# Seminar Talk

**Dr. Jian Wu**
**Assistant Professor**
**Department of Computer Science**
**Old Dominion University**

**Friday, April 05, 2019**
**3:00 p.m. KH 224**

**Title:** Mining Scholarly Big Data

**Abstract:**

In the past decade, much research has been conducted to develop models aiming at automatically extracting metadata from scholarly big data, including millions of scientific papers, such as conference proceedings and journal articles. The advent of automatic indexing techniques allows millions of articles to be indexed and searchable by digital library search engines such as Microsoft Academic and CiteSeerX. However, with the unprecedented growth rate of scientific papers, more efficient services are required to reduce the effort researchers retrieve and read large volumes of papers. To this end, natural language processing combined with artificial intelligence techniques is adopted to train new models to read, extract, and classify semantic information from free text, as well as to augment data by linking records in external data sources. This will lead to a knowledge graph powering a new digital library search engine interface featuring direct answers to users questions. CiteSeerX is an ideal platform to implement and make these features available in the near future.

**Bio:**

Dr. Jian Wu is an assistant professor in the Department of Computer Science at the Old Dominion University. He is in the Web Science Digital Library Group (WS-DL) directed by Dr. Michael Nelson. Before joining ODU, Dr. Wu was an assistant teaching professor in the College of Information Sciences and Technology at the Pennsylvania State University. Dr. Wu has been working with Dr. C. Lee Giles on the CiteSeerX project since 2013. He has been Co-PIs of 4 grants awarded by NSF or NASA. Dr. Wu's research interests include data mining, scholarly big data, and search engines. He uses Machine Learning, Information Retrieval, and Natural Language Processing techniques to mine knowledge from scholarly big data. He has published 30+ peer-reviewed papers in ACM, IEEE, and AAAI conferences and magazines, in addition to articles published early in astrophysical journals. Dr. Wu made critical improvements to the architecture, web crawling, and extraction modules of CiteSeerXSpers in ACM, IEEE, and A(,)-Bq.0.00000p4BT/ques,-7(r)