The Effect of Social Media Trending Topics Related to Cultural Venues' Content

Vassilis Poulopoulos*, Costas Vassilakis*, Manolis Wallace*, Angela Antoniou† and George Lepouras†
* Knowledge and Uncertainty Research Laboratory
Department of Informatics and Telecommunications, University of the Peloponnese, Tripoli, Greece
Email: {vacilos, costas, wallace} @uop.gr
†Department of Informatics and Telecommunications, University of the Peloponnese, Tripoli, Greece
Email: {angelant, gl} @uop.gr

***Abstract*—Social media have gained the majority of attention on the Internet having an extreme number of daily visitors worldwide. The amount of information exchanged is vast, while users have become equally producers and consumers of data. Words like "trending", "influencers", "likes" and "viral" are in the daily agenda of data analyzers, as they are associated with factors that play an important role in the influence of social media content to its audience. These aspects are nowadays strongly taken into account by organizations that want to draw the public's attention to the content they deliver, and in this context cultural institutions have already started to take under great consideration not only their presence in social media, but also the monitoring and exploitation of social media dynamics. In this paper we propose a method that can enable cultural venues to benefit from matches between their own content and ongoing discussions on social media. More specifically we extract trending topics that can be related semantically with the content of a cultural institute and examine how a venue can benefit by exploiting these matches. The proposed approach has been developed in the scope of the "CrossCult" H2020 project, and has been experimentally tested by analyzing the case of Twitter in Greek language.**

***Index Terms*—trending topics, cultural venues, cultural informatics, social media, personalizationtrending topics, cultural venues, cultural informatics, social media, personalization**

I. Introduction

The importance of cultural heritage is now recognized, not only by historians and scholars; cultural heritage is now placed within the core of the citizens' rights to information and education, white its preservation is also established as a responsibility towards future generations. In this context, cultural information sites, typically maintained by cultural venues and educational organizations, are continuously developed and enriched, while the European Union has declared the year 2018 as year European Year of Cultural Heritage[18]. An important goal of this declaration is to inform people, to approach new audiences, to note that cultural heritage can be an integral part of our present and future, and contribute to the process of shaping it. In this context, the role of technology and the Internet is central, since both technology and the Internet have become an indispensable part of peoples' everyday life. Social media, in particular, offer unprecedented ease of access to information sharing and consumption, being thus one of the major technological and social developments to be considered for the whole lifecycle of cultural data management, including production, enrichment and dissemination. All major cultural venues nowadays exploit social media, at least for dissemination and outreach purposes, while additional methods for harvesting benefits from the use of social media by cultural venues are examined within the discipline of *cultural informatics*.

978-1-5386-8225-8/18/$31.00 ©2018 European Union

More specifically, a large number of cultural organizations already base a part of their activities and allocate budget on social media related tasks, by trying to locate potential customers and people that can contribute to the spreading of their message. The procedure includes concepts such as "trends", "viral" and "influencers", which are central in the world of social media. Analysts from several different sectors are involved in the aforementioned procedure, including social media experts, online marketing specialists, data analysts and big data scientists, all of which leads to a simple conclusion: the effect of social media on people is such that it is worth investing in their use for marketing and outreach purposes. A combination of the aforementioned - organizations and social media - is a matter of extensive research and when it comes to interconnection of information technology with culture and more specifically cultural heritage, a number of interesting projects can be identified.

CrossCult, an Horizon 2020 EU project [17], is one such project that aims to alter the way people view history and supports multiple interpretations of the European past in a cross-border fashion. Within the scope of CrossCult project we examine the stimulation of reflections to people by enabling several different interactions within a museum visit. Such interactions will allow the visitors to engage in deeper levels of reflection, compared to traditional means of history presentation (e.g., type of a statue, or its construction date), including social aspects of life in antiquity, power structures etc. Starting from the premise that social media play an important role to the shaping of peoples' opinion and angles of view, we research on the effect of social media trending topics when related to cultural venues' content.

The main hypothesis of our research presented in this paper, is that cultural venues can benefit from the social media, by

identifying trending topics or discussions that are related to their content, either permanent or temporary (eg. seasonal exhibition). When such a relationship is established, cultural venues will need to react by participating actively in social media posts, seizing the opportunity to highlight the linkage between the developments in the world of social media and the contents they offer. We assert that in this way, social media users will be intrigued and stimulated to examine the content offered by the venue; this will promote the dissemination of the venue's cultural content and ultimately lead to an increase of digital or physical visitors.

In this paper we present and discuss our approach regarding the establishment of relationships between social media trending topics and venues' cultural content. We present a system that examines content posted on Twitter in Greece (greek language) and identifies cultural venue topics that are most relevant to the ongoing discussions. This can enable cultural venues to promote their content, as discussed above, but can also assist in identifying conversations in which the cultural venue could (or perhaps should) actively participate due to their relevance to the venue's cultural content and message.

The rest of the paper is structured as follows: section II presents the related work while section III analyses our system in detail. Section III presents the experimental evaluation and we conclude with remarks and future work in section.

II. Related Work

Museums and cultural venues have recognized the added value of social media and already put their focus in them. Actually, several studies are concentrating on the effect of social media in different types of users, many of them concentrating on cultural content. According to Fletcher and Lee [2], museums are currently trying to increase their use of social media for multi-way communication strategies. Nevertheless, cultural venues need to concentrate on interaction on social media rather than one way communications.

In cultural venues people can find interconnections of the past with their own experiences. Social media can enable sideways of achieving the interaction by providing tools for engagement, knowledge acquire and communication. In social media, users/visitors can organize ideas and interpretations to create meaningful associations between their own and others experiences [8]. Social media can become a powerful tool in the hands of museums and cultural sites. In [8] a framework for museum visitors experiences through social networks is proposed in order to enrich physical learning. This model includes rapid publication of museums, personalization of the museums content, content sharing and content creation by the audiences. In another study, Instragram was researched in order to explore how it can be used to communicate the experience of the user visit [9]. In parallel, as the social media become a part of the procedures of cultural sites, it is considered as necessary to train the employees to understand the nature of the social media landscape in order to understand limitations and opportunities [10].

A number of recent projects are researching on the combination of technology and cultural heritage and more precisely they are focusing on the visitor's experience enhancement. PLUGGY project's scope is to create an application that will act as a social platform through which visitors will be able to act as "storytellers". Participants will be able to create fascinating personalized stories and share them through social networking with friends, associates and professionals. The content will be both crowd-sourced and retrieved from digital collections, allowing users to create links between seemingly unrelated facts, events, people and digitized collections [11]. The crowd-sourcing contribution to virtual museums and digital cultural heritage, through a platform, is one of the main outcomes of ViMM project [12] while EMOTIVE project intends to provide tools create engaging, memorable stories for cultural heritage, and to enable the public to enjoy and share these stories [16].

One of the goals of the CrossCult EU project is the examination of stimulation of reflection discussion in museum visits through the use of social media. As the project research implies, information deriving from twitter's trending topics allows the preparation of personalized cultural experiences, tailored to the cognitive style (long term parameter) and current interests (short-term parameter) of museum visitors and enhances the online presence of the museum and user engagement [1], [4].

In our study, we made the choice of twitter for examining the trending topics for two main reasons. Firstly, twitter offers the possibility of direct access to lists of trending topics while in parallel studies have shown that there are indication that Twitter's feature set and multi-platform presence play major roles in mediating cultural performances by Twitter users [3]. The latter, implies that it could be considered as an ideal platform for usage for cultural venues, as it seems to be suitable for the purpose they serve [5].

III. System Analysis

For the purpose of the research project we developed a system that is able to perform analysis on Twitter's posts. The system intends to recognize tweets that are relevant to the museum content, or to keywords that are selected by each cultural venue. The interaction of the museum managers to the presented content is only restricted to selecting if it is useful or not, and, of course, be able to see it; in a more sophisticated analysis they select if they will interact with the specific tweet or not.

In many cases the content or keyword target may change from time to time. The main factor of the changes are the alteration of the museum exhibitions or special exhibitions. While in the generic case, a simple search procedure can reveal a large number of interesting information we need to focus on a more refined procedure through which we will not only establish a simple keyword relevance. Going a step beyond, we extract qualitative information related to the content relevance to trending topics as well as hashtags or tweets originating from users with large audience. By doing this, the topics of

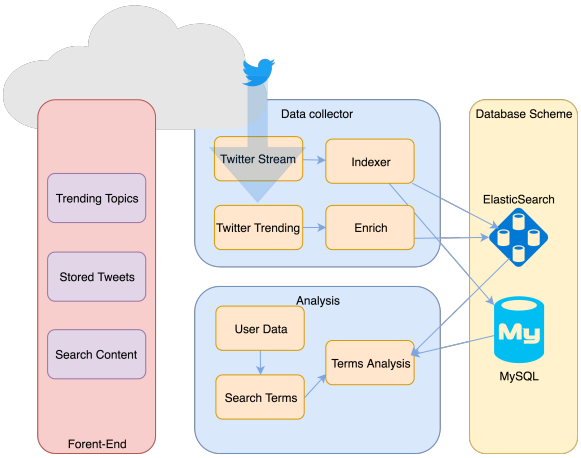

Fig. 1. System Architecture.

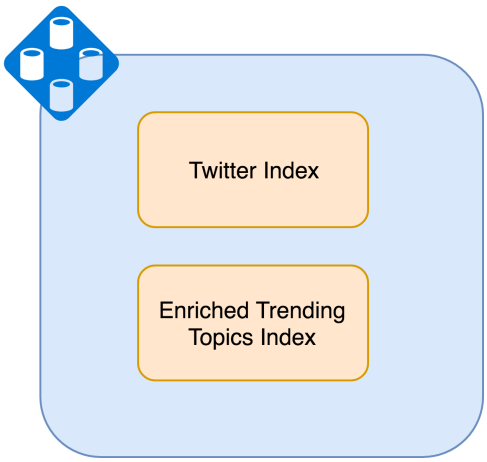

Fig. 2. ElasticSearch Index.

interest are not only restricted to a small audience but have higher possibilities to be involved in trending topics or viral conversations.

To achieve the aforementioned the provided by Twitter API, endpoints [13], [15] and [14] are not sufficient as there is a strong need to dig information into the tweets. The proposed solution is based on the collection and lexical analysis of large amounts of twitter posts. A modular system is constructed to perform the procedures, which consists of the collector, the storage engine, the analysis and the front-end. The system architecture is presented in figure 1.

The collector is responsible for collecting data from twitter. The Twitter stream [15] API is used in order to obtain the tweets posted. According to its API definition, the stream is a non-stopping communication link that fetches in real-time the tweets that are posted and are relevant to a list of search terms (filters). Our system requires that all the tweets posted need to be obtained in order to be further analysed. To achieve that we use a common technique that is based on the fact that all tweets for a specific language (which in some cases, such as the case of Greek, maps very closely to a country) can be obtained by using as search terms the most important stop-words of the specific language and as parameter filter the language itself. This enables us to create an ”off-line” copy of the tweets on which we need to perform analysis.

Apart from the main twitter stream data collector we are utilizing the Twitter Trending API endpoint [13] in order to collect information about the trending topics. The trending topics API definition implies that the topics have an interval of update of at least 5 minutes. We are executing the trending topics search procedure every hour in order to check the changes in topics, and their volumes. The trending topics are usually related to single terms or hashed terms, which are possibly difficult to locate in museum content. Our analysis is furthermore based on significant terms frequency and in the parallel search of these terms in our infrastructure.

While a part of the back-end is responsible for collecting data while a second subsystem is responsible for user data analysis. The second subsystem collects information from users and formulates the search queries to be executed. Finally, it collects user responses to the presented results and permanently stores the information for future use. The analysis in terms of digging within the tweets collected from the stream is done using the search capabilities of ElasticSearch [1], while in terms of searching for trending topics the procedure is more complex. Each trending topic extracted from twitter is enriched in keywords by searching on the local pool of tweets for “significant terms”. The complete list of trending keywords together with their significant terms are stored in a separate index and searching for museum content related to trending topics is done on the second index.

The storage engine is a hybrid database structure solution. After obtaining the tweets from Twitter stream API, they are saved in two different databases. The first one is an RDBMS (MariaDB [2]) for permanent storage of the tweet accompanied with meta-data (user, number of retweets, number of favorites, etc.) while in parallel the information related to the tweet body is stored to a noSQL database, more specifically ElasticSearch. The data is not stored to an arbitrary ElasticSearch index. The index is specially constructed in order to perform NLP on the added text, a functionality natively supported by ElasticSearch, which is using Lucenes NLP engine to achieve this. As such, stemming, stop-word removal, word delimiting, sentence delimiting is performed by the database engine itself using the icu-analysis plugin [3]. The engine is able to perform high volume complex query semantic search that is essential for the performance of our system.

The front-end solution is minimal, as its main scope is to exchange information with users that are usually not accustomed to complex environments. As such the system provides few different functionalities; adding new search terms (words, sentences or even complete texts), viewing the current search

[1]ElasticSearch, https://www.elastic.co/, Open Source Search and Analytics
[2]MariaDB, https://mariadb.org/ Open Source Database
[3]ICU - Analysis Plugin for ElastiSearch https://www.elastic.co/guide/en/elasticsearch/plugins/current/analysis-icu.html

terms and viewing the results on search term analysis on twitter. On the results presentation the users are able to rate based on the usefulness of the presented result (tweet).

IV. Experimental Evaluation

The experimental evaluation is done in relation to two factors,

- relevance of the fetched results to the museum content,
- analysis on the rating of the usage of each presented result

While the first procedure can be done automatically and extracted by the algorithmic procedure applied by our system, we furthermore score the relevance as it can be noted by a human. On the other side the rating of usefulness is totally subjective procedure and is based on the human factor.

While the system is running it is able to collect a number of 200k-300k tweets per day considering exclusively tweets written in greek. In parallel, the trending topics that are collected daily vary from 50-100, which are stored into the database alongside their enriched terms that are 5 at most (limited by our query). Enrichment is done by using ElasticSearch's capabilities of locating similar terms. This leads to a number of 250 to 500 "trending terms" per day. The trending topics are not totally recycled on a daily basis but they usually remain active as much as a subject is in the news.

According to the aforementioned analysis it is clearly difficult to approach the collected data using only the human factor. As such, we apply the advanced search procedures provided by ElasticSearch in order to obtain the data with high relevance. By relevance ElasticSearch notes that the standard similarity algorithm used in Elasticsearch is known as term frequency/inverse document frequency, or TF/IDF in order to provide a score of relevance. In any search perform we apply the time factor considering that data in the social media are not considered "fresh" even only a few hours after they are produced, as such the time of our searches is limited to the last 48 hours.

The testing procedure is conducted from data collected in one month time sizing more than 7M tweets and topics in the index - still being updating and offering results in real time.

A. Relevant Tweets

The search terms are expected to be large texts deriving from cultural venues. Thus, it is expected that a large number of tweets will be returned as relevant during search; leading to limitation to the number of search results to 20 per search, which are ordered by similarity score. As a first approach, texts from several sources of a cultural venue are extracted and used as the search terms. This approach leads to results that can be considered irrelevant, still having a high relevance in score. This can be considered as expected as long as the texts are using generic terminology. So approaching the system usage from the view of using texts as a whole leads to the need of refinement. Figure 3 presents a list of results.

As such, we perform text selection and removal of sentences or words that include irrelevant terms. We repeat the procedure of searching on data. From the results it is more than obvious

ID	Text	Score
974207251505340416	Δωρεάν σπιρομετρήσεις στην Αρχαία Ολυμπία	26.726303
972094680291790848	Περιήγηση στην αρχαία Ολυμπία Φίλες και φίλοι καλησπέρα. Σήμερα θα μεταφερθούμε στην Ολυμπία, το πιο δοξασμένο...	24.430382
971882331974914048	Δίαθλο την Κυριακή στην Αρχαία Ολυμπία	24.36227
973805202892574720	Τα αστέρια του μέλλοντος έλαμψαν στην Αρχαία Ολυμπία	24.194458
971777053874847745	Πολιτικός μηχανικός στην Αρχαία Ολυμπία Γεωργακόπουλος Βασίλειος	24.161346
973854027900444672	Τα αστέρια του μέλλοντος έλαμψαν στην Αρχαία Ολυμπία-Φωτό	23.817137
974228009443123200	Δωρεάν σπιρομετρήσεις τη Δευτέρα 19 Μάρτη στην Αρχαία Ολυμπία	22.211445
972559273594118144	Ένας φωτογραφικός περίπατος στην Αρχαία Ολυμπία Ηλείας από το φακό του Dimitris Zogokis Skiouros !!!!!!	21.405624
971853099831906304	Αετός, το Σύμβολο του Δία: Στην Αρχαία Ελλάδα και τον κόσμο.	20.987387

Fig. 3. Search results on term: "The statue of Zeus in Ancient Olympia (in greek)".

that the results can now be considered as more relevant in subject matter than the previous procedure.

An important outcome is that the score of the newly formatted results is lower. This is expected as we narrow the search terms and by means of terminology our search is much more specific and precise.

B. Results usefulness

The next procedure includes human intervention and more accurately requires the scoring of the presented results, regarding how useful can be for the museum online marketing and dissemination purposes. The idea of evaluation is based on both the relevance and usefulness as a whole. It seems that trending topics that cannot be considered directly as relevant, can be extremely useful while in parallel some very relevant results can be proved not at all useful. It is indeed a subjective procedure; in order to achieve recording the usefulness we add an action link indicating a tag of "Useful", "Not Useful", "Neutral" as well as "Remove".

As already mentioned, it is important to describe that through the experimental evaluation it is easy to understand the existence of different views of the result. Table I presents the different situations followed by an explanation.

The tweets that are considered to be neutral are not at all examined but still remain relevant. In this category we observe texts that more likely include facts rather than opinions, news or statements. They are clearly not tending to become trending topics or have high view-ability. Finally, we can observe a number of tweets that are totally irrelevant and the scope is to totally remove them from the presented results.

In fact, the experimental evaluation leads to the observation that there seems to be a strong need to enhance two-way communication in the social media. Even in cases of irrelevant information - not semantically connected to the content - there is strong possibility to interact, gain audience, attract more people and spread messages.

TABLE I
TAGGING EXPLANATION

Relevance	Usefulness	Explanation
High	High	A topic/tweet that is directly related to the museum's content and can possibly help for dissemination or spreading a cultural related message
High	Low	A topic/tweet that is directly related to the museum's content but is impossible to intervene to the conversation either because of the people participating or the context
Low	Low	Expected for most of irrelevant tweets
Low	High	Irrelevant tweets but very useful due to the fact that participants or context of tweet/conversation is such that the message can give added value to the conversation

V. CONCLUSIONS AND FUTURE WORK

Online presence of cultural venues, as well as online presence of digital heritage related data, can be useful from the scopes of preservation, people awareness and for information purposes.

We presented a module acting as a useful tool for museum managers or curators within the scope of the CrossCult project and more precisely as part of pilot 3 of the project. The scope of the "Twitter trending topics module" is to serve as an online management assistant for cultural venues and more specifically help them reach broader audience, spread their message more accurately and put the focus on conversations to data related to the museum or exhibition.

We presented its hybrid scheme that can help us achieve high level of text analysis, having in parallel high performance of the system. The modules produced are fully implemented and tested, and they are already provided to related cultural venues for testing and integration to their procedures.

We evaluated the results within the scope of the project and evaluated the complete system against more than 7M tweets collected in one month time. We conclude to the fact that copying data from the cultural venue sources could be proved to be disorientating, so a first data filtering must be applied. Due to the size of these texts the procedure can easily be done manually. Furthermore, we realized that from the remaining texts we can obtain information even in cases of low relevance as long as the context can be proved suitable. As suitable context can be considered cases that the users writing the tweets are influencing personalities or the conversation includes a large number of people, answers, retweets and favorites.

The system is focused mainly on the content and does not currently analyse qualitative factors related to Twitter's characteristics including personality of the users (followers, number of tweets, number of retweets, favorites, etc), the characteristics of a tweet (favorites, retweets) and their evolution in time. To perform such an analysis every possible tweet has to be collected multiple times a procedure that is out of scope of the current project's module.

For our research purposes we will proceed with the system evolution in terms of collecting meta-data that could help us improve the system. Furthermore, as the system is language dependent we have already started the collection of data in english language in order to support more cultural venues related to the project. The dependence of the language is based on the supported languages of the system text indexer and analyser which is the languages that are supported by ElastiSearch's icu-analysis plugin, which are delending on Lucene's NLP tools.

ACKNOWLEDGMENT

This work has been realized within project CrossCult: "Empowering reuse of digital cultural heritage in context-aware crosscuts of European history", funded by the European Unionś Horizon 2020 research and innovation program, Grant #69315.

A part of this work has been realized within project PaloAnalytics:"A platform for large-scale analysis of web mentions, in a multilingual context, for identification of business opportunities", funded by the "Competitiveness, Entrepreneurship and Innovation" (EPAnEK) Operational Programme of the Hellenic Republic.

REFERENCES

[1] C. Vassilakis, A. Antoniou, G. Lepouras, V. Poulopoulos, M. Wallace, S. Bampatzia and I. Bourlakos, Stimulation of reflection and discussion in museum visits through the use of social media, Social Network Analysis and Mining, Springer, https://doi.org/10.1007/s13278-017-0460-3, 2017.

[2] A. Fletcher and MJ. Lee, Current social media uses and evaluations in American museums. Mus Manag Curatorship 27(5):505521, 2012.

[3] A. Brock, From the Blackhand Side: Twitter as a Cultural Conversation, Journal of Broadcasting & Electronic Media, 56:4, 529-549, DOI: 10.1080/08838151.2012.732147, 2012.

[4] S. Bampatzia, A. Antoniou, G. Lepouras, C. Vassilakis and M. Wallace M Using social media to stimulate history reflection in cultural heritage. In: 11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP 2016), 2021 October, Thessaloniki, Greece, 2016.

[5] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World wide web (WWW '10). ACM, New York, NY, USA, 591-600. DOI=http://dx.doi.org/10.1145/1772690.1772751, 2010.

[6] L. M. Aiello et al., "Sensing Trending Topics in Twitter," in IEEE Transactions on Multimedia, vol. 15, no. 6, pp. 1268-1282, Oct. 2013.

[7] J. Kidd, "Enacting engagement online: framing social media use for the museum", Information Technology & People, Vol. 24 Issue: 1, pp.64-77, 2011

[8] A. Russo, J. Watkins and S. Groundwater-Smith, The impact ofsocial media on informal learning in museums. Educ Media Int46(2):153166, 2009.

[9] A. Weilenmann, T. Hillman, B. Jungselius, Instagram at themuseum: communicating the museum experience through socialphoto sharing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI13), ACM, pp 18431852, 2013

[10] AS. Wong, Ethical issues of social media in museums: a casestudy. Mus Manag Curatorship 26(2):97112, 2011

[11] Pluggable Social Platform for Heritage Awareness and Participation (PLUGGY), EU Project ID: 726765. https://www.pluggy-project.eu/. Accessed 3 May 2018

[12] Virtual Multimodal Museum (ViMM), EU Project ID: 727107. http://www.vi-mm.eu/. Accessed 5 May 2018

[13] Twitter GET trends/place documentation page. https://developer.twitter.com/en/docs/trends/trends-for-location/overview. Accessed 17 Apr 2018

[14] Twitter GET search/tweets documentation page. https://developer.twitter.com/en/docs/api-reference-index. Accessed 17 Apr 2018
[15] Twitter POST statuses/filter documentation page https://developer.twitter.com/en/docs/tweets/filter-realtime/overview/statuses-filter. Accessed 17 Apr 2018
[16] Emotive Virtual cultural Experiences through personalized story-telling (EMOTIVE), EU Project ID: 727188. http://www.emotiveproject.eu/. Accessed 6 May 2018
[17] CrossCult: Empowering reuse of digital cultural heritage in context-aware crosscuts of European history, EU Project. http://www.crosscult.eu/. Accessed 6 May 2018
[18] European Year of Cultural Heritage. https://europa.eu/cultural-heritage/ Accessed 6 May 2018