather than concept-based algorithms. Sentic Album is a novel content-, concept-, and context-based online personal photo management system that exploits both data and metadata of online personal pictures to intelligently annotate, organize, and retrieve them. Many salient features of pictures, in fact, are only noticeable in the viewer’s mind, and the cognitive ability to grasp such features is a key aspect for accordingly analyzing and classifying personal photos. To this end, Sentic Album exploits not just colors and texture of online images (content), but also the cognitive and affective information associated with their metadata (concept), and

their relative timestamp, geolocation, and user interaction metadata (context).

Keywords Human computer interaction · Cognitive and affective information processing · Image affect · Image classification · Image features · Emotional semantic image retrieval · Sentic computing

Introduction

Thanks to the many social networking websites that allow users to easily upload and share personal pictures online, today most of the social interaction between web users is expressed through personal digital photos and the metadata associated with these. Publishing, adding descriptions, commenting, tagging, linking pictures online are among the most common activities performed on the Web, not just on specific photo-sharing websites such as Picasa¹ and Flickr,² but also on social networking websites like Facebook³ or MySpace.⁴ For the principle that a picture is worth a thousand words, in fact, the most popular user-generated content (UGC) is represented by images and their relative metadata rather than text, audio, or video. But, for the same principle, annotating, organizing, and retrieving these images in a way that they can be easily queried and visualized are very difficult tasks. In the past, content-based image retrieval (CBIR) systems and image meta search engines applied different techniques to extract meaning from image data and metadata, but none of these

E. Cambria (✉)
Temasek Laboratories, National University of Singapore,
117411 Singapore, Singapore
e-mail: cambria@nus.edu.sg
URL: <http://sentic.net>

A. Hussain
Department of Computing Science and Mathematics,
University of Stirling, Stirling FK9 4LA, UK
e-mail: ahu@cs.stir.ac.uk

¹ <http://picasa.google.com>.

² <http://flickr.com>.

³ <http://facebook.com>.

⁴ <http://myspace.com>.

so far have managed to ‘bridge the semantic gap’ between the low-level data representation and the high-level concepts the user associates with images, as human perception and understanding of images is subjective and operates rather on the semantic level [72].

Sentic Album is a multi-tier architecture that exploits AI and Semantic Web techniques to process image data and metadata at content, concept, and context level, in order to grasp the salient features of online personal photos, and hence find intelligent ways of annotating, organizing, and retrieving them. In this work, in particular, we focus on bridging the gap at concept level by exploiting semantics and sentics [5], that is, the cognitive and affective information, associated with online pictures. We use sentic computing [7], a multi-disciplinary approach to opinion mining and sentiment analysis, to process image metadata, and define the perceived quality of online pictures. We then exploit different web ontologies to encode the results in a semantic aware format and, eventually, represent this information as an interconnected knowledge base, which is browsable through a multi-faceted classification website.

The structure of the paper is as follows: “[Online Personal Photo Management](#)” section presents the state of the art of online personal photo management; “[Importance of Semantics and Sentics in Personal Photos](#)” section discusses the importance of the cognitive and affective information associated with personal pictures; “[Sentic Computing](#)” section explains in detail the sentic computing tools and techniques adopted within this work; “[Annotation Module](#)”, “[Storage Module](#)” and “[Search and Retrieval Module](#)” sections illustrate the annotation module, the storage module, and the search and retrieval module, respectively; “[Evaluation](#)” section presents an evaluation of the overall system; “[Conclusions and Future Work](#)” section, eventually, comprises concluding remarks and a description of future work.

Online Personal Photo Management

Efficient access to online personal pictures requires the ability to properly annotate, organize, and retrieve the information associated with them. While the technology to search personal documents has been available for some time, the technology to manage personal images is much more challenging.

This is mainly due to the fact that, even if images can be roughly interpreted automatically, many salient features exist only in the user’s mind. The only way for a system to accordingly index personal images, hence, is to try to capture and process such features. Existing CBIR systems such as QBIC [23], Virage [1], MARS [59], ImageGrouper [52], MediAssist [54], CIVR [63], EGO [68], ACQUINE

[17], and K-DIME [2] have attempted to build intelligent user interfaces (UIs) capable of retrieving pictures according to their intrinsic content through statistics, pattern recognition, signal processing, computer vision, support vector machines and neural networks, but these techniques are still too weak to bridge the gap between the data representation and the images’ conceptual models in the user’s mind.

Image meta search engines such as Webseek [64], Webseer [24], PicASHOW [44], IGroup [36] or Google,⁵ Yahoo⁶ and Bing⁷ Images, on the other hand, rely on tags associated with online pictures but, in the case of personal photo management, users are unlikely to expend substantial effort to manually classify and categorize images in the hopes of facilitating future retrieval. Moreover, since these techniques mainly depend on keyword-based rather than concept-based algorithms, they often miss potential connections between keywords expressed through different vocabularies or concepts that exhibit implicit semantic connectedness. In order to effectively deal with photo metadata and hence effectively annotate images, it is, in fact, necessary to work at a semantic, rather than syntactic level.

A good effort in this sense has been made within the development of ARIA [46], a software agent that aims to facilitate the storytelling task by opportunistically suggesting photos that may be relevant to what the user is typing. ARIA goes beyond the naïve approach of suggesting photos by simply matching keywords in a photo annotation with keywords in the story. Finally, ARIA applies natural language techniques to the annotation process to extract concepts rather than keywords from the text. A similar approach has been followed by Raconteur [14], a system for conversational storytelling that encourages people to make coherent points, by instantiating large-scale story patterns and suggesting illustrative media. It exploits a large common sense knowledge base to perform natural language processing in real-time on a text chat between a storyteller and a viewer and recommends appropriate media items from a library. Both these approaches present a lot of advantages since concepts, unlike keywords, are not sensitive to morphological variation, abbreviations, or near synonyms. However, simply relying on a semantic knowledge base is not enough to infer the salient features that make different pictures more or less relevant in each user’s mind.

To this end, the proposed Sentic Album exploits AI and Semantic Web techniques to perform reasoning on different knowledge bases and, hence, infer both the cognitive

⁵ <http://google.com/images>.

⁶ <http://images.search.yahoo.com>.

⁷ <http://bing.com/images>.

and the affective information associated with photo metadata. The system further supports this concept-level analysis with content- and context-based techniques, in order to capture all the different aspects of online pictures and, hence, provide users with an IUI that is navigable in real-time through a multi-faceted classification website, since much of what we call cognitive problem-solving intelligence is really the ability to identify what is relevant and important in a context and to subsequently make that knowledge available just in time [47].

Importance of Semantics and Sentic in Personal Photos

Cognitive and affective processes are tightly intertwined in everyday life [16]. The affective aspect of cognition and communication is recognized to be a crucial part of human intelligence and has been argued to be more fundamental in human behavior for ensuring success in social life than intellect [56, 70]. Emotions, in fact, influence our ability to perform common cognitive tasks, such as forming memories and communicating with other people. A psychological study, for example, showed that people asked to conceal emotional facial expressions in response to unpleasant and pleasant slides remembered the slides less well than control participants [3]. Similarly, a study of conversations revealed that romantic partners who were instructed to conceal both facial and vocal cues of emotion while talking about important relationship conflicts with each other, remembered less of what was said than did partners who received no suppression instructions [62]. Many studies have indicated that emotions both seem to improve memory for the gist of an event and to undermine memory for more peripheral aspects of the event [4, 15, 61, 73]. The idea, broadly, is that arousal causes a decrease in the range of cues an organism can take in. This narrowing of attention leads directly to the exclusion of peripheral cues, and this is why emotionality undermines memory for information at the event's edge. At the same time, this narrowing allows a concentration of mental resources on more central materials, and this leads to the beneficial effects of emotion on memory for the event's center [40].

Hence, rather than assigning particular cognitive and affective valence to a specific visual stimulus, we more often balance the importance of personal pictures according to how much information contained in them is pertinent to our lives, goals, and values (or perhaps, the lives and values of people we care about). For this reason, a bad-quality picture can be ranked high in the mind of a particular user, if it reminds him/her of a notably important moment or person of his/her life. Events and situations, in fact, are likely to be organized in the human mind as interconnected concepts and most of the links relating such

concepts are probably weighted by affect, as we tend to better recall memories associated with either very positive or very negative emotions, just as we usually tend to more easily forget about concepts associated with very little or null affective valence [11]. The problem, when trying to emulate such cognitive and affective processes, is that while cognitive information is usually objective and unbiased, affective information is rather subjective and argumentative. For example, while 'car' is always a car, and there is usually not much discussion about the correctness of retrieving an image showing a tree in an African savanna under the label 'landscape', there might be some discussion about whether the retrieved car is 'cool' or just 'nice' or whether the found landscape is 'peaceful' or 'dull' [28].

In order to properly handle the ambiguousness of both emotions and natural language, Sentic Album exploits an ensemble of affective computing and common sense computing techniques to analyze picture data and metadata and, hence, infer what really matters to each user in different online photos. In particular, as the semantic content of an image has usually the greatest impact on the emotional influence it conveys, sentics are built on the top of semantics and processed pairwise with these. In this way, the ensemble of cognitive and affective information associated with personal pictures can be accordingly inferred by means of sentic computing, a recently proposed concept-level opinion mining paradigm that has been hereby adopted, for the very first time, in the field of personal photo management, in combination with other content- and context-level techniques for a comprehensive analysis of online images.

Sentic Computing

Sentic computing is a multi-disciplinary approach to sentiment analysis that exploits both computer and social sciences to better recognize, interpret, and process sentiments in natural language. In sentic computing, whose term derives from the Latin *sentire* (root of words such as sentiment and sentience) and *sensus* (intended both as capability of feeling and as common sense), the analysis of natural language is based on affective ontologies [5] and brain-inspired techniques [13], which enable the analysis of text not only at document, page, or paragraph level but also at sentence and clause level.

In particular, sentic computing involves the use of AI and Semantic Web techniques, for knowledge representation and inference; mathematics, for carrying out tasks such as graph mining and multi-dimensionality reduction; linguistics, for discourse analysis and pragmatics; psychology, for cognitive and affective modeling; sociology, for understanding social network dynamics and social

influence; and finally ethics, for understanding-related issues about the nature of mind and the creation of emotional machines. In this work, in particular, we exploit three sentic computing tools, namely:

1. a language visualization and analysis system (see “[AffectiveSpace](#)” section)
2. a novel emotion categorization model (see “[The Hourglass of Emotions](#)” section)
3. a web ontology for human emotions (see “[The Human Emotion Ontology](#)” section)

and three sentic computing techniques, that is:

1. a technique for clustering concepts in a multi-dimensional space (see “[Sentic Medoids](#)” section)
2. a statistical method for the identification of common semantics (see “[CF-IOF Weighting](#)” section)
3. a technique that expands semantics through spreading activation (see “[Spectral Association](#)” section)

Most of such tools and techniques have been developed by the authors in previous works and are only briefly reported here for the sake of clarity.

AffectiveSpace

AffectiveSpace [8] is a multi-dimensional vector space representation of AffectNet, a semantic network built upon ConceptNet [30], a directed graph representation of common sense knowledge, and WordNet-Affect (WNA) [66], a linguistic resource for the lexical representation of affective knowledge. In particular, AffectNet exploits the ‘blending’ technique [32] to perform inference over ConceptNet and WNA simultaneously, taking advantage of the overlap between them. The alignment operation operated over these two knowledge bases yields a matrix, A , in which common sense and affective knowledge coexist, that is, a matrix $14,301 \times 117,365$ whose rows are concepts (e.g., ‘dog’ or ‘bake cake’), columns are either common sense and affective features (e.g., ‘isA-pet’ or ‘hasEmotion-joy’), and whose values indicate truth values of assertions.

Therefore, in A , each concept is represented by a vector in the space of possible features whose values are positive for features that produce an assertion of positive valence (e.g., ‘a penguin is a bird’), negative for features that produce an assertion of negative valence (e.g., ‘a penguin cannot fly’) and zero when nothing is known about the assertion. The degree of similarity between two concepts, then, is the dot product between their rows in A . The value of such a dot product increases whenever two concepts are described with the same feature and decreases when they are described by features that are negations of each other.

In particular, we use truncated singular value decomposition (TSVD) [71] in order to obtain a new matrix containing both hierarchical affective and common sense knowledge. The resulting matrix has the form $\tilde{A} = U_k * \Sigma_k * V_k^T$ and is a low-rank approximation of A , the original data. This approximation is based on minimizing the Frobenius norm of the difference between A and \tilde{A} under the constraint $\text{rank}(\tilde{A}) = k$. For the Eckart–Young theorem [20], it represents the best approximation of A in the mean-square sense, in fact:

$$\begin{aligned} \min_{\tilde{A}|\text{rank}(\tilde{A})=k} |A - \tilde{A}| &= \min_{\tilde{A}|\text{rank}(\tilde{A})=k} |\Sigma - U^* \tilde{A} V| \\ &= \min_{\tilde{A}|\text{rank}(\tilde{A})=k} |\Sigma - S| \end{aligned}$$

assuming that \tilde{A} has the form $\tilde{A} = USV^*$, where S is diagonal. From the rank constraint, that is, S has k non-zero diagonal entries, the minimum of the above statement is obtained as follows:

$$\begin{aligned} \min_{s_i} \sqrt{\sum_{i=1}^n (\sigma_i - s_i)^2} &= \min_{s_i} \sqrt{\sum_{i=1}^k (\sigma_i - s_i)^2 + \sum_{i=k+1}^n \sigma_i^2} \\ &= \sqrt{\sum_{i=k+1}^n \sigma_i^2} \end{aligned}$$

Therefore, \tilde{A} of rank k is the best approximation of A in the Frobenius norm sense when $\sigma_i = s_i$ ($i = 1, \dots, k$), and the corresponding singular vectors are the same as those of A . If we choose to discard all but the first k principal components, common sense concepts, and emotions are represented by vectors of k coordinates: these coordinates can be seen as describing concepts in terms of ‘eigenmoods’ that form the axes of AffectiveSpace, that is, the basis e_0, \dots, e_{k-1} of the vector space (Fig. 1). For example, the most significant eigenmood, e_0 , represents concepts with positive affective valence.

That is, the larger a concept’s component in the e_0 direction is, the more affectively positive it is likely to be. Concepts with negative e_0 components, then, are likely to have negative affective valence. Thus, by exploiting the information sharing property of TSVD, concepts with the same affective valence are likely to have similar features—that is, concepts conveying the same emotion tend to fall near each other in AffectiveSpace. Concept similarity does not depend on their absolute positions in the vector space, but rather on the angle they make with the origin. For example, we can find concepts such as ‘beautiful day’, ‘birthday party’, ‘laugh’, and ‘make person happy’ very close in direction in the vector space, while concepts like ‘sick’, ‘feel guilty’, ‘be laid off’, and ‘shed tear’ are found in a completely different direction (nearly opposite with respect to the center of the space).

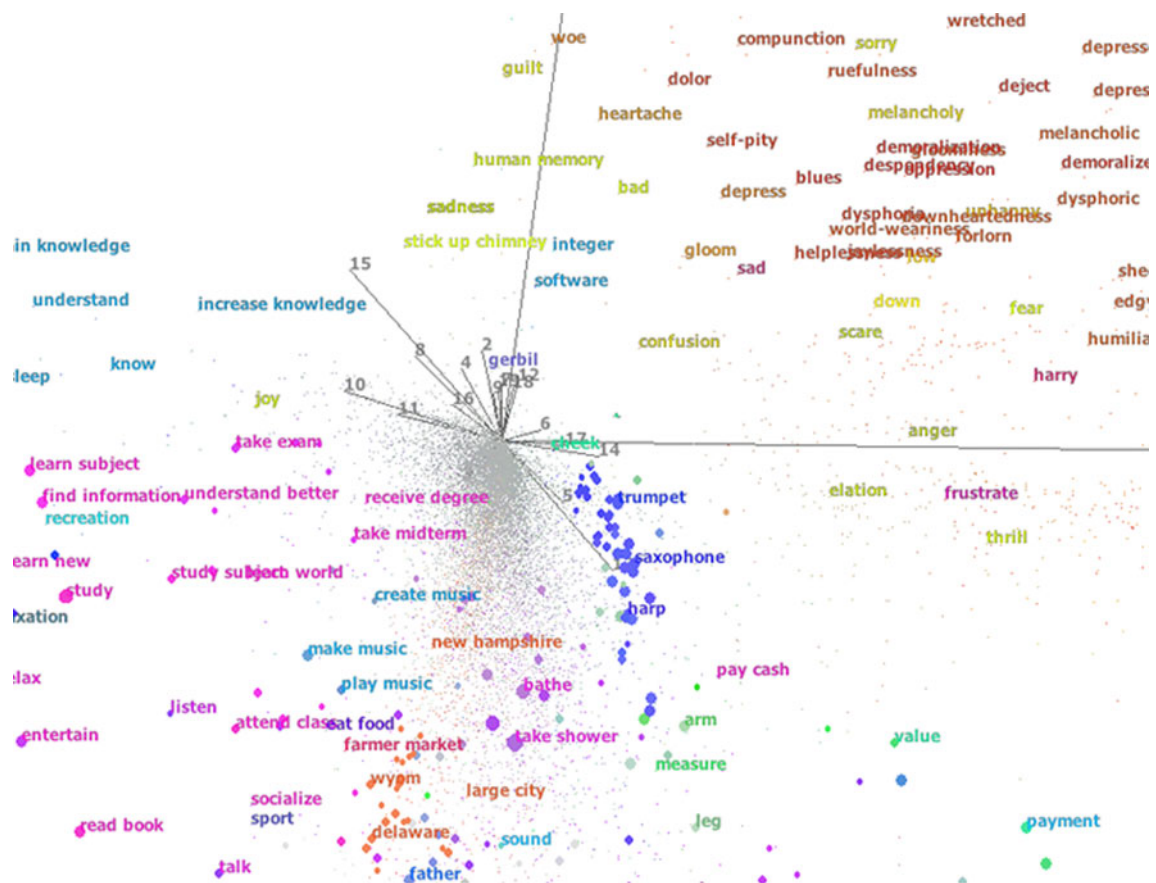


Fig. 1 AffectiveSpace

The Hourglass of Emotions

To reason on the disposition of concepts in AffectiveSpace, we use the Hourglass of Emotions [10], a novel affective categorization model in which sentiments are organized around four independent—but concomitant—dimensions, whose different levels of activation are argued to make up the total emotional state of the mind. The Hourglass model, in fact, is based on the idea that the mind is made of different independent resources and that emotional states result from turning some set of these resources on and turning another set of them off [50]. Each such selection changes how we think by changing our brain's activities: the state of anger, for example, appears to select a set of resources that help us react with more speed and strength while also suppressing some other resources that usually make us act prudently.

The primary quantity we can measure about an emotion we feel is its strength. But, when we feel a strong emotion, it is because we feel a very specific emotion. And, conversely, we cannot feel a specific emotion like fear or amazement without that emotion being reasonably strong. Mapping this space of possible emotions leads to a

hourglass shape (Fig. 2). The Hourglass of Emotions, in particular, can be exploited in the context of HCI to measure how much, respectively, the user is amused by interaction modalities (Pleasantness), interested in interaction contents (Attention), comfortable with interaction dynamics (Sensitivity), or confident in interaction benefits (Aptitude). Each affective dimension, in particular, is characterized by six levels of activation (measuring the strength of an emotion), termed 'sentic levels', which determine the intensity of the expressed/perceived emotion as an $int \in [-3, 3]$.

These levels are also labeled as a set of 24 basic emotions [58], six for each of the affective dimensions, in a way that allows the model to specify the affective information associated with text both in a dimensional and in a discrete form. The dimensional form, in particular, is termed 'sentic vector' and is a four-dimensional *float* vector that can potentially synthesize any human emotion in terms of Pleasantness, Attention, Sensitivity, and Aptitude. Some particular sets of sentic vectors have special names as they specify well-known compound emotions. For example, the set of sentic vectors with a level of Pleasantness $\in (1,2]$ (joy), null Attention, null Sensitivity

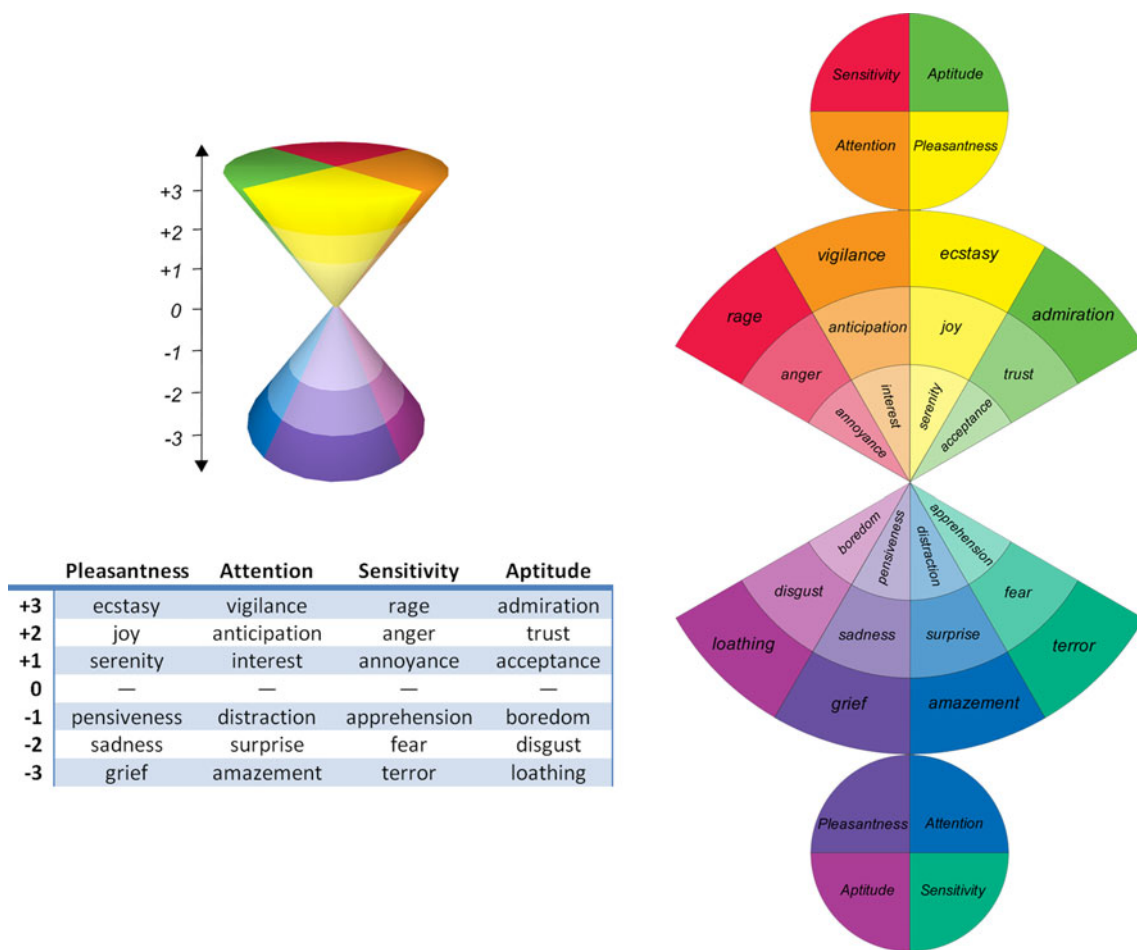


Fig. 2 The hourglass of emotions

and a level of Aptitude $\in (1,2]$ (trust) are called ‘love sentic vectors’ since they specify the compound emotion of love.

The Human Emotion Ontology

The human emotion ontology⁸ (HEO) [27] (Fig. 3) is conceived as a high-level ontology for human emotions that supplies the most significant concepts and properties which constitute the centerpiece for the description of every human emotion. If necessary, these high-level features can be further refined using lower-level concepts and properties related to more specific descriptions or linked to other more specialized ontologies. The main purpose of HEO is to create a description framework that could grant flexibility (by allowing the use of a wide and extensible set of descriptors to represent all the main features of an emotion) and interoperability (by allowing the mapping of concepts and properties belonging to different emotion representation models) at the same time.

⁸ <http://semedia.dibet.univpm.it/heo/heo.owl>.

The ontology web language description logic (OWL DL) [51] was chosen for the development of HEO, in order to exploit its expressiveness and inference power to map the different models used in the emotion description. OWL DL, in fact, allows a taxonomical organization of emotion categories and properties restriction, in order to link emotion description made by category and dimension. In HEO, for example, Ekman’s ‘joy’ archetypal emotion represents a superclass for the emotions ‘ecstasy’, ‘joy’, and ‘serenity’ of the Hourglass model.

Using property restriction, the Plutchik’s ‘joy’ emotion can also be defined as an emotion that ‘has Pleasantness some float $\in (1,2]$ ’, ‘interest’ as an emotion that ‘has Attention $\in [0,+1]$ ’ and ‘love’ as an emotion that ‘has Pleasantness some float $\in (1,2]$, and Aptitude some float $\in (1,2]$ ’. In this way, querying a database that supports OWL DL inference for basic emotions of type ‘joy’ will return not only the emotions expressly encoded as Ekman archetypal emotions of type ‘joy’, but also the emotions encoded as Hourglass basic emotion of type ‘joy’ and the emotions that ‘have Pleasantness some float $\in (1,2]$ ’.

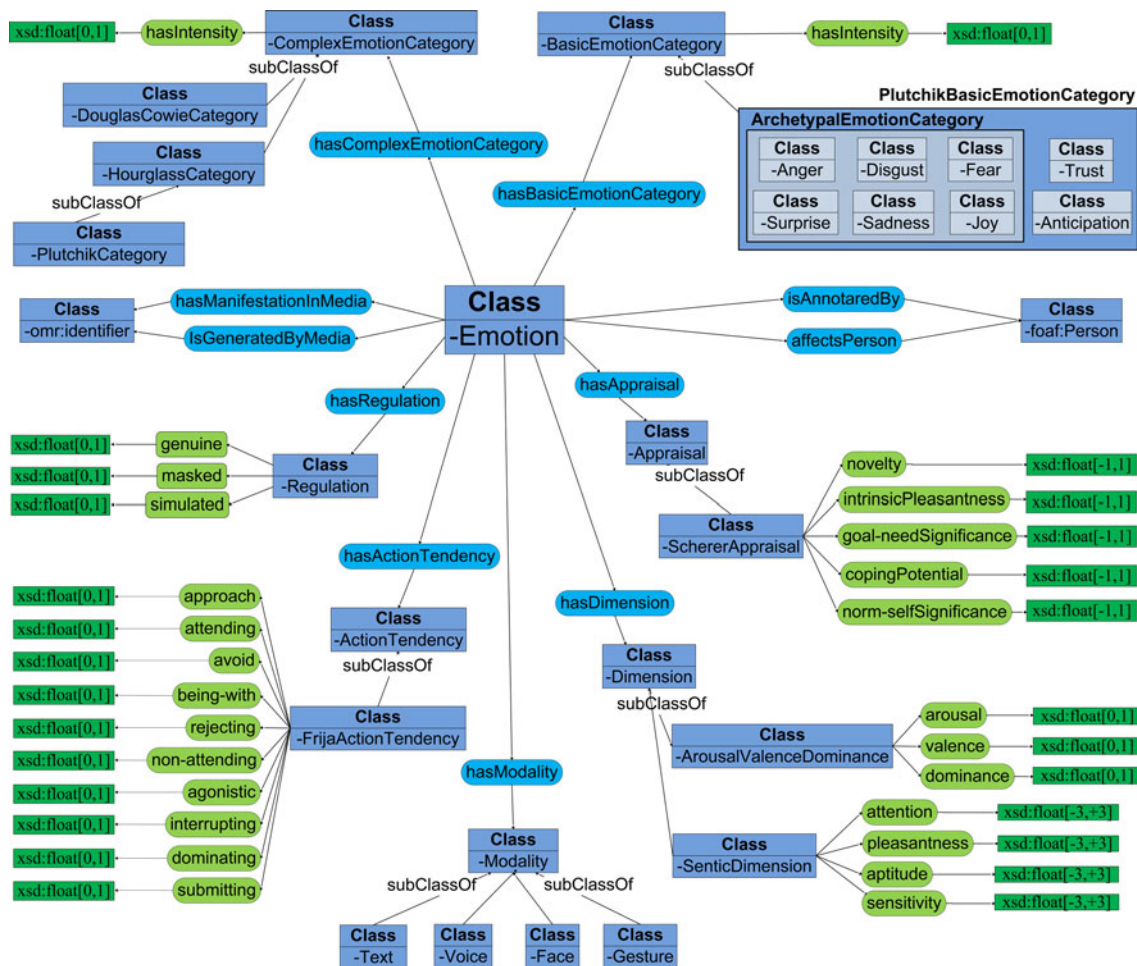


Fig. 3 The human emotion ontology

Sentic Medoids

Sentic medoids [11] is a technique that adopts a *k*-medoids approach [37] to partition the given observations into *k* clusters around as many centroids, trying to minimize a given cost function. Differently from the *k*-means algorithm [29], which does not pose constraints on centroids, *k*-medoids do assume that centroids must coincide with *k* observed points. The most commonly used algorithm for finding the *k* medoids is the Partitioning Around Medoids (PAM) algorithm. The PAM algorithm determines a medoid for each cluster selecting the most centrally located centroid within the cluster.

After selection of medoids, clusters are rearranged so that each point is grouped with the closest medoid. Since *k*-medoids clustering is a NP-hard problem [25], different approaches based on alternative optimization algorithms have been developed all of which carry the risk of being trapped around local minima. We use a modified version of the algorithm recently proposed by Park and Jun [57], which runs in a similar way to the *k*-means clustering

algorithm. This has been shown to have similar performance when compared to the PAM algorithm while taking a significantly reduced computational time. Specifically, we have *N* concepts (*N* = 14,301) encoded as points $x \in \mathbb{R}^p$ (*p* = 50). We want to group them into *k* clusters and, in our case, we can fix *k* = 24 as we are looking for one cluster for each sentic level *s* of the Hourglass model.

Generally, the initialization of clusters for clustering algorithms is a problematic task as the process often risks getting stuck in local optimum points, depending on the initial choice of centroids [19]. In this work, we are able to conveniently use, as initial centroids, the concepts that are currently used as centroids for clusters—since they specify the emotional categories we want to organize Affective-Space into. For this reason, what is usually seen as a limitation of the algorithm can be seen as an advantage for this particular approach, since we are not looking for the 24 centroids leading to the best 24 clusters but indeed for the 24 centroids identifying the required 24 sentic levels (i.e., the centroids should not be ‘too far’ from the ones currently used).

In particular, as the Hourglass affective dimensions are independent but concomitant, we need to cluster AffectiveSpace four times, once for each dimension. According to the Hourglass categorization model, however, each concept can convey, at the same time, more than one emotion (which is why we get compound emotions), and this information can be expressed via a sentic vector specifying the concept’s affective valence in terms of Pleasantness, Attention, Sensitivity, and Aptitude.

Therefore, given that the distance between two points in AffectiveSpace is defined as $D(a, b) = \sqrt{\sum_{i=1}^p (a_i - b_i)^2}$ (note that the choice of Euclidean distance is arbitrary), the employed algorithm, applied for each of the four affective dimensions, can be summarized as follows:

1. Each centroid $C_n \in \mathbb{R}^{50} (n = 1, 2, \dots, k)$ is set as one of the six concepts corresponding to each s in the current affective dimension
2. Assign each record x to a cluster Ξ so that $x_i \in \Xi_n$ if $D(x_i, C_n) \leq D(x_i, C_m) \quad m = 1, 2, \dots, k$
3. Find a new centroid C for each cluster Ξ so that $C_j = x_i$ if $\sum_{x_m \in \Xi_j} D(x_i, x_m) \leq \sum_{x_m \in \Xi_j} D(x_h, x_m) \quad \forall x_h \in \Xi_j$
4. Repeat step 2 and 3 until no changes on centroids are observed

This clusterization of AffectiveSpace allows us to calculate, for each common sense concept x , a four-dimensional sentic vector that defines its affective valence in terms of a degree of fitness: $\mathbf{f}(x)$ where $f_a = D(x, C_j) \quad C_j | D(x, C_j) \leq D(x, C_k)$
 $a = 1, 2, 3, 4 \quad k = 6a - 5, 6a - 4, \dots, 6a$

CF-IOF Weighting

CF-IOF (concept frequency—inverse opinion frequency) [7] is a technique that identifies common domain-dependent semantics in order to evaluate how important a concept is to a set of opinions concerning the same topic.

Firstly, the frequency of a concept c for a given domain d is calculated by counting the occurrences of the concept c in the set of available d -tagged opinions and dividing the result by the sum of number of occurrences of all concepts in the set of opinions concerning d . This frequency is then multiplied by the logarithm of the inverse frequency of the concept in the whole collection of opinions, that is:

$$CF-IOF_{c,d} = \frac{n_{c,d}}{\sum_k n_{k,d}} \log \sum_k \frac{n_k}{n_c}$$

where $n_{c,d}$ is the number of occurrences of concept c in the set of opinions tagged as d , n_k is the total number of concept occurrences, and n_c is the number of occurrences of c in the whole set of opinions. A high weight in CF-IOF is reached by a high concept frequency in a given domain and a low frequency

of the concept in the whole collection of opinions. Therefore, as a result of using CF-IOF weights, it is possible to filter out common concepts and detect relevant topic-dependent semantics.

Spectral Association

Spectral association [31] is a technique that involves assigning values, or activations, to ‘seed concepts’ and applying an operation that spreads their values across the ConceptNet graph. This operation, which is an approximation of many steps of spreading activation, transfers the most activation to concepts that are connected to the key concepts by short paths or many different paths in common sense knowledge.

In particular, we build a matrix C that relates concepts to other concepts, instead of their features, and add up the scores over all relations that relate one concept to another, disregarding direction. Applying C to a vector containing a single concept spreads that concept’s value to its connected concepts. Applying C^2 spreads that value to concepts connected by two links (including back to the concept itself). But what we would really like is to spread the activation through any number of links, with diminishing returns, so perhaps the operator we want is:

$$1 + C + \frac{C^2}{2!} + \frac{C^3}{3!} + \dots = e^C$$

We can calculate this odd operator, e^C , because we can factor C . C is already symmetric, so instead of applying Lanczos’ method to CC^T and getting the SVD, we can apply it directly to C and get the spectral decomposition $C = V\Lambda V^T$. As before, we can raise this expression to any power and cancel everything but the power of Λ . Therefore, $e^C = Ve^\Lambda V^T$. This simple twist on the SVD lets us calculate spreading activation over the whole matrix instantly.

As with the SVD, we can truncate these matrices to k axes and, therefore, save space while generalizing from similar concepts. We can also rescale the matrix so that activation values have a maximum of 1 and do not tend to collect in highly connected concepts such as ‘person’, by normalizing the truncated rows of $Ve^{\Lambda/2}$ to unit vectors, and multiplying that matrix by its transpose to get a rescaled version of $Ve^\Lambda V^T$.

Annotation Module

Today manual image annotation is still the most common practice for indexing and then later retrieving personal image collections. However, manual image annotation is an expensive and labor-intensive procedure, and hence, there has been great interest in coming up with automatic

ways to retrieve images based on their associated information. In order to make the most of both photo data and metadata, Sentic Album aims to annotate online personal pictures either at content, concept, and context level.

In particular, the annotation module mainly exploits metadata such as descriptions, tags, and comments, which we call ‘conceptual metadata’, associated with each image to extract its relative semantics and sentics and, hence, enhance the picture specification with its intrinsic cognitive and affective information. This concept-level annotation procedure is performed through an ensemble of sentic computing tools and techniques, and it is supported with a parallel content and context-level analysis. User’s personal photo data and metadata are currently pulled from Picasa (through Google Data API⁹) but, in the future, we plan to expand the breadth of the system by interfacing it with more sources, for example, other online photo-sharing services, blogs and social networks.

A Three-Level Architecture

The annotation module works at three different levels: content, context, and concept. The content-based annotation, in particular, is performed through Python Imaging Library¹⁰ (PIL), an external library for the Python¹¹ programming language that adds support for opening, manipulating, and saving many different image file formats. For every online personal picture, in particular, we exploit PIL to extract luminance and chrominance information and other image statistics, for example, the total, mean, standard deviation, and variance of the pixel values.

The context based annotation, in turn, exploits information such as timestamp, geolocation, and user interaction metadata. Such metadata, which we call ‘contextual metadata’, are processed by the Context Deviser, a submodule that extracts small bits of information suitable for storing in a relational database for re-use at a later time, that is, time, date, city, and country of caption plus all the relevant user interaction metadata such as number and IDs of friends who viewed, commented or liked the picture.

The concept-based annotation represents the core of the module and is designed by means of sentic computing, which allows the system to go beyond a mere syntactic analysis of the metadata associated with pictures. A big problem of manual image annotation, in fact, is the different vocabulary that different users (or even the same user) can use to describe the content of a picture.

The different expertise and purposes of tagging users, in fact, may result in tags that use various levels of abstraction to describe a resource: a photo can be tagged at the ‘basic level’ of abstraction [39] as ‘cat’ or at a superordinate level as ‘animal’ or at various subordinate levels below the basic level as ‘Persian cat’ or ‘*Felis silvestris catus longhair Persian*’.

To overcome this problem, Sentic Album extends the set of available tags (if any) with related semantics and sentics and, to further expand the cognitive and affective metadata associated with each picture, it extracts additional common sense and affective concepts from its description and comments (if any). In particular, the conceptual metadata are processed by four submodules: a pre-processing submodule, which performs a first skim of the textual data, a semantic parser, whose aim is to extract concepts from the lemmatized text, AffectNet, for the inference of the semantics associated with the given concepts, and AffectiveSpace, for the extraction of sentics (Fig. 4).

The pre-processing submodule firstly interprets all the affective valence indicators usually contained in opinionated text such as special punctuation, complete upper-case words, onomatopoeic repetitions, exclamation words, negations, degree adverbs and emoticons. Secondly, it converts text to lower-case and, after lemmatizing it, splits the opinion into single clauses according to grammatical conjunctions and punctuation. The semantic parser deconstructs text into concepts using a lexicon based on sequences of lexemes that represent multiple-word concepts extracted from ConceptNet, WordNet and other linguistic resources.

These n -grams are not used blindly as fixed word patterns but exploited as reference for the module, in order to extract multiple-word concepts from information-rich sentences. So, differently from other shallow parsers, the module can recognize complex concepts also when irregular verbs are used or when these are interspersed with adjective and adverbs, for example, the concept ‘buy christmas present’ in the sentence ‘I bought a lot of very nice Christmas presents’. The semantic parser, additionally, provides, for each retrieved concept, the relative frequency, valence and status, that is the concept’s occurrence in the text, its positive or negative connotation and the degree of intensity with which the concept is expressed.

The AffectNet submodule finds matches between the retrieved concepts and those previously calculated using CF-IOF and spectral association. In particular, CF-IOF weighting is exploited to find seed concepts for a set of a-priori categories, extracted from Picasa’s popular tags, meant to cover common topics in personal pictures, for example, art, nature, friends, travel, wedding or holiday. Spectral association is then used to expand this set with semantically related common sense concepts. The AffectiveSpace submodule projects the retrieved concepts into

⁹ <http://code.google.com/apis/gdata>.

¹⁰ <http://pythonware.com/products/pil>.

¹¹ <http://python.org>.

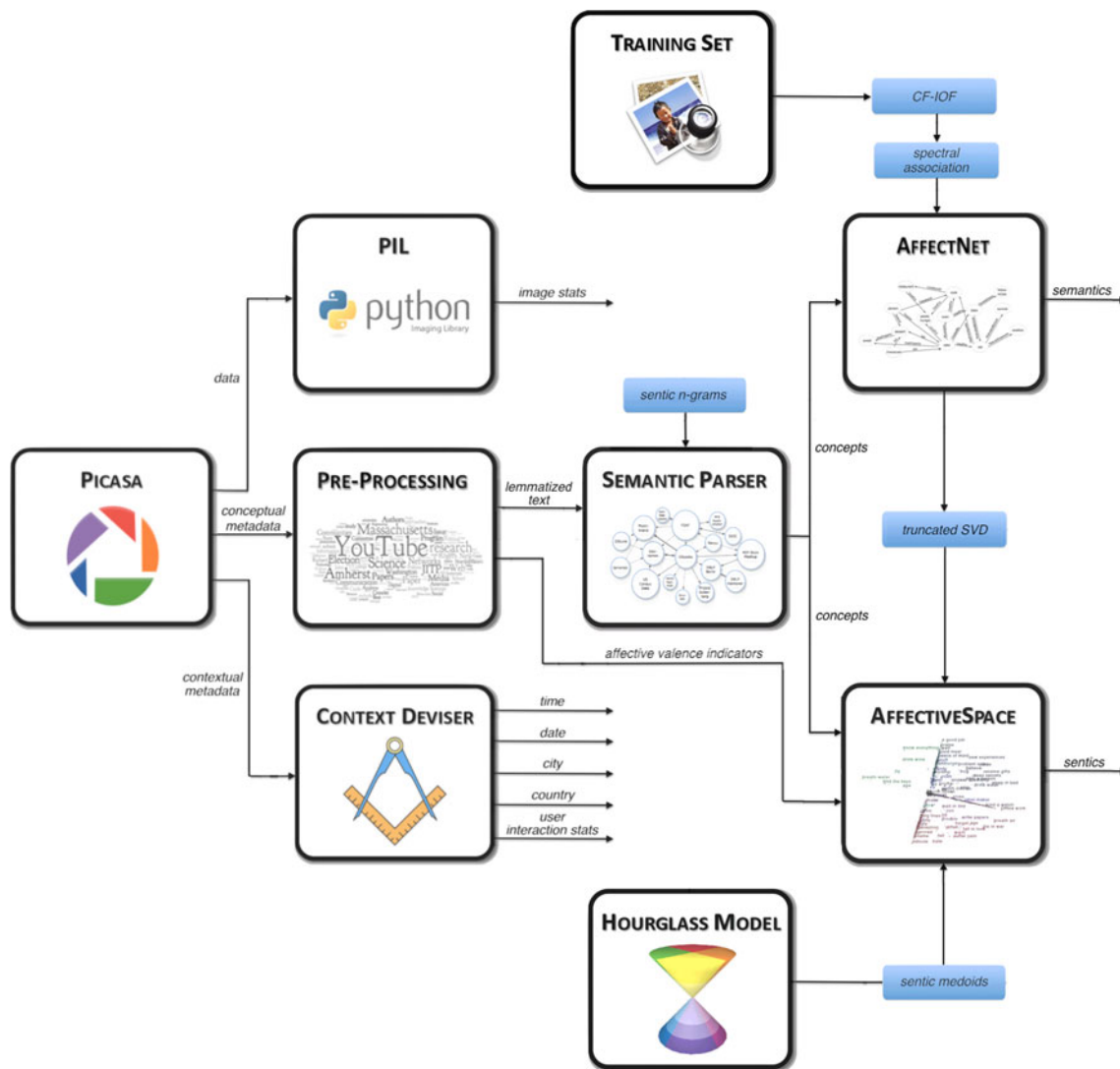


Fig. 4 Annotation module

the vector space representation of AffectNet. The multi-dimensional space, clustered with respect to the Hourglass model using sentic medoids, is then exploited to infer the affective valence of the retrieved concepts, in terms of Pleasantness, Attention, Sensitivity and Aptitude, according to the relative position they occupy in the space. This information, finally, is also exploited to calculate the overall polarity associated with pictures, which is calculated according to the sentics relative to each retrieved concept, that is:

where c_i is an input concept, N the total number of retrieved concepts and 9 the normalization factor (as the Hourglass dimensions are defined as float $\in [-3,+3]$). In the formula, Attention is taken in absolute value since both its positive and negative intensity values correspond to positive polarity values (e.g., ‘surprise’ is negative in the sense of lack of Attention but positive from a polarity point of view). Similarly, Sensitivity is taken as a negative absolute value since both its positive and negative intensity values correspond to negative polarity values (e.g., ‘anger’

$$p = \sum_{i=1}^N \frac{\text{Pleasantness}(c_i) + |\text{Attention}(c_i)| - |\text{Sensitivity}(c_i)| + \text{Aptitude}(c_i)}{9N}$$

is positive in the sense of level of activation of Sensitivity but negative in terms of polarity).

Perceived Quality of Online Pictures

Providing a satisfactory visual experience is one of the main goals for present-day electronic multimedia devices. All the enabling technologies for storage, transmission, compression, rendering should preserve, and possibly enhance, image quality; and to do so, quality control mechanisms are required. Systems to automatically assess visual quality are generally known as objective quality metrics. The design of objective quality metrics is a complex task because predictions must be consistent with human visual quality preferences. Human preferences are inherently quite variable and, by definition, subjective; moreover, in the field of visual quality, they stem from perceptual mechanisms that are not fully understood yet. A common choice is to design metrics that replicate the functioning of the human visual system to a certain extent, or at least that take into account its perceptual response to visual distortions by means of numerical features [38]. Although successful, these approaches come with a considerable computational cost, which makes them impractical for most real-time applications. Computational intelligence paradigms allow to tackle the quality assessment task from a different perspective, since they aim at mimicking quality perception instead of designing an explicit model of the human visual system [48, 53, 60]. In the special case of personal pictures, perceived quality metrics can be computed not only at content level, but also at concept and context level. One of the primary reasons why people take pictures is to remember the emotions they felt on special occasions of their lives.

Extracting and storing such affective information can be a key factor in improving future searches, as users seldom want to find photos matching general requirements. Users' criteria in browsing personal pictures, in fact, are more often related to the presence of a particular person in the picture and/or its perceived quality (e.g., to find a good photo of your mother). Satisfying this type of requirement is a tedious task as chronological ordering or classification by event does not help much. The process usually involves repeatedly trying to think of a matching picture, and then looking for it. An exhaustive search (looking through the whole collection for all of the photos matching a requirement) would normally only be carried out in exceptional circumstances, such as following a death in the family. In order to accordingly rank personal photos, Sentic Album exploit data and metadata associated with them to extract useful information at content, concept, and context level and, hence, calculate the perceived quality of online pictures (PQOP), defined as:

$$PQOP(p, u) = 3 \frac{Content(p) * Concept(p, u) * Context(p, u)}{Content(p) + Concept(p, u) + Context(p, u)}$$

where $Content(p)$, $Concept(p, u)$, and $Context(p, u)$ are float $\in [0, 1]$ representing image quality assessment values associated with picture p and user u , in terms of visual, conceptual, and contextual information, respectively. The proposed formula is not meant to be a rigid definition, but rather a qualitative indication (drawn from our experimental usability tests) of which features (or levels of analysis) should be taken into account when calculating the perceived quality of online pictures.

In our specific case, $Content(p)$ is computed from numerical features extracted through a reduced-reference framework for objective quality assessment exploiting a circular extreme learning machine (C-ELM) [18] and the color correlogram [34] of p . $Concept(p, u)$, in turn, specifies how much the picture p is relevant to the user u in terms of cognitive and affective information, that is, the semantic and sentic similarity (degrees of separation in the AffectNet graph and the dot products in AffectiveSpace, respectively) between the concepts associated with p and the concepts that characterize the user u (defined by means of CF-IOF). $Context(p, u)$, eventually, defines the degree of relevance of picture p for user u in terms of time, location, and user interaction, for example, time elapsed between capture date of p and dates relevant for u , geographic distance between location of p and places relevant for u , and frequency of interaction between u and users who have viewed/commented on p . The 3C (Content, Concept, and Context) are all equally relevant for measuring how good a personal picture is to the eye of a user. According to the formula, in fact, if any of the 3C is null the PQOP is null as well, even though the remaining elements of the 3C have both maximum values, for example, a perfect quality picture ($Content(p) = 1$) taken in the hometown of the user on the date of his birthday ($Context(p, u) = 1$) but depicting people he/she does not know and objects/places that are totally irrelevant for him/her ($Concept(p, u) = 0$).

Storage Module

The storage module is the middle-tier in which the outputs of the annotation module are stored, in a way that these can be easily accessible by the search and retrieval module at a later time. The module stores information relative to photo data and metadata redundantly at three levels:

1. in a relational database fashion
2. in a Semantic Web format
3. in a matrix format

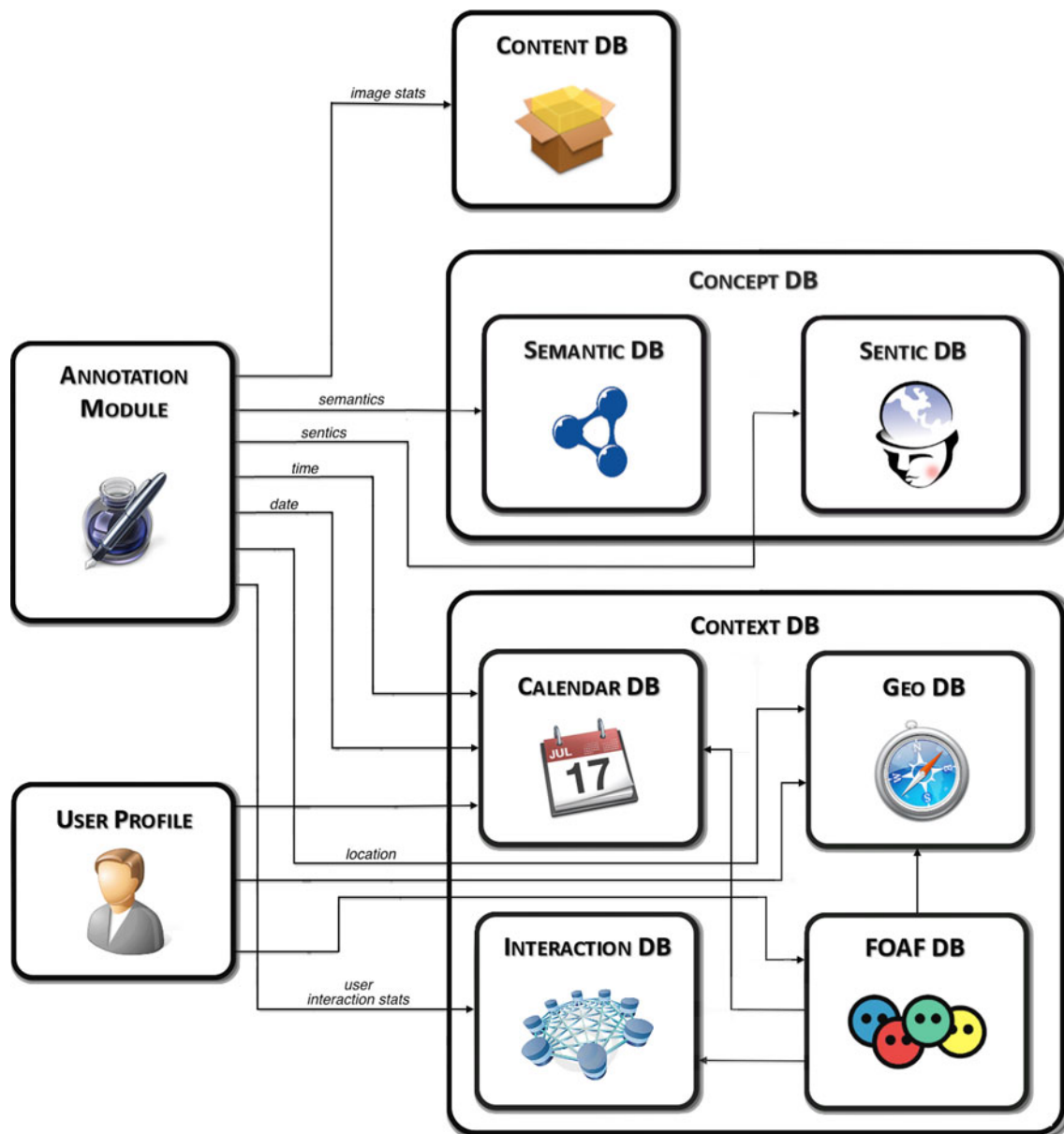


Fig. 5 Storing image data and metadata in a relational database fashion

Storing Information in a Relational Database Fashion

Sentic Album stores information in three main SQL databases (Fig. 5), that is a Content DB, for the information relative to data (image statistics), a Concept DB, for the information relative to conceptual metadata (semantics and sentics), and a Context DB, for the information relative to contextual metadata (timestamp, geolocation and user interaction metadata). The Concept DB, in particular, consists of two databases, the Semantic DB and the Sentic DB, in which the cognitive and the affective information associated with photo metadata, respectively, are stored. The Context DB, in turn, is divided into four databases, the Calendar, Geo,

FOAF (Friend Of A Friend) and Interaction DBs, which contain the information relative to timestamp, geolocation, social links and social interaction, respectively.

These databases are also integrated with information coming from the web profile of the user such as user's DOB (for the Calendar DB), user's current location (for the Geo DB) or user's list of friends (for the FOAF DB). The FOAF DB, in particular, plays an important role within the Context DB since it provides the other peer databases with information relative to user's social connections, for example, relatives' birthdays or friends' hometowns. Moreover, the Context DB receives extra contextual information from the inferred semantics.

Personal names in the conceptual metadata are recognized by building a dictionary of first names from the Web and combining them with regular expressions to recognize full names. These are added to the database (in the FOAF DB) together with geographical places (in the Geo DB), which are also mined from databases on the Web and added to the parser's semantic lexicon.

Storing Information in a Semantic Web Format

As for the Semantic Web format [43], all the information related to pictures' metadata is stored in RDF/XML according to a set of predefined web ontologies. This operation aims to make the description of the semantics and santics associated with pictures applicable to most online images coming from different sources, for example, online photo-sharing services, blogs, social networks. To further this aim, it is necessary to standardize as much as possible the descriptors used in encoding the information about multimedia resources and people to which the images refer, in order to make it univocally interpretable and suitable to feed other applications.

For this reason, we encode the information relative to image metadata and people by using the descriptors provided by OMR (Ontology for Media Resources)¹² and the FOAF¹³ ontology, respectively. OMR represents an important effort to help circumvent the current proliferation of audio/video metadata formats, currently carried on by the W3C Media Annotations Working Group.

It offers a core vocabulary to describe media resources on the Web, introducing descriptors such as 'title', 'creator', 'publisher', 'createDate', and 'rating'. It defines semantic-preserving mappings between elements from existing formats. This ontology is supposed to foster the interoperability among various kinds of metadata formats currently used to describe media resources on the Web.

FOAF represents a recognized standard in describing people, providing information such as their names, birthdays, pictures, blogs, and especially other people they know, which makes it particularly suitable for representing data that appears on social networks and communities. OMR and FOAF together supply most of the vocabulary we need for describing media and people and we add other descriptors only when necessary. For example OMR, at least in the current version, does not supply vocabulary for describing comments, that we analyze to extract the affective information relative to media. We extend this ontology introducing the 'Comment' class, and define for it the 'author', 'text', and 'publicationDate' properties.

¹² <http://w3.org/TR/mediaont-10>.

¹³ <http://www.foaf-project.org>.

In HEO, we introduce properties to link emotions to multimedia resources and people. In particular, we defined 'hasManifestationInMedia' and 'isGeneratedByMedia' to describe emotions that occur and are generated in media, respectively, and the property 'affectPerson' to connect emotions to people. Moreover, to improve the hierarchical organization of emotions in HEO, we exploit WNA, the linguistic resource for the lexical representation of affective knowledge we use to build AffectiveSpace. WNA is built by assigning to a number of WordNet [22] synsets one or more affective labels (a-labels) and then by extending the core with the relations defined in WordNet. In particular, the affective concepts representing emotional states are identified by synsets marked with the a-label 'EMOTION', but there are also other a-labels for concepts representing moods, situations eliciting emotions or emotional responses. Thus, the combination of HEO with WNA, OMR and FOAF (Fig. 6) provides a complete framework to describe not only the image metadata and the users connected with them, but also the cognitive and affective information carried by the images and the way they are perceived by people.

In particular, this information is encoded in RDF/XML and stored in a Sesame triplestore, a purpose-built database for the storage and retrieval of RDF metadata, using the descriptors defined by HEO, WNA, OMR and FOAF. Sesame can be embedded in applications and used to conduct a wide range of inferences on the information stored, based on RDFS and OWL type relations between data. In addition, it can also be used in a standalone server mode, much like a traditional database with multiple applications connecting to it. In this way, all the pieces of knowledge stored inside Sesame can be queried and the results can also be retrieved in a semantic aware format and used for other applications.

Storing Information in a Matrix Format

As for the storage of photo data and metadata in a matrix format, we build a dataset, which we call '3CNet', integrating the information from the 3C in a unique knowledge base. The aim of this representation is to exploit principal component analysis (PCA) to later organize online personal images in a multi-dimensional vector space (as for AffectiveSpace) and, hence, reason on their similarity. 3CNet, in fact, is an $n \times m$ matrix whose rows are user's personal pictures IDs, whose columns are either content, concept, and context features (e.g., 'contains cold colors', 'conveys joy' or 'located in Italy'), and whose values indicate truth values of assertions.

Therefore, in 3CNet, each image is represented by a vector in the space of possible features whose values are +1, for features that produce an assertion of positive

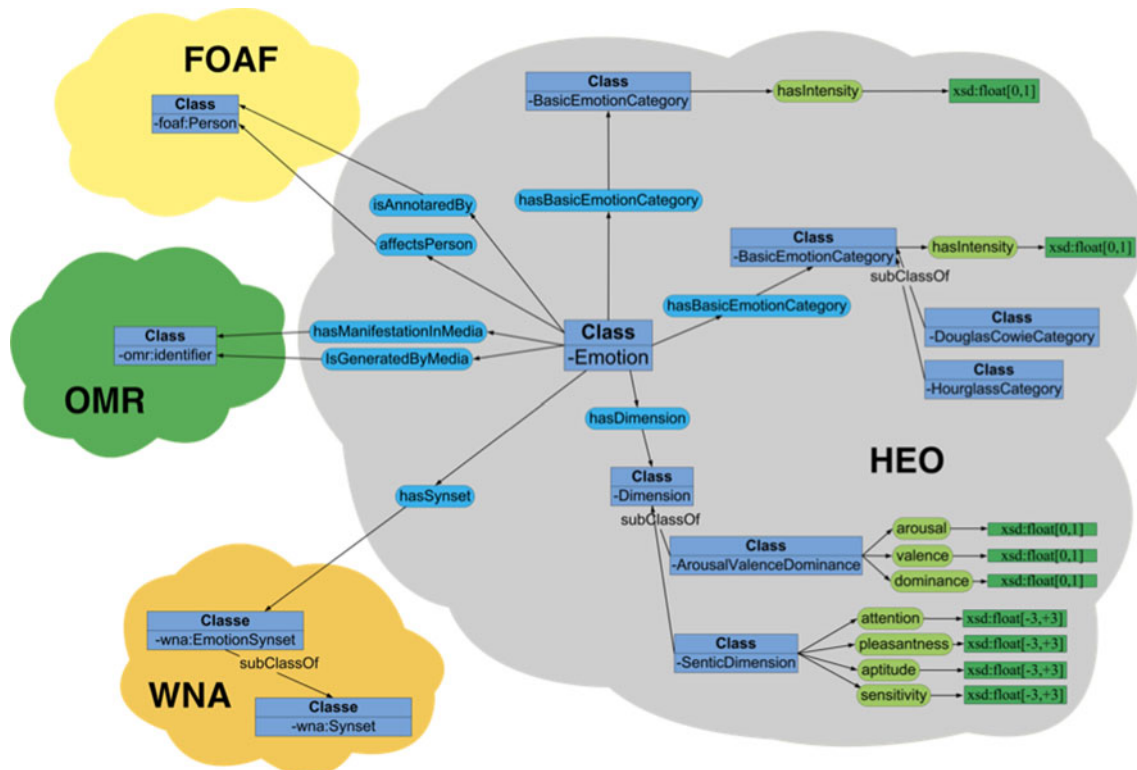


Fig. 6 Merging different ontologies in the storage module

valence, -1 , for features that produce an assertion of negative valence, and 0 when nothing is known about the assertion. The degree of similarity between two images, then, is the dot product between their rows in 3CNet. The value of such a dot product increases whenever two images are described with the same feature and decreases when they are described by features that are negations of each other.

Search and Retrieval Module

The main aim of the search and retrieval module is to provide users with an UI that allows them to easily manage, search and retrieve their personal pictures online. Most of the existing photo management systems let users search for pictures through a keyword-based query, but results are hardly ever good enough since it is very difficult to come up with an ideal query from the user's initial request. The initial idea of an image the user has in mind before starting a search session, in fact, often deviates from the final results he/she will choose [67].

In order to let users start from a sketchy idea and then dynamically refine their search, we exploit a multi-faceted classification paradigm. Faceted classification allows the assignment of multiple categories to an object, enabling the classifications to be ordered in multiple ways, rather than in

a single, pre-determined, taxonomic order. This makes it possible to perform searches combining the textual approach with the navigational one. Faceted search enables users to navigate a multi-dimensional information space by concurrently writing queries in a text box and progressively narrowing choices in each dimension.

Sentic Album specifically uses the SIMILE Exhibit¹⁴ API, a set of Javascript files that allows easily the creation of rich interactive webpages including maps, timelines and galleries, with very detailed client-side filtering. Exhibit pages use the multi-faceted classification paradigm to display semantically structured data stored in a Semantic Web aware format, for example, RDF or JavaScript object notation (JSON). One of the most relevant aspects of Exhibit is that, once the page is loaded, the web-browser also loads the entire data set in a lightweight database and performs all the computations (sorting, filtering, etc.) locally on the client-side, providing high performances (Fig. 7). The search and retrieval module exports all the information contained in the storage module's sesame triplestore into a JSON file in order to feed the Exhibit interface and, hence, make the data available for browsing as a unique knowledge base.

Personal images are displayed in a dynamic gallery that can be ordered according to different parameters, either

¹⁴ <http://simile-widgets.org/exhibit>.



Fig. 7 Exhibit IUI

textual or numeric, that is visual features (e.g., color balance, hue, saturation, brightness and contrast), semantics (i.e., common sense concepts such as ‘go jogging’ and ‘birthday party’ but also people and objects contained in the picture), sentic (i.e., emotions conveyed by the picture and its polarity) and contextual information (e.g., time of caption, location and social information such as users who viewed/commented on the picture). By using such an interface, it is possible to explore such information both by using the search box, to perform keyword-based queries, and by adding or removing constraints on the facet properties, to filter results accordingly. Further, natural language processing techniques similar to those used to process the image conceptual metadata are employed to analyze the text typed in the search box and, hence, perform queries on the SQL databases of the storage module.

The order of visualization of the retrieved images is given by the PQOP, so that images containing more relevant information at content, concept, and context level are first displayed. If, for example, the user is looking for pictures of his/her partner, Sentic Album initially proposes photos representing important events such as first date, first child birth or honeymoon, that is pictures with high PQOP.

Storage module’s 3CNet is also exploited in the IUI, in order to find similar pictures. Toward the end of a search, the user sometimes may be interested in finding pictures similar to one of those so far obtained, even if this does not fulfill the constraints currently set via the facets. To serve this purpose, every picture is provided with a ‘like me’ button that opens a new Exhibit window displaying

content-, concept-, and context-related images, independently of any constraint. Picture similarity is calculated by means of PCA and, in particular, through TSVD, as for AffectiveSpace. The number of singular values to be discarded (in order to reduce the dimensionality of 3CNet and hence reason on picture similarity) is chosen according to the total number of user’s online personal pictures and the amount of available metadata associated with them, that is, according to size and density of 3CNet.

Thus, by exploiting the information sharing property of TSVD, images specified by similar content, concept, and context are likely to have similar features and, hence, tend to fall near each other in the built-in vector space. Finally, the IUI, also offers to display images according to date of caption on a timeline. Chronology, in fact, is a key categorization concept for the management of personal pictures. Having the collection in chronological order is helpful for locating particular photos or events, since it is usually easier to remember when an event occurred relative to other events, as opposed to remembering its absolute date and time [41].

Evaluation

Many works dealing with object detection, scene categorization or content analysis on the cognitive level have been published, trying to bridge the semantic gap between represented objects and high-level concepts associated with them [45], however, where affective retrieval and classification of digital media is concerned, publications, and

especially benchmarks, are very few [49]. To overcome the lack of availability of relevant datasets, we evaluate, in this preliminary study, both the user-friendliness and the performance of Sentic Album through a usability test on a pool of 18 Picasa regular users and an evaluation of the system's annotation capabilities on a topic and mood-tagged dataset. For the usability test, users were asked to freely browse their online personal collections using Sentic Album IUI and to retrieve particular sets of pictures, in order to judge both usability and accuracy of the interface.

Common queries included 'find a funny picture of your best friend', 'search for the shots of your last summer holiday', 'retrieve pictures of you with animals', 'find an image taken on Christmas 2009', 'search for pictures of you laughing', and 'find a good picture of your mom'. From the test, it emerged that users really appreciate being able to dynamically and quickly set/remove constraints in order to display specific sets of pictures (which they cannot do in Picasa). After the test session, participants were asked to fill-in an online questionnaire in which they were asked to rate, on a five-level scale, each single functionality of the interface according to its perceived utility. Concept facets and timeline, in particular, were found to be the most used by participants for search and retrieval tasks (Table 1).

Users also really appreciated the 'like me' functionality, which was generally able to propose very relevant (semantically and affectively related) pictures (again not available in Picasa). When freely browsing their collections, users were particularly amused by the ability to navigate their personal pictures according to the emotion these conveyed, even though they did not always agree with the results. Additionally, participants were not very happy with the accuracy of the search box, especially if they searched for one particular photo out of the entire collection. However, they always very much appreciated the order in which the pictures were proposed, which allowed them to quickly have all the most relevant pictures available as first results. 83.3 % of test users declared that, despite not being as nifty as Picasa, Sentic Album is a very good photo management tool (especially for its novel semantic faceted search and PQOP functionalities) and

they hope they could still be using it because, in the end, what really counts when browsing personal pictures is to find best matches in the shortest time.

As for the evaluation of the system's annotation capabilities, we calculated statistical classifications, specifically precision, recall and F-measure rates of the semantics and sentics extraction process by using a corpus of topic and mood-tagged blogs from LiveJournal (LJ), respectively. LJ is a virtual community of more than 23 millions users who keep a blog, journal or diary. One of the interesting features of this website is that LJ bloggers are allowed to label their posts not only with a topic tag but also with a mood label, by choosing from more than 130 predefined moods or by creating custom mood themes. Since the indication of the affective status is optional, the mood-tagged posts are likely to reflect the true mood of the authors and, hence, form a good test-set for Sentic Album. However, since LJ mood themes do not perfectly match the sentic levels, we had to consider a reduced set of 10 moods, specifically, 'ecstatic', 'happy', 'pensive', 'surprised', 'enraged', 'sad', 'angry', 'annoyed', 'scared' and 'bored'. Moreover we could not consider non-affective webposts since mood-untagged blog entries do not necessarily lack emotions.

As for the topic tags, in turn, we selected the LJ labels that match Picasa popular tags, for example, 'friends', 'travel' or 'holiday', in order to collect natural language text that is likely to have the same semantics as the conceptual metadata usually associated with online personal pictures. All LJ accounts have Atom, RSS and other data feeds which show recent public entries, friend relationships and interests. Unfortunately the current LJ API allows retrieval of posts by topic only, so, in order to also get mood-tagged posts, we had to design our own web crawler (Fig. 8).

After retrieving and storing relevant data and metadata from 10,000 LJ posts, we extracted semantics and sentics through Sentic Album's annotation module and compared the output with the relative topic and mood tags, in order to calculate precision, recall and F-measure rates. On average, each post contained around 140 words, from which about 12 affective valence indicators and 60 concepts were extracted.

Table 1 Perceived utility of the different interface features by 18 Picasa regular users

Feature	Not at all (%)	Just a little (%)	Somewhat (%)	Quite a lot (%)	Very much (%)
Concept facets	0	0	5.6	5.6	88.8
Content facets	77.8	16.6	5.6	0	0
Context facets	16.6	11.2	5.6	33.3	33.3
Search box	0	11.2	16.6	33.3	38.9
Like me	0	5.6	5.6	16.6	72.2
Timeline	0	0	0	16.6	83.4
Sorting	11.2	33.3	33.3	16.6	5.6

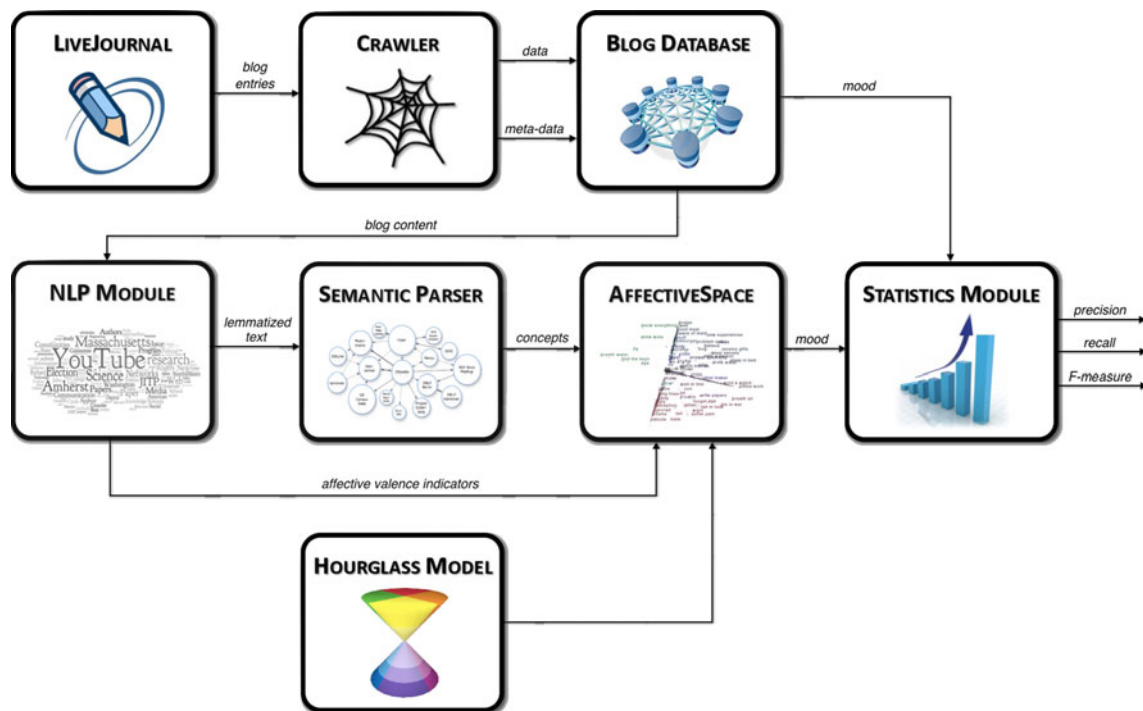


Fig. 8 Evaluation of the sentsics extraction process

From the retrieved concepts we inferred semantics and sentsics associated with each of the selected posts and, hence, tagged them with topic and mood labels. We then compared these labels with the corresponding topic and mood LJ tags, obtaining very good accuracy in terms of both semantics and sentsics extraction. As for the detection of moods, for example, ‘happy’ and ‘sad’ posts were identified with particularly high precision (89.2 and 81.8 %, respectively) and good recall rates (76.5 and 68.4 %), as shown in Table 2.

The F-measure values obtained, hence, were significantly good (82.4 and 74.7 %, respectively), especially when compared to the corresponding F-measure rates, calculated on the same dataset, by using a number of

conventional commonly employed approaches to automatic identification of emotions in text, namely: keyword spotting [21, 55, 74], in which text is classified into categories based on the presence of fairly unambiguous affect words (53.7 % F-measure for ‘happy’ posts and 51.4 % for ‘sad’ posts); lexical affinity [65, 75], which assigns arbitrary words a probabilistic affinity for a particular mood (63.2 and 58.1 % F-measure rates, respectively); and, statistical methods [26, 33], which calculate the valence of keywords and word co-occurrence frequencies on the base of a large training corpus (69.5 and 62.9 % F-measure for ‘happy’ and ‘sad’ posts, respectively).

As for the detection of topics, the classification of ‘travel’ and ‘friends’ posts was performed with a precision of 75.6 and 69.1 % and recall rates of 65.3 and 58.4 %, respectively. The total F-measure rates, hence, were considerably good (70.4 % for ‘travel’ posts and 63.8 % for ‘friends’ posts) in comparison with the corresponding F-measure rates of the baseline methods (44.7 and 35.5 % for keyword spotting, 53.1 and 39.8 % for lexical affinity, 61.9 and 52.6 % for statistical methods).

Conclusions and Future Work

Managing digital photos is a huge problem with no good solutions. Looking for a robust way to dynamically tag photos based not just on words but focusing on emotions (which photos capture) seems a good direction. With the advent of

Table 2 Evaluation results of the sentsics extraction process

Mood	Precision (%)	Recall (%)	F-measure (%)
Ecstatic	73.1	61.3	66.6
Happy	89.2	76.5	82.3
Pensive	69.6	52.9	60.1
Surprised	81.2	65.8	72.6
Enraged	68.9	51.6	59.0
Sad	81.8	68.4	74.5
Angry	81.4	53.3	64.4
Annoyed	77.3	58.7	66.7
Scared	82.6	63.5	71.8
Bored	70.3	55.1	61.7

digital photography, taking a lot of pictures is no longer an issue today, both in terms of development costs and timings. Indeed, thanks to the many social networking websites that allow users to easily upload and share personal pictures online, photography has become an important part of our social life. For the same reasons, however, the volume of online personal pictures is growing so much that users often lose control of them.

Since the manual annotation of images is an expensive and labor-intensive procedure, users tend to upload hundreds of non-annotated pictures and, because it takes growing effort to retrieve images from their personal collections, they sometimes even lose track of them. In the past, CBIR systems and image meta search engines applied different techniques to automatically extract meaning from image data and metadata but none of these so far have managed to bridge the semantic gap between the low-level data representation and the high-level concepts the user associates with images.

In this paper, we described Sentic Album, a novel content-, concept-, and context-based online personal photo management system that exploits both data and metadata of online personal pictures to intelligently annotate, organize, and retrieve them. Sentic Album exploits not just colors and texture of online images (content), but also the cognitive and affective information associated with their metadata (concept), and their relative timestamp, geolocation and user interaction metadata (context). Sentic Album fuses different AI and Semantic Web techniques to extract semantics and sentsics associated with online personal pictures and, hence, enhance their specification with intrinsic cognitive and affective information. Moreover, Sentic Album exploits the concept that behind every personal image there is always an emotion to define PQOP, the perceived quality of online pictures, and hence develop a method to automatically rate pictures for search and retrieval purposes.

The main limitations of the proposed tool currently reside in the incapability of the system to effectively organize pictures when there is a small amount of metadata associated with them (since content-level analysis on its own is not enough) and when the pictures belong to a context about which there are few concepts available in AffectNet (e.g., if the user is a fan of buzkaishi, the Afghan national sport). What is now needed is the use of additional datasets for carrying out a more thorough evaluation of the application, which is currently underway. We are also planning to carry out broader usability tests and to assess the system also at content and context level, in addition to exploiting semantics and sentsics (which was the focus of this preliminary study), in order to understand the role each annotation level plays in the overall picture analysis and management process.

In the future, we also plan to explore different techniques for the analysis of images at content and context level, which is currently performed mainly as a support to the concept-

level analysis. In particular, we aim to improve Sentic Album capabilities to exploit image visual contents by analyzing not just color and texture but also shape and layout features. We will also exploit theoretical and empirical concepts from psychology and art theory to define image features that are specific to the domain of artworks with emotional expression. Specifically, the results of psychological experiments on emotional response to color [69], as well as work on color in art [35], will be used. Despite objective methods for quality assessment that typically analyze the luminance component of the color information, in fact, recent studies have shown that chrominance also plays a relevant role in quality perception [58].

To this end, we are currently developing a human semiotics ontology (HSO) [42], to be merged later with HEO, in order for the system to reason on colors and other visual features and connect these to human emotions. Moreover, we will explore multi-dimensionality reduction techniques, similar to those employed within Affective-Space and 3CNet, in order to reason on the user interaction metadata and, hence, exploit them for defining PQOP and for search and retrieval tasks. We are also working on making the system adaptive. The user feedback on auto-categorized images, in fact, is extremely important. We plan to assign to every piece of information stored in Sentic Album a confidence score, which will be increased/decreased according to user's feedback. We plan to improve the UI by adding new functionalities, such as the option to display pictures according to their location caption on a world map, and to make its design adaptive to the user's current emotional state. Eventually, we plan the development of a multi-modal system capable of perceiving user's attitudes and feelings when looking at personal pictures, for example, by analyzing facial expressions, gestures and speech or by taking into account how much time and how many times the user stares at specific photos. In conclusion, Sentic Album can be seen to represent a first step toward the development of Sentic Interfaces, that is, next-generation emotion-sensitive systems, capable of perceiving and expressing the cognitive and affective information associated with multi-modal user interaction.

Acknowledgments This work has been part-funded by Hewlett-Packard Labs India, the UK Engineering and Physical Sciences Research Council (EPSRC Grant Reference: EP/G501750/1) and Sitekit Solutions Ltd. (UK). We would like to thank Praphul Chandra for providing the application case study and the HP industrial placement opportunity.

References

1. Bach J, Fuller C, Gupta A, Hampapur A, Horowitz B, Humphrey R, Jain R, Shu C. Virage image search engine: an open framework for image management. In: Sethi I, Jain R, editors. Storage

- and retrieval for still image and video databases, vol. 2670. Bellingham: SPIE; 1996. p. 76–87.
2. Bianchi-Berthouze N. K-DIME: an affective image filtering system. *IEEE Multimedia*. 2003; 10(3):103–6.
 3. Bonanno G, Papa A, O’Neill K, Westphal M, Coifman K. The importance of being flexible: the ability to enhance and suppress emotional expressions predicts long-term adjustment. *Psychol Sci*. 2004; 15:482–7.
 4. Burke A, Heuer F, Reisberg D. Remembering emotional events. *Memory Cogn*. 1992; 20:277–90.
 5. Cambria E, Benson T, Eckl C, Hussain A. Sentic PROMs: application of sentic computing to the development of a novel unified framework for measuring health-care quality. *Expert Systems with Applications*, Elsevier. 2012. doi:10.1016/j.eswa.2012.02.120.
 6. Cambria E, Grassi M, Hussain A, Havasi C. Sentic computing for social media marketing. *Multimedia Tools Appl*. 2011. doi:10.1007/s11042-011-0815-0.
 7. Cambria E, Hussain A. Sentic computing: techniques, tools, and applications. Berlin: Springer; 2012.
 8. Cambria E, Hussain A, Durrani T, Havasi C, Eckl C, Munro J. Sentic computing for patient centered application. In: *Proceedings of IEEE ICSP*. Beijing; 2010. p. 1279–82.
 9. Cambria E, Hussain A, Havasi C, Eckl C. AffectiveSpace: blending common sense and affective knowledge to perform emotive reasoning. In: *Proceedings of CAEPIA*. Seville; 2009. p. 32–41.
 10. Cambria E, Livingstone A, Hussain A. The Hourglass of Emotions. *LNCS, Cognitive Behavioral Systems*. Springer-Verlag, Berlin Heidelberg; 2012.
 11. Cambria E, Mazzocco T, Hussain A, Eckl C. Sentic medoids: organizing affective common sense knowledge in a multi-dimensional vector space. In: *Advances in neural networks*. Lecture notes in computer science, vol. 6677. Berlin: Springer; 2011. p. 601–10.
 12. Cambria E, Olsher D, Kwok K. Sentic panalogy: swapping affective common sense reasoning strategies and foci. In: *Proceedings of CogSci*. Sapporo; 2012.
 13. Cambria E, Olsher D, Kwok K. Sentic activation: a two-level affective common sense reasoning framework. In: *Proceedings of AAAI*. Toronto; 2012.
 14. Chi P, Lieberman H. Intelligent assistance for conversational storytelling using story patterns. In: *IUI*. Palo Alto 2011.
 15. Christianson S, Loftus E. Remembering emotional events: the fate of detailed information. *Cogn Emot*. 1991;5:81–108.
 16. Damasio A. *Descartes’ error: emotion, reason, and the human brain*. New York: Grossett/Putnam; 1994.
 17. Datta R, Wang J. ACQUINE: aesthetic quality inference engine—real-time automatic rating of photo aesthetics. In: *Proceedings of the international conference on multimedia information retrieval*. Philadelphia; 2010.
 18. Decherchi S, Gastaldo P, Redi J, Zunino R, Cambria E. Circular-ELM for the reduced-reference assessment of perceived image quality. *Neurocomputing*. (in press).
 19. Duda R, Hart P. *Pattern classification and scene analysis*. New York: Wiley; 1973.
 20. Eckart C, Young G. The approximation of one matrix by another of lower rank. *Psychometrika*. 1936;1(3):211–8.
 21. Elliott CD. The affective reasoner: A process model of emotions in a multi-agent system. Ph.D. thesis, Northwestern University, Evanston; 1992.
 22. Fellbaum C. *WordNet: An electronic lexical database (language, speech, and communication)*. Cambridge: The MIT Press; 1998.
 23. Flickner M, Sawhney H, Niblack W, Ashley J, Huang Q, Dom B, Gorkani M, Hafner J, Lee D, Petkovic D, Steele D, Yanker P. Query by image and video content: the QBIC system. *Computer*. 1995;28(9):23–32.
 24. Frankel C, Swain MJ, Athitsos V. *WebSeer: an image search engine for the world wide web*. Technical report. Chicago: University of Chicago; 1996.
 25. Garey M, Johnson D. *Computers and intractability: a guide to the theory of NP-completeness*. San Francisco: Freeman; 1979.
 26. Goertzel B, Silverman K, Hartley C, Bugaj S, Ross M. The baby webmind project. In: *Proceedings of AISB*. Birmingham; 2000.
 27. Grassi M. Developing HEO human emotions ontology. In: Fierrez J, Ortega-Garcia J, Esposito A, Drygajlo A, Faundez-Zanuy M, editors. *Biometric ID management and multimodal communication*. Lecture notes in computer science, vol. 5707. Berlin: Springer; 2009. p. 244–51.
 28. Hanjalic A. Extracting moods from pictures and sounds: towards truly personalized TV. *IEEE Signal Process Mag*. 2006;23(2):90–100.
 29. Hartigan J, Wong M. Algorithm AS 136: a k-means clustering algorithm. *J R Stat Soc*. 1979;28(1):100–8.
 30. Havasi C, Speer R, Alonso J. ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge. In: *Proceedings of RANLP*. Borovets; 2007.
 31. Havasi C, Speer R, Holmgren J. Automated color selection using semantic knowledge. In: *Proceedings of AAAI CSK*. Arlington; 2010.
 32. Havasi C, Speer R, Pustejovsky J, Lieberman H. Digital intuition: applying common sense using dimensionality reduction. *IEEE Intell Syst*. 2009;24(4):24–35.
 33. Hu M, Liu B. Mining opinion features in customer reviews. In: *Proceedings of AAAI*. San Jose; 2004.
 34. Huang J, Ravi S, Mitra M, Zhu W, Zabih R. Image indexing using color correlograms. In: *Proceedings of IEEE CVPR*, 1997. p. 762–8.
 35. Itten J. *The art of color: the subjective experience and objective rationale of color*. New York: Wiley; 1973.
 36. Jing F, Wang C, Yao Y, Deng K, Zhang L, Ma WY. IGroup: web image search results clustering. In: *Proceedings of ACM Multimedia*. Santa Barbara; 2006.
 37. Kaufman L, Rousseeuw P. *Finding groups in data: an introduction to cluster analysis*. London: Wiley; 1990.
 38. Keelan B. (2002) *Handbook of image quality*. New York: Marcel Dekker; 2002.
 39. Lakoff G. *Women, fire, and dangerous things*. Chicago: University of Chicago Press; 1990.
 40. Laney C, Campbell H, Heuer F, Reisberg D. Memory for thematically arousing events. *Memory Cogn*. 2004;32(7):1149–59.
 41. Lansdale M, Edmonds E. Using memory for events in the design of personal filing systems. *Int J Man Mach Stud*. 1992;36(1):97–126.
 42. Lascu A, Cambria E, Grassi M. Human semiotics ontology. In: *Proceedings of ICMC*. Venice; 2011. p. 152.
 43. Lee B, Hendler J, Lassila O. *The semantic web*. Scientific American 2001.
 44. Lempel R, Soffer A. PicASHOW: pictorial authority search by hyperlinks on the web. In: *Proceedings of WWW*. Hong Kong; 2001.
 45. Lew M, Sebe N, Djeraba C, Jain R. Content-based multimedia information retrieval: state of the art and challenges. *ACM Trans Multimedia Comput Commun Appl*. 2006;2(1):1–19.
 46. Lieberman H, Rosenzweig E, Singh P. ARIA: an agent for annotating and retrieving images. *IEEE Comput*. 2001;34(7):57–62.
 47. Lieberman H, Selker T. Out of context: computer systems that adapt to, and learn from, context. *IBM Syst J*. 2000;39(3):617–32.
 48. Lu W, Zeng K, Tao D, Yuan Y, Gao X. No-reference image quality assessment in contourlet domain. *Neurocomputing*. 2012; 73(4–6):784–94.
 49. Machajdik J, Hanbury A. Affective image classification using features inspired by psychology and art theory. In: *Proceedings of the international conference on multimedia*. Florence; 2010.

50. Minsky M. *The emotion machine: commonsense thinking, artificial intelligence, and the future of the human mind*. New York: Simon & Schuster; 2006.
51. Motik B, Sattler U, Studer R. Query answering for OWL-DL with rules. 2004. p. 549–63.
52. Nakazato M, Manola L, Huang T. ImageGrouper: Search, annotate and organize images by groups. In: Chang S, Chen Z, Lee S editors. *Recent advances in visual information systems. Lecture notes in computer science*, vol. 2314. Berlin: Springer; 2002. p. 93–105.
53. Narwaria M, Lin W. Objective image quality assessment based on support vector regression. *IEEE Trans Neural Netw*. 2010;12(3): 515–19.
54. O'Hare N, Lee H, Cooray S, Gurrin C, Jones G, Malobabic J, O'Connor N, Smeaton A, Uscilowski B. MediAssist: Using content-based analysis and context to manage personal photo collections. In: *Proceedings of CIVR*, Tempe; 2006. p. 529–32.
55. Ortony A, Clore G, Collins A. *The cognitive structure of emotions*. Cambridge: Cambridge University Press; 1988.
56. Pantic M. Affective computing. In: *Encyclopedia of multimedia technology and networking*, vol. 1. Idea Group Reference; 2005. p. 8–14.
57. Park H, Jun C. A simple and fast algorithm for k-medoids clustering. *Exp Syst Appl*. 2009;36(2):3336–41.
58. Plutchik R. The nature of emotions. *Am Sci*. 2001;89(4):344–50.
59. Porkaew K, Chakrabarti K. Query refinement for multimedia similarity retrieval in MARS. In: *Proceedings of ACM international conference on multimedia*. New York: ACM; 1999. p. 235–8.
60. Redi J, Gastaldo P, Heynderickx I, Zunino R. Color distribution information for the reduced-reference assessment of perceived image quality. *IEEE Trans Circuits Syst Video Technol*. 2012; 20(12):1757–69.
61. Reisberg D, Heuer F. Memory for emotional events. *Memory Emot*. 2004;3–41.
62. Richards J, Butler E, Gross J. Emotion regulation in romantic relationships: the cognitive consequences of concealing feelings. *J. Social Pers. Relation*. 2003;20:599–620.
63. Sebe N, Tian Q, Loupias E, Lew MS, Huang TS. Evaluation of salient point techniques. In: *Proceedings of the international conference on image and video retrieval*. London: Springer; 2002. p. 367–77.
64. Smith J, Chang S. An image and video search engine for the world-wide web. In: *Symposium on electronic imaging: science and technology 1997*.
65. Somasundaran S, Wiebe J, Ruppenhofer J. Discourse level opinion interpretation. In: *Proceedings of COLING*. Manchester; 2008.
66. Strapparava C, Valitutti A. WordNet-Affect: An affective extension of WordNet. In: *Proceedings of LREC*. Lisbon; 2004.
67. Urban J, Jose J, Van Rijsbergen C. An adaptive approach towards content-based image retrieval. *Multimedia Tools Appl*. 2006;31: 1–28.
68. Urban J, Jose JM. EGO: a personalized multimedia management and retrieval tool. *Int J Intell Syst*. 2006;21(7):725–45.
69. Valdez P, Mehrabian A. Effects of color on emotions. *J Exp Psychol General*. 1994;123(4):394–409.
70. Vesterinen E. Affective computing. In: *Digital media research seminar*. Helsinki; 2001.
71. Wall M, Rechtsteiner A, Rocha L. Singular value decomposition and principal component analysis. In: Berrar D, Dubitzky W, Granzow M (eds) *A practical approach to microarray data analysis*. Berlin: Springer; 2003. pp. 91–109.
72. Wang W, He Q. A survey on emotional semantic image retrieval. In: *Proceedings of IEEE ICIP*, 2008. p. 117–20.
73. Wessel I, Merckelbach H. The impact of anxiety on memory for details in spider phobics. *Appl Cogn Psychol*. 1997;11:223–31.
74. Wiebe J, Wilson T, Cardie C. Annotating expressions of opinions and emotions in language. *Lang Resour Eval*. 2005;39(2): 165–210.
75. Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceedings of HLT/EMNLP*. Vancouver; 2005.