Modeling and Reconciling Nightlife Events from Public Event Databases for the Automatic Generation of Magazines

Thomas Steiner[1], Ruben Verborgh[2], Raphaël Troncy[3],
Giuseppe Rizzo[3], José Luis Redondo Garcia[3],
Joaquim Gabarro[1], and Rik Van de Walle[2]

[1] Universitat Politècnica de Catalunya – Department LSI
08034 Barcelona, Spain
{tsteiner,gabarro}@lsi.upc.edu
[2] Ghent University – IBBT, ELIS – Multimedia Lab
Gaston Crommenlaan 8 bus 201, B-9050 Ledeberg-Ghent, Belgium
{ruben.verborgh,rik.vandewalle}@ugent.be
[3] EURECOM, Sophia Antipolis, France
{troncy,rizzo,redondo}@eurecom.fr

Abstract. The chorus of the popular song *TiK ToK* by the artist *Ke$ha*[1] goes *"Don't stop, make it pop. DJ, blow my speakers up. Tonight, I'mma fight. 'Til we see the sunlight. Tik tok on the clock. But the party don't stop, no"*. We all know, however, that each nightlife event, be it a party, concert, or bar evening, comes to an end eventually. With NiteOutMag, we present a Chrome Web application that can help people revive nightlife events in the recent past. Among the younger generation, nightlife activities—just like any other activity—together with related multimedia data get shared online on social networks. The problem is that for one and the same event, the event-related user-generated data may be shared on a plethora of social networks. Therefore, with this paper, we introduce an application that extracts, reconciles, and models events from several event databases or calendars, social data from multiple social networks, and media from some photo and video sharing platforms. The collected data is attached to events held in a given area and further processed to generate an event-centric magazine where each page represents an event illustrated by media items.

Keywords: Event Illustration, Event Reconciliation, Media Finder, Magazine

1 Introduction

In March 2012, 901 million monthly active users of the social networking site (SNS) Facebook have uploaded more than 300 million photos on average per day [4]. Many of those photos are event-related and often illustrate nightlife

[1] http://www.mtv.com/videos/keha/445445/tik-tok.jhtml

events such as music concerts that a user has attended. Facebook, however, albeit the biggest, is only one among a plethora of social networks that people use to share event-related content. A social network is an online service or media platform that focuses on building and reflecting social relationships among people sharing interests and/or activities.

In this paper, we consider eleven different social networks that represent all together most of the Western world's market share: Google+ (`google.com/+`), Myspace (`myspace.com`), Facebook (`facebook.com`), Twitter (`twitter.com`), Instagram (`instagram.com`), Flickr (`flickr.com`), YouTube (`youtube.com`), yfrog (`yfrog.com`), Twitpic (`twitpic.com`), and MobyPicture (`mobypicture.com`), img.ly (`img.ly`).

The actual detection of events based on information from SNSs is out of scope for this paper. We observe that this has become a very active research field. For example, Petkos *et al.* propose to detect events using a multimodal clustering approach [16]. In this paper, we will consider event databases or calendars that maintain listings of past and upcoming events using crowd sourcing techniques and that can be queried programmatically. More precisely, we consider four event databases, namely Eventful (`eventful.com`), Foursquare (`foursquare.com`), Upcoming (`upcoming.org`), and Google Places (`google.com/places`). In this paper, we propose an approach to fully automatically generate an event-centric magazine where each page corresponds to a known event that is illustrated with media shared on the Web. The application is available online for testing purposes at `http://users.restdesc.org/rgverbor/tmp/NiteOutMag.crx`. A screencast showcasing the application is available at `http://youtube.googleapis.com/v/5pEiy2RHBjY`.

The remainder of this paper is organized as follows. Section 2 presents some related work aiming at creating summaries and narratives of events using social media. Section 3 details our approach: we start from a given location and we collect the events and their illustrative media in a recent past before assembling a digital magazine. In Section 4, we evaluate our system and we discuss the results of user study. Finally, we give our conclusions and outline future work in Section 5.

2 Related Work

Capturing life moments and building narratives using social networks is addressed in [1], where the authors investigated about the interaction between event stories and the use of social networks to tell them. They proposed Storify [21], a Web application which supports users to perform story telling and in particular: (1) sorting and organizing the items of an experience similar to the elements of a story, (2) communicating and discussing strategies on how to guide a user towards an intended experience. The overall storytelling creation is supervised by the user, who describes the story as a crafted experience [10]. Streams of news navigate through the social platforms such as Twitter, YouTube. Getting the big picture from them is the objective of Storyful [22]. This application allows the user to navigate through the story created by other users or to create his own aggregating the content available from Twitter, Youtube and adding more.

Events have a key role in the human life and illustrating events with media items stemming from social networks is a very active research topic. In [2], Brenner and Izquierdo present an approach to detect social events and retrieve associated photos in collaboratively annotated photo collections combining data of various modalities such as time, location, and textual and visual features. In [14], Liu *et al.* present a method for combining semantic inferencing and visual analysis for automatically gathering photos and videos illustrating events. Regarding aesthetic aspects of media compilation, in [18], Sandhaus *et al.* consider visual and aesthetic features for the automatic creation of photo books. Obrador *et al.* use visual and aesthetic features for a category-based approach to automatically assess the aesthetic appeal of photographs [15].

3 NiteOutMag Implementation

In this section, we provide the implementation details of the NiteOutMag application. It is implemented as a Chrome application since this technique provides a transparent statement about the required resources of the application, *i.e.*, cross-domain access to event databases and search APIs.

3.1 Geocoding

The application focuses on events that occurred in a certain area. While event databases tend to work based on GPS points like (40.7590110, -73.98447220), *i.e.*, a pair composed of a latitude and a longitude, human beings use more human-friendly place names such as *Times Square, New York*. Geocoding is the process of finding the geographic coordinates (generally expressed as latitude and longitude) associated to other geographic data such as an address composed of a street name and a zip code. The user can enter an exact address, like *618 W. 46th St, New York, NY 10036*, or a very broad area like *Las Vegas*. The returned latitude and longitude pair is always the center of the particular area of interest. In our implementation, we use the Google Geocoding API [7].

3.2 Event Search and Event Categories

Based on a latitude and longitude pair, we search for events that took place near this location (in a radius of five kilometers) in the recent past using the respective event search APIs from Eventful, Foursquare,Google Places, and Upcoming [3,5,8,23].

Event databases or calendars typically classify events in different categories. Our focus with NiteOutMag being nightlife events, wherever possible we filter events to match a manually compiled list of nightlife-type categories. For Foursquare, this includes the categories *nightlife* and *arts & entertainment*.

For Eventful, this includes the categories *music*, *festivals_parades*, and *singles_social*. For Upcoming, this includes the categories *Music*, *Performing/Visual Arts*, *Media*, *Festivals*, and *Comedy*. For Google Places, no such event filtering is possible. Google Places, as the name of the service suggests, focuses on places. If an event is known to take place at a place, its event data are returned by the service. In consequence, for Google Places, rather than filtering on event categories, we filter on place categories to include typical nightlife places such as clubs, discothèques, bars, movie theaters, etc.

3.3 Event Modeling with Schema.org

The event databases we are harvesting use different event modeling schemas. In order to have a common internal event representation, we use `schema:Event` properties [9] from the schema.org vocabulary, which are also mapped to `dbo:Event` from the DBpedia Ontology [11] or the LODE ontology [20]. We maintain a straight-forward alignment of the particular event database properties from Eventful, Foursquare, Upcoming, and Google Places to schema.org properties. Concretely, we support the following properties:

- `name` The name of the event.
- `description` A short description of the event.
- `location` The location of the event.
- `startDate` The start date and time of the event (in ISO 8601 date format).
- `endDate` The end date and time of the event (in ISO 8601 date format).
- `url` The URL of the event.
- `image` The URL of an image of the event (typically, the poster).

At this stage of NiteOutMag, we do not expose these event representations externally. However, we plan to make a better use of the EventMedia dataset in the future [13].

3.4 Event Reconciliation

We sanitize event titles by removing HTML tags, punctuation, and common abbreviations such as *w/*, which becomes *with*; or *ft./feat.*, which becomes *featuring*. Event directories can overlap in their coverage. Therefore, we reconcile the returned set of events based on the simple, yet effective Levenshtein distance of the event titles with an empirically determined editing distance of five. Furthermore, if two event start times are also within a window of one hour, we regard the events as the same. We recall that due to our event search approach outlined in Subsection 3.2, events are already geographically close. In consequence, we are capable to detect that two events with similar event titles such as *Titanic* and *Titanic 3D* and start times of 8:00PM and 8:30PM correspond most probably to the same movie event.

3.5 Media Search

Unlike event search, media search in our application is *not* based on geolocation. According to a study performed in May 2012 by Seline [19], only roughly 1% of all microposts on Twitter are geotagged. In addition to that, as our approach covers more SNSs than Twitter, we need to consider that not all SNSs support geotagging of content. In consequence, we use a two-tier approach for media search. We illustrate this by a real event of a Beach Boys concert[2] in the Merriweather Post Pavilion in Columbia, MD. The event title is *Beach Boys*, the venue name is *Merriweather Post Pavilion* and the area is *Columbia, Maryland (MD)*. We define as *human-friendly address* the first part up to the comma of the `formatted_address` field in the Google Geocoding API [7], *i.e.*, *Columbia* in this case. Human beings aware of the local context do not need to disambiguate, *i.e.*, there is no need to explicitly state *MD*. The two-tier approach consists thus of a full-text search for:

- *(i)* the (event title) + the (venue name), and
- *(ii)* the (event title) + the (human-friendly address).

We have implemented a media collector consisting of media extractors for all SNSs listed in Section 1. This media collector has been described in [12,17]. It takes a search term as input, *e.g.*, *Beach Boys Columbia* for the example above, and then performs a parallel full-text search using the search APIs of all SNSs. When all the media extractors have responded, a unified SNS-agnostic output is delivered. We propose a common alignment schema, illustrated in Listing 1, which shows the resulting metadata for an exemplary media item. Fig. 1 depicts the overall architecture of the media collector.

```
{
  "mediaurl": "http://t.co/ubagkiR9",
  "storyurl": "https://t.co/dau5pBWD",
  "message": {
    "text": "My wife, who once said her vision of
        heaven was being on the lawn bouncing
        around beach balls at a Beach Boys
        concert [...]
        <a href=\"http://t.co/pH2OV1fW\" >
        http://t.co/pH2OV1fW</a> [...]",
    "clean": "My wife, who once said her vision of
        heaven was being on the lawn bouncing
        around beach balls at a Beach Boys
        concert [...]"
  },
  "user": "https://t.co/gIf2UVZR",
  "type": "photo",
  "timestamp": 1329342797000,
  "published": "2012-02-15T21:53:17.000Z"
}
```

Listing 1. Sample output of the media collector showing a Google+ post (edited for legibility, URLs shortened).

[2] `http://www.thebeachboys.com/#tour`

3.6 Magazine Layout

In order to create the illusion of a real magazine, the ability of flipping from page to page needs to be as credible as possible. The open-source library `turn.js` by Emmanuel García [6] allows for the dynamic creation of magazines solely based on HTML5 technologies with a very realistic page flip effect. We treat each event as one page of the magazine, add the event metadata as headline and subheading of the page, and arrange the images and accompanying microposts to fill up the page as illustration. For each event, we use the image with the highest resolution as background image of the page in order to create a lifestyle magazine appearance. For additional print-like look and feel, we use drop caps (Fig. 2). The title page gets dynamically created based on a Google image search for the (human-friendly address) + the keyword (nightlife).

3.7 Installation Instructions

The NiteOutMag application requires the Chrome browser. Download the application from the URL `http://goo.gl/fjuE5` to the local file system. Drag and drop it into the `chrome://chrome/extensions/` page in the browser. When you drop it on the extensions page, you will notice an install option popping up. When you agree to install, you will see the standard installation dialog that informs you about the rights that the application is requesting. It needs access to the event databases and media search APIs we have mentioned before.

4 Evaluation

We have conducted a short user study via an online questionnaire[3]. Overall, 12 participants (all male) with an average social media experience of 3.9 (all values are between 1 and 5) have evaluated NiteOutMag. We asked participants to evaluate the application for the cities of Las Vegas and Hong Kong. Las Vegas is known to have good event coverage in the event databases we are harvesting,

[3] `https://docs.google.com/spreadsheet/viewform?formkey=dEY1aVJfSk5aQnVPRzBlSmVIUEVTa1E6MQ`

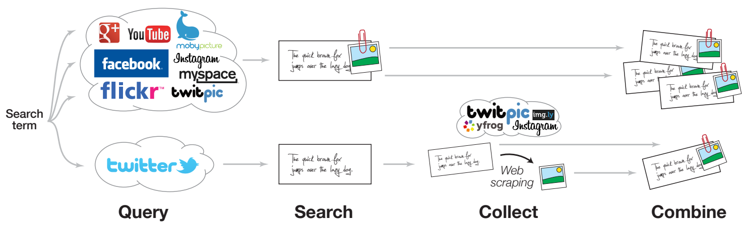

Fig. 1. Overview of the media collector: hybrid approach for the media item extraction process using a combination of API access and Web scraping

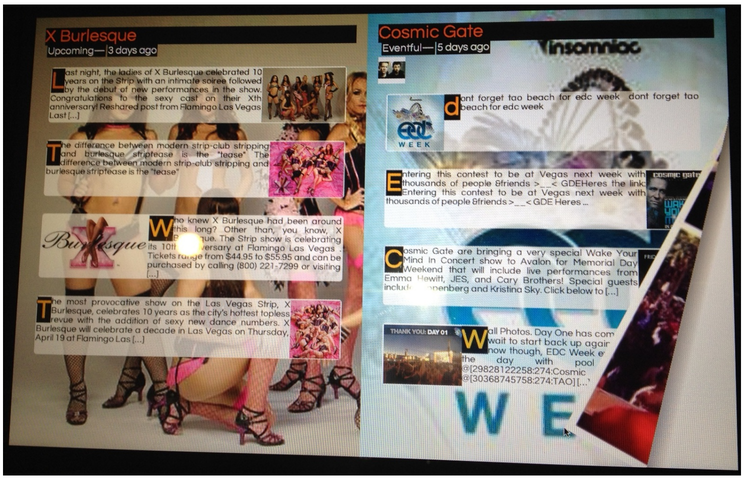

Fig. 2. Screenshot of the NiteOutMag application with two events from Las Vegas and the righthand-side page about to be flipped

whereas Hong Kong is known to have bad coverage. Further, we asked them to evaluate the application with a city of choice. Our evaluation criteria were *speed* at which a magazine for a given city was generated, *usefulness* of the selected events, *relevancy* of the selected images for these events, and *attractiveness* of the events and media. Further, we allowed for optional free-form answers to the following questions: “How could NiteOutMag be improved?”, “How many pages did you count in NiteOutMag?”, and “How many event sources did you count in NiteOutMag?”. The last two questions enable us to check whether the application worked as expected.

4.1 Speed, Usefulness, Relevancy, and Attractiveness

We have evaluated the criteria *speed*, *usefulness*, *relevancy*, and *attractiveness* for all cities in an accumulated way. The raw anonymized user feedback from our user study is available for further analysis[4]. The results are analyzed in details below.

Speed: 2.3
: Participants were not convinced by the magazine generation speed. Unfortunately, due to the limited number of HTTP requests, the only way to address this issue is to generate pages more on demand rather than all at once, with drawbacks in the page flipping speed.

Usefulness: 1.9
: The very low rating of the usefulness of the selected events suggests that we need to improve the event filtering and be more selective with the categories introduced in Subsection 3.2.

[4] `https://docs.google.com/spreadsheet/ccc?key=0AtLlSNwL2H8YdFR3ZDd4YVp4WE5Ed3FwTThLU3I5RUE`

Relevancy: 2.3
Participants rated the relevancy of the selected media items for the events quite low. Upon further observation and after checking back with the participants, we could trace this down to a negative bandwagon effect with false positive media items. When only one media item was obviously bad, participants punished the whole event media item set with a negative rating, albeit not the whole set was bad.

Attractiveness: 2.3
What most limits the impression of a real magazine is the limited amount of coherent text for many events. More advanced language analysis is required to potentially generate coherent summaries out of several microposts. In a first step, non-English microposts can be easily removed. According to the full-text responses, some users also had concerns with the selected page background image whose implementation is detailed in Subsection 3.6.

4.2 Number of Pages and Number of Event Sources

For Las Vegas, on average, participants counted 10.3 pages in their magazines and 2.0 event sources, while for Hong Kong, participants counted only 1 page and 0.9 event sources. For the city of choice, participants counted 4.2 pages from 1.3 event sources. This reflects the initial observation that Hong Kong has bad, and Las Vegas good event coverage, with the random cities of choice ranging in between.

4.3 Suggested Improvements

The most demanded feature was to add more content to the magazine. More event sources and more social networks are ways to address this request. As pages are generated and removed dynamically, sometimes, when not enough media content is found for an event, pages get removed when a user is looking at them. This causes confusion, however, is owed to the inflexible API of `turn.js` [6]. We will investigate ways to add pages only once it is known that enough media content is available. A recurrent suggested feature is to load page contents only on demand rather than all at once for the whole magazine, which is the current implementation. Again the API of `turn.js` and the desire to allow for fast page flipping makes it hard to find a good compromise. However, a re-implementation of the page flipping code can take this request into account. Participants clearly expected a back cover and a table of contents, however, the current implementation does not provide those. We will add a back cover that takes an even and odd number of pages into account, and create a table of contents and also add headlines to the magazine's front cover.

5 Conclusion and Future Work

The results stand and fall with the quality of the events in the event databases. One of the problem we encounter are event titles such as *Kandyland 2012 @*

the Playboy Mansion – 1.877.VIP.MANSION[5] which combine the event title with the venue name and a vanity contact phone number. The detection of such composed titles is difficult and requires advanced linguistic processing. Event reconciliation is a second issue. The two movie events *The Avengers* and *Marvel's The Avengers 3D* are the same for a human being. However, for a machine, the duplication is harder to detect. Media item reconciliation improvement is a task that we have left for future work.

The system's response time is improvable. The main issue here is the browser-enforced maximum number of simultaneous HTTP requests. With our current four event sources that we have limited to ten events per source, we already have up to forty events that we need to search media items for. As outlined before in Section 3.5, we have opted for a two-tier approach in order to improve the recall, which effectively means two search requests per event. Assuming that for each search request we get on average only two media items, we have roundabout $5 + 40 * 2 * 2 = 165$ HTTP requests for only one magazine. We will investigate ways to improve the application's responsiveness in the future.

In this paper, we have reported on a Chrome Web application that illustrates events via four event databases and eleven social networks and media sharing platforms. The application harvests event-related data on-the-fly for a user-determined center of interest and compiles the data in an aesthetic lifestyle magazine. Using popular party destinations, big cities, and places of interests, we have evaluated the retrieved results for both relevancy and visual appeal, with a special focus on optimized recall. While there are actionable issues for future work, we are quite happy with the outcome so far.

Acknowledgments

This work was partially supported by the European Commission under Grants No. 248296 FP7 I-SEARCH, No. 287911 FP7 LinkedTV and the French National Agency under contract ANR.11.EITS.006.01, "Open Innovation Platform for Semantic Media" (OpenSEM).

References

1. B. Atasoy and J.-B. Martens. STORIFY: a tool to assist design teams in envisioning and discussing user experience. In *Annual Conference on Human Factors in Computing Systems*, pages 2263–2268, 2011.
2. M. Brenner and E. Izquierdo. Social Event Detection and Retrieval in Collaborative Photo Collections. In *2nd ACM International Conference on Multimedia Retrieval (ICMR'12)*, Hong Kong, 2012.
3. Eventful, Inc. API Method /events/search, 2012. `http://api.eventful.com/docs/events/search`.
4. Facebook, Inc. Key Facts – Facebook Newsroom, 2012. `http://newsroom.fb.com/content/default.aspx?NewsAreaId=22`.

[5] `http://t.co/YUm7FtmE`

5. Foursquare Labs, Inc. Venue Events, 2012. https://developer.foursquare.com/docs/venues/events.
6. E. García. turn.js: Make a Flipbook with HTML5, 2012. http://www.turnjs.com/.
7. Google, Inc. The Google Geocoding API, 2012. https://developers.google.com/maps/documentation/geocoding/.
8. Google, Inc. The Google Places API (Experimental), 2012. https://developers.google.com/maps/documentation/places/#PlaceSearches.
9. Google, Inc. and Yahoo, Inc. and Microsoft Corporation. Event – schema.org, 2012. http://schema.org/Event.
10. M. Hassenzahl. Experience Design: Technology for All the Right Reasons. *Synthesis Lectures on Human-Centered Informatics*, 3, 2010.
11. M. Hausenblas and R. Cyganiak. DBpedia to Schema.org terms, 2012. http://mappings.dbpedia.org/server/ontology/export.
12. H. Khrouf, G. Atemezing, G. Rizzo, R. Troncy, and T. Steiner. Aggregating Social Media for Enhancing Conference Experiences. In *1st International Workshop on Real-Time Analysis and Mining of Social Streams (RAMSS'12)*, pages 34–37, Dublin, Ireland, 2012.
13. H. Khrouf and R. Troncy. EventMedia: a LOD Dataset of Events Illustrated with Media. *Semantic Web journal, Special Issue on Linked Dataset descriptions*, 2012.
14. X. Liu, R. Troncy, and B. Huet. Finding Media Illustrating Events. In *1st ACM International Conference on Multimedia Retrieval (ICMR'11)*, Trento, Italy, 2011.
15. P. Obrador, M. Saad, P. Suryanarayan, and N. Oliver. Towards Category-Based Aesthetic Models of Photographs. In *18th International Conference on Advances in Multimedia Modeling (MMM'12)*, pages 63–76, 2012.
16. G. Petkos, S. Papadopoulos, and Y. Kompatsiaris. Social Event Detection Using Multimodal Clustering and Integrating Supervisory Signals. In *2nd ACM International Conference on Multimedia Retrieval (ICMR'12)*, Hong Kong, 2012.
17. G. Rizzo, T. Steiner, R. Troncy, R. Verborgh, and J. R. Garcia. What Fresh Media Are You Looking For? Extracting Media Items from Multiple Social Networks. In *1st International Workshop on Socially-Aware Multimedia (SAM'12)*, Nara, Japan, 2012.
18. P. Sandhaus, M. Rabbath, and S. Boll. Employing Aesthetic Principles for Automatic Photo Book Layout. In *17th International Conference on Advances in Multimedia Modeling (MMM'11)*, pages 84–95, 2011.
19. C. Seline. What percentage of Twitter photos are geotagged?, 2012. http://www.quora.com/What-percentage-of-Twitter-photos-are-geotagged.
20. R. Shaw, R. Troncy, and L. Hardman. LODE: Linking Open Descriptions Of Events. In *4th Asian Semantic Web Conference (ASWC'09)*, 2009.
21. Storify. Storify – Find the best of social media, 2012. http://storify.com/.
22. Storyful. Storyful, 2012. http://storyful.com/.
23. Yahoo!, Inc. Upcoming API Documentation—event.search [HTTP GET], 2012. http://upcoming.yahoo.com/services/api/event.search.php.