BIOGRAPHICAL SKETCH

Rina Singh was born in Mahottari, Nepal. She is a graduate student in the Department of Computer Science at Tennessee Tech University. She is currently a Ph.D. candidate, and her research focus is on developing algorithms for event-oriented analysis in sequential pattern mining. Prior to joining the Ph.D. program, she received a Master's in Computer Science from Delft University of Technology (TU Delft), Netherlands. Her other research interests include machine learning, deep learning, data mining, and graph mining.

EDUCATION

Ph.D., Engineering

Tennessee Tech University (Expected graduation: December 2018)

M.Sc., Computer Science

Delft University of Technology, 2012

B.E., Software Engineering

Nepal College of Information Technology, 2006

College of Engineering

TENNESSEE TECH

The Department of

Computer Science

Announces the Dissertation Defense

of

Rina Singh

In Partial Fulfillment of the Requirements

For the degree of

Ph.D., Engineering

Thursday, November 8, 2018 at 12:00PM

Held in

313 Bruner Hall

Tennessee Tech University

FIELD OF STUDY

Sequential Pattern Mining

DISSERTATION TOPIC

EVENT-ORIENTED ANALYSIS AND MITIGATING THE IMPACT OF REDUNDANCY IN SEQUENTIAL PATTERN MINING

EXAMINING COMMITTEE

Dr. Douglas Talbert, Chairperson, Computer Science

Dr. William Eberle, Computer Science

Dr. Martha Kosa, Computer Science

Dr. Allan Mills, Mathematics

Dr. Rachel Hall, School of Nursing

ABSTRACT

Sequential pattern mining is a challenging problem that has received much attention. It is, however, notorious for producing large result sets. Motivated by developing specific capabilities associated with a clinical application, this dissertation presents research that addresses the problem of efficiently finding a small set of interesting and non-redundant sequential patterns that either begin or end with a specific event or pattern of events. This type of pattern mining, called prefix mining or suffix mining, has received little attention. To improve the efficiency of such mining, we present algorithms that are both provably correct and more efficient than the existing algorithm for this type of mining. Additionally, we develop an adaptation of a heuristic search for use in this type of mining and present experiments that show only a small degradation in the quality of the resulting patterns. To address concerns about overly large result sets, we adapt and apply two techniques from subgroup discovery to prefix and suffix mining. First, we enable the inclusion of interestingness metrics that allow us to perform top-k mining to focus the results on patterns that best satisfy the specified definition of "interesting.'" Second, we incorporate multiplicative-weighted coverage into top-k mining, enabling users to control the trade-off between pattern interestingness and pattern diversity, thereby, allowing for a reduction of redundancy in the result set. Using multiple real-world datasets, we demonstrate the ability of our algorithms to use both interestingness and diversity to refocus and reduce the results.