Cogn Comput (2011) 3:480–489
DOI 10.1007/s12559-011-9101-8

Sentic Web: A New Paradigm for Managing Social Media Affective Information

Marco Grassi · Erik Cambria · Amir Hussain · Francesco Piazza

Received: 3 May 2010 / Accepted: 7 May 2011 / Published online: 21 May 2011

Abstract The recent success of media-sharing services caused an exponential growth of community-contributed multimedia data on the Web and hence a consistent shift of the flow of information from traditional communication channels to social media ones. Retrieving relevant information from this kind of data is getting more and more difficult, not only for their volume, but also for the different nature and formats of their contents. In this work, we introduce Sentic Web, a new paradigm for the management of social media affective information, which exploits AI and Semantic Web techniques to extract, encode, and represent opinions and sentiments over the Web. In particular, the computational layer consists in an intelligent engine for the inference of emotions from text, the representation layer is developed on the base of specific domain ontologies, and the application layer is based on the faceted browsing paradigm to make contents available as an interconnected knowledge base.

Keywords Sentic computing · AI · Semantic web · Ontologies · NLP · Emotion and affective UI

M. Grassi (✉) · F. Piazza
Department of Biomedical, Electronic and Telecommunication Engineering, Università Politecnica delle Marche, Ancona 60131, Italy
e-mail: m.grassi@univpm.it

E. Cambria · A. Hussain
Department of Computing Science and Mathematics, University of Stirling, Stirling FK94LA, UK

Introduction

Differently from early web development, retroactively labeled Web 1.0, today Internet is a dynamic being in which information is no more the core—the user is now at the center of it. This sort of Copernican revolution brought us to the Web 2.0 era, in which the first simple web sites evolved to become more and more interactive, from static to dynamically generated, from handcrafted to CMS driven, from purely informative to more and more social.

The efforts to understand this new cultural and social phenomenon gave birth to Web Science [1], a new discipline that brings computer scientists and social scientists together across the disciplinary divide, to explore the development of the Web across different areas of everyday life and technological development. In particular, the passage from a read-only to a read-write Web made webusers more inclined to express their opinions and sentiments via blogs, wikis, fora, chats, social networks, video, and photo sharing. In this context, we introduce Sentic Web, a new paradigm which combines AI and Semantic Web techniques to manage social media affective information and make it effectively accessible by allowing crossed information retrieval and classification through a simple and intuitive user interface.

The term 'sentic' derives from the Latin 'sentire', the root of words like sentiment and sensation, and it is hereby adopted for the use of Sentic Computing [2], a multidisciplinary approach to opinion mining and sentiment analysis, described in the computational layer (Sect. 3). The encoding of the extracted affective information is done on the base of the human emotion ontology (HEO) [3] and it is illustrated in the representation layer (Sect. 4). The resulting knowledge base can be finally browsed through a multifaceted classification web site, described in the application layer (Sect. 5), which allows users to jointly manage the information about media, users and their emotional content, and state, respectively.

The Importance of Social Media

The World Wide Web represents one of the most revolutionary applications in the history of computing and human communication, which is keeping on changing how information is disseminated and retrieved, how business is conducted and how people communicate with each other.

As the dimension of the Web increases, the technologies used in its development and the services provided to its users are developing constantly. Even if just few years have passed, in fact, Web 1.0's static and read-only HTML pages seem now just an old memory. Today, the Web has become a dynamic and interactive reality in which more and more people actively participate by creating, sharing, and consuming contents. In this way, Internet configures itself not only as a 'Web of data' but also as a 'Web of people' where data and users are interconnected in an unbreakable bond.

Virtual communities have been constantly attracting more and more users and have become an alternative form of communication. Social networks, in particular, represent today virtual *agorai* where people meet and talk to each others using instant messaging or posting messages, organize events, keep their friends and relatives informed of what they are doing and how they feel. Blogs and fora have become tools of common use for users to express and share their own opinions.

As the Web plays a more and more significant role in people's social lives, it contains more and more information concerning their feelings and mood. The distillation of knowledge from this huge amount of unstructured information, also known as opinion mining and sentiment analysis, is a task which has recently raised more and more interest for reasons such as marketing and financial market prediction.

Several commercial tools for social media marketing have been recently developed by Umbria [4], Cymfony [5], Neilsen [6], and Evolve24 [7] (among many others) to provide companies with a way to analize the blogosphere on a large scale to extract information about the trend of the opinions about their products . Under the push of this growing interest in social media marketing, great advances have been done in the field of sentiment analysis since its relatively recent start. Many systems have been developed for the classification of sentiments (typically positive, negative, or neutral) in a target document [8–10], using machine learning approaches [11] as well as approaches for the selection of most relevant text fragments in subjective opinions [12] and feature selection [13]. Nevertheless, most of the existing tools are limited to a polarity evaluation or a mood classification according to a very limited set of emotions. In addition, such methods mainly rely on parts of text in which emotional states are explicitly expressed and are unable to capture emotions that are expressed implicitly.

Another problem for the analysis of social media affective information is that today web contents (and social media contents in particular) are perfectly suitable for human consumption but they remain hardly accessible to machines. The Web mostly owes its success to the development of search engines like Google and Yahoo, which represent the starting point for information retrieval. Such engines, which base their searches on keyword-based algorithms relying on the textual representation of the web page, are very good in retrieving texts, splitting them into parts, checking the spelling, counting their words. But when it comes to interpreting sentences and extracting useful information for users, their capabilities result still very limited [14].

To overcome these problems, Sentic Web adopts a concept-based approach to infer opinions and emotions from text and then exploits Semantic Web techniques to encode and represent the extracted information as an interconnected knowledge base.

Computational Layer

Today, text is one of the most important modalities for affective analysis and generation because the bulk of computer user interfaces are still text based.

In the past, emotion inference from text involved the implementation of different techniques such as handcrafted models, keyword spotting, fuzzy logic, lexical affinity, and statistical methods. These methods turned out to be semantically weak since they mainly rely on parts of text in which emotional states are explicitly expressed, i.e., verbs, adjectives, and adverbs of emotions. In fact, emotions are more often expressed implicitly through concepts with an affective valence such as *play a game*, *be laid off*, or *go on a first date*.

Sentic Computing overcomes this problem using a common sense reasoning [15] approach and a novel emotion categorization born from the idea that our mind consists of four independent emotional spheres, whose different levels of activation make up the total emotional state of the mind.

Common Sense

When people communicate with each other, they rely on similar background knowledge, e.g., the way objects relate to each other in the world, people's goals in their daily lives, the emotional content of events or situations.

This taken for granted information is what we call common sense—obvious things people normally know and usually leave unstated. The Open Mind Common Sense project has been collecting this kind of knowledge from volunteers on the Internet since 2000 to provide intuition to AI systems and applications.

AffectiveSpace

AffectiveSpace [16] is a vector space of affective concepts built by applying multidimensionality reduction techniques over the blend of ConceptNet [17], a directed graph representation of the Open Mind corpus, and WordNet-Affect [18], a linguistic resource for the lexical representation of affective knowledge. This alignment operation yields a new dataset in which common sense and affective knowledge coexist.

After performing singular value decomposition (SVD) on the resulting matrix, we use a trial and error approach to discard those components representing relatively small variations in the data, i.e., all but the first 50 principal components of the original matrix. This technique is termed truncated SVD and yields a fifty-dimensional space (illustrated in Fig. 1) in which different vectors represent different ways of making binary distinctions among concepts and emotions.

In AffectiveSpace, everyday life concepts like *have breakfast*, *meet people,* or *watch tv* are linked to a hierarchy of affective domain labels. By exploiting the information sharing property of truncated SVD, concepts with the same affective valence are likely to have similar features, i.e., concepts concerning the same emotion tend to fall near each other in the vector space.

For example, we can find separated groups of affectively related concepts such as *love*, *satisfaction*, *laugh* and *sing* or *sick*, *isolation*, *frustration* and *depression*. However, similarity and analogy in AffectiveSpace do not depend on concepts' absolute positions in the vector space but only on their positions relative to each other. Concepts and emotions are represented by vectors of 50 coordinates: these coordinates can be seen as describing concepts in terms of *eigenmoods* that form the axes of AffectiveSpace, i.e., the basis $e_0, \ldots, e_{49}$ of the vector space.

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. Consequently, concepts with negative e_0 components have negative affective valence.

The Sentics Extraction Process

The complete process for inferring emotions from text numbers a Natural Language Processing (NLP) module, which performs a first skim of the document, a Semantic Parser, whose aim is to extract concepts from the processed text, and eventually AffectiveSpace, for the analysis of concepts' affective valence (Fig. 2).

The NLP module interprets all the affective valence indicators usually contained in text such as special punctuation, complete upper-case words, onomatopoeic repetitions, exclamation words, negations, degree adverbs, and emoticons. The Semantic Parser then deconstructs text into concepts and provides, for each of them, the relative

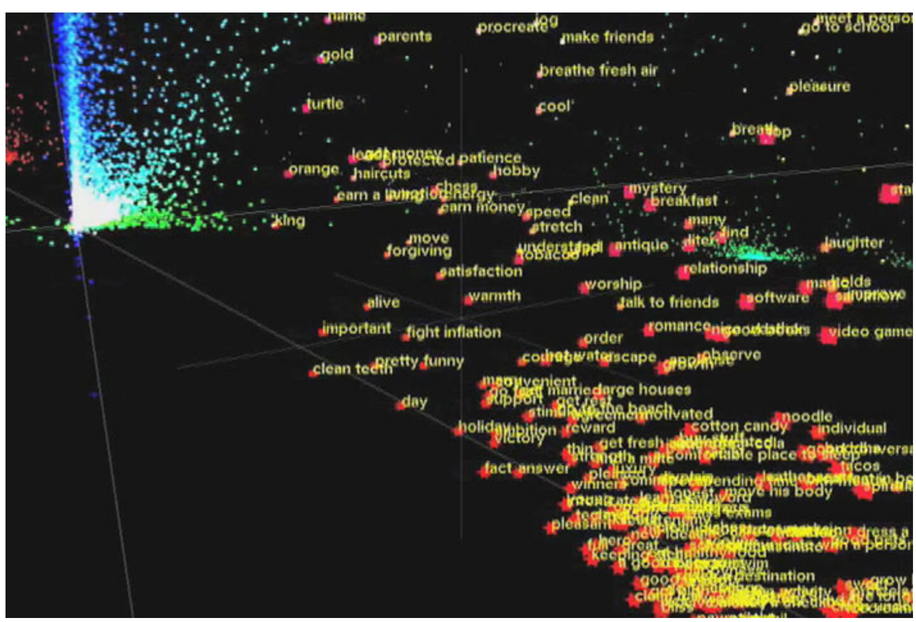

Fig. 1 AffectiveSpace

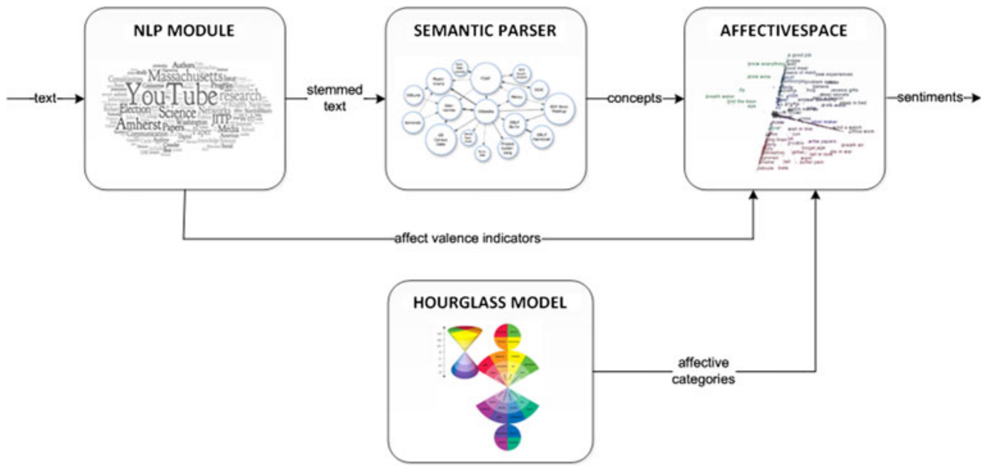

Fig. 2 Sentics extraction process

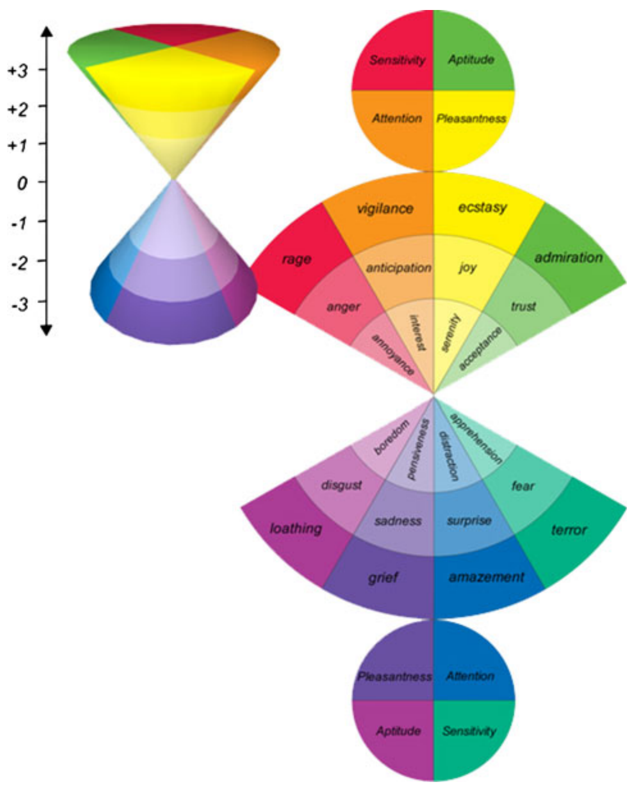

Fig. 3 The hourglass of emotions

frequency, valence, and status, i.e., the concept's occurrence in the text, its positive or negative connotation, and the degree of intensity with which the concept is expressed.

AffectiveSpace finally extracts, from the set of concepts so far obtained, a list of four-dimensional vectors, called 'sentic vectors', which contain the affective information of each concept in terms of Pleasantness, Attention, Sensitivity, and Aptitude, i.e., the sentic dimensions of the Hourglass of Emotions [19] (Fig. 3).

This model is a variant of Plutchik's emotion categorization [20] and constitutes an attempt to emulate Marvin Minsky's conception of emotive reasoning. Minsky sees the mind as made of different independent resources and believes that our emotional states result from turning some set of these resources on and turning another set of them off [21]. 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.

We exploit this model to cluster AffectiveSpace around 24 centroids (that is the labels of the Hourglass) using a k-nearest neighbor (k-NN) approach. Hence, every time the Semantic Parser detects a concept, this is projected into the vector space and its affective valence is calculated, for each sentic dimension, according to the position the concept occupies in the space (i.e., according to the dot product between this and the nearest centroid).

Representation Layer

Web resources and social media resources, in particular, represent a peculiar kind of data that are characterized for its deeply interconnected nature. The Web itself in fact is based on links that bound together different data and information and social media resources characterize themselves for the collaborative way in which they are created and maintained. An effective description of such resources needs therefore to capture and manage their interconnected nature, allowing to encode information not only about the resource itself but also about the linked resources into an unique knowledge base. To achieve this purpose, we exploit Semantic Web techniques.

The Semantic Web

The Semantic Web uses uniform resource identifiers (URIs) to univocally identify entities and the resource description framework (RDF) to express the information in an univocally interpretable format, whose basic building block is an object-attribute-value triple, i.e., a statement. Resources may be authors, books, publishers, places, people, hotels, rooms, search queries, and so on. Properties are a special kind of resources that describe relations between resources such as *writtenBy*, *age*, *title*, and so on. Statements assert the properties of resources. To provide machine-accessible and machine-processable representations, it is usual to encode the RDF triples using a XML syntax. However, RDF does not make assumptions about any particular application domain, nor does it defines the semantics of any domain. For this purpose, it is necessary the introduction of ontologies.

Ontologies basically deal with knowledge representation and can be defined as formal explicit descriptions of concepts in a domain of discourse (named classes or concepts), properties of each concept describing various features and attributes of the concept (roles or properties), and restrictions on property (role restrictions).

An ontology, together with a set of individual instances of classes, constitutes a knowledge base. Ontologies make possible the sharing of common understanding about the structure of information among people or software agents. In addiction, encoding the semantic of the information through the definition of a taxonomical organization of concepts and properties and the formalization of their relationships, ontologies make possible reasoning. This means that starting from the data and the semantic of data added by an ontology, it is possible to infer new relationship between data by revealing the knowledge implicitly contained in data. This information can then be used to answer more advanced user queries or as input data to other applications.

Within the Semantic Web different languages have been developed for designing ontologies, in particular the RDF schema (RDFS) and the ontology web language (OWL). RDFS can be seen as an RDF vocabulary and a primitive ontology language since it offers certain modeling primitives with fixed meaning.

Key concepts of RDFS are class, subclass relations, property, subproperty relations, domain, and range restrictions. OWL is a language more specifically conceived for ontologies creation. It is built upon RDFS and uses XML-based RDF syntax. OWL introduces a number of features that are missing in RDFS like local scope of property, disjointness of classes, cardinality restriction, Boolean combination of classes, etc.

Once encoded in a semantic aware language on the base of a specific domain ontology, the relevant information can be stored in a triple-store, a purpose-built database for the storage and retrieval of RDF data, to allow powerful queries that exploit the inference capabilities provided by the semantic data modeling.

The Human Emotion Ontology

In the last years, the study of emotions has attracted a growing attention from different research fields, ranging from advanced signal processing to psychology, from AI to linguistics. Such great research effort has led to great advances in the understanding of emotions and in the development of many different models and classification techniques.

As a result, a standardization of the knowledge about emotions is getting more and more important but at the same time more difficult. Within the scientific community, in fact, the debate over human emotions is still open and there is still not common agreement about which features are the most relevant in the definition of an emotion and which are the basic emotions and their names.

Dealing with the emotion extraction from media, the kind of used descriptors and techniques strictly rely on the considered medium. It is evident, for example, that the characteristics extracted from an audio sample rely basically on speech prosody, while features extracted from a video sample focus on face expressions. For this reason, there exists an unavoidable and intrinsic heterogeneity in the emotion descriptors that make impossible to define a standard and unique set of such descriptors that could grant at the same time flexibility and interoperability.

Very interesting works in the definition of proper languages for emotions description have been recently done by the Humaine Project [22], the COST Action 2102 [23], and the W3C Emotion Markup Language Incubator Group [24]. The latter, in particular, has recently published a first report about the preliminary work for the definition of EmotionML, a markup language for the description and annotation of emotions.

The main issue of such languages, as EmotionML and EARL, is represented by the limited semantic expressiveness of the language in which they have been encoded. XML in fact lacks of formal semantics and provides only a strict syntactic representation of information based on preexisting schema. This does not allow an effective mapping between different concepts or properties and does not supply inference power that can be exploited to enhance system intelligence in processing and managing affective data.

To overcome such limitations, we have developed a Human Emotion Ontology (HEO), which standardizes the main sets of existing human emotion descriptors, introduced by the existing languages for emotion description, into a computational ontology. HEO (illustrated in Fig. 4) is conceived as an high level ontology for human emotions, which supplies the most significant concepts and properties that are necessary to provide accurate human emotion descriptions, whose semantic is univocally defined and machine interpretable. These high level features can also

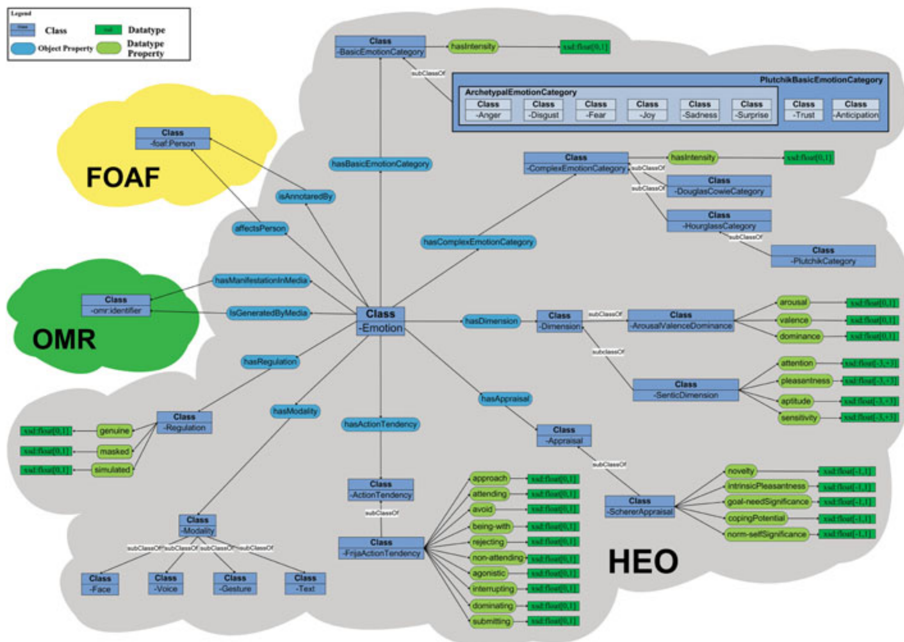

Fig. 4 Merging HEO with other ontologies for representing social media affective

be further refined using lower level concepts and properties related to more specific descriptions or linked to other more specialized ontologies.

In fact, the rich semantic expressiveness of OWL language has been exploited in encoding HEO in order to univocally provide interpretable semantics to its descriptors and hence allow its extensibility and its interoperability with other existing ontologies. In this way, HEO provides at the same time flexibility in emotion descriptions, by allowing the use of a wide and extensible set of descriptors to define all the main features of an emotion, and interoperability for different applications that use different vocabularies or emotion models, by allowing to map concepts and properties belonging to such different emotion representations. Also, OWL inference capabilities are exploited to provide a more effective management of the information encoded in the knowledge base.

In HEO, an emotion can be described both in a discrete way, using the *hasCategory* property, and in a dimensional way, using the *hasDimension* property.

HEO introduces two main disjoint classes for describing emotions by category: *BasicEmotionCategory* and *ComplexEmotionCategory*. The reason for this distinction is two-fold. Firstly, there are some emotion models that define complex emotions as a combination of two or more basic emotions. Secondly the same descriptor is often used as a basic emotion or as complex emotion, according to different models, but the represented concept is different. Different models can be used both for expressing the basic emotions, e.g., the 6 archetypal emotions by Ekman (anger, disgust, fear, joy, sadness, surprise) or the 8 basic emotions by Plutchik (acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise), and for the complex emotions using wider emotion sets, e.g., the 48 descriptors by Douglas-Cowie or the 40 emotions of the Hourglass model.

Instead, to describe emotions by dimension, HEO uses the *hasDimension* property which includes the arousal-valence-dominance model and the sentic dimensions (Pleasantness, Attention, Sensitivity, and Aptitude). HEO has been developed in OWL description logic (OWL DL) to take advantage of its expressiveness and its inference power in order to map the different models used in the emotion description. OWL DL, in fact, allows a taxonomical organization of emotion categories and properties restriction to link emotion description made by category and dimension.

In HEO, for example, Ekman's 'joy' archetypal emotion represents a superclass for Plutchik's 'ecstacy', 'joy', and 'serenity' emotions. Using property restriction, the basic Plutchik's 'joy' emotion can also be defined as an emotion that 'has Pleasantness some float $\{+1, +2\}$', 'interest' as an emotion that 'has Attention some float $\{0, +1\}$ ' and 'love' as an emotion that 'has Pleasantness some float $\{0, +3\}$ and Aptitude some float $\{0, +3\}$'. In this way, for example, querying a database that support 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 Plutchik basic emotion of type 'joy' and the emotions that have Pleasantness values in $\{+1, +2\}$.

Describing Media, People and Their Emotions

A serious barrier in managing social media has been represented till now by the lack of a standardization for media resources descriptors. An important effort to help circumventing the current proliferation of audio/video metadata formats is currently carried on by the W3C Media Annotations Working Group, which developed the ontology for media resource (OMR) [25]. OMR offers a core vocabulary to describe media resources on the Web, introducing descriptors such as *title*, *creator*, *publisher*, *createDate*, and *rating*. It defines semantics-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. A recognized standardization for people description is defined in the friend of a friend ontology (FOAF) [26]. FOAF represents information about people, 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 supply most of the vocabulary that we need for describing media and people and represent an authoritative reference. For this reason, we encode our annotation using their descriptors and adding others only when necessary. For example, OMR, at least in the current realization, 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. In addition, inside HEO, we have introduced some properties that allow to connect this ontology with OMR and FOAF, i.e., to connect an emotion with a person and a media. In particular, we have defined the *hasManifestationInMedia* and *isGeneratedByMedia*, to describe emotions that respectively occur and are generated in a media, and the property *affectPerson* to connect emotions to people. Thus, the combined use of HEO, OMR, and FOAF gives a complete framework to describe jointly not only multimedia contents and the users that have created, uploaded, or interacted with it but also the affective content carried by the media and the way it is perceived by people.

Application Layer

Due to the way they are created and maintained, community-contributed multimedia resources are very different from standard web data. One fundamental aspect is constituted by the collaborative way in which such data are created, uploaded, and annotated.

A deep interconnection emerges in the nature of these data and metadata, allowing for example to associate videos of completely different genre but uploaded by the same user or different users, even living in opposite sides of the world, who have appreciated the same pictures. All the processes involving data connected to people's social lives are often strongly driven by emotions. Emotions, in fact, play a fundamental role in our life, influencing cognition, perception, and everyday tasks such as learning, communication, and decision making. Commonly, users upload videos or photos that are linked to particular moments of their lives and post comments, or rate videos and photos, according to the emotions arisen while watching them.

We believe that, to create effective applications dealing with social media data, it is necessary to treat conveniently their peculiarities. Thus, it is necessary to manage all the wide sets of associated information, and particularly their emotional content, as an interconnected knowledge base and hence make them available through intuitive interfaces.

Affective Faceted Classification

As proof of concept, we created a demonstrative web site [27] in which 100 YouTube videos, taken from an online publicly available dataset [28], can be displayed and navigated using the faceted browsing paradigm. The purpose was two-fold. On one side to show how the affective information can be exploited for the affective classification of YouTube videos. On the other, to demonstrate how the semantic encoding of data can be used to improve the productivity of the search queries and to provide more engaging interfaces for their fruition.

We chose YouTube as source for our data because it nowadays is the world's most popular on line video community, with more than 2 billion videos watched everyday and more than 24 h of video uploaded every minute. We also wanted to remark how affective information extraction from text can be used not only for textual resource classification but also for every kind of community-contributed resource (as video, images, products, etc). In addition, YouTube supplies powerful DATA APIs that provide access to a wide set of information about every video, such as its genre, duration, rating, and upload date (also visible on a timeline), about their uploaders, such as their name, gender, and country and also to retrieve all the comments added to a video.

The purpose of our demo was also to show how the application of Semantic Web techniques allows heterogeneous information coming from different sources (e.g., affective information extracted from user comments and video information retrieved through YouTube APIs) to be merged together and encoded in a unique interconnected knowledge base that can be exploited to enhance data management and fruition. For example, information becomes available for being accessed and displayed according to the faceted browsing paradigm that allows to browse and query the information in an intuitive and engaging fashion.

The 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, predetermined, taxonomic order. This makes possible to perform searches combining the textual approach with the navigational approach. Faceted search, in fact, enables users to navigate a multidimensional information space by concurrently writing queries in a text box and progressively narrowing choices in each dimension.

In our application, we implemented the faceted browsing paradigm using SIMILE Exhibit API [29], a set of JavaScript files that allows to easily create rich interactive web pages including maps, timelines, and galleries with very detailed client-side filtering. Exhibit pages allow to display semantically structured data stored in a file in a Semantic Web aware format (as RDF or JSON). One of the most relevant aspect 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.

Among all the data available, we extracted 100 comments associated with each video and processed them using Sentic Computing to extract the related affective information. We then encoded all the retrieved information in RDF/XML, using the descriptors defined by HEO, OMR, and FOAF, and stored it in a Sesame [30] triple-store. Sesame can be embedded in applications and used to conduct a wide range of inferences on the information stored, based on RDFS 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 knowledge stored inside Sesame can be queried and the results can also be retrieved in a semantic aware format and used for other applications. For our demo, we exported all the information contained into Sesame in a JSON file to feed our Exhibit. The videos are dynamically embedded into the page and, jointly with their associate information, they can be visualized both as a gallery and on a timeline (to be browsed according to their upload date) (Fig. 5).

Using faceted menus, it is possible to explore such information both using the search box, to perform keyword-based queries, and filtering the results using the facet menus, i.e., by adding or removing constraints on the facet properties. Several properties are available for filtering the social media, combining constraints about the video, the uploader and the affective content carried by the video. In this way, it becomes very easy and intuitive to accomplish even complex queries, such as 'find all the music video uploaded by a certain user with an associated emotion of type *Joy*'.

Evaluation

One of the main issues in evaluating the performances of an affect recognition system is the difficulty in finding datasets of emotionally tagged text excerpts whose tags are reliable and whose dimension is large enough to be statistically significant. Unfortunately, YouTube does not provide any affective tags for its videos hence, in order to test Sentic Web's affect recognition capabilities, we

Fig. 5 Screenshot of a Sentic Web application powered by exhibit

evaluated the system with a corpus of mood-tagged blogs from LiveJournal (LJ) [31].

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 with a mood tag, 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 Web.

However, since LJ mood themes do not perfectly match sentic levels, we had to consider a reduced set of 10 moods, i.e., 'ecstatic', 'happy', 'pensive', 'surprised', 'enraged', 'sad', 'angry', 'annoyed', 'scared', and 'bored'. Moreover, we could not consider non-affective web posts since untagged blog entries do not necessarily lack emotions. All LJ accounts have Atom, RSS and other data feeds which show recent public entries, friend relationships, and interests. Unfortunately, there is no possibility to get mood-tagged blog-posts via data feeds so we had to design our own crawler. After retrieving and storing relevant data and metadata from 5,000 LJ posts, we extracted sentics through the Sentics Extraction Process and compared the output with the relative mood-tags, in order to calculate statistical classifications such as precision and recall.

On average, each post contained around 123 words and, from it, about 3 affective valence indicators and 42 sentic vectors were extracted. According to this information, we assigned mood-labels to each post and compared these with the corresponding LJ mood-tags, even if a direct comparison with other affect extraction methods cannot be performed because of the different datasets used and the different affective states that are taken into account, evaluation results evidence a very good accuracy for each of the 10 selected moods. Among these, 'happy' and 'sad' posts were identified with particularly high precision (81 and 76%, respectively) and decorous recall rates (67 and 62%), for a total F-measure value of 73 and 68%, respectively, which outperforms the classification results obtained using the baseline methods (44 and 35% for keyword spotting, 53 and 39% for lexical affinity, 61 and 52% for statistical methods).

Conclusion and Future Efforts

As the Web plays a more and more significant role in people's social lives, it contains more and more information concerning their feelings and opinions but most of this informative content slips from all the searches that are everyday performed in the Web. In this paper, we have shown how AI and Semantic Web techniques can be used in conjunction to manage the affective information associated with community-created data and metadata.

This information can be accessed as a whole knowledge base allowing advanced features selection, both in querying the information and in displaying it. We are currently working on redesigning Sentic Web as a web service, so that it can be dynamically used and embedded in other applications.

In this way, it will be possible to use the affective information contained in web pages to trigger the behavior of applications and to really implement what we envision as Sentic Web—a tomorrow's Web in which emotion-based search queries will be possible and in which applications will be aware of the user's emotional state and will offer an ad hoc service according to it.

Acknowledgments This work was undertaken during the first author's research visit to Sitekit Labs and partly funded with the aid of a COST 2102 Short Term Scientific Mission.

References

1. Hendler J, Shadbolt N, Hall W, Berners-Lee T, Weitzner D. Web science: an interdisciplinary approach to understanding the world wide web. Comm ACM 2008;51(7):60–69.
2. Cambria E, Hussain A, Havasi C, Eckl C. Sentic computing: exploitation of common sense for the development of emotion-sensitive systems. LNCS, vol. 5967. Berlin: Springer; 2010. pp. 148–156.
3. Grassi M. Developing HEO human emotion ontology. LNCS, vol. 5707. Berlin: Springer; 2009. pp. 244–251.
4. Umbria – http://umbrialistens.co.
5. Cymfony – http://www.cymfony.co.
6. Nielsen – http://nielsen-online.co.
7. Evolve24 – http://evolve24.co.
8. Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. In: EMNLP, Philadelphia 2002.
9. Sood S, Owsley S, Hammond K, Birnbaum L. Reasoning through search: a novel approach to sentiment classification. Northwestern University Tech Report Number NWU-EECS-07-05 2007.
10. Turney P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: ACL, Philadelphia 2002.
11. Alm C, Roth D, Sproat R. Emotions from text: machine learning for text-based emotion prediction. In: HLT/EMNLP, Vancouver 2005.
12. Pang B, Lee L. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: ACL, Barcelona 2004.
13. Sood S, Vasserman L. Esse: exploring mood on the web. In: ICWSM, Seattle 2009.
14. Antoniou G, Van Harmelen F. A semantic web primer. Cambridge: Mit Press; 2004.
15. Cambria E, Hussain A, Havasi C, Eckl C. Common sense computing: from the society of mind to digital intuition and beyond. LNCS, vol. 5707. Berlin: Springer 2009. pp. 252–259.
16. Cambria E, Hussain A, Havasi C, Eckl C. AffectiveSpace: blending common sense and affective knowledge to perform emotive reasoning. In: WOMSA at CAEPIA, Seville 2009.

17. Havasi C, Speer R, Alonso J. ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge. In: RANLP, Borovets 2007.
18. Strapparava C, Valitutti A. WordNet-Affect: an affective extension of WordNet. In: LREC, Lisbon 2004.
19. Cambria E, Hussain A, Havasi C, Eckl C. SenticSpace: visualizing opinions and sentiments in a multi-dimensional vector space. LNAI, vol. 6279. Berlin: Springer; 2010. pp. 385–393.
20. Plutchik R. The nature of emotions. Am Scient 2001;89(4): 344–350.
21. Minsky M: The emotion machine. New York: Simon and Schuster; 2006.
22. The Humaine Project – http://emotion-research.ne.
23. COST Action 2102: Cross-Modal Analysis of Verbal and Non-Verbal Communication – http://cost2102.cs.stir.ac.u.
24. W3C Emotion Markup Language – http://w3.org/2005/Incubator/emotio.
25. Ontology for Media Resource – http://w3.org/TR/mediaont-1.
26. Friend of a Friend Project – http://foaf-project.or.
27. Sentic Web demo – http://semedia.dibet.univpm.it/senticwe.
28. Cheng X, Dale C, Liu J. Statistics and Social Network of YouTube Videos. In: IEEE IWQoS, Enschede 2008.
29. SIMILE Exhibit API – http://simile-widgets.org/exhibi.
30. Sesame – http://www.openrdf.or.
31. LiveJournal – http://livejournal.co.