



# DomainSenticNet: An Ontology and a Methodology Enabling Domain-Aware Sentic Computing

Damiano Distante<sup>1</sup> · Stefano Faralli<sup>1</sup> · Steve Rittinghaus<sup>2</sup> · Paolo Rosso<sup>3</sup> · Nima Samsami<sup>4</sup>

Received: 22 June 2020 / Accepted: 12 January 2021 / Published online: 4 February 2021  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

## Abstract

In recent years, SenticNet and OntoSenticNet have represented important developments in the novel interdisciplinary field of research known as sentic computing, enabling the development of a variety of Sentic applications. In this paper, we propose an extension of the OntoSenticNet ontology, named DomainSenticNet, and contribute an unsupervised methodology to support the development of domain-aware Sentic applications. We developed an unsupervised methodology that, for each concept in OntoSenticNet, mines semantically related concepts from WordNet and Probase knowledge bases and computes domain distributional information from the entire collection of Kickstarter domain-specific crowdfunding campaigns. Subsequently, we applied DomainSenticNet to a prototype tool for Kickstarter campaign authoring and success prediction, demonstrating an improvement in the interpretability of sentiment intensities. DomainSenticNet is an extension of the OntoSenticNet ontology that integrates each of the 100,000 concepts included in OntoSenticNet with a set of semantically related concepts and domain distributional information. The defined unsupervised methodology is highly replicable and can be easily adapted to build similar domain-aware resources from different domain corpora and external knowledge bases. Used in combination with OntoSenticNet, DomainSenticNet may favor the development of novel hybrid aspect-based sentiment analysis systems and support further research on sentic computing in domain-aware applications.

**Keywords** Sentic computing · SenticNet · OntoSenticNet · Kickstarter · Interpretability · Opinion mining · Marketing

## Introduction

In recent decades, the Internet has become the preferred communication channel for novel forms of everyday human activities. As recently highlighted by the unfortunate global situation caused by the COVID-19<sup>1</sup> pandemic, people are now able to perform new activities online to replace or complement traditional behaviors. Popular examples of the new forms of activity domains include e-learning, e-commerce, telehealth, telemedicine, social media, and e-government. Within this context, the majority of the above-mentioned sectors and fields of research benefit from the analyses of popular opinions and sentiments that are massively and extensively conveyed over the Internet via user-generated contents. To support this, researchers are investigating and developing methodologies for aspect-based sentiment analysis (ABSA). As reported by recent surveys [10, 12, 13], the literature on ABSA has identified

---

✉ Stefano Faralli  
stefano.faralli@unitelmasapienza.it

Damiano Distante  
damiano.distante@unitelmasapienza.it

Steve Rittinghaus  
steverittinghaus@gmail.com

Paolo Rosso  
proso@dsic.upv.es

Nima Samsami  
nima.samsami@outlook.com

<sup>1</sup> University of Rome Unitelma Sapienza, Rome, Italy

<sup>2</sup> Independent researcher, Freelancer Digital Transformation, Baden-Wurttemberg, Germany

<sup>3</sup> Universitat Politcnica de Valncia, Valencia, Spain

<sup>4</sup> Independent researcher, Software Architect, Baden-Wurttemberg, Germany

<sup>1</sup> Coronavirus disease 2019 (COVID-19) is a contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). [https://en.wikipedia.org/wiki/Coronavirus\\_disease\\_2019](https://en.wikipedia.org/wiki/Coronavirus_disease_2019)

many open challenges to be solved. The authors of [14] hold that state-of-the-art ABSA approaches can be broadly categorized into *symbolic* and *sub-symbolic* approaches. *Symbolic* approaches “consist of machine learning techniques that perform sentiment classification based on word co-occurrence frequencies.” *Sub-symbolic* approaches, on the other hand, “include the use of lexicons, ontologies, and semantic networks to encode the polarity associated with words and multiword expressions.” In both cases, ABSA “is a suitcase research problem” [10] that requires many natural language processing (NLP) challenges to be overcome.

In this paper, we introduce DomainSenticNet, an extension of the OntoSenticNet ontology [14] to aid the development of hybrid ABSA systems by leveraging the advantages of both *symbolic* and *sub-symbolic* approaches. DomainSenticNet is a resource written in the Web Ontology Language (OWL) standard that, for each of the 100,000 OntoSenticNet concepts, provides a set of semantically related concepts and domain distributional information. Specifically, to build DomainSenticNet, for each of the concepts in OntoSenticNet, we mined semantically related concepts from the knowledge bases WordNet [18] and Probase [34] and obtained domain distributional information by computing the distribution of occurrences and co-occurrences of the concept across domain-specific texts extracted from textual descriptions of the entire collection of Kickstarter<sup>2</sup> crowdfunding campaigns.

The present paper describes the unsupervised methodology we designed to build our resource, which can be replicated to generate similar resources from different domain corpora and external knowledge bases. Therefore, DomainSenticNet, used in combination with OntoSenticNet, can support future investigations of sentic computing [7] for domain-aware research and applications. Moreover, in this paper, we discuss the practical usage of our resource and present an example of a real application that provides a high level of interpretability of sentiment intensities expressed for domain aspects.

The remainder of the paper is organized as follows. Section 2 states our research objectives. Section 3 describes DomainSenticNet and the unsupervised methodology we designed to construct it from the external knowledge bases WordNet [18] and Probase [34], and the textual description of Kickstarter crowdfunding campaigns. Section 4 describes an example of a real application that, drawing on DomainSenticNet, demonstrates improved interpretability of aspect-based sentiment analysis outcomes. Section 5 summarizes the existing literature related to our work. Finally, Section 6 provides concluding remarks.

The DomainSenticNet project page is available at <https://github.com/needindex/domainsenticnet>. The related resources are publicly available under *Attribution 4.0 International (CC BY 4.0)*.<sup>3</sup>

## Research Objectives

OntoSenticNet [14] is a commonsense ontology for sentiment analysis based on SenticNet, a semantic network of 100,000 concepts. In this paper, our main research objective is to provide an extension (not a substitution) of OntoSenticNet to:

- RO1: provide a wider coverage of domain-specific concepts (not yet included in SenticNet) to support the development of novel hybrid (symbolic and sub-symbolic) domain-specific SenticNet-based ABSA systems;
- RO2: include, for each concept, effective and human-readable information on the domain pertinence; and
- RO3: use a standard knowledge representation language to ease the adoption and reuse of our OntoSenticNet extension.

Additionally, with respect to the methodology, we had one further research objective:

- RO4: to define a replicable (and generalized) methodology that could be adapted with minimal efforts to cover additional concepts and domains.

In Section 3, we describe the resource and the methodology.

## DomainSenticNet Resource and Methodology

In this section, we introduce DomainSenticNet and describe the unsupervised methodology we defined to create the resource.

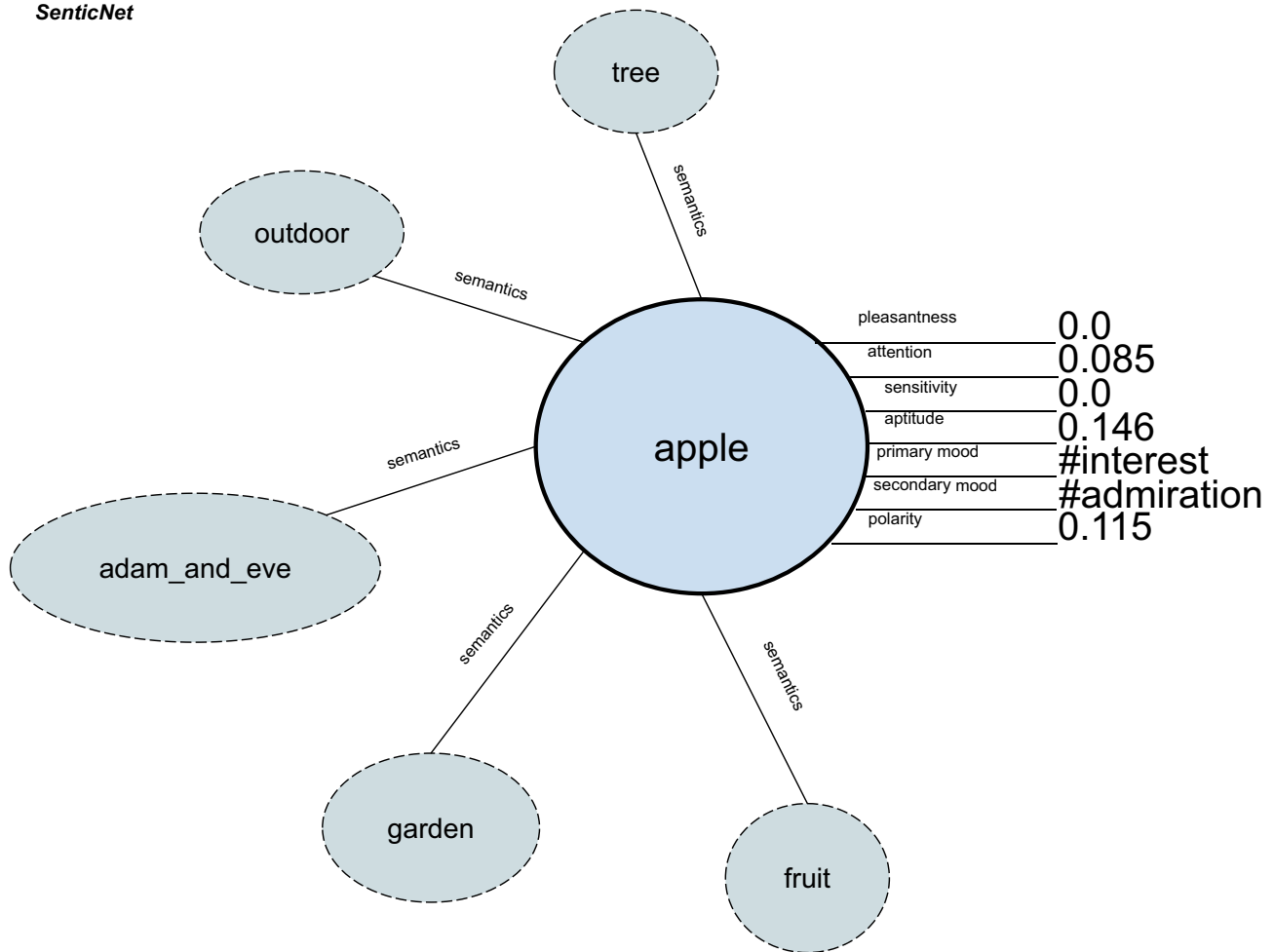
DomainSenticNet is a resource that extends OntoSenticNet with:

1. additional related concepts harvested from external knowledge bases;
2. distributional information, i.e., occurrences and co-occurrences of each SenticNet concept and related concepts, in domain-related texts.

<sup>2</sup> <https://www.kickstarter.com/>

<sup>3</sup> <https://creativecommons.org/licenses/by/4.0/deed.en>

## SenticNet



**Fig. 1** Representation of the SenticNet concept “apple”

To illustrate the characteristics of our resource, in Fig. 1, we visually represent the original SenticNet concept “apple” as a graph. In this graph, nodes represent SenticNet concepts and edges represent semantic relatedness between pairs of concepts. Figure 2 shows a visual representation of the corresponding “apple” concept in DomainSenticNet. In this figure, additional nodes represent the semantically related concepts mined from external knowledge bases and edges are complemented by domain distributional information about occurrences and co-occurrences in domain texts.

Figure 3 depicts the methodology workflow we designed and performed to generate the DomainSenticNet resource. The methodology included four main steps:

- Step 1: expansion (see Section 3.1);
- Step 2: mining of domain corpora (see Section 3.2);
- Step 3: domain weighting (see Section 3.3);
- Step 4: OWL translation (see Section 3.4).

In the following sections, we describe each of the four steps and, without loss of generality, make explicit reference to the external knowledge bases and corpora used to generate DomainSenticNet.

### Expansion

To address our first research objective (see Section 2, RO1), in the first step of our workflow, for each concept  $\in$  SenticNet, we searched for semantically related concepts in the external knowledge bases WordNet [18] and Probase [36]. In both knowledge bases, we first identified all concepts corresponding to those in SenticNet. Then, to collect all neighborhood concepts, for each identified concept, we performed a 1-hop visit on the corresponding knowledge graphs, following the hypernymy (“is a”) and synonymy relationships. Figure 4 shows an excerpt of the semantically related concepts we found for the “apple” SenticNet concept. For this concept, we first identified the concepts “apple#1” and “apple#2” in WordNet and “apple” in Probase. Subsequently,

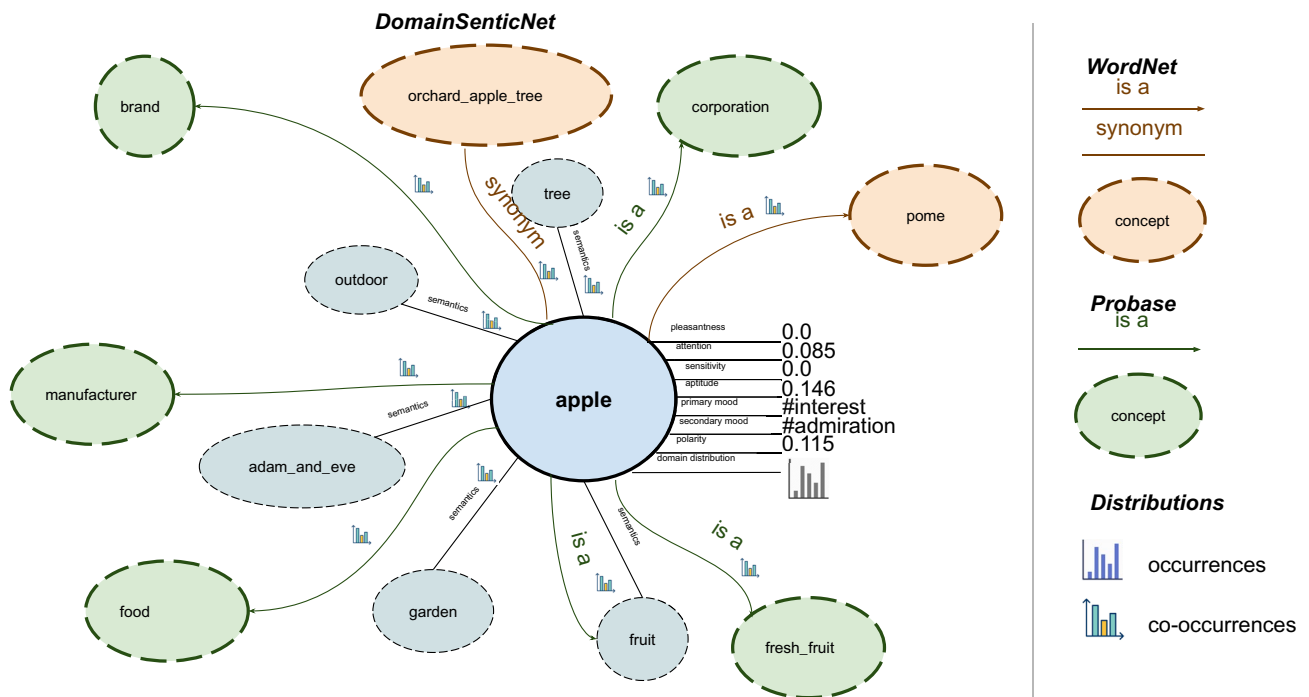


Fig. 2 Excerpt of the DomainSenticNet concept “apple”

we collected two synonyms (i.e., “malus pumila” and “orchard apple tree”) and four hypernyms (i.e., “apple tree,” “edible fruit,” “false fruit,” and “pome”) from WordNet, and ~4.6K hypernyms (e.g., “brand,” “corporation,” “company,” “crop,” “firm,” “food,” “fresh fruit,” “fruit,” “fruit tree,” “manufacturer,” etc.) from Probase.

### Mining of Domain Corpora

Distributional information was at the base of our second research objective (see Section 2, RO2). To tackle this objective, we applied standard text mining techniques on domain-specific corpora, to compute: i) the number of occurrences of concepts belonging to SenticNet and ii) the number of co-occurrences of each concept in SenticNet and the semantically related external concepts we previously harvested in Step 1 (see Section 3.1). As a medium-sized collection of domain-specific texts, Kickstarter was chosen as a data source.<sup>4</sup>

Kickstarter, a popular source for data scientists, includes approximately 480K campaign descriptions<sup>5</sup> in the form of hypertexts, including text, images, videos, and hyperlinks.<sup>6</sup> To identify the domains of interest of each campaign, we leveraged the labels available on the Kickstarter platform to categorize each campaign description. In Table 1, we present an excerpt of the 15 main domain categories of Kickstarter, with related subcategories.<sup>7</sup> The number of occurrences and co-occurrences was computed in four substeps:

- Step 2.1: Starting from the campaign uniform resource locators (URLs), we retrieved campaign textual descriptions by means of a custom-made crawler;
- Step 2.2: For each word  $w$  corresponding to one of the concepts generated in Step 1 (see Section 3.1) and for each textual campaign description  $t$ , we computed the number of occurrences  $occ(w, t)$  of word  $w$  in  $t$ ;
- Step 2.3: For each campaign description  $t$  and for each pair of words  $\{w_1, w_2\}$  s.t.  $occ(w_1, t) > 0$  and  $occ(w_2, t) > 0$ , we computed the number of co-occurrences  $co\_occ(w_1, w_2, t)$  of words  $w_1$  and  $w_2$  in the description  $t$  as  $co\_occ(w_1, w_2, t) = occ(w_1, t) * occ(w_2, t)$ ;
- Step 2.4: Since Kickstarter campaigns are labeled with two domain categories (i.e., a main category and an optional subcategory), we leveraged this labeling to compute the distributions of occurrences and co-occurrences of concepts across domains.

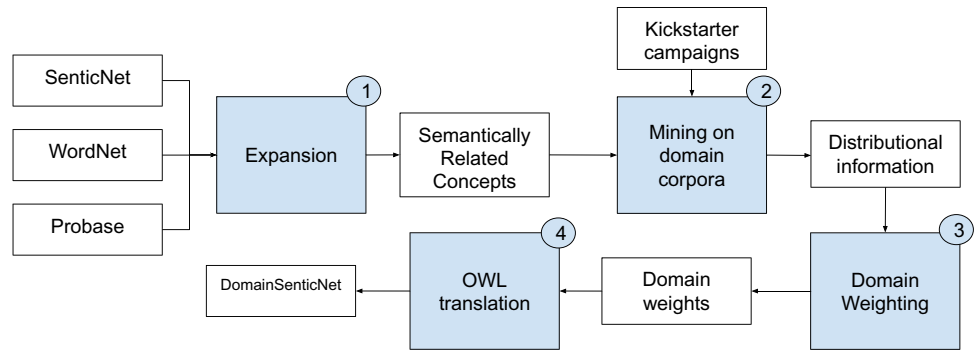
<sup>4</sup> Monthly updated dataset of the Kickstarter campaign URLs is available at: <https://webrobots.io/kickstarter-datasets/>

<sup>5</sup> Real-time statistics are accessible at: <https://www.kickstarter.com/help/stats>

<sup>6</sup> We were able to crawl a total of ~230K Kickstarter descriptions from the original ~480K campaigns.

<sup>7</sup> An overview of the respective domains and related statistics is available at: <https://www.kickstarter.com/help/stats>

**Fig. 3** DomainSenticNet construction workflow



Returning to the “apple” concept example, Fig. 5 depicts the distribution of occurrences of the word “apple” over each resulting domain corpus; Fig. 6 presents the co-occurrences distribution for the pair of words “apple” and “brand.”

**Domain Weighting**

Since most distributional methodologies perform better using normalized weights, to complete our second research objective (see Section 2, RO2), we defined a proper transformation to obtain correct domain distributional information in the third step of our workflow. To this end, we defined a domain relevance function that assigned each SenticNet concept  $w$  a domain relevance with respect to a corpus  $C_d$ . The function is defined as follows:

$$domainOccScore(w, C_d) = \frac{\sum_{t \in C_d} occ(w, t)}{|C_d|} \tag{1}$$

where  $C_d$  includes all textual descriptions of the Kickstarter campaigns labeled with a specific domain category  $d$ .

Additionally, in order to represent the domain relevance of a pair of related concepts  $\{w_1, w_2\}$  we defined:

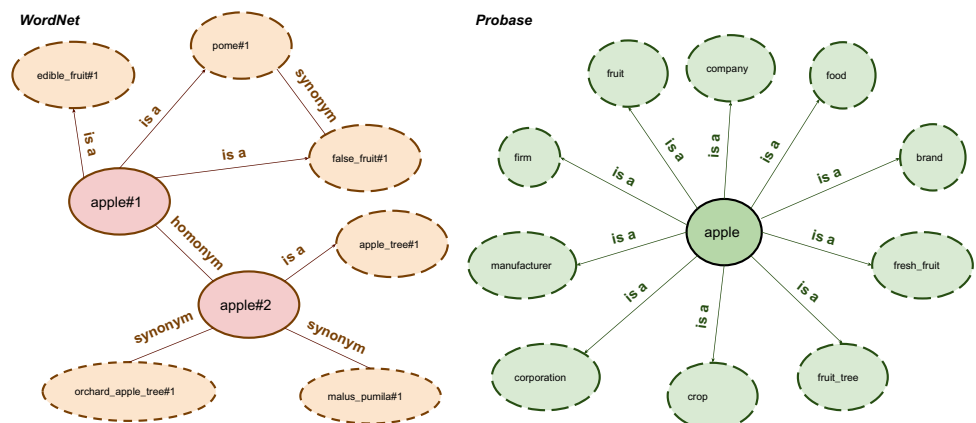
$$domainCooccScore(w_1, w_2, C_d) = \frac{\sum_{t \in C_d} co\_occ(w_1, w_2, t)}{|C_d|} \tag{2}$$

Continuing the “apple” concept example, Fig. 7 shows the domain distribution of the  $domainOccScore$  for the concept “apple,” and Fig. 8 presents the domain distribution of  $domainCooccScore$  for the two semantically related concepts “apple” and “brand.” Finally, in Table 2, we provide the top 40 most co-occurring concepts with “apple” across domains.

**OWL Translation**

To address the third research objective (see Section 2, RO3), in the fourth step of our workflow (see Fig. 3, block 4), we translated all collected domain distributional information into an OWL representation. As shown in the ontology schema depicted in Fig. 9, DomainSenticNet refers to the original definition of SenticConcept, thus enabling reference to all original OntoSenticNet facts.

**Fig. 4** Excerpt example of the semantically related concepts and relations considered during the expansion step (step 1) for the SenticNet concept “apple”



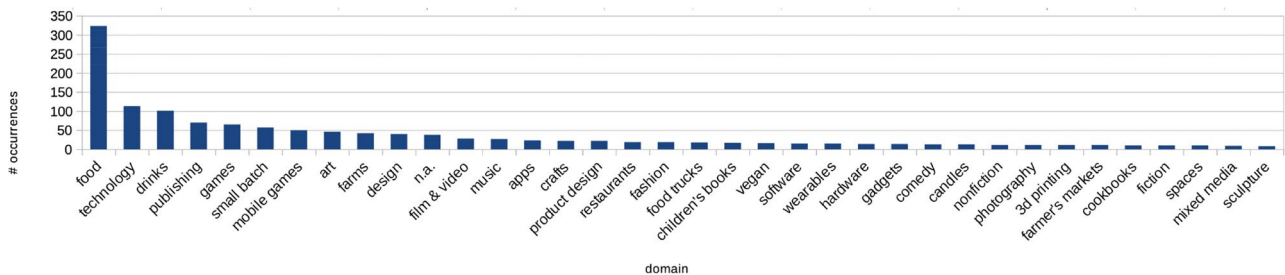
**Table 1** Excerpt of the Kickstarter campaign domains of interest (categories) and subdomains (subcategories) (February 2020)

Main category	Subcategories
Art	Ceramics, Conceptual Art, Digital Art, Illustration, Installations, Mixed Media, Painting, Performance Art, Public Art, Sculpture, Textiles, Video Art
Comics	Anthologies, Comic Books, Events, Graphic Novels, Webcomics
Crafts	Candles, Crochet, DIY, Embroidery, Glass, Knitting, Pottery, Printing, Quilts, Stationery, Taxidermy, Weaving, Woodworking
Dance	All Dance Projects, Performances, Residencies, Spaces, Workshops
Design	Architecture, Civic Design, Graphic Design, Interactive Design, Product Design, Typography
Fashion	Accessories, Apparel, Childrenswear, Couture, Footwear, Jewelry, Pet Fashion, Ready-to-wear
Film and Video	Projects, Action, Animation, Comedy, Documentary, Drama, Experimental, Family, Fantasy, Festivals, Horror, Movie Theaters, Music Videos, Narrative Film, Romance, Science Fiction, Shorts, Television, Thrillers, Web series
Food	Bacon, Community Gardens, Cookbooks, Drinks, Events, Farmer’s Markets, Farms, Food Trucks, Restaurants, Small Batch, Spaces, Vegan
Games	Gaming Hardware, Live Games, Mobile Games, Playing Cards, Puzzles, Tabletop Games, Video Games
Journalism	Audio, Photo, Print, Video, Web
Music	Blues, Chiptune, Classical Music, Comedy, Country & Folk, Electronic Music, Faith, Hip-Hop, Indie Rock, Jazz, Kids, Latin, Metal, Pop, Punk, R&B, Rock, World Music
Photography	Animals, Fine Art, Nature, People, Photobooks, Places
Publishing	Academic, Anthologies, Art Books, Calendars, Children’s Books, Comedy, Fiction, Letterpress, Literary Journals, Literary Spaces, Nonfiction, Periodicals, Poetry, Radio \$ Podcasts, Translations, Young Adult, Zines
Technology	3D Printing, Apps, Camera Equipment, DIY Electronics, Fabrication Tools, Flight, Gadgets, Hardware, Makerspaces, Robots, Software, Sound, Space Exploration, Wearables, Web
Theater	Comedy, Experimental, Festivals, Immersive, Musical, Plays, Spaces

As an example, in OntoSenticNet [14], the concept “apple” is defined as follows:

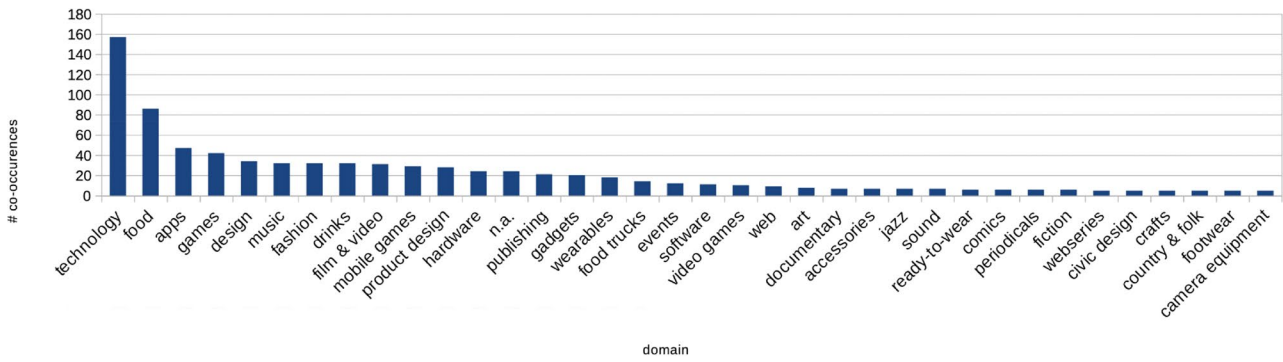
```

<owl:NamedIndividual rdf:about="urn:absolute:ontosenticnet#apple">
  <rdf:type rdf:resource="urn:absolute:ontosenticnet#SenticConcept"/>
  <aptitude rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">0.146</aptitude>
  <attention rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">0.085</attention>
  <pleasantness rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">0</pleasantness>
  <polarity rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">0.077</polarity>
  <polarityText>positive</polarityText>
  <primitiveURI rdf:resource="urn:absolute:ontosenticnet#admiration"/>
  <primitiveURI rdf:resource="urn:absolute:ontosenticnet#interest"/>
  <semantics rdf:resource="urn:absolute:ontosenticnet#adam_and_eve"/>
  <semantics rdf:resource="urn:absolute:ontosenticnet#fruit"/>
  <semantics rdf:resource="urn:absolute:ontosenticnet#garden"/>
  <semantics rdf:resource="urn:absolute:ontosenticnet#outdoor"/>
  <semantics rdf:resource="urn:absolute:ontosenticnet#tree"/>
  <sensitivity rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">0</sensitivity>
  <text>apple</text>
</owl:NamedIndividual>
    
```



**Fig. 5** Occurrence distribution (top 36 domains) of the word “apple” in domain corpora extracted from Kickstarter campaigns





**Fig. 6** Co-occurrence distribution (top 36 domains) of the words “apple” and “brand” in the domain corpora extracted from the Kickstarter campaigns

where: i) *aptitude*, *attention*, *pleasantness*, and *sensitivity* are defined as SenticValues for the corresponding *Hourglass of Emotions* model dimensions; ii) *polarity* is the overall sentiment polarity; iii) *semantics* are properties representing five semantically related concepts (e.g., *adam\_and\_eve*, *fruit*, *garden*, *outdoor*, and *tree*); and iv) *primitiveURI* refers to two primitive moods (e.g., *admiration* and *interest*).

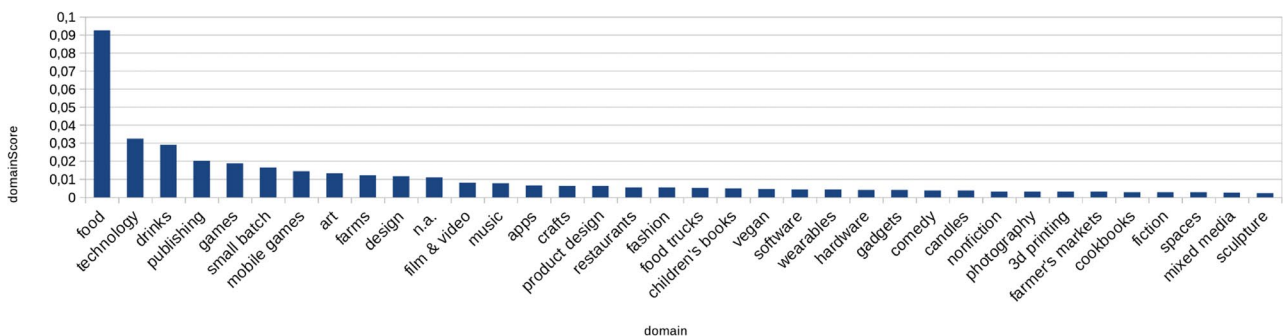
To represent all of the concepts mined from the external knowledge bases in the first step (see Fig. 3, block 1), we defined the “ExternalConcept” class as follows:

```
<owl:Class rdf:about="urn:absolute:domainsenticnet#ExternalConcept"/>
```

The above class enables the model to reference concepts such as the “malus pumila,” in which WordNet presents as a synonym of the SenticNet concept “apple.” Instances of the “ExternalConcept” class have two annotation properties, namely *provenance* and *text*, which represent the source knowledge base and the lexeme, respectively:

```
<owl:AnnotationProperty rdf:about="urn:absolute:ontosenticnet#provenance"/>
<owl:AnnotationProperty rdf:about="urn:absolute:ontosenticnet#text"/>
```

As an example, the external concept “malus pumila” is defined as follows:



**Fig. 7** The  $domainOccScore(w, C_d)$  distribution ( $d \in$  set of the top 36 domains) for the DomainSenticNet concept  $w =$ “apple”

**Table 2** Top 40 most co-occurring concepts across domains (DCS = *domainCooccScore*)

Rank	Concept	Domain	DCS	Rank	Concept	Domain	DCS
1	product	technology	0.46	21	product	hardware	0.11
2	work	technology	0.34	22	choice	technology	0.10
3	device	technology	0.34	23	device	hardware	0.09
4	food	food	0.28	24	being	games	0.09
5	being	technology	0.19	25	work	mobile games	0.09
6	system	technology	0.17	26	name	games	0.09
7	work	games	0.15	27	product	food	0.09
8	case	technology	0.15	28	store	technology	0.09
9	name	technology	0.14	29	work	publishing	0.09
10	good	technology	0.14	30	device	gadgets	0.08
11	platform	technology	0.13	31	service	technology	0.08
12	website	technology	0.12	32	work	food	0.08
13	player	games	0.12	33	product	wearables	0.08
14	product	gadgets	0.12	34	good	food	0.08
15	idea	technology	0.12	35	player	mobile games	0.08
16	business	food	0.12	36	work	hardware	0.08
17	business	technology	0.11	37	fruit	food	0.07
18	company	technology	0.11	38	food	food trucks	0.07
19	computer	technology	0.11	39	name	food	0.07
20	work	film and video	0.11	40	product	design	0.07

```

<owl:NamedIndividual rdf:about="urn:absolute:domainsenticnet#malus_pumila">
  <rdf:type rdf:resource="urn:absolute:domainsenticnet#ExternalConcept"/>
  <provenance>WordNet</provenance>
  <text>malus pumila</text>
  <semanticallyRelatedTo rdf:resource="urn:absolute:ontosenticnet#apple"/>
  ...
</owl:NamedIndividual>

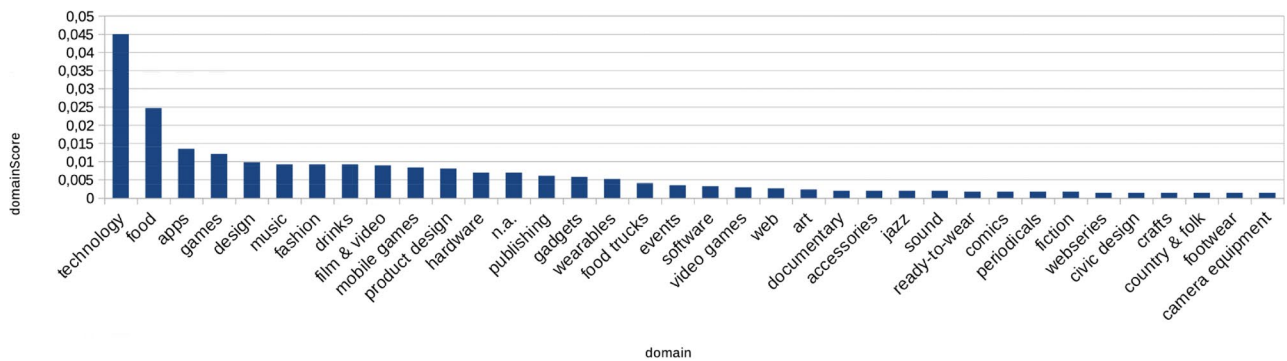
```

where *semanticallyRelatedTo* is an *ObjectProperty* defined as follows:

```

<owl:ObjectProperty rdf:ID="semanticallyRelatedTo">
  <rdfs:domain rdf:resource="#ExternalConcept"/>
  <rdfs:range rdf:resource="#SenticConcept"/>
</owl:ObjectProperty>

```



**Fig. 8** The *domainCooccScore*( $w_1, w_2, C_d$ ) distribution ( $d \in$  set of the top 36 domains) for the DomainSenticNet concepts  $w_1$  =“apple” and  $w_2$  = “brand”



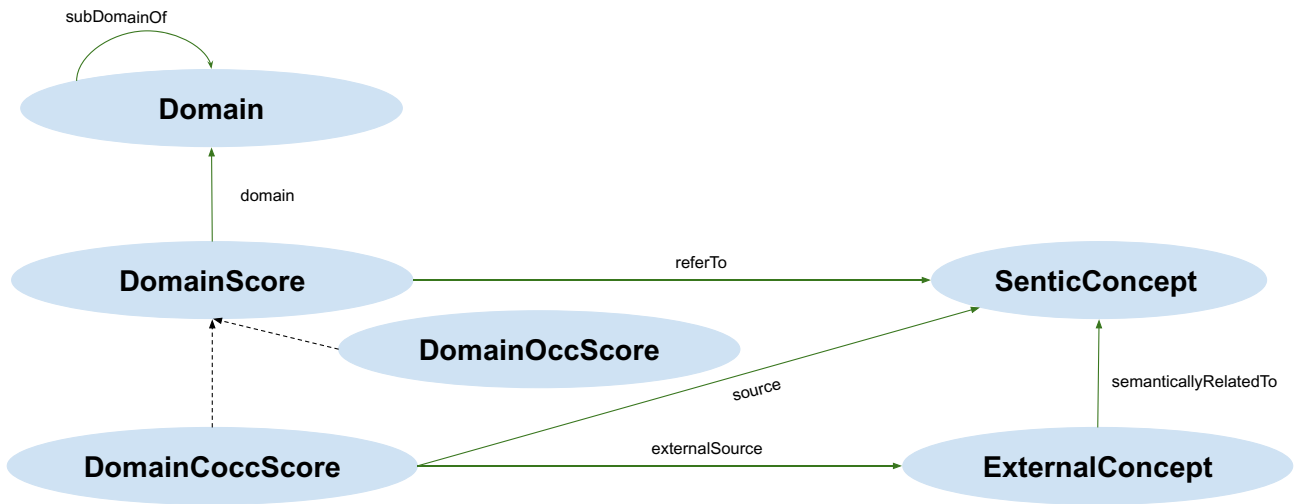


Fig. 9 Overview of the DomainSenticNet scheme

To represent each of the 176 considered domains, we defined the following *Domain* class:

```
<owl:Class rdf:about="urn:absolute:domainsenticnet#Domain"/>
```

The 15 main categories and 161 subcategories were then defined as subclasses of the *Domain* class.

```
<owl:ObjectProperty rdf:about="urn:absolute:domainsenticnet#subDomainOf">
  <rdfs:domain rdf:resource="urn:absolute:domainsenticnet#Domain"/>
  <rdfs:range rdf:resource="urn:absolute:domainsenticnet#Domain"/>
</owl:ObjectProperty>
```

As an example, the resulting definition for the domain “Ceramics” includes the annotation property *subDomainOf*, representing the fact that “Ceramics” is a subdomain of “Art.”

```
<owl:NamedIndividual rdf:about="urn:absolute:domainsenticnet#domain_ceramics">
  <rdf:type rdf:resource="urn:absolute:domainsenticnet#Domain"/>
  <subDomainOf rdf:resource="urn:absolute:domainsenticnet#domain_art"/>
  <text>Ceramics</text>
</owl:NamedIndividual>
```

To represent the domain weights described in Section 3.3, we provided the definitions for the *DomainScore*, *DomainOccScore*, and *DomainCooccScore* classes, as follows:

```

<owl:Class rdf:about="urn:absolute:domainsenticnet#DomainScore"/>

<owl:Class rdf:about="urn:absolute:domainsenticnet#DomainOccScore">
  <rdfs:subClassOf rdf:resource="urn:absolute:domainsenticnet#DomainScore"/>
</owl:Class>

<owl:Class rdf:about="urn:absolute:domainsenticnet#DomainCooccScore">
  <rdfs:subClassOf rdf:resource="urn:absolute:domainsenticnet#DomainScore"/>
</owl:Class>

```

The datatype property *score* represents a numeric weight:

```

<owl:DatatypeProperty rdf:about="urn:absolute:domainsenticnet#score">
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#double"/>
</owl:DatatypeProperty>

```

The following object property *domain* represents the domain related to a score:

```

<owl:ObjectProperty rdf:about="urn:absolute:domainsenticnet#domain">
  <rdfs:domain rdf:resource="urn:absolute:domainsenticnet#DomainScore"/>
  <rdfs:range rdf:resource="urn:absolute:domainsenticnet#Domain"/>
</owl:ObjectProperty>

```

Finally, the object properties *referTo*, *source*, and *externalSource* bind a *DomainScore* to one or more *SenticConcepts* or *ExternalConcepts*:

```

<owl:ObjectProperty rdf:about="urn:absolute:domainsenticnet#referTo">
  <rdfs:domain rdf:resource="urn:absolute:domainsenticnet#DomainScore"/>
  <rdfs:range rdf:resource="urn:absolute:ontosenticnet#SenticConcept"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="urn:absolute:domainsenticnet#source">
  <rdfs:domain rdf:resource="urn:absolute:domainsenticnet#DomainCooccScore"/>
  <rdfs:range rdf:resource="urn:absolute:ontosenticnet#SenticConcept"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="urn:absolute:domainsenticnet#externalSource">
  <rdfs:domain rdf:resource="urn:absolute:domainsenticnet#DomainCooccScore"/>
  <rdfs:range rdf:resource="urn:absolute:domainsenticnet#ExternalConcept"/>
</owl:ObjectProperty>

```

As an example, the *domainOccScore*("apple,"  $D_{food}$ ), defined in Section 3.3, is represented as follows:

```

<owl:NamedIndividual rdf:about="urn:absolute:domainsenticnet#dos_apple_51">
  <rdf:type rdf:resource="urn:absolute:domainsenticnet#DomainOccScore"/>
  <referTo rdf:resource="urn:absolute:ontosenticnet#apple"/>
  <domain rdf:resource="urn:absolute:domainsenticnet#domain_food"/>
  <score>0.0924971363115693</score>
</owl:NamedIndividual>

```

Additionally, the *domainCooccScore*("apple," "company,"  $D_{technology}$ ), defined in Section 3.3, is represented as follows:

```

<owl:NamedIndividual rdf:about="urn:absolute:domainsenticnet#dcs_apple_2047">
  <rdf:type rdf:resource="urn:absolute:domainsenticnet#DomainCooccScore"/>
  <referTo rdf:resource="urn:absolute:ontosenticnet#apple"/>
  <externalSource rdf:resource="urn:absolute:domainsenticnet#company"/>
  <domain rdf:resource="urn:absolute:domainsenticnet#domain_technology"/>
  <score>0.1145475372279496</score>
</owl:NamedIndividual>

```

## Results

DomainSenticNet was the result of our investigations aimed at achieving research objectives RO1, RO2, and RO3 (see Section 2).

The proposed approach was the result of RO4 (see Section 2), which primarily defined a generalized methodology that could be easily adapted to cover additional concepts and domains. In fact, the methodology can generate similar resources by simply using different domain corpora and external knowledge bases as input (see Fig. 3). Moreover, the methodology can be used to provide both domain distributional information and OWL representations for semantic networks other than OntoSenticNet, such as *DBpedia* and *WebIsADB* [17].

DomainSenticNet can be enhanced as a dynamic resource<sup>8</sup> in two ways:

1. by integrating significant variations in the concept collections and domain distribution of occurrences and co-occurrences linked to future releases of the domain corpora and external knowledge bases; and
2. by including timestamps (e.g., campaign start times) of the domain corpora (e.g., dumps of Kickstarter campaign URLs<sup>9</sup>) or other references to specific time in a temporal dimension in domain distributional information.

<sup>8</sup> Real-time data are widely recognized as the life blood of a variety of applications (e.g., [10])

<sup>9</sup> <https://webrobots.io/kickstarter-datasets/>

<sup>10</sup> <https://github.com/needindex/domainsenticnet>

<sup>11</sup> <https://github.com/needindex/gameon>

<sup>12</sup> It is worth noting that the tool can also process the human-crafted partitions of the domain aspects.

To address the above-mentioned dynamicity, we created a project Web page<sup>10</sup> and established a maintenance schedule for the generation of time-based update releases.

## Domain-Aware Kickstarter Campaign Success Prediction with DomainSenticNet

In this section, we present an example application of DomainSenticNet.

*GameOn* [16] is a prototype application designed to support the authoring of successful crowdfunding campaigns in Kickstarter.

The main characteristics of *GameOn*<sup>11</sup> are:

- It automatically induces (by means of clustering) a partition of semantically related domain aspects mined from user-generated product and service reviews, with each cluster representing an “influencing factor” for the campaign success;<sup>12</sup>
- It employs SenticNet to perform an ABSA and to identify emotional intensities expressed in textual campaign descriptions for the above-mentioned domain aspects;
- It aggregates the above-mentioned emotional intensities into a statistical index (NeedIndex), which: i) identifies the most influencing factors of the campaign success and ii) calibrates an objective and key result (OKR)<sup>13</sup> scale to interpret NeedIndexes, through the identification of the low and high emotional intensity bounds, delimiting *low*, *medium*, and *high* emotional intensity states, respectively;

<sup>13</sup> OKR models are commonly used by very successful companies such as Amazon, Facebook, and Google. <https://www.whatmatters.com/faqs/how-to-grade-okrs> <https://conceptboard.com/blog/okr-google-goal-setting-success/>

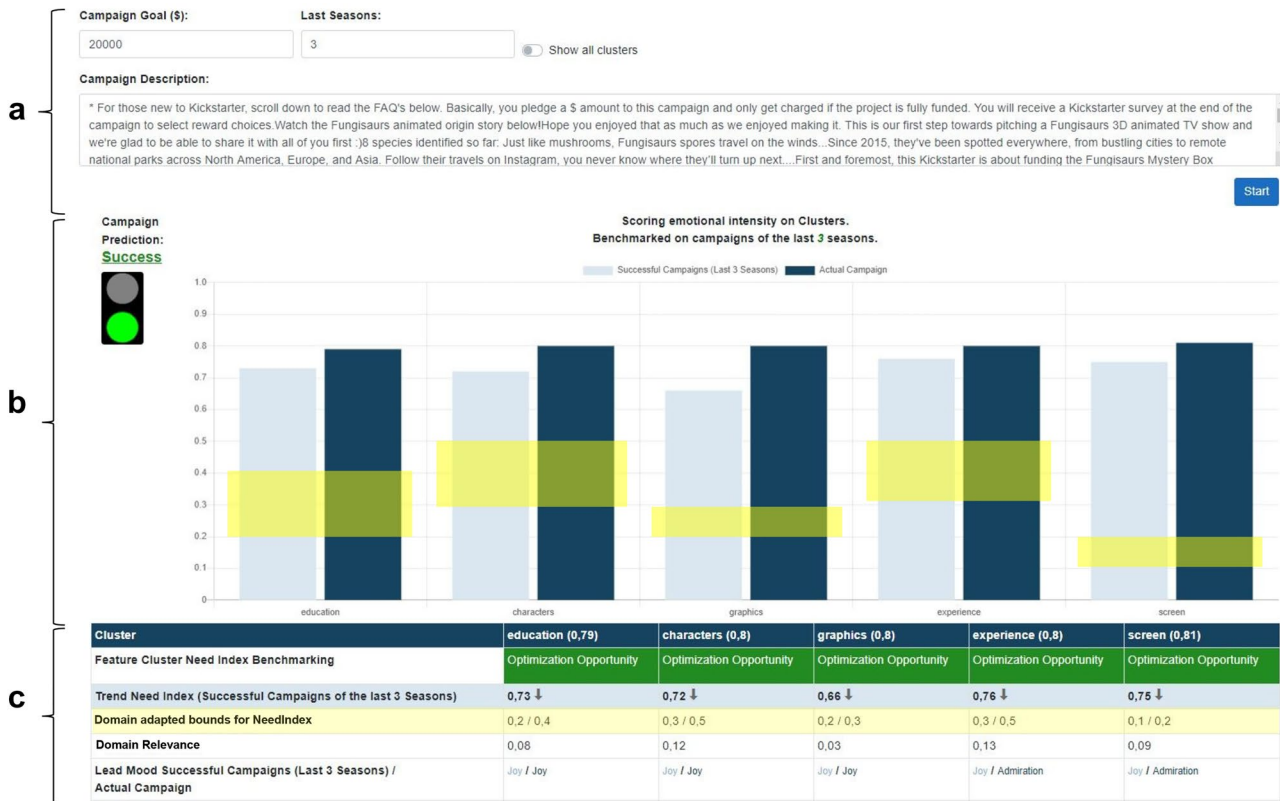


Fig. 10 A screenshot of the *GameOn* user interface

- It leverages DomainSenticNet to further tune (for a given domain of interest) the OKR scale for the interpretations of the emotional intensities.

Finally, the application compares the computed NeedIndexes with the average of the corresponding indexes of the successful “mobile games” campaigns during the past 3 seasons (see Fig. 10, parts B and C). Therefore, in this application, Need-Indexes are used both to train the model for campaign success forecasting and to provide highly interpretable explanations of the prediction outcomes. NeedIndexes are thus effective indicators used by the application to suggest actions to be performed on the textual descriptions to refine the emotional intensities expressed with respect to influencing factors (i.e., clusters).

Using DomainSenticNet, the application can also provide a domain adaption (at a cluster level) of the NeedIndex OKR scales of interpretation, whereby the resulting states of emotional intensities are calibrated with respect to the *domainOccScore* (defined in Section 3) for the “mobile games” domain.

To convey the previously mentioned calibration of the OKR scales, Fig. 11 presents two OKR scales related to the interpretation of the emotional intensities. The top part of the figure shows the original OKR scale (not adapted to the domain of interest), wherein two threshold values (i.e.,

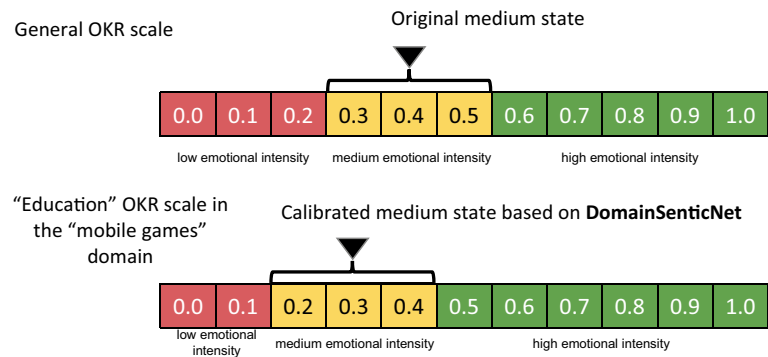
0.3 and 0.5) represent the lower and upper bounds used to identify the range of the NeedIndexes values corresponding to a *medium* emotional intensity level. In contrast, the bottom part of the figure depicts the domain-adapted scale with the corresponding bounds for the cluster labeled “education,” wherein the adapted medium level is bounded by the thresholds 0.22 and 0.43. For each cluster, the relevant bounds are obtained by computing the average *domainOccScores* of the concepts (in the cluster) occurring in the unsuccessful and successful campaign descriptions, respectively.

Figure 10 part C shows both the “domain adapted Need-Index bounds” and “domain relevance.” The domain-adapted emotional intensity states reflect both the average emotional intensity and the domain relevance for successful and unsuccessful campaigns, respectively.

In the “education” cluster<sup>14</sup>, the *medium* emotional intensity state produced lower values for two main reasons: i) in the considered Kickstarter dataset, the emotional intensities provided for the corresponding influential factors in

<sup>14</sup> The education cluster groups the following aspects: “education,” “student,” “school,” “college,” “instruction,” “classroom,” “brain,” “growth,” “level,” “course,” “knowledge,” “career,” “tutorial,” “education,” “lecture,” “tutor,” “teacher,” “learning,” “teaching,” and “skill.”

**Fig. 11** The original (top) and domain-adapted (bottom) versions of the OKR scale for the influencing factor “education” in the “mobile games” domain



the “mobile games” domain were lower than the average observed over the previous three seasons with respect to other aspects; and ii) the average *domainOccScore* of the corresponding aspects indicated a lower domain pertinence.

## Related Works

In this paper, we have presented DomainSenticNet as a resource to extend OntoSenticNet, a state-of-the-art commonsense ontology [14].

OntoSenticNet is an ontological representation of SenticNet [11], which is a resource resulting from the combined application of symbolic and sub-symbolic artificial intelligence methodologies to automatically discover conceptual primitives from text and link them to commonsense concepts and named entities. SenticNet includes the definition of 100K concepts (called SenticConcept).<sup>15</sup> Each SenticConcept (see Fig. 1 for a visual representation of the concept “apple”) is defined by: i) a multiword expression; ii) the weights for the four dimensions of the *Hourglass of Emotions* model [29] (i.e., *pleasantness*, *attention*, *sensitivity*, and *aptitude*); iii) primary and secondary mood labels (e.g., “#interest,” “#admiration”); iv) a polarity score; and v) a collection of five semantically related SenticConcepts.

OntoSenticNet is an ontological definition of the semantic network induced by the 100K SenticConcepts. Its main characteristic is its ability to provide a precise conceptual hierarchy, including associated concepts and sentiment values. Hence, OntoSenticNet is a preferential resource for developing state-of-the-art applications of sentiment analysis based on SenticNet.

In recent years, SenticNet and OntoSenticNet have represented important research developments. In particular, the findings from Cambria’s research group have enabled a novel interdisciplinary field of research known as sentic computing [7]. Within sentic computing, many successful investigations have generated novel insights in the domains of knowledge

representation [2], deep learning-based ABSA [24], business intelligence [19], social media marketing [6], recommender systems [3], and financial forecasting [38], to name only a few.

In the remainder of this section, we summarize the relevant literature pertaining to the key aspects of the definition and construction of the DomainSenticNet resource.

In constructing the proposed resource, with the aim of collecting neighborhood semantically related concepts from external knowledge graphs, we applied basic graph mining techniques (as described in Section 3.1). In general, the task of collecting semantically related concepts from affordable or noisy automatically acquired external knowledge graphs can be performed by sophisticated approaches (e.g., [26]). As an example (see [27] for a recent survey), the authors of [30] experimented with similarity expansion-based techniques and obtained high levels of efficiency and precision with regard to the task of extending new concepts in a given knowledge base.

As already mentioned, the backbone of DomainSenticNet is the OntoSenticNet ontological description of SenticNet. One of the key characteristics of SenticNet is that all concepts are defined with valued attributes derived from the *Hourglass of Emotions* model [9].<sup>16</sup> Therefore, SenticNet is considered an appropriate knowledge base for the development of human interpretable sentiment analysis approaches.

The availability of the above-mentioned resources is beneficial for all ontology-driven sentiment analysis (ODSA)-based applications. Specifically, the authors of [4] recently surveyed studies applying ODSA to customer reviews. Furthermore, as an example of an ODSA-based approach, the authors of [25] presented a hybrid solution for sentence-level ABSA using a lexicalized domain ontology in combination with neural attention networks.

Researchers in this field are also exploring the creation of new resources to be leveraged in ODSA-based applications. As an example, in [23], the authors presented a methodology to extend ontologies in the “Materials Science” domain. The presented approach leveraged the titles and abstracts of 600 domain publications and complemented a given ontology with additional concepts and axioms by means of a phrase-based

<sup>15</sup> SenticNet 6 has recently been released. This updated resource now contains 200K concepts [8]

<sup>16</sup> A recent model revision is described in [35]

topic model approach. In a similar direction, the authors of [39] proposed the addition of SOBA—a semiautomated methodology to generate ontologies—to ODSA applications.

In contrast to the works mentioned above, our methodology (see Section 3) is unsupervised and can be easily adapted to include other external knowledge bases and multiple domain corpora. In this way, our approach automatically generates a high coverage of domain-relevant concepts (not included in OntoSenticNet) and related distributional information for an arbitrarily defined set of domains of interest. Additionally, in the present paper, we discuss a real application that benefited from the availability of DomainSenticNet, in terms of both sentiment analysis performance and ease of interpretation (see Section 4).

As discussed in the Introduction (see Section 1), DomainSenticNet is suitable for use in domain-aware sentiment analysis applications. Such applications have recently been improved due to advancements in semisupervised learning [15] and, more specifically, in semisupervised learning for social data analysis [5, 20]. Researchers are experimenting with semisupervised learning as a potentially more robust solution to problems such as word polarity disambiguation [37] and the extraction of actionable information from unstructured text [21]. As an example, in [22], the authors presented a deep learning approach named *ConvNet – SVM<sub>BoVW</sub>* for fine-grained sentiment analysis. The model combined textual and visual features built on a convolution neural network (ConvNet) enhanced with the contextual scoring mechanism of SentiCircle [31]. The proposed model performed sentiment polarity classification with 91% accuracy. Moreover, in [1], the authors recently provided a stacked ensemble-based methodology to assess the emotional intensities in texts related to a general domain and performed sentiment analysis in the financial domain. With respect to the two above-mentioned studies, and in line with the findings of [33], the distributional information of DomainSenticNet may be coupled with contextual semantic features to address the problem of word polarity disambiguation. Finally, our resource may also be leveraged to improve the interpretability and explainability of sentiment analysis outcomes (see Section 4, in which we discuss these two properties through a real application).

## Conclusions

This paper has presented DomainSenticNet—a resource that extends the OntoSenticNet commonsense ontology with: i) additional related concepts harvested from external knowledge bases and ii) distributional information on the occurrences and co-occurrences of each OntoSenticNet concept and related concepts in domain corpora. The paper also describes the methodology we adopted to generate DomainSenticNet. This methodology can be easily adapted to process different

domain corpora and external knowledge bases to generate domain-aware resources similar to ours and to extend semantic networks other than OntoSenticNet. Therefore, this methodology also enables the computation of domain-adapted scales of interpretation to benchmark domain ABSA application outcomes (as shown in Section 4).

To provide a concrete example of the benefit of DomainSenticNet to a variety of applications, we described a prototype tool for successful Kickstarter campaign authoring and campaign success prediction. Specifically, we discussed the high human interpretability level of both the prediction outcomes and the changes suggested for campaign descriptions to improve the likelihood of success. Moreover, the domain distributional information provided by DomainSenticNet enables it to produce domain-adapted scales of interpretation for predictive features at the level of influencing factors.

Regarding resource dynamicity (discussed in Section 3.5), we identify two opportunities: i) integrating updated releases (including new portions of the domain corpus) and ii) extending the current DomainSenticNet ontology schema with the inclusion of a time dimension. Additional dynamicity can be further leveraged by means of applying the proposed methodology (see Section 3) to other application-specific corpora. For instance, in the e-commerce domain, product and service reviews can be leveraged to capture the dynamics and trends of emotional intensities within customer opinion statements. Therefore, DomainSenticNet provides a basis for further interdisciplinary research within *behavioral economics*, *applied data sciences* and *applied mathematics*, with the aim of increasing the resource “dynamicity” to apply to an unlimited range of applications.

Additionally, to address the above-mentioned interdisciplinary investigations, we aim to study the effectiveness of *causal inference* approaches such as the *DoWhy* [32] framework. The *DoWhy* framework can be leveraged to gain insight into cause-and-effect relationships when domain adaption is applied. Such insights can then support the development and the interpretation of calculated domain-aware emotional intensity weights. Specifically, we are interested in the ability of the *DoWhy* approach to identify the correlation magnitude of unexploited features in classification models [28], thus enabling, for example, the magnitude of missing domain concepts to be determined.<sup>17</sup>

The current version of DomainSenticNet does not include sentiment polarities for *ExternalConcepts*; instead, it references OntoSenticNet for SenticConcept sentiment polarities. Therefore, other possible future research might aim at “propagating” the *Hourglass of Emotions* dimension weights and polarities to a collection of added external concepts. In addition, similar to [8, 11], our resource opens an avenue for further research on

<sup>17</sup> [https://microsoft.github.io/dowhy/dowhy\\_confounder\\_example.html](https://microsoft.github.io/dowhy/dowhy_confounder_example.html)



the generation of contextual domain embeddings in deep neural network-based applications. Finally, as discussed in Section 5, approaches such as [1, 21, 22] can leverage DomainSenticNet as an effective resource to improve the interpretability and explainability of domain-aware sentic applications.

**Acknowledgements** The work of Paolo Rosso was partially funded by the Spanish MICINN under the project PGC2018-096212-B-C31.

## Compliance with Ethical Standards

**Conflicts of Interest** The authors declare that they have no conflict of interest.

**Ethical Approval** The present work did not involve any research with human participants or animals performed by any of the authors.

## References

- Akhtar MS, Ekbal A, Cambria E. How intense are you? predicting intensities of emotions and sentiments using stacked ensemble [application notes]. *Computer Intelligence Magazine*. 2020;15(1):64–75. <https://doi.org/10.1109/MCI.2019.2954667>
- Alhussien I, Cambria E, NengSheng Z. Semantically enhanced models for commonsense knowledge acquisition. In: 2018 IEEE International Conference on Data Mining Workshops (ICDMW), p. 1014–1021. November 17–20, Singapore (2018). <https://doi.org/10.1109/ICDMW.2018.00146>
- Angulo C, Falomir IZ, Anguita D, Agell N, Cambria E. Bridging cognitive models and recommender systems. *Cogn Comput* 12(2), 426–427 (2020). <https://doi.org/10.1007/s12559-020-09719-3>
- Bandari S, Bulusu VV. Survey on ontology-based sentiment analysis of customer reviews for products and services. In: K.S. Raju, R. Senkerik, S.P. Lanka, V. Rajagopal (eds.) *Data Engineering and Communication Technology*, vol. 1079, pp. 91–101. Springer Singapore, Singapore (2020). [https://doi.org/10.1007/978-981-15-1097-7\\_8](https://doi.org/10.1007/978-981-15-1097-7_8)
- Billal B, Fonseca A, Sadat F, Lounis H. Semi-supervised learning and social media text analysis towards multi-labeling categorization. In: 2017 IEEE International Conference on Big Data (Big Data), pp. 1907–1916. December 11–14, Boston, MA, USA (2017). <https://doi.org/10.1109/BigData.2017.8258136>
- Cambria E, Grassi M, Hussain A, Havasi C. Sentic computing for social media marketing. *Multimed Tools Appl* 59(2), 557–577 (2012). <https://doi.org/10.1007/s11042-011-0815-0>
- Cambria E, Hussain A, Havasi C, Eckl C. Sentic Computing: Exploitation of Common Sense for the Development of Emotion-Sensitive Systems, *Lecture Notes in Computer Science*, vol. 5967, pp. 148–156. Springer Berlin Heidelberg, Berlin, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-12397-9\\_12](https://doi.org/10.1007/978-3-642-12397-9_12)
- Cambria E, Li Y, Xing FZ, Poria S, Kwok K. Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM '20*, p. 105–114. Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3340531.3412003>
- Cambria E, Livingstone A, Hussain A. The hourglass of emotions. In: A. Esposito, A.M. Esposito, A. Vinciarelli, R. Hoffmann, V.C. Müller (eds.) *Cognitive Behavioural Systems, COST 2012 International Training School*, vol. 7403, pp. 144–157. Springer Berlin Heidelberg (2012). [https://doi.org/10.1007/978-3-642-34584-5\\_11](https://doi.org/10.1007/978-3-642-34584-5_11)
- Cambria E, Poria S, Gelbukh A, Thelwall M. Sentiment analysis is a big suitcase. *IEEE Intell Syst* 32(06), 74–80 (2017). <https://doi.ieeecomputersociety.org/10.1109/MIS.2017.4531228>
- Cambria E, Poria S, Hazarika D, Kwok K. Senticnet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In: S.A. McIlraith, K.Q. Weinberger (eds.) *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18)*, the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), pp. 1795–1802. AAAI Press, New Orleans, Louisiana, USA (2018). <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16839>
- Chakraborty K, Bhattacharyya S, Bag R. A survey of sentiment analysis from social media data. *IEEE Transactions on Computational Social Systems* 7(2), 450–464 (2020). <https://doi.org/10.1109/TCSS.2019.2956957>
- Chauhan GS, Meena YK. Domsent: Domain-specific aspect term extraction in aspect-based sentiment analysis. In: A.K. Somani, R.S. Shekhawat, A. Mundra, S. Srivastava, V.K. Verma (eds.) *Smart Systems and IoT: Innovations in Computing*, vol. 141, pp. 103–109. Springer Singapore, Singapore (2020). [https://doi.org/10.1007/978-981-13-8406-6\\_11](https://doi.org/10.1007/978-981-13-8406-6_11)
- Dragoni M, Poria S, Cambria E. Ontosenticnet: A commonsense ontology for sentiment analysis. *IEEE Intell Syst* 33, 77–85 (2018). <https://doi.org/10.1109/MIS.2018.033001419>
- van Engelen JE, Hoos HH. A survey on semi-supervised learning. *Mach Learn* 109(2), 373–440 (2020). <https://doi.org/10.1007/s10994-019-05855-6>
- Faralli S, Rittinghaus S, Samsami N, Distant D, Rocha E. Emotional intensity-based success prediction model for crowd-funded campaigns. *Inf Process Manag* 58(1), article ID 102394 (2021). <https://doi.org/10.1016/j.ipm.2020.102394>
- Faralli S, Velardi P, Yusifli F. Multiple knowledge GraphDB (MKGDB). In: *Proceedings of The 12th Language Resources and Evaluation Conference*, pp. 2325–2331. European Language Resources Association, Marseille, France (2020). <https://www.aclweb.org/anthology/2020.lrec-1.283>
- Fellbaum C. (ed.): *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA (1998)
- Fernandez-Breis JT, Qazi A, Raj RG, Tahir M, Cambria E, Syed KBS. Enhancing business intelligence by means of suggestive reviews. *Sci World J* vol. 2014, article ID 879323 (2014). <https://doi.org/10.1155/2014/879323>
- Hussain A, Cambria E. Semi-supervised learning for big social data analysis. *Neurocomputing* 275, 1662 – 1673 (2018). <https://doi.org/10.1016/j.neucom.2017.10.010>
- Khatua A, Cambria E. A tale of two epidemics: Contextual word-2vec for classifying twitter streams during outbreaks. *Inf Process Manag* 56(1), 247 – 257 (2019). <https://doi.org/10.1016/j.ipm.2018.10.010>
- Kumar A, Srinivasan K, Cheng WH, Zomaya AY. Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. *Inf Process Manag* 57(1), article ID 102141 (2020). <https://doi.org/10.1016/j.ipm.2019.102141>
- Li H, Armiento R, Lambrix P. A method for extending ontologies with application to the materials science domain. *Data Science Journal* 18, 1–21 (2019). <https://doi.org/10.5334/dsj-2019-050>
- Ma Y, Peng H, Cambria E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm. In: *AAAI Conference on Artificial Intelligence*, pp. 5876–5883 (2018). <https://aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16541>

25. Me D, Frasinca F, Aldonar: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model. *Inf Process Manag* 57(3), article ID 102211 (2020). <https://doi.org/10.1016/j.ipm.2020.102211>
26. Nguyen HT, Duong PH, Cambria E. Learning short-text semantic similarity with word embeddings and external knowledge sources. *Knowledge-Based Systems* 182, article ID 104842 (2019). <http://www.sciencedirect.com/science/article/pii/S095070511930317X>
27. Paulheim H. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web* 8(3), 489–508 (2017). <https://doi.org/10.3233/SW-160218>
28. Pearl J, Mackenzie D. *The Book of Why*. Basic Books, New York (2018). <https://dl.acm.org/doi/book/10.5555/3238230>
29. Plutchik R. The nature of emotions. *Am Sci* 89(4), 344–350 (2001). <https://www.jstor.org/stable/27857503>
30. Rajagopal D, Cambria E, Olsher D, Kwok K. A graph-based approach to commonsense concept extraction and semantic similarity detection. In: *Proceedings of the 22nd International Conference on World Wide Web, WWW '13 Companion*, p. 565–570. Association for Computing Machinery, New York, NY, USA (2013). <https://doi.org/10.1145/2487788.2487995>
31. Saif H, Fernandez M, He Y, Alani H. Senticircles for contextual and conceptual semantic sentiment analysis of twitter. In: V. Presutti, C. d'Amato, F. Gandon, M. d'Aquin, S. Staab, A. Tor-dai (eds.) *The Semantic Web: Trends and Challenges*, pp. 83–98. Springer International Publishing, Cham (2014). [https://doi.org/10.1007/978-3-319-07443-6\\_7](https://doi.org/10.1007/978-3-319-07443-6_7)
32. Sharma A, Kiciman E. DoWhy: A Python package for causal inference (2019). <https://github.com/microsoft/dowhy>
33. Shiller R. Narrative economics. *Am Econ Rev* 107, 967–1004 (2017). <https://doi.org/10.1257/aer.107.4.967>
34. Song Y, Wang H, Wang Z, Li H, Chen W. Short text conceptualization using a probabilistic knowledgebase. In: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Three, IJCAI'11*, p. 2330–2336. AAAI Press, Barcelona, Catalonia, Spain (2011)
35. Susanto Y, Livingstone AG, Ng BC, Cambria E. The hourglass model revisited. *IEEE Intell Syst* 35(5), 96–102 (2020). <https://doi.org/10.1109/MIS.2020.2992799>
36. Wu W, Li H, Wang H, Zhu KQ. Probbase: A probabilistic taxonomy for text understanding. In: *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, SIGMOD*, p. 481–492. Association for Computing Machinery, New York, NY, USA (2012). <https://doi.org/10.1145/2213836.2213891>
37. Xia Y, Cambria E, Hussain A, Zhao H. Word polarity disambiguation using bayesian model and opinion-level features. *Cogn Comput* 7(3), 369–380 (2015). <https://doi.org/10.1007/s12559-014-9298-4>
38. Xing FZ, Cambria E, Welsch RE. Natural language based financial forecasting: a survey. *Artif Intell Rev* 50(1), 49–73 (2018). <https://doi.org/10.1007/s10462-017-9588-9>
39. Zhuang L, Schouten K, Frasinca F. Soba: Semi-automated ontology builder for aspect-based sentiment analysis. *Journal of Web Semantics* 60, article ID 100544 (2019). <https://doi.org/10.1016/j.websem.2019.100544>