Mauro Dragoni, Marco Federici, Andi Rexha, **An unsupervised aspect extraction strategy for monitoring real-time reviews stream**, INFORMATION PROCESSING & MANAGEMENT, Volume: 56, Issue: 3, pp. 1103-1118, May 2019, DOI: 10.1016/j.ipm.2018.04.010

An Unsupervised Aspect Extraction Strategy For Monitoring Real-Time Reviews Stream

Mauro Dragoni

Fondazione Bruno Kessler, Trento, Italy

Marco Federici

University of Amsterdam, Netherlands

Andi Rexha

Know-Center, Graz, Austria

Abstract

One of the most important opinion mining research directions falls in the extraction of polarities referring to specific entities (aspects) contained in the analyzed texts. The detection of such aspects may be very critical especially when documents come from unknown domains. Indeed, while in some contexts it is possible to train domain-specific models for improving the effectiveness of aspects extraction algorithms, in others the most suitable solution is to apply unsupervised techniques by making such algorithms domain-independent and more efficient in a real-time environment. Moreover, an emerging need is to exploit the results of aspect-based analysis for triggering actions based on these data. This led to the necessity of providing solutions supporting both an effective analysis of user-generated content and an efficient and intuitive way of visualizing collected data. In this work, we implemented an opinion monitoring service implementing (i) a set of unsupervised strategies for aspect-based opinion mining together with (ii) a monitoring tool supporting users in visualizing analyzed data. The aspect extraction strategies are based on the use of an open information extraction strategy. The effectiveness of the platform has been tested on benchmarks provided by

Email addresses: `dragoni@fbk.eu` (Mauro Dragoni), `marco.federici@student.uva.nl` (Marco Federici), `arexha@know-center.at` (Andi Rexha)

the SemEval campaign and have been compared with the results obtained by domain-adapted techniques.

Keywords: Real-time Opinion Mining, Aspect-based Sentiment Analysis, Decision Support System

1. Introduction

Online services like booking platforms, shops, and social media are becoming widely used by an increasing percentage of population. Each user of *the Internet* can express own opinions regarding products, services, or even other people thoughts. Opinions expressed by masses can lead other consumers to different choices, give direct feedback to the producers, or underline problematics of a service. For all these reasons, in the last few years, a lot of effort has been invested in understanding and extracting valuable data from user's reviews.

Opinion Mining and Sentiment Analysis are Natural Language Processing (NLP) tasks that aims to extract opinions from texts and to classify them with a value representing the overall polarity (*positive*, *negative*, or *neutral*) associated with a given subject [1, 2]. This research field attracted a lot of interest due to the possibility of applying developed strategies to a wide set of applications in different domains like marketing, politics, and social sciences. In the beginning, built applications aimed to compute overall polarity values then associated with a document. By using this strategy was not possible to distinguish which were the subject of each opinion and how such a subject was judged by users. This issue led to focus on the extraction of all subjects, namely *aspects*, from texts in order to equip developed systems with the possibility of computing aspects' polarities independently [3].

Let us consider the following example:

Last weekend, I tried a new restaurant in downtown.
The place was awesome, but the quality of the food was quite poor.

The proposed example contains three aspects, *restaurant*, *place*, and *food*, and each aspect is associated with a specific opinion:

- *place* → *awesome*
- *food* → *quite poor*
- *restaurant* → no opinions. In this case, the polarity can be computed by averag-ing the polarities associated with the other aspects contained in the document.

For obtaining the opinion-based structure of the sentence, it is necessary to address two tasks: (i) the detection of the aspects, and (ii) the computation of the associated polarities. While the latter is easily supported by using opinion-based dictionaries (Section 3), the former requires different strategies. Many approaches presented in the literature, and discussed in Section 2, proposed supervised models for extracting aspects from text. Unfortunately, the use of a supervised approach clashes with real-world requirements. Firstly, the creation of a model requires annotated datasets containing aspects annotations for all possible domains. Nowadays, these datasets are not available except for a limited number of domains. Secondly, a document can have sentences belonging to many of these domains. Hence, the use of a single model is not feasible.

In light of these challenges, the development of approaches able to provide effective aspect extraction, polarity computation, and data visualization procedures is of interest for contexts where it is necessary to provide dashboards showing a real-time summary of opinion-based data-streams containing documents belonging to domains unknown a-priori. The use of an open information extraction strategy can be suitable for a real-time scenario. Indeed, here the system has to extract and to analyze information coming from all possible domains in an efficient way.

This work focuses on the creation of an opinion-based support system built upon the following three pillars.

- the design and the development of an open information extraction approach for supporting the detection of aspects within texts;
- the design and the development of a scalable platform able to process a high volume of opinion-based documents in real-time; and,
- the development of a data visualization interface supporting an easy access to the processed data.

The innovative aspect of this solution focuses on the combination of these three pillars in order to position this work as a state of the art platform for the real-time management of complex opinion-based documents.

One of the aim of the proposed system is to support different kind of users (managers, buyers, customers, etc.) with a multi-facet analysis of products' features. Indeed, the main issue when a product is judged with a single metric (e.g. overall document polarity) is that it does not allow users to obtain results tailored to their specific needs. For example, customers of an online shop could be interested in the *battery life* of a laptop rather than its overall quality. Currently, they do not have the possibility of obtaining this kind of information directly from the reviews, if users do not have the possibility of rating the specific *battery life* aspect.

The paper is structured as follows. In Section 2, we provide an overview of the opinion mining field with a focus on aspects extraction approaches. Section 3 introduces the background knowledge bases integrated into the proposed platform. Section 4 provides an overview of platform's components, while in Sections 5 and 6 we describe the strategy used for extracting aspects and the client application we developed for supporting users in monitoring the real-time data stream, respectively. Section 7 discusses the overall performance of our platform. Section 8 concludes the paper.

2. Related Work

In this Section, we briefly review the main contributions in the field of sentiment analysis and opinion mining, firstly from a general standpoint and then with a particular attention to the social media scenario. A brief overview of significant recent contributions in the open information extraction field is also provided.

2.1. Sentiment Analysis and Opinion Mining

The topic of sentiment analysis has extensively been studied in the literature [4, 2], where several techniques have been proposed and validated.

Machine learning techniques are the most common approaches used for addressing the sentiment analysis problem. For instance, in [1] and [5] the authors compared the performance of Naive-Bayes, Maximum Entropy, and Support Vector Machines classifiers in sentiment analysis, using different features like considering only unigrams, bigrams, combination of both, incorporating parts of speech and position information, or considering only adjectives.

The recent massive growth of online product reviews paved the way for using sentiment analysis techniques in marketing activities. The issue of detecting the different opinions concerning the same product expressed in the same review emerged as a challenging problem. This task has been carried out by identifying the aspect of the product that a sentence in the opinion may refer to. In the literature, many approaches have been proposed: conditional random fields (CRF) [6], hidden Markov models (HMM) [7], sequential rule mining [8], dependency tree kernels [9], and clustering [10].

Recently, the application of sentiment analysis approaches attracted a lot of interest also in the social networks research field [11]. The use of social networks for expressing opinions and comments about products, political or social events, significantly increased in the last years. However, the analysis of the social network environment brought to light new challenges mainly related to (i) the different ways people express their opinions (i.e. *multi-modality*) and to (ii) the management of noisy data contained in social network texts [12].

The social dimension of the Web fostered the development of multi-disciplinary approaches combining computer and social sciences to improve the interpretation, recognition, and processing of opinions and sentiments expressed in social networks. The synergy between these approaches has been called sentic computing [13]. Sentic computing has been employed for addressing several cognitive-inspired problems like the classification of natural language text [14] and the extraction of emotions from images [15].

Real-world solutions have been also developed. For example the authors of SENTILO [16, 17] presented a semantic-based solution for extracting opinion frames from texts.

Different level of granularities have been considered: while some approaches operate at document level [18, 19], other focus their goal on opinion classification by the means of a fine-grained analysis of the text at sentence level [20, 21]. Other approaches propose the use of fuzzy logic [22, 23] or other aggregation techniques [24] to compute the score of each single word. In the case of sentence-level opinion classification, two different sub-tasks have to be addressed. The first one, called *subjectivity classification*, consists in detecting if the sentence is subjective or objective, while the second one focuses on determining if the expressed opinion is positive, negative, or neutral. *Subjectivity classification* rose great interest in the community [20, 21]. Systems implementing the capabilities of identifying opinion's holder, target, and polarity have been discussed [25].

Recent work on text modality used Convolutional Neural Network (CNN) [26] for sentiment related tasks such as sarcasm detection [27] and aspect-based opinion mining [28]. Several deep learning based approaches have been evaluated in Sentiment Analysis tasks. In [29], Recursive Neural Networks are used to handle the syntactic tree structure of a sentence: following the generated parse tree, the different distributed representations of sentence parts are recursively built. The model is trained on the Stanford Sentiment Treebank, which has annotations on the whole parse tree. Authors of [30] learn a distributed representation of reviews through Convolutional Neural Networks which are subsequently feed into Recurrent Neural Networks to learn distributed representation of the viewed products and of the opinion holders. In [28] a 7-layer deep Convolutional Neural Network has been trained to identify the target of an opinion within a text fragment, in conjunction with some linguistic patterns. An Extreme Learning Machine approach implemented over the data analytics framework Apache Spark[1] has been proposed in [31]. The approach deals with large amount of natural language text coming from the Social Web.

2.2. Opinion Mining in Social Media

The application of opinion mining approaches in social media became attractive by opening up new challenges due to the different ways people express their opinions [12]. People use social networks to express their moods and opinion about recently purchased items or new products available on the marketplaces.

[1]https://spark.apache.org/

One of the first studies on opinion mining on micro-blogging websites has been discussed in [11], where the authors presented a distant supervision-based approach for opinion classification on Twitter. In [32] the authors presented SentiStrength. The described algorithm focuses on the detection of emotion strength in a social context. SentiStrength implements a machine learning approach aiming to optimize opinion words weightings used for inferring the polarity of each message. Moreover, the approach implements a spelling correction method used to address the misspelling issue which often occurs on user-generated content.

However, the research in social media analytics has only recently started to employ aspect-based sentiment analysis in the process of attracting users. In particular, aspects extracted from opinions are exploited to attract users to follow the links related to products which have been judged interesting by users communities. A first attempt to exploit extracted aspects for better orienting advertisements content is discussed in [33]. While in [34], the authors focused on tips instead of reviews. Their objective was recommending the right tips to the right people via the Foursquare platform, by taking into consideration the timeliness of user-provided tips and the users' tastes and social connections.

The increasing number of online product reviews enhanced the development of new opinion mining techniques due to their value in marketing activities. The detection of opinions regarding a specific product emerged as a real challenge. In this context, aspect extraction approaches achieved interesting results. The aspect extraction literature is divided into two distinct paths: supervised and unsupervised methodologies. The first one requires manually annotated data and it is mainly based on Conditional Random Fields [6, 35, 36, 37], while the latter is focused on topic modeling [38, 39, 40, 41] and dependency relations [42]. Other approaches propose hidden Markov models (HMM) [43, 7], sequential rule mining [8], dependency tree kernels [9], clustering [10], and genetic algorithms [44]. With respect to these works, our approach relies on a scalable and unsupervised technique for detecting domain-specific aspects from opinion documents. This way, we are able to cope the challenge of deploying a light system into real-world general purpose scenarios.

In this paper, we bridge the aspect-based opinion mining and the user engagement areas by providing a smart way to exploit the knowledge extracted from user reviews. This work has been implemented and validated in a specific context. However, the approach described in Section 5 can be easily deployed in different scenarios.

2.3. Open Information Extraction

In the past years, a lot of research has been dedicated to constantly improve the performance of Open Information Extraction (OpenIE) systems. In the beginning, shallow syntactical features such as part-of-speech tags were employed: *TextRunner* [45], *WOE*^{pos} [46], and *ReVerb* [47] making these systems highly efficient but poor in quality.

To improve the extraction quality, complex features, like dependency tree information, started to be exploited: *Kraken* [48], *Ollie* [49], *ClausIE* [50], and *CSD-IE* [51].

So far, the majority of the research focused on the English language, but other languages such as Spanish [52], Chinese [53], and German [54] recently attract interest from the research community. The word presented in [55] showed that OpenIE based on dependency trees is suitable for various languages besides English. They used a multilingual parser with a common output tag-set for the supported languages (English and Romance).

The multilingual OpenIE system *ArgOE* [56] tries to be more open for different dependency parsers by using the CoNLL-X format. It manages to extract tuples in several languages with the same rule set, relying on a dependency parser that uses a common tag-set for five European languages. In [52] the Spanish system *ExtrHech* has been described. It works with part-of-speech-tagged input and semantic constraints, demonstrating that this approach achieves similar results for Spanish and English as well.

SCOERE [53] is an OpenIE system for the Chinese language. It uses a semi-supervised approach and focused on a fixed set of entities, namely person, organization, location and time. [54] introduced *PropDE*, an OpenIE system for the German language. The *PropDE* system transfers the available set of extraction rules (*PropS* [57]) from English to German.

3. Material

Before presenting the system architecture and the approach designed for the specific aspect extraction task, we introduce here the resources we used for supporting the whole text analysis activity. We exploited four different resources: a stopwords list,[2] sentiment lexicons, a linguistic knowledge base, and a general-purpose natural language processing library.

3.1. Sentiment Lexicons

Sentiment Lexicons are used for associating each term with a polarity value. Terms having such an association are called *opinion words* and they are used for estimating the polarity of a given sentence. Associating a polarity value to a specific word is a task that has been addressed by different perspectives. The results have been the availability of different resources that can be easily integrated within real-world systems. In our platform, we decided to aggregate polarity values coming from three resources freely available: SenticNet [58], the General Inquirer vocabulary [3] [59], and the MPQA dictionary[4] [60].

SenticNet is a publicly available resource for opinion mining that exploits both artificial intelligence and semantic Web techniques to infer the polarities associated with common-sense concepts and to represent them in a semantic-aware format. The development of SenticNet was inspired by SentiWordNet [61], a lexical resource in which each WordNet synset is associated to three numerical scores describing how objective, positive, and negative the terms contained in each synset are. The differences between SenticNet and SentiWordNet are basically three: (i) in SentiWordNet, each synset is associated to a three-valued representation (the objectivity of the synset, its positiveness, and its negativeness), while in SenticNet there is only one value belonging to the $[-1, 1]$ interval for representing the polarity of the concept; (ii) SenticNet provides the sentiment model of more complex common-sense concepts, while SentiWordNet

[2]The used stopwords list is available at http://www.lextek.com/manuals/onix/stopwords1.html
[3]http://www.wjh.harvard.edu/ inquirer/spreadsheet_guide.htm
[4]http://mpqa.cs.pitt.edu/corpora/mpqa_corpus/

is focused on assigning polarities to WordNet synsets: for instance, in SenticNet, complex concepts like *make good impression*, *look attractive*, *show appreciation* , *being fired*, *leave behind*, or *lose control* are used for defining positive or negative situations; and (iii) completely neutral concepts are not reported. SenticNet contains almost 40,000 polarity concepts and it may be connected with any kind of opinion mining application. For example, after the de-construction of the text into concepts through a semantic parser, SenticNet can be used to associate polarity values to these and, hence, to infer the overall polarity of a clause, sentence, paragraph, or document by averaging such values.

The *General Inquirer* is an English-language dictionary containing almost 12,000 elements associated with their polarities in different contexts. Such dictionary is the result of the integration between the *Harvard* and the *Lasswell* general-purpose dictionaries as well as a dictionary of categories defined by the dictionary creators. When necessary, for ambiguous words, specific polarity for each sense is specified. For every word, a set of tags is provided in the dictionary. Only a subset of them are relevant to the opinion mining topic and, thus, exploited in this work:

- Valence categories: the two well-known *positive* and *negative* classifications.
- Semantic dimensions: these tags reflect semantic differential findings regarding basic language universals. These dimensions are: *hostile*, *strong*, *power*, *weak*, *submit*, *active*, and *passive*. A word may be tagged with more than one dimension, if appropriate.
- Words of pleasure: these tags are usually also classified positive or negative, with virtue indicating strength and vice indicating weakness. They provide more focus than the categories in the previous two bullets. Such categories are *pleasure*, *pain*, *feel*, *arousal*, *emotion*, *virtue*, *vice*.
- Words reflecting presence or lack of emotional expressiveness: these tags indicate the presence of overstatement and understatement; trivially, such tags are *overstated* and *understated*.

Other categories indicating ascriptive social tags rather than references to places have been considered out of the scope of the opinion mining topic and have not been considered in the implementation of the approach.

Finally, *MPQA* is a sentiment lexicon built for Multi-Perspective Question Answering purposes. The lexicon contains around 8,222 terms annotated with their polarity (*positive*, *negative*, and *neutral*) and with their intensity level (*strong* and *weak*) and a set of 10,000 sentences manually annotated through the proposed annotation scheme. Indeed, besides the classic association $\langle word, polarity \rangle$, the MPQA lexicon implements a detailed annotation scheme that identifies key components and properties of opinions, emotions, sentiments, speculations, evaluations, and private states [62]. This annotation scheme covers a broad and useful subset of the range of linguistic expressions and phenomena employed in naturally occurring text to express opinion and emotion. The proposed annotation scheme is relatively fine-grained, annotating text at the word- and phrase-level rather than at the level of the document or sentence. For every expression of a private state in each sentence, a private state frame is defined. A private state frame includes the source of the private state (i.e., that whose private state is being expressed), the target (i.e., what the private state is about), and various properties involving intensity, significance, and type of attitude. An important property of sources in the annotation scheme is that they are nested, reflecting the fact that private states and speech events are often embedded in one another. The representation scheme also includes frames representing material that is attributed to a source, but is presented objectively, without evaluation, speculation, or other type of private state by that source.

The lists of terms contained in the resources presented above do not overlap completely. The strategy implemented within our platform considers words with a non-zero polarity value in at least one of the integrated resources. For example, the word *third* is not present neither in MPQA nor in SenticNet and has a polarity of 0 according to the General Inquirer. Consequently, it is not a valid opinion word. On the other hand, the word *huge* has a positive value of 0.069 in SenticNet, a negative value of -1 in MPQA and a value of 0 in the General Inquirer, therefore, it is evaluated as opinion word even if lexicons express contrasting values. SenticNet already implements a continuous representation of polarity values. MPQA uses a discrete scale $[-1, 0, 1]$ that has been extended to $[-1, -0.5, 0, 0.5, 1]$ by halving -1 and 1 when the *weak* intensity level is present. For the General Inquirer the same strategy adopted for the MPQA lexicon has been adopted by exploiting the semantic dimension of the dictionary for halving the -1 and 1 values. Finally, the three values are aggregated by using the arithmetic average.

3.2. WordNet

WordNet[5] [63] is a large lexical database of English nouns, verbs, adjectives, and adverbs grouped into sets of cognitive synonyms called *synsets*, where each synset expresses a distinct concept. In particular, each synset represents a list of synonyms, intended as words that denote the same concept and that are interchangeable in many contexts. WordNet contains around 117,000 synsets linked to each other by a small set of *conceptual relations*, i.e., synonymy, hypernymy, hyponymy, etc.. Additionally, a synset contains a brief definition (*gloss*) and, in most cases, one or more short sentences illustrating the use of the synset members. Words having several distinct meanings are represented in as many distinct synsets. Even if WordNet superficially resembles a thesaurus, there are some important distinctions with respect to it. Firstly, WordNet does not define links between words, but between specific senses of words; this way, words that are found in close proximity to one another in the network are semantically disambiguated. Secondly, WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus does not follow any explicit pattern other than the similarity of their meanings. In the implemented system, Wordnet's compound names list has been used to detect word sequences that represent a single noun.

3.3. Stanford Core NLP

The preliminary textual analysis, consisting in converting the raw input text in an annotated and structured representation, is performed through the Stanford Core Natural Language Processing Library [64]. Stanford CoreNLP is an integrated framework

[5] https://wordnet.princeton.edu/

providing a wide range of natural language analysis tools. Each functionality is provided by a specific module. Below, we show the four modules of the CoreNLP library adopted within our system.

The *Pos Tagger* (Part Of Speech Tagger) is a software module aiming to assign a part of speech tag (such as noun, verb, adjective, etc.) to every word of a given sentence [65]. The *Coref Annotator* (Co-reference resolution Annotator) generates co-reference Chain Annotations representing groups of words referring to the same entity [66]. Chains are used to resolve pronoun references. The *Parse Annotator* (Parser Annotator) [67] provides full syntactic analysis generating a tree grammar dependencies structure. Finally, the *Depparse Annotator* (Dependency Parser Annotator) [68] provides a representation of grammatical relations between words in a sentence producing graphs like the one shown in Figure 4.

4. System Architecture

The system presented in this work implements a set of modules for supporting the gathering, the processing, and the analysis of opinion-based document stream. In particular, we focused on the Amazon website. Figure 1 shows an abstract overview of these modules. Reviews collected in real-time from the Amazon website are given as input to the ***Data Manager Module*** that is responsible of parsing raw documents and of enriching them with further metadata. Processed data are saved into a knowledge repository in order to make them available to a ***Web Service*** that is responsible of exposing the structured knowledge as result of client requests.

The workflow works in the following way. The stream of reviews are given as input to the *Data Manager Module*. This module is composed by two components: the Document Analyzer Pipeline and the Document Enricher. This former is responsible of applying the open information extraction strategy, together with other natural language processing tools, for extracting tuples containing the aspects mentioned in the text and the associated polarities. Details about this module are provided in Section 5.

Once a review is analyzed, the result is sent to the Document Enricher component that is responsible of linking information extracted from text to the product for which

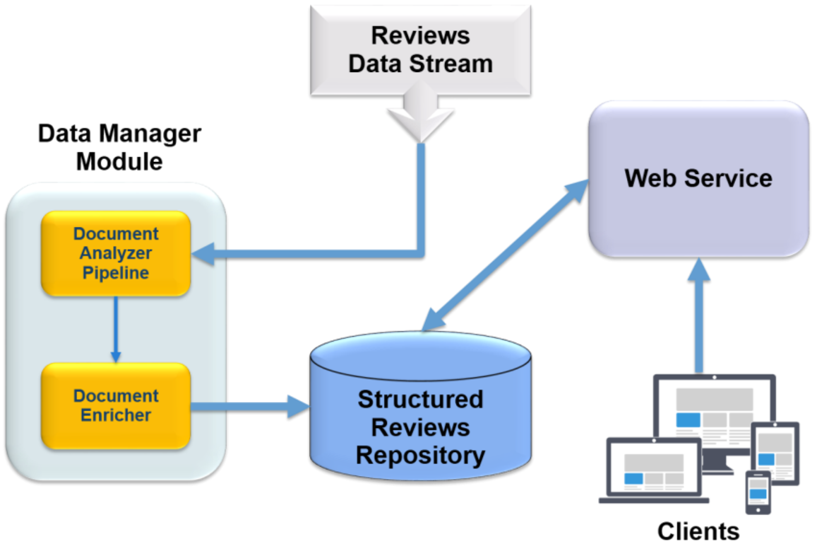

Figure 1: Overview of the implemented platform.

the review has been provided. The linking operation consists of retrieving the product name, the domain, the category, and the review score. This operation has been implemented on top of the Amazon Product API. Here, given the product's id contained in the review's metadata, it is possible to retrieve the product's information mentioned above.

The output of the *Data Manager Module* is the structured representation shown in Figure 2. Each object is then stored into the repository.

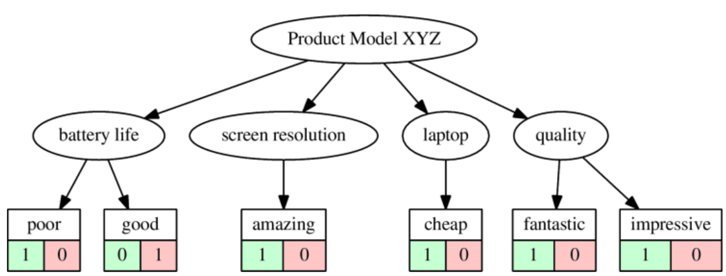

Figure 2: Example of the tree structure extracted by NLP Module

Leaves of the tree contain the label of the opinion word and its polarity. These are associated with the respective aspects contained within the connected upper-layer. All aspects are finally associated with the product entity.

The content of the repository is then exploited by final users through the *Web Service* integrated into the platform. The *Web Service*, and the client application briefly presented in Section 6, enables users to access data and to have a real-time visualization about the opinion trends associated with products' categories, items, or specific features of items.

5. Document Analyzer Pipeline

Here, we present the approach implemented for extracting aspects from document content. The overall aspect extraction approach relies on the NLP pipeline shown in the middle layer of Figure 3.

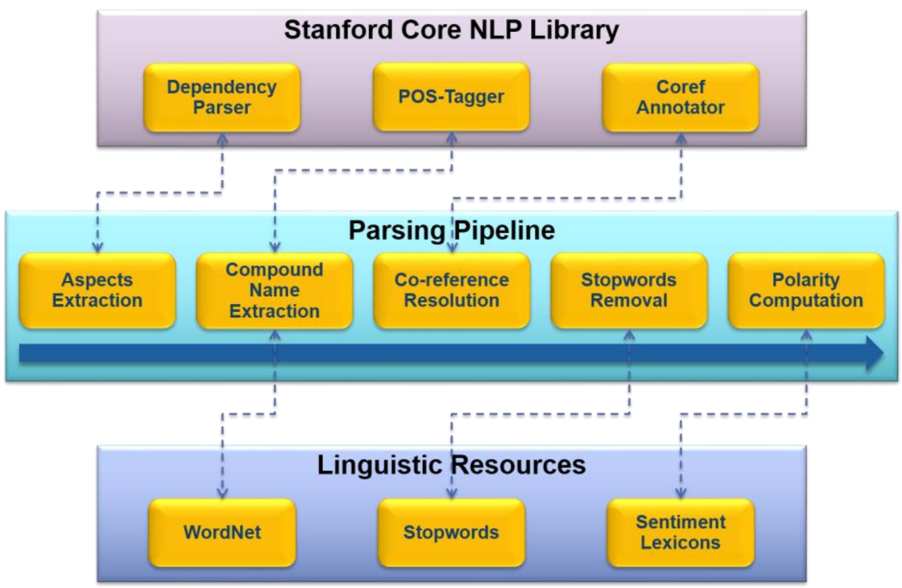

Figure 3: The NLP pipeline implemented within the proposed platform aiming to extract aspects and compute their polarity from the analyzed textual resources.

As it is shown on the bottom layer, the implemented pipeline exploits the three linguistic resources introduced in Section 3. These resources are used by the Stanford Core NLP Library [64] shown in the top layer of Figure 3.

The pipeline is composed by the following five phases:

- Aspect Extraction. This first step is the most important one and it consists on detecting the correct aspects contained in the text and the associated opinion words. Details about this step are provided below where we present the open information extraction algorithm adopted for analyzing provided text.

- Compound Noun Extraction. The second step consists in detecting the presence of compound names. This step is supported by the use of the POS-Tagger module provided by the Stanford Core NLP library and by WordNet (both introduced in Section 3). When two consecutive words are tagged as nouns by the POS-Tagger, their composition is searched within the WordNet dictionary. If the compound expression is found, it is tagged as compound name and used as a unique token, otherwise not.

- Co-reference Resolution. This step consists in associating pronouns with the related noun (or compound noun). This is necessary for detecting all associations between opinion words and aspects. This operation is completely supported by the Coref Annotator. Refinements of the adopted algorithm are out of scope of this paper and they are part of future work.

- Stopwords Removal. Once compound names have been detected and pronouns have been replaced with the right terms, the pipeline removes all stopwords from the text by exploiting the list mentioned above.

- Polarity Computation. Finally, this step is responsible for computing the polarity associated with each aspect extracted during the previous steps. The overall polarity of an aspect A is computed by aggregating the single polarities of the opinion words associated with A. Single polarity values are extracted and aggregated from the sentiment lexicons as described in Section 3.

5.1. *The Open Aspect Extraction Strategy*

The Open Aspect Extraction component uses a generic solution for identifying possible aspects in the user's opinion. This component implements an OpenIE strategy for supporting this task. OpenIE is a NLP branch of research that tries to determine meaningful patterns over parsing structure of a sentence and morphological characteristics.

The developed algorithm analyzes the structure of the grammar dependencies graph generated by the CoreNLP library for extracting the connections between aspects and opinions. Each dependency extracted by the CoreNLP library can be expressed by a triple: {*Relation_Type, Governor, Dependant*}[6].

The generated dependency graph is then processed by applying a set of rules for determining if the content of each node is supposed to be an aspect, an opinion, or nothing. These rules can be considered as a representation of the most common patterns that can be used for detecting pairs of the type *aspect-opinion_word*. The choice of these three rules allows at the same time to have a system that is efficient in processing document content and effective in covering content structure. Indeed, results of an in-vitro experiment shown the by disabling one of the rules the effectiveness of the system dramatically decreases. Hence, given a dependency node *n*, the algorithm checks if one of the following rules subsists:

Rule 1: If the relation type is an adjectival modifier ("amod"), a connection between an aspect and an opinion word persists if and only if the governor is an aspect and the dependant has a polarity value in at least one of the sentiment lexicons.

Rule 2: If the relation type is a nominal subject ("nsubj"), a connection between an aspect and an opinion word persists if and only if the governor has a polarity value in at least one of the sentiment lexicons and the dependant is an aspect.

Rule 3: If the relation type is a direct object ("dobj"), a connection between an aspect and an opinion word persists if and only if the governor has a polarity value in at least one of the sentiment lexicons and the dependant is an aspect.

Figure 4 shows the result obtained by applying these three rules to our running example and Figure 5 summarizes only the valid relationships extracted from the grammar dependencies graph. We reported with a dotted line also the relationship between *I* and *enjoyed*. Actually, this relationship is not valid because *I* is tagged as *personal*

[6]The meaning of each element of the triple together with all the possible relation type, can be found in the official Stanford Document available at http://nlp.stanford.edu/software/dependencies_manual.pdf

pronoun but within the sentence such a pronoun is not resolved.

The color code used in the figures is the following: light red nodes are nouns that have not been detected as *aspect* by the system; red nodes are nouns that have been detected as *aspect* by the system; green nodes are verbs for which a polarity value is present in the sentiment lexicons; and, blue nodes are adjectives for which a polarity value is present in the sentiment lexicons.

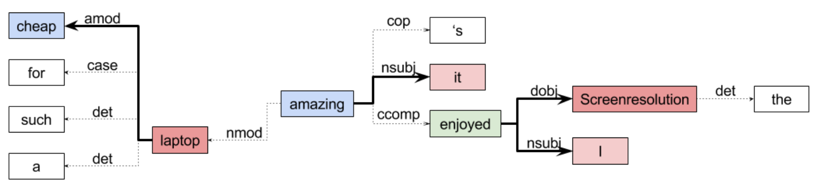

Figure 4: Dependency graph generated by the implemented approach.

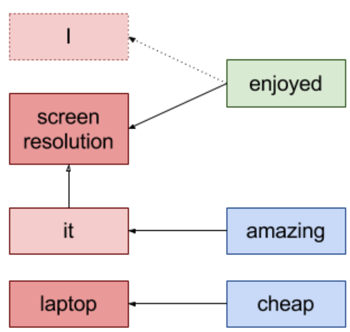

Figure 5: Relationships generated by the implemented approach.

Finally, the extracted relationships can be summarized as follows:

$$\text{laptop} \leftarrow \{\text{cheap}\}$$

$$\text{screenresolution (it)} \leftarrow \{\text{amazing, enjoyed}\}$$

These associations allow our system to infer, for both aspects *laptop* and *screen_resolution*, a positive polarity.

6. Client User Interface

The platform has been equipped with a web-based client application for supporting users during the analysis of the processed data. Users can query the repository by means of a controlled query interface built on a single-page web application that provides all data visualization functionalities.

The client interface has been designed with the aim of being very simple and intuitive. First of all, users have to select a category from the related list. They can eventually specify an aspect of their interest and, then, submit their request. The web service will provide the list of products according to the specified category and the requested aspect. The client application is then in charge of organizing the response data as shown in Figure 6. The harmonization of the terms provided by the users (plural forms, synonyms, etc.) is performed by using the WordNet [63] lexicalizer.

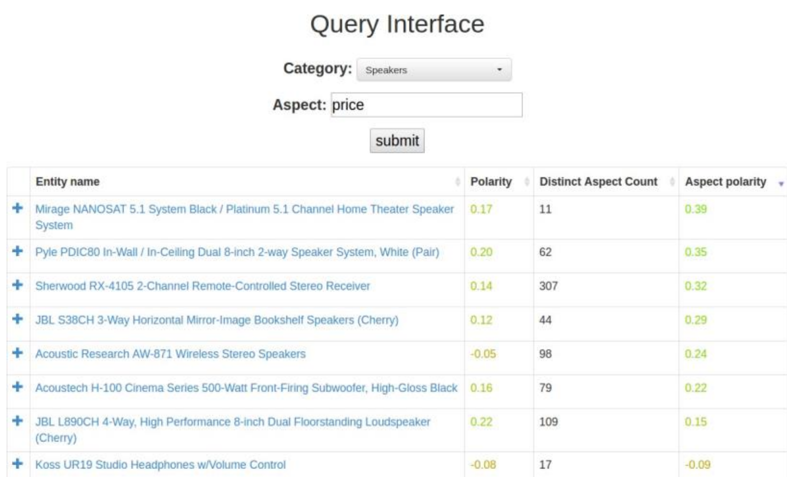

Figure 6: Example of query results.

Each row can be ordered according to the product name, the number of reviewed aspects, the average polarity, or the polarity of the aspect provided by the user, if any. These two last metrics are particularly useful because they represent, respectively, the customers' overall opinion of the product and their appreciation of the selected aspect.

A complete visualization of the product's opinion hierarchy is generated by clicking on its name. Figure 7 shows the editable tree-view obtained by selecting a specific product. Users can hide opinions for a better visualization of larger trees. Moreover, colors have been added to give an immediate feedback on polarity values.

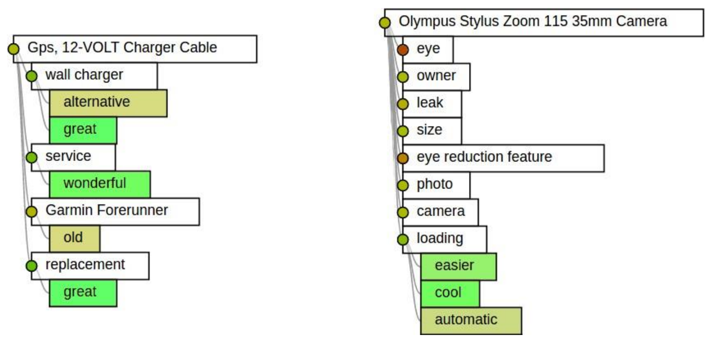

Figure 7: Example of the tree visualization of the data regarding two products

For retrieving further product details, each row can be expanded for showing the details of every single aspect extracted from the reviewed entity. The aspects sub-table shows aspect's name, average polarity, and the number of related opinion.

Single opinions can be visualized by expanding the aspect's row as shown in Figure 8.

−	reception	0.02	4

Opinion name	Polarity	Positive Count	Negative Count	Inferred	General Inquirer	MPQA	SenticNet
great	0.49	1	0	0.07	0.50	0.75	0.86
good	0.20	2	0	0.03	0.50	0.25	0.88
better	-0.03	1	0	-0.01	0.50	0.25	0.13
terrible	-0.61	1	0	-0.10	-0.50	-0.75	-0.90

Figure 8: Example of the opinion subtable

Once displayed, opinion's table presents a slightly different structure. Each row reports opinion's name and its polarity as well as the values associated with the same word in the other lexicons resources. This way, it is possible to do an immediate comparison of such values. The count of positive and negative occurrences of the specific opinion are shown in two separates columns.

From the technological perspective, the main component of the web interface has been developed with Spark Micro Framework[7]. This Java library facilitate the creation of a simple REST service to manage client's requests. Each call is binded to a specific query function which automatically maps the results to a JSON serializable object. The client side JavaScript code has been written following MVC pattern. AJAX requests and responses are handled by JQuery, Bootstrap[8], and Bootstrap-table[9] JavaScript libraries. These libraries are in charge of managing the presentation layer. Finally, the D3.js[10] library has been used to represent complex product data.

7. Evaluation

In this Section, we present the evaluation of the proposed system. Such a system is evaluated under different perspectives aiming to show the efficiency and effectiveness of the implemented modules:

- *Aspect extraction*. The OpenIE approach is in charge of detecting aspects within text. Such a task is important for defining, later in the analysis process, which aspects are the most significant ones and which are the opinion words associated with them. This evaluation task focused on measuring the effectiveness of the aspect-extraction approach.

- *Polarity detection*. The computation of the aspect's polarity enables the detection of which product features are strong or weak. The Sentiment Module is in charge of inferring the polarity of each aspect given the context in which such an aspect is included. Here, we measured the capability of our approach to infer the correct polarity.

- *Lessons Learned*. Besides the effectiveness of the technological components, we provide a discussion about the usability of the user interface and the direction

[7]http://sparkjava.com/
[8]http://getbootstrap.com/
[9]https://github.com/wenzhixin/bootstrap-table
[10]https://d3js.org/

we intend to follow for the evolution of the platform presented in this article. Here, the web-based tool has been evaluated by a group of 42 users of different expertises that answered to a survey. We report the most important feedback we collected.

The OpenIE module has been evaluated on two benchmarks: the SemEval 2015 Task 12 [11] and SemEval 2016 Task 5 [12] datasets [13]. Both benchmarks required the detection of aspects from text belonging to the *Restaurant* and *Laptop* domains and the computation of the associated polarity. Then, concerning the SemEval 2015 Task 12 dataset, the polarity computation was requested also for the *Hotel* domain. In Section 7.1, we report the results obtained on the aspect detection task. For the SemEval 2015 Task 12 dataset the precision, recall, and f-measure metrics were available, while for the SemEval 2016 Task 5 dataset only the f-measure was reported in the official evaluation report. Then, in Section 7.2, we report the results obtained on the polarity computation task. Here, the system accuracy has been reported. Finally, besides the systems participated to the SemEval challenges, we included also a comparison with other two approaches available in the state of the art that are particularly relevant for our use case. Technical details about such approaches are presented in [3] and [69].

We applied the paired t-test for measuring if the obtained results were significant or not. In each table, we used the following notation near the results obtained by the systems we compared: $--$ and $-$ means that the gap is significantly worse with a p-value of 0.01 or 0.05 respectively. While, $++$ and $+$ means that the gap is significantly better with a p-value of 0.01 or 0.05 respectively.

7.1. Evaluation on Aspect Extraction

Tables 1, 1, and 3 report the results obtained by our system on the SemEval benchmarks. As mentioned above, the algorithm has been tested on both the *Restaurant* and

[11] http://alt.qcri.org/semeval2015/task12/

[12] http://alt.qcri.org/semeval2016/task5/

[13] Concerning the evaluation on the SemEval 2016 Task 5 dataset, we applied our system to the Subtask 1, Slots 2 and 3 only. We worked in this way because the other tasks and slots aimed to verify system facilities that were out of scope of this paper.

Laptop domains.

Table 1: Results obtained on the aspect extraction task, for the *Restaurant* domain on the SemEval 2015 Task 12 benchmark. For each dataset, we reported Precision, Recall, and F-Measure. Acronyms refer to the systems participated in the SemEval 2015 Task 12 competition. Technical details about the participant systems can be found in the SemEval 2015 Proceedings (http://aclweb.org/anthology/S/S15/)

	Restaurant		
System Acronym	**Precision**	**Recall**	**F-Measure**
IHS-RD-Belarus	**0.7095**	0.3845^{--}	0.4987^{--}
LT3 pred	0.5154^{--}	0.5600	0.5367^{--}
NLANGP	0.6385^{-}	**0.6154^{+}**	**0.6268**
sentiue	0.6332^{-}	0.4722^{-}	0.5410^{--}
SIEL	0.6440^{-}	0.5135^{-}	0.5714^{-}
TJUdeM	0.4782^{--}	0.5806	0.5244^{--}
UFRGS	0.6555^{-}	0.4322^{--}	0.5209^{--}
UMDuluth	0.5697^{--}	0.5741^{+}	0.5719^{-}
V3	0.4244^{--}	0.4129^{--}	0.4185^{--}
[3]	0.5795^{--}	0.5287	0.5529^{-}
[69]	0.6182^{-}	0.5329	0.5724^{-}
System Results	0.6895	0.5368	0.6036

The overall performance are in line with the best systems participating in the evaluation campaigns and, in both cases, our approach obtained the best F-measure on the *Laptop* domain. It is also important to highlight that all the systems we compared to, apply a domain-specific supervised approaches for extracting aspects, while our approach implements an unsupervised technique. This peculiarity enables the possibility of implementing the system in any environment without the requirement of training a new model.

While on the *Laptop* domain our system outperforms the others, a different scenario occurs for the *Restaurant* domain where our system loses around 3% and 6% on the two datasets, respectively. A more in depth analysis of the results obtained on the

Table 2: Results obtained on the aspect extraction task, for the *Laptop* domain on the SemEval 2015 Task 12 benchmark. For each dataset, we reported Precision, Recall, and F-Measure. Acronyms refer to the systems participated in the SemEval 2015 Task 12 competition. Technical details about the participant systems can be found in the SemEval 2016 Proceedings (http://aclweb.org/anthology/S/S15/)

	Laptop		
System Acronym	**Precision**	**Recall**	**F-Measure**
IHS-RD-Belarus	0.5548^{--}	0.4483	0.4959^{-}
NLANGP	0.6425^{-}	0.4208	0.5086
sentiue	0.5773^{--}	0.4409	0.5000
TJUdeM	0.4489^{--}	**0.4820**$^{+}$	0.4649^{-}
UFRGS	0.5066^{--}	0.4040	0.4495^{--}
V3	0.2710^{--}	0.2310^{--}	0.2494^{--}
[3]	0.6247^{-}	0.3589^{--}	0.4559^{--}
[69]	0.6412^{-}	0.3773	0.4751^{--}
System Results	**0.6702**	0.4157	**0.5131**

Restaurant datasets highlighted how around the 90% of the errors are caused by the extraction of aspects resulted as false positive. This observation was not unexpected. Indeed, one of the most common issue in unsupervised aspect-based approaches is the extraction of false positive elements [70]. By analyzing possible consequences of this weakness, we suppose that this may lead to a poor effectiveness of components that exploit the outcomes of the aspect extraction module.

Concerning the results of the paired t-test, in the cases where our system did not obtain the best result, we performed the test by taking as reference the best system. While, in the cases where our system obtained the best result, we performed the test by taking as reference the runner-up system. Concerning the results show in Table 1, the precision value has been compared with the *IHS-RD-Belarus* system by resulting not significant (p-value = 0.159), while the recall value has been compared with the *NLANGP* system and in this case the difference resulted significant (p-value = 0.020). Results presented in Table 2 were compared with the *NLANGP* system for the precision value and with the *TJUdeM* system for the recall value. Both differences resulted significant at the t-test with p-values of 0.043 and 0.016, respectively. The same happened for the results shown in Table 3. Here, the comparison has been performed only against the *NLANGP* system and for both domain the improvements resulted significant with p-values of 0.041 and 0.047. Overall, by considering the F-Measure values, our system obtained significant improvement on the *Laptop* domain, while for the *Restaurant* and *Hotel* domains both positive and negative differences with respect to the other systems did not result statistical significant. However, by considering the unsupervised nature of our approach, with respect to the compared systems that are all supervised, we may consider our strategy feasible for being implemented in a real-world general purpose environment.

7.2. Evaluation on Polarity Computation

Tables 4 and 5 report the results obtained on the polarity computation task. The approach has been evaluated on the two benchmarks described in the preamble of this section. Here, we measured the accuracy of the polarity computation algorithm: given the set of opinion words associated with an aspect, the polarity is computed by aggregating the polarity values of each opinion words.

Results demonstrated the effectiveness of the polarity computation strategy implemented into the proposed system. The system obtained the best performance on the *Laptop* domain in both benchmarks, while the gap with the best systems on the other domains is always lower than 2%. After a detailed analysis of the results, we noticed that the reason for which our approach performs better on the *Laptop* dataset is due to the simple language used for describing product features. Indeed, in the *Restaurant* dataset opinions are expressed in a more articulated way and sometimes the approach fails to detect the right polarity. Part of future effort will be dedicated to improve our system in this direction.

We performed the same t-test described in the previous subsection also to the polarity computation results. Concerning the results reported in Table 4, we compared our system with *SENTIUE* on the *Restaurant* and *Laptop* domains, while for the *Hotel* domain the comparison has been done with the *LSISLIF* system. For both the *Restau-*

rant and *Hotel* domains the differences did not result significant (p-values of 0.218 and 0.227, respectively), while for the *Laptop* domain the improvement is significant (p-value of 0.029). Finally, concerning the results shown in Table 5, we compared our system with *XRCE* for the *Restaurant* domain and with *INSIG* for the *Laptop* domain. In the first case the difference was not significant (p-value = 0.189), while in the second case the improvement obtained by our system was significant (p-value = 0.038). Overall, in almost all cases our approach significantly improved the other systems. In particular, on the *Laptop* domain in both cases all improvements are significant for a p-value of at least 0.05.

7.3. Lessons Learned

Early in this section, we demonstrated the suitability of the components integrated within the proposed platform. Besides such validation tasks, we interviewed a group of 42 users for collecting feedback about possible improvements on the client side. In particular, what we collected from users can be recognized in two main aspects: (i) efficient management of data streams, and (ii) understandability of the user interface [14].

Architecture Efficiency. The scenario used in this first prototype focused on using document sets having a limited number of items. By switching from a test environment to a more complex one, we noticed that the time needed for extracting all aspects increased significantly. This issue was related to the necessity of detecting, for each aspect that was already extracted, the presence of further opinions connected to him. While a possible solution might be the parallelization of this task, some tricks have to be applied. Indeed, the constraint of analyzing documents by keeping the timing order in which they have been generated, requires to perform some checks based on the number of documents that we want to analyze at a certain time. Thus, by having, for example, a window of n documents that we want to parallelize, a possible strategy is to verify if there are conflicts between the aspects extracted from such documents. This way,

[14]To Reviewers: here, we did not provided a detailed analysis of the user evaluation of the tool because we considered it out of scope of the paper. However, if the Reviewers consider it useful, we can integrate in a revision more details about how the user evaluation has been conducted

we would be able to safely update aspect-based information without losing potential edges. For completeness, we report data concerning the scalability of the system. We run the scalability test on a server equipped with a double Xeon X5650 and 32Gb of RAM and we measured the time necessary for processing the 1,000,000 documents contained within the DRANZIERA dataset [71]. We tested three systems: the one we propose, the approach presented in [3], and the one presented in [69]. Our system completed the processing operation in 23 minutes and 27 seconds, the algorithm of [3] in 63 minutes and 45 seconds, and the algorithm of [69] in 92 minutes and 31 seconds. Thus, we may state that our system is definitely more efficient.

User Interface Improvement. The second lesson we learned from this work is related to which improvements should be carried out to the user interface for making the platform more appealing from the user's perspective. Users interviewed for judging the tool provided feedback that can be summarized in the following two issues:

- Contextual information into the aspect visualization: in this prototype we did not take into account the possibility of having different kind of users: Basic and Advanced. While basic users can be satisfied from a simple graphical information supporting the detection of the most interesting aspects, advanced users wanted to see detailed information associated with them, i.e. the polarity value, a summary of supporters and opponents associated with each aspect, etc. This functionality will be included in the next version of the platform.

- Animate the evolution of single aspect: the second issue raised by the users was related to the impossibility of observing how each aspect *evolves* during the analysis of the data stream. In particular, a desiderata is the possibility of focusing on a single aspect and to observe how such aspect is judged through time. This feature has been considered as a valuable support for associating peaks of supporters or opponents based on contextual events that cannot be tracked through the proposed system.

The two issues brought to light from users' feedbacks will be used as a starting point for improving the infrastructure of the presented platform. Thus it will be possible to employ such a platform in a larger scale context with the aim of increasing its technological readiness level.

8. Conclusions and Future Work

Results reported in the previous sections revealed the feasibility of the proposed architecture and of the implement techniques. Even if aspect recognition procedure presented in Section 5 may *lack* of precision and recall due to the adopted unsupervised techniques, the results reported in Section 7 shows that the effectiveness of the system is comparable with the supervised systems participated in the SemEval challenges. Thus, few changes in the proposed approach could result in significantly better performances. Examples of actions focus on the improvement of precision that could be achieved by adding a semantic clustering phase in the parsing pipeline shown in Figure 3. Then, by detecting the semantic distances between extracted aspect might help to discard uncorrelated aspects that may not refer to the reviewed product. Recall values could be increased as well by applying less strict rules than the ones presented in Section 5. These possibilities will be taken into account for possible future developments.

Another important part of this work focuses on aspect and opinion polarization. Tables 4 and 5 shown that the presented technique works well during the polarity inference phase. These results suggest that by using an aggregation of (i) general purpose sentiment lexicons and (ii) specific ones, the polarity evaluation phase is positively affected.

Concerning the overall architecture, the presented solution provides a wide range of functionalities that can be applied to provide useful facilities for both customers and producers. For instance, the three level-tree structure shown in Figure 2 can be used to produce both a flexible ranking system and an effective representation of each expressed opinion that can highlight specific qualities and problematics (Figure 7) of each reviewed product.

In the future, efforts will be focused on several different perspectives. The first one concerns the developing of semantic clustering approach for extracting aspects This way, search would be based on inserted words' semantics rather than their syntax. As a result, ranking products by *screen* would also organize *display* and *screen_resolution* aspects rather than discarding them because of their different form.

Other possible progresses regard the application of different aspect extraction techniques on the implemented framework, the refinement of the user interface described in Section 6 and a more detailed comparison between multiple domain-specific lexicon. This last perspective could result in interesting developments concerning the production of a domain-distance metric and the integration of fuzzy membership for unclassified reviews or automatic domain labeling. Once domain-specific lexicons have been produced, they can be used alongside aspect extraction techniques to give a score value to portion of texts which are not provided with that additional information. Such an application could be easily benchmarked with existing reviews.

References

[1] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? sentiment classification using machine learning techniques, in: Proceedings of EMNLP, Association for Computational Linguistics, Philadelphia, 2002, pp. 79–86.

[2] E. Cambria, Affective computing and sentiment analysis, IEEE Intelligent Systems 31 (2) (March-April 2016) 102–107. `doi:10.1109/MIS.2016.31`.
URL `http://dx.doi.org/10.1109/MIS.2016.31`

[3] M. Hu, B. Liu, Mining and summarizing customer reviews, in: W. Kim, R. Kohavi, J. Gehrke, W. DuMouchel (Eds.), Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004, ACM, 2004, pp. 168–177. `doi:10.1145/1014052.1014073`.
URL `http://doi.acm.org/10.1145/1014052.1014073`

[4] B. Liu, L. Zhang, A survey of opinion mining and sentiment analysis, in: C. C. Aggarwal, C. X. Zhai (Eds.), Mining Text Data, Springer, 2012, pp. 415–463.

[5] B. Pang, L. Lee, A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, in: ACL, 2004, pp. 271–278.

[6] N. Jakob, I. Gurevych, Extracting opinion targets in a single and cross-domain setting with conditional random fields, in: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP 2010, 9-11 October 2010, MIT Stata Center, Massachusetts, USA, A meeting of SIGDAT, a Special Interest Group of the ACL, ACL, 2010, pp. 1035–1045.
URL http://www.aclweb.org/anthology/D10-1101

[7] W. Jin, H. H. Ho, R. K. Srihari, Opinionminer: a novel machine learning system for web opinion mining and extraction, in: KDD, 2009, pp. 1195–1204.

[8] B. Liu, M. Hu, J. Cheng, Opinion observer: analyzing and comparing opinions on the web, in: WWW, 2005, pp. 342–351.

[9] Y. Wu, Q. Zhang, X. Huang, L. Wu, Phrase dependency parsing for opinion mining, in: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009, 6-7 August 2009, Singapore, A meeting of SIGDAT, a Special Interest Group of the ACL, ACL, 2009, pp. 1533–1541.
URL http://www.aclweb.org/anthology/D09-1159

[10] Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, Z. Su, Hidden sentiment association in chinese web opinion mining, in: WWW, 2008, pp. 959–968.

[11] A. Go, R. Bhayani, L. Huang, Twitter sentiment classification using distant supervision, CS224N Project Report, Standford University (2009).

[12] L. Barbosa, J. Feng, Robust sentiment detection on twitter from biased and noisy data, in: COLING (Posters), 2010, pp. 36–44.

[13] E. Cambria, A. Hussain, Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis, Springer, 2015.

[14] Q. F. Wang, E. Cambria, C. L. Liu, A. Hussain, Common sense knowledge for handwritten chinese recognition, Cognitive Computation 5 (2) (2013) 234–242.

[15] E. Cambria, A. Hussain, Sentic album: Content-, concept-, and context-based online personal photo management system, Cognitive Computation 4 (4) (2012) 477–496.

[16] A. Gangemi, V. Presutti, D. R. Recupero, Frame-based detection of opinion holders and topics: A model and a tool, IEEE Comp. Int. Mag. 9 (1) (2014) 20–30. doi:10.1109/MCI.2013.2291688.
URL https://doi.org/10.1109/MCI.2013.2291688

[17] D. R. Recupero, V. Presutti, S. Consoli, A. Gangemi, A. G. Nuzzolese, Sentilo: Frame-based sentiment analysis, Cognitive Computation 7 (2) (2015) 211–225. doi:10.1007/s12559-014-9302-z.
URL https://doi.org/10.1007/s12559-014-9302-z

[18] M. Dragoni, Shellfbk: An information retrieval-based system for multi-domain sentiment analysis, in: Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval ’2015, Association for Computational Linguistics, Denver, Colorado, 2015, pp. 502–509.

[19] G. Petrucci, M. Dragoni, An information retrieval-based system for multi-domain sentiment analysis, in: F. Gandon, E. Cabrio, M. Stankovic, A. Zimmermann (Eds.), Semantic Web Evaluation Challenges - Second SemWebEval Challenge at ESWC 2015, Portorož, Slovenia, May 31 - June 4, 2015, Revised Selected Papers, Vol. 548 of Communications in Computer and Information Science, Springer, 2015, pp. 234–243. doi:10.1007/978-3-319-25518-7_20.
URL http://dx.doi.org/10.1007/978-3-319-25518-7_20

[20] E. Riloff, S. Patwardhan, J. Wiebe, Feature subsumption for opinion analysis, in: EMNLP, 2006, pp. 440–448.

[21] T. Wilson, J. Wiebe, R. Hwa, Recognizing strong and weak opinion clauses, Computational Intelligence 22 (2) (2006) 73–99.

[22] M. Dragoni, A. G. Tettamanzi, C. da Costa Pereira, Propagating and aggregating fuzzy polarities for concept-level sentiment analysis, Cognitive Computation 7 (2) (2015) 186–197. doi:10.1007/s12559-014-9308-6.
URL http://dx.doi.org/10.1007/s12559-014-9308-6

[23] M. Dragoni, A. G. B. Tettamanzi, C. da Costa Pereira, A fuzzy system for concept-level sentiment analysis, in: V. Presutti, M. Stankovic, E. Cambria, I. Cantador, A. D. Iorio, T. D. Noia, C. Lange, D. R. Recupero, A. Tordai (Eds.), Semantic Web Evaluation Challenge - SemWebEval 2014 at ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014, Revised Selected Papers, Vol. 475 of Communications in Computer and Information Science, Springer, 2014, pp. 21–27. `doi:10.1007/978-3-319-12024-9_2`.
URL `http://dx.doi.org/10.1007/978-3-319-12024-9_2`

[24] C. da Costa Pereira, M. Dragoni, G. Pasi, A prioritized ”and” aggregation operator for multidimensional relevance assessment, in: R. Serra, R. Cucchiara (Eds.), AI*IA 2009: Emergent Perspectives in Artificial Intelligence, XIth International Conference of the Italian Association for Artificial Intelligence, Reggio Emilia, Italy, December 9-12, 2009, Proceedings, Vol. 5883 of Lecture Notes in Computer Science, Springer, 2009, pp. 72–81. `doi:10.1007/978-3-642-10291-2_8`.
URL `http://dx.doi.org/10.1007/978-3-642-10291-2_8`

[25] A. P. Aprosio, F. Corcoglioniti, M. Dragoni, M. Rospocher, Supervised opinion frames detection with RAID, in: F. Gandon, E. Cabrio, M. Stankovic, A. Zimmermann (Eds.), Semantic Web Evaluation Challenges - Second SemWebEval Challenge at ESWC 2015, Portorož, Slovenia, May 31 - June 4, 2015, Revised Selected Papers, Vol. 548 of Communications in Computer and Information Science, Springer, 2015, pp. 251–263. `doi:10.1007/978-3-319-25518-7_22`.
URL `http://dx.doi.org/10.1007/978-3-319-25518-7_22`

[26] I. Chaturvedi, E. Cambria, D. Vilares, Lyapunov filtering of objectivity for spanish sentiment model, in: 2016 International Joint Conference on Neural Networks, IJCNN 2016, Vancouver, BC, Canada, July 24-29, 2016, IEEE, 2016, pp. 4474–4481. `doi:10.1109/IJCNN.2016.7727785`.
URL `https://doi.org/10.1109/IJCNN.2016.7727785`

[27] S. Poria, E. Cambria, D. Hazarika, P. Vij, A deeper look into sarcastic tweets using deep convolutional neural networks, in: N. Calzolari, Y. Matsumoto, R. Prasad (Eds.), COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan, ACL, 2016, pp. 1601–1612.
URL http://aclweb.org/anthology/C/C16/C16-1151.pdf

[28] S. Poria, E. Cambria, A. F. Gelbukh, Aspect extraction for opinion mining with a deep convolutional neural network, Knowl.-Based Syst. 108 (2016) 42–49. doi:10.1016/j.knosys.2016.06.009.
URL https://doi.org/10.1016/j.knosys.2016.06.009

[29] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Stroudsburg, PA, 2013, pp. 1631–1642.

[30] T. Chen, R. Xu, Y. He, Y. Xia, X. Wang, Learning user and product distributed representations using a sequence model for sentiment analysis, IEEE Comp. Int. Mag. 11 (3) (2016) 34–44. doi:10.1109/MCI.2016.2572539.
URL https://doi.org/10.1109/MCI.2016.2572539

[31] L. Oneto, F. Bisio, E. Cambria, D. Anguita, Statistical learning theory and ELM for big social data analysis, IEEE Comp. Int. Mag. 11 (3) (2016) 45–55. doi:10.1109/MCI.2016.2572540.
URL https://doi.org/10.1109/MCI.2016.2572540

[32] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas, Sentiment in short strength detection informal text, JASIST 61 (12) (2010) 2544–2558. doi:10.1002/asi.21416.
URL http://dx.doi.org/10.1002/asi.21416

[33] T. Fan, C. Chang, Sentiment-oriented contextual advertising, Knowl. Inf. Syst.

23 (3) (2010) 321–344. doi:10.1007/s10115-009-0222-2.
URL http://dx.doi.org/10.1007/s10115-009-0222-2

[34] M. Sklar, K. J. Concepcion, Timely tip selection for foursquare recommendations, in: L. Chen, J. Mahmud (Eds.), Poster Proceedings of the 8th ACM Conference on Recommender Systems, RecSys 2014, Foster City, Silicon Valley, CA, USA, October 6-10, 2014, Vol. 1247 of CEUR Workshop Proceedings, CEUR-WS.org, 2014.
URL http://ceur-ws.org/Vol-1247/recsys14_poster18.pdf

[35] Y. Choi, C. Cardie, Hierarchical sequential learning for extracting opinions and their attributes, in: Proceedings of the ACL 2010 Conference Short Papers, ACLShort '10, Association for Computational Linguistics, Stroudsburg, PA, USA, 2010, pp. 269–274.
URL http://dl.acm.org/citation.cfm?id=1858842.1858892

[36] M. Zhang, Y. Zhang, D. Vo, Neural networks for open domain targeted sentiment, in: L. Màrquez, C. Callison-Burch, J. Su, D. Pighin, Y. Marton (Eds.), Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015, The Association for Computational Linguistics, 2015, pp. 612–621.
URL http://aclweb.org/anthology/D/D15/D15-1073.pdf

[37] M. Mitchell, J. Aguilar, T. Wilson, B. V. Durme, Open domain targeted sentiment, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL, ACL, 2013, pp. 1643–1654.
URL http://aclweb.org/anthology/D/D13/D13-1171.pdf

[38] Q. Mei, X. Ling, M. Wondra, H. Su, C. Zhai, Topic sentiment mixture: Modeling facets and opinions in weblogs, in: Proceedings of the 16th International Conference on World Wide Web, WWW '07, ACM, New York, NY, USA, 2007, pp.

171–180. doi:10.1145/1242572.1242596.
URL http://doi.acm.org/10.1145/1242572.1242596

[39] I. Titov, R. McDonald, A joint model of text and aspect ratings for sentiment summarization, in: PROC. ACL-08: HLT, 2008, pp. 308–316.

[40] F. Li, M. Huang, X. Zhu, Sentiment analysis with global topics and local dependency, in: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010, 2010.
URL http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/view/1913

[41] A. Mukherjee, B. Liu, Aspect extraction through semi-supervised modeling, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1, ACL '12, Association for Computational Linguistics, Stroudsburg, PA, USA, 2012, pp. 339–348.
URL http://dl.acm.org/citation.cfm?id=2390524.2390572

[42] Y. Wu, Q. Zhang, X. Huang, L. Wu, Phrase dependency parsing for opinion mining, in: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3, EMNLP '09, Association for Computational Linguistics, Stroudsburg, PA, USA, 2009, pp. 1533–1541.
URL http://dl.acm.org/citation.cfm?id=1699648.1699700

[43] W. Jin, H. H. Ho, A novel lexicalized HMM-based learning framework for web opinion mining, in: Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09, ACM, New York, NY, USA, 2009, pp. 465–472. doi:10.1145/1553374.1553435.
URL http://doi.acm.org/10.1145/1553374.1553435

[44] M. Dragoni, A. Azzini, A. Tettamanzi, A novel similarity-based crossover for artificial neural network evolution, in: R. Schaefer, C. Cotta, J. Kolodziej, G. Rudolph (Eds.), Parallel Problem Solving from Nature - PPSN XI, 11th International Conference, Kraków, Poland, September 11-15, 2010, Proceedings,

Part I, Vol. 6238 of Lecture Notes in Computer Science, Springer, 2010, pp. 344–353. doi:10.1007/978-3-642-15844-5_35.
URL http://dx.doi.org/10.1007/978-3-642-15844-5_35

[45] A. Yates, M. Cafarella, M. Banko, O. Etzioni, M. Broadhead, S. Soderland, Textrunner: Open information extraction on the web, in: Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, NAACL-Demonstrations '07, Association for Computational Linguistics, Stroudsburg, PA, USA, 2007, pp. 25–26.
URL http://dl.acm.org/citation.cfm?id=1614164.1614177

[46] F. Wu, D. S. Weld, Open information extraction using wikipedia, in: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL '10, Association for Computational Linguistics, Stroudsburg, PA, USA, 2010, pp. 118–127.
URL http://dl.acm.org/citation.cfm?id=1858681.1858694

[47] A. Fader, S. Soderland, O. Etzioni, Identifying relations for open information extraction, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '11, Association for Computational Linguistics, Stroudsburg, PA, USA, 2011, pp. 1535–1545.
URL http://dl.acm.org/citation.cfm?id=2145432.2145596

[48] A. Akbik, A. L"oser, Kraken: N-ary facts in open information extraction, in: Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction, AKBC-WEKEX '12, Association for Computational Linguistics, Stroudsburg, PA, USA, 2012, pp. 52–56.
URL http://dl.acm.org/citation.cfm?id=2391200.2391210

[49] Mausam, M. Schmitz, R. Bart, S. Soderland, O. Etzioni, Open language learning for information extraction, in: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural

Language Learning, EMNLP-CoNLL '12, Association for Computational Linguistics, Stroudsburg, PA, USA, 2012, pp. 523–534.
URL http://dl.acm.org/citation.cfm?id=2390948.2391009

[50] L. Del Corro, R. Gemulla, Clausie: Clause-based open information extraction, in: Proceedings of the 22Nd International Conference on World Wide Web, WWW '13, ACM, New York, NY, USA, 2013, pp. 355–366. doi:10.1145/2488388.2488420.
URL http://doi.acm.org/10.1145/2488388.2488420

[51] H. Bast, E. Haussmann, Open information extraction via contextual sentence decomposition, in: Proceedings of the 2013 IEEE 7th International Conference on Semantic Computing, ICSC '13, IEEE, 2013.

[52] A. Zhila, A. F. Gelbukh, Open information extraction for spanish language based on syntactic constraints., in: ACL (Student Research Workshop), 2014, pp. 78–85.

[53] M. Wang, L. Li, F. Huang, Semi-supervised chinese open entity relation extraction, in: Proceedings of the 3rd IEEE International Conference on Cloud Computing and Intelligence Systems, IEEE, 2014.

[54] T. Falke, G. Stanovsky, I. Gurevych, I. Dagan, Porting an open information extraction system from english to german, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016, 2016, pp. 892–898.
URL http://aclweb.org/anthology/D/D16/D16-1086.pdf

[55] P. Gamallo, M. Garcia, S. Fern'andez-Lanza, Dependency-based open information extraction, in: Proceedings of the Joint Workshop on Unsupervised and Semi-Supervised Learning in NLP, Association for Computational Linguistics, 2012.

[56] P. Gamallo, M. Garcia, Multilingual Open Information Extraction, Springer International Publishing, Cham, 2015, pp. 711–722. doi:10.1007/

978-3-319-23485-4_72.
URL http://dx.doi.org/10.1007/978-3-319-23485-4_72

[57] G. Stanovsky, J. Ficler, I. Dagan, Y. Goldberg, Getting more out of syntax with props, CoRR abs/1603.01648.
URL http://arxiv.org/abs/1603.01648

[58] E. Cambria, S. Poria, R. Bajpai, B. W. Schuller, Senticnet 4: A semantic resource for sentiment analysis based on conceptual primitives, in: COLING, ACL, 2016, pp. 2666–2677.

[59] S. P.J, D. Dunphy, S. Marshall, The General Inquirer: A Computer Approach to Content Analysis, Oxford, England: M.I.T. Press, 1966.

[60] L. Deng, J. Wiebe, MPQA 3.0: An entity/event-level sentiment corpus, in: R. Mihalcea, J. Y. Chai, A. Sarkar (Eds.), NAACL HLT 2015, The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, Colorado, USA, May 31 - June 5, 2015, The Association for Computational Linguistics, 2015, pp. 1323–1328.
URL http://aclweb.org/anthology/N/N15/N15-1146.pdf

[61] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in: LREC, 2010, pp. 2200–2204.

[62] R. Quirk, S. Greenbaum, G. Leech, J. Svartvik, D. Crystal, A comprehensive grammar of the English language, Vol. 397, Cambridge Univ Press, 1985.

[63] C. Fellbaum, WordNet: An Electronic Lexical Database., MIT Press, Cambridge, MA, 1998.

[64] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, D. McClosky, The Stanford CoreNLP natural language processing toolkit, in: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System

Demonstrations, Association for Computational Linguistics, Baltimore, Maryland, 2014, pp. 55–60.
URL http://www.aclweb.org/anthology/P/P14/P14-5010

[65] C. M. Kristina Toutanova, Dan Klein, Y. Singer, Feature-rich part-of-speech tagging with a cyclic dependency network, in: In Proceedings of HLT-NAACL 2003, 2003, pp. 252–259.

[66] K. Clark, C. D. Manning, Entity-centric coreference resolution with model stacking, in: Association for Computational Linguistics (ACL), 2015.

[67] E. W. Hinrichs, D. Roth (Eds.), Accurate Unlexicalized Parsing, 2003.

[68] D. Chen, C. D. Manning, A fast and accurate dependency parser using neural networks, in: Empirical Methods in Natural Language Processing (EMNLP), 2014.

[69] G. Qiu, B. Liu, J. Bu, C. Chen, Opinion word expansion and target extraction through double propagation, Computational Linguistics 37 (1) (2011) 9–27. doi:10.1162/coli_a_00034.
URL http://dx.doi.org/10.1162/coli_a_00034

[70] Q. Liu, Z. Gao, B. Liu, Y. Zhang, Automated rule selection for aspect extraction in opinion mining, in: Q. Yang, M. Wooldridge (Eds.), Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015, AAAI Press, 2015, pp. 1291–1297.
URL http://ijcai.org/Abstract/15/186

[71] M. Dragoni, A. Tettamanzi, C. da Costa Pereira, Dranziera: An evaluation protocol for multi-domain opinion mining, in: N. C. C. Chair), K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), European Language Resources Association (ELRA), Paris, France, 2016.

Table 3: Results obtained on the aspect extraction task, for both the *Restaurant* and *Laptop* domains on the SemEval 2016 Task 5 benchmark. For each dataset, we reported the F-Measure. Acronyms refer to the systems participated in the SemEval 2016 Task 5 competition. Technical details about the participant systems can be found in the SemEval 2016 Proceedings (http://aclweb.org/anthology/S/S16/)

System Acronym	**F-Measure *Restaurant***	**F-Measure *Laptop***
NLANGP	$\mathbf{0.7234}^{+}$	0.5194^{-}
AUEB	0.7044^{+}	0.4911^{--}
UWB	0.6709	0.4790^{--}
GTI	0.6655	n.a
Senti	0.6655	n.a
bunji	0.6488	0.3959^{--}
DMIS	0.6350^{-}	n.a
XRCE	0.6198^{--}	n.a
UWate	0.5707^{--}	n.a
KnowC	0.5682^{--}	n.a
TGB	0.5505^{--}	n.a
BUAP	0.5025^{--}	0.2679^{--}
basel	0.4407^{--}	0.3748^{--}
IHS-R	0.4381^{--}	0.3902^{--}
IIT-T	0.4260^{--}	0.4391^{--}
SeemGo	0.3433^{--}	0.4150^{--}
SYSU	n.a	0.4907^{--}
BUTkn	n.a	0.4840^{--}
NileT	n.a	0.4720^{--}
INSIG	n.a	0.4586^{--}
LeeHu	n.a	0.4375^{--}
UFAL	n.a	0.2698^{--}
CENNL	n.a	0.2691^{--}
[3]	0.6321	0.5141^{-}
[69]	0.6427	0.5187^{-}
System Results	0.6687	**0.5692**

Table 4: Results obtained on the computation of polarities associated with single aspects on the SemEval 2015 Task 12 benchmark. For each dataset, we reported the accuracy obtained in computing polarities (*positive*, *negative*, or *neutral*). Acronyms refer to the systems participated in the SemEval 2015 Task 12 competition. Technical details about the participant systems can be found in the SemEval 2015 Proceedings (http://aclweb.org/anthology/S/S15/)

System Acronym	**Acc. *Restaurant***	**Acc. *Laptop***	**Acc. *Hotel***
ECNU	0.7810	0.7829^{-}	n.a
EliXa	0.7005^{--}	0.7291^{--}	0.7965^{--}
lsislif	0.7550	0.7787^{-}	**0.8584**
LT3	0.7502	0.7376^{--}	0.8053^{-}
sentiue	**0.7869**	0.7934^{-}	0.7876^{--}
SIEL	0.7124^{-}	n.a	n.a
SINAI	0.6071^{--}	0.6585^{--}	0.6372^{--}
TJUdeM	0.6887^{--}	0.7323^{--}	n.a
UFRGS	0.7171^{-}	0.6733^{--}	0.6578^{--}
UMDuluth	0.7112^{-}	n.a	0.7139^{--}
V3	0.6946^{-}	0.6838^{--}	0.7109^{--}
wnlp	0.7136^{-}	0.7207^{--}	0.5546^{--}
[3]	0.6936^{-}	0.7587^{-}	0.7896^{--}
[69]	0.6997^{-}	0.7654^{--}	0.7947^{--}
System Results	0.7794	**0.8589**	0.8524

Table 5: Results obtained on the computation of polarities associated with single aspects on the SemEval 2016 Task 5 benchmark. For each dataset, we reported the accuracy obtained in computing polarities (*positive*, *negative*, or *neutral*). Acronyms refer to the systems participated in the SemEval 2016 Task 5 competition. Technical details about the participant systems can be found in the SemEval 2016 Proceedings (http://aclweb.org/anthology/S/S16/)

System Acronym	**Acc. *Restaurant***	**Acc. *Laptop***
XRCE	**0.8813**	n.a
IIT-T	0.8673	0.7840^{-}
NileT	0.8545	0.7740^{-}
IHS-R	0.8394	0.7790^{-}
ECNU	0.8359^{-}	0.7815^{-}
AUEB	0.8324^{-}	0.7690^{-}
INSIG	0.8207^{-}	0.7840^{-}
UWB	0.8184^{-}	0.7378^{-}
SeemGo	0.8114^{-}	0.7216^{--}
bunji	0.8102^{-}	0.7029^{--}
TGB	0.8091^{--}	n.a
UWate	0.8033^{--}	0.7129^{--}
DMIS	0.7998^{--}	n.a
Senti	0.7811^{--}	0.7428^{-}
LeeHu	0.7811^{--}	0.7591^{-}
basel	0.7648^{--}	0.7004^{--}
AKTSKI	0.7171^{--}	n.a
COMMIT	0.7055^{--}	0.6754^{--}
SNLP	0.6997^{--}	n.a
GTI	0.6997^{--}	0.6729^{--}
CENNL	0.6391^{--}	0.5993^{--}
BUAP	0.6089^{--}	0.6280^{--}
[3]	0.8318^{-}	0.7184^{--}
[69]	0.8162^{-}	0.7458^{-}
System Results	0.8710	**0.8108**