# Twindex Fuorisalone: Social Listening of Milano during Fuorisalone 2013

Marco Balduini[1], Emanuele Della Valle[1], Daniele Dell'Aglio[1], Mikalai Tsytsarau[2], Themis Palpanas[2], and Cristian Confalonieri[3]

[1] DEIB – Politecnico di Milano, Italy
{marco.balduini,emanuele.dellavalle,daniele.dellaglio}@polimi.it
[2] DISI – Universitá degli Studi di Trento, Italy
themis@disi.unitn.eu, tsytsarau@disi.unitn.it
[3] Studiolabo, Italy
cristian.confalonieri@studiolabo.com

**Abstract.** Fuorisalone during Milano Design Week, with almost three thousands events spread around more than six hundreds venues, attracts half a million visitors: what do they say and feel about those events? Twindex Fuorisalone is a mash-up that listens what all those visitors posted on Twitter and Instragram in that week. In this paper, we briefly report on how Twindex Fuorisalone works and on its ability to listen in real-time the pulse of Fuorisalone on social media.

## 1 Introduction

The Salone Internazionale del Mobile[1] is the largest furniture fair in the world. It is held in Milano every spring and it lasts a week. In the same week, thousands of satellite events are scheduled; they are grouped under the name of Fuorisalone[2]. Those industrial design events have been gaining more and more attention in the last years. As a result Fuorisalone is one of the most important events for industrial design worldwide. Now, this week dedicated to design is called the Milano Design Week (MDW), and it attracts every year more than 500.000 visitors. This year (i.e., 2013), MDW was held in April 9-14.

While the Salone Internazionale del Mobile is held in one specific location (the Milano fair area), Fuorisalone is held in different venues (more than 600) around the city. Due to this decentralization of the event, the Fuorisalone organisers are interested in monitoring in real time the reaction of the visitors:

Q.1 What are the most attractive events?
Q.2 What do visitors say about the events they join?
Q.3 What is their mood before, during and after the events they join?

Manually collecting the information to answer those questions is complex and expensive. The state of the art consists in analysing mobile network data [1],

[1] Cf. http://www.cosmit.it/en/salone_internazionale_del_mobile
[2] Cf. http://fuorisalone.it/2013/

P. Cimiano et al. (Eds.): ESWC 2013, LNCS 7955, pp. 327–336, 2013.

but only big event organisers (e.g., Olympic games' hosters) can afford its cost. Nowadays, the Web offers a cheaper way to collect the data to answer them: the rise of the social Web, e.g., Twitter[3] and Instagram[4], provides a continuous flow of information in the form of social text streams (shortly, social streams). Being able to process social streams in real-time allows to develop new services, useful not only for the aforementioned organisers (real time monitoring), but also for visitors (find the more popular events) and other subjects (e.g., Milano municipality, municipal police, etc.).

In this paper, we present Twindex Fuorisalone (TF), a mash-up that gathered, processed, and analysed social streams related to Milano Design Week using the fuorisalone.it repository of events as a source of information to make sense of the continuous flow. The techniques are similar to those previously used for disastrous events (e.g., earthquakes [2]), but are applied in the new setting of event management. In particular, this work builds on the results of the experiments we performed in the previous projects in Politecnico di Milano (i.e, BOTTARI [3], Social Listening of London Olympics 2012 [4]) and University of Trento (i.e., sentiment mining and contradiction detection on the Web [5]).

The remainder of this paper is organised as follows. Section 2 describes the data sources that TF integrates, the architecture of TF and how it processes the social streams from Twitter and Instagram against the information stored in fuorisalone.it repository. Section 3 describes the front end of TF as it was made available during MDW 2013. In Section 4, we report on the ability of TF to answer the three questions listed in this section. Finally, Section 5 concludes.

## 2 Data and System Architecture

In this section, we first analyse the data sets that TF consumes, then we explain its architecture and how social streams are processed.

### 2.1 Data Sources

As illustrated in Figure 1, TF gathers data from three sources: fuorisalone.it, Twitter and Instagram.

The *fuorisalone.it* repository is a proprietary data set that collects information about the schedule of the events. It contains about 2730 events, located in 676 venues. For each event, it describes its duration, its category and who sponsors it. The data are accessed through a Web service interface.

*Twitter* is the most famous microblogging service in the Web: it allows to post short text messages. The messages (tweets) have an identifier and a time stamp; additionally they can be annotated with topics (a word starting with #), user names (strings starting with @), and geo-references (WGS84 coordinates). TF accesses Twitter data through the streaming API[5]: it is a push service that

[3] Cf. `http://www.twitter.com`
[4] Cf. `http://www.instagram.com`
[5] Cf. `https://dev.twitter.com/docs/streaming-apis`

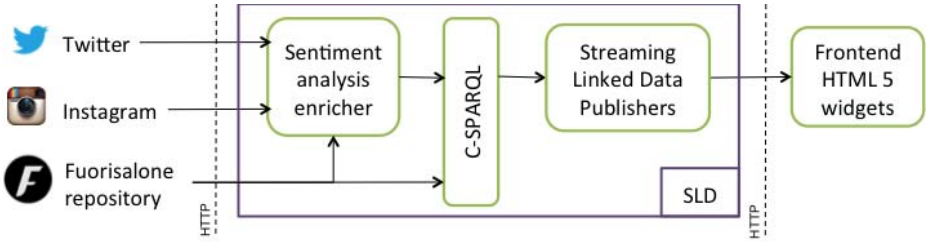


**Fig. 1.** Architecture of Twindex Fuorisalone

allows to obtain one or more data streams (sequences of tweets ordered by time stamp).

*Instagram* is another microblogging service; unlike Twitter, in this service users post images optionally associated to a short message that, in most of the cases, is a list of hashtags. Due to the fact that Fuorisalone is (mainly) composed of industrial design events, we are interested in monitoring this source as well: people taking pictures and sharing them on Instagram. The data is also accessed through a streaming API[6].

TF subscribes three kinds of social streams on Twitter and Instagram:

- A *geo-bound stream* is associated to a rectangular area and it contains all the micro posts georeferenced with a coordinate inside the area up to a given maximum rate. The geo-bound stream used by TF contains all the micro posts with a coordinate in Milano. Note that both in Twitter and Instagram the georeference is optional, so this stream captures only a part of the micro posts related to the Fuorisalone events.
- A *topic-based stream* contains a subset of the micro posts about a given topic (a syntactic match is performed and a max rate filter is applied). Using the textual descriptions of Fuorisalone's locations, events, and sponsors, which are stored in fuorisalone.it repository, we defined a list of 300 keywords to ask Twitter and Instagram about.
- A *user-centric stream* is associated to a social network accounts and it contains all the messages posted from those accounts. We manually inserted relevant Twitter users (e.g., @fuorisalone) whose micro posts are processed by TF.

### 2.2 Mash-Up Architecture and Processing

The processing of TF relies on four data-driven steps: data integration, enrichment, analysis, and visualisation. TF exploits Semantic technologies to achieve the first three steps (i.e., Semantic Web technologies to integrate and analyse the data, and opinion mining [6] to enrich it) and on HTML5 for the fourth step.

RDF is used as common data model. As vocabulary, to which the data from the different sources can be translated, TF uses the extension of Semantically-Interlinked Online Communities ontology [7] proposed in [3]. It is worth to note

[6] Cf. `http://instagram.com/developer/realtime/`

that TF handles social streams, which are characterised by a (temporal) order and by high dynamics; to treat these data, TF uses RDF streams, an extension of the RDF model proposed in [8] and adopted by a growing community [9,10].

For instance, hereafter, we represent in RDF one of the tweets[7] that was posted during Fuorisalone:

```
[] sioc:content "Fuori salone Milano 2013 #imanartist #milanodesignweek #mdw2013 pic.twitter.com/GqR7KZc2RP";
   sioc:has_creator <https://twitter.com/WHOISBUMIN> ;
   sioc:topic :imanartist, :milanodesignweek, :mdw2013 ;
   sioc:links_to "http://pic.twitter.com/GqR7KZc2RP".
```

Data enrichment is performed through a dictionary-based sentiment classifier [5], which includes positive and negative emotion patterns. Given a micro post in English or in Italian, this technique allows to infer information about the opinion of the author (e.g., she likes or dislikes the main topic of the message). TF uses a dictionary-based sentiment classifier, because this type of classifier is known to be efficient for large-scale analysis of short texts concentrating on a single topic, such as micro posts. Moreover, since many sentiments are domain-specific, this kind of classifier is easy to adapt to the particular domain of analysis, i.e., industrial design and high tech products' launches. While this method is very suitable for large-scale analysis thanks to its minimal performance requirements, some sentiments (e.g., sarcasms, idioms) require more sophisticated methods.

The third step is the data analysis of the data collected and enriched in the previous steps. In TF, all analyses are encoded in C-SPARQL [11], an extension of SPARQL for continuous querying RDF streams as well as static RDF data sets[8]. For instance, the following C-SPARQL query counts for each hashtag the number of micro posts in a time window of 15 minutes that slides every minute.

```
1  REGISTER STREAM HashtagAnalysis AS
2  CONSTRUCT { [] sld:about ?tag ; sld:count ?n . }
3  FROM STREAM <http://.../fuorisalone2013> [RANGE 15m STEP 1m]
4  WHERE { { SELECT ?tag (COUNT(?tweet) AS ?n) WHERE { ?tweet sioc:topic ?tag . }  GROUP BY ?tag } }
```

Complex analyses can be performed combining C-SPARQL queries in a network. The continuous results of the queries are published as Streaming Linked Data [12], and are used by HTML 5 visual widgets that compose the user interface. The component that supports networking of C-SPARQL queries and Streaming Linked Data publishing is named SLD [3]. Solutions alternative to SLD are presented in [13,14].

[7] See `https://twitter.com/WHOISBUMIN/status/335490229429493760`

[8] Similar results could have been obtained using continuous RDF stream processors like SPARQL$_{stream}$ [9] and CQELS [10].

## 3 The Mash-Up

Figure 2 shows a screenshot of the Twindex Fuorisalone mash-up[9]. The top-most visual widget is a heatmap that illustrates in real time where the social streams originate from. The default view covers the entire Milano area. The users can zoom in each of the nine districts interested by Fuorisalone events by clicking on the names of the districts in the side bar on the heatmap. The users can also add to the map an overlay showing the position of the micro posts received by TF in the last 15 minutes. Clicking on the Twitter's and Instagram's icons, they can read the tweets and view the images.

Moving down, the second widget displays the tweets received on the user-centric stream. The third one is a bar chart that presents the volume of micro posts received on the geo-bound stream every 15 minutes in the last two hours. The fourth one uses an area chart to show the same information, but on a six hours time window that slides every 15 minutes. The yellow area highlights the number of tweets that refers to MDW 2013. The fifth one displays the top-10 hashtags used in the micro posts. The content of this graph depends on the zoom level of the heatmap. When zooming on a specific district, e.g., Brera Design District, it displays the top-10 hashtags used in that in the micro posts originating from that area. The last two graphs show the number of micro posts originating from the nine districts interested by Fuorisalone events using a bar char and a dot chart.

## 4 Evaluation

Twindex Fuorisalone, between April $8^{th}$ and April $17^{th}$ 2013, analysed 106,770 micro posts, and it was viewed by more than twelve thousands distinct users.

Before reporting on the ability of TF to answer the three questions proposed in Section 1, we point the reader to the area chart illustrated in Figure 2.(a). In this graph, it is straightforward to see that MDW 2013 is visible in the change of volume of micro posts. On April $8^{th}$, 2013 at 18.00 the number of tweets moves from 90/150 every 15 minutes to 180/210. For the entire duration of MDW 2013 the volume of tweets is larger than 100 tweets every 15 minutes, while normally is less than 100. During the week the number of tweets after midnight is much larger than in the normal days. The April $14^{th}$, 2013 at 20.00 MDW 2013 ends and the volume of tweets rapidly goes back under 100 tweets every 15 minutes. Moreover, the yellow area (the number of tweets that refers to MDW 2013) is more visible during the event than in the following days.

Our mashup is able to answer question Q.1: "What are the most attractive events?". The heatmap allows to visually identify hotspots where the social stream activity concentrates. Figures 3.(b) and (c) show the *hot spots* in Brera design district during a night of MDW 2013 and in a night after MDW,

[9] At the time we write this paper the mashup is still running at `http://twindex.fuorisalone.it`. Video recording of what was visible during MDW 2013 are available at `http://www.streamreasoning.org/demos/mdw2013`.

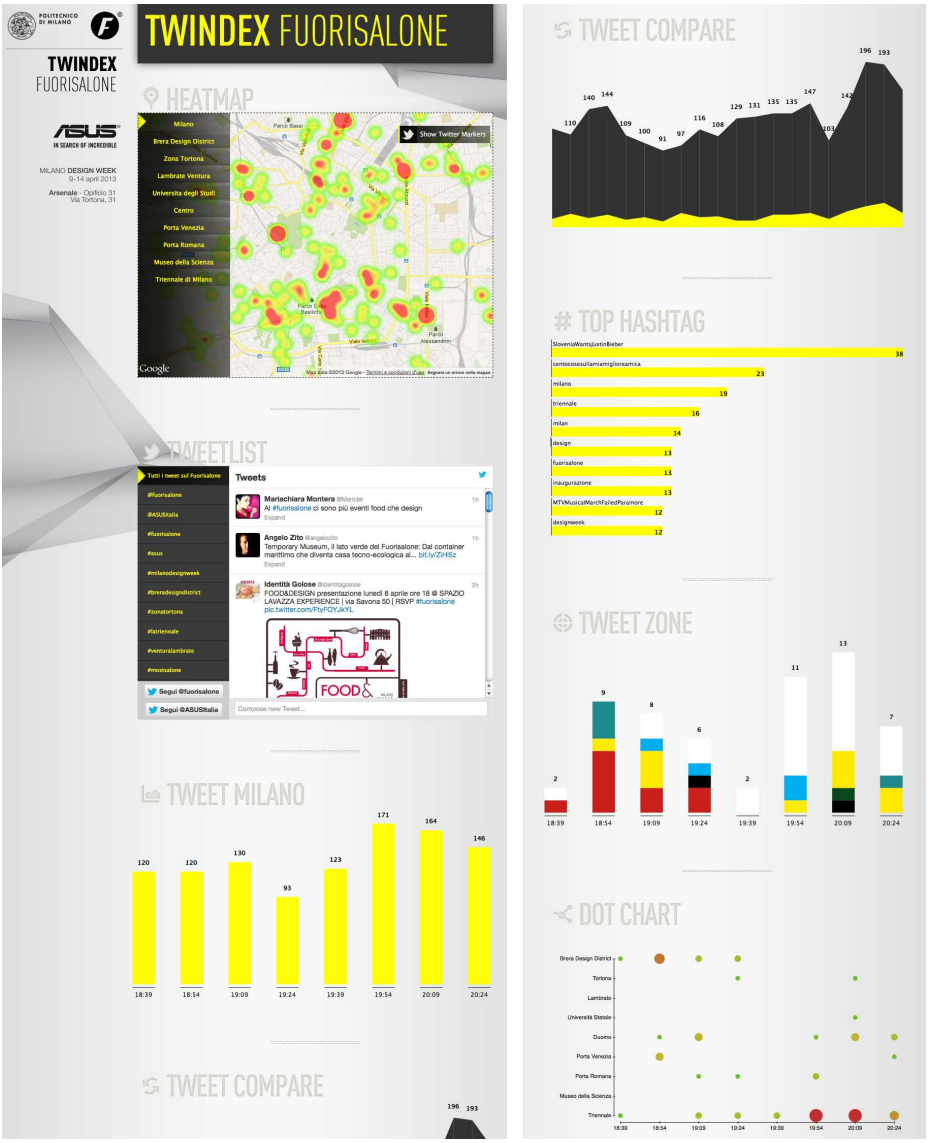


**Fig. 2.** The mashup published on the official Fuorisalone website by Studiolabo

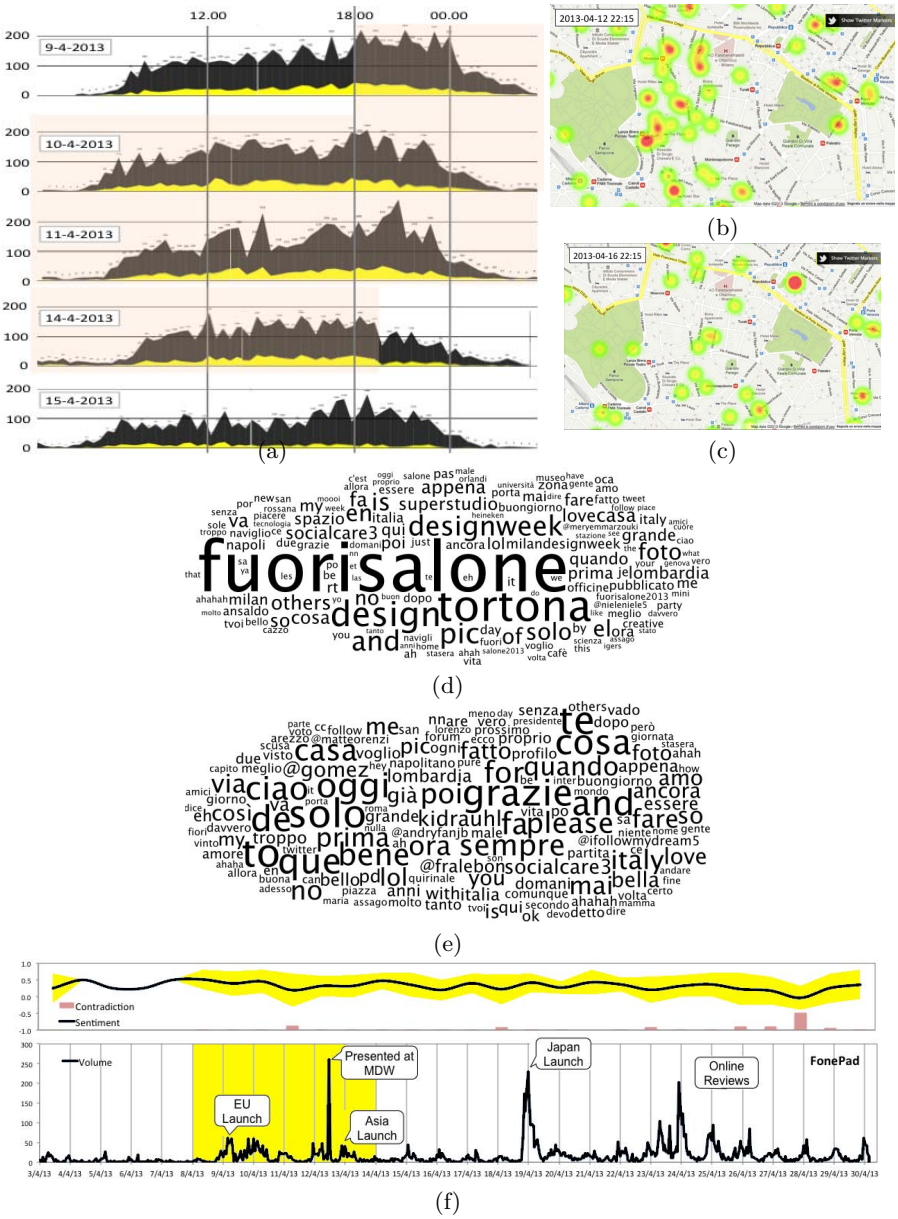


**Fig. 3.** A visual presentation of the results obtained with Twindex Fuorisalone: MDW 2013 is visible in the volume of micro posts (a); the hot spots Twindex Fuorisalone's heatmap are all in proximity of Fuorisalone venues (b) and are different from those observable in a night after MDW 2013 (c); what the visitor say (d) around the venues, from which most of the geo-tagged micro posts originate, during Fuorisalone is different from what Milano's citizens normally write in their micro posts (e); and changes were observed in public sentiment before, during and after the launch of an ASUS product during one of Fuorisalone events (f).

respectively. In a normal night, as the one in (c), few geo-tagged tweets are posted from Brera, whereas during MDW a number of hot points are visible on the map. The two most popular venues were Cesati antiques & works of art and Porta nuova 46/b; 16,653 and 13,416 tweets were, respectively, posted in their proximity. Thousands of tweets were posted in the proximity of a group of 6 other venues. Hundreds of tweets were posted around another group of 10 venues. Tens of tweets were posted on 62 venues. Around the remaining 81 only few tweets were posted.

The second question TF aims at answering is: "hat do visitors say about the events they join?". TF displays in real time the top-10 hash tags appearing in the micro posts using a bar chart as the one illustrated in Figure 3. Normally the geo-tagged micro posts related to Milano cover a variety of topics (see Figure 3.(d)), whereas during MDW 2013 the most frequently used hashtags were all related to the ongoing event. Figure 3.(e) shows a tag cloud obtained using the words[10] that appears in the 16,653 micro posts TF linked to the events hosted in Cesati antiques & works of art in Brera design district.

The number of geo-located micro posts per event was not enough to answer in real-time question Q.3: "What is the mood of the visitors before, during and after the event they join?". However, as explained in Section 2.1, TF also listens to a topic-based stream related to MDW 2013. Between April $3^{rd}$ and April $30^{th}$, TF analysed 107,044,487 micro posts and put us in the condition to answer a variant of Q.3: "What is the mood of the micro posts before, during and after an event where a new product is launched?"

Figure 3.(f) illustrates the results we obtained analysing the tweets related to *FonePad* – a product ASUS launched during MDW 2013 in one of the events in Brera Design District. Before MDW 2013, few micro posts every 15 minutes talk about FonePad; the peaks reach at most 25 micro posts every 15 minutes. During MDW 2013, ASUS announced the pre-sales in EU and presented the product. Those two facts are visible in the volume of micro posts: both the average number of micro posts and the height of the peaks are an order of magnitude larger than before MDW. In particular, the presentation[11] on April $12^{th}$ is associable to the peak of 250 micro posts in 15 minutes. Similar peaks are visible also for the Japan launch[12] and when the first online reviews[13] were published.

The sentiment expressed in the micro posts about FonePad was mostly positive. During this period some tweets about FonePad contained sentences like *"wanna buy it so bad!"*, which were classified as negative sentiments, but in reality were expressing positive sentiment. The contradiction level during such periods was also high due to a concern expressed by some users. For instance,

[10] As often done in Natural Language Processing, we filtered out typical stop words in Italian and in English. Additionally, we also excluded Milano, which is the most commonly used word in micro posts originating from Milano.

[11] See `http://www.youtube.com/watch?v=vhyktTroDTw`

[12] See `http://www.asus.co.jp/News/JWtqmBbuQsxEkukt/`

[13] See `http://www.expertreviews.co.uk/laptops/1299202/asus-FonePad` and `http://reviews.cnet.com/tablets/asus-FonePad/4505-3126_7-35619221.html`

reviews of FonePad, although very positive, caused a lot of discussions in the social media, where mixtures of positive and negative sentiments were mentioned, resulting in more contradicting distributions.

## 5 Conclusions

In this paper, we presented Twindex Fuorisalone, a mash-up that makes sense of social streams obtained from Twitter and Instagram using fuorisalone.it repository as a source of information about the events and the venues of Fuorisalone.

To cope with the streaming nature of micro posts, TF uses RDF streams and C-SPARQL within the Streaming Linked Data framework. The 106,770 tweets received during MDW 2013 as well as the 1,136,052 invocations to the Streaming Linked Data publishers were processed in main memory on a €25 per month cloud share using at most 2 CPU and 2 GB of RAM.

TF appears to be an effective solution to socially listen to Fuorisalone. TF found the events that attract the most of the visitors and observed what visitors say about them. The geo-bounded social stream lacked the volume to detect sentiment patterns, but the topic-based stream was rich enough to allow for observing changes in public sentiment before, during and after some events.

We are currently evolving TF in a generic solution to listen to city-scale events by blending social stream with mobile telecom data (e.g, number of phone calls, text messages and data connects originating from a given part of the city).

## References

1. Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., Ratti, C.: Real-time urban monitoring using cell phones: A case study in rome. IEEE Transactions on Intelligent Transportation Systems 12(1), 141–151 (2011)
2. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: WWW, pp. 851–860 (2010)
3. Balduini, et al.: BOTTARI: An augmented reality mobile application to deliver personalized and location-based recommendations by continuous analysis of social media streams. J. Web Sem. 16, 33–41 (2012)
4. Balduini, M., Della Valle, E.: Tracking Movements and Attention of Crowds in Real Time Analysing Social Streams – The case of the Open Ceremony of London 2012. In: Semantic Web Challenge at ISWC 2012 (2012)
5. Tsytsarau, M., Palpanas, T., Denecke, K.: Scalable Detection of Sentiment-Based Contradictions. In: DiversiWeb workshop, WWW, Hyberabad, India (2011)
6. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2), 1–135 (2007)
7. Breslin, J.G., Harth, A., Bojars, U., Decker, S.: Towards semantically-interlinked online communities. In: Gómez-Pérez, A., Euzenat, J. (eds.) ESWC 2005. LNCS, vol. 3532, pp. 500–514. Springer, Heidelberg (2005)
8. Della Valle, E., Ceri, S., Barbieri, D.F., Braga, D., Campi, A.: A first step towards stream reasoning. In: Domingue, J., Fensel, D., Traverso, P. (eds.) FIS 2008. LNCS, vol. 5468, pp. 72–81. Springer, Heidelberg (2009)

9. Calbimonte, J.-P., Corcho, O., Gray, A.J.G.: Enabling ontology-based access to streaming data sources. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) ISWC 2010, Part I. LNCS, vol. 6496, pp. 96–111. Springer, Heidelberg (2010)
10. Le-Phuoc, D., Dao-Tran, M., Xavier Parreira, J., Hauswirth, M.: A native and adaptive approach for unified processing of linked streams and linked data. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (eds.) ISWC 2011, Part I. LNCS, vol. 7031, pp. 370–388. Springer, Heidelberg (2011)
11. Barbieri et al.: C-SPARQL: A Continuous Query Language for RDF Data Streams. Int. J. Semantic Computing 4(1), 3–25 (2010)
12. Barbieri, D.F., Della Valle, E.: A proposal for publishing data streams as linked data - a position paper. In: LDOW (2010)
13. Phuoc, D.L., Nguyen-Mau, H.Q., Parreira, J.X., Hauswirth, M.: A middleware framework for scalable management of linked streams. J. Web Sem. 16, 42–51 (2012)
14. Gray, A.J.G., et al.: A semantically enabled service architecture for mashups over streaming and stored data. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) ESWC 2011, Part II. LNCS, vol. 6644, pp. 300–314. Springer, Heidelberg (2011)