SOFTWARE

Twitter integration of chemistry software tools

Naruki Yoshikawa[1†], Ryuichi Kubo[3†] and Kazuki Z Yamamoto[2,3,4*]

*Correspondence:
kazuki@ric.u-tokyo.ac.jp
[2]Isotope Science Center, The
University of Tokyo, Tokyo, Japan
Full list of author information is
available at the end of the article
†Equal contributor

Abstract

Social media activity on a research article is considered to be an altmetric, a new measure to estimate research impact. Demonstrating software on Twitter is a powerful way to attract attention from a larger audience. SNS integration of software can also lower the barriers to trying the tools and make it easier to save and share the output. We present three case studies of Twitter bots for cheminformatics: retrosynthetic analysis, 3D molecule viewer, and 2D chemical structure editor. These bots make software research more accessible to a broader range of people and facilitate the sharing of chemical knowledge, concepts, and ideas.

Keywords: Twitter; Social media; Retrosynthesis; Visualization; Molecule editor; Chemical space; Knowledge sharing; Open science

ORCIDs
- Naruki Yoshikawa: 0000-0003-1546-8709
- Ryuichi Kubo: 0000-0001-8733-9135
- Kazuki Z Yamamoto: 0000-0001-6231-0475

Introduction

Social media such as Twitter has become an important communication tool for scientists. *Twitter* is a social networking service on which users post short messages called *tweet*, interact with each other by sending a *reply* and share other users' tweets by *retweet*. Scientists are using *hashtags* such as `#chemtwitter` and `#ScienceTwitter` to create communities on Twitter for networking and information exchange [1]. Twitter activities about a research paper are considered to be one of the altmetrics [2], which is an alternative to citation counting for measuring scientific impact. Some studies reported that the number of tweets on a research article is positively correlated to the article citations [3, 4, 5]. Attracting attention on Twitter is getting more important to promote scientific works.

A Twitter bot is an automated program on Twitter controlled through Twitter API. Bots are common on Twitter, and they usually deliver tweets like news and blog updates [6]. Twitter bots are also used for scientific purposes, such as posting updates of journal publications or preprints. While most bots automatically post tweets, they can also respond to replies from other users. We can construct an interactive application using Twitter API. In this paper, we report case studies of cheminformatics tools demonstrated on Twitter.

Case Study 1: Retrosynthesis Bot

In this section, we introduce a Twitter bot `@retrosynthchan` [7], which conducts retrosynthetic analysis by AiZynthFinder [8]. The bot responds to a request given as

a SMILES [9] string and returns the results of retrosynthetic analysis for the given molecule. Users are expected to send a reply containing only a SMILES string right after the bot's screen name (e.g., `@retrosynthchan c1ccccc1Br`).

The simplest version of a cheminformatics tool bot has been operating on `@souyakuchan` account [10] since 2017 in the form of a bot that draws a 2D structural formula image for replies containing a SMILES or an IUPAC name [11]. The retrosynthesis bot was implemented in December 2020 for the blog event called Drug Discovery (dry) Advent Calendar 2020 [12] hosted by `@souyakuchan`.

Implementation

The workflow of the bot program is as follows: (1) retrieve a reply to the bot via Twitter API, (2) check whether the given string is valid as SMILES using RDKit [13] and send a request confirmation, (3) run retrosynthesis analysis by AiZynthFinder [8], (4) send the retrosynthesis result images via Twitter API. The details of each step will be discussed.

Retrieve reply   A reply in Twitter is a message toward a specific account, and it usually starts with the screen name of the receiver (e.g., `@jack How are you today?`). At the time of this writing, the Twitter API provides a method to retrieve tweets that contain a specific keyword in real-time (`POST statuses/filter` [14]). The bot receives a reply to the bot by retrieving tweets that contain the bot's screen name with this API. A Python library Tweepy [15] is used to handle Twitter API.
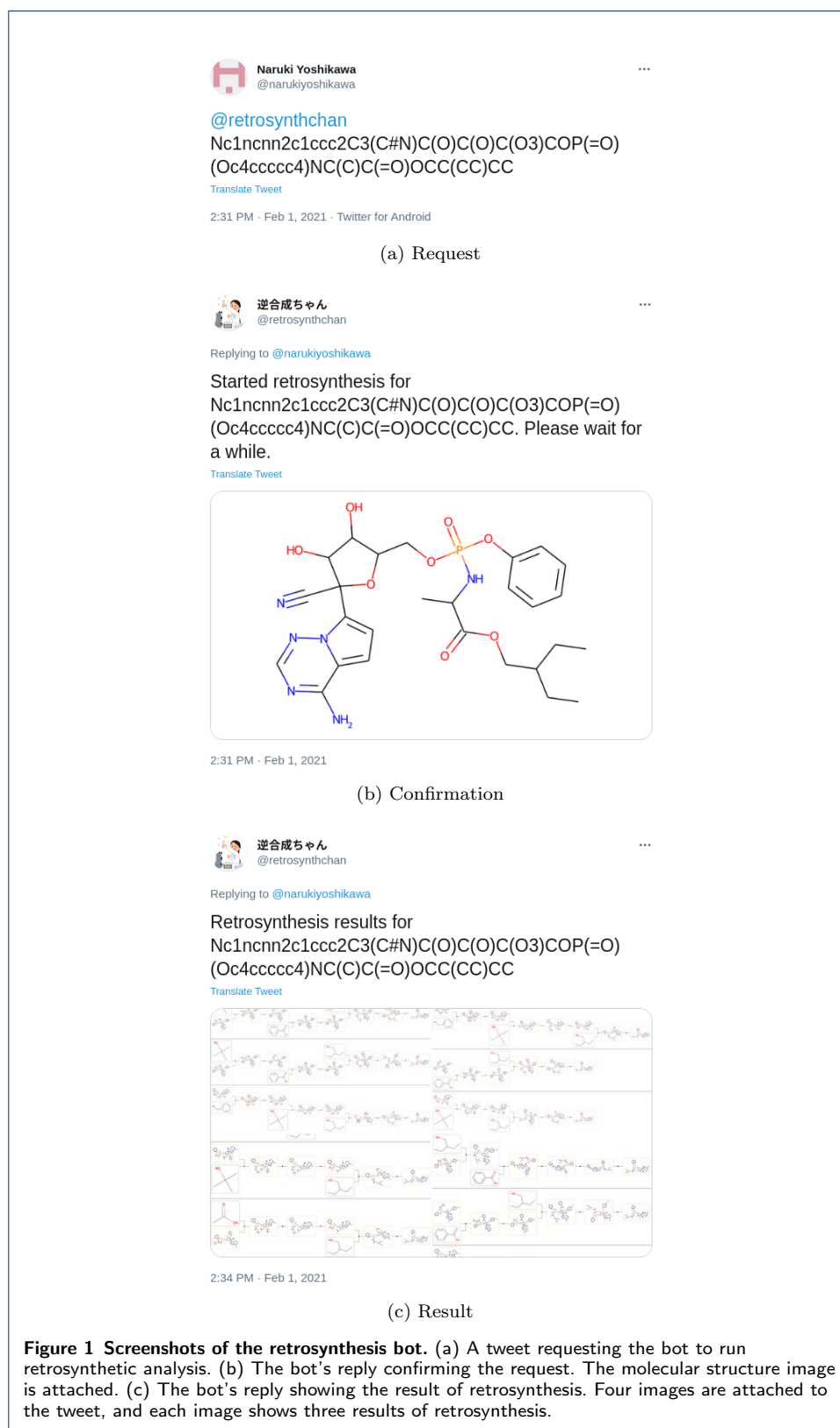
Confirm request   After receiving a reply, the string is separated into parts by a space. The second part is assumed to be a SMILES string since the first part is the bot's screen name. The SMILES string is processed by RDKit [13] to check the validity. The bot returns the input molecule's chemical structure if the input is valid, otherwise, send back an error message to the sender and finish. The confirmation is sent as a reply to the users by a Twitter API `POST statuses/update` [16] using Tweepy.

Run retrosynthetic analysis   Retrosynthetic analysis of the molecule represented by the given SMILES is conducted by AiZynthFinder [8], open-source software for retrosynthetic planning. The input molecule is broken down into purchasable precursors with the algorithm based on a Monte Carlo tree search guided by a neural network. The synthesis route images are generated for individual routes.

Send synthetic pathway images   The generated synthetic route images are sent to the user as a reply with images. Twitter allows to attach only up to four images in one post, so three images are connected into one image using Pillow [17] to send a maximum of twelve synthesis routes. The generated images are sent back to the users with the Twitter API using Tweepy.

Results

A screenshot of the bot in action is shown in Figure 1. A request confirmation with a chemical structure image was sent right after the SMILES was sent to the bot.

(a) Request

(b) Confirmation

(c) Result

Figure 1 Screenshots of the retrosynthesis bot. (a) A tweet requesting the bot to run retrosynthetic analysis. (b) The bot's reply confirming the request. The molecular structure image is attached. (c) The bot's reply showing the result of retrosynthesis. Four images are attached to the tweet, and each image shows three results of retrosynthesis.

The result of the retrosynthetic analysis is sent several minutes later since it takes a while to run AiZynthFinder.

The bot attracted attention from Twitter users interested in chemistry, and it gained more than one hundred followers in the first two weeks. The followers include chemists in academia, researchers in the drug industry, undergraduate or graduate chemistry students, etc. The Twitter bot helped the retrosynthesis software outreach to a broader range of people.

Case Study 2: 3D molecule viewer embedded in Twitter timeline

Next, we introduce a 3D molecule viewer which can be played on a Twitter timeline inlinely via our web server `chemical.space` [18]. This feature is based on the Twitter Player Card system [19]. The Player Card system is used as a video/audio player in the Twitter timeline, wherein the 3D molecule viewer works as interactive video content. This feature is currently available on web browsers, while the Twitter mobile application can not load Player Card directly in the timeline due to the limitation of Twitter's specifications. Users can view any Protein Data Bank (PDB) structures interactively without leaving their timeline on the browser. To embed the viewer in a tweet, users are expected to post a tweet containing a PDB ID string right after the bot's screen name with a `#pdb` hashtag (e.g., `@souyakuchan 1ema #pdb`).

Implementation

As a 3D molecule viewer, we adopted LiteMol [20], which is a JavaScript application written in TypeScript. The way of receiving requests is the same as in Case 1. After receiving the PDB ID via users' tweets, the bot makes a reply in which the URL of our hosting server of LiteMol is indicated. Also, to provide Twitter Player Card, our viewer web page serves meta information in HTML header named `twitter:card`.

Results

A screenshot of the viewer played on a Twitter timeline is shown in Figure 2. After receiving a PDB code by a *mention* tweet, the bot immediately replies with an URL in which the 3D viewer Player Card can be expanded. Users can view the 3D structure interactively as it is in the Twitter timeline.

Case Study 3: Shareable 2D chemical structural formula editor

The 2D chemical structural formula editor integrated with tweet posting function (available at `chemical.space/editor`) is one of the easiest ways to share one's own idea of chemical structures on Twitter. Users can share their manually edited structural formula as a SMILES or Chemical Markup Language (CML) [21] format. By using CML, users can include the information of coordinate and rotation of each object. This is a better way to share complex structures, whereas SMILES can only represent the molecular graph information, which has the advantage of the smallest data size. This editor has one more useful feature, in which users can send their manually drawn chemical structure to `@retrosynthchan` bot (shown in Case Study 1) by just clicking the `RETROSYNTHESIZE` button.

(a) Request and reply
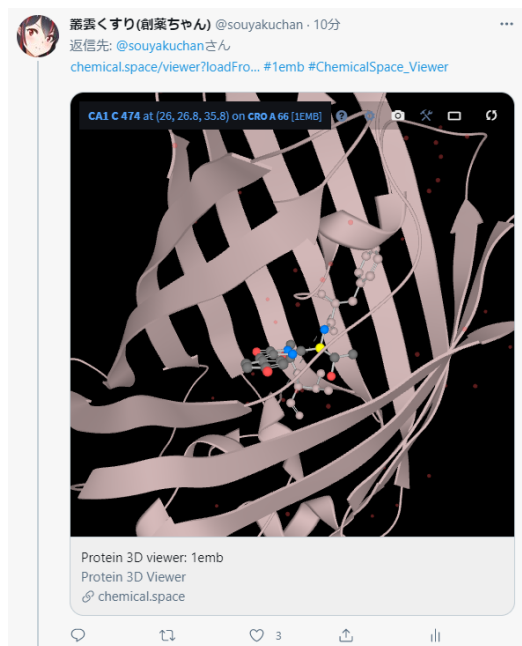
(b) View

Figure 2 Screenshots of the Twitter integrated 3D molecule viewer. (a) A tweet requesting the bot to return an URL of a 3D molecular viewer and the subsequent reply with the viewer embedded in the Twitter timeline. Users can expand the viewer by clicking the play button. (b) The viewer expanded in the Twitter timeline. Users can interactively manipulate the view.

Implementation

Among the existing well-designed traditional browser-based 2D molecule editors such as JSME [22] and MarvinSketch [23], we chose Kekule.js [24] for its ease of integration as a web application and its sufficient functionality. Its JavaScript-based editor can input and output SMILES strings and CML strings of manually drawn structural formulas. While SMILES only stores simple characters to express atoms, bonds, rings, and their orders, CML has coordinate and rotation information of each atom and bond. Furthermore, CML contains many XML markup tags in addition to that information. Meanwhile, CML strings become too long to tweet because one tweet accepts only 280 characters. However, URL strings are not included in this limitation. To embed CML strings into an URL, `chemical.space/editor` compresses CML strings by gzip using pako [25] library, then encodes it into Base64url using brianloveswords/base64url [26].

Results

The 2D molecule editor hosted on `chemical.space/editor` is shown in Figure 3. Users can utilize it to draw structural formulas from scratch in their browser and post them to Twitter in SMILES or CML format using the button at the bottom of the editor. We have also prepared a button that allows users to submit their structural formula to the retrosynthetic analysis bot introduced in Case Study 1. Structural formulas posted in CML format can be re-edited by opening the URL in the tweet, which is useful for open discussion on compounds on the SNS.

Discussion

As presented in these case studies, it is technically possible to make any cheminformatics tool accessible and noticeable from social networking services in the form of interfacing with Twitter bots. However, to make effective impressions on the scientific community and stimulate discussion, it is necessary to use appropriate hashtags that are widely prevailing and be tweeted by accounts with a large number of followers and strong influence. It would be helpful for scientists to follow accounts of mutual interest regularly and grow their follow/follower count appropriately. Tweets with high-impact and sensational content are likely to generate more impressions, but the information described in the tweet must have a certain level of accuracy. As with the sanitization of structural formulas [27], it is also beneficial for peers of scientists to ensure each other's scientific soundness on the SNS timeline. An attempt to socially share ideas for compounds has also emerged very recently in the fight against the pandemic coronaviral disease COVID-19 (e.g., the COVID Moonshot project [28, 29, 30]), which is an example of the trend toward open science in drug discovery. Looking ahead to applications in drug discovery, we also need widgets in which users can interactively view and discuss the interaction and activity of compounds with target proteins, which is a future challenge. As a semi-permanently maintained database infrastructure, Twitter has the potential to store an unlimited amount of chemical information.

Conclusions

We presented three examples of SNS integrated cheminformatics tools in this study. Twitter integration can offer a convenient interface to run cheminformatics software
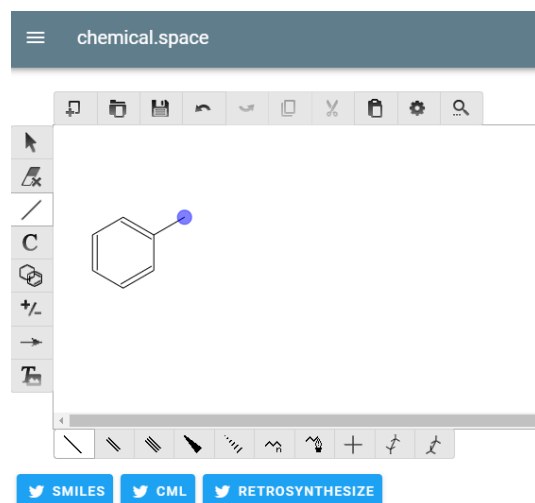
(a) Edit

(b) Share

(c) Display

Figure 3 Screenshots of the Twitter-integrated 2D editor. (a) The 2D molecule editor integrated with tweet buttons. (b) The tweet input window to share the manually drawn structural formula. The CML strings are encoded in the URL. (c) The 2D molecule viewer expanded on the Twitter timeline.

without a laborious environment setup, and users can easily share the result by tweet and retweet. Providing a demonstration on social media will be a new approach to present one's research and stimulate discussion in a more accessible manner. If you would like to add your own tool to our bot functions, please contact the authors for support.

Availability and requirements

Case Study 1

- Project name: Twitter Retrosynthesis Bot
- Project home page: https://github.com/n-yoshikawa/retrosynthesis_bot
- Operating system(s): Platform independent
- Programming languages: Python
- Other requirements: AiZynthFinder, RDKit, Tweepy, Pillow
- License: MIT License

Case Study 2 and 3

- Project name: chemical.space
- Project home page: https://github.com/ddquest/chemical.space
- Operating system(s): Platform independent
- Programming languages: Python
- Other requirements: LiteMol, RDKit SpreadsheetApp.openByUrl, Tweepy
- License: MIT License

Authors' information
Author details
[1]Department of Computer Science, University of Toronto, Toronto, Canada. [2]Isotope Science Center, The University of Tokyo, Tokyo, Japan. [3], SHaLX Inc., Tokyo, Japan. [4]Department of Computer Science, Tokyo Institute of Technology, Yokohama, Japan.

References
1. Reeser, D., Vincent-Ruz, P., Ashwell, S., Kemsley, J., Yarnell, A.: The secret silos of #ChemTwitter. https://cen.acs.org/sections/the-secret-silos-of-chemtwitter.html. Accessed 19 March 2021.
2. Priem, J., Groth, P., Taraborelli, D.: The altmetrics collection. PloS one **7**(11), 48753 (2012). doi:[10.1371/journal.pone.0048753](https://doi.org/10.1371/journal.pone.0048753)
3. Eysenbach, G.: Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. Journal of medical Internet research **13**(4), 123 (2011). doi:[10.2196/jmir.2012](https://doi.org/10.2196/jmir.2012)
4. Peoples, B.K., Midway, S.R., Sackett, D., Lynch, A., Cooney, P.B.: Twitter predicts citation rates of ecological research. PloS one **11**(11), 0166570 (2016). doi:[10.1371/journal.pone.0166570](https://doi.org/10.1371/journal.pone.0166570)

5. Klar, S., Krupnikov, Y., Ryan, J.B., Searles, K., Shmargad, Y.: Using social media to promote academic research: Identifying the benefits of twitter for sharing academic work. PloS one **15**(4), 0229446 (2020). doi:10.1371/journal.pone.0229446

6. Chu, Z., Gianvecchio, S., Wang, H., Jajodia, S.: Who is tweeting on Twitter: human, bot, or cyborg? In: Proceedings of the 26th Annual Computer Security Applications Conference, pp. 21–30 (2010). doi:10.1145/1920261.1920265

7. @retrosynthchan. https://twitter.com/retrosynthchan. Accessed 29 March 2021.

8. Genheden, S., Thakkar, A., Chadimová, V., Reymond, J.-L., Engkvist, O., Bjerrum, E.: AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. Journal of cheminformatics **12**(1), 1–9 (2020). doi:10.1186/s13321-020-00472-1

9. Weininger, D.: SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Journal of chemical information and computer sciences **28**(1), 31–36 (1988). doi:10.1021/ci00057a005

10. @souyakuchan. https://twitter.com/souyakuchan. Accessed 29 March 2021.

11. Statistics of souyakuchan magic. https://kubor.github.io/posts/souyakuchan_stats.html. Accessed 11 January 2021.

12. Drug Discovery (dry) Advent Calendar 2020. https://adventar.org/calendars/5119. Accessed 29 March 2021.

13. RDKit: Open-source cheminformatics. http://www.rdkit.org

14. POST statuses/filter. https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/api-reference/post-statuses-filter. Accessed 29 March 2021.

15. Tweepy: An easy-to-use Python library for accessing the Twitter API. https://www.tweepy.org/. Accessed 29 March 2021.

16. POST statuses/update. https://developer.twitter.com/en/docs/twitter-api/v1/tweets/post-and-engage/api-reference/post-statuses-update. Accessed 29 March 2021.

17. Pillow (PIL Fork) 8.1.2 documentation. https://pillow.readthedocs.io/. Accessed 29 March 2021.

18. Twitter-integrated 3D molecule viewer. https://chemical.space/viewer. Accessed 28 February 2021.

19. Developer Document for Twitter Player Card. https://developer.twitter.com/en/docs/twitter-for-websites/cards/overview/player-card. Accessed 28 February 2021.

20. Sehnal, D., Deshpande, M., Vařeková, R.S., Mir, S., Berka, K., Midlik, A., Pravda, L., Velankar, S., Koča, J.: LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. Nature Methods **14**(12), 1121–1122 (2017). doi:10.1038/nmeth.4499

21. Murray-Rust, P., Rzepa, H.S.: Chemical markup, xml, and the worldwide web. 1. basic principles. Journal of Chemical Information and Computer Sciences **39**(6), 928–942 (1999). doi:10.1021/ci990052b

22. Bienfait, B., Ertl, P.: JSME: a free molecule editor in JavaScript. Journal of cheminformatics **5**, 24 (2013). doi:10.1186/1758-2946-5-24

23. Csizmadia, P.: MarvinSketch and MarvinView: Molecule applets for the world wide web. In: Proceedings of The 3rd International Electronic Conference on Synthetic Organic Chemistry. MDPI, Basel, Switzerland (1999). doi:10.3390/ecsoc-3-01775

24. Jiang, C., Jin, X., Dong, Y., Chen, M.: Kekule.js: An open source javascript chemoinformatics toolkit. Journal of Chemical Information and Modeling **56**(6), 1132–1138 (2016). doi:10.1021/acs.jcim.6b00167

25. pako: zlib port to javascript, very fast! http://nodeca.github.io/pako/. Accessed 29 March 2021.

26. brianloveswords/base64url: Converting to, and from, base64url. https://github.com/brianloveswords/base64url. Accessed 29 March 2021.

27. Molecular Sanitization - The RDKit Book: The RDKit 2021.03.1 documentation. https://www.rdkit.org/docs/RDKit_Book.html. Accessed 29 March 2021.

28. The COVID Moonshot Consortium: COVID Moonshot: Open Science Discovery of SARS-CoV-2 Main Protease Inhibitors by Combining Crowdsourcing, High-Throughput Experiments, Computational Simulations, and Machine Learning (2020). doi:10.1101/2020.10.29.339317

29. Chodera, J., Lee, A.A., London, N., von Delft, F.: Crowdsourcing drug discovery for pandemics. Nat. Chem. **12**(7), 581–581 (2020). doi:10.1038/s41557-020-0496-2

30. PostEra: COVID Moonshot: An international effort to DISCOVER A COVID ANTIVIRAL. https://covid.postera.ai/covid. Accessed 29 March 2021.