Introduction toSocialRadio Case Studies and Perspectives

Amosse Edouard, Nhan Le-Thanh

Introduction toSocialRadio

Case Studies and Perspectives

Amosse EDOUARD
WIMMICS - The I3S Laboratory - CNRS - INRIA
University of Nice Sophia Antipolis
Sophia Antipolis, France
Amosse.Edouard@unice.fr

Nhan LE-THANH
WIMMICS - The I3S Laboratory - CNRS - INRIA
University of Nice Sophia Antipolis
Sophia Antipolis, France
nhan.le-thanh@unice.fr

***Abstract*— In this paper, we introduce a Social Radio platform designed to support contextual communication between communities. Radio communication was no longer used for information broadcast using different communication channels. The same concept is used in this work but for different purposes. We set up a microblogging community channel to report, comment or simply vote some events. In order to understand the concept we detail a scenario in the epidemic field.**

Keywords:social networks, online communities, social radio, short messages.

I. INTRODUCTION

Since people moved an important part of their time on Internet [8], microblogs such as Twitter or social networks such Facebook, Google+ become more and more popular and can be considered as important tools for communication among people, sharing expertise and knowledge enrichment. It is important to note that people have different motivations in using these tools. Some are interested in sharing observations while others are concerned with consuming information. Peoples' needs depend on whether they are consumers or producers. While information are more and more available, finding the relevant information among all those context-dependant observations is an important issue for consumers and also a challenge for researchers or service providers. For example, a candidate can be interested in what people think of his candidacy for the next election.

Handlinglarge quantities of comments from all users on a given topic can be difficult or even impossible without appropriate tools. Such tools must be able to retrieve the common trends or the main idea among a set of observations.In this work we tackle this problem by introducing the social radio, a platform enabling users to extract the common trends on a specific subject (theme) by analysing reported observations within an online community.

The term radio is commonly used for information broadcasting using radio waves or frequencies. In this study, we use the same concept but for different purposes. The main parts of a radio are: frequency (its identity),information sources, a redaction service and an audience (people interested in this radio). Information sources can be employees (journalist, reporters), music producers (for a music radio) etc.

Activities on the web are quite similar toa radio becauseinformation is constantly shared between producers and consumers. The main differences between the web and a radio come from the organisation of the information. Numerous online services provide multidisciplinary information at large scale to an open audience. Finding relevant information is an important challenge for consumers.

Data analysis in online social networks or microblogging platform is largely explored since the past decade in many domains [1,4,5,7,13,14,15]. Lots of studieshave focused on motivation [14], member's activities and behaviours [19,20], data filtering and enrichment [15,21]. In [31], the goal was to extract common thematic used in Twitter posts based on an n-dimensional analysis taking into accounttheme, space and time. The authors in [32] proposed an approach for event detection from social text stream by exploring the content as well as temporal and social dimension. F. Abel has proposed in [15] an approach to enrich status messages from Twitter based on external links referred in the messages.In [21], the authors used hash tags and linked data to enrich post from twitter.

In this paper we study the topology of the social radio, a platform for information synthesis within an online community. We analyse the main needs that can make the platform generic. By generic we mean, anyone can configure it according to his need with multiple information sources such as online social networks, microblogging, websites etc.

The rest of this paper is organized as follows; Section 2 presents background about online communities and microblog messages;Section 3 introducesa generic platform for broadcasting information in online communities, which we call the social radio. We then detail a use case for the social radio in Section 4. Finally, we discuss future work and make conclusions in Section 5.

II. BACKGROUND

A. Online communities, structures and objectives

In an "online community", members do not necessarily meet and communicate with each other face-to-face. Online community can be built in two ways: implicitly from user behaviour (e.g. linking users who have the same hobbies or who took similar actions an a website) or

explicitly (e.g. users link themselves to others: friends or co-workers).
According to [17], online communities can be broken down into overlapping classes as shown in Figure 1. Generally, members of online communities can be split into two categories [12]:

- Active Participants (Posters): Users who participate actively in the community by posting messages
- Inactive Participants (Consumers): Members who are interested in the information shared within the community.

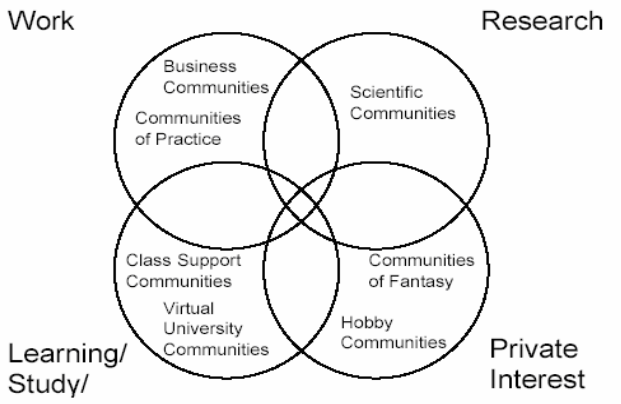


Figure 1: Different Online Communities Profiles (extracted from [17]).
A community can be at the intersection of one or more classes.

As pointed out by J. Nielsen [33], user participation follows more or less a 90-9-1 rule. That is, 90% of users read/observe but don't contribute; 9% contribute from time to time; and most contributions come from just 1% of the users, who participate frequently.
How information is used depends onwhether a participant is active or inactive.For instance, onan e-commerce platformusers (customers) comment or rate products independently of other users while buyerslurk for most rated products to buy. Users post information regarding their observations or personal opinions on fact they observe. While many users can report observations on the same fact, event or theme;how to present the relevant information to consumers is an important challenge.

B. Short messages

Microblogging and social networks differ from blogsandgeneralweb sites becausethey enable anyone to sharetheir thoughts online to an open audience. This *freedom of the crowd* distributes a large amount of data on social platforms. Microblogginghas also been successful in corporate environments in facilitating informal communication, learning and knowledge exchange (almost real-time) and fast propagation of new information.
However, data frommicroblogs is substantially differentfrom other that found in other information systems such as theenterprise web site.In particular, (1) itrepresentsrich social connections (between the information senders/authors and recipients/reviewers); and (2) the content is strongly context sensitive. That is, not only within the text piece content, the meaning of words is dependent from the context of the communication as well as relationships among social actors.
In a communication process between humans, the interlocutors know the context.That's why they can understand the meaning even when a message is very short. As an example "*This traffic jam bothers me*"has a meaning for the recipientbecause he has the context (e.g. who the sender is, where the sender is, why he's there, his relationship with the sender). Inferring context is essentialwhen processing such messages.Hence, in our work, we willprocess messages based on both the thematic content and on the community in which the message was produced.

III. TYPOLOGY OF THE SOCIAL RADIO AND DEFINITION

This section defines terms and concept of the studied platform.
In other to define the Social Radio for information broadcasting within online communities, we consider the following concepts according to member'sbehaviours and needs:

- A *theme*defines the context of shared information.
- A *community*is composed of*posters* and *consumers*.
- Information processing mechanisms definehow to structure information based onconsumers'needs and information privacy.

Definition III.1 Informationtheme
The theme of information is what that defines the context 0of the information being shared. s the main part of a message, that what which defines the context of the information being shared. Because of the brevity of post in online social platforms, the theme is an important key for context grasping. Context inferring is widely explored in computer science [1,34]and semantic web [15]. Information context is the first step for information processing within online communities.

Definition III.1 Posters and Consumers
Sincemembers of a community have different information needs and behaviours; information processing it is important to we deliberately splitthe members who form a radio's community intoposters and consumers [12]. Posters are those who post messages on the radio while *consumers* are those who consume information.This approach allows us to define how information should be presented to users according to their profiles.

Definition III.3 Information synthesis
When seeking for information, to begin with, users are interested in anoverview instead of details [33]. For example, the most important part of a meeting is the

summary of points that have been discussed and decisions that have been taken. Details can come up if users ask for them. A radio social shall be able to summarize information in other to present the most relevant information to the user.

A. *Types of radio*

Consideringcommunity structures, behaviour and expectations [12], information is processingfrom two perspectives:posters (who post information) and consumers (who seek andconsume information). For posting and consuming there are three possibilities: *public access, private access and hybrid access*. A radio with public access for posting or consuming enable any users to access information on the radio; a radio with private access restrict access to some members of the radio; and radio with hybrid access manage access to information according to privacy and type of information. Considering the three possibilities for consuming and posting, the Table 1 summarizes the different types of radio including appropriate examples.

Posters	Consumers	Example
Public	Public	**Traffic event report:***Anyone is allowed to report or consume information within the community*
Public	Private	**Customer satisfaction survey:***Anyone is allowed to post information while only employees of the service providers are allowed to consume reported information.*
Public	Hybrid	**E-Commerce platform:** *Anyone is allowed to buy and comment on product they have bought, rating and comments are visible to other customers while information on a customer such as address or phone number is accessible only by employees of the company.*
Private	Public	**An online blog:** *Only manager (s) can publish information on a blog while anyone is granted to consume published information.*
Private	Private	**Instant Messaging:** *Only member can post information that can be consumed only by other members*
Private	Hybrid	**Twitter:** *Only account owners can twitt; some information is available for public access while other is available for (some) followers.*
Hybrid	Public	**TODO:**
Hybrid	Private	**Air pollution survey:** *in an air pollution survey, although anyone can report observations but only authorized members can access the reported information*
Hybrid	Hybrid	**Epidemic survey:** *This community allows anyone to report epidemic symptoms, but information reported by a health worker is more reliable than one reported by an unknown user. In the output only confirmed epidemics are reported with public access. Information such as unverifiedcases is granted to some members (e.g. doctors).*

Tableau 1. Type of radio and examples.

B. *Characteristics of the social radio*

In a survey on Twitter in [14], Java et al. found that members' activity can group as:1) Daily Chatter (what a user is actually doing), 2) Conversations (direct exchange between members), 3) Sharing Information and 4) Reporting News. In [34], Kate et al. extend those categories into:status, provide information, forward information, ask question direct post anddirect question. Although those studies are on the Twitter platform, the categories describe the main user activities within online communities. In the radio social, more than the categories defined in [14, 34], we attempt to capture main types of user's activities within online communities: *reporting, evaluation, and opinion*.In Table 2 we summarize our codingactivitiessupported by examples.

Category	Example
Reporting: *Users report news or event based on observation in their environment. Mostly such actions are objective; users report what they observe in the universe. There is no personal appreciation or interpretation of the fact that the user is being described. In twitter, when reporting an event users use hash tag to highlight the most important word in the post, in most cases those keywords define the context of the post [1, 15].*	**Trafic jam on a road:** *The is any personal appreciation when reporting traffic jam, the user observes that this event exist right now in this place.* *@BLTEDI: "This #traffic bothers me". This means(somewhere in the world) a traffic jam exists.*
Opinion: *The post describes the user's appreciation on the given subject.Activities such as evaluation, notation and comment fall into user's opinion. An opinion is subjective and strongly depends on the user point of view. More generally, opinion groups appreciation and evaluation of facts, services and events. Opinion can be qualitative (the set of value is finite; e.g. like vs. dislike) or descriptive (the set of value that a user can use is indefinite).*	***Descriptive Opinion:*** *"#enjoying #holidays with family and friends."* ***Qualitative Opinion:****@MATITA: "#S5 like this phone"*
Questioning: *Posts that are looking for feedback. The user looks for a response within the community.*	*@BKFD: "Anyone knows how to import data from a #RDBMS to #RDF?"*

Tableau 2. User activities and examples. Examples are extracted from Twitter; usernames are removed for privacy;

C. *Discussion*

In the previous section, we have definedtypes and characteristics of information in online social services; in this section we discuss the needs for the social radio to be a generic service for information broadcasting within online communities.

Consistent to previous research, the social radio shall take in consideration type of communities [5] as information theme depends on the group in which it is sharing [20]. On the other hand, information processing shall consider consumers need. As pointed out in [7], in a community

members have different expectations: 1) content presentation, 2)Interactive Realism, 3) privacy, 4) leadership and 5) decision making. Consequently, withina community information can be processed differently according to the consumers needs (e.g. service providers and customers in a E-commerce platform).
According to E. José et al.,*a way to reduce the gap between real and virtual communities is through a proper intervention by one or various person in charge of the community*. We can adopt this management approach to allow users define the structure of a social radio.

Information theme can exploit semantic web technics [1, 15] to define characteristics of shared data on the radio. Indeed, web semantic and ontologies enable definition of concept and relation among concepts in an information-sharing environment. Defining a vocabulary will benefit for the radio in this it defines clearly the radio thematic and his relation to other thematic on the web as well as dependant concept and entities.

Information sources are the sources that provide information to the radio. A radio can define external as well as internal information sources according to the user needs. Internal sources can be dedicated applications that providedata directly to the radio (e.g. users with smart devices, sensors etc.). Existing social services platform expose their data in API such external application can use them. Based on its thematic a radio can connect to an external source and gather information shared within the community. Because of the brevity of information in microblogs and social networks, data retrieved from external sources must be enriched in other to grasp their context. Enrichment mechanism can rely on previous research in the field [1, 15, 21].

The main part of the social radio is automatic processing of information according to user needs, types of information and characteristics. According to that a radio shall provide generic algorithms for information processing and enable a user to configure them according to his needs.

IV. RASE "RADIO SOCIAL FOR EPIDEMIOLOGY", A CASE STUDY

In this section, the focused use case is illustrated by an application developed with health workers in Biot to report epidemic symptom based on the social radio concept.
The system is composed of two parts:an administrative part that enables management and configuration settings and the client part that enable users to interact the system. Interaction can be report new epidemic symptom or consume information that has been shared within the community. In the current version only health workers (e.g. doctors, nurses, pharmacists etc.)are allowed to postwhile other users can consume shared information. The Figure 2 presents a use case of the application.

In this application we are facing two constraints:a) time constraint since event reporting is not the primary part of health workers' activity and b) medical confidentialities.

a) Medical confidentialities

Patient information is not required for the system; posters report symptoms without any personal information regarding the patient.

b) Dealing with time constraint

In order to make the application less time consuming, interfaces shall be user friendly and intuitive. In this scenario, information theme is well known; members of the community exchange information on epidemic symptom and diseases.
Unfortunately, existing medical ontologies did not applicable in our case because they are domain-related. Diseases or symptoms are treated differently from a medical domain to another. Pathology of diseases or symptoms are not relevantto our system, we do not need a complete ontology. For that, we have built anontology to that defines a shared vocabulary for the actors of the system. The ontology has two concepts:Epidemic and Symptoms and a set of properties as shown in Figure 4. Instances of the ontology are created using and administrative interface (Figure 5).
A message is a quintuplet $\theta = < P, O, S, I, A >$. [4]

a) P is the phenomenon (e.g. influenza) an instance of the ontology;
b) O is the observer (e.g. a doctor);
c) S is the location of the observation;
d) I the time (instant) of the observation and
e) A content of the observation (e.g. cough, headache) instance of the ontology related to the phenomenon.

To reportan epidemic a user pick an instance of epidemic in the vocabulary and the system retrieves automatically symptoms related; the user chooses one or many symptoms.
Space and timeare inherent to epidemiology [9]. Asking the user to provide those properties will overload the user actions. Hopefully, smart devices enable to retrieve user location, current time and information about the user.We consider a "broadcast delay" that consists in postponing the message over a short period while we enrich the message.

There exist two main approaches for spatial representation: a quantitative approach that represents spatial entities using numerical values (e.g. longitude, latitude, geometric form); and a qualitative approach that represents spatial entities using lexical term close to the human perception of spatial notions [22]. The main advantage of the qualitative approach on the quantitative approach is the ability to deal with incomplete information. Because the user location is not always available (e.g. GPS disabled) and the system does not require fine-grained location; spatial entities are represented qualitatively. For example, when a user reports and epidemic case, we associate the nearest locality to the message he reports. Spatial settings are defined according to the Geonames ontology[1]in which geographic spatial entities are grouped in a subsumption relationship. We have restricted ourselves to five levels in the ontology

[1] http://www.w3.org/2005/Incubator/geo/XGR-geo-ont-20071023/

as presented in table 3. To retrieve spatial entities, we have used the Google Reverse Geocoding API to map the user location with named spatial entities. Mapping between Google spatial entities and Geonames are resumed in Table 3.

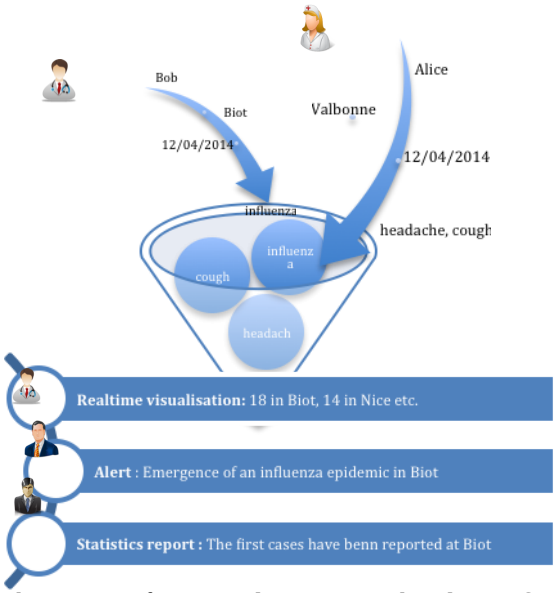
Google parameters	Geonames Properties	Name
country	A	Country
administrative_area_level_1	ADM1	Region
administrative_area_level_2	ADM2	County
administrative_area_level_3	ADM3	City
locality	ADM4	Locality

Tableau 3 Mapping Google reverse geocoding attributes and Geonames ontology

Data processing

Because an epidemiologic symptom is reported in a region does not mean that an epidemic exists. To detect the emergence of an epidemic in a region, we have defined analgorithmthat process periodically information based on their spatiotemporal properties. At this stage, we cannot confirm automatically that an epidemic exist but instead weraise a notification when the incidence of an epidemic exceeds a given threshold. For each period, the algorithm computes the number of cases reported in each distinct region and determines a ratio according to the population. If for a region the incidence is greater than the threshold,the system raises an alert to notify the administrators.

Given T a time interval defined by$T = [t_i, t_j]$.The set of observations in this interval is $\Omega = \{\theta_1, \theta_2, \theta_3, \ldots\ldots, \theta_n\}$, where $\theta_i$is a message.



Let S be a set of unique location in the dataset $S = \{s_1, s_2, \ldots, s_m \ / \ m \leq n\}$. For each region in S the number of reported observations is determined by:

$$\Delta_s = \sum_{i=1}^{i=card\,(\Omega)} \begin{matrix} 1 \text{ if } \theta_i \in \Omega \text{ and } \theta_i(S) = \text{ s} \\ 0 \text{ else} \end{matrix}$$

The incidence is a probabilistic value given by: $\kappa_s = \frac{\Delta_s}{\eta_S}$where $\eta_S$is the population of the region.

The dataset do not contain information about the population of a region. DBpedia is a crowd-source effort to make structured information publicly available on the Web. It links resources from Wikipedia with other datasets and enables asking complex queries using SPARQL, a query language for the semantic Web. The following request to the DBpedia SPARQL endpoint retrieves the population of the region S. Note we assume S is Nice in the request.

```
SELECT ?q{
SERVICE <http://dbpedia.org/sparql> {
<http://dbpedia.org/resource/Nice>
<http://dbpedia.org/property/population>  ?q. }}
```

Request 1: Request to the DBpedia SPARQL endpoint retrieving the population of the city of Nice

From the previous SPARQL request we have the population of the region as:

$\eta_S = q$(Where q is the response obtained from DBpedia)

Then the incidence for a region is:

$$\kappa_s = \frac{\Delta_s}{\eta_S}$$

According to the INSERM vector scale, we have associated a colour to each region regarding the position of the incidence in the scale. If the incidence is greater than the threshold we raise an alert to the system managers.

Conclusion

We have presented the RaSE platform an application built on the concept of the social radio. This application provided tools and interfaces that enable health workers to exchange information on epidemic symptoms in a distributed environment. Data reported by health workers are processed periodically according to their spatiotemporal properties and displayed on base map in other to make them available for consumers.

In France, there exist the "Sentinels Network " that provide information about epidemic incidence on the territory. However in their approach “sentinels” using paper and phone calls to report information.Furthermore "sentinels" are dedicated users that journey over the country to report information. In our approach, data are real-time reported and can be reported by several health workers such as (doctors, pharmacists, nurses etc.). Moreover, “sentinel” can use the platform to report information to the "Sentinels Network".

V. CONCLUSION &PERSPECTIVES

In this paper we have introduced the concept of "Social Radio", a platform for information broadcasting in online communities. Online communities' members have different expectations and we have found that information cannot process in the same way for all members in a community. The social radio aims to provide tools enabling members of online communities building their own "radio" according to their needs. We have defined the types and characteristics for a social to be generic.

To prove the concept we have presented RaSE an application that enables health workers to report epidemiologic symptoms. Data reported on the platform are automatically processed and enrich before sharing them to consumers.

The following points resumed the direction for future works:

- Data enrichment: Short messages are widely used in online communities and can be from several sources,in future works we will define an approach for data enrichment and reconciliation. We believe that spatiotemporal properties are a good alternative for data enrichment.
- Information processing: We have identified in the Section III.C that algorithm for data processing can be: 1) probabilistic, 2) statistical and 3) evidential.

REFERENCES

[1] A. Passant, T. Hastrup, U. Bojars, and J. Breslin “Microblogging: A Semantic and Distributed Approach”

[2] M. C. Kindsmuller, J. Milz, and J. Schimdt, "Instant Online Communities as a Means to Foster Conferences".

[3] G. Marsh, "The Community of Circoumstance - A tale of three cities: community participation in St Kilda, Knox, and Lewisham".

[4] N. Le-Thanh “Introduction à l'Univers des BRIEFS”

[5] E.J da Silva, and S. Sallaume "Online Communities Administration: Defining Tools for Different Profiles"

[6] S. Seufert, "Design and Management of Online Learning Communities", In: 2nd Annual Conference on Innovative Research in management.

[7] E. J. da Silva, C. S. de Souza, R. O. Prates, and A. M. Nicolaci-da-Costa "What they want and what they get: A study of light-weight technologies for online communities"

[8] L. Backstrom, E. Sun, C; Marlow, "Find Me If You Can: Improving Geographical Prediction with Social and Spatial proximity"

[9] G. Camara, "Un système de veilleépidémiologique à base d’ontologie", pp. 44-54

[10] "Epidemiology Key to Prevention", chapter 1, pp 1-9

[11] "Epidemiology Key to Prevention", chapter 2, pp 28-56

[12] B.Nonnecke, D. Andrews, J. Preece, "Non-public and public online community participation: Needs, attitudes and behavior"

[13] M. Naaman, J. Boase, C. Lai, "Is it Really About Me? Message Content in Social Awareness Streams"

[14] A. Java, T. Finin, X. Song, B. Tseng "Why We Twitter: Understanding Microblogging Usage and Communities"

[15] F. Abel, Q. Gao, G. Houben, and K. Tao, "Semantic Enrichment of Twitter Posts for UserProfile Construction on the Social Web".

[16] M. S. Ackerman, "The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility"

[17] S. Seufert "Design and Management of Online Learning Communities"

[18] C. Mounia, A. Martin, B. B. Yaghlane "Estimation de la fiabilité des sources des bases de données évidentielles"

[19] M. Stankovic, P. Laublet, A. Passant "Directing Status Messages to their Audience in Online Communities"

[20] T. Schoberth, J. Preece, A. Heinzl " Online Communities: A Longitudinal Analysis of Communication Activities"

[21] A. Passant, P.Laublet, "Meaning Of A Tag: A Collaborative Approach to Bridge the Gap Between Tagging and Linked Data"

[22] J. Renz : "Qualitative spatial reasoning using constraint calculi", Australian National University

[23] M. J. Franklin, D. Kossmann, T. Kraska, S. Ramesh, and R. Xin. "Crowddb: answering queries with crowdsourcing." In SIGMOD, 2011.

[24] A. Marcus, E. Wu, D. Karger, S. Madden, and R. Miller. "Crowdsourced databases: Query processing with people." In CIDR, 2011.

[25] A. Parameswaran and N. Polyzotis. "Answering queries using humans, algorithms and databases." In CIDR, 2011.

[26] A. Parameswaran et al. " CrowdScreen: Algorithms for Filtering Data with Humans"

[27] M. Germain, J. Boucher, G. B. Béniéet É. Beaudry, "Fusion évidentielle multisource baséesurune nouvelle approchestatistiqueflous".

[28] Dempster, A. P. (1968) A generalisation of Bayesian inference, Journal of the Royal Statistical Society, p. 205-247.

[29] Shafer, G. (1976) A Mathematical "Theory of Evidence", vol. Princeton, NJ.

[30] D. Deng, G. Mai, T. Chuang, R. Lemmens and K. Shao, "Social Web Meets Sensor Web: From User-Generated Content to Linked Crowdsourced Observation Data"

[31] M. Nagarajan, K. Gomadam, A. P. Sheth, A. Ranabahu, R. Mutharaju, A. Jadhav "Spatio-Temporal-Thematic Analysis of Citizen Sensor Data: Challenges and Experiences"

[32] Q. Zhao,P. Mitra, B. Chen "Temporal and Information Flow Based Event Detection From Social Text Streams"

[33] J. Nielsen, "Participation Inequality: Encouraging More Users to Contribute" [online] accessed on April, 17th 2014

[34] K. Ehrlich, N. S. Shami, "Microblogging Inside and Outside the Workplace"