EmoTube: A Sentiment Analysis Integrated Environment for Social Web Content

Evangelia Polymerou
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
polyevan@gmail.com

Despoina Chatzakou
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
deppych@csd.auth.gr

Athena Vakali
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
avakali@csd.auth.gr

ABSTRACT

Mashup technologies offer numerous capabilities for innovative applications and services development and integration. Content accessing via mashups' interoperable and dynamic interfaces facilitates new enterprise models, while technological tools and smart techniques contribute to the development of integrated platforms. This work presents the principles and characteristics of the so-called "EmoTube" mashup, which is an integrated Web environment suitable for capturing and summarizing users' opinions expressed in their comments on YouTube videos. The main goal of this implementation is the visualization of users' opinions on a geo-located map for a better positioning of peoples' attitudes about various issues. Such summarization can be beneficial for several services such as recommendations and policy and decision making.

Categories and Subject Descriptors

D.2.6 [**Software Engineering**]: Programming Environments– *Integrated environments*

General Terms

Human Factors, Measurement

Keywords

Web 2.0, mashups, social networks, sentiment analysis, user-generated content

1. INTRODUCTION

The Internet is ubiquitous in peoples' life. With the existence of abundant sources and services of interconnected hypertext documents, users have the opportunity to seek information of their interest and cite their own views by participating in social networks, blogs, etc. from everywhere and in real time. Moreover, web users have access to large amount of data and as a consequence they can be informed of what is happening around them at any time. Web is considered as a vital tool for businesses since it facilitates the promotion of their products and/or services globally without effort and time via users contributed valuable information, such as relevant personal opinions, reviews and customer experiences.

WIMS '14, June 02 - 04 2014, Thessaloniki, Greece

http://dx.doi.org/10.1145/2611040.2611081

Multimedia content sharing has increased in a wide extent since web services, such as YouTube[1] (social video sharing platform) and Flickr[2] (social photo sharing platform) make users active contributors to the production of web content, which often lacks in structure. The need for a better organization and summarization of the available content leads to the appearance of mashups, websites or applications that use content from more than one sources. Mashups are organized in three general categories based on the targeted users and/or the provided content: customer, business, and data mashups. The integration of data from different sources in a unified way can lead to the creation of attractive and well-organized applications and websites able to satisfy users' needs without spending valuable time and effort. Finally, due to the abundance of freely distributed APIs, the development of mashups is considered in many cases as an inexpensive and straight-forward procedure.

This paper presents the EmoTube mashup, a web tool focused on the recognition and understanding of humans' behavior about daily issues based on the content provided from YouTube. The contribution of this work is linked to the up-to-date information of the end users about personal opinions as these are expressed in YouTube videos through the combination of multimedia content with geographical information. Here, we aim to overcome simplistic rating users' declarations by further analyzing YouTube metadata (video comments). Also, this application can be useful for understanding and identifying users' opinion fluctuations on different topics and in various contexts (e.g., geographical regions). Finally, EmoTube offers crowd pulse summaries to facilitate many services at a second level, through offering recommendations, proceeding to decision making and forming policies for social impact.

2. AN OVERVIEW OF MASHUP DESIGN AND DEVELOPMENT

The World Wide Web explosion along with the subsequent rise of the Semantic Web and the increase in user participation resulted in the development of new services, websites, technologies and protocols. As a consequence, the need for hybrid platforms capable of combining complicated and sometimes disconnected data and services (written in diverse languages and derived from two or more different sources) was born. Additionally, the interest in exploring users' opinions has increased rapidly and thus it has inspired the research community. The evaluation and the quantification of opinions and emotions expressed by users in social media is suitable for capturing the wisdom of crowds, the crowd pulse and the emergent trends.

[1] http://www.youtube.com/

[2] http://www.flickr.com/

In this section we briefly provide some basic principles and technologies related to the mashup applications and websites, while in the next section we discuss some important issues related to the sentiment analysis procedure. Table 1 provides a summarization of the most popular technologies used, based on the mashup technological domain.

Table 1: Mashups' categories and technologies

Principles	Technologies
Presentation-oriented mashups	**Connection with API services** XMLHTTPRequest objects, XML-RPC, JSON-RPC, REST and SOAP. **Presentation and Functionality** HTML/XHTML, CSS, JavaScript, AJAX.
Process-oriented mashups	Java, Python, PHP
Data-oriented mashups	**In-process** JavaScript, Jscript, DOM, XML, JSON **Out-of-process** Java, Python, XML

Mashups can be organized in three general categories: Presentation-oriented, Process-oriented and Data-oriented [9]. To ensure the integrating process of web applications' functionalities, the Service-Oriented Architecture (SOA), which is independent of any product or technology, is usually followed. In spite of the variable user interfaces and data resources, a common architectural pattern is feasible. For instance, all mashups can follow the principles of Representation State Transfer (REST) network protocol [6].

The presentation-oriented mashups constitute the integration of different user interfaces for the creation of a new unified application or web page. HTML/XHTML, CSS, JavaScript and AJAX technologies are used for the effective presentation and functionality of the presentation-oriented mashups. Presentation-oriented mashups mostly rely on data and user interface objects retrieved from websites through APIs, data feeds, etc. Libraries of Web 2.0 Application Programming Interfaces (APIs) are considered essential for the mashup development as they provide services for data collection from a wide range of sources and scripts for connecting web applications and instructions whenever they are needed. XMLHTTPRequest objects, XML-RPC, JSON-RPC, REST and SOAP are utilized to connect the mashup functionality with API services.

The process-oriented mashups target in the combination of functionalities from one or more external processes using programming languages such as Java, Python and PHP [9]. This kind of mashups allows sharing and reuse of services from diverse sources [1], often referred to as a Software-as-a-Service model (SaaS), a software delivery method that provides access to users via their web browser.

The data-oriented mashups combine data from two or more difference sources for creating an advanced web page or application suitable for providing consolidated and more in depth knowledge. In this type of mashups usually scripting languages and techniques such as JavaScript and Jscript are used. The data processes (e.g., sending, storing and retrieving) are usually carried out with the use of the XML structure, the JSON format and KML tools. They are also distinguished in two general categories: in-process data-oriented mashups and out-of-process data-oriented mashups. In the first subcategory data processing is conducted within the application or a web-page process (client-based). The results of data processing and analysis are usually depicted in a tree format via a Document Object Model (DOM) allowing in that way scripting techniques to have dynamic access, update and process of content. The out-of-process model is related to the interconnection of data from multiple sources (using Java, Python and XML) in order to create new data models within a remote host or server process (server-based). In contrast to the in-process model, the out-of process model is capable of applying data transformations, evaluating data validity, storing in a mashup server and returning the results directly to a client. An intermediate model, called hybrid, applies in- and out- process methods for data manipulation (e.g. enterprise mashups).

Regarding mashups applications and websites Lopez J. et al. [17] developed a CRM enterprise-oriented mashup based on the reuse of different source components such as databases, and SOAP and REST protocols for searching new customers. In [3] a methodology is proposed through which the end users can participate in the mashup development so as to self-define applications which satisfy their situational needs. Finally, [4] presents a work that its goal is to assess mashup quality in terms of the way that it is developed and propagated from the basic components to the final mashup application.

3. SENTIMENT ANALYSIS TECHNIQUES IN SOCIAL NETWORKS

The Web has entirely changed the way people communicate and express their thoughts and feelings. The increase of reviews, recommendations, comments, opinions and experiences – User-Generated Content (UGC) [11] – resulted in the need for understanding human opinions. Sentiment analysis is guided to this direction as it tries to translate, interpret, comprehend and analyze human opinions.

As a computational study of views, feelings (sentiments, emotions [12], [14]), assessments and attitudes expressed in texts, Sentiment Analysis [17] known as Opinion Mining [16] combines Natural Language Processing (NLP), Computational Linguistics and Text Mining to recognize contextual polarity (positive, negative, neutral) and classify texts based on sentiment. Classification [16], [17] has been broadly studied in order to aggregate sentiment-oriented sentences using techniques such as Polarity Classification, a methodology used for the removal of the objective sentences [14]. The subjectivity detectors are responsible for the automated division of opinionated sentences (subjective sentences) from those which describe only facts (objective sentences) [15]. Other approaches aim at the identification of the strength or weakness of the expressed opinions expressed in texts and examine the positive or negative aspect of them via a scaling system [8].

Two very popular methodologies have been shown to be usually tailored for the recognition of human's opinions, namely machine learning and lexicon-based techniques, the combination of which can lead to better results. In lexicon-based techniques the analysis is based on the use of specific lexicons, either targeted ones that are related to a specific subject of interest (for example medicine, movies, reviews) or more general lexicons that can be used in all cases resulting, in some cases, in negative impact on performance. Many research approaches [8], [13], [18], [19] rely on lexicons so as to recognize the Semantic Orientation of words or phrases in texts taking into consideration the Part-of-Speech (POS) parsers. For the recognition of opinions through lexicon-based techniques, tags, meta-tags and polarity rules are usually used. With the use of lexicons the identification of linguistics phenomena such as the intensity (e.g., good, quite good, very good) or the orientation of the words (e.g., good, not good) is feasibly.

The machine learning methodologies utilize the machine learning principles (subfield of Artificial Intelligent) for classifying texts by the overall sentiment. In [15] a text classification methodology is proposed based on machine learning methodologies and more specifically the Support Vector Machines (SVMs) and Naïve Bayes. Machine learning methodologies are relied on the extraction of suitable features such as unigrams, bigrams and POS tags, a word tagging system which marks up words in texts based on the corresponding parts of speech (e.g., adjectives [10], adverbs, nouns, verbs [2], [7]).

In our case, a lexicon-based methodology was considered more appropriate as one important disadvantage of the machine learning methodologies is the need for a large amount of data for conducting better results. Due to limitations in the YouTube API the number of comments sometimes can be quite small. Additionally, lexicon-based techniques can perform better over patterns that do not appear frequently, while machine learning methodologies are based on static trained datasets.

In general, Sentiment Analysis can be useful in many ways. A great amount of information can be analyzed in order to extract opinions or emotions automatically. Thus, businesses can identify at early stages the success of their products or services by monitoring customer opinions as they are continuously changing. As a consequence, they can improve their response time for better customer services and quality of their products and/or services. The understanding of human opinions/emotions, as these are imprinted in texts, encompasses many challenges. The social and educational background, culture and diverse experiences make even more difficult the procedure of recognizing the opinions expressed in texts. Particularly in social media, sentiments are not always obvious as the users usually express their opinions in an informal way, for instance with the use of abbreviations (e.g., the use of word "lol" instead of the expression "laugh out loud"). An additional difficulty is the ambiguity, namely the interpreting of a text with contradictory ways (the same phrase can concurrently express a positive and negative opinion). Although there are several efforts in understanding the way that people express their views using key-words or groups of key-words, the analysis mainly focuses on simple terms and there are no standards for sentiment classification until now.

4. THE EmoTube FRAMEWORK

In this section we present the EmoTube framework, a web page mashup, which integrates Web 2.0 multimedia content (YouTube) with geographical information (via Google Maps API) applying at the same time sentiment analysis for capturing users' opinions.

Table 2 outlines an overview of the used technologies and methodologies.

The goal of EmoTube is the provision of a unified framework suitable for demonstrating (in an organized and attractive way) peoples' opinions regarding specific topics around different regions -cities- worldwide. This is realized by utilizing the valuable information inherent in the metadata extracted from the YouTube videos. The proposed mashup constitutes an out-of-process data-oriented mashup and according to the approach depicted in Figure 1 a three-step procedure is proposed at the back-end to offer the EmoTube web interface at the front-end. Next, we briefly describe the following steps.

Table 2: Overview of technologies and methodologies

Procedure/Entity	Technologies/Methodologies
Data sources	YouTube API, Google Maps API
Data collection	Python crawler
Data cleaning (Back-end)	Python, NLTK library
Data retrieval/formulation (Back-end)	XML format
Sentiment analysis (Back-end)	Lexicon-based methodology (SentiWordNet)
Data/Results presentation (Front-end)	JavaScript, HTML, CSS

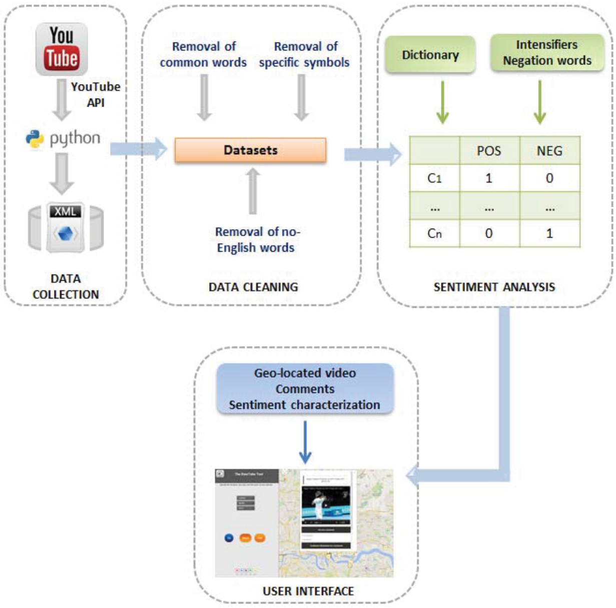

Figure 1: The proposed framework.

4.1 The EmoTube Architecture

The first step of the whole procedure is the data collection using the YouTube API and a Python crawler. Data collection focuses on the retrieval of specific categories of YouTube videos (e.g., news, sports). Although YouTube API provides an abundance of information for each video, in our case we retain specific information (video id, related URL, title, publication date, geo-location, view count, rating and relevant comments), necessary for the concept of the built mashup. So far, EmoTube collects YouTube videos of five different categories for two popular cities, London and New York (based on the geolocation of the uploaded video). The location list is easily extensible. The collected data are structured in XML format.

The preprocessing step involves the data cleaning procedure, which is performed in order to remove the semantically irrelevant information for reducing the noise in the available data (any portion which does not contribute to sentiment detection). More specifically, this step includes the removal of some common words, such as "a", "about", "is", "was", "for", "by", etc and the numbers. Words that do not appear in any formulation of the

English dictionary are being removed (using the NLTK[3] library). Finally, some punctuation symbols, such as “!”, “;” are also being removed as they do not offer any useful information.

4.2 Sentiment Analysis Approach

After the dataset’s cleaning up step, the final step is the recognition of opinions expressed in the videos’ comments. For extracting the opinion orientation of the comments, a lexicon-based methodology is being applied. Table 3 provides the necessary notation.

Table 3: Basic symbols notation

Symbol	Definition
ci	Comment ci, where i=1,..,n
ECi	Set of opinionated words of comment ci
ELi	Set of emoticons of each comment ci

To capture the opinionated words we use a widely known lexicon, the SentiWordNet[4], which assigns scores to each word on the basis of how positive or negative they are. For a more accurate result, in the sentiment analysis procedure the negation words (such as not, no, isn’t, can’t, wouldn’t) and the intensifiers (for example less, hardly, almost, very, quite) are taken into consideration. For example the expression “very good” expresses stronger opinion in relation to the word “good”, while the expression “not good” has different meaning from the word “good”.

The consideration of the intensifiers in the calculation of the overall score of an opinionated word ec_{i1} is carried out with the following formula:

$$SCI(ec_{i1}) = (1 + score(intens_j)) \times score(ec_{i1}) \qquad (1)$$

where $intens_j$ is the intensity score of an intensifier and $score(ec_{i1})$ is the sentiment score of an opinionated word ec_{i1} based on the SentiWordNet lexicon.

Regarding the consideration of the negation words in the total score of an opinionated word the next approach is followed:

$$SCV(ec_{i1}) = 1 - score(ec_{i1}) \qquad (2)$$

Another important parameter for the calculation of the sentiment score of each comment is the consideration of the emoticons. Emoticons are the pictorial representation of facial expressions that visualize peoples’ mood. Specifically, the set $El_i = \{el_{i1},\ldots,el_{ir}\}$ contains the emoticons extracted from each comment c_i. For example, in the comment “The worst game ever :-(”, the symbol “:-(” is considered as emoticon, and thus El_i = {:-(}. The score of each emoticon is based on a proposed lexicon from the University of Maryland, Baltimore[5]. This methodology is a simplified version of the methodology proposed in [5].

Algorithm 1 provides the pseudo-code for the calculation of the sentiment score of all the comments retrieved from the YouTube videos. The first step involves a data preprocessing phase, where the non-valid words from YouTube comments (line 2) and the words that do not carry any opinionated information based on the emotional lexicon (SentiWordNet) are removed (line 3). In the next step, the calculation of the total score of each comment is conducted (line 8) taking into consideration the way that the intensifiers (line 5) and the valence shifters (line 6) affect the words’ meaning.

Algorithm 1

Input: The set C of n comments, the dictionary.

Output: The opinion orientation for each comment.

1: /*Preprocessing of data*/
2: C* = CleanData(*C*)
3: EC = FindOpinionWords(*C**)
4: /*Calculation of the emotional score for each word based on intensifiers and valence shifters*/
5: SCI = CalculateScoreIntensifiers(*EC, intensifiers*)
6: SCV = CalculateScoreValenceShifters(*EC, valence shifters*)
7: /*Calculation of the total score for each comment*/
8: SC = CalculateTotalScore(*C*, SCI, SCV*)

5. IMPLEMENTATION AND SCENARIOS

After the analysis of the necessary content the final step is the integration and the visualization of all the available and necessary information in a unified way. This entails the dynamic presentation of the multimedia content on a map taking into consideration the location, the category and the time period of each video. The EmoTube front-end environment was developed in a simplified way in order to ensure an easy navigation through the available information.

Users have to specify three criteria in order to retrieve the needed information in the map: area, topic and year. After they make their choices all the related information is imprinted on the map: for each video a related pin appears on the map based on its geo-location (the place from which the video was uploaded). Finally, the user can “zoom in” each video and see related information (video, recent comments, sentiment distribution of comments).

The EmoTube front-end environment was developed using widely known web technologies, such as JavaScript, HTML, CSS and by leveraging the Google Maps API for the creation of the map, which permits the provision of a friendly interface for pleasant and easy navigation (Figure 2).

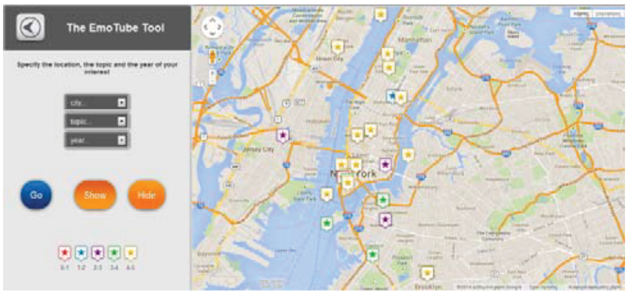

Figure 2: The front-end environment of the EmoTube.

As already mentioned, the system encompasses three mandatory options for the appearance of the desired results on the map: location, topic and time period. Until now the application covers five topics (comedy, entertainment, music, news, sports from 2008 to 2012), however their extension is a quite straight-forward procedure. The datasets collected so far are related to two different regions, London and New York. Users have the opportunity to see each video's title by hovering on the corresponding marker, and then they can “zoom in” in it in order

[3] http://nltk.org/

[4] http://sentiwordnet.isti.cnr.it/

[5] http://www.csee.umbc.edu/

to see its contents. Except from the title and the video itself, for each video the five most recent comments are presented in a pop-up window and if there are more, users have the opportunity to navigate through them with the "more comments" option. They can also see the polarity of the related video based on its comments. The results of the polarity detection are depicted on a pie chart for a better visualization. Finally, the popularity of each video is based on the ratings retrieved from the YouTube API. Each video rating ranges from low (0) to high (5) and it is illustrated via the corresponding pin color.

5.1 Usage Scenarios

In this section, two use case scenarios are provided for a better demonstration of the proposed framework's usefulness. Additionally, we analyze the users' behavior based on the results extracted via the sentiment analysis process.

In the first use case scenario, London and Sports in 2012 were selected from the corresponding drop-down lists (see Figure 3). Videos with yellow markers that outweigh have very high ratings (4-5). This indicates that the users' attitude towards the videos related to sports at that time period was positive (which is consistent with the results of the sentiment analysis process in the related comments).

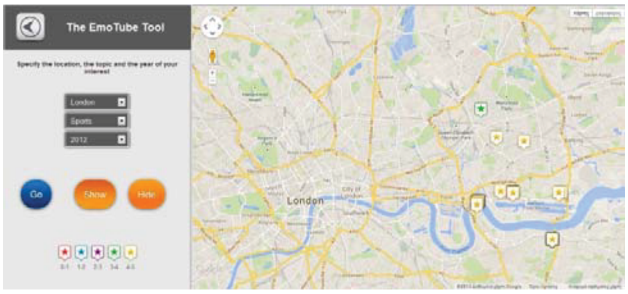

Figure 3: First use case.

By entering the info-window of some yellow markers, we can access the sentiment analysis results which reveal that the majority of comments are positively inclined. For example, in Figure 4 we observe the results of the sentiment analysis process of the video entitled "Squash: HotShots – Saurav Ghosal - Skills – EP5" and uploaded in the O2 Arena (8 miles far from the city center).

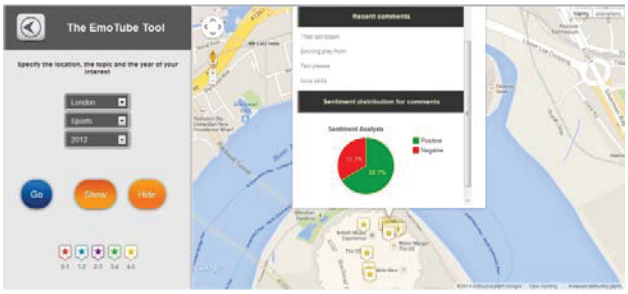

Figure 4: Results in the first use case - video 1.

It is obvious that most of the comments posted by YouTube users for this video are positive (or neutral which is not presented in the pie chart). In particular, the positive emotional state of users is at 66.7%, while the negative one stands at 33.3%. This means that the above video is clearly positive as a whole. For proof of concept, some related comments are the following: "Exciting play from Saurav!", "Nice skills" and "That last boast was a thing of beauty".

On the other hand, if we "zoom in" videos with the lowest ratings at the same region, we observe that the related ratings are negative. This fact is also expressed in the results derived by the sentiment analysis procedure. For instance, if we see the video content of a red marker named "Olympic Boxing Scandal Azerbaijan. Video of Decision and Apology", we observe that the overall positive score is 44.1%, while the negative score is 55.9%. Some related comments with negative emotional tension are: "The most corrupt event ever" and "How did they ever think they could get away with that it was so blatantly obvious, even by the way he stood. Good on the crowd for booing" (Figure 5).

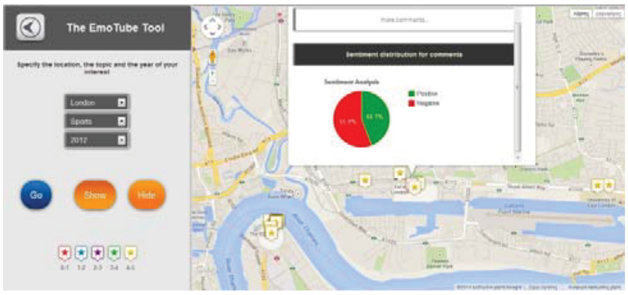

Figure 5: Results in the first use case – video 2.

In the second use case scenario, New York and News in 2010 were chosen from the corresponding drop-down lists (see Figure 6). In that case, we observe that the ratings vary in the different videos (gold, green, purple, blue pins), unlike the previous case where the most videos were highly rated. Studying the videos with News content we observe that in most cases they are related with the dissemination of negative news. As a result, the topic inherent subject (e.g. news, outbreaks) drives users' emotions in expressing more negative opinions. Thus, in that case we expect that videos with high ratings correspond mostly to negative news.

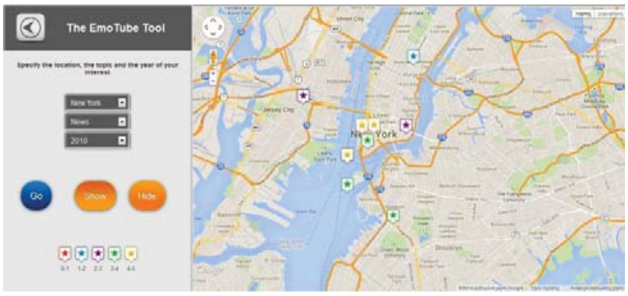

Figure 6: Second use case.

Next, we study a video with 3-4 rating (green marker) which was uploaded in New York. The video with the title "WTC ***HARD EVIDENCE*** explain this if you can" reveals that most users express their dissatisfaction for the related event. This is clearly shown in the written comments such as "WHAT HE DONE I Believe should be classed as murder" and "…the plane got too scared it crapped itself" (Figure 7).

According to the sentiment analysis procedure, negative comments amount to 66.7%. On the contrary, some users express their concerns regarding the provided news (e.g., "yes I agree a plane was crashed into that building but was it really a terrorist act or just a lame excuse to go war…”) and as a result positive ones stand at 33.3%. When observing the pie chart it is easy to figure out that the above video is more negative than positive as a whole.

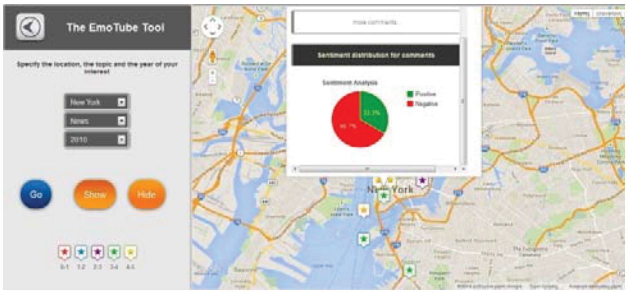

Figure 7: Results in the second use case.

6. CONCLUSIONS AND FUTURE WORK

The development of Web 2.0 has led to the emergence of new services, applications and technologies suitable for sharing and distributing user-generated content. At the same time, the need for well-presented information leads to the development of services and applications suitable for presenting the available content in a more organized manner (via mashups).

In this paper we propose a framework that utilizes the YouTube content and Google Maps, along with its metadata to provide a platform which summarizes users’ emotional dispersion. EmoTube manages to integrate topic, time and location, as well as to effectively summarize and visualize latent information acquired via the content's analysis, such as the people’s opinions distribution. The recognition of people’s opinions is conducted with the use of a sentiment analysis methodology, which seems to successfully capture the polarity expressed in the users’ comments.

The proposed framework is beneficial for either individuals or companies who intend to monitor consumers’ trends. Enterprises will be able to examine the users’ opinions regarding their products, helping them in that way to adjust their communication strategies so as to increase the users’ satisfaction.

In the future we intend to provide a more in depth analysis of the YouTube videos using specific emotions (e.g, anger, disgust and joy) for capturing the crowd pulse in a more fine-grained way. Finally, we aim at the development of a more automatic mechanism suitable for enriching the available content based on the users’ needs.

7. ACKNOWLEDGMENTS

The research leading to these results has received funding from the EU FP7 EINS under grant agreement No 288021.

8. REFERENCES

[1] S. Baden, D. Lefebvre, J. Dar, M. Langner, M. T. Yu, “Web application hybrid”.

[2] F. Benamara, C. Cesarano and D. Reforgiato, “Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone”, *ICWSM-07 Boulder*, CO USA, 2007.

[3] C. Cappiello M. Matera, M. Picozzi, G. Sprega, D. Barbagallo, C. Francalanci, “DashMash: A Mashup Environment for End User Development”, 11th International Conference, ICWE 2011, Paphos, Cyprus, June 20-24, 2011.

[4] C. Cappiello, F. Daniel, M. Matera, C. Pautasso, “Information Quality in Mashups”, Internet Computing, Volume 14, Issue 4, pp. 14-22, July-Aug 2010.

[5] D. Chatzakou, V. Koutsonikola, A. Vakali and K. Kafetsios, “Micro-Blogging Content Analysis via Emotionally-Driven Clustering”, *ACII*, Geneva, 2013.

[6] L. Clarkin and J. Holmes, “Enterprise Mashups”, *MSDN Architecture Center*, Journal 13.

[7] Y. Dang, Y. Zhang and H. Chen, “A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews”, *IEEE Computer Science, IEEE Intelligent Systems*, Vol. 25. 4, pp 46-53, July.-Aug. 2010).

[8] A. Esuli and F. Sebastiani, “SentiWordNet: A Publicly Available Lexical Recource for Opinion Mining”, In *Proceedings of 5th Conference on Language Resources and Evaluation, LREC-06*. Genova, Italy, pp 417-422, 2006.

[9] J. J. Hanson, “Mashups: Strategies for the Modern Enterprise”, *Mashups Styles, Techniques and Technologies*, Chapter 1, pp 28-25, May 2009.

[10] V. Hatzivassiloglou and J. Wiebe, “Effects of Adjective Orientation and Gradability on Sentence Subjectivity”, In *Proceedings of the International Conference on Computational Linguistics (COLING-00)*, 2000.

[11] B. Liu, “Sentiment Analysis and Subjectivity”, Handbook of Natural Language Processing, Second Edition, Taylor and Francis Group, Boca, Chicago, USA, 2010.

[12] B. Liu, “Sentiment Analysis and Opinion Mining”, Chicago, USA, May 2012.

[13] Y. Mejova, “Sentiment Analysis: An Overview”, Nov. 2009.

[14] B. Pang and L. Lee, “A Sentimental Education: Sentiment Analysis using Subjectivity Summarization based on Minimum Cuts”, In *Proceedings of the Association for Computational Linguistics* (ACL), pp. 271-278, 2004.

[15] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis”, Foundations *and Trends in Information Retrieval*, Vol. 2, pp. 1-135, 2008.

[16] K. Scherer, “Trends and Developments: Research on Emotions”, *Social Science Information, Sage Publications*, Vol. 44. 4, pp. 695-729, 2005.

[17] C. Strapparava and R. Mihalcea, “Learning to Identify Emotions in Text”, SAC '08 Proceedings of the 2008 ACM symposium on Applied computing, pp. 1556-1560, 2008.

[18] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, “Lexicon-based methods for Sentiment Analysis”, Computational Linguistics, Volume 37 Canada, pp. 268-307, Sept. 2010.

[19] K. Tsagkalidou, V. Koutsonikola, A. Vakali and K. Kafetsios, “Emotional Aware Clustering on Micro-blogging Sources”, ACII'11 Proceedings of the 4th international conference on Affective computing and intelligent interaction, pp. 387-396.