OntoSenticNet: A Commonsense Ontology for Sentiment Analysis

Mauro Dragoni
Fondazione Bruno Kessler

Soujanya Poria
Nanyang Technological University

Erik Cambria
Nanyang Technological University

Editor:
Erik Cambria

In this work, we present OntoSenticNet, a commonsense ontology for sentiment analysis based on SenticNet, a semantic network of 100,000 concepts based on conceptual primitives. The key characteristics of OntoSenticNet are: (i) the definition of precise conceptual hierarchy and properties associating concepts and sentiment values; (ii) the support for connecting external information (e.g., word embedding, domain information, and different polarity representations) to each individual defined within the ontology; and (iii) the capability of associating each concept with annotations contained in external resources (e.g., documents and multimodal resources).

In recent years, sentiment analysis has become increasingly popular for processing social media data on online communities, blogs, wikis, microblogging, platforms, and other online collaborative media.[1] Sentiment analysis is a branch of affective computing research[2] that aims to classify text (but sometimes also audio and video)[3] as either positive or negative (but sometimes also neutral).[4] Most of the literature is on English language but recently an increasing number of publications are tackling multiple languages in their research.[5]

Sentiment analysis techniques can be broadly categorized into symbolic and sub-symbolic approaches: the former include the use of lexicons, ontologies, and semantic networks to encode the polarity associated with words and multiword expressions; the latter consist of machine learning techniques that perform sentiment classification based on word co-occurrence frequencies. While most works approach it as a simple categorization problem, sentiment analysis is actually a suitcase research problem[6] that requires tackling many natural language processing (NLP) tasks, e.g., named-entity recognition,[7] personality recognition,[8] sarcasm detection,[9] and aspect extraction.[10]

Sentiment analysis has raised growing interest, both within the scientific community, leading to many exciting challenges, as well as in the business world, due to the remarkable benefits to be had from financial[11] and political[12] forecasting, user profiling[13] and community detection,[14] computational advertising[15] and dialogue systems,[16] etc. However, mining opinions and sentiments from multimodal resources (texts, images, videos, audio-recordings, etc.) is an extremely difficult task because it requires a deep understanding of the explicit and implicit, regular and irregular, features (linguistic, visual, or audio) of a resource.

Existing approaches to multimodal sentiment analysis rely mainly on mapping multimodal information to parts of text in which opinions are explicitly described, such as polarity terms, affect words, and their co-occurrence frequencies. However, opinions and sentiments associated with these parts of text are often conveyed implicitly through latent semantics, which make purely syntactic approaches ineffective. The task of associating polarities to these features suffers from the limitation of not being able to perform inference operations on the concepts extracted (or mapped in case of multimodal resources) from the resource to analyze.

For example, it is very difficult to extract all the possible inflections of an example like "buy_drink," as this can be expressed in countless ways by using different combinations of the many synonyms of the verb buy (e.g., purchase, acquire, obtain, get, etc.) and the several synonyms of the concept drink (e.g., water, beer, booze, liquor, cocktail, etc.) and all its instances, e.g., tonic_water, mojito, coke, vodka_lemon, etc.

In this article, we present OntoSenticNet, a commonsense ontology for sentiment analysis based on SenticNet,[17] a semantic network of 100,000 concepts based on conceptual primitives (Figure 1). The characteristics that distinguish OntoSenticNet (available for download at http://sentic.net/ontosenticnet.zip) from previous versions of SenticNet are: (i) the definition of precise conceptual hierarchy and properties associating concepts and sentiment values; (ii) the support for connecting external information (e.g., word embeddings, domain information, different polarity representations, etc.) to each individual defined within the ontology; and (iii) the capability of associating each concept with annotations contained in external resources (e.g., documents, multimodal resources, etc).

OntoSenticNet does not blindly use keywords and word co-occurrence counts, but instead relies on the implicit meaning associated with commonsense concepts. Unlike purely syntactic techniques, OntoSenticNet can detect subtly expressed sentiments by enabling the analysis of multiword expressions that do not explicitly convey emotion but are instead related to concepts that do. Moreover, the provided representation supports the integration of reasoning engines able to infer implicit sentiment information.

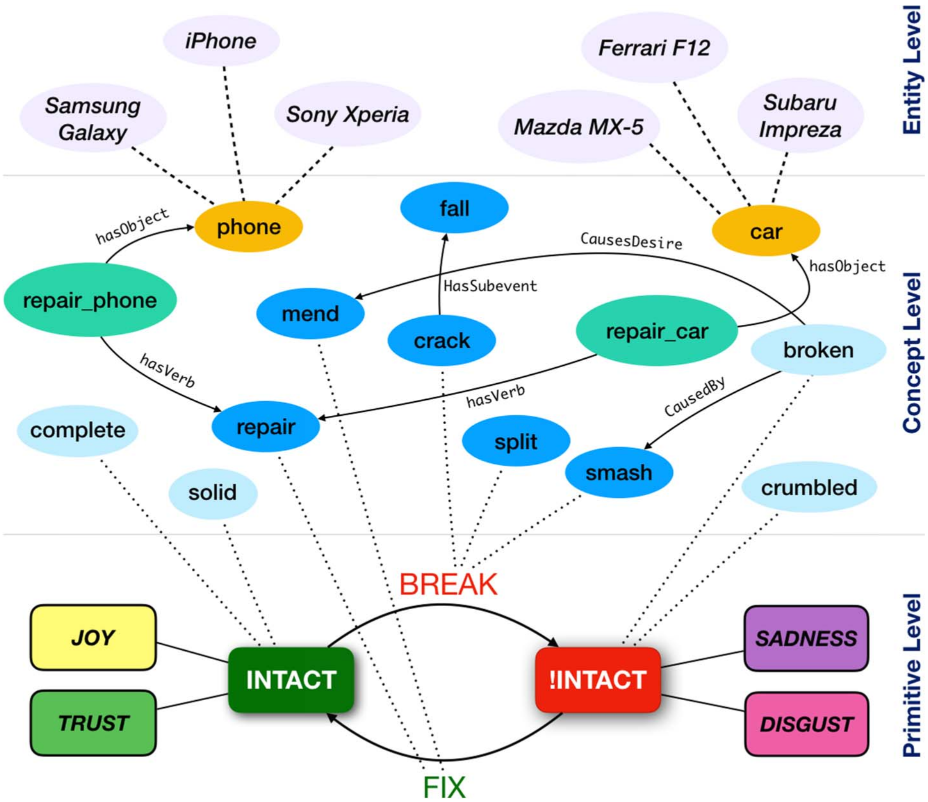

Figure 1. Overview of SenticNet. Commonsense knowledge is organized at three levels: entities link down to concepts, which in turn link down to conceptual primitives, where meaning is encoded in terms of polarity and emotions.

SENTIMENT ONTOLOGIES

Despite the rise of sentiment analysis, there is a lack of sentiment ontologies. In particular, there are only two general models, and they are limited in terms of functionalities and possibility of being integrated into real-world applications. The Emotion Markup Language (EML) was created for supporting the task of annotating documents with tags extracted from customized vocabularies (http://w3.org/TR/emotionml). On the one hand, this language is useful for creating emotional dictionaries for a specific domain. On the other hand, the effort necessary for creating a new resource is significant, and, at the same time, the promotion of a markup language fosters the proliferation of resources that often have a high linguistic and semantic overlap. This way, reusability is strongly penalized.

The other model is the Emotion Ontology (MFOEM), which was developed for supporting a structured representation of mental functioning, including mental processes such as cognition and traits such as intelligence (http://bioportal.bioontology.org/ontologies/MFOEM). This ontology can complement SenticNet in a sense that, while OntoSenticNet is specifically thought for describing the emotional domain, MFOEM can be considered an upper level ontology that can be aligned with the top-level concepts of SenticNet. This way, OntoSenticNet would benefit from categorizations and properties describing the human brain from a more general perspective and, at the same time, the MFOEM ontology can exploit the granularity of OntoSenticNet for accessing real-world emotional items (documents, videos, images, etc.). We leave this alignment task for future work.

BUILDING ONTOSENTICNET

The modeling of OntoSenticNet requires the ability of representing a set of entities able to summarize not only the basic concepts for describing the polarity associated with opinions, but also support a semantic representation of such opinions, their similarity, and the relationships between each opinion's words and real-world multimodal resources. The construction of OntoSenticNet is therefore driven by typical questions that arise when building ontological representations of a domain, that is:

- "Which are the entities that exist, or can be said to exist, in sentiment analysis and opinion mining domain?"
- "How can such entities be grouped, related within a hierarchy, and subdivided according to similarities and differences?"
- "How can such entities be modeled in order to easily support the task of annotating real-world resources?"

These questions are motivated from the philosophical process of ontology building, where ontology engineers have to investigate the essence and the nature of being of each entity. Instead, we answer these questions from a computer science point of view, where conceptual representations are used for fostering clarity, reuse, and mutual understanding of information.

The process of building OntoSenticNet follows the METHONTOLOGY methodology.[18] This methodology proposes a general method for building any kind of ontology or meta-ontology, and it is based on the experience acquired in developing ontologies in the domain of chemicals. METHONTOLOGY provides a set of guidelines of how the activities identified in the ontology development process should be carried out, what kinds of techniques are the most appropriate in each activity, and what products each one produces. The methodology is split in seven phases. For space reasons, we are not able to provide an in-depth description of all phases, but we briefly report the aspects that guided the process of building OntoSenticNet.

Specification. OntoSenticNet has been thought of for filling the gap between fundamental emotion ontologies (like EML and MFOEM) that cannot be easily integrated into real-world applications and sentiment (or opinion) words dictionaries that, instead, do not support a semantic representation of the emotion domain. Moreover, OntoSenticNet aims to bridge concepts and resources in order to enable its integration into complex annotation and reasoning frameworks. OntoSenticNet is represented by using a natural-language semi-formal format due to the necessity of adopting concept names expressing specific meanings through their labels. The granularity level is classified as *high* thanks to the rich set of terminologies and commonsense expressions represented in the ontology.

Knowledge Acquisition. OntoSenticNet is not built by manually labelling pieces of knowledge coming from general-purpose resources such as WordNet or DBpedia. Instead, it is automatically constructed by applying graph-mining and multi-dimensional scaling techniques on the affective commonsense knowledge collected from three different sources, namely: WordNet-Affect, Open Mind Common Sense, and GECKA.[18] This knowledge is represented redundantly at three levels: semantic network, matrix, and vector space. Subsequently, semantics and sentics are calculated though the ensemble application of spreading activation, neural networks, and an emotion categorization model.

Conceptualization. The conceptualization of OntoSenticNet was split in two steps. The first step was covered by the knowledge acquisition phase, where most of the terminology is collected and directly modeled into the ontology. The second step, instead, consisted in defining the concept and the properties (both object and data) used for providing a detailed, as well as complex, representation of concept polarities and for supporting annotation tasks. Details about the modeled concepts are provided in the next section, where the rationale of each concept is described.

Integration. No specific integrations have been performed during the developing process of OntoSenticNet. As mentioned earlier, OntoSenticNet is not based upon existing meta-ontologies.

Implementation. The implementation of OntoSenticNet is provided in two programming languages. RDF/XML provides a formal representation enabling the check of inconsistencies, the visualization of the ontology structure, and the ease of maintenance. Python, instead, provides an easier support for the integration of OntoSenticNet into real-world applications. The two versions are always synchronized.

Evaluation. OntoSenticNet has been verified by using the verification framework proposed in METHONTOLOGY. Based on the criteria proposed in the framework, OntoSenticNet has been assessed as *correct*, *complete*, *consistent*, and *not redundant*.

Documentation. During the knowledge acquisition phase, we collected documentation about the information sources used for modeling OntoSenticNet. As for the conceptual documentation, we produced the set of guidelines we followed, for each phase, to model all concepts, objects properties, and data properties. The rationale behind the modeling of each concept is presented in the next section.

ENTITIES OF ONTOSENTICNET

OntoSenticNet is composed of four main branches (Figure 2): *SenticConcept, Domain, PolarityInstance*, and *Resource*.

SenticConcept. The *SenticConcep*t entity models what in classic sentiment analysis are called "opinions words." This entity represents the basic concept grouping all concepts that can be used for representing linguistic elements that can be associated with a sentiment expression. As children of *SenticConcept*, we modeled three further concepts representing the three kinds of sentiment elements that can be found in natural language text and that can be used for annotating multimodal resources: *SingleToken, CommonsenseExpression*, and *Emotion*. The *SingleToken* concept embodies the well-known "opinion word" element generally adopted in sentiment analysis and used for performing single computations of text orientations (positive, neutral, or negative). With the *CommonsenseExpression* concept, we introduce the conceptualization of lexical expressions used for representing a complex sentiment status. A sample instance of this concept is "buy_christmas_gift." By modeling this kind of sentiment concepts, it is possible to support multimodal annotation activities where the single sentiment concepts are not expressive enough for providing a complete conceptual description of the sentiment status. The third concept is *Emotion*, used for representing primitive moods like "joyful" or "sad" and used for supporting a clustered representation of both *SingleToken* and *CommonsenseExpression* instances.

Domain. The second branch is described only by the concept *Domain* but it represents an important step of the process of modeling sentimental status in real-world applications. Most of the literature concerning sentiment analysis and opinion mining does not care about the emotional differences that the same lexical expressions may have within different contexts, or domains. Instead, there are plenty of adjectives or complex lexical expressions that assume different sentiment values based on their contexts. An example is given by the adjective "small" where, if we talk about an item's ability to hold or contain objects it assumes a negative connotation. While, if we talk about an item's ease of portability, it conveys a positive polarity. With this branch, we support the instantiation of domains and contexts that are of interest for the application where OntoSenticNet is deployed into.

PolarityInstance. The third branch, *PolarityInstance*, contains the conceptualizations of the different kinds of polarity representation supported by the ontology. In this version of OntoSenticNet, we foresee two kinds of polarity representations: *CrispPolarity* and *FuzzyPolarity*. *CrispPolarity* instances are represented by double value types used for associating a single-value representation to the instances of the *SenticConcept* entities. For the instances of the *FuzzyPolarity* concept, instead, we want to support a more complex representation of polarity values by integrating in such a representation the uncertainty that is associated with them. Indeed, the assignment of a specific value to a *SenticConcept* instance is a subjective task that may result in a collection of several polarity values. The use of fuzzy logic for representing these values allows real-world applications to properly interpret these polarity values during the inference task.[20]

Resource. The last branch is associated with the *Resource* entity. The use of a concept for describing resources' identifiers enable the linking operation between OntoSenticNet and external artifacts (associated with persistent identifiers) in order to create collections of concrete annotated entities that can be exploited for further reasoning activities. Within OntoSenticNet, we identified four kinds of resources modeled by the following concepts: *AnnotatedTextResource*, *AnnotatedMultimodalResource*, *EmbeddingResource*, and *ExternalReference*. The *AnnotatedTextResource* concept is instantiated when a textual document is annotated with one or more instances of *SenticConcept*. Similarly, *AnnotatedMultimodalResource* is instantiated when the annotated resources are images or videos. The support of this kind of conceptualization enables the possibility of refining, or learning, how instances of the *SenticConcept* entity are used for annotating content and, at the same time, to trigger machine learning activities on annotated resources for improving inference capabilities. The third concept is *EmbeddingResource*. Due to the relevant use of feature embeddings for machine learning purposes, we wanted to provide a way for associating such embeddings within OntoSenticNet.Instances of this concept are represented by an array of double values that can be directly associated with instances of type *SenticConcept*. These associations can be exploited for inference purposes and for supporting internal representation of documents within data repositories. The last concept of this branch is *ExternalReference*. Instances of this concept are used for linking purposes. Terminologies defined in external linguistic resources (for instance WordNet) can be linked with instances of type *SenticConcept* through these mappings. Once such mappings are defined, OntoSenticNet can be used as an entry point for acquiring further information for the linked external resource.

Additional Annotations. Besides the four branches presented above, we also included a further concept used for instantiating particular scenarios: the *ComplexSenticEntity* concept. This concept allows for the modeling of situations where to an instance of type *SenticConcept*, it is necessary to associate a specific domain (instance of type *Domain*) and a specific polarity (instance of type *PolarityInstance*). Besides concepts, OntoSenticNet defines a list of *ObjectProperty* and *DataProperty* modeling relationships between entities. Due to space limits, we do not provide a detailed description of each of them. However, we report them in Table 1 and 2, respectively. Finally, we included two annotations for supporting unique identification of entities: *id* and *resourceIRI*.

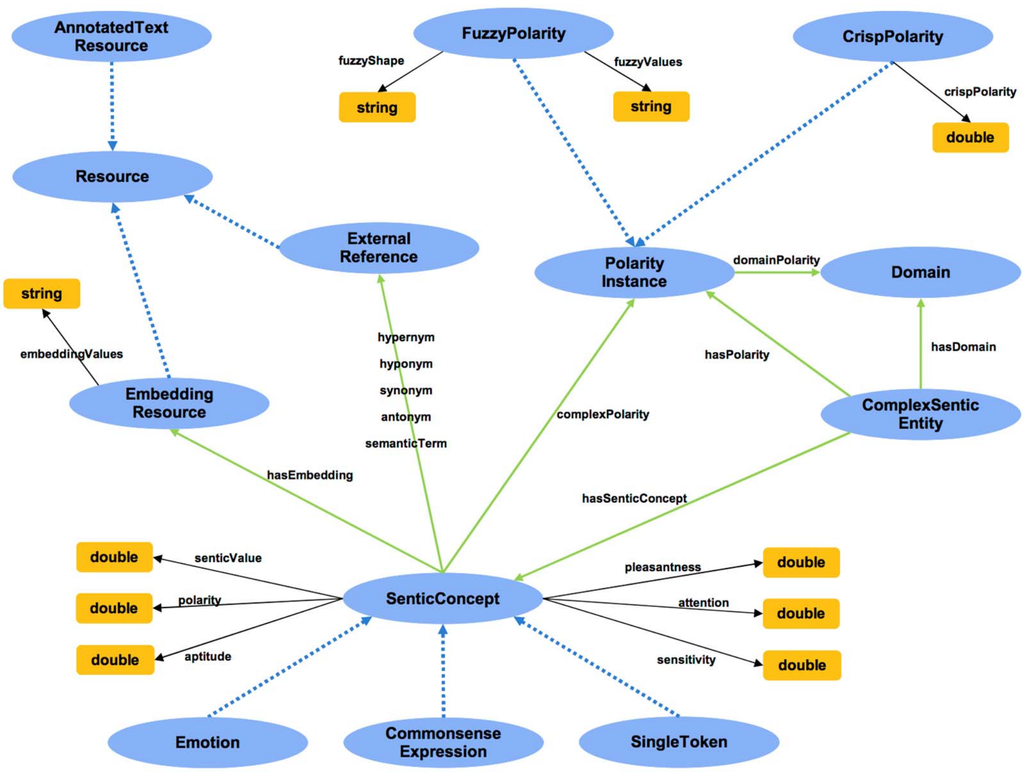

Figure 2. Overview of OntoSenticNet: each ellipse corresponds to one of the concepts composing OntoSenticNet; dashed arrows represent subsumption relationships; green arrows show the ObjectProperties; finally, black arrows describe the DatatypeProperties associated with each concept.

Table 1. List of the Object Properties included in OntoSenticNet.

Object Property	Domain	Range
complexPolarity	SenticConcept	PolarityInstance
domainPolarity	PolarityInstance	Domain
hasAnnotation	AnnotatedTextResource AnnotatedMultimodalResource ExternalReference	SenticConcept
hasDomain	ComplexSenticEntity	Domain
hasEmbedding	SenticConcept	EmbeddingResource
hasPolarity	ComplexSenticEntity	PolarityInstance
hasSenticConcept	ComplexSenticEntity	SenticConcept
semanticTerm (antonym, hypernym, hyponym, synonym)	SenticConcept	SenticConcept ExternalReference

Table 2. List of the Data Properties included in OntoSenticNet.

Data Property	Domain	Range
crispPolarity	CrispPolarity	decimal
embeddingSize	EmbeddingResource	double
embeddingValues	EmbeddingResource	string
fuzzyShape	FuzzyPolarity	string
fuzzyValues	FuzzyPolarity	string
senticValue (aptitude, attention, pleasantness, polarity, sensitivity)	SenticConcept	double

Capturing public opinions has raised increasing interest within both the scientific and business communities because of the remarkable benefits offered by marketing and financial prediction, which have led to many exciting open challenges. While there are many lexicons and knowledge bases available for sentiment analysis; however, there is a lack of sentiment ontologies.

In this paper, we proposed OntoSenticNet, a conceptual model supporting the structuring analysis of emotions from multimodal resources based on SenticNet, a commonsense knowledge base for sentiment analysis. We discussed the methodology implemented for creating the resource and the rationale behind the main classes and properties modeled into the ontology.

OntoSenticNet is freely available for download, and it can be easily integrated into business platforms and real-world applications. Future work will focus on the development of an ecosystem of services and data that will be directly integrated into OntoSenticNet in order to support the construction of smart emotion-based applications.

REFERENCES

1. E. Cambria et al., *A Practical Guide to Sentiment Analysis*, Springer, 2017.
2. S. Poria et al., "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, vol. 37, 2017, pp. 98–125.
3. A. Zadeh et al., "Tensor fusion network for multimodal sentiment analysis," *Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing* (EMNLP), 2017, pp. 1114–1125.
4. I. Chaturvedi et al., "Distinguishing between facts and opinions for sentiment analysis: Survey and challenges," *Information Fusion*, vol. 44, 2018, pp. 65–77.
5. S.L. Lo et al., "Multilingual sentiment analysis: From formal to informal and scarce resource languages," *Artificial Intelligence Review*, vol. 48, no. 4, 2017, pp. 499–527.
6. E. Cambria et al., "Sentiment analysis is a big suitcase," *IEEE Intelligent Systems*, vol. 32, no. 6, 2017, pp. 74–80.
7. Y. Ma et al., "Label embedding for zero-shot fine-grained named entity typing," *The 26th International Conference on Computational Linguistics: Technical Papers* (COLING), 2016, pp. 171–180.
8. N. Majumder et al., "Deep learning-based document modeling for personality detection from text," *IEEE Intelligent Systems*, vol. 32, no. 2, 2017, pp. 74–79.
9. S. Poria et al., "A deeper look into sarcastic tweets using deep convolutional neural networks," *The 26th International Conference on Computational Linguistics: Technical Papers* (COLING), 2016, pp. 1601–1612.
10. Y. Ma et al., "Targeted aspect-bsaed sentiment analysis via embedding commonsense knowledge into an attentive LSTM," *The 32nd AAAI Conference on Artificial Intelligence* (AAAI-18), 2018, pp. 5876–5883.
11. F. Xing et al., "Natural language based financial forecasting: A survey," *Artificial Intelligence Review*, vol. 50, no. 1, 2018, pp. 49–73.
12. M. Ebrahimi et al., "Challenges of sentiment analysis for dynamic events," *IEEE Intelligent Systems*, vol. 32, no. 5, 2017, pp. 70–75.
13. R. Mihalcea and A. Garimella, "What men say, what women hear: Finding gender-specific meaning shades," *IEEE Intelligent Systems*, vol. 31, no. 4, 2016, pp. 62–67.
14. S. Cavallari et al., "Learning community embedding with community detection and node embedding on graphs," *Proc. of the 2017 ACM Conference on Information and Knowledge Management* (CIKM), 2017, pp. 377–386.
15. M. Dragoni, "A Three-Phase Approach for Exploiting Opinion Mining in Computational Advertising," *IEEE Intelligent Systems*, vol. 32, no. 3, 2017, pp. 21–27.
16. T. Young et al., "Augmenting end-to-end dialog systems with commonsense knowledge," *The 32nd AAAI Conference on Artificial Intelligence* (AAAI-18), 2018, pp. 4970–4977.
17. E. Cambria et al., "SenticNet5: Discovering conceptual primitives for sentiment analysis by means of context embeddings," *The 32nd AAAI Conference on Artificial Intelligence* (AAAI-18), 2018, pp. 1795–1802.
18. M. Fernández-Lopez et al., "METHONTOLOGY: From Ontological Art Towards Ontological Engineering," *AAAI Spring Symposium Series*, 1997, pp. 33–40.
19. E. Cambria and A. Hussain, *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*, Springer, 2015.
20. M. Dragoni et al., "Propagating and Aggregating Fuzzy Polarities for Concept-Level Sentiment Analysis," *Cognitive Computation*, vol. 7, no. 2, 2015, pp. 186–197.

ABOUT THE AUTHORS

Mauro Dragoni is a researcher scientist at Fondazione Bruno Kessler. His research interests include knowledge management, information retrieval, and machine learning by focusing on the development of real-world prototypes. Dragoni received a PhD in computer science from the University of Milan.

Soujanya Poria is a research scientist at Temasek Laboratories, Nanyang Technological University. His research interests include computational linguistics, deep learning, and multimodal sentiment analysis. Poria holds a PhD in computer science and mathematics from the University of Stirling.

Erik Cambria is an assistant professor at Nanyang Technological University. He is a recipient of the AI's 10 to Watch award by IEEE Intelligent Systems and associate editor of several journals. His research interests include commonsense reasoning, multimodal interaction, and sentiment analysis.