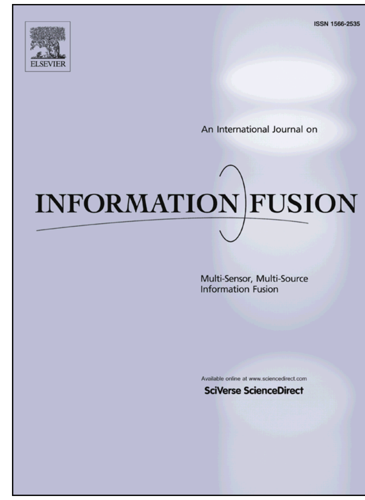
Accepted Manuscript



Opinion Mining and Information Fusion: A Survey

Jorge A. Balazs, Juan D. Velásquez

PII: S1566-2535(15)00053-6
DOI: http://dx.doi.org/10.1016/j.inffus.2015.06.002
Reference: INFFUS 712

To appear in: *Information Fusion*

Please cite this article as: J.A. Balazs, J.D. Velásquez, Opinion Mining and Information Fusion: A Survey, *Information Fusion* (2015), doi: http://dx.doi.org/10.1016/j.inffus.2015.06.002

Opinion Mining and Information Fusion: A Survey

Jorge A. Balazs, Juan D. Velásquez*

Department of Industrial Engineering
Universidad de Chile
Av. República 701 - P.O. Box: 8370439
Santiago, Chile

Abstract

Interest in Opinion Mining has been growing steadily in the last years, mainly because of its great number of applications and the scientific challenge it poses. Accordingly, the resources and techniques to help tackle the problem are many, and most of the latest work fuses them at some stage of the process. However, this combination is usually executed without following any defined guidelines and overlooking the possibility of replicating and improving it, hence the need for a deeper understanding of the fusion process becomes apparent. Information Fusion is the field charged with researching efficient methods for transforming information from different sources into a single coherent representation, and therefore can be used to guide fusion processes in Opinion Mining. In this paper we present a survey on Information Fusion applied to Opinion Mining. We first define Opinion Mining and describe its most fundamental aspects, later explain Information Fusion and finally review several Opinion Mining studies that rely at some point on the fusion of information.

Keywords: Information Fusion, Survey, Opinion Mining, Sentiment Analysis

*Corresponding author. Juan D. Velásquez, Tel.: +56-2-2978-4834 ; fax: +56-2-2689-7895

Email addresses: jabalazs@ug.uchile.cl (Jorge A. Balazs), jvelasqu@dii.uchile.cl (Juan D. Velásquez)

1. Introduction

With the advent of the Web 2.0 and its continuous growth, the amount of freely available user-generated data has reached an unprecedented volume. Being so massive, it is impossible for humans to make sense of its whole in a reasonable amount of time, which is why there has been a growing interest in the scientific community to create systems capable of extracting information from it.

Moreover, the diversity of available data in terms of content, format and extension is huge. Indeed, the data available in microblogs such as Twitter are short and written without much concern for grammar, while review-related data are more extensive and follow stricter grammatical rules [1]. So it is also necessary to bear these differences in mind when attempting to perform any kind of analysis.

In this work, we will focus on two fields charged with dealing with the aforementioned problems, Opinion Mining (OM) and Information Fusion (IF). Opinion Mining (also known as *Sentiment Analysis* [2, 3]) is a sub-field of text mining in which the main task is to extract opinions from content generated by Web users. Opinions play a fundamental role in the decision-making process of both individuals and organizations since they deeply influence people's attitudes and beliefs [4]. Such is the interest in harnessing the power to automatically detect and understand opinions that today this field is one of the most popular areas of research in the Natural Language Processing (NLP) and Computer Science communities, with more than 7000 articles published [5].

To mention some examples, mining opinions enables e-commerce businesses to gain deeper knowledge of their customers and products without having to pay for surveys [6], it allows politicians to understand the political sentiment of the community towards them without having to rely on polls [7], lets companies anticipate their stock trading volumes and financial returns [8], and helps strengthening the deliberation process in the public policy context [9].

Additionally, extracting opinions from reviews, blogs and microblogs, combined with the fusion of different sources of information presents several advantages such as higher authenticity, reduced ambiguity and greater availability [10]. Information Fusion is defined as "the study of efficient methods for automatically or semi-automatically transforming information from different sources and different points in time into a representation that provides effective support for human or automated decision making" [11]. Most of the research in Information Fusion has been done in fields related to the military where data is generated by electronic sensors, however there is growing interest in the fusion of data generated by humans (also called *soft* data) [10, 12].

In this paper we attempt to review the state of the art in Opinion Mining studies that explicitly or implicitly use the fusion of information. Our aim is to provide both new and experienced researchers with insights on how to better perform the fusion process in an Opinion Mining context while also supplying enough information to help them understand both fields separately.

The remainder of this work is structured as follows: In section 2 we show an overview of Opinion Mining by formally defining it, describing the usual process pipeline, explaining the different levels of analysis at which it performs, the different approaches that it uses and the most common challenges it faces. In section 3 we review the state of the art in Opinion Mining combined with Information Fusion and present a simple framework for guiding the fusion process in the Opinion Mining context. Finally, in section 4 we present some of the reviews that have been published both for Opinion Mining and Information Fusion.

2. Opinion Mining

Merriam-Webster's Online Dictionary[1] defines an opinion as a belief, judgement or way of thinking about something. Opinions are formed by the experiences lived by those who hold them. A consumer may look for another's opinion before buying a product or deciding to watch a movie, to gain insights into the potential experiences they would have depending on the decisions they make. Moreover, businesses could benefit from knowing the opinions of their customers by discovering cues on what aspects of a certain service to improve, which features of a determined product are the most valued, or which are new potential business opportunities [13, 14]. In essence, a good Opinion Mining system could eliminate the need for polls and change the way traditional market research is done.

[1] *http://www.merriam-webster.com/dictionary/opinion* (Visited May 11, 2015)

2.1. Definition

Opinion Mining is the field charged with the task of extracting opinions from unstructured text by combining techniques from NLP and Computer Science.

Bing Liu [15] defines an opinion as a 5-tuple containing the target of the opinion (or *entity*), the attribute of the target at which the opinion is directed, the sentiment (or polarity) contained in the opinion which can be positive, negative or neutral, the opinion holder and the date when the opinion was emitted. Formally, an opinion is defined as a tuple:

$$(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$$

where e_i is the i-th opinion target, a_{ij} is the j-th attribute of e_i, h_k is the k-th opinion holder, t_l is the time when the opinion was emitted and s_{ijkl} is the polarity of the opinion towards the attribute a_{ij} of entity e_i by the opinion holder h_k at time t_l.

Note that we described the sentiment contained in an opinion as positive, negative or neutral, notwithstanding it could also be numerically represented. For instance -5 could denote a very negative opinion while 5 a very positive one. Also, in case the analysis did not require much level of detail, the attributes of an entity could be omitted and denoted by $GENERAL$ instead of a_{ij}.

Therefore the main objective of Opinion Mining is to find all the opinion tuples $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ within a document, collection of documents (called *corpus*) or across many corpora. Other works define Opinion Mining as "the task of identifying positive and negative opinions, emotions and evaluations" [16], "the task of finding the opinions of authors about specific entities" [5], "tracking the mood of the public about a particular product or topic" [17], or simply "the task of polarity classification" [18]. These definitions present different scopes and levels of granularity, however all of them can be adapted to fit Liu's opinion model.

There are other approaches, like the one presented in [19], in which the authors attempt to classify emotional states such as "anger", "fear", "joy", or "interest" instead of just positive or negative. In this case, Liu's model could be enriched by adding another element to the opinion tuple model to represent this information.

2.2. Opinion Mining Process: Previous steps

The usual Opinion Mining process or pipeline usually consists of a series of defined steps [20, 21, 22]. These correspond to corpus or data acquisition, text preprocessing, Opinion Mining core process, aggregation and summarization of results, and visualization. In this paper we will give an overview of the first three. Particularly, in this section we will briefly review the two first steps previous to the core OM process: data acquisition and text preprocessing.

2.2.1. Data Acquisition

The first step of any Opinion Mining pipeline is called corpus or data acquisition and consists of obtaining the corpus that is going to be mined for opinions. Currently there are two approaches to achieving this task. The first is through a website's Application Programming Interface (API) being Twitter's[2] one of the most popular [22, 23, 24, 25]. The second corresponds to the use of Web crawlers in order to scrape the data from the desired websites [26, 27, 28]. Olston and Najork portray a robust survey of Web crawling in [29].

Both approaches present some advantages and disadvantages so there is a trade-off between using either. In [30] the authors briefly compare them.

With the API-based approach the implementation is easy, the data gathered is ordered and unlikely to change its structure, however it presents some limitations depending on the provider. For instance search queries to the Twitter REST API are limited to 180 per 15-minute time window.[3] Additionally, the Streaming API has no explicit rate limits for downloading tweets, but is limited in other aspects such as the number of clients from the same IP address connected at the same time, and the rate at which clients are able to read data.[4] This approach is also subject to the availability of an API since not all websites provide one, and even if they do it might not present every needed functionality.

In contrast, crawler-based approaches are more difficult to implement, since the data obtained is noisier and its structure is prone to change, but have the advantage of being virtually unrestricted. Still, using these approaches requires to respect some good etiquette protocols such as the *robots*

[2] *https://dev.twitter.com/rest/public* (Visited May 11, 2015)
[3] *https://dev.twitter.com/rest/public/rate-limiting* (Visited May 11, 2015)
[4] *https://dev.twitter.com/streaming/overview/connecting* (Visited May 11, 2015)

exclusion standard,[5] not issuing multiple overlapping requests to the same server and spacing these requests to prevent putting too much strain on it [29]. Furthermore, Web crawlers can prioritize the extraction of subjective and topically-relevant content. In [31], the authors propose a focused crawler that collects opinion-rich content regarding a particular topic and in [32] this work is further developed by proposing a formal definition for sentiment-based Web crawling along with a framework to facilitate the discovery of subjective content.

2.2.2. Text Preprocessing

The second step in the OM pipeline is Text Preprocessing and is charged with common NLP tasks associated with lexical analysis [33]. Some of the most common techniques are:

Tokenization: task for separating the full text string into a list of separate words. This is simple to perform in space-delimited languages such as English, Spanish or French, but becomes considerably more difficult in languages where words are not delimited by spaces like in Japanese, Chinese and Thai [34].

Stemming: heuristic process for deleting word affixes and leaving them in an invariant canonical form or "stem" [35]. For instance, *person*, *person's*, *personify* and *personification* become *person* when stemmed. The most popular English stemmer algorithm is Porter's stemmer [36].

Lemmatization: algorithmic process to bring a word into its non-inflected dictionary form. It is analogous to stemming but is achieved through a more rigorous set of steps that incorporate the morphological analysis of each word [37].

Stopword Removal: activity for removing words that are used for structuring language but do not contribute in any way to its content. Some of these words are *a*, *are*, *the*, *was* and *will*[6].

[5] *http://www.robotstxt.org/robotstxt.html* (Visited May 11, 2015)

[6] For a more complete list, visit: *http://snowball.tartarus.org/algorithms/english/stop.txt* (Visited May 11, 2015)

Sentence Segmentation: procedure for separating paragraphs into sentences [38]. This step presents its own challenges since periods are often used to mark the ending of a sentence but also to denote abbreviations and decimal numbers [39].

Part-of-Speech (POS) Tagging: is the step of labeling each word of a sentence with its part of speech, such as *adjective*, *noun*, *verb*, *adverb* and *preposition* [40, 41, 42], either to be used as input for further processing like dependency parsing [43] or to be used as features for a machine learning process [44, 45].

Note that all of these steps are not always necessary and have to be selected accordingly for every Opinion Mining application. For example, a machine-learning-based system that relies on a bag-of-words approach will probably use all of the mentioned methods in order to reduce dimensionality and noise [46], while an unsupervised approach might need some of the stopwords' parts of speech to build the dependency rules later used in the Opinion Mining core process [43] therefore omitting the stopword removal process. We present a more detailed analysis of supervised versus unsupervised OM approaches in section 2.3.2.

Moreover, there are other steps that depend heavily on the data source and acquisition method. In particular, data obtained through a Web crawler will have to be processed to remove HTML tags and nontextual information (images and ads) [14, 30, 47], and text extracted from Twitter will need special care for hashtags, mentions, retweets, poorly written text, emoticons, written laughs, and words with repeated characters [46, 48, 49].

2.3. Opinion Mining Process: Core

The third phase in the pipeline is the Opinion Mining core process. In this section we will review the levels of granularity at which it is performed and the different approaches utilized.

2.3.1. Levels of Analysis

Since Opinion Mining began to rise in popularity, the sought-after level of analysis has passed through several stages. First it was performed at the document level where the objective was to find the general polarity of the whole document. Then, the interest shifted to the sentence level and finally to the entity and aspect level. It is worth noting that the analyses that are more fine-grained can be aggregated to form the higher levels. For example an aspect-based Opinion Mining process could simply calculate the average sentiment in a given sentence to produce a sentence-level result.

Document Level: Opinion mining at this level of analysis attempts to classify an opinionated document into positive or negative. The applicability of this level is often limited and usually resides within the context of review analysis [4]. Formally, the objective in the document-level Opinion Mining task can be defined as a modified version of the one presented in section 2.1 and corresponds to finding the tuples:

$$(-, GENERAL, s_{GENERAL}, -, -)$$

where the entity e, opinion holder h, and the time when the opinion was stated t are assumed known or ignored, and the attribute a_j of the entity e corresponds to $GENERAL$. This means that the analysis will only return the generalized polarity of the document. To give a few examples, in [47], Pang and Lee attempted to predict the polarity of movie reviews using three different machine learning techniques: Naïve Bayes, Maximum Entropy classification and Support Vector Machine (SVM). Similarly, in [50] the same authors tried to predict the rating of a movie given in a review, instead of just classifying the review into a positive or negative class.

Sentence Level: This level is analogous to the previous one since a sentence can be considered as a short document. However, it presents the additional preprocessing step consisting of breaking the document into separate sentences, which in turn poses challenges similar to tokenization in languages not delimited by periods. In [51] Riloff and Wiebe used heuristics to automatically label previously unknown data and discover extraction patterns to extract subjective sentences. In [52] the authors achieved high recall and precision (80-90%) for detecting opinions in sentences by using a naïve Bayes classifier and including words, bigrams, trigrams, part-of-speech tags and polarity in the feature set.

Entity and Aspect Level: This represents the most granular level at which Opinion Mining is performed. Here, the task is not only to find the polarity of the opinion but also its target (entity, aspect or both), hence the 5-tuple definition described in section 2.1 fully applies. Both document-level and sentence-level analyses work well when the text being examined contains a single entity and aspect, but they falter when more are present [5].

Aspect-based Opinion Mining attempts to solve this problem by detecting every mentioned aspect in the text and associating them with an opinion.

The earliest work addressing this problem is [6] in which Hu and Liu detect product features (aspects) frequently commented on by customers, then identify the sentences containing opinions, assess their polarity and finally summarize the results. Likewise, in [53] the process to perform the aspect-based Opinion Mining task is to first identify product features, then identify the opinions regarding these features, later estimate their polarity and finally rank them based on their strength.

Marrese-Taylor et al. [54] extend the opinion definition provided by Bing Liu by incorporating *entity expressions* and *aspect expressions* into the analysis. Later they follow the steps of aspect identification, sentiment prediction and summary generation and apply their methodology to the tourism domain by mining opinions from TripAdvisor reviews. They achieved high precision and recall (90%) in the sentiment polarity extraction task but were only able to extract 35% of the explicit *aspect expressions*. In [55], the authors further developed their methodology and integrated it into a modular software that considers all of the previous steps with the addition of a visualization module.

2.3.2. Different Approaches

There are two well-established approaches to carry out the OM core process. One is the unsupervised lexicon-based approach, where the process relies on rules and heuristics obtained from linguistic knowledge [43], and the other is the supervised machine learning approach where algorithms learn underlying information from previously annotated data, allowing them to classify new, unlabeled data [47]. There have also been a growing number of studies reporting the successful combination of both approaches [44, 56, 57]. Furthermore there is an emerging trend that uses ontologies to address the Opinion Mining problem. This is called concept-based Opinion Mining.

Unsupervised Lexicon-based Approaches: Also called semantic-based approaches, attempt to determine the polarity of text by using a set of rules and heuristics obtained from language knowledge. The usual steps to carry them out are first, to mark each word and phrase with its corresponding sentiment polarity with the help of a lexicon, second, to incorporate the analysis of sentiment shifters and their scope (intensifiers and negation), and finally, to handle the adversative clauses (*but-clauses*) by understanding how they affect polarity and reflecting this in the final sentiment score [4]. Later steps could include opinion summarization and visualization.

The first study to tackle Opinion Mining in an unsupervised manner was [58], in which the author created an algorithm that first extracts bigrams abiding certain grammatical rules, then estimates their polarity using the Pointwise Mutual Information (PMI) and finally, computes the average polarity of every extracted bigram to estimate the overall polarity of a review. In [6], Hu and Liu created a list of opinion words using WordNet [59] to later predict the orientation of opinion sentences by determining the prevalent word orientation. Later, in [60], Taboada et al. incorporated the analysis of intensification words (*very*, *a little*, *quite*, *somewhat*) and negation words (*not*) to modify the sentiment polarity of the affected words. In [43], Vilares et al. further incorporated the analysis of syntactic dependencies to better assess the scope of both negation and intensification, and to deal with adversative clauses (given by the adversative conjunction: *but*).

Supervised Learning-based Approaches: Also known as machine-learning-based approaches or statistical methods for sentiment classification, consist of algorithms that learn underlying patterns from example data [61], meaning data whose class or label is known for each instance, to later attempt to classify new unlabeled data [62]. Usually the steps in a machine-learning approach consist of engineering the features to represent the object whose class is to be predicted, and then using its representation as input for the algorithm. Some features frequently used in Opinion Mining are: term frequency, POS tags, sentiment words and phrases, rules of opinion, sentiment shifters and syntactic dependency, among others [4, 44].

In [47] the authors were the first to implement such an approach. They compared the results of using the Naïve Bayes, Maximum Entropy classification and SVM approaches, and found that using unigrams as features (bag-of-words approach) yielded good results.

In [63], Pak and Paroubek rely on Twitter happy and sad emoticons to build a labeled training corpus. They later train three classifier algorithms: Naïve Bayes Classifier, Conditional Random Fields (CRF) and SVM, and find that the first yielded the best results. In [64], Davidov, Tsur and Rappoport in addition to emoticons also use hashtags as labels to train a clustering algorithm similar to k-Nearest Neighbors (kNN) to predict the class of unlabeled tweets.

In [65] the authors attempt to predict sentiment dynamics in the media by using 80 features extracted from tweets with two different machine-learning approaches, Dynamic Language Model (DynamicLM) [66] and a Constrained Symmetric Nonnegative Matrix Factorization (CSNMF) [67], achieving a 79% sentiment prediction accuracy with the latter, whereas only 60% with the former. This is caused mainly because DynamicLM performs better in long texts and tweets are limited to 140 characters.

Concept-based Approaches: These approaches are relatively new and consist of using ontologies for supporting the OM task. An *ontology* is defined as a model that conceptualizes the knowledge of a given domain in a way that is understood by both humans and computers. Ontologies are usually presented as graphs where concepts are mapped to nodes linked by relationships. The study presented in [68] displays a good background study on ontologies, their applications and development. It also describes how the authors incorporated them into an Opinion Mining system to extract text segments containing concepts related to the movie domain to later classify them. In [69], Cambria et al. present a semantic resource for Opinion Mining based on common-sense reasoning and domain-specific ontologies, and describe the steps they took to build it. This resource is improved in [70], where it is enriched with affective information by fusing it with WordNet-Affect [71], another semantic resource, to add emotion labels such as *Anger*, *Disgust*, *Joy* and *Surprise*. In [72], the author presents a new method to classify opinions by combining ontologies with lexical and syntactic knowledge. The work in [73] describes the steps in creating what the authors call a "Human Emotion Ontology" (HEO) which encompasses the domain of human emotions, and shows how this resource can be used to manage affective information related to data issued by online social interaction.

One of the advantages of using unsupervised methods is in not having to rely on large amounts of data for training algorithms, nevertheless it is still necessary to obtain or create a sentiment lexicon. Unsupervised methods are also less domain-dependent than supervised methods. Indeed, classifiers trained in one domain have consistently shown worse performance in other domains [74, 75].

Furthermore it is worth noting that there are several other facets of Opinion Mining that are beyond the scope of this survey such as the lexicon creation problem, comparative opinions, sarcastic sentences, implicit features, cross-lingual adaptation, co-reference resolution, and topic modeling, among others. To get more information on these topics refer to the surveys [1] and [4].

Finally, in Table 1 we provide a brief overview on some of the most popular datasets used for training and validating Opinion Mining systems.

3. Information Fusion applied to Opinion Mining

3.1. An Overview of Information Fusion

Information Fusion has many definitions, indeed some define it as the process of integrating information from multiple sources, others as the process of combining large amounts of dissimilar information into a more comprehensive and easily manageable form. Boström et al. [11] integrate these and several other definitions to create a single and universal one: "Information Fusion is the study of efficient methods for automatically or semi-automatically transforming information from different sources and different points in time into a representation that provides effective support for human or automated decision making." The authors further explain that by "transformation" they mean any kind of combination and aggregation of data. They also state that the sources of data can be of many kinds such as databases, sensors, simulations, or humans, and the data type might also vary (numbers, text, graphics, ontologies).

The benefits of fusing information as opposed to using data from a single source are many. Khalegi et al. [10] compile some of the benefits of applying Information Fusion in the military context and then generalize them to be applied into other fields. The main advantages are increased data authenticity and availability. The first implies improved detection, confidence, reliability and reduction in data ambiguity, and the second means a wider spatial and temporal coverage. In section 3 we will show specific examples issuing from the application of Information Fusion to the OM task.

Another important fact is that Information Fusion deals with two kinds of fusion, the fusion of data generated by electronic sensors, called *hard data*, and data generated by humans, called *soft data* [10]. The main differences between both reside fundamentally in the accuracy, bias, levels of observation and inferences provided by each [108]. A sensor will be better than a human in measuring the velocity of a missile or the electric current passing through a cable, while a human will be better at recognizing relationships between entities and inferring underlying reasons for observed phenomena.

Dataset	References	Languages	Used In	Description
SemEval Twitter Dataset	[76, 77, 78]	English	NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets [79] NRC-Canada-2014: Recent Improvements in the Sentiment Analysis of Tweets [80] UNITN: Training Deep Convolutional Neural Network for Twitter Sentiment Classification [81]	Dataset containing regular and sarcastic tweets, SMS and LiveJournal entries, all of which are tagged with their polarity (positive, negative or neutral).
SemEval Aspect-Based OM Dataset	[82, 83]	English	NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews [84] DLIREC: Aspect Term Extraction and Term Polarity Classification System [85] Sentiue: Target and Aspect based Sentiment Analysis in SemEval-2015 Task 2 [86]	Dataset composed of restaurant and laptop reviews. Each review sentence is tagged with the target of the opinion, its category and the polarity towards it (positive, negative or neutral).
Movie Review Data	[47, 50, 66]	English	Lexicon-Based Methods for Sentiment Analysis [60] Learning to Shift the Polarity of Words for Sentiment Classification [87]	Dataset containing movie reviews tagged at the document level as positive or negative.
OpinRank Dataset	[88]	English	Good Location, Terrible Food: Detecting Feature Sentiment in User-Generated Reviews [89] CONSENTO: A New Framework for Opinion Based Entity Search and Summarization [90]	Dataset containing reviews on cars and hotels. The former are composed of the full textual review and a "favorite" field where each reviewer wrote what he deemed positive about the car. The latter are composed of unlabeled hotel reviews from various major cities, along with TripAdvisor metadata from each hotel such as its overall rating, cleanliness, service and value, among others.
English Product Reviews	[6, 91]	English	Movie Review Mining and Summarization [92]	Corpus composed of several product reviews tagged at the aspect level with the polarity and intensity towards it (-3 very negative; +3 very positive).
Pressrelations Dataset	[93]	German	Integrating viewpoints into newspaper opinion mining for a media response analysis [94]	Dataset containing German news articles tagged at the document level as positive, negative or neutral.
Chinese Product Reviews	[95]	Chinese	Incorporating sentiment prior knowledge for weakly supervised sentiment analysis [96]	Corpora containing Chinese reviews on different products tagged at the document level as positive or negative.
CLEF Replab Dataset	[97, 98]	English, Spanish	LyS at CLEF RepLab 2014: Creating the State of the Art in Author Influence Ranking and Reputation Classification on Twitter [99] LIA@Replab 2014: 10 methods for 3 tasks [100]	Collection of tweets comprising several entities from the automotive, banking, universities and music domains. Each tweet is annotated with a tag showing whether it is related to the entity, a tag with its polarity (positive, negative or neutral), one depicting the topic to which it belongs and another representing the topic's priority.
TASS Corpora	[101, 102, 103]	Spanish	TASS: A Naive-Bayes strategy for sentiment analysis on Spanish tweets [104] Elhuyar at TASS 2013 [105] LyS at TASS 2013: Analysing Spanish tweets by means of dependency parsing, semantic-oriented lexicons and psychometric word-properties [106] LyS at TASS 2014: A Prototype for Extracting and Analysing Aspects from Spanish tweets [107]	Dataset containing Spanish tweets about personalities concerning politics, economy, communication, mass media and culture. Each tweet is tagged with its polarity (very positive, positive, neutral, negative, very negative), both at the global and entity levels Additionally if a tweet does not contain sentiment it is tagged as "NONE." Furthermore, each tweet contains an agreement tag detailing whether its sentiment agrees with its content and, finally, a tag representing the topics to which the tweet belongs. Similar datasets exist exclusively for the political domain and for a discussion concerning a football championship final.

Table 1: Datasets for Opinion Mining

Some of these datasets are available in Kavita Ganesan's Blog,[a] Lillian Lee's homepage[b] and Bing Liu's website.[c]

[a] `http://www.text-analytics101.com/2011/07/user-review-datasets_20.html` (Visited May 28, 2015)
[b] `http://www.cs.cornell.edu/home/llee/data/` (Visited May 28, 2015)
[c] `http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#datasets` (Visited May 28, 2015)

Additionally, most of the research in Information Fusion has been concerned with hard data and very little with soft data [12]. However, the number of roles humans are playing in this field is growing. With the fast expansion of the Web, humans are acting as soft sensors to generate input for traditional fusion systems, and collaborating between them to perform distributed analysis and decision-making processes through multiple digitized mediums (like social media or review sites) [109]. Take a review site like Yelp for instance[7] where users comment on various services such as restaurants, pubs and healthcare, by describing their experiences when using them. Here, each human plays the role of a soft sensor giving its impressions on a given number of aspects of the service, some of which could be quality of service, tastiness of food or overall ambience. By fusing or aggregating their opinions, it would be possible to obtain an accurate depiction of the service being evaluated and its aspects. Hence, aspect-based Opinion Mining could be considered as a form of soft, high-level information fusion.

Furthermore, Khalegi et al. [10] introduce the work done by Kokar et al. [110] as the first step towards a formalization of the theory of information fusion. The proposed framework captures every type of fusion, including data fusion, feature fusion, decision fusion and fusion of relational information. They also state that the most important novelty of the work is that it is able to represent both the fusion of data and the fusion of processing algorithms, and it allows for consistent measurable and provable performance. Finally, Wu and Crestani [111] present a geometric framework for Information Fusion in the context of Information Retrieval. The purpose of this framework is to represent every component in a highly dimensional space so that data fusion can be treated with geometric principles, and the Euclidean Distance can be used as a measure for effectiveness and similarity.

Now that we have explained both Opinion Mining and Information Fusion, we focus on reviewing studies that apply these fields jointly, either explicitly, meaning the authors state that they used Information Fusion techniques, or implicitly, indicating they used some form of fusion without acknowledging it. The remainder of this section is structured similarly to the typical Opinion Mining pipeline described in sections 2.2 and 2.3. We will first review those studies in which the fusion was performed within the data sources, and later those in which it was applied during the main process,

[7] *http://www.yelp.com* (Visited May 11, 2015)

either by fusing lexical resources or techniques from different fields.

3.2. Fusion of Data Sources

The studies that fuse information in this step are those that use raw data from different sources, such as for example, those that combine information coming from tweets and reviews from an e-commerce site.

The work by Shroff et al. [112] presents an "Enterprise Information Fusion" framework that exploits many techniques to provide a better understanding of an enterprise's context, including client feedback and important news about events that could affect it. This framework relies on numerous sources of information for news and feedback, Twitter being the source for the former, and emails, comments on discussion boards and RSS feeds from specific blogs, sources for the latter. They also include the analysis of corporate data to understand how the events and opinions mined from external sources could impact the enterprise's business. To perform the fusion of information they use a "blackboard architecture" described in [113]. Basically, a blackboard system is a belief network in which nodes represent propositions with associated probability distributions and edges denote conditions on the nodes. The authors finally report that they observed a dip in sales of a given product after a raise in negative feedback, and state that even though their analysis was ex post, the mining of unstructured data synchronized with sales data could have provided insights to perform better marketing campaigns and find a better market niche for this product.

Dueñas-Fernández et al. [114] describe a framework for trend modeling based on LDA and Opinion Mining consisting of four steps. The first corresponds to crawling a set of manually-selected seed sources, the second to finding new sources and extracting their topics, the third and fourth to retrieving opinionated documents from social networks for each detected topic and then extracting the opinions from them. They later used a set of 20 different Rich Site Summary (RSS) feeds discussing technology topics as seed documents, and discovered 180 "feasible" feeds utilized for discovering additional information. By mining these newly found feeds, the authors extracted more than 200.000 opinionated tweets and factual documents containing 65 significant events. Finally, they were able to depict the overall polarity of these events over a period of 8 months. All things considered, the authors were able to consistently fuse information from different sources bound together by their topics, which represents a clear example of Information Fusion applied in the data extraction process of an OM application.

3.3. Fusion in the Opinion Mining Core Process

In this section we focus on the studies that fuse either the resources or the techniques necessary to execute the OM core process. By resources, as opposed to the data sources mentioned in section 3.2, we mean knowledge bases that influence the OM process directly. Resources for Opinion Mining consist of lexicons, ontologies, or any annotated corpus.

3.3.1. Fusion of Resources

In this section we review a few of the latest studies that apply the fusion of resources in the OM core process.

In [70] the authors fused two semantic resources to create a richer one. They enhanced the SenticNet resource [69] with affective information from WordNet-Affect (WNA) [71]. To accomplish this task, the authors assigned one of the six WNA emotion labels (*surprise*, *joy*, *sadness*, *anger*, *fear* and *disgust*) to each SenticNet concept. Further, they performed two sets of experiments, one relying only on features based on similarity measures between concepts and another considering these features with the addition of statistical features from the *International Survey of Emotion Antecedents and Reactions* (ISEAR),[8,9]containing statements associated with a particular emotion. They also experimented with three machine learning approaches, Naïve Bayes, Neural Networks and Support Vector Machines, and found the best results when using ISEAR-based features with a SVM. The final product of this work is a new resource that combines polar concepts with emotions.

Hai et al. [115] present a new method to identify opinion features from online reviews by taking advantage of the difference between a domain-specific corpus and a domain-independent one. Their methodology is first to obtain a set of candidate features based on syntactic rules, then compare these candidates with the domain-specific corpus to calculate the *intrinsic-domain relevance* (IDR) and with the domain-independent corpus to obtain the *extrinsic-domain relevance* (EDR). Those candidates with high IDR scores and low EDR scores are accepted as opinion features. Therefore, fusion occurs in the feature-extraction process of the unsupervised Opinion Mining approach, by combining information close to the domain of the review being analyzed,

[8] *http://www.affective-sciences.org/system/files/webpage/ISEAR_0.zip* (Visited May 11, 2015)

[9] *http://www.affective-sciences.org/researchmaterial* (Visited May 11, 2015)

with more general domain-independent information. This allows for obtaining a better estimation of the degree of membership a candidate feature has with the review's domain. Finally by pruning those candidates that are not strongly related to the domain and accepting those with a high degree of relevance, the authors obtain a better set of opinion features.

The work by Xueke et al. [116] exhibits a new methodology to expand sentiment lexicons. The authors propose a generative topic model based in Latent Dirichlet Allocation (LDA) [117], to extract aspect-specific opinion words and their correspondent sentiment polarity. More specifically, their model enriches words from already existing sentiment lexicons by incorporating contextual sentence-level co-occurrences of opinion words under the assumption that usually only one sentiment is present in a sentence. They also compare the performance of their expanded lexicon on three aspect-based Opinion Mining tasks, *implicit aspect identification*, *aspect-based extractive opinion summarization* and *aspect-level sentiment classification*, and find it performs better overall than a non-expanded lexicon. To summarize, the authors found a methodology to fuse the contextual information of a given word with the sentiment prior of said word, thus incorporating new information to it and producing better results.

In [118] the authors present a domain-independent opinion relevance model based on twelve features characterizing the opinion. It is worth noting that the model considers different relevancies of an opinion for different users depending on different parameters. For example, if a certain user is looking for opinions, those authored by a friend will have higher relevance than those of a stranger, since it is natural to consider a friend's opinion as more important. Additional parameters considered to assess the relevance of an opinion are the *author experience*, given by the amount of opinions the author has expressed, *age similarity*, which gives a notion of the differences in age between the opinion author and the opinion consumer, and *interest similarity*, among others. Evidently the more experience, age similarity and interest similarity an author has with a user, the more relevant the opinion will be. The novelty presented in this work is the fact of fusing information concerning the opinion's author and his network of contacts to obtain the opinion relevance metric. This would enable a generic opinion-search engine to provide better search results.

Similarly, the work presented in [119] combines the information given by the activities and relationship networks of the opinion authors to assess the opinion relevance in a social commerce context. The purpose of this analysis is to reflect the honesty, expertise and influence level of the author in the opinion domain. This work, akin to [118], presents a methodology that fuses the information concerning the author's activities and social network with the opinion information in order to estimate its relevance, veracity and objectivity, and to enhance the trust of consumers in providers within an e-commerce setting.

Schuller and Knaup [120] designed a method for Opinion Mining applied to reviews that relies on the combined knowledge of three online resources: The General Inquirer [121], WordNet [59] and ConceptNet [122]. The General Inquirer returns the sentiment valence of a given verb or adjective with 1 corresponding to a positive valence and -1 to a negative valence. If the given word is not found there, they use WordNet to look for synonyms until a match is found. Finally they rely on ConceptNet to identify features toward which the sentiments are directed. All these extracted features are then used as an input for a machine learning algorithm that will classify the review as positive, negative or neutral. Moreover, the authors test the impact of applying early fusion and late fusion methods. Early fusion corresponds simply to the aggregation of scores given by the online knowledge sources as an additional feature for the input feature vector, whereas late fusion corresponds to the combination of the output of several methods on a semantic layer. They found that early fusion yielded a slightly better accuracy and negative recall than the baseline approach at the expense of neutral recall, while late fusion for a given set of parameters, significantly increased accuracy and positive recall at a cost of a significant decrease in negative and neutral recall.

Karamatsis et al. [123] used more than 5 lexicons for creating a system that performs subjectivity detection and polarity classification in social network messages. Each lexicon provides seven features for each message, later used as inputs for a SVM classifier. They tested their system with several datasets containing data from different sources and obtained good results with LiveJournal entries, Twitter messages and sarcastic texts. Likewise, in [80] the authors used features issued from three manually constructed and two automatically generated lexicons. However, in neither work were the lexicons technically combined. The fusion took place in a higher level of abstraction, when the corresponding machine learning algorithms "learned" underlying patterns from features coming from different sources.

3.3.2. Fusion of Techniques

Here we will review some of the studies that combine Opinion Mining techniques with other disciplines.

In [124], the authors jointly extract opinion targets and words by using a word-alignment model. First they find opinion targets and word candidates and later use an *Opinion Relation Graph* to assess their confidence. Finally those candidates with a confidence superior to a certain threshold are accepted as opinion targets/words. The fusion occurs when they use information given by the word-alignment model together with that given by the opinion-relation graphs to find the opinion targets and words. Finally the authors applied their method to three different corpora and found that it outperformed state-of-the-art techniques.

Duan and Zeng [125] propose a method to forecast stock returns by mining opinions from web forums. First they extract the sentiment of a post with a purely lexical approach, meaning they use only a sentiment lexicon to obtain the polarity of sentiment-bearing words, and aggregate their scores as they appear without incorporating syntactic or semantic information. Later they use a Bayesian inference model to predict the stock returns according to the previously obtained sentiments. Here the authors fuse Opinion Mining techniques with stock prediction techniques to obtain better prediction results than those obtained by using purely numerical methods. They also propose to fuse different prediction methods, such as time series, to further improve their model.

Miao et al. [72] merged the product feature extraction and opinion extraction into one single task by using Conditional Random Fields [126]. Later, they "propagated" the found features and opinions by looking for their synonyms and antonyms, and estimated the strength of association between opinion words and product features to generate a domain-specific lexicon. This lexicon is later used to identify the polarity of opinion words in a text by following heuristic rules.

In [127], the authors present an Opinion Mining system that utilizes a supervised machine-learning approach with n-gram and lexicon features. They explicitly state "The main novelty in our system lies not in the individual techniques but rather in the way they are combined and integrated". Certainly, they not only combine four different lexicons (*MPQA* [16], *SentiWord-*

Net [128], *General Inquirer*,[10] and *Bing Liu's Opinion Lexicon*[11,12]) but also present new ways to combine unsupervised semantic-based techniques with supervised machine learning techniques. Specifically, they build a rule-based system which relies only on lexicon information to classify polarity, to later explore different approaches for transforming it into features for the machine-learning algorithm. They report that the combination of both approaches performs better than the systems being implemented separately, and propose to further investigate the individual contribution of each component to the overall system.

Similarly, Rosenthal et al. [129] combined two systems to obtain better results than by using each system individually. The first phrase-based sentiment-detection system relies on lexicon-based knowledge from the *Dictionary of Affect in Language* (DAL) [130], *WordNet* [131], *SentiWordNet* [128] and *Wikitionary* [132]. These and some other features are used as input for a logistic-regression classifier first presented in [133], to obtain the overall polarity of the whole input phrase. The second system uses an emoticon and acronym dictionary, as well as the DAL. The emoticon dictionary contains emoticons labeled as extremely negative, negative, neutral, positive and extremely positive, whereas the acronym dictionary presents the expansions for many internet terms such as *lol* and *fyi*. By using this information they classify the polarity of each tweet. Finally the authors found that the first system had better recall while the second presented higher precision, so they decided to combine both. To implement this they simply created the rule to use the second system when the first presented a precision lower than 70%. With this they achieved better results than when using each system individually.

In [134], Mudinas et al. showcase an Opinion Mining system that integrates both lexicon-based and learning-based techniques. Lexicon-based techniques are used for the detection of common idioms and emoticons, and for the generation of features such as *negations*, *intensifiers*, *sentiment words*, *lexicon-based sentiment scores* and for the detection of new adjectives. Later, learning-based techniques rely on a linear implementation of SVM to measure sentiment polarity. The authors state "The main advantage of our

[10] *http://www.wjh.harvard.edu/~inquirer/inqtabs.txt* (Visited May 11, 2015)

[11] *http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar* (Visited May 11, 2015)

[12] *http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon* (Visited May 11, 2015)

hybrid approach using a lexicon/learning symbiosis, is to attain the best of both worlds," and later specify that they successfully combined the stability and readability from a lexicon with the high accuracy and robustness from a machine-learning algorithm. Their results show that the performance of their system is higher than the state of the art.

Wu et al. [135] propose an Opinion Mining system to evaluate the usability of a given product. After the usual Opinion Mining process they use factor analysis to extract those feature-opinion pairs related to usability. Here, the fusion occurs between the usual lexicon-based OM process and some additional statistical techniques to obtain metrics related to usability.

Table 2 summarizes the papers described in this section and categorizes them according to the type of fusion they display.

Type of Fusion	Study	Year
Fusion of Data Sources	Enterprise Information Fusion for Real-Time Business Intelligence [112]	2011
	A Bayesian Blackboard for Information Fusion [113]	2004
	Detecting Trends on the Web: A Multidisciplinary Approach [114]	2014
Fusion of OM Resources	Enhanced SenticNet With Affective Labels for Concept-Based Opinion Mining [70]	2013
	Identifying Features in Opinion Mining Via Intrinsic and Extrinsic Domain Relevance [115]	2014
	Aspect-Level Opinion Mining of Online Customer Reviews [116]	2013
	A Graph-Based Comprehensive Reputation Model: Exploiting the Social Context of Opinions to Enhance Trust in Social Commerce [119]	2014
	SORM: A Social Opinion Relevance Model [118]	2014
	Learning and Knowledge-Based Sentiment Analysis in Movie Review Key Excerpts [120]	2011
	AUEB: Two Stage Sentiment Analysis of Social Network Messages [123]	2014
	NRC-Canada–2014: Recent Improvements in the Sentiment Analysis of Tweets [80]	2014
Fusion of OM Techniques	Mining Fine Grained Opinions by Using Probabilistic Models and Domain Knowledge [72]	2010
	Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model [124]	2014
	Mining Opinion and Sentiment for Stock Return Prediction Based on Web-Forum Messages [125]	2013
	Aspect-Based Polarity Classification for SemEval Task 4 [127]	2014
	Columbia NLP: Sentiment Detection of Sentences and Subjective Phrases in Social Media [129]	2014
	Combining Lexicon and Learning Based Approached for Concept-Level Sentiment Analysis [134]	2012
	A Novel Approach Based on Review Mining for Product Usability Analysis [135]	2013

Table 2: Summary of papers exemplifying different types of Information Fusion

3.4. A Conceptual Framework for Applying Information Fusion to the Opinion Mining Process

In this section we provide a simple framework for applying Information Fusion techniques to the Opinion Mining pipeline. The most popular fusion model is the one presented by the Joint Directors of Laboratories (JDL) [136], which has been proposed as a fusion model in other fields such as Intrusion Detection [137]. The JDL Fusion Model was originally designed for addressing the combined effects of different levels of abstraction and problem-space complexity, and was divided in 5 levels at which fusion could be performed [137, 138]. Below, these levels are described and linked to the Opinion Mining pipeline depicted in section 2:

Level 0 - Data Refinement: Just as its name suggests, this level deals with data at the lowest level of abstraction by filtering and calibrating them. In the Opinion Mining pipeline, this fusion level would be used while combining different data sources in the Data Acquisition step, as presented in section 3.2. Furthermore, according to Dasarathy's model [139] this step is analogous to *Data In-Data Out Fusion*, meaning data is fed to this level as input and data is received as output. Dueñas-Fernández et al. [114] implicitly executed this step by filtering feeds that did not add valuable information to the process.

Level 1 - Object Refinement: In this level, data must be aligned to a common frame of reference or data structure. This step is the logical successor to level 0, indeed, after having gathered, calibrated and filtered raw data it is necessary to correlate them in order to process them jointly. In the Opinion Mining context this step corresponds to obtaining features from raw text through processes such as POS tagging and lemmatization in the data preprocessing step. This concept is consistent with the *Data In-Feature Out Fusion* presented in Dasarathy's study. For example, if we wanted to align a blog post and a review to a common representation, it would be necessary to depict both types of text according to the features they share, like sentences and the corresponding POS tags of their tokens. In general, this step will be composed of a feature extraction process which will transform data in a set of features, thus allowing to represent different documents in a common frame of reference, such as a vector space [140].

Level 2 - Situation Refinement: This level is executed at a higher level of abstraction, farther from the data and closer to the knowledge. Here, the objects represented as a set of features in a common frame of reference are evaluated according to their coordinated behavior or other high-level attribute. In Dasarathy's model this level corresponds to *Feature In-Feature Out Fusion*. In OM, this step is analogous to the Opinion Mining core process in which features are fed to an algorithm which returns other features such as the target aspects of a given opinion, along with their associated polarity.

Level 3 - Threat Assessment: Here, situation knowledge is used to analyze objects and aggregated groups against *a priori* data to provide an assessment of the current situation and suggest or identify future external conditions. In Dasarathy's model, this type of fusion is called *Feature In-Decision Out Fusion* since refined features are fed to the process and the resulting output corresponds to decisions made either by an expert system or a human at an even higher level of abstraction. For example, a manager could use a summarized opinion report to make better-informed decisions, or alternatively, an expert system could detect a negative trend concerning a specific product and alert those in charge of handling the situation.

Level 4 - Resource Management: In this final stage, the previous levels are further refined by using the information on the current situation and performing a more thorough analysis.

To summarize, level 0 of the JDL could be used to fuse different data sources in the data acquisition step of the Opinion Mining process. Further, level 1 of the JDL model could be used to obtain features from these different data sources and locate them in the same frame of reference in the data preprocessing step. Additionally, a different level 1 process could be used to fuse different sentiment lexicons as in the studies presented in section 3.3.1. Likewise, the OM core process would take the features produced by level 1 and combine them in level 2 of the JDL model by producing opinion-related output. Moreover, both the summarization and visualization step of the OM process correspond to level 3 since they further aggregate the output created by level 2 in order to support decision making by processes in a higher level of abstraction (See Figure 1).

Additionally, in order to categorize the level at which the fusion of a particular set of techniques occurs, a deeper analysis has to be performed since the category will depend on their characteristics. For example, in the work by Duan and Zeng [125] the authors fused the output generated by an OM system and the one produced by a Bayesian inference model in a level of abstraction higher than any of these two, meaning the fusion took place at level 3. Furthermore, Miao et al. [72] merged product feature extraction and opinion extraction into a single process which implies fusion took place at level 2.

Finally, it is worth mentioning that there are other, more complex Information Fusion frameworks, such as the one presented by Kokar et al. [110], that would enable researchers to represent the integration of Information Fusion techniques to Opinion Mining more formally.

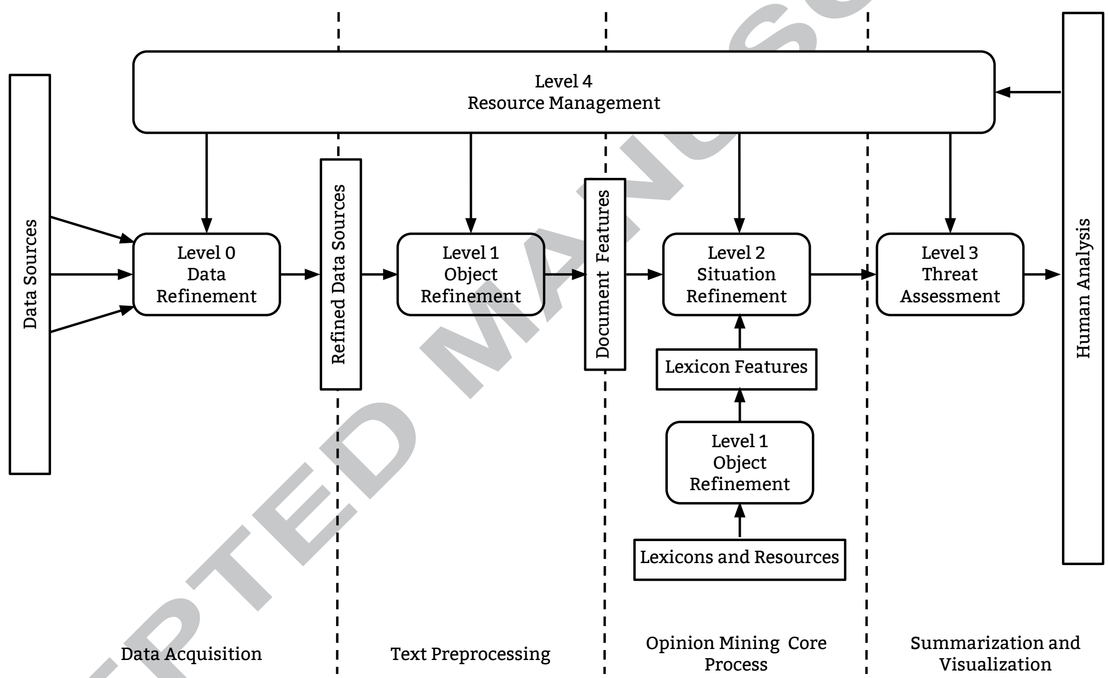


Figure 1: Framework for applying Information Fusion to Opinion Mining

4. Related Work

In this final section we present surveys related to both the Opinion Mining and Information Fusion fields.

4.1. Opinion Mining

There are several surveys that cover Opinion Mining thoroughly. The work by Pang and Lee [1] considers more than 300 publications and presents diverse applications and challenges, as well as the OM problem formulation and the different approaches to solving it. The authors also mention opinion summarization, study the economic implications of reviews and comment on a plethora of publicly available resources.

A more recent review was written by Bing Liu and covers more than 400 studies [4]. Here the author covers the OM subject more exhaustively by defining an opinion model and giving a stricter definition of Sentiment Analysis. He also addresses the different levels at which OM systems are implemented (document, sentence and aspect level), deals with sentiment lexicon generation, opinion summarization, comparative and sarcastic opinions, opinion spam detection, and the quality of reviews, among others.

In [18], Cambria et al., review the Opinion Mining task in general terms, describe its evolution, and discuss the direction the field is taking. In a similar fashion, Feldman [5] describes the task and places greater emphasis on its applications and some of the common issues faced by the research community, such as sarcasm and noisy texts.

More specific OM reviews include the work by Vinodhini and Chandrasekaran [17] in which they cover subjects such as commonly employed Sentiment Analysis data sources as well as different approaches like machine learning and unsupervised learning, or as they call it, "Semantic Orientation approach". They also explain some of the challenges faced in the field such as negation handling and mention some of the applications and tools available. They finish their work by presenting a table comparing different studies, the mining techniques used in them, their feature selection approaches, data sources utilized and performance metrics (accuracy, recall and F-measure).

Khozyainov et al. [141] direct their study towards the difficulties often encountered in OM such as multidimensionality, indirect opinions, bad spelling and grammar, feature interinfluence in feature-based approaches, and the temporal dependency of opinions. Similarly, [142] studies the challenges encountered in developing sentiment analysis tools in the social media context, and covers additional concepts such as relevance, contextual information and volatility over time.

In [143] the authors survey the state of the art in opinion summarization in which they describe the background of Opinion Mining, define a conceptual framework for opinion summarization, and deepen their analysis in aspect-based and non-aspect-based opinion summarization. Finally they discuss how to evaluate summarization methods and mention some of the open challenges in this field.

Martínez-Cámara et al. [144] focus on the latest advancements in Sentiment Analysis as applied to Twitter data. They begin by giving an overview of this microblogging site mentioning some of its sociological aspects as well as the importance of the word of mouth, and later discuss the research concerning polarity classification, temporal prediction of events and political opinion mining. In a similar fashion, Marrese-Taylor et al. [145] present an overview of Opinion Mining, describe some of the most popular sources for extracting opinionated data, discuss summarization and visualization techniques, and finally exhibit an example of a document-level Opinion Mining application for finding the most influential users on Twitter.

Medagoda et al. [146] focus on recent advancements in Opinion Mining achieved in Hindi, Russian and Chinese. Guo et al. [30] define the concept of "Public Opinion Mining," compare different approaches used in each step of the OM pipeline and propose future directions for the field. In [20] the authors propose a faceted characterization of Opinion Mining composed of two main branches, namely *opinion structure* which deals with the relation between unstructured subjective text and structured conceptual elements, and *Opinion Mining tools and techniques* which are the means to achieve the OM task. They also tackle the problems of entity discovery and aspect identification, lexicon acquisition and sarcasm detection. Finally [147] covers some of the usual OM tasks and presents a table similar to the one presented in [17] but instead of using known metrics it just shows an arbitrary "performance" metric without clarifying whether if it represents accuracy, precision, recall, F-measure or some other measure.

Table 3 presents a summary of Opinion Mining reviews presented in this section.

4.2. Information Fusion

One of the most recent surveys on Information Fusion corresponds to the work by Khalegi et al. [10]. In it, the authors focus on reviewing the state of the art in multisensor data fusion. They begin by explaining the potential benefits of implementing an information fusion system and the usual challenges faced while doing so. They also present the work done by Kokar et al. [110] and describe it as one of the first attempts to formally define the Information Fusion theory. They later review the techniques for the fusion of *hard data* (generated by sensors), namely by describing the algorithms used for data fusion in detail, and classifying them according to the challenges they tackle. Finally, the authors mention some of the efforts

Depth	Scope	Study	Year	# Refs.	Main Discussed Topics
Exhaustive	General	Opinion Mining and Sentiment Analysis [1]	2008	333	Different Approaches to OM, OM Applications, OM Challenges, Opinion Summarization, OM Resources, Economic Impact of Product Reviews.
		Sentiment Analysis and Opinion Mining [4]	2012	403	OM at the Three Levels of Granularity, Opinion Summarization, Lexicon Generation, Comparative Opinions, Sarcastic Opinions, Opinion Spam Detection, Quality of Reviews.
	Focused	Comprehensive Review of Opinion Summarization [143]	2011	66	Aspect-based Summarization, Non-Aspect-based Summarization, Topic Modeling, Opinion Visualization, OM Challenges.
		Sentiment Analysis in Twitter [144]	2012	65	Twitter Overview, Twitter Sociological Aspects, Word-of-Mouth Importance, Latest OM Studies Applied to Twitter, Temporal Prediction of Events, Political OM, Open Research Issues.
Brief	General	Techniques and Applications for Sentiment Analysis [5]	2013	40	OM at the Three Levels of Granularity, Comparative Opinions, Lexicon Generation, OM Applications, Open Research Issues.
		Sentiment Analysis and Opinion Mining: A survey [17]	2012	45	Data Sources for OM, Different Approaches to OM, OM Challenges, OM Applications.
		New Avenues in Opinion Mining and Sentiment Analysis [18]	2013	33	OM at the Three Levels of Granularity, Different Approaches to OM Concept-level OM, Multimodal Sentiment Analysis, Future Tendencies.
		A Faceted Characterization of the Opinion Mining Landscape [20]	2014	30	OM at the Three Levels of Granularity, OM Pipeline, Lexicon Generation, Sarcastic Opinions.
		Web Opinion Mining and Sentimental Analysis [145]	2013	25	Data Sources for OM, Document-level OM, Opinion Summarization, Opinion Visualization.
		Opinion Mining and Analysis: A Literature Review [147]	2014	40	Document-level OM, Sentence-level OM, Learning-based approaches to OM, OM Data Sources.
	Focused	A Survey of Internet Public Opinion Mining [30]	2014	47	Data Acquisition, Preprocessing, Topic Modeling, Opinion Tendency, Future Directions.
		Spelling out Opinions: Difficult Cases of Sentiment Analysis [141]	2013	20	Available Tools for OM, Opinion Characteristics, OM Challenges.
		Challenges in Developing Opinion Mining tools for Social Media [142]	2012	34	Specific Challenges for Applying OM in a Social Media Context (Relevance Target Identification, Negation, Context and Volatility)
		A Comparative Analysis of Opinion Mining and Sentiment Classification in Non-English Languages [146]	2013	20	Different Approaches to OM, Latest OM Studies in Hindi, Russian and Chinese.

Table 3: Summary of Opinion Mining Reviews.

Reviews are categorized either as *Exhaustive* or *Brief*, the former meaning surveys cover their main topics in a thorough way, while the latter implies they just mention the topic and explain it briefly. Furthermore, *General* reviews are those that present Opinion Mining as a whole whereas *Focused* reviews focus on a particular Opinion Mining sub-topic. Finally, # *Refs.* represent the amount of studies cited by each survey (references).

made towards the fusion of *soft data* (generated by humans) and the new tendency of attempting to fuse them with hard data.

General surveys include the work by Bloch [148], in which she compares and classifies the different operators used to combine the data gathered by multiple sensors in information fusion systems. She classifies these operators as "Context Independent Constant Behavior Operators (CICB)", "Context Independent Variable Behavior Operators (CIVB)" and "Context Dependent Operators (CD)," and describe the theory underlying each one of them. Furthermore, Hall et al. [149] review both the military and non-military applications for Information Fusion, describe a data fusion process model and some of the architectures for data fusion (Centralized, Autonomous and Hybrid Fusion). Additionally, Smith et al. [150] comment on several methods for target tracking through sensor data fusion. The authors structure their work according to the Joint Directors of Laboratories (JDL) model [136] by reviewing the advancements for each one of its levels: object refinement, situation assessment, threat assessment and process assessment.

More specific studies include the survey by Wache et al. [151] in which the authors review the use of ontologies for the fusion of data issued from different sources. Specifically, they define the role of ontologies, their representations, the use of mappings designed to integrate them into the fusion systems and their engineering process. In [152] the authors introduce the concept of *reliability* and discuss the theory and approaches for incorporating it into common IF operators. They define reliability coefficients as the measure of how well each belief model represents reality. Yao et al. [153] define "Web Information Fusion" as the task of combining all kinds of information on the Web. They give an overview of the advances in this field by reviewing some of the contributions made to it by the Artificial Intelligence (AI) and database communities to it. Furthermore, they comment on the role that ontologies and the "Semantic Web" play in Web Information Fusion.

Additionally, there are other surveys reviewing the application of Information Fusion in specific fields. The work in [154] presents the state of the art in image fusion. The authors begin by describing this field, then review its history, categorize the most common image fusion algorithms into low, mid and high level, describe some of the applications, and finish by mentioning some emerging technologies and future directions for the field. Corona et al. [155] review the state of the art of Information Fusion applied to computer security. They first define computer security as the quantitative evaluation of three qualities of an information flow: availability, confidentiality and integrity. They then describe the intrusion-detection problem, state that it corresponds to a pattern recognition task and define the role Information Fusion plays in it. Later, the authors present a high-level framework for information fusion, comment on the current applications, and finish by proposing a new approach for data fusion in computer security. Faouzi et al. [156] provide a survey of the application of Information Fusion in different areas of Intelligent Transport Systems (ITS). First, they describe the background on data fusion, secondly, they enumerate the opportunities and challenges of ITS Information Fusion, and finally review the applications in which IF is applied to ITS. In [157] the authors review the role of IF in data privacy [158]. They begin by defining data privacy, next they comment on several protection methods used in the literature, such as *microaggregation* which provides privacy by clustering data and representing it as the clusters' centroids, and *record linkage* which in the context of data privacy represents a way to provide disclosure risk assessment of protected data. The authors also demonstrate how both of these methods are greatly benefited from the use of Information Fusion. Finally, Sun et al. [159] exhibit a survey on multi-source domain adaptation, in which they comment on the latest advancements concerning the problem of adapting training data to test data from a different domain. Their work includes the review of algorithms, theoretical results and the discussion on open problems and future work.

The Information Fusion reviews described in this section are summarized in Table 4.

5. Conclusions

In this paper we presented a short survey of the most popular Opinion Mining techniques, defined the Information Fusion field, proposed a simple framework for guiding the fusion process in an Opinion Mining system and reviewed some of the studies that have successfully implemented Information Fusion techniques in the Opinion Mining context. Indeed, the future of Opinion Mining relies on creating better and deeper sources of knowledge, which can be achieved by fusing already existing knowledge bases such as ontologies and lexicons. Nevertheless, few studies have done so by explicitly applying well-established techniques. In fact, studies in which authors fuse different lexical resources or techniques without following any standard procedure are the most common.

Depth	Scope	Study	Year	# Refs.	Main Discussed Topics
Exhaustive	General	Multisensor Data Fusion [10]	2013	197	Multisensor Data Fusion, Hard-Data Fusion, Soft-Data Fusion, IF Challenges, IF Algorithms, Emerging Paradigms, Open Research Issues.
		Formalizing Classes of Information Fusion Systems [110]	2004	36	Fusion Definition, IF Theory, Multi-Source IF, Single-Source IF, Effectiveness of IF Systems
		Information Combination Operators for Data Fusion: A Comparative Review with Classification [148]	1996	20	Multi-Source IF, Behavior of IF Operators, Image Fusion.
		An Introduction to Multisensor Data Fusion [149]	1997	130	IF Terminology, Military IF Applications, Non-Military IF Applications, JDL Model, IF Process Model, IF Architectures.
		Approaches to Multisensor Data Fusion in Target Tracking: A survey [150]	2006	195	JDL Model Stages (Object Refinement, Situation Assessment, Threat Assessment, Process Assessment), Multisensor-Tracking Challenges.
	Focused	Ontology-based Integration of Information: A Survey of Existing Approaches [151]	2001	60	Role of Ontologies, Representation of Ontologies, Creation of Ontologies, Use of Mappings to Integrate Ontologies and IF Systems.
	Applications	Information Fusion for Computer Security: State of the Art and Open Issues [155]	2009	59	Computer Security, Intrusion Detection, JDL Model, Current IF Applications in Computer Security, Proposal of new IF approach applied to Computer Security, Open Research Issues.
		Data Fusion in Intelligent Transportation Systems: Progress and Challenges – A Survey [156]	2011	94	IF Approaches, Opportunities of Applying IF to ITS, Challenges of Applying IF to ITS, Current IF Applications in ITS, Future Directions.
		Information Fusion in Data Privacy: A Survey [157]	2012	139	Data Privacy Basic Concepts, Data Protection Approaches, IF Applied to Data Privacy, Record Linkage.
		A Survey of Multi-Source Domain Adaptation [159]	2014	43	Domain Adaptation Basic Concepts, Domain Adaptation Algorithms, IF Applied to Multi-Source Domain Adaptation, Datasets for Domain Adaptation, Open Research Issues.
Brief	General	Data Fusion Lexicon [136]	1991	N/A	Data Fusion Terms, JDL Model Proposition.
	Focused	Reliability in Information Fusion: Literature Survey [152]	2004	40	*Reliability* Definition, Incorporating Reliability into IF Operators, Reliability Coefficients, Reliability of Fusion Results.
		Web Information Fusion: A Review of the State of the Art [153]	2008	33	The Web, IF Overview, Ontologies, Semantic Web, Relationship Between IF and the Web, Web-Based Support Systems.
	Applications	Image Fusion: Advances in the State of the Art [154]	2007	66	Image Fusion Basic Concepts, Image-Fusion-Algorithms Classification, Image Registration, Image Fusion Applications, Emerging Image Fusion Technologies, Future Directions.

Table 4: Summary of Information Fusion Reviews.

The categories for *Depth* and *Scope* are equal to those presented in Table 3, with the addition of *Applications*, which represents those surveys that review the latest advancements of Information Fusion applied to a specific field.

However, even if a fusion process does not follow a strict framework, the results of applying it are consistently better than not doing so. From this it follows that both fields could greatly benefit from a more standardized and consistent way to fuse opinion-related data. This is why the knowledge generated in the Information Fusion field becomes essential. Broadening the knowledge on soft fusion for instance, would facilitate the fusion of data from different online sources such as Twitter and review sites, increasing its authenticity and availability, which would in turn allow the production of higher-quality Opinion Mining systems. Furthermore, advancements in the fusion of soft data with hard data would make possible the combination of audiovisual content with textual data and push forward the Multimodal Sentiment Analysis field [18].

Admittedly, using Information Fusion jointly with Opinion Mining would allow for a better understanding of the effects of every fused component in the final system while enabling researchers to improve the fusion process and ultimately lay the foundations for creating better systems.

Acknowledgements

The authors would like to acknowledge the continuous support of the Chilean Millennium Institute of Complex Engineering Systems (ICM: P-05-004-F, CONICYT: FBO16). This work was partially funded by *Corporación de Fomento de la Producción* (CORFO) under project number 13IDL2-223170 entitled *OpinionZoom* (`www.opinionzoom.cl`).

References

[1] B. Pang, L. Lee, Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval 2 (1-2) (2008) 1–135. `doi:10.1561/1500000011`.

[2] P. A. Tapia, J. D. Velásquez, Twitter sentiment polarity analysis: A novel approach for improving the automated labeling in a text corpora, in: Active Media Technology, Springer International Publishing, 2014, pp. 274–285.

[3] G. L. Rodrigo Dueńas-Fernández, J. D. Velásquez, Sentiment polarity of trends on the web using opinion mining and topic modeling, in: Proceedings of the First Workshop On Social Web Intelligence (WOSWI)

in conjunction with 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology, WI-IAT Í13, IEEE Computer Society, Washington, DC, USA, 2013.

[4] B. Liu, Sentiment analysis and opinion mining, Synthesis Lectures on Human Language Technologies 5 (1) (2012) 1–167. doi:10.2200/S00416ED1V01Y201204HLT016.

[5] R. Feldman, Techniques and applications for sentiment analysis, Communications of the ACM 56 (4) (2013) 82–89. doi:10.1145/2436256.2436274.

[6] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2004, pp. 168–177. doi:10.1145/1014052.1014073.

[7] A. Tumasjan, T. O. Sprenger, P. G. Sandner, I. M. Welpe, Predicting elections with Twitter: What 140 characters reveal about political sentiment, in: Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM 2010), Vol. 10, AAAI Press, Washington, USA, 2010, pp. 178–185.
URL https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1441

[8] W. Zhang, S. Skiena, Trading strategies to exploit blog and news sentiment, in: Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM 2012), Vol. 10, AAAI Press, Washington, USA, 2010, pp. 375–378.
URL http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1529

[9] J. D. Velásquez, P. González, Expanding the possibilities of deliberation: The use of data mining for strengthening democracy with an application to education reform, The Information Society 26 (1) (2010) 1–16. doi:10.1080/01972240903423329.

[10] B. Khaleghi, A. Khamis, F. O. Karray, S. N. Razavi, Multisensor data fusion: A review of the state-of-the-art, Information Fusion 14 (1) (2013) 28–44. doi:10.1016/j.inffus.2011.08.001.

[11] H. Boström, S. F. Andler, M. Brohede, R. Johansson, A. Karlsson, J. Van Laere, L. Niklasson, M. Nilsson, A. Persson, T. Ziemke, On the definition of information fusion as a field of research, Tech. Rep. HS-IKI-TR-07-006, Informatics Research Centre, University of Skövde (2007).
URL http://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-1256

[12] K. Sambhoos, J. Llinas, E. Little, Graphical methods for real-time fusion and estimation with soft message data, in: Proceedings of the 11th International Conference on Information Fusion (FUSION 2008), IEEE, Cologne, Germany, 2008, pp. 1–8.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4632405

[13] G. L. Urban, J. R. Hauser, "Listening in" to find and explore new combinations of customer needs, Journal of Marketing 68 (2) (2004) 72–87. doi:10.1509/jmkg.68.2.72.27793.

[14] O. Netzer, R. Feldman, J. Goldenberg, M. Fresko, Mine your own business: Market-structure surveillance through text mining, Marketing Science 31 (3) (2012) 521–543. doi:10.1287/mksc.1120.0713.

[15] B. Liu, Sentiment analysis and subjectivity, in: R. Dale, H. Moisl, H. Somers (Eds.), Handbook of natural language processing, 2nd Edition, Machine Learning and Pattern Recognition, CRC Press, New York, NY, USA, 2010, pp. 627–666.

[16] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, in: Proceedings of the Joint Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP 2005), Association for Computational Linguistics, Vancouver, Canada, 2005, pp. 347–354. doi:10.3115/1220575.1220619.

[17] G. Vinodhini, R. Chandrasekaran, Sentiment analysis and opinion mining: a survey, International Journal of Advanced Research in Computer Science and Software Engineering 2 (6) (2012) 282–292.

[18] E. Cambria, B. Schuller, Y. Xia, C. Havasi, New avenues in opinion mining and sentiment analysis, IEEE Intelligent Systems 28 (2) (2013) 15–21. doi:10.1109/MIS.2013.30.

[19] A. Neviarouskaya, H. Prendinger, M. Ishizuka, Recognition of fine-grained emotions from text: An approach based on the compositionality principle, in: T. Nishida, L. C. Jain, C. Faucher (Eds.), Modeling Machine Emotions for Realizing Intelligence, Vol. 1 of Smart Innovation, Systems and Technologies, Springer Berlin Heidelberg, 2010, pp. 179–207. doi:10.1007/978-3-642-12604-8_9.

[20] R. Arora, S. Srinivasa, A faceted characterization of the opinion mining landscape, in: Proceedings of the 6th International Conference on Communication Systems and Networks (COMSNETS 2014), IEEE, Bangalore, India, 2014, pp. 1–6. doi:10.1109/COMSNETS.2014.6734936.

[21] L. Dey, S. M. Haque, Opinion mining from noisy text data, International Journal on Document Analysis and Recognition (IJDAR) 12 (3) (2009) 205–226. doi:10.1007/s10032-009-0090-z.

[22] F. H. Khan, S. Bashir, U. Qamar, TOM: Twitter opinion mining framework using hybrid classification scheme, Decision Support Systems 57 (2014) 245–257. doi:10.1016/j.dss.2013.09.004.

[23] A. Bakliwal, J. Foster, J. van der Puil, R. O'Brien, L. Tounsi, M. Hughes, Sentiment analysis of political tweets: Towards an accurate classifier, in: Proceedings of the NAACL Workshop on Language Analysis in Social Media (LASM 2013), Association for Computational Linguistics, Atlanta, GA, USA, 2013, pp. 49–58.
URL http://doras.dcu.ie/19962/

[24] L. Jiang, M. Yu, M. Zhou, X. Liu, T. Zhao, Target-dependent Twitter sentiment classification, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011), Vol. 1, Association for Computational Linguistics, Portland, OR, USA, 2011, pp. 151–160.
URL http://dl.acm.org/citation.cfm?id=2002472.2002492

[25] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, Sentiment analysis of Twitter data, in: Proceedings of the Workshop on Languages in Social Media (LSM 2011), Association for Computational Linguistics, Portland, OR, USA, 2011, pp. 30–38.
URL http://dl.acm.org/citation.cfm?id=2021109.2021114

[26] B. Krishnamurthy, P. Gill, M. Arlitt, A few chirps about Twitter, in: Proceedings of the 1st Workshop on Online Social Networks (WOSN 2008), ACM, Seattle, WA, USA, 2008, pp. 19–24. doi:10.1145/1397735.1397741.

[27] Y. H. Gu, S. J. Yoo, Rules for mining comparative online opinions, in: Proceedings of the 4th International Conference on Computer Sciences and Convergence Information Technology (ICCIT 2009), IEEE, Seoul, Korea, 2009, pp. 1294–1299. doi:10.1109/ICCIT.2009.16.

[28] X. Hu, J. Tang, H. Gao, H. Liu, Unsupervised sentiment analysis with emotional signals, in: Proceedings of the 22nd International Conference on World Wide Web (WWW 2013), International World Wide Web Conferences Steering Committee, Rio de Janeiro, Brazil, 2013, pp. 607–618.
URL http://dl.acm.org/citation.cfm?id=2488388.2488442

[29] C. Olston, M. Najork, Web crawling, Foundations and Trends in Information Retrieval 4 (3) (2010) 175–246.
URL http://dl.acm.org/citation.cfm?id=1734789

[30] K. Guo, L. Shi, W. Ye, X. Li, A survey of internet public opinion mining, in: Proceedings of the International Conference on Progress in Informatics and Computing (PIC 2014), IEEE, Shanghai, China, 2014, pp. 173–179. doi:10.1109/PIC.2014.6972319.

[31] T. Fu, A. Abbasi, D. Zeng, H. Chen, Sentimental spidering: leveraging opinion information in focused crawlers, ACM Transactions on Information Systems 30 (4) (2012) 24:1–24:30. doi:10.1145/2382438.2382443.

[32] A. G. Vural, Sentiment-focused Web crawling, Ph.D. thesis, Middle East Technical University (2013).
URL http://etd.lib.metu.edu.tr/upload/12616409/index.pdf

[33] A. Hippisley, Lexical analysis, in: R. Dale, H. Moisl, H. Somers (Eds.), Handbook of natural language processing, 2nd Edition, Machine Learning and Pattern Recognition, CRC Press, New York, NY, USA, 2010, pp. 31–58.

[34] D. Palmer, Text preprocessing, in: R. Dale, H. Moisl, H. Somers (Eds.), Handbook of natural language processing, 2nd Edition, Machine Learning and Pattern Recognition, CRC Press, New York, NY, USA, 2010, pp. 9–30.

[35] B. Liu, Web data mining: exploring hyperlinks, contents, and usage data, 1st Edition, Springer Berlin Heidelberg, New York, NY, USA, 2007, Ch. 6.

[36] M. F. Porter, An algorithm for suffix stripping, Program 40 (3) (2006) 211–218. doi:10.1108/00330330610681286.

[37] C. D. Manning, P. Raghavan, H. Schütze, Introduction to information retrieval, Vol. 1, Cambridge University Press, 2008, Ch. 2.

[38] T. Kiss, J. Strunk, Unsupervised multilingual sentence boundary detection, Computational Linguistics 32 (4) (2006) 485–525. doi:10.1162/coli.2006.32.4.485.

[39] S. Bird, E. Klein, E. Loper, Natural language processing with Python, 1st Edition, O'Reilly Media, Inc., 2009.

[40] E. Brill, A simple rule-based part of speech tagger, in: Proceedings of the Workshop on Speech and Natural Language, Association for Computational Linguistics, Harriman, New York, 1992, pp. 112–116. doi:10.3115/1075527.1075553.

[41] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, N. A. Smith, Part-of-speech tagging for Twitter: Annotation, features, and experiments, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011), Vol. 2, Association for Computational Linguistics, Portland, Oregon, 2011, pp. 42–47.
URL http://dl.acm.org/citation.cfm?id=2002736.2002747

[42] T. Güngör, Part-of-speech tagging, in: R. Dale, H. Moisl, H. Somers (Eds.), Handbook of natural language processing, 2nd Edition, Machine Learning and Pattern Recognition, CRC Press, New York, NY, USA, 2010, pp. 9–30.

[43] D. Vilares, M. A. Alonso, C. Gómez-Rodríguez, A syntactic approach for opinion mining on Spanish reviews, Natural Language Engineering 21 (01) (2015) 139–163. doi:10.1017/S1351324913000181.

[44] M. Joshi, C. Penstein-Rosé, Generalizing dependency features for opinion mining, in: Proceedings of the 4th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2009), Association for Computational Linguistics, Suntec, Singapore, 2009, pp. 313–316. URL http://dl.acm.org/citation.cfm?id=1667583.1667680

[45] J. D. Velásquez, V. Palade, Adaptive Web Sites, Vol. 170, IOS Press, 2008.

[46] V. Hangya, R. Farkas, Target-oriented opinion mining from tweets, in: Proceedings of the 4th International Conference on Cognitive Infocommunications (CogInfoCom 2013), IEEE, Budapest, Hungary, 2013, pp. 251–254. doi:10.1109/CogInfoCom.2013.6719251.

[47] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), Vol. 10, Association for Computational Linguistics, Philadelphia, PA, USA, 2002, pp. 79–86. doi:10.3115/1118693.1118704.

[48] D. Vilares, M. Á. Alonso, C. Gómez-Rodríguez, Supervised polarity classification of Spanish tweets based on linguistic knowledge, in: Proceedings of the 13th Symposium on Document Engineering (DocEng 2013), ACM, Florence, Italy, 2013, pp. 169–172. doi:10.1145/2494266.2494300.

[49] A. Bakliwal, P. Arora, S. Madhappan, N. Kapre, M. Singh, V. Varma, Mining sentiments from tweets, in: Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2012), Association for Computational Linguistics, Jeju, Korea, 2012, pp. 11–18. URL http://dl.acm.org/citation.cfm?id=2392963.2392970

[50] B. Pang, L. Lee, Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, in: Proceedings

of the 43rd Annual Meeting on Association for Computational Linguistics (ACL 2005), Association for Computational Linguistics, Ann Arbor, MI, USA, 2005, pp. 115–124. doi:10.3115/1219840.1219855.

[51] E. Riloff, J. Wiebe, Learning extraction patterns for subjective expressions, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2003), Association for Computational Linguistics, Sapporo, Japan, 2003, pp. 105–112. doi:10.3115/1119355.1119369.

[52] H. Yu, V. Hatzivassiloglou, Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2003), Association for Computational Linguistics, Sapporo, Japan, 2003, pp. 129–136. doi:10.3115/1119355.1119372.

[53] A.-M. Popescu, O. Etzioni, Extracting product features and opinions from reviews, in: A. Kao, S. R. Poteet (Eds.), Natural Language Processing and Text Mining, Springer London, 2007, pp. 9–28. doi:10.1007/978-1-84628-754-1_2.

[54] E. Marrese-Taylor, J. D. Velásquez, F. Bravo-Marquez, Y. Matsuo, Identifying customer preferences about tourism products using an aspect-based opinion mining approach, Procedia Computer Science 22 (2013) 182–191. doi:10.1016/j.procs.2013.09.094.

[55] E. Marrese-Taylor, J. D. Velásquez, F. Bravo-Marquez, Opinion Zoom: A modular tool to explore tourism opinions on the Web, in: Proceedings of the the IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT 2013), IEEE, Atlanta, GA, USA, 2013, pp. 261–264. doi:10.1109/WI-IAT.2013.193.

[56] 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), Vol. 3, Association for Computational Linguistics, Singapore, 2009, pp. 1533–1541. URL http://dl.acm.org/citation.cfm?id=1699648.1699700

[57] T. Nakagawa, K. Inui, S. Kurohashi, Dependency tree-based sentiment classification using CRFs with hidden variables, in: Proceedings of The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2010), Association for Computational Linguistics, Los Angeles, CA, USA, 2010, pp. 786–794.
URL `http://dl.acm.org/citation.cfm?id=1857999.1858119`

[58] P. D. Turney, Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, in: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL 2002), Association for Computational Linguistics, Philadelphia, PA, USA, 2002, pp. 417–424. `doi:10.3115/1073083.1073153`.

[59] G. A. Miller, WordNet: A lexical database for English, Communications of the ACM 38 (11) (1995) 39–41. `doi:10.1145/219717.219748`.

[60] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Lexicon-based methods for sentiment analysis, Computational Linguistics 37 (2) (2011) 267–307. `doi:10.1162/COLI_a_00049`.

[61] V. L. Rebolledo, G. L'Huillier, J. D. Velásquez, Web pattern extraction and storage, in: Advanced Techniques in Web Intelligence-I, Springer Berlin Heidelberg, 2010, pp. 49–77.

[62] T. Mitchell, Machine Learning, McGraw-Hill, 1997.

[63] A. Pak, P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, in: Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010), Vol. 10, European Language Resources Association, Valletta, Malta, 2010, pp. 1320–1326.
URL `http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf`

[64] D. Davidov, O. Tsur, A. Rappoport, Enhanced sentiment learning using Twitter hashtags and smileys, in: Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010), Association for Computational Linguistics, Beijing, China, 2010, pp.

241–249.
URL http://dl.acm.org/citation.cfm?id=1944566.1944594

[65] L. T. Nguyen, P. Wu, W. Chan, W. Peng, Y. Zhang, Predicting collective sentiment dynamics from time-series social media, in: Proceedings of the 1st International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM 2012), no. 6, ACM, Beijing, China, 2012, pp. 6:1–6:8. doi:10.1145/2346676.2346682.

[66] B. Pang, L. Lee, A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, in: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-2004), Association for Computational Linguistics, Barcelona, Spain, 2004, pp. 271–278. doi:10.3115/1218955.1218990.

[67] W. Peng, D. H. Park, Generate adjective sentiment dictionary for social media sentiment analysis using constrained nonnegative matrix factorization, in: Proceedings of the 5th International Conference on Weblogs and Social Media (ICWSM 2011), AAAI Press, Barcelona, Spain, 2011, pp. 273–280.
URL http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2723

[68] L. Zhou, P. Chaovalit, Ontology-supported polarity mining, Journal of the American Society for Information Science and Technology 59 (1) (2008) 98–110. doi:10.1002/asi.20735.

[69] E. Cambria, R. Speer, C. Havasi, A. Hussain, SenticNet: A publicly available semantic resource for opinion mining, in: Proceedings of the Fall Symposium on Computational Models of Narrative, AAAI, Arlington, VA, USA, 2010, pp. 14–18.
URL https://www.aaai.org/ocs/index.php/FSS/FSS10/paper/view/2216

[70] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, S. Bandyopadhyay, Enhanced SenticNet with affective labels for concept-based opinion mining, IEEE Intelligent Systems 28 (2) (2013) 31–38. doi:10.1109/MIS.2013.4.

[71] C. Strapparava, A. Valitutti, et al., WordNet Affect: An affective extension of WordNet, in: In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), Vol. 4, European Language Resources Association, Lisbon, Portugal, 2004, pp. 1083–1086.
URL http://www.lrec-conf.org/proceedings/lrec2004/pdf/369.pdf

[72] Q. Miao, Q. Li, D. Zeng, Mining fine grained opinions by using probabilistic models and domain knowledge, in: Proceedings of the IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technologies (WI-IAT 2010), Vol. 1, IEEE, Toronto, Canada, 2010, pp. 358–365. doi:10.1109/WI-IAT.2010.193.

[73] M. Grassi, E. Cambria, A. Hussain, F. Piazza, Sentic Web: A new paradigm for managing social media affective information, Cognitive Computation 3 (3) (2011) 480–489. doi:10.1007/s12559-011-9101-8.

[74] A. Aue, M. Gamon, Customizing sentiment classifiers to new domains: A case study, in: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2005), Borovets, Bulgaria, 2005.
URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.90.3612

[75] J. Blitzer, M. Dredze, F. Pereira, Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification, in: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL 2007), Association for Computational Linguistics, Prague, Czech Republic, 2007, pp. 440–447.
URL http://www.aclweb.org/anthology/P07-1056

[76] P. Nakov, S. Rosenthal, Z. Kozareva, V. Stoyanov, A. Ritter, T. Wilson, SemEval-2013 Task 2: Sentiment analysis in Twitter, in: Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013), Association for Computational Linguistics, Atlanta, GA, USA, 2013, pp. 312–320.
URL http://www.aclweb.org/anthology/S13-2052

[77] S. Rosenthal, A. Ritter, P. Nakov, V. Stoyanov, SemEval-2014 Task 9: Sentiment analysis in Twitter, in: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014, pp. 73–80.
URL https://www.aclweb.org/anthology/S/S14/S14-2009.pdf

[78] S. Rosenthal, P. Nakov, S. Kiritchenko, S. Mohammad, A. Ritter, V. Stoyanov, SemEval-2015 Task 10: Sentiment analysis in Twitter, in: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado, 2015, pp. 451–463.
URL http://alt.qcri.org/semeval2015/cdrom/pdf/SemEval078.pdf

[79] S. Mohammad, S. Kiritchenko, X. Zhu, NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets, in: Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013), Association for Computational Linguistics, Atlanta, GA, USA, 2013, pp. 321–327.
URL http://www.aclweb.org/anthology/S13-2053

[80] X. Zhu, S. Kiritchenko, S. M. Mohammad, NRC-Canada-2014: Recent improvements in the sentiment analysis of tweets, in: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014, pp. 443–447.
URL http://www.aclweb.org/anthology/S14-2077

[81] A. Severyn, A. Moschitti, UNITN: Training deep convolutional neural network for Twitter sentiment classification, in: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado, 2015, pp. 464–469.
URL http://alt.qcri.org/semeval2015/cdrom/pdf/SemEval079.pdf

[82] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, Semeval-2014 task 4: Aspect based sentiment

analysis, in: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014, pp. 27–35.
URL http://www.aclweb.org/anthology/S14-2004

[83] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, I. Androutsopoulos, SemEval-2015 Task 12: Aspect based sentiment analysis, in: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado, 2015, pp. 486–495.
URL http://alt.qcri.org/semeval2015/cdrom/pdf/SemEval082.pdf

[84] S. Kiritchenko, X. Zhu, C. Cherry, S. Mohammad, NRC-Canada-2014: Detecting aspects and sentiment in customer reviews, in: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014, pp. 437–442.
URL http://www.aclweb.org/anthology/S14-2076

[85] Z. Toh, W. Wang, DLIREC: Aspect term extraction and term polarity classification system, in: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014, pp. 235–240.
URL http://www.aclweb.org/anthology/S14-2038

[86] J. Saias, Sentiue: Target and aspect based sentiment analysis in semeval-2015 task 12, in: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado, 2015, pp. 767–771.
URL http://alt.qcri.org/semeval2015/cdrom/pdf/SemEval130.pdf

[87] D. Ikeda, H. Takamura, L.-A. Ratinov, M. Okumura, Learning to shift the polarity of words for sentiment classification, in: Proceedings of the 3rd International Joint Conference on Natural Language Processing (IJCNLP 2008), Vol. 1, Association for Computational Linguistics, Hyderabad, India, 2008, pp. 296–303.
URL https://www.aclweb.org/anthology/I/I08/I08-1039.pdf

[88] K. Ganesan, C. Zhai, Opinion-based entity ranking, Information Retrieval 15 (2) (2011) 116–150. doi:10.1007/s10791-011-9174-8.

[89] M. Cataldi, A. Ballatore, I. Tiddi, M.-A. Aufaure, Good location, terrible food: detecting feature sentiment in user-generated reviews, Social Network Analysis and Mining 3 (4) (2013) 1149–1163. doi:10.1007/s13278-013-0119-7.

[90] J. Choi, D. Kim, S. Kim, J. Lee, S. Lim, S. Lee, J. Kang, Consento: A new framework for opinion based entity search and summarization, in: Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM 2012), ACM, Maui, Hawaii, USA, 2012, pp. 1935–1939. doi:10.1145/2396761.2398547.

[91] X. Ding, B. Liu, P. S. Yu, A holistic lexicon-based approach to opinion mining, in: Proceedings of the 1st International Conference on Web Search and Data Mining (WSDM 2008), ACM, Stanford, CA, USA, 2008, pp. 231–240. doi:10.1145/1341531.1341561.

[92] L. Zhuang, F. Jing, X.-Y. Zhu, Movie review mining and summarization, in: Proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM 2006), ACM, Arlington, Virginia, USA, 2006, pp. 43–50. doi:10.1145/1183614.1183625.

[93] T. Scholz, S. Conrad, L. Hillekamps, Opinion mining on a german corpus of a media response analysis, in: P. Sojka, A. Horák, I. Kopeček, K. Pala (Eds.), Text, Speech and Dialogue, Vol. 7499 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2012, pp. 39–46. doi:10.1007/978-3-642-32790-2_4.

[94] T. Scholz, S. Conrad, Integrating viewpoints into newspaper opinion mining for a media response analysis, in: Proceedings of the 11th Conference on Natural Language Processing (KONVENS 2012), Vienna, Austria, 2012, pp. 30–38.
URL http://www.oegai.at/konvens2012/proceedings/08_scholz12o/08_scholz12o.pdf

[95] T. Zagibalov, C. John, Automatic seed word selection for unsupervised sentiment classification of Chinese text, in: Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008),

Manchester, UK, 2008, pp. 1073–1080.
URL http://www.aclweb.org/anthology/C/C08/C08-1135.pdf

[96] Y. He, Incorporating sentiment prior knowledge for weakly supervised sentiment analysis, ACM Transactions on Asian Language Information Processing (TALIP) 11 (2). doi:10.1145/2184436.2184437.

[97] E. Amigó, J. Carrillo de Albornoz, I. Chugur, A. Corujo, J. Gonzalo, T. Martín, E. Meij, M. de Rijke, D. Spina, Overview of RepLab 2013: Evaluating online reputation monitoring systems, in: Proceedings of the 4th International Conference of the CLEF Initiative (RepLab 2013), Valencia, Spain, 2013.
URL http://ceur-ws.org/Vol-1179/CLEF2013wn-RepLab-AmigoEt2013.pdf

[98] E. Amigó, J. Carrillo-de Albornoz, I. Chugur, A. Corujo, J. Gonzalo, E. Meij, M. de Rijke, D. Spina, Overview of RepLab 2014: Author profiling and reputation dimensions for online reputation management, in: Proceedings of the 5th International Conference of the CLEF Initiative (RepLab 2014), Sheffield, UK, 2014, pp. 1438–1457.
URL http://ceur-ws.org/Vol-1180/CLEF2014wn-Rep-AmigoEt2014.pdf

[99] D. Vilares, M. Hermo, M. A. Alonso, C. Gómez-Rodríguez, J. Vilares, LyS at CLEF RepLab 2014: Creating the state of the art in author influence ranking and reputation classification on Twitter, in: Proceedings of the 5th International Conference of the CLEF Initiative (RepLab 2014), Sheffield, UK, 2014, pp. 1468–1478.
URL http://ceur-ws.org/Vol-1180/CLEF2014wn-Rep-VilaresEt2014.pdf

[100] J.-V. Cossu, K. Janod, E. Ferreira, J. Gaillard, M. El-Bèze, LIA@Replab 2014: 10 methods for 3 tasks, in: Proceedings of the 5th International Conference of the CLEF Initiative (RepLab 2014), Sheffield, UK, 2014, pp. 1458–1467.
URL http://ceur-ws.org/Vol-1180/CLEF2014wn-Rep-CossuEt2014.pdf

[101] J. Villena Román, S. Lana Serrano, E. Martínez Cámara, J. C. González Cristóbal, TASS-Workshop on sentiment analysis at SEPLN,

Procesamiento del Lenguaje Natural 50 (2013) 37–44.
URL http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/4657/2759

[102] J. Villena Román, J. García Morera, S. Lana Serrano, J. C. González Cristóbal, TASS 2013 - A second step in reputation analysis in Spanish, Procesamiento del Lenguaje Natural 52 (2014) 37–44.
URL http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/4901/2915

[103] J. Villena Román, E. Martínez Cámara, J. García Morera, S. M. Jiménez Zafra, TASS 2014 - The challenge of aspect-based sentiment analysis, Procesamiento del Lenguaje Natural 54 (2015) 61–68.
URL http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5095/2976

[104] P. Gamallo, M. García, S. Fernández-Lanza, TASS: A naive-bayes strategy for sentiment analysis on Spanish tweets, in: Proceedings of the Workshop on Sentiment Analysis at SEPLN (TASS 2013), 2013, pp. 126–132.
URL http://www.daedalus.es/TASS2013/papers/tass2013-submission1-CITIUS-CILENIS.pdf

[105] X. Saralegi Urizar, I. San Vicente Roncal, Elhuyar at TASS 2013, in: Proceedings of the Workshop on Sentiment Analysis at SEPLN (TASS 2013), 2013, pp. 143–150.
URL http://www.daedalus.es/TASS2013/papers/tass2013-submission3-Elhuyar.pdf

[106] D. Vilares, M. A. Alonso, C. Gómez-Rodrıguez, LyS at TASS 2013: Analysing Spanish tweets by means of dependency parsing, semantic-oriented lexicons and psychometric word-properties, in: Proceedings of the Workshop on Sentiment Analysis at SEPLN (TASS 2013), 2013, pp. 179–186.
URL http://www.daedalus.es/TASS2013/papers/tass2013-submission8-LYS.pdf

[107] D. Vilares, Y. Doval, M. A. Alonso, C. Gómez-Rodríguez, LyS at TASS 2014: A prototype for extracting and analysing aspects from Spanish tweets, in: Proceedings of the Workshop on Sentiment Analysis at

SEPLN (TASS 2014), 2013.
URL http://www.daedalus.es/TASS2014/papers/2.LyS.pdf

[108] D. L. Hall, M. McNeese, J. Llinas, T. Mullen, A framework for dynamic hard/soft fusion, in: Proceedings of the 11th International Conference on Information Fusion (FUSION 2008), IEEE, Cologne, Gemrany, 2008, pp. 1–8.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4632196

[109] D. L. Hall, M. D. McNeese, D. B. Hellar, B. J. Panulla, W. Shumaker, A cyber infrastructure for evaluating the performance of human centered fusion, in: Proceedings of the 12th International Conference on Information Fusion, IEEE, Seattle, WA, USA, 2009, pp. 1257–1264.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5203798

[110] M. M. Kokar, J. A. Tomasik, J. Weyman, Formalizing classes of information fusion systems, Information Fusion 5 (3) (2004) 189–202. doi:10.1016/j.inffus.2003.11.001.

[111] S. Wu, F. Crestani, A geometric framework for data fusion in information retrieval, Information Systems 50 (2015) 20–35. doi:10.1016/j.is.2015.01.001.

[112] G. Shroff, P. Agarwal, L. Dey, Enterprise information fusion for realtime business intelligence, in: Proceedings of the 14th International Conference on Information Fusion (FUSION 2011), IEEE, Chicago, IL, USA, 2011, pp. 1–8.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5977516

[113] C. Sutton, C. T. Morrison, P. R. Cohen, J. Moody, J. Adibi, A Bayesian blackboard for information fusion, in: Proceedings of the 7th International Conference on Information Fusion (FUSION 2004), Vol. 2, International Society of Information Fusion, Stockholm, Sweden, 2004, pp. 1111–1116.
URL http://www.fusion2004.foi.se/papers/IF04-1111.pdf

[114] R. Dueñas-Fernández, J. D. Velásquez, G. L'Huillier, Detecting trends on the web: A multidisciplinary approach, Information Fusion 20 (2014) 129–135. doi:10.1016/j.inffus.2014.01.006.

[115] Z. Hai, K. Chang, J.-J. Kim, C. C. Yang, Identifying features in opinion mining via intrinsic and extrinsic domain relevance, IEEE Transactions on Knowledge and Data Engineering 26 (3) (2014) 623–634. doi:10.1109/TKDE.2013.26.

[116] X. Xueke, C. Xueqi, T. Songbo, L. Yue, S. Huawei, Aspect-level opinion mining of online customer reviews, China Communications 10 (3) (2013) 25–41. doi:10.1109/CC.2013.6488828.

[117] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, Journal of Machine Learning Research 3 (2003) 993–1022.
URL http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

[118] A. D. S. Lima, J. S. Sichman, SORM: A social opinion relevance model, in: Proceedings of the IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technologies (WI-IAT 2014), Vol. 1, IEEE, Warsaw, Poland, 2014, pp. 78–85. doi:10.1109/WI-IAT.2014.19.

[119] S.-R. Yan, X.-L. Zheng, Y. Wang, W. W. Song, W.-Y. Zhang, A graph-based comprehensive reputation model: Exploiting the social context of opinions to enhance trust in social commerce, Information Sciences (2014). doi:10.1016/j.ins.2014.09.036.

[120] B. Schuller, T. Knaup, Learning and knowledge-based sentiment analysis in movie review key excerpts, in: A. Esposito, A. M. Esposito, R. Martone, V. Mller, G. Scarpetta (Eds.), Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces. Theoretical and Practical Issues, Vol. 6456 of Lecture Notes in Computer Science, Springer, 2011, pp. 448–472. doi:10.1007/978-3-642-18184-9_39.

[121] P. J. Stone, D. C. Dunphy, M. S. Smith, The General Inquirer: A Computer Approach to Content Analysis, MIT press, Oxford, England, 1966.

[122] C. Havasi, R. Speer, J. Alonso, ConceptNet 3: A flexible, multilingual semantic network for common sense knowledge, in: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2007), Borovets, Bulgaria, 2007.
URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.115.2844

[123] R. M. Karampatsis, J. Pavlopoulos, P. Malakasiotis, AUEB: Two stage sentiment analysis of social network messages, in: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014, pp. 114–118.
URL http://www.aclweb.org/anthology/S14-2015

[124] K. Liu, L. Xu, J. Zhao, Co-extracting opinion targets and opinion words from online reviews based on the word alignment model, IEEE Transactions on Knowledge and Data Engineering 27 (3) (2015) 636–650. doi:10.1109/TKDE.2014.2339850.

[125] J. Duan, J. Zeng, Mining opinion and sentiment for stock return prediction based on web-forum messages, in: Proceedings of the 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2013), IEEE, Shenyang, China, 2013, pp. 984–988. doi:10.1109/FSKD.2013.6816338.

[126] J. Lafferty, A. McCallum, F. C. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, in: Proceedings of the 18th International Conference on Machine Learning (ICML 2001), Morgan Kaufmann Publishers Inc., Williamstown, MA, USA, 2001, pp. 282–289.
URL http://dl.acm.org/citation.cfm?id=645530.655813

[127] J. Wagner, P. Arora, S. Cortes, U. Barman, D. Bogdanova, J. Foster, L. Tounsi, DCU: Aspect-based polarity classification for semeval task 4, in: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014, pp. 223–229.
URL http://www.aclweb.org/anthology/S14-2036

[128] S. Baccianella, A. Esuli, F. Sebastiani, SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in: Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010), Vol. 10, European Language Resources Association, Valletta, Malta, 2010, pp. 2200–2204.
URL http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf

[129] S. Rosenthal, A. Agarwal, K. McKeown, Columbia NLP: Sentiment detection of sentences and subjective phrases in social media, in: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014, pp. 198–202.
URL http://www.aclweb.org/anthology/S14-2031

[130] C. M. Whissell, The dictionary of affect in language, Emotion: Theory, research, and experience 4 (113-131) (1989) 94.

[131] C. Fellbaum (Ed.), WordNet: An electronic lexical database, MIT Press, 1998.

[132] S. Rosenthal, K. McKeown, Columbia NLP: Sentiment detection of subjective phrases in social media, in: Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013), Association for Computational Linguistics, Atlanta, GA, USA, 2013, pp. 478–482.
URL http://www.aclweb.org/anthology/S13-2079

[133] A. Agarwal, F. Biadsy, K. R. Mckeown, Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams, in: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009), Association for Computational Linguistics, Athens, Greece, 2009, pp. 24–32.
URL http://dl.acm.org/citation.cfm?id=1609067.1609069

[134] A. Mudinas, D. Zhang, M. Levene, Combining lexicon and learning based approaches for concept-level sentiment analysis, in: Proceedings of the 1st International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM 2012), no. 5, ACM, Beijing, China, 2012, pp. 5:1–5:8. doi:10.1145/2346676.2346681.

[135] M. Wu, L. Wang, L. Yi, A novel approach based on review mining for product usability analysis, in: Proceedings of the 4th International Conference on Software Engineering and Service Science (ICSESS 2013), IEEE, Beijing, China, 2013, pp. 942–945. doi:10.1109/ICSESS.2013.6615461.

[136] F. E. White, Data fusion lexicon, Tech. rep., Joint Directors of Laboratories, Data Fusion Sub-Panel, Naval Ocean Systems Center (1991). URL http://www.dtic.mil/dtic/tr/fulltext/u2/a529661.pdf

[137] T. Bass, Intrusion detection systems and multisensor data fusion, Communications of the ACM 43 (4) (2000) 99–105. doi:10.1145/332051.332079.

[138] J. Llinas, C. Bowman, G. Rogova, A. Steinberg, E. Waltz, F. White, Revisiting the JDL data fusion model II, in: Proceedings of the 7th International Conference on Information Fusion (FUSION 2004), Stockholm, Sweden, 2004, pp. 1218–1230.
URL http://www.fusion2004.foi.se/papers/IF04-1218.pdf

[139] B. V. Dasarathy, Sensor fusion potential exploitation–Innovative architectures and illustrative applications, Proceedings of the IEEE 85 (1) (1997) 24–38. doi:10.1109/5.554206.

[140] C. D. Manning, P. Raghavan, H. Schütze, Introduction to information retrieval, Vol. 1, Cambridge University Press, 2008, Ch. 6.

[141] I. Khozyainov, E. Pyshkin, V. Klyuev, Spelling out opinions: Difficult cases of sentiment analysis, in: Proceedings of the International Joint Conference on Awareness Science and Technology and Ubi-Media Computing (iCAST-UMEDIA 2013), IEEE, Aizu-Wakamatsu, Japan, 2013, pp. 231–237. doi:10.1109/ICAwST.2013.6765439.

[142] D. Maynard, K. Bontcheva, D. Rout, Challenges in developing opinion mining tools for social media, in: Proceedings of the LREC workshop @NLP can u tag #usergeneratedcontent?! (LREC 2012), Istambul, Turkey, 2012, pp. 15–22.
URL http://lrec-conf.org/proceedings/lrec2012/workshops/21.LREC2012%20NLP4UGC%20Proceedings.pdf#20

[143] H. D. Kim, K. Ganesan, P. Sondhi, C. Zhai, Comprehensive review of opinion summarization, Tech. rep., University of Illinois at Urbana-Champaign (2011).
URL http://hdl.handle.net/2142/18702

[144] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Urena-López, A. Montejo-Ráez, Sentiment analysis in Twitter, Natural Language Engineering 20 (01) (2014) 1–28. doi:10.1017/S1351324912000332.

[145] E. Marrese-Taylor, C. Rodríguez, J. D. Velásquez, G. Ghosh, S. Banerjee, Web opinion mining and sentimental analysis, in: Advanced Techniques in Web Intelligence-2, Studies in Computational Intelligence, Springer Berlin Heidelberg, 2013, pp. 105–126. doi:10.1007/978-3-642-33326-2_5.

[146] N. Medagoda, S. Shanmuganathan, J. Whalley, A comparative analysis of opinion mining and sentiment classification in non-English languages, in: Proceedings of the International Conference on Advances in ICT for Emerging Regions (ICTer 2013), IEEE, Colombo, Sri Lanka, 2013, pp. 144–148. doi:10.1109/ICTer.2013.6761169.

[147] V. Singh, S. K. Dubey, Opinion mining and analysis: A literature review, in: Proceedings of the 5th International Conference – Confluence The Next Generation Information Technology Summit (CONFLUENCE 2014), IEEE, Noida, India, 2014, pp. 232–239. doi:10.1109/CONFLUENCE.2014.6949318.

[148] I. Bloch, Information combination operators for data fusion: a comparative review with classification, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 26 (1) (1996) 52–67. doi:10.1109/3468.477860.

[149] D. L. Hall, J. Llinas, An introduction to multisensor data fusion, Proceedings of the IEEE 85 (1) (1997) 6–23. doi:10.1109/5.554205.

[150] D. Smith, S. Singh, Approaches to multisensor data fusion in target tracking: A survey, IEEE Transactions on Knowledge and Data Engineering 18 (12) (2006) 1696–1710. doi:10.1109/TKDE.2006.183.

[151] H. Wache, T. Voegele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, S. Hübner, Ontology-based integration of information: a

survey of existing approaches, in: Proceedings of the IJCAI Workshop on Ontologies and Information Sharing (IJCAI 2001), International Joint Conferences on Artificial Intelligence, Seattle, WA, USA, 2001, pp. 108–117.
URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.12.8073

[152] G. L. Rogova, V. Nimier, Reliability in information fusion: literature survey, in: Proceedings of the 7th International Conference on Information Fusion (FUSION 2004), Vol. 2, International Society of Information Fusion, Stockholm, Sweden, 2004, pp. 1158–1165.
URL http://www.fusion2004.foi.se/papers/IF04-1158.pdf

[153] J. Yao, V. V. Raghavan, Z. Wu, Web information fusion: a review of the state of the art, Information Fusion 9 (4) (2008) 446–449. doi:10.1016/j.inffus.2008.05.002.

[154] A. A. Goshtasby, S. Nikolov, Image fusion: advances in the state of the art, Information Fusion 8 (2) (2007) 114–118. doi:10.1016/j.inffus.2006.04.001.

[155] I. Corona, G. Giacinto, C. Mazzariello, F. Roli, C. Sansone, Information fusion for computer security: State of the art and open issues, Information Fusion 10 (4) (2009) 274–284. doi:10.1016/j.inffus.2009.03.001.

[156] N. E. El Faouzi, H. Leung, A. Kurian, Data fusion in intelligent transportation systems: Progress and challenges – a survey, Information Fusion 12 (1) (2011) 4–10. doi:10.1016/j.inffus.2010.06.001.

[157] G. Navarro-Arribas, V. Torra, Information fusion in data privacy: A survey, Information Fusion 13 (4) (2012) 235–244. doi:10.1016/j.inffus.2012.01.001.

[158] J. D. Velásquez, Web mining and privacy concerns: Some important legal issues to be consider before applying any data and information extraction technique in web-based environments, Expert Systems with Applications 40 (13) (2013) 5228–5239.

[159] S. Sun, H. Shi, Y. Wu, A survey of multi-source domain adaptation, Information Fusion 24 (2014) 84–92. doi:10.1016/j.inffus.2014.12.003.

A survey on Information Fusion applied to Opinion Mining

Information Fusion techniques in the Opinion Mining context.

Text mining algorithms to extract relevant opinions from documental data bases.