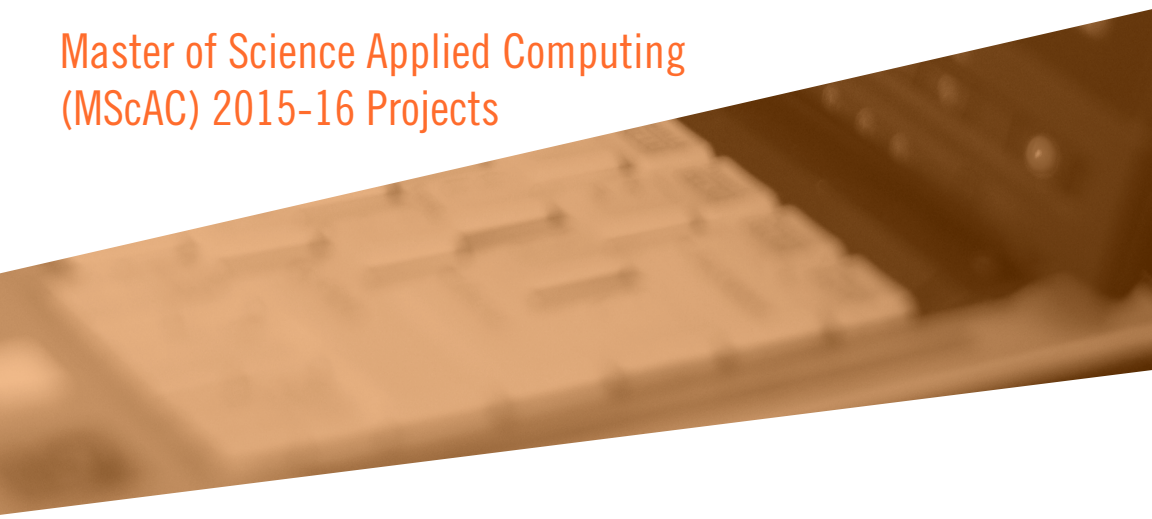
# APPLIED RESEARCH IN ACTION 2016

## Master of Science Applied Computing (MScAC) 2015-16 Projects



Computer Science
UNIVERSITY OF TORONTO

## MESSAGE FROM THE CHAIR

"Innovation" is a buzzword these days, sometimes defined simply as a new idea, product, or method.

Universities are seen as innovators of ideas, but not every new idea reaches past the walls of the academe to benefit society.

This is why our Master of Science in Applied Computing (MScAC) program is so vital. In partnership with industry, our MScAC students broaden the reach of our innovation, applying the research skills our graduate program is known for to real world problems.

This is just one avenue through which companies can collaborate with us. Whether it's supporting a joint research project, or commercializing research, together we can solve even more problems. I encourage you to contact our Associate Chair of Research, Professor **Khai Truong**, to learn more (acrir@cs.toronto.edu).

Computing is the key to solving some of industry's biggest questions, and furthermore, some of the world's biggest problems. The Applied Research in Action (ARIA) projects are an example of our boundless potential.

**Ravin Balakrishnan**
Professor & Chair

## MESSAGE FROM THE PROGRAM DIRECTOR

Congratulations to our 2015/2016 cohort on the completion of their internships! These projects represent the breadth of skills and talent our applied computing students apply to the comprehensive range of research problems in a variety of industries.

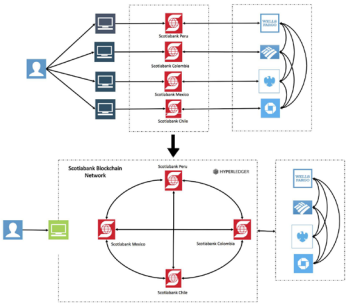
Thank you to our partners without whom these projects would not have been possible. The MScAC internship is an experience that is both educational and professionally rewarding for both the student and internship host. Thank you for being a part of our program.

Year after year, interest in our program grows and our enrolment has risen to match. If you are new to learning about our applied computing program, I welcome the opportunity to meet with you further. Our MScAC team is happy to answer prospective student questions, and companies are encouraged to reach out if they have any interest in partnering with us (mscac@cs.toronto.edu).

We hope all our guests at ARIA enjoy learning about this year's internship projects.

**Matt Medland** (MScAC 2012)
Director of Professional Programs & External Relations
Assistant Professor, Teaching Stream

## **BLOCKCHAIN RESEARCH** – Scotiabank

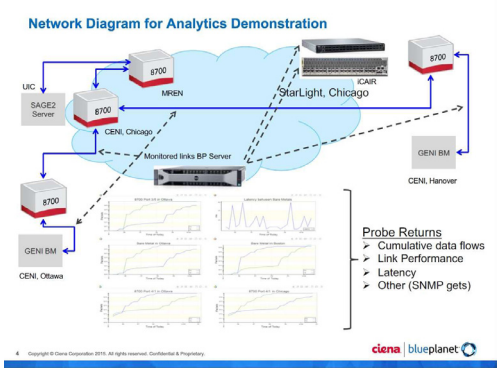


The lab's main focus is to identify and build proof of concept applications using blockchain and distributed ledger technologies. Inherently, this requires the evaluation of public and private protocols being developed in the blockchain community. This evaluation allows the team to select the correct protocol based on the financial or business use case in question. Throughout the development process, the lab tries to educate business line experts about blockchain technology and collaborate to identify potential use case and opportunities for efficiency, transparency, and mutualization.

One of the proof of concepts they helped build was Project Guava, which involved the creation of an instant cross-currency, cross-borders transfer and payments system for Scotiabank's Latin American subsidiaries. This project allows medium to small business owners, operating in Colombia, Chile, Mexico, and Peru, to easily and instantly transfer funds between accounts in different countries and in different currencies. Project Guava can potentially automate and remove friction from manual processes in the current system, provide a more secure and efficient platform for their customers, and help further the bank in its digital transformation.

**Project Team**
**Sukrit Handa & Stephanie Huynh**, MScAC students
**Henry Kesisyan**, Scotiabank
Professor **Ding Yuan**

## **BLUE PLANET NETWORK ANALYTICS RESEARCH** – Ciena




Ciena's CTO External Research Program has built an international testbed network connecting its Ottawa R&D facility to researchers in the United States and Europe. Through 2014 this Testbed was used for software defined networking (SDN), switching experiments, and demonstrations. Ciena Enviroment for Network Innovation (CENI) adds virtual cloud and compute resources to allow experiments in Network Function Virtualization (NFV) and protocols needed for distributed virtual infrastructures.

Ciena is also interested in Interns that can perform research on a compelling additional area of Investigation concerning Software Defined switching exchanges (SDX).

Ciena believes that SDN will become the standard method to provision networks, but the duration of provisioned services is becoming very temporary and fluid. The term "orchestration" is being applied more and more. As network domains (layers 1, 2, and 3) converge on network POP's. Methods to control the multi-domain, multi-layer networks

are being researched.

Ciena's Blue Planet product is expected to support this kind of function, but how this will work is an area of research on which I was working. It is on the cutting edge of next generation network architecture and we intend to use the CENI testbed as a vehicle to experiment with potential solutions to this problem.

**Project Team**
**Mohsen Norouzi**, MScAC student
**Rodney Wilson**, Ciena
Professor **Alberto Leon-Garcia**

## **CARDIOVASCULAR DISEASE IDENTIFICATION** – Analytics for Life

ANALYTICS 4 LIFE™

Analytics for Life is a medical device company focused on non-invasive cardiac monitoring. We are investigating machine learning and signal processing techniques for the purpose of assessing and diagnosing cardiac diseases, with an initial focus on Coronary Artery Disease (CAD). The objective is to develop an efficient and cost-effective solution for disease diagnosis.

The feature vectors are extracted from raw physiological signals as topological properties of the constituent phase energy and noise subspace. These transformations project the time series data into a high dimensional feature space. Genetic algorithms are used for objective maximisation search within the high dimensional feature space to generate non-linear combinations of features, which are most likely to correspond to the intrinsic mechanisms of the observed system. Other machine learning techniques are used to improve the predictive performance and generalization on the input data.

The techniques developed are being utilized in a clinical study to assess significant coronary artery disease.

**Project Team**
**Abhinav Doomra**, MScAC student
**Ian Shadforth**, Analytics for Life
Professor **Anthony Bonner**

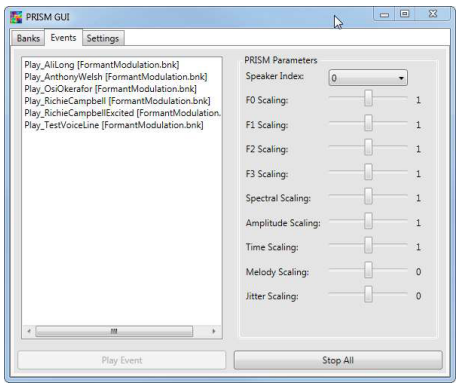
## **ENHANCED RECOMMENDATIONS USING "DOUBLE-FILTERING"**– Rakuten Kobo Inc.

Recommendations, both item-to-item ("similar items") and user-to-item ("personal recommendations"), are one of the main avenues for discovery of content at KOBO. Approximately 30% of sales are driven by recommendations. The Big Data team works to continually expand and improve upon these algorithms, which are mainly based on co-purchase patterns of items bought on Kobo's website and through Kobo eReaders and free mobile apps. One pressing weakness of the current algorithm (and one faced by most collaborative filtering-based recommendation systems) is that popular books

frequently appear in “similar titles” since they are frequently purchased with other books, even though these books may not be what we may consider “similar”. This is especially problematic for books with low purchase numbers. In this project, “shelves” in each user’s library are created via an unsupervised greedy clustering algorithm that uses book-to-book similarities for grouping similar books into shelves (and for the stopping criterion). Once each user books are grouped into shelves, we recompute book-to-book similarities using shelves instead of original user libraries - hence “double (collaborative) filtering”. In the new version, two books cannot be similar unless they were found together on at least one shelf, thus leading to more accurate and meaningful similarities. We’ve seen impressive off-line results; improvements are currently being A/B tested (results pending).

**Project Team**
**Sasa Milic**, MScAC student
Dr. **Darius Braziunas**, Rakuten Kobo Inc.
Professor **Suzanne Stevenson**

## PITCH-SHIFTING FOR REAL-TIME AND INTELLIGIBLE SPEECH MODULATION – Ubisoft




Ubisoft Toronto is a growing and diverse team on a collective mission to create what’s next in AAA games. Since opening its studio in 2010, the Ubisoft Toronto team has shipped the critically-acclaimed Tom Clancy’s Splinter Cell Blacklist as lead studio, and Assassin’s Creed Unity, Far Cry 4, Far Cry Primal and Watch Dogs 2 in collaboration with Ubisoft Montreal. The studio is currently hard at work on For Honor, as well as several top-secret, unannounced projects. Ubisoft Toronto is proud to be recognized as one of Canada’s Top 100 Employers for 2017.

PRISM is a system that modifies waveforms of recorded speech to produce multiple different sounding voices for characters within a video game. This is done by using digital signal processing to extract the base components of the signal, which are then modified to change the prosody of the waveform – attributes such as: fundamental frequency, formants, duration, intensity, melody, etc. The modified components are then used to synthesize a new waveform that differs from the original source.

For PRISM to be useful in a video game setting it must be able to generate multiple different sounding voices from one source speaker. All of these voices need to sound realistic, as video games need to have a high level of quality to avoid ruining the player’s experience and immersion. PRISM also needs to perform the voice transformation in real-time to save storage space and to be able to be synched up with the character animations.

**Project Team**
**Alex Halliwushka**, MScAC student
**Gavin Whitlock**, Ubisoft
Professor **Frank Rudzicz**

## PREDICTION OF EMPLOYEE TURNOVER WITH MACHINE LEARNING – PricewaterhouseCoopers (PwC)

In human resource management, retaining valuable employees is as important as attracting talent. PwC is constantly seeking to better understand how to engage and support its employees. The success of the project could improve employee's satisfaction and decrease instability. Moreover, it also provides a potential new offering to PwC clients, such as commercial banks suffering from the high turnover rate.

However, traditional personal trait analysis could not yield satisfactory results. Machine learning, evolved from the study of computational learning theory in artificial intelligence, can learn from and make predictions on data. It has been widely used in various industry applications, such as the Uber self-driving car and Google Translate.

In this project, we explored the effectiveness of machine learning in retention analytics. Various machine learning classifiers, feature selection methods, and data cleaning techniques were examined in the project to gain meaningful insights. The study results generate values in both retention analytics and data analytics, and prove that machine learning could predict employee turnover accurately. From a data analytic perspective, we found tree-based ensemble classifiers have decent performances while dealing with unbalanced data. Moreover, the ability to visualize classification rules makes tree-based classifier a good choice in the business setting.

**Project Team**
**Yue Zhao**, MScAC student
**Shak Parran**, PricewaterhouseCoopers
Professor **Anthony Bonner**

## A LOOK AT BLOCKCHAIN BEYOND DIGITAL CURRENCIES – SecureKey Technologies Inc.


SecureKey is designing a blockchain-based digital identity management service, running on consumer devices, to securely store personally identifiable information, such as name and home address, health card number. This service derives the next-generation identity and digital asset sharing initiative.

The solution enables users to utilize verified information from well-known trusted parties, such as banks and governments, directly from their own personal devices. It puts the consumer back in the middle, enabling them to take control over their digital assets, deciding whether to share them and with whom, always with informed consent.

SecureKey's approach provides an entirely new technology for building decentralized information management systems, enabling the creation of a nation-scale digital information marketplace. It revolutionizes the software engineering of data management systems by launching a new distributed blockchain-based technology that, instead of

providing a single database, provides a secure distributed storage system with privacy enhancing integrity controls. This may be accessed by different entities at different locations.

SecureKey's decentralized solution hampers cyber-attacks and is resistant to unauthorized data access by relying on powerful cryptography techniques, ensuring privacy and security.

**Project Team**
**Caroline Rodrigues**, MScAC student
**Dmitry Barinov**, SecureKey Technologies Inc.
Professor **Mariano Consens**

## A MOBILE APPLICATION FOR IDENTITY MANAGEMENT – SecureKey Technologies Inc.


The rise in adoption of online services for sensitive tasks such as banking and accessing public services, has led to increased security and data privacy requirements. Many institutions use third-party services for logging in and handling security. So you might enter a single username and password, connecting only once to access a number of different government services. This is convenient for a user, but there are some security risks in the way these third-party services are currently implemented. There's a single point at which an unauthorised user may be able to break into all the services. This project has involved implementing more secure ways of authenticating a user's identity to third parties by utilising blockchain technology and distributed storage. This means that data gets stored in multiple locations, removing any single point of vulnerability.

The goal for the mobile team has been to create an iOS application for iPhone, which allows a user to securely store their identity (such as name and address from a bank) and use this information when a third party requires authentication (such as a mobile provider). The research has focused on finding the most secure and reliable ways to transfer identity data between the application and the systems which store the data (blockchain and distributed storage). All the data needs to be encrypted and tamper proof, in order to withstand any kind of attack. It has also been important to create a user friendly design, with visual appeal and clear affordances.

**Project Team**
**Charlie Needham**, MScAC student
**Dmitry Barinov**, SecureKey Technologies Inc.
Professor **Mariano Consens**

## A RESTFUL API FOR A LARGE COLLECTION OF COMPLEX CLIENT SETTINGS – Goldman Sachs

In this project, we built a RESTful API that allows users to view and manage a large and complex collection of client settings. In the broker-dealer industry, the protocol for electronic communications between clients and broker-dealers is rapidly evolving and not strongly enforced. Also, clients often request bespoke functionality. This necessitates a large number of complex client-specific configuration settings. The overall goal of this project is to improve the visibility and ease of management for the large collection of client settings. This RESTful API achieves the following goals:

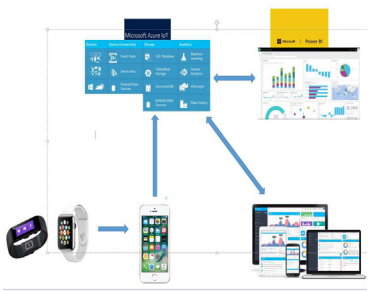
Provide a RESTful API that allows users to answer the following questions easily

- A summary view of client settings within a region or globally.
- What setting values are used by a specific client.
- How many and which clients are using a specific setting value
- Transform the client setting data to reflect their semantic meaning so users can query and retrieve the data in an intuitive way.
- Make the RESTful API easy to extend for new use-cases (eg. if the user wants a new view or if the user wants to expose new settings).
- Provide a programmatic interface that allows client applications to easily retrieve and collate client settings data with other data of interest.
- Bonus: allow concurrent access to the RESTful API.

This project is implemented using the Java Vert.x framework. In this presentation we outline the design decisions we made in order to achieve the project goals.

**Project Team**
**Yuan Wang**, MScAC student
**JP Lyall**, Goldman Sachs
Professor **Yashar Ganjali**

## ANXIETY METER FOR CHILDREN WITH AUTISM – Dymaxia



Autism is a developmental disability. Children with autism, also known as autism spectrum disorder or ASD, have social, communication and language problems. They also have restricted and repetitive patterns of behavior, interests, or activities, such as flipping objects, echolalia, or excessive smelling or touching of objects. Not all children with autism have the exact same problems. Parents and therapists are currently unable to objectively measure anxiety in children with autism to prevent emotional meltdown. They rely on subjective questions like “Do you feel like your volcano is erupting?” Up to 84% of children with autism experience anxiety which can cause emotional meltdowns that disrupts school performance, social development and self-confidence.

Anxiety Meter is a mobile health app which uses clinically validated technology with the ease and convenience of wearable wristbands. The app provide real-time feedback for the child and caregiver on a smart device. The core function is to collect the data from

wearable device and detect heart rate and relates it to anxiety levels on a smart device. The purpose is to empower children and caregivers to detect anxiety and take proactive measures to alleviate an emotional meltdown. It also uploads the data to the Microsoft Azure, and allows the parents and therapists monitor the anxiety level from the remote devices, such as mobile phone, laptop or desktop. Thus, teachers, therapists to parents, know how your child's anxiety levels are throughout the day.

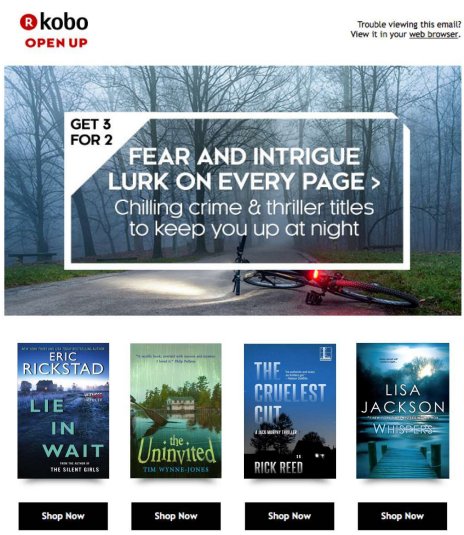
**Project Team**
**Guang Yang**, MScAC student
**Hussam Malek**, Dymaxia
Professor **Jonathan Rose**

## BUILD GOOGLE HOME SEARCH FEATURE WITH KNOWLEDGE GRAPH – Google


Google Home is a voice-activated device powered by Google Assistant. It acts like a personal assistant that can understand your questions, search for answers and make responses. It relies on the technical accumulations of search, and is more effective and human friendly in interaction. I worked on playing movie or TV shows music on Google Home. Users may ask for the same movie's theme song with various queries. The intention needs to be identified for different natural language descriptions. I built data driven models to improve the precision and recall of the triggering. Before I worked on the project, the response came from the web search results of the query, which was not always reliable. In the production, the response needs to reach high precision and recall. Knowledge graph is a knowledge graph built by Google, which contains high quality semantic search information. The knowledge graph helps search understand the real world as human beings do. Things, people, places and their related information are all included in the knowledge graph. In this project, I firstly identified the movie the user was interested in, and then retrieved the song of it from the knowledge graph. The data quality is also an important issue in the project. I evaluated the quality and coverage of the data in knowledge graph, and explored ways to improve it.

**Project Team**
**Amelia Gao**, MScAC student
**Xiaoming Chen**, Google
Professor **Raquel Urtasun**

## **DATA DRIVEN CRM FOR OPTIMISED EREADING** – Rakuten Kobo Inc.




Engaging customers via email communication is an important aspect of customer relationship management (CRM). In addition to driving sales, the user experience can be enhanced by cleverly targeting users with relevant and personalized content in a context and time - sensitive manner. At Kobo, email marketing campaigns may be launched for various reasons, such as to promote a specific author, to inform customers of a sale,or to advertis e a set of newly released books. Generally, these marketing campaigns are focused around a specific list of promoted books. This raises the question of how to properly select customers from the user base who are likely to be interested in the items present in a given book list.

Furthermore, many of these lists may consist of hundreds of books, and only a small subset may be shown to the user within an email; therefore, finding the optimal ordering of a book list for each user is also a concern. In this project,various data mining and machine learning techniques were explored for intelligently targeting existing customers with personalized lists of promoted books.

Vector - based methods to compute the “affinity” of a customer to a given list were investigated for both targeting and personalization. Additionally, targeting approaches that model the probability of a user purchasing an item from a given email campaign were also explored.

**Project Team**
**Jake Stolee**, MScAC student
**Dr. Darius Braziunas**, Rakuten Kobo Inc.
Assistant Professor, Teaching Stream, **Nathan Taback**

## DETECTING FINANCIAL FRAUD USING SUPERVISED MACHINE LEARNING – CaseWare International, Inc.

The primary aim of our project is to use computational techniques to analyze the financial reports that companies release to private investors and the public. Specifically, our goal is to automate the detection of financial statements that do not accurately represent the issuing firm. To achieve this, we investigate methods to represent the textual and numeric data contained in financial statements. Using these representations, we employ machine learning (ML) and natural language processing (NLP) techniques to i) construct models to predict the likelihood that a particular financial statement will later be modified, ii) lead to an investigation by governing authorities, or iii) both. Past research in this area has primarily focused on the numeric data of the statements. In contrast, a major aspect of our work uses NLP techniques to make predictions about likelihood of fraud or misrepresentation using the textual financial notes.

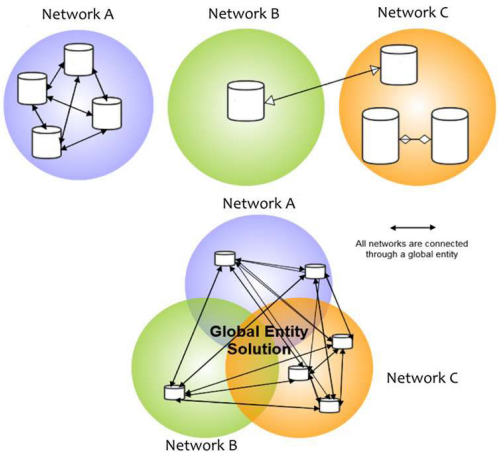
**Project Team**
**Tom LaMantia & Abraham Escalante**, MScAC students
**Brett Kennedy**, CaseWare International, Inc.
Professor **Frank Rudzicz**

## DYNAMIC MEMORY ALLOCATOR FOR CUDA DYNAMIC PARALLELISM – NVIDIA

Memory required for any CUDA thread is typically pre-allocated by the CUDA driver. The introduction of CUDA Dynamic Parallelism, however, challenger this behavior. Dynamic parallelism allows threads to recursively launch any number of CUDA kernels and synchronize on them. Synchronization is particularly very memory expensive on GPUs as the states of thousands of threads may need to be saved out. A static pre-allocation scheme not just falls short in meeting the memory needs of such threads for huge GPUs, but is also extremely memory inefficient. In this project, we designed a dynamic memory allocation scheme for the GPU. Our parallel allocator consistently offers high-performance even in the most demanding cases and is also memory-efficient.

**Project Team**
**Pradeep D Moorthy**, MScAC student
**Melvin McGee**, NVIDIA
Professor **Eyal de Lara**

## **ENTITY RESOLUTION** – Scotiabank




Scotiabank has developed and built a data ecosystem that contains Scotiabank proprietary data and data from 3rd parties. This ecosystem contains data that has been transformed into network data structures, and entity reference data. Currently, analysis is performed on each unique network data structure, and there is no global entity that maps these data networks together. Scotiabank would like to create a global entity management process, which takes data from the ecosystem to generate a unique entity. The definition of entity is broad and may include retail, commercial, machines, financial institutions etc.

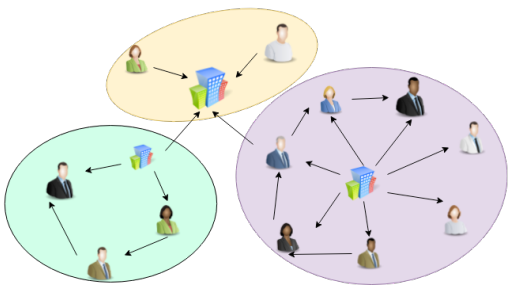
In this project, we are dealing with unstructured and structured data in the scale of billions of records. We use both classic and state-of-the-art research results from parallel computing, natural language processing, data mining and machine learning to develop a novel entity solution system.

The network data structures contain rich entity information. This entity information is complex and has a geographical scope that contains financial routing information, competitor information, and government information. On a daily basis the volume of data grows and new entity information is captured. Scotiabank would like to build an engine that maintains existing entity IDs and generates new entity IDs from the data

ecosystem. This engine must be flexible enough to add new entity information with newly acquired data.

**Project Team**
**Zili Chen**, MScAC student
**David King & Junjie Zhu**, Scotiabank
Professor **Peter Marbach**

## **GRAPH-BASED MEASUREMENT FRAMEWORK FOR FINANCIAL NETWORK ANALYTICS** – Scotiabank



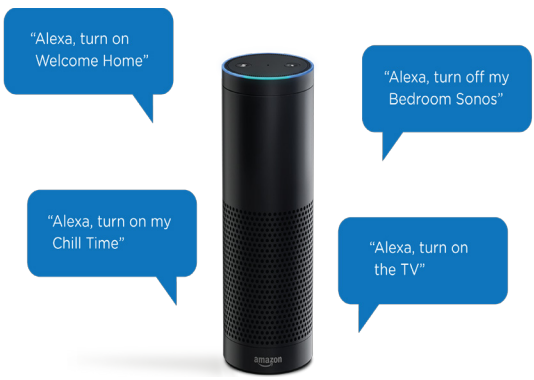
Network theory is becoming more present in finance because it provides solutions to problems that cannot be solved by traditional analysis. Scotiabank has built a data ecosystem that contains various networks of Scotiabank proprietary data and data from 3rd parties. The network data structures contain nodes that represent individuals or institutions, and edges that represent interactions between the entities. The transaction data is collected on a daily basis and the data is growing substantially day by day. As a result, traditional algorithms become computationally expensive for the large scale data. On the

other hand, an analytical framework that is applicable to different types of networks is needed, based on which different approaches can be developed according to business objectives.

The objective of this project is to develop a generic measurement framework based on graph theory for large scale financial network analytic problems. Based on the standardized transaction data, we applied our analysis on two levels. On structure level, we applied community detection to the network, identified clusters of nodes that are highly interconnected together, and studied the evolution of communities over time. On the node level, using traditional graph theory, we implemented a series of centrality metrics which provide multi-dimensional features of the nodes. These features are used for various applications such as identifying nodes with specific characteristics, detecting topology in communities, etc.

**Project Team**
**Weiwei Li**, MScAC student
**David King & Alex Jeon**, Scotiabank
Professor **Peter Marbach**

## **HAMMURABI** – Amazon



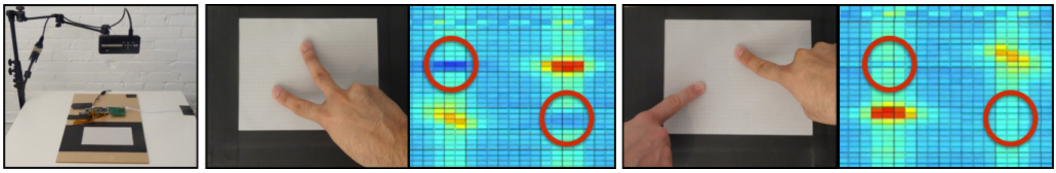

Alexa is Amazon's intelligent service that allows the user to voice-enable any connected product with a microphone and speaker.

Alexa employs thousands of heuristic rules, finite state transducers, and a variety of statistical models to classify speakers' intent. The thousands of heuristic rules reside in flat text files and are managed manually via file editing. This workflow is hard to maintain and scale, and consequently makes the entire process of deployment very slow. Additionally, this workflow also entails a steep learning curve - which means that only a selected group of individuals are able to effectively maneuver through the text files and make the required changes (an unneeded dependency).

Hammurabi, my internship project, is a tool that tends to solve all of the above mentioned problems through an intuitive user interface, that is powered by a very specialized backend service. It is a tool that consumes all the rules that span across multiple text files, and builds a representation that is easy to understand. It removes the dependency on text files, and allows the user to easily create, update, prioritize, save, and retire rules. It flattens out the learning curve even further by allowing the user to test & deploy the updated set of rules right from the interface. Hammurabi will eventually evolve into a tool that will also help with tracking the actual usage of every rule, and in turn will help with the overall management of the system.

**Project Team**
**Syed Rizwan Gilani**, MScAC student
**Kelly Vanee**, Amazon
Professor **Frank Rudzicz**

## INPUT DISCRIMINATION ON CAPACITIVE MULTI-TOUCH SENSORS – Tactual Labs



Everything from phones to cars, are now touch enabled. While we can easily detect when someone touches a device, we cannot tell what exact finger is touching the interface, i.e. thumb or index finger. Even the most advanced technologies in production fail to differentiate touch as coming from the same or different users, if it is a left or right hand, or what finger is used without any external hardware.

**Project Team**
**Sidharth Sahdev,** MScAC student
**Ricardo Jota**, Tactual Labs
Professor **Eugene Fiume**

## LEGAL COGNITION: AN UNSUPERVISED PASSAGE RANKER FOR LEGAL QUESTIONS – ROSS Intelligence

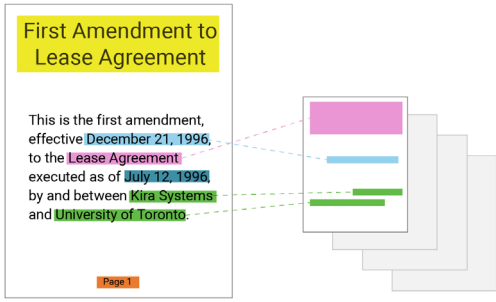
ROSS is an A.I. lawyer that helps human lawyers to find accurate answers to legal questions asked in a natural language. ROSS is used by industry leading law firms including Dentons, the world´s largest firm by headcount, and Latham & Watkins the largest by revenue.

The project was to build a framework that could allow ROSS to understand legal text in more depth and, therefore, answer more complicated legal questions than before. The framework gets passages from a passage retriever, found in U.S. case law, and ranks each of the passages based on how well the passage answers the question.

The framework uses a text paraphrasing algorithm, based on an entity-relation analysis, to score the passages. The algorithm uses a core of natural language processing techniques to detect entities (legal phrases) and syntactic relations between every two related entities in a text. The passage is then scored in accordance with how well the entity-relations for the passage match the entity-relations for the question.

**Project Team**
**Elías Jónsson**, MScAC student
**Jimoh Ovbiagele**, ROSS Intelligence
Professor **Graeme Hirst**

## LINKING AMENDED DOCUMENTS TO THEIR CONTRACT FAMILY – Kira Systems




Lawyers spend a significant amount of time performing manual labour tasks. One of these manual tasks is tracking amended documents passing between various stakeholders and maintaining the link to the original document. Various factors contribute to the complexity of the aforementioned endeavour, including large scale logistical and coordination challenges, and a lack of standard formatting for documents. In addition, a large proportion of documents are digital image versions converted by optical character recognition. Consequently, law firms end up with databases of documents with no metadata or links between related documents. The purpose of this project is to create a method to automate this linkage between referring documents. This project provides a chronological display of families of documents allowing a better understanding and simpler analysis and review process.

**Project Team**
**Eda Doko**, MScAC student
**Alexander Hudek**, Kira Systems
Professor **Frank Rudzicz**

## MARKET SEGMENTATION WITH MULTINOMIAL MIXTURE MODELS – Ipsos Group

Ipsos is a global leader in research, specializing in advertising, loyalty , marketing, media and public affairs research. It's market research organization conducts detailed surveys on customer opinions and consumption behaviour for a given market. Of primary interest to analysts and clients is "how is this market segmented in terms of brands and consumers?"

The purpose of this project is to answer exactly that type of question. Given a large amount of market research survey data, a machine learning clustering algorithm known as a Binomial Mixture Model will be trained via the Expectation-Maximization algorithm. Survey respondents will be clustered based on their brand consumption behaviour, and each cluster will be subsequently mined for demographic insights.

**Project Team**
**Neill Patterson**, MScAC student
**Daniel Gahan**, Ipsos Group
Dr. **Radu Craiu**

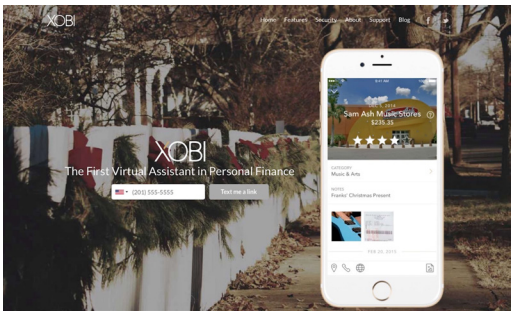
## MEASURING FEATURE SUCCESS IN FIREFOX – Mozilla

Currently Mozilla does not have a reliable way to make decisions about bringing new features into Firefox. Common practice is to simply release new features to the entire

Firefox population without measuring or testing the feature's impact until after release. A major new feature, Activity Stream, replaces the functionality of the current New Tab page in Firefox. In order to feel confident releasing a major change to all users, a rigorous testing and analysis process must take place. If this new roll-out process is successful in Activity Stream, it will be automated and applied to all new features going into Firefox.

**Project Team**
**Marina Samuel**, MScAC student
**Tim Spurway**, Mozilla
Professor **Mariano Consens**

## MERCHANT RESOLUTION OF FINANCIAL TRANSACTION RECORDS – SmartFinance



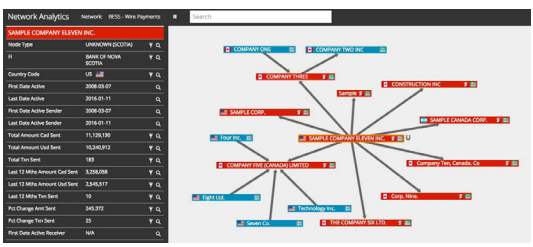

With the development of e-finance, most transactions utilize debit or credit cards, and this makes it possible to help people monitor their financial history in a more user-friendly way.

SmartFinance is developing an iOS app called Xobi. It allows people to link any debit or credit cards, and provides users with personalized insights for informed financial decisions with advanced features. Xobi also offers users the ability to ask questions in natural language about their spending, as well as upload receipts to match the appropriate transactions.

Merchant resolution is an important component to achieve the goals. It applies sophisticated string processing and matching procedures to identify the relevant merchant information from the messy financial transactions, which are often uncategorized, incomplete and inconsistent. Merchant resolution API is now ready to release, with the purpose of giving both the banks and their customers a more complete picture of financial transactions, and to lay a proper foundation for an artificial intelligent semantic search platform at high speed.

**Project Team**
**Yiqing Yang**, MScAC student
**Richard Zemel**, SmartFinance
Professor **Raquel Urtasun**

## NETWORK VISUALIZATION PLATFORM – Scotiabank




Scotiabank has developed and built a data ecosystem that contains Scotiabank proprietary data and data from third parties. This ecosystem contains data that has been transformed into network data structures and entity reference data.

We have developed a big data architecture in the back-end and a web application

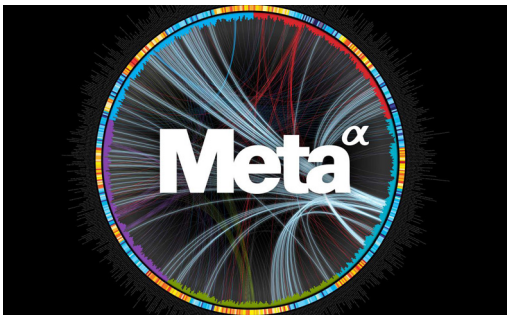
in the front-end, to allow access to analyses on these structures performed by the data sciences team, as well as enabling exploration of the results by internal downstream partners.

After the data has been ingested from our different data sources, we process and analyze the data on our data ecosystem, before serving it to the visualizer. Users are then provided with a networked view of the data and are able to modify the structure, explore relationships over time, search the dataset and create reports from their customized view.

Visualizing large scale graphs in an aesthetically pleasing manner is challenging. To address some of the challenges, we employed several techniques such as, breaking large graphs into highly connected sub-graphs (communities), enabling the egocentric analysis and exploration of nodes of interest and finally, presenting networks in a force-directed layout.

**Project Team**
**Kyriakos Georgiou**, MScAC student
**Andrew Kujtan**, Scotiabank
Professor **Peter Marbach**

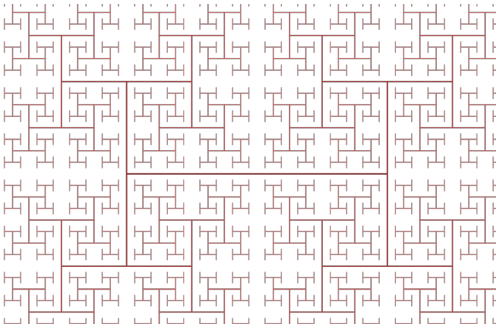
## PREDICTING IMPACT OF BIOLOGICAL RESEARCH PAPERS – Meta Inc.




Meta's Bibliometric Intelligence (BI) is a service platform to provide automated and in-depth manuscript assessment to authors and publishers. My internship project is to develop a new version of BI algorithm to predict the future impact indicators of biological manuscripts. Predicted indicators include Eigenfactors® , citations and whether the paper belongs to the top 1% / top 5% / top 10% classes among all the manuscripts. Both content- and metadata-based features are used to train the model. The new model increases both the accuracy and consistency of manuscript impact prediction. Accurate assessment and profiling of manuscripts by the new BI algorithm will allow researchers and editors to pre-rank research papers efficiently, and to reduce the unnecessary painful submission waiting time. Success of this project not only brings in business value to the partner organization, but also greatly benefits the research publication eco-system.

**Project Team**
**Liu Yang**, MScAC student
**Ofer Shai**, Meta Inc.
Professor **Frank Rudzicz**

## **ROUTING LOCAL CLOCKS BEFORE DATA** – Altera



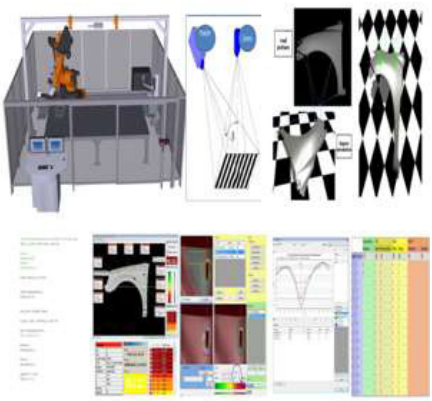
Altera designs and manufactures field-programmable gate arrays (FPGAs), which are integrated circuits that can be programmed after manufacturing. Altera also develops computer-aided design (CAD) software that allows users to program the FPGAs. Within the CAD team, the Routing Group is tasked with mapping user circuit connections onto the physical FPGA fabric of horizontal and vertical wires and logic elements.

Modern FPGAs have pre-fabricated clock grids that utilize balanced tree topologies to distribute clock signals to many thousands of registers with extremely low skew. Sometimes, however, clock signals will not use the global clock grid for many reasons. The goal of this project is to research and implement synthesis algorithms for mapping clock connections in FPGA designs to the programmable routing fabric, such that setup is met with an optimal circuit speed and hold constraints are not violated. We have added an additional local clock routing phase to the software flow of the router. Tis was done by simply routing the clock connections according to minimum delay and congestion resolution, and then locking down those connections along with their min/max delay budgets. This is so that slack allocation can assign valid delay budgets to the rest of the connections knowing that the clock connections will not change. This should improve the QoR for non-global clock domains.

Eventually, the algorithm to route the local clocks will improve QoR for cross clock connections and same clock connections by optimizing three goals: clock skew, wire length, and routing delay.

**Project Team**
**Alvin Leung**, MScAC student
**Nish Sinnadurai**, Altera
Professor **Paul Chow**

## **SURFACE DEFORMATION DETECTING SYSTEM USING COMPUTER VISION** – Shanghai Winners Automation Tech Co., Ltd



The goal of this project is to develop a system for detecting defects on the surface of a steel part, which will be used in an automobile manufacturing product line for quality control. This system is intended to improve on the precision and speed of an existing 3D computer vision system that detects surface deformations that indicate defects. Using new structured light sensing technology, combined with recent machine learning technology, the new system aims to improve workflow on the assembly line.

In the first stage of our project, the position and orientation of a test part with respect to a 3D geometric model part is computed,

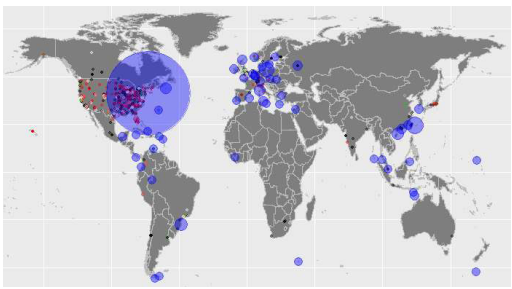
allowing the test part to be modeled in a simulator. Next, the simulator can define a non-uniform search grid overlaid on the test part, in which higher-curvature subsurfaces are localized, such as the deformations surrounding a car door handle. These subsurfaces represent "fields of view" for a structured light scanner that is automatically positioned to acquire 3D images of the grid cells. Moreover, the simulator can also model unwanted lighting effects which can be masked out during image acquisition. From the resulting range images, features are extracted and fed to a classifier trained on positive and negative (defective) subsurfaces. The results are automatically captured in a QA report, providing critical diagnostics on individual parts as well as entire processes.

**Project Team**
**Peng Xu**, MScAC student
**Tianming Zhao**, Shanghai Winners Automation Tech Co.,Ltd
Professor **Sven Dickinson**

## **SWIFT/OMR JUST-IN-TIME PROJECT** – IBM Canada Lab

This project explores how reusable compiler componentry from the open source Eclipse OMR project could be employed for the open source Swift programming language in server side deployments. The goal is to evaluate how well the dynamic and speculative optimization framework from a Java Just In Time (JIT) compiler may be able to accelerate server side applications written in the Swift language. One of the key initial focus areas for this work was to be able to translate the Swift Intermediate Language (SIL) into the intermediate language of the OMR JIT compiler to enable the evaluation of the dynamic optimization technology. As IBM decided to extend its already-considerable support for the Swift programming language, the project is going to enhance the efficiency of Swift compilation by applying robust OMR technologies. The OMR project consists of a highly integrated set of open source C and C++ components that can be used to build language runtimes that support many different hardware and operating system platforms. These components include but are not limited to: memory management, threading, platform port (abstraction) library, diagnostic support, monitoring support, garbage collection, and native Just In Time compilation.

**Project Team**
**Shuyu Li & Zhongtian Qiu**, MScAC students
**Mark Stoodley, Vijay Sundaresan**, IBM Canada Lab
Professor **Angela Demke Brown &** Assistant Professor, Teaching Stream, **Mathew Zaleski**



## **TAX DATA ANALYTICS: EXPLORATORY ANALYSIS FOR CORPORATE TAX BENCHMARKING AND PLANNING** – PricewaterhouseCoopers (PwC)

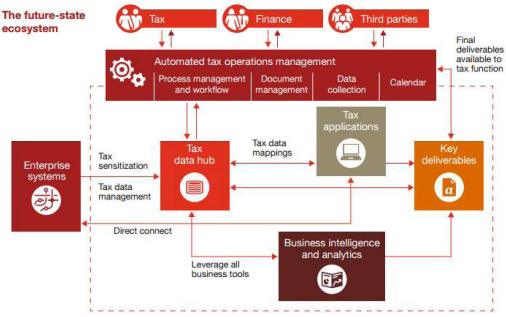
PwC's Tax Transfer pricing team offers Value Chain Transformation (VCT) services

designed to help multi-national corporations to transform business operations in a tax-efficient manner that increases shareholder value and improves cash flow. To facilitate this service line, population of public and private company financial statements, tax returns, and detail transaction level information such as intercompany transactions and debt issuance data. Currently, the data analytics team and Tax team are jointly building a tax data analytics platform which will integrate disparate data sources in order to automate tax related provision and compliance processes, self-serve reporting and advanced analytics.

The project focuses on 1) searching for tax efficiency benchmarks and key performance indicators 2) measuring its impact on individual company's operating efficiency and profitability with a tax "lens" 3) scoring individual performances within each industry peer groups or conversely identifying clusters with similar characteristics, and 4) combining the results with internal resources to apply them in various areas such as detection of risk, opportunity identification, prediction and planning, and overall business support. The project itself consists of multiple modular components, with each ideas built upon another by testing different hypothesis generated by inputs from tax professionals within the team. Statistical and data mining techniques are applied to identify patterns from third party public financial data that may not be easily seen through standard reporting and visualization.

**Project Team**
**Sooa Lim**, MScAC student
**Jeff Rogers**, PricewaterhouseCoopers
Professor **Ken Jackson**

## **TECHNICAL SOLUTIONS IN EQUITY DERIVATIVES** – Goldman Sachs




An equity derivative is a financial instrument whose value is derived from the price of an equity security. Futures and options are the most common equity derivatives; others are equity swaps, structured products and convertible bonds.

A swap is an over-the-counter (OTC) financial instrument in which two parties agree to make a series of payments to each other over a period of time. In an equity swap, one party agrees to pay the return on some fixed amount (notional) of equity while the other party agrees to pay an agreed upon rate such as LIBOR. Equity swaps have become very popular since their introduction in the late 1980s as they allow firms to more precisely control their exposure to equity risk. Two projects are presented here. The first is a web app used to describe and price equity swaps. The second explores the use of synthetic products during netting procedures at a large broker-dealer

**Project Team**
**Matthew Shepherd**, MScAC student
**Cameron Brien**, Goldman Sachs
Professor **Kenneth Jackson &** Assistant Professor, Teaching Stream, **Matt Medland**

## CONTACT US

Master of Science in Applied Computing (MScAC)
Department of Computer Science
University of Toronto
40 St. George Street
Toronto, ON
M5S 2E4
+1 416 946-5292
mscac@cs.toronto.edu

**www.cs.toronto.edu/mscac**