



# UNDERGRADUATE RESEARCH SHOWCASE

2015



Computer Science  
UNIVERSITY OF TORONTO

# **AUTOMATIC CATEGORIZATION OF DISCUSSION FORUM POSTS**

A forum thread consists of posts in the form of conversations between multiple users, often including categories like questions, answers, feedback and off-topic discussions. Identifying these categories automatically can help summarize a long forum thread and make information retrieval more efficient. However, this categorization depends on complex linguistic factors which makes it a difficult task. This project aims to develop statistical techniques for automatically clustering forum posts into these categories.

**Mohamed Abdalla**

[mohamed.abdalla@mail.utoronto.ca](mailto:mohamed.abdalla@mail.utoronto.ca)

**Krish Perumal**

[krish@cs.toronto.edu](mailto:krish@cs.toronto.edu)

Professor **Graeme Hirst**

[gh@cs.toronto.edu](mailto:gh@cs.toronto.edu)

## 3-D OBJECT RECOGNITION



Developing a 3D Object Recognition pipeline using RGBD images which provide unambiguous shape information. Generate multi-scale region/surface segmentation proposals from which viewpoint-invariant features can be extracted, match to database of object models.

**Katie Datsenko** (*photo left*)

[katiedatsenko@mail.utoronto.ca](mailto:katiedatsenko@mail.utoronto.ca)

**Kamyar Seyed Ghasemipour**

[kamyar.seyedghasemipour@mail.utoronto.ca](mailto:kamyar.seyedghasemipour@mail.utoronto.ca)

Professor **Sven Dickinson**

[sven@cs.toronto.edu](mailto:sven@cs.toronto.edu)

**Faisal Qureshi**, Visiting Professor

[faisal@cs.toronto.edu](mailto:faisal@cs.toronto.edu)



## LIFTING A MODEL TRANSFORMATION LANGUAGE



Model Driven Engineering (MDE), the use of models to represent software artifacts, helps manage complexity in software engineering by raising the level of abstraction. Model transformations take in one model as input and produce another as output, making the changes specified by a set of transformation rules. DSLTrans is a visual language in which such transformations can be performed.

Software Product Lines (SPLs) help manage complexity in software engineering by concisely representing a set of similar product variants, explicitly capturing product commonalities and differences. While model transformations and SPLs are powerful on their own, the transformation of SPLs is traditionally inefficient and computationally expensive.

Lifted transformations operate on entire SPLs instead of individual models. In a lifted transformation, the transformation rules themselves do not change, but the way in which they are applied to the input does. This project successfully completes the lifting of the DSLTrans language engine so that it can support transformations on entire SPLs instead of just individual models.

**Kawin Ethayarajh** (*photo left*)  
[kawin.ethayarajh@mail.utoronto.ca](mailto:kawin.ethayarajh@mail.utoronto.ca)

Professor **Marsha Chechik**  
[chechik@cs.toronto.edu](mailto:chechik@cs.toronto.edu)

# COMPUTATION OF LOSS DISTRIBUTION FOR CREDIT PORTFOLIOS



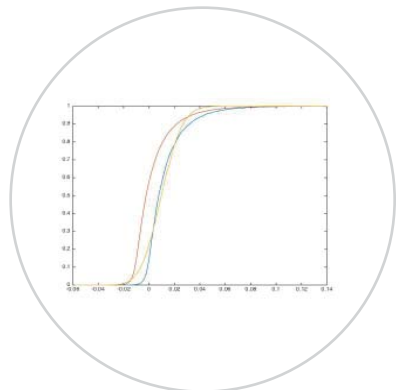
Models for the dynamic behavior of credit portfolio are crucial in the field of credit portfolio analysis and risk management, especially with the implementation of Basel III, which requires more up-to-date liquidity examination and risk exposure control. Our team aims to develop credit portfolio valuation models in a dynamics scheme by cutting long time range into small sub-periods. Among all the properties, cumulative loss probability is of the most concern. Furthermore, several plausible algorithms are proposed and compared using artificially simulated portfolio data. In the first part of our model, inter-period relationship is assumed to be independent which indicates that all credit products are replaced with kindred assets at the end of each sub-period. In this case, Monte Carlo method, convolution method and conditional Monte Carlo method are performed on the same sample data set. In the second part, cumulative loss probability will be calculated under the dependent case. From the industry perspective, this scenario is of more importance since the liquidation at the end of each period is far from realistic.

**Zhenan Fan** (*photo left*)  
[zhenan.fan@mail.utoronto.ca](mailto:zhenan.fan@mail.utoronto.ca)

**Yuehao (Josh) Qin**  
[yuehao.qin@mail.utoronto.ca](mailto:yuehao.qin@mail.utoronto.ca)

**Meng Han**  
[m7han@cs.toronto.edu](mailto:m7han@cs.toronto.edu)

Professor **Ken Jackson**  
[krj@cs.toronto.edu](mailto:krj@cs.toronto.edu)



## **VISUALLY SIMULATING GOAL MODELS OVER TIME**

Goal models in the i\* language are a useful tool for modelling software system requirements. Analysis techniques exist on these models to assist stakeholders in choosing between alternatives. However, these techniques are limited by the fact that they use static models. A method has been developed to add dynamism to these models by using functions of time. This project's goal was to develop an i\* modelling tool that incorporated these dynamic functions and allowed users to simulate changes of their models over time, thereby leading to more informed decisions.

**Jake Fear**

[jake.fear@mail.utoronto.ca](mailto:jake.fear@mail.utoronto.ca)

**Alicia Grubb**

[amgrubb@cs.toronto.edu](mailto:amgrubb@cs.toronto.edu)

Professor **Marsha Chechik**

[chechik@cs.toronto.edu](mailto:chechik@cs.toronto.edu)

## COLLABORATIVE DISEASE DIAGNOSTICS



Collaborative Disease Diagnostics is a computer vision and machine learning company who is in the medical diagnostics market. We are developing an image pattern recognition software to be used in conjunction with current microscopy procedures to diagnose malaria. Malaria is one of the leading causes of death in developing countries particularly afflicting children under five and pregnant women.

We have validated that manual microscopy diagnosis of malaria within red blood cells is significantly prone to human error which translates to negative consequences economically and in the health of the population. Automation of microscopy diagnosis through software designed by CDDx will eliminate the afore listed issues.

Our architecture consists of four components: image processing and rendering, detection, feature extraction, and classification. We are currently focusing on cell detection using difference of Gaussians. This is done by subtracting a blurred version of the image from the original image, which enhances the edges of the original image. Next, the watershed segmentation technique is implemented to detect overlapping cells. By determining the local minimums of the image, this method is able to segment the original image, producing sub-differentiated objects the original image is composed of. This is incredibly important for identifying red blood cells since these cells tend to overlap-especially on slides which have not been well-prepared. Using these techniques allowed us to accurately count the number of cells as well as then crop the image so that we can then proceed with feature extraction.

**Yannie Lai, Victoria Bukta, Fatema Chowdhury and Tien-Che Lee**  
[yannie.lai@mail.utoronto.ca](mailto:yannie.lai@mail.utoronto.ca)

**Mario Grech**, Director, Innovation Lab (DCSIL)  
[mario.grech@utoronto.ca](mailto:mario.grech@utoronto.ca)

# HIGH-ACCURACY 3D SCANNING BY STRUCTURED LIGHT TRANSPORT

The key point in 3D scanning is to acquire accurate correspondence of the two views, where structured light is often used. However, many previous approaches using structured light are likely to cause systematic errors due to the assumption that the light should get reflected only once when traveling between the light source, which is often a projector, and the camera, and we call such light direct light. The error gets huger when there is specular surface or transparent objects in the scene to be scanned. To address the problem, we present an approach to capture only direct light part of the projected structured light, using primal-dual coded camera, which consists of a conventional camera, a DMD mask in front of the camera, and a projector.

**Zhaowei Liu**

[tod.liu@mail.utoronto.ca](mailto:tod.liu@mail.utoronto.ca)

Professor **Kyros Kutulakos**

[kyros@cs.toronto.edu](mailto:kyros@cs.toronto.edu)



# SQL VISUALIZER



An online visualizer for SQL queries, providing intermediate tables and meta information to help understand how SQL queries get executed.

**Martin Petkov** (photo left)  
[martin.petkov@mail.utoronto.ca](mailto:martin.petkov@mail.utoronto.ca)

**Katherine Ge**  
[k.ge@mail.utoronto.ca](mailto:k.ge@mail.utoronto.ca)

**Sophia Huynh**  
[sophia.huynh@mail.utoronto.ca](mailto:sophia.huynh@mail.utoronto.ca)

Associate Professor, Teaching Stream, **Diane Horton**  
[dianeh@cs.toronto.edu](mailto:dianeh@cs.toronto.edu)

Associate Professor, Teaching Stream, **Andrew Petersen** (UTM)  
[petersen@cs.toronto.edu](mailto:petersen@cs.toronto.edu)

The screenshot displays the SQL Visualizer interface. On the left, a SQL query is shown: `SELECT sid, dept || cnum AS course, grade FROM Took NATURAL JOIN Offering WHERE cnum IN (SELECT cnum FROM Student WHERE (first_name = 'Sophia' OR first_name = 'Kathy'));`. The main area shows two intermediate tables, both titled 'Source: Took NATURAL JOIN Offering'. The first table has 4 rows, with the second row highlighted in green. A tooltip titled 'Reasons' points to this row, showing the subquery result: `SELECT cnum FROM Student WHERE (first_name = 'Sophia' OR first_name = 'Kathy'))`. The second table has 4 rows, with the first row highlighted in blue. At the bottom, there are sections for 'Namespace' and 'Tables and Views'.

```
SELECT sid, dept || cnum AS course, grade FROM Took NATURAL JOIN Offering WHERE cnum IN (SELECT cnum FROM Student WHERE (first_name = 'Sophia' OR first_name = 'Kathy'));
```

1. FROM Took NATURAL JOIN Offering

2. WHERE ( cnum IN ( SELECT cnum FROM Student WHERE ( ( first\_name = 'Sophia' ) OR ( first\_name = 'Kathy' ) ) ) ) )

3. SELECT sid, dept || cnum AS course, grade

Source: Took NATURAL JOIN Offering

oid	aid	grade	dept	cnum	instr
2	1	87	csc	343	D. H
4	1	73	ger	100	E. I
2	2	92	csc	343	D. H
1	3	80	csc	209	K. F
1	4	60	csc	209	K. F

Reasons

```
SELECT cnum FROM Student WHERE (first_name = 'Sophia' OR first_name = 'Kathy'))
```

oid	aid	grade	dept	cnum	instr
2	1	87	csc	343	D. H
2	2	92	csc	343	D. H
1	3	80	csc	209	K. F
1	4	60	csc	209	K. F

Namespace

- Took: sid, oid, grade, course, cnum
- Offering: oid, dept, cNum, instructor, course, cnum

Tables and Views

- Student (sid, firstName, email, cpga, cnum)
- Offering (oid, dept, cNum, instructor, course, cnum)
- Course (dept, cNum, name)
- Took (sid, oid, grade, course, cnum)

## CALCULATION OF BLOOD FLOW RATES FROM CT SCANS



The goal of this project is to calculate blood flow rates in arteries from CT scans. The motivation is to calculate the flow in coronary by-pass blood-vessel grafts, as well as the amount of blood that flows backwards around a particular type of artificial heart valve. Being able to compute accurate blood flow rates will be very useful in assessing patients and deciding on potential treatments.

The method developed is based on injecting contrast agent into a patient and using a CT scan to measure the time-varying level of contrast agent in a blood vessel segment. Computing the blood flow rate from these CT measurements of time-varying contrast-agent levels is an inverse problem. To solve this problem, a fairly simple numerical model for the change of contrast agent in a blood vessel that depends on the blood flow rate,  $Q$ . Then using a minimization method that runs this model many times with different  $Q$ 's to find the  $Q$  that best fits the measurements from the CT scan. Additional methods include function analysis of the changes in concentration from the CT scan. We are now in the process of testing this method on (1) “synthetic” data created by a computer model, (2) data from scans of water in pipes and (3) actual patient data.



**Bogdan Scaunasu** (*photo left*)  
[bogdan.scaunasu@mail.utoronto.ca](mailto:bogdan.scaunasu@mail.utoronto.ca)

**Lin-Liang (Bill) Wu** (*photo right*)  
[linliang.wu@mail.utoronto.ca](mailto:linliang.wu@mail.utoronto.ca)

Professor **Ken Jackson**  
[krj@cs.toronto.edu](mailto:krj@cs.toronto.edu)

## A FAST WAIT-FREE SINGLE-EXTRACTOR PRIORITY QUEUE



A single-extractor priority queue is a concurrent priority queue, with  $n$  processes that can perform Insert operations and one process that can perform Extract-Max operations. We present a single-extractor priority queue implementation from LL/SC objects that is wait-free with  $O(\log n + \log m)$  step complexity for both Insert and Extract-Max, where  $m$  is the maximum number of elements in the priority queue

during the operation. Our implementation uses a single-inserter single-deleter (SISD) ordered multiset, which supports insertion and deletion with  $O(\log r)$  step complexity, and finding the maximum element with constant step complexity, where  $r$  is the maximum number of elements in the multiset during the operation.

**Yuanhao Wei** (*photo left*)

[yuanhao.wei@mail.utoronto.ca](mailto:yuanhao.wei@mail.utoronto.ca)

**Kenneth Donald Hoover** (*photo right*)

[khoover@cs.toronto.edu](mailto:khoover@cs.toronto.edu)

Professor **Faith Ellen**

[faith@cs.toronto.edu](mailto:faith@cs.toronto.edu)



## MarkUs



MarkUs (pronounced “mark us”) is an open-source tool which recreates the ease and flexibility of grading assignments with pen on paper, within a web application. It also allows students and instructors to form groups, and collaborate on assignments.

**Lawrence Wu** (*photo left*)

[lawrence@lawrencewu.me](mailto:lawrence@lawrencewu.me)

**Jesse Gleave Riemann** (*photo right*)

[jesse.riemann@gmail.com](mailto:jesse.riemann@gmail.com)

Associate Professor, Teaching Stream, **Karen Reid**

[reid@cs.toronto.edu](mailto:reid@cs.toronto.edu)

**David Liu**, Lecturer

[david@cs.toronto.edu](mailto:david@cs.toronto.edu)

# SEMI-SUPERVISED GENOMIC PATTERN DISCOVERY USING SEGWAY



The human genome is packaged into chromatin, a structure of DNA and thousands of associated proteins. The ENCODE Project has produced maps of where chromatin proteins localize in specific human cell types. Using a software called Segway, which can annotate genome semi-automatically through machine learning, we can transform these maps into annotations of regions important to genome regulation.

However, most of what have been done in Segway uses unsupervised learning. Here we explore method of using semi-supervised machine learning techniques on human genome pattern discovery, and try to apply it on Transcription Start Site finding.

**Xing Zeng**

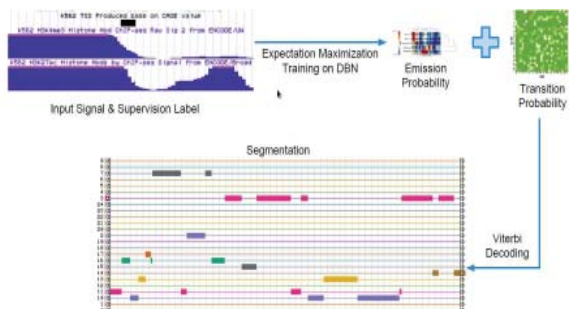
[xpress.zeng@mail.utoronto.ca](mailto:xpress.zeng@mail.utoronto.ca)

**Eric Roberts**

[eroberts@uhnresearch.ca](mailto:eroberts@uhnresearch.ca)

Professor **Michael Hoffman**

[michael.hoffman@utoronto.ca](mailto:michael.hoffman@utoronto.ca)



## DCSUPP

DCSUPP is a student to project matching portal developed to both centralize and simplify the process of matching students with research projects available on campus. DCSUPP is developed as a Single-Page application using the popular web frameworks Angular.js and Ruby on rails.

**Alan Yuan** and **Christina Chung**

[alan.yuan@mail.utoronto.ca](mailto:alan.yuan@mail.utoronto.ca)

Associate Professor, Teaching Stream **Francois Pitt**

[fpitt@cs.utoronto.ca](mailto:fpitt@cs.utoronto.ca)

## ADDITIONAL RESEARCH PROJECTS PRESENTED BY

Christine Murad

## 2015 VIDEO COMPETITION WINNERS

In 2014, we introduced a video competition for all undergraduate summer researchers. The videos are a way for our students to show-off their research and scholarly activities, and overall passion for computer science. We also gain a fun and insightful look in to our students' experiences this summer.

A committee of 4 faculty members and 9 graduate students/Postdoctoral Fellows evaluated and ranked the this year's submissions. The 2015 winners are:

**First place** (\$300 prize)

**Seyed Kamyar Seyed Ghasemipour**

**Second place** (\$200 prize)

**Mohamed Abdalla**

**Third place** (\$100 prize)

**Roman Polyanovsky**

## **COMPUTER SCIENCE**

### **SUMMER RESEARCH INTERNSHIPS**

#### **ABOUT THE PROGRAM**

Computer Science offers our undergraduate summer research interns a unique experience at the University of Toronto. They participate in a summer-long agenda that offers social events, student mentoring and opportunities to present their research.

Interns are funded through NSERC's Undergraduate Summer Research Award (USRA), the University of Toronto Excellence Award (UTEA), and the Undergraduate Toronto Research Experience (UTRECS) programs.

Contact [ug@cs.toronto.edu](mailto:ug@cs.toronto.edu) to learn more.

**DEPARTMENT OF COMPUTER SCIENCE**

University of Toronto  
10 King's College Road  
Toronto, ON M5S 3G4

[www.cs.toronto.edu](http://www.cs.toronto.edu)