Computer Standards & Interfaces 35 (2013) 454–469

Contents lists available at SciVerse ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

OpinionMining-ML

Livio Robaldo *, Luigi Di Caro

Department of Computer Science, University of Turin, Italy

ARTICLE INFO

Available online 23 November 2012

Keywords:
Opinion mining
Sentiment analysis

ABSTRACT

In this paper we propose OpinionMining-ML, a new XML-based formalism for tagging textual expressions conveying opinions on objects that are considered relevant in the state of affairs. The need of such a formalism is motivated by the lack of standards for Opinion Mining (a.k.a. Sentiment Analysis) that obey to certain requirements of efficiency, ease of manual annotation, scalability, and, most of all, that aim at satisfying the real goal of Sentiment Analysis applications. Opinion Mining is an Information Retrieval task, so that its output should be designed for being usable and fruitful from the perspective of a search engine. Our contribution is twofold. First, we present a standard methodology for the annotation of affective statements in text that is strictly independent from any application domain. The second and orthogonal part of the approach regards instead the domain-specific adaptation that relies on the use of an ontology of support, that is domain-dependent by definition. We finally evaluate our proposal by means of fine-grained analyses of the disagreement between different annotators.

1. Introduction

Opinion Mining, or Sentiment Analysis,[1] can be generally defined as the extraction of users' opinions from textual data. Given the huge and ever-growing amount of data made available thanks to the "sharing-age" of the Web 2.0, it became increasingly worth to investigate high-level Information Retrieval tasks like the extraction of users' intentions and feelings about "objects" (henceforth called "facets"), at different levels of granularity. In our opinion, Pang and Lee [43], Tang et al. [59], Liu [30] and Tsytsarau and Palpanas [61] represent very accurate and complete surveys on this topic.

The most relevant motivations behind the recent attraction on this task have to do with its interesting range of applications. For example, a product seller may be interested in knowing the customers' opinions about its products. More generally, enterprises can automatically gain customer feedback [70]. In a completely different scenario, a politician may want to understand what people think about her or him [40] with a view to the next elections, and so forth.

Indeed, the discovery of *sentiments* and *opinions* that are contained in texts is strictly related to many open problems in different research areas, from Psychology to Computer Science. In the first domain, the concepts of emotion and *expressiveness* have to do with many cognitive processes and it is therefore complex by nature. What are the limits of subjectivity? How much the user's experiences change the way of perceiving things and expressing opinions? Wiebe and Mihalcea [65] and Barrett [9] respectively try to answer these questions. On the other hand, Computer Science is involved on the use of Natural Language Processing (NLP) techniques. At the current state of the art, NLP partially provides methods and approaches that can fit with these *emotion-based* kinds of information. Several electronic dictionaries for Sentiment Analysis like Senti-Wordnet [6] have been proposed so far. Nevertheless, the aggregation of simple associations <*word-sentiment*> does not take into account the high complexity of whole sentences, where the use of deep syntactic parsing becomes crucial in that sense.

In addition, in our opinion, the concepts of *sentiment* and *opinion* only cover one part of a bigger set of interesting information that can be relevant at the same level. The speaker/writer could point out details without ascribing any sentiment to them. For instance, he could point out that a certain restaurant made the take-away service available, without commenting anything about its efficiency, quality, and so on. Such objective information, that is clearly precious from the perspective of an Information Retrieval system, is usually denoted as "neutral" [23]. Finally, it seems that other kinds of information should be integrated in such models. For example, texts can contain suggestions, comparisons, questions, and so forth.

Therefore, from a computer scientist's perspective, Sentiment Analysis should be seen as an information extraction subtask. It directly involves the concept of *query*, i.e., a sort of skeleton that has to match with some existing information within a corpus. It may then be argued that the concept of *emotion* becomes less important than the concept of *facet that caused the emotion*. Furthermore, facets can have relations connecting them and so they can be organized into an ontology (cf. [73,72,45]). Then, sentiments, opinions, observations, suggestions and comparisons can refer to different concepts in the ontology, at different level of specificity.

* Corresponding author. Tel.: +39 011 6706812; fax: +39 011 751603.
E-mail addresses: robaldo@di.unito.it (L. Robaldo), dicaro@di.unito.it (L. Di Caro).

[1] Following [43], in the present paper the locutions "Sentiment Analysis" and "Opinion Mining" are assumed to be synonymous.

0920-5489/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csi.2012.10.004

In the light of this, it would be rather useful to have at disposal a formalism that allows to tag all relevant information and to organize them by decoupling relevant textual expressions from the facets those expressions refer to, and relate the former to the latter possibly collocating them within an ontology.

In the industry, there are some attempts to define such a formalism.[2] But, to our knowledge, so far no one has ever tried to systematize and generalize the solutions found in order to share them with the scientific community, by making such solutions contextually independent, easy to extend, easy to integrate within heterogeneous computational systems, etc.

An exception is perhaps Emotion-ML[3] [15,52]. Although Emotion-ML's authors claim that the formalism could be applied for Opinion Mining tasks, in our view the latter cannot be carried out so felicitously via Emotion-ML. The formalism has been designed by focusing on the emotions conveyed by textual expressions[4]rather than on the facets that receive such emotions. In the next section, we will try to sharpen the limits of Emotion-ML, whose overtaking can be considered as the goal of the present proposal.

In this paper we propose a standard formalism called OpinionMining-ML. On the one hand, it would fix some first key points on the relevant information needed to manage emotion-based assertions in texts according to Information Retrieval-based standard requirements. On the other hand, it would allow computer scientists to experiment and evaluate already-existing rather than novel automatic techniques for Sentiment Analysis. It is important to note that our contribution is twofold, since we present a general methodology for the annotation of affective statements in text, showing how it can fit with every domain-specific adaptation relying on an ontology of support, that is domain-dependent by definition.

The evaluation of an annotation scheme is always a challenging process that needs to be faced with care. The validity of such a formalism should take into account the ease of the manual process. In order to enhance the quality of the annotations, and to reduce the time needed to carry them out, the formalism should be as modular as possible, so that the annotation process can be divided into distinct and mostly-independent phases. In this paper, we will explain how our formalism supports quasi-independent modular steps with straightforward guidelines for the annotators.

Finally, we think that this paper can put some basis for the creation of a standard in this field. Standards usually take strong efforts to be built and maintained, with a joint work of several experts in different domains. In spite of this, with the converging of existing models and technologies, the research can obtain significant gains in terms of both quantity and quality.

All these facts suggest the research questions that we will face in this work:

- The construction of a formalism for structuring sentiments and opinions, as well as observations, suggestions and comparisons, in textual documents.
- The construction of a formalism that needs to be strongly application-oriented, i.e., we aim at obtaining annotated textual documents that can be easily queried according to classic Information Retrieval requirements.
- The construction of a formalism based on an ontology that accurately defines the world of objects that can be annotated.
- The proposal of some basis for future standards in the field of Opinion Mining.

In Section 2 we will give a brief overview of the literature, introducing the work that has been done on the different disciplines that Sentiment Analysis covers.

Then, Section 3 presents the input data from which our work began, showing the most significant cases that have been managed in the design of the formalism. The input data come from 2Spaghi,[5] one of the biggest Web2.0-based Italian sites that collects and shows comments about restaurants and pizzerias. Users of the site can leave their comments about the restaurants they attended and read comments left by others in order to understand which new restaurants they would like to try. At the moment, comments are simply gathered and shown, while no analysis of their sentiment is carried out. This paper represents a first step towards an evolution of the site in that sense.

Section 4 presents OpinionMining-ML, while Section 6 illustrates its possible extensions. It will be shown that the design of OpinionMining-ML has been thought to fit with other kind of data as well as with further extensions. Section 7 concludes the paper.

The present work is part of an industrial proposal aiming at providing modules to tag 2Spaghi's comments in OpinionMining-ML. Besides the definition of the formalism, we built a corpus including 1000 annotated comments. The corpus will be used to train and evaluate our future applications. In Section 5, we present some tools that we used to build the corpus as well as an evaluation of the results.

2. Related work

In this section we provide an overview of the related works in this topic from different points of view. On the one hand, Sentiment Analysis includes social and psychological analyses. On the other hand, with respect to the actual scenarios where it can be applied, it is instead related to Computer Science, in particular with Natural Language Processing (NLP) and Information Retrievial (IR) techniques.

2.1. Social aspects

There is a large literature on Social Sciences and Psychology talking about emotions. For instance, several studies demonstrated that the perception of the emotions changes with respect to the experience of a person and her personal issues [9], and to the gender [55]. Many studies have also tried to understand a sort of primary set of emotions although without a perfect agreement [18,47,62]. Still, it has also been demonstrated that people assume different and even independent emotions within the same assertion [21]. Other aspects are the roles of the emotions within general discourses [7,46,53], and within on-line communications [20,25,36].

In general, there are several granularities regarding this type of analysis. A common one is to understand whether a text represents an *objective* or a *subjective* thought [41,26]. A deeper approach is based on the analysis of the *polarity* (or *valence*) of the text, i.e., positive or negative. Most of the approaches concentrate on this task. Going further, another problem is to figure out the *intensity* (also called *arousal* or *strength*) of an emotional state underlying a text [60,57]. A more complex problem is finding the exact *emotions* covered by a textual expression, like "happy", "sad", "angry" and so on [69,16]. Finally, the most challenging problem is represented by the extraction of users' intentions, arguments and speculations [68].

From a procedural point of view, Sentiment Analysis can be done at different levels, i.e., at word level, at sentences level, or at document level. Moreover, the analysis has to be carefully adapted to the context and the source of the data. Texts could be news articles (substantially well-written and formal), tweets (short text messages coming from the web service Twitter[6] [12,13], phone messages [24], bulletin boards, chatrooms, and sites [8,17]. In this sense, with the advent of the Web 2.0

[2] http://research.celi.it/JSPWiki/Wiki.jsp?page=CELI%20Sentiment%20analysis%20%28Linguagrid%20WS%29.

[3] http://www.w3.org/TR/2010/WD-emotionml-20100729/.

[4] Indeed, Emotion-ML has more general aims and a broader range of potential applications. It has been designed in order to allow the "Manual annotation of material involving emotionality, such as annotation of videos, of speech recordings, of faces, of texts, etc."

[5] http://www.2spaghi.it.

[6] http://twitter.com/.

and its related social services, other aspects of subjectivity came out like the concept of *mood* (i.e., an emotional state that the users can attach with the text). [33,35] are a few attempts on the analysis of such data. Still, *emoticons* play an important role in the expressiveness of this kind of texts [19,63,37,50].

2.2. Standards and formalisms

The World Wide Web Consortium[7] (W3C) already proposed the standard Emotion-ML [15,52] which aims at providing an interchange language for applications that deal with emotion and affect. While this work seems to be very close to ours, the goals we want to achieve with OpinionMining-ML are significantly different. Generally speaking, while Emotion-ML has been created with the goal of representing emotions in their general meaning and along their theoretical fundaments, our aim is more narrow in that sense, and it is focused on the efficient applicability on real scenarios. In detail, the differences between EmotionML and OpinionMining-ML are the following:

- Emotion-ML has been thought to cover all the literature on the meaning of the emotions, along with all the theoretical backgrounds coming from the social sciences. OpinionMining-ML is instead designed for Information Retrieval systems.
- Emotion-ML is a markup language that covers a huge range of variabilities and parameters, thus its complexity is consequently high. OpinionMining-ML has been thought to be as much easy as possible in terms of complexity and usability in Information Retrieval's systems and domains.
- Emotion-ML's focus is the concept of emotion, while that of OpinionMining-ML's are the objects under evaluation (that we call *facets* in our formalism). This is because IR systems usually work with object-based queries for the extraction of associated information (e.g., *sentiments*).
- It seems rather hard to build and/or refine annotations in Emotion-ML. Emotion-ML tags allow to represent very specific details, but this should be not necessarily seen as an advantage, for the task at hand. Building a corpus in Emotion-ML could be difficult, as it could be hard to modularize the annotation process or to specify details while leaving others underspecified or at a lower level of specificity. No corpus for Sentiment Analysis in Emotion-ML exists, to our knowledge. Conversely, we aim at designing a formalism where it is possible to move from coarse annotations to fine-grained ones while minimizing the effort.

Another relevant work is the one presented in Wiebe et al. [66] and Wilson [68], where an annotation scheme has been defined to produce the MPQA Opinion Corpus, a freely available corpus that contains manually-annotated news articles. As the authors state in their work, the goal of the annotation scheme is "to represent internal mental and emotional states". Contrariwise, we are more oriented on an Information Retrieval perspective where the main part of the corpus lies on the quality and the usefulness of the results for a user query-based system. In general, the problem of finely representing private states is instead out of the scope of our contribution.

In Stoyanov and Cardie [56], the authors propose a methodology for the manual annotation of opinions that is more in-line with our Information Retrieval-oriented approach. However, our methodology results to be more modular, thus it is likely to lead to more accurate and less time-consuming results. Moreover, OpinionMining-ML presents a simpler definition of the parts that constitute an expression conveying opinions, since it only works with text fragments that can be composed to represent facets and expressions, whereas they define the topic (our facet), the topic span (i.e., the closest minimal span of text that mentions the topic), the target span (i.e., the span of text that covers the syntactic surface of the opinion), and other aspects like the holder of the opinion. Finally, they don't take into account the possibility of using facets organized into ontologies as well as expressions that are in the form of observations, suggestions, comparisons, and questions, as OpinionMining-ML does.

Finally, in Boldrini et al. [11] the authors started from Wiebe et al. [66] illustrating an annotation scheme to capture subjectivity. They proposed to annotate each word in a sentence with different features like confidence of the annotator, emotion, phenomenon, polarity, source, target, and so forth. Even if we think that such a strategy can carry to have highly informative corpora, in our paper we concentrate on annotating pieces of text conveying opinions on domain-specific ontological facets or properties that can profusely serve Information Retrieval systems in practical scenarios.

2.3. Automatic approaches for opinion retrieval and mining

In literature there exist several automatic (or semi-automatic) approaches that have to do with Sentiment Analysis aspects. The main distinction here is between Opinion Retrieval and Opinion Mining. While the first task deals with the general identification and extraction of opinions from text, the latter aims at capturing their sentiment salience or polarity. Even if these two views are theoretically different, we think that they share both the motivations and the adopted technology in practice. For instance, the works published in Zhang et al. [71], Mishne [34] and Java et al. [27] all present automatic techniques for Opinion Retrieval, though they use sentiment lexicons as well as standard classification algorithms, like Opinion Mining-oriented approaches.

Indeed, there exist two main approaches (and hybrid solutions) on these tasks: the use of sentiment lexicons (eventually with Natural Language Processing techniques), and the use of Machine Learning classification algorithms. A Sentiment lexicon is a list of words that are associated to polarity values (positive rather than negative). One of the most known is ANEW [14], developed for linguistic studies even before the concept of Sentiment Analysis in Computer Science. Other resources are General Inquirer [54], Opinion Finder [64], Wordnet-Affect [58], Senti-Wordnet [6], Q-Wordnet [3], Senti-strength [60] and a few others [39,4]. NLP techniques represent a deeper level of analysis since they take into account the context in which the words appear (it is well known that words in natural languages can have different meaning with respect to the context). Some of the approaches presented in literature include the use of n-grams [2,38,59], lexical and syntactic patterns [51], and rule-based systems [48,69].

On the other hand, Sentiment Analysis can be viewed as a Machine Learning classification problem. Generally speaking, the task becomes to classify the opinion as falling under one of two opposing sentiment polarities, or locate its position on the continuum between these two polarities, given an opinionated piece of text [43]. There exist many works that face Sentiment Analysis under this view, like [44] which uses Support Vector Machines (SVM) over a dataset of movies.[8] In our opinion, relevant papers on this direction are Abbasi, Chen, and Salem [1], Abbasi et al. [2], Argamon et al. [5], Gamon [22], Mishne [33], Wilson et al. [67], and Pang and Lee [42].

While the main contribution of this work is to define a standard formalism for the annotation of sentiments in text corpora, we plan to develop systems for automatically building annotations obeying the proposed formalism, OpinionMining-ML. We think that this work can incentivate the community moving the attention on direct comparisons among the existing technologies and the sharing of data.

[7] http://www.w3.org/.

[8] http://www.cs.cornell.edu/people/pabo/movie-review-data/.

3. Analysis of the input data

As pointed out in Section 1, the analysis below is restricted to a set of comments taken from the website 2Spaghi.[9] Since the data contains Italian texts, in the paper we will only report the English translation. The comments that we analyzed convey a great variety of appraisals, suggestions, observations, and comparisons, with different degrees of intensity, on properties and features relevant in the domain of restaurants. In this paper, we use the term Expressive Statement (ES) to refer to the smallest class that includes them all. The identification of the Expressive Statements (ESs) occurring in the comments is a rather difficult task. A comment can denote multiple ESs, which are spread out and intertwined within complex linguistic realizations.

In (1), we start presenting some simple comments. Expressive Statements are enclosed between indexed square brackets. Furthermore, comments have been made anonymous.[10] For instance, in (1.c), the name of the restaurant that occurred in the original comment has been substituted with the string "Restaurant1".

(1)

a. [Excellent the choice of wines]$_{a1}$, [erudite the cuisine]$_{a2}$, [punctual the service]$_{a3}$.
b. [It's better to make a reservation]$_{b1}$ but [most of all be punctual!]$_{b2}$.
c. [It somehow reminds the Restaurant2]$_{c1}$ [but here you can eat better]$_{c2}$
d. [15€ for buffet including beverages]$_{d1}$ …[dishes typical of Sri Lanka.]$_{d2}$

In (1.a), it is rather easy to identify several positive comments, as they are bounded by (elliptical) sentences. Note also that each ES expresses a judgment about a particular *facet* of the restaurant: the wines ([…]$_{a1}$), the cuisine ([…]$_{a2}$), and the service ([…]$_{a3}$).

(1.b) Contains two suggestions[11] about the facet "average waiting time": in order to reduce it, the commentator suggests to make a reservation and to be punctual, presumably because the restaurant is usually very crowded.

In (1.c), […]$_{c1}$ compares two restaurants, i.e. the one the comment refers to and Restaurant2. On the other hand, […]$_{c2}$ compares the *cuisine* of the two restaurants.

Finally, the ESs in (1.d) appear to belong to another category of comments. They are not really judgments, as they do not seem to have, positive rather than negative, polarity. Does the commentator appreciates the fact that the total cost is 15€ and that the dishes are typical of Sri Lanka? In our view, […]$_{d1}$ and […]$_{d2}$ do not provide enough information to give an answer. The semantic/pragmatic differences among the examples such as the last three ones considered are very small, strongly subjective, and contextually dependent. As pointed out in Section 1, in the literature ESs like […]$_{d1}$ and […]$_{d2}$ are usually classified as "appraisals with neutral polarity". We do not like such a term, as we prefer to think of an appraisal as an ES that always has a polarity. Instead, we classify examples such as […]$_{d1}$ and […]$_{d2}$ as simple "observations".

In formal terms, we distinguish between an appraisal and an observation by stipulating that the former includes at least one word expressing the polarity, while the latter does not include any. Therefore, for instance, […]$_{a1}$, […]$_{a2}$, and […]$_{a3}$ are appraisals because they respectively contain the words "excellent", "erudite", and "punctual". […]$_{d1}$ and […]$_{d2}$ are observations because they do not contain any word as such. In case the comment would have been "I really *like* the restaurant serves dishes typical of Sri Lanka", it would have been an appraisal.

3.1. Facets

In the beginning of the section, we discussed some examples taken from the data set. It has been observed that each ES identified in (1) refers to a particular *facet* of a restaurant, e.g. the "cuisine", the "price", the "service", the "wines", or, in the most general case, the whole restaurant the comment is about.

In this subsection, we investigate more in depth what facets are, and try to clarify to the reader what we exactly mean by using this term.

First of all, it is important to understand that the facet is a *concept*, not a *linguistic expression*. The same facet can be identified via different linguistic expressions. For instance, it is clear that (2.a-b) and (2.c-d) respectively refer to the facets "cuisine" and "live music", even if only (2.a) and (2.c) explicitly contain these terms.

(2)

a. [The cuisine got definitely worse with time.]$_a$
b. [The tastes are excessive, nauseating.]$_b$
c. [Live music every evening]$_c$
d. [At night there is the minstrel.]$_d$

The facet could also be denoted by an anaphorical expression. For instance, in (1), the pronoun "here" occurring in […]$_{c2}$ deictically refers to the restaurant the ES is about.

Now, how can we identify the list of relevant concepts, i.e. facets? We are trying to define an XML-based formalism for annotating the sentiment of textual expressions. We want such annotations to be fruitful for search engines that are able to give useful suggestions to the customers (e.g. where it is possible to eat good dishes made of fish, and what restaurants have the smoking room) or to indicate a restaurant's owner what aspects of his restaurant are more/less appreciated by the customers. In our view, the list of relevant facets must be found by keeping in mind this very final goal. In the light of this, consider the following ESs:

(3)

a. [Be always sure the pizza maker is not overloaded.]$_a$
b. [If you are not equipped with a discount card the check can be expensive.]$_b$

In (3.a), it is likely to say that […]$_a$ refers to the pizza maker. However, in our view, ascribing the ES to such a facet would lead to a wrong, or at least useless, annotation. No user of the search engine will query the system by entering the keyword "pizza maker".

The concept of 'facet' we are trying to adopt must be distinguished from the concepts of 'topic' or 'focus' (a.k.a. 'theme' and 'rheme') of the sentence, as well as from its syntactic actants ('subject', 'object', etc.). Having identified the facets as the objects or characteristics of the domain that are relevant with respect to the *use* of the final annotated corpus, […]$_a$ is taken to be a suggestion about the "average waiting time" of the restaurant: in order to avoid long waiting time, the commentator suggests to check that the pizza maker is reasonably free before making an order.

[9] http://www.2spaghi.it.

[10] In the final annotated corpus every comment is indexed, and X is taken to be the index of the comment. For instance, the restaurant referred to by the comment with index 187 is called "Restaurant187". In case the comment mentions other restaurants, those are referred to by the strings "Restaurant187 _ 2", "Restaurant187 _ 3", etc. Of course, it could be the case that two different comments with, respectively, index X and Y, are about the same restaurant. In such a case, it would be sufficient to assert the identity "RestaurantX = RestaurantY" in the ontology. Nevertheless, the annotated corpus does not specify such identities, since they are not relevant for the present work.

[11] As it will be extensively explained below, discourse connectives, e.g. "but" in (1.b), are left outside the annotation. In other words, only the text between […]$_{b1}$ and […]$_{b2}$ will be enclosed between OpinionMining-ML tags. However, it is clear that the discourse connective contributes to identify the proper `ESs`, both during manual annotation and (in particular) during an automatic one. For instance, by assuming the classification in the Penn Discourse Treebank (http://www.seas.upenn.edu/pdtb/), "but" may denote several semantic relations among which "Contrast" and "(Pragmatic) Concession" (cf. also [32]). Therefore, once the proper sense of a discourse connective is identified, it may be exploited to relate the annotations of the two clauses it connects, i.e. finding the proper tagging of a clause depending on the annotation of the other one.

In several cases, it is somehow hard to identify the facets referred to by the ESs. For instance, both the words "discount card" and "check" appear to denote valid facets for $[\ldots]_b$. At the end, we decided that $[\ldots]_b$ is an observation about the "(average) price" of the restaurant, which is taken to be the facet denoted by the word "check". We achieve such a conclusion by considering again the perspective of a search engine that uses the final annotated corpus. In our view, it is likely that a customer wants to know information about the prices of the restaurants. $[\ldots]_b$ provides a (general) information about the price, and so it must be displayed when the system is so queried. Conversely, the use of a discount card (or any other way that could help the customer to save some money) appears to be immaterial from this point of view.

A related question is: how much fine-grained must be the list of available facets? And, is it really a plain list or is it a richer data structure? Consider the following examples:

(4)
a. [Big and good pizza]$_a$
b. [The service was excellent (zealous and polite).]$_b$
c. [They always offer a flower to women…]$_c$
d. [A note for the desserts, I tasted the little cheese cake. Excellent!]$_d$

It may be argued that each ES in (4) has to be splitted into two sub-ESs. In fact, in (4.a), the adjectives "big" and "good" refers to two different *features* of the pizza: its size and its taste. Analogously, in (4.b), the commentator specifies why he found the service excellent: he appreciated both its speed and its politeness.

Examples (4.a–b) highlight that for many facets it is possible to identify specific *features* of the facets. Ontologically speaking, the latter are again facets, which may be related with the "main" ones via some semantic relations.

Of course, the level of granularity strongly depends on the context and (again) on the usefulness of the annotations by the final applications. For instance, (4.c) is ascribed to the facet "service" rather than to facets like "the restaurant's attitude of making presents" or "the restaurant's attitude of being gentle with women". The latter do not appear to be really useful from the perspective of a search engine, and so we do not think that such a refinement will be done in future versions of the corpus.

Conversely, in the domain of the input data used in this paper, it seems worthy adopting a deeper level of granularity for what concerns the kinds of food served at the restaurant, and even the names of the dished. A potential customer could be interested in eating a particular dish, as well as the owner could be interested in knowing what specialities offered by their restaurants are more/less appreciated. Therefore, for instance, (4.d) is taken to refer to the facet "little cheese cake".

To summarize, in order to provide proper and useful annotations, the facets should be semantically organized into an *ontology*. In the simplest case, the ontology is a plain list of concepts. Furthermore, the list may be then enriched by introducing new semantic relations that inter-connect the facets, thus obtaining a more complex data structure. The corpus we developed in the present research includes only two semantic relations: 'is-a-feature-of' (relating restaurants with their "cuisine", "service", "average price", etc.) and 'is-served-at' (relating restaurants with the kinds of food or specific dishes they serve, e.g. "pizza", "pasta", and "little cheese cake").

Concerning the granularity of the list of concepts, for the first version of the corpus we adopted a very low level of granularity. Such a choice led to a rather high inter-annotator agreement. Nevertheless, it is obvious that too coarse annotations are more useless than fine-grained ones. In other words, coarse annotations favor the inter-annotator agreement, a property that is necessary to produce consistent and reliable corpora, but they also tend to decrease the usefulness of the final corpus. In order to overcome the trade-off, OpinionMining-ML will be designed to facilitate further refinements of the annotations. For the first version of the corpus, we will ask the annotators to produce as coarse as possible annotations. But in our view substituting such coarse annotations with more specific ones will take little effort.

3.2. Bound of the ESs

As discussed in the previous subsection, OpinionMining-ML will be a strongly *facet-oriented* annotation formalism. In our corpus, facets are contextually relevant concepts about which the customers/owners of the restaurant could be interested in knowing what the commentators say.

We aim at annotating every portion of text that conveys an appraisal, observation, suggestion, or comparison about a facet. In some cases it is easy to do so, because the bounds of the ESs are the same of the clauses of the input sentences. In all examples seen above the bounds of the ESs match those of the clauses.

On the other hand, sometimes it is rather hard to identify such bounds, because multiple ESs occur within the same clause.

The corpus includes many comments where the modifiers of a sentence refer to facets different from the one the whole sentence is about. Consider examples in (5).

(5)
a. [Excellent pizza]$_{a1}$ [without gluten]$_{a2}$
b. [The guesthouse (upstairs) offers]$_{b1}$ [few]$_{b2}$ [great]$_{b3}$ [rooms]$_{b4}$ [at very reasonable prices]$_{b5}$

In (5.a) it is useful to identify *two* facets: "pizza" and "gluten". Both should be tagged because it is likely that some of the potential customers are looking for restaurants serving pizzas, while some others, being celiac,[12]for restaurants serving food without gluten. The facet "gluten", however, is referred to by a prepositional modifier of the clauses.

There are of course comments much more complex than (5.a). For instance, (5.b) includes modifiers, about the same facet ("guesthouse"), that denote appraisals with different polarities. The adjective in $[\ldots]_{b2}$ conveys a negative appraisal, while the adjective in $[\ldots]_{b3}$ and the prepositional modifier in $[\ldots]_{b5}$ are two positive ones.

How do we annotate examples in (5)? A solution could be the one of separating the modifiers (referring to relevant facets) from the clauses where they occur. Such a solution is not attractive because it leads to the definition of guidelines that are hard to follow. Conversely, we decided to adopt a solution where clauses are *always* annotated as a whole and ascribed to their facet. In case some of their modifiers convey additional ESs, they are separately annotated. The advocated solution is achieved by separating the annotation of the input text from the annotation of the ESs. The input text is firstly split into fragments. The latter are then assembled into ESs, so that a fragment may occur within multiple ESs. The tools that have been developed for building the corpus, described in Section 5, allow the annotator to easily carry out such a two-level tagging.

For instance, in the corpus the fragmentation of (5.a–b) coincides with the one identified via the notation '[…]'. Then, fragments are separately assembled[13]as in (6).

(6)
a. $[\ldots]_{a1}+[\ldots]_{a2}=$"Excellent pizza without gluten" (positive appraisal about the facet "pizza")
b. $[\ldots]_{a2}=$"without gluten" (observation about the facet "gluten")

[12] Celiac disease (also called Coeliac) is an autoimmune disorder of the small intestine caused by a reaction to gluten.

[13] Note that (6.c) has been tagged as 'positive appraisal', although it includes a modifier ("few") that expresses a negative comment about the guesthouse. Of course, the polarity of the main ES has to be taken as the *average* of the polarities of the several modifiers. The present version of the corpus adopts a qualitative classification of the appraisals. The latter are tagged as either positive or negative, with three qualitative values of intensity (low, medium, high).In the future, such a classification could be replaced by a finer-grained one that assigns each appraisal a numerical value estimating its degree of polarity.

c. $[\ldots]_{b1}+[\ldots]_{b2}+[\ldots]_{b3}+[\ldots]_{b4}+[\ldots]_{b5}$ = "The guesthouse (upstairs) offers great few rooms at very reasonable prices" (positive appraisal about the facet "guesthouse")
d. $[\ldots]_{b2}$ = "great" (positive appraisal about the facet "guesthouse")
e. $[\ldots]_{b3}$ = "few" (negative appraisal about the facet "guesthouse")
f. $[\ldots]_{b5}$ = "at very reasonable prices" (positive appraisal about the facet "guesthouse")

Other problems arise while trying to annotate coordinations. The conjuncts in a coordination could refer to different facets or ESs. For instance, in (7.a), the object of the verb "to suggest" includes two facets, i.e. two different dished served at the restaurant. Similarly, in (7.b) "good" denotes a positive appraisal about the (quality of the) fish while "expensive" a negative one about the (price of the) fish.

(7)
a. [I suggest]$_{a1}$, [the spaghetti coi muscoli]$_{a2}$ or [the leccia alla ligure]$_{a3}$
b. [Good]$_{b1}$ and [expensive]$_{b2}$ [for the fish]$_{b2}$

Our solution splits (7.a) into the fragments $[\ldots]_{a1}$, $[\ldots]_{a2}$, and $[\ldots]_{a3}$, and reshuffles the fragments so that it is possible to identify the clauses $[\ldots]_{a1}+[\ldots]_{a2}$ = "I suggest the spaghetti coi muscoli" and $[\ldots]_{a1}+[\ldots]_{a3}$ = "I suggest the leccia alla ligure". Similarly, in (7.b), the text has to be fragmented in $[\ldots]_{b1}$, $[\ldots]_{b2}$, and $[\ldots]_{b3}$ so that the clauses $[\ldots]_{b1}+[\ldots]_{b3}$ = "good for the fish" and $[\ldots]_{b2}+[\ldots]_{b3}$ = "expensive for the fish" can be built. Note that the conjunctions "or" and "and" are left outside the fragments.

Finally, the corpus includes comments where modifiers and conjuncts referring to different ESs interact to each other. This may be done in two ways: either the coordination occurs within a modifier, e.g. (8.a), or a modifier has to be attached to both conjuncts of a coordination, e.g. (8.b).

(8)
a. [It's a wonderful place, modern and essential,]$_{a1}$ [with]$_{a2}$ [a great service]$_{a3}$ [and]$_{a4}$ [a great menu for who is a lover like me of the raw food]$_{a5}$
b. [Since he is Sardinian]$_{b1}$ [he is specialized in Sardinian cuisine]$_{b2}$ and [he has several excellent Sardinian wines]$_{b3}$

(8.a) Includes a coordination within a prepositional modifier. Since each of the two conjuncts refers to a different facet (the "service" and the "raw food") they need to be inserted within separate fragments. Note that also the conjunction "and" is inserted within a fragment. As pointed out above, we require clauses denoting an ES to be tagged as a whole, including all their modifiers. In case modifiers convey additional ESs, they are also separately tagged. In case these modifiers include coordinations whose conjuncts refer to different facets or ESs, the annotation must allow to build different modifiers, each one involving a different conjunct. With respect to example (8.a), the fragmentation should then allow the annotation of the following portions of text:

(9)
a. $[\ldots]_{a1}+[\ldots]_{a2}+[\ldots]_{a3}+[\ldots]_{a4}+[\ldots]_{a5}$ = "It's a wonderful place, modern and essential, with a great service and a great menu for who is a lover like me of the raw food" (positive appraisal about the facet "restaurant")
b. $[\ldots]_{a1}+[\ldots]_{a3}$ = "with a great service" (positive appraisal about the facet "service")
c. $[\ldots]_{a2}+[\ldots]_{a5}$ = "with a great menu for who is a lover like me of the raw food" (positive appraisal about the facet "raw food")

On the other hand, in (8.b) the modifier in $[\ldots]_{b1}$ modifies the coordination "$[\ldots]_{b2}$ and $[\ldots]_{b3}$". $[\ldots]_{b2}$ and $[\ldots]_{b3}$ are respectively about the facets "cuisine" and "wine". Furthermore, $[\ldots]_{b1}$ conveys in turn an observation about the owner of the restaurant (which is taken to refer to the facet "service"). In our solution, the modifier $[\ldots]_{b1}$ is both annotated and assembled with both $[\ldots]_{b2}$ and $[\ldots]_{b3}$.

(10)
a. $[\ldots]_{b1}+[\ldots]_{b2}$ = "Since he is Sardinian he is specialized in Sardinian cuisine" (positive appraisal about the facet "cuisine").
b. $[\ldots]_{b1}+[\ldots]_{b3}$ = "Since he is Sardinian he has several excellent Sardinian wines" (positive appraisal about the facet "wine").
c. $[\ldots]_{b1}$ = "Since he is Sardinian" (observation about the facet "service")

In the next section, we present the definion of OpinionMining-ML, a new XML-based annotation formalism for annotating the sentiment of textual expressions. OpinionMining-ML provides a solution to properly tag the examples presented above.

4. OpinionMining-ML

We have seen in the previous sections that tagging textual expressions conveying sentiment is not straightforward. In all examples we analyzed, sentiments are ascribed to facets, i.e. objects or features that are assumed to be relevant in the domain. For this reason, we are moving towards the definition of a strongly facet-oriented formalism.

We aim at devising an XML-complaint solution that is able to cover the range of comments we found in our input data, but that also makes easy:

(11)
a. The definition of clear and sharp guidelines for the annotators. Of course, without such guidelines, it is likely that the inter-annotator agreement becomes very low, i.e. that the annotations are not reliable.
b. The future extensions of the formalism. The design of OpinionMining-ML must allow the substitution of coarse annotations with more fine-grained ones, while minimizing the effort.

We present now the set of tags of OpinionMining-ML.

4.1. Facets and fragments

As pointed out above, the annotations of the input data should both lead to the identification of the relevant facets in the domain and the one of the ESs about these facets.

4.1.1. Facets

Since the annotations with OpinionMining-ML are strongly facet-oriented, we start from presenting the tags that allow to build the ontology. The ontology will be almost a plain list of facets. The list is identified by the tag ONTOFACETS, while each facet by the tag FACET. Each facet has an attribute id, whose value is unique within the ontology.

Facets are concepts; therefore the tag FACET is not used to annotate portions of text. The text among <FACET> …</FACET> is a mere informal textual description of the facet for making the ontology human-readable. It is not part of the input data.

Typical facets of our corpus are "cuisine", "price", "location", "service", etc. of a certain restaurant. Restaurants are also facets, as well as specific dishes, e.g. "little cheese cake", or kinds of food ("pizza", "fish", etc.).

Facets can be related to each other via semantic relations. At the moment, only two semantic relations are introduced: SERVED-AT and FEATURE-OF. They respectively relate the restaurants with the kinds of food or the specific dishes they serve, and with their features (e.g. "quality/price ratio at the restaurant", and "parking at the restaurant").

Some of the latter facets are mandatory and some are not; for instance, every restaurant has a cuisine, but not every restaurant has the live music. However, we do not distinguish below between mandatory and optional facets. Moreover, we list in the ontology only the facets for which there is at least one ES in the input data that refers to it.

The semantic relations SERVED-AT and FEATURE-OF have an attribute facetId that specifies the restaurant that is the agent of the semantic relation. The patients of the semantic relation are listed within the two tags by means of the sub-tag FACETREFERENCE, that specifies the id of the referred FACET. An example of ontology is shown in (12).

The annotators, while they are tagging the ESs, are in charge of populating the ontology with the facets each ES refers to. In the first version of the corpus, facets are kept at the low level of specificity described above. Only the restaurants, the kinds of food, the specific dishes, and the common "abstract" features of the restaurants (e.g. "cuisine", and "service") are taken to be facets.

```
(12) <ONTOFACETS id="1">
       <FACET id="1"> Restaurant1</FACET>
       <FACET id="2"> Restaurant2</FACET>
                 …
     <FACET id="3"> Pizza served at Restaurant1</FACET>
       <FACET id="4"> Pizza served at Restaurant2</FACET>
       <FACET id="5"> Fish served at Restaurant1</FACET>
       <FACET id="6"> Little cheese cake served at
         Restaurant2</FACET>
                 …
       <FACET id="7">Service at Restaurant1</FACET>
       <FACET id="8">Service at Restaurant2</FACET>
       <FACET id="9">Price at Restaurant1</FACET>
       <FACET id="10">Price at Restaurant2</FACET>
       <FACET id="11">Live Music at Restaurant2</FACET>
       <FACET id="12">Quality/Price ratio at Restaurant1
       </FACET>
       <FACET id="13">Quality/Price ratio at Restaurant2
       </FACET>
                 …
       <SERVED-AT facetId="1">
        <FACETREFERENCE>3</FACETREFERENCE>
        <FACETREFERENCE>5</FACETREFERENCE>
                 …
       </SERVED-AT>
       <SERVED-AT facetId="2">
        <FACETREFERENCE>4</FACETREFERENCE>
        <FACETREFERENCE>6</FACETREFERENCE>
                 …
       </SERVED-AT>
       <FEATURE-OF facetId="1">
        <FACETREFERENCE>7</FACETREFERENCE>
        <FACETREFERENCE>9</FACETREFERENCE>
        <FACETREFERENCE>12</FACETREFERENCE>
                 …
     </FEATURE-OF>
       <FEATURE-OF facetId="2">
        <FACETREFERENCE>8</FACETREFERENCE>
        <FACETREFERENCE>10</FACETREFERENCE>
        <FACETREFERENCE>11</FACETREFERENCE>
        <FACETREFERENCE>13</FACETREFERENCE>
                 …
     </FEATURE-OF>
  </ONTOFACETS>
```

4.1.2. Fragments

As shown above, ESs may span over non-contiguous portions of text. In order to annotate them in XML, we must decouple the annotation of the text from the real semantic annotation. The text is split into *fragments*, each of which is indexed. Sets of fragments are then assembled into ESs by inserting the ids within semantic tags specifying whether the ES is an *appraisal*, *suggestion*, *observation*, or *comparison*.

Such an architectural solution is not new in the literature of the XML-based annotation formalisms. It has been implemented[14] in several existing ones, e.g. TigerXML [31] and Time-ML [49].

The tag FRAGMENT is used for annotating portions of text. Each FRAGMENT has an attribute id that is unique within the comment. The id is used as an external reference to the FRAGMENT. The tag FRAGMENT is the only tag for annotating text; all other tags are meta-tags that implement the semantic annotation. The fragmentation of some examples discussed in the previous section are shown in (13) and (14). They do not need further explanations because they simply identify, in terms of the FRAGMENT, the same portions of text identified above in terms of the notations $[\ldots]_a$, $[\ldots]_b$, $[\ldots]_c$, etc.

```
(13) <FRAGMENT id="1"> I suggest</FRAGMENT>
       <FRAGMENT id="2"> the spaghetti coi muscoli
       </FRAGMENT>or
       <FRAGMENT id="3"> la leccia alla ligure </FRAGMENT>
```

```
(14) <FRAGMENT id="1"> It's a wonderful place, modern and
       essential, </FRAGMENT>
       <FRAGMENT id="2"> with </FRAGMENT>
       <FRAGMENT id="3"> a great service </FRAGMENT>
       <FRAGMENT id="4"> and </FRAGMENT>
       <FRAGMENT id="5"> a great menu for who is a lover like
  me of the raw food </FRAGMENT>
```

4.2. Tags APPRAISAL, OBSERVATION, SUGGESTION, and COMPARISON

OpinionMining-ML provides four semantic meta-tags for annotating the sentiment of portions of text: APPRAISAL, SUGGESTION, OBSERVATION, and COMPARISON.

These tags specify the type of ES conveyed by the portion of text. As pointed out in the previous subsection, they do not span over text. Rather, they include references to fragments, specified via the sub-tag FRAGMENTREFERENCE. For this reason, we say they are (semantic) *meta*-tags.

Furthermore, they include at least one tag FACETREFERENCE that specifies the facets they are about. In particular, in the first version of the corpus, the tags APPRAISAL, OBSERVATION, and SUGGESTION contain exactly one FACETREFERENCE each, while the tag COMPARISON exactly two. However, it is basically possible to allow any number of FACETREFERENCE tags within the four meta-tags. We leave open that possibility for future extensions of the corpus.

A final remark concerns the tag APPRAISAL. It includes two attributes: polarity and intensity. The former specifies whether the APPRAISAL is positive rather than negative. We remind that, contrary to most current theories, we prefer to avoid the polarity "neutral" and we make use of the tag OBSERVATION. The tag intensity indicates the degree of the APPRAISAL, i.e. the emphasis used by the commentator. It has three possible values: low, medium and high. Contrary to what is done in other

[14] Tiger XML allows to annotate syntactic trees. It tags each word as a node of the tree, then the nodes are connected via special tags that implement the syntactic relations. Time-ML allows to annotate events and time-expressions in the text, then relate the former with the latter via separate tags.

current theories and systems for Sentiment Analysis, we do not express the intensity of an appraisal in terms of a numerical value. We do not think that such a precise estimate of the intensity is really reliable. In our view, such a value should be weighted on the personality of the commentator, an information that is not (currently) available. In fact, a commentator having a very extrovert and vivacious personality is likely to insert very emphatic comments, with a lot of smiles, exclamation marks, or even vulgar words. Another commentator, who appreciated (or did not appreciate) the same restaurant with the “same” intensity, but who has a much more introvert and mild personality, could instead post more quiet and contained comments. We would then assign a “high” intensity to the APPRAISAL made by the first commentator and a “low” one to the APPRAISAL made by the second commentator. Nevertheless, it could be reasonably argued that they actually have the same intensity. Therefore, we do not consider the value of the attribute intensity so reliable, so that we think it is fine to adopt a rough qualitative scale for its values.

Below we report some examples. The ontology, its id, and the ones of the fragments are replicated for each example, i.e. they are unique within the example only. On the contrary, the real corpus contains a single ontology that is referred in every comment.

(15) Be always sure the pizza maker is not overloaded. (=(3.a))

```
<ONTOFACETS id="1">
    <FACET id="1">Restaurant1</FACET>
    <FACET id="2">average waiting time at Restaurant1
    </FACET>
    <FEATURE-OF facetId="1">
        <FACETREFERENCE>2</FACETREFERENCE>
    </FEATURE-OF>
</ONTOFACETS>

<COMMENT id="1" ontologyreference="1">
    <FRAGMENT id="1">Be always sure the pizza maker is not
overloaded.</FRAGMENT>

    <SUGGESTION>
        <FACETREFERENCE>2</FACETREFERENCE>
        <FRAGMENTREFERENCE>1
        </FRAGMENTREFERENCE>
    </SUGGESTION>
</COMMENT>
```

(16) It somehow reminds the Restaurant2 but here you can eat better (=(1.c))

```
<ONTOFACETS id="1">
    <FACET id="1">Restaurant1</FACET>
    <FACET id="2">Restaurant2</FACET>
    <FACET id="3">cuisine feature of Restaurant1
    </FACET>
    <FACET id="4">cuisine feature of Restaurant2
    </FACET>
    <FEATURE-OF facetId="1">
        <FACETREFERENCE>3</FACETREFERENCE>
    </FEATURE-OF>
    <FEATURE-OF facetId="2">
        <FACETREFERENCE>4</FACETREFERENCE>
    </FEATURE-OF>
</ONTOFACETS>

<COMMENT id="1" ontologyreference="1">
    <FRAGMENT id="1">It somehow reminds the
     Restaurant2
    </FRAGMENT>
     but
     <FRAGMENT id="2">here you can eat better
     </FRAGMENT>
  <COMPARISON>
    <FACETREFERENCE>1</FACETREFERENCE>
    <FACETREFERENCE>2</FACETREFERENCE>
    <FRAGMENTREFERENCE>1</FRAGMENTREFERENCE>
  </COMPARISON>
  <COMPARISON>
    <FACETREFERENCE>3</FACETREFERENCE>
    <FACETREFERENCE>4</FACETREFERENCE>
    <FRAGMENTREFERENCE>2</FRAGMENTREFERENCE>
  </COMPARISON>
</COMMENT>
```

(17) Since he is Sardinian he is specialized in Sardinian cuisine and he has several excellent Sardinian wines (=(8.b))

```
<ONTOFACETS id="1">
    <FACET id="1">Restaurant1</FACET>
    <FACET id="2">service feature of Restaurant1
    </FACET>
    <FACET id="3">cuisine feature of Restaurant1
    </FACET>
    <FACET id="4">wine served at Restaurant1</FACET>
    <FEATURE-OF facetId="1">
        <FACETREFERENCE>2</FACETREFERENCE>
        <FACETREFERENCE>3</FACETREFERENCE>
    </FEATURE-OF>
    <SERVED-AT facetId="1">
        <FACETREFERENCE>4</FACETREFERENCE>
    </SERVED-AT>
</ONTOFACETS>

<COMMENT id="1" ontologyreference="1">
    <FRAGMENT id="1">Siccome è sardo</FRAGMENT>
    <FRAGMENT id="2">è specializzato in cucina sarda
    </FRAGMENT>e
    <FRAGMENT id="3">ha tutta una serie di ottimi vini
sardi</FRAGMENT>
    <OBSERVATION>
      <FACETREFERENCE>2</FACETREFERENCE>
      <FRAGMENTREFERENCE>1
      </FRAGMENTREFERENCE>
    </OBSERVATION>
    <APPRAISAL polarity="positive" intensity="medium">
        <FACETREFERENCE>3</FACETREFERENCE>
        <FRAGMENTREFERENCE>1
        </FRAGMENTREFERENCE>
        <FRAGMENTREFERENCE>2
        </FRAGMENTREFERENCE>
    </APPRAISAL>
    <APPRAISAL polarity="positive" intensity="medium">
        <FACETREFERENCE>4</FACETREFERENCE>
```

<FRAGMENTREFERENCE>1
</FRAGMENTREFERENCE>
<FRAGMENTREFERENCE>3
</FRAGMENTREFERENCE>
</APPRAISAL>
</COMMENT>

5. Building an OpinionMining-ML corpus

In this section, we describe some tools developed for building a corpus in OpinionMining-ML. So far, the tools were used to annotate 1000 comments about restaurants that have been given us by the administrators of 2Spaghi. All 1000 comments have been doubly-annotated by the authors of this paper.

The tools have been developed for facilitating the tagging of the comments. We modularized the annotation into three phases in order to achieve more accurate and less time-consuming results. Each tool implements one of the phases.

1. Tool1: It allows to split the text into fragments. It builds a partial document in OpinionMining-ML by enclosing portions of text within FRAGMENT tags.
2. Tool2: It allows to assemble the fragments annotated via the first tool into ESs. It adds to the partial OpinionMining-ML document the tags APPRAISAL, SUGGESTION, OBSERVATION, and COMPARISON.
3. Tool3: It allows to associate each ES with a FACET. It creates the ontology of the facets and insert a FACETREFERENCE for each ES annotated via the second tool.

Of course, the gold standard built via Tool1 is used as input for Tool2, while the gold standard built via Tool2 is used as input for Tool3. The three tools, and the main guidelines for annotating the comments, are presented in the next subsections.

5.1. Tool1: fragmenting the text

The first tool allows to split the text into fragments. Comments are loaded one by one; then, for each of them, the annotator can select the (contiguous) words that are part of a same fragment. Fig. 1 shows a screenshot of the first tool.

By using the mouse, the annotator can select the words; once selected, they are colored in orange. In Fig. 1, the annotator selected the words "per la qualità". It is possible to delete the identified fragments and redo the annotation. "Previous Fragment" and "Next Fragment" allow to select the fragments done so far (the currently selected one is shown in yellow), and by pressing the button "Delete Fragment" the currently selected fragment is deleted. In Fig. 1, two fragments have been identified: "la mangiata di pesce che ho fatto a Pasqua me la ricordo ancora!" and "è molto famoso in zona". The latter is the currently selected one. It will be deleted in case the button "Delete Fragment" is pressed.

Finally, the radio button "TULE chunking" allows to select fragments that correspond to the chunks identified by the TULE parser [28]. The latter is a rule-based dependency parser developed at the University of Turin. It is currently one of the most effective dependency parsers for Italian [29]. Tool1 parses the input text via the TULE parser and collects the chunks identified by the latter. For instance, in Fig. 1, by parsing the text colored in yellow and orange the TULE parser identifies the chunks "è", "molto famoso", "in zona", and "per la qualità".

In case the radio button is set on "Yes", chunks are selectable as a whole. In other words, when an annotator selects a word belonging to a chunk, all words belonging to that chunk are highlighted. Obviously, the functionality has been added in order to reduce the time needed to annotate fragments. The identification of chunks via the TULE parser is rather precise, so that the latter revealed to be rather helpful in that sense.

TULE chunking may be deactivated by switching the radio button on "No". Some input comments are rather ungrammatical and contain several typos. As a consequence, most chunks identified by the TULE parser are wrong. In such cases, the annotator deactivates TULE chunking and carries out the annotation by selecting words one by one.

We provide the following guideline for tagging fragments via Tool1:

(18) Guidelines Tool1
- (a) Every clause, including all its modifiers, that evaluates, provides suggestions about, compares, makes observations, or even simply mentions a relevant facet must be tagged.
  - (I) An exception concerns hypothetical or strongly elliptical sentences. In such cases, two or more clauses can belong to the same fragment.
- (b) Every modifier (of a clause) that evaluates, provides suggestions about, compares, makes observations, or even simply mentions a relevant facet must be separately tagged.
  - (I) An exception concerns modifiers that denote observations about the same facet referred to by their main clauses. Currently, such modifiers are not separately tagged; their annotation is left as future work.
- (c) In case different facets or different ESs occur within a coordination, those are inserted into separate fragments. The conjunctions connecting them are left outside FRAGMENT tags. In case such coordinations are the arguments of a clause, fragments must be identified so that it is possible to compose each conjunct with the main clause, including its modifiers (cf. (7) and (8) above).

The exceptions mentioned in (18) are exemplified in (19.a–b). In (19.a), the meaning of "spending something more" is taken to be

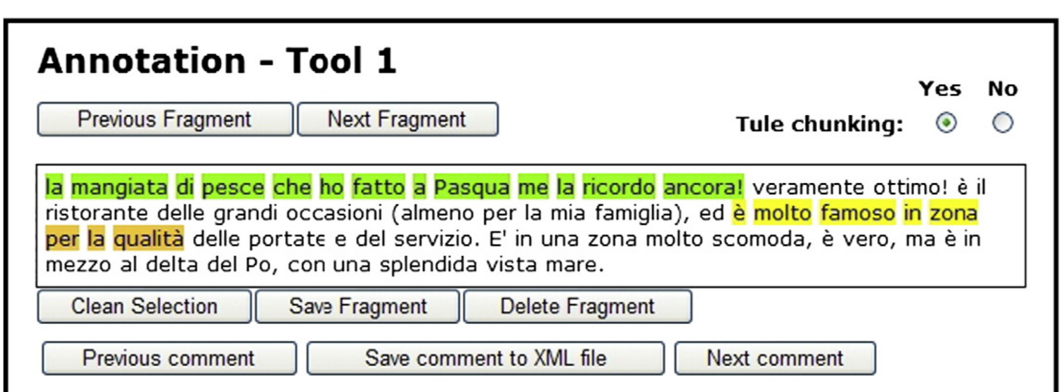

Fig. 1. Tool1 - Screenshot.

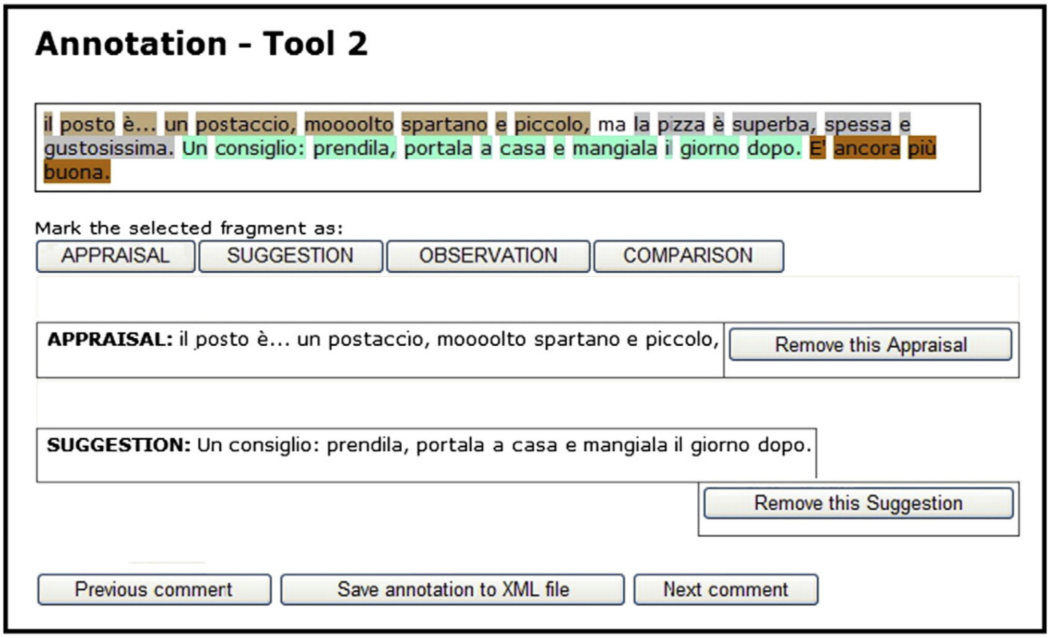

Fig. 2. Tool2 - Screenshot.

hypothetical (the sentence means that a customer should go to eat in another restaurant *even if* that restaurant is more expensive), and so it is not separately tagged. In (19.b), the fact that the pizza is baked in a wood oven is an OBSERVATION about the "pizza". Since the latter is the same facet referred to by the whole clauses, the modifier conveying that OBSERVATION is not separately annotated.

(19)
(a) [It is better to spend something more and go to eat "for real".]$_b$
(b) [The pizza baked in the wood oven is not bad.]$_c$

5.2. Tool2: assemblying fragments into ESs

The second tool takes as input the gold fragmentation and allows to assemble the fragments into ESs. Fig. 2 shows a screenshot of Tool2.

The tool shows each fragment in a different color (but yellow). The user can select one or more fragments with the mouse. All selected fragments are colored in yellow. By pressing either "appraisal", "suggestion", "observation", or "comparison", the tool registers the group of fragment as the corresponding ES.

In Fig. 2, two ESs have been built so far: the appraisal "il posto è un postaccio, moooolto spartano e piccolo" and "un consiglio: prendila, portala a casa e mangiala il giorno dopo", each of which is composed by a single fragment. In the main window, the fragment "la pizza è superba, spessa e gustosissima" is selected.

In case the "appraisal" button has been pressed, a little pop-menu appears and asks the annotator to specify the polarity and the intensity of the appraisal (cf. Fig. 3).

In order to help the user annotate, the system automatically provides a suggestion of the polarity of the selected "appraisal". Of course, the user is able to eventually change the suggestion. The computation of the polarity is done by using SentiWordNet [6] in conjunction with MultiWordnet [10]. While the first resource contains a large dictionary of polarity-annotated WordNet synsets, the latter allows to map them to Italian words. The polarity is then chosen by aggregating (and then comparing) both negative and positive polarity of all the words in the selected "appraisal".

It is worth spending some further words on few recurrent[15] patterns that we found in the comments, and for which we decided to adopted ad-hoc solutions. There are three cases, exemplified in (20), that appear to denote a certain ES from a syntactic or semantic point of view, but that, in our opinion, pragmatically denote another one.

(20)
(a) [I have been a patron at Restaurant3 for almost ten years.]$_a$
(b) [I strongly suggest you the potatoes cooked in a wrapper.]$_b$

Although comment (20.a) looks like an observation, we take it to be an appraisal, by assuming that if someone *often* goes to eat in a restaurant he really likes it. The input data contains several comments as such, e.g. comments in the form "I went in that restaurant since I was child", etc. We annotate them all as positive appraisals.

On the other hand, the comment in (20.b) appears to be a suggestion, as it contains the verb "to suggest". Conversely, we take it to be a positive appraisal about the dish "potatoes cooked in a wrapper" by assuming that if a commentator suggests to try a certain dish or food, he really likes it.

Finally, the input data contain a class of comments that is really borderline between appraisal, suggestion, and observation. Two instances of this class are shown in (21).

(21)
(a) [It is suitable for a romantic dinner.]$_a$
(b) [It is excellent for dinners with friends.]$_c$

(21.a) Can be likely classified as an observation about the restaurant. The comment seems to point out a feature of the restaurant. Nevertheless, (21.a) is semantically very similar to (21.b), but it does not seem that the latter can be felicitously classified as observations. (21.b) More likely appears to be a suggestion or an appraisal.

It is important to annotate (21.a–b) with the same tag, because they actually convey the same kind of ES. We decided to tag them as

[15] We found few other comments with very specific and thorny syntactic structure and/or meaning. They required us some discussions in order to agree the best annotation in OpinionMining-ML. However such examples are rather hard to generalize so that we avoid to discuss them here.

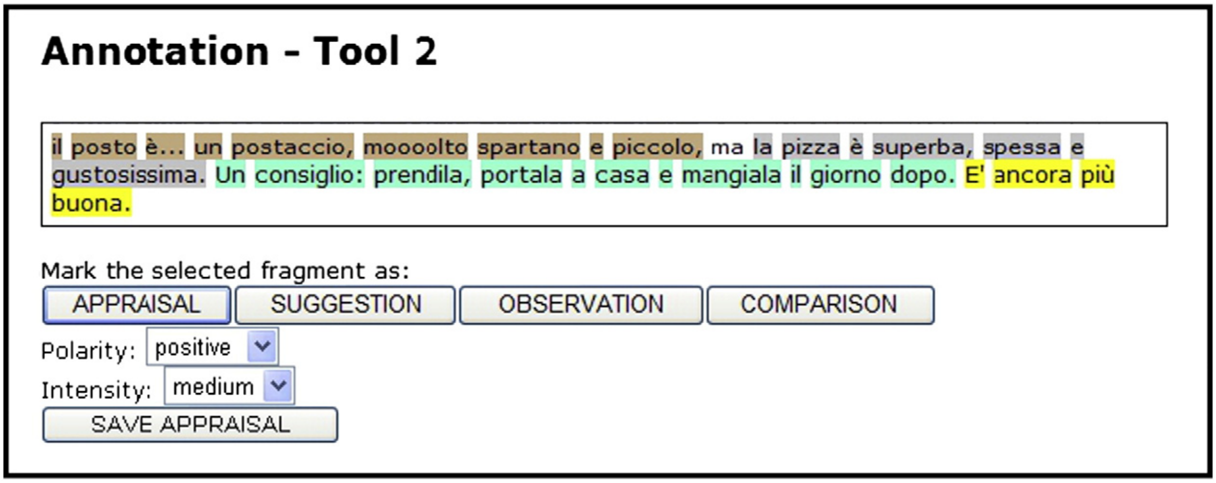

Fig. 3. Tool2 - Automatic Polarity selection.

suggestions. In commentator's view, the appreciation of the restaurant should increase together with a particular company, but it is obviously possible to go there together with other persons. In other words, the commentator is suggesting the kind of fellowship for that restaurant.

Such an ad-hoc solution is not fully satisfying, as it does not allow to properly cover the exact meaning of all similar comments. However, at the moment, the solution suffices for providing a (rough) classification of the ESs. In future works, we could decide to introduce a new tag and use it for annotating all comments as such, or to add special attributes to the tag SUGGESTION in order to discriminate several different subtypes.

We report in (22) the guidelines for Tool2.

(22) Guidelines Tool2

- (a) Fragments are assembled into groups in the following way:
  - (I) Clauses, including modifiers, are inserted within a single group.
  - (II) In case a modifier occurs within a fragment different from the one enclosing its clause, the fragment is inserted in a new different group.
  - (III) Fragments that are conjuncts of a coordination are inserted each within different groups. Each group so built includes the fragments enclosing the clause as well as those other modifiers of the latter.
- (b) Each group of fragment is classified as:
  - (I) Appraisal, if the text in the fragments conveys an evaluation, either positive or negative, about a relevant facet.
  - (II) Suggestion, if the text in the fragments describes actions that should be carried out in order to use more efficiently, appreciate more, etc. a relevant facet.
  - (III) Observation, if the text in the fragments points out or simply mentions a relevant facet, without making any substantial claim about it.
  - (IV) Comparison, if the text in the fragments compares two relevant facets on a certain feature or from a certain point of view.

For the exceptions exemplified in (20) and (21), the present guidelines are overridden by the ad-hoc ones described in the text above.

5.3. Tool3: ascribing ESs to facets

The third and final tool shows each ES built via Tool2 and asks the annotator to specify the facet(s) it refers to. In the first version of the corpus, all ESs, but the comparisons, are associated with a single facet. Comparisons are associated with two facets.

Fig. 4 shows a screenshot of Tool3.

The main window shows the whole comment. The ESs are selected one by one; the words of the currently selected ES, e.g. "il personale gentile" in Fig. 4, are shown in yellow. The annotator must select[16] a facet from the list on the left (or two, in case of comparisons), press either radio buttons "feature-of" or "served-at", and finally the button "Save to XML file". Once "Save to XML file" is pressed, Tool3 adds the corresponding FACETREFERENCE(s) in the OpinionMining-ML tag denoting the ES, and creates the corresponding semantic relations in the ontology (FEATURE-OF or SERVED-AT tags, depending on the semantic relation chosen by the annotator).

Obviously, in case the list does not contain the facet(s) that the annotator wants to select, s/he can add a new facet by writing it in the textbox on the left of "Write a facet and add it to the list". By pressing the button "Insert", the facet is added to the list.

The third tool is the simplest to use. Some difficulties arise from the fact that facets are concepts, but the tool requires to write and select *linguistic expressions* identifying these concepts. It may be the case that, although two or more annotators decide to ascribe an ES to the same facet, they choose two different linguistic realizations of that facet. For instance, suppose Tool3 shows an appraisal about the food served at a certain restaurant. One annotator can associate it with the linguistic expression "food" and another one with the linguistic expression "cuisine". Obviously, the facet identified by the two linguistic expressions is the same. In order to properly calculate the inter-annotators' agreement, we should build the mapping between the facets and the linguistic expressions identifying them. Such a solution would require a lot of further work, so that we decided to adopt an alternative one. One annotator starts annotating and, meanwhile, creates the list of the normalized linguistic realizations he wants to associate with the facets. The list is then reshuffled and given to the other annotator. Obviously, the latter look it up whenever they want to know what normalized linguistic expression is associated with a certain facet. We do not think that this simpler solution jeopardizes the reliability of the inter-annotators' agreement. We provide a single guideline for annotating via Tool3:

(23) Guideline Tool3

- (a) The ESs are ascribed to facets. For each ES, the annotator must select one or more normalized linguistic expressions identifying the facet(s). Appraisals, suggestions, and

[16] In case the annotator thinks that the `ES` is a generic `ES` about the restaurant, he selects the string "RESTAURANT" from the list.

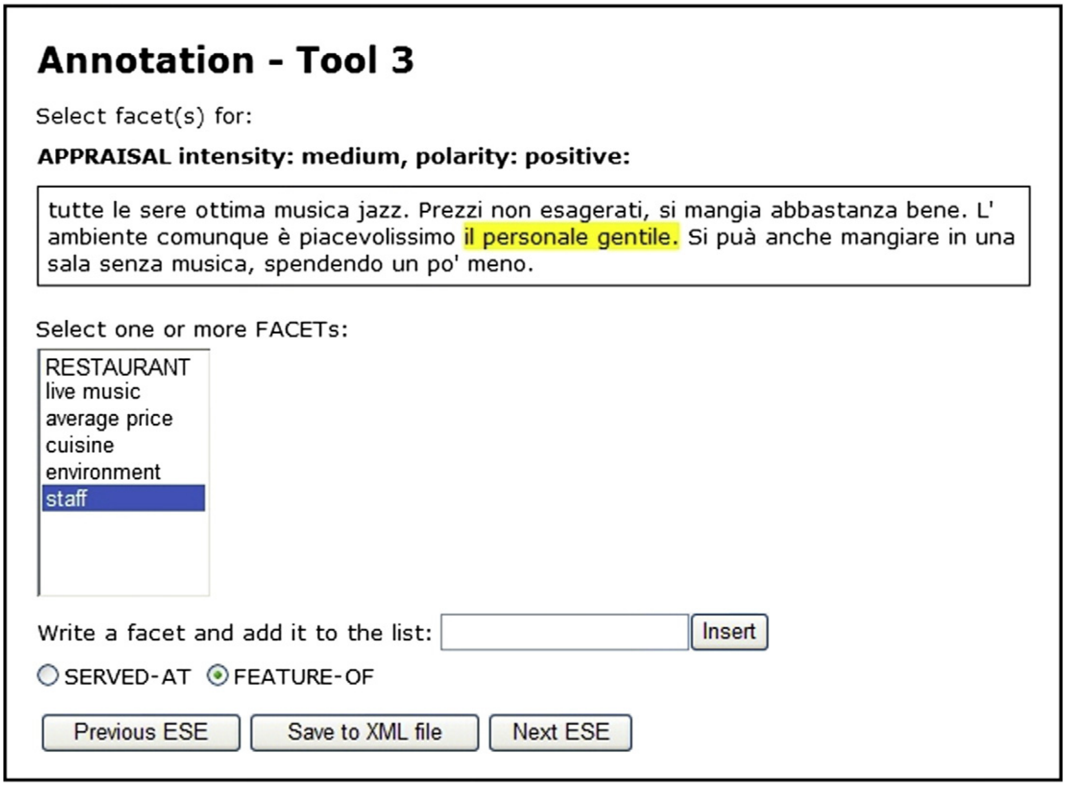

Fig. 4. Tool3 - Screenshot.

observations are associated with a single facet, while comparisons with two ones. The available facets are:

(I) The restaurants referred to by the comments.

(II) Typical features of restaurants: the "cuisine", the "average price", the "quality–price ratio", etc. In the ontology, these are related with their restaurants via the tag FEATURE-OF.

(III) Typical kinds of food served at restaurants: "pizza", "pasta", "fish", "desserts", etc. In the ontology, these are related with the restaurant that serves them via the tag SERVED-AT.

(IV) Specific dishes served at the restaurants: the "Mistua", the "Tortino al formaggio", the "Brodetto", etc. In the ontology, these are related with the restaurant that serves them via the tag SERVED-AT.

5.4. Inter-annotation agreement

In this section we present an evaluation of the agreement between the two annotators, i.e. the two authors of this paper, that built the corpus according to the OpinionMining-ML formalism presented in Section 4. As already mentioned in the paper, the design of OpinionMining-ML has been guided by the principle of the ease of the annotation procedure. Moreover, the very low degree of specificity of the proposed version of OpinionMining-ML allowed us to achieve high values of inter-annotation agreement.

Since the annotation phase has been divided in three parts, we analyzed the agreement for each one of them, carrying out the evaluation at different degrees of granularity.

Table 1
The results of the inner-annotation agreement between the annotators, for Tool1.

Comment-level	First annotator	Second annotator	Gold
First annotator	–	–	–
Second annotator	73.82%	–	–
Gold	89.12%	81.91%	–
Fragment-level	First annotator	Second annotator	Gold
First annotator	–	–	–
Second annotator	91.05%	–	–
Gold	90.13%	92.86%	–

5.4.1. Agreement for Tool1

As expected, the first phase of the annotation process resulted to be the most difficult one in terms of agreement between the annotators. This is due to the high complexity of the procedure. As explained in the previous subsections, the three phases (and so the three tools) are not to be assumed as independent. For this reason, once an annotator created the fragments of phase 1, she or he had to think also to the next phases.

This step has been evaluated using two levels of granularity. On the one hand, we calculated the percentage of comments that have been annotated in the same way. On the other hand, given the fact that each comment can contain multiple fragments, we alsocomputed the accuracy at fragment level. Table 1 shows the accuracy levels.

As can be noticed, the inner-annotation agreement between the two annotators ranges around the 74% of accuracy when considering the comparison at comment-level. While this value could lead one to believe it represents an unfavorable result, we think the most important result to consider should be instead the most fine-grained one. Indeed, considering the agreement at fragment-level, the agreement reaches about the 91% of the total accuracy. Given that this phase represents the most complex part of the whole annotation, these already-high accuracy levels emphirically provide an at-a-glance evaluation of the validity of our scheme of annotation on the whole.

5.4.2. Agreement for Tool2

While the output of Tool1 involves the composition of the text into fragments, the aim of Tool2 is to create different types of ESs. More in detail, the complete list of ESs that can be annotated at this point has

Table 2
The results of the inner-annotation agreement between the annotators for ESE composition via Tool2.

ESs composition	First annotator	Second annotator	Gold
First annotator	–	86.18%	94.12%
Second annotator	90.85%	–	90.51%
Gold	96.11%	91.94%	–

Table 4
The results of the inner-annotation agreement between the annotators for appraisal intensity via Tool2.

Intensity of type appraisal	First annotator	Second annotator	Gold
First annotator	–	93.38%	97.14%
Second annotator	95.18%	–	96.19%
Gold	97.22%	97.13%	–

cardinality 9: six types of appraisal (with polarity *positive* or *negative* and intensity *low*, *medium*, or *high*), observation, suggestion, and comparison. In addition to this, fragments of Tool1 can be grouped in different ways. For this reason, the evaluation of the agreement at this phase has several levels of granularity to consider. Tables 2–4 below show the accuracy values reached in this phase.

The most coarse-grained evaluation only considers how the fragments are grouped into ESs, and it calculates how many fragment-compositions in an annotation *A* are contained also in a second annotation *B*. For this reason, in this case the results are different depending on the direction of the evaluation process ($A \rightarrow B$ rather than $B \rightarrow A$). As shown in Table 2 the accuracy levels range around 88% between the two annotators while they reach more than 90% with the Gold annotation.

A second level of analysis concerns the type of ESs that has been annotated for each ESs (equally-composed by the annotators). For this step, we considered 5 different types out of the total: positive appraisal, negative appraisal, suggestion, observation, and comparison. We did not count at this step the intensity values of the appraisals, leaving them out for the next evaluation.

Finally, the finest evaluation only considers the intensity values of the appraisals that have been annotated with the same polarity. The accuracy values are close to 100%.

5.4.3. Agreement for Tool3

The last phase of the annotation process is about the association between ESs and facets. As for Tool1, the granularity level can be set at comment-level or at fragment-level. In the first case, only equally-annotated comments are considered as correct, whereas at fragment-level the evaluation takes into account the number of correct fragments on the whole. Table 5 shows the results of this evaluation, and, as for Tool2, the results can be different depending on the direction of the evaluation process ($A \rightarrow B$ rather than $B \rightarrow A$). This evaluation reaches the 90% of accuracy value most of the times, thus it clearly demonstrates the validity of this phase.

6. Extending OpinionMining-ML

The design of OpinionMining-ML is grounded on two main architectural choices: facets are organized into a reference ontology, and the annotation of the textual content has been decoupled from the actual semantic annotation. These choices strongly facilitate the achievement of objective (11.b), i.e. the future refinements of the annotations.

Each semantic annotation specifies at least one FACETREFERENCE. It is then possible to consistently add new specific facets in the ontology, connect them to the more abstract ones, and assert new annotations (or modify the existing ones) so that they refer to the new facets. The other annotations can be left unchanged. For instance, consider (24):

(24) If you take pizza, it's really cheap but you start eating fish, it comes to be quite expensive.
<ONTOFACETS id="1">
<FACET id="1">Restaurant1</FACET>
<FACET id="2">average price at Restaurant1</FACET>
<FEATURE-OF facetId="1">
<FACETREFERENCE>2</FACETREFERENCE>
</FEATURE-OF>
</ONTOFACETS>
<FRAGMENT id="1">If you take pizza, it's really cheap,
</FRAGMENT>
but
<FRAGMENT id="2">you start eating fish it comes to be quite expensive.</FRAGMENT>
<APPRAISAL polarity="positive" intensity="medium">
<FACETREFERENCE>2</FACETREFERENCE>
<FRAGMENTREFERENCE>1</FRAGMENTREFERENCE>
</APPRAISAL>
<APPRAISAL polarity="negative" intensity="medium">
<FACETREFERENCE>2</FACETREFERENCE>
<FRAGMENTREFERENCE>2</FRAGMENTREFERENCE>
</APPRAISAL>

In the corpus, both ESs have been ascribed to the facets "average price" because, as said above, the level of granularity has been kept as low as possible. Therefore, for instance, every ES that somehow specifies how it is expensive eating at the restaurant is ascribed to "average price". However, it would be clearly worth introducing two more specific facets for the example under examination, i.e. "price of the pizza (served at Restaurant1)" and "price of the fish (served at Restaurant1)".

We could then refine the annotation in (24) by inserting these new facets and by changing the FACETREFERENCE of the two appraisals as in (25).

(25) If you take pizza, it's really cheap but you start eating fish, it comes to be quite expensive.
<ONTOFACETS id="1">
<FACET id="1">Restaurant1</FACET>
<FACET id="2">average price at Restaurant1</FACET>
<FACET id="3">pizza served at Restaurant1</FACET>
<FACET id="4">fish served at Restaurant1</FACET>

Table 3
The results of the inner-annotation agreement between the annotators for ESE type via Tool2.

ESs type	First annotator	Second annotator	Gold
First annotator	–	91.77%	89.17%
Second annotator	89.47%	–	95.71%
Gold	92.29%	95.91%	–

Table 5
The results of the inner-annotation agreement between the annotators, for Tool3.

Comment-level	First annotator	Second annotator	Gold
First annotator	–	91.11%	93.55%
Second annotator	95.12%	–	94.15%
Gold	92.51%	93.15%	–
Fragment-level	**First annotator**	**Second annotator**	**Gold**
First annotator	–	91.88%	94.11%
Second annotator	94.30%	–	95.01%
Gold	93.65%	95.12%	–

```
        <FACET id="5">average price of pizza at Restaurant1
        </FACET>
        <FACET id="6">average price of fish at Restaurant1
        </FACET>
        <FEATURE-OF facetId="1">
                <FACETREFERENCE>2</FACETREFERENCE>
        </FEATURE-OF>
        <SERVED-AT facetId="1">
                <FACETREFERENCE>3</FACETREFERENCE>
                <FACETREFERENCE>4</FACETREFERENCE>
        </SERVED-AT>
        <FEATURE-OF facetId="3">
                <FACETREFERENCE>5</FACETREFERENCE>
        </FEATURE-OF>
        <FEATURE-OF facetId="4">
                <FACETREFERENCE>6</FACETREFERENCE>
        </FEATURE-OF>
  </ONTOFACETS>
 <FRAGMENT id="1">If you take pizza, it's really cheap,
 </FRAGMENT>
  but
  <FRAGMENT id="2">you start eating fish it comes to be quite
     expensive.</FRAGMENT>
  <APPRAISAL polarity="positive" intensity="medium">
        <FACETREFERENCE>5</FACETREFERENCE>
        <FRAGMENTREFERENCE>1</FRAGMENTREFERENCE>
  </APPRAISAL>
  <APPRAISAL polarity="negative" intensity="medium">
        <FACETREFERENCE>6</FACETREFERENCE>
        <FRAGMENTREFERENCE>2</FRAGMENTREFERENCE>
  </APPRAISAL>
```

Of course, even if all other annotations are left unchanged, the consistency of the corpus is preserved. Suppose the corpus includes other ESs that also refer to the price of the pizza (but that still continue to refer to the average price of Restaurant1 as they have not been updated). These ESs are now vague and approximate with respect to the current ontology, but they are not inconsistent.

To conclude, although the degree of specificity of the current corpus is low, in our future research we plan to refine the annotations by detecting the specific facets that could be useful for a search engine.

The ease of refining the corpus is further guaranteed by decoupling the annotation of the textual content, via the tag FRAGMENT *only*, and the actual semantic annotation, via additional tags whose content are references to FRAGMENTs and FACETs. Fragments simply mark sets of contiguous words. A fragment can be further splitted into sub-fragments; in the limit case, each fragment will include a single word. In order to preserve the semantic annotation, the id of the initial fragment must be substituted, in every semantic tag where it occurs, by the list of all ids of the sub-fragments it has been splitted into.

For instance, in the real market, companies often want to know what people ask about their products. Suppose a bank issues a new kind of credit card. The bank is likely to be aware of what people ask about the new credit card among blogs and forums, in order to plan an effective advertising campaign. OpinionMining-ML may be easily extended to satisfy that goal, as it is sufficient to introduce a new semantic tag <QUESTION> to annotate the set of fragments conveying a question.

On the other hand, as pointed out above, most recent works in Opinion Mining aim at developing dictionaries where lemmas are classified according to the sentiment they convey. The corpus may be then augmented with the lexical knowledge of such dictionaries by enclosing within a fragment each word that has an entry therein, and by separately associating the word with its information in the dictionary via new semantic tags. For instance, with respect to the simple sentence “I like pizza”, we could add in the corpus the score that the word “like” has in Senti-Wordnet [6] in the following way:

(26) Mi piace la pizza (*I like pizza*)

```
Mi<FRAGMENT id="1">piace</FRAGMENT>la pizza
 <SENTIWORDNETENTRY id="01777210" polarity="positive"
 score="0.625">
        <FRAGMENTREFERENCE>1</FRAGMENTREFERENCE>
 </SENTIWORDNETENTRY>
```

7. Conclusions

Sentiment Analysis is receiving high interest because of its direct application in real scenarios and the challenges it generated in several research areas and communities. While most of the works in this field have put its efforts on the introduction of dictionaries and techniques for the extraction of polarity values from texts (negative rather than positive, or value ranges of polarity like $[-1,1]$), only a few moved on the direction of some standards for benchmarking. In fact, in the current state of affairs, we think that Sentiment Analysis needs some key points to be fixed: 1) a standard formalism for the annotation of sentiments and opinions in textual documents and 2) a corpus annotated with such formalism that can enable direct comparisons among the efforts in the field.

In this paper we proposed a proposal for such a formalism called OpinionMining-ML for the annotation of sentiments and opinions within textual documents that avoids cumbersome and complex schemes coming from the studies of the *emotions* in their deepest and sociological meanings.

Contrariwise, our main aim was to develop tools and technologies that can help the community build and compare systems for Sentiment Analysis in an efficient and reliable manner. In order to do so, we directly started from a dataset of restaurant reviews and created an annotation scheme that was as easy and flexible as possible. Still, since our initial idea was to propose a concrete model that could be useful for real systems, we focused our attention on the concept of *facet*, that can be seen as the user query in standard information retrievial approaches. Indeed, we think that Sentiment Analysis, in the field of Computer Science, must be faced as an Information Extraction process, so it needs to be *query-oriented* (or *facet-oriented*).

Once introduced to the data and the tags of OpinionMining-ML, we were able to build an annotated corpus, including so far 1000 comments from spaghi, and run inner-annotation agreements evaluations on it. The construction of such a corpus is described in Section 5 of the present paper.

It is well known that Sentiment Analysis is nothing but trivial since it has to take into account the subjectivity of people expressing opinions. This is due to different causes, like the intrinsic ambiguity of natural languages as well as the different ways that the humans express and perceive things. For this reason, the agreement is an important evaluation step for any formalism in general.

In our case, the results of such evaluation clearly indicate that our proposal represents an effective annotation scheme that is able to cover high complexity while preserving good agreement among different people. In our view, this is one of the key points when introducing supporting techniques in the field of Sentiment Analysis.

In future works, we plan to extend the corpus and compare state-of-the-art techniques for automatic annotations with OpinionMining-ML as well as propose novel methods that can combine the use of statistics with deep syntactic analyses of the input.

Acknowledgments

We gratefully acknowledge the support of the administrators of spaghi, that provided us the input data of the present work. We thank the two anonymous reviewers for their useful comments and observations.

References

[1] A. Abbasi, H. Chen, A. Salem, Sentiment analysis in multiple languages: feature selection for opinion classification in web forums, ACM Transactions on Information Systems (TOIS) 26 (3) (2008) 1–34.
[2] A. Abbasi, H. Chen, S. Thoms, T. Fu, Affect analysis of web forums and blogs using correlation ensembles, IEEE Transactions on Knowledge and Data Engineering 20 (9) (2008) 1168–1180.
[3] R. Agerri, A. García-Serrano, Q-wordnet: extracting polarity from wordnet senses. , Retrieved May in: Seventh Conference on International Language Resources and Evaluation, Malta, vol. 25, 2010, (page 2010).
[4] A. Andreevskaia, S. Bergler, Mining wordnet for fuzzy sentiment: sentiment tag extraction from wordnet glosses, in: Proceedings of EACL, vol. 6, 2006, pp. 209–216.
[5] S. Argamon, C. Whitelaw, P. Chase, S.R. Hota, N. Garg, S. Levitan, Stylistic text classification using functional lexical features, Journal of the American Society for Information Science and Technology 58 (6) (2007) 802–822.
[6] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. Retrieved May in: Seventh Conference on International Language Resources and Evaluation, Malta, vol. 25, 2010, (page 2010).
[7] A. Balahur, Z. Kozareva, A. Montoyo, Determining the polarity and source of opinions expressed in political debates, Computational Linguistics and Intelligent Text Processing (2009) 468–480.
[8] N.S. Baron, Language of the internet, in: The Stanford Handbook for Language Engineers, 2003, pp. 59–127.
[9] L.F. Barrett, Valence is a basic building block of emotional life, Journal of Research in Personality 40 (1) (2006) 35–55.
[10] L. Bentivogli, E. Pianta, C. Girardi, Multiwordnet: developing an aligned multilingual database, in: First International Conference on Global WordNet, Mysore, India, 2002.
[11] E. Boldrini, A. Balahur, P. Martínez-Barco, A. Montoyo, Emotiblog: a finer-grained and more precise learning of subjectivity expression models, in: Proceedings of the Fourth Linguistic Annotation Workshop, Association for Computational Linguistics, 2010, pp. 1–10.
[12] D. Boyd, Why youth (heart) social network sites: the role of networked publics in teenage social life,[w:] buckingham d.(red.), Mc Arthur Foundation on Digital Learning–Youth, Identity, and Digital Media Volume, 2007.
[13] D.M. Boyd, Taken out of context: American teen sociality in networked publics. Ph.D. thesis, University of California, Berkeley, 2009.
[14] M.M. Bradley, P.J. Lang, Affective Norms for English Words (anew): Instruction Manual and Affective Ratings, University of Florida: The Center for Research in Psychophysiology, 1999.
[15] F. Burkhardt, M. Schröder, Emotion markup language: requirements with priorities, in: W3C Incubator Group report, 2008.
[16] F.R. Chaumartin, Upar7: a knowledge-based system for headline sentiment tagging, in: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), 2007, pp. 422–425.
[17] D. Crystal, Language and the Internet, Cambridge Univ. Press, 2006.
[18] C. Darwin, P. Ekman, P. Prodger, The Expression of the Emotions in Man and Animals, Oxford University Press, USA, 2002.
[19] D. Derks, A.E.R. Bos, J. Von Grumbkow, Emoticons and online message interpretation, Social Science Computer Review 26 (3) (2008) 379.
[20] D. Derks, A.H. Fischer, A.E.R. Bos, The role of emotion in computer-mediated communication: a review, Computers in Human Behavior 24 (3) (2008) 766–785.
[21] E. Fox, Emotion Science: Cognitive and Neuroscientific Approaches to Understanding Human Emotions, Palgrave Macmillan, 2008.
[22] M. Gamon, Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis, in: Proceedings of the 20th International Conference on Computational Linguistics, Association for Computational Linguistics, 2004, p. 841-es.
[23] A. Go, L. Huang, R. Bhayani, Twitter sentiment analysis, Entropy (2009).
[24] R. Grinter, M. Eldridge, Wan2tlk?: everyday text messaging, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2003, pp. 441–448.
[25] J.T. Hancock, K. Gee, K. Ciaccio, J.M.H. Lin, I'm sad you're sad: emotional contagion in CMC, in: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, ACM, 2008, pp. 295–298.
[26] T.I. Jain, D. Nemade, Recognizing contextual polarity in phrase-level sentiment analysis, International Journal of Computer Applications IJCA 7 (5) (2010) 5–11.
[27] A. Java, P. Kolari, T. Finin, A. Joshi, J. Martineau, J. Mayfield, The BlogVox opinion retrieval system, in: The Fifteenth Text REtrieval Conference (TREC 2006) Proceedings, 2007.
[28] L. Lesmo, The Rule-Based Parser of the NLP Group of the University of Torino, Intelligenza Artificiale 2 (4) (June 2007) 46–47.
[29] L. Lesmo, The Turin University parser at Evalita 2009, in: Proc. of EVALITA, 9, 2009.
[30] B. Liu, Sentiment analysis and subjectivity, in: Handbook of Natural Language Processing, 2010, pp. 627–666.
[31] A. Mengel, W. Lezius, An XML-based representation format for syntactically annotated corpora, in: Proc. of the International Conference on Language Resources and Evaluation, 2000, pp. 121–126.
[32] E. Miltsakaki, L. Robaldo, A. Lee, A. Joshi, Sense annotation in the Penn discourse treebank, in: Proc. of Computational Linguistics and Intelligent Text Processing, LNCS, vol. 4919, 2008, pp. 275–286.
[33] G. Mishne, Experiments with mood classification in blog posts, in: Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access, Citeseer, 2005, p. 19.
[34] G. Mishne, Multiple ranking strategies for opinion retrieval in blogs, in: Online Proceedings of TREC, 2006.
[35] G. Mishne, M. De Rijke, Capturing global mood levels using blog posts, in: AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs, 2006, pp. 145–152.
[36] B.A. Nardi, Beyond bandwidth: dimensions of connection in interpersonal communication, Computer Supported Cooperative Work (CSCW) 14 (2) (2005) 91–130.
[37] A. Neviarouskaya, H. Prendinger, M. Ishizuka, Textual affect sensing for sociable and expressive online communication, Affective Computing and Intelligent Interaction (2007) 218–229.
[38] V. Ng, S. Dasgupta, S.M. Arifin, Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews, in: Proceedings of the COLING/ACL on Main Conference Poster Sessions, Association for Computational Linguistics, 2006, pp. 611–618.
[39] F.Å. Nielsen, A new anew: evaluation of a word list for sentiment analysis in microblogs, *Arxiv preprint arXiv:1103.2903*, 2011.
[40] B. O'Connor, R. Balasubramanyan, B.R. Routledge, N.A. Smith, From tweets to polls: linking text sentiment to public opinion time series, in: Proceedings of the International AAAI Conference on Weblogs and Social Media, 2010, pp. 122–129.
[41] B. Pang, L. Lee, A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, in: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2004, p. 271.
[42] 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, Association for Computational Linguistics, 2005, pp. 115–124.
[43] B. Pang, L. Lee, Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval 2 (1–2) (2008) 1–135.
[44] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, in: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10, Association for Computational Linguistics, 2002, pp. 79–86.
[45] I. Peñalver-Martínez, R. Valencia-García, F. García Sánchez, Ontology-guided approach to feature-based opinion mining, in: NLDB, 2011, pp. 193–200.
[46] J.W. Pennebaker, M.R. Mehl, K.G. Niederhoffer, Psychological aspects of natural language use: our words, our selves, Annual Review of Psychology 54 (1) (2003) 547–577.
[47] R. Plutchik, Emotion: A Psychoevolutionary Synthesis, Harper & Row, New York, 1980.
[48] R. Prabowo, M. Thelwall, Sentiment analysis: a combined approach, Journal of Informetrics 3 (2) (2009) 143–157.
[49] J. Pustejovsky, J. Castao, R. Saurì, R. Ingria, R. Gaizauskas, A. Setzer, G. Katz, TimeML: robust specification of event and temporal expressions in text, in: In Proc. of Fifth International Workshop on Computational Semantics (IWCS-5), Tilburg, The Netherlands, 2003.
[50] J. Read, Using emoticons to reduce dependency in machine learning techniques for sentiment classification, in: Proceedings of the ACL Student Research Workshop, Association for Computational Linguistics, 2005, pp. 43–48.
[51] E. Riloff, J. Wiebe, Learning extraction patterns for subjective expressions, in: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing-Volume 10, Association for Computational Linguistics, 2003, pp. 105–112.
[52] M. Schröder, I. Wilson, W. Jarrold, D. Evans, C. Pelachaud, E. Zovato, K. Karpouzis, What is most important for an emotion markup language? in: Proc. Third Workshop Emotion and Computing, KI 2008, 2008.
[53] J.C. Short, T.B. Palmer, The application of diction to content analysis research in strategic management, Organizational Research Methods 11 (4) (2008) 727.
[54] P.J. Stone, D.C. Dunphy, M.S. Smith, The general inquirer: a computer approach to content analysis, 1966.
[55] J.M. Stoppard, G.D.G. Gruchy, Gender, context, and expression of positive emotion, Personality and Social Psychology Bulletin 19 (2) (1993) 143.
[56] V. Stoyanov, C. Cardie, Annotating topics of opinions, in: Proceedings of LREC, 2008.
[57] C. Strapparava, R. Mihalcea, Learning to identify emotions in text, in: Proceedings of the 2008 ACM Symposium on Applied Computing, ACM, 2008, pp. 1556–1560.
[58] C. Strapparava, A. Valitutti, Wordnet-affect: an affective extension of WordNet, in: Proceedings of LREC, vol. 4, Citeseer, 2004, pp. 1083–1086.
[59] H. Tang, S. Tan, X. Cheng, A survey on sentiment detection of reviews, Expert Systems with Applications 36 (7) (2009) 10760–10773.
[60] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas, Sentiment strength detection in short informal text, Journal of the American Society for Information Science and Technology 61 (12) (2010) 2544–2558.
[61] M. Tsytsarau, T. Palpanas, Survey on mining subjective data on the web, Data Mining and Knowledge Discovery (2011) 1–37.

[62] J.H. Turner, On the Origins of Human Emotions: A Sociological Inquiry into the Evolution of Human Affect, Stanford Univ. Press, 2000.
[63] J.B. Walther, K.P. D'Addario, The impacts of emoticons on message interpretation in computer-mediated communication, Social Science Computer Review 19 (3) (2001) 324.
[64] B. Wang, B. Spencer, C.X. Ling, H. Zhang, Semi-supervised self-training for sentence subjectivity classification, in: Advances in Artificial Intelligence: 21st Conference of the Canadian Society for Computational Studies of Intelligence, Canadian AI 2008, Windsor, Canada, May 28–30, 2008: proceedings, vol. 5032, Springer-Verlag New York Inc., May 28–30, 2008, p. 344.
[65] J. Wiebe, R. Mihalcea, Word sense and subjectivity, in: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2006, pp. 1065–1072.
[66] J. Wiebe, T. Wilson, C. Cardie, Annotating expressions of opinions and emotions in language, Language Resources and Evaluation 39 (2–3) (2005) 165–210.
[67] T. Wilson, J. Wiebe, R. Hwa, Recognizing strong and weak opinion clauses, Computational Intelligence 22 (2) (2006) 73–99.
[68] T.A. Wilson, Fine-grained subjectivity and sentiment analysis: Recognizing the intensity, polarity, and attitudes of private states. Ph.D. thesis, Citeseer, 2008.
[69] C.H. Wu, Z.J. Chuang, Y.C. Lin, Emotion recognition from text using semantic labels and separable mixture models, ACM Transactions on Asian Language Information Processing (TALIP) 5 (2) (2006) 165–183.
[70] H.L. Yang, C.L. Liu, A new standard of on-line customer service process: integrating language-action into blogs, Computer Standards and Interfaces 31 (1) (2009) 227–245.
[71] W. Zhang, C. Yu, W. Meng, Opinion retrieval from blogs, in: Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management, ACM, 2007, pp. 831–840.
[72] Lili Zhao, Chunping Li, Ontology based opinion mining for movie reviews, in: KSEM, 2009, pp. 204–214.
[73] L. Zhou, P. Chaovalit, Ontology-supported polarity mining, JASIST 59 (1) (2008) 98–110.