2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Database Knowledge Enrichment Utilizing Trending Topics from Twitter

Costas Vassilakis, Dimitra Maniataki, George Lepouras, Angeliki Antoniou, Dimitris Spiliotopoulos
Department of Informatics and Telecommunications
University of the Peloponnese
Tripolis, Greece
costas@uop.gr, dimitra_mnt@hotmail.com, gl@uop.gr, angelant@uop.gr, dspiliot@uop.gr

Vassilis Poulopoulos, Manolis Wallace
Knowledge and Uncertainty Research Laboratory
Department of Informatics and Telecommunications
University of the Peloponnese
Tripolis, Greece
vacilos@uop.gr, wallace@uop.gr

Dionisis Margaris
Department of Informatics and Telecommunications
University of Athens
Athens, Greece
margaris@di.uoa.gr

***Abstract*—Every day, many people use at least one social network (or social media) account. This development has been boosted by the rapid growth of technology, making both smartphones and mobile data much more accessible and inexpensive. Therefore, the number of social networks users is growing rapidly, accounting more than 1 billion active users worldwide. The ease of use, as well as the ability to communicate without spatial and temporal restrictions underpinned the rapid increase of the popularity of social networks, as well as their wide acceptance by the general public. This popularity influences people's opinion on many issues, shapes consumer habits and behaviour, mood, etc. The work of many scientists across multiple disciplines has focused on studying social media from various perspectives, including marketing, journalism and sociology. This paper investigates how trending information from social media can be used to match topics of interest from cultural database indices. Matches identified in this process are then presented to cultural venue curators, who can then review matches, mark them as useful or reject them, and exploit them for various tasks, and most notably for the promotion of the venue and its content. More specifically, we have developed an application, which collects the 10 most popular twitter trends and then matches their content with the contents of a given cultural database. Using the results of this match, items from the database that may be related to current issues may be recommended to the user. As a result, these matches, after being inspected and approved by the administrator, can be used to attract the interest of the target audience, highlighting the correlation of current issues with the database's items.**

Keywords—Social Networks, Business Intelligence, Trending Topics, Knowledge Enrichment, Extraction, Processing, Application

I. Introduction

Technology plays nowadays an important role in several fields. Entertainment, infotainment and information, are three of them, in which new technological developments continuously emerge [1]. The phenomenon of social networking has become an integral part of the lives of people belonging to all age groups. Social networking is no longer just a fad, but goes one step further to the point where it is considered to be consolidating as a new social phenomenon, which is considered to be spread even faster than a "virus" [2]. Like all social phenomena, this one has many implications, including economic, political, personal with some aspects however being ominous (e.g. addiction) [3]. Scientists from many disciplines conduct research on all aspects of social networks.

This explosion of information was brought about by the advent of Web 2.0 which allows for multimedia sharing, easy website creation -usually in the form of blogs and web applications with content- thus giving a more interesting form to the Internet that aroused the curiosity of the people and triggered their will to participate [4]. The potential of developing web applications helped in the development of social networking. Another factor that has influenced the development of social networks (SNs) is that mobile phones acquired new features, including the ability to take photos and videos, which, in combination with applications that allow easy publication and sharing of multimedia material, provide all the conditions for the flow of material to Web 2.0 applications, and SNs in particular [5,6].

Among SNs, Twitter, which this paper focuses on, became known for the ability to instantly share information not only with friends, but also with any Twitter user throughout in the world [7]. The basic information unit is the *tweet*, i.e. a simple short text, in which the keywords are marked as hashtags, thus allowing them to be located by other Twitter users [8]. Keywords that appear frequently in the recent period are termed as *trending topics*, and indicate the subjects that are mostly discussed by users.

In this paper, we present the development an application that utilizes the most trending Twitter topics, in order to locate items in a database that are related to these topics. The tracking is done by matching the texts of the tweets that correspond to the popular topics with the texts of the database items (titles and descriptions). To enrich the search and make it more

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00 © 2020 IEEE

effective, synonyms are used from a thesaurus, which is available as a web application. The best matches are suggested to the user of the application for further examination and possible utilization.

The remainder of the paper is presented as follows: Section 2 overviews related work. Section 3 presents the requirements analysis and the design of the application. In Section 4 we present the operation of the application and, finally, Section 5 concludes the paper and outlines the future work.

II. Related Work

Over the last years many works were based on the Twitter social network data, especially for mainstream interests, such as the news, culture, education.

Kim et al. [9] present TNIE (Twitter news in education), a platform which utilizes Twitter for newspaper content in education. Its main advantages, when compared to traditional NIE systems, are that it is able to classify the latest news into various topics, it uses a hierarchical visualization scheme in order to browse the classified news content and it exploits the learners' news selection in order to enable the formation of effective discussion and debate groups.

Kotzias et al. [10] address the informative geostamps sparsity issue via an approach which is able to identify users attached to a given location of interest (e.g. a city). Afterwards, they focus on retrieving finer granularity tweets within a radius of a location, such as a city's block. This approach leverages the correlation between strong connectivity in the social graph and proximity in the real world, while utilizing both textual tweet content and Twitter's underlying social graph.

Deng et al. [11] introduce a dynamic user modelling strategy that exploits information from the Twitter platform in order to upgrade the video recommendation quality in the YouTube platform. More specifically, the proposed strategy firstly extracts real-time hot topics from users' tweets, that are utilized to retrieve related videos from YouTube. Then, in order to generate the final recommendation list, it ranks the obtained YouTube videos, by considering the YouTube user profiles, time factors, as well as qualitative factors. In this way, both the long-term (user profile) interests of users and the short-term (hot topics) are effectively considered in the recommendation formulation.

Yigitcanlar et al. [12] evaluate the utilization of smart cities technologies and concepts. The proposed methodology includes a social media analysis approach (systematic geo-Twitter analysis) which contains content, descriptive, spatial, and policy analyses. As far as the empirical investigation is concerned, they selected the Australian context as their testbed. The results reveal, among others, that the most popular smart city concepts are governance, innovation and sustainability; the most popular technologies are artificial intelligence, internet-of-things and autonomous vehicle technology; the leading Australian smart cities are Melbourne, Sydney and Brisbane and, finally, one of the most useful methodological approach for investigating perceptions and utilization of smart city concepts and technologies is systematic geo-Twitter analysis.

Yaqub et al. [13] investigate two main US presidential candidates of 2016 (Donald Trump and Hillary Clinton) tweet sentiments. They analysed these texts aiming to evaluate how accurately Twitter represented the public opinion and real-world events of significance related with the aforementioned elections. They also analysed the behaviour of distinct Twitter users to identify whether specific opinions were repeated, or original opinions were mostly shared. Last, they evaluated the tweets' sentiments by the two candidates, as well as their impact on the election related discourse on Twitter. Their most important findings are that Twitter was primarily used mainly for retweeting already present opinions, rather than creating original content by users. Furthermore, it was found that public opinion can be predicted by sentiments and topics expressed on Twitter.

Zubiaga et al. [14] explore the trigger types which spark trends on Twitter, introducing a typology with ongoing events, news, commemoratives and memes types. Contrary to earlier research, which analysed trending topics over the long term, the authors look at the earliest tweets that produce a trend, with the aim of categorizing trends, which allows them to provide a filtered subset of trends to end users. They use a set of straightforward language-independent features, based on the social spread of trends. They categorize them using the aforementioned typology, which provides a way to accurately categorize trending topics without requiring additional information. Furthermore, patterns associated with each type of trend are revealed by analysing social features (e.g. short tweets about ongoing events are most probably to be from using mobile devices).

Bao et al. [15] introduce an emerging topic detection and elaboration method, which uses multimedia streams cross different online platforms. They selected Flickr, New York Times and Twitter to represent the image sharing platform, news portal and the microblog, respectively. They use aging theory to extract the emerging Twitter keywords and introduce the Robust Cross-Platform Multimedia Co-Clustering method in order to detect emerging topics. Finally, each emerging topic was enriched with the samples from Flickr and New York Times by computing the implicit links between samples from selected news and Flickr image clusters, obtained by the aforementioned method, and social topics.

Yang and Rim [16] present the Trend Sensitive-Latent Dirichlet Allocation, a topic model which is able to extract latent topics from contents by modelling temporal trends on Twitter over time.

Marchetti and Ceccobelli [17] evaluate the discussions between Italian Twitter users by analysing all the hashtags and keywords that became trending topics in the Italian Twittersphere, during both a non-electoral and an electoral period. They found that entertainment was the main discussion field, consisting of topics concerning actors, singers, etc., while politics is a secondary topic that elicits minor arguments within users' conversations. Furthermore, they empirically confirmed the logic of breaking news events on Twitter, a microblogging site that primarily focuses on exceptional occurrences.

Cai et al. [18] use various visual features in order to study the impact of tweet images on social event detection for different event categories. They design a topic model that jointly models text, location, image, hashtag and timestamp from Twitter texts, to discover events. Finally, they visualize each event by representative images selected on predefined criteria.

Yang et al. [19] present a Country-Level Micro-Blog model which analyses microblogging user activity and behaviour across different countries in the Cyber-Physical-Social System applications. The presented model considers characteristics of user behaviour in microblogging data, including user emotion index, content of microblogging messages and user relationship network. Furthermore, they evaluate the presented model using a microblog dataset from the sixteen countries that are the origins of the largest number of active users. The experimental results show that users pay more attention to social functionalities of microblogging service in countries with small population, but strong cohesion. They use microblogging services as a news dissemination platform for countries containing mostly large loose social groups. They also use more linguistic elements to express happiness in microblogging services in countries whose social network structure exhibits reciprocity rather than hierarchy.

Kowald et al. [20] study temporal Twitter hashtag usage practices aiming to design a cognitive-based recommendation algorithm. Their main idea is to design a predictive model that incorporates the effect of time on both social hashtag reuse and individual hashtag reuse. The proposed algorithm is validated using two crawled Twitter datasets, the first one containing only temporal usage patterns of past hashtag assignments, while in the second one these patterns are combined with a content-based analysis of the current tweet. It is found that only the temporal effects play an important role for both individual and social hashtag reuse.

Aswani et al. [21] analysed more than a million tweets, in order to explore the factors of rapid propagation of misinformation. It was found that the authenticity of a shared content can be determined by the tweet emotion, as well its polarity. Furthermore, it was determined that such tweets contained a higher element of surprise combined with other emotions, while tweets that show case-neutral content often lack the possibilities of virality when it comes to misinformation. Last, the propagation of misinformation is catalysed by network attributes, including community and topological and centrality analysis.

Poulopoulos et al. [22] present an approach that enriches cultural venues with relative ongoing discussions on SNs. More specifically it extracts trending topics that can be semantically related with the content of a cultural institute and examine how a venue can benefit by exploiting these relative discussions. The aforementioned approach has been experimentally tested by analysing the case of Twitter in Greek language.

However, none of those works present an application that takes advantage of the most popular twitter topics per country level, in order to locate objects in a database that are related to these topics, by matching the texts of the tweets that correspond to the popular topics with the corresponding texts of the database objects (titles, descriptions, etc.) and suggests the best matches to the user of the application for further control and possible utilization.

III. Application requirements, analysis and design

In this section, we present the requirements of the application and its design. The requirements of the applications were mainly gathered from the experience gained in the CrossCult project, where the potential of social media to stimulate history reflection in cultural heritage was surveyed [22,23,24].

The main requirements for the application are the following:

- The application should be able to edit the 10 most popular Twitter trending topics.
- Since popular trending topics contain a spatial reference (different trending topics are popular in different regions), the application must allow the users to select the region from which the trending topics will be sourced. In the present work we considered the following locations: UK, USA, Greece, and the whole world.
- The application should be able to utilize thesauri, to find synonyms and enhance thus the efficiency of the matching process.
- The application should allow the consideration of items that have recently been entered in the database against which trending topics are matched.
- The application should check if some item in the database matches against any trending topic, considering synonyms in the matching process; any matches are returned to the user for further inspection.
- The application should finally allow the user to select one of the trending topics and search on the internet and get relevant results.
- Regarding non-functional requirements, emphasis should be placed on the usability and intuitiveness of the interface. A good level of performance should also be ensured.

Fig. 1 depicts the application's architectural design and the main processing workflow. The different activities in the workflow are detailed in the next section.

The data used for the application may pertain in any domain; in the experiments conducted, the data used described the content of the History Museum of the University of Athens (http://hist-museum.mm.di.uoa.gr/). The data was extracted and stored in a database, comprising of the following main entities:

- *exhibit*, comprising (among other fields) a short and a long description, as well as a creation place and one or more creators.
- *collections*, which are sets of related exhibits. Each exhibit is part of one or more collections. A collection includes a short and a long description.

- *creator*, who is a contributor to the creation of an exhibit. Each exhibit can have one or more creators.

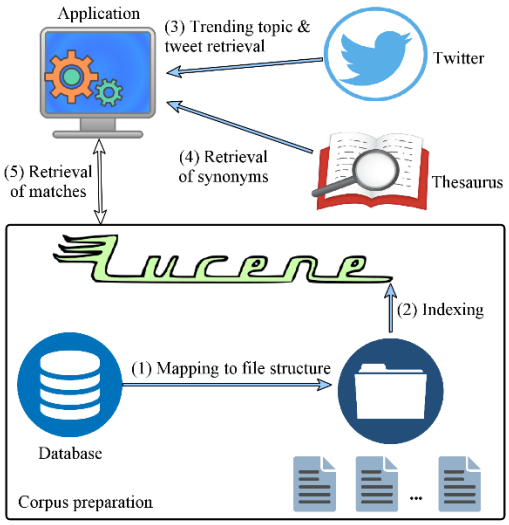

Fig. 1. The application architectural design

IV. The Application

This section elaborates on the application functionality, as implemented in the application. The different steps in the functionality workflow are detailed and illustrated through screenshots, while control flow across different functionalities is also presented.

In the application homepage (Fig 2), the user must first use the "Database indexing" functionality. In this process, each exhibit is first extracted from the database and mapped to a file in the file system, i.e. the contents of the database related to the exhibit (name and title) are retrieved and written to a file. The name of the exhibit is the id of the exhibit from the database and the contents of the file are the name and description of the exhibit. If the name and description of the exhibit are available in multiple languages, a separate file is created for each language, which is placed in a folder corresponding to the language. This mapping is done by the application, and the resulting file and directory structure is shown in Fig. 3.

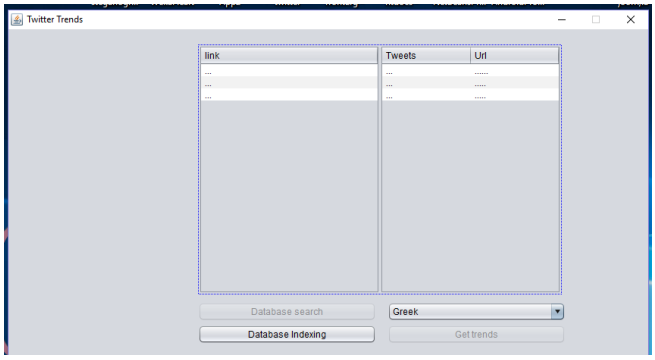

Fig. 2. The application's homepage

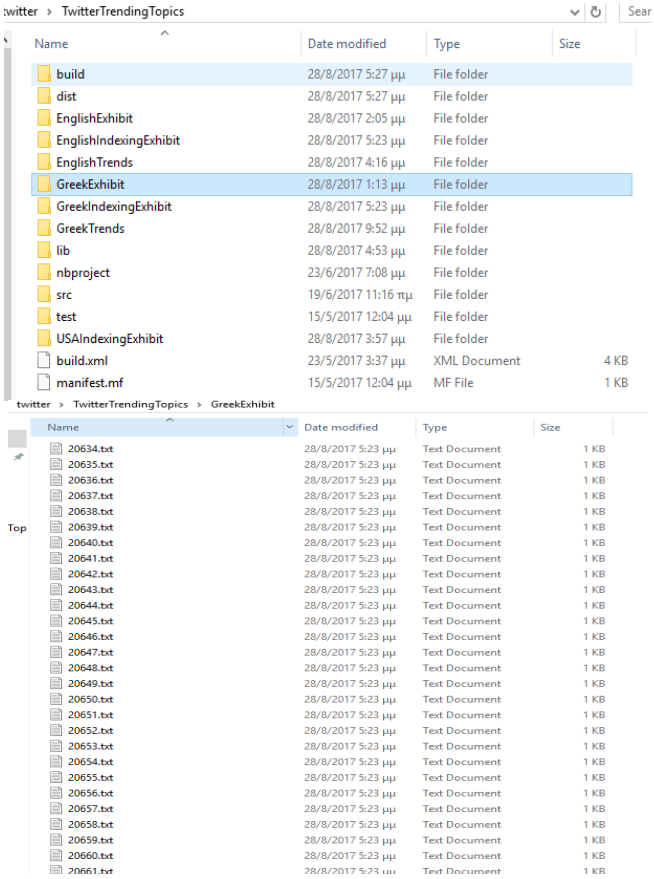

Fig. 3. The GreekExhibit and EnglishExhibit folders (top) and the files of the GreekExhibit folder, where we can see that the name of each file corresponds to the ID in the database (bottom)

Once the mapping of the entire database has concluded, the Apache Lucene search engine (https://lucene.apache.org/) undertakes the task to create the index. This process creates a distinct folder for each language where index data for the files of the language-specific subfolder are written (e.g. *GreekIndexingExhibit* and *EnglishIndexingExhibit*; c.f. Fig. 4). Index data will be used later to calculate the relevance score between the submitted queries and individual items. Note that this procedure should be repeated when any changes made to the database need to be reflected to the search corpus.

At this stage the corpus has been prepared and indexed and the user may proceed to retrieve trending topics and match them against the database. To this end, the user firstly selects the country from which the ten most popular trending topics will be retrieved. Depending on the user choice, the corresponding results from the designated location (specific country or the world) will be retrieved and presented. In order to retrieve the most popular trending topics, a request is submitted to the Twitter API; the retrieved results are displayed in the right-hand side pane (c.f. Fig. 5). Subsequently, the user may select the trending topics that will be matched against database items, possibly excluding irrelevant or inappropriate ones.

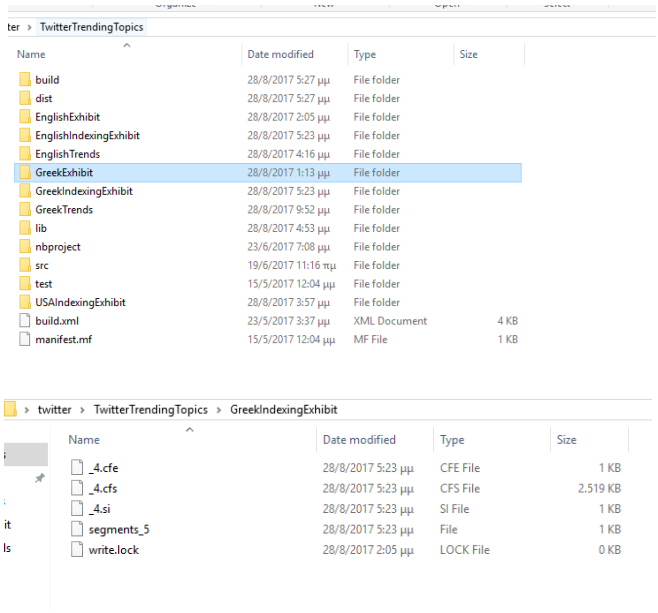

Fig. 4. Folders country_IndexingExhibit (top) and index data files created by Apache Lucene (bottom)

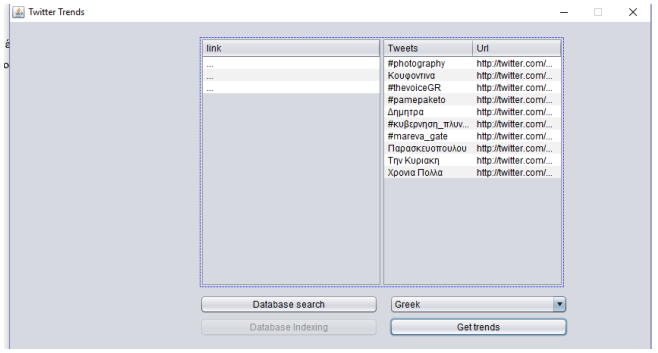

Fig. 5. The application's homepage listing the retrieved trending topics

The search process proceeds as follows: firstly, a query is made to the Twitter API, to retrieve tweets that are related to the selected trending topic. As a response, Twitter returns a list of the top tweets that refer to the particular topic. The result list is then pruned to retain only up to 15 tweets, and then the text of these tweets is processed. Links, emoticons and special characters are removed, because they create noise and lead to the production of a suboptimal score by Apache Lucene [25, 26] within the search procedure. The bodies of the tweets are merged into one file and synonyms for the terms within this file are retrieved from a thesaurus and added to the file. The addition of the thesauri data enables the calculation of a more accurate matching score by Apache Lucene. Finally, the text of the popular trending topic itself is added to the file but without the hashtag (if present), and the Lucene engine is asked to match the content of the newly created file against its corpus. For each item match (corresponding to a file) identified by Lucene, a matching score is obtained indicating the matching degree. Only matches whose score surpasses a user-specified threshold are displayed to the user; if some trend produces no matches surpassing the threshold, an appropriate error message is displayed.

For the matches that qualify to be shown to the user, it is possible to retrieve the complete item data from the database, since by virtue of the protocol followed for the construction of the corpus, where each exhibit is mapped to a file named after the exhibit's id; and since the name of the file is returned by Lucene, it is possible to directly extract the item id and use it to query the database.

Fig. 6 depicts indicative search results, illustrating the correspondence of files, matching scores and the presence of the item id within the filename. Note that this form of results is only used internally; users are presented with a more user-friendly view.

```
Total Results :: 0
Total Results :: 2856
Path : GreekExhibit\20965.txt, Score : 46.84993
Path : GreekExhibit\20964.txt, Score : 41.643505
Path : GreekExhibit\20997.txt, Score : 41.600517
Path : GreekExhibit\21997.txt, Score : 41.22058
Path : GreekExhibit\21242.txt, Score : 40.283077
Path : GreekExhibit\21060.txt, Score : 39.124447
Path : GreekExhibit\21546.txt, Score : 38.888054
Path : GreekExhibit\22510.txt, Score : 38.796173
Path : GreekExhibit\23778.txt, Score : 38.796173
Path : GreekExhibit\20951.txt, Score : 37.806004
Total Results :: 62
Path : GreekExhibit\21388.txt, Score : 19.812954
Path : GreekExhibit\21438.txt, Score : 17.049006
Path : GreekExhibit\21455.txt, Score : 16.586771
Path : GreekExhibit\22169.txt, Score : 16.586771
Path : GreekExhibit\21428.txt, Score : 16.419104
Path : GreekExhibit\20828.txt, Score : 14.920809
Path : GreekExhibit\22459.txt, Score : 14.920809
Path : GreekExhibit\23727.txt, Score : 14.920809
Path : GreekExhibit\22570.txt, Score : 14.535269
Path : GreekExhibit\21429.txt, Score : 13.907235
Total Results :: 2872
Path : GreekExhibit\21312.txt, Score : 95.26228
Connecting to database...
```

Fig. 6. Scores produced by the Apache Lucene

The user interface additionally provides the user with the functionality to perform a Google search for some trending topic, in order to obtain more information on it. This is accomplished by simply selecting the trending topic on the right pane, and then the application queries Google and presents the results in the left pane as links (Fig. 7). Clicking on a link will result to opening the default browser to display the link contents.

V. Conclusion And Future Work

The advent of SN and the abundance of content they offer, open new opportunities for the realization of value added services over the SN data. In this paper, we have presented the requirements, design and implementation of an application that users Twitter data to assist database curators to match their content against current trends and therefore identify suitable items to be used for promoting their content and/or organization.

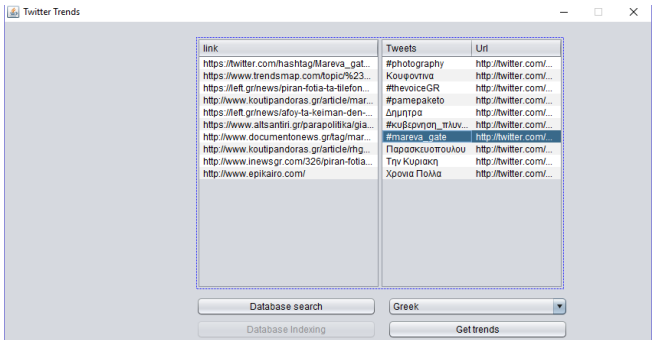

Fig. 7. Selecting the trending topic

More specifically, in this paper, we have presented the requirements, design and implementation of application that utilizes the 10 most popular Twitter trending topics, in order to identify items in a given database that are related to these topics. The objects are identified by matching the texts of the top-rated tweets that correspond to each trending topics with the corresponding texts of the database items (titles and descriptions). In order to enrich the search, synonyms are used from a thesaurus of words. Finally, the top matches exceeding a user-specified threshold are suggested to the user of the application for examination and -potentially- further utilization.

Our future work will focus on incorporating other SNs in order to combine information, divide into categories, compare and combine the most popular trending topics [27-29]. Furthermore, we will conduct experiments to test how the removal of remove stop words (e.g. “are”, “is”, “the” and other similar words) affects matching score calculation [30-32]. Finally, future work includes combination with personalization techniques, so as to adapt both the result returned to the user and the ranking of this result based on each user’s either explicitly or implicitly gathered attributes [33-36].

VI. References

[1] D. Floegel and K.L. Costello, “Entertainment media and the information practices of queer individuals,” Library & Information Science Research, vol. 41(1), pp. 31-38, 2019.

[2] P. Pourghomi, F. Safieddine, W. Masri and M. Dordevic, “How to stop spread of misinformation on social media: Facebook plans vs. right-click authenticate approach,”,. Proceedings of the 2017 IEEE International Conference on Engineering & MIS (ICEMIS), pp. 1-8, 2017.

[3] T.T.Goh, Z. Xin and D. Jin, “Habit formation in social media consumption: a case of political engagement. Behaviour & Information Technology,” vol. 38(3), pp. 273-288, 2019.

[4] R. Newman, V. Chang, R.J. Walters and G.B. Wills “Web 2.0-The past and the future,” International Journal of Information Management, vol. 36(4), pp. 591-598, 2016.

[5] E. Karahan and G. Roehrig, “Use of web 2.0 technologies to enhance learning experiences in alternative school settings,” International Journal of Education in Mathematics, Science and Technology, vol. 4(4), pp. 272-283, 2016.

[6] M. Aivazoglou, A.O. Roussos, D. Margaris, C. Vassilakis, S. Ioannidis, J. Polakis and D. Spiliotopoulos, “A fine-grained social network recommender system,” Social Network Analysis and Mining, vol. 10(1), 8, 2020.

[7] S. Martinčić-Ipšić, E. Močibob and M. Perc, “Link prediction on Twitter,” PloS one, vol.12(7), e0181079, 2017.

[8] G. Enli and C.A. Simonsen, “Social media logic’meets professional norms: Twitter hashtags usage by journalists and politicians,” Information, Communication & Society, vol. 21(8), pp. 1081-1096, 2018.

[9] Y. Kim, E. Hwang and S. Rho, “Twitter news-in-education platform for social, collaborative, and flipped learning,” The Journal of Supercomputing, vol. 74, pp. 3564-3582, 2018.

[10] D. Kotzias, T. Lappas, and D. Gunopulos, “Home is where your friends are: Utilizing the social graph to locate twitter users in a city,” Information Systems, vol. 57, pp. 77-87, 2016.

[11] Z. Deng, M. Yan, J. Sang and C. Xu, “Twitter is Faster: Personalized Time-Aware Video Recommendation from Twitter to YouTube,” ACM Transactions on Multimedia Computing, Communications, and Applications vol. 11(2), Article 31, 2015.

[12] T. Yigitcanlar, N. Kankanamge and K. Vella, “How Are Smart City Concepts and Technologies Perceived and Utilized? A Systematic Geo-Twitter Analysis of Smart Cities in Australia,” Journal of Urban Technology, pp. 1-20, 2020.

[13] U. Yaqub, S.A. Chun, V. Atluri and J. Vaidya, “Analysis of political discourse on twitter in the context of the 2016 US presidential elections,” Government Information Quarterly, vol. 34(4), pp. 613-626, 2017.

[14] A. Zubiaga, D. Spina, R. Martínez, and V. Fresno, “Real-time classification of twitter trends,” Journal of the Association for Information Science and Technology, vol. 66(3), pp. 462-473, 2015.

[15] B.K. Bao, C. Xu, W. Min and M.S. Hossain, “Cross-Platform Emerging Topic Detection and Elaboration from Multimedia Streams,” ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 11(4), Article 54, 2015.

[16] M.C. Yang and H.C. Rim, “Identifying interesting Twitter contents using topical analysis,” Expert Systems with Applications, vol. 41(9), pp. 4330-4336, 2014.

[17] R. Marchetti and D. Ceccobelli, “Twitter and television in a hybrid media system: the 2013 Italian election campaign,” Journalism Practice, vol. 10(5), pp. 626-644, 2016.

[18] H. Cai, Y. Yang, X. Li and Z. Huang. “What are Popular: Exploring Twitter Features for Event Detection, Tracking and Visualization,” Proceedings of the 23rd ACM international conference on Multimedia, pp. 89-98, 2015.

[19] P. Yang, J. Liu, J. Qi, Y. Yang, X. Wang and Z. Lv, “Comparison and Modelling of Country-level Microblog User and Activity in Cyber-physical-social Systems Using Weibo and Twitter Data,” ACM Transactions on Intelligent Systems and Technology, vol. 10(6), Article 65, 2019.

[20] D. Kowald, S.C. Pujari and E. Lex, “Temporal Effects on Hashtag Reuse in Twitter: A Cognitive-Inspired Hashtag Recommendation Approach,” Proceedings of the 26th International Conference on World Wide Web (WWW '17), pp. 1401-1410, 2017.

[21] R. Aswani, A.K. Kar and P.V. Ilavarasan, “Experience: Managing Misinformation in Social Media-Insights for Policymakers from Twitter Analytics,” Journal of Data and Information Quality, vol. 12(1), Article 6, 2019.

[22] V. Poulopoulos, C. Vassilakis, M. Wallace, A. Antoniou and G. Lepouras, “The Effect of Social Media Trending Topics Related to Cultural Venues’ Content,” Proceedings of the 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), 2018, pp. 7-12, 2018.

[23] Bampatzia, S., Antoniou, A., Lepouras, G., Vassilakis, C. and Wallace, M., “Using social media to stimulate history reflection in cultural heritage,”. Proceedings of the 11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), 2016.

[24] Vassilakis, C., Antoniou, A., Lepouras, G., Wallace, M., Lykourentzou, I. and Naudet, Y., “Interconnecting Objects, Visitors, Sites and (Hi)Stories Across Cultural and Historical Concepts: The CrossCult Project”. In Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection, 2016, pp. 501-510.

[25] E.L. Lydia, S. Satyanarayan, K.V. Kumar and D. Ramya, “Indexing documents with reliable indexing techniques using Apache Lucene in Hadoop,” International Journal of Intelligent Enterprise, vol. 7(1-3), pp. 203-214, 2020.

[26] S. Lakhara, and N. Mishra, “Desktop full-text searching based on Lucene: A review,” Proceedings of the 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), pp. 2434-2438, 2017.

[27] P. Munjal, M. Narula, S. Kumar and H. Banati, “Twitter sentiments based suggestive framework to predict trends,” Journal of Statistics and Management Systems, vol. 21(4), 685-693, 2018.

[28] W. Wang, I. Hernandez, D.A. Newman, J. He and J. Bian, “Twitter analysis: Studying US weekly trends in work stress and emotion,” Applied Psychology, vol. 65(2), pp. 355-378, 2016.

[29] A. Deshwal and S.K. Sharma, “Twitter sentiment analysis using various classification algorithms,” Proceedings of the 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), pp. 251-257, 2016.

[30] D. Margaris, C. Vassilakis and D. Spiliotopoulos, “Handling uncertainty in social media textual information for improving venue recommendation formulation quality in social networks,” Social Network Analysis and Mining, vol. 9(1), 64, 2019

[31] R. Urena, G. Kou, Y. Dong, F. Chiclana and E. Herrera-Viedma, “A review on trust propagation and opinion dynamics in social networks and group decision making frameworks,” Information Sciences, vol. 478, pp. 461-475, 2019.

[32] R. Ureñaa, G. Koub, Y. Dongc, F. Chiclanaad, E. Herrera-Viedmaden, “A review on trust propagation and opinion dynamics in social networks and group decision making frameworks,” Information Sciences, vol. 478, pp. 461-475, 2019.

[33] D. Margaris, C. Vassilakis and D. Spiliotopoulos, “What makes a review a reliable rating in recommender systems?,” Information Processing and Management, vol. 57(6), Article 102304, 2020.

[34] D. Margaris, A. Kobusinska, D. Spiliotopoulos and C. Vassilakis, “An Adaptive Social Network-Aware Collaborative Filtering Algorithm for Improved Rating Prediction Accuracy.” Access, vol. 8(1), pp. 68301-68310, 2020.

[35] D. Spiliotopoulos, D. Margaris and C. Vassilakis, “Data-Assisted Persona Construction Using Social Media Data,” Big Data and Cognitive Computing, vol. 4(3), 21, 2020.

[36] V. Poulopoulos, C. Vassilakis, A. Antoniou, G. Lepouras, A. Theodoropoulos and M. Wallace, “The Personality of the Influencers, the Characteristics of Qualitative Discussions and Their Analysis for Recommendations to Cultural Institutions,” Heritage, vol. 1(2), pp. 239-253, 2018.