BIOGRAPHICAL SKETCH

Lenin Mookiah has worked in the Knowledge Discovery Lab at TTU since May 2014. He has around 12 years of experience, including eight in industry and four in academia. As part of these experiences, he has developed models for anomaly detection (monitoring real-time databases), traffic prediction, context mining in news network, and entity resolution in social networks. His research interests are in News Mining, Graph Mining, Anomaly Detection, and Text Mining. His thesis work focuses on mining and extracting knowledge from news networks in a personalized manner. He has been investigating problems such as ranking journalists for a topic and filtering out unimportant news articles with graph-based approaches and natural language processing techniques.

EDUCATION

Ph.D., Engineering
Tennessee Technological University
Cookeville, TN, expected 2017

M.Sc., Computer Science and Technology
Tsinghua University
Beijing, China, 2013

B.E., Electrical and Electronics Engineering
Madras University
Tamil Nadu, India, 2003

College of Engineering
TENNESSEE TECH

College of Engineering
TENNESSEE TECH

The Department of

Computer Science

Announces the Dissertation Defense

Of

Lenin Mookiah

In Partial Fulfillment of the Requirements

For the degree of

Doctor of Philosophy

Tuesday, March 21, 2016 at 11:00 a.m.

Held at

Bruner Hall, Room 126

110 University Drive

Tennessee Tech University

Cookeville, TN 38505

FIELD OF STUDY

Data Mining

DISSERTATION TOPIC

Personalized Context Mining of News Streams using Graph-Based Approach

EXAMINING COMMITTEE

Dr. William (Bill) Eberle, Committee Chairperson (CSC)

Dr. Martha Kosa (CSC)

Dr. Allan Mills (MATH)

Dr. Ambareen Siraj (CSC)

Dr. Doug Talbert (CSC)

ABSTRACT

Over the past decade, there has been a proliferation of online news articles. News articles can contain rich content and contextual information pertaining to groups in societies such as senior citizens, child rights groups, religious minority groups, or environmentalist groups. In addition, news articles contain different object types such as people, organization names, statistical (numerical) information, country names, author names, or events. Thus, it is possible to create a complex heterogeneous graph containing multi-type objects (vertices) and multi-type linkages (edges) among the objects, such as common keywords found between two news articles. We call such a graph a Heterogeneous News Graph (HNG). Currently, it is possible to extract rich information and knowledge from an HNG. It is our belief that one could use an HNG to resolve the bias and visibility issues found in many news sources, also capturing important news articles. First, due to the amount of news feeds currently available in this digital age, readers want a filtered view of relevant news articles allowing them to focus on important (breaking) news that contain rich contextual information for their particular societal group. For example, senior citizen groups might want to know new safety measures taken by police for elderly people. Second, visibility is another problem in the world of journalism, where there are multiple objects in the news articles such as author names, and organization names. In this case, for example, readers might need to know who are the relevant authors, or experts, for particular topics, such as Libya, Afghanistan, and Climate Change. Third, there is an inherent bias in the narrative of the news by the authors, whereby they try to align their opinion to a perceived general public sentiment in order to increase readership. For example, news agencies often tend to link minorities to the rise in crime, and terrorism to specific religious groups. To address the issues of determining importance, visibility of objects, and bias in news narrative, we propose novel graph-based approaches using HNGs that will (1) rank the expertness of an article's author on a specific topic, (2) identify articles of particular interest and value, and (3) classify articles as unbiased. In summary, our approach will use a novel graph-based, streaming approach for determining context and content in big data so that more personalized recommendations can be realized.