Retrieve Volunteered Geographic Information for Forest Fire

Laura Spinsanti[1], Frank Ostermann[1]

[1] Spatial Data Infrastructures Unit
European Commission - Joint Research Center
Institute for Environment and Sustainability
Via E. Fermi, T.P. 262, I-21027 Ispra (VA), Italy
{laura.spinsanti, frank.ostermann}@Springer.com

Abstract. The proposed contribution describes a methodology we are testing in the framework of an exploratory research project on the contribution of Volunteered Geographic Information (VGI) in the case-study of forest fires. The purpose of the research is to assess the value of VGI at different stages of a fire event. The first step of the methodology focuses on the retrieval of geographical information related to forest fire from social networks. VGI is intrinsically heterogeneous as it is provided by different people, using different media such as photographs, text, and video. Moreover, geographical information can be explicit and expressed as coordinates or implicit as expressed in the text as location names. In this stage of the project we are extracting text data from Twitter and photographs from Flickr.

Keywords: VGI, Forest Fire, natural hazards, geographic information retrieval, social network, Twitter, Flickr.

1 Introduction

In the framework of an exploratory research project in partnership with Google and the Land Management and Natural Hazards Unit (JRC-EU), we are developing and testing a methodology to semi-automatically harvesting and analyzing Volunteered Geographic Information (VGI) in a case-study on forest fires. The purpose of the research is to assess the value of VGI at different stages of a fire event and integrate it with the flow of information from "official" sources as used by the European Forest Fire Information System (effis.jrc.ec.europa.eu).

The background of the research is the concept of citizens as sensors for a next generation digital earth [1, 2, 3]. The amount of volunteered geographic information is ever increasing, with many social media platforms adding the ability to geographically reference any volunteered information (Twitter GeoTagging, Facebook Places). The wide-spread use of GPS sensors in portable devices like digital cameras and smartphones facilitates this development.

The expected amount of volunteered data makes it inevitable to automatize the tasks of retrieve and filter using context aware methodologies to help in the assessment of the information relevance.

How to integrate this data with existing spatial data infrastructures will be a major future challenge.

In the next sections, we give some background information on VGI, we focus on the extension in a forest fire context of a work flow originally proposed in [2] and report on our preliminary findings of retrieved data for 2010 summer fire season.

2 Volunteered geographic information and forest fire management

Today, there is already a substantial amount of information provided by the public on disasters, which has always been strongly involved [5]. This VGI is intrinsically heterogeneous as it is provided by different people, using different media and authors often overcome device and software limitations in imaginative and unpredictable ways. Moreover, the expected future amount of near-real time, geographically referenced volunteered information will increase manifold during the coming years because of the next generation smart phones and device. Experiences like the Ushahidi platform [6] usage in Haiti 2010 heart quake disaster management cannot be replicate unless the analysis of the data flow will be (semi)-automatic.

This development is going to change the way information is collected, distributed and used. In the past, there was only a uni-directional vertical flow of information from officials to public, with horizontal peer-to-peer communication limited to a small reach. With the web 2.0 the amount of bi-directional horizontal peer-to-peer communication among the public has been increasing. Another boost has been the recent possibilities to reach an even larger number of people in real-time through social media platforms like Twitter and Facebook. It is expected that also the traditional uni-directional vertical flow of information is going to be affected, possibly becoming bi-directional soon. The lines between public and official already blur when official administrative agencies like British Columbia that use regularly Facebook or Twitter for communication services.

Despite the difficulties to integrate with traditional established emergency response protocols and opposition, it is a fact that people in crisis context tried to get more information, to help by organizing the information and making portals available to others. This activity was motivated by the observation that traditional media outlets were often too inaccurate, too global, biased towards large agglomeration/urban areas, and late.

3 VGI work flow: from harvesting to integration with official data

The entire work flow starts with the retrieval of data from various social media sources in the Web. Then the data are processed, integrated, analyzed and evaluated. At the end, clustered and quality assessed information is integrated with the official data and produce an enriched representation of the event. The core part of the process is a complex knowledge discovery process that involves several steps and cycles, especially for data source that are continuously updated during the event.

The data can be analyzed and assessed as single piece of information, but also as grouped information. The dimensions of analysis are numerous and can be complex, as for space and time dimension. For these reasons, analysis and discovery methods are domain driven, including natural language processing (NLP) and ontology, and data driven, including spatio-temporal data mining. Moreover, the grouped information can lead to a new knowledge that feeds back into the assessment of the single information.

The work flow is completed with the integration of the quality-controlled, event-related VGI into the official spatial data infrastructures used during the crisis.

3.1 Geographic information retrieval: summer 2010

We harvested Twitter using the provided streaming API and 12 wildfire related keywords (e.g. fire, helicopter, evacuation) in 8 different languages. We collected 24,5 GB of data for more than 8 million messages and 700 thousand images. Around 1% of the dada is geocoded with coordinates (even fewer in Europe). Many of these contain geographic information such as: *"Two very Large forest fire in the mountains behind Funchal clouds of smoke covered the sun turn sunlight deep yellow ash coming down"*. It is worth notice, that this kind of information must be carefully analyzed. At first, it refers to the city of Funchal. We used YahooPlacemaker[1] that is a freely available geoparsing Web service. More precisely, the text refers to the mountain behind the city. This level of natural language analysis has to be performed in the future of the project.

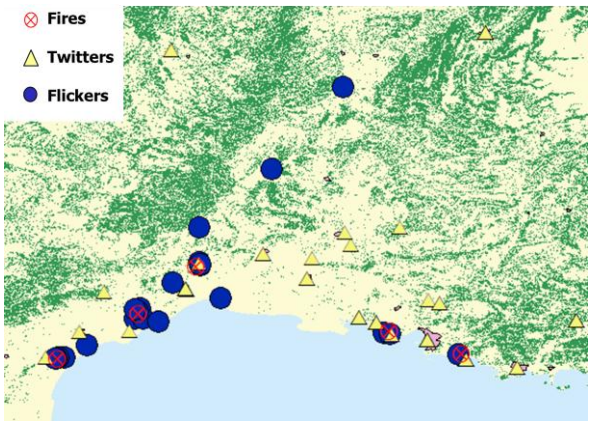

Fig. 2. *'Incendie'* Twitter and Flickrs events in south of France, summer 2010 and Effis registered forest fires

A first broader analysis was performed by Country. As an example we show the 2010 French Fires located in the South part of France near the city of Marseille. We

[1] http://developer.yahoo.com/geo/placemaker/

selected the Tweets and the photographs posted on Flickr in the season containing the word "*incendie"*. Only few of them had coordinates. For the others, we used the YahooPlacemaker service and filtered the ones located in France. Then we clustered (minimum 2 related information) by day and location, resulting in spatio temporal point we call 'events'. As it is possible to observe in Fig. 2, there is a correspondence between the official fires registered and the social networks activity. In the map it is possible to observe only the spatial dimension of the events, but temporal clustering also confirms the temporal correspondence. The next steps of the project consist in applying knowledge discovery methods to refine the retrieval and the analysis of the data.

4 Conclusion

In this paper, we have argued that emerging use of social media by members of the public will become an important pathway for communicating during a crisis event. Almost all volunteered information has a geographic component. VGI has the potential to dramatically change the traditional top-down, uni-directional communication pathway from official to public via broadcasting media. Recent examples have shown that the public and official institutions embrace the new platforms for communicating volunteered information. However, the increasing usage will also amplify two main challenges: the volume of information will need some sort of filtering, and in order to be useful for official emergency response, its quality, relevance and credibility needs to be assessed. We propose a work flow for the retrieval, formatting, filtering, and assessment of volunteered geographic information from social media. The integration of VGI with official spatial data infrastructures, and its use by the public and decision-makers, are the final steps in the proposed work flow.

References

1. Craglia, M. et al., 2008. Next-Generation Digital Earth: A position paper from the Vespucci Initiative for the Advancement of Geographic Information Science. International Journal of Spatial Data Infrastructures Research, 3, 146-167.
2. De Longueville, B., Annoni, A. et al., 2010. Digital Earth's Nervous System for crisis events: real-time Sensor Web Enablement of Volunteered Geographic Information. International Journal of Digital Earth, 3(3), 242 - 259.
3. Goodchild, M.F., 2007. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. International Journal of Spatial Data Infrastructures Research, 2, 24-32.
4. Goodchild, M.F., 2010. Twenty years of progress: GIScience in 2010. Journal of spatial information science, Number 1 (2010), pp. 3–20.
5. Palen, L. & Liu, S.B., 2007. Citizen Communications in Crisis: Anticipating a Future of ICT-Supported Public Participation. In CHI 2007 Proceedings. Computer Human Interaction 2007. San Jose, USA, pp. 727-726.
6. http://haiti.ushahidi.com/.