My Desktop
Prepare & Submit Proposals
New! Prepare Proposals (Limited proposal types)
Prepare Proposals in FastLane
Proposal Status
Awards & Reporting
Notifications & Requests
Project Reports
Award Functions
Manage Financials
Program Income Reporting
Grantee Cash Management Section Contacts
Administration
Lookup NSF ID

Preview of Award 1619028 - Final Project Report

Cover |
Accomplishments |
Products |
Participants/Organizations |
Impacts |
Changes/Problems

Cover

Federal Agency and Organization Element to Which Report is Submitted:	4900
Federal Grant or Other Identifying Number Assigned by Agency:	1619028
Project Title:	III: Small: Collaborative Research: Global Event and Trend Archive Research (GETAR)
PD/PI Name:	Edward A Fox, Principal Investigator Andrea L Kavanaugh, Co-Principal Investigator Chandan K Reddy, Co-Principal Investigator Donald J Shoemaker, Co-Principal Investigator
Recipient Organization:	Virginia Polytechnic Institute and State University
Project/Grant Period:	01/01/2017 - 12/31/2019
Reporting Period:	01/01/2019 - 12/31/2019
Submitting Official (if other than PD\PI):	N/A
Submission Date:	N/A
Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions)	N/A

Accomplishments

*** What are the major goals of the project?**

We will ingest tweets and Web-based content from social media and the general Web, including news and governmental information. In addition to archiving materials found, we will build an information system that includes related metadata and knowledge bases, consistent with the 5S (Societies, Scenarios, Spaces, Structures, Streams) framework, along with results from our intelligent focused crawler, to support comprehensive access to event related content. With the support of key partners, the GETAR team will undertake research, education, and dissemination efforts, to achieve three complementary

objectives:
1. Collecting: We will spot, identify, and make sense of interesting events and trends. We also will accept specific or general requests about types of events. Given resource and sampling constraints, we will integrate methods to identify appropriate URLs as seeds, and specify when to start crawling and when to stop, with regard to each event or subevent. We will integrate focused crawling and filtering approaches in order to ingest content and generate new collections, with high precision and recall.
2. Archiving & Accessing: Permanent archiving, and access to those archives, will be ensured by our partner, Internet Archive (IA). Immediate access to ingested content will be facilitated through big data software built on top of our Hadoop cluster.
3. Analyzing & Visualizing: We will provide a wide range of integrated services beyond the usual (faceted) browsing and searching, including: classification, clustering, summarization, text mining, topic identification, trend analysis, and visualization.

*** What was accomplished under these goals (you must provide information for at least one of the 4 categories below)?**

Major Activities:

The GETAR project, with more than 50 collaborators and 12 collaborating institutions, developed tweet and webpage collections, datasets, services, software, systems, and methods. The related activities include: collecting event-related content, software and system development and refinement, experimentation, evaluation, and working with diverse users (representing key stakeholder groups). Studies have proceeded of important events, including integration of survey and analysis approaches, and publishing findings. Organizations studied included corporations, hotels, and tourism agencies. Technologies included deep learning and reinforcement learning, as well as natural language processing and information retrieval.

The Internet Archive (see also its separate report for IIS-1619371) has expanded its collections and technology support, as well as outreach activities. It has expanded its R&D staffing. It hosts, preserves, and provides public access with attribution to web collections created by the project team through its public Wayback Machine interface and Archive-It service. The latter may be browsed by descriptive metadata and searched through Archive-It's full-text Elasticsearch engine. New or updated Internet Archive and Archive-It API documentation and workshops provide project stakeholders with several means to query the data from and about these collections, and to derive datasets for further textual and visual analyses. It also provided multiple on-request collections of webpages for studies at Virginia Tech, as well as supported programatic and other direct access to archives.

Specific Objectives:

Objectives included integrating digital libraries with archives, developing methods and systems to work with events and trends, and connecting stakeholders and subject-matter experts with a broad team of researchers and students, including those in courses or volunteering. Project infrastructure (including a Hadoop cluster) was supplemented by department and university facilities, to efficiently and effectively handle large collections and deep learning requirements. Leveraging roughly 5 billion tweets and hundreds of millions of webpages, in over 1700 collections, scalable methods were validated.

Regarding collections, prior collections were extended, new ones were launched as events occurred or requests were made by users, the event focused crawler was deployed and released through Code Ocean, and diverse related curation efforts proceeded. Student (independent study, undergraduate. and graduate courses) research led to improved tweet and webpage techniques for content cleaning and processing, information extraction, classification, clustering, topic analysis, sentiment analysis, indexing, searching, browsing, and visualization.

Significant Results:

Advances have been made in big data handling, computational linguistics, digital libraries, information retrieval, information visualization, machine learning, and Web archiving. These have been integrated into a large system built around a Hadoop cluster, that works with growing numbers of expanding collections of tweets and

webpages, supplemented by cleaning, information extraction, and adding value through advanced analysis.

The GETAR project has developed novel methodology and workflows, tailored to addressing the challenging problem of working with events and trends. At a high level, for collection building, is a workflow to collect tweets about each event or event class, extract URLs, use the URLs present therein as seeds to our event focused crawler, and add resulting webpages to our Web collection. The event focused crawler workflow uses the extracted URLs as seeds to construct an event model that guides the selection and focused crawling for webpages. A new method to detect events, with a more elaborate event model, was devised using both tweets and news. Key new methods were developed to analyze and accordingly add value (and metadata) to the collected content, such as through multi-document summarization.

Regarding our processing of tweets, a new framework was extended to streamline a variety of tweet analysis and transformation workflows. This led to a new methodology for building digital libraries, integrating UX and system development processes by linking a knowledge graph with workflows.

Publications described improved digital library methods and services including: classification, crawling, hashing, question answering, recommendation, summarization, topic modeling, and visualization. They covered a variety of content sources, including: biomedical/healthcare information, documents (e.g., theses/dissertations), images, news, reviews, tweets, and webpages. Some focused on specific types of events, such as: crime, disasters, elections, and shootings. Analyses described communication, diffusion/propagation, moods, and patterns, with both spatial and temporal aspects.

Key outcomes or Other achievements:

A number of new methods were developed for text topic modeling and summarizing reviews. A variety of methods were evaluated with multiple events to automatically construct summaries from large linked collections of tweets and related webpages.

Collection building and analysis (of both tweets and webpages) has improved through advances in classification, big data workflows, focused crawling (to identify webpages focused on an event of interest), determining the location of tweets, topic analysis, and natural language processing (including Arabic). Insights gained have been shared regarding crime, school shootings, and the use of information during conflicts, crises, elections, and uprisings.

Collections are available to support other research and exploration regarding important events since 2007 such as the above, as well as attacks, bombings, celebrations, climate change, collapses, community activities, crashes, disease outbreaks, earthquakes, eclipses, environmental disruptions, erosion, explosions, fires, floods, hurricanes, innovations, judicial decisions, pollution, power outages, protests, revolutions, shootings, sports, storms, summits, tornadoes, transportation failures, tsunamis, typhoons, and veteran activities. Collaboration has expanded to support a broad set of researchers.

*** What opportunities for training and professional development has the project provided?**

Two doctoral dissertations were completed in early 2020, and two more will be completed later in the year. In Fall 2019, the advanced graduate course on Digital Libraries (CS6604) had archiving as a key focus, and two of the team projects related to GETAR (see from https://vtechworks.lib.vt.edu/handle/10919/47780).

In Spring 2019, two teams in CS4624 (Multimedia, Hypertext, and Information Access) worked on projects related to GETAR, also uploading deliverables (e.g., reports, presentations, data, code), accessible from https://vtechworks.lib.vt.edu/handle/10919/18655. These joined those of projects from prior years, and will be supplemented by 6 more in May 2020.

*** How have the results been disseminated to communities of interest?**

Dissemination has been through the reported publications and presentations. Further dissemination was through the project website (http://eventsarchive.org) and the website connected to the tweet collections and descriptions. In addition, we helped lead the 2019 Web Archiving and Digital Libraries (WADL) workshop.

Products

Books

Book Chapters

Inventions

Journals or Juried Conference Papers

Aman Ahuja, Ashish Baghudana, Wei Lu, Edward A. Fox, and Chandan K. Reddy (2019). Spatio-Temporal Event Detection from Multiple Data Sources. *Proc. 23rd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2019), 14-17 April 2019, Macau, China, http://dmkd.cs.vt.edu/papers/PAKDD19.pdf*. 293. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1007/978-3-030-16148-4_23

Andrea Kavanaugh, Ziqian Song, Liuqing Li, Edward A. Fox, Bethany Hsaio (2020). Too Small to Fail: Information Sharing Behavior in a US Municipal Election. *Proc. 21st Annual International Conference on Digital Government Research (dg.o 2020), Seoul, South Korea, ACM, 14 pages*. . Status = AWAITING_PUBLICATION; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Andrea L. Kavanaugh, Ziqian Song, Liuqing Li, Edward A. Fox (2019). Communication Behavior in an Emerging Democracy: Political Expression via Tweets during the 2014 Tunisian Elections. *Proceedings of the 20th Annual International Conference on Digital Government Research (dg.o 2019) June 18-20, Dubai, United Arab Emirates, ACM, 2019*. 445. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1145/3325112.3325263

Florian J. Zach, Yufeng Ma, and Edward A. Fox (2019). A Preliminary Analysis of Images in Online Hotel Reviews. *ENTER 2019: The 26th Annual eTourism Conference, Nicosia, Cyprus, 30 January - 1 February, 2019; also in Review of Tourism Research (eRTR), Springer journal special issue*. 16 (2/3), 156. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: https://journals.tdl.org/ertr/index.php/ertr/article/view/328

Khoa D. Doan, Pranjul Yadav, and Chandan K. Reddy (2019). Adversarial Factorization Autoencoder for Look-alike Modeling. *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM), Beijing, China, November 2019*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: http://dmkd.cs.vt.edu/papers/CIKM19.pdf

Khoa Doan and Chandan K. Reddy (2020). Efficient Implicit Unsupervised Text Hashing using Adversarial Autoencoder. *Proceedings of The Web Conference (WWW), Taipei, Taiwan, April 2020*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: http://dmkd.cs.vt.edu/papers/WWW20a.pdf

Khoa Doan, Guolei Yang, and Chandan K Reddy (2019). Attentive Spatio-Temporal Neural Model for Successive Point of Interest Recommendation. *Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Macau, China, April 2019*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: http://dmkd.cs.vt.edu/papers/PAKDD19a.pdf

Li, Liuqing; Fox, Edward A. (2019). Understanding patterns and mood changes through tweets about disasters. *Proc. ISCRAM 2019, 16th International Conference on Information Systems for Crisis Response and Management, Valencia, Spain, 19-22 May 2019: ISCRAM*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: http://idl.iscram.org/files/liuqingli/2019/1863_LiuqingLi+EdwardA.Fox2019.pdf

Liuqing Li and Edward Fox (2020). Disaster Response Patterns across Different User Groups on Twitter: A Case Study during Hurricane Dorian. *WiP/Practitioner paper in Proceedings ISCRAM 2020, May 2020, also in the ISCRAM Digital Library, to be presented in May 2021 at ISCRAM 2021, Blacksburg, VA, 11 pages*. . Status = AWAITING_PUBLICATION; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Liuqing Li, Jack Geissinger, William A. Ingram, Edward A. Fox (2020). Teaching Natural Language Processing through Big Data Text Summarization with Problem-Based Learning. *Data and Information Management, ISSN:2543-9251*. 4 (4), . Status

= PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.2478/dim-2020-0003

Ming Zhu, Aman Ahuja, Wei Wei, and Chandan K. Reddy (2019). A Hierarchical Attention Retrieval Model for Healthcare Question Answering. *Proceedings of The Web Conference (WWW), San Francisco, CA, May 2019*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: http://dmkd.cs.vt.edu/papers/WWW19.pdf

Ming Zhu, Busra Celikkaya, Parminder Bhatia, and Chandan K. Reddy (2020). LATTE: Latent Type Modeling for Biomedical Entity Linking. *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), New York, NY, February 2020*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: http://dmkd.cs.vt.edu/papers/AAAI20.pdf

Ping Wang, Tian Shi, and Chandan K. Reddy (2020). Tensor-based Temporal Multi-Task Survival Analysis. *IEEE Transactions on Knowledge and Data Engineering (TKDE), http://dmkd.cs.vt.edu/papers/TKDE20.pdf*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1109/TKDE.2020.2967700

Ping Wang, Tian Shi, and Chandan K. Reddy (2020). Text-to-SQL Generation for Question Answering on Electronic Medical Records. *Proceedings of The Web Conference (WWW), Taipei, Taiwan, April 2020*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: http://dmkd.cs.vt.edu/papers/WWW20.pdf

Sheetz, Steven; Kavanaugh, Andrea; Fox, Edward; Hassan, Riham; Yang, Seungwon; Magdy, Mohamed; Donald, Shoemaker (2019). Information Uses and Gratifications Related to Crisis: Student Perceptions since the Egyptian Uprising. *Proc. ISCRAM 2019, 16th International Conference on Information Systems for Crisis Response and Management*. 674-690. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: http://idl.iscram.org/files/stevensheetz/2019/1862_StevenSheetz_etal2019.pdf

Shuo Niu, D. Scott McCrickard, Julia Nguyen, Derek Haqq, Lindah Kotut, Timothy L. Stelter, Edward A. Fox (2020). Investigating Paradigms of Group Territory in Multiple Display Environments. *Proceedings of the ACM (PACM) on Human-Computer Interaction, January 2020, Article No. 13*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1145/3375193

Steven D. Sheetz, Andrea Kavanaugh, Hamida Skandrani, Edward A. Fox (2019). Uses and Gratifications of Political Information: Student Perceptions of Information from the 2014 Tunisian Elections. *Proc. Deep Transformations and the Future of Organisations, 4th International Conference of the Interdisciplinary Laboratory of University-Business Management, University of Manouba, Tunisia (December 6-7, 2019)*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Tian Shi, Vineeth Rakesh, Suhang Wang, and Chandan K. Reddy (2019). Document-Level Multi-Aspect Sentiment Classification for Online Reviews of Medical Experts. *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM), Beijing, China, November 2019*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: http://dmkd.cs.vt.edu/papers/CIKM19a.pdf

Ting Hua, Chang-Tien Lu, Jaegul Choo, and Chandan K. Reddy (2020). Probabilistic Topic Modeling for Comparative Analysis of Document Collections. *ACM Transactions on Knowledge Discovery and Data Mining (TKDD), pp. 24:1-24:27, March 2020*. 14 (2), 24:1. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: http://dmkd.cs.vt.edu/papers/TKDD20.pdf

Xuan Zhang, Zhilei Qiao, Aman Ahuja, Weiguo Fan, Edward Fox, and Chandan K. Reddy (2019). Discovering Product Defects and Solutions from Online User Generated Contents. *Proceedings of The Web Conference (WWW), San Francisco, CA, May 2019*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: http://dmkd.cs.vt.edu/papers/WWW19a.pdf

Yaser Keneshloo, Naren Ramakrishnan, and Chandan K. Reddy (2019). Deep Transfer Reinforcement Learning for Text Summarization. *Proceedings of SIAM International Conference on Data Mining (SDM), Calgary, Canada, May 2019*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: http://dmkd.cs.vt.edu/papers/SDM19.pdf

Yaser Keneshloo, Tian Shi, Naren Ramakrishnan, and Chandan K. Reddy (2019). Deep Reinforcement Learning For Sequence-to-Sequence Models. *IEEE Transactions on Neural Networks and Learning Systems (TNNLS), https://arxiv.org/pdf/1805.09461.pdf*. 1. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1109/TNNLS.2019.2929141

Licenses

Other Conference Presentations / Papers

Ziqian Song, Austin Spencer, Taylor Thackaberry, Kayley Bogemann, Shane Burchard, Jessie Butler, Liuqing Li, Kris Wernstedt, Pamela Murray-Tuite, Edward A. Fox (2020). *A comparison of people's use of Twitter in Puerto Rico Earthquake and Hurricane Maria*. Poster, with abstract in Proceedings International Conference on Information Systems for Crisis Response and Management (ISCRAM 2020), May 2020, to be published in 2020 and presented in May 2021 at ISCRAM 2021. Blacksburg, VA. Status = AWAITING_PUBLICATION; Acknowledgement of Federal Support = Yes

Jason Callahan (2019). *Framing Crime: An Analysis of News Media Twitter Data*. Southern Criminal Justice Association Annual Meeting, September 11. Nashville, TN. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Jason Callahan (2019). *Framing Crime: An Analysis of News Media Twitter Data*. American Society of Criminology Annual Meeting, November 13. San Francisco, CA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Edward A. Fox (2020). *How should one explore the digital library of the future? Invited keynote*. ACM/IEEE Joint Conference on Digital Libraries, JCDL 2020, Aug. 1-5. virtual conf. hosted by Wuhan U., Wuhan, PRC. Status = ACCEPTED; Acknowledgement of Federal Support = Yes

Prashant Chandrasekar, Kris Wernstedt, Edward A. Fox, Pamela Murray-Tuite (2020). *Hurricane Irma: Multiple Avenues of Study*. Poster, with abstract in Proceedings International Conference on Information Systems for Crisis Response and Management (ISCRAM 2020), May 2020, to be published in 2020 and presented in May 2021 at ISCRAM 2021. Blacksburg, VA. Status = AWAITING_PUBLICATION; Acknowledgement of Federal Support = Yes

Prashant Chandrasekar, Edward A. Fox (2019). *Interactive Digital Library Support for Workflows: Applying Machine Learning in Network Science (poster)*. Algorithms That Make You Think, Fourth Annual Virginia Tech Workshop on the Future of Human-Computer Interaction, April 11-12, 2019. Blacksburg, VA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Edward A. Fox (2019). *Introduction to Digital Libraries, https://doi.org/10.1109/JCDL.2019.00111*. ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2019). UIUC, Urbana-Champaign, Illinois, USA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Tian Shi, Ping Wang, and Chandan K. Reddy (2019). *LeafNATS: An Open-Source Toolkit and Live Demo System for Neural Abstractive Text Summarization*. Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), June 2019 (demo), https://www.aclweb.org/anthology/N19-4012. Minneapolis, MN. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Ziqian Song, Wenqi Shen, Weiguo (Patrick) Fan, Edward A. Fox (2019). *Measuring the Impact of Corporate Crisis News Propagation via Twitter*. Proc. Informs 11th Conference on Information Systems and Technology (CIST 2019), Oct. 19-20. Seattle, WA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

William A. Ingram and Edward A. Fox (2019). *Preparing ETD code and data for reproducible publication: a hands-on workshop*. ETD 2019: Fruits of knowledge, 22nd International Symposium on Electronic Theses and Dissertations. Porto, Portugal. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Liuqing Li, Rishabh Anand, and Edward A. Fox (2019). *Users, User Roles, and Topics in School Shooting Collections of Tweets*. WADL 2019, Web Archiving and Digital Libraries Workshop at JCDL 2019, June 6, 2019, http://fox.cs.vt.edu/talks/2019/20190606LiAnandFoxWADL2019.pdf. UIUC, Urbana-Champaign, Illinois, USA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Other Products

Software or Netware.

Liuqing Li, Ziqian Song, Xuan Zhang, and Edward A. Fox. TwiRole: A Hybrid Model for Role-related User Classification on Twitter. Code Ocean reproducible software capsule, Nov. 2019, https://codeocean.com/capsule/9584745/tree/v4

Software or Netware.

Mohamed Magdy Gharib Farag. Event Focused Crawler. Code Ocean reproducible software capsule, Oct. 2019, https://codeocean.com/capsule/8475497/tree/v2

Other Publications

Viet Doan, Matt Crawford, Aki Nicholakos, Robert Rizzo, Jackson Salopek (2019). *Tourism Destination Websites*. CS4624 team term project archive http://hdl.handle.net/10919/92622. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Xinyue Wang, Naman Ahuja, Nathaniel Llorens, Ritesh Bansal, and Siddharth Dhar (2019). *Toward an Intelligent Crawling Scheduler for Archiving News Websites Using Reinforcement Learning*. CS6604 team term project archival submission, http://hdl.handle.net/10919/96482. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Ola Karajeh, Chidubem Arachie, Edward Powell, Eslam Hussein (2019). *Tweet Analysis and Classification: Diabetes and Heartbleed Internet Virus as Use Cases*. CS6604 team term project archive http://hdl.handle.net/10919/96396. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Gregory Pickett, Kenneth Worden, and Adam Wilborn (2019). *Twitter Role Classification*. CS4624 team project submission, http://hdl.handle.net/10919/93954. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Patents

Technologies or Techniques

Thesis/Dissertations

Supritha Basavaraj Patil. *Analysis of Moving Events Using Tweets, http://hdl.handle.net/10919/90884*. (2019). Virginia Tech. Acknowledgement of Federal Support = Yes

Prashant Chandrasekar. *Continuously Extensible Information Systems: Extending the 5S Framework by Integrating UX and Workflows. Doctoral dissertation to appear in 2020 in https://vtechworks.lib.vt.edu*. (2020). Virginia Tech. Acknowledgement of Federal Support = Yes

Abigail Bartolome. *Describing Trail Cultures through Studying Trail Stakeholders and Analyzing their Tweets, http://hdl.handle.net/10919/84528*. (2018). Virginia Tech. Acknowledgement of Federal Support = Yes

Liuqing Li. *Event-related Collections Understanding and Services, dissertation at http://hdl.handle.net/10919/97365*. (2020). Virginia Tech. Acknowledgement of Federal Support = Yes

Jason Callahan. *Framing Crime: An Analysis of News Media Twitter Data, dissertation to appear in https://vtechworks.lib.vt.edu*. (2020). Virginia Tech. Acknowledgement of Federal Support = Yes

Ziqian Song. *The Impact of Operational Crisis on Firm Equity Value: An Event-driven Approach. Doctoral dissertation to appear in 2020 in https://vtechworks.lib.vt.edu*. (2020). Virginia Tech. Acknowledgement of Federal Support = Yes

Websites

Events Archiving
http://eventsarchive.org/

Homepage for GETAR, as well as related prior NSF-funded projects including CTRnet and IDEAL.

Participants/Organizations

What individuals have worked on the project?

Name	Most Senior Project Role	Nearest Person Month Worked
Fox, Edward	PD/PI	1
Kavanaugh, Andrea	Co PD/PI	1

Name	Most Senior Project Role	Nearest Person Month Worked
Reddy, Chandan	Co PD/PI	1
Shoemaker, Donald	Co PD/PI	1
Bailey, Jefferson	Co-Investigator	1
Agozino, Onwubiko	Faculty	0
Angermeier, Paul	Faculty	0
Coleman, Shane	Faculty	0
Deligiannis, Nikolaos	Faculty	0
Elmongui, Hicham	Faculty	0
Farag, Mohamed	Faculty	0
Horning, Mike	Faculty	0
Jelesko, John	Faculty	0
Kanan, Tarek	Faculty	0
Krometis, Leigh	Faculty	0
Lee, Sunshin	Faculty	0
Murray-Tuite, Pamela	Faculty	0
Nesbitt, Sterling	Faculty	0
Niu, Shuo	Faculty	0
North, Chris	Faculty	0
Pereira, Denilson	Faculty	0
Salehi-Isfahani, Djavad	Faculty	0
Sandoval-Almazan, Rodrigo	Faculty	0
Sheetz, Steven	Faculty	0
Skandrani, Hamida	Faculty	0
Smith, Eric	Faculty	0

Name	Most Senior Project Role	Nearest Person Month Worked
Tedesco, John	Faculty	0
Wernstedt, Kris	Faculty	0
Wimberley, Dale	Faculty	0
Xie, Zhiwu	Faculty	0
Yang, Seungwon	Faculty	0
Zach, Florian	Faculty	0
Moneim, Riham	Other Professional	0
Holzmann, Helge	Staff Scientist (doctoral level)	0
Klein, Martin	Staff Scientist (doctoral level)	0
Ma, Yufeng	Staff Scientist (doctoral level)	0
Mather, Paul	Staff Scientist (doctoral level)	0
Sforza, Peter	Staff Scientist (doctoral level)	0
Zhang, Xuan	Staff Scientist (doctoral level)	0
Ahuja, Aman	Graduate Student (research assistant)	0
Alazmi, Huda	Graduate Student (research assistant)	0
Callahan, Jason	Graduate Student (research assistant)	1
Chakravarty, Saurabh	Graduate Student (research assistant)	0
Chandrasekar, Prashant	Graduate Student (research assistant)	2
Do, Tien	Graduate Student (research assistant)	0
Li, Liuqing	Graduate Student (research assistant)	5
Malpani, Ashish	Graduate Student (research assistant)	0
Patil, Supritha	Graduate Student (research assistant)	0
Song, Ziqian	Graduate Student (research assistant)	3
Wang, Xinyue	Graduate Student (research assistant)	1

Name	Most Senior Project Role	Nearest Person Month Worked
Conte, Philip	Undergraduate Student	1
Hsiao, Bethany	Undergraduate Student	0
Karajeh, Ola	Other	1

Full details of individuals who have worked on the project:

Edward A Fox
Email: fox@vt.edu
Most Senior Project Role: PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: PI/PD, coordinating all aspects of the work, and supervising GRA efforts

Funding Support: This project

International Collaboration: No
International Travel: Yes, Portugal - 0 years, 0 months, 6 days

Andrea L Kavanaugh
Email: kavan@vt.edu
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: Guiding multiple GRA efforts, researching and publishing regarding digital government and social/community behavior involving analysis and visualization based on tweets and surveys

Funding Support: This project

International Collaboration: Yes, Egypt, Tunisia
International Travel: Yes, United Arab Emirates - 0 years, 0 months, 4 days

Chandan K Reddy
Email: reddy@cs.vt.edu
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: Supervising research related to machine learning, deep learning, and analysis of tweets and other data

Funding Support: This project

International Collaboration: No
International Travel: No

Donald J Shoemaker
Email: shoemake@vt.edu
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: Supervising doctoral dissertation and work based in sociology

Funding Support: This project

International Collaboration: No
International Travel: No

Jefferson Bailey
Email: jefferson@archive.org
Most Senior Project Role: Co-Investigator
Nearest Person Month Worked: 1

Contribution to the Project: Internet Archive is a collaborative partner, also receiving funds on this project from NSF, through IIS-1619371. We use their equipment and services and data, and collaborate on research.

Funding Support: This project, i.e., IIS-1619371

International Collaboration: No
International Travel: No

Onwubiko Agozino
Email: agozino@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Faculty collaborating on collecting, curating, and analyzing data.

Funding Support: Local support

International Collaboration: No
International Travel: No

Paul Angermeier
Email: biota@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Faculty collaborating on collecting, curating, and analyzing data.

Funding Support: Local support

International Collaboration: No
International Travel: No

Shane Coleman
Email: shanec4@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Faculty collaborating on collecting, curating, and analyzing data.

Funding Support: Local support

International Collaboration: No
International Travel: No

Nikolaos Deligiannis

Email: ndeligia@etrovub.be
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Collaborating regarding geolocating tweets.

Funding Support: His local funding. He is an assistant professor at the Electronics and Informatics department at Vrije Universiteit Brussel (VUB) and principal investigator in Data Science at the imec institute in Belgium.

International Collaboration: Yes, Belgium
International Travel: No

Hicham Galal Elmongui
Email: elmongui@alexu.edu.eg
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Collaboration regarding data re Egypt

Funding Support: Local support

International Collaboration: Yes, Egypt
International Travel: No

Mohamed Farag
Email: mohamedmagdy@gmail.com
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Collaborating regarding the software he developed as part of his earlier doctoral work supported by a prior NSF project, IDEAL, related to this research, as well as this project. Was in USA during part of the year; earlier was teaching in Egypt.

Funding Support: Local support

International Collaboration: Yes, Egypt
International Travel: No

Mike Horning
Email: mhorning@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Faculty collaborating on collecting, curating, and analyzing data.

Funding Support: Local support

International Collaboration: No
International Travel: No

John Jelesko
Email: jelesko@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Faculty collaborating on collecting, curating, and analyzing data.

Funding Support: Local support

International Collaboration: No
International Travel: No

Tarek Kanan
Email: tarek.kanan@gmail.com
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Collaborate regarding NLP and handling of Arabic texts, extending his doctoral work completed earlier at VT

Funding Support: Local support

International Collaboration: Yes, Jordan
International Travel: No

Leigh Anne Krometis
Email: lehenry@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Faculty collaborating on collecting, curating, and analyzing data.

Funding Support: Local support

International Collaboration: No
International Travel: No

Sunshin Lee
Email: sslee777@gmail.com
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Worked as GRA on this project, then as postdoc, then as faculty at Radford U. and now at U. Illinois at Springfield, collaborating to extend his doctoral research.

Funding Support: Local support

International Collaboration: No
International Travel: No

Pamela Murray-Tuite
Email: pmmurra@clemson.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Formerly a VT faculty member, now on faculty at Clemson, she is PI on another NSF project in which PI Fox serves at co-PI, and is helping with curation and analysis of data related to disasters that effect both transportation and power systems.

Funding Support: Local support

International Collaboration: No
International Travel: No

Sterling Nesbitt
Email: sjn2104@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Faculty collaborating on collecting, curating, and analyzing data.

Funding Support: Local support

International Collaboration: No
International Travel: No

Shuo Niu
Email: ShNiu@clarku.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Graduate student collaborating on project research, now a faculty member at Clark U.

Funding Support: Local support

International Collaboration: No
International Travel: No

Chris North
Email: north@cs.vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Faculty collaborating on software, collecting, curating, and analyzing data.

Funding Support: Local support

International Collaboration: No
International Travel: No

Denilson Pereira
Email: denilsonpereira@dcc.ufla.br
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Collaboration on publications and related research connected with analysis of tweets, classification, disambiguation, and text analysis

Funding Support: Local support

International Collaboration: Yes, Brazil
International Travel: No

Djavad Salehi-Isfahani
Email: salehi@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Faculty collaborating on collecting, curating, and analyzing data.

Funding Support: Local support

International Collaboration: No
International Travel: No

Rodrigo Sandoval-Almazan
Email: rsandovuaem@gmail.com
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Collaborating regarding publication, analysis, and collection/curation of data related to events in Mexico.

Funding Support: Local support

International Collaboration: Yes, Mexico
International Travel: No

Steven D. Sheetz
Email: sheetz@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Collaborator on IDEAL, prior related project, as co-PI. Continuing collaboration.

Funding Support: Local support

International Collaboration: No
International Travel: Yes, Tunisia - 0 years, 0 months, 4 days

---

Hamida Skandrani
Email: hamida.skandrani@gmail.com
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Collaborated on publications and studies related to events in Tunisia and the region.

Funding Support: Local support

International Collaboration: Yes, Tunisia
International Travel: No

Eric Smith
Email: epsmith@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Faculty collaborating on collecting, curating, and analyzing data.

Funding Support: Local support

International Collaboration: No
International Travel: No

John Tedesco

Email: tedesco@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Faculty collaborating on collecting, curating, and analyzing data.

Funding Support: Local support

International Collaboration: No
International Travel: No

Kris Wernstedt
Email: krisw@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: He is co-PI on another NSF project in which PI Fox also serves at co-PI, and is helping with curation and analysis of data as well as surveys related to disasters that effect both transportation and power systems.

Funding Support: Related NSF project

International Collaboration: No
International Travel: No

Dale Wimberley
Email: wimberly@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Faculty collaborating on collecting, curating, and analyzing data.

Funding Support: Local support

International Collaboration: No
International Travel: No

Zhiwu Xie
Email: zhiwuxie@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Library researcher collaborating on web archiving; co-organizer of Web Archiving and Digital Libraries (WADL) workshop at JCDL.

Funding Support: Local support

International Collaboration: No
International Travel: No

Seungwon Yang
Email: seungwonyang@lsu.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Collaborate on collecting and analyzing tweets related to events, especially related to the Gulf region. Led effort for MOU between GETAR and his group at LSU.

Funding Support: Local support

International Collaboration: No
International Travel: No

Florian Zach
Email: florian@vt.edu
Most Senior Project Role: Faculty
Nearest Person Month Worked: 0

Contribution to the Project: Collaborating on research related to tourism

Funding Support: Local support

International Collaboration: No
International Travel: Yes, Cyprus - 0 years, 0 months, 4 days; - 0 years, 0 months, 0 days; - 0 years, 0 months, 0 days

---

Riham Abdel Moneim
Email: riham@aucegypt.edu
Most Senior Project Role: Other Professional
Nearest Person Month Worked: 0

Contribution to the Project: Continuing collaboration related to collection, curation, and analysis of data re Egypt, and on technologies like crawling.

Funding Support: Works at Microsoft in Egypt

International Collaboration: Yes, Egypt
International Travel: No

Helge Holzmann
Email: mail@helgeholzmann.de
Most Senior Project Role: Staff Scientist (doctoral level)
Nearest Person Month Worked: 0

Contribution to the Project: Collaborating regarding ArchiveSpark and related web archiving technologies

Funding Support: L3S and then Internet Archive

International Collaboration: Yes, Germany
International Travel: No

Martin Klein
Email: martinklein0815@gmail.com
Most Senior Project Role: Staff Scientist (doctoral level)
Nearest Person Month Worked: 0

Contribution to the Project: Collaborate on collecting data (e.g., tweets) and undertaking related analysis; co-organizer of Web Archiving and Digital Libraries (WADL) workshop at JCDL.

Funding Support: Local support

International Collaboration: No
International Travel: No

Yufeng Ma
Email: yufeng.ma@verizonmedia.com
Most Senior Project Role: Staff Scientist (doctoral level)
Nearest Person Month Worked: 0

Contribution to the Project: Collaborating on deep learning research, and on data related to tourism. After Ph.D. at VT, at Verizon Media

Funding Support: Local support

International Collaboration: No
International Travel: No

Paul Mather
Email: pmather@vt.edu
Most Senior Project Role: Staff Scientist (doctoral level)
Nearest Person Month Worked: 0

Contribution to the Project: Collaborate on managing our equipment and software, connecting with Library efforts.

Funding Support: Local support

International Collaboration: No
International Travel: No

Peter Sforza
Email: psforza@vt.edu
Most Senior Project Role: Staff Scientist (doctoral level)
Nearest Person Month Worked: 0

Contribution to the Project: Director of local center on GIS collaborating on spatial location.

Funding Support: Local support

International Collaboration: No
International Travel: No

Xuan Zhang
Email: xuancs@vt.edu
Most Senior Project Role: Staff Scientist (doctoral level)
Nearest Person Month Worked: 0

Contribution to the Project: Graduate student collaborating on project research, then as scientist at Walmart / Sams Club Research

Funding Support: Local support

International Collaboration: No
International Travel: No

Aman Ahuja
Email: aahuja@vt.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 0

Contribution to the Project: Graduate student collaborating on project research.

Funding Support: Local support

International Collaboration: No
International Travel: Yes, Macau - 0 years, 0 months, 4 days

Huda Alazmi
Email: ahuda1@vt.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 0

Contribution to the Project: Curating and analyzing datasets and collaborating toward publication

Funding Support: Local support

International Collaboration: Yes, Kuwait
International Travel: No

Jason Callahan
Email: jcallaha@vt.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 1

Contribution to the Project: Graduate student collaborating on project research.

Funding Support: Local support

International Collaboration: No
International Travel: No

Saurabh Chakravarty
Email: saurabc@vt.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 0

Contribution to the Project: Graduate student completed thesis in support of this project; provided follow-on assistance.

Funding Support: Local support

International Collaboration: No
International Travel: No

Prashant Chandrasekar
Email: peecee@vt.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 2

Contribution to the Project: GRA on project part of the year; GRA on a related project collaborating with this project

Funding Support: This project, a related project

International Collaboration: No
International Travel: No

Tien Do
Email: thdo@etrovub.be

Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 0

Contribution to the Project: Graduate student collaborating about geo-location of tweets

Funding Support: At his university

International Collaboration: Yes, Belgium
International Travel: No

Liuqing Li
Email: liuqing@vt.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 5

Contribution to the Project: GRA supported by this project, working on all aspects.

Funding Support: This project

International Collaboration: No
International Travel: Yes, Spain - 0 years, 0 months, 4 days

Ashish Malpani
Email: ashish76@vt.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 0

Contribution to the Project: Collaborating on software development and application

Funding Support: Local support

International Collaboration: No
International Travel: No

Supritha Patil
Email: patil93@vt.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 0

Contribution to the Project: Graduate student collaborating on project research.

Funding Support: Local support

International Collaboration: No
International Travel: No

Ziqian Song
Email: ziqian@vt.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 3

Contribution to the Project: Graduate student collaborating on project research, funded in part by this project.

Funding Support: Local support, this project

International Collaboration: No
International Travel: No

Xinyue Wang
Email: xw0078@vt.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 1

Contribution to the Project: Graduate student collaborating on project research.

Funding Support: Local support

International Collaboration: No
International Travel: No

Philip Conte
Email: pconte@vt.edu
Most Senior Project Role: Undergraduate Student
Nearest Person Month Worked: 1

Contribution to the Project: Starting as volunteer, leading up to doing undergraduate research in 2019 to help with project.

Funding Support: Self

International Collaboration: No
International Travel: No

Bethany Hsiao
Email: bhsiaoburg@gmail.com
Most Senior Project Role: Undergraduate Student
Nearest Person Month Worked: 0

Contribution to the Project: Volunteer helping with project, then student at U. Penn

Funding Support: Self

International Collaboration: No
International Travel: No

Ola Karajeh
Email: okarajeh@vt.edu
Most Senior Project Role: Other
Nearest Person Month Worked: 1

Contribution to the Project: Analyzing data sets, curating and classifying, such as related to heart attacks.

Funding Support: Self (as doctoral student in CS@VT)

International Collaboration: No
International Travel: No

What other organizations have been involved as partners?

Name	Type of Partner Organization	Location
Al Zaytonah University of Jordan	Academic Institution	Jordan
Arab Academy for Science and Technology	Academic Institution	Alexandria, Egypt
University of Illinois at Springfield	Academic Institution	Springfield, Illinois
University of Tunis - Manouba Campus	Academic Institution	Tunisia
Vrije Universiteit Brussel (VUB)	Academic Institution	Brussels, Belgium
Clemson University	Academic Institution	Clemson, SC
George Washington University	Academic Institution	Washington, D.C.
Internet Archive	Other Nonprofits	San Francisco, CA
Los Alamos National Laboratory	State or Local Government	Los Alamos, New Mexico
Louisiana State University	Academic Institution	Baton Rouge, LA
Radford University	Academic Institution	Radford, VA
Universidad Autónoma del Estado de México (UAEM)	Academic Institution	Mexico
Universidade Federal de Lavras (UFLA)	Academic Institution	Lavras, MG, Brasil

Full details of organizations that have been involved as partners:

Al Zaytonah University of Jordan

Organization Type: Academic Institution
Organization Location: Jordan

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: Tarek Kanan continues to collaborate regarding NLP and handling of Arabic texts, extending his doctoral work completed earlier at VT.

Arab Academy for Science and Technology

Organization Type: Academic Institution
Organization Location: Alexandria, Egypt

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: Mohamed Farag, on the faculty, is collaborating regarding the software he developed as part of his earlier doctoral work supported by a prior NSF project, IDEAL, related to this research, as well as this project.

Clemson University

Organization Type: Academic Institution
Organization Location: Clemson, SC

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: Dr. Pamela Murray-Tuite, formerly a VT faculty member, now on faculty at Clemson, is PI on another NSF project in which PI Fox serves at co-PI, and is helping with curation and analysis of data related to disasters that effect both transportation and power systems.

George Washington University

Organization Type: Academic Institution
Organization Location: Washington, D.C.

Partner's Contribution to the Project:
In-Kind Support

More Detail on Partner and Contribution: We use the Social Feed Manager software from GWU Libraries, which they continue to support and enhance through collaboration.

Internet Archive

Organization Type: Other Nonprofits
Organization Location: San Francisco, CA

Partner's Contribution to the Project:
Facilities
Collaborative Research

More Detail on Partner and Contribution: Internet Archive is a collaborative partner, also receiving funds on this project from NSF, through IIS-1619371. We use their equipment and services and data, and collaborate on research. Jefferson Bailey is co-PI on GETAR.

Los Alamos National Laboratory

Organization Type: State or Local Government
Organization Location: Los Alamos, New Mexico

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: We collaborate on collecting data (e.g., tweets) and undertaking related analysis. This work is led by Dr. Martin Klein.

Louisiana State University

Organization Type: Academic Institution
Organization Location: Baton Rouge, LA

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: We collaborate on collecting and analyzing tweets related to events, especially related to the Gulf region. This is led by Dr. Seungwon Yang, whose doctoral work was supported in part by prior NSF-funded related projects at VT. See for example http://www.rsgis.envs.lsu.edu/icar/datasets/

Radford University

Organization Type: Academic Institution
Organization Location: Radford, VA

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: Dr. Sunshin Lee, whose Ph.D. work was supported by this project before he started as a faculty member at Radford, continues to collaborate with our research. He served also as volunteer postdoc aiding GETAR before going to Radford.

Universidad Autónoma del Estado de México (UAEM)

Organization Type: Academic Institution
Organization Location: Mexico

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: Dr. Sandoval Almazan is collaborating regarding publication, analysis, and collection/curation of data related to events in Mexico.

Universidade Federal de Lavras (UFLA)

Organization Type: Academic Institution
Organization Location: Lavras, MG, Brasil

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: Dr. Pereira is continuing collaboration on publications and related research connected with analysis of tweets, classification, disambiguation, and text analysis.

University of Illinois at Springfield

Organization Type: Academic Institution
Organization Location: Springfield, Illinois

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: Dr. Sunshin Lee, whose Ph.D. work was supported by this project before he started as a faculty member at Radford, continues to collaborate with our research. He served also as volunteer postdoc aiding GETAR before going to Radford. He moved to Illinois in 2019.

University of Tunis - Manouba Campus

Organization Type: Academic Institution
Organization Location: Tunisia

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: Hamida Skandrani has collaborated on publications and studies related to events in Tunisia and the region.

Vrije Universiteit Brussel (VUB)

Organization Type: Academic Institution
Organization Location: Brussels, Belgium

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: Nikos Deligiannis is an assistant professor at the Electronics and Informatics department at Vrije Universiteit Brussel (VUB) and principal investigator in Data Science at the imec institute in Belgium, as well as the director of the Master in Applied Computer Science at Vrije Universiteit Brussel. He and his students are collaborating regarding tweet data analysis and geolocation.

What other collaborators or contacts have been involved?
Nothing to report

Impacts

What is the impact on the development of the principal discipline(s) of the project?

The integration of processing of tweets and webpages, all related to important events, in one system with linked workflows, should broaden the scope of studies that largely just use only one of these two sources for digital library and Web archiving research. Our new methodology for integrating UX and system development efforts, using knowledge graphs and workflows, should improve the construction of digital libraries. The broad range of digital library methods and services adapted to aid stakeholders interested in events and trends, should provide better support for such users, including those working across the diversity of content types explored.

What is the impact on other disciplines?

Tweet and webpage collections are of interest to many disciplines studying recent history and current events, including history, sociology, political science, economics, environmental science, linguistics, communications, government, etc. As a result of this project, scores of Virginia Tech scholars, from a variety of departments as well as University Libraries, have expressed interest in our methods and activities, and a number have worked with us on focused studies. We have collected information and shared that with them, as well as helped with related analysis.

This shows how broadly the impact is likely to spread to a number of other disciplines.

What is the impact on the development of human resources?

Two doctoral students working on the project, Liuqing Li and Jason Callahan, completed their dissertations early in 2020. Two others who were GRAs, Ziqian Song and Prashant Chandrasekar, will complete their dissertations later in 2020. Other dissertations with overlap with the project were completed by Yufeng Ma and Yaser Keneshloo.

Master's students Ashish Malpani, Pranavi Rambhakta, and Supritha Patil volunteered and completed related research studies.

Two team term projects collaborating with GETAR were carried out in each of CS6604 (Digital Libraries) and CS4624 (Multimedia, Hypertext, and Information Access).

In all of these education-related efforts, students learned about digital libraries, archives, tweets, webpages, and a wide range of scientific methods, concepts, and technologies -- preparing them for future R&D, building upon computer science.

The project has involved more than 50 collaborators and 12 collaborating institutions. People in diverse fields have been exposed to advanced data analytics and visualization, enhancing their appreciation of science and understanding about working with data.

What is the impact on physical resources that form infrastructure?

Virginia Tech Advanced Research Computing has provided essentially unrestricted access to its Cascades and Huckleberry clusters with GPUs, expanding the use of deep learning methods. The State Council of Higher Education for Virginia (SCHEV) provided some equipment in 2019 that will provide further help.

What is the impact on institutional resources that form infrastructure?

In addition to stimulating support from the Deparment of Computer Science for our infrastructure, such as providing a number of VMs, now also including some with GPUs, University Libraries has built a very similar infrastructure, and the campus IT groups have expanded the size and access to clusters to support other similar types of investigations. University Libraries has agreed to taking over handling and support for data/content from our project, and to providing a range of services through technology transfer with our project team.

What is the impact on information resources that form infrastructure?

Aided in part by the Web Archiving and Digital Libraries workshops, and other dissemination of project activities and accomplishments, other teams involved in Web archiving have engaged in related studies and efforts to devise software and methods, as well as build collections. There is a growing movement for collecting and archiving tweets and/or webpages, and to broaden the support for working with those archives. The enormous collection of over 400 billion webpages at the Internet Archive, as well as other archives, has stimulated broad interest in these information resources. Our methods to add value through analysis, and to support event-oriented studies and access, shows promise to expand the utility of the expanding information resources. An IMLS-funded effort was launched that ran a 2019 workshop that helped disseminate GETAR results to aid librarians and archivists across the nation, and more will be done in the coming years to extend that effort, including through WADL 2020.

What is the impact on technology transfer?

The Internet Archive is a partner, working with the GETAR team, and has access to our technology, software, and data. Its actions broadly influence the rest of the worldwide Web archiving community. There is a plan for former project GRA Liuqing Li to visit the Internet Archive after COVID-19 travel limitations have been removed, that should enhance the effects of GETAR on Internet Archive activities.

Internet Archive staff have updated or written new documentation for stakeholders and the general public to query and use data from and about these collections and their contents through its general and collection-specific Wayback index (CDX) APIs, OpenSearch API, and “WASAPI” web archive data transfer API. Improvements to derivative dataset generation and analysis processes that the project team may use to mine and/or visualize these archives were likewise documented and formed the basis of workshops in the United States and abroad to train further librarians, archivists, and researchers to use similar resources and tools.

What is the impact on society beyond science and technology?

The collections developed can be used by any interested groups. As our software and systems mature, open access to suitable portions of our collections will be provided to the public.

Changes/Problems

Changes in approach and reason for change
Nothing to report.

Actual or Anticipated problems or delays and actions or plans to resolve them
Nothing to report.

Changes that have a significant impact on expenditures
Nothing to report.

Significant changes in use or care of human subjects
Nothing to report.

Significant changes in use or care of vertebrate animals
Nothing to report.

Significant changes in use or care of biohazards
Nothing to report.