Louisiana Tech University
2016 Student Research Symposium

February 26, 2016
Oral Presentations: 8:00 – 12:00 p.m.
Poster Presentations: 8:30 – 12:00 p.m.
Scotty Robertson Memorial Gym

Student Contact Information

Name: Ramakrishna Prasad Sakhamuri
Telephone: (385) 237-7339
Email: rps015@latech.edu

Type of Presentation (check one): ☒ **oral presentation** ☐ **poster presentation**

Classification (check one): ☐ **Undergraduate** ☒ **Graduate**

College: College of Engineering and Science

Title of Presentation: An unsupervised framework for effective indexing of Big Data.

Faculty Research Advisor: Pradeep Chowriappa
Faculty Telephone: (318) 257-4612
Faculty Email: pradeep@latech.edu

Abstract (250 words maximum – <u>required</u> for oral presentations; optional for posters):

With the burgeoning realization of effective data analytics and Big Data management tools, there is a dearth of techniques that support effective contextual query processing. Existing HADOOP Distributed File System (HDFS) file index supports storage using Apache Lucene©. The Lucene file index widely supports contextual searching and indexing of text documents alone. In this context, there is a need for better file indexing based on patterns inherent to the non-text data stored. We propose an indexing and retrieval schema using clustering features. The clustering features serve as profiles of know clusters. These profiles evolve incrementally with the insertion of newer data.

The objective of this work is to develop a framework that exploits big data storage infrastructure like HDFS, while capturing inherent data patterns for effective query processing. The proposed framework employs (a) data cleaning and data transformation to ensure consistency in data formats before storage into HDFS. (b) We prose a profile based aggregation schematic to store data in a hierarchical manner for effective indexing. Furthermore, (c) the created index will then be exploited for pattern based data retrieval over large volumes of data. We showcase the effectiveness of the proposed framework using data from simulated biological experiments.

Please return completed forms by **Jan 29, 2016 (graduate oral presentations)** or **Feb 12, 2016 (undergraduate oral presentations)** to the following individuals, depending on your College affiliation. **All poster abstracts are due Feb 12, 2016.**

College of Applied & Natural Science: Bill Campbell (Box #37 or Campbell@latech.edu)
College of Business: Rob Blackstock (Box #8 or rblackst@latech.edu)
College of Education: Dawn Basinger (Box #9 or dbasing@latech.edu)
College of Engineering and Science: Ramu Ramachandran (Box #10 or ramu@latech.edu)
College of Liberal Arts: Steven Wehre (Box #7 or swehre@latech.edu)