Virtual Celebrator Machine

Loi Huynh[1], Peter Stanchev[1,2]

[1]Kettering University, Flint, USA
[2]Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
`{huyn2714,pstanche}@kettering.edu`

Abstract. There has been a huge growth of social network in the recent years. This trend does not only allow us to get connected and share the information in an efficient way, but also reveals some potential beneficial in dealing with several social issues, such as earthquake detection, social spam detection, flu pandemic tracking, media monitoring, etc.
In this paper, we propose a new way of utilizing social network. By implementing what is called a Virtual Celebrator Machine (VCM), we are able to let everyone who has connection with this machine in term of social networking be able to share their cultural experience and points of view about certain social events locally or globally. In that way, we provide a way to reinforce the relationship and connection between people virtually, which, we believe, would help to flourish cultural heritage preservation.

Keywords: Social Network, Event Celebrator, Twitter, Culture Preservation.

1   Introduction

Undergoing a tremendous development in the past decade, Social Network appears to public as a tool to get connected with others and the community effectively. Social Network has shown a strong impact on many different disciplines of human-beings in the 21st century. One of the fastest growing social network communities, Facebook, with 1.11 billion subscribers up to today [6], helps to add the increments of €15.3 billion into European economy by providing 232,000 jobs of making apps and of apps distributors [4]. Social media like LinkedIn, one of the most popular networks with about 168 million users, helps to create the connections between potential employers and employees around the world [5].

Asides from Facebook, Twitter reveals its potential with totally 500 million users and 135,000 new tweets are posted every day [4] [6]. Its API (Application Programming Interface) is open source, with full of documentations, and is free to use. The other feature of Twitter is that it shows the diversity of the world. Whenever someone tweets something in his Twitter homepage or to other people, the tweet contains the geographic location and other information about the sender (nationality, language, etc.).

The social tagging has also been utilized in cultural heritage domain. It is one of the methods that people use to stimulate the connection between social network and

cultural heritage, helping to boost people's cultural awareness and responsibility in preservation [7]. In addition, cultural heritage can also be appreciated via serious game, when people use high-quality gaming (combined with real-time graphics and artificial intelligence) to support people explore the culture [8].

In this paper, we underscore a new aspect of using Social Network in human communication and in cultural preservation. We utilize Twitter in organizing a Virtual Celebration Machine (VCM) to help people around the world stay tuned in certain social event.

The idea of celebrating social event together comes from the fact that human's nature makes them to be curious of the way foreign cultures interact and communicate in a certain activity. Because everyone has different background, experience, and culture, it is fair to say that they have a variety of ways to celebrate the same social event.

The rest of the paper is organized as follow. We explain the model for the VCM that we implement in part 2, illustrate the realization of the model in part 3, show the experimental results in part 4, and conclude in part 5.

2   The Virtual Celebrator Machine Model

The way that we will host a virtual social event celebration on the internet is by creating a Virtual Celebrator Machine (VCM) named `@Social_event` where `Social_event` can be replaced with any real events. `@Social_event` is a regular Twitter account, created to serve a particular social purpose. Essentially, all messages that are sent to `@Social_event` will be recorded by this VCM. In other words, VCM, by using the Twitter APIs, gains control over this account and collects the event information from all relating users. The machine uses that information to create responses to user that are querying those questions. The VCM interface is shown in figure 1.
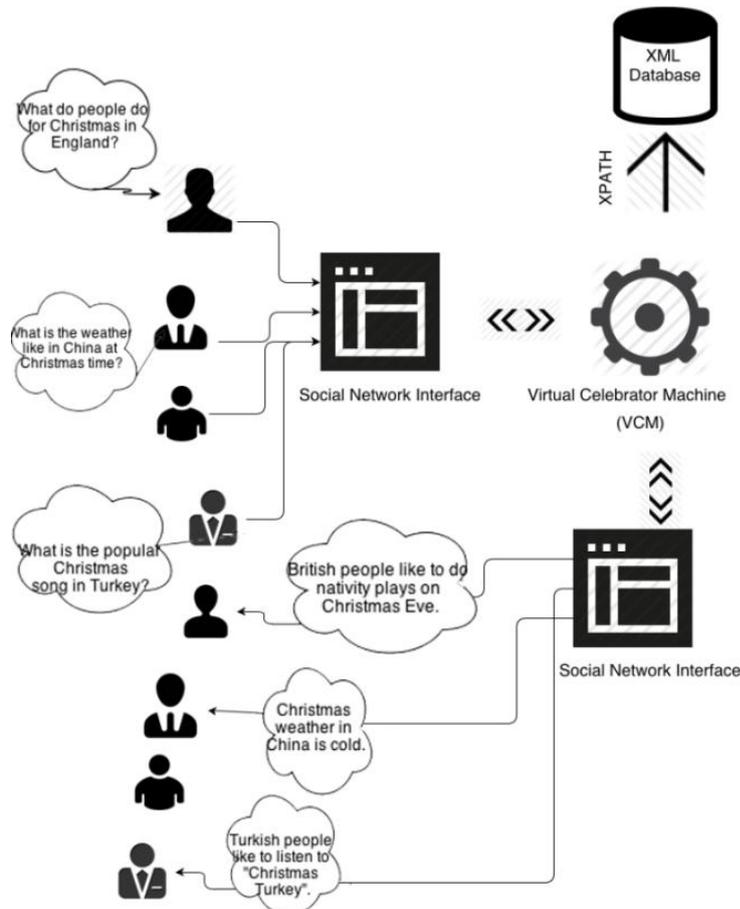
Fig. 1. VCM Interface

2.1 Social Tweeting Definitions and Formats

We will use the following definitions and formats:

Tag: The group of consecutive non-white-space characters in a tweet.

Social Tweet: Is the tweet that makes sense to VCM (etc. the tweet that are match with our pre-defined prototype) and it is presented in the following format: `<@Social_event> <#query> <location> <arg1> <arg2>...<argn>`, where `@Social_event` is the *Twitter* message format that starts in every tweet sent to others; `<#query>` is defined as a string starts with the pound key (#) and followed by the keyword, such as `#ask, #define, #vote`. As Twitter also use `#key` to highlight the keyword for searching purposes (hash-tag), we decided to use it to format our **Social Tweet**; `<location>` specify the location (town, city,

country, continent, etc.) where the users who send the social tweet represent for; `<arg1> <arg2> <arg3> ... <argn>`:form the list of arguments that can be passed in a *Social Tweet*. There are usually 1-2 arguments in the current implementation. Examples for the list of `<argn>` are: `<nativity plays> <Flint> <America> <British> <London> <shopping location> <Christmas song>`, etc.

2.2   Virtual Celebrator Machine Database Structure

We manage this huge database using appropriate database model. The database serve for two basics functionality: receive queries from users and make responses to those queries. Our challenge is to find the best matched response to the queries posted by the users. Hence, it is necessary that we classify the input into the right slot of information.

We use XML (eXtensible Markup Language) and XPATH (XML Path Language) to solve a complicated database structure in implementing VCM.

The Current model supported the structure:

```
<VCM> -> <Topic_0> ->  <tag_0>...</tag_0> -> </Topic_0>
...   <tag_n>...</tag_n>
       -> <Topic_m> -> <tag_0>...</tag_0> -> </Topic_m>
..   <tag_n> ...</tag_n>
</VCM>
```

Voting for the best Social Tweet: Every tweet that has been entered into the VCM should be evaluated by other users, so that the system knows to keep the significant information and get rid of the uninformative one. After every #ask query has been answered, VCM also asks the users to evaluate that response. A score from each individual user will be accumulated in evaluating the overall score of the response. The VCM also has a score threshold of deleting the tweet if the information is rated as "uninformative" by the users.

Periodical Information Broadcasting: Because the data structure we use is based on location path into the property tags, it is possible to combine those tags and retrieve the content of each tag against the location. For example, users might want to know the popular Christmas song of different regions in the world, by filter the data against tag "Christmas song", we can come up with the list that look like:

```
<Christmas song>
<New York, N.Y> Fairytale of New York
<Los Angeles, CA> Christmas in Los Angeles
<Philadelphia, PA> A Christmas Carol
<Istanbul, Turkey> Christmas Turkey
<Moscow, Russia> Russian Male Choir of Moscow
</Christmas song>
```

To help individuals find interest in a certain social event, we broadcast the above information once over a certain period of time on the VCM account home page.

Automatic Story Recording and Telling: VCM knows who are the active users, among those who send queries that have significant contribution to the community of that social event. VCM might then send an invitation to those prominent members to write a Christmas story about their representative regions. The story will be stored internally in the VCM machine, and retrieved on its homepage as the headlines. Users again have the chance to exchange experience which is even more meaningful through story telling. The only drawback of this method is the limitation for the length of a single tweet (140 characters). There are several suggestions in the Social Net-working community to split the long tweet into smaller lengths via a third party appli-cation.

Addresses conformity: The locations that people specify in the tweet to `@Social_event` are not uniform. There are many ways that people mean Flint in Michigan, United States:

```
flint, mi; flint Genesee, Michigan; Flint , mi, US; flint
mi united state of America; flint, mi USA
```

In order to yield an appropriate reply, VCM interpret those addresses as one. In this regard, we choose *Google MAP APIs* because it is globally considered as one of the prominent geographical helpers. A call to **http://maps.googleapis.com/ maps/api/geocode/json** asking a conformity way to refer to the address **Flint, Gene-see MI United State** yield unified couples of options (Table 1).

Table 1. Two ways to refer to Flint, Genesee MI USA

Option #1	Option #2
{} 0 ■ long_name : "Flint" ■ short_name : "Flint" ⊞ [ ] types {} 1 ■ long_name : "Genesee" ■ short_name : "Genesee" ⊞ [ ] types {} 2 ■ long_name : "Michigan" ■ short_name : "MI" ⊞ [ ] types {} 3 ■ long_name : "United States" ■ short_name : "US" ⊞ [ ] types formatted_address : "Flint, MI, USA"	{} 0 ■ long_name : "Flint Township" ■ short_name : "Flint Charter Town: ⊞ [ ] types {} 1 ■ long_name : "Genesee" ■ short_name : "Genesee" ⊞ [ ] types {} 2 ■ long_name : "Michigan" ■ short_name : "MI" ⊞ [ ] types {} 3 ■ long_name : "United States" ■ short_name : "US" ⊞ [ ] types formatted_address : "Flint Township, MI

Each node of each solution has a short name and a long name variation. By picking one of them as the prototype, we are able to come up with a unified tree that describes all of the possible variations of the unique place "Flint, Michigan". In short, by using

Google MAP API, we get rid of confusion caused by: comma in the location, capitalization, use of abbreviation, use of other name for the same place.

We use the tree structure of the XML to manage navigating from the most global places to the most local places. For example, from USA to Flint, the places that we are going through in the middle of the tree structure are Michigan and Genesee. In case, when there is more than one option for the same address, for example "Flint, MI" and "Flint Township, MI" as described in the Address Conformity section above, we reuse the node Michigan and Genesee. So, in the tree structure, 'Flint' and "Flint Township" are the siblings and they have the common parent Genesee and the common grandparent Michigan. For uniformity, in the tree structure we use *long_name* notation for the place. To obtain flexibility, the above diagram suggests us that the suitable implementation for the data is tree-like structure XML (eXtensible Markup Language).

Tagging order: Users list the parameters tags (`<location>`, `<keyword>`, `<content>`) are in the tweet not in a unique order. XPATH (XML Path Lan-guage), which is the query language for selecting nodes from XML documents, help us with XML tree navigation. By changing the order of the tags in a full XPATH statement, we can navigate to every single location of the database tree without any barriers in tags-ordering. We generate the XPATH for each of those permutations. Then we can retrieve whether or not there is information associated with those XPATH in the tree.

3 Virtual Celebrator Machine Realization

For the current implementation, users are free to enter any characters into the tweet and, we have decided to use double quote (") in the Tweet to separate our tags. In case the tag is only one word, the users do not have to include the double quote in the tweet. The event that we pick for the implementation is Christmas and the Social Network System is Twitter. The VCM allows people to share their personal or region-wise Christmas knowledge in any categories they prefer to share, such as activities, shopping places, traditional Christmas movie or music, etc.

`@Christmas_Celeb` is the account that we created in Twitter, which acts as the VCM. Suppose someone in Flint, MI, USA would like to know what the favorite Christmas song of people in Beijing China is, he can find the answer by sending a *Social Tweet* to `@Christmas_Celeb` and get a response in his Interaction Pane of the Twitter Window. Below are the two current query types that are embedded in the VCM for `@Christmas_Celeb`.

#define Statement. This statement uses the information from the users' tweets and adds that information into the database. `@Christmas_Celeb`: `#define <location> <keyword> <content>`.

Sample Statement: **@Christmas_Celeb**: **#define "London, England" "activities" "nativity plays; Christmas music"**. *Note:* In the example above, the tag "activities" does not have to include the double quote (Table 2).

Table 2. Examples for tags in the #define statement

Tag	Example
Location	Flint, MI; Shanghai, China; Tokyo, Japan; London, England
Keyword	activities, shopping places, traditional Christmas movie
Content	"sending cards", "Nativity Plays", "Joyous Tidings"

#ask Statement. This statement acts like a query that pulls the information out of the database, and retrieves it back as a social tweet and put it into the Interaction Pane for the users. `@Christmas_Celeb: #ask <location> <keyword>`.

Sample Statement: `@Christmas_Celeb: #define "London, England" "activities"`.

3.1 Implementation Environment for the Virtual Celebrator Machine

We have implemented the VCM machine in MATLAB, because MATLAB has highly well-functioning tools in dealing with URL signal (posting and getting method). Moreover, MATLAB has the user command line interface (similar to Command Shell), that allow intermediate interaction. On the other hand, **AOuth** service is used to embed the secure authorization process which interfaces with the Twitter APIs, and URL-like command line **Curl** is to post requests to Twitter. The limitation for reading message from Twitter API is 15 times/15 minute [9]. We have set our update and response rate to be 1 every 1 minute. Hence, the machine is able to answer the Social Tweet request from user in the range of 1 minute.

4 Experiments with the Virtual Celebrator Machine

The experiment has been conducted and tested by 10 people with different backgrounds: USA, China, United Kingdom, and Vietnam. The survey is conducted about VCM's functionality, the friendliness of the Twitter interface between users and VCM, and lastly people's eagerness to see VCM developed in the future.

Here are some of the testimonies that we received from the users:

- "I like the speed of responding from **@Christmas_Celeb**, it takes less than a minute to see the update tweet in my interaction pane."
- "When there was something wrong with my syntax in my tweet to **@Christmas_Celeb**, it gave me some directions on how to fix the syntax rather than just spitting out the non-readable error messages."
- "It was cool when I entered two different formats for the addresses. I entered Flint MI for the first time, and enter FLINT Michigan for the second time, the message that I got back from the queries to **@Christmas_Celeb** are the same."
- "The thing that I prefer most about **@Christmas_Celeb** is the flexibility of the addresses."

Here are some statistics from our survey. About 90% of the people satisfy with the functionality of the two social tweet statements #define and #ask. Around 80% of people believe that the responses from VCM's interface (syntax, feedback, tweeting style) are user-friendly. And almost 100% of the people would like to see VCM developed in the future.

5   Conclusion and Future Work

Noise word in the tags of the tweets handling: As the tag in a tweet can be misspelled or annotated differently. "3people" instead of "three people"; "NEW YORK METS" instead of "NEW YORK MEATS". Our jobs is to make the system smart enough that it will interpret those tags as one or two things that are highly related. This problem will be in our future research. In the meantime, some of the online API likes Fuzzy-Wuzzy: Fuzzy String Matching in Python or Dictionary.com API seem to be a good starting point to get rid of those noise in searching.

The paper discusses the new aspect of using social network in celebrating a social event. Ultimately, we are trying to create a database communication gateway between social networking and traditional database system. We introduce the new conception of VCM (Virtual Celebrator Machine) embedded into the social network system, present several problems in dealing with a natural language as the main inputs. The VCM has been implemented and tested on Matlab. Future research will add more functionality to VCM, and make it a powerful in cultural information exchange machine.

References

1. Semertzidis/ CERTH/ITI, T. (2010). Social Networks Overview: Current Trends and Research Challenges. Publications Office of the European Union.
2. Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th international conference on World wide web (WWW '10). ACM, New York, NY, USA, 851-860. DOI=10.1145/1772690.1772777 http://doi.acm.org/10.1145/1772690.1772777
3. Ryong Lee and Kazutoshi Sumiya. 2010. Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks (LBSN '10). ACM, New York, NY, USA, 1-10. DOI=10.1145/1867699.1867701 http://doi.acm.org/10.1145/1867699.1867701
4. Brussels. (2012, Jan 24). Facebook. Measuring Facebook's economic impact in Europe. Retrieved May 11, 2013, from http://www.facebook.com/notes/facebook-brussels/measuring-facebooks-economic-impact-in-europe/309416962438169
5. LinkedIn statistics, number of LinkedIn users & demographics, http://www.socialbakers.com/linkedin-statistics/
6. Smith, C. (2013, 06). Expanded Ramblings. How Many People Use the Top Social Media,

7. Apps & Services? Retrieved May 17, 2013, from http://expandedramblings.com/index.php/resource-how-many-people-use-the-top-socialmedia/

8. Falk Anderson, E., McLoughlin, L., Liarokapis, F., Peters, C., Petridis, P., & De Freitas, S. (2009). Academia. Serious Games in Cultural Heritage. Retrieved from http://www.academia.edu/191398/Serious_Games_in_Cultural_Heritage

9. Oomen, J.,Belice Baltussen, L. Limonard, S., van Ees, A., Brinkerink, M., Aroyo, L., Vervaart, J., Asaf, K., Gligorov, R. Emerging practices in the cultural heritage domain-social tagging of audiovisual heritage. Retrieved from https://www.zotero.org/susannaanas/items/itemKey/6FQ8UVSQ

10. Twitter Developers. REST API v1.1 Limits per window by resource-Twitter Developers. Retrieved June 01, 2013, from http://dev.twitter.com/docs/rate-limiting/1.1/limits