



The TACC Workshop will be held as part of the ICSTI 2017 General Assembly and Workshops event on Thursday 26th October 2017 in Washington D.C.

The event is hosted by the Library of Congress.

TACC Workshop

ICSTI's **Technical Activities Coordinating Committee (TACC)** typically focuses on exploring and communicating technical aspects of innovative trends in information science-based tools that help make STI more useable and accessible. The **Chair of TACC is Brian Hitson**, Director US DOE/Office of Scientific and Technical Information (DOE/OSTI).

Workshop Description

Machine Learning and Its Applications to Scientific and Technical Information

Machine learning may be the next great innovation in knowledge search and discovery. Machine learning describes what happens in machines that get trained to perform a task by exposure to examples of what they're supposed to learn. It's already happening all around us in the development of facial and object recognition; self-driving cars; instant language translation; and speech recognition. This workshop will explore machine learning and its applications relevant to science and, more specifically, to various forms of scientific and technical information, including images, data, and text.

Workshop Program



Andrew Fast – CounterFlow AI

Presentation Title: *A Brief History of Machine Learning*

Abstract: Since Arthur Samuel published a computer program that learned how to play checkers in 1959, intelligent machines have grown in prominence, both in our collective imaginations and in fielded implementations. In recent years, the combination of the rise of "Big Data" and advances in machine learning have led to an emergence of technologies previously limited to science fiction including self-driving cars, intelligent assistants, and game playing 'robots'. At the core of these systems are inductive learners that can

generalize from multiple examples. This talk will introduce the types of problems that machine learning can solve and highlight popular approaches that are enabling the recent triumphs of machine learning.

Bio: Dr. Andrew Fast is the Chief Data Scientist and co-founder of CounterFlow AI, a company that builds artificial intelligence systems for breach detection and network forensics. Previously, he served as the Chief Scientist at Elder Research, Inc., a leading data science consulting firm, where he helped hundreds of companies expand their data science capabilities. In 2012, Dr. Fast co-authored a book titled Practical Text Mining that was published by Elsevier and won the PROSE Award for top book in the field of Computing and Information Sciences for that year. His work analyzing NFL coaching trees was featured on ESPN.com in 2009. Dr. Fast received Ph.D. and M.S. degrees in Computer Science from the University of Massachusetts Amherst and B.S. in Computer Science from Bethel University. He has published on data science and text analytics in both academic and industrial venues and was invited to keynote the Text Analytics World conference in 2014.



Ray Shan – Yewno

Presentation Title: *Using AI to enhance Biomedical research*

Abstract: *This presentation will explore how the use of AI technology can help deal with fragmented and dispersed sources and handle the interplay of disparate factors by using mathematical techniques to analyze the local and globally-emergent properties of the Biomedical domain.*

Bio: Ray Shan leads product design, product management and human interface engineering at Yewno. Ray has 11 years of broad experience in business, engineering and design. He has helped grow teams, customer satisfaction and revenue at startups and established corporations, including Shutterfly, Merrill Lynch and HSBC. In his free time, Ray enjoys photography and contributing to the open source software community."



Andrew McCallum – University of Massachusetts

Presentation Title: *AI for Representation and Reasoning in Knowledge Bases of Science*

Abstract: *Google, Amazon, and Bloomberg have "knowledge graphs" that include celebrities, products and companies. In many ways science could be accelerated if we had a comprehensive knowledge graph of scientific entities---genes, proteins, symptoms, diseases, drugs, chemical compounds, materials, equipment, procedures, problems---and the relations among them. In this talk I will describe advances in machine learning and deep neural networks for extracting entity-relations from the natural language full text of millions scientific papers, as well as for representing and reasoning about the resulting knowledge base. I will introduce "universal schema," our approach that embeds many database schema and natural language expressions into a common*

semantic vector space. I will also discuss our work in chains of reasoning that employ deep learning to reason about implicit knowledge and scientific "missing links." I will briefly describe our applications of this work to building scientific knowledge bases, including biomedicine (with multiple collaborators), and material science (with Elsa Olivetti at MIT). If there is time and interest, I will also discuss our efforts to revolutionize peer review with our system <http://OpenReview.net>.

Bio: Andrew McCallum is a Professor and Director of the Information Extraction and Synthesis Laboratory, as well as Director of Center for Data Science in the College of Information and Computer Science at University of Massachusetts Amherst. He has published over 250 papers in many areas of AI, including natural language processing, machine learning and reinforcement learning; his work has received over 50,000 citations. He obtained his PhD from University of Rochester in 1995 with Dana Ballard and a postdoctoral fellowship from CMU with Tom Mitchell and Sebastian Thrun. In the early 2000's he was Vice President of Research and Development at WhizBang Labs, a 170-person start-up company that used machine learning for information extraction from the Web. He is a AAAI Fellow, the recipient of the UMass Chancellor's Award for Research and Creative Activity, the UMass NSM Distinguished Research Award, the UMass Lilly Teaching Fellowship, and research awards from Google, IBM, Microsoft, and Yahoo. He was the General Chair for the International Conference on Machine Learning (ICML) 2012, and currently serves as the Past-President of the International Machine Learning Society, as well as member of the editorial board of the Journal of Machine Learning Research. For the past ten years, McCallum has been active in research on statistical machine learning applied to text, especially information extraction, entity resolution, social network analysis, structured prediction, semi-supervised learning, and deep neural networks for knowledge representation. His work on open peer review can be found at <http://openreview.net>. McCallum's web page is <http://www.cs.umass.edu/~mccallum>



Lee Angelelli – IBM

Presentation Title: *Applying Cognitive Computing to Understand and Answer Users Information Needs*

Abstract: *This presentation will discuss the challenges organizations are facing with the explosive growth of Big Data; and complexity it adds to the fundamental need to extract implied/tacit information and search organizations' knowledge management systems (KMS). The presenter will provide an overview of the cognitive computing technologies used to develop Watson that enable it to understand unstructured data, interact naturally with people, and learn and reason, at scale. The presenter will discuss applications of cognitive computing systems listed below – to help organizations collect, reason, and exploit knowledge from their KMS to meet users import information needs.*

- *Cognitive exploration - help users find and understand the information they need.*
- *Apply deep learning approach to produce short summarization of documents.*
- *Cognitive systems enabling humans to make better decision making.*

Bio: Mr. Angelelli joined the IBM Federal CTO group in 2012. Mr. Angelelli specializes in applying Big Data, advance analytics, and cognitive computing technologies to develop strategic solutions for United States Department of Defense (DoD) clients. Currently, Mr. Angelelli is developing a solution to apply Watson to cyber security. Lee Angelelli is working with US Army on developing a Human Water Stability Index (WSI) using Big Data Analytics and Social Cognitive technologies to understand the impacts of water and food security on regional socioeconomics, sociopolitical, and ecological stability.

Mr. Angelelli has a Computer Science degree from Virginia Polytechnic Institute and State University and is a trained Lean Six Sigma Black Belt. Lee is a member of the Association for the Advancement of Artificial Intelligence (AAAI).



Georgia Tourassi – Oak Ridge National Laboratory

Presentation Title: ***Deep Learning Enabled National Cancer Surveillance***

Abstract: *Pathology reports are a primary source of information for cancer registries which process high volumes of free-text reports annually. Information extraction and coding is a manual, labor-intensive process. In this talk I will discuss the latest deep learning technology, presenting both theoretical and practical perspectives that are relevant to natural language processing of clinical pathology reports. Using different deep learning architectures, I will present benchmark studies for various information extraction tasks and discuss their importance in supporting a comprehensive and scalable national cancer surveillance program.*

Bio: **Dr. Georgia Tourassi** is Group Leader of Biomedical Science, Engineering, and Computing and Director of the Health Data Sciences Institute at the Oak Ridge National Laboratory (ORNL). She holds a B.S. in Physics from the University of Thessaloniki, Greece and a Ph.D. in Biomedical Engineering from Duke University. She received the Young Investigator's Award from NIH and the Whitaker Foundation. Before joining ORNL, Dr. Tourassi was Associate Professor of Radiology and the Medical Physics Graduate Program at Duke University Medical Center, where she currently holds an Adjunct Professor position. Her research interests include biomedical informatics, scalable data-driven biomedical discovery, medical imaging, and computer-aided decision support, and clinical human-computer interaction. Her medical imaging research has been featured in several publications including *The Economist* and won an R&D 100 award in 2014. Latest research work is focused on the use of cyber-informatics for cancer related epidemiological discovery, which was selected as finalist for R&D 100 award in 2015. Her research has been funded by NIH, DOD, DOE, and the Komen and Whitaker Foundations. She has authored over 200 peer-reviewed journals, conference proceedings papers, and book chapters. She serves regularly on NIH grant review study sections as charter and ad hoc member. She also serves on the FDA advisory committee on computer-aided diagnosis devices. She is a senior member of IEEE and INNS. She is Fellow of the American Institute of Medical and Biological Engineering (AIMBE), the American Association of Physicists in Medicine (AAPM), and the International Society for Optics and Photonics (SPIE). In 2017 she received the DOE Secretary's Appreciation Award for leading ORNL in the Joint Design of Advanced Computing Solutions for Cancer, a DOE-NCI partnership in response to the Vice President's Cancer Moonshot.