

Varish Mulwad, Ph.D.

✉ vmulwad@gmail.com

🌐 <http://varish.net>

🌐 <http://www.linkedin.com/in/varish/>

📅 Last Updated: Jan. 2022

Summary

- ~ 14 years of (*grad. school + industry*) research experience in developing novel algorithms and production-ready solutions in the areas of **information extraction** and **knowledge graph population** from semi-structured and unstructured text.
- Authored **18 peer-reviewed publications**, **825+ citations**, and **7 issued patents**.
- Experience in working across the spectrum of technology readiness level research programs, including deployed applications with \$XX M+ of estimated impact.
- Leadership experience managing & leading 3-4 member project teams. Secured \$10M+ in US govt. funding as Principal Investigator/Co-Investigator.

Education

- May 2015 📖 **Ph.D., Computer Science, University of Maryland, Baltimore County (UMBC)**
Dissertation: *TABEL* – A domain independent and extensible framework for inferring the semantics of tables and representing them as RDF Linked Data
Committee: Drs. **Tim Finin** (advisor), Anupam Joshi, Tim Oates, Yun Peng (UMBC), Drs. L V Subramaniam, & Indrajit Bhattacharya (IBM Research, India)
- August 2010 📖 **Masters of Science, Computer Science, University of Maryland, Baltimore County (UMBC)**
Thesis: An automatic framework for extracting, interpreting and representing tables as linked data.
Committee: Drs. **Tim Finin** (advisor), Anupam Joshi, Tim Oates (UMBC), Dr. Evelyn Viegas (Microsoft Research)
- December 2007 📖 **Bachelor of Engineering in Computer Science and Engineering, University of Mumbai**
Final Year Project: The Interview Scheduler – A web based application for scheduling interviews

Employment History

- Nov. 2020 – *present* 📖 **GE Research (Niskayuna, NY)**
Senior Scientist
 - Principal Investigator leading a GE Research + UMBC team of researchers developing information extraction and knowledge graph population algorithms from tables and data charts and data discovery methods grounded in automatically computed data reliability funded as part of an IARPA seedling.
 - Led a (*\$XXX K budget*) research project developing and deploying information extraction, query generation, and information retrieval algorithms for document recommendation to assist in resolution of engineering & maintenance tickets for GE's assets. Estimated impact: \$X M of internal productivity savings and \$XX M in customer value.

Employment History (continued)

Mar. 2018 – Nov. 2020

📌 **GE Research** (*Niskayuna, NY*)

Lead Scientist

- Developed novel algorithms for extraction of scientific models (equations), concepts, and model variables' context from textual sources as part of DARPA's ASKE program. Co-led the project execution defining sprints and iterative thin-threads for the overall system.
- Led the development of information extraction algorithms for concept and relation extraction from aviation maintenance records and contracts as part of prototype systems to improve internal productivity. Additionally, mentored/guided early-career employees and interns as part of these project(s).
- Participated and contributed in building a vision for the development and use of NLP and Knowledge Graph technologies with GE Research and GE businesses leadership teams and U.S. govt. agencies program managers.
- Contributed to external research proposals as Principal Investigator/Co-Investigator resulting in more than \$10 M funding for GE Research.

Aug. 2015 – Mar. 2018

📌 **GE Research** (*Niskayuna, NY*)

Knowledge Discovery Researcher

- Led the development of [Mylo](#) – an AI agent for IT Tech agents. Mylo assists in classifying IT tickets into categories, recommends possible resolutions for new problems, and recommends experts that can assist in resolving new issues. Learns from implicit and explicit feedback. Deployed in production, Mylo has contributed to tens of millions of dollars of productivity savings across GE.
- Contributed to the development of NLP algorithms for concept extraction from medical documents and recommendation of relevant documents from patient history to radiologists; Contributed to the productionization of algorithms into GE Healthcare's medical imaging software product. Work resulted in 1 publication and 3 patents.
- Mentored summer interns.

Apr. 2015 – Jul. 2015

📌 **Ebiquity Research Lab, UMBC** (*Baltimore, MD*)

Postdoctoral Research Associate

- Mentored students (and helped) to develop techniques for information extraction and knowledge graph population from semi-structured (table and table-like sources) and unstructured (text) data from the cybersecurity domains resulting in 2 publications. Contributed to research funding proposals.






Sep. 2009 – Mar. 2015

📌 **Ebiquity Research Lab, UMBC** (*Baltimore, MD*)

Research Assistant

- Developed novel techniques for information extraction and knowledge graph population from semi-structured (table and table-like sources) and unstructured (text) data from the web, medical, and cybersecurity domains resulting in 10+ publications.
- Contributed to research proposals resulting in \$200K funding.
- Peer-mentored Masters and new Ph.D. students. Managed lab meetings and visits from industry and funding agencies.

Employment History (continued)

- May 2012 – Aug. 2012  **Microsoft Research** (*Redmond, WA*)
Research Intern
– Information extraction from apps' text descriptions in the Windows Phone Store. Mentor: Dr. Evelyne Viegas. 1 patent issued based on internship work.
- Jun. 2011 – Aug. 2011  **Microsoft Bing** (*Bellevue, WA*)
Software Development Engineer Intern
– Developed a prototype to demonstrate the use of entity disambiguation to improve search results. Two disclosures filed with Microsoft for patent consideration.
- May 2009 – Aug. 2009  **Symantec Corporation** (*Columbia, MD*)
Intern
– Acquired knowledge of Symantec's internal IPC tool and provided inputs to the design team. Developed a tool for automated testing of the SEPEX Proxy DLL. Proposed a design of a more generic automated test tool to test any given DLL.
- Jan. 2008 – Jun. 2008  **Ness Technologies India** (*Mumbai, MH, India*)
Intern (Project Trainee)
– Prepared and presented a Test–Automation prototype for a Ness client. Developed an application independent Test–Automation framework.
- Jun. 2006 – Mar. 2007  **Mastek India** (*Mumbai, MH, India*)
Final year undergraduate project
– Developed The Interview Scheduler – a web based application for the Human Resource department to schedule interviews between candidates and interviewers.

Research Publications

Conference Proceedings

- 1 Bandi, A., Joshi, K. P., & **Mulwad, V.** (2020). Affinity Propagation Initialisation Based Proximity Clustering For Labeling in Natural Language Based Big Data Systems. In *6th IEEE International Conference on Big Data Security on Cloud*.
- 2 McHugh, J., Cuddihy, P. E., Williams, J. W., Aggour, K. S., Kumar, V. S., & **Mulwad, V.** (2017). Integrated access to big data polystores through a knowledge-driven framework. In *International Conference on Big Data (Big Data)* (pp. 1494–1503). IEEE.
- 3 Mittal, S., Das, P. K., **Mulwad, V.**, Joshi, A., & Finin, T. (2016). CyberTwitter: Using Twitter to generate alerts for Cybersecurity Threats and Vulnerabilities. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 860–867). IEEE.
- 4 Nimbalkar, P., **Mulwad, V.**, Puranik, N., Joshi, A., & Finin, T. (2016). Semantic Interpretation of Structured Log Files. In *17th IEEE International Conference on Information Reuse and Integration (IRI)*. IEEE.
- 5 **Mulwad, V.**, Finin, T., & Joshi, A. (2014). Interpreting Medical Tables as Linked Data for Generating Meta–Analysis Reports. In *15th IEEE International Conference on Information Reuse and Integration*. IEEE Computer Society.

- 6 **Mulwad, V.**, Finin, T., & Joshi, A. (2013). Semantic Message Passing for Generating Linked Data from Tables. In *Proceedings of the 12th International Semantic Web Conference*. Springer.
- 7 **Mulwad, V.** (2011). DC Proposal: Graphical Models and Probabilistic Reasoning for Generating Linked Data from Tables. In *Doctoral Consortium Track, Int. Semantic Web Conf.* (pp. 317–324). Springer Berlin/Heidelberg.
- 8 Syed, Z., Finin, T., **Mulwad, V.**, & Joshi, A. (2010). Exploiting a Web of Semantic Data for Interpreting Tables. In *Proceedings of the Second Web Science Conference*.

Workshops, Posters & Demos

- 1 Dixit, S., **Mulwad, V.**, & Saxena, A. (2021). Extracting Semantics from Maintenance Records. *Workshop on Applied Semantics Extraction and Analytics (at IJCAI)*.
- 2 Cuddihy, P., McHugh, J., Williams, J. W., **Mulwad, V.**, & Aggour, K. (2018). SemTK: A Semantics Toolkit for User-friendly SPARQL Generation and Semantic Data Management. *Poster & Demo Track, International Semantic Web Conference*.
- 3 Cuddihy, P., McHugh, J., Williams, J. W., **Mulwad, V.**, & Aggour, K. S. (2017). SemTK: An Ontology-first, Open Source Semantic Toolkit for Managing and Querying Knowledge Graphs. *arXiv preprint arXiv:1710.11531*.
- 4 Tari, L., **Mulwad, V.**, & von Reden, A. (2016). Interactive Online Learning for Clinical Entity Recognition. *Proceedings of the Workshop on Human-In-the-Loop Data Analytics (at SIGMOD)*. ACM.
- 5 Yus, R., **Mulwad, V.**, Finin, T., & Mena, E. (2014). Infoboxer: Using Statistical and Semantic Knowledge to Help Create Wikipedia Infoboxes. *Poster & Demo Track, 13th International Semantic Web Conference (ISWC 2014), Riva del Garda (Italy)*.
- 6 **Mulwad, V.**, Finin, T., & Joshi, A. (2011a). Automatically Generating Government Linked Data from Tables. *Working notes of AAAI Fall Symposium on Open Government Knowledge: AI Opportunities and Challenges*.
- 7 **Mulwad, V.**, Finin, T., & Joshi, A. (2011b). Generating Linked Data by Inferring the Semantics of Tables. *Proceedings of the First International Workshop on Searching and Integrating New Web Data Sources (at VLDB)*.
- 8 **Mulwad, V.**, Li, W., Joshi, A., Finin, T., & Viswanathan, K. (2011). Extracting Information about Security Vulnerabilities from Web Text. In *Proceedings of the Web Intelligence for Information Security Workshop at IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 3, 257–260. IEEE.
- 9 **Mulwad, V.**, Finin, T., Syed, Z., & Joshi, A. (2010). T2LD: Interpreting and Representing Tables as Linked Data. *Proc. Poster and Demonstration Session at the 9th Int. Semantic Web Conf.*
- 10 **Mulwad, V.**, Finin, T., Syed, Z., & Joshi, A. (2010). Using linked data to interpret tables. *Proc. 1st Int. Workshop on Consuming Linked Data, Shanghai (at Int. Semantic Web. Conf.)*

Books and Chapters

- 1 **Mulwad, V.**, Finin, T., & Joshi, A. (2012). A Domain Independent Framework for Extracting Linked Semantic Data from Tables. In *Search Computing - Broadening Web Search*. LNCS, Springer.

Patents

Issued





- 1 Cuddihy, P. E., Williams, J. M. W., Aggour, K. S., Kumar, V. S., & **Mulwad, V.** (2022). *System and method for performing semantically-informed federated queries across a polystore*. US Patent 11,216,477.

- 2 **Mulwad, V.**, & Aggour, K. S. (2020). *Systems and methods for learning to extract relations from text via user feedback*. US Patent 10,606,953.
- 3 **Mulwad, V.**, Bueno, I., & Mutharaju, V. R. (2019). *Apparatus, system and method for providing an agent that intelligently solves information technology issues*. US Patent 10,430,517.
- 4 Von Reden, A., **Mulwad, V.**, Fluharty, E., Winograd, E., & Jing, T. (2019). *Display screen or portion thereof with graphical user interface*. US Patent D870,132.
- 5 Tari, L. B. N., & **Mulwad, V.** (2018). *System and method for entity recognition and linking*. US Patent 10,146,859.
- 6 Von Reden, A., **Mulwad, V.**, Fluharty, E., Winograd, E., & Jing, T. (2018). *Display screen or portion thereof with graphical user interface*. US Patent D821,423.
- 7 Viegas, E., **Mulwad, V.**, & Pantel, P. (2017). *Action broker*. US Patent 9,558,275.

Applications


- 1 Santamaria-Pang, A., Tu, P. H., Iyer, N. S., **Mulwad, V.**, & Zhao, G. (2021). Emergent language based data encryption. US Patent App. 16/835,683.
- 2 Crapo, A. W., Virani, N., & **Mulwad, V.** (2020). Method and system for principled approach to scientific knowledge representation, extraction, curation, and utilization. US Patent App. 16/791,617.
- 3 Sellmann, M., Wang, T., Cuddihy, P. E., & **Mulwad, V.** (2020). Decision-Support System for Aircraft Requiring Emergency Landings. US Patent App. 17/134,139.
- 4 Gustafson, S. M., Aggour, K. S., Royval, A. G., & **Mulwad, V.** (2018). Methods and systems for programmatically selecting predictive model parameters. US Patent App. 15/347,844.

Funding

- 2021  Principal Investigator: *ProCure: Enhancing Information Assimilation for COVID-19 like Events*, [IARPA COVID-19 Seedling](#); (~ \$1,000,000) (GE Research)
- 2019  Co-Investigator: *RACK: Rapid Assurance Curation Kit*, [DARPA ARCOS](#); (~ \$10,000,000); PI – Kit Siu (GE Research)
- 2018  Co-Investigator: *Augmented Bayesian Networks Integrating Semantics With Extraction and Readability*, [DARPA ASKE](#); (~ \$1,000,000); PI – Dr. Andy Crapo (GE Research)
- 2012  Significant contributions: *EAGER: T2K: From Tables to Knowledge*, National Science Foundation proposal; (\$200,000); PI – Dr. Anupam Joshi; CO-PI – Dr. Tim Finin (UMBC).

Academic Activities

Summary



-  Co-organized two workshops; reviewed papers as a Program Committee member at eight top-ranked (A*, A)¹ Artificial Intelligence, Natural Language Processing, and Semantic Web/Knowledge Graph conferences, workshops co-located at these conferences, and journals.

Organization and Leadership Roles














- 2019  Co-organizer, “Knowledge Graph Technology and Applications” workshop at the Web Conf. [A*]

¹CORE Conference Rankings in [square brackets] <http://portal.core.edu.au/conf-ranks/>






Academic Activities (continued)

- 2017  Co-organizer, “Industrial Knowledge Graphs” workshop at the ACM Web Science Conf
- 2012–2013  President, [UMBC ACM student chapter](#)









Program Committee, Conferences

- 2022  Research Track, ACM IKDD CODS and COMAD (CODS-COMAD)
- 2021, 2014  AAAI Conf. on Artificial Intelligence (AAAI) [A*]
- 2021  Research track, Extended Semantic Web Conf. (ESWC) [A]
- 2021, 2020  Research track, Int. Semantic Web Conf. (ISWC) [A]
-  Conf. on Empirical Methods in Natural Language Processing (EMNLP) [A]
- 2020  Applied Research track, ACM Int. Conf. on Information and Knowledge Management (CIKM) [A]
-  Conf. of the Asia-Pacific Chapter of the Association for Computational Linguistics and the Int. Joint Conf. on Natural Language Processing (AACL-IJCNLP) [B]
-  European Conf. on Artificial Intelligence (ECAI) [A]
- 2019  Annual Meeting of the Association for Computational Linguistics (ACL) [A*]
- 2020, 2019  International Joint Conf. on Artificial Intelligence (IJCAI) [A*]
- 2016 – 2021  Poster & Demo track, Int. Semantic Web Conf. (ISWC) [A]
- 2017 – 2019  Poster & Demo track, Extended Semantic Web Conf. (ESWC) [A]
- 2017  In-Use track, Int. Semantic Web Conf. (ISWC) [A]



Program Committee, Workshops

- 2018  Int. Workshop on Natural Language Interfaces for Web of Data (NLIWoD) @ ISWC
-  Workshop on Knowledge Base Construction, Reasoning and Mining
- 2014–2017  Int. Workshop on Linked Data for Information Extraction @ ISWC
- 2012  Mid-Atlantic Student Colloquium on Speech, Language and Learning
-  Int. Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data @ ESWC

Journal Reviews; Sub-reviews

- 2022  Communications of the ACM
- 2017  Semantic Web Journal (SWJ)
-  Special Issue on Linked Data for Information Extraction, Semantic Web Journal (SWJ)
- 2016  External reviewer, ACM Conf. for Human Factors in Computing Systems (CHI) [A*]
-  Sub reviewer, AAAI Conference on Artificial Intelligence (AAAI) [A*]
- 2013  IEEE Transactions on Knowledge and Data Engineering (TKDE))
- 2012  VLDB journal’s special issue on Structured, Social and Crowd-sourced Data (VLDBJ)
- 2011  IEEE Intelligent Systems special issue on Linked Open Government Data

Honors and Awards

- 2020  5 under 5, CTO Tech. Leadership Award, GE Research
-  Impact award (x3), GE Research

Honors and Awards (continued)

- 2019 **📌** Impact award (x4), GE Research
- 2015 **📌** Strategic Initiative Award, GE Research
- 📌** Above & Beyond Bronze Award (Impact Award), GE Research
- 2010,11,13,14 **📌** Int. Semantic Web Conf. Travel Award (via Semantic Web Science Assoc. and NSF)
- 2013 **📌** Best PhD Research Award, UMBC CSEE Annual Research Review
- 2012 **📌** First Place, Poster Presentation Competition, UMBC CSEE Annual Research Review
- 2011 **📌** Third Place, Poster Presentation Competition, UMBC CSEE Annual Research Review
- 2010,11 **📌** Outstanding Oral Presentation Award, UMBC Graduate Research Conference
- 2008 **📌** Spot Award, Test Automation Prototype, Ness Technologies
- 2007 **📌** Felicitation by leadership team for the success of The Interview Scheduler project (as college trainees), Mastek,

Mentoring

Committee

- 2020 **📌** Adithya Bandi, M.S. Thesis Co-advisor, UMBC

Summer Interns

- 2021 **📌** Hanna Tischer (co-mentor)
- 2020 **📌** Kris Pan (co-mentor)
- 2019 **📌** Sharad Dixit, Adithya Bandi
- 2017 **📌** Raka Dalal
- 2016 **📌** Sudip Mittal, Yue Liu, Emily LeBlanc (co-mentor)

Thesis Mentoring

- 2015 **📌** Piyush Nimbalkar (M.S. Computer Science, UMBC)
- 2012 **📌** Nikhil Puranik (M.S. Computer Science, UMBC)

References

Available on Request