George Papapanos, Computer Scientist, MSc in Applied Geo-informatics
Department of Geography, University of the Aegean
gpapapanos@aegean.gr

Sofia Karampela, Economist-Regionalist, MSc in Tourism
Department of Geography, University of the Aegean
karampela@aegean.gr

"What Do You See in the Landscape of Lavrio?" Photography and Text Analysis of Geotagged Photos in Flickr.

Abstract

Flickr website has been widely used from a large number of users, who share their travel experiences and everyday life by uploading photos and writing textual comments. In this paper, we report a study of extracting textual comments and photos, tagged with geographical information uploaded by users for the area of Lavrio which has once been a flourishing industrial city with a multitude of industries, but after the definitive closure of the mines and most of the industries as a consequence of the general de-industrialization of the country, a period of economic crisis and increased unemployment took place. The results of (a) the textual comments in English from title, description and tags are analyzed with text mining procedures and the most "popular" words are presented, (b) the geo-tagged photos (except of those licensed as all rights reserved) produced maps with the most visited areas and a story map, (c) the statistics for the views and date taken per photo determines the most "popular" one and (d) one hundred (100) random sample photos categorized. This information provides opportunities to understand behaviors and the potential attractions in the case study area. The photographic evidence reveals quite different stories, memories, attitudes, gives a sense of the place and the local identity of Lavrio however this evidence is not related with the economic depression and crisis that the region has faced during the previous years.

Keywords:
Geo-information, Geo-tagged photos, Flickr, Text Mining, Lavrio

Introduction

The proliferation of digital photo and video-capture devices equipped with global positioning system (GPS) and the growing practices of sharing photos and videos online using social media sites, such as Flickr (flicker. com) and YouTube (youtube. com), have resulted in huge volumes of geo-tagged photos and videos available on the Web (Majid et al., 2015). Flickr website has been widely used from a large number of users, who share their travel experiences and everyday life by uploading photos and writing textual comments. These individuals are in many cases in the best position to provide information that requires indigenous experience, esoteric understanding of a particular physical environment, and current information about local conditions (Flanagin & Metzger, 2008). Each photo may

contain metadata added by its photographer, such as tags that describe either its visual content, its geo-location, or a free text description somehow related to the photo contents and the roots of this annotation process lies within the analog photo era, where users wrote some "metadata" information, such as place and date, behind paper photos (Spyrou & Mylonas, 2016).
The analysis of the textual information provided by the user is called text mining which can be broadly defined as a knowledge-intensive process in which a user interacts with a document collection over time to extract useful information from unstructured textual data through the identification and exploration of patterns (Feldman & Sanger, 2007).
In this paper the textual comments in English from title, description and tags are analyzed with text mining procedures and the most "popular" words are presented; the geo-tagged photos (except of those licensed as all rights reserved) produced maps with the most visited areas and a story map; the statistics for the views and date taken per photo determines the most "popular" one and one hundred (100) random sample photos categorized. The categorization was based to a previous study of Beerli & Martin (2004) for factors influencing destination image, recently adopted by Stoleriu & Ibanescu (2017) for online destination image projected by visitor photos which were posted on the TripAdvisor website.

Objectives

The main objective was to find out information about the area of Lavrio and its landscape through Flickr's photos and their metadata and to investigate if someone could get a sense of the place through this process. Moreover a question arouse whether this piece of information either the photo or its metadata was accurate according to its content or to where it was placed on a map, as the user has the freedom (no administrative control over the information provided by the user) or the unawareness of how and where to upload it to Flickr.

Methodology

The textual content associated with a photo consists of a set of tags, a title and a description. Moreover textual metadata often serve as a reminder of the context of the image for the photographer and his social circle (Nov et al., 2010). In this case the model used for text mining is called 'bag-of-words' because any information about the order or structure of words in the document is discarded and the model is only concerned with whether known words occur in the document, not where in the document (Goldberg, 2017). In order to have a complete picture as possible about the history and the identity of the place considering both textual and geographical metadata the Flickr's API was used to query for photos (except of those licensed as all rights reserved) that their field tag included most of the possible written combinations of the word Lavrio such as: "Lavrio", "Lavrion", "Layrio", "Layrion", "Λαύριο", "Λαύριον", "Λαυριο", "Λαυριον", "Λάυριο", "Λάυριον".
The output of the above query was about 130 photos and their metadata which were used for the following procedures of various visualizations and results:

- Points were inserted into QGIS for verification that the photos were geo-tagged accordingly,

- Removal of irrelevant geo tagged points,
- Text mining for finding the most "popular" words in English from the fields of title, tags and description with the use of R which is an open source language (R Project for Statistical Computing, n.d.) presented in a word cloud,
- A cluster map as a way of representing in any scale a set of points of close proximity into clusters; it was developed in Google maps so as to find out the aggregation of points,
- A heat map as a representation of data in the form of a map in which data values are represented as colors; it was developed in Google maps in order to find the concentration of points at the area of Lavrio,
- Creation of a story map in ESRI ArcGis online for presenting this piece of information in the form of a story which gives the user the freedom to combine authoritative maps with narrative text, images, and multimedia content and generally they make it easy to harness the power of maps and geography to tell any story (ESRI, n.d.). In this case the story was visualized into four categories, analyzed in a following step,
- Statistics for the views and date taken per photo determines the most "popular" one,
- Categorization of one hundred (100) random sample photos according to their visual content based on four categories derived from Beerli & Martin (2004) and developed by the authors as: a) Culture and Tourism, b) Social Environment and Atmosphere of the Place, c) Natural Resources and Natural Environment and d) General Infrastructure,
- A collage of twenty five (25) random photos to reveal the sense of the place and the local identity.

Case Study Area

The conference for which this article was written took place at the Lavrio Technological Cultural Park, at the area of Lavrio, in the Municipality of Lavreotiki (Figure 1). Lavreotiki extends to the southeast of Attica with 176.87 km^2 and has signs of habitation from the Neolithic period, as people were always attracted by the rich subsoil. Probably the very name «Lavrion» comes from the word «Lavra» or «lavri», which means alley, paved street. The word can also be found in the great ancient Greek poems by Homer (8th century B.C.), meaning the corridor, or a passage. In fact, the town of Lavrio is a modern town, founded in the late 19^{th} century by European exploiters (Guide of Lavrio, 2018) and has once been a flourishing industrial city with a multitude of industries, but after the definitive closure of the mines and most of the industries as a consequence of the general de-industrialization of the country, a period of economic crisis and increased unemployment took place.

More specifically, the French Lavrio Mining Company, which was the longest mining and metallurgical industry in Greece, operated until 1992. The factory of the company at the Kyprianos area of Lavrio, after its closure, was bought by the Greek state and was granted to the Ministry of Culture, which declared it as a preserved monument. It was then assigned to the National Technical University of Athens, which founded the Lavrio Technological Cultural Park.

Figure 1: Municipality of Lavreotiki

Source: Google Maps APIs, processed by the authors

According to the census of the year 2011 the Municipality of Lavreotiki had about 25.000 residents (Table 1) where 7.000 of them in Lavrio. During the decade 2001 - 2011 the population and the economically active population increased but on the other hand the unemployment rate was also increased as a result of the economic crisis of the country (Hellenic Statistical Authority, 2018).

Table1: Characteristics of Lavreotiki Municipality

Variables \ Year	**2011**	**2001**
Area (km^2)	176.87	
Population	25,102	20,199
Population change	24.3%	
Economically Active Population	10,192	8,872
Unemployed Population	2,445	1,387
Unemployment Rate (Unemployed Population/Economically Active Population)	24.0%	15.6%

Source: Hellenic Statistical Authority, processed by the authors

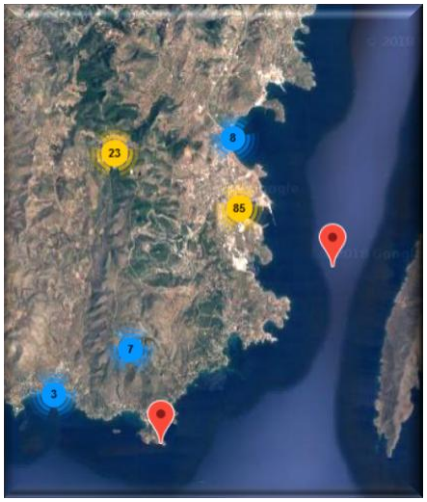
Results

The output of the text mining procedures produced a matrix with the most "popular" words and their frequency values which represent how often each word occurred. The final result is a world cloud (see Figure 2a) of the twenty most frequently found words in the data set where in the first ten the history of Lavrio as a mining site is presented ("mining", "silver", "mines"). The overall image of the words also provides geographical information with the words "greece", "attica", "athens", "sea", "konstantinos" and about history with the words "empire", "bce", "classical", "era", "persian".

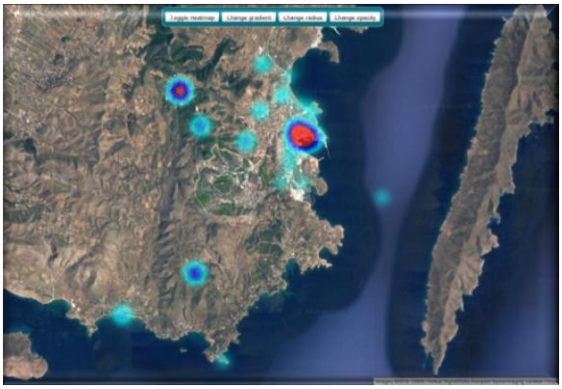
As far as the cluster map most of the points (approximately 85 from 130, 65%) are located into the city of Lavrio (see Figure 2 b1). Similarly the heat map shows with colors that the concentration of points is at the city of Lavrio (Figure 2 b2). The story map for Lavrio (Figure 2 b3) was created including all the available information from Flickr and the four categories.

Figure 2: Results of word cloud, cluster map, heat map and story map

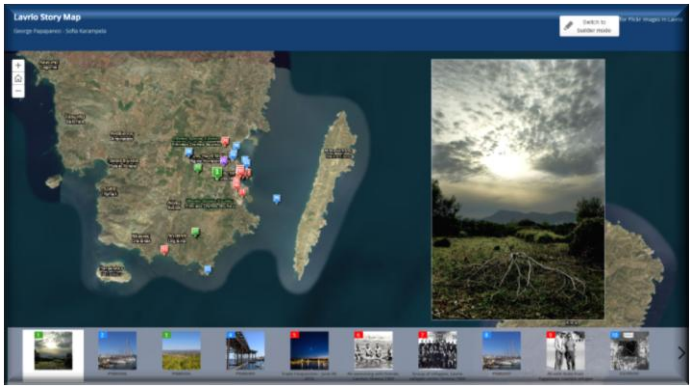
(a) Word cloud



(b1) Cluster map



(b2) Heat map



(b3) Story map

Source: WordClouds.com, Google Maps APIs, ESRI Story maps processed by the authors

The most viewed photo (Figure 3c, CC BY-NC 2.0 by Ali Eminov) from the derived data set had 4.323 views. As the owner states it is a "group of refugees sit for a picture with camp administrators at Camp Lavrion before their departure to Athens and eventually to the USA". This photo reveals the aspect of people leaving Greece after the civil war and specifically the history of the refugee camp of Lavrio which is still active nowadays.

Regarding the categorization of one hundred (100) random sample photos, the first category of culture and tourism was the biggest one with 52%, followed by social environment and atmosphere of the place with 28%, natural resources and natural environment with 14% and general infrastructure with 6%. The same percentages occur in the city of Lavrio where the first two categories (culture and tourism & social environment and atmosphere) are presented the most. More specifically the culture related to the high numbers of historical buildings and abandoned mine sites while the social environment related to "underprivileged people" who where refugees working at the mines after the Greek civil war (the most popular photo described above, Figure 3c, belongs to this category). A random photo collection of 25 photos from the area of Lavrio is presented in the form of a collage (Figure 3d, CC BY-NC 2.0 by Nasos Efstathiadis, Ali Eminov, Nikolas, Thanasis Papathanasiou, Der Wunderbare Mandarin, Costas Tavernarakis, Ava Babili, Cptn11, Stephen Pougas, Panos Kritsonis), where all the categories are represented and the piece of information about the area is comprehensive and easily identifiable by the viewer.

Figure 3: The most popular photo and a collage of collected photos

(c) The most popular photo

(d) Collage of collected photos

Source: Flickr.com, photocollage.com processed by the authors

Conclusions

This study provides opportunities to understand behaviors and the potential attractions in the case study area. The photographic evidence reveals quite different stories, memories, attitudes, gives a sense of the place and the local identity of Lavrio however this evidence is not related with the economic depression and crisis that the region has faced during the previous years. It might be a case that people want to capture and share joyful moments of their life or photos that are pieces of art.

Flickr and other photo sharing sites tend to be or are already visual archives where the exchange of photos, experiences and interpretations in the present naturally lead to a communal sense of the past (Van Dijck, 2011). In this context, a question of high importance is when these photos were taken or their metadata uploaded to Flickr for someone to have a sense of the identity of the place in present as the development and the environment of places is dynamic. For example if someone would take a look of the photos taken in 1960 which are the most photos belonging to category of social environment would have a different perspective of the place than taking a look to the whole data set or those photos taken after 2010. However the local identity of the place which is based in historical, cultural and social aspects is revealed up to some degree to someone, even if he has not visited the place and even though that the photos which present social aspects of the area of Lavrio have been taken many years ago.

Implications & Recommendations

The search or the query for photos and their metadata from Flickr's api service represent significant new benefits to the study of geography. However these queries for specific areas into the field of tags (or where free text is allowed) might not return all the possible values, as users may have misspelled it or totally omitted it. These issues of credibility on the volunteered user-generated geo-tagging should become of broader research interest in various areas (Spyrou & Mylonas, 2016). They also suggest serious threats to the veracity of

geodata, and the degree to which information thus provided can and should be trusted (Flanagin & Metzger, 2008). There could be an automatic proposal for geo-tags based on the coordinates of each photo so that there is a homogeneity in the information and metadata that pertains to each area and the avoidance of defining the wrong areas by the user.

References

Beerli, A., & Martin, J. D. (2004). Factors influencing destination image. *Annals of tourism research, 31*(3), 657-681.

ESRI (n.d.) Story maps. Retrieved from storymaps.arcgis.com/en/

Feldman, R., & Sanger, J. (2007). *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. USA: Cambridge University Press.

Flanagin, A.J., & Metzger, M.J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72(3-4), 137-148.

Flickr (n.d. a) The app garden. Retrieved from www.flickr.com/services/api

Flickr (n.d. b) Website. Retrieved from www.flickr.com

Goldberg, Y. (2017). Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies*, *10*(1), 65.

Google Maps APIs (n.d.) Retrieved from https://developers.google.com/maps

Guide of Lavrio (2018). Retrieved from http://www.lavrioguide.gr/english/

Hellenic Statistical Authority. (2018). Various reports. Retrieved from http://www.statistics.gr/home

Majid, A., Chen, L., Mirza, H.T., Hussain, I., & Chen, G. (2015). A system for mining interesting tourist locations and travel sequences from public geo-tagged photos. *Data & Knowledge Engineering*, 95, 66-86.

Nov, O., Naaman, M., & Ye, C. (2010). Analysis of participation in an online photo-sharing community: A multidimensional perspective. *Journal of the Association for Information Science and Technology*, 61(3), 555-566.

R Project for Statistical Computing (n.d.) Retrieved from https://www.r-project.org/

Spyrou, E., & Mylonas, P. (2016). A survey on Flickr multimedia research challenges. *Engineering Applications of Artificial Intelligence*, 51, 71-91.

Stoleriu, O.M., & Ibanescu, B. (2017). The online destination image of Danube delta (Romania) projected by visitor photos. *4th International Multidisciplinary Scientific Conference SOCIAL SCIENCES & ARTS SGEM 2017.*

Van Dijck, J. (2011). Flickr and the culture of connectivity: Sharing views, experiences, memories. *Memory Studies, 4*(4), 401-415.

WordClouds.com (n.d.) Retrieved from https://www.wordclouds.com/