Twittener: An Aggregated News Platform

Chan Wei Chang
Department of Computer Science
Nanyang Technological University
Singapore
wchan027@e.ntu.edu.sg

Owen Noel Newton Fernando
Department of Computer Science
Nanyang Technological University
Singapore
ofernando@ntu.edu.sg

Abstract—**The Internet offers an abundance of online sources for trending topics and news. However, this gives rise to the issue of content overload, where users must filter through large amount of content to find those that are of relevance or interest to them. This project aims to solve this issue by creating a web application called Twittener. Twittener aims to improve users' experience and time-efficiency when reading news online. Methods used include text-to-speech technology, sentiment analysis and recommender system. Text-to-speech technology enables users to listen to tweets and news without paying attention to their screens. This could also be useful for populations with visual impairments. Sentiment analysis on Twitter trends provides useful information on general sentiment towards each trend and a hybrid recommender system is deployed to recommend news that would likely be of interest to users. This paper seeks to document the development, implementation, design and implications of Twittener.**

Keywords- twitter; news aggregation; news feeds; sentiment analysis; personalised content; text-to-speech

I. INTRODUCTION

The rise of the Internet has significantly changed the landscape of the media industry as online news sources become easily available. With news constantly being updated online, readers can have prompt access to global news, which often comes at little or no cost. This increased accessibility to news online has enticed readers to switch from traditional newspapers to digital media. As a result, there has been a decline in the print media industry over time. According to a research, U.S. Weekday and Sunday circulation decreased by 11% and 10% respectively in 2017 [3]. With an abundance of news sources, readers may feel overwhelmed when finding news articles which spark their interest. Furthermore, news sites generally do not offer personalised content. Readers are required to search for relevant topics, which can be time-consuming. To facilitate their search, many news aggregators have emerged over the past decade. News aggregators are web applications which aggregate news articles from various sources and are subsequently presented on one site [10].

Popular news aggregators such as Google News and Yahoo! News have come into existence since the late 1990s. Some news aggregators such as Google News offers personalised content, mainly in two areas: topics related to a reader's genuine interests and interesting topics outside of his usual interests [4]. However, current news aggregators only source from official news sites and exclude posts on social media platforms, despite the possibility that these posts may be accurate news. Two prominent examples are the Mumbai terrorist attacks in 2008 and the U.S. plane crash in the Hudsons River in 2009, which were initially distributed on Twitter [6]. Therefore, one enhancement that could be made to news aggregators is to enable it to source for information from both official news sites and social media platforms.

The aim of this project was to develop a web application, Twittener, that combines information from existing news aggregators (Google News and Yahoo! News) with social media platforms (Twitter and Reddit). Google News and Yahoo! News were chosen as both are popular news aggregators which tend to be apolitical [11]. Twitter was chosen as it was identified as the social media platform where breaking news often first appeared on [13], while Reddit was chosen as it provides a holistic overview of the latest news. In addition to news personalization, text-to-speech and sentiment analysis technologies were employed to enhance the user's experience. Text-to-speech technology was used to convert news summaries and tweets into audio. In this way, readers can choose between reading and listening to tweets or news abstract. Sentiment analysis technology was used to determine the general sentiment towards a trend. All in all, the improved news aggregator Twittener improves user experience by providing a comprehensive one-stop site for highly relevant and interesting news from various platforms.

II. RELATED WORK

Related work falls into three main areas: (1) News Aggregators; (2) Voice synthesis on Twitter and (3) Recommender System.

A. News Aggregators

News aggregators aim to help readers find new and interesting news articles. Some news aggregators provide additional features beyond news aggregation to increase the value of its service. One such feature is personalizing the content shown to each user. However, existing news aggregators do not take into consideration information on social media platforms, where citizen journalism can occur. Considering that online media is gaining popularity and social media platform users can freely share information, it is

no surprise that citizen journalism can deliver breaking news much faster than traditional news sites.

Twittener aims to leverage on the advantages of current news aggregators and further enhance it by incorporating Twitter and text-to-speech technology. A study comparing three social media platforms Facebook, Google Plus, and Twitter has shown that Twitter dominates the other two in providing timely news [8]. Therefore, Twitter was chosen as the social media platform to be integrated into Twittener.

B. Voice synthesis on Twitter

Multiple applications have been developed to allow people to listen to tweets instead of reading them. However, an application that utilizes analytic to help users identify high-valued content has yet to be introduced. Twisten [2] is a third-party Twitter client on Windows that users can download to enable them to listen to their Twitter feed. It uses Window's built-in text-to-speech to convert tweets to its audio format. Given that the role of Twitter in Twittener is to provide users with the latest breaking news, a simple audio feature would not be enough. Analytics need to be applied so that it can help users identify high-valued content. The first and most important step is to understand the sentiment of each tweet, which is a complicated task, since tweets are multi-lingual. Studies found that text preprocessing is essential when working with Twitter's corpus [8]. As such, Twittener will preprocess text from Twitter before performing sentiment analysis.

C. Equations

Recommender systems recommend contents or products that might be of interest to a user, reducing effort needed to browse through massive amounts of content or products on online services [14].

There are 3 main filtering methods: Content-based, Collaborative, and Hybrid (Content-based with Collaborative). Content-based filtering analyses past contents read by the user and make a recommendation based on them [13]. It can manage situations where different users have different interests and can account for a user's change in interest in a short span of time [9]. However, it faces the issues of limited content analysis, overspecialization and sparsity of data [1]. Collaborative filtering makes recommendations based on content read by other users with similar interests [12]. It recommends content which are outside the usual interest of a group of users with a certain set of interest, but which appears popular among them, to other users with the similar set of interest. However, this has two main constraints. Firstly, new content that has not been read by other users could not be recommended. This is often referred to as the first-rater problem [7] and impedes the sites' ability to deliver the latest news in the fastest possible time. Secondly, it does not take individual preferences into account [5]. Hybrid recommender system makes use of a combination of recommendation techniques. This is more useful as one technique is able to overcome the issues faced by another technique [8]. In a study performed with very similar context, the hybrid recommendation method improved the quality of news recommendations and attracted

more frequent visits to the site when compared to collaborative filtering [13].

III. TWITTENER: DESIGN AND IMPLEMENTATION

A. System Architectures

The backend server is implemented using JavaScript based frameworks (Node.js and Express.js). A nonrelational database (MongoDB) is used to enable robust deployment and scaling. A cache (Redis) is also used to stored fetched content for certain period so that content can be served to users as fast as possible.

The server can be accessed using a RESTful API. Figure 1 depicts the backend architecture of Twittener.
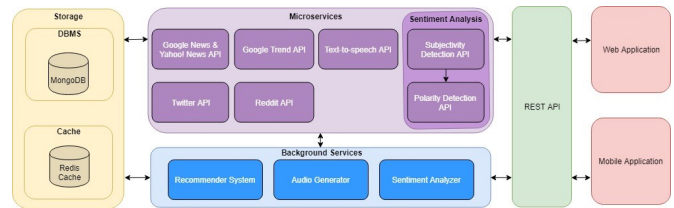

Figure 1: Twittener architecture.

B. Sentiment Anaysis on Twitter Trends

For each trending hashtag or keyword (e.g. #MCU), sentiment analysis will be performed on one thousand tweets that contains the hashtag or keyword to determine the general sentiment of a trend. Sentiment analysis is performed using APIs provided by SenticNet, which has three sentiment classification: positive, negative, and neutral. Multiple calls are made to the SenticNet API concurrently and Twittener server will compute the quantity of each sentiment based on the result of each API call. This process is depicted in Figure 2.
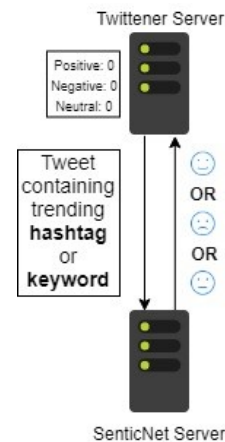

Figure 2: Sentiment analysis logic.

C. Text-to-speech

The text-to-speech conversion for tweets and news abstracts are performed using IBM Watson's text-to-speech API. The content of the text were preprocessed to remove all

characters that could not be translated to audio (e.g. emojis) and checked that it is English before being sent to IBM Watson' text-to-speech API, which will generate a MP3 audio file.

D. Hybrid Recommender

The hybrid recommender system was developed using LightFM, which implements the Weighted Approximate-Rank Pairwise (WARP) loss for implicit feedback learning-to-rank. WARP generally performs better than the more popular Bayesian Personalised Ranking (BPR) loss by a large margin [10]. Each time a user clicks on an article from News page, the topic of the article will be stored along with user's unique Twittener ID. This data is then used to build a model that recommends 3 topics to the user when given the unique Twittener ID

IV. TWITTENER: FEATURES

Twittener website has 4 pages: Home page, Trends page, Topic page, and News page. Each page contains a specific type of content

A. Home

The Home page delivers updated news on Twitter and provides basic Twitter functionalities. Content from Twitter are retrieved through the official Twitter APIs. Tweet refers to a post on Twitter, which can be text alone or together with multimedia such as an image or video. The Home page is split into 3 tabs: Twitter Timeline, Twitter Trends, and Twitter Trends Timeline. The Home page is shown in Figure 3.
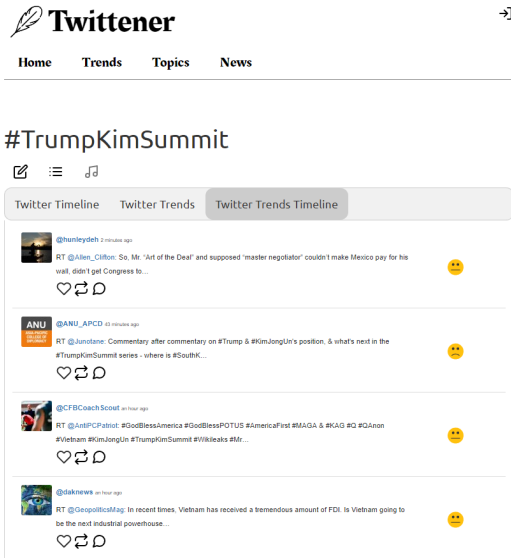

Figure 3: Home page of Twittener. Shown here is the Twitter Trends Timeline

Twitter Timeline is where tweets from the user's Twitter Timeline appear. It is incorporated into Twittener as it can be a great place to discover breaking news, which tends to be shared by multiple people. On top of the basic Twitter features, users have the option to listen to tweets instead of reading them. Users can choose to either listen to individual tweets or their entire timeline.

Twitter Trends displays trending hashtags or keywords that keep appearing in tweets in a particular country, enabling users to learn about popular news. The general sentiment of each trend is determined using sentiment analysis and shown beside it to help users easily identify interesting trends. For example, when the general sentiment of a trend does not match the user's expectation, it would pipe his interest. Twitter Trends Timeline displays the tweets that contain the trending hashtags or keywords from Twitter Trends tab. The sentiment of each tweet is also displayed.

B. Trends

The Trends page aims to provide users with an overview of events that are trending globally. To provide a variety of insights and content, data are sourced from 3 different platforms: Google News, Yahoo! News and Reddit. Each of these has its own tab. The trends page is shown in Figure 4.
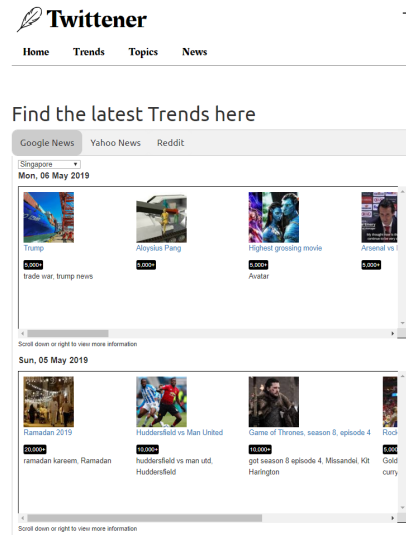

Figure 1. : Trends page of Twittener. Shown here is the Google News.

Google Trends tab shows popular search terms on Google along with the number of searches performed and other related terms. Clicking on the search term will redirect users to a news article related to it. Yahoo! News tab shows the daily trending articles from Yahoo! News. Articles are also categorized by trending dates. Three days' worth of trends and articles will be available on these tabs. Reddit tab shows the top five daily post from the different subreddits. Subreddits are like a sub-forum. Users can personalize their own set of subreddits. Twittener will save the subreddit setting of each user in the database. Users can be identified by their Twitter ID, since they login to Twittener using their Twitter account.

C. Topics

The Topic page allows users to customize the content they want to see on Twitter based on topics. This differs from Twitter Timeline which contains a mix of tweets from family, friends, and people of interest. This makes it easier

for users to find specific content as they filter through less non-related content. The Topics page is shown in Figure 5.
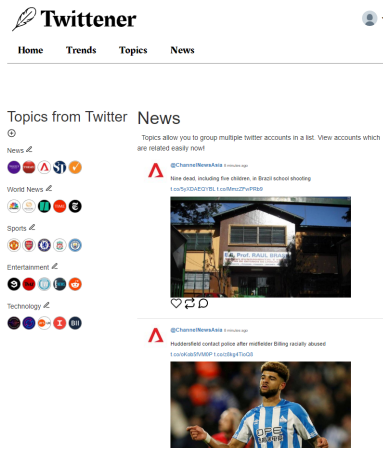

Figure 2.   Topic page of the Twittener.

This is done using a Twitter provided feature, "List". Users can group Twitter accounts with similar content together and view a separate timeline containing tweets only from these accounts. A topic on Twittener is essentially a list from Twitter. Users can create, delete or modify their own topic. All functionalities are provided through Twitter's API.
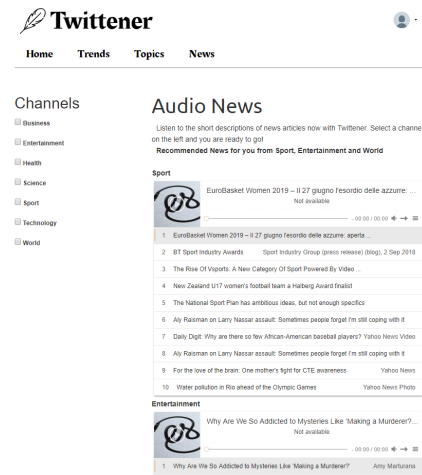
D. News

Figure 3.   News page of the Twittener.

The News page provides users with content from traditional news sites but uses a different filtering method. Users can choose a channel and listen to summaries of all the articles in it. Should any summary interest them, they can read the full article by clicking on the article title on the audio player. A hybrid recommender system recommends 3 topics to the user. News will be refreshed hourly to ensure that they are kept up-to-date. The News page is shown in Figure 6.

V.   CONCLUSION AND FUTURE WORK

Twittener is a news aggregator that delivers up-to-date news from both official news sites and social media. Technologies for text-to-speech conversion, sentiment analysis and news recommendation improve user time-efficiency in consuming content or filtering news.

In the future, we are planning to conduct an in-depth user study using the Technology Acceptance Model (TAM) model. The purpose of this study is to elicit user's feedback and evaluate the current system. Factors such as the perceived ease of use and perceived usefulness of the system will be evaluated.

VI.   ACKNOWLEDGEMENT

REFERENCES

[1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender system. A survey of the state-of-the-art and possible extensions, " IEEE Trans Knowl Data Eng;17(6):734–49.

[2] A. Allen-Wagner, "Twisten - Listen to Twitter – the alternative," Retrieved from https://nathan.alner.net/twisten/.

[3] M Barthel, "*Newspapers Fact Sheet*. Pew Research Center," Retrieved from http://www.journalism.org/fact-sheet/newspapers/

[4] J. Beel, B Gipp, S. Langer and C. Breitinger, "Research-paper recommender systems: a literature survey," *International Journal on Digital Libraries*, 17(4): 305-338.

[5] D. Billsus and M. J. Pazzani, "User Modeling for Adaptive News Access," *User Modeling and User-Adapted Interaction*, 10(2-3): 147-180.

[6] M. Broersma and T. Graham, "SOCIAL MEDIA AS BEAT," *Journalism Practice*, 6(3): 403-419.

[7] A. S. Das, M. Datar, A. Garg and S. Rajaram, "Google news personalization: Scalable Online Collaborative Filtering," In *Proceedings of the 16th international conference on World Wide Web - WWW 07*, 285-295.

[8] D. Inouye and J. K Kalita, "Comparing Twitter Summarization Algorithms for Multiple Post Summaries," *IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, Boston, MA. 298-306.

[9] F. Isinkaye, Y. Folajimi and B. Ojokoh, "Recommendation systems: Principles, methods and evaluation," *Egyptian Informatics Journal*, 16(3) 261-273. Available: 10.1016/j.eij.2015.06.005.

[10] Maciej Kula, "Learning-to-rank using the WARP loss," Retrieved from https://lyst.github.io/lightfm/docs/examples/warp_loss.html

[11] A. Lee and H. Chyi, "The Rise of Online News Aggregators: Consumption and Competition," *International Journal on Media Management* 17(1): 3-24.

[12] J. Liu, P. Dolan and E. R. Pedersen, "Personalized news recommendation based on click behavior," In *Proceedings of the 15th international conference on Intelligent user interfaces - IUI 10*.

[13] M. Osborne and M Dredze, "Facebook, Twitter and Google Plus for breaking news: Is there a winner?," In *International Conference on Weblogs and Social Media (ICWSM)*, 2014.

[14] W. Zhang, H. Sun, X. Liu and X Guo, "Temporal QoS-aware web service recommendation via non-negative tensor factorization," In *Proceedings of the 23rd international conference on World wide web - WWW 2014*.