International Journal of Cultural and Digital Tourism
Volume 2 Number 1, Spring 2015

ISSN (Online): 2241-9705
ISSN (Print): 2241-973X

Perspectives on Cognitive Linguistics And Sentiment Analysis of Blogs: The Case of Monterosso 2011 Floods

Miriam Bait [1*], Raffaella Folgieri [2], Jean Paul Medina Carrion [3]

[1*] *Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano*
[2] *Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano*
[3] *Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano*

Abstract

The aim of this study is to examine the use of web resources with a view to tracing the discursive strategies enacted to restore the image of a tourist destination. We focus, in particular, on the case of Monterosso which, along with other parts of Cinque Terre, Italy, was hit by a flood in 2011.
The innovative aspect of this paper consists in its twofold approach: a linguistic approach within the framework of Discourse Analysis, and a sentiment analysis approach realized through tools available on the Internet and specific procedures we have developed in the R environment.
The findings are interesting and encourage to refine our approach in the future.

Keywords: *Discourse Analysis, Cognitive Linguistics, Artificial Intelligence, Sentiment Analysis.*

JEL Classification: M15

1. Introduction

This study is part of an ongoing research project on tourism communication on the web. Specifically, it reports on an exploratory analysis of the communicative strategies displayed by blogs to restore the image of a top tourist destination of the Italian Riviera, Monterosso al Mare, where a serious flood took place on 25th October 2011.

The flood hit a wide coastal area from Liguria to Tuscany, but in particular the Cinque Terre, a UNESCO World Heritage site, and a National Park. Our study started on the discursive forms of (inter)action enacted to rebuild the image of a tourist destination collecting web texts taken from two websites maintained by the city council of Monterosso1, from online newspapers, and from blogs. We soon realized that blogs are more worthy of attention because of their paramount global potential as participatory communication tools and their interesting language traits..

The analysis performed in this study is based on a cognitive linguistic approach because we analyze not only texts, but their impact on readers' emotions and cognition mechanisms to evaluate their general involvement.

To complete our analysis, we used tools and methods of Artificial Intelligence and performed a sentiment analysis of the contents of the two selected websites, and of the general feelings generated on the web by the event, that were detected by users' tweets.

This paper is thus divided as follows: in section two, we introduce the reasons why we decided to consider blogs our main starting point for the analysis. We then present methodology, which is based on a cognitive linguistic approach. In the last section results are shown and discussed.

The third paragraph is devoted to the sentiment analysis. Here, we introduce an overview of this type of analysis and the general principles we have followed. Then, we present the tools and methods we selected, and, finally, we discuss the obtained results.

In the last section we draw some general conclusions as a result of the two performed analyses.

2. The Cognitive Linguistic Approach

2.1 Blogs: What And Why

The starting point is, not surprisingly, the Internet, a virtual public sphere (Habermas 2006), which provides a shared pool of key information and instructions about a place hit by a natural disaster, contributing therefore to constructing and reconstructing its image.

The blog seems to be the fastest growing computer-mediated communication genre. The three main constitutive features are the reverse chronology of its entries, the frequent updating and the combination of links with personal commentary (Miller /Shepherd 2004:4; Herring/Kouper et al. 2005:1). Scholars have long debated the origin of blogs. Some authors – mainly bloggers themselves - claim they are an example of new web-native genres (Blood 2000), just as social networks. Others, in line with Todorov who claimed that "a new genre is always a transformation of an earlier one" (1990: 15), recognize 'antecedents' and therefore talk about 'genre migration' of pre-existing genres, adapted to fully exploit the opportunities offered by the electronic/hypertextual format, or better a "remediation" (Bolter 2001), in the sense that a newer medium takes the place of an older one, borrowing and reorganizing its features.

Possible antecedents can be personal diaries and journals, if not the curio collection or scrapbooks. Whatever the case, it follows that blogs manage to convey a paramount sense of personal authenticity and trigger expectations from the public. The blog allows the author to share his/her views on a variety of subjects directed to a potentially global, but more often local, public. But at the same time, it engages its audience and invites them to respond and interact with the content produced by the blogger.

Our case study illustrates how bloggers establish their own independent perspective on destinations and provide neutral and more reliable information which can be effective in (re)building a destination's image (Johnson /Kaye 2007).

Through the case-study of Monterosso, we will try to track possible origins and derivations of the genre examining the features of blogs informing the web community of the disaster, supporting rebuilding and contributing to image reconstruction.

2.2 The Linguistic Analysis

Firstly, the study aims at investigating the discursive strategies used in blogs to respond to the crisis and restore the image of the area, drawing on principles and analytical tools in Discourse Analysis tradition (Fairclough 1989, 1992, 1995) and on the narrative theory of Labov (Labov and Waletzky 1967, Labov 1997).

To address our research issue, we have collected a dataset of web texts from 27 blogs in the period spanning from October 25th, 2011 to February 29th, 2012.

We will start by presenting who the participants (the voices) are in the texts selected and to what extent they may be interpreted as active social agents in the interaction. As our dataset is too limited (15245 words) to allow a quantitative analysis using frequency measurement, we decided to explore the quantitative distribution of words through a visualization realized using an on-line application called Wordle[2] generating word clouds based on the frequency of the words in a text.

In the following paragraph, we will discuss the results obtained by this first analysis.

2.3 Discussion And Obtained Results

Figure 1. The cloud obtained using Wordle.

As you see from the "cloud" shown in Figure 1, the main voices are Monterosso (also named "the town", "the village"), the people of Monterosso (also referred to as "the population", "the Monterossini", "the townspeople", "the residents", "the community"), the people who helped ("trained emergency workers", "mud angels", "volunteers"), and the administration (the "township" of Monterosso, "the mayor").

Not surprisingly, Monterosso, is the most frequent word used, together with The Cinque Terre, and Vernazza another village hit by the flood. But, on close sight, what is interesting to note is that Monterosso is mostly represented as performing an active role, or better, is personified:

(1) Monterosso has not slept
(2) Monterosso is gradually getting back on its feet

The representation of Monterosso as a 'doer' is counterbalanced by other examples where the city features in object position, that is to say it is 'acted upon':

(3) Monterosso was buried
(4) …backbreaking work that is being done to get Monterosso back on its feet.

Monterosso's active response is well depicted in bloggers' texts which support the constant activity of fundraising but also give account of the tragic consequences of the flood,

(5) Many of you have asked for more information concerning the floods that hit Liguria last week (thanks again, And for yesterday's note!). Your best bet at the moment is to follow the blogs and blog posts by people on the ground.

and remind tourists of the attractiveness of this region:

(6) The village of Monterosso possesses that element that is lacking in modern society. The people here know each other. The people here even care for each other. Many a wave and a smile are exchanged by the locals throughout the day.

Moreover – and more surprisingly – despite being non-professional blogs, they provide the most detailed 'technical' information concerning the area:

(7) Monterosso is recovering quickly and will almost certainly be in perfect working order by the time the tourists start trickling in spring/summer 2012.

Blogs are obviously multy-party spoken text genres characterized by a high degree of what Bakhtin defines 'dialogism' (1981), the simultaneous and often dialectic presence of the voice of the writer and of the reader. Therefore, blogs are characterized by the frequent use of interpersonal pronouns, i.e. I, we and you. In some cases, blogs are more similar to written monologue, in other cases the dialogic interaction prevails because most blogs feature comments which react to or comment on the content of relevant posts. The use of posts, commentary and links contribute to realizing a peculiar kind of interaction, favouring a sense of community among bloggers.

The first person "I" is used to share personal memories about Monterosso and one gets the impression that these people are "speaking" to us and "speaking" with us.

(8) …this group of five seaside villages is probably one of Italy's most scenic travel destinations. I can confirm this as I have visited the Cinque Terre area on a few occasions, such as when I stayed at the Soviore sanctuary which lies in the hills above Monterosso.

(9) I have stayed in Monterosso three times and it saddens me to see this.

The "I" characterizing the online environment, in the form of blogs, is relevant as it influences interactions aimed at supporting participation, partnership, and interconnectedness.

The call for solidarity and partnership is confirmed and enhanced by the strategic use of the personal pronouns "we".

"We" obviously stands out as the ideal 'solidarity builder', together with "you", the addressees of this call for help, as this excerpt exemplify:

> (10) We've updated the following section, and will continue to do so as we find out about more ways you can help this devastated area!

"You" and the related possessive adjective "your" can be interpreted in the texts not only as addressed to the implied audience, the visitors interested in the tragic event, but also as an example of the use of direct speech. These types of texts are in fact aimed at involving the readers: on the one hand, they document personal experiences. On the other hand, they provide factual, useful information motivating people to action, i.e. encouraging commitment to help one another:

> (11) Any of you have asked for more information concerning the floods that hit Liguria last week… Your best bet is to follow the blogs and blog posts by people on the ground.

> (12) If you have not heard of the Cinque Terre, then you may not know that this group of five seaside villages is probably one of Italy's most scenic travel destinations.

At various points, special emphasis is given to the personal dimension, and the symmetrical activation of an interpersonal component. The use of "you" throughout the whole dataset introduces a strong interpersonal component (Halliday 1994) into the text, initiating a dialogue with readers and sometimes using questions

> (13) Yes, October 25, 2011 was a terrible day. But do you remember the good ones in Cinque Terre? (D3)

or exhortations

> (14) Make sure you don't miss these beautiful coastal villages in the Liguria region of Italy.
> (15) Please don't give up on coming here. (D3)

We have seen that the most distinguishing feature of blogs is their 'blogness', i.e. the quality of writing independently and without obligations. This emphasis on the personal dimension has allowed some scholars to identify the blog's generic antecedents in the diary and personal journal genre (cf. McNeill 2003). The blog is a narrative form optimized for the web and, in our specific case, as travelling also means images, memories and their narrative enactments, prominence is given to different contributions from bloggers to make the message more credible and reliable. The narrative component plays a central role from a discursive point of view as the exchange of information is likely to establish a value-laden reality to be shared by all participants. Narratives enhance communication and create a sense of empathy which helps create a common ground.

In this specific case, text analysis reveals that the narrative element is realized in three different forms:

1. there are narratives referring to the past, to events and situations before the flood,

> (16) When I moved to the seaside Italian village of Monterosso al Mare, I was ready for a lot of things.[…] I was ready to stretch out on the beach in the morning, splashing in the clear as glass water.[…]My first summer in paradise went exactly as planned in my little slice of the Italian Riviera

> (17) On October 24, 2011, I visited Cinque Terre while on a day trip with my group. We had a pleasant outing, although it did sprinkle a little.

2. there are narratives of events occurred during the dramatic hours of the flood,

(18) On October 25th, 2011, the heaviest flash flooding in the history of Liguria devastated parts of the region. Monterosso was one of the worst hit. […] The damage Monterosso sustained was worse than all previous disasters combined, including the devastations of World War II. I looked out the window and saw my street turn into a deadly torrent of mud and water as I saw all the cars, parked helplessly, swept into the sea. […]
(19) The next morning I learned that Liguria, the region in which Cinque Terre is located, had received more than 20 inches of rain in three hours. […] Monterosso al Mare, had been inundated by massive flash floods of water, mud, and debris flowing uncontrolled through their steep and narrow streets to the sea. The ground floors of buildings – shops, restaurants, markets, schools, and homes – were filled with water and mud to their ceilings.

3. and finally there are narratives of actions performed after the flood, and in reaction to it

(20) Emergency crews arrived a few days later, Long before the hordes of tourists, Liguria was a poor region that could sustain itself only on what it could produce. […] Reconstruction work is going ahead in Monterosso, so while staying there might not be advisable, nor possible, today, a visit in the near future should not be too problematic, although there will be building works going on. Please keep reading for more information on the travel situation […]Hotels in Monterosso will start reopening in early March 2012, and more will open in April.
(21) The latest reports are promising. Most of the mud is gone. People are repairing and repainting their buildings. Drainage systems are being cleaned out and refurbished (see photos), and everything is on track to be ready for tourists by Easter.

All types of narrative are discursively realized on two different but intertwining planes: a public one giving the 'objective' accounts of the event, providing information and suggestions, and a private one made of personal experiences and memories.

Labov and Waletzky (1972) propose a model of narrative analysis that identifies “the invariant structural units which are represented by a variety of superficial forms.” Although this framework is nearly forty years old, and is focused on oral narrative instead of written text, it continues to influence language studies and maybe – as another scholar suggested (Toolan 1982) – it becomes an ideal tool when analyzing Internet writing, which is often less formal than other types of writing, but more structured than spoken language.

Labov defines narrative as “one method of recapitulating past experience by matching the verbal sequence of clauses to the sequence of events (it is inferred) actually occurred” (1972:359).

His original theory of narrative structure, identifies six main parts of a narrative (abstract, orientation, complicating action, evaluation, resolution, and coda). They are not all necessarily present and not necessarily in this specific order. However, if we apply this model to the texts under examination, we can discover interesting things.

The so-called abstract, i.e. a summary or introductory part of the narrative which serves the purpose of attracting the reader may correspond to the lead of a newspaper article and the introductory paragraph of a blog text, as these examples show:

(22) Three months have passed since the October 25th flood hit Monterosso, causing the village to experience the worst natural disaster of its entire history.
(23) Monterosso was devastated by heavy flooding and mudslides on October 25, 2011.
(24) Flood was horrible last October, but recovery went fast.

The next stage called orientation is also always present and it is aimed at providing the necessary information about the setting of the story that means locating events in time and space and introducing participants, as you see in the following:

(25) I have been speaking everyday with our ex-next door neighbor Bruna. She is an 85 year young sparkplug who never stays still. Friday she took me into her cantina and showed me where her son-in-law was forced to cut a hole in the roof and help Bruna's daughter and granddaughter (who happens to be 8 months pregnant) escape to Bruna's apartment above. They had been in the cantina when the waters in Via Buranco rose so high, and with such force, that they slammed the door shut. The force knocked the keys out of the door and they were closed in.

These lines work as catalysers that lead up to the events and contribute to creating suspense and raising interest.

The complicating action introduces the threatening, descriptive action. It refers to the events told that make the narrative proceed. They provide the referential function of the narrative and report about the next occurrences representing the real backbone of the story:

(26) Fire crews and civil protection teams worked their way through the flood-ravaged towns of Vernazza and Monterosso where cars were washed into the sea and roads turned into rivers.Officials said that within a 24-hour period, 500mm of rain had fallen and this had led to houses collapsing and roads and train lines subsiding.

The texts also offer many examples of evaluation, i.e. statements that tell the reader what to think about a place, a person, an event. Evaluation indicates why the story has been told and brings out the significant elements of the story, for the narrator and the readers, and make the story worth telling, as you these examples show:

(27) Monterosso occupies a very special place in my heart. Most of my family is from there. I grew up playing soccer on the beautiful beach in Fegina or having a fresh granita while talking a stroll through the town with my friends. It's a little paradise populated with a lot of good hearted people. […] The streets, shops and bars I went to are buried in feet and feet of mud. Monterosso and its people are on their knees, praying for a miracle!

The evaluative clauses may contain irrealistic clauses – negative, conditionals, futures – which refer to events that did not happen or might have happened or had not yet happened. But usually the narrative may be interrupted by a subjective report of the writer's feelings, emotions or judgements, as in the following two excerpts:

(28) Now I know this is a small disaster in the scheme of things, in a part of the world most of us will never see, but that doesn't mean we cannot care about the people there, nor offer at least a short prayer for their recovery.
(29) It's heartbreaking....everything that was on via roma is in the same condition. i only know of 3 restaurants that survived unscathed.We were in Monterosso in June. This is so tragic. Does anyone know where donations can be sent to benefit the people of this devastated region?
(30) If you have not heard of the Cinque Terre, then you may not know that this group of five seaside villages is probably one of Italy’s most scenic travel destinations

The latter excerpt on the one hand weakens the power of the narrative by making the listeners aware that this is the narrator’s experience and not their own. On the other hand, sharing thoughts and impressions about the Cinque Terre may encourage interaction and add realistic details to the whole picture.

The *coda*, provides general view of the action, signals the closure of the narrative and takes back to the present moment. The narrator connects the story to actual everyday life:

(31) For all our friends abroad who love Monterosso and want it to go back to it's old splendor here is the English version http://www.rebuildmonterosso.com/ of the italian blog www.buongiornomonterosso.blogspot.com , useful pages full of info that keep us updated on reconstruction of the village. LET'S HELP MONTEROSSO NOW !

(32) In the meantime reservations for all the area are going nicely. People from all over the world love this land so much that they do not want to make Cinque Terre people lack their support even this year. It is rewarding for all the inhabitants and workers of the entire area. […] There's an intense work going on the trails to rebuild all the vast net; much has been done already […]

The strength of narrative as a communication method lies in the fact that it manages to establish a common ground among all participants and provide a faster method of creating a social relationship. The exchange of information is therefore value laden, but in a manner that creates credibility.

3. The Sentiment Analysis

a. An Overview

To complete our analysis, we performed a double sentiment analysis (Pang and Lee, 2008; Godbole, Srinivasaiah and Skiena, 2007): firstly, on the contents of the selected websites; secondly, on the echo produced by the news related to the Monterosso's disaster on the social media in the period October 2011–January 2015.

Not surprisingly, individuals have an innate sense of determining sentiments, therefore we could simply read the text delivered on the two websites and decide if the general mood was positive, neutral or negative. However, human beings are often influenced by personal opinion or one's own life experiences, therefore this approach might lack objectivity. For this reason, we have decided to detect feelings emerging from the contents, using sentiment analysis tools and some procedures that have been specifically developed in the R environment.

However, an extensive analysis of the feelings generated by the event on the web (the second type of analysis) would have resulted in an excessive, time-consuming and probably inconsistent outcome. In fact, it is not realistic for an individual to read thousands of possible comments, reviews, tweets and social communication on the web (Pak and Paroubek, 2010).

That is where tools derived from Artificial Intelligence (A.I.) comes in (Nasukawa and Yi, 2003; Lohr, 2012). Thanks to cloud-based tools and procedures based on Natural Language Processing A.I. algorithms, it is possible to extract sentiment from content and from the web just as any individual would do, but much faster and allowing to obtain more consistent results, considering that usually individuals hardly come to an agreement

2.2. Tools, Materials And Methods

To perform a sentiment analysis of the content of the two selected websites, we used Semantria and Socialmention, two free online tools.

Semantria[3] is a cloud-based tool adopting a Natural Language Processing algorithm to perform the analysis. Following the software technical specification, the algorithm operates on the selected text through the following steps:

1. breaks the sentences of the document into their structural elements (e.g. nouns, adjectives, verbs, and adverbs), on the basis of the selected language;
2. identifies sentiment-bearing phrases (for example "terrible experience" or "unpredictable disaster";
3. each of the identified phrases obtains a logarithmic score, ranging from -10 to 10;

4. the software combines the obtained scores and determines the overall sentiment of the text, assigning a score in the range of -2 to 2.

The queries used by Semantria to calculate the sentiment of a sentence?, are based on the closeness of a word to others. Once done, a hit count is updated with each result and then combined through the log odds ratio mathematical operation. Thus, a sentence receives its final score.

Semantria knowledge base takes advantage from Wikipedia, and creates a Concept Matrix containing the indication of the closeness of the considered terms.

For a deeper understanding of the Semantria tool, you can consult the dedicated webpage[4]. Here we will limit our scope to a list of the outputs we can obtain from the sentiment analysis:

- highlighting the text, using different colours (red – negative; green – positive; gray – neutral) of significant text. These correspond and give form to the words cloud shown at the top of the results, giving a graphical interpretation of the analysis.
- the general sentiment of the document (positive, negative or neutral).
- a text summarization, extracting the most significant sentences and giving a concise synopsis of the original text.
- entities: significant terms/phrases extracted from the text
- themes: relevant themes from the source text
- categories (automatic or created by the user)

For each of the latter three, the results show:

- sentiment: positive, negative or neutral mood of all the mentions of an entity
- evidence: number of sentiment-bearing phrases associated to an entity (score 1-7 where 7 is the most)

We also performed our analysis using procedures we have specifically developed in R environment. R[5] is a free software environment for statistical computing and graphics that we used in Windows O.S.

To develop our sentiment analysis procedures, we have included the following R packages: library (plyr) and library(stringr).

We performed our evaluation considering the databases of positive, negative or neutral words (Minqing and Bing, 2004), first, without grouping the evaluations in the range -5, +5; then, grouping the 10 levels into four (Very Positive, Positive, Negative, Very Negative).

The pseudo-code of the procedures is the following:

```
Import text
Clean text from punctuation marks and abnormal characters
Divide text into strings separed by a blank character
Match text with db_positive_words
sum_pos = sum of detected correspondences
Match text with db_negative_words
sum_neg = sum of detected correspondences
return sum_pos, sum_neg
###considering all the positive and negative levels
import db_AFINN
create categories for each evaluation level
score = c()
for each evaluation level
match text with level (pos/neg)
sum_level = sun detected correspondences
score = score, sum_level
return score
###considering only 4 groups for levels
import db_AFINN
create categories
```

Vpos = db_AFINN with evaluation 5 and 4
pos = db_AFINN with evaluation 3, 2, 1
neg = db_AFINN with evaluation -1, -2, -3
Vneg = db_AFINN with evaluation -4, -5
score = c()
for each evaluation level
match text with level (pos/neg)
sum_level = sum detected correspondences
score = score, somma_livello
return score

As mentioned before, we also performed a sentiment analysis of the general feelings generated by the event on the web. For this purpose, we used the free online tool Social Mention, Topsy and backtweets.

Social Mention[6] is a platform that enables a search on social media performing an analysis based on the aggregation of user generated content, considering what people say across media platforms like Twitter, Facebook, FriendFeed, YouTube, Digg, Google, and so forth. The result is shown in an easy-to-interpret way, on a web page where we can read the detected sentiment, the top keywords occurring in communications, the top users influencing the general feelings. In addition, on the right, we have the possibility to see and explore the retrieved messages and to export data in CSV format.

To conclude, we have also created a script in R environment to perform sentiment analysis on tweets retrieved from the page https://twitter.com/RBMonterosso. In this case, we used the libraries: twitteR, ROAuth, plyr, stringr, colorspace, ggplot2, wordcloud.

We linked to Twitter through the Application Programming Interface (API). We then selected a specific search mode according to what we wish to analyze (a user or a word).

In our case, the pseudo code has been the following:

Assume "RBMonterosso" as the name of the user of the tweets we are searching for
Compose a list of all the tweet retrieved from the selected user
Assume "monterosso" as the string to search
Compose a list of all the occurrence (tweets) of the word “monterosso”
Load positive and negative words from the reference dataset
for every retreived tweet
remove punctuation
remove redundant white spaces
transform text in lower case letters
transform the considered tweet in a list of words
search all the positive or negative words in the considered tweet
score = sum(positive words) – sum(negative words)
end

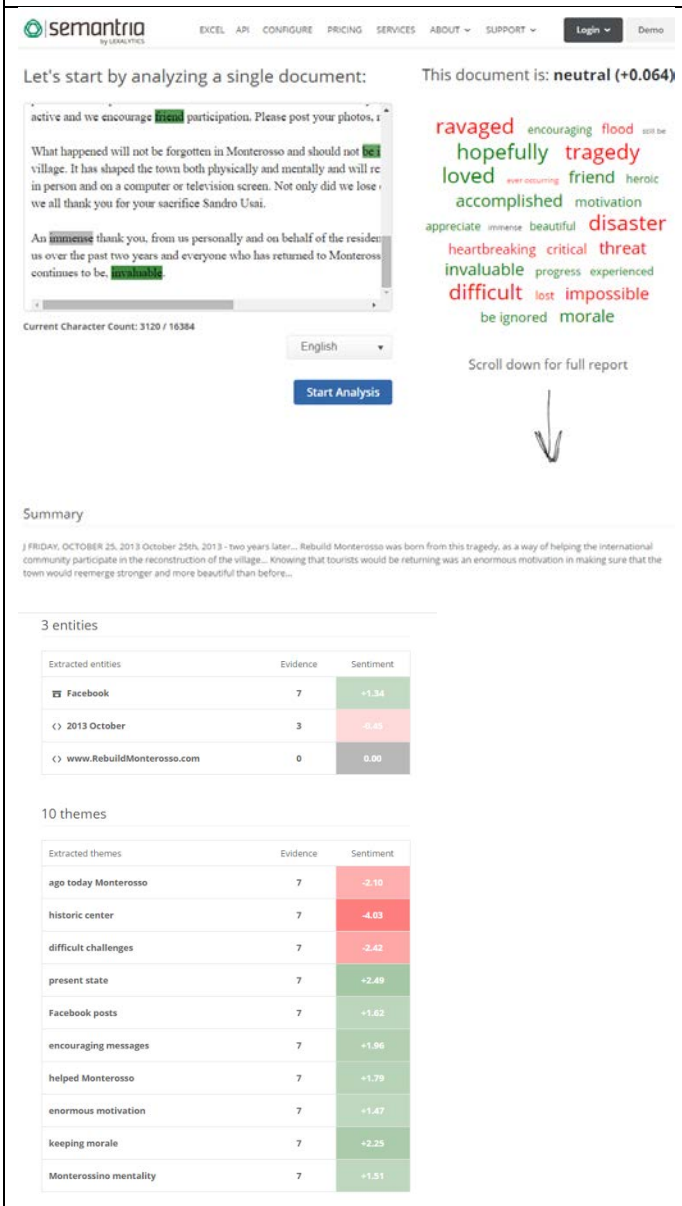
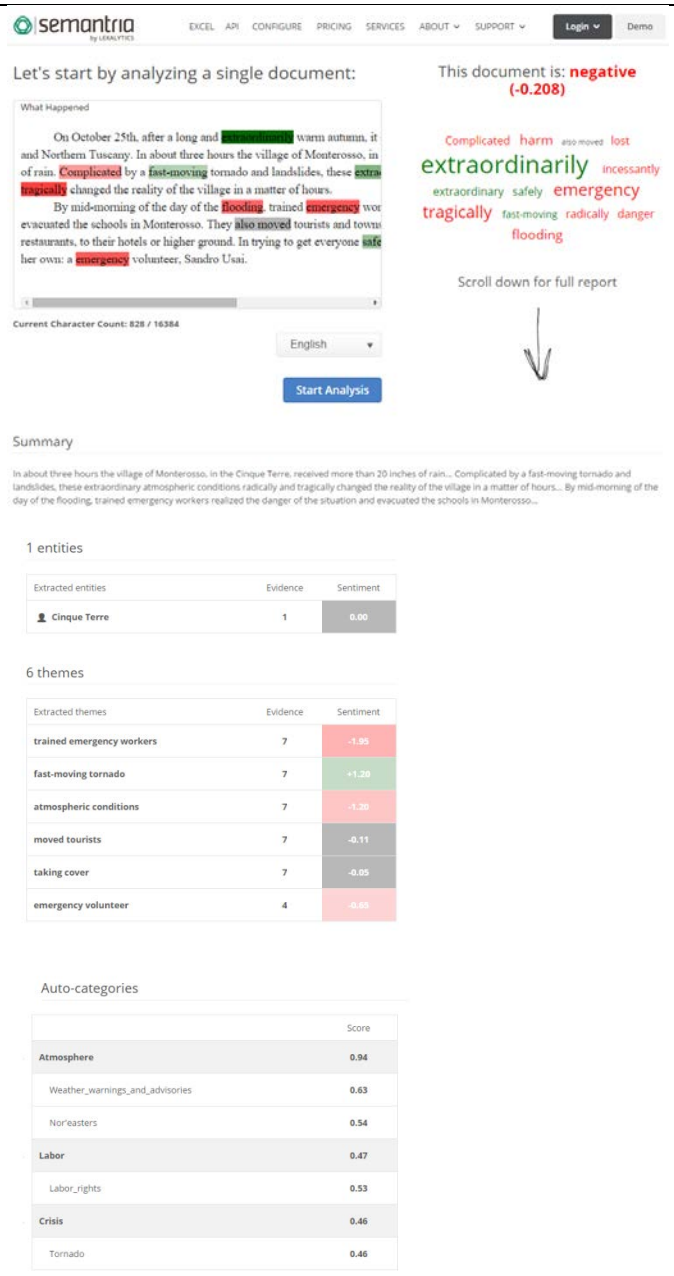
To create the graphic cloud of most recurrent words, considering all the tweets, we also set a procedure. Its pseudo-code is the following:

Consider the text of all the retrieved tweets
Remove punctuation and redundant white spaces
Transform text in lower case
Create a frequency table for each word
Create a column considering the length of the counted strings
Remove all the words having a length less than 3 characters
Remove all the words having a frequency less than 2
Create, using the package wordcloud(), a cloud of the remaining words

2.3. Results And Discussion

In the following section, we present the results obtained by the sentiment analysis performed on the contents of the website www.Rebuildmonterosso.com using Semantria software.

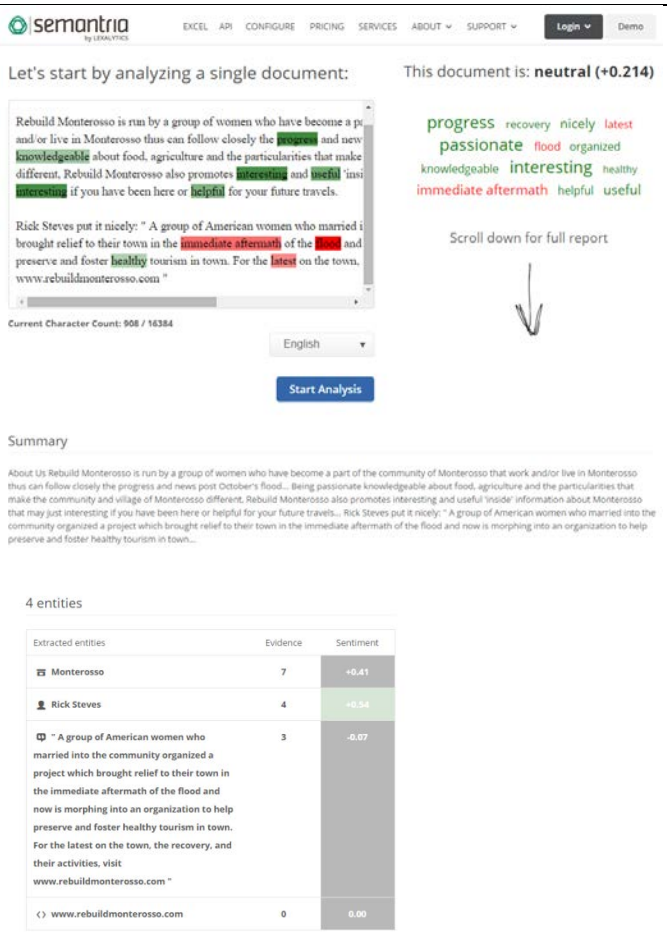
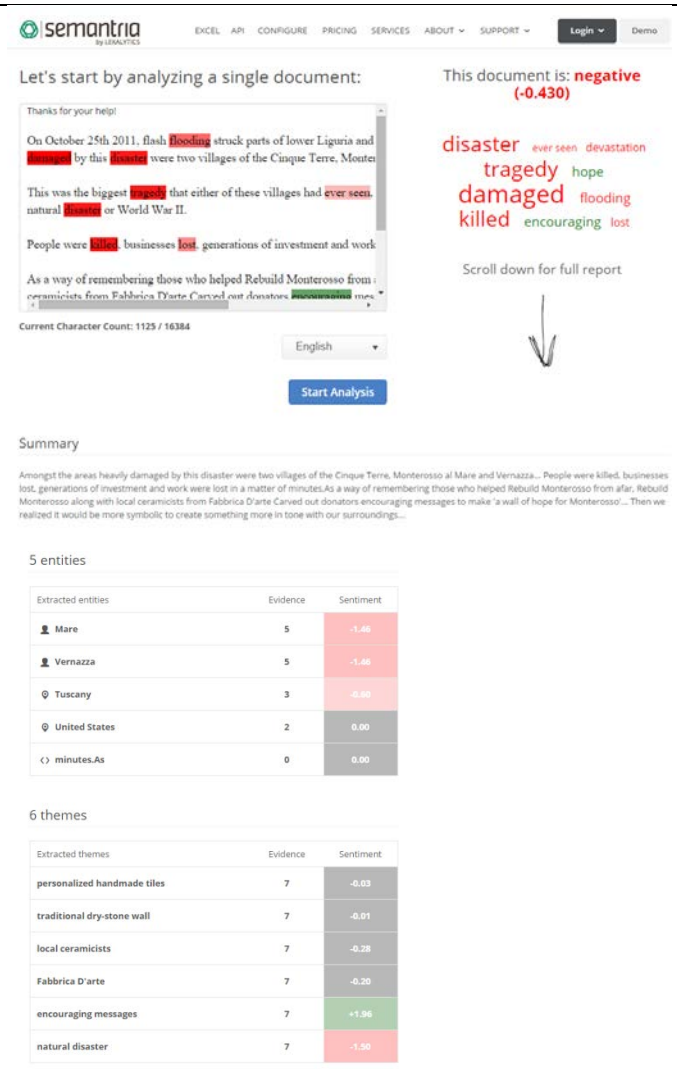
Figure 2. Sentiment analysis of the content of the "home" and "what happened" webpages from the website www.rebuildmonterosso.com performed using Semantria software.

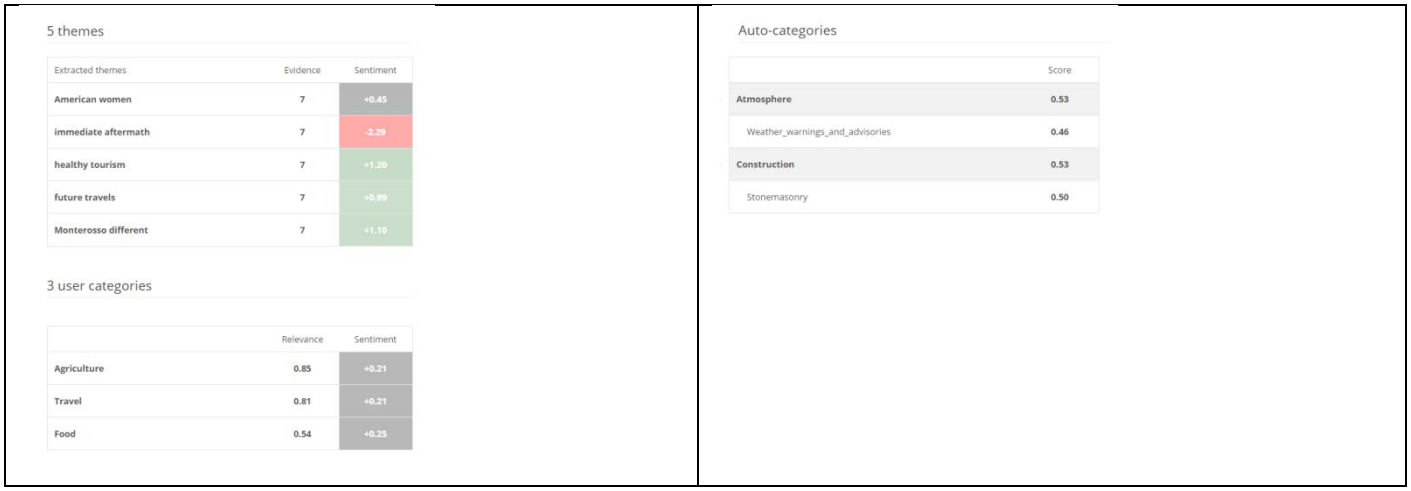
home	What happened
	

All texts delivered on the website are in English.

As you can see in figure 2, while the general sentiment of the “home” page is neutral, the mood of the page “what happened” is, as expected, negative. On the contrary, one might expect that the “Thanks for your help” web page would give a positive sentiment, while, as you can see in the following figure 3, the terms chosen seem to assign a general negative sentiment to the text, while the content of the “about us” web page is, rather obviously, neutral.

Figure 3. Sentiment analysis of the content of the "about us" and "thanks for your help" webpages from the website www.rebuildmonterosso.com, using Semantria software

About us	Thanks for your help
	



We have also considered the contents of the website www.Buongiornomonterosso.com, in the Italian version. The results obtained using Semantria software are shown and commented in figure 4.

Figure 4. Sentiment analysis of the content of the "about us" and "thanks for your help" webpages from the website www.rebuildmonterosso.com, using Semantria software

4 entities

Extracted entities	Evidence	Sentiment
Protezione Civile	4	-0.71
25 ottobre 2011	4	-0.40
Brugnato	1	+1.45
Sandro Usai	1	0.00

10 themes

Extracted themes	Evidence	Sentiment
anno intero	7	+0.56
immaginava la tragica evoluzione	4	+0.02
fronte ad una vera	4	-0.03
ansia per la conta	4	0.00
ore a Brugnato	4	+0.01
stazione di Monterosso	4	+0.15
mm il quantitativo	4	+0.38
è un giorno	4	+0.27
iniziata nella mattinata	4	+0.14
annullava il contorno	4	+0.03

2 user categories

	Relevance	Sentiment
Tempo	0.57	0.00
Elezioni	0.51	0.00

The sentiment analysis of the contents (in Italian) of the website "buongiornomonterosso.com" has been less significant because the website is rich with images and contains few textual parts.

We have to consider, in fact, that the website was created by the municipality of Monterosso to inform citizens and interested people about the reconstruction of the sites after the disaster. Therefore, most of the website is devoted to images, short comments and news.

The only page related to the actual disaster that occurred is called "Diario del disastro (25 ottobre 2011)" that is, in fact, a diary recalling all the events that happened on the day of the flood. As a result, the sentiment analysis could only be performed on this page as it would lead to a neutral result.

The result obtained with our procedures written in R confirm what we obtained with the described free tools. In particular, we have analyzed the webpage from the website www.rebuildmonterosso.com both using the database used in Minqing and Bong, 2004 and the scores obtained using the AFINN list of negative and positive words. Results are shown in table 1.

Table 1. Write summary table obtained by the sentiment analysis performed on the webpages of "www.rebuildmonterosso.com".

	Results obtained by Minqing and Bing's database of words	Scores obtained using the AFINN list of words			
Considered webpage	Score	Vneg o negV?	Neg	Pos	Vpos o posV
About us	neg: 0 ; pos: 12	0	0	9	0
Home	neg: 15 - pos: 20	0	12	19	0
Thanks for your help	neg: 9 - pos: 5	0	8	6	0
What happened	neg: 26 - pos: 25	0	21	20	0

The page "About us" generally resulted, with both procedures, more positive than negative.

Also the "Home" page resulted, in general, positive, but, considering the slight difference between the two positions, that could be seen as a neutral result, thus confirming the results obtained with the free tool Semantria.

The page "Thanks for your help" gave us a general negative sentiment, while in the page "What happened" we can observe a more negative mood even if, also in these two cases, the slight difference between the two positions could be seen as a neutral general feeling. This latter analysis seems to be more accurate than those performed using the tool Semantria.

Due to the use of Italian, on the page "Diario di un disastro (19 ottobre 2011)" from the website "www.buongiornomonterosso.com" we performed the sentiment analysis only using the Minqing and Bong database, translating terms into Italian.

In this case, we obtained the score: pos 6 and neg 19

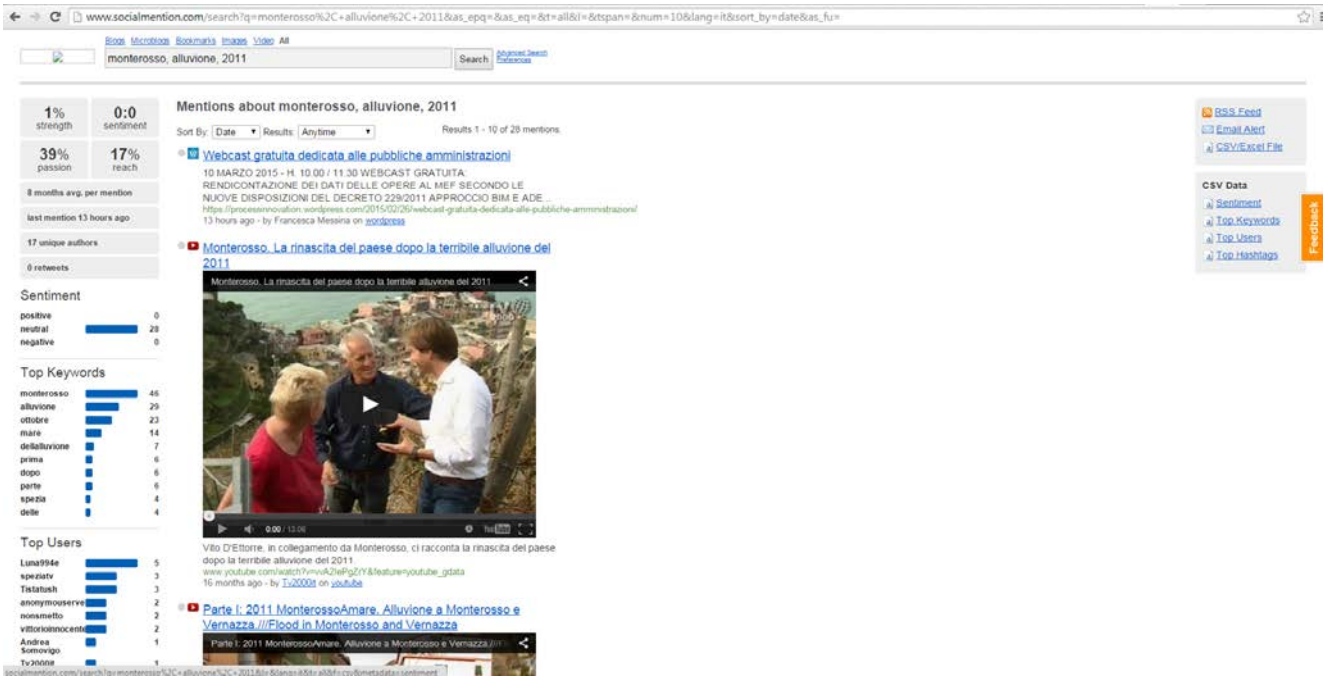
Unlike the neutral score obtained with the tool Semantria, in this case the general mood results negative.

In the following section we will discuss the results obtained performing the sentiment analysis using the tool Socialmention[7] on the web sphere.

We tried different keywords to perform the search all over the web, but the most significant ones where "monterosso", "alluvione" (flood), "2011".

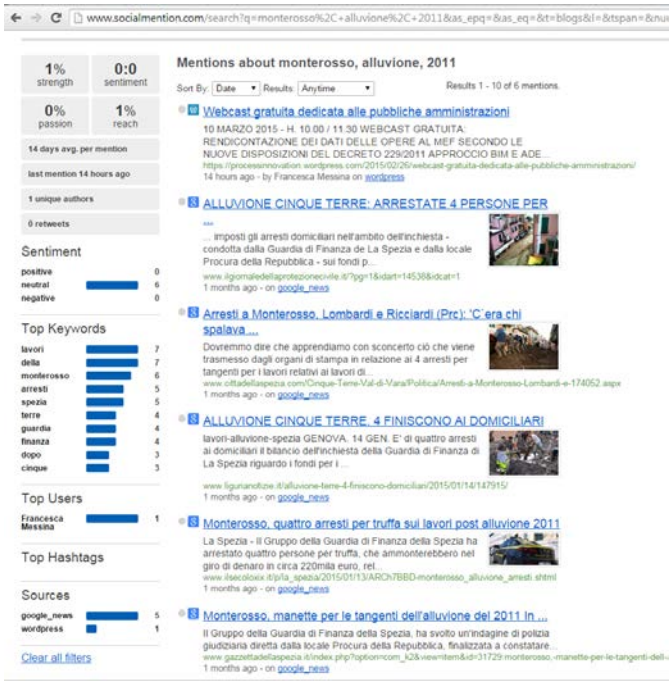
The most recurrent keywords linked to the event and extracted from the messages on the web are shown in figure 5 and are, in fact, strictly related to the event.

Figure 5. Global results obtained using the Socialmention tool on the social sphere.



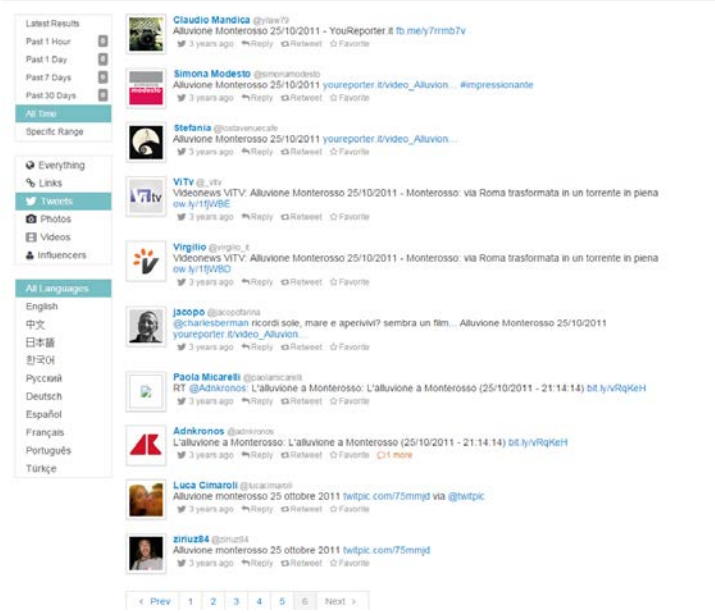
The detected general sentiment is neutral and most of the messages are in the form of videos on YouTube and derived from blogs (see figure 6).

Figure 6. Most comments consist in videos from the web.



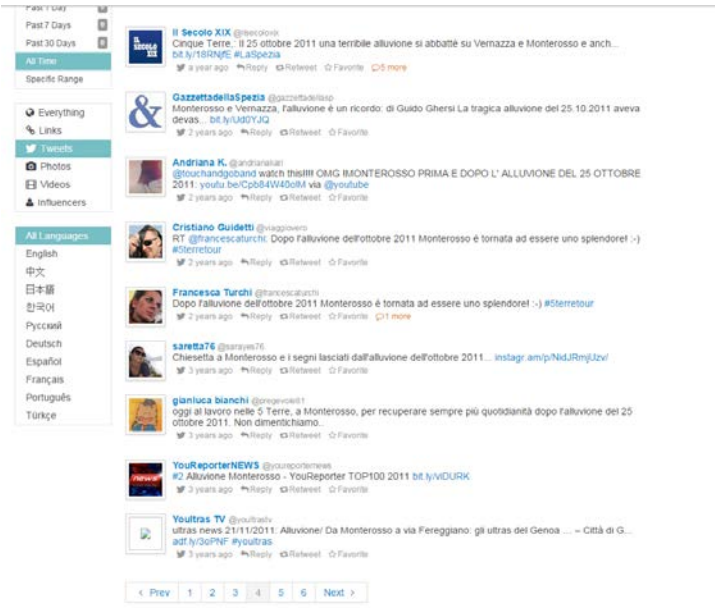
To complete our analysis, we show the results obtained analyzing tweet by Topsy. We have retreived all the tweets from 25 October 2011 to January 2015. As you can see in the following figure, the largest number of tweets has been posted in the first month after the flood, were mainly related to the damages caused by the disaster, and expressed a general negative sentiment.

Figure 7. Tweets delivered immediately after the flood.



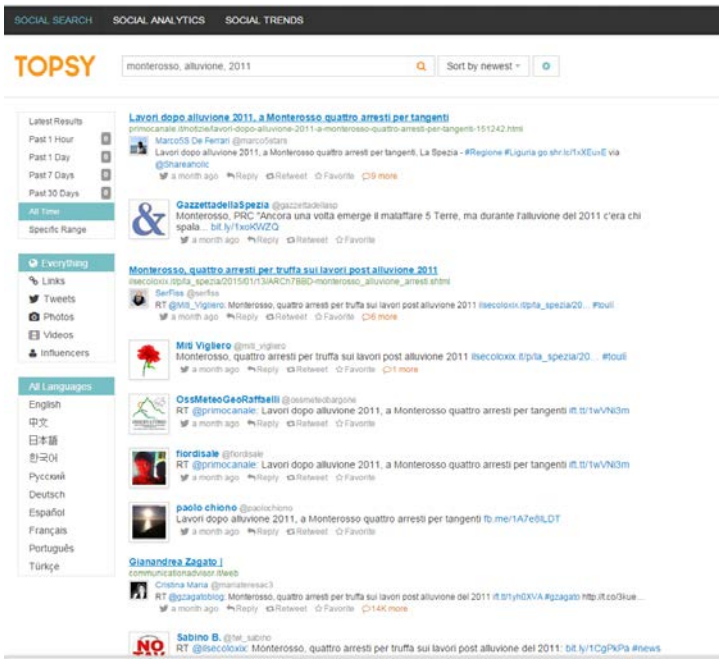
One year after the flood, tweets were much fewer and were divided into those dealing with recovery works and those aiming at reminding that the flood was not to be forgotten. The general mood is, in this case, positive.

Figure 8. Trends of tweets one, two and three years after the flood.



More recently, six months ago, the number of tweets drastically declined, while it became again high in the last month in the wake of the corruption news related to the reconstruction of Monterosso's sites. The sentiment is, therefore, generally negative.

Figure 9. More recent tweets.



As mentioned before, we have also performed sentiment analysis on tweets retrieved from the page https://twitter.com/RBMonterosso using a procedure that is specifically written in R environment. We considered all the 147 tweets from the user RBMonterosso from the day of the flood to present and analyzed them with the procedure. The results were the following:

3 tweets with score -2
6 tweets with score -1
102 tweets with score 0
31 tweets having score 1
5 tweets scored 2

Where the sign "-" indicates a negative feeling, the sign "+" stands for a positive feeling, while if the score is 0 that means we have a neutral feeling.

Our procedure also provides the cloud of words that are mostly used by the user @RBMonterosso, shown in figure 10.

Figure 10. The cloud of the most frequently used words by the user @RBMonterosso on Twitter obtained with our R written procedure.

4. Detection And Categorization Of Comments: Towards A New Model

In the following we propose a model to recognize and categorize positive, negative or neutral comments, based on machine learning methods. This solution guarantees higher reliability and presents the advantage of constant refining by simply increasing the number of cases used for the training of the methods.

4.1 Tools and methods

Considering the eterogeneicity of the textual contents present on social media, we used a pre-processing system for texts based on the recognition of "parts of the speech" (grammatical content), allowing us not to refer just to the frequency of the most recurrent words for the considered categories (positive, negative).

The procedure has been implemented using the statistical software R-cran. To implement the model for the algorithm extracting textual contents from the web and for the pre-processing of the obtained texts, we used several R-package, such as RTextTools; e1071; koRpus; NLP; tm; qdap; data.table; stringr; NLP; foreach; doParallel; compiler; lattice; ggplot2; caret; kernlab; plyr; textcat.

First of all, we needed to find a dataset of categorized comments, either for Italian and for English. In the latter case, our basis was a dataset of comments from several movies, categorized in positive and negative, extracted from the polarity dataset 2.0 (http://www.cs.cornell.edu/people/pabo/movie-review-data/).

As to the Italian language, considering the absence of a similar dataset, we needed to create and use an automatic algorithm extracting comments from Tripadvisor. This choice has been encouraged by the presence of a preference level (from 5 stars to 1 stars) accompanying each comment. 4 and 5 stars were categorized as positive comments, 1 and 2 as negative comments while three stars was regarded as neutral comment.

The extraction algorithm of comments from Tripadvisor is based on the manipulation of the content of the web pages, structured using markup languages, allowing us to access text and metadata (evaluation, date...) through the recall of their respective tags.

To encode the text content in the corresponding grammatical categories, we used TreeTagger, a tool for annotating text that recognizes the "part of the speech", created by the Institute for Computational Linguistics of the University of Stuttgart.

The dataset used for the English language contains 1,000 positive comments and 1,000 negative comments, 800 of which, for each category, have been used for the training of the models, while the remaining 200 were used for testing.

The Italian dataset derives from an extraction process of the comments delivered on Tripadvisor, using, the following tourist areas/sights: Museo della scienza e della tecnologia Leonardo da Vinci – Milan; Colosseo – Rome; Duomo – Milan; La Scala – Milan; Pompei – Naples; Palazzo dei Normanni – Palermo;

Museo egizio – Turin; Acquario di Genova – Genoa; Porto antico – Genoa; Museo della scienza – Trento; L'ultima cena – Milan; Torre di Pisa – Pisa; Canal Grande – Venice; Basilica di San Marco – Venice; Piazzale Michelangelo – Florence; Duomo – Florence; Campo Felice – l'Aquila; Stazione Centrale – Bologna; Stazione Centrale – Milan; Stazione Torino porta Susa – Turin; Stazione Termini – Rome; Stazione Centrale – Naples.

Also in this case we used 800 positive e 800 negative comments for the training of the supervised model.

Each comment has been divided into token and every token has been transformed into its grammatical component through the treetag (http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/; http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/treetagger.en.html). Each sentence has been therefore considered for the frequency of the distribution of the grammatical elements.

Each sentence was analyzed considering the length of the different token. From this analysis, we created a matrix containing the frequency of the different lengths of each sentence, arbitrarily choosing a maximum length of 14 characters.

We also used the amount of "positive" or "negative" words within comments as an input variable (Minqing and Bing, 2004).

The aim of this pre-processing is to avoid depending on the context in which the textual content is produced, and then on the frequency that a certain kind of word can assume in a certain situation and not in another.

4.2. Results And Discussion

Both for the English and for the Italian version, we created different models of svm and neural networks, whose precision (accuracy) is summarized in the following table.

Table 2. Precision (accuracy) obtained by the different models of svm used.

model	Italian accuracy	english accuracy
nnet	0,7258318	0,6317475
PCAnnet	0,7350409	0,6789958
svm linear	0,752004	0,6427962
lssvm radial	0,680904	0,6607845
svm radial	0,7508794	0,7347118
svm radial C	0,7463426	0,7278892
svm radial W	0,7553901	0,7352255

The results show a level of confidence that ranges between 70 and 76%. These findings provide hope for wide margins for improvement, through a rethinking of the variables to exclude those of low relevance in building the model.

The following figure shows the relevant variables detected using the best models in both languages.

Figure 10. Relevant variables detected using svm radial W in Italian.

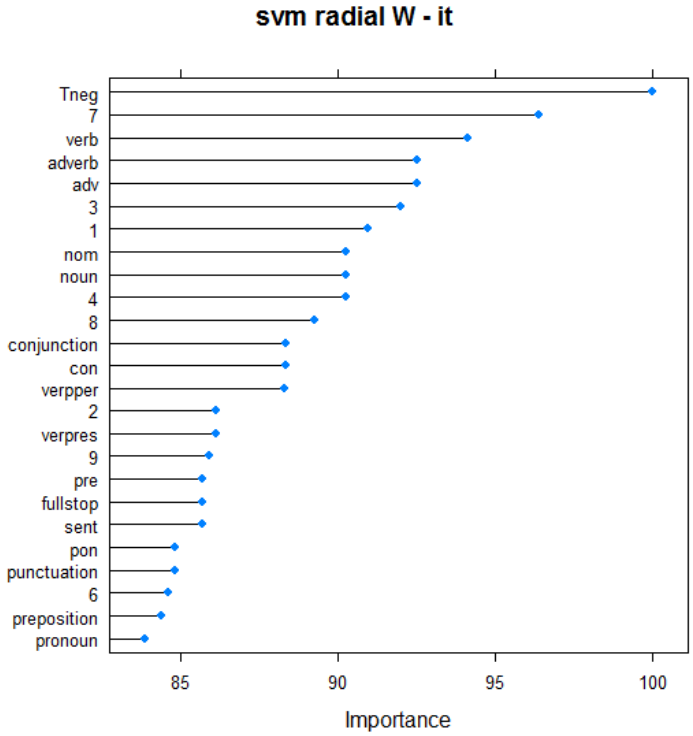
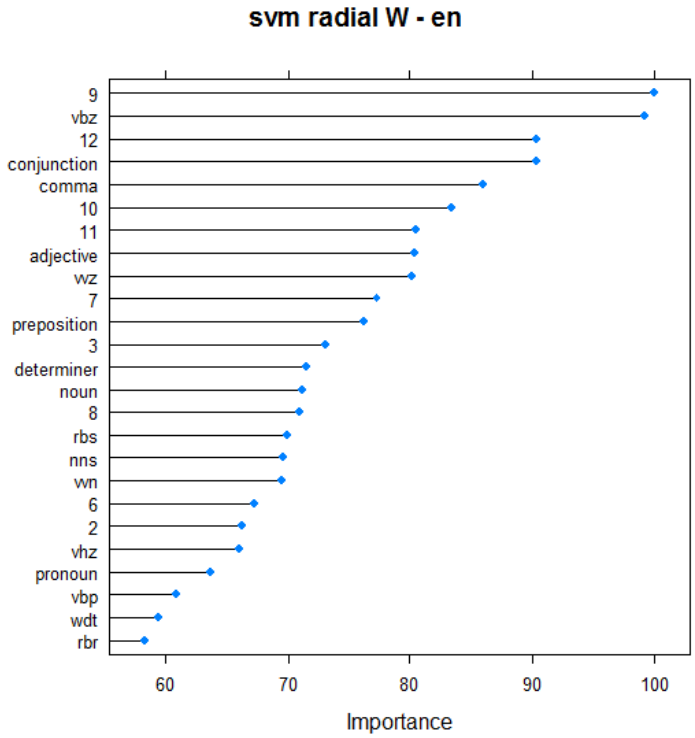


Figure 11. Relevant variables detected using svm radial W in English.



This methodology can be easily adapted to the text content delivered in the web, especially for the evaluation of texts freely written, without any limit to comply with.

To apply this method to the analysis of text content delivered on Twitter and other social media, it is necessary to adapt the model, considering, in the case of Twitter, only comments with 140 characters or less.

The analysis of the grammatical content of comments becomes a less reliable tool where neologisms, abbreviations, symbolic characters instead of words, or emoticons appear in the text, because these words and symbols contextualize the text within a frame that enhances or weakens or overturns the expressed meaning.

Comments made through web-scraping, limited to those whose content is below 140 characters, were little more than 100 for the negatives, and a few hundred for positives. These numbers limit the use of machine learning methods which require a much higher amount of data to reach satisfactory levels of reliability.

5. Conclusions And Further Developments

The aim of this study has been to explore the use of web resources to trace the discursive strategies enacted in the image restoration of the tourist destination of Monterosso. In our opinion, the findings are sufficient to be indicative of certain patterns in the data.

The use of the Internet in general, and of web 2.0 applications and services in particular, seems to confirm their functional role in delivering information and contributing to managing and – possibly recovering from – a crisis. And this appears even more striking when web-based communications from institutions (State, Region, Province, Tourist boards) are missing.

Bloggers stand out as expert testimonials, as members of communities to which they provide with updated information and suggestions. They send out persuasive and promotional messages expressing strong affection and thus a high level of involvement and empathy.

Moreover, weblogs constantly update information and follow the evolution of the situation: crisis, effects, emergency, rebuilding. This reporting is interestingly accompanied by story-telling, i.e. the narration of personal memories which contribute to depict Monterosso as appealing as it was before the flood and as appealing as it is now.

Similar choices in discursive strategies result in different and opposite suggestive effects. On the one hand, blogs underline personal experiences (especially bloggers' own experiences as visitors) and the commitment of the local community, together with volunteers and supporters engaged in the effort to take this land back to its beauty. On the other hand, blogs have evolved into more complex discursive forms and are delivered on the websites of institutions, organizations, newspapers and magazines. So in this multiplicity of voices, there are professional bloggers and individual bloggers, but the distinction between public and professional and private and spontaneous is increasingly blurring. The blogosphere has become the public sphere where opinions and consensus are realized through communicative actions, through the exchange of opinions and information. A blog "empowers individuals on many levels" (Blood 2002) and therefore one can argue that questions of power and social control come into play.

Our findings have also been supported by the sentiment analysis performed on the contents of the two considered websites.

The sentiment analysis showed that the predominant feeling was neutral over time, with a greater emphasis of the negative feeling in the stages immediately after the flood.

It must be said that little attention has been paid to texts on the two websites in terms of social web-oriented communication, and that the terms chosen by webpage authors do not seem to contribute to eliciting the expected feelings.

In addition, both sites are not updated. In fact, the latest changes date back to 2013.

Similar impressions emerged by analyzing web communication and in particular the tweets related to the flood. The tweets were numerous and showed a negative connotation during the first months after the flood, while they have dramatically decreased in the following periods, where we registered generally more positive messages, related to the reconstruction of the damaged areas.

In more recent months, the general sentiment delivered by tweets, which have increased in number again, is negative because it is linked to corruption news related to the processes of reconstruction.

In future works, we aim to refine our procedures using AI algorithms that seem more reliable (such as, for example, artificial neural networks) and also suitable to follow the sentiment over time. We also aim to refine the database of terms used to perform the sentiment analysis, with a view to creating a tool able to distinguish not only among different languages, but also different contexts and writing styles.

References

Bakhtin M.M. (1981). The Dialogic Imagination: Four Essays, University of Texas Press.

Blood, R. (2000, September 7). Weblogs: A History and Perspective. Rebecca's Pocket. Retrieved from http://www.rebeccablood.net/essays/weblog_history.html. Accessed the 5th of June 2006.

Bolter, J.D. (2001). Writing Space: Computers, Hypertext, and the Remediation of Print, Second Edition. Mahwah: Lawrence Erlbaum Associates.

Fairclough, N. (1989). Language and Power. London-New York: Longman.

Fairclough, N. (1992). Discourse and Social Change, Polity, Cambridge.

Fairclough N. (1995). Critical Discourse Analysis, London-New York: Longman.

Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. ICWSM, 7, 21.

Habermas, J. (2006). Contemporary political philosophy: an anthology. In Goodin R. E., Pettit P., (eds.). 2nd ed., Malden, MA, Blackwell.

Halliday, M.A.K. (1994). An Introduction to Functional Grammar, Edward Arnold, London, 94-58

Herring, S. C., Kouper, I., Paolillo, J. C., Scheidt L. A., Tyworth, M., Welsch, P., Wright, E. &
Yu, N. (2005). Conversations in the blogosphere: An analysis 'from the bottom up'. In proceedings of the Thirty-Eighth Hawaii International Conference on System Sciences. Los Alamitos, CA: IEEE Press.

Johnson, T.J., Kaye, B.K., et als.(2007). Every blog has its day: Politically-interested Internet users' perceptions of blog credibility. Journal of Computer-Mediated Communication, 13(1), article 6.

Labov, W. (1997). Some further steps in narrative analysis. Journal of Narrative and Life History 7,395-415.

Labov, W. Waletsky, J. (1972). Narrative Analysis. Journal of Narrative and Life History, 7: 138, p. 12.

Lohr, S. (2012). The age of big data. New York Times, 11.

McNeill, L. (2003). Teaching an Old Genre New Tricks: The Diary on the Internet. Biography: An Interdisciplinary Quarterly, 26, 24-48.

Miller, C.R., Shepherd, D. (2004). Blogging as a Social Action: A Genre Analysis of the Weblog. Into the Blogsphere, University of Minesota.

Minqing, H., & Bing, L. (2004). Mining and Summarizing Customer Reviews. In proceedings of the ACM SIGKDD International Conference on Knowledge, Discovery and Data Mining (KDD-2004), Aug 22-25, Seattle, Washington, USA

Nasukawa, T., & Yi, J. (2003, October). Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the 2nd international conference on Knowledge capture (pp. 70-77). ACM.

Pak, A., & Paroubek, P. (2010, May). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In LREC (Vol. 10, pp. 1320-1326).

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2), 1-135.

Tzvedan, T. (1990). The Origin of Genres. In Genres in Discourse, Cambridge, Cambridge U.P.
Toolan M. (1980). Narrative. A Critical Linguistic Introduction, London. Routledge.

Endnotes

1. http://www.buongiornomonterosso.com; http://www.rebuildmonterosso.com (accessed on 1/3/2015)
2. http://www.wordle.net/
3. http://www.semantria.com
4. https://semantria.com/support/resources/technology
5. http://www.r-project.org/
6. http://www.socialmention.com
7. http://www.socialmention.com