TweeNLP: A Twitter Exploration Portal for Natural Language Processing

Viraj Shah, Shruti Singh and **Mayank Singh**
Indian Institute of Technology Gandhinagar, Gujarat, India
{shah.viraj, singh_shruti, singh.mayank}@iitgn.ac.in

Abstract

We present TWEENLP, a one-stop portal that organizes Twitter's natural language processing (NLP) data and builds a visualization and exploration platform. It curates 19,395 tweets (as of April 2021) from various NLP conferences and general NLP discussions. It supports multiple features such as TweetExplorer to explore tweets by topics, visualize insights from Twitter activity throughout the organization cycle of conferences, discover popular research papers and researchers. It also builds a timeline of conference and workshop submission deadlines. We envision TWEENLP to function as a collective memory unit for the NLP community by integrating the tweets pertaining to research papers with the NLPExplorer scientific literature search engine. The current system is hosted at URL.

1 Introduction

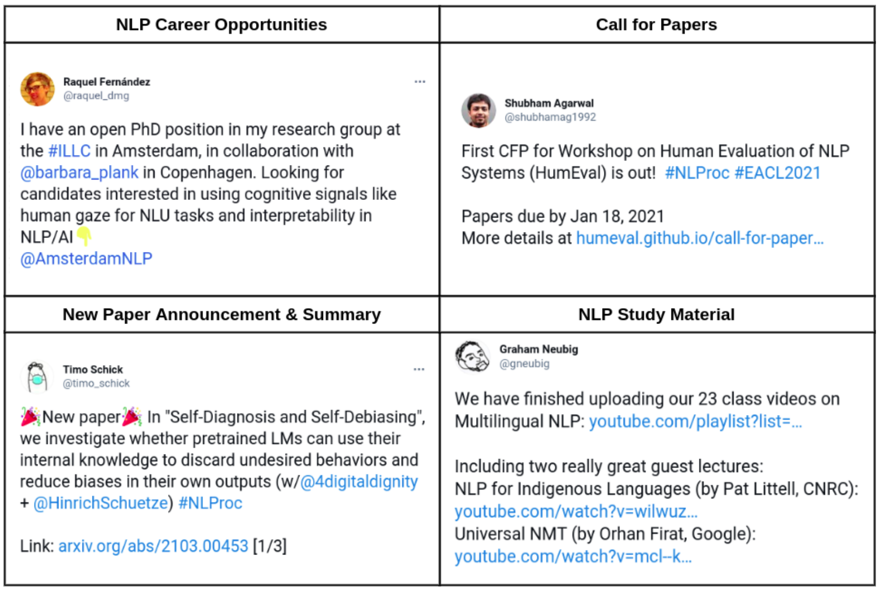
Online communication channels have become popular in the Internet era, and several online communities of like-minded people have evolved around these channels. For example, communities such as Stack Overflow and AskUbuntu are question-answering forums; Twitter and Reddit are content-sharing forums. These forums over the years have provided a platform for novice users to learn from the experts, facilitated discussions among the community members, and have over the years accumulated a rich database of questions, answers, and discussions.

According to the theory of diffusion of innovation proposed by Rogers (2003), the communication channel is one of the four main elements influencing the spread of a new idea. Notably, the communication channel serves as a collective long-term memory or a knowledge archive of the community, which any member can access to study the community's stance on diverse topics at any point in time.

Although several mailing lists, slack channels, and subreddits exist for communication, most natural language processing (henceforth NLP) community discussions are primarily carried out on Twitter due to its open accessibility and wider reach. Announcements of calls for papers and submission deadlines, recently accepted papers, interesting talks and seminars, lecture videos, and tutorials on various topics are often posted on Twitter. These are a great medium to stay updated on the recent developments in the NLP field. It is also a medium for researchers to engage in informal research discussions which might be unreported in official publications. We present a sample of diverse NLP tweets in Figure 1 to emphasise the utility of the platform.

However, unlike subreddits or communities like Stack Overflow and AskUbuntu, Twitter is not an exclusive channel for NLP discussions. Exclusive channels provide users a one-stop destination for their interests and allow extremely topic-specific exploration. While Twitter allows search by hashtags to narrow down to specific topics, the usage of hashtags is highly irregular. Furthermore, Twitter is more suited to live discussions and less suitable for maintaining a snapshot of the discussions taking place in the online community. Relevant Twitter discussions about specific research papers are often forgotten in the long run because there is no infrastructure to link these discussions with the papers on the proceedings archives or research paper search engines. In an attempt to address these issues, we extend the functionality of NLPExplorer (Parmar et al., 2020) platform by integrating TWEENLP with it. NLPExplorer is a portal for searching, and visualizing NLP research volume based on the ACL Anthology (ACL Anthology). In our current work, we build an automatic pipeline for curating NLP tweets and build a one-stop portal - TWEENLP, for the search and browsing of

Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations, pages 265–271, August 1st - August 6th, 2021.



NLP Career Opportunities

Raquel Fernández
@raquel_dmg

I have an open PhD position in my research group at the #ILLC in Amsterdam, in collaboration with @barbara_plank in Copenhagen. Looking for candidates interested in using cognitive signals like human gaze for NLU tasks and interpretability in NLP/AI👇
@AmsterdamNLP

Call for Papers

Shubham Agarwal
@shubhamag1992

First CFP for Workshop on Human Evaluation of NLP Systems (HumEval) is out! #NLProc #EACL2021

Papers due by Jan 18, 2021
More details at humeval.github.io/call-for-paper...

New Paper Announcement & Summary

Timo Schick
@timo_schick

🎉New paper🎉 In "Self-Diagnosis and Self-Debiasing", we investigate whether pretrained LMs can use their internal knowledge to discard undesired behaviors and reduce biases in their own outputs (w/@4digitaldignity + @HinrichSchuetze) #NLProc

Link: arxiv.org/abs/2103.00453 [1/3]

NLP Study Material

Graham Neubig
@gneubig

We have finished uploading our 23 class videos on Multilingual NLP: youtube.com/playlist?list=...

Including two really great guest lectures:
NLP for Indigenous Languages (by Pat Littell, CNRC): youtube.com/watch?v=wilwuz...
Universal NMT (by Orhan Firat, Google): youtube.com/watch?v=mcl--k...

Figure 1: A sample of diverse natural language processing tweets.

NLP discussion on Twitter. The system has curated 19,395 NLP tweets as of April 2021.

TWEENLP organizes NLP tweets into topics: (i) New paper announcements, (ii) Call for Paper announcements, (iii) Reading Materials & Tutorials, (iv) Career Opportunities, (v) Talks & Seminars, and (vi) Others. Topic-wise tweets are presented via dashboards for easy exploration. TWEENLP supports dashboards to browse through popular NLP tweets in the previous week and the month. We construct a CFP Timeline from ‘Call for Papers’ announcements on Twitter and arrange it according to the upcoming submission deadlines of various workshops and conferences. We link the research paper tweets to the research paper’s metadata, accessible via the NLPExplorer paper discovery feature. We also build live Conference Visualization dashboards, which curate tweets about the conference schedule, ongoing talks, poster sessions, and interesting papers at the conference, and present statistics such as popular hashtags, users, tweet languages, etc.

We integrate TWEENLP with NLPExplorer (Section 2) to build a joint-portal that aims to bridge the gap between published research and its informal communication on the social media platform Twitter. Our automatic data curation pipeline and the architecture of the system is described in Section 3 and Section 4 respectively. We describe the features of TWEENLP in detail in Section 5. In Section 6, we discuss previous works in organizing the NLP literature and visualization of research papers.

2 NLPExplorer

NLPExplorer[1] (Parmar et al., 2020) is an automatic portal for indexing, searching, and visualizing Natural Language Processing research volume. It presents multiple paper, venue, and author statistics, including paper citation distribution, paper topic distribution, authors, their field of study, their citation distributions, etc. It also presents category information of research papers into various topics broadly arranged in five categories: (i) Linguistic Target (Syntax, Discourse, etc.), (ii) Tasks (Tagging, Summarization, etc.), (iii) Approaches (unsupervised, supervised, etc.), (iv) Languages (English, Chinese, etc.) and (v) Dataset types (news, clinical notes, etc.). The current snapshot consists of 75k research papers and 50k authors. Since its inception, it has been accessed by more than 7.3k users having a close to 9.7k sessions.

3 Dataset

We curate the dataset from two primary sources:

3.1 Twitter

We curate the Twitter data using the open-source library *Twint*[2] by retrieving tweets with the hashtag NLProc. We also curate tweets with NLP conference hashtags such as #acl2020, #emnlp2020, etc. The list of NLP conferences is compiled via ACL

[1] http://nlpexplorer.org/

[2] https://github.com/twintproject/twint

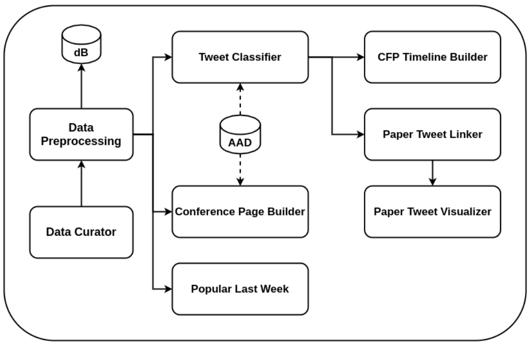


Figure 2: The architecture of TWEENLP. Arrow directions denote the flow of data. AAD represents the ACL Anthology Dataset which is the other data source apart from Twitter.

Anthology. Our system is scheduled to download the Twitter data for each day automatically. For ongoing conferences, our system curates new tweets every hour to continually update the *Conference Visualizer* page. The current snapshot (as of April 2021) contains data since October 2017 (around 1300 days) and consists of 19,395 tweets.

3.2 ACL Anthology

We curate the conference and journal names and URLs from the ACL Anthology github repository[3]. We also curate the paper titles and their links. Tweets are collected periodically every day, and the system checks for paper mentions in the tweets by substring matching the paper URLs collected from the ACL Anthology github repository.

4 Architecture

We present the pipeline of our system in Figure 2. The Data Curator module curates tweets daily. The curated tweets are processed before we perform further steps. The following modules process tweets: (i) Tweet Classifier, (ii) Conference Page Builder, (iii) CFP Timeline Builder, and (iv) Paper Tweet Linker. We describe the tweet processing modules in detail below:

1. *Tweet Classifier*: The Tweet Classifier module classifies a tweet into one of the six topics: (i) New Paper Announcements, (ii) Call for Paper announcements, (iii) Reading Materials & Tutorials, (iv) Career Opportunities, (v) Talks & Seminars, and (vi) Others. The Tweet Explorer feature utilizes these tweet categories. The detailed description of each topic is presented in Section 5.1. We experiment by fine-tuning a BERT-base(Devlin et al., 2019) classifier and twitter-roberta-base(Barbieri et al., 2020) to predict the tweet topics. The BERT-base model[4] obtains the best test accuracy of 75% on a small manually annotated dataset[5].
2. *Conference Page Builder*: The Conference Page Builder classifies a tweet either as discussing an ongoing conference or other topics. The module builds specific conference pages using such tweets.
3. *CFP Timeline Builder*: The module processes 'Call for Papers' tweets identified by the Tweet Classifier module. It extracts the conference (and workshop) name by regex-based keyword matching against a pre-compiled list of venues. The submission date are extracted from the tweets by labeling dates using the Spacy[6] library. The tweets are arranged in a timeline sorted by the submission deadline.
4. *Paper Tweet Linker*: The Paper Tweet Linker module maps specific tweets to research papers using regex matching of the paper title and paper URL. The Paper Tweet Visualizer uses these mappings to embed the tweets on the research paper page on NLPExplorer.

The pipeline then stores the tweets in the database after processing by the above modules. We schedule our system to automatically curate the Twitter data daily and increase it to an hourly frequency during ongoing conferences.

5 TWEENLP Features

5.1 Tweet Explorer

We present a Tweet Explorer dashboard that allows a user to browse tweets from specific topics such as:

1. *New paper announcements:* This topic organizes tweets about recent papers, which often involve the summary or a short introduction of the research paper. These twitter threads facilitate other researchers to communicate informally with the paper authors. These also contain interesting discussions by the community on the insights, merits, and critiques of the research paper, and post questions about

[3] https://github.com/acl-org/acl-anthology

[4] We also experimented with a zero-shot classifier but it underperformed the BERT-base classifier.

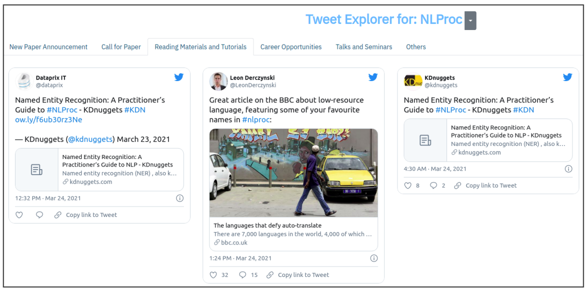
[5] each tweet was annotated by two ML/NLP students and inter annotator agreement computed using Cohen's κ=0.68

[6] https://spacy.io/

the work. The authors’ short introductions offer an informal account of the paper compared to the paper alert services that usually present the title and the abstract of the research paper.

2. *Call for Papers (CFPs) by various conferences and workshops:* Users can view the announcements for call for papers and submission deadlines by various workshops and conferences.
3. *Reading Materials & Tutorials:* It lists various study material, such as lecture slides and videos, tutorials, online courses, and blog posts.
4. *NLP Career Opportunities:* Individuals frequently advertise opportunities for various positions such as interns, full-time, Ph.D., postdoctoral fellows, and research fellows on Twitter.
5. *NLP Talks & Seminars:* Various online NLP talks and seminars can be accessed using the NLP Talks & Seminars filter on the Tweet Explorer dashboard.
6. *Others:* This category contains the NLP tweets which do not belong to any of the above topics.

The Tweet Explorer feature allows users to specifically browse through tweets by topics and filter them based on their immediate interests. A snapshot of the same is presented in Figure 3. We present the distribution of tweets in the six categories from tweets curated by the system in the last 1,300 days in Table 1.



Dataprix IT @dataprix

Named Entity Recognition: A Practitioner’s Guide to #NLProc - KDnuggets #KDN ow.ly/f6ub30rz3Ne

— KDnuggets (@kdnuggets) March 23, 2021

Named Entity Recognition: A Practitioner’s Guide to NLP - KDnuggets
Named entity recognition (NER) , also k...
kdnuggets.com

12:32 PM · Mar 24, 2021

Leon Derczynski @LeonDerczynski

Great article on the BBC about low-resource language, featuring some of your favourite names in #nlproc:

The languages that defy auto-translate
There are 7,000 languages in the world, 4,000 of which ...
bbc.co.uk

1:24 PM · Mar 24, 2021

32 15

KDnuggets @kdnuggets

Named Entity Recognition: A Practitioner’s Guide to #NLProc - KDnuggets #KDN

Named Entity Recognition: A Practitioner’s Guide to NLP - KDnuggets
Named entity recognition (NER) , also k...
kdnuggets.com

4:30 AM · Mar 24, 2021

8 2

Figure 3: Tweet Explorer feature of TWEENLP which facilitates browsing tweets by six different topics.

5.2 Conference Visualizer – Near real-time view for conferences

TWEENLP supports real-time statistics for multiple top conferences and the popular #NLProc hashtag. The information is updated hourly for live events and weekly for past events. Some of the statistics presented are top mentions, top hashtags, top linked URLs, and top discussed papers in tweets. We present the most popular hashtags, mentions, URLs, and highly discussed papers for ACL2020 in Table 2. A summary of Twitter activity from the Conference Visualizer page for ACL 2020 is presented in Table 3. Apart from Twitter discussions about a conference in a specific month, we also show insights from the conferences across the year. The insights from ACL conference over time is presented in Figure 4. We also present other conference-specific statistics such as the number of tweets per month, daily distribution of tweets in the conference month, most active users tweeting about the conference, and a distribution of the tweet languages other than English.

Topic	Tweet Count
New Paper Announcements	6,337
Call for Papers	972
Reading Materials & Tutorials	1,400
NLP Career Opportunities	681
NLP Talks & Seminars	2,382
Others	7,623
Total Tweets	**19,395**

Table 1: Distribution of tweets (curated since October 2017) into various topics.

5.3 Popular Paper Visualizer

We showcase widely discussed papers on Twitter in the Popular Paper Visualizer dashboard. It presents the titles and provides direct links to the full-text of the top discussed papers for quick reference. The system extracts tweets mentioning research papers and assigns a popularity score to each paper based on the count of tweets that mention it, and the likes, retweets, and replies on the paper tweets. We present a snapshot of few popular papers identified by our platform in Figure 5. It also presents the most active users tweeting about #NLProc on Twitter. Popular Paper Visualizer dashboard also supports exploration of most liked and retweeted #NLProc tweets of all times and in the last month.

5.4 CFP Timeline

TWEENLP presents a timeline of the upcoming submission deadlines. The timeline is created by identifying ‘Call for Papers’ tweets using keyword based filtering of tweets and also lists the conference/workshop website. The details are described in the CFP Timeline Builder module 3. We present a snapshot of the timeline in Figure 6.

Top Hashtags	Top Mentions	Top URLs	Top Papers Discussed
#acl2020nlp	@aclmeeting	virtual.acl2020.org/socials.html	Beyond Accuracy: Behavioral Testing of NLP models with CheckList
#acl2020en	@emilymbender	virtual.acl2020.org/plenary_session_keynote_kathy_mckeown.html	Photon: A Robust Cross-Domain Text-to-SQL System
#nlproc	@akoller	virtual.acl2020.org/paper_main.701.html	Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data
#acl2020zht	@winlpworkshop	virtual.acl2020.org/workshop_W1.html	Language Models as an Alternative Evaluator of Word Order Hypotheses: A Case Study in Japanese
#acl2020hi	@xandaschofield	www.aclweb.org/anthology/2020.acl-main.442/	Don't Stop Pretraining: Adapt Language Models to Domains and Tasks
#mt	@gneubig	virtual.acl2020.org/workshop_W10.html	The State and Fate of Linguistic Diversity and Inclusion in the NLP World

Table 2: ACL 2020 Twitter Coverage: Top discussed papers, mentions and URLs and popular hashtags.

Tweet Counter	Likes Counter	Retweet Counter	Unique Mentions	Unique Paper Mentions
5,343	58,160	11,440	907	251

Table 3: ACL 2020 Twitter Statistics of various activities.

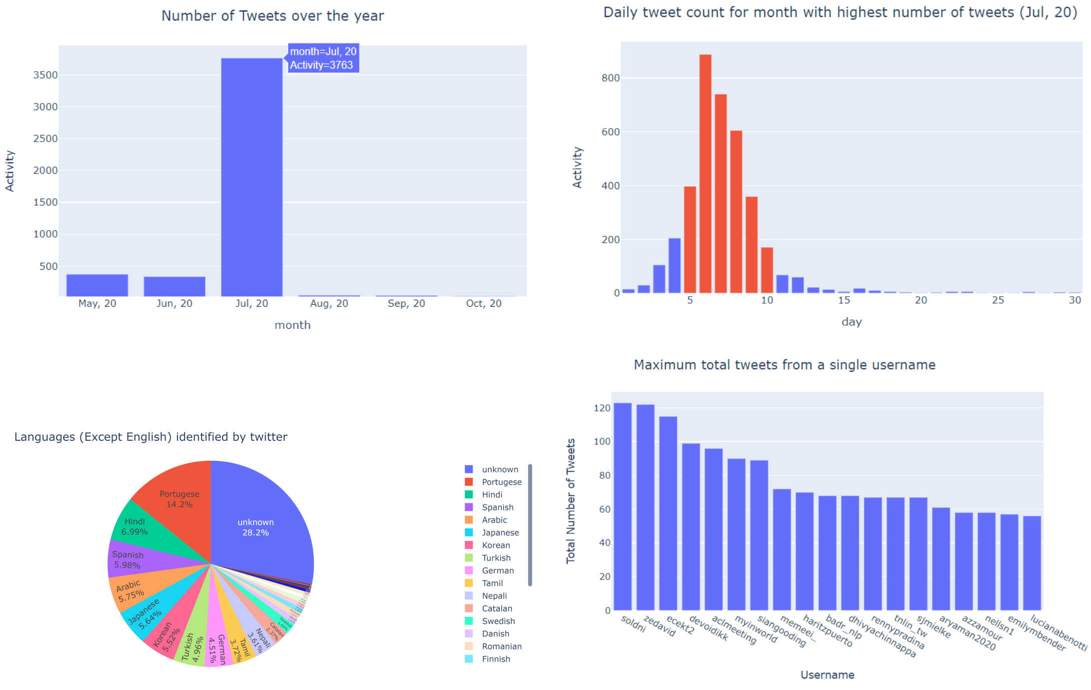
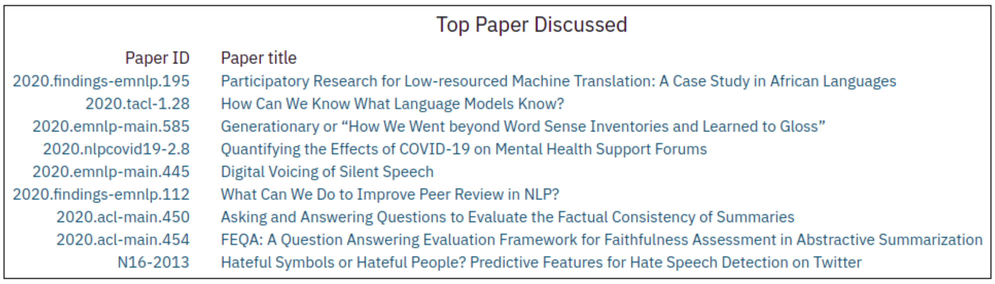


Figure 4: Conference Visualizer: ACL2020 Statistics. (a) Distribution of tweets across different months. (b) Daily distribution of ACL2020 tweets in July 2020. (c) Distribution of tweet languages except English. (d) Twitter users with highest ACL2020 tweets.

5.5 Paper Tweet Visualizer

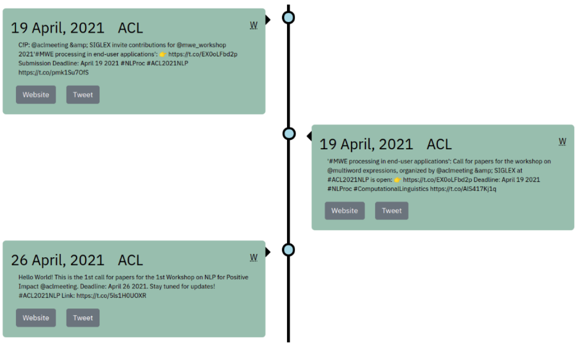
NLPExplorer supports a research paper search interface and builds research paper pages which showcase standard paper related statistics such as the publication year and venue, author information, citations, citation distribution over the years and the link to the corresponding PDF article. Additionally, it also provides interesting insights like similar papers, topical distribution and mentioned URLs. We map research paper discussion tweets on Twitter to the NLPExplorer paper page. This feature allows users to browse through discussions about the paper along with the metadata of the paper. We present a snapshot of the feature in Figure 7.



Top Paper Discussed

Paper ID	Paper title
2020.findings-emnlp.195	Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages
2020.tacl-1.28	How Can We Know What Language Models Know?
2020.emnlp-main.585	Generationary or "How We Went beyond Word Sense Inventories and Learned to Gloss"
2020.nlpcovid19-2.8	Quantifying the Effects of COVID-19 on Mental Health Support Forums
2020.emnlp-main.445	Digital Voicing of Silent Speech
2020.findings-emnlp.112	What Can We Do to Improve Peer Review in NLP?
2020.acl-main.450	Asking and Answering Questions to Evaluate the Factual Consistency of Summaries
2020.acl-main.454	FEQA: A Question Answering Evaluation Framework for Faithfulness Assessment in Abstractive Summarization
N16-2013	Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter

Figure 5: Popular papers identified TWEENLP based on twitter activity.



19 April, 2021 ACL

W

CfP: @aclmeeting &amp; SIGLEX invite contributions for @mwe_workshop 2021'#MWE processing in end-user applications': 👉 https://t.co/EX0oLFbd2p Submission Deadline: April 19 2021 #NLProc #ACL2021NLP https://t.co/pmk1Su7OfS

19 April, 2021 ACL

W

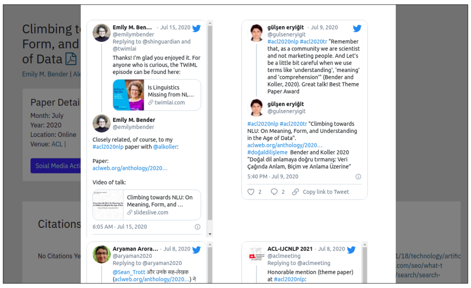
'#MWE processing in end-user applications': Call for papers for the workshop on @multiword expressions, organized by @aclmeeting &amp; SIGLEX at #ACL2021NLP is open: 👉 https://t.co/EX0oLFbd2p Deadline: April 19 2021 #NLProc #ComputationalLinguistics https://t.co/AlS417Kj1q

26 April, 2021 ACL

W

Hello World! This is the 1st call for papers for the 1st Workshop on NLP for Positive Impact @aclmeeting. Deadline: April 26 2021. Stay tuned for updates! #ACL2021NLP Link: https://t.co/5ls1H0UOXR

Figure 6: CFP Timeline built from tweets. 'W' on top-right denotes Workshop.



Emily M. Ben... · Jul 15, 2020
@emilymbender
Replying to @shinguardian and @twimlai

Thanks! I'm glad you enjoyed it. For anyone who is curious, the TWiML episode can be found here:

Is Linguistics Missing from NL...
twimlai.com

Emily M. Bender
@emilymbender

Closely related, of course, to my #acl2020nlp paper with @alkoller:

Paper:
aclweb.org/anthology/2020...

Video of talk:

Climbing towards NLU: On Meaning, Form, and ...
slideslive.com

6:05 AM · Jul 15, 2020

gülşen eryiğit · Jul 9, 2020
@gulseneryigit
#acl2020nlp #acl2020tr "Remember that, as a community we are scientist and not marketing people. And Let's be a little bit careful when we use terms like 'understanding', 'meaning' and 'comprehension'" (Bender and Koller, 2020). Great talk! Best Theme Paper Award

gülşen eryiğit
@gulseneryigit

#acl2020nlp #acl2020tr "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data".
aclweb.org/anthology/2020...
#doğaldilişleme Bender and Koller 2020 "Doğal dil anlamaya doğru tırmanış: Veri Çağında Anlam, Biçim ve Anlama Üzerine"

5:40 PM · Jul 9, 2020

2 2 Copy link to Tweet

Aryaman Arora... · Jul 8, 2020
@aryaman2020
Replying to @aryaman2020
@Sean_Trott और उनके सह-लेखक (aclweb.org/anthology/2020...) ने

ACL-IJCNLP 2021 · Jul 8, 2020
@aclmeeting
Replying to @aclmeeting
Honorable mention (theme paper) at #acl2020nlp:

Figure 7: Paper Tweet Visualizer curates tweets and metadata of a research paper on a joint portal. The image background is a 'Paper' page from NLPExplorer which lists paper metadata, citing papers, field-of-study tags, and similar papers alongwith the associated tweets.

5.6 Popular Last Week

Lastly, we present popular tweets in the NLP community on Twitter (also referred as NLP Twitter). This feature allows researchers to catch up with the recent NLP-related Twitter discussions in a single dashboard without searching for them specifically in the Twitter feed.

6 Related Works

Bird et al. (2008) curated the *ACL Anthology Reference Corpus (ACL ARC)* of research papers in NLP and CL. Radev et al. (2009, 2013) constructed the ACL Anthology Network (AAN) by manual annotation of the references to complete the citation network and analysed the network to present central papers, authors and other network statistics (Radev et al., 2016). Works by Schäfer et al. (2011) and Parmar et al. (2020) provide a comprehensive search interface to browse through the NLP based on parameters such as author, full text, year of publication, title, and the field of study. Mohammad (2020) built the NLPScholar platform which consists of interactive dashboards that present various aspects of NLP research papers. The platform uses ACL Anthology and Google Scholar as the information source.

Few works have analysed Twitter data to predict scholarly impact. Shuai et al. (2012) report a statistical correlation between high volume of Twitter mentions and arXiv downloads and early citations (i.e., citations occurring less than seven months after the publication of a preprint). However, they also point out that Twitter mentions cannot be directly concluded to be causative of higher levels of download and early citations. Several other works such as Eysenbach (2011), Thelwall et al. (2013), and Haustein et al. (2014) have tried to analyze whether tweets correlate with citations.

However, to the best of our knowledge, no prior work has tried to curate NLP discussions data from Twitter in an attempt to organize it and link it to research papers via a search engine or a visualization portal.

7 Future Scope and Extensions

Currently, the system is implemented only for NLP papers present in the ACL Anthology. The system could be extended to papers from NeurIPS, ICLR, and CVPR as the data for these conferences is available publicly. The system is versatile and can be easily extended to other domains. TWEENLP provides basic visualization graphs over Twitter activity. Over time, these discussions could be used to build a timeline of evolution of research in various domains of NLP based on the Twitter activity of researchers. Tweets by popular users attain likes and retweets at a higher rate in comparison to new users (or users with less followers) of the community. TWEENLP currently only presents popular tweets based on retweets and likes count which can bias the conversations, understanding and presentation of ideas by emphasising the tweets of a small set of popular users. Future work includes identifying novel alternative ideas and perspectives by adjusting user popularity to create an inclusive space for the community.

References

ACL Anthology. ACL Anthology. https://www.aclweb.org/anthology/. Accessed: 2021-03-01.

Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa-Anke, and Leonardo Neves. 2020. TweetEval:Unified Benchmark and Comparative Evaluation for Tweet Classification. In *Proceedings of Findings of EMNLP*.

Steven Bird, Robert Dale, Bonnie J Dorr, Bryan Gibson, Mark Thomas Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir R Radev, and Yee Fan Tan. 2008. The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics. *Language Resources and Evaluation Conference*.

J. Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*.

Gunther Eysenbach. 2011. 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):e123.

Stefanie Haustein, Isabella Peters, Cassidy R Sugimoto, Mike Thelwall, and Vincent Larivière. 2014. Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology*, 65(4):656–669.

Saif M. Mohammad. 2020. NLP scholar: An interactive visual explorer for natural language processing literature. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 232–255, Online. Association for Computational Linguistics.

Monarch Parmar, Naman Jain, Pranjali Jain, P Jayakrishna Sahit, Soham Pachpande, Shruti Singh, and Mayank Singh. 2020. NLPExplorer: exploring the universe of nlp papers. In *European Conference on Information Retrieval*, pages 476–480. Springer.

Dragomir R Radev, Mark Thomas Joseph, Bryan Gibson, and Pradeep Muthukrishnan. 2016. A bibliometric and network analysis of the field of computational linguistics. *Journal of the Association for Information Science and Technology*, 67(3):683–706.

Dragomir R Radev, Pradeep Muthukrishnan, and Vahed Qazvinian. 2009. The ACL Anthology Network Corpus. *ACL-IJCNLP*, page 54.

Dragomir R Radev, Pradeep Muthukrishnan, Vahed Qazvinian, and Amjad Abu-Jbara. 2013. The ACL anthology network corpus. *Language Resources and Evaluation*, 47(4):919–944.

Everett M. Rogers. 2003. *Diffusion of innovations*, 5th edition. Free Press, New York, NY [u.a.].

Ulrich Schäfer, Bernd Kiefer, Christian Spurk, Jörg Steffen, and Rui Wang. 2011. The ACL Anthology searchbench. In *Proceedings of the ACL-HLT 2011 System Demonstrations*, pages 7–13, Portland, Oregon. Association for Computational Linguistics.

Xin Shuai, Alberto Pepe, and Johan Bollen. 2012. How the scientific community reacts to newly submitted preprints: Article downloads, twitter mentions, and citations. *PloS one*, 7(11):e47523.

Mike Thelwall, Stefanie Haustein, Vincent Larivière, and Cassidy R Sugimoto. 2013. Do altmetrics work? twitter and ten other social web services. *PloS one*, 8(5):e64841.