

Using Social Media Posts as Knowledge Resource for Generating Life Stories

Robee Khyra Mae J. Te
De La Salle University
Manila, Philippines
robee_te@dlsu.edu.ph

Janica Mae M. Lam
De La Salle University
Manila, Philippines
janica_lam@dlsu.edu.ph

Ethel Ong
De La Salle University
Manila, Philippines
ethel.ong@delasalle.ph

Abstract

People thrive in storytelling as a means of sharing their experiences with one another and learning about the world through narratives. With the advancement of technology, social media platforms have become a common medium for people to recount and to share stories about their life events anytime and anywhere as they happen. Automated story generation systems can be designed to utilize this vast knowledge resource. In this paper, we describe the approach we used in sourcing storytelling knowledge from social media posts, specifically Facebook, to extract events that can be used to form a life story. We also describe briefly our strategies in classifying the posts based on their textual content to find snippets of life events. These posts are represented in an event model that is then used by our story generator to produce a personal life story that chronicles travelling, dining and celebrating activities. Results from user evaluation showed that Facebook data can provide enough information to generate a person's life story which introduces him/her to others through the narration of some life events that happened in his/her life.

1 Introduction

People are natural storytellers. We use stories to recount events in our lives, and to share them with our families and friends. The advent of mobile and online technologies has moved this practice to social media platforms, enabling one to post their adventures and discoveries anytime and anywhere as they are taking place.

The past decade has also seen the development of software agents designed to mimic human abilities in story generation. While bulk of the challenge in story generation is in planning the content and sequencing the events that will comprise the story (Solis et al., 2009; Ang et al., 2011), the amount of information available for a computer to source its knowledge from is also a major consideration on the variances of stories that computers can generate. As Meehan (1977) described it, “the lack of knowledge, often of the most mundane variety” gave rise to the production of “wrong” stories. This led researchers in story generation to employ varying strategies to provide story generators with the knowledge base they need, as described in Cua et al. (2010), Yu and Ong (2012), and Ong et al. (2018).

Various studies have identified the different types of knowledge needed for story generation, which include domain-based knowledge about our world, operational knowledge on story structures and how to write stories (Rishes et al., 2013), and linguistic knowledge on how to express the abstract story representation into surface text in a form that humans can understand. Domain-based knowledge comprises attributes and behaviors that story characters may embody, descriptions of objects that may exist in the story world; as well as events and their causal, temporal and spatial relations.

Finding appropriate resources to supply story generation systems with domain-based knowledge continues to be a challenge despite the availability of large-scale semantic knowledge bases such as Cyc (Lenat et al., 1985), ConceptNet (Singh et al., 2004) and SUMO (Cua et al., 2010). Alternatives

are corpora of human-crafted stories as sources of story ideas, as described in (McIntyre and Lapata, 2009) and (Daza et al., 2015).

Volumes of short snippets of events one encounters in their daily activities, from personal triumphs and failures, to wishes and goals, are being shared online through social media platforms. This poses an opportunity to utilize user data as resource for the generation of a person’s life story. However, user generated posts tend to be noisy, are varied, and may or may not provide sufficiently relevant data. To be usable, a series of tasks need to be performed on the posts that involve the collection, preprocessing, classification and extraction of relevant information.

The main contribution of our work is in utilizing social media posts that people share in their Facebook pages to generate one’s personal life story. Satish, Jain and Gupta (2009) posits that “a storyteller has lots of event-related data in his collection, but must select only a few that are most relevant considering his audience”. Among the myriad of posts that a user shares in their Facebook account, our life story generator needs to be able to select those that it can use. We describe briefly our strategy in classifying the posts based on their textual content to find those containing traveling, dining and celebrating events. Relevant information are then extracted from these posts and represented into an event frame for use by our story generator to produce stories that more resemble a life story. We end our paper with an analysis of the feasibility of social media posts as a potential source of storytelling knowledge.

2 Life Stories and Events

A story is “a pattern within which events or series of events can be ordered, understood and communicated” (Livo and Rietz, 1986). This organized narrative of events enable readers to make a meaning and to understand the world around them. A life story is a non-fictional story that contains a personal narrative on the significant events and experiences in one’s life (Titon, 1980), as well as his/her personal details that include birthday, family members, educational background and work experiences. Preferences and interests may also be described.

Facebook contains various data ranging from texts to photos and videos. It presents numerous sto-

ries, facts, and events happening from users all over the world. Facebook requires users to create visible profiles providing their name, gender, date of birth, and email address. In addition to this, other information such as their contact details, personal interests, educational and work background, family members, and favorites can also be added at their own discretion (Nadkarni and Hofmann, 2012). Facebook also allows interaction among its users through simple status updates, posts or shares that inform their whereabouts and actions.

Stories are comprised of sequences of events describing character actions to achieve some story goals and to effect some changes in the story world. In the case of a life story, events represent any activity that a user, the protagonist, has shared in his/her social media account. This event has occurred at a specific location, during a particular time period, and may involve other characters and objects.

Events are usually denoted in stories as verbs, e.g., *read* (a book), *see* (a movie), *dine* (at a restaurant), and *wait* (at an airport). Nouns can also be used to denote events that take place in stories, such as celebrations and vacations, and naturally-occurring phenomena like *flood* and *earthquake*.

A post can contain more than one sentence, and each of these may be associated with zero to several events. In a single post, it is possible that one may describe who he/she travelled with, how he/she celebrated his/her birthday, and what he/she ate with his/her friends, summing up to three different events. Such posts are split into independent sentences to extract the individual event details and are represented in an event frame of the form:

verb (doer, object, tagged, date, location)

3 Gathering Storytelling Knowledge

Given the target personal life story, textual content from 21,412 social media posts of 216 user accounts that include descriptions of activities, locations and time, and named entities were gathered by utilizing Facebook’s Graph API. No specific sampling techniques and keywords were used in extracting the posts. All posts from one’s account, excluding the shared posts, were extracted with full disclosure from the users. Ethical considerations from the use of the participant’s personal data were taken into ac-

count. Specifically, an informed consent form was prepared to provide participants with details on the procedures for data collection, storage and disposal, and how confidentiality of data and anonymity of the participant will be preserved during the study.

The gathered posts were annotated using all 20 predefined activities of Facebook. The top four (4) most frequently used categories, namely *Celebrating*, *Travelling*, *Eating*, and *Drinking*, were then chosen as labels. Posts that fall under the other 16 categories were automatically tagged as *Others*. The resulting dataset is comprised of 193 posts about *Eating*, 53 on *Drinking*, 409 on *Travelling* and 643 on *Celebrating* events. Manual inspections of the dataset showed the abundance of posts that conform to the characteristics described in the work of Kinsella, Passant, and Breslin (2011), containing foreign characters, emoticons, laughter expressions and hashtags. These currently have no bearing to the event classifier model and were thus removed during preprocessing.

Posts with multiple sentences were split into their constituent sentences using Stanford CoreNLP and classification was performed on the individual sentences. However, Stanford CoreNLP highly depends on the use of periods as end-of-sentence markers. Thus, a post written in the form of a list (e.g. “1. Hi 2. Hello”) instead of “1. Hi” and “2. Hello” is split into three sentences: “1.”, “Hi 2.”, and “Hello”.

3.1 Event Classification

Stanford CoreNLP is also used to identify POS tags and to generate a constituent and dependency representation, which is used for syntactic analysis to extract elements from the post. A verb is used to signify the activity described in the post and the object represents the receiver of the action. Lemmatization is also performed in order to increase the efficiency of the classifier. However, these do not work for posts that do not have explicit verbs. Such posts will have to rely on the event classification algorithm to determine its category and to identify the verb that can be used in the text generation module.

A simple classifier model based on keywords is used to classify events. A reference table containing the keywords related to each event category was derived from different knowledge bases such as WordNet (Miller, 1995) and ConceptNet (Liu and

Singh, 2004) and through the manual inspection of the dataset. The keyword matching algorithm makes use of a scoring system with a threshold value of 2. The event category with the highest score is then used to label the post.

If a token matches a word in the keywords list, then a score of 1 will be added to the total score of the keyword’s category. Consider the post “*We are going on a trip to check off an item from our bucket list.*”, it has a score of 3 in the *travelling* label because of the presence of the lemmatized keywords “*go*”, “*trip*”, and “*bucket list*”. The threshold value was set to 2 because majority of the posts contained at least 2 keywords under the same category such as “*Merry Christmas*” and “*eat pizza*”. Setting the threshold to 1 would increase the likelihood of classification as many posts such as “*Ready, set, go!*” only contain a single keyword. However, increasing the threshold value to 3 would result in a high false negative.

3.2 Performance of the Classifier

The posts with no verbs were fed to the classifier to assess their performances. Table 1 shows the performance of the score-based classifier for each of the four categories of events.

Classifier	Precision	Recall
Travelling	8.47%	83.33%
Eating	30.77%	100.00%
Drinking	20.00%	80.00%
Celebrating	65.47%	98.91%

Table 1: Score-based Classification Performance Results on Event Categories

The *Celebrating* category achieved the highest precision because the posts stated the events more explicitly compared to the other categories. Setting the threshold value to 2 did not affect the performance as *Celebrating* posts contained at least 2 keywords. *Drinking* and *Travelling* categories achieved lower precision because posts in these categories have implied events through the use of proper nouns such as the name of the restaurant, dishes and the like. In the example posts in Listing 1, post #1 contains the keyword “Happy” from the restaurant name “Happy Lemon” and the keyword “dessert”,

thus can be classified as either a *Celebrating* or an *Eating* post. Post #2, on the other hand, contains the keyword “At” and the word “Go” from the restaurant name “*Cake 2 Go*”, thus, led to its classification as a *Travelling* event.

Post # 1 - I am satisfied with my Happy Lemon
for dessert.
Post # 2 - At Cake 2 Go #KainPa
Post # 3 - Happy friendversary thesismate!

Listing 1: Sample posts that led to misclassification because of noisy data.

3.3 Event Frame Generation

Given a sample post “Going to Japan for a short vacation.”, the following steps are performed to extract the information that will be needed during the generation of life stories:

1. Extract the verb(s) that signifies the action described in the post, as well as the subject or doer of the action, the object(s) or recipient of the action, which may be another Facebook user, and the location where the event took place.
2. Apply lemmatization to transform the verb to its lemma for use in event classification.
3. Generate an event frame of the form *verb (doer, object, tagged, date, location)* containing details extracted from the post.

If the *doer* is not mentioned in the post, similar to the given example, the user who made the post is assumed to be the doer. The *tagged* and *location* can be left blank (N/A) if these are not indicated in the original post. The corresponding event frame for the given example is

travel(Mae, Japan, N/A, 04/26/2018, N/A)

For posts that do not explicitly contain verbs, the verb is determined based on the post’s classification. *Drinking* posts is assigned the verb “*drink*”, “*eat*” is used for *Eating* posts, “*travel*” for *Travelling* posts, and “*celebrate*” for *Celebrating* posts. Consider the post - “*Happy birthday to my best friend!*”, the assigned verb is “*celebrate*” and the object is “*birthday*”. Assuming the user tagged her friend “*Ja*” and the location “*Manila*” in the original post, then the resulting event frame is:

celebrate(Mae, birthday, Ja, 05/01/2018, Manila)

4 Generating Life Stories

A life story has three components, namely introduction, body and conclusion. The content of each component was based on the article of Davis (n.d.). The introduction contains facts or direct knowledge, such as details about one’s name, birthday location, family, educational background and work experiences that introduces the user. The body contains the life events extracted from the posts that tells a story based on the significant events that happened in the user’s life. The conclusion contains one’s likes and interested events to describes the person’s preferences and interests.

Two approaches were considered in story generation - template-based and grammar-based. Template-based generation is rigid and requires effort in the creation of the templates (Tuffield et al., 2006). With the use of RDF data to construct descriptive sentences dynamically (Sleimi and Gardent, 2016), grammar-based text generation provided greater flexibility and scalability.

4.1 RDF Triples

Resource Description Framework (RDF) data consists of (*subject-predicate-object*) triples such as (*John, nationality, American*) (Lassila and Swick, 1999). The *subject* indicates the resource, while the *predicate* indicates the trait or aspect of the resource to describe the relationship between the *subject* and the *object*. As illustrated in Figure 1, RDF data can be represented as a graph where edges are labelled with predicates, and vertices with subject and object resources.

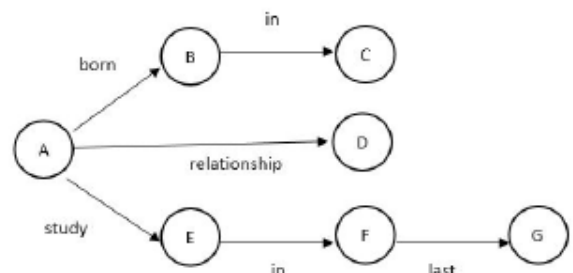


Figure 1: An example of a graph representation of RDF data.

During story content planning, RDF triples are

derived from assertions representing data sourced from a user's account. Example assertions are shown in Listing 2. The RDF triples are then used to form the basic messages comprising a story text.

```
(1) person(lastName, firstName, middleName)
(2) livingIn(obj)
(3) family(relationship, names<>)
(4) education(institution, type, yearGrad,
  course)
(5) work(institution, startDate, endDate,
  location)
(6) eventGoing(name, location)
(7) eventInterested(name, location)
(8) likes(category, page<>)
```

Listing 2: Example assertions for the *Introduction* (#1-5) and the *Conclusion* (#6-8).

4.2 Grammar-Based Generation

A set of grammar rules were defined for each component of the life story. Listing 3 shows the grammar rules for the *introduction* part.

```
<INTRODUCTION> -> <SENTENCE>+
<SENTENCE> -> <subject> <PREDICATE>
<PREDICATE> -> <verb> <OBJECT>
<OBJECT> -> <noun> [<prep-phrase>*]
  || <article> <noun> [<prep-phrase>*] ||<prep-phrase>
```

Listing 3: Grammar rules for the *Introduction*.

A bottom-up strategy is used to fill-up the grammar rules with data. Those with no content are then removed or excluded from the resulting story text.

The process proceeds as follows. First, each of the assertions shown in Listing 2 is checked to determine if they can be filled with data from the user's Facebook account. The data is retrieved from the *About Me* section which includes the user's name, gender, birthday, family members, educational and work background, and current location. Filled assertions, such as those shown in Listing 4, are then transformed to corresponding RDF triples as shown in Listing 5.

```
(1) person("Smith", "Mary", "")
(2) gender("Female")
(3) birth("1992-06-15", "New York")
(4) livingIn("New York")
(5) work("Pizzeria", "2013-01-20",
  "2018-02-10", "New York")
```

Listing 4: Assertions filled with Facebook data for the *Introduction*.

```
(1) ("Mary Smith", "is", "female")
(2) ("Mary Smith", "born", "1992-06-15")
(3) ("Mary Smith", "lives", "New York")
(4) ("Mary Smith", "works", "Pizzeria")
```

Listing 5: RDF triples for the *Introduction*.

The grammar rules shown in Listing 3 are then applied to generate individual sentences comprising the story text, as illustrated in Listing 6. Sentence aggregation concatenates the resulting sentences to form the *introduction* paragraph, while pronouns are generated to replace references to the user's name in succeeding sentences.

```
(1) Mary Smith is female.
(2) She born on 1992-06-15.
(3) She lives in New York.
(4) She works at Pizzeria from 2013-01-20 to
  2018-02-10.
```

Listing 6: Sentences formed from RDF triples.

The resulting *introductory* paragraph for the given example is shown in Listing 7.

```
Mae Ang, born on May 25, 1996, got her high school diploma from Midtown High School last 2013. She has yet to get her college diploma from Xavier Institute. She worked from September 01, 2013 to April 30, 2014 at University Student Council. She is from Manila, Philippines. She is the daughter of Ian Quintin and Ann Ang.
```

Listing 7: Sample *Introductory* paragraph.

The *conclusion* provides a summary of the user's preferences, such as the pages they *Liked* (e.g. *Local Business, Company Organization, Brand, Artist, Music, Movies*, among others), and *Going and Interested Events*. Still using *Mae*, a sample *conclusion* paragraph is shown in Listing 8.

```
Mae likes Communities such as Technology Impact Summit 2017, RVR COB Week 2017, and Ann-yeong Oppa; Artists such as Calleftgraphy, Park Shin Hye, and Song Hye Kyo; and TV shows such as The Flash, Moonlight Drawn by Clouds, and Descendants of the Sun.
```

```
Mae attended Cybersecurity and International Relations, Publication Writing Workshop, Study Assembly Christmas Party at Tower Residences in Manila, and Technology Summit 2016 Forum at Xavier Institute in Manila.
```

Listing 8: Sample text for the *Conclusion*.

4.3 Generating the Body of the Story

Bulk of the work of the story generator lies in formulating the contents for the *body* of the life story. The contents are sourced from the event frames that were previously derived from the classified posts, as detailed in Section 3.3. These events are organized and sequenced according to their topical and temporal relations. Topical relations are based on event categories; one paragraph in the story represents one event category. Within a paragraph, the narrative is presented based on the most recent event to the least recent event determined by the timestamp linked to the original post.

In the early iterations of the story generator, a sequence of event frames produced stories that contain text with redundant sentences of the form:

On <date>, <subject> <verb> <object> with <tagged friends> in <location>.

To address this, sentence aggregation groups together events with closer temporal relations (i.e., same date, same month or same year), or with the same set of tagged people. Given two events that happened on February 2017, instead of generating “*On February 14, 2017, Mae celebrated friendversary thesismate with Cam. On February 07, 2017, Mae celebrated with Jamie.*”, the story text would instead contain “*On February 2017, she celebrated with Jamie and Cam.*”

Listing 9 shows the resulting *body* paragraph for the *Celebrating* activities of user *Mae*, using the posts found in Table 2.

On February 2017, Mae celebrated with Cam and Jamie. On August 16, 2016, She celebrated with Shane in Manila, Philippines. On October 03, 2014, She celebrated Angie.

Listing 9: Sample text for the *Body* paragraph.

4.4 Challenges in Generation

A number of errors present in the generated story text are caused by mistagged words. This is because the Stanford POS Tagger encountered challenges in properly tagging non-dictionary words, gerunds, mixed languages, and colloquial language.

Consider post # 3 in Listing 1 that contains the non-dictionary word “*friendversary*”, a word invented by Facebook to describe the yearly recurrence of the date that two people first became friends

Original Post	Metadata	
Happy 18th Angie!	date created	10/03/14
Party party!	date created	08/16/16
	user tagged	Shane
	location	Manila, Philippines
Happy anniversary Jamie HAHA	date created	02/07/17
	user tagged	Jamie
Happy friendversary thesismate!	date created	02/14/17
	user tagged	Cam

Table 2: Sample Facebook posts classified as celebrating, and their Metadata

on the social media app. Due to this, the text generator considered the phrase “*friendversary thesismate*” instead of just “*friendversary*” as an object of the verb and yielded the sentence “*Mae celebrated friendversary thesismate with Cam.*”

Gerunds or words ending with “-ing”, such as “*eating*” in the post “*I love eating.*”, are often tagged as verbs instead of nouns. Posts containing mixed languages is also common in countries where English is not the first language, as shown in post #2 in Listing 1, with the word *KainPa* (keep eating). The use of onomatopoeia or words formed from the imitation of associated sounds (e.g. “*oink oink*” and “*hahaha*”), and colloquial language such as posts containing abbreviation (e.g. “*c u l8r*” – originally “*see you later*”), further compounded the problems.

Social media posts also tend to be brief and written in informal language, with users posting snippets of incomplete or context-based glimpses of their life events. This may lead to event frames with missing information. To address this, verbs are assigned for posts with implicit activities, and the Facebook user is assigned as the default doer or subject of the post. However, when the object is missing, the sentence is automatically omitted without making any assumption regarding the recipient of the action. Posts with missing tagged friends and location are tolerated as the generated story text can still convey the life event even without these two information.

Some users also insert hyperlinks and attachments to provide additional context to their posts (Kinsella et al., 2011). These are currently ignored by the classifier and the event frame generator.

5 Evaluation of Resulting Stories

The evaluation of the resulting life stories proceeded as follows. 12 participants, aged 18 years old and above, were briefed about the features of the system. Each participant then logged in his/her Facebook credential to start the extraction process from his/her account. Once the life story has been generated, the participant evaluates the language composition of the resulting life story to assess its overall sentence structure, flow, grammar and readability.

Separate criteria for evaluating each of the components of the story were also used. For the introduction and conclusion paragraphs, the criteria focus on the correctness of the extracted information (i.e., user’s profile, hobbies and interests) and the flow of the paragraphs. For the body paragraphs, the focus of the evaluation is on the correctness of the classification of events and extracted event details, and the temporal sequence of events. Each criterion was rated with a score from 1 to 4; 1 being the lowest and 4 being the highest. Lastly, comments and suggestions for improvement were gathered for future improvement of the stories.

The average scores for the language composition criteria is shown in Table 3. The average scores showed how diverse Facebook is in terms of the use of mixed languages making it difficult for the POS tagger to identify the event details. Another problem was the improper use of upper and lower cases (e.g. “GoIng To The MaLL.”) causing the POS tagger to not adapt to the capitalization of the words. Life event categories which only contain few sentences, some of which do not even make sense, caused the paragraphs to have sudden shifts in topic, thus affecting the flow of the story. The overall scores for the introduction, body and conclusion paragraphs were 3.83, 3.38 and 3.79, respectively. The introduction and conclusion-specific scores were better than the body-specific scores because these two parts used facts that require little to no text understanding and analysis compared to the data used in the body. The low average score for the body-specific criteria was caused by the misclassification and extraction of incorrect event details of the posts, producing sentences that are difficult to comprehend and may even contain incomplete thoughts.

The story that garnered the highest evaluation

Criteria	Average Scores
Grammar	2.50
Correct usage of punctuation	3.17
Correct usage of pronoun	3.17
Capitalizations	3.58
No redundant Information	2.92
Flow of Sentences	2.58
The story can be considered as an autobiography	2.83

Table 3: Average Scores for the Language Composition Section

score was from a Facebook user who does not post any life event. Thus, the user found no missing information and all the data extracted from his/her profile, Liked pages, and interested events were properly used in the *introduction* and *conclusion* paragraphs. The story flow is also easy to understand because the assertions were simpler and facts were easily inserted to the grammar rules.

The challenges encountered in generating the *body* paragraph(s) due to the misclassification of the posts were prevalent during user evaluation. In one of the cases, user posts contain promotional details such as inviting his/her friends to a particular activity, and sentimental posts such as insights to or rants about a product. These kinds of posts often contain keywords that are common among all four event categories, resulting in the misclassification of the posts. Most of these posts also do not contain any check-ins, tagged friends and other information that could help the story generation know more about the event.

The context influencing the comprehension of the original post also affected the generated stories. Posts with sarcastic tones and double meanings (or double entendre) imply different denotations. The post “*Enjoy your vacation!*” can both have a positive and a negative sentiment which can only be recognized by the sender and the receiver. Terms like “*cooking*” can also be conveyed in different ways such as the act of preparing food and an informal basketball term.

It is also evident that users share opinions, quotes, sentiments or lyrics more than they share actual life event posts. An observation by Cavalin, Moyano, and Miranda (2015) showed that only a small percentage, specifically, 4.65% of their dataset is classified as event posts. Similarly, our dataset only recorded 6.06% life event posts under the categories of eating, drinking, celebrating and traveling.

6 Conclusion and Further Work

People use their vast repository of daily experiences and interactions to share stories with one another. For computer systems to exhibit the same behavior with their human users, they must be able to utilize a large knowledge resource containing descriptions of things found in our everyday activities and our world. In recent years, people resort to the use of social media platforms to share events and things that interest them. This poses an opportunity to investigate the sufficiency and extent of information that can be derived from these systems.

In this paper, we studied Facebook's potential as a knowledge resource for story generation. In particular, we look at a user's personal information, likes, events, and posts in providing the needed knowledge, specifically about story characters, their attributes, events (activities and actions), and the locations where these take place. Utilizing natural language processing approaches and tools, we performed a sequence of steps to collect, pre-process, classify and extract data from the posts to transform them into a form that our story generator can use.

While Facebook was considered as a candidate resource for story generation due to the free-form nature and amount of data present, it was evident that the content used in producing a life story is highly dependent on the output from the event classifier and event frame generator modules. The performance of the classifier was affected by several challenges that were encountered in working with the noisy characteristics of user-generated content. These include missing information, use of mixed languages and colloquial expressions, and the presence of hyperlinks and emoticons alongside textual content.

Analysis of the resulting life stories showed that the quality of the content in the body paragraphs is dependent on a number of factors, including the

amount of event-related posts found in a user's account, the presence of the needed information that can be extracted from these posts, and the linguistic form that a user adheres to when composing a post. Only 6.06% of the dataset contained descriptions of events, and only a small fraction of these posts provide sufficient data from which useful details about the event could be extracted. Future work can consider processing the metadata from embedded objects that typically accompany a textual post. Online resources can also be referenced to determine if a given text is a quote, an excerpt from some documents or books, or a song lyric. Doing so can provide additional context regarding the posts that may help the classifier in correctly identifying an event's category. The story generator can also take advantage of this knowledge to improve the narrative.

Further analysis of the posts show users commonly split a single activity or event into different posts as they constantly post updates regarding their status. Thus, instead of classifying a post in isolation, collective classification strategies should be investigated to take advantage of this noticeable dependency between the new post to previous posts. This dependency can highlight the topical and temporal relations that exist between events, and which the sentence aggregation rely on to improve the coherence of the story.

References

- Karen Ang, Sherie Yu, and Ethel Ong. 2011. Theme-Based Cause-Effect Planning for Multiple-Scene Story Generation. *International Conference on Computational Creativity*, 48–53, April 27-29 2011, Mexico City, Mexico.
- Paulo Cavalin, Luis Moyano, and Pedro Miranda. 2015. A Multiple Classifier System for Classifying Life Events on Social Media. *Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. Institute of Electrical and Electronics Engineers, 1332–1335. <https://doi.org/10.1109/ICDMW.2015.182>
- Jeffrey Cua, Ruli Manurung, Ethel Ong and Adam Pease. 2010. Representing Story Plans in SUMO. *Proceedings of the NAACL Human Language Technology 2010 Workshop on Computational Approaches to Linguistic Creativity*, 40–48, June 5 2010, Los Angeles, USA. ACL.

- Spencer Hope Davis. (n.d.). Contents of an Autobiography. Retrieved from <https://penandthepad.com/contents-autobiography-8627982.html>
- Angel Daza, Hiram Calvo, and Jess Figueroa-Nazuno. 2016. Automatic Text Generation by Learning from Literary Structures. *Proceedings of the NAACL Human Language Technology 2016 5th Workshop on Computational Linguistics for Literature*, 9–19.
- Sheila Kinsella, Alexandre Passant, and John G. Breslin. 2011. Topic Classification in Social Media Using Metadata from Hyperlinked Objects. *European Conference on Information Retrieval, ECIR 2011 (LNCS 6611)*, P. et al. Clough (Ed.). Springer, Berlin Heidelberg, 201–206.
- Ora Lassila and Ralph R. Swick. 1999. Resource Description Framework (RDF) Model and Syntax Specification.
- Douglas B. Lenat, Mayank Prakash, and Mary Shepherd. 1985. CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks. *AI magazine*, 6(4), 65–85.
- Hugo Liu and Push Singh. 2004. Commonsense Reasoning In and Over Natural Language. *International Conference on Knowledge-Based Intelligent Information and Engineering Systems (LNCS 3215)*. Springer, Verlag Berlin Heidelberg, 293–306. https://doi.org/10.1007/978-3-540-30134-9_40
- Norma J. Livo and Sandra A. Rietz. 1986. *Storytelling: Process and Practice*. Littleton, Colo.: Libraries Unlimited.
- Neil McIntyre and Mirella Lapata. 2009. Learning to Tell Tales: A Data-driven Approach to Story Generation. *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1*. Association for Computational Linguistics, 217–225.
- James R. Meehan. 1977. TALE-SPIN, An Interactive Program that Writes Stories. *International Joint Conference on Artificial Intelligence*, Vol. 77, 91–98.
- George A. Miller. 1995. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11), 39–41.
- Ashwini Nadkarni and Stefan G. Hofmann. 2012. Why Do People Use Facebook?. *Personality and individual differences*, 52(3), 243–249.
- Dionne Tiffany Ong, Christine Rachel De Jesus, Luisa Katherine Gilig, Junlyn Bryan Alburo and Ethel Ong. 2018. Building a Commonsense Knowledge Base for a Collaborative Storytelling Agent. In Yoshida K., Lee M. (Eds.), *Knowledge Management and Acquisition for Intelligent Systems, PKAW 2018, LNCS 11016* 1–15, August 28-29, 2018. Springer, Cham.
- Ehud Reiter and Robert Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press, New York.
- Elena Rishes, Stephanie M. Lukin, David K. Elson, and Marilyn A. Walker. 2013. Generating Different Story Tellings from Semantic Representations of Narrative. In H. Koenitz, T.I. Sezen, G. Ferri, M. Haahr, D. Sezen, and G. Catak (Eds.), *International Conference on Interactive Digital Storytelling (ICIDS 2013) (LNCS 8230)*, 192–204. Springer International Publishing, Switzerland. https://doi.org/10.1007/978-3-319-02756-2_24
- Arjun Satish, Ramesh Jain, and Amarnath Gupta. 2009. Tolkien: An Event Based Storytelling System. *Proceedings of the VLDB Endowment*, 2(2), 1630–1633.
- Push Singh, Barbara Barry, and Hugo Liu. 2004. Teaching Machines About Everyday Life. *BT Technology Journal*, 22(4), 227–240.
- Amin Sleimi and Claire Gardent. 2000. Generating Paraphrases from DBpedia using Deep Learning. *Proceedings of the 2nd International Workshop on Natural Language Generation and the Semantic Web (WebNLG 2016)*, 54–57.
- Candice Solis, Joan Tiffany Siy, Emerald Tabirao, and Ethel Ong. 2009. Planning Author and Character Goals for Story Generation. *Proceedings of the NAACL Human Language Technology 2009 Workshop on Computational Approaches to Linguistic Creativity*, 63-70, Colorado, USA.
- Jeff T. Titon. 1980. The Life Story. *The Journal of American Folklore*, 93(369), 276–292.
- Mischa M. Tuffield, Dave E. Millard, and Nigel R. Shadbolt. 2006. Ontological Approaches to Modelling Narrative. *2nd AKT DTA Symposium 2nd AKT DTA Symposium*.
- Sherie Yu and Ethel Ong. 2012. Using Common-Sense Knowledge in Generating Stories. In P. Anthony, M. Ishizuka and D. Lukose (Eds.), *12th Pacific Rim International Conference on Artificial Intelligence (PRICAI 2012)*, LNAI 7458, 838–843, September 5-7 2012, Malaysia. Springer, Heidelberg, 2012.