A crowdsourcing approach to advance collective awareness and social good practices

Ilias Dimitriadis
School of Informatics
Aristotle University, Thessaloniki, Greece, idimitriad@csd.auth.gr

Vasileios G. Psomiadis
School of Informatics
Aristotle University, Thessaloniki, Greece, vpsomiadis@csd.auth.gr

Athena Vakali
School of Informatics
Aristotle University, Thessaloniki, Greece, avakali@csd.auth.gr

ABSTRACT

Contemporary societies are challenged by many social issues, lots of which relate with environmental threats. Plastic waste's impact on societal well being, economy, and environment has already triggered the need for its revaluation and its transformation into an asset which can inspire and mobilize innovative solutions. This work envisions trusted "wisdom of the crowds" analytics and a grassroots-driven approach to facilitate citizens and stakeholders participatory and engagement practices, under an inspiring crowdsourcing framework and model. A detailed evaluation of the filtering methodology is employed. This three-layer approach for the Data Filtering enables a Crowdsourcing Component implementation to be used as a plastic-topics insightful barometer, able to detect new trends, identify interesting content, spot influential users and generally raise awareness regarding the problem of plastic overuse.

CCS CONCEPTS

• Information systems~Crowdsourcing • Information systems~RESTful web services • Computing methodologies~Topic modeling

KEYWORDS

Web 2.0 analytics, crowdsourcing tools and models, social innovation, social media innovation

ACM Reference format:

Ilias Dimitriadis, Vasileios G. Psomiadis and Athena Vakali. 2019. A crowdsourcing approach to advance collective awareness and social good practices. In *Proceedings of WI '19: IEEE/WIC/ACM International Conference on Web Intelligence (WI '19 Companion), October 14-17, 2019, Thessaloniki, Greece.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3358695.3361104

WI '19 Companion, October 14-17, 2019, Thessaloniki, Greece

ACM ISBN 978-1-4503-6988-6/19/10...$15.00
https://doi.org/10.1145/3358695.3361104

1 Introduction

Contemporary societies are challenged by many social issues, lots of which relate with environmental threats. Plastic waste poses such a huge threat to society and the environment, from overflowing landfills to the ever-growing ocean gyres since it's damaging fragile ecosystems and affecting the food chain. A huge waste of plastic resources, threats environment but also lives of organisms and ever humans. Plastics drawbacks become a large-scale problem which needs a global response. Plastics waste impact on societal wellbeing, economy, and environment has already triggered the need for a novel and innovative Plastics Economy[1] movement.

Novel solutions are now needed for enabling social innovation in revaluing plastics. This work is motivated by the capacity of social Web technologies to make social change and innovation under a plastics-as-an-asset paradigm. Since social Web technologies have evolved with capturing the so called "wisdom of the crowd", in the proposed work Citizens to Communities (C2C) associations are detected and shared by harnessing crowdsourcing via a social media observatory with favors plastics economy participatory motivation by active user's engagement and plastics social innovation co-sharing. This work envisions trusted "wisdom of the crowds" analytics and a grassroots-driven approach to facilitate citizens and stakeholders participatory and engagement practices, under an inspiring crowdsourcing framework and model.

Data produced in abundance in online social media sources offer a fertile ground for harvesting users' feedback and views concerning their reactions and opinions on evolving plastic recycling and reusing key issues. The exploitation of such resources can lead to the delivery of innovative and more human-centered services for the public, leveraging the collective intelligence of the crowd. Social media data mining as a collective intelligence approach, has tremendously evolved during the last decade. It involves extracting latent information and insights from the: (i) unstructured social media User Generated Content (UGC) and (ii) users' interaction in social media, such as communities or groups of similar behaviour. These facts (massive data sizes and low-quality content, e.g. poorly formatted, short textual entities, etc.) pose significant challenges to typical data mining algorithms.

By a proper design and implementation of a social media ideas sharing terminology and an application release with crowdsourcing mechanisms plastics reuse actions can be

[1] The New Plastics Economy: Catalysing action; World Economic Forum, Jan 2017. https://www.weforum.org/reports/the-new-plastics-economy-catalysing-action

accelerated and time for plastics reuse designing decisions may be improved. Since social innovation is targeted, visualization of social media wisdom of the crowds is proposed to enable plastics reuse, recycling, and lifestyle perception for the wider public under the free form of expression in popular social media interactions and content threads. User friendly data visualizations are used (such as tag clouds, charts, etc.) to allow users to gain an evidence of the social data correlations providing zoom and filtering capabilities, topic-level details etc.

The proposed work targets at automatically collecting plastic-related content which will be visible over an application's layer of a crowdsourced topics observatory. The crowdsourced data will be monitored and analyzed at a middleware layer along with further classification approaches to maintain a plastics topics hierarchy which will enable topics and open data classification and standards alignment. The proposed crowdsourcing observatory allows users to understand plastics thematic relationships and impacts their awareness and plastics reuse motivation. Furthermore, the proposed crowdsourcing approach views and correlates plastics crowdsourced analytics in order to open knowledge and data sharing. Twitter as a most popular social media platform offers an appropriate testbed for crowdsourcing and additional social media sources can be exploited (such as Pinterest) which are most suitable for spotting and further exploiting plastic reuse and revalue ideas.

In this context, crowdsourcing acts as a plastic-topics insightful barometer, opening knowledge of people's views and contributing to societal change and innovation. Social media data are collected and filtered based on cutting edge crowdsourcing methods and algorithms and appropriate.

The remaining of this paper is organized as follows: Section 2 discusses the preliminaries and highlights the relevant related work with emphasis on social media data processing and analyzing. Section 3 contains the methodology of the proposed approach with emphasis on the multi-type filtering, while Section 4 discusses the evaluations of the different filtering types. Finally, Section 5 includes the conclusions of the article and future potentials.

2 Preliminaries and Related Work

A preliminary crowdsourcing analysis has been conducted for identifying the types and volume of plastic-related content shared in public social media sources and some insightful details are summarized next. Focusing on the Twitter platform, content contributed by users has been collected based on specific terms/phrases related to either plastic waste in general (e.g. plastic rubbish, plastic waste, and plastic pollution) or to specific plastics' substances (e.g. polystyrene, polyethylene, and polypropylene). Twitter content has been collected for a 3-weeks period (around 54,386 posts created by 8,100 unique users).

Getting immediate insight, out of a simple term-based searching on the multi-threaded Twitter's content production, offers an initial input for plastic-related crowd's standing, since the relevant content may span several topic areas and be of diverse quality. To identify qualitative and interesting and most relevant to PTwist's targeted issues content, we've applied simple techniques for assessing the users' reputation and influence within the social networks, as well as the content's popularity among social network users. Within such posts, we've identified interesting content such as articles for new campaigns raising awareness about plastics in oceans and overuse of plastics in daily life, accompanied by photos illustrative of the problem. Interestingly, in the collected posts we've discovered almost 40 occurrences of the word "reuse" (not included in the tracked terms list), which is indication that ideas and articles for plastic reuse are detected in Twitter (top-40 terms and hashtags are depicted in Figure 1).

Figure 1: Tag clouds generated based on data collection from Twitter: (a) and (c) depict the most frequent terms and hashtags, respectively; (b) and (d) similar illustrations.

Existing relevant approaches mainly focus on the classification of Twitter content into general categories such as e.g. news, memes, ongoing events, and commemoratives [1], or substance, status style, social [2]. Mostly NLP approaches for assessing the quality of the underlying texts have been utilized, in terms of syntax and other structural features [3], as well as the expressed sentiment which may be more engaging for readers and can indicate their engagement capacity [4], [5]. Moreover, the differences in privacy perceptions between users of different OSNs were studied and both Facebook and MySpace OSNs are concerned about privacy, yet this does not prevent them from sharing information online [6]. It is believed that often the perceived benefits of users outweigh the risk of personal privacy [7], [8]. However, in several tasks a more fine-grained and possible context-specific (such as in relation to plastic waste in PTwist) categorization is needed. Such an approach requires the utilization of external domain specific knowledge provided by online sources (such as DBpedia and Wikipedia) or ontologies. In [9] there is a relevant effort that utilized a Wikipedia-based global "real-time" knowledge base to automatically classify streaming data from Twitter. Therefore, social media topics related either to the company, the products and/or product types can be linked to these open classification standards in order to provide an improved semantics-aware repository for plastics-related terms. Since social media analytics will dynamically evolve, these terms could be updated by exploratory analysis occurring at the social media analytics part of the project.

3 A social media innovation methodology

The proposed work leverages on the plastics related terms crowdsourcing by building on advanced data mining approaches, to establish a robust social media content processing pipeline

that will be able to identify qualitative content, and further classify topics in a set of context-specific categories. Analysis of the metadata and information around the collected UGC via topic modelling approaches is accompanied by textual-based classification approaches to extract context along with features that will enable categorization in appropriate, predefined classes (such as *reuse ideas*, *initiatives*, etc.). An advanced machine learning methodology is used to provide a much richer representation of text for more robust and accurate social data analysis. The proposed machine learning approach exploits quantitative measures (number of comments, 'likes', etc.), as well as qualitative analytics building on analytics approaches [10] and other plastics contextual feedback such in case of recycling [11] and green social behaviour [12]. In the proposed approach, a methodology for automatically collecting and identifying content relevant to specific topic areas (in this case *plastic-as-an asset* and relevant mitigation approaches) is used inline with a preliminary analysis (outlined above in Section 2). The proposed methodology steps include:

- acquiring coarse-grained streams of content (such as posts and annotated multimedia and text content, as well as user reactions to them) via appropriate data collection mechanisms;
- the identification and characterization of qualitative relevant information and the fine-tuning of the data collection parameters;
- exploiting and advancing the state of the art in the areas of qualitative crowdsourcing and on dynamic topic categorization.

The proposed data filtering methodology during the process of the Crowdsourced Data Collection and Analysis involves three main filtering types:

1. Primitive Filtering, to identify the Social media sources which will be used for the collection of data;
2. Dynamic content filtering, to identify terms related to plastic and plastics reuse;
3. Data streams & User filtering, for
    a. Identifying and detecting influential – expert - users regarding plastic re-use thematology.
    b. Filtering incoming data in order to produce high quality content.
    c. Updating the keywords used to collect data.

Apart from the whole filtering process, a detailed evaluation of the filtering methodology is employed. This three-layer approach for the Data Filtering enables a Crowdsourcing Component implementation to be used as a plastic-topics insightful barometer, able to detect new trends, identify interesting content, spot influential users and generally raise awareness regarding the problem of plastic overuse.

3.1 Primitive Filtering

Data collection from social media is rapidly getting progressively limited since many concerns are raised regarding privacy issues and the exploitation of user generated data through the OSNs while recent legislation frameworks are becoming increasingly more rigid towards this sensitive subject. Thus, very few of the popular social networks allow developers to access data, even if these are meant as public posts by their owners. The social networks that are used as sources are the following:

1. Twitter[2]: Twitter is the main data source for the crowdsourcing tool.
2. Flickr[3]: Provides access to images that have been described with certain tags relevant to plastic related terms.
3. Thingiverse[4]: Acts as the main source of the open 3D printer designs repository and is used as an interlink, where the users are able to upload their own designs.
4. Instagram[5]: Used as a secondary source, because of the access limitations to its data.

3.2 Dynamic Content Filtering

Collecting information from various sources, implies the use of specific filters so that the content of the collected data remains as close to the intended thematic as possible. For the purpose of this work the primary goal was the identification and detection of simple and sophisticated content about plastic pollution, reuse and recycling along with their correlations, trends and phenomena reflected in the society.

To identify content in Twitter a group of human experts provided an extensive list of specific keywords covering a very wide scope of the plastic thematic. These keywords were then classified using a taxonomy provided by the same expert group, forming an initial set of different categories and topics that include:

- General terms related to plastic;
- Plastic innovations;
- Plastic machines (for re-valuating plastic waste);
- Products made of plastic);
- Plastic processes (for re-valuating plastic waste);
- Plastic pollution.

Along with English this list was further translated to German, Dutch and Greek in order to capture trends and interesting topics in local level. Twitter's API was queried for all tweets that contain any on the keywords in the four languages.

Since a significant percentage of the tweets may be noisy, spam or offensive, a profanity filter was used in order to ensure the quality of the final delivered content. In our initial evaluation of the collected data, except some swearing language, no abusive content was spotted. This is mainly due to the fact that the plastic thematic does not offer a fertile ground for adult content. However, a typical text filter has been applied to filter tweets with inappropriate words included in their text. After detecting such tweets, the next step is identifying users as possibly "malicious" and blacklisting them so they are ignored by the Twitter crawler.

Apart from collecting streaming tweets, data has been collected for a number of specific user accounts that are considered to be experts in the particular field of plastic waste reuse and recycling.

Flickr is used to deliver related image content by exploiting the intelligence that has been captured from the Twitter posts

[2] https://twitter.com
[3] https://www.flickr.com
[4] https://www.thingiverse.com
[5] https://www.instagram.com

analysis. A module has been developed based on the official Flickr API, which given a word or set of words, returns images whose description matches these keywords. In a similar fashion Thingiverse acts as the main source for public designs (under the Creative Commons license) feeding an open data repository, part of the developed crowdsourcing toolkit.

Although Instagram, a very popular photo and video-sharing social networking service owned by Facebook, has an official developers' API to collect data, it provides very limited access to them. Thus, is used only to collect posts that include a set of preselected hashtags related to specific activities.

3.3 Data Streams & User Filtering

This section describes the filtering methodology that have been implemented on the captured social media data streams and is responsible for producing the final content available to the end-user of the crowdsourcing toolkit. This content is assigned to the following sub components: Visualizations (Wordclouds, Locations, Topic Modelling); Repositories; Influential Content (Top Posts, Top URLs, Influencers); Activities.

The filtering process begins when receiving a document (tweet). If the content is profane the document is discarded. Else, it is checked against the known taxonomy and tagged with the according label. From this point a set of parallel running filtering functions, each one responsible for a specific sub-component, ensure the delivery of the data to the storage component.

3.3.1 Visualizations (Wordclouds, Locations, Topic Modelling). Wordclouds is a novel visual representation of text data where the importance of each word is depicted by increased font size and different color. This kind of representation offers users the opportunity to visually identify trending terms associated with the thematic groups that interest them. The Wordclouds filtering process begins with removing all the white space and punctuation, URLs, hashtags, numbers, mentions, stop-words, social network reserved words and emoticons present in the tweet's text to reduce noise and produce more reliable and qualitative results. Then the remaining text is tokenized and the frequency of each word is calculated for all the tweets in the collection. Finally, a JSON document is created that includes a word frequency map along with from and to timestamps.

Topic modelling is an unsupervised machine learning method used for the discovery of latent topic information from large document collections. In our approach we utilized a probabilistic topic modelling method called Latent Dirichlet Allocation (LDA) [13]. LDA uses a bag of words approach to transform user's corpus into a vector of word counts. It uses two probability values: P(word | topics) and P(topics | documents). These values are calculated based on an initial random assignment, after which they are repeated for each word in each document, to decide their topic assignment. In an iterative procedure, these probabilities are calculated multiple times, until the convergence of the algorithm. After the tweets' text is pre-processed in a similar cleanup procedure as described previously, a stemming process is applied to reduce the words in their base root. Then the training of the model occurs, while the next and final step is the evaluation of the results. A topics-to-words matrix for each collection is visualized and the end-user can examine the words with the highest weight inside a topic.

Estimating the real location of a social network user is a challenge, especially when it comes to Twitter that supports only 140 characters and contains such sparse data. Previous research has been conducted on user geolocation exploiting all sources of available information [14]. In our approach we collect tweets from users that share their exact location (100% accurate but only 1-2% of the total dataset) or their exact city (high level accuracy – low granularity) or state their location in their profile (research has shown a 50-60% accuracy on city level). Tweets that can not be associated with any geographical information are omitted. Then we aggregate the total number of tweets collected along with the geospatial information (which is grouped over city level) and the results are displayed over a dynamic map. Users are able to select a geographical region of their choice and view the popular terms and hashtags of that location.

3.3.2 Repositories. Under this section the following media streams are captured and presented to the end-users in real time utilizing the official APIs of the respective services:

- open 3D printer designs hosted in the Thingiverse platform sorted by popularity or age, featured collections and also any user designs uploaded under a specific group;
- Flickr images whose description matches the keywords that the user selects from a list of top terms and hashtags extracted from the collected Twitter posts.

3.3.3 Influential Content (Top Posts, Top URLs, Influencers). Twitter, is based on a social-networking model, in which users can choose who to follow or interact with (via retweets, mentions or replies). Based on that notion we can think of Twitter as a large graph $\boldsymbol{G(U,E)}$ where $\boldsymbol{U}$ stands for the whole set of users (nodes) and $\boldsymbol{E}$ all the possible interactions between users (edges). Our aim is to use a filtering methodology capable of detecting the users that diffuse information related to our subject more efficiently throughout the network. These users, called influencers, are expected to have established credibility in the thematic we are investigating, have access to a wide audience and share content of higher quality in comparison with other users.

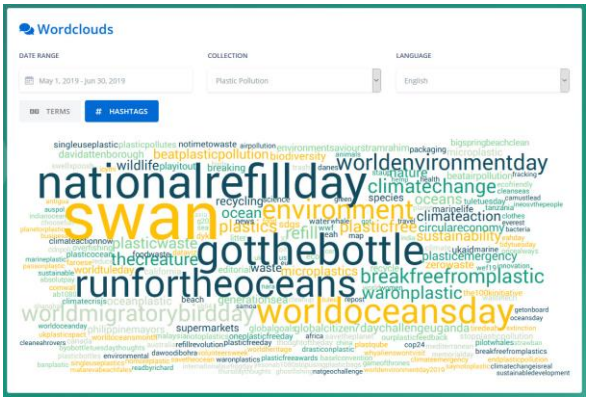

Figure 2: Tagcloud from the plastic pollution collection (May - June 2019)

In our case, the graph consists of all the users that have posted a tweet regarding plastic and have been collected by the Twitter crawler. Every retweet, mention or reply to a certain user will create an edge between the author and the accounts interacting with. Tong et al [15] presented an epidemic

immunization approach, called NetShield, which is able to detect the nodes that if isolated, one could immunize the whole network. In other words, it is able to detect those nodes that are able to spread the virus, i.e. information, throughout the whole network. The application of NetShield algorithm in our approach can be described in brief as providing a solution to the following problem: *Given a graph **G** of Twitter accounts, find the top x users that are expected to propagate a tweet or information in general better and faster.* A second algorithm, based on the betweenness centrality metric which has been proven to be quite efficient in detecting influencers [16], is also used.

The quality of social media content can be tracked by filtering out posts, based on other users' interactions. For Twitter, these interactions refer to the number of times a post has been retweeted, replied or marked as favorite. Thus, to select the top collected tweets a ranking by retweets, favorites and replies count is performed. Since Twitter posts often contain web links, these are also filtered by calculating their appearance frequency in the tweets collections of every month and then sorted in descending order to determine the top URLs.

3.3.4 Activities. In this sub component an Instagram crawler collects all posts that contain specific preselected hashtags. These hashtags are associated with events and schemes that are immediately linked with the thematic of interest. Thus, it offers handpicked content allowing users direct access to high quality information while promoting participation in focused social activities

4 Crowdsourced data collection and analysis evaluations

The initial analysis and evaluation of the collected data revealed that using only the keywords provided by experts could lead to consider a significant percentage of social media posts that were out of context. This was the result of some search terms that were quite generic and commonly used in other contexts. To overcome this the keyword 'plastic', which was the thematic basis of this application, was added to every search leading to much improved results. Figure 2, Figure 3 and Figure 4 showcase some results from the visualizations component as are presented in the developed web application.

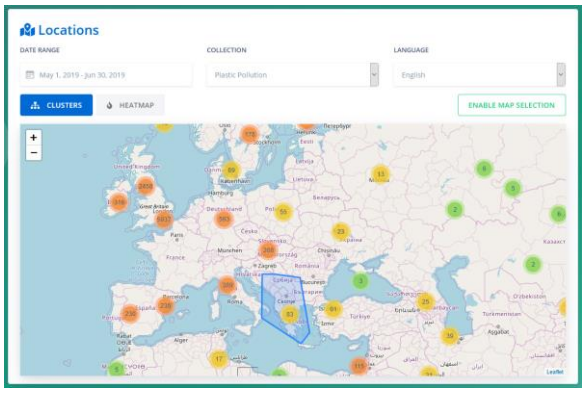

Figure 3: Dynamic map the harvested terms density per geographical region

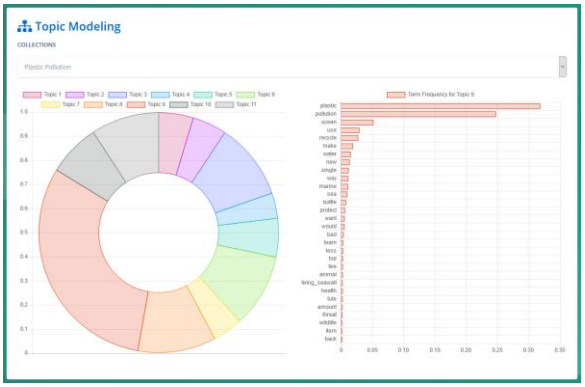

Figure 4: Identified topics from the plastic pollution collection

Another important point was to be able to capture in detail how a social media post affects the network (especially crucial when assessing the influential users). When crawling posts very close to the moment they were published, you could miss their actual impact since it might require some time before a post spreads into the network. In order to address this, an iterative process was developed that makes sure that up to one-week old posts are tracked to assess accurately the impact they generate.

Focusing on the evaluation of the Influencer filtering process a test was made to identify the top influential users in a graph that contained 629,787 nodes and 994,376 edges. The top six accounts regarding the quality of the content and their influence upon other users based on the NetShield algorithm in this dataset were: *National Geographic*, *Coral Reef Fish*, *UN Environment*, *Seal Scotland*, *Ellen MacArthur Foundation* and *BBC News*. Someone can notice that all these accounts belong to well known organizations that are very active regarding plastic and its environmental impacts, thus validating that the process used produces accurate and reliable results. Output from the filtering process for determining influential content is presented in Figure 5, Figure 6 and Figure 7.

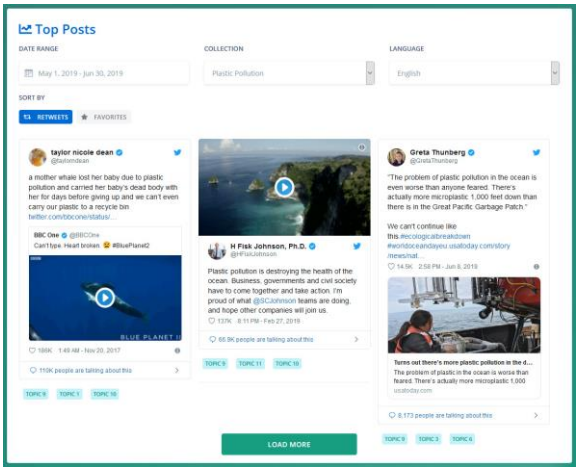

Figure 5: Top posts by retweets in the plastic pollution collection (May - June 2019)

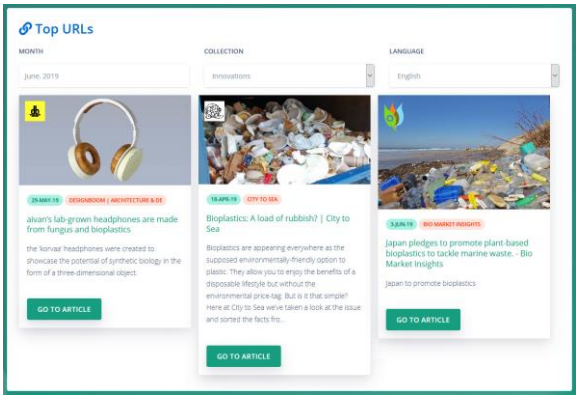

Figure 6: Top URLs in the plastic innovations collection (June 2019)

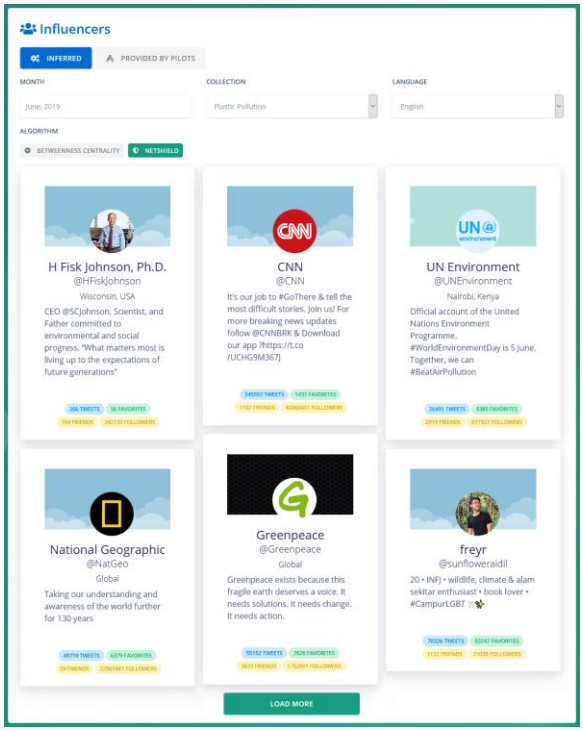

Figure 7: Top influencers regarding plastic pollution by NetShield (June 2019)

The proposed crowdsourcing approach offers an exploitable implementation since it is generically designed under a data scheme capable of integrating social media data streams with plastics classification terms hierarchies. Results of crowdsourcing are visible to the open public by an applications level observatory interface. This open Crowdsourcing API has been developed with Flask, a Python framework for building RESTful APIs. Under this interoperability scheme the research community can integrate open crowdsourced data and wisdom of the crowd's knowledge to their current research and experimentation, especially in social media analytics. At the same time, authorities can gain insights of most trendy plastics relevant topics discussed in social media. This insight impacts decision making with respect to crowd's pulse detection. The following Figure 8 and Figure 9, contain examples of related external content that is discovered and presented to the user in real time.

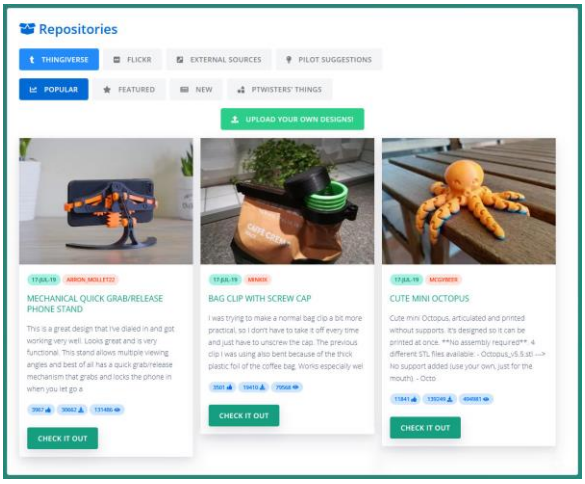

Figure 8: Popular open 3d printer designs from Thingiverse

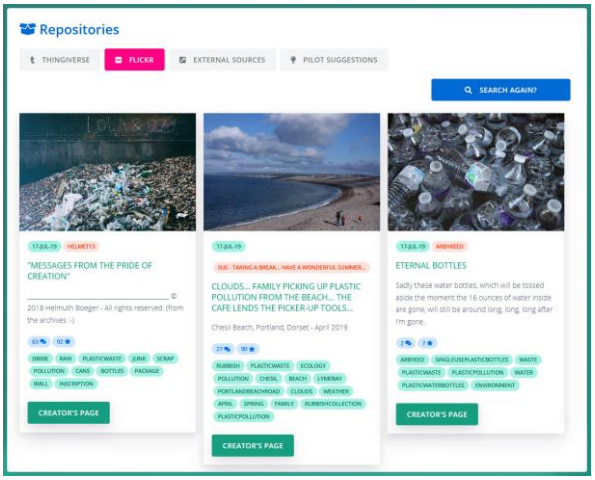

Figure 9: Photos related with the "PLASTICWASTE" term fetched from Flickr

Figure 10 contains social media posts from Instagram collected using a pre-selected hashtag, connected to a specific activity regarding plastic waste revaluation, that was chosen to spread word in social media.

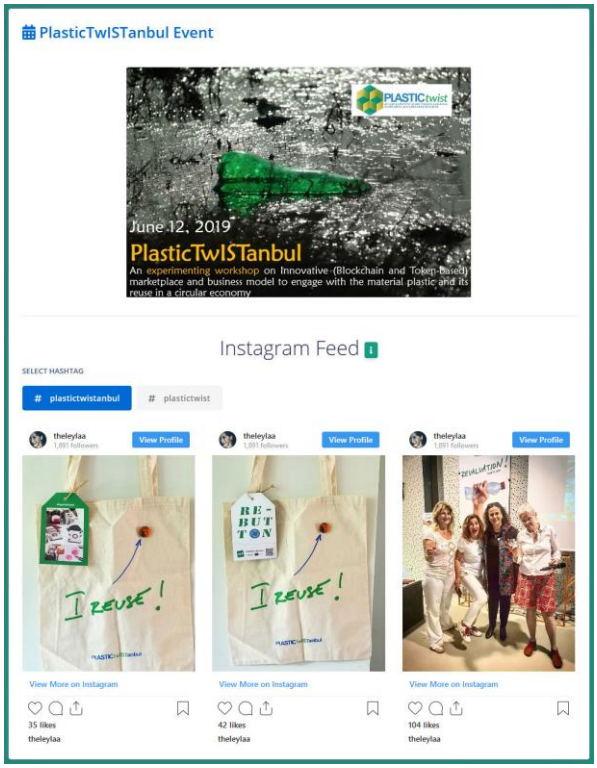

Figure 10: Social media posts of a specific event fetched from Instagram

5 Conclusions and Future Work

This work has addressed the potential of crowdsourcing and social media analysis for cases of societal and environmental threats, such as in the case of plastics revaluing. The publicly-available developed components along with the open API managed to serve a tri-fold purpose successfully:

- reflect public opinion and raise awareness around a critical issue that concerns today's society;
- inform and educate on the thematic of plastic waste pollution and reuse; and
- disseminate novel solutions and provide inspiration for further grass-rooted innovation.

The followed approach, has focused on a multi-type filtering process which has been used during the data collection and analysis phase. Extracting high quality content from the users that have been identified as influencers by our system and use it to train an LDA model is a future approach, which can be used to classify other users, and extract topics using topic modelling per location. An iterative methodology can also be applied in the future work and it can be built upon the intelligence extracted by the already available high-quality content (top tweets – top URLs) to identify new trends and dynamically update the keywords used to track tweets of specific content.

ACKNOWLEDGMENTS

This work has been supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 780121.

REFERENCES

[1] A. Zubiaga, D. Spina, R. Martínez, V. Fresno (2015). Real-Time Classification of Twitter Trends. Journal of the Association for Information Science and Technology, 66(3), 462–473.
[2] D. Ramage, S. Dumais, D. Liebling (2010). Characterizing microblogs with topic models. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.
[3] Ya. Duan, Z. Chen, F. Wei, M. Zhou, H.-Y. Shum (2012). Twitter Topic Summarization by Ranking Tweets using Social Influence and Content Quality. Proceedings of COLING 2012: 763-780.
[4] T.-A. Hoang, W.W. Cohen, E.-P. Lim, D. Pierce, and D.P. Redlawsk (2013). Politics, sharing and emotion in microblogs. In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13). ACM, New York, NY, USA, 282-289.
[5] A. Kanavos, I. Perikos, P. Vikatos, I. Hatzilygeroudis, C. Makris and A. Tsakalidis (2014). Modeling ReTweet Diffusion using Emotional Content, IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI 2014), Rodos, Greece.
[6] J. Zhang (2015). Voluntary information disclosure on social media. Decision Support Systems, 73, 28–36.
[7] Y. Li, A.F. Smeaton (2014). From Smart Cities to Smart Neighborhoods: Detecting Local Events from Social Media. In: Information Access in Smart Cities (i-ASC) 2014 Workshop in conjunction with ECIR 2014, Amsterdam, Netherlands.
[8] B. Debatin, J.P. Lovejoy, A.-K. Horn, M.A., B.N. Hughes (2009). Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. Journal of Computer-Mediated Communication, 15(1), 83–108.
[9] A. Gattani, D.S. Lamba, N. Garera, M. Tiwari, X. Chai, S. Das, S. Subramaniam, A. Rajaraman, V. Harinarayan, and A. Doan (2013). Entity extraction, linking, classification, and tagging for social media: a wikipedia-based approach. Proc. VLDB 6(11), 1126–1137.
[10] D. Chatzakou, N. Passalis, A. Vakali (2015). Multispot: Spotting sentiments with semantic aware multilevel cascaded analysis. International Conference on Big Data Analytics and Knowledge Discovery. Springer International Publishing, 337–350.
[11] M. Dupré, S. Meineri (2016). Increasing recycling through displaying feedback and social comparative feedback. Journal of Environmental Psychology, 48, 101–107.
[12] A. Biswas, R. Mousumi (2016). Impact of Social Media Usage Factors on Green Choice Behavior Based on Technology Acceptance Model. Journal of Advanced Management Science, 4(2), 92–97.
[13] D. M. Blei, A. Y. Ng and M. I. Jordan (2003). Latent dirichlet allocation. Journal of Machine Learning Research. 3, 993–1022.
[14] D. Jurgens, T. Finethy, J. McCorriston, Y. T. Xu and D. Ruths (2015). Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice. ICWSM, 15, 188–197.
[15] H. Tong, B. A. Prakash, C. Tsourakakis, T. Eliassi-Rad, C. Faloutsos and D. H. Chau (2010). On the vulnerability of large graphs. In Proceedings - IEEE International Conference on Data Mining, ICDM, 1091-1096.
[16] M.Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H.A. Makse (2010). Identification of influential spreaders in complex networks. Nature Physics, 6(11), 888–893.