



Department of Informatics Aristotle University of Thessaloniki



“EmoTube: A Sentiment Analysis Integrated Environment for Social Web Content”

Evangelia Polymerou, Despoina Chatzakou and Athena Vakali

4th International Conference on Web, Intelligence, Mining and Semantics
Thessaloniki, Greece, June 2014

Contents

- Introduction
- Design and Development of Mashups
 - Principles and Technologies
- Sentiment Analysis Approaches in Social Networks
- The EmoTube Framework
 - Mashup Architecture
 - Lexicon-based methodology
- The EmoTube Website
 - Use case scenario
- Conclusions and future extensions

Purposes

- The development of an integrated web environment – **mashup**
- The visualization of users' **opinions**
 - expressed in their comments on YouTube videos
 - depicted on a **geo-located map**
- The recognition of personal opinion in **multimedia content**

Theoretical Background

- **Mashups** – Integrated platforms that combine complicated data from more than one sources
- Customer, business and data mashups
- Design and development

Table 1. Mashup categories and technologies

Principles	Presentation-oriented		Process-oriented	Data-oriented	
Technologies	XMLHTTP Request objects XML-RPC JSON-RPC REST SOAP	HTML XHTML CSS JavaScript AJAX	Java Python PHP	JavaScript Jscript DOM XML JSON	Java Python XML

- Service-Oriented Architecture (SOA)
- Representation State Transfer (REST) network protocol

Theoretical Background

- **Mashups** – Integrated platforms that combine complicated data from more than one sources
- Customer, business and data mashups
- Design and development

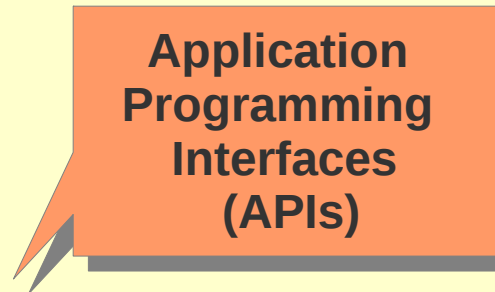


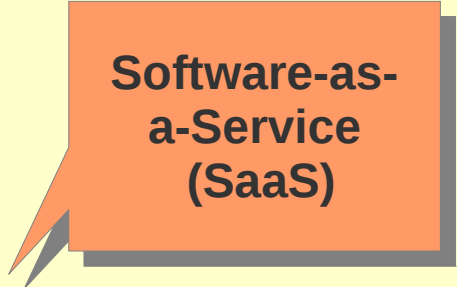
Table 1. Mashup categories and technologies

Principles	Presentation-oriented		Process-oriented	Data-oriented	
Technologies	XMLHTTP request objects XML-RPC JSON-RPC REST SOAP	HTML XHTML CSS JavaScript AJAX	Java Python PHP	JavaScript Jscript DOM XML JSON	Java Python XML

- Service-Oriented Architecture (SOA)
- Representation State Transfer (REST) network protocol

Theoretical Background

- **Mashups** – Integrated platforms that combine complicated data from more than one sources
- Customer, business and data mashups
- Design and development



Software-as-a-Service (SaaS)

Table 1. Mashup categories and technologies

Principles	Presentation-oriented		Process-oriented	Data-oriented	
Technologies	XMLHTTP Request objects XML-RPC JSON-RPC REST SOAP	HTML XHTML CSS JavaScript AJAX	Java Python PHP	JavaScript Jscript DOM XML JSON	Java Python XML

- Service-Oriented Architecture (SOA)
- Representation State Transfer (REST) network protocol

Theoretical Background

- **Mashups** – Integrated platforms that combine complicated data from more than one sources
- Customer, business and data mashups
- Design and development

Document
Object Model
(DOM)

Table 1. Mashup categories and technologies

Principles	Presentation-oriented		Process-oriented	Data-oriented	
Technologies	XMLHTTP request objects XML-RPC JSON-RPC REST SOAP	HTML XHTML CSS JavaScript AJAX	Java Python PHP	JavaScript Jscript DOM XML JSON	Java Python XML

- Service-Oriented Architecture (SOA)
- Representation State Transfer (REST) network protocol

Sentiment Analysis

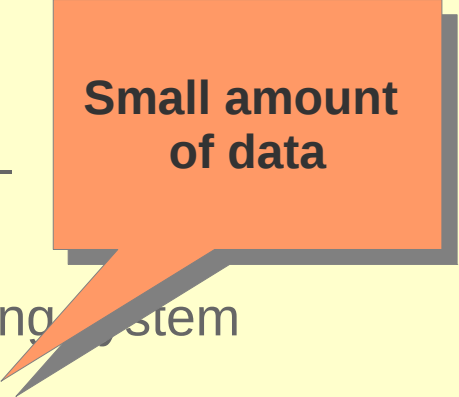
- **SA** – Opinion Mining (OM) → Natural Language Processing (NLP), Computational Linguistics and Text Mining
- Approaches
 - **Polarity Classification** removes objective sentences – subjectivity detector
 - **Identification of strength** or weakness in texts – scaling system
 - **Lexicon-based techniques**
 - Semantic Orientation (SO) – Part-of-Speech (POS) parsers
 - the intensity and the orientation of words
 - **Machine Learning methodologies**
 - Support Vector Machines (SVMs)
 - Naïve Bayes
 - unigrams, bigrams, and POS tags

Sentiment Analysis

- **SA** – Opinion Mining (OM) → Natural Language Processing (NLP), Computational Linguistics and Text Mining

- Approaches

- **Polarity Classification** removes objective sentences – subjectivity detector
- **Identification of strength** or weakness in texts – scaling system



Small amount
of data

➔ **Lexicon-based techniques**

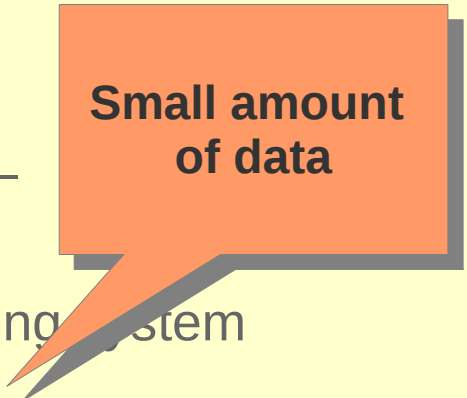
- Semantic Orientation (SO) – Part-of-Speech (POS) parsers
 - the intensity and the orientation of words
- **Machine Learning methodologies**
 - Support Vector Machines (SVMs)
 - Naïve Bayes
 - unigrams, bigrams, and POS tags

Sentiment Analysis

- **SA** – Opinion Mining (OM) → Natural Language Processing (NLP), Computational Linguistics and Text Mining

- Approaches

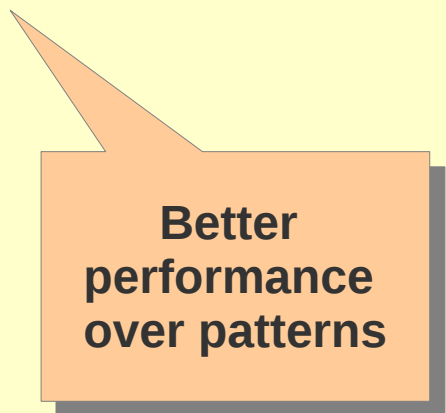
- **Polarity Classification** removes objective sentences – subjectivity detector
- **Identification of strength** or weakness in texts – scaling system



**Small amount
of data**

- ➔ **Lexicon-based techniques**

- Semantic Orientation (SO) – Part-of-Speech (POS) parsers
 - the intensity and the orientation of words
- **Machine Learning methodologies**
 - Support Vector Machines (SVMs)
 - Naïve Bayes
 - unigrams, bigrams, and POS tags



**Better
performance
over patterns**

The EmoTube Framework

- Web 2.0 multimedia content
- Geo-information
- **Sentiment Analysis** in users' opinions



out-of-process
data-oriented mashup

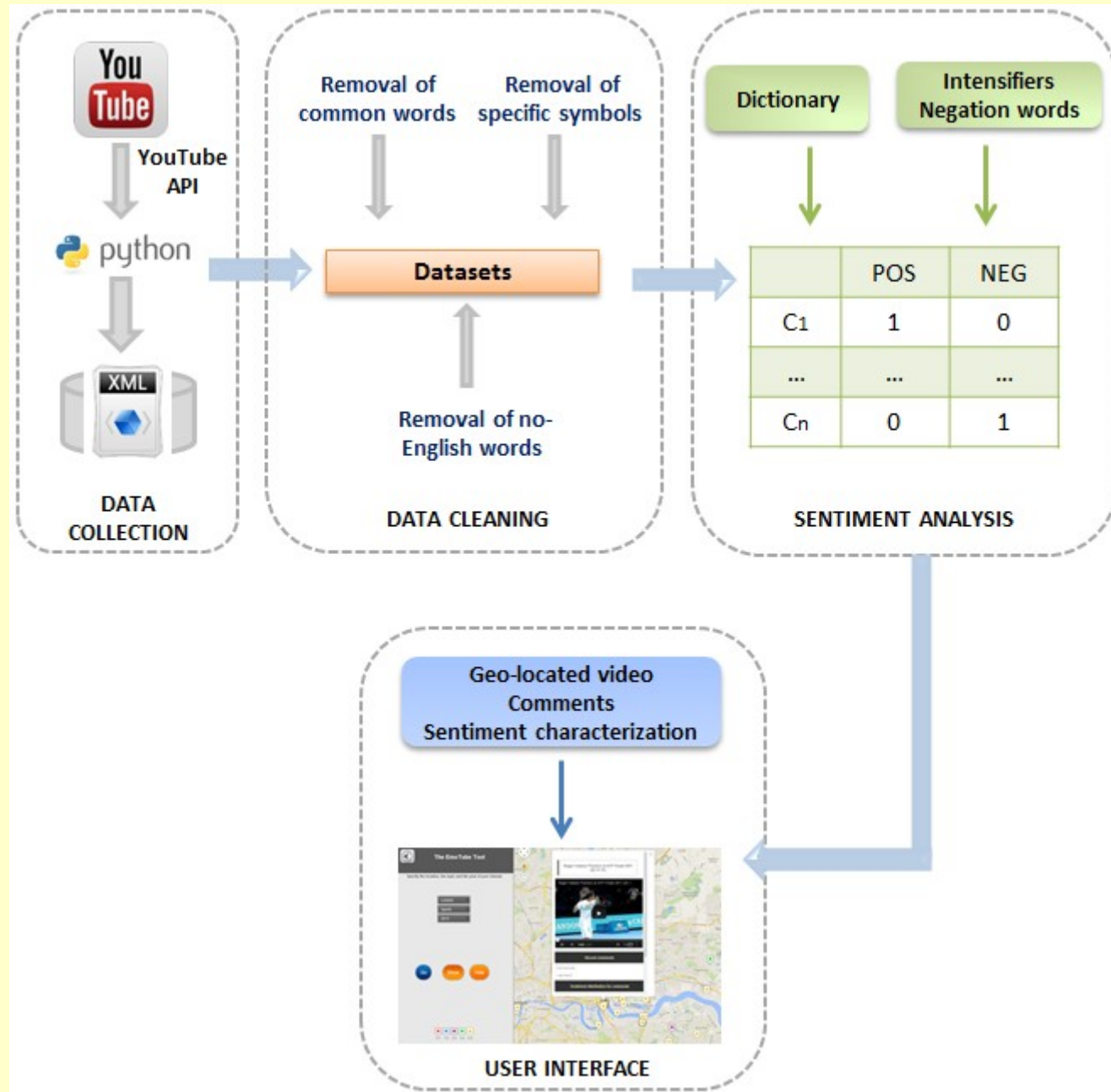


Figure 1. The proposed framework

The EmoTube Architecture

Table 2. Overview of methodologies, entities, technologies and types of data

Processing	Environment	Web Tools	Data
Collection	Back-end	YouTube API Python crawler	video id – URL – title publication date – geo-location view counts – ratings – comments
Cleaning	Back-end	Python NLTK library	common words (e.g., “a”, “is”) no-English words numbers – punctuations
Retrieval	Back-end	XML	10 datasets
Sentiment Analysis	Back-end	SentiWordNet lexicon	video comments
Integration/ Visualization	Front-end	Google Maps API JavaScript HTML – CSS	video content – geo-located comments – semantic results

Algorithm 1

- **Semantic Orientation** of each analyzed comment
 - **Negation words** (e.g., “not”, “can't”)
 - **Intensifiers** (e.g., “less”, “hardly”)
 - **Emoticons** (e.g., “:-”) - lexicon from University of Maryland, Baltimore
- The calculation of the total sentiment score of comments

Table 3. Pseudo-code of Algorithm 1

```
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)
```

Algorithm 1

- **Semantic Orientation** of each analyzed comment
 - **Negation words** (e.g., “not”, “can't”)
 - **Intensifiers** (e.g., “less”, “hardly”)
 - **Emoticons** (e.g., “:-”) - lexicon from University of Maryland, Baltimore
- The calculation of the total sentiment score of comments

Table 3. Pseudo-code of Algorithm 1

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)

Algorithm 1

- **Semantic Orientation** of each analyzed comment
 - **Negation words** (e.g., “not”, “can't”)
 - **Intensifiers** (e.g., “less”, “hardly”)
 - **Emoticons** (e.g., “:-”) - lexicon from University of Maryland, Baltimore
- The calculation of the total sentiment score of comments

Table 3. Pseudo-code of Algorithm 1

1: /*Preprocessing of data*/

2: $C^* = \text{CleanData}(C)$

➔ **3: $EC = \text{FindOpinionWords}(C^*)$**

4: /*Calculation of the emotional score for each word based on intensifiers and valence shifters*/

5: $SCI = \text{CalculateScoreIntensifiers}(EC, \text{intensifiers})$

6: $SCV = \text{CalculateScoreValenceShifters}(EC, \text{valence shifters})$

7: /*Calculation of the total score for each comment*/

8: $SC = \text{CalculateTotalScore}(C^*, SCI, SCV)$

Algorithm 1

- **Semantic Orientation** of each analyzed comment
 - **Negation words** (e.g., “not”, “can't”)
 - **Intensifiers** (e.g., “less”, “hardly”)
 - **Emoticons** (e.g., “:-”) - lexicon from University of Maryland, Baltimore
- The calculation of the total sentiment score of comments

Table 3. Pseudo-code of Algorithm 1

```
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)
```



Algorithm 1

- **Semantic Orientation** of each analyzed comment
 - **Negation words** (e.g., “not”, “can't”)
 - **Intensifiers** (e.g., “less”, “hardly”)
 - **Emoticons** (e.g., “:-”) - lexicon from University of Maryland, Baltimore
- The calculation of the total sentiment score of comments

Table 3. Pseudo-code of Algorithm 1

```
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)
```

Algorithm 1

- **Semantic Orientation** of each analyzed comment
 - **Negation words** (e.g., “not”, “can't”)
 - **Intensifiers** (e.g., “less”, “hardly”)
 - **Emoticons** (e.g., “:-”) - lexicon from University of Maryland, Baltimore
- The calculation of the total sentiment score of comments

Table 3. Pseudo-code of Algorithm 1


```
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)
```



The EmoTube Website

- A multiple-choice web environment
 - Location
 - Topic
 - Time period
- **Dynamically**-presented information
 - Video title
 - Video content
 - Most recent comments
 - More comments
 - The **polarity score** of comments on **pie charts**

The EmoTube Website




 **The EmoTube Tool**


Specify the location, the topic and the year of your interest

city...

topic...

year...

 0-1 1-2 2-3 3-4 4-5



Scenario



The EmoTube Tool

Specify the location, the topic and the year of your interest

New York

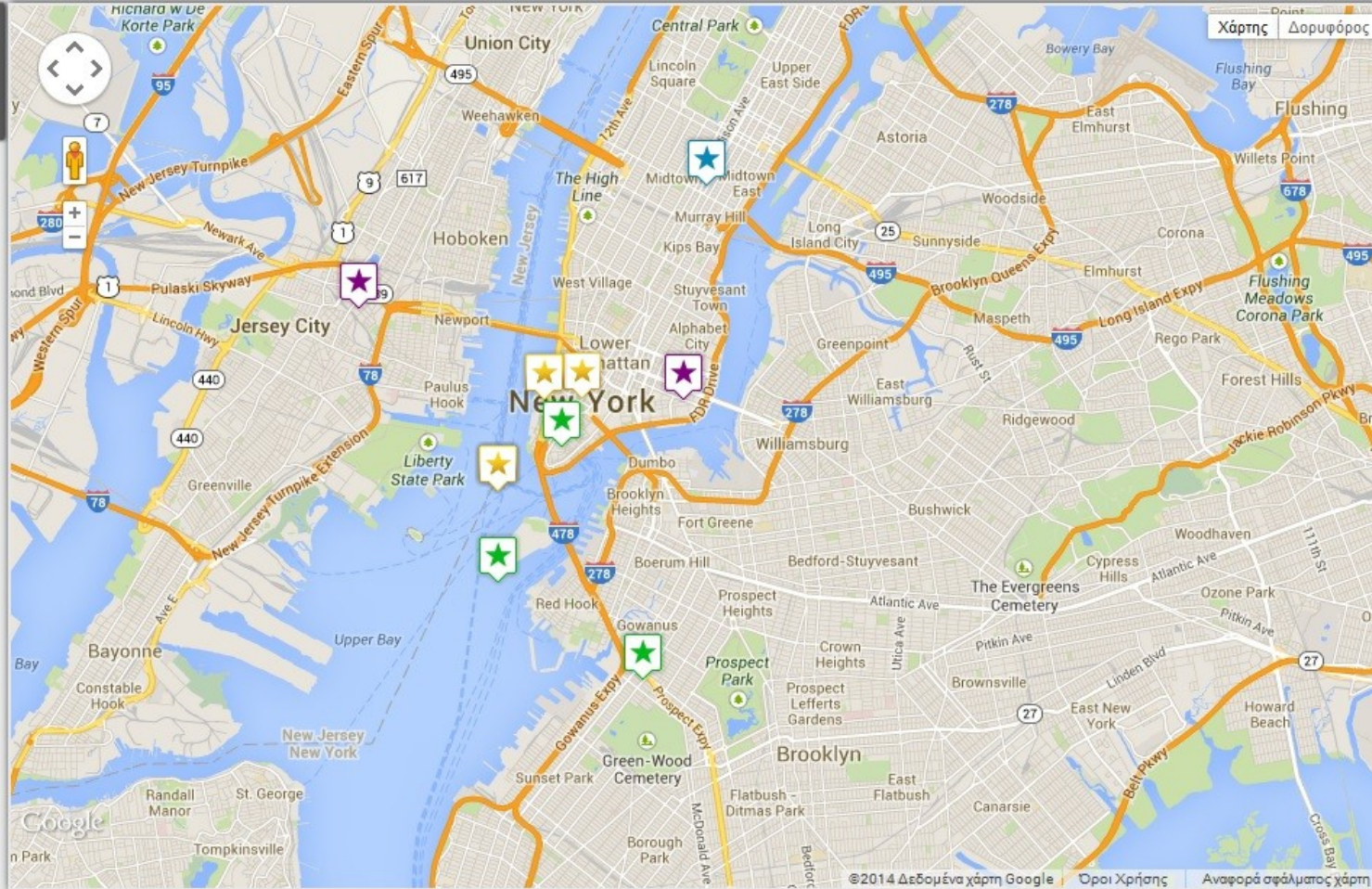
News

2010

Go

Show

Hide



Scenario



The EmoTube Tool

Specify the location, the topic and the year of your interest

New York

News

2010

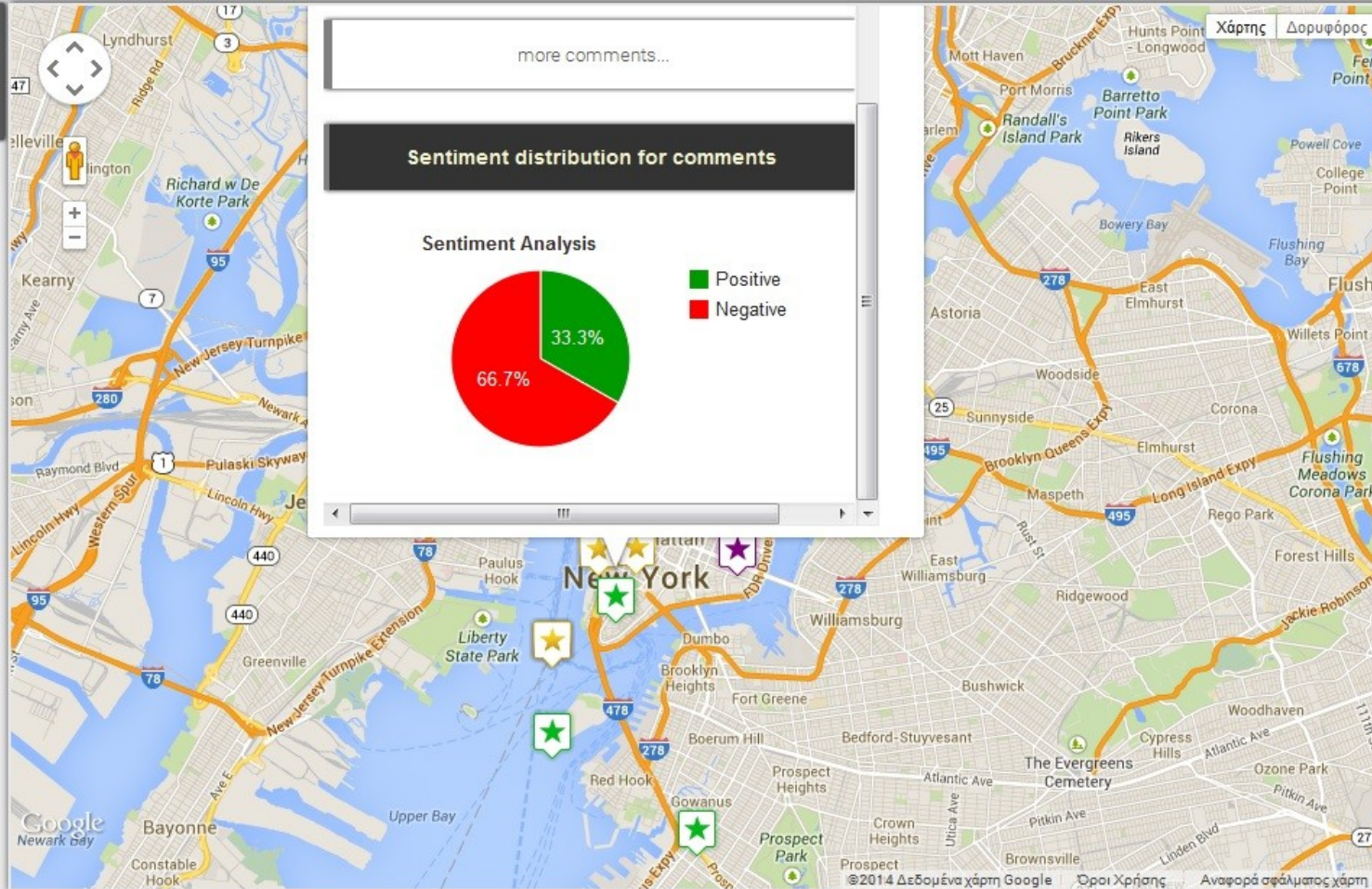
Go

Show


Hide



0-1 1-2 2-3 3-4 4-5



Scenario

 **The EmoTube Tool**

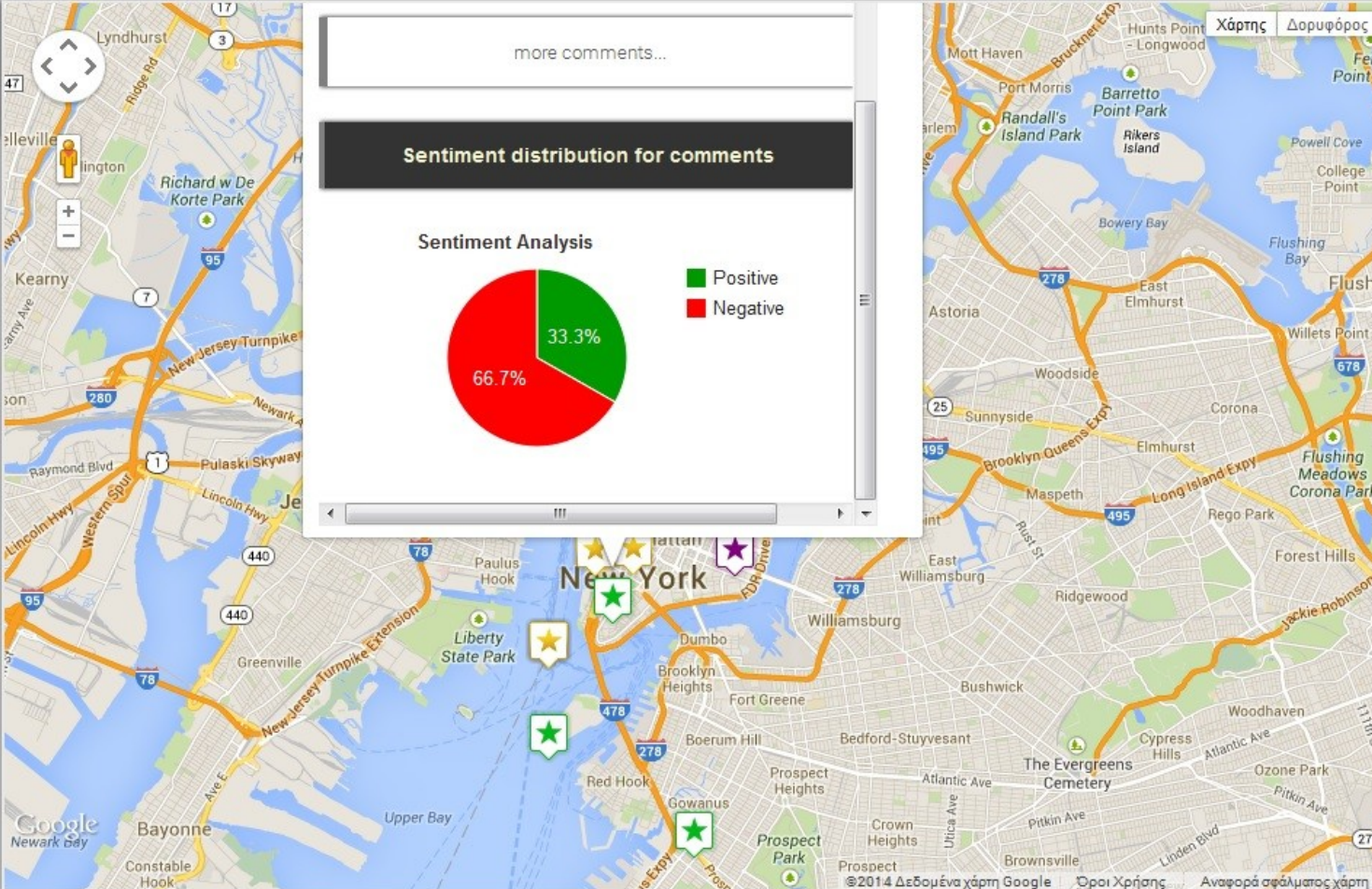
Specify the location, the topic and the year of your interest

New York

News

2010

0-1 1-2 2-3 3-4 4-5



more comments...

Sentiment distribution for comments


Sentiment Analysis

- Positive: 33.3%
- Negative: 66.7%

© 2014 Δεδομένα χάρτη Google | Όροι Χρήσης | Αναφορά σφάλματος χάρτη

“...HARD EVIDENCE”
“...should be classed as murder”
“...the plane got too scared...”

Scenario




 **The EmoTube Tool**


Specify the location, the topic and the year of your interest

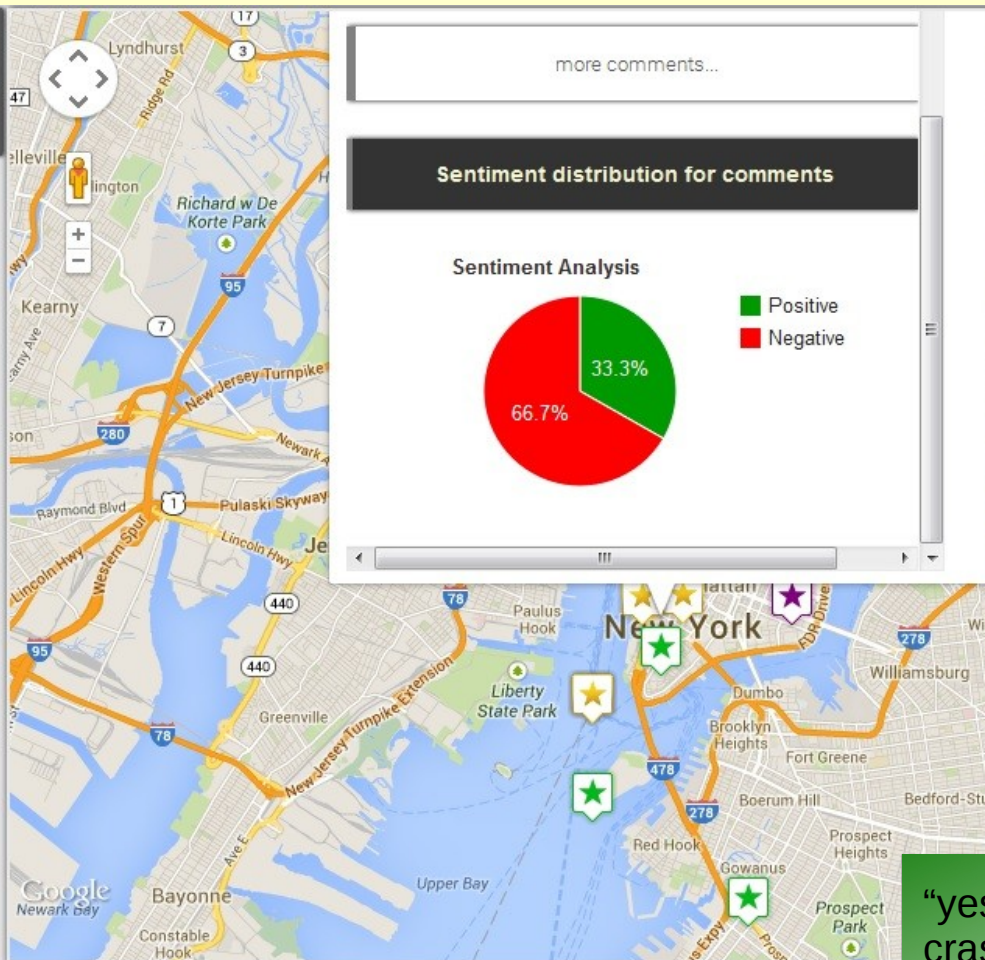
New York

News

2010


0-1 1-2 2-3 3-4 4-5

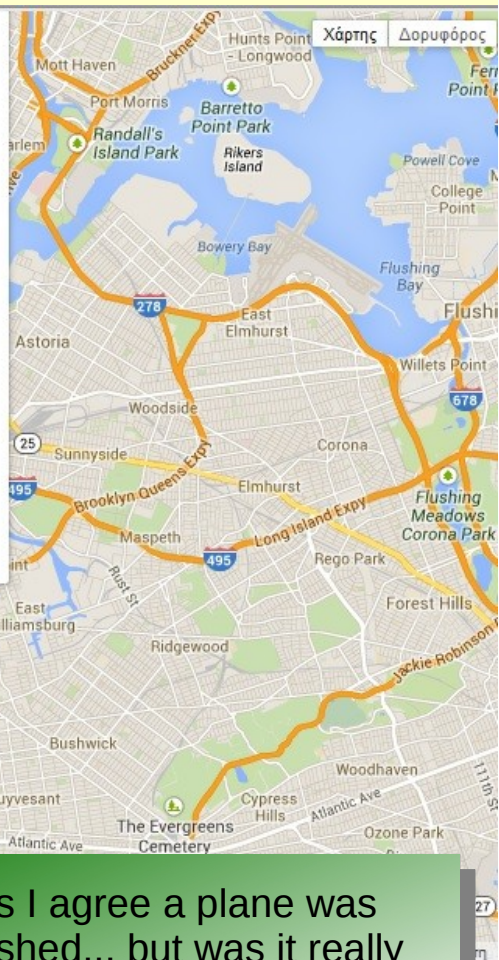


more comments...

Sentiment distribution for comments

Sentiment Analysis

- Positive: 33.3%
- Negative: 66.7%



Χάρτης Δορυφόρος

“...HARD EVIDENCE”
“...should be classed as murder”
“...the plane got too scared...”

“yes I agree a plane was crashed... but was it really a terrorist act or just a lame excuse...”

Scenario

 **The EmoTube Tool**

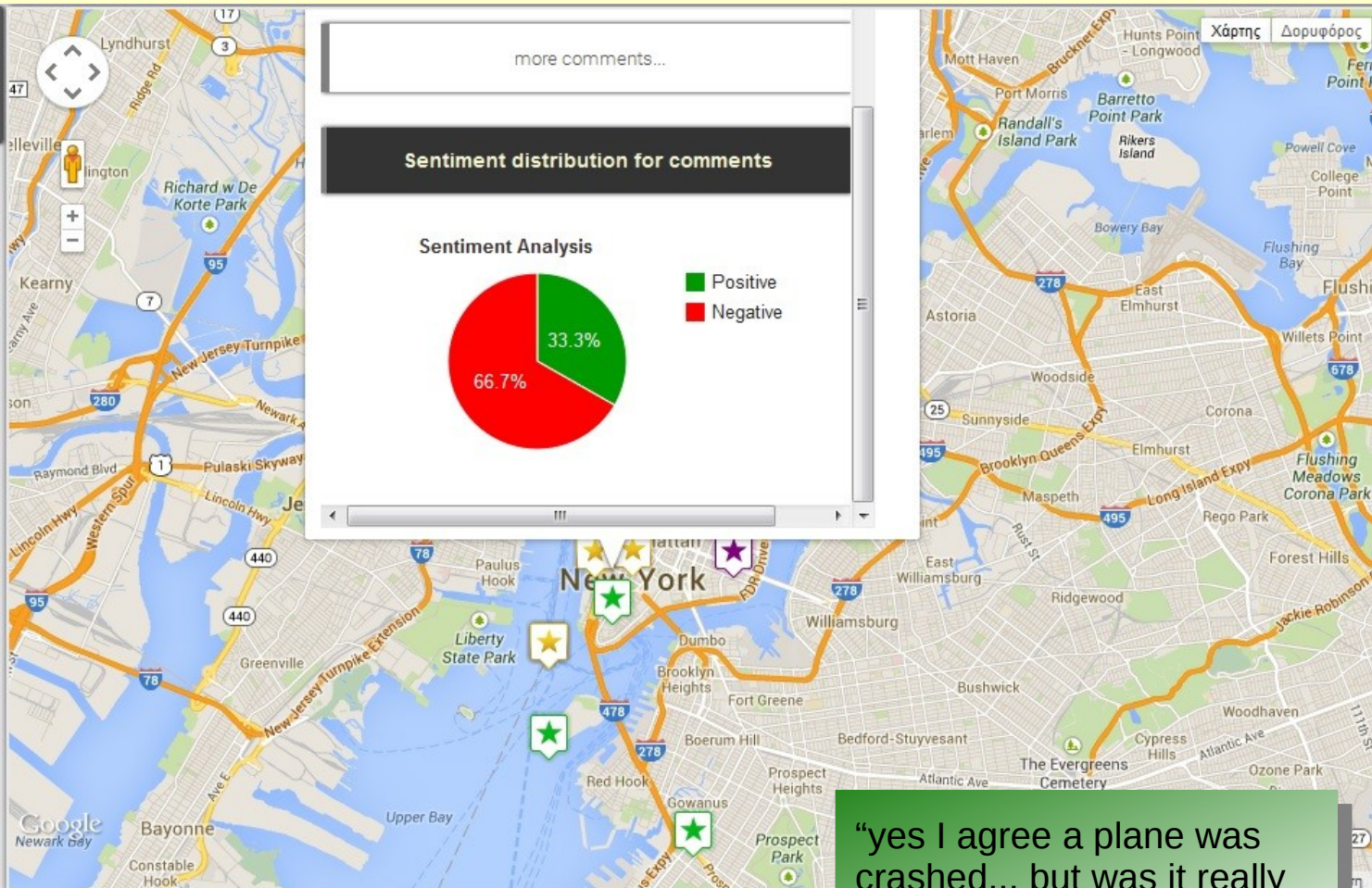
Specify the location, the topic and the year of your interest

New York

News

2010

0-1 1-2 2-3 3-4 4-5



more comments...

Sentiment distribution for comments

Sentiment Analysis

- Positive: 33.3%
- Negative: 66.7%

“yes I agree a plane was crashed... but was it really a terrorist act or just a lame excuse...”

Positive: 33.3%

Negative: 66.7%

“...HARD EVIDENCE”
“...should be classed as murder”
“...the plane got too scared...”

Conclusions

- **The EmoTube Tool**
 - Friendly interface for easy navigation
 - Opinionated information of “YouTubers” in a unified way
 - Motivation to enterprises
- **Difficulties** in sentiment detection of user comments
 - Abbreviations (e.g., “lol”)
 - Positive and negative meaning of the same phrase
- **Future extensions**
 - A more in depth analysis with the use of specific emotions (e.g., anger)
 - A more automatic mechanism that enriches the available content according to users' needs

References

- S. Baden, D. Lefebvre, J. Dar, M. Langner, M. T. Yu, "Web application hybrid".
- S. Baden, D. Lefebvre, J. Dar, M. Langner, and Yu Min-Tse, "ICT approaches to integrating institutional and non-institutional data services for better understanding of hydro- meteorological phenomena", Savona, Genoa, Italy, June 2012.
- F. Benamara, C. Cesarano and D. Reforgiato, "Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone", *ICWSM-07 Boulder, CO USA*, 2007.
- A. Benitez, S. Paek, S. Chang, A. Puri, Q. Huang, J. Smith, C. Li, L. Bergman and C. Judice, "Object-based Multimedia Content Description Schemes and Applications for MPEG-7", Volume 16, Issues 1-2, pp. 235-269, September 2000.
- 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.
- C. Cappiello, F. Daniel, M. Matera, C. Pautasso, "Information Quality in Mashups", *Internet Computing*, Volume 14, Issue 4, pp. 14-22, July-Aug 2010.
- D. Chatzakou, V. Koutsonikola, A. Vakali and K. Kafetsios, "Micro-Blogging Content Analysis via Emotionally-Driven Clustering", *ACII*, Geneva, 2013.
- L. Clarkin and J. Holmes, "Enterprise Mashups", *MSDN Architecture Center*, Journal 13.
- 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).
- A. Esuli and F. Sebastiani, "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining", In *Proceedings of 5th Conference on Language Resources and Evaluation, LREC-06*. Genova, Italy, pp 417-422, 2006.
- J. J. Hanson, "Mashups: Strategies for the Modern Enterprise", *Mashups Styles, Techniques and Technologies*, Chapter 1, pp 28-25, May 2009.

References

- 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.
- B. Liu, "Sentiment Analysis and Subjectivity", Handbook of Natural Language Processing, Second Edition, Taylor and Francis Group, Boca, Chicago, USA, 2010.
- B. Liu, "Sentiment Analysis and Opinion Mining", Chicago, USA, May 2012.
- J. Lopez, F. Bellas A. Pan and P. Montoto, "A component-based approach for engineering enterprise mashups", ICWE '9 Proceedings of the 9th International Conference on Web Engineering, pp. 30-34 Spain.
- Y. Mejova, "Sentiment Analysis: An Overview", Nov. 2009.
- J. Palfrey and U. Gasser, "Mashups Interoperability and elnnovation. *Berkman Publication Series*", Nov. 2007.
- 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.
- B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis", *Foundations and Trends in Information Retrieval*, Vol. 2, pp. 1-135, 2008.
- K. Scherer, "Trends and Developments: Research on Emotions", *Social Science Information, Sage Publications*, Vol. 44. 4, pp. 695-729, 2005.
- 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.
- 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.
- 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.

Thank you for your attention!