HistoChatbot: Educating History by Generating Quizzes in Social Network Services

Yasunobu Sumikawa[1] and Adam Jatowt[2]

[1] Dept. of Computer Science, Takushoku University, Japan
[2] Dept. of Computer Science, University of Innsbruck, Austria
ysumikaw@cs.takushoku-u.ac.jp, adam.jatowt@uibk.ac.at

Abstract. Microblogging platforms can provide novel, attractive opportunities for communicating and disseminating content about important events from the past. We propose a novel framework for building interactive chatbot systems that post history-related content including automatic quizzes related to current temporal context and that take and assess user responses. The chatbot is currently available on Twitter sharing history-related quizzes in English[3]. We are the first to propose guidelines for designing history-focused chatbot systems that aim at fulfilling educational and entertaining objectives in microblogging platforms.

Keywords: social media analysis · public history · collective memory

1 Introduction

History is commonly believed to play a significant role in our society letting us understand the processes which shape the present and giving meaning and orientation with regard to the past. Education of history starts from elementary schools onwards. Many initiatives aim however at educating and sharing historical knowledge in parallel to the traditional framework of schooling in order to provide complementary information and to raise our interest in history. Public history in particular is a research area defined as communication between history and ordinary users or "non-researchers" [2] that aims to let historical knowledge reach the wide public. Social media in particular can serve as a convenient and effective venue for disseminating historical knowledge. Previous studies have already shed light on such sharing activities taking place on Twitter in the context of commemorating World War I [5], or based on the categorization of history-related tweets into 6 fundamental categories of commemoration [17].

In this paper, we propose several retrieval and recommendation models of extracting and sharing history-related content. Based on our findings we report design guidelines for developing effective chatbots to provide quizzes that attract and encourage users to learn more about history. We also showcase a working

[3] The code, quiz data and evaluation results are available at https://github.com/sumilab/programs/tree/master/histo_chatbot_quiz

prototype that we implemented based on our framework called `HistoChatbot` which periodically communicates history-related content. The purpose of the proposed chatbot system is to discover, collect and process historical content accumulated in Wikipedia and one that is circulating on Twitter, and then to communicate such content in the form of quizzes, either automatically or in response to user tweets, for actively disseminating historical knowledge. In order to realize this objective, we propose the following work modes:

- Providing quizzes on past events that occurred on the same calendar day as the interaction time
- Providing quizzes on past events related to the current news
- Providing quizzes on past events related to the current trending topics
- Providing quizzes on past events related to entities specified by users (i.e., events, people and organizations)

The proposed chatbot is an effective tool to actively share content related to history in a rather novel fashion - within online social networks and using automatically generated questions. As mentioned above, it generates and provides quizzes as well as evaluates answers as correct/incorrect based on different types of prompts which are either implicit (based on calendar, recent news, or trending topics) or explicit (based on user input). Note that while at present the chatbot is designed to work with historical event synapses contained in Wikipedia and with history-related tweets, it can be adapted for sharing short heritage content from collections maintained by memory institutions (archives, libraries, or museums). This could present an interesting opportunity for GLAM institutions (Galleries, Libraries, Archives, and Museums) to appeal to the wider public, especially to younger users.

Related work. Our proposal is to use chatbots for automatizing the effective dissemination of historical content in online social networks. Chatbots used for education started also to appear recently. The common application is the study of foreign languages [9, 11] under the assumption that learning languages need to be done in an interactive manner. For details, readers can refer to a general survey of chatbot technologies [1]. The chatbot designed in this study is oriented towards learning guidance and aims to provide education. It tweets past events collected from Wikipedia and Twitter [18], yet it can be flexibly extended to other content (even one from other domains like biology or chemistry). Question answering on historical document collections [19] or on history-related content [4, 6] as well as temporal information retrieval [12, 13] have been recently actively researched, and we believe it is good time now to employ such technologies for educating history in social networks.

2 History-related Data Collection

To populate the historical event database that underlies our chatbot system we need to find short descriptions of a large number of important historical events. As automatically extracting such event synopses from open text collections (e.g., Web or historical textbooks) is rather still error-prone, we harvested the data

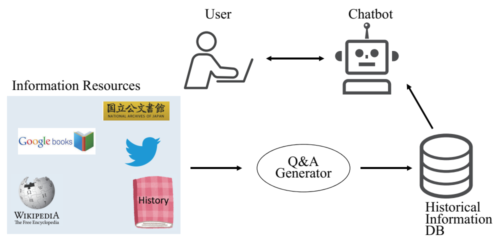

Fig. 1. High-level overview of the proposed chatbot system.

on important past events from dedicated list pages in Wikipedia. Wikipedia describes important events for each date in the dedicated year and month-day articles. We used year pages[4], as well as day pages[5] (i.e., each day from January 1 to December 31). In total, the dataset we compiled contains the descriptions of 71,374 events from all years spanning from 1AD to 2019AD.

To generate history-related quizzes, we apply a question generation method [3] based on the recurrent BERT model to the collected Wikipedia data. This method takes a pre-selected answer and the text passage containing it as input to generate a matching question for the chosen answer by considering its passage context. This is done with pre-trained transformers (seq-2-seq models) using a straight-forward end-to-end method.

Before generating questions, we first extracted all named entities to be used as answers. For simplicity, we considered as the named entities any linked texts pointing to other Wikipedia articles. Finally, we input the prepared answers and their passages into the question generation method. In total, we have generated in this way 22,883 history-related questions from events that were previously collected from Wikipedia.

3 History-focused Quiz Posting

Fig. 1 shows the conceptual diagram of `HistoChatbot`. The current chatbot posts quizzes on Twitter from our Historical Information DB according to server-side settings that run the program; in the current implementation the quizzes are posted periodically and users can provide answers whenever they notice the questions. The chatbot evaluates tweets as correct or wrong according to the lexical match between the gold answer and the user tweet that provides an answer. The user then receives a feedback tweet informing whether the answer was correct or not. In this section, we describe the key work modes of the proposed chatbot and the way in which it selects and recommends history-related content.

In all the modes, we need to apply ranking since several event (question) candidates could be used by the chatbot. We rank candidates according to their importance based on the assumption that Wikipedia articles which are often

[4] E.g.,https://en.wikipedia.org/wiki/1998
[5] E.g.,https://en.wikipedia.org/wiki/February_25

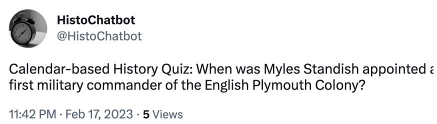

Fig. 2. Calendar-based quiz mode's example.

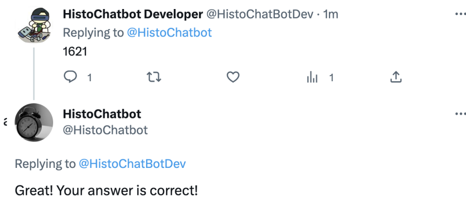

Fig. 3. Example of user–chatbot interaction being a response to the question shown in Fig. 2.

linked by many other articles are important. We use the following approach to rank events, and indirectly, the questions that were generated from them.

$$Importance(evt) = \sum_{ett \in Entity(evt)} Link(ett) \tag{1}$$

$$result(E) = \arg\max_{evt \in E} Importance(evt) \tag{2}$$

where E is a set of events. The above equations count the numbers of Wikipedia articles linking to the entities mentioned in the description of a given event as a way to approximate the event's importance. The function $Link(ett)$ counts the number of Wikipedia articles linking to a given entity, while evt and ett denote events and entities, respectively. $Entity(evt)$ denotes the set of entities extracted from the content of event description evt.

Calendar-based Quiz Mode. In the most basic mode, calendar-based mode, the chatbot posts quizzes about past events that occurred on the same calendar day in the past. This is in the same spirit as some newspapers which report (often at the end of the printed volume) important events that happened in the past on the same day, or ones that were reported by the same newspaper on that particular day. Fig. 2 shows an example of results generated through the application of this mode. According to Wikipedia, Myles Standish was elected as the first commander of the Plymouth Colony militia in 1621. Our chatbot posted this quiz on Feb 17, 2023, which is exactly the same day as the date of that event. Fig. 3 shows how the chatbot evaluates the user's answer[6].

With the calendar-based mode, the chatbot can actually provide several updates to its followers at regular times of the day, since there were usually multiple events that occurred on a particular calendar day in the past, throughout human history. The selection of an event to be tweeted is done as follows:

$$result(d, E) = \arg\max_{evt \in E} \{Importance(evt) \mid date(evt) == d\} \tag{3}$$

where d is the current date (represented by day and month information) and $date(evt)$ is the date when the event evt, which was not yet shared, occurred.

[6] We assume here that the HistoChatbot Developer denotes a real user.

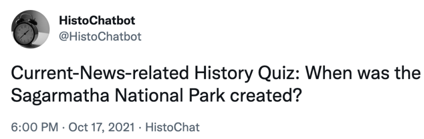

Fig. 4. Current-news-based quiz mode example.

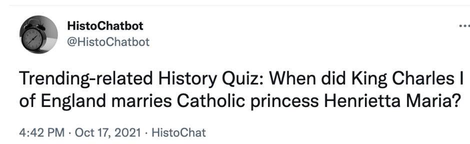

Fig. 5. Trending-words-based quiz mode example.

Current-news-based Quiz Mode. This mode outputs content about past events which are similar to the current events. To collect data on the current events, we created a list that includes Twitter official accounts of multiple news companies including CNN, BBC, NYT, and others[7]. Fig. 4 shows an example of the chatbot's output in this work mode. This quiz asks about the year when a particular national park was created because its trigger was the recent CBC British Columbia's tweet [8] about the latest event that happened in a park.

Generating quizzes in this mode is executed each time there is news shared from one of the above-mentioned Twitter accounts of news companies. The chatbot first collects a newly issued tweet from the media companies. It then extracts nouns from the tweet's text to be used for extracting relevant event descriptions from our database. We use logical OR search based on the extracted nouns to find events most similar to the target utilized tweet. Next, the chatbot creates bag-of-words feature vectors for representing the selected event descriptions and the news company's tweet text. It then applies cosine similarity to these feature vectors and calculates the events' ranking scores considering both the cosine similarity and event importance scores. The calculation is as follows:

$$Impr(evt, E) = \frac{\sum_{ett \in Entity(evt)} Link(ett)}{Max(\{\sum_{ett \in Entity(evt')} Link(ett)) \mid evt' \in E\})}$$

$$score(txt, evt, E) = \alpha CosSim(txt, evt) + (1-\alpha) Impr(evt, E) \tag{4}$$

$$result(txt, E) = \underset{evt \in E}{\arg\max}\; score(txt, evt, E) \tag{5}$$

where txt is the text of a news and α is a hyper-parameter to adjust the scores of $CosSim$ and $Impr$. In this mode, we use a news event description as a text. We set 0.5 as a default score to α. As the range of values of the cosine similarity in our case is [0, 1], the function $Impr$ used in this equation returns the normalized value of importance, so that the importance of the entities has also the same range of values. The normalization is done by dividing the importance score of each event by the maximum value of the importance scores of all the events in our database.

Trending-words-based Quiz Mode. Twitter lists words that many users were recently interested in by indicating them as trending. We use the trend-

[7] `https://mobile.twitter.com/i/lists/1256794745512185857`
[8] `https://mobile.twitter.com/cbcnewsbc/status/1449609372296953859`

ing contents to post quizzes that users may be potentially interested in. Fig. 5 shows an example of the output in this work mode. Since the word "maria" was included in the New Zealand's[9] trending word list on 17 October 2021, the chatbot generated the quiz that includes this word.

The procedure for generating a quiz in this mode is as follows. The chatbot first collects the trending words through the Twitter trends/place API. After removing # to extract text from hashtags, the chatbot loads event descriptions from our database. If there is at least one event description that includes this trending text, it is used for quiz selection. The chatbot then creates feature vectors for both the found event descriptions and for the collected trending words. Next, it applies cosine similarity using the feature vectors, and calculates the ranking score by combining the cosine similarity and importance scores. This algorithm uses the same equations as the current-news-based work mode.

Entity-based Quiz Mode. We describe now our last work mode. Its idea is to post quizzes corresponding to users' requests. If a user wants to receive questions about particular entity, she can request a related quiz about that entity. For example, if a user requested a quiz about Japan by sending the tweet "Give me a quiz about japan" to the chatbot, the chatbot extracts the word "japan" from the text, loads events related to Japan, and outputs a top-scored quiz based on the result of applying Eq. 5.

For simplicity, we currently apply the template-based entity detection in this mode[10]. In particular, we utilize a template "Give me a quiz about". The chatbot then extracts all the words following this template. Next, the historical information database is searched for events that involve all the detected entities to be used as candidate events for constructing chatbot's reply. If none such event exists, the events that include at least one entity are considered as candidates.

4 Conclusions & Future Work

In this paper we have described a framework for designing responsive chatbot systems that post history-related quizzes in SNSs. We proposed several work modes for recommending the events relevant to users or to the current context (same day, popular news, trending words). The proposed framework can serve as initial step for designing similar chatbot systems to actively disseminate knowledge from other domains besides history for educational purposes.

In future, we will measure event popularity (e.g., with statistical approaches similar to [21]) to better select content for dissemination. Estimating educational value (e.g., [8]), interestingness degree [10], bias degree [15] or contemporary relevance [14, 16] of historical events will be further investigated for increasing the attractiveness and credibility of shared content. Finally, we will experiment with incorporating large language models [20, 22] (e.g., ChatGPT) to provide more human-like conversational capabilities and commonsense reasoning [7] and by this to increase responsiveness and believably of the chatbot.

[9] Assuming for this example that the user lives in New Zealand.

[10] In the future, we plan to apply named entity recognition tools to allow users to write arbitrary texts.

References

1. Adamopoulou, E., Moussiades, L.: An overview of chatbot technology. In: Maglogiannis, I., Iliadis, L., Pimenidis, E. (eds.) Artificial Intelligence Applications and Innovations. pp. 373–383. Springer International Publishing, Cham (2020)
2. Cauvin, T.: The rise of public history: An international perspective. Historia Crítica No.40 **68**, 3–26 (04 2018)
3. Chan, Y.H., Fan, Y.C.: A recurrent BERT-based model for question generation. In: Proceedings of the 2nd Workshop on Machine Reading for Question Answering. pp. 154–162. ACL, Hong Kong, China (Nov 2019)
4. Chen, W., Wang, X., Wang, W.Y.: A dataset for answering time-sensitive questions (2021)
5. Clavert, F., Majerus, B., Beaupré, N.: #ww1. twitter, the centenary of the first world war and the historian. Twitter for Research 2015
6. Cole, J.R., Chaudhary, A., Dhingra, B., Talukdar, P.: Salient span masking for temporal understanding. arXiv preprint arXiv:2303.12860 (2023)
7. Davis, E.: Benchmarks for automated commonsense reasoning: A survey. arXiv preprint arXiv:2302.04752 (2023)
8. Hong, J.C., Cheng, C.L., Hwang, M.Y., Lee, C.K., Chang, H.Y.: Assessing the educational values of digital games. Journal of Computer Assisted Learning **25**(5), 423–437 (2009)
9. Huang, W., Hew, K.F., Fryer, L.K.: Chatbots for language learning—are they really useful? a systematic review of chatbot-supported language learning. Journal of Computer Assisted Learning **38**(1), 237–257 (2022)
10. Jatowt, A., Hung, I., Färber, M., Campos, R., Yoshikawa, M., et al.: Exploding tv sets and disappointing laptops: Suggesting interesting content in news archives based on surprise estimation. In: European Conference on Information Retrieval. pp. 254–269. Springer (2021)
11. Jia, J.: Csiec: A computer assisted english learning chatbot based on textual knowledge and reasoning. Knowledge-Based Systems **22**(4), 249–255 (2009)
12. Joho, H., Jatowt, A., Roi, B.: A survey of temporal web search experience. In: Proceedings of the 22nd International Conference on World Wide Web. pp. 1101–1108 (2013)
13. Kanhabua, N., Nguyen, T.N., Niederée, C.: What triggers human remembering of events?: A large-scale analysis of catalysts for collective memory in wikipedia. pp. 341–350. JCDL '14, London, United Kingdom (2014)
14. Kanhabua, N., Niederée, C., Siberski, W.: Towards concise preservation by managed forgetting: Research issues and case study. iPres'13, Lisbon, Portugal (2013)
15. Lim, S., Jatowt, A., Färber, M., Yoshikawa, M.: Annotating and analyzing biased sentences in news articles using crowdsourcing. In: Proceedings of the Twelfth Language Resources and Evaluation Conference. pp. 1478–1484 (2020)
16. Sato, M., Jatowt, A., Duan, Y., Campos, R., Yoshikawa, M.: Estimating contemporary relevance of past news. In: ACM/IEEE, JCDL 2021. pp. 70–79. IEEE (2021)
17. Sumikawa, Y., Jatowt, A.: Analyzing history-related posts in twitter. Int. J. Digit. Libr. **22**(1), 105–134 (2021)
18. Sumikawa, Y., Jatowt, A.: Designing chatbot systems for disseminating history-focused content in online social networks. D - IEICE TRANSACTIONS on Information and Systems (Japanese Edition) **J104-D**(5), 486–497 (2021), (in Japanese)

19. Wang, J., Jatowt, A., Yoshikawa, M.: Archivalqa: A large-scale benchmark dataset for open-domain question answering over historical news collections. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 3025–3035 (2022)
20. Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., et al.: A survey on large language model based autonomous agents. arXiv preprint arXiv:2308.11432 (2023)
21. Yamamoto, Y., Tezuka, T., Jatowt, A., Tanaka, K.: Supporting judgment of fact trustworthiness considering temporal and sentimental aspects. In: Web Information Systems Engineering-WISE 2008: 9th International Conference, Auckland, New Zealand, September 1-3, 2008. Proceedings 9. pp. 206–220. Springer (2008)
22. Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.Y., Wen, J.R.: A survey of large language models (2023)