Journal of Environmental Protection and Ecology 16, No 4, 1344–1352 (2015)

Marine ecology

USING GEOSOCIAL ANALYSIS FOR REAL-TIME MONITORING THE MARINE ENVIRONMENTS

V. JOVANOVIC[a]*, D. VUKELIC[b]

[a]*Singidunum University, Belgrade, Serbia*
E-mail: *vjovanovic@singidunum.ac.rs*
[b]*Retail Management Solutions (RMS) Ltd., Tower Road, Sliema, Malta*
E-mail: *dvukelic68@gmail.com*

Abstract. Web 2.0 is bringing the ability to share information online and create the different forms of the collaboration. Replacing a static nature with the dynamics and real-time content, the social nature of the World Wide Web changed. Accordingly, social media feeds are becoming increasingly geosocial in the sense that they often have a substantial geographic content. This information contributes additional feature to social media (location) and provides additional context to analyse these data (topics). The geographic content of social media content represents a new type of geographic information. Due to the particularities of the analysis that these data support, the model presented use a hybrid mix of spatial and social analysis with the central role of GIS, with possibility of applying in the real-time monitoring of the marine environments. It expounds the enquiry carried out for the occurrence of Jellyfish in coastal and marine waters of Malta through the use of GIS and social analysis. Appliances of both tools should ensure that the results obtained will have a major impact on environmental monitoring outcomes. Such outcomes would provide affirmation of the efficiency the technologies posses as being a reliable means for the overcoming of marine environmental monitoring issues.

Keywords: GIS, geosocial, modelling, marine, environments.

AIMS AND BACKGROUND

Marine environment monitoring is an indispensable number of systematic process and activities which take place to observe and characterise the contemporary affects of sea terrestrials on the environment and the natural order. It has attracted an increasingly great deal of research and development attention due to its essentiality and issues which have not yet been resolved. During the past decade, various marine environment monitoring systems have been developed. However, most of the environment monitoring systems still continue to possess two disadvantageous aspects, which are expensiveness and time-consumption and has low responses both in time and space, with very limited or sometimes no presence to the general public.

---

* For correspondence.

With the appearance of Web 2.0, ubiquitous computing and corresponding technological advancements social networks have become massively popular, especially during the last decade. The term social networks refers to a different spectrum of digital interaction and information platforms. It includes blogs (e.g. Twitter, Blogger, WordPress,), digital social services (e.g. Facebook, Google+) and multimedia content sharing services (YouTube). Although they are distinguishable, these social media services share the common goal of enabling the general public to contribute, disseminate and exchange information[1]. At the same time, social media content is rapidly increasing. Facebook announced in 2012 that its system deals with petabyte scale data as it processes 2.5 billion content elements and over 500 TB of data daily[2].

The geographic content of social media content represents a new type of geographic information. It that can be 'collected' from social media feeds and can be referred to as ambient geographic information – AGI (Ref. 3). Although, it might be related with the originator of feed, most of the time it is embedded in the content of these feeds, often across the content of numerous entries rather than within a single one and has to be somehow extracted.

In this paper, the model for real-time monitoring of marine environments will be outlined using geosocial analysis. Based on this model the software application is developed and tested in the pilot project.

MODEL DESCRIPTION

Currently, social media services offer a wide range of platforms using various technologies. As a result, their content tends to be very diverse both in the content itself – ranging from text to photos and videos – and in form, ranging from structured content to semi- or non-structured content. In addition, the form of raw social media data tends to be unstructured or ill-defined, making valuable knowledge hidden and limiting the capability to process it through automation[4].

Geosocial data are also differentiated from 'traditional' geospatial information due to their complexity. Specifically, they are predominantly linked information; links are created among users to establish a social network and among words to define a 'topic' that is communicated through pieces of information. It is analogous to information aggregation in a Geosensor network[5], where each sensor contributes a piece of information, but it is through across an aggregation that the observed event is revealed in all its complexity. When it comes to social media, people act as sensors too, reporting their observations in the form of multimedia feeds, and the challenge is to compose these fragmented contributions into a bigger picture, overcoming the limitations of individual perception. However, the Social network feeds allow for the first time to explore the physical presence of people together with their online activities, enabling us to link the cyber and physical spaces on a massive scale.

In the following model (Fig. 1) GIS has the central role as it uses spatio-temporal (space-time) location as the key index variable. GIS software is now usually marketed as combination of various interoperable applications and API (Ref. 6). The manipulation of data can be made through: browsing, altering data structure, querying and analysis[7].

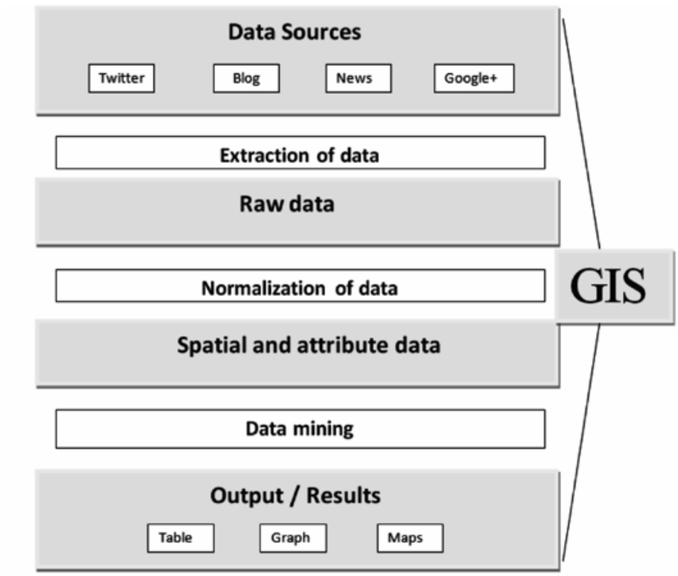

Fig. 1. Model description

Data source serves as a directory of the different source types that can provide data/information to the analysis framework. This information is uniquely identified. Such source-specific attributes, which are driven by the characteristics of each social media source[8], are regarded as source dependent and therefore after the extraction of the data the process of the normalisation data will be necessary.

Spatial information for social media feeds can be inferred indirectly from content analysis, or it can be extracted directly from the data itself. It is important to highlight various forms of this Geolocation content, as it was already mentioned in the works of Croitoru et al.[9] and Crooks et al.[10]

Attribute data (Date and Time) can typically be found in all social media platforms. In this model, date and time information is embedded with each entry instance along with a time stamp-type identifier.

Attribute data (Keywords and meta data) are the parts of a social media entry. Users contribute keywords or tags, like hashtags (#) to emphasise views and ideas and engage other users[11]. Hashtags also support the building of semantic networks by allowing individual tweets to be linked thematically based on their content.

GIS with the capability to operate with the different type of the data (attribute and spatial), has the central role of 'converting' social media data into structured geosocial information from which knowledge can be extracted through further analysis. GIS enables the different outputs of the results like: map, tables, graphs.

Based on this model, new software application 'GeoSocioMonitor' was created. It includes extraction of the data from the different sources, processing raw data to get relevant spatial and attribute data, normalisation of the data and output of the results.

RESULTS AND DISCUSSION

Over the last decade, due to increased fishing pressure, anthropogenic eutrophication and global warming, many marine zooplankton communities have changed toward increasing dominance of harmful gelatinous species[12]. This, along with the occurrence of some outbreaks, has resulted in a greater degree of attention being devoted recently to jellyfish[13]. Such an avid interest has generally revolved around stinging gelatinous species, in view of their impact on the bathing amenity of coastal areas[14] and other effects[15].

However, very few studies on the occurrence of such species in Maltese coastal waters have been published to date, with most recent sightings being reported in media portals, mostly newspapers.

The Spot the Jellyfish initiative was launched by the staff at the IOI-Malta Operational Centre of the University of Malta in June 2010 as a citizen science initiative targeting the sighting of jellyfish species by the public in Maltese coastal waters. Although, it did not attract a big number of people, it is a valid source which was used to compare the results with 'GeoSocioMonitor' pilot project.

The pilot project used 'GeoSocioMonitor' software to monitor Twitter, Google+, News and Blogs for any activity related with jellyfish inhabiting or approaching the Maltese shores from 01.07.2015 to 15.08.2015. The search engine is based on the different meta data with date/time information. The attribute data and spatial data are extracted based on the combination of the following keywords: Jellyfish, Malta, Tuffieha, Golden Bay, Mellieha, Sliema. Tuffieha, Golden Bay, Mellieha and Sliema are the geographical locations related with bathing areas and beaches in Malta. Twitter data, in cases where geocoding information was available, were tagged with the corresponding, above-mentioned, geographical areas.

In the first level of the processing data, the raw data are collected. The software is scheduled to run as a background process in order to pick up the real time information. The second level of processing data is applied to get relevant information about spatial and attributable information, normalising the data in the forms of the tables and output the results in the form of graphs and maps.

The results are clearly indicating the ‘quantity supremacy’ of Twitter compared to other sources of data (Fig. 2). Limited number of the characters and possible ‘involvement’ in a ‘discussion’ makes Twitter one of the ideal social network media for the real time exchange of the information. Presenting the results in the percentage of the number of feeds, it provides further justification for the evidence of the Twitter grandness (Fig. 2). Google+ obtained an interesting position of close to 30% of total feeds – still finding the right place in the areas of the social networks – gaining the popularity. Blogs are the source of the data which can not be expected to have many feeds, but are a very valuable source of information and could be used to compare the validity of previously mentioned data. Even more so, the category of the news will be obtaining only an increasing number of the feeds if there is an event which might cause a big impact on the public.

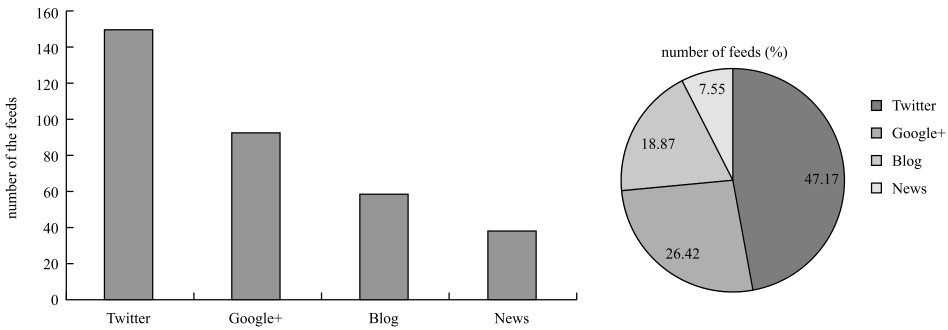

Fig. 2. Number and the percentage of the feeds from the different data sources

By comparing the results of the different data sources in the complete timeline, it could be seen there are corresponding trends, especially when the score of the events were at their highest (Fig. 3). This is an important part of the model where the analysis of the different data sources gives the indication of the reliability of the results. The results for the ‘pick’ period (27.07.2015–03.08.2015) showed the common corresponding trends for the different data sources (Fig. 4).

Comparing the results between GeoSocioMonitor and the Spot the Jellyfish for the ‘pick’ period – the highest number of the events – showed the close matches between these two (Fig. 5). However, it should be highlighted there are still a small number of inputs and further tests which need to be carried out.

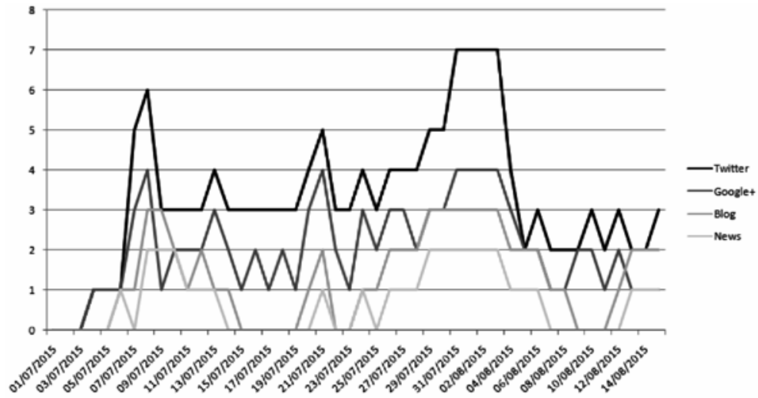

Fig. 3. The feeds in the timeline from the different data sources

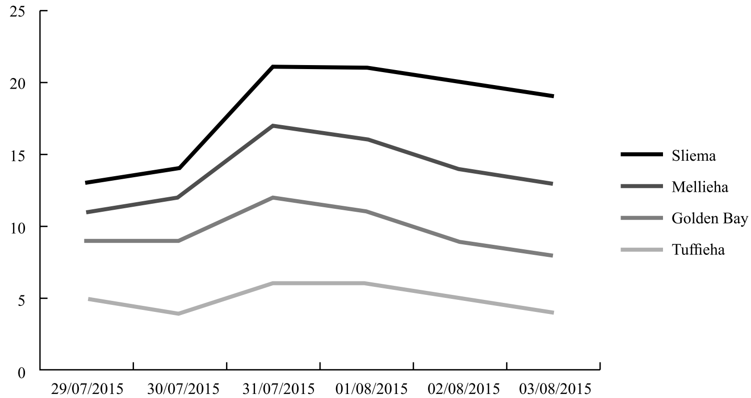

Fig. 4. The pick of the data and the trends related with the different geographic locations

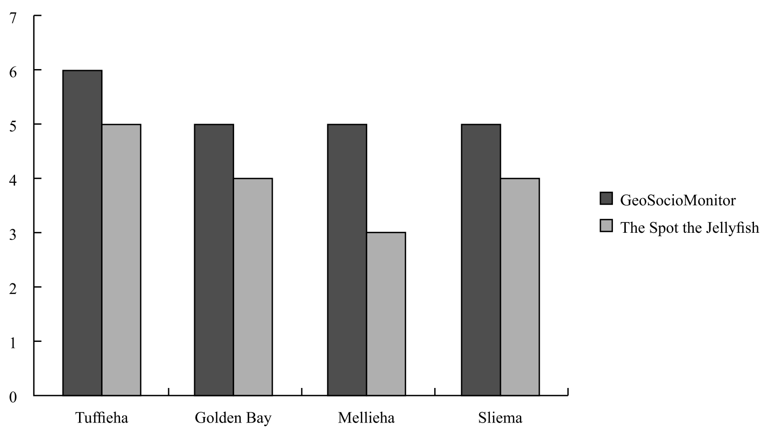

Fig. 5. Comparing the results of GeoSocioMonitor project and the Spot the Jellyfish for the pick period

GeoSocioMonitor software (as part of GIS) includes the possibility to create the maps with the real-time monitoring the marine environments. Figure 6 shows the distribution of the Jellyfish for 01.08.2015 at 10.00 h, Fig. 7 – the same day at 13.00 h, and Fig. 8 – at 16.00 h. The maps could be created at any point in time during the observation, but here the results were presented for three ‘snapshots’ for the same day. In this way, the main features of the time evolution of the monitored variable are evidenced and useful information could be obtained for further predictions of the running process[16]. This visualisation gives clear indication about the trends in the real-time monitoring the marine environments.

Fig. 6. Maps of spotted Jellyfish from GeoSocioMonitor for 01.08.2015 at 10.00 h

Fig. 7. Maps of spotted Jellyfish from GeoSocioMonitor for 01.08.2015 at 13.00 h

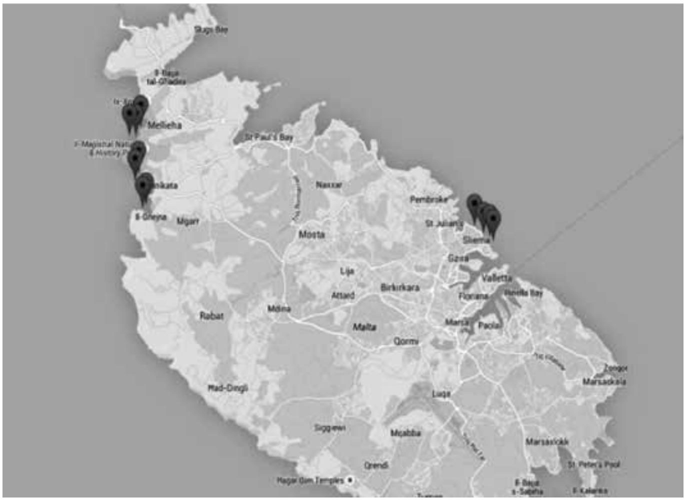

Fig. 8. Maps of spotted Jellyfish from GeoSocioMonitor for 01.08.2015 at 16.00 h

CONCLUSIONS

The emergence of social media is putting a different type of big data challenge to the geoinformatics community especially in the meaning of the new volumes of the data. Due to the particularities of the analysis that these data support, differentiate them from traditional geospatial datasets. At the same time, it is important to note that the ability to monitor human observations at a massive scale and to cross-reference such data across a variety of sources and modalities (e.g. text, imagery, video, and audio) presents a unique opportunity to validate information regarding the different events (in space and time). The model exhibited in this paper was utilised for the creation of the software application, which enables real-time monitoring of maritime environments using geosocial analyses. Observations carried out in the experimental pilot project of monitoring the appearance of jellyfish in the Maltese sea were proven to be positive results and applicable results. Further research should include text mining and sentiment analyses.

REFERENCES

1. A. M. KAPLAN, M. HAENLEIN: Users of the World, Unite! The Challenges and Opportunities of Social Media. ESCP Europe, F-75011 Paris, France.
2. D. BORTHAKUR: Petabyte Scale Data at Facebook. Accessed 03/07/2015. http://www.infoq.com/presentations/Data-Facebook, 2012.
3. A. STEFANIDIS, A. COTNOIR, A. CROITORU, A. T. CROOKS, J. RADZIKOWSKI, M. RICE: Statehood 2.0: Mapping Nation-States and Their Satellite Communities through Social Media Content. Cartogr Geogr Inform, **40** (2), 116 (2013).

4. F. SAHITO, A. LATIF, W. SLANY: Weaving Twitter Stream into Linked Data: A Proof of Concept Framework. In: Proc. of the 7th International Conference on Emerging Technologies, Islamabad, Pakistan, 2011, 1–6.
5. A. STEFANIDIS, S. NITTEL: GeoSensor Networks. CRC Press, Boca Raton, FL, 2004.
6. I. S. HERBAN, C. GRECEA, C. C. MUSAT: Using a Geographic Information System (GIS) to Model, Manage and Develop Urban Data of the Timisoara City. J Environ Prot Ecol, **15** (2), 558 (2014).
7. M. GOLUMBEANU, S. AVDIMIOTIS: Destination Management, Coastal Spatial Analysis and GIS. The Case of Chalkidiki, Greece. J Environ Prot Ecol, **10** (4), 1100 (2009).
8. J. SCOTT: Social Network Analysis. London, U.K.: SAGE Publications., 2012.
9. A. CROITORU, A. STEFANIDIS, J. RADZIKOWSKI, A.T. CROOKS, J. STAHL, N. WAYANT: Towards a Collaborative GeoSocial Analysis Workbench. COM-Geo, Washington, DC, 2012.
10. A. T. CROOKS, A. CROITORU, A. STEFANIDIS. J. RADZIKOWSKI: Earthquake: Twitter as a Distributed Sensor System. Transactions in GIS, **17** (1), 124 (2013).
11. D. ROMERO, B. MEEDER, J. KLEINBERG: Differences in the Mechanics of Information Diffusion across Topics: Idioms, Political Hashtags and Complex Contagion on Twitter. In: Proc. of the 20th International Conference on World Wide Web, Hyderabad, India, 2011, 695–704.
12. T. SHIGANOVA, A. MALEJ: Native and Non-native Ctenophores in the Gulf of Trieste, Northern Adriatic. J Plankton Res, **31** (1), 61 (2009).
13. M. PALOMARES, D. PAULY: The Growth of Jellyfishes. Hydrobiologia, **616** (1), (2009).
14. P. SCHEMBRI, A. DEIDUN, P. VELLA: The First Record of *Cassiopea andromeda* (Scyphozoa: Rhizostomeae: Cassiopeidae) from the Central Mediterranean Sea. Marine Biodiversity Records, 3: e6. doi: 10.1017/S1755267209990625. Published online by Cambridge University Press, 2010.
15. E. OZGUR, B. OZTURK: A Population of the Alien Jellyfish *Cassiopea andromeda* (Forsskal, 1775) (Cnidaria: Scyphozoa: Rhizostomea) in the Oludeniz Lagoon, Turkey. Aquatic Bioinvasions, **3** (4), 423 (2008).
16. I. MEGHEAA, M. MIHAIB, I. LACATUSUB, T. APOSTOLC: Time Series Model Applied to Environmental Monitoring Data Analyses. J Environ Prot Ecol, **13** (2), 426 (2012).

Received 10 October 2015
Revised 19 November 2015