THE DEPARTMENT OF COMPUTER SCIENCE PRESENTS: **DR. CHRIS BAKER**

New Frontiers in Occupational Health Informatics

In 2012 the Conference Board of Canada reported direct costs of workplace absenteeism averaged 2.4 % of gross annual payroll resulting in an estimated $16.6 billion loss to the Canadian economy. Correlations between occupational exposures and health have served to inform Workplace safety policies and contribute to reduce absenteeism. Despite this importance of occupation health correlations many essential patient data sets have yet to be integrated with occupation information.

CanPATH https://canpath.ca/ Canadian Partnership for Tomorrow Project has gathered data on the current and longest held job titles from over 300,000 project participants linked to information about health, lifestyle, environment and behaviour and positions Canada amongst the world's leaders in longitudinal cancer and chronic disease research. The first essential step to leveraging this data is the linking of the reported job descriptions with the Canadian National Occupation Classification (NOC). e.g., Job Title: "manager of dog grooming business" codes to NOC code: 0651- Managers in customer and personal services. Currently achieved by manual look-up and coding to the NOC, which is a time-consuming, error prone and costly activity.

Motivated by the need for a scalable approach we developed the algorithms for auto-coding of job titles provided CanPATH and other local cohorts including patients with Diabetes and Pneumonia. The first algorithm we introduced (ACANOC) used iterative several search strategies to match textual input strings with NOC titles and job descriptions and a filtering steps to select candidates.

We benchmarked our coding on manually coded data sets achieving an overall coding fourth digit accuracy of 63% and an accuracy of 86% in specific categories of occupation, e.g., Truck Drivers. In the second algorithm (ENENOC) Ensemble Multi-class Classifier approach, was leveraged using both doc2vec neural network and TFIDF term frequency used to train support vector machine (SVM), random forest (RF) and K-nearest neighbor (KNN) machine learning classifiers. Initial benchmarking showed similar performance to ACANOC however subsequent A-B tests identified it has superior performance. Additionally, significant speed improvement over ACANOC and manual coding was shown. ENENOC can code 65,000 records in approximately 5 hours whereas manual coding took 2 years to do.

Furthermore, accurate prediction of NOC codes makes it possible to link coded patient data coded with other classifications namely, disease (ICD), functional impairment (ICF), occupation and job attributes (NOC Career Handbook). Based on semantic integration with these resources we have developed a job transition "Return to Work" recommendations engine for patients with specific functional impairments. Sample queries to the model include "List the NOC Codes for all jobs a Truck Driver (with Diabetes) can transition to, that have visual requirements other than [Total visual field (V-4)]". This approach can be useful to supplement career counseling regarding job transition following acquired disability or immigration.

ICD - International Classification of Diseases https://www.who.int/classifications/icd/en/
ICF - International Classification of Functioning, Disability and Health (ICF) https://www.who.int/classifications/icf/en/
NOC Career Handbook - https://www.canada.ca/en/employment-social-development/services/noc.html

Speaker Bio: Professor Chris Baker

Dr. Chris Baker is an award-winning scientist with a career spanning both industry, government, and academia. Dr Baker is inventor of 2 Patents filed with the USPTO and has published 90 papers in journals and conference proceedings. He holds a full professorship at the University of New Brunswick, Canada and is currently Chair of the Computer Science Department in Saint John. He has 20 years of expertise with Biomedical Informatics, Data Integration, and Interoperability. In 2019 his research team was awarded first prize for Outstanding Research Articles in Bio-surveillance by International Society for Disease Surveillance (ISDS).

DATE: FRIDAY, FEB. 12TH, 2021
BUILDING: ONLINE VIA TEAMS
TIME: 10:30 AM

FOR MORE INFORMATION, VISIT: unb.ca/sj-computerscience

Join Microsoft Teams Meeting

UNB Saint John