INTERNATIONAL
HELLENIC
UNIVERSITY

Text mining in social media for participatory sensing data

Georgios Keikoglou

SID: 3301100005

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in ICT Systems

OCTOBER 2011

THESSALONIKI – GREECE

INTERNATIONAL
HELLENIC
UNIVERSITY

Text mining in social media for participatory sensing data

Georgios Keikoglou

SID: 3301100005

Supervisor: Prof. Kostas Karatzas

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in ICT Systems

OCTOBER 2011

THESSALONIKI – GREECE

DISCLAIMER

This dissertation is submitted in part candidacy for the degree of Master of Science in ICT Systems, from the School of Science and Technology of the International Hellenic University, Thessaloniki, Greece. The views expressed in the dissertation are those of the author entirely and no endorsement of these views is implied by the said University or its staff.

This work has not been submitted either in whole or in part, for any other degree at this or any other university.

Signed:

Name: Georgios Keikoglou

Date: 21/10/2011

ABSTRACT

This dissertation was written as a part of the MSc in ICT Systems at the International Hellenic University. The main goal of the dissertation was to discover text mining tools in order to use them in social media and social networks for participatory sensing reasons. Environmental issues are around us each time of day and for that reason it is necessary to identify them using available text mining tools. It is a fact that there is a lack of available text mining software for social networks for use because companies use them for personal or research purposes only. For that reason two internet text mining tools have been used for the completion of this dissertation.

The objective of the current dissertation is to monitor social media and a specific social network for participatory sensing issues. The main idea of participatory sensing is to gather information from individuals and then try to derive/create/disseminate useful knowledge about issues that concern all of us and most of the times about the environment. In line with the adopted methodology, specific keywords have been searched followed by geographical assessment of the town of Thessaloniki in order to acquire information of that specific area. An internet web tool called *"Social Mention"* has been used for all social media and another web tool called *"Trending"* has been used for Twitter. The results of both web internet tools generated interesting information about environmental issues and problems showing that quite a few issues like waste and garbage management, concern Greek society and especially the town of Thessaloniki. With the participation of citizens and access in environmental information, a more safe and healthy environment is possible in the future.

At this point, I would like to thank my supervisor Professor Kostas Karatzas for his confidence in me and for the valuable guidance throughout the preparation of my dissertation. I would like also to thank Ms. Marina Riga which was by my side the entire time of the dissertation and help me out with her patience. Last but not least, I would like to thank my family for their love and support.

Georgios Keikoglou
21/10/2011

CONTENTS

1 Introduction

Through the past decades, technology has entered into people’s lives with the speed of light. Besides the Internet, advanced mobile phones called smart phones have begun their breakthrough to everyday society. People are able to find and share information using technology like they were able to do so since the day they were born. Information is very important, even more when it becomes knowledge. This has resulted to the combination of technology such as the Internet and the usage of smart phones, which introduced a very nowadays popular approach called *Participatory Sensing*.

With the rise of technology, social media and especially social networks establish a very important connection between information and knowledge. In this particular dissertation, text mining tools are used in social media, in order to extract useful information for participatory sensing matters.

The structure of the dissertation is summarized in 7 chapters. The content of the first chapters covers the area of participatory sensing with references to its process and its use. In the next chapters, social media and social networks categorization is mentioned as well as mining techniques in order to identify all the important information out of social media. One step before the text mining in real participatory situations, web mining methods are mentioned. As we proceed to the next chapters, text mining tools are being used in social media as well as in Twitter with specific keywords that concern environmental issues. The discussion of the results is followed, giving us the conclusions as well as future work about participatory sensing and mining in social networks.

2 Participatory sensing

Participatory sensing is a growing area of research that has its origins in research on wireless sensor networks. John Burke was the first to introduced participatory sensing with a paper at 2006 in which he described it as “a sensing architecture to enhance data credibility quality, privacy and “shareability””. [2]

The main idea of participatory sensing is to gather information from people’s activities, and share that information in order to monitor and improve situations such as health behaviors, adopt sustainable practices in resource consumption, and participate in civic processes. This is achieved from the participation of citizens in the process of sensing and documenting everyday activities and habits such as where they live, work, and spend their time. The data that are gathered can range from one single person to the combination of a group of individuals that has shared their everyday knowledge. [1]

Participatory sensing is a revolutionary new approach and its main concerns have to do most of the times with environmental issues and that comes from the voluntary decision of individuals to sense and share everything that surrounds them. This approach has enormous potential because it takes power of individuals that collect sensor data for applications that has to do with environmental monitoring, intelligent transportation, and public health, which are matters that affect people every day. [6]

Participatory sensing can draw on a variety of data collection devices, but the main technological devices that are used for this purpose are the mobile phones and especially smart phones, since their use is ubiquitous and practically universal. That is happening because, a smart phone embodies several features that can be used for participatory sensing such as its sensors. Despite that, Web 2.0 has a crucial role in the whole concept. With the combination of the Internet, people in a city or in the whole world can easily gather data about their everyday activities and then upload them to servers that can process and integrate them with other important data, such as GIS map layers and weather reports. [1]

2.1 Participatory sensing process

There are different ways to perform participatory sensing. The basic process though can be shown in the following steps: coordination, capture, transfer, storage, access, analysis and visualization (Figure 1). [1]

First of all, individuals around the world, who are interested in performing a participatory sensing project, must get in touch with other participants in order to **coordinate** the roles of every person to determine the goals and data collection plan. Such planning can be accomplished with a lot of ways and these are through social networks, via computer or mobile phones communication or even better, by face to face gatherings. [1]

One of the most important steps in the participatory sensing process is the **capture** of data. Data that are needed for a successful participatory sensing project must be gathered through mobile phones or other devices. Software of mobile phones is nowadays popular in order to capture and gather the data that you need with your mobile phone. This way participants can collect data automatically (location logging) or manually (pictures and sound or video recordings). [1]

After the data that are necessary for the participation sensing project are collected, they are **transferred** everywhere in the world and to all the participants that are involved, using a mobile phone or a wireless network. There are nowadays, mobile phone applications that can make the data transparent to the participants and tolerant of inevitable network interruptions. [1]

After the successful data transfer, a **storage** location is specified and at most of the times these are servers distributed across the Internet. Beyond private servers, commercial internet storage locations can be used such as Google, or even sharing oriented services such as Facebook. [1]

The next step of the process is to deal with crucial issues such as security and privacy on data **access**. Many people nowadays feel safer if they entrust their private e-mail and other data to website providers. A common characteristic is to share information with other trusted members of a network according to a specific and user controlled set of rules and regulations that can be found especially in social networks. In participatory sensing projects, there is much sensitive information such as images of families or friends. This information is extremely important because it can reveal a participant’s identity. That is why, any participant must be extremely careful about the information that shares and reveals. While many privacy mechanisms can already exist, this is clearly a crucial issue that requires continual attention and improvement to reduce the risks associated with abuse and misuse. [1]

The data that are gathered and stored must be **analyzed** with a variety of data processing methods. First of all, the data need to be aggregated in order to be more understandable for the participant, and then a more sophisticated analysis will determine the activity of the participant in the project. There is also the image analysis, which automatically eliminates blurry or poorly exposed images. Analysis also includes the calculation of group statistics and the integration of contributed data into statistical and spatial models that can be used to determine patterns in space and time. [1]

The final step of the participatory sensing process is the **visualization** of the data. The visualization can take many forms and that is determined by the type of the project and the nature of the participants. The effectiveness of any project depends on how well its results are understood by the target audience. Excellent methods for mapping, graphing, and animation make this a rich area to develop in the context of participatory sensing. [1]

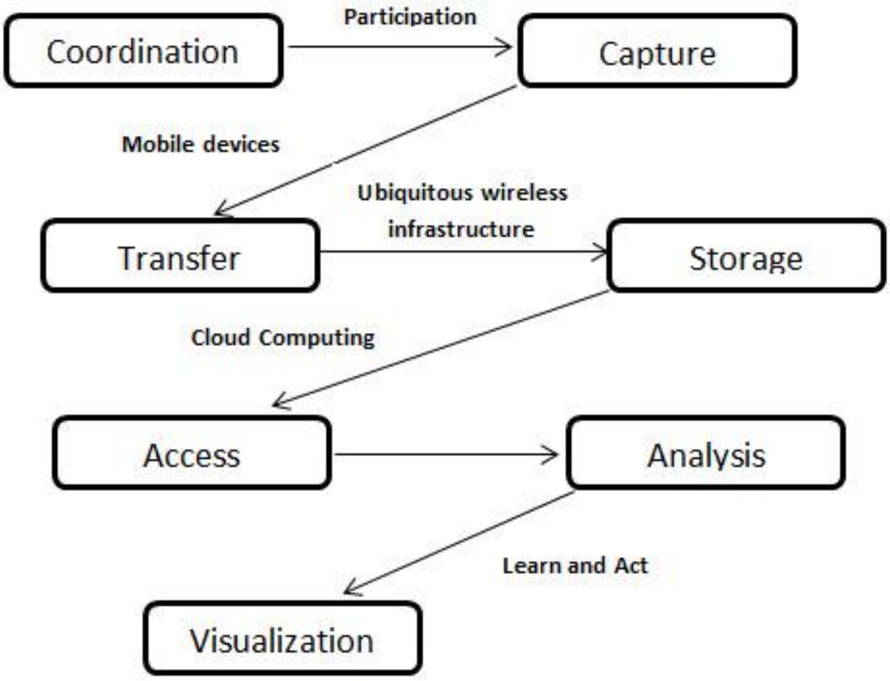


Figure 1: Participatory sensing process

2.2 Mobile phones and Sensors

We live in world where mobile phones have the first and the last word in our everyday activities. More than two billion people carry mobile phones. Participatory sensing is becoming quite popular because it embraces this technology and especially the area of smart phones. Smart phones are able to complete a variety of tasks such as gathering information using their sensors. These devices have a lot of capabilities and some are capturing, classifying and transmitting image, acoustic, location and other data. They can also play important roles because they can act as sensor nodes and location-aware data collection instruments.

A mobile phone is equipped with sensor technology which is capable of hosting a variety of applications that are highly important to the participatory sensing process. The main applications that participatory sensing is focusing are the Health Monitoring (Figure 2) and the Environmental Monitoring. Mobile phones provide an opportunity to monitor everything that has to do with the environment in order to detect and reduce pollution, to help for better medical applications and to try to reduce other problems as well. [5]

The sensors of a mobile device give the opportunity to individuals to gather, analyze and share local knowledge. The microphones and cameras that a smartphone has on board can record environmental data at any time. Besides that, cell tower localization, GPS and other technologies that are embodied to a smartphone, can provide location and time-synchronization data. Wirelesses radios and on board processing enable human interaction with both local data processing and remote servers. [2]

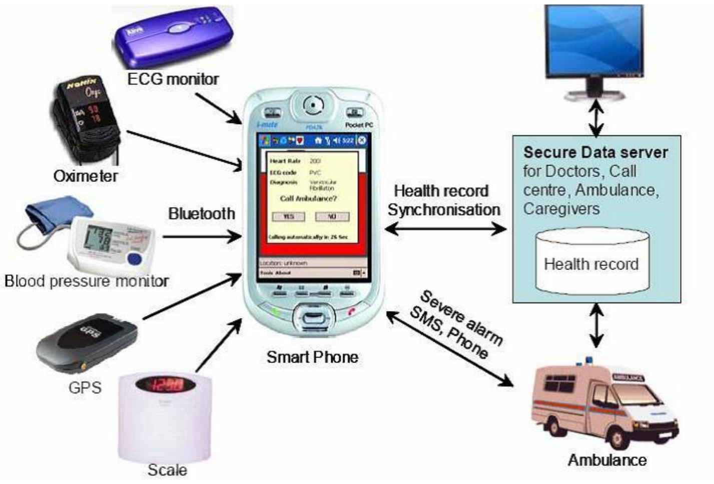


Figure 2: The use of smart phones in Health Monitoring

2.3 Participation of citizens

Participatory sensing would be a meaningless approach without the participation of citizens. They are the key component that holds the entire idea. There are three main approaches for the participation of citizens in participatory sensing and these are the following:

- *Collective Design and Investigation*. This approach includes individuals that form a group and collaborate together in order to achieve certain things like defining what, where, and why to sense something, to make data collection systems, to make an investigation to collect the desirable data, to analyze and take valuable information from the data, and make conclusions from the results. By combining local knowledge and individual empowerment with technology, this approach develops a community's potential for self-determination. Because it is a community based approach, individuals play the most crucial role in the investigative process rather than serving merely as research subjects.
- *Public Contribution*. The individuals on this approach collect data in response to a research approach that is defined by another individual or organization. Participants' main responsibilities are restricted in the collection of the data, and not in the definition of research questions or use of the results. Organizations acquire information and knowledge by finding individuals who can contribute to an effort they find meaningful.
- *Personal Use and Reflection*. Despite the fact that participatory sensing has to do mostly with a group of people, there can be individuals who can collect information about themselves and use the results for personal discovery. Any information, from images, sounds to video records, can be analyzed and visualized, and reveal valuable information about health, safety, and other important issues about a person's life. In this sense, participatory sensing can be characterized as a very valuable process because it allows someone to evaluate his life and change anything that was previously overlooked. A person may wish to keep these newly discovered patterns private or, like reflections written in a blog, share them with family, friends, and the public. [1]

2.4 Participatory Sensing Projects

Even though Participatory sensing is a new field, there are many applications and online services that are around. Some of them are not yet completed and they are in an immature level, while others are becoming widely successful.

One of the most well-known projects for monitoring urban environments basically for noise pollution is the NoiseTube project. NoiseTube enables citizens to measure their personal exposure to noise with their mobile phones which are equipped with GPS and use them as noise sensors. The system allows participants to share data though a website (http://www.noisetube.net) in order to facilitate collective monitoring initiatives. NoiseTube is a mobile application which gathers and visualizes measured data from the server side, and from the client side, the users measure sound level values using the microphone of their mobile phones, which are presented on the screen as a color-coded number and as a variable in time. Each of these values are tagged with time and location stamps and are uploaded to the website of the project, where the results are processed and a Google Earth map representation is produced. [4]

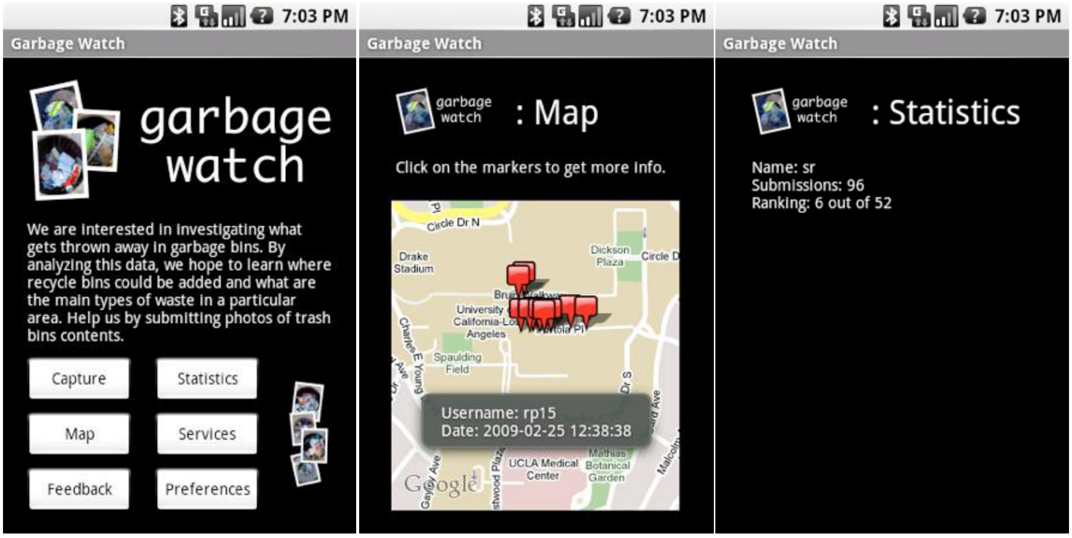
Other projects can be found in a website called Urban Sensing. A variety of participatory sensing projects that run successfully are gathered together for the convenience of the crowd. Few of them are the following: [3]

- **Cycle sense**: An application that runs on mobile phones and helps bikers find good routes and collect data to improve them. There are many features that can be complicated for a biker in Los Angeles like road and path availability, air quality, traffic and accidents, bright sunlight, all of which that affect the quality of the ride. UCLA's Center for Embedded Networked Sensing (CENS) is collaborating with Los Angeles bikers to create an application for the convenience of the bikers in Los Angeles. This application enables bike commuters to log their bike route using GPS and provide geo-tagged annotations along with automatic sensor data to infer the roughness and traffic density of the road. Currently a pilot called Biketastic (http://biketastic.com/) is running, in which bikers can share their routes which are automatically annotated by noise level, roughness, variation in elevation and duration of stops.

- **Diet sense**: An online service that allows people to self-monitor their food choices and further request comments from dietary specialists. Mobile phones with CENS participatory sensing platform record every day meals, either automatically or by sensible notifications.
- **Family dynamics**: An application that enables families to explore their own dynamics with mapping and coaching tools. Tools that embedded on a phone can collect data otherwise invisible to wellness professionals who most commonly rely on family member self-reporting. The first coaching tool is a prototype called Andwellness. It is a personal health self-management application for the Android phones that supports flexible geo-spatial, social and activity triggered reminders and ecological momentary assessment.
- **Footstep**: The footstep project measures walking activity by leveraging cell phone, GIS, and sensor technologies. The philosophy is that an accurate and individual feedback is essential in addressing and improving awareness of exercise patterns. The first-generation system targets to provide an easy to understand heat-map of walking traces, a comprehensive data histogram, and a trend-analyzer. The website of the particular project is http://footstep.cens.ucla.edu.
- **PEIR**: The Personal Environmental Impact Report (PEIR) is a new kind of online tool that allows the use of a mobile phone to explore and share the impact of the environment to an individual and measure the impact of him in the environment. Information is gathered and then analyzed with published scientific models that produce estimates of the exposure and impact in four categories:
    - Smog Exposure
    - Fast food exposure
    - Carbon impact
    - Sensitive Site Impact
- **Remapping LA**: A project that aims at facilitating fluid and inclusive expressions of Los Angeles as communities explore their environments, culture and identities and retell their histories with technology built in a process they shape. The process uses mobile devices to help communities in discovering, mapping and documenting the city and adding to this “collective memory.” Mapping of the histories and cultural identities by communities is a way of community asset-mapping.

- **Garbage watch**: A project that asks members of the UCLA community to perform a coordinated waste audit using their mobile phones. Individuals use their phones to collect and upload geo-tagged images of the contents of garbage bins to help UCLA Facilities determine where new recycle bins should be placed, the effectiveness of existing recycling infrastructure, and to learn more about when, where, and what materials get thrown away on campus.

Flowingly, screen shots taken from two different participatory applications-projects are given (Figures 3-4).



We are interested in investigating what gets thrown away in garbage bins. By analyzing this data, we hope to learn where recycle bins could be added and what are the main types of waste in a particular area. Help us by submitting photos of trash bins contents.

Click on the markers to get more info.

Username: rp15
Date: 2009-02-25 12:38:38

Name: sr
Submissions: 96
Ranking: 6 out of 52

Figure 3: Garbage Watch

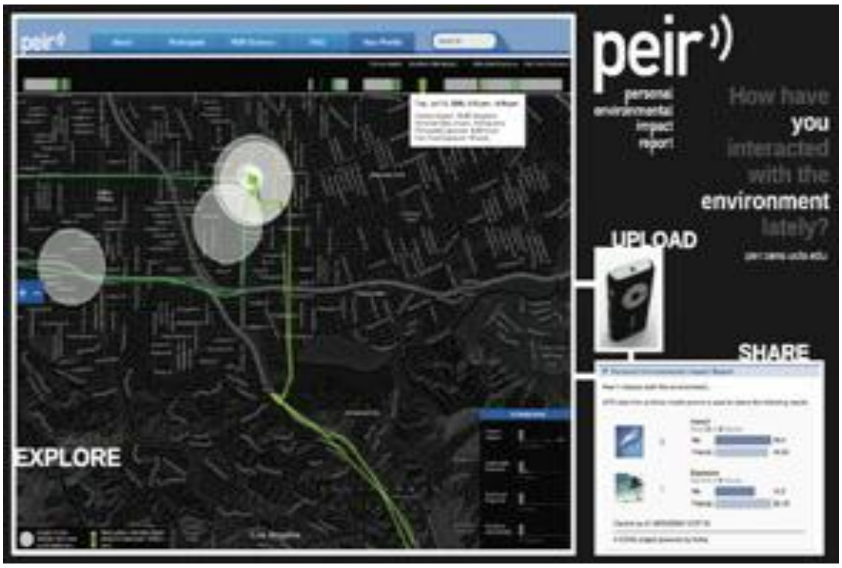

peir
personal environmental impact report
How have you interacted with the environment lately?
UPLOAD
SHARE
EXPLORE

Figure 4: PEIR

3 Social Media

Last year was a big year for Social Media in many ways. Social networking services have grown rapidly over the past 12 months, and also social media marketing has been introduced. Many people thought that social media was a fad and would simply disappear, however after seeing many large enterprise organizations such as Dell, Starbucks, and Ford implement very successful social media campaigns, many smaller and bigger companies are now using social media to create an online presence and engage with new and existing customers on various social platforms.

It is true that Web 2.0 has become a successful tool nowadays for methods such as participatory information sharing, interoperability and collaboration on the World Wide Web. A site that is based on Web 2.0 gives privileges to users to interact and collaborate with many others in a social media dialogue as creators of user-generated content in a virtual community. Examples of Web 2.0 include social networking sites, blogs, wikis, video sharing sites, web applications, mashups and folksonomies.

3.1 Social Media Description

Internet has entered our lives and since that day everything has been different. With Web 2.0 and the rise of social media, internet became more and more popular. Two-thirds of the world’s Internet population visit social networking or blogging sites, accounting for almost 10% of all internet time.

Social media is a type of online media that ease conversations as opposed to traditional media, which doesn't allow everyone to participate in the creation or development of the content. Social media has a unique characteristic to encourage contributions and feedback from everyone that uses them. Media and audience are becoming one. Most social media services are open to feedback and participation. Actions like voting, leave comments and share information are available to everyone. There are rarely any barriers to accessing and making use of content. [9]

In order to understand the importance and the effect that social media have in today's world, below there are some interesting statistics of 2011 about social media usage in the United States.

- 38 million people in the US, from age 13 to 80, said that they are influenced by social media in order to make decisions, a 14% increase in the past six months [8]
- 1 million people view customer service related tweets every week, with 80% of them being critical or negative in nature [10]
- 132.5 million people in the US use Facebook in the year of 2011. It is estimated that by the year of 2013 the number will increase to 152.1 million [11]
- 59% of Internet users use at least one social networking service, compared to 34% who did in 2008 [12]
- 81 minutes was the average daily use of mobile apps in June 2011, compared to 74 minutes for the Web [13]
- 750 million Facebook active users per month, up from 500 million active monthly users last year [14]

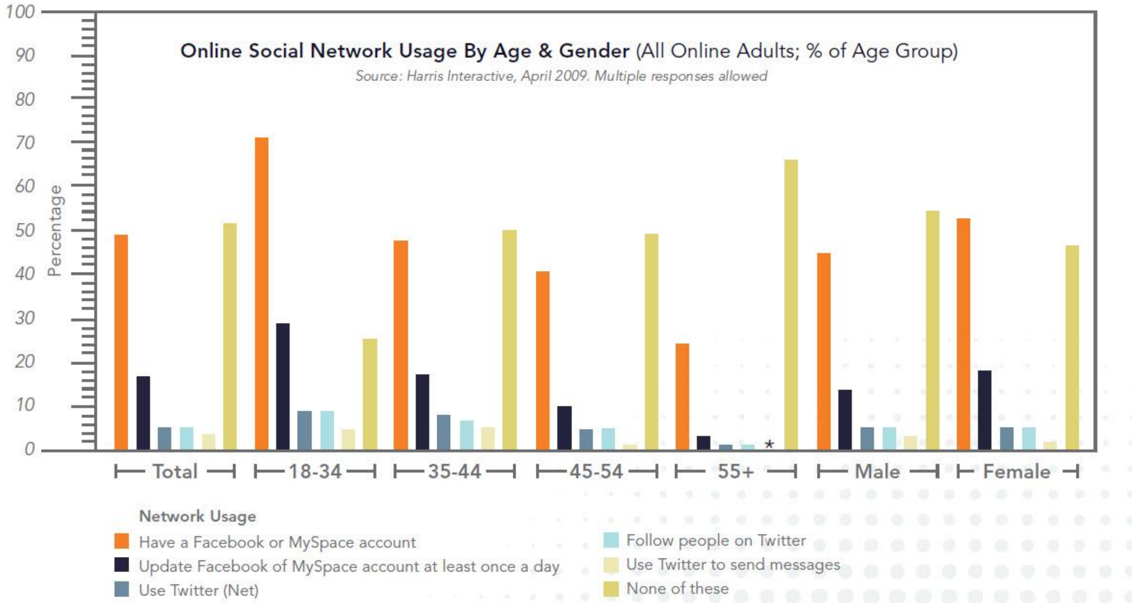


Figure 5: Online Social Network Usage by Age & Gender

3.2 Categorization of social media

Social media are becoming very popular nowadays. But when someone is referring to social media usually is referring to the various kinds of them. At this time, there are basically six kinds of social media. Each of them can be characterized as tools which play a vital role in the success of an event. Tools such as online social communities, blogs, videos and podcasts can further enhance the show's messaging and brand while also delivering promotional information on services. The social media kinds are listed below:

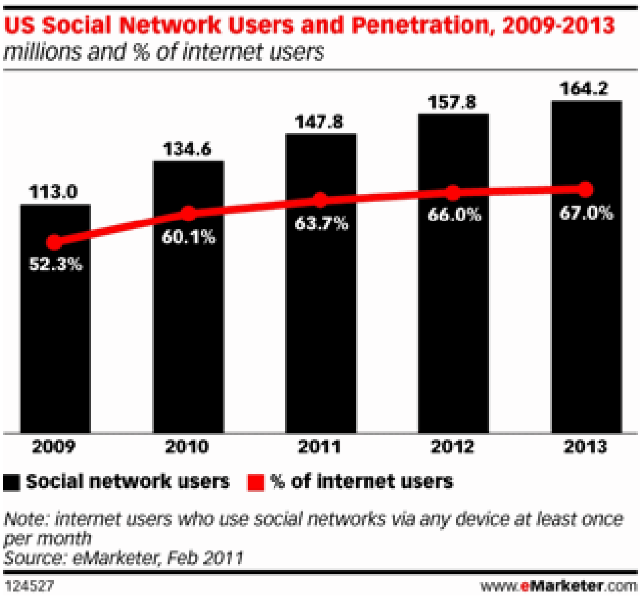
3.2.1 Online Social Communities

Online Social Communities or Social Networks which they are known are web sites which allow people to form personal web pages and then connect with other individuals to share content and to communicate. They can connect to people they already know or establish connections with others because the social networking sites help them have the ability to match the interests of individuals and bring them closer. This is happening not only within a limited network between individuals but everywhere in the world. Inside these social network sites, people collaborate, create new content and establish friendships. The most known global social networking sites are Facebook and Twitter. [15] [16]

There are various benefits from a social network. Social networks give the opportunity to individuals or groups to express themselves and to be in contact with people with similar issues or challenges. They also help individuals to organize better events and their plans. [16]

Social networking penetrates nowadays into people's life more easy than ever. Especially in the United States the growth of social networks are extremely fast. There are estimations that nearly 150 million US web users will use social networks either by computer or a mobile device at least monthly this year, bringing the reach of such sites to 63.7% of the online population. By 2013, 164.2 million Americans will use social networks, or 67% of internet users as shown in the following graphical representation (Figure 6). [26]

Besides communication between individuals, social networks can be beneficial for other reasons as well. Online social networks offer free means of promoting meeting planner's events pre, during and post show. Facebook gives that opportunity for discussion for attendees to share best practices and relevant content. Also, specific Twitter event pages are extremely helpful and provide an excellent channel for real-time announcements and event promotions that engage attendees and exhibitors. [15]



Note: internet users who use social networks via any device at least once per month
Source: eMarketer, Feb 2011

Figure 6: US Social Network Users and Penetration, 2009-2013

3.2.2 Blogs

Blogs are considered to be one of the most well-known forms of social media. The original term is "Web log" that means an online, chronological collection of personal commentary and links. The author of a blog is usually an individual or a group. It is extremely easy for someone to create a blog and to publish anything and for that reason blogs have become a successful established communication tool. [17]

Blogs have a unique characteristic of helping people in knowledge sharing, and this is the main reason why they attract a large and dedicated group of audience. Especially in women, blogs have a unique effect on them. They have reported that they influenced a lot by blogs in order to purchase something. Most of the times, blogs have positive influence on women in purchasing things. The following graphical representation (Figure 7) shows more about the decisions of women based on blogs. We can see that Blogs are becoming an important component of the Internet landscape, because they provide to both the authors of the blog and the readers space to express themselves and connect with each other, without the need of chat rooms or formal media outlets. [17]

As mentioned before, the creation and maintenance of a blog is really easy and that makes blogs an ideal place for discussions among the Internet community on new or timely topics. Typically, the entries of a blog are dated with the most recent on top, and there is usually an opportunity for readers to make comments. It's a fact that there are more than 110 million blogs. Blogs can be useful for promoting meetings as well and for setting up blog channels around events in order to create interest, to get feedback and to build communities. [15] [17]

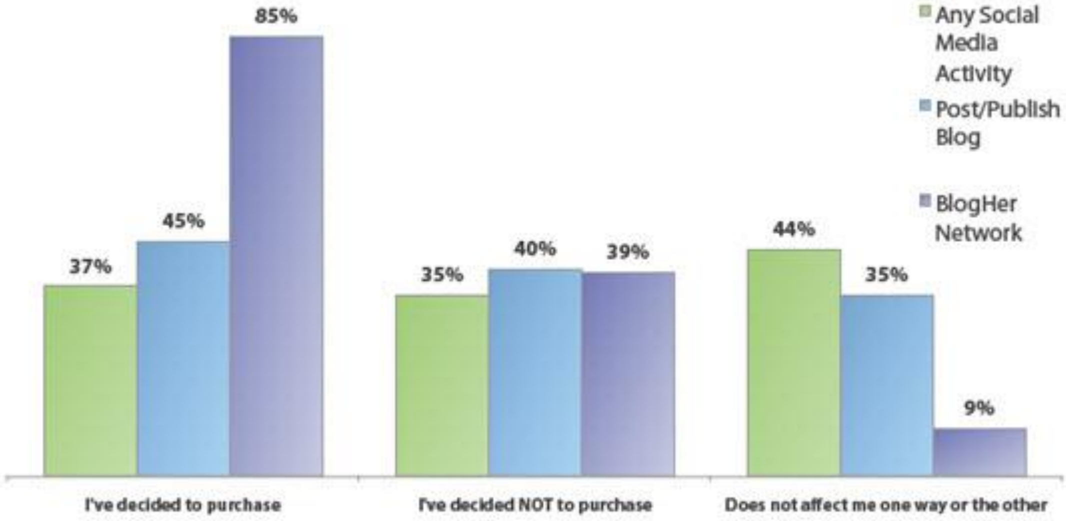


Figure 7: How blogs influence women's decisions

3.2.3 Podcasts

Podcasting is a term inspired by the Apple Computer Corporation and at first it was related to the iPod which is a portable digital audio player with which the users are capable to download their music from their computer directly to the device in order to listen to it. The term is no longer related to the iPod but refers also to any software and hardware combination that allows automatic downloading of audio file. Podcasts make use of the Internet's Real Simple Syndication (RSS) standard. It differs from broadcasting and webcasting in the way that content is published and transmitted via the Web. Instead of a central audio stream, podcasting sends audio content directly to an iPod or other MP3 player. [18]

Podcasts are also great promotional tools for events. Meetings are content generators. For example, interviewing upcoming speakers about their subject matter to create podcasts is one way to promote meeting attendance. If done properly, podcasts will bring traffic to the website, generate business and attract attendance to meetings. [15]

Podcasting is evolving at a rapid rate. New features come to light such as categorizing, navigating, and indexing. Consequently, designers and producers of podcasts are seeking new ways to add layers of richness to simple audio files, creating audio experiences that are both entertaining and instructive. [18]

3.2.4 Forums

An Internet forum is an online site where individuals can make conversations by posting messages. Every message posted, has to be approved first by the moderator of the site before it appears. The structure of a forum is hierarchical or tree-like. In a forum there are different topics and each the new discussion that takes place is called a thread, and can be replied to by as many people as wish to.

In some forums there are restrictions such as users that want to post a message or even to read existed ones, have to register first. On most forums, users do not have to log in to read existing messages. Even though forums introduced before the term “social media”, nowadays are a powerful and popular element of online communities. [19]

3.2.5 Wikis

A Wiki is a collaborative Web page or even better a collection of web pages that have been designed to give the ability to anyone to create a web page which will allow visitors to search its content and edit it in real time, as well as view updates since their last visit. The most well-known function of a Wiki is the collaboration of users within them and the manageability they offer. Additional features include calendar sharing, live AV conferencing, RSS feeds, and more. [20]

Wikis have been taken a lot of names over the years. They have been described as a composition system, as a discussion medium, as a repository, as a mail system, and most of all as a tool for collaboration. Besides that, wikis are able to incorporate sounds, movies, and pictures; they may prove to be a simple tool to create multimedia presentations and simple digital stories. [21]

3.2.6 Content communities

Content communities are a form of social media and from the term only, someone can imagine that they look a little bit like social networks, but they main focus of them is on a particular type of content. A social media can be identified by five main characteristics and these are Connecting, Community, Conversation, Openness and Participation. In these online communities we can find a lot of social networking characteristics such as conversations, comments, invitations, and groups. [22]

The content communities are web sites that organize and share particular kinds of content. One of the best-known video distribution sites is YouTube. Mobile data usage grew by 77% for the first half of 2011 and YouTube accounts for 22% of that (Figure 8). Additionally, Flickr is another image and video hosting website to share photographs and is also used widely by bloggers as a photo repository. Except from that, there is also Delicious (http://www.delicious.com/) for bookmarked links. [9]

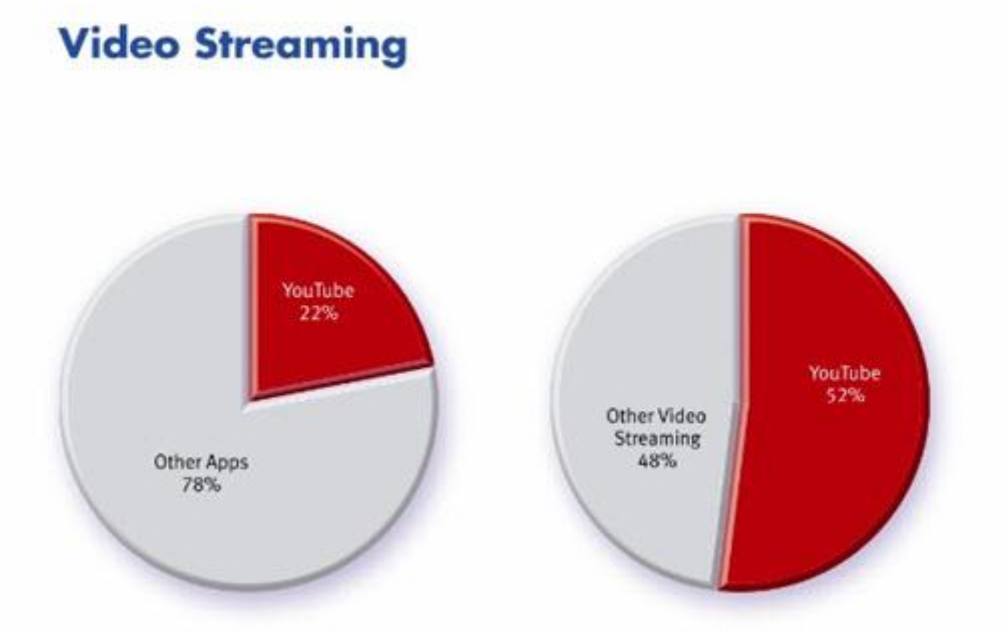


Figure 8: YouTube accounts keep rising

4 Information Analysis

As times passes, people and technology changes constantly, and new kind of information must be gathered in order to cope with these changes. There are enormous amount of data that are stored in files, databases, and other repositories around the world and it is increasingly important to develop powerful means for analysis and sometimes interpretation of such data in order to extract important information and knowledge that can be used in decision-making. In that case, there are mining techniques that can be the key for analyzing important information. [23]

4.1 Data Mining

Data mining, also known as Knowledge Discovery in Databases (KDD), carries a lot of definitions about its substance. Professor Osmar R. Zaïane from the University of Alberta refers to Data Mining as *"the extraction of implicit, not yet known and potentially useful information from data in databases".* Jeffrey W. Seifert on the other hand, who is Analyst in Information Science and Technology Policy Resources, adds that *"Data Mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets"*. Consequently, data mining consists of more than collecting and managing data, it also includes analysis and prediction. [23] [24]

There are opinions that data mining and knowledge discovery in databases is the same thing, but data mining is actually part of the knowledge discovery process. The following figure shows data mining as a step in an iterative knowledge discovery process. [23]

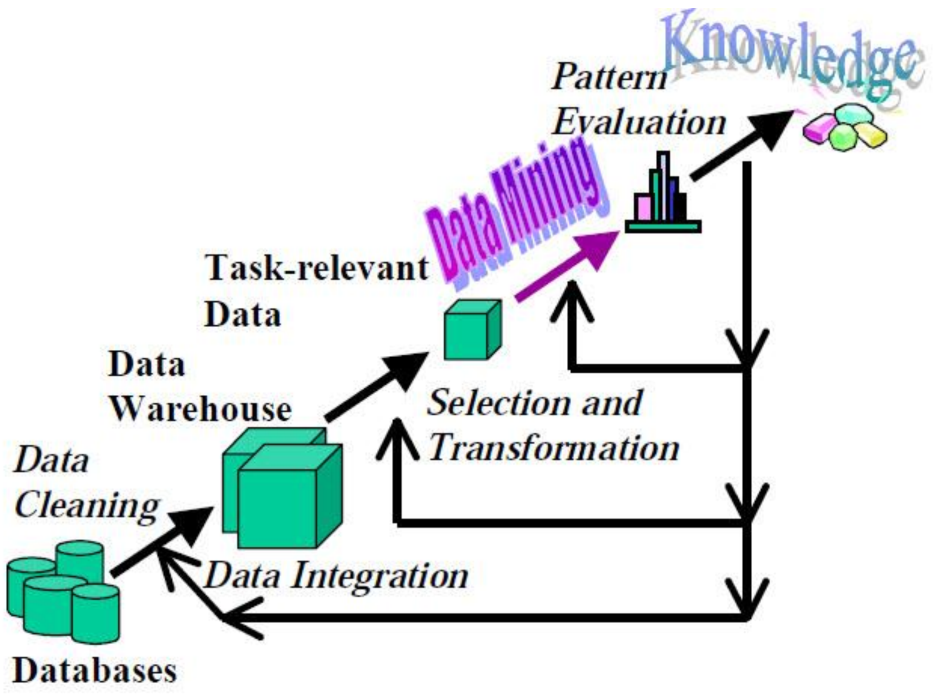


Figure 9: Data Mining: The core of Knowledge Discovery process

4.1.1 Data Mining Process

The Data Mining process comprises of a few steps leading from initial raw data collections to some form of new knowledge. The iterative process consists of the following steps:

- *Data cleaning*: it is the first phase of the process and in which noise data and data that are not needed are removed from the collection.
- *Data integration*: at this stage, multiple data sources, many times heterogeneous, may be combined in a common source.
- *Data selection*: one of the most important phases of the process. In this stage the relevant data are decided and then retrieved from the data collection.
- *Data transformation*: it is the phase in which the data that were selected in the previous phase are transformed into forms appropriate for the mining procedure.
- *Data mining*: it is the most crucial phase in which techniques are applied to extract patterns potentially useful.
- *Pattern evaluation*: in this phase, strictly interesting patterns representing knowledge are identified based on given measures.

- *Knowledge deployment*: it is the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand and interpret the data mining results. [23]

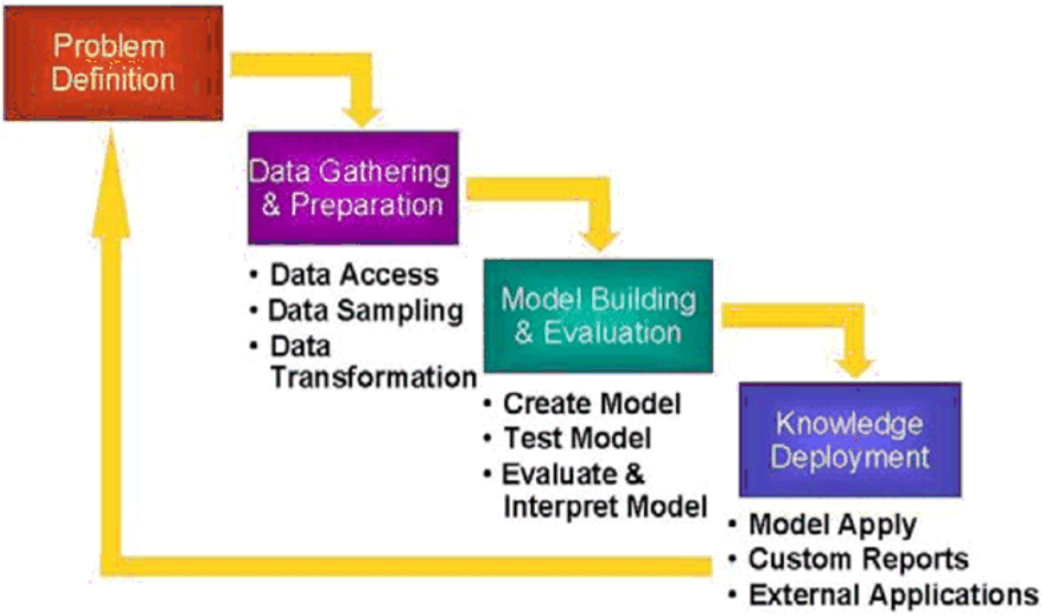


Figure 10: Data Mining Process

It is true that many times some of the steps of the Data Mining process are combined together. For instance, data cleaning and data integration can be performed together as a pre-processing phase to generate a data warehouse. Data selection and data transformation can also be combined where the consolidation of the data is the result of the selection, or, as for the case of data warehouses, the selection is done on transformed data. [23]

Data Mining is an iterative process. Once the discovered knowledge is presented to the user, the evaluation measures can be enhanced, the mining can be further refined, new data can be selected or further transformed, or new data sources can be integrated, in order to get different, more appropriate results. [23]

4.1.2 Mined Data

Data mining has to do with all kinds of data and not with one type specifically. Data mining should be applicable to any kind of information repository. But different types of data may require different algorithms and approaches. In the next paragraphs, different kinds of data that Data Mining can be processed and acquire vital information and knowledge are presented. [23]

- **Flat files:** These kinds of files are the most common data source when it comes to data mining algorithms, especially at the research level. They are simple data and as an example of flat files can be transactions, time-series data, scientific measurements, etc. [23]
- **Relational Databases**: This kind of databases consists of a set of tables that contains either values of entity attributes, or values of attributes from entity relationships. Relationship databases can be more useful for data mining than flat files because of the databases structure. While data mining can benefit from SQL for data selection, transformation and consolidation, it goes beyond what SQL could provide, such as predicting, comparing, detecting deviations, etc. [23]
- **Data Warehouses:** A data warehouse can be characterized as a repository of data that has been collected from multiple data sources. A data warehouse gives the option to analyze data from different sources under the same roof. [23]
- **Transaction Databases:** A transaction database is a set of records which represents transactions, each with a time stamp, an identifier and a set of items. Because relational databases do not allow nested tables, transactions are usually stored in flat files or stored in two normalized transaction tables, one for the transactions and one for the transaction items. [23]
- **Multimedia Databases:** As a multimedia database can be characterized a database which includes video, image, audio, and text media. These data can be stored on extended object-relational or object-oriented databases, or simply on a file system. Data mining from multimedia repositories may require computer vision, computer graphics, image interpretation, and natural language processing methodologies. [23]
- **Spatial Databases:** A spatial database is a database that is a little different in content than the other databases. That kind of database store geographical information like

maps, and global or regional positioning. Such spatial databases present new challenges to data mining algorithms. [23]

- **Time-Series Databases:** A time-series database contains data that changes over time, such as stock market data or logged activities. The main challenge when it comes to data mining in these databases is the case of continuous flow of new data, and the fact that everything is happening in real time. Studying trends and correlations between evolutions of different variables, as well as the prediction of trends and movements of the variables in time are the main steps in this case. [23]
- **World Wide Web:** The internet represents the largest and most dynamic repository available. It continuously changes and this is happening because people from all over the world are contributing to the growth of its content every day. Data is organized in inter-connected documents and these can be either text, audio, video, raw data, and even applications. There are three main components of the World Wide Web. The content of the Web, the structure of the Web and the usage of the web. Data mining in the World Wide Web, or web mining, tries to address all these issues and is often divided into web content mining, web structure mining and web usage mining. [23]

4.2 Text Mining

Text mining, also known as Knowledge Discovery from Text (KDT), deals with retrieving information from text. To achieve this, the method uses techniques different from information retrieval, information extraction as well as natural language processing, and connects them with algorithms and methods of data mining, machine learning and statistics. The procedure that is followed is similar to the Data Mining process, except that in this case Text Mining doesn't deal with all types of data but with text, and tries to analyze it. [25]

4.2.1 Text Mining Process

The purpose of text mining is to identify the useful information from documents and texts. Text mining can be characterized as an empirical tool that has a capacity of identifying new information that is not apparent from a document collection.

The text mining process (Figure 11) uses Information retrieval and Natural Language Processing to mine various datasets and infer the knowledge available in the dataset. Other actions of the text mining process are searching, extracting, and categorization only where the themes are readable and the meaning is obvious. [25]

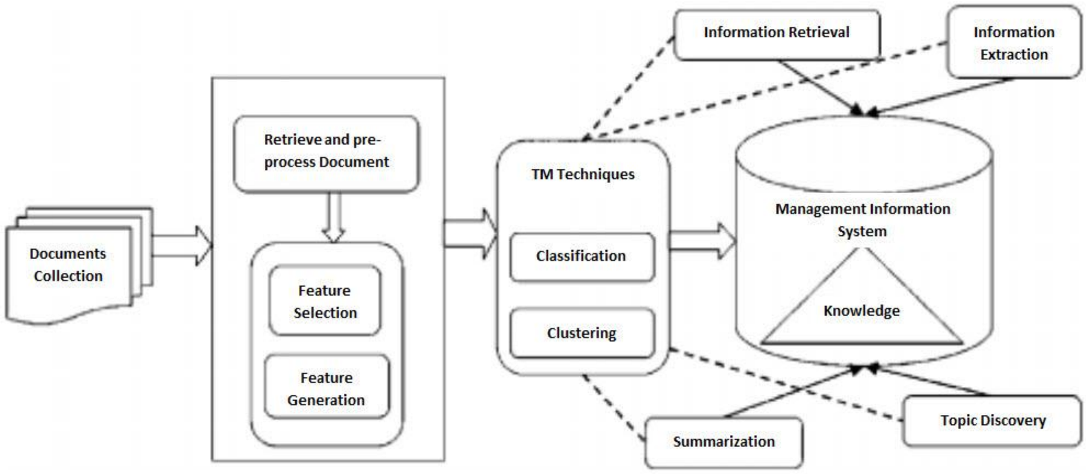


Figure 11: Text Mining Process

The first phase of the process is the document collection; that is set of files with any extension like PDF, txt or even flat file extension. Each of them are usually collected from online chats, SMS, emails, message boards, newsgroups, blogs, wikis and web pages. Also, text data set could be created by processing spontaneous speech, printed text and handwritten text, but this data may contain processing noise. [25]

The document collection is an unstructured dataset of documents which are preprocessed using the following three rules:

1. Tokenize the file into individual tokens using space as the delimiter.
2. Removing the stop word which does not convey any meaning.
3. Use porter stemmer algorithm to stem the words with common root word. [25]

Next in line is the selection of features that are appropriate for the text mining process. After the appropriate selection of them, the text mining techniques are incorporated for the applications like Information retrieval, Information Extraction, Summarization and Topic Discovery for necessary knowledge discovery process. The process is using also Data Mining or else, Knowledge Discovery in Database, which is a fundamental step in Text Mining. [25]

Finally, the knowledge that has been acquired is stored in the management information system and it can be and retrieved when the system needs to access this particular knowledge. [25]

5 Mining in Social Media

Social media represent a huge amount of information and provide challenges for researchers and analysts to identify the useful knowledge behind the data. The enormous amount of information has been the major problem thought the past few years because it is extremely difficult to extract the right information when you have to deal with a lot of TB of data. The above characteristics along with the combination of the heterogeneity of the data and the possibility that some information may be incorrect, makes even harder the search for good web results. A very trustworthy method which gives the user reliable information in short period of time is called Social Mining or Web Mining. [28]

Web mining main responsibility is to extract specific knowledge from the World Wide Web and more specifically from Social Networks. It provides the algorithms and methodology required for the analysis of user behavior on the Internet. Depending on the type of the data that is applied, according to Kosala and Blockeel (2000), web mining is categorized as: [29]

- Web Content Mining
- Web Structure Mining
- Web Usage Mining

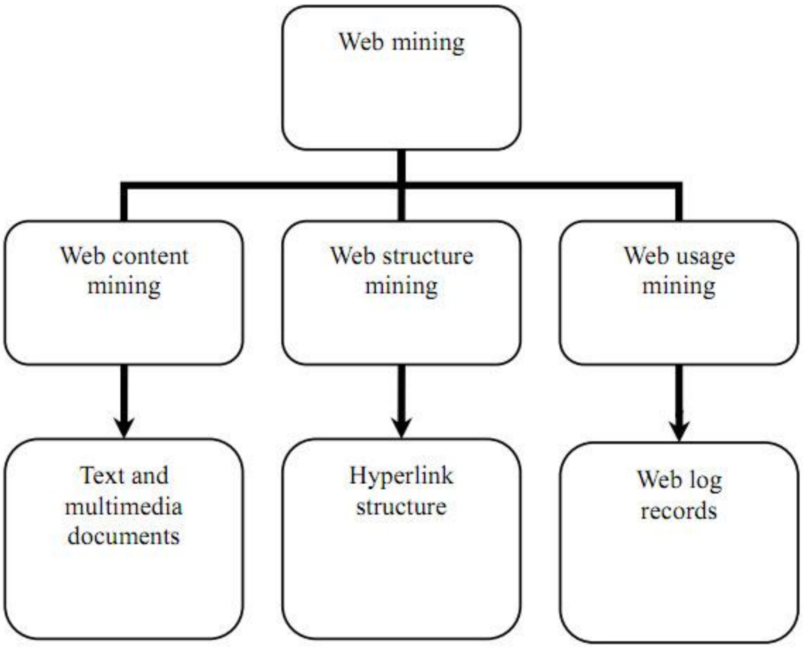


Figure 12: Web Mining and its contents

5.1 Web Content Mining

The data that are extracted in this category is based on the contents of web pages and on search results. The content can be text, image, sound or video. The aim of web content mining is to extract information from websites based on some inherent characteristics. In addition, to export data based on the content, metadata is used, like for example XML files. The challenges of web content mining are too many because the size of the web sites is too large and they have heterogeneous structure. There are also numerous texts in multiple versions and incorrect and incomplete information. This makes the need to use techniques for more correct and proper results even greater. [28]

5.2 Web Usage Mining

Web usage mining aim is the prediction of users' behavior when they are visiting the Internet. It collects data from log records in order to reveal the access patterns that a user uses in a web page. The results of extraction from such data are particularly important for administrators of web pages, to provide better services to their users and information relating to their interests. Application data containing a rating mechanism, combined with the recording of the behavior of users through the navigation, favor the mining techniques. [28]

This particular method of web mining depends on the collaboration of the user to allow the access in his log records. This is highly important because privacy is becoming a new issue to web usage mining. Users must be informed about privacy policies and then to be able to reveal their personal data by allowing the access in their web log records. [27]

5.3 Web Structure Mining

Web structure mining main goal is the hyperlink structure of the Internet. This particular method tries to identify the different links between objects. Because there are cases where traditional approaches can lead to wrong conclusions, appropriate handling of the links must be applied. This could benefit and lead to potential correlations, and then improve the predictive accuracy of the learned models. [27]

The process of knowledge discovery from data structure of the Internet can be used by search engines to assess the similarity and correlation between sites. Another interesting implementation of this method is also the social networks, which are the main focus of the dissertation. Members of these networks interact and share common interests. Because there are a lot of issues around social networks, the knowledge discovery is extremely important for designing efficient mechanisms of personalization and recommendations. [28]

6 Mining citizen’s observations in Social Network

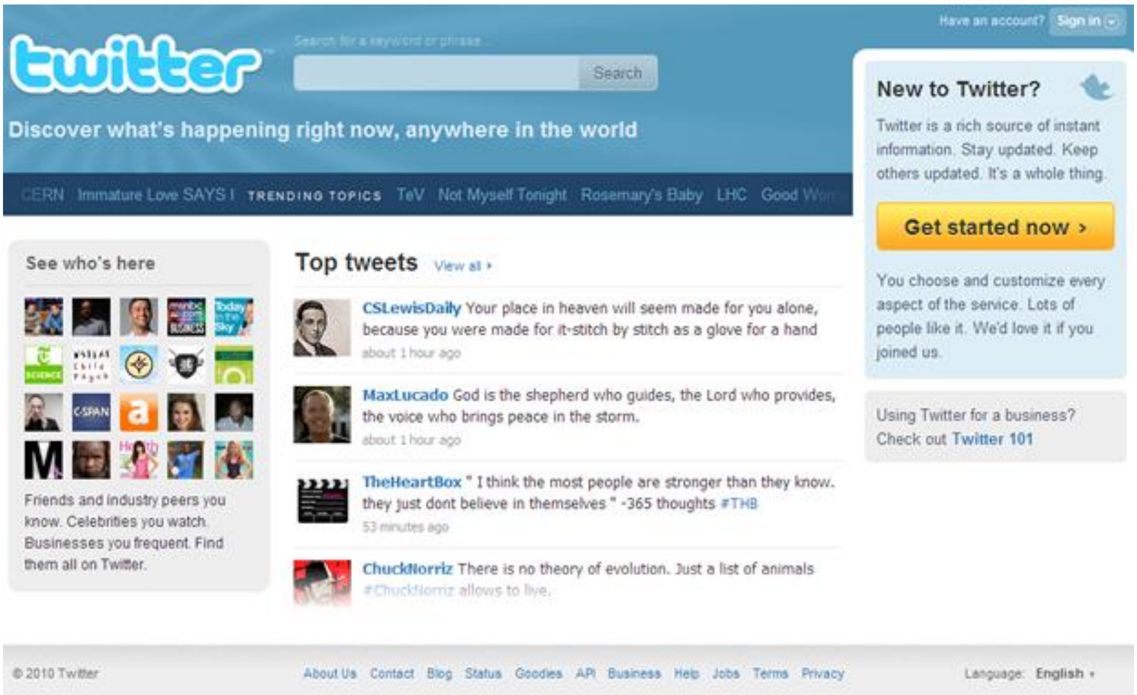
Social networks are the trend of the century. Most of the persons that have Internet experience have a profile in one or more social networks. Each and every social network has millions of users in their database. This means that opinions for everyday situations are posted quite often at their profiles. When it comes to participatory sensing matters, this information is quite interesting and important. The users within a social network, that records conditions or opinions related to the quality of the environment in an unstructured way, are called “soft sensors”, and their information regarding to the environment needs can be collected in order to important decisions to be made.

In this particular dissertation, we are going to get familiar and work with a specific social network and that is Twitter. The main reason that Twitter was the only choice is that Twitter is an “open” social network, where the information is easier to get. Facebook is the largest and most successful social network in the world. But the choice to collect information only from Twitter is because on Facebook most accounts don’t allow to be investigated without the permission of their users. Twitter users have “open” accounts which make the mining process even better to be done. Also, Twitter has been used again in the past as a source for data miming research.

6.1 Twitter

Twitter is an online social network and a micro-blogging service, which allows users, to produce short messages (140 characters) and read messages of other users of the service, known as tweets. It's more an information network and a news source. [34]

Twitter is written on a platform of open source called “Ruby on Rails”, based on the Ruby language, and has its own API (Application programming interface). The first idea of Twitter came in 2005 by Jack Dorsey. He wanted to get informed about his friends’ habits and what were their actions. Thus one day Twitter was created by a corporation named “Obvious” which has its headquarters in San Francisco. The first edition came in March 2006 and the first official appearance on the web was in August 2006. Because of Twitter's success, in May 2007 the company "Twitter Incorporated" was created. Nowadays, Twitter has more than 100 million users while on a daily basis more than 55 million tweets are exchanged. [34]


twitter
Have an account? Sign in
Search for a keyword or phrase...
Search
CERN Immature Love SAYS I TRENDING TOPICS TeV Not Myself Tonight Rosemary's Baby LHC Good Wo
See who's here
Top tweets View all ›
Get started now ›
About Us Contact Blog Status Goodies API Business Help Jobs Terms Privacy
Language: English

Discover what's happening right now, anywhere in the world

New to Twitter?

Twitter is a rich source of instant information. Stay updated. Keep others updated. It's a whole thing.

You choose and customize every aspect of the service. Lots of people like it. We'd love it if you joined us.

Using Twitter for a business? Check out Twitter 101

Friends and industry peers you know. Celebrities you watch. Businesses you frequent. Find them all on Twitter.

CSLewisDaily Your place in heaven will seem made for you alone, because you were made for it-stitch by stitch as a glove for a hand
about 1 hour ago

MaxLucado God is the shepherd who guides, the Lord who provides, the voice who brings peace in the storm.
about 1 hour ago

TheHeartBox " I think the most people are stronger than they know. they just dont believe in themselves " -365 thoughts #THB
53 minutes ago

ChuckNorriz There is no theory of evolution. Just a list of animals #ChuckNorriz allows to live.

© 2010 Twitter

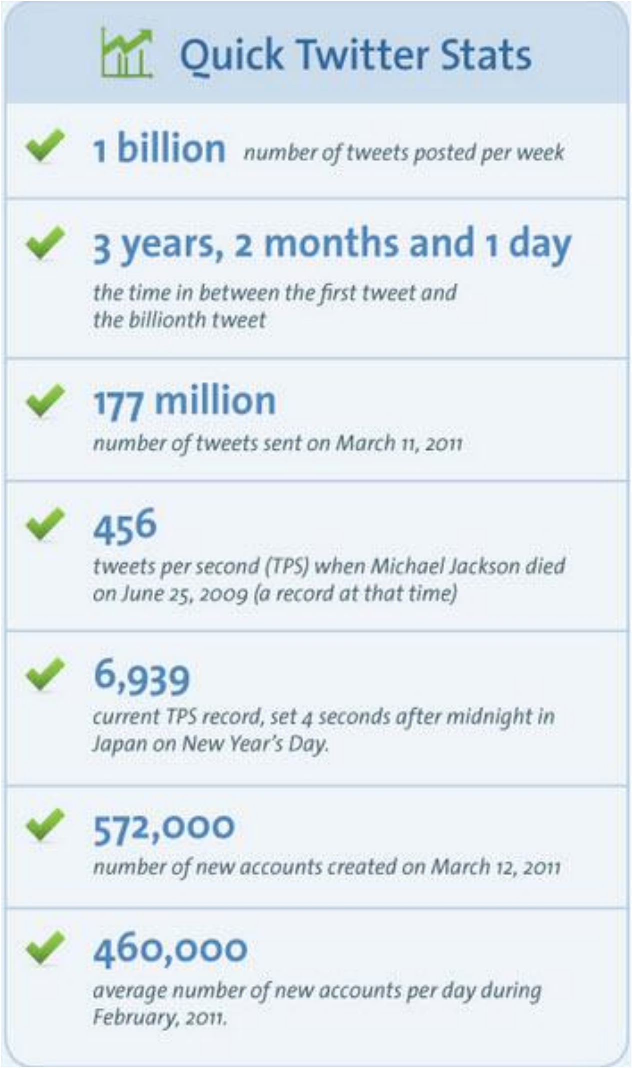
Figure 13: Twitter's main page

Twitter is based upon its small size massages, called tweets, which are similar to the Short Message Service - SMS. The only difference is the public notification. The idea of a tweet is that a user can make known any type of message like for instance what he feels or his thoughts, in a single message at any moment. In order for this to be achieved, the users must generate a network of people. These so called followers can "follow" a user by being notified for each message and the following, which is the reverse of the followers. [34]

Twitter has gained a lot of ground between the other social networks. Today, it has embodied in Twitter's main page its own search engine that can be used to search any kind of information or relevant topics. This is the main reason why Twitter has been the social network that the dissertation is referring to, because the information in this micro blogging network is quickly accessed. The right key words or phrases are enough for a successful query search. The search function on Twitter can also bring a lot of new followers, especially when people are exposed to new trends and common interests. Twitter is the latest social networking scene and is the preferred choice of many. [35]

6.2 Twitter Statistics

Twitter is been around for more than 4 years now, and its success is enormous. The statistics of 2011 speak for themselves.



Quick Twitter Stats

- ✓ **1 billion** *number of tweets posted per week*
- ✓ **3 years, 2 months and 1 day** *the time in between the first tweet and the billionth tweet*
- ✓ **177 million** *number of tweets sent on March 11, 2011*
- ✓ **456** *tweets per second (TPS) when Michael Jackson died on June 25, 2009 (a record at that time)*
- ✓ **6,939** *current TPS record, set 4 seconds after midnight in Japan on New Year's Day.*
- ✓ **572,000** *number of new accounts created on March 12, 2011*
- ✓ **460,000** *average number of new accounts per day during February, 2011.*

Figure 14: Twitter Statistics 2011

- In a week, 1 billion tweets are sent
- One year ago, 50 million tweets were sent per day
- In the last month, 140 million tweets were sent per day
- During the earthquake in Japan on March 11, 2011 more than 177 million tweets were sent
- When Michael Jackson died on June 25 2009, 456 tweets per second were sent (a record at that time).
- Worldwide record was in Japan, just 4 seconds after midnight on New Year's Day, 6,939 tweets were sent
- On March 12 2011, 572,000 new Twitter accounts were created.
- Over the last month more than 460,000 new Twitter accounts were created per day
- Over the past year, there were 182% of increase in the number of mobile Twitter users [32]

Some other interesting Twitter statistics are related to the types of the Twitter messages. The content of the tweets varies 27% of it have to do with private conversations of the users and 30% about their current status. The remaining proportion has to do with links, spams or advertisements. It has great importance the fact that 11 % of the publications have to do with advertising messages, which were noticed in Twitter in August 2009. [33]

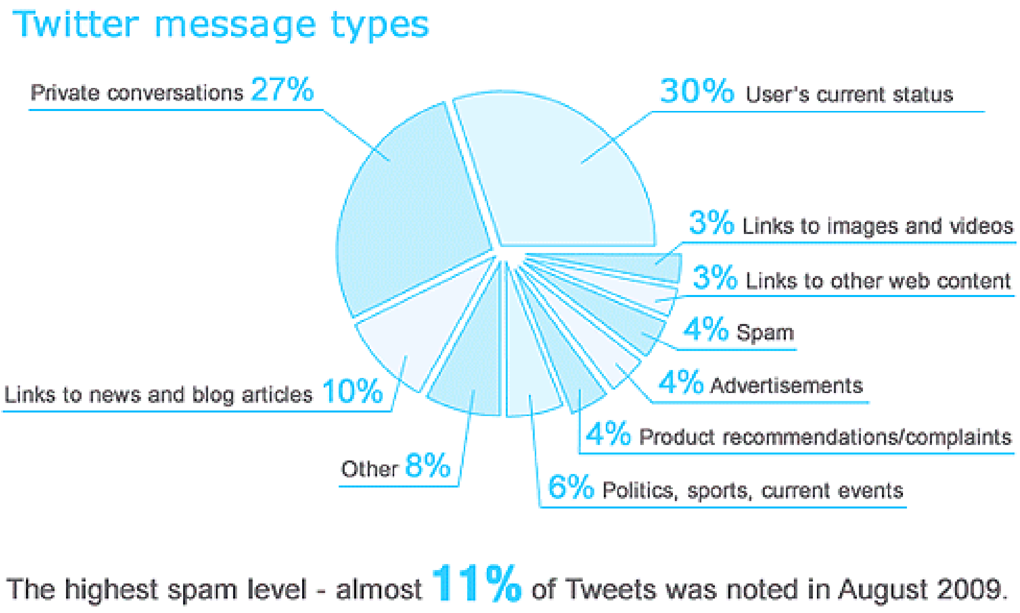


Figure 15: Twitter message types

7 Text mining methodology

The main goal of the dissertation is to use text mining techniques in order to retrieve useful information about participatory sensing matters. In order to get a clearer picture of the web posts, the idea is to monitor all social media with an internet web tool, following the monitor of the last posts in Twitter with another web tool. The main steps are the following:

- Queries based on text annotations will be used. Specific keywords will be searched with a selection of internet web tools that search within Twitter and social media in order for us to get informed about environmental matters in a specific area. That specific area will be the town of Thessaloniki in Greece.
- The specific keywords that we are going to use will be both in Greek and In English language. These keywords are the following:
    - “Ρύπανση” (pollution)
    - “Ατμοσφαιρική ρύπανση” (Air pollution)
    - “Ποιότητα αέρα” (Air Quality)
    - “Αιωρούμενα σωματίδια” (Particulates)
    - “Απορρίμματα” (Waste)
    - “Σκουπίδια” (Garbage)
    - “Συγκοινωνιακό” (Transportation)
    - “Θόρυβος” (Noise)
    - “Ηχορύπανση” (Noise pollution)
    - “Περιβαλλοντική πληροφορία” (Environmental information)
    - “Περιβαλλοντική πληροφόρηση” (Environmental informing)
- Also, the Greek keywords will be searched without the word Thessaloniki in order to distinguish the frequency of the terms with or without the specific area.
- After the appropriate query searches with the internet web tools, the results will be quoted in tables revealing useful information about participatory matters. The idea is to make use of the information concerning the searched terms, and the way that they are used, in order to investigate the possibility to extract useful information about the state of the environment. On this basis, we will be able to estimate the

potential success of the use of social networks as sources and platforms for participatory environmental sensing information.

7.1 Available Text Mining tools

There are a lot of web pages that can perform text mining and give back to the user a certain amount of information in response to the given request. This kind of pages take as input keywords or phrases and the result is that they search the internet across the globe to match this words that are given and return a positive feedback. In this dissertation, two specific internet web tools will be used and those are Social Mention and Trending.

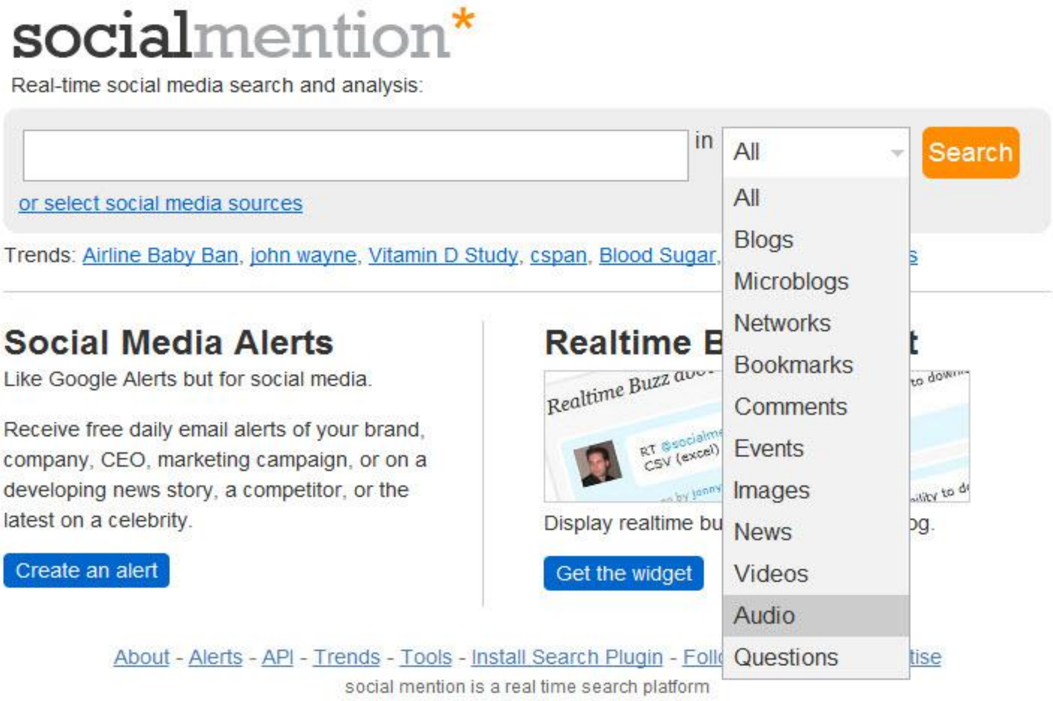
Except from online text mining tools, there are also software that performs text mining techniques with similar functionalities and sometimes even more. It is really hard to acquire a text mining program because they are not freeware and their use is restricted to companies for internal use or research. Usually, companies make their own text mining programs in order to retrieve information about them through the web. The only program that we were able to acquire was TwitMiner.

TwitMiner is a desktop application that allows gathering public information from various tweets adding criteria. These criteria are geographical points like longitude and latitude of a place, as well as keywords that can be searched. Because the protocol that the streamer is using, examines only 10% of the posts, we cannot gather useful information about environmental issues. For that reason, only a reference for the software is being made in this section.

Text mining and text searching or information retrieval are two concepts that are usually confusing. For that reason, it is highly important to differentiate them before we start perform text mining using available tools. The main goal of information retrieval is to assist people find documents that satisfy their information needs. For example, this is what Google search engine does by providing links of the text search that has been queried. Text mining on the other hand, is a huge area compared to information retrieval. Text mining performs mining tasks like document classification, document clustering, building ontology, sentiment analysis, document summarization, information extraction etc. For example, a text mining tool will return except of the documents or links, extra information that has to do with the sentiment of the results, the users, the sources and other important information. [38] [39]

7.1.1 Social Mention

One of the most popular text mining web pages is the socialmention.com. This particular web page offers real-time social media search and analysis. Social Mention was introduced to the public in September 2008 by Jon Cianciullo. The site averages about 40,000 unique visitors per month. It's a social media platform which gathers content that is generated by users into a single stream of information. It has the ability to search the internet through social media and more specific it searches blogs, micro blogs, networks, bookmarks, comments, events, images, news, videos, audio and questions. Besides that, the most popular trends appear in the main page, some social media alerts and also the option to download a widget and display real time information on someone's blog or page.



Real-time social media search and analysis.

Trends: Airline Baby Ban, john wayne, Vitamin D Study, cspan, Blood Sugar,

Social Media Alerts

Like Google Alerts but for social media.

Receive free daily email alerts of your brand, company, CEO, marketing campaign, or on a developing news story, a competitor, or the latest on a celebrity.

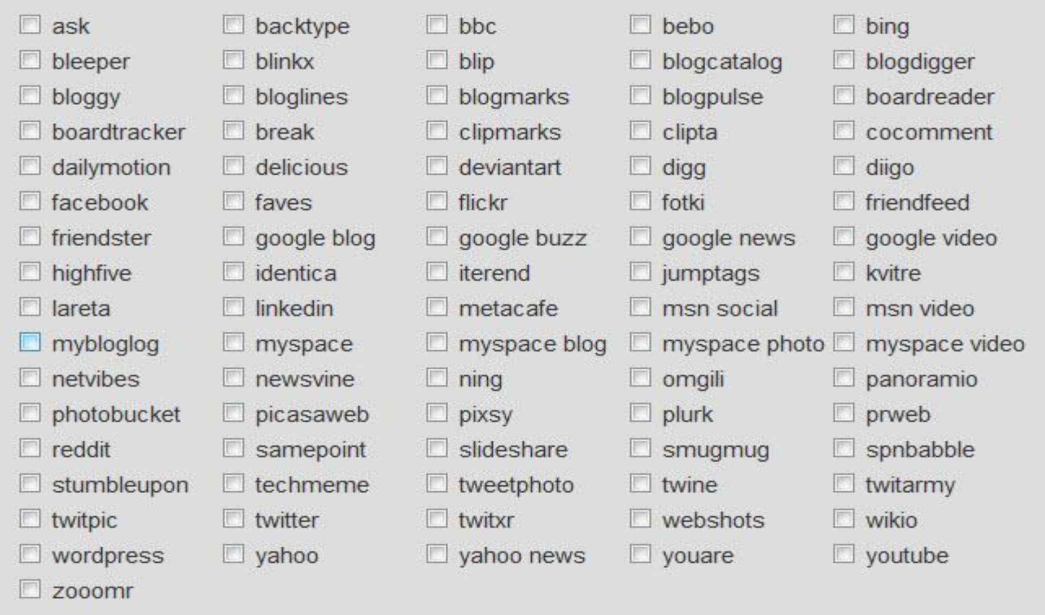
Realtime B

Display realtime bu

social mention is a real time search platform

Figure 16: Social Mention main page

When there is the need to search a specific social media source, there is also the option to select the social media source that is desirable. It allows easy track to what people are saying by monitoring 80+ social media. In the following figure, there are many different social media sources and there are all available through the web site for a more specific query.



- ☐ ask
- ☐ backtype
- ☐ bbc
- ☐ bebo
- ☐ bing
- ☐ bleeper
- ☐ blinkx
- ☐ blip
- ☐ blogcatalog
- ☐ blogdigger
- ☐ bloggy
- ☐ bloglines
- ☐ blogmarks
- ☐ blogpulse
- ☐ boardreader
- ☐ boardtracker
- ☐ break
- ☐ clipmarks
- ☐ clipta
- ☐ cocomment
- ☐ dailymotion
- ☐ delicious
- ☐ deviantart
- ☐ digg
- ☐ diigo
- ☐ facebook
- ☐ faves
- ☐ flickr
- ☐ fotki
- ☐ friendfeed
- ☐ friendster
- ☐ google blog
- ☐ google buzz
- ☐ google news
- ☐ google video
- ☐ highfive
- ☐ identica
- ☐ iterend
- ☐ jumptags
- ☐ kvitre
- ☐ lareta
- ☐ linkedin
- ☐ metacafe
- ☐ msn social
- ☐ msn video
- ☐ mybloglog
- ☐ myspace
- ☐ myspace blog
- ☐ myspace photo
- ☐ myspace video
- ☐ netvibes
- ☐ newsvine
- ☐ ning
- ☐ omgili
- ☐ panoramio
- ☐ photobucket
- ☐ picasaweb
- ☐ pixsy
- ☐ plurk
- ☐ prweb
- ☐ reddit
- ☐ samepoint
- ☐ slideshare
- ☐ smugmug
- ☐ spnbabble
- ☐ stumbleupon
- ☐ techmeme
- ☐ tweetphoto
- ☐ twine
- ☐ twitarmy
- ☐ twitpic
- ☐ twitter
- ☐ twitxr
- ☐ webshots
- ☐ wikio
- ☐ wordpress
- ☐ yahoo
- ☐ yahoo news
- ☐ youare
- ☐ youtube
- ☐ zooomr

Figure 17: Social Media sources available

When a user wants to search for something, simply enters the terms for tracking. The process in order to be completed takes a few minutes and then a results stream is generated. After the process is completed, the results that are returned contain a stream of the mentions of each keyword as well as several analyses. At the top of the page several features are appeared. First of all is the "strength" attribute which calculates mentions within the last 24 hours divided by the total possible mentions. This particular attribute indicates the likelihood that the search terms are being discussed. The next attribute is the "sentiment" ratio which determines the positive to negative mentions. The "passion" attribute shows the probability that individuals will continue to talk about the query keyword. Finally, the "reach" attribute determines the measure of influence. It divides the number of unique authors by the total mentions. [37]

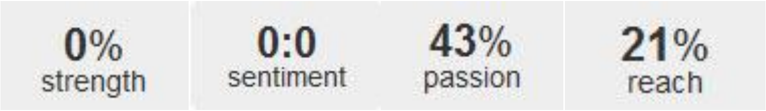
In that point it is important to understand more about the sentiment attribute, because it has a crucial role in a text mining analysis. Sentiment analysis or opinion mining as it is known, aims to identify the opinion and the feeling of a user with respect to a topic or a document. There are quite a few methods of how to perform sentiment analysis. Computers have the ability to do that, using elements from machine learning such as latent semantic analysis, support vector machines, "bag of words" and Semantic Orientation. There are more sophisticated methods that try to identify the holder of a sentiment and the target. In order to get the feeling out of a text or a post, the grammatical relationships of words are used. Grammatical dependency relations are obtained by deep parsing of the text. [40]

In order for a query to start, a key word or a phrase must be typed in the search box and then press search. The specific keywords will be listed in the following paragraphs followed by the keyword of the town Thessaloniki as well as the most important results. The query search will be done with two ways for the specific internet tool: only for Twitter and for all the social media that Social Mention has. The date that the query search took place is August 17, 2011. The following pages show the results and the useful information that are gathered from each query search.

“Ρύπανση”

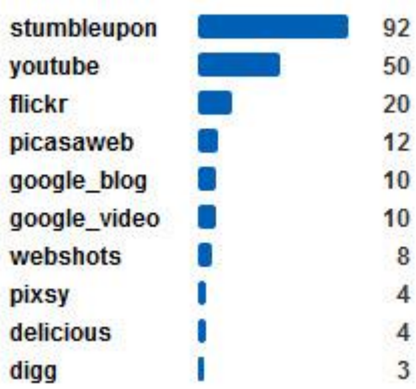
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 213 mentions concerning “Ρύπανση”
  - The last mention was 8 days ago



0%	0:0	43%	21%
strength	sentiment	passion	reach

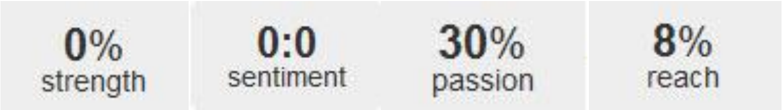
    - There is 0% strength that the keyword “Ρύπανση” is being discussed in social media
    - The sentiment of the users is neutral for all mentions
    - There is 43% that individuals talk about the specific keyword repeatedly in social media
    - There is 21% range of influence by the users
  - There are 102 unique authors who are talking about this particular subject
  - The sources are mentioned where the key phrase was found and also the number of items in the results. The mentions for that keyword are from:



“Ρύπανση Θεσσαλονίκη”

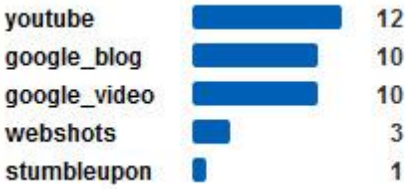
- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 36 mentions concerning “ρύπανση Θεσσαλονίκη”
    - The last mention was 2 months ago



0% strength | 0:0 sentiment | 30% passion | 8% reach

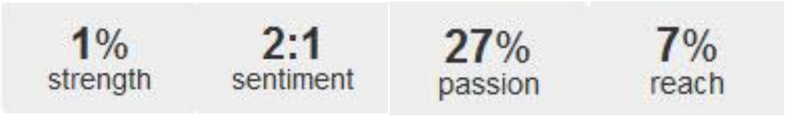
-
    -
        - There is 0% strength that the keyword “ρύπανση Θεσσαλονίκη” is being discussed in social media
        - The sentiment of the users is neutral for all mentions
        - There is 30% that individuals talk about the specific keyword repeatedly in social media
        - There is 8% range of influence by the users
    - There are 17 unique authors who are talking about this particular subject
    - The sources are mentioned where the key phrase was found and also the number of items in the results. The mentions for that keyword are from:



“Pollution Thessaloniki”

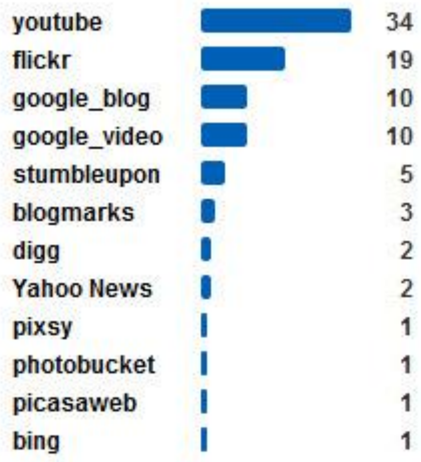
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 89 mentions concerning “Pollution Thessaloniki”
  - The last mention was 2 days ago



1%	2:1	27%	7%
strength	sentiment	passion	reach

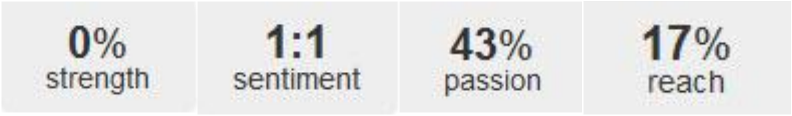
    - There is 1% strength that the keyword “Pollution Thessaloniki” is being discussed in social media
    - The sentiment of the users is 2:1. With more detail the sentiments of the users are: 12 positive, 71 neutral and 6 negative
    - There is 27% that individuals talk about the specific keyword repeatedly in social media
    - There is 7% range of influence by the users
  - There are 38 unique authors
  - The mentions for that keyword are from:



“Ατμοσφαιρική ρύπανση”

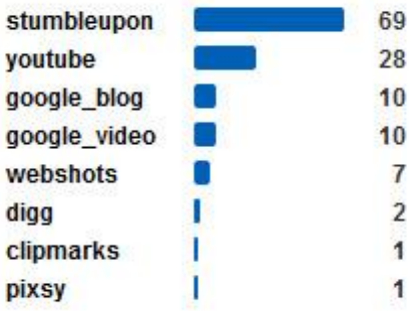
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 128 mentions concerning “Ατμοσφαιρική ρύπανση”
  - The last mention was 12 days ago



0% strength | 1:1 sentiment | 43% passion | 17% reach

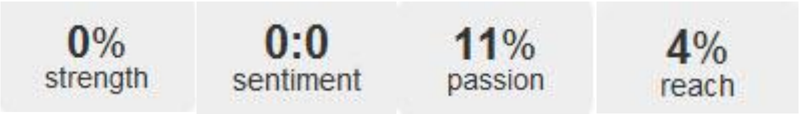
    - There is 0% strength that the keyword “Ατμοσφαιρική ρύπανση” is being discussed in social media
    - The sentiment of the users is 1:1. With more detail, the sentiments of the users are: 1 positive, 126 neutral and 1 negative
    - There is 43% that individuals talk about the specific keyword repeatedly in social media
    - There is 17% range of influence by the users
  - There are 66 unique authors
  - The mentions for that keyword are from:



“Ατμοσφαιρική ρύπανση Θεσσαλονίκη”

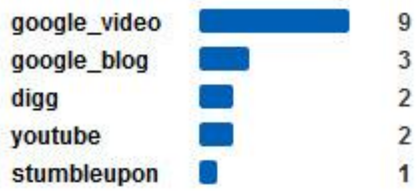
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 18 mentions concerning “Ατμοσφαιρική ρύπανση Θεσσαλονίκη”
  - The last mention was 2 month ago



0% strength | 0:0 sentiment | 11% passion | 4% reach

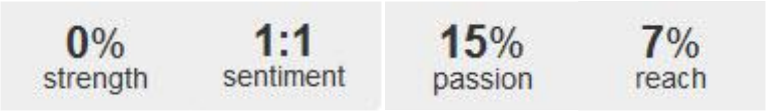
    - There is 0% strength that the keyword “Ατμοσφαιρική ρύπανση Θεσσαλονίκη” is being discussed in social media
    - The sentiment of the users is neutral for all mentions
    - There is 11% that individuals talk about the specific keyword repeatedly in social media
    - There is 4% range of influence by the users
  - There are 8 unique authors
  - The mentions for that keyword are from:



“Air pollution Thessaloniki”

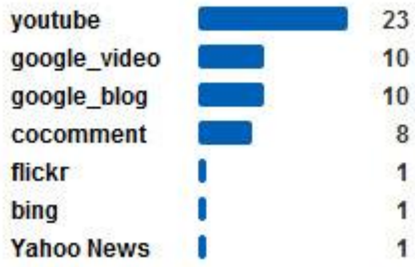
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 55 mentions concerning “Air pollution Thessaloniki”
  - The last mention was 3 days ago



0%	1:1	15%	7%
strength	sentiment	passion	reach

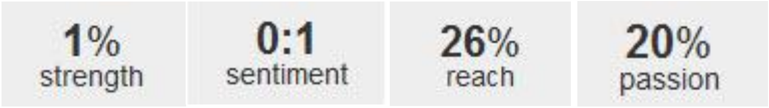
    - There is 0% strength that the keyword “Air pollution Thessaloniki” is being discussed in social media
    - The sentiment of the users is 1:1. With more detail, the sentiments of the users are: 12 positive, 34 neutral and 9 negative
    - There is 15% that individuals talk about the specific keyword repeatedly in social media
    - There is 7% range of influence by the users
  - There are 28 unique authors
  - The mentions for that keyword are from:



“Ποιότητα αέρα”

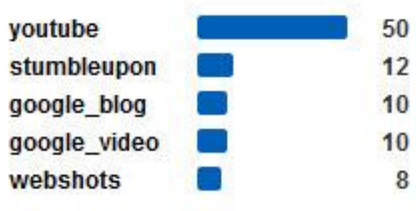
- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 90 mentions concerning “Ποιότητα αέρα”
    - The last mention was 21 hours ago



1%	0:1	26%	20%
strength	sentiment	reach	passion

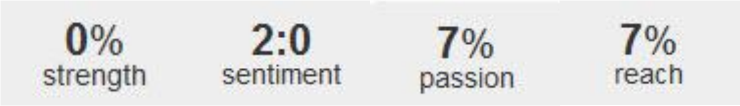
        - There is 1% strength that the keyword “Ποιότητα αέρα” is being discussed in social media
        - The sentiment of the users is 1:0. With more detail, the sentiments of the users are: 0 positive, 89 neutral and 1 negative
        - There is 26% that individuals talk about the specific keyword repeatedly in social media
        - There is 20% range of influence by the users
    - There are 64 unique authors
    - The mentions for that keyword are from:



“Ποιότητα αέρα Θεσσαλονίκη”

- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 24 mentions concerning “Ποιότητα αέρα Θεσσαλονίκη”
  - The last mention was 2 month ago



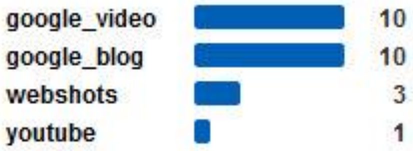
0% strength

2:0 sentiment

7% passion

7% reach

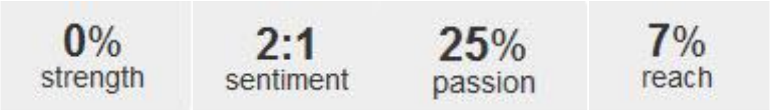
    - There is 0% strength that the keyword “Ατμοσφαιρική ρύπανση Θεσσαλονίκη” is being discussed in social media
    - The sentiment of the users is 2:0. With more detail, the sentiments of the users are: 2 positive, 22 neutral and 0 negative
    - There is 7% that individuals talk about the specific keyword repeatedly in social media
    - There is 7% range of influence by the users
  - There are 13 unique authors
  - The mentions for that keyword are from:



“Air quality Thessaloniki”

- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 51 mentions concerning “Air quality Thessaloniki”
    - The last mention was 8 days ago



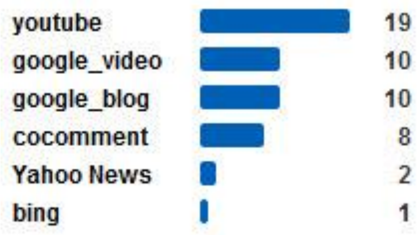
0% strength

2:1 sentiment

25% passion

7% reach

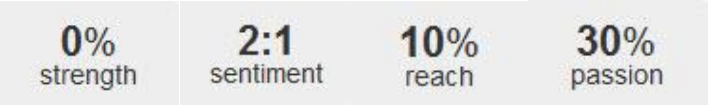
        - There is 0% strength that the keyword “Air quality Thessaloniki” is being discussed in social media
        - The sentiment of the users is 2:1. With more detail, the sentiments of the users are: 12 positive, 33 neutral and 6 negative
        - There is 25% that individuals talk about the specific keyword repeatedly in social media
        - There is 7% range of influence by the users
    - There are 22 unique authors
    - The mentions for that keyword are from:



“Αιωρούμενα σωματίδια”

- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 42 mentions concerning “Αιωρούμενα σωματίδια”
    - The last mention was 28 days ago



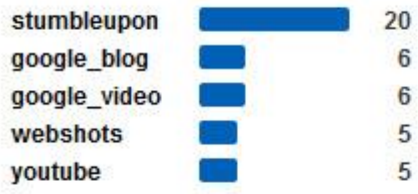
0% strength

2:1 sentiment

10% reach

30% passion

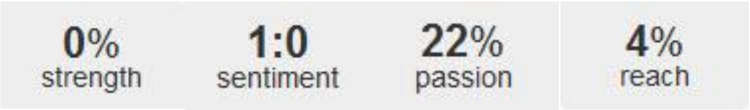
        - There is 0% strength that the keyword “Αιωρούμενα σωματίδια” is being discussed in social media
        - The sentiment of the users is 2:1. With more detail, the sentiments of the users are: 2 positive, 39 neutral and 1 negative
        - There is 10% that individuals talk about the specific keyword repeatedly in social media
        - There is 30% range of influence by the users
    - There are 25 unique authors
    - The mentions for that keyword are from:



“Αιωρούμενα σωματίδια Θεσσαλονίκη”

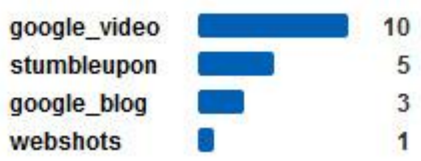
- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 19 mentions concerning “Αιωρούμενα σωματίδια Θεσσαλονίκη”
    - The last mention was 2 months ago



0% strength | 1:0 sentiment | 22% passion | 4% reach

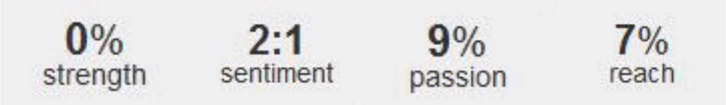
-
    -
        - There is 0% strength that the keyword “Αιωρούμενα σωματίδια Θεσσαλονίκη” is being discussed in social media
        - The sentiment of the users is 1:0. With more detail, the sentiments of the users are: 1 positive, 18 neutral and 0 negative
        - There is 22% that individuals talk about the specific keyword repeatedly in social media
        - There is 4% range of influence by the users
    - There are 7 unique authors
    - The mentions for that keyword are from:



“Particulates Thessaloniki”

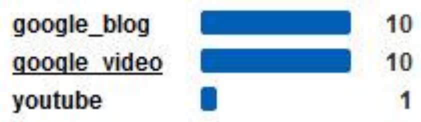
- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 21 mentions concerning “ Particulates Thessaloniki ”
    - The last mention was 2 days ago



0%	2:1	9%	7%
strength	sentiment	passion	reach

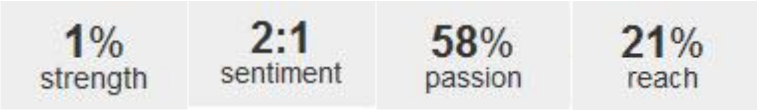
        - There is 0% strength that the keyword “Particulates Thessaloniki” is being discussed in social media
        - The sentiment of the users is 2:1. With more detail, the sentiments of the users are: 4 positive, 15 neutral and 2 negative
        - There is 9% that individuals talk about the specific keyword repeatedly in social media
        - There is 7% range of influence by the users
    - There are 10 unique authors
    - The mentions for that keyword are from:



“Απορρίμματα”

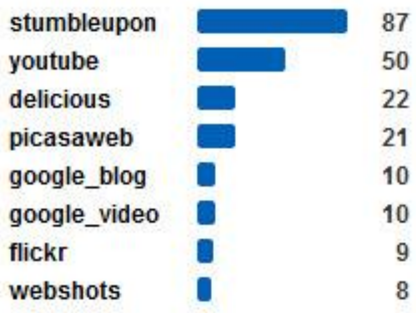
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 217 mentions concerning “Απορρίμματα”
  - The last mention was 10 hours ago



1%	2:1	58%	21%
strength	sentiment	passion	reach

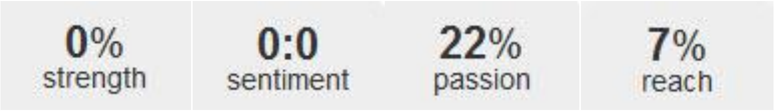
    - There is 1% strength that the keyword “Απορρίμματα” is being discussed in social media
    - The sentiment of the users is 2:1. With more detail, the sentiments of the users are: 2 positive, 214 neutral and 1 negative
    - There is 58% that individuals talk about the specific keyword repeatedly in social media
    - There is 21% range of influence by the users
  - There are 82 unique authors
  - The mentions for that keyword are from:



“Απορρίμματα Θεσσαλονίκη”

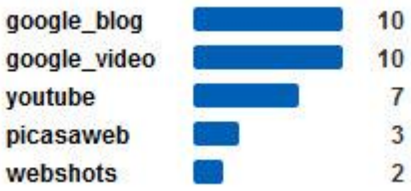
- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 32 mentions concerning “Απορρίμματα Θεσσαλονίκη”
    - The last mention was 26 days ago



0% strength | 0:0 sentiment | 22% passion | 7% reach

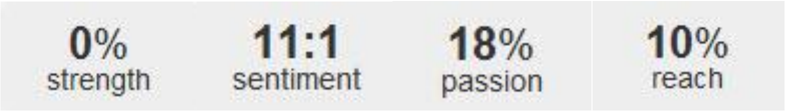
        - There is 0% strength that the keyword “Απορρίμματα Θεσσαλονίκη” is being discussed in social media
        - The sentiment of the users is neutral for all mentions
        - There is 22% that individuals talk about the specific keyword repeatedly in social media
        - There is 7% range of influence by the users
    - There are 17 unique authors
    - The mentions for that keyword are from:



“Waste Thessaloniki”

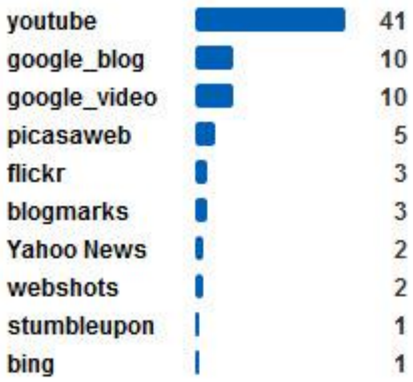
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 78 mentions concerning “Waste Thessaloniki”
  - The last mention was 3 days ago



0%	11:1	18%	10%
strength	sentiment	passion	reach

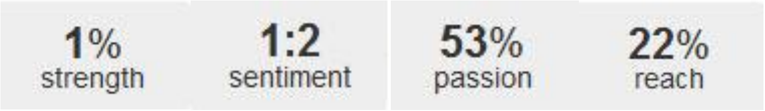
    - There is 0% strength that the keyword “Waste Thessaloniki” is being discussed in social media
    - The sentiment of the users is 11:1. With more detail, the sentiments of the users are: 11 positive, 66 neutral and 1 negative
    - There is 18% that individuals talk about the specific keyword repeatedly in social media
    - There is 10% range of influence by the users
  - There are 47 unique authors
  - The mentions for that keyword are from:



“Σκουπίδια”

- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 268 mentions concerning “Σκουπίδια”
  - The last mention was 26 days ago



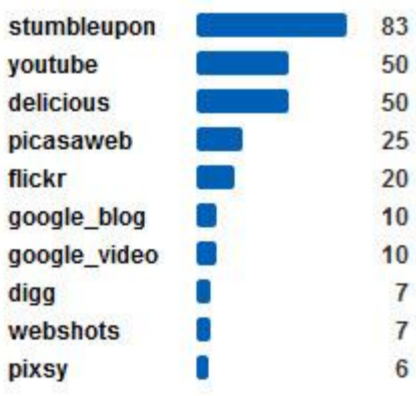
1% strength

1:2 sentiment

53% passion

22% reach

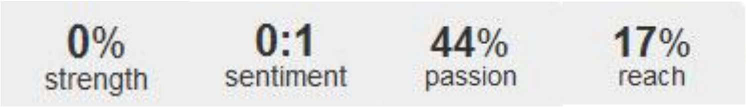
    - There is 1% strength that the keyword “Σκουπίδια” is being discussed in social media
    - The sentiment of the users is 1:2. With more detail, the sentiments of the users are: 3 positive, 260 neutral and 5 negative
    - There is 53% that individuals talk about the specific keyword repeatedly in social media
    - There is 22% range of influence by the users
  - There are 108 unique authors
  - The mentions for that keyword are from:



“Σκουπίδια Θεσσαλονίκη”

- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 131 mentions concerning “Σκουπίδια Θεσσαλονίκη”
  - The last mention was 26 days ago



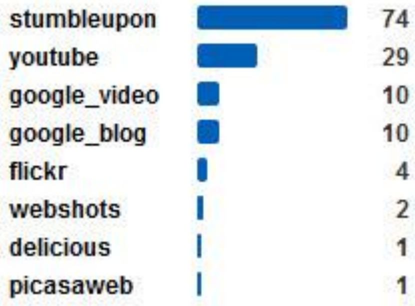
0% strength

0:1 sentiment

44% passion

17% reach

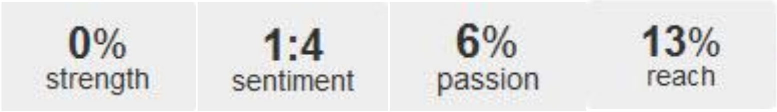
    - There is 0% strength that the keyword “Σκουπίδια Θεσσαλονίκη” is being discussed in social media
    - The sentiment of the users is 0:1. With more detail, the sentiments of the users are: 0 positive, 130 neutral and 1 negative
    - There is 44% that individuals talk about the specific keyword repeatedly in social media
    - There is 17% range of influence by the users
  - There are 65 unique authors
  - The mentions for that keyword are from:



“Garbage Thessaloniki”

- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 81 mentions concerning “Garbage Thessaloniki”
    - The last mention was 5 day before



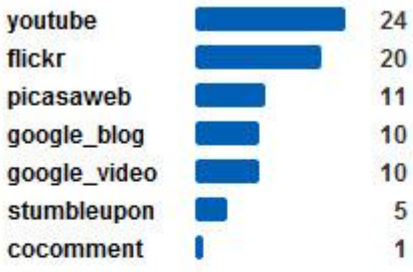
0% strength

1:4 sentiment

6% passion

13% reach

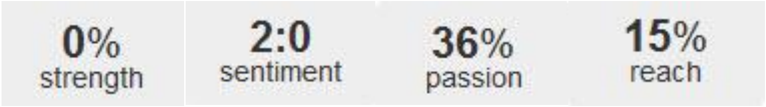
-
    -
        - There is 0% strength that the keyword “Garbage Thessaloniki” is being discussed in social media
        - The sentiment of the users is 1:4. With more detail, the sentiments of the users are: 5 positive, 54 neutral and 22 negative
        - There is 6% that individuals talk about the specific keyword repeatedly in social media
        - There is 13% range of influence by the users
    - There are 45 unique authors
    - The mentions for that keyword are from:



“Συγκοινωνιακό”

- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 68 mentions concerning “Συγκοινωνιακό”
    - The last mention was 11 days ago



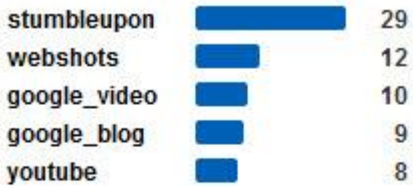
0% strength

2:0 sentiment

36% passion

15% reach

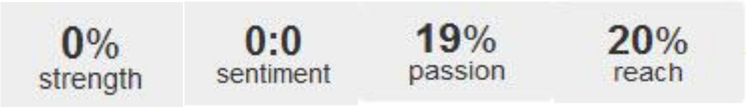
-
    -
        - There is 0% strength that the keyword “Συγκοινωνιακό” is being discussed in social media
        - The sentiment of the users is 2:0. With more detail, the sentiments of the users are: 2 positive, 66 neutral and 0 negative
        - There is 36% that individuals talk about the specific keyword repeatedly in social media
        - There is 15% range of influence by the users
    - There are 37 unique authors
    - The mentions for that keyword are from:



“Συγκοινωνιακό Θεσσαλονίκη”

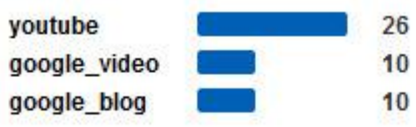
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 46 mentions concerning “Συγκοινωνιακό Θεσσαλονίκη”
  - The last mention was 14 days ago



0% strength | 0:0 sentiment | 19% passion | 20% reach

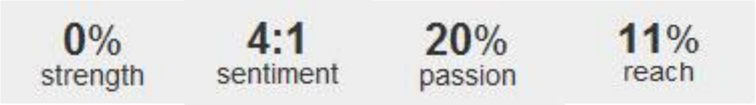
    - There is 0% strength that the keyword “Συγκοινωνιακό Θεσσαλονίκη” is being discussed in social media
    - The sentiment of the users is neutral for all mentions
    - There is 19% that individuals talk about the specific keyword repeatedly in social media
    - There is 20% range of influence by the users
  - There are 29 unique authors
  - The mentions for that keyword are from:



“Transportation Thessaloniki”

- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 131 mentions concerning “Transportation Thessaloniki”
  - The last mention was 1 day ago



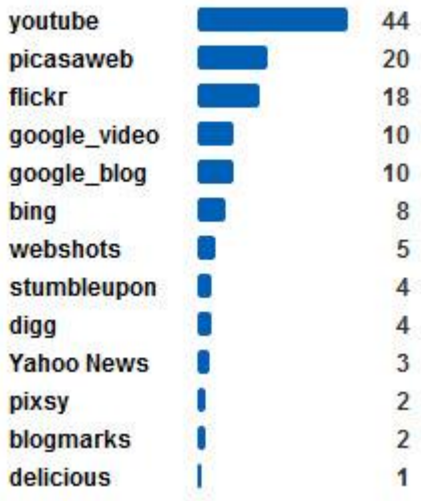
0% strength

4:1 sentiment

20% passion

11% reach

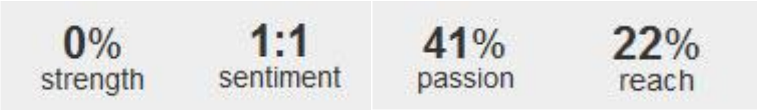
    - There is 0% strength that the keyword “Transportation Thessaloniki” is being discussed in social media
    - The sentiment of the users is 4:1. With more detail, the sentiments of the users are: 19 positive, 107 neutral and 5 negative
    - There is 20% that individuals talk about the specific keyword repeatedly in social media
    - There is 11% range of influence by the users
  - There are 70 unique authors
  - The mentions for that keyword are from:



“Θόρυβος”

- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 159 mentions concerning “Θόρυβος”
  - The last mention was 6 days ago



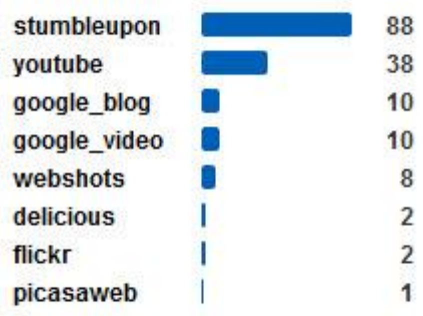
0% strength

1:1 sentiment

41% passion

22% reach

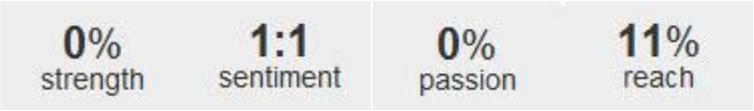
    - There is 0% strength that the keyword “Θόρυβος” is being discussed in social media
    - The sentiment of the users is 1:1. With more detail, the sentiments of the users are: 6 positive, 147 neutral and 6 negative
    - There is 41% that individuals talk about the specific keyword repeatedly in social media
    - There is 22% range of influence by the users
  - There are 86 unique authors
  - The mentions for that keyword are from:



“Θόρυβος Θεσσαλονίκη”

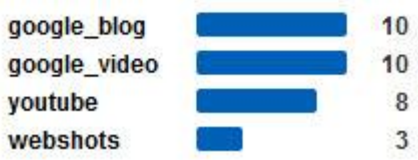
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 31 mentions concerning “Θόρυβος Θεσσαλονίκη”
  - The last mention was 2 days ago



0% strength
1:1 sentiment
0% passion
11% reach

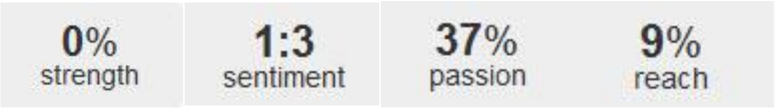
    - There is 0% strength that the keyword “Θόρυβος Θεσσαλονίκη” is being discussed in social media
    - The sentiment of the users is 1:1. With more detail, the sentiments of the users are: 1 positive, 29 neutral and 1 negative
    - There is 0% that individuals talk about the specific keyword repeatedly in social media
    - There is 11% range of influence by the users
  - There are 21 unique authors
  - The mentions for that keyword are from:



“Noise Thessaloniki”

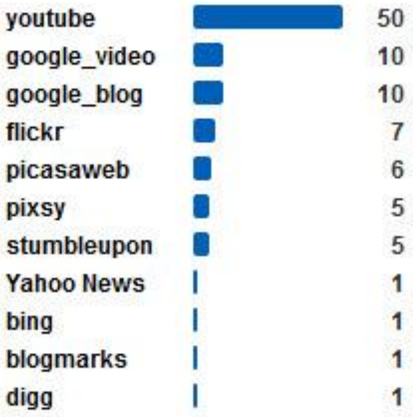
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 97 mentions concerning “Noise Thessaloniki”
  - The last mention was 3 days ago



0%	1:3	37%	9%
strength	sentiment	passion	reach

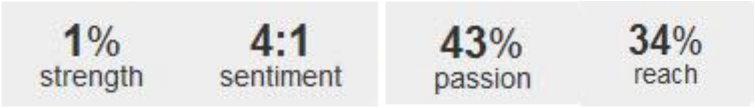
    - There is 0% strength that the keyword “Noise Thessaloniki” is being discussed in social media
    - The sentiment of the users is 1:3. With more detail, the sentiments of the users are: 15 positive, 43 neutral and 39 negative
    - There is 37% that individuals talk about the specific keyword repeatedly in social media
    - There is 9% range of influence by the users
  - There are 43 unique authors
  - The mentions for that keyword are from:



“Ηχορύπανση”

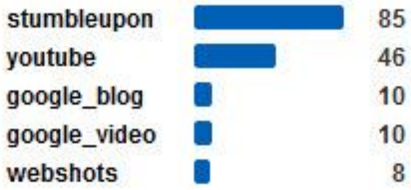
- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 159 mentions concerning “Ηχορύπανση”
    - The last mention was 2 days ago



1% strength 4:1 sentiment 43% passion 34% reach

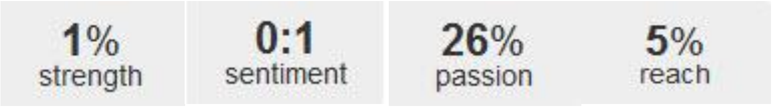
        - There is 1% strength that the keyword “Ηχορύπανση” is being discussed in social media
        - The sentiment of the users is 4:1. With more detail, the sentiments of the users are: 4 positive, 154 neutral and 1 negative
        - There is 43% that individuals talk about the specific keyword repeatedly in social media
        - There is 34% range of influence by the users
    - There are 84 unique authors
    - The mentions for that keyword are from:



“Ηχορύπανση Θεσσαλονίκη”

- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 25 mentions concerning “Ηχορύπανση Θεσσαλονίκη”
  - The last mention was 16 hours ago



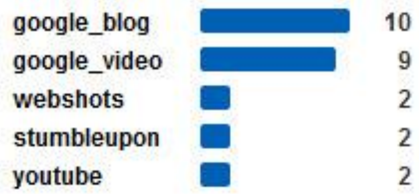
1% strength

0:1 sentiment

26% passion

5% reach

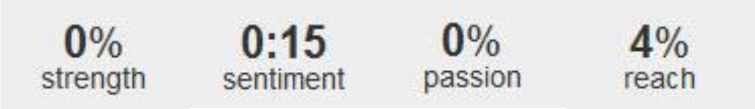
  - There is 1% strength that the keyword “Ηχορύπανση Θεσσαλονίκη” is being discussed in social media
  - The sentiment of the users is 0:1. With more detail, the sentiments of the users are: 0 positive, 24 neutral and 1 negative
  - There is 26% that individuals talk about the specific keyword repeatedly in social media
  - There is 5% range of influence by the users
  - There are 11 unique authors
  - The mentions for that keyword are from:



“Noise pollution Thessaloniki”

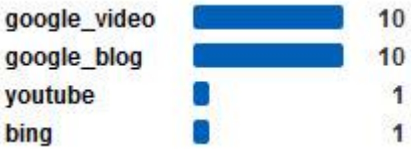
- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 22 mentions concerning “Noise pollution Thessaloniki”
    - The last mention was 3 days ago



0% strength | 0:15 sentiment | 0% passion | 4% reach

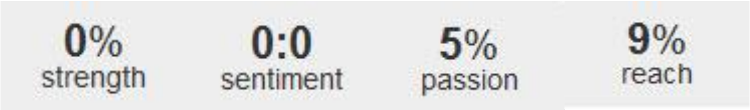
-
    -
        - There is 0% strength that the keyword “Noise pollution Thessaloniki” is being discussed in social media
        - The sentiment of the users is 0:15. With more detail, the sentiments of the users are: 0 positive, 7 neutral and 15 negative
        - There is 0% that individuals talk about the specific keyword repeatedly in social media
        - There is 4% range of influence by the users
    - There are 7 unique authors
    - The mentions for that keyword are from:



“Περιβαλλοντική πληροφορία”

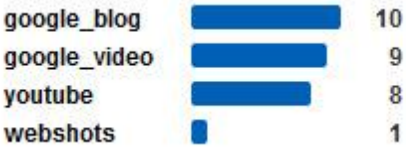
- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 28 mentions concerning “Περιβαλλοντική πληροφορία”
    - The last mention was 22 days ago



0% strength | 0:0 sentiment | 5% passion | 9% reach

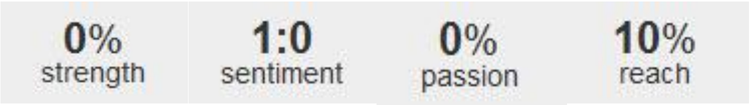
-
    -
        - There is 0% strength that the keyword “Περιβαλλοντική πληροφορία” is being discussed in social media
        - The sentiment of the users is neutral for all mentions
        - There is 5% that individuals talk about the specific keyword repeatedly in social media
        - There is 9% range of influence by the users
    - There are 17 unique authors
    - The mentions for that keyword are from:



“Περιβαλλοντική πληροφορία Θεσσαλονίκη”

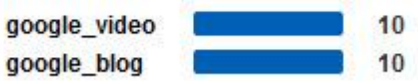
- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 20 mentions concerning “Περιβαλλοντική πληροφορία Θεσσαλονίκη”
    - The last mention was 12 days ago



0% strength | 1:0 sentiment | 0% passion | 10% reach

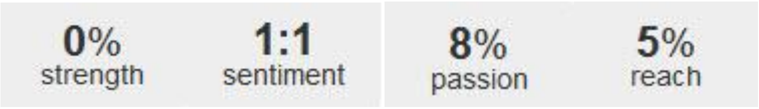
-
    -
        - There is 0% strength that the keyword “Περιβαλλοντική πληροφορία Θεσσαλονίκη” is being discussed in social media
        - The sentiment of the users is 1:0. With more detail, the sentiments of the users are: 1 positive, 19 neutral and 0 negative
        - There is 0% that individuals talk about the specific keyword repeatedly in social media
        - There is 10% range of influence by the users
    - There are 10 unique authors
    - The mentions for that keyword are from:



“Environmental information Thessaloniki”

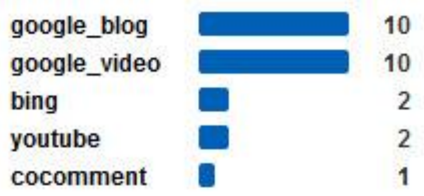
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 25 mentions concerning “Environmental information Thessaloniki”
  - The last mention was 1 day ago



0% strength | 1:1 sentiment | 8% passion | 5% reach

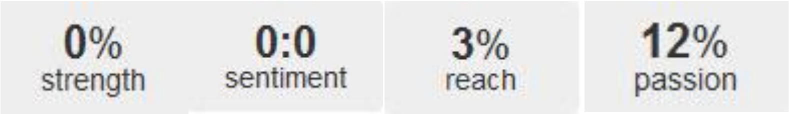
  - There is 0% strength that the keyword “Environmental information Thessaloniki” is being discussed in social media
  - The sentiment of the users is 1:1. With more detail, the sentiments of the users are: 4 positive, 16 neutral and 5 negative
  - There is 25% that individuals talk about the specific keyword repeatedly in social media
  - There is 28% range of influence by the users
  - There are 11 unique authors
  - The mentions for that keyword are from:



“Περιβαλλοντική πληροφόρηση”

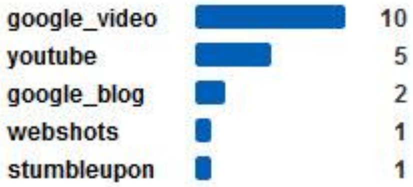
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 19 mentions concerning “Περιβαλλοντική πληροφόρηση”
  - The last mention was 15 days ago



0% strength 0:0 sentiment 3% reach 12% passion

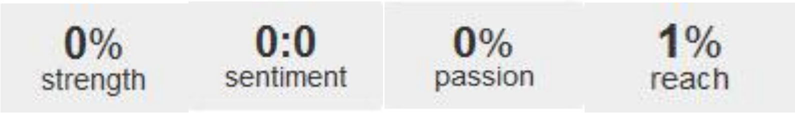
    - There is 0% strength that the keyword “Περιβαλλοντική πληροφόρηση” is being discussed in social media
    - The sentiment of the users is neutral for all mentions
    - There is 3% that individuals talk about the specific keyword repeatedly in social media
    - There is 12% range of influence by the users
  - There are 7 unique author
  - The mentions for that keyword are from:



“Περιβαλλοντική πληροφόρηση Θεσσαλονίκη”

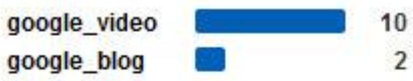
- Twitter:
  - Zero mentions in Twitter for that keyword

- All Social Media:
  - There have been 12 mentions concerning “Περιβαλλοντική πληροφόρηση Θεσσαλονίκη”
  - The last mention was 27 days ago



0% strength | 0:0 sentiment | 0% passion | 1% reach

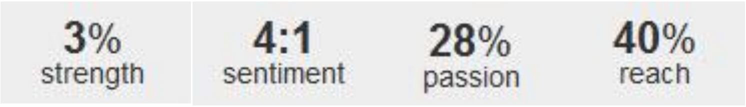
    - There is 0% strength that the keyword “Περιβαλλοντική πληροφόρηση Θεσσαλονίκη” is being discussed in social media
    - The sentiment of the users is neutral for all mentions
    - There is 0% that individuals talk about the specific keyword repeatedly in social media
    - There is 1% range of influence by the users
  - There are 1 unique author
  - The mentions for that keyword are from:



“Environmental informing Thessaloniki”

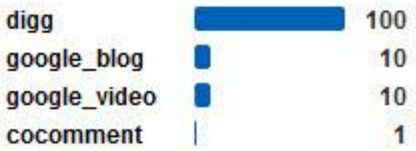
- Twitter:
    - Zero mentions in Twitter for that keyword

- All Social Media:
    - There have been 121 mentions concerning “Environmental informing Thessaloniki”
    - The last mention was 10 hours ago



3% strength 4:1 sentiment 28% passion 40% reach

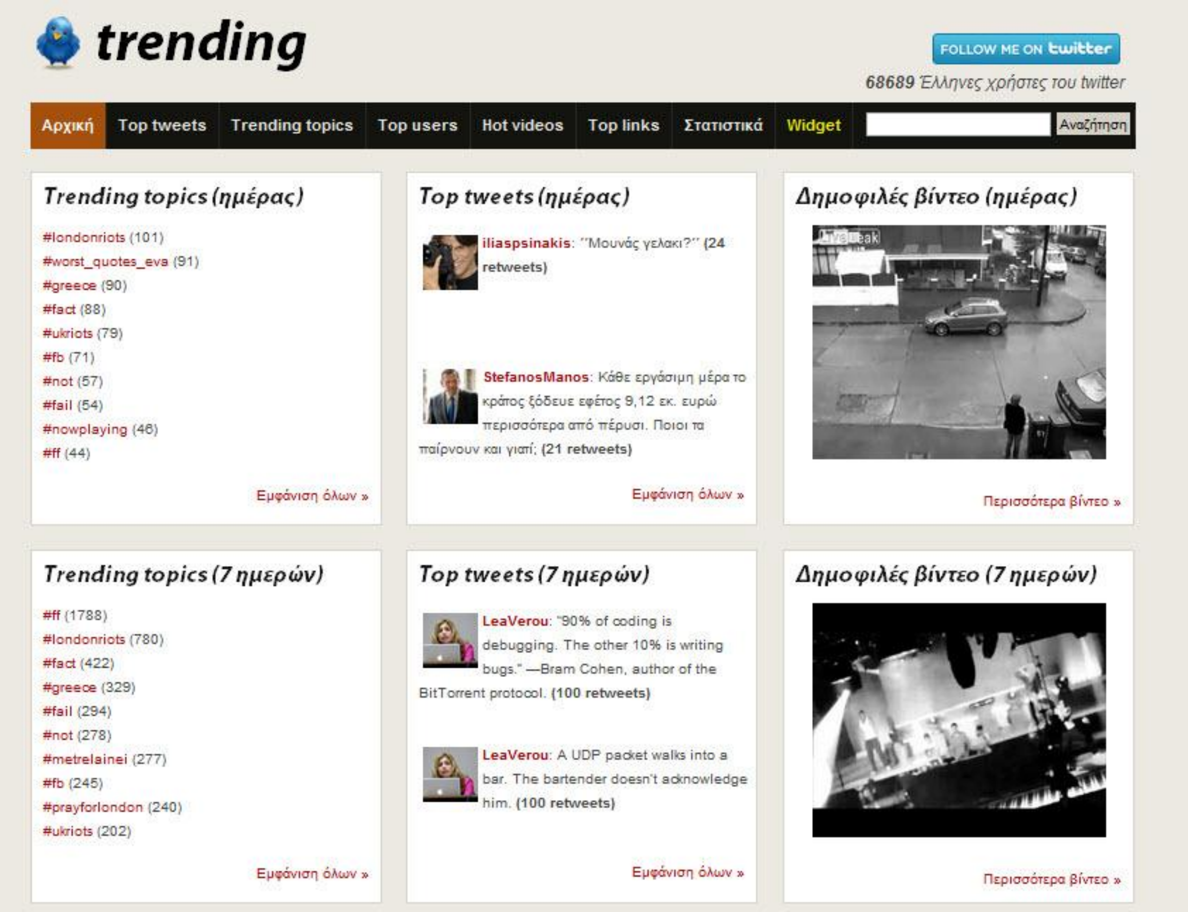
-
    -
        - There is 3% strength that the keyword “Environmental informing Thessaloniki” is being discussed in social media
        - The sentiment of the users is 4:1. With more detail, the sentiments of the users are: 62 positive, 45 neutral and 14 negative
        - There is 28% that individuals talk about the specific keyword repeatedly in social media
        - There is 40% range of influence by the users
    - There are 79 unique author
    - The mentions for that keyword are from:



Gathering the results of the query searches, interesting information came to light. First of all, it is obvious that not a single tweet is returned for all the queries. That means that Social Mention does not have good database of Twitter. We can also see that not a few mentions are being posted about environmental matters and issues. There are more than enough authors that look these issues with serious minds and are ready to act in order to problem solving. In the following chapter we will try to identify the most useful information out of them.

7.1.2 Trending

Trending.gr is another successful Greek web site that analyzes on a daily basis the tweets of all Greek people and records interesting statistics like the most re-tweeted tweets, trending topics, top users, hot videos and top links. It is the only web site - text mining tool available in the internet which analyzes and collects the tweets of the Greek users.



68689 Έλληνες χρήστες του twitter

Trending topics (ημέρας)

#londonriots (101)
#worst_quotes_eva (91)
#greece (90)
#fact (88)
#ukriots (79)
#fb (71)
#not (57)
#fail (54)
#nowplaying (46)
#ff (44)

Εμφάνιση όλων »

Top tweets (ημέρας)

iliaspsinakis: "Μουνάς γελακι?" (24 retweets)

StefanosManos: Κάθε εργάσιμη μέρα το κράτος ξόδευε εφέτος 9,12 εκ. ευρώ περισσότερα από πέρυσι. Ποιοι τα παίρνουν και γιατί; (21 retweets)

Εμφάνιση όλων »

Δημοφιλές βίντεο (ημέρας)

Περισσότερα βίντεο »

Trending topics (7 ημερών)

#ff (1788)
#londonriots (780)
#fact (422)
#greece (329)
#fail (294)
#not (278)
#metrelainei (277)
#fb (245)
#prayforlondon (240)
#ukriots (202)

Εμφάνιση όλων »

Top tweets (7 ημερών)

LeaVerou: "90% of coding is debugging. The other 10% is writing bugs." —Bram Cohen, author of the BitTorrent protocol. (100 retweets)

LeaVerou: A UDP packet walks into a bar. The bartender doesn't acknowledge him. (100 retweets)

Εμφάνιση όλων »

Δημοφιλές βίντεο (7 ημερών)

Περισσότερα βίντεο »

Figure 18: Trending.gr main page

This idea to monitor the daily tweets of Greek people came from a company named Sidebar. They started making a social monitoring tool and in the process discovered that the statistics generated by the twitter have enough interest, so they decided to publicize it. Besides the trending topics of Greek society, they embody the site with new ideas. The main site, as shown in the above figure (Figure 18), it has also the top tweets, top users, hot videos, top links etc. [30]

All the stats and search results are generated by the database that trending.gr has. In practice, the company that runs the database records almost all the tweets of Greek users. The process consists of two different processes running simultaneously and continuously during the day. The first process is the discovery of Greek users. The second process examines the Greek people that the site has already in the database, which at this point are 68520 (August 10, 2011), bringing their new tweets at regular intervals. [30]

Trending.gr is becoming nowadays a useful tool even for big companies like Vodafone or Cosmote who can identify their weaknesses or strengths through the tweets of Greek people. It is also a useful tool for quantifying the effectiveness of advertising and PR campaigns. Furthermore watching the trending topics / tweets / links, someone can possibly draw conclusions about new trends and interests in Greece. [30]

The same methodology is followed here as before with the previous internet tool. This web tool also allows the query search with Greek words followed by the word Thessaloniki. The results of each of the desirable keywords are the following:

“Ρύπανση”

- There are 778 mentions concerning “Ρύπανση”
- Most of the mentions took place between March and August 2011

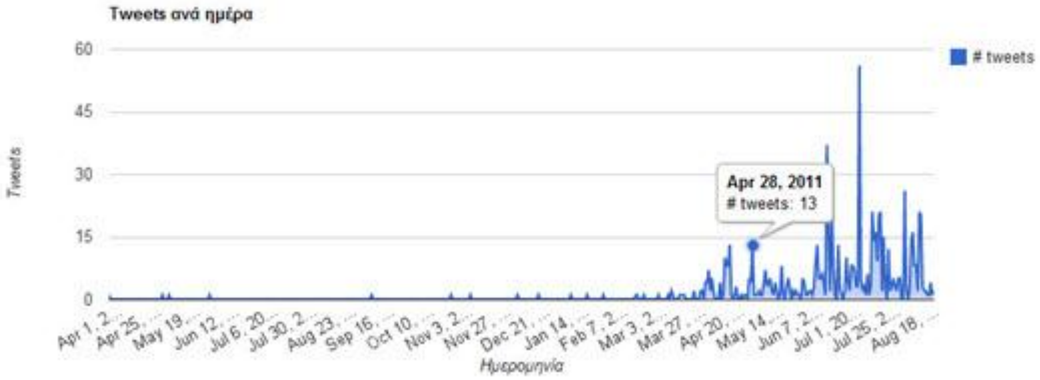


Figure 19: Results of the word “Ρύπανση” in Trending.gr

“Ρύπανση Θεσσαλονίκη”

- There are 7 mentions concerning “Ρύπανση Θεσσαλονίκη”
- Most of the mentions took place on May 2010, April and June 2011

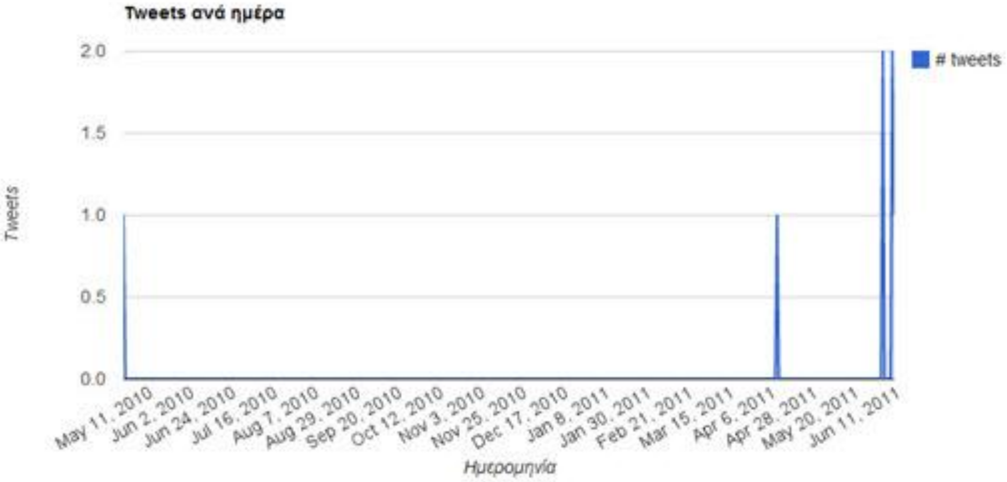


Figure 20: Results of the word “Ρύπανση Θεσσαλονίκη” in Trending.gr

“Pollution Thessaloniki”

- There are no mentions concerning “Pollution Thessaloniki”

“Ατμοσφαιρική ρύπανση”

- There are 3 mentions concerning “Ατμοσφαιρική ρύπανση”
- Most of the mentions took place between April and August 2011

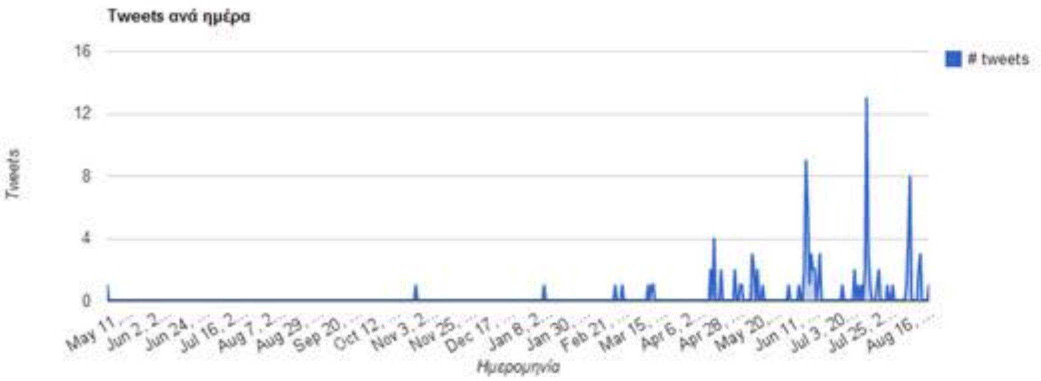


Figure 21: Results of the word “Ατμοσφαιρική ρύπανση” in Trending.gr

“Ατμοσφαιρική ρύπανση Θεσσαλονίκη”

- There are 3 mentions concerning “Ατμοσφαιρική ρύπανση Θεσσαλονίκη”
- Most of the mentions took place on May 2010 and June 2011

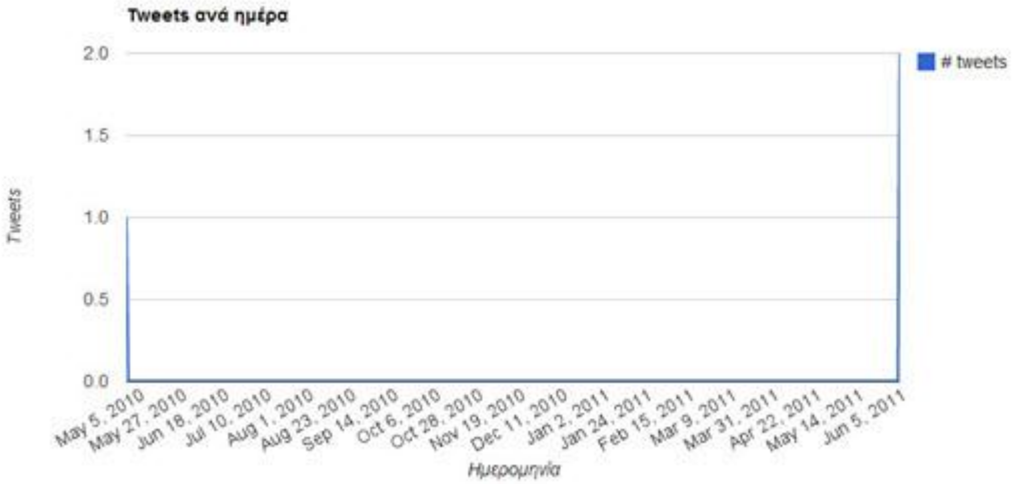


Figure 22: Results of the word “Ατμοσφαιρική ρύπανση Θεσσαλονίκη” in Trending.gr

“Air pollution Thessaloniki”

- There are no mentions concerning “Air pollution Thessaloniki”

“Ποιότητα αέρα”

- There are 13 mentions concerning about “Ποιότητα αέρα”
- Most of the mentions took place on July 2011

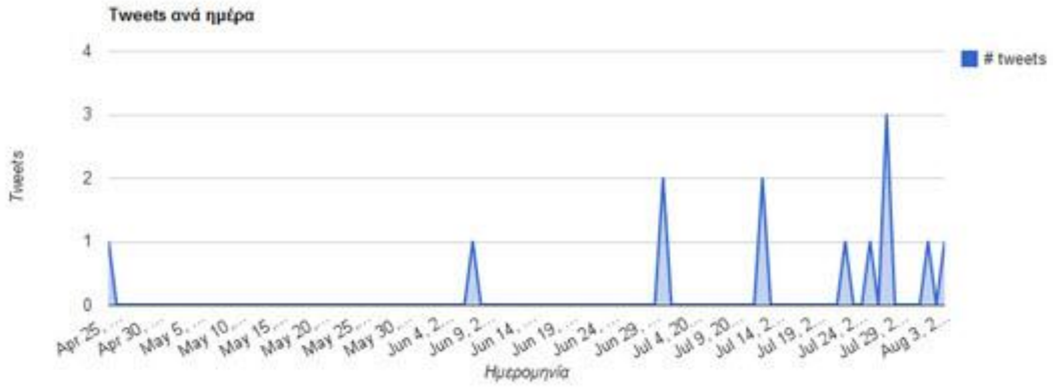


Figure 23: Results of the word “Ποιότητα αέρα” in Trending.gr

“Ποιότητα αέρα Θεσσαλονίκη”

- There are no mentions concerning “Ποιότητα αέρα Θεσσαλονίκη”

“Air Quality Thessaloniki”

- There is 1 mentions concerning “Air Quality Thessaloniki”
- The mention took place on January 2010

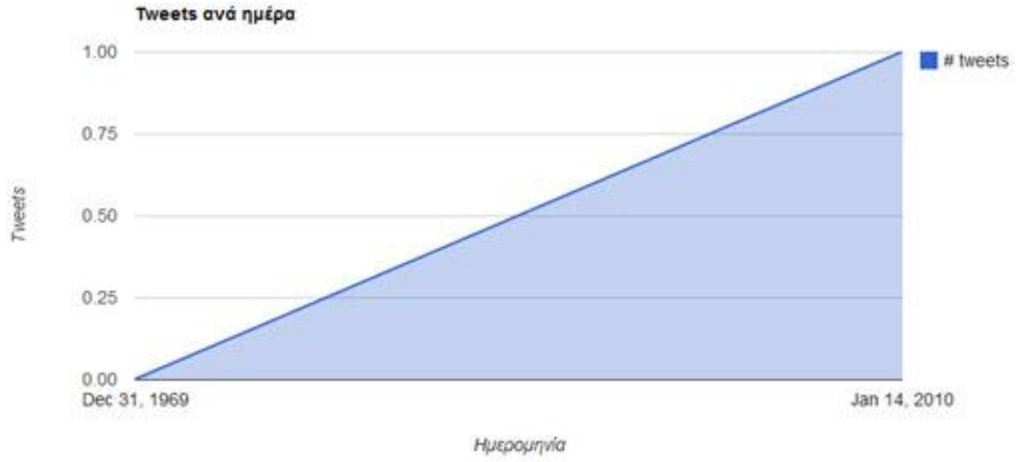


Figure 24: Results of the word “Air Quality Thessaloniki” in Trending.gr[1]

“Αιωρούμενα σωματίδια”

- There are 6 mentions concerning “Αιωρούμενα σωματίδια”
- Most of the mentions took place on June 2011

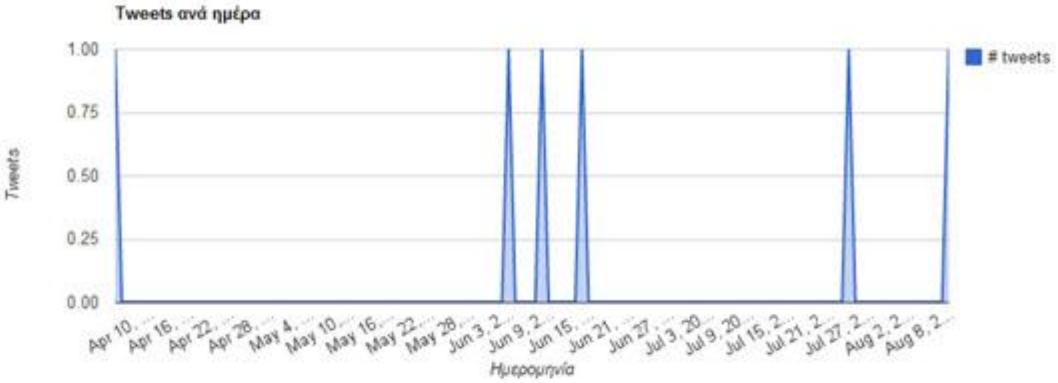


Figure 25: Results of the word “Αιωρούμενα σωματίδια” in Trending.gr

[1] There seems to be an error of Trending because it returns a tweet that starts from 1969 and ends in 2010.

“Αιωρούμενα σωματίδια Θεσσαλονίκη”

- There are no mentions concerning “Αιωρούμενα σωματίδια Θεσσαλονίκη”

“Particulates Thessaloniki”

- There are no mentions concerning “Particulates Thessaloniki”

“Απορρίμματα”

- There are 233 mentions concerning “Απορρίμματα”
- Most of the mentions took place between April and August 2011

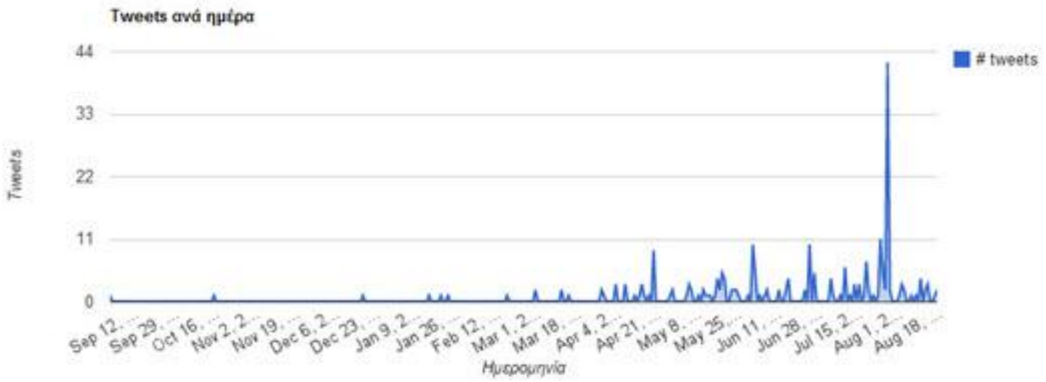


Figure 26: Results of the word “Απορρίμματα” in Trending.gr

“Απορρίμματα Θεσσαλονίκη”

- There are no mentions concerning “Απορρίμματα Θεσσαλονίκη”

“Waste Thessaloniki”

- There is 1 mentions concerning “Waste Thessaloniki”
- The mention took place on July 2010

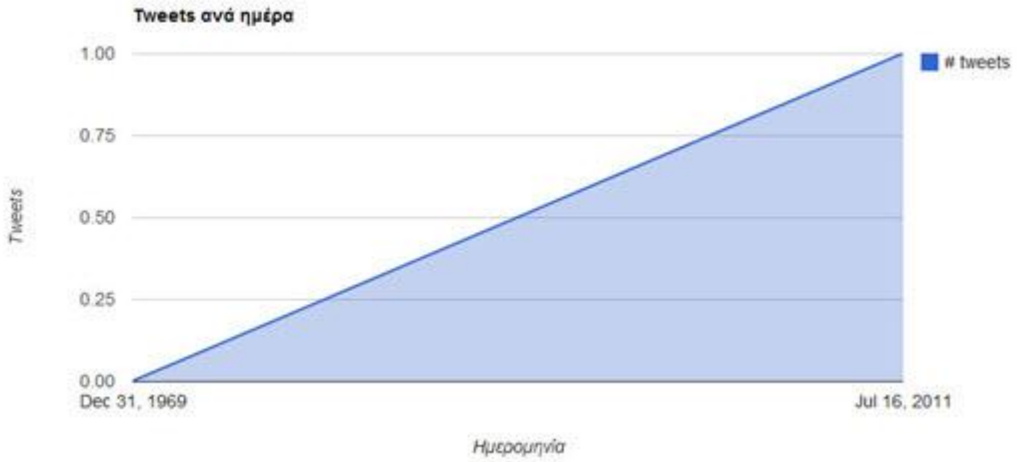


Figure 27: Results of the word “Waste Thessaloniki” in Trending.gr[2]

“Σκουπίδια”

- There are 4763 mentions concerning “Σκουπίδια”
- Most of the mentions took place between April and August 2011

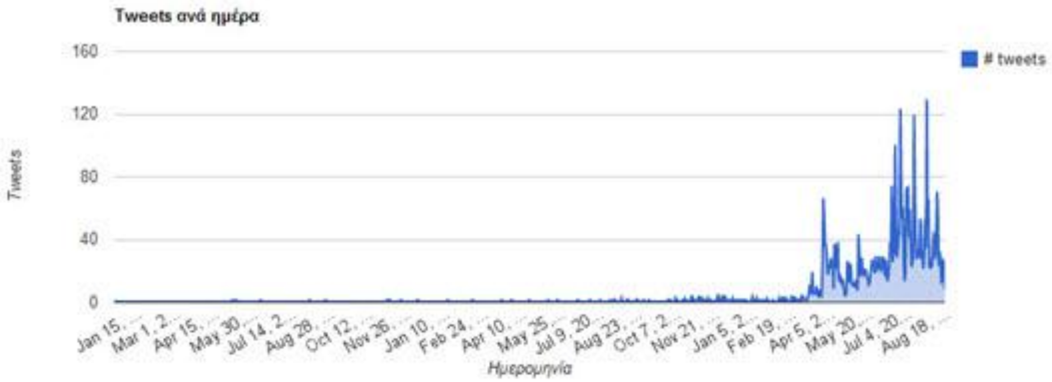


Figure 28: Results of the word “Σκουπίδια” in Trending.gr

[2] There seems to be the same error as before because it returns a tweet that starts from 1969 and ends in 2010

“Σκουπίδια Θεσσαλονίκη”

- There are 194 mentions concerning “Σκουπίδια Θεσσαλονίκη”
- Most of the mentions took place between March and April 2011 and June and July of the same year

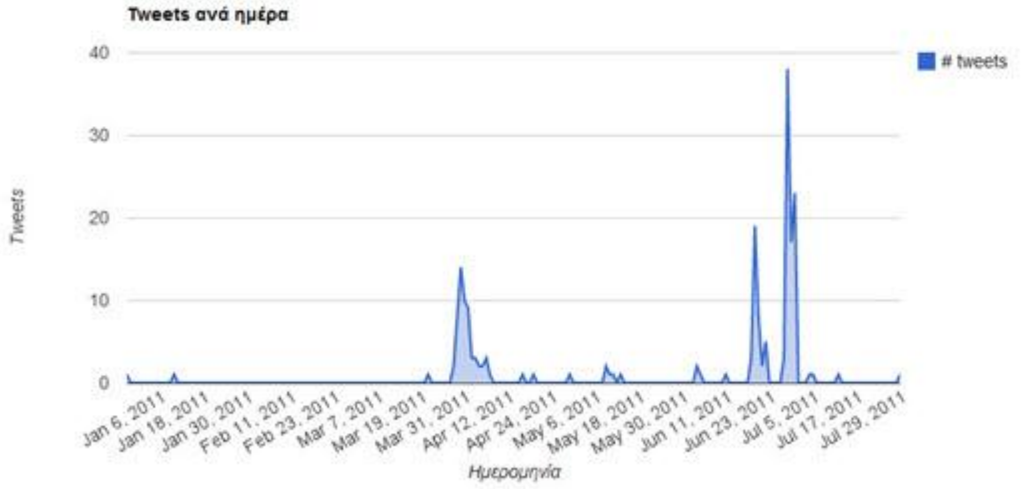


Figure 29: Results of the word “Σκουπίδια Θεσσαλονίκη” in Trending.gr

“Garbage Thessaloniki”

- There are 2 mentions concerning “Garbage Thessaloniki”
- The mentions took place on April 2011 and May of the same year

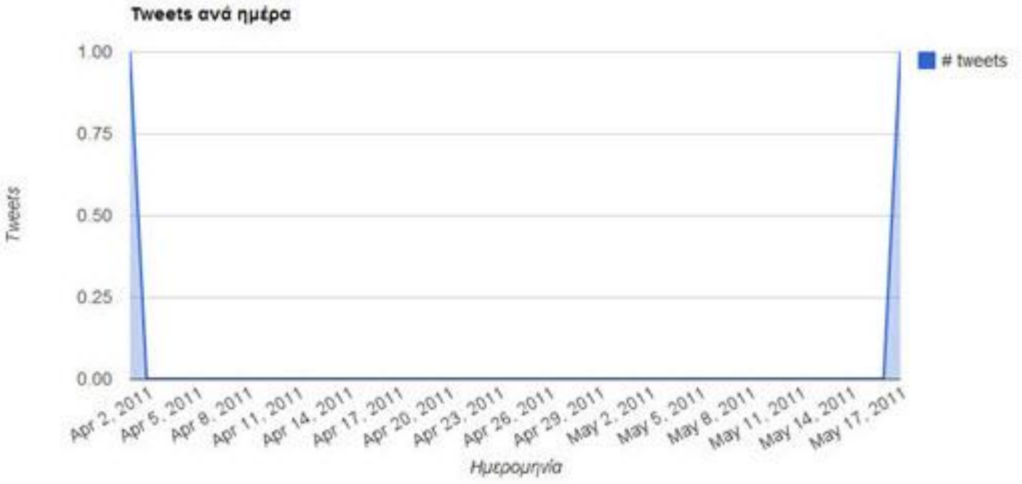


Figure 30: Results of the word “Garbage Thessaloniki” in Trending.gr

“Συγκοινωνιακό”

- There are 8 mentions concerning “Συγκοινωνιακό”
- Most of the mentions took place on April 2011

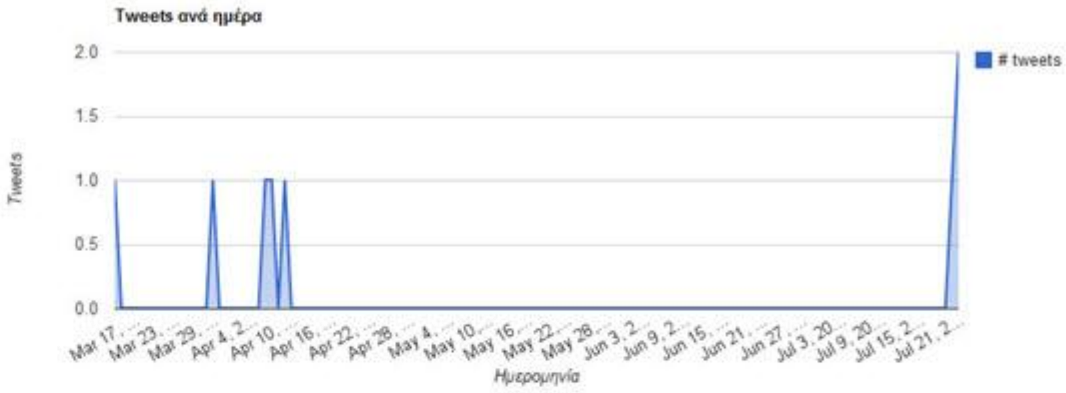


Figure 31: Results of the word “Συγκοινωνιακό” in Trending.gr

“Συγκοινωνιακό Θεσσαλονίκη”

- There are no mentions concerning “Συγκοινωνιακό Θεσσαλονίκη”

“Transportation Thessaloniki”

- There are 7 mentions concerning “Transportation Thessaloniki”
- The mentions took place on May and July 2011

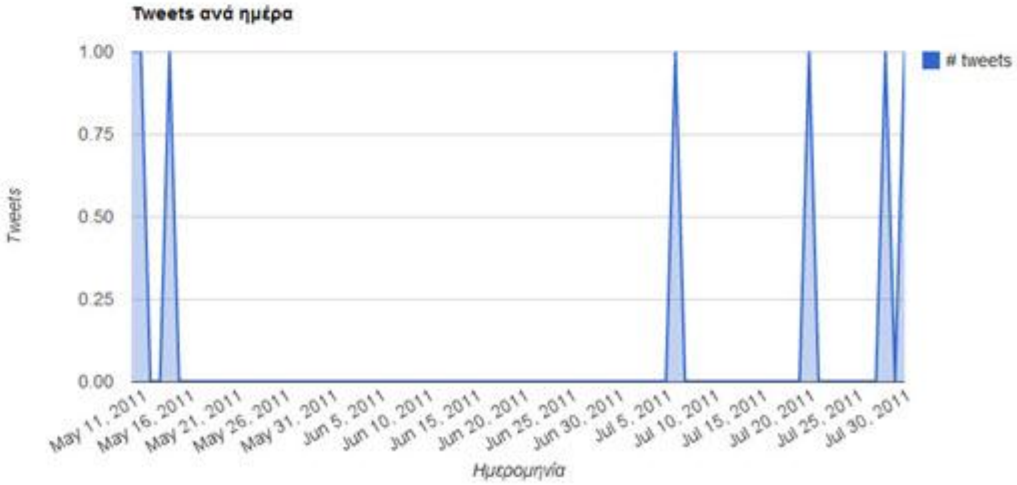


Figure 32: Results of the word “Transportation Thessaloniki” in Trending.gr

“Θόρυβος”

- There are 423 mentions concerning “Θόρυβος”
- Most of the mentions took place between April and August 2011

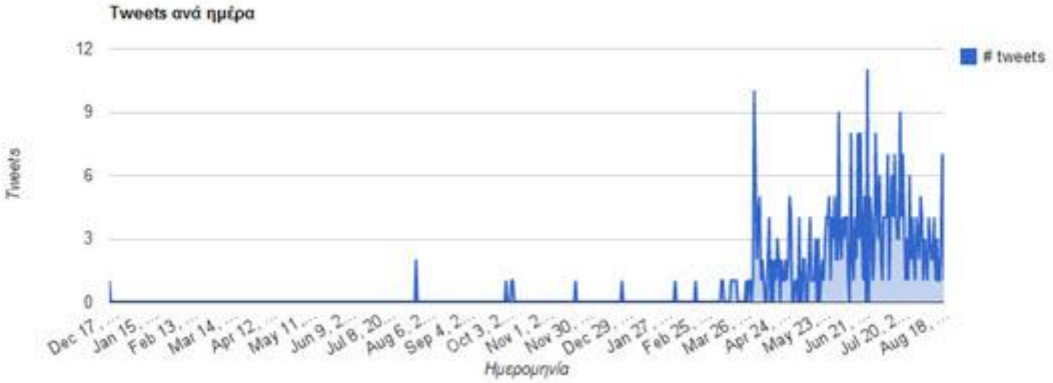


Figure 33: Results of the word “Θόρυβος” in Trending.gr

“Θόρυβος Θεσσαλονίκη”

- There are 3 mentions concerning “Θόρυβος Θεσσαλονίκη”
- The mentions took place on October 2010 and August 2011

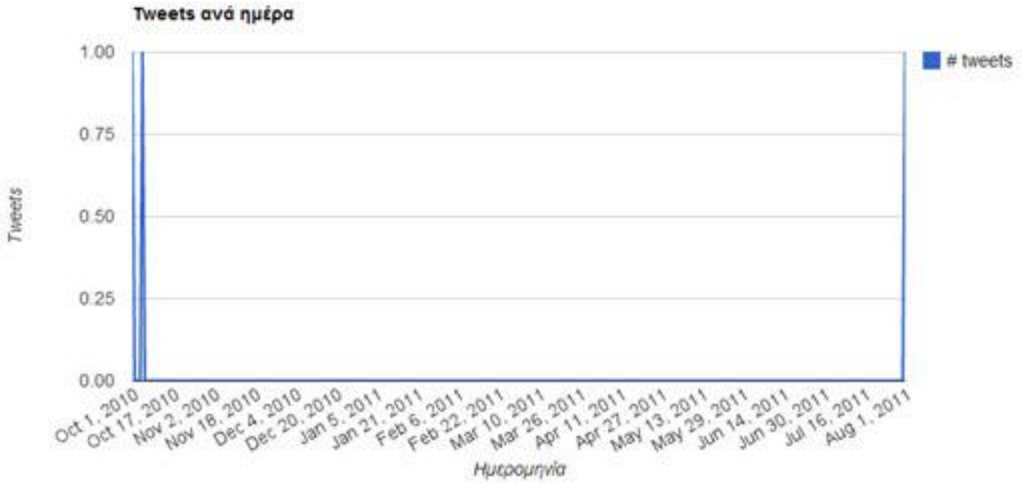


Figure 34: Results of the word “Θόρυβος Θεσσαλονίκη” in Trending.gr

“Noise Thessaloniki”

- There are 8 mentions concerning “Noise Thessaloniki”
- Most of the mentions took place between May and July 2011

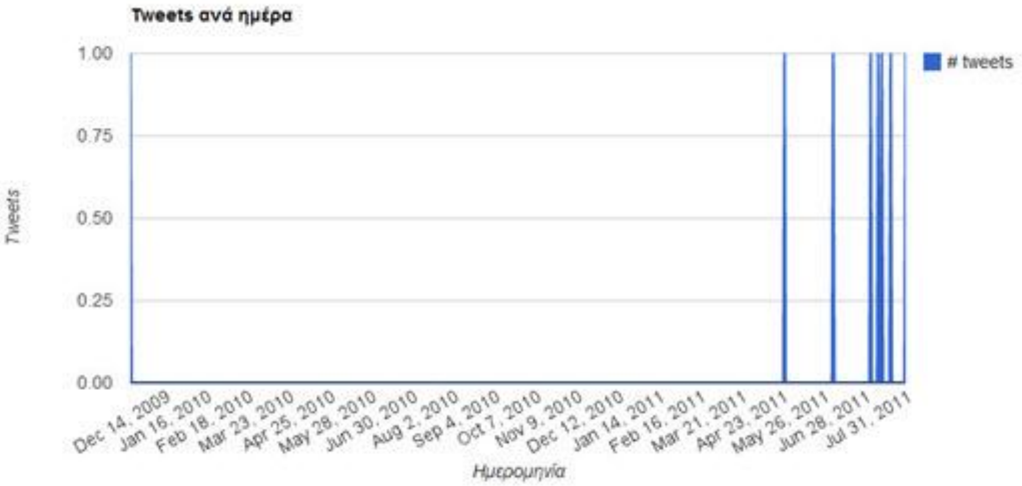


Figure 35: Results of the word “Noise Thessaloniki” in Trending.gr

“Ηχορύπανση”

- There are 189 mentions concerning “Ηχορύπανση”
- Most of the mentions took place between April and August 2011

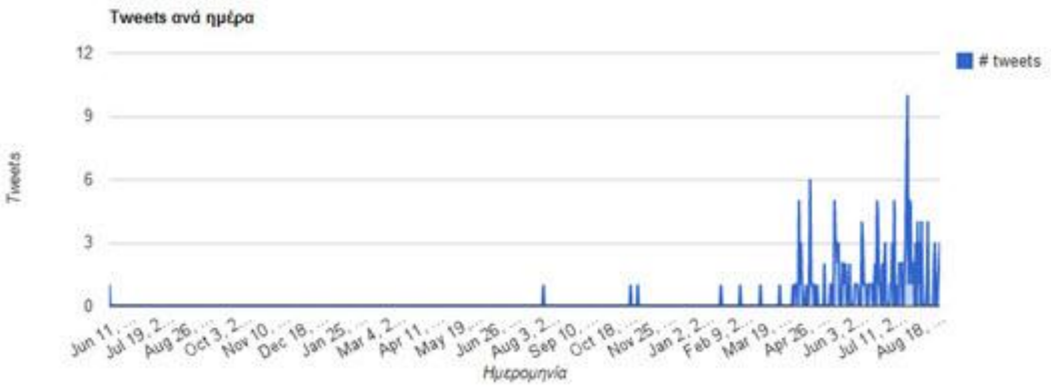


Figure 36: Results of the word “Ηχορύπανση” in Trending.gr

“Ηχορύπανση Θεσσαλονίκη”

- There are 5 mentions concerning “Ηχορύπανση Θεσσαλονίκη”
- The mentions took place on July 2011

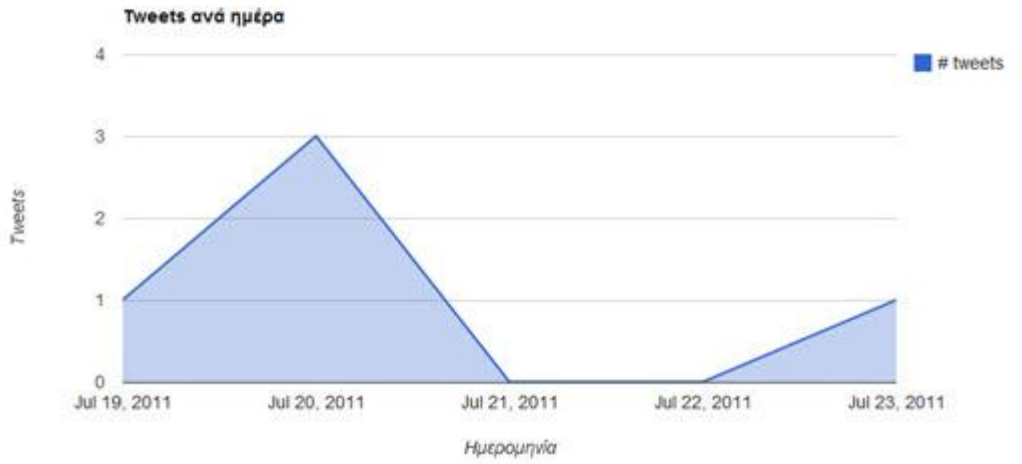


Figure 37: Results of the word “Ηχορύπανση Θεσσαλονίκη” in Trending.gr

“Noise pollution Thessaloniki”

- There are no mentions concerning “Συγκοινωνιακό Θεσσαλονίκη”

“Περιβαλλοντική πληροφορία”

- There is 1 mention concerning “Περιβαλλοντική πληροφορία”
- The mention took place on August 2011

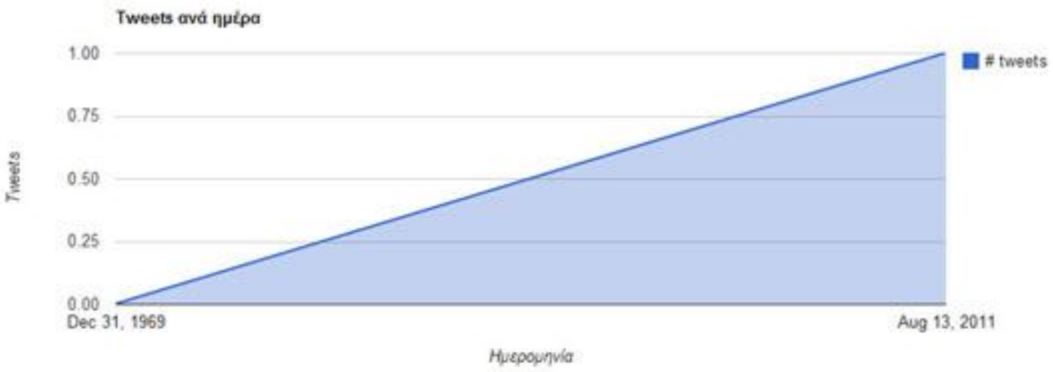


Figure 38: Results of the word “Περιβαλλοντική πληροφορία” in Trending.gr[3]

“Περιβαλλοντική πληροφορία Θεσσαλονίκη”

- There are no mentions concerning “Περιβαλλοντική πληροφορία Θεσσαλονίκη”

“Environmental information Thessaloniki”

- There are no mentions concerning “Environmental information Thessaloniki”

[3] There seems to be the same error as before because it returns a tweet that starts from 1969 and ends in 2010

“Περιβαλλοντική πληροφόρηση”

- There are no mentions concerning “Περιβαλλοντική πληροφόρηση”

“Περιβαλλοντική πληροφόρηση Θεσσαλονίκη”

- There are no mentions concerning “Περιβαλλοντική πληροφόρηση Θεσσαλονίκη”

“Environmental informing Thessaloniki”

- There are no mentions concerning “Environmental informing Thessaloniki”

Interesting information and knowledge about the number of posts as well as the time line of each post is given by Trending. We can identify that for some keywords we have no posts at all. Most of the keywords that don’t have posts are in the English language, and that make sense because Trending gathers information and tweets only from Greek users. Once again, it is obvious that the database of that tool is not fully updated. In the following chapter, that raw information will be gathered in a table and we will try to extract useful information out of them.

8 Discussions of results

All text mining tools, which have been used for the attempt to discover useful information from social media and from Twitter, have returned some very interesting and also important results which have to do with the environment and especially with the city of Thessaloniki. The following tables show each of the desirable keywords that have been used and also useful information that has been extracted with the use of the aforementioned text mining tools.

First of all, we need to separate each keyword and create categories with same characteristics. The categories for that purpose are: Greek keywords without Thessaloniki, Greek keywords with Thessaloniki and keywords with Thessaloniki. By merging same characteristics, it is easier to extract more information and then to become knowledge.

8.1 Social Mention

Social Mention can be characterized as a successful text mining internet tool because it has the ability to discover information from all social media at once. For that reason, it is important the information that has been returned after the query searches to be categorized into videos, pictures and text (short messages, blogs etc.). The number of authors is another important aspect that needs attention. Also important is to identify the time line of information in order to see the first and last indication of the each specific keyword. All the derived information is shown in the following table.

Table 1: Social Mention Results

Keywords	*Results*	*Authors*	*Type of information*	*Time Line of information*
Ρύπανση	213	102	Videos: 28% Pictures: 23% Text: 49%	August 2011 – 2006
Ρύπανση Θεσσαλονίκη	36	17	Videos: 61% Pictures: 8% Text: 31%	June 2011 – 2006

Pollution Θεσσαλονίκη	89	38	Videos: 51% Pictures: 2% Text: 47%	August 2011 – 2007
Ατμοσφαιρική ρύπανση	128	66	Videos: 30% Pictures: 6% Text: 64%	August 2011 – 2007
Ατμοσφαιρική ρύπανση Θεσσαλονίκη	18	8	Videos: 61% Text: 39%	June 2011 – 2007
Air pollution Θεσσαλονίκη	55	28	Videos: 60% Pictures: 2% Text: 38%	August 2011 – 2008
Ποιότητα αέρα	90	64	Videos: 67% Pictures: 9% Text: 24%	August 2011 – 2007
Ποιότητα αέρα Θεσσαλονίκη	24	13	Videos: 46% Pictures: 12% Text: 42%	June 2011 – 2007
Air Quality Θεσσαλονίκη	51	22	Videos: 57% Text: 43%	August 2011 – 2008
Αιωρούμενα σωματίδια	42	25	Videos: 26% Pictures: 2% Text: 72%	July 2011– 2007
Αιωρούμενα σωματίδια Θεσσαλονίκη	19	7	Videos: 53% Pictures: 5% Text: 42%	June 2011 – 2008
Particulates Θεσσαλονίκη	21	10	Videos: 52% Text: 48%	August 2011 – 2005
Απορρίμματα	217	82	Videos: 28% Pictures: 18% Text: 54%	August 2011 – 2006
Απορρίμματα Θεσσαλονίκη	32	17	Videos: 53% Pictures: 16% Text: 31%	July 2011– 2008
Waste Θεσσαλονίκη	78	47	Videos: 65%	August 2011 – 2007

			Pictures: 13% Text: 22%	
Σκουπίδια	268	108	Videos: 22% Pictures: 22% Text: 56%	August 2011 – 2007
Σκουπίδια Θεσσαλονίκη	131	65	Videos: 30% Pictures: 5% Text: 65%	July 2011– 2008
Garbage Θεσσαλονίκη	81	45	Videos: 42% Pictures: 38% Text: 20%	August 2011 – 2007
Συγκοινωνιακό	68	37	Videos: 29% Pictures: 18% Text: 53%	August 2011 – 2007
Συγκοινωνιακό Θεσσαλονίκη	46	29	Videos: 78% Text: 22%	August 2011 – 2007
Transportation Θεσσαλονίκη	131	70	Videos: 41% Pictures: 34% Text: 25%	August 2011 – 2006
Θόρυβος	159	86	Videos: 30% Pictures: 7% Text: 63%	August 2011 – 2006
Θόρυβος Θεσσαλονίκη	31	21	Videos: 58% Pictures: 10% Text: 32%	August 2011 – 2007
Noise Θεσσαλονίκη	97	43	Videos: 62% Pictures: 19% Text: 19%	August 2011 – 2006
Ηχορύπανση	159	84	Videos: 35% Pictures: 5% Text: 60%	August 2011 – 2006
Ηχορύπανση Θεσσαλονίκη	25	11	Videos: 44% Pictures: 8% Text: 48%	August 2011 – 2006

Noise pollution Θεσσαλονίκη	22	7	Videos: 50% Text: 50%	August 2011 – 2006
Περιβαλλοντική πληροφορία	28	17	Videos: 61% Pictures: 3% Text: 36%	July 2011 – 2009
Περιβαλλοντική πληροφορία Θεσσαλονίκη	20	10	Videos: 50% Text: 50%	August 2011 – 2008
Environmental information Θεσσαλονίκη	25	11	Videos: 48% Text: 52%	August 2011 - 2005
Περιβαλλοντική πληροφόρηση	19	7	Videos: 79% Pictures: 5% Text: 16%	August 2011 – 2009
Περιβαλλοντική πληροφόρηση Θεσσαλονίκη	12	1	Videos: 83% Text: 17%	July 2011 – 2009
Environmental informing Θεσσαλονίκη	121	79	Videos: 8% Text: 92%	August 2011 – 2007

8.1.1 Type of information

Since we deal with all social media with this particular text mining tool, we have to identify at which point the information we get is video, picture or just text. During the query search for all desirable keywords, the data that Social Mention has returned were pretty interesting.

As discussed before we will seek knowledge according to each category. The first category is Greek keywords, “Thessaloniki” being excluded. We can see in the following figure that when the query has to do with environmental issues without the word Thessaloniki, half of the results are in the form of text. Followingly, videos are in the second highest level in the list, while pictures are the least frequent form of information used. We can understand from this figure, that when someone is referring to an environmental issue in general without any geographical specification, most of the times they do that by text (50%) by posting a short message or write on a blog. There is of course a high percentage of video posts (39%) as well, and only 11% of the results are pictures.

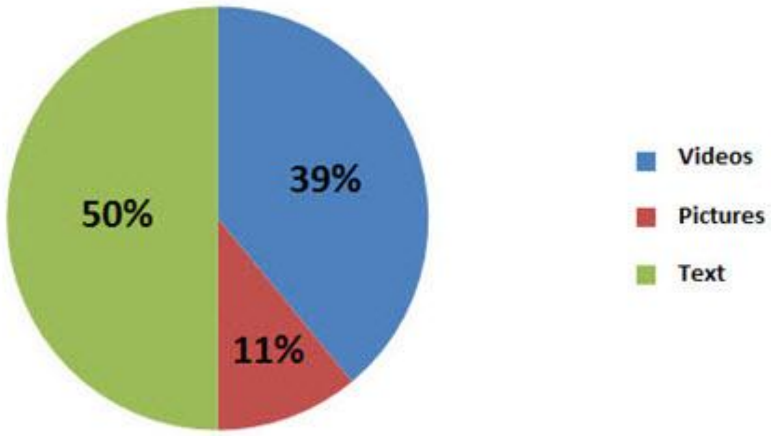


Figure 39: Greek Keywords percentage without Thessaloniki

Interesting information is extracted from the next category which is the Greek keywords, now including Thessaloniki. In the following figure, it is clearly that the highest percentage in results takes the videos and then the text and pictures. We can see that when someone mentions the word Thessaloniki in his post with environmental matter, most of the results that are returned are in the form of videos (56%). Great percentage takes the text results (38%) followed by the pictures with only 6% of the times.

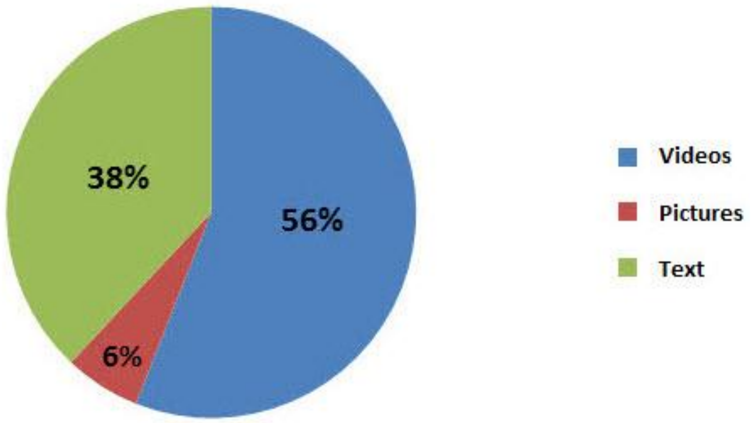


Figure 40: Greek Keywords percentage with Thessaloniki

Last but not least is the category of the results of the keywords in English with the word Thessaloniki. In this category, again the video posts are more in numbers (49%), but the text results are not so far away (41%). Once again, the pictures have the last word in a query search with only 10%.

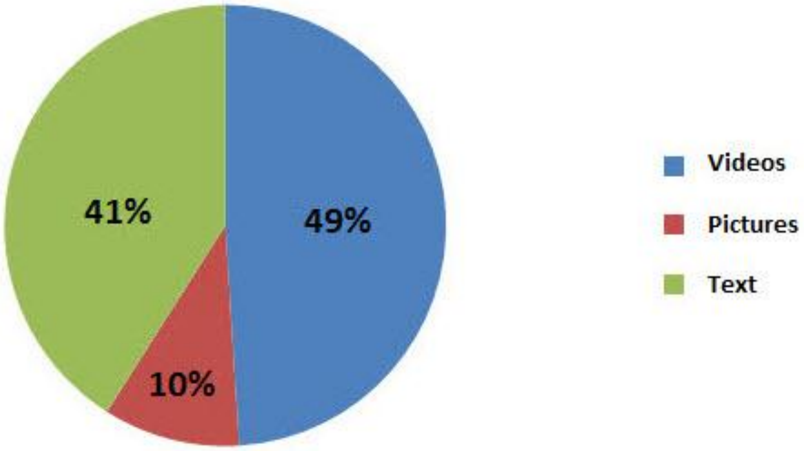


Figure 41: Keywords percentage with Thessaloniki

According to the above categorization, another important result came up and that is the relation of the posts retrieved with the type of information regarding to each post. As it has been mentioned before, posts that have been found in social media ("Soft" sensor data or citizens observations) are divided either to text or video. It is interesting to see that the types of posts in relation to environmental issues are the following:

- *Video relating posts:* pollution, air quality, waste, transportation and noise
- *Text relating posts:* air pollution, particulates, garbage and noise pollution
- *Picture relating posts:* waste and garbage

The above information reveals that most of the problems are preferably mentioned through videos and secondly through a short message or a post. In fact, when there is a reference for Thessaloniki, video type of information is more commonly used. The only time that people record into pictures an environmental issue, is when they are referring to waste or garbage matters. And that is clearly understandable because waste and garbage create visual annoyance.

8.1.2 Authors

It is important to consider the number of unique authors that dedicate time in order to post a video, picture or text and refer to an environmental matter. Based on the above categories and the above figure, we are in position to say that in the first category (Greek keywords without Thessaloniki) there were in average 62 unique authors that referred to environmental issues in general, in the second category (Greek keywords with Thessaloniki) there were in average 18 unique authors that referred specifically in the town of Thessaloniki, and last but not least in the third category (keywords with Thessaloniki) there were in average 36 unique authors that referred to the town of Thessaloniki in the English language. That means that most of the authors refer to environmental issues in general and few of them they refer to the town of the Thessaloniki.

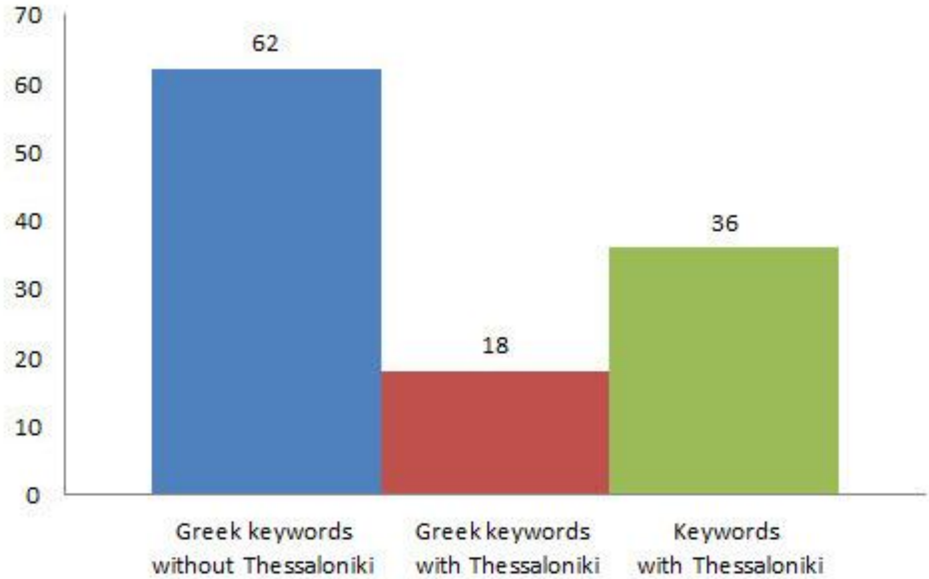


Figure 42: Average number of unique authors per category

8.1.3 Sentiment

Sentiment analysis is another important aspect of text mining and Social Mention offers that. When a query search is finished and the results are returned, sentiment information is mentioned on the side also. Because there are many keywords and many sentiments for each of them, there was no point to check each one individually. So, a total count of the sentiments was made in order to get a clearer image of what is the feeling and the opinions of the users that post to social media. As we can observe in the following graphical representation, posts with positive sentiment are 200, with neutral sentiment are 2232 and with negative sentiment are 116. It is obvious that 88 % of the posts have neutral sentiment, which means that most users don't provide with any positive or negative view in their participatory environmental observations, as reported via social networks. That could mean two possible things: either that sentiment analysis is not accurate enough (not adequate/proper sentiment analysis dictionary) or second there is no enough interest in environmental issues. It is interesting also to note that the positive sentiments are more than the negative ones. There is also the fact that someone can mock this kind of things by adding positives features to his posts.

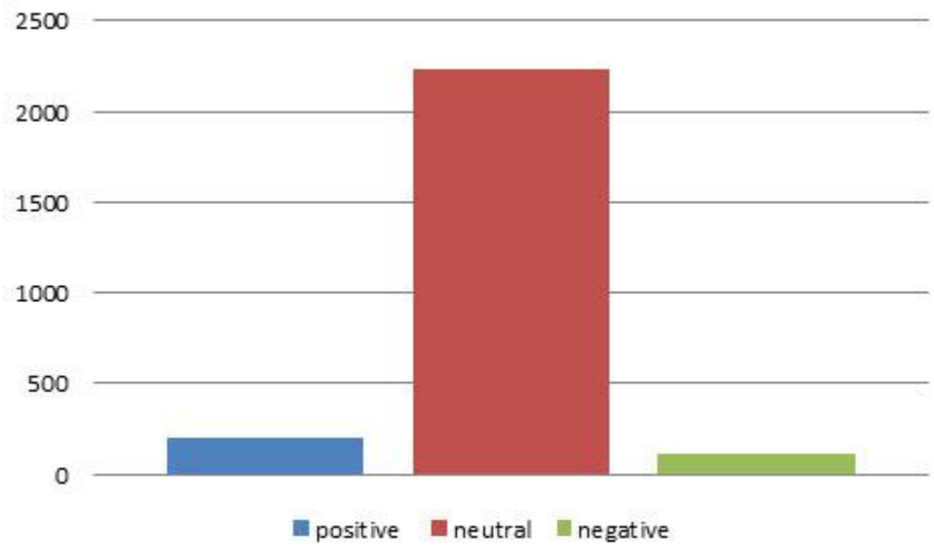


Figure 43: Number of sentiments

8.2 Trending

Trending is one of the few Greek web sites that deal with Twitter messages originating from Greece only. Since we live in Greece, it is important to extract information from Greek accounts in order to identify the environmental needs in Greece generally and in the town of Thessaloniki specifically. Because we deal at this point only with Twitter, Trending gives the opportunity to study the number of results derived from the query search followed by the time line of that information. That information is shown in the following table.

Table 2: Trending Results

Keywords	*Results*	*Time Line of information*
Ρύπανση	778	August 2011 – April 2010
Ρύπανση Θεσσαλονίκη	7	June 2011 – May 2010
Pollution Θεσσαλονίκη	-	-
Ατμοσφαιρική ρύπανση	3	August 2011 – May 2010
Ατμοσφαιρική ρύπανση Θεσσαλονίκη	3	June 2011 – May 2010
Air pollution Θεσσαλονίκη	-	-
Ποιότητα αέρα	13	August 2011 – April 2011
Ποιότητα αέρα Θεσσαλονίκη	-	-
Air Quality Θεσσαλονίκη	1	January 2010
Αιωρούμενα σωματίδια	6	August 2011 – April 2011
Αιωρούμενα σωματίδια Θεσσαλονίκη	-	-
Particulates Θεσσαλονίκη	-	-
Απορρίμματα	233	August 2011 – September 2010
Απορρίμματα Θεσσαλονίκη	-	-
Waste Θεσσαλονίκη	1	July 2011

Σκουπίδια	4763	August 2011 – January 2010
Σκουπίδια Θεσσαλονίκη	194	August 2011 – January 2010
Garbage Θεσσαλονίκη	2	May 2011 – April 2011
Συγκοινωνιακό	8	July 2011 – March 2011
Συγκοινωνιακό Θεσσαλονίκη	-	-
Transportation Θεσσαλονίκη	7	July 2011 – May 2011
Θόρυβος	423	August 2011 – December 2009
Θόρυβος Θεσσαλονίκη	3	August 2011 – October 2010
Noise Θεσσαλονίκη	8	July 2011 – December 2009
Ηχορύπανση	189	August 2011 – June 2009
Ηχορύπανση Θεσσαλονίκη	5	July 2011
Noise pollution Θεσσαλονίκη	-	-
Περιβαλλοντική πληροφορία	1	August 2011
Περιβαλλοντική πληροφορία Θεσσαλονίκη	-	-
Environmental information Θεσσαλονίκη	-	-
Περιβαλλοντική πληροφόρηση	-	-
Περιβαλλοντική πληροφόρηση Θεσσαλονίκη	-	-
Environmental informing Θεσσαλονίκη	-	-

8.2.1 Posts

One of the information that Trending gives is the number of post that have been made upon the query request. It is important to take a look on the average number of posts that have been made per category in order to identify the reference to environmental issues with our without the word Thessaloniki.

In the following graphical representation, the average number of results per category is shown. In the first category (Greek keywords without Thessaloniki) there were in average 583 posts that referred to environmental issues in general without the word Thessaloniki, in the second category (Greek keywords with Thessaloniki) there were in average 19 posts that referred specifically in the town of Thessaloniki, and in the third category (keywords with Thessaloniki) there were in average only 2 posts that referred to the town of Thessaloniki in the English language. This means that although the numbers of posts are large, there are few mentions about environmental issues for the town of Thessaloniki either in the Greek or the English language.

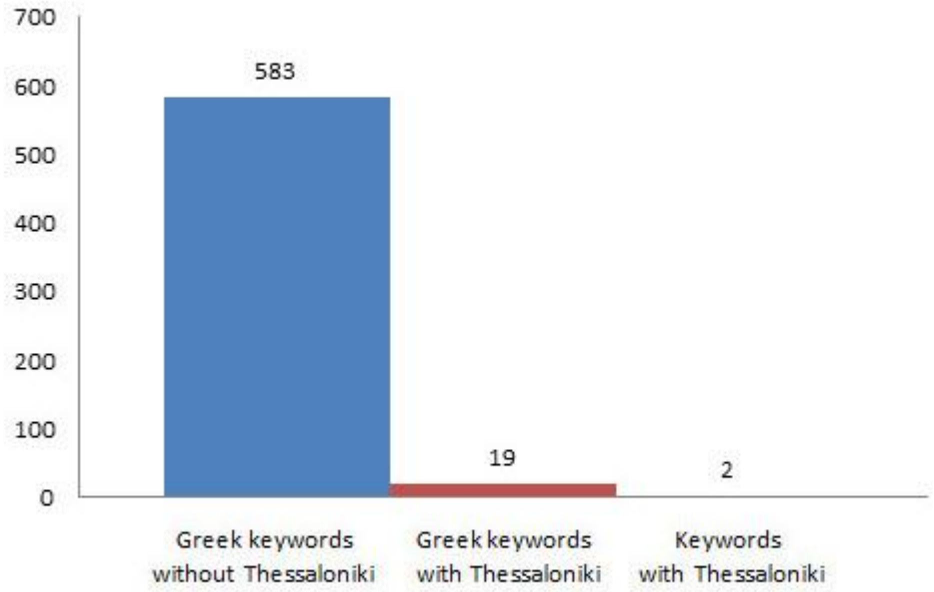


Figure 44: Average number of results per category

8.2.2 Time Line

The next and last information that Trending returns has to do with the attribute of time. In every query search a chart of time and posts appears giving us a very good opinion of what we are dealing with. In order to find out useful information of each of the queries, all posts had to be counted individually for each month of 2011. It is obvious that it is impossible to count all posts individually, so the following graphical representation shows most of them. The point remains the same though.

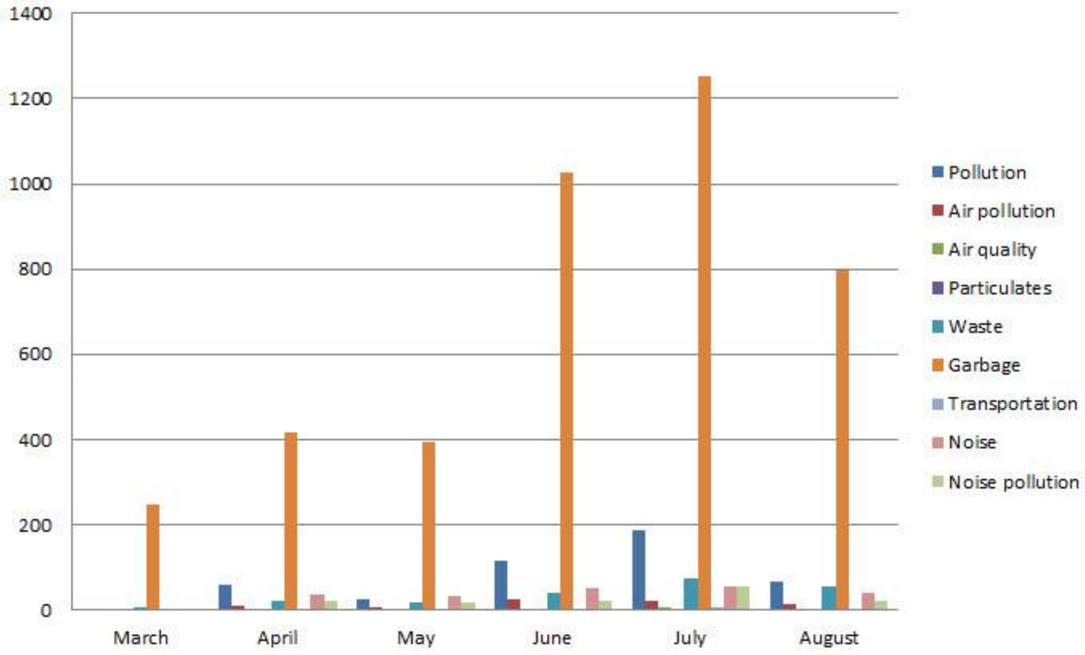


Figure 45: Trending posts per month of 2011

From the above graphical representation, we can see that most of the environmental matters concentrate in the months of summer (June, July and August). The largest environmental issue is the garbage issue by far. That may be interpreted as an indication that Thessaloniki has a problem with garbage and waste management most of the months of the year.

The mayor of Thessaloniki, Yiannis Boutaris, is aware of the problem with garbage and waste management of the town and he said that 4,500 recycling bins would be set up by the summer to encourage city dwellers to recycle household waste and reduce the amount of trash going to landfills. In addition to that another 15 new garbage collection trucks will be added to the current fleet of 25. That is a hopeful start for a town with limited garbage and waste. [41]

There are a couple of reasons why all the above environmental issues happen during summer. First of all, a lot of pollution is being concentrated to cities during warm periods where dust, soot and fine particles are being aggregated in the city. Additionally, heat and sunlight reinforces photochemical reactions between primary air pollutants, forming ozone, which is toxic in the lower bounds of the atmosphere and aggravates the quality of the atmosphere. There is also the fact that a lot of people are active outdoors, being exposed more frequently to environmental problems.

8.3 Correlation of results based on surveys

There are quite a few surveys that study environmental issues based upon citizens' opinions. In order to correlate some of their information, we must highly the environmental issues that appear the most. The following graphical representation shows the most popular participatory sensing issues that appear general in Greece and in Thessaloniki combined, based on the number of unique tweets, You Tube posts, blog posts etc.

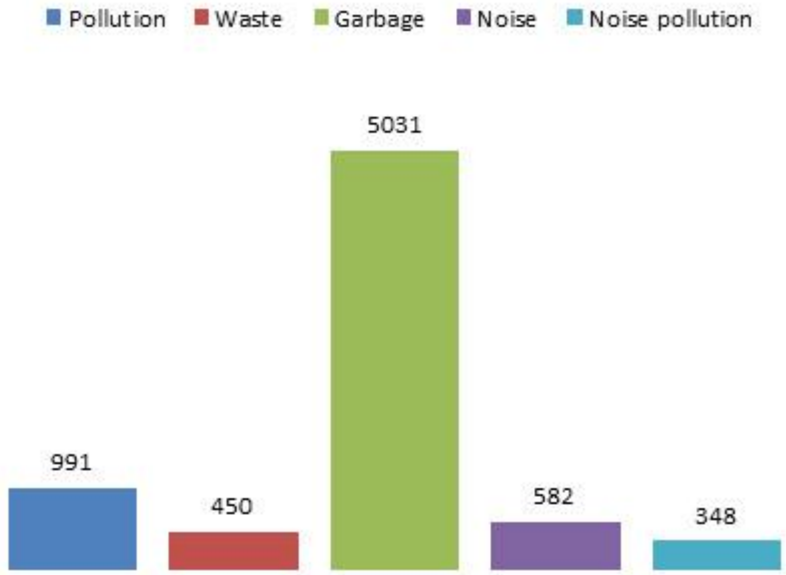


Figure 46: Most popular environmental issues

The findings presented in Fig. 46 correlate quite well with those of a survey that was conducted by the University of London on the environmental problems that citizens prioritize as high in their concerns about environmental quality. [42] We can thus argue that two cities like Thessaloniki and London, suffer almost by the same environmental issues. Garbage and waste management are the number one issue in both cases, and especially for Thessaloniki and Greece, as we can identify from the above graphical representation and through the analysis made in the current dissertation based on citizens' observations. In Greece and in Thessaloniki, pollution in general concerns the Greek society as well as noise pollution. Even though, the above environmental issues occur the most, that doesn't mean that there are no other issues that need our attention. [42]

In addition, the European Commission has released a report in 2007, in which interesting information about environmental matters have been found. A survey showed that in Greece the main environmental issues that the population worries about are the ones that are showed in the following graphical representation. We can identify that three out of five mentioned environmental issues are the same with the results that we discovered applying text mining in social media and networks. According to the same survey, over 80% of Greek people say that environmental problems have a direct effect on their daily life, and almost all Europeans consider the protection of the environment to be highly important. [43]

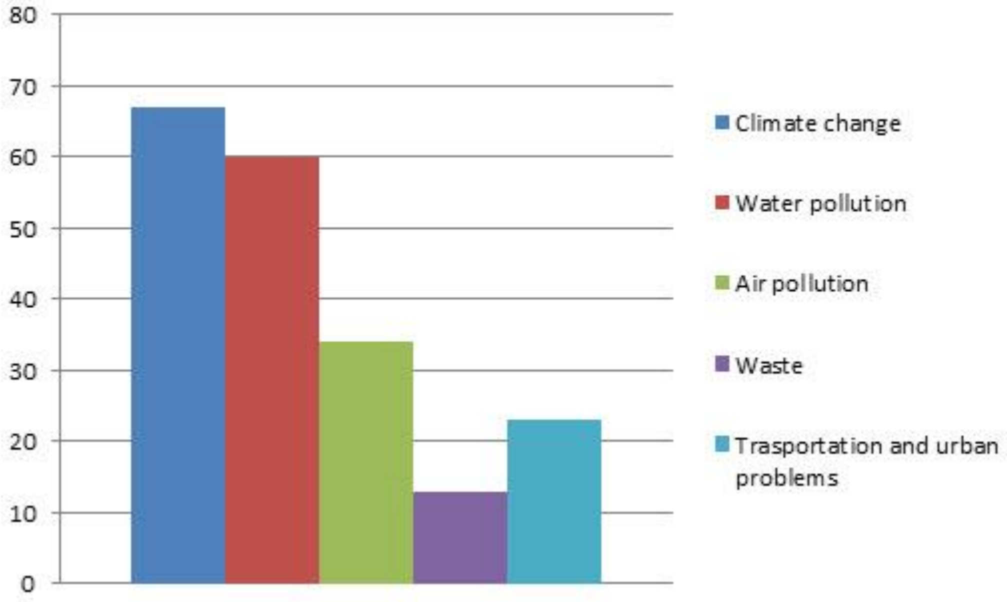


Figure 47: Percentage distribution of main environmental issues in Greece

9 Conclusions

As mobile phone technology evolves and social media and especially social networks keep rising, participatory sensing is growing in terms of concepts and applications. The environment needs everybody’s attention and with the right combination of technology tools, measures, disseminated information and of course with the participation of citizens, we can accomplish a better quality of life for everyone.

In the current dissertation, mining tools were employed in order to perform text mining in social media, with the aid to investigate the potential of using Participatory Sensing-based, unstructured information (“soft” sensors) for investigating environmental pressures and conditions. The main subject was the research of environmental issues for Greece and more specific for the town of Thessaloniki. The text mining tools that have been selected for the process were Social Mention and Trending, which are both web tools.

The dissertation’s aim was to investigate the web and especially social media for participatory sensing data in order to identify environmental issues that concern Greece and especially Thessaloniki. The research has been successfully contacted and the results that have been returned by the text mining tools outline that the most important environmental issues that people suffer from, are mainly the garbage and waste management, mostly during summer. The results also pointed that people tend to refer to environmental problems with neutral sentiment and only in some cases they refer either positively or negatively. The research showed last, that video and text are the most commonly used ways to express an opinion about environmental issues through social media.

With the right mining tools it’s easier to scan the web and identify environmental problems, in order to solve them accurately, effectively and on-time. Although the specific research was successful, few limitations have been found. First of all, text mining tools that are available on the web are limited in numbers. Also, there is not enough accuracy when it comes to text mining on the basis of text annotations/keywords. Finally, the semantic analysis of the research is not adequate because there is a lack of proper semantic dictionary. For that reasons, in the future, a text mining tool could be developed with more capabilities and fewer limitations, for people to get informed more frequently and with more accuracy about environmental issues for a more safe and healthy environment.

BIBLIOGRAPHY

[1] Jeffrey Goldman, Katie Shilton, Jeff Burke, Deborah Estrin, Mark Hansen, Nithya Ramanathan, Sasank Reddy, Vids Samanta, Mani Srivastava, and Ruth West. *Participatory Sensing: A citizen-powered approach to illuminating the patterns that shape our world*. 2009

[2] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, M. B. Srivastava. *Participatory Sensing,* Center for Embedded Networked Sensing (CENS), University of California, Los Angeles, 2006

[3] Urban Sensing CENS/UCLA. [online], (n.d.)
Available from: <http://urban.cens.ucla.edu/projects>

[4] Ellie D'Hondt, Matthias Stevens. *Empowering technologies for environmental participatory sensing.* BrusSense Team, Vrije Universiteit Brussel, Brussels, Belgium, (n.d.)

[5] Dr Eiman Kanjo. *Mobile Phones as Sensors*. University of Cambridge Horizon Seminar, Kaetsu Centre, New Hall, Cambridge, (n.d.)

[6] Akshay Dua, Nirupama Bulusu, Wu-chang Feng, Wen Hu. *Towards Trustworthy Participatory Sensing*. Portland State University, CSIRO ICT Centre, Australia, 2009

[7] Susan Ward, About.com Guide. *Social media definition*. [online], (n.d.),
Available from: <http://sbinfocanada.about.com/od/socialmedia/g/socialmedia.htm>

[8] Knowledge Networks. *Social Media Now Influences Brand Perceptions, Purchase Decisions of 38 Million in U.S.,* [online], 2011,
Available from: <http://www.businesswire.com/news/home/20110614005666/en/Social-Media-Influences-Brand-Perceptions-Purchase-Decisions>

[9] Antony Mayfield. *What is Social Media?* iCrossing. 2008

[10] Lauren Dugan. *More Than 1 Million People Per Week View Customer Service Tweets.* [online], 2011, Available from: <http://www.mediabistro.com/alltwitter/more-than-1-million-people-per-week-view-customer-service-tweets-study_b10173>

[11] Stephanie Reese. *Quick Stat: Facebook to Reach 132.5 Million Users in the US This Year.* [online], 2011, Available from: <http://www.emarketer.com/blog/index.php/quick-stat-facebook-reach-1325-million-users-year/>

[12] Keith Hampton, Lauren Sessions Goulet, Lee Rainie, Kristen Purcell. *Social networking sites and our lives.* [online], 2011,
Available from: <http://www.pewinternet.org/Reports/2011/Technology-and-social-networks/Summary.aspx?view=all>

[13] Azam Khan. *Average User Spends 9% More Time Using Mobile Apps Than The Internet. .* [online], 2011, Available from: <http://socialtimes.com/average-user-spends-9-more-time-using-mobile-apps-than-the-internet_b67349>

[14] Jason Kincaid. *Facebook Now Has 750 Million Users.* [online], 2011,
Available from: <http://techcrunch.com/2011/06/23/facebook-750-million-users/>

[15] Freeman, Corbin Ball, CSP, CMP. *Social Media: Extending & Growing Your Brand*. 2009

[16] Clicks and Links Ltd. *Online Social Networks*. Department for Communities and Local Government, Research Report, 2008

[17] Educause Learning Initiative. *7 things you should know about… Blog.* 2005

[18] Educause Learning Initiative. *7 things you should know about… Podcasting.* 2005

[19] Wikipedia. *Internet Forum.* [online], (n.d.),
Available from: <http://en.wikipedia.org/wiki/Internet_forum>

[20] Daniel Burrus. *Social Networks That Boost Your Business.* [online], 2010, Available from: <http://web2.sys-con.com/node/1352612>

[21] Educause Learning Initiative. *7 things you should know about… Wikis.* 2005

[22] Allison Fine. *Content Communities for Coalitions, Getting Started with Social Media*. A CADCA Institute Webinar Series, 2008

[23] Osmar R. Zaïane. *Principles of Knowledge Discovery in Databases*. University of Alberta, Department of Computing Science, 1999

[24] Jeffrey W. Seifert. *Data Mining: An Overview.* Information Science and Technology Policy, Resources, Science, and Industry Division, CRS Report for Congress, 2004

[25] Vidhya. K. A & G. Aghila. *Text Mining Process, Techniques and Tools: an Overview*. International Journal of Information Technology and Knowledge Management, 2010

[26] Iab.net. *April 2011: The Days of Double-Digit Growth in Social Network Users Are Over.* [online], 2011, Available from: <http://www.iab.net/insights_research/industry_data_and_landscape/1675/1644724>

[27] Miguel Gomes da Costa Júnior, Zhiguo Gong. *Web Structure Mining: An Introduction*. Department of Computer and information, Science Faculty of Science and Technology, University of Macau, China, 2005

[28] Nikolaos Mantas. *Implementation of data mining techniques in Web data*. Thesis, Aristotle University of Thessaloniki, 2010

[29] Raymond Kosala, Hendrik Blockeel. *Web Mining Research: A Survey*. Department of Computer Science, Katholieke Universiteit Leuven, 2000

[30] Christos Syllas. *Trending.gr: Για τι μιλάμε στο twitter;* [online], (n.d.), Available from: <http://www.journalism.gr/el/home/themata/64-people/221-trendinggr-twitter.html>

[31] *Trending main page* [online], (n.d.), Available from: <http://trending.gr/>

[32] Kissmetrics. *Twitter Statistics.* [online], (n.d.),
Available from: <http://blog.kissmetrics.com/twitter-statistics/>

[33] Online Marketing Trends. *Twitter statistics on its 5th Anniversary*. [online], 2011, Available from: <http://www.onlinemarketing-trends.com/2011/03/twitter-statistics-on-its-5th.html>

[34] Aikaterini Tsagalidou. *Semantic definition of views and subjective classification of posts in social networks, Case study Twitter*. Aristotle University of Thessaloniki, 2011

[35] Provecuador. *Twitter search engine.* [online], (n.d.),
Available from: <http://www.provecuador.com/>

[36] *Social Mention main page.* [online], (n.d.),
Available from: <http://www.socialmention.com/>

[37] AppAppeal. *Social Mention.* [online], (n.d.),
Available from: (http://www.appappeal.com/app/social-mention/)

[38] José María Gómez Hidalgo. *The difference between Information Access and Information Retrieval*. [online], 2009,
Available from: <http://jmgomezhidalgo.blogspot.com/2009/01/difference-between-information-access_26.html>

[39] Quora. *What is the difference between information retrieval and text mining?* [online], 2011, Available from: <http://www.quora.com/What-is-the-difference-between-information-retrieval-and-text-mining>

[40] Wikipedia. *Sentiment Analysis.* [online], (n.d.),
Available from: <http://en.wikipedia.org/wiki/Sentiment_analysis>

[41] Kathimerini. *Thessaloniki swamped by garbage.* [online], 2011,
Available from: (http://www.neoskosmos.com/news/en/Garbage-crisis-Thessaloniki)

[42] Mordechai Haklay. *Public Environmental Information – Understanding requirements and patterns of likely public use*. Department of Geomatic Engineering and Centre for Advanced Spatial Analysis, University College London, 2002

[43] European Commission, Eurobarometer. *Attitudes of European citizens towards the environment.* Directorate General Environment, DirectorateGeneral Communication, 2008