



CONTACT US

For more information on any of these projects:

SARA FRANCA

Communications & Liaison Officer Department of Computer Science 10 King's College Road, Room 3302 Toronto, Ontario M5S 3G4 Telephone: 416-978-3619 E-mail: sfranca@cs.toronto.edu Website: www.cs.toronto.edu

OUR NEXT RESEARCH IN ACTION EVENT IS SCHEDULED FOR NOVEMBER 2009. MORE DETAILS WILL BE AVAILABLE IN THE FALL.

RESEARCH IN ACTION RESEARCH HIGHLIGHTS

CONTENTS

- 1 Friend Forecaster: Cell Phone Software Aiding Memory for Names
- 2 Interactive Performance Control -A New Puppetry of Real and Virtual Robots
- **3** Using Language to Learn Structured Appearance Models for Image Annotation
- 4 SketchPad Sketching Interface for Conceptual Design of Floor Plans
- 5 Towards a Comparative Database of Dysarthric Articulation
- 6 Linked Movie Database
- 7 Analyzing Ranking Algorithms on Web Graphs
- 8 Distributed Recommender Systems
- 9 Interactive Synthesis of Motion Using a Generative Model
- 10 Automated Ligand-Based Active Site Alignment
- **11** Restricted Dead-End Elimination: Pruning for a Small Number of Mutations
- 12 How Do Scientists Develop and Use Scientific Software?
- 13 Stylization of Character Motion
- 14 Using a Large High-Resolution Display to Process Daily Work
- 15 SnowFlock
- 16 Audience Measurement Using Computer Vision Techniques
- 17 Proximity-Based Authentication
- 18 A Granular Dynamics Solver
- 19 ILoveSketch
- 20 Buffer Sizing in Internet Routers
- 21 Making User-Submitted Reviews More Useful by Using Novel Extraction and Visualization
- 22 JSCOOP: A High-Level Concurrency Framework for Java
- 23 Video Browsing by Direct Manipulation
- 24 Precise Packet Generator
- 25 Spam Detection in IP Telephony
- **26** Criteria for Problems in Information Retrieval and the Community Extraction Problem
- 27 What's on the Grapevine?
- 28 Network Measurements Using OpenFlow
- 29 Abstract Model Checking for Verification and Refutation
- 30 Extracting Keywords from Documents
- 31 Parallel Building Blocks
- 32 Learning Foreign Language Vocabulary with Contextual Translation
- **33** Customizing the Composition of Actions, Programs, and Web Services with User Preferences
- 34 Using Statistical Information to Solve Logical Constraints
- **35** Recovering Related Artifacts in Software Projects' History: a Comparison of Information Retrieval Based Methods
- **36** A User Interface with Semantic Tactile Feedback for Mobile Touch-screen Devices

MESSAGE from the chair

The Department of Computer Science (DCS) has been a leading international research department for over forty years. Our roughly 60 faculty and 300 graduate students are engaged in cutting-edge investigations at the forefront of numerous areas of information and communications technology. We also place great emphasis on exposure to, and direct engagement in, research projects among our undergraduates, thus integrating research deeply into our curriculum.



We are pleased to highlight a small cross-section of the many fascinating research projects currently underway in DCS. The research described in this booklet gives a sense of the fundamental advances in computing taking place in the department, from new database and networking technologies to new tools for software development or interacting with a variety of media. It also demonstrates the wide-ranging impact of computer science on areas as diverse as document analysis, communication, memory aids, biological discovery, architecture, cloud computing, language learning, and even puppetry. While the projects described here are, to a degree, applications-oriented, faculty and graduate students in DCS also engage in groundbreaking theoretical research that influences the entire discipline. This year, we are also proud to showcase the work of several start-ups that grew out of research conducted in DCS.

Much of our innovative research is shaped by collaborations with research partners in industry and in other academic disciplines – in the sciences, health sciences, social sciences and humanities. Connections and collaborations with industry are especially important to ensure that our research has genuine impact. As part of our commitment to strengthen these ties, I'm extremely pleased to introduce Professor Ravin Balakrishnan, our new Associate Chair for Research and Industrial Relations. Ravin has already had a tremendous impact in bringing together our faculty with potential industrial research partners, and I encourage you to contact him at acrir@cs.toronto.edu or 416-978-8715 to explore ways that we can work together.

On another front, we're developing a new Professional Master's Program, which we plan to roll out in the Fall of 2010. Students in the 16-month program would spend their first eight months taking classes, including new courses in technical entrepreneurship and communication skills, and the subsequent eight months in an industrial internship. We hope that these internships will serve as a catalyst to increase collaboration between our faculty and local industry.

We're proud of our deep interdisciplinary connections and our strong ties with the industrial and business communities, but are always striving to do more. I encourage potential partners to explore the breadth and depth of the internationally renowned research in DCS, and propose novel ways to benefit from our expertise, creativity and drive. Have you ever recognized someone but forgot their name? Friend Forecaster is a new context-aware mobile phone application that uses information such as timing and location to predict who the user will see at a given place and time. Developed for a variety of platforms, Friend Forecaster was developed using both participatory and user-centred design practices, and has

benefited from research conducted with a variety of populations: from businesspeople to senior citizens with clinically-diagnosed memory problems. Early user testing has revealed a variety of need areas and highlighted issues important to the design process.



PROJECT TEAM

Ron Baecker Faculty (rmb@kmdi.utoronto.ca) Kent Fenwick Graduate Student Mike Massimi Graduate Student

3 Using Language to Learn Structured Appearance Models for Image Annotation

Manual annotation of new images in large image collections is prohibitively expensive for commercial databases, and overly time-consuming for the home photographer. However, lowcost imaging, storage and communication technologies have already made accessible millions of images that are meaningfully associated with text in the form of captions or keywords. It is tempting to see these pairings of visual and linguistic representations as a kind of distributed Rosetta Stone from which we may learn to automatically translate between the names of things and their appearances. Our algorithm uses the repetition of appearance across an unstructured collection of captioned

images and a measure of correspondence with caption words to learn to recognize named objects.



PROJECT TEAM

Sven Dickinson Faculty (sven@cs.toronto.edu) Yulia Eskin Undergraduate Student Afsaneh Fazly Postdoctoral Fellow Mike Jamieson Graduate Student Suzanne Stevenson Faculty

A New Puppetry of Real and Virtual Robots

From video games to feature films, the laws of physics play an increasing role in creating the compelling animation modern audiences demand. Animating agile, physically realistic characters, however, is as much art as it is science, whether the characters are inhabitants of fantastical virtual worlds, real robotic competitors running in a soccer stadium, or mechanical family pets.

Current approaches to making such creatures move about nimbly focus on the use of large motion databases, geneticallyinspired learning algorithms, or meticulously crafted, but often brittle, custom programming. In contrast, our approach puts the reins directly into the hands of the animator, who becomes an interactive performer, vicariously flexing a character's muscles or motors, moment-to-moment, to make it run and jump about under its own power, a "stringless puppet" in a potential new performance art form.

We present a variety of simple but compelling examples of animated characters "performed" in this fashion, and argue the approach to be a viable form of physics-based animation, engaging and entertaining to performer and audience alike.



PROJECT TEAM Eugene Fiume Faculty Joe Laszlo Graduate Student (jflaszlo@dgp.toronto.edu)

SketchPad - Sketching Interface for Conceptual Design of Floor Plans

SketchPad is a tool for the conceptual design of architectural floor plans. Our unique sketch-based interface allows an architect to quickly iterate through many sketches early in the design process before progressing to more detailed drawings. The interface of SketchPad is designed for use on a tablet, making use of techniques such as marking menus and gesture recognition to provide a natural interface for users who normally spend much of their time sketching on pen and paper. To allow the user to quickly setup furniture, we provide a catalogue of rooms with accurate product information the user can drop into the scene.



PROJECT TEAM Ravin Balakrishnan Faculty Colin Graham Sketch2 Inc. Karan Singh Faculty Ladislav Stejskal Sketch2 Inc. (ladislav@chandlergraham.com) Eron Steger Graduate Alumnus, Sketch2 Inc. (eron@sketch2.com)

5 Towards a Comparative Database of Dysarthric Articulation

This work describes ongoing data collection of synchronized acoustic and kinematic features of dysarthric speech. Recordings consist of electromagnetic articulographic measurements and 3D reconstructions from video over a variety of prompts read by individuals with cerebral palsy and matching nondysarthric subjects. Preliminary results show a preponderance of mispronounced plosives in dysarthric speech, and greater articulatory variability. An experiment with a standard hidden Markov model for speech recognition resulted in 84% triphone recognition accuracy for non-dysarthric speech, but only 6%

accuracy for individuals with cerebral palsy.



PROJECT TEAM

Graeme Hirst Faculty (gh@cs.toronto.edu) Gerald Penn Faculty Frank Rudzicz Graduate Student Fraser Shein Scientist, Bloorview Research Institute Pascal van Lieshout Director, Oral Dynamics Lab in the Department of Speech Language Pathology at the Faculty of Medicine

Analyzing Ranking Algorithms on Web Graphs

Ranking algorithms bring order to various types of items (such as web pages, pictures and videos) on the World Wide Web, and provide web users with interesting and relevant contents. For the sake of designing efficient ranking algorithms with provable performance guarantees, it is necessary to be able to analyze ranking algorithms on a rigorous basis. We propose a formal framework for the analysis of ranking algorithms in terms of their ability to discover the correct rankings of the online items. With this framework, it becomes possible to formally investigate the subtleties underlying ranking algorithms as well as the natural structure of the WWW.



Linked Movie Database

Movies are highly popular on the Web. There are several web resources dedicated to movies and many others containing movie-related information. Creating a single source of information about movies that contains information from existing open web data sources and links to other related data sources is a challenging task and the goal of the Linked Movie Data Base (LinkedMDB) project. LinkedMDB provides a demonstration of the first open semantic web movie data source connecting several major existing (and highly popular) movie web resources. LinkedMDB uses a novel way of creating and maintaining large quantities of high-quality links by employing state-of-the-art

approximate join techniques for discovering links, and providing additional metadata about the quality of the links and the techniques used for deriving them.



PROJECT TEAM

Mariano Consens Faculty, Computer Science and Mechancial & Industrial Engineering Oktie Hassanzadeh Graduate Student (oktie@cs.toronto.edu)

8 Distributed Recommender Systems

Inspired by how people interact in a social network to spread information, we aim to design a distributed recommender system in which every user links to a controllable number of "friends", and recommendations are made by friends forwarding queries and replies. We have produced theoretical results which provide insights on the tradeoff between the number of friends and the depth of queries. The next step is to develop an application based on the insights gained from the theoretical analysis.



6

9

The use of 3D character animation in the entertainment industry has exploded over the past decade. The dominant methods for creating 3-D animations are to either manually "keyframe" a character using software tools or acquire data from real humans or animals using a process called motion capture. Unfortunately both of these methods are time-consuming and expensive. We propose a method for the synthesis of novel high-quality motion using a probabilistic model that has been trained from motion capture data. Our model is able to synthesize and transition between diverse motion styles while responding to stylistic cues provided by the user.



PROJECT TEAM

Geoff Hinton Faculty Sam Roweis Faculty Graham Taylor Graduate Student (gwtaylor@cs.toronto.edu)

| 1 | Restricted Dead-End Elimination: |
|---|---|
| | Pruning for a Small Number of Mutations |

Novel biological function can be engineered into a target protein with the use of computational algorithms for protein redesign. Given a protein with N mutable residues, the aim of the redesign algorithms is to search for the lowest energy protein sequence. Such a solution can contain as many as N mutations compared to the wild type sequence. As one measure of success for a computational redesign experiment is the ease of wet-lab preparation of the suggested sequence, a minimal number of mutations in the final solution is desirable to aid preparation by site-directed mutagenesis. As for most redesigns, the number of mutable resides N can be quite large, solutions suggested by existing protein design algorithms can pose problems during wetlab synthesis. We present a novel solution for protein redesign where the number of mutations in the target can be restricted, thus preventing wet-lab synthesis problems. Our algorithm is provably correct and avoids redundant computation inherent to traditional redesign algorithms when an upper bound on the number of allowed mutation exists.

10 Automated Ligand-Based Active Site Alignment

State-of-the-Art drug discovery relies on a precise understanding of how a drug and its target protein interact. Unfortunately, it is typically difficult to determine the most critical features of these interactions.

Fortunately, patterns may emerge when we compare the interactions of several different proteins with the same drug. The goal of this project is to automate these comparisons by developing freely available, easy to use visualization software. We demonstrate LigAlign, a PyMOL-based, structure visualization tool, which performs drug-based protein alignments.



PROJECT TEAM Abraham Heifets Graduate Student Ryan Lilien Faculty (lilien@cs.toronto.edu)

12 How Do Scientists Develop and Use Scientific Software?

New knowledge in science and engineering relies increasingly on results produced by scientific software. Therefore, knowing how scientists develop and use software in their research is critical to assessing the necessity for improving current development practices and to making decisions about the future allocation of resources. We, as software engineering researchers, want to understand this community to better help them achieve their goals.

There is a wide chasm between the general computing community and the scientific computing community. As a result, there has been little exchange of ideas relevant to scientific application software. Much of scientific computation research has focused on high performance computing, to the exclusion of all else. This ongoing research project seeks to give a better understanding of the needs and practices of scientists with respect to software development. A large-scale online survey in late 2008 has given us an initial look at the type of software development scientists take part in. Our follow-up work seeks to better understand scientists conceptions of software development and reproducibility.

PROJECT TEAM Carolyn MacLeod Graduate Student (carolyn.macleod@gmail.com) Greg Wilson Faculty Most animation tools allow users to do amazing work, but they require animators to work creatively in terms of how the software thinks, rather than how they think. We are looking at a variety of approaches to formally model aesthetic concepts used by animators when they plan, talk about, create, edit, or analyze movement. Using these formal models we can create new character animation tools that let users work with motion the way they think about it, allowing them to make large stylistic changes without having to express them in terms of the algorithms used by common software tools.



14 Using a Large High-Resolution Display to Process Daily Work

With the ever increasing amount of digital information, users desire more screen real estate to process their daily computing work, and might well benefit from using a wall-size large high-resolution display instead of a desktop one. Unfortunately, we know very little about users' behaviors when using such a display for daily computing. Moreover, current software UIs are designed based on normal desktop screens, which are unsuited to a large display. We conducted a longitudinal study that investigates large display use in a personal desktop computing context by comparing it with single and dual desktop monitor use. Results show users' unanimous preference for using a large display, and reveal distinct usage patterns in partitioning screen real estate and managing windows. Based on the study,

a windows management prototype, called WallTop, is designed and implemented to facilitate windows management tasks on a large display.



CognoVision

PROJECT TEAM Ravin Balakrishnan Faculty Xiaojun Bi Graduate Student (xiaojun@dgp.toronto.edu)

16 Audience Measurement Using Computer Vision Techniques

Online retailers and advertisers have a rich set of tools that allow them to measure how well their websites and banner ads are performing, to sell advertising based on the number of actual impressions, and the power to target content based on audience demographics. Unfortunately, this same level of measurability and accountability is currently missing in the traditional brickand-mortar world of business. CognoVision, a start-up company founded by DCS alumni, has developed a suite of applications that address this need. Using sophisticated computer vision object detection and tracking algorithms, our Anonymous Impression Metric (AIM) software can provide retailers and advertisers with the ability to collect data, such as counting the number of visitors entering and leaving a store and extracting visitor demographics such as gender and age bracket.

In addition to providing businesses with a rich set of data about their visitors, the AIM technology ensures that the data detected and aggregated is anonymous, to preserve individuals' privacy.

PROJECT TEAM

Phil Hubert CognoVision Faizal Javer CognoVision Shahzad Malik Graduate Alumnus, CognoVision (shahzad.malik@cognovision.com) Haroon Mirza CognoVision Umesh Patel CognoVision Abhishek Ranjan Graduate Alumnus, CognoVision

PROJECT TEAM Patrick Coleman Graduate Student (patrick@dgp.toronto.edu) Karan Singh Faculty

15 SnowFlock

Scientists and engineers are developing applications that interact with terabyte-sized datasets, yet their users are constrained by the inability to purchase and maintain the equipment necessary to use these applications. Cloud computing offers an attractive solution to this problem, but developing applications for cloud environments is not a trivial task. SnowFlock introduces a novel paradigm, VM fork, that leverages virtual machine (VM) technology to revolutionize the way in which applications harness resources in a cluster. VM fork is a cloud computing abstraction that instantaneously clones a VM into multiple replicas running on different hosts. SnowFlock provides sub-second VM cloning, scales to hundreds of workers, consumes few cloud I/O resources, and has negligible runtime overhead. All replicas share the same initial state, matching the intuitive semantics of stateful worker creation. This paradigm thus enables the straightforward creation and efficient deployment of many applications capable of

PROJECT TEAM

Michael Brudno Faculty

exploiting cloud resources on demand, found in fields such as bioinformatics, finance, and rendering.

Roy Bryant Graduate Student Eyal de Lara Faculty (delara@cs.toronto.edu) Olga Irzak Graduate Student Andres Lagar-Cavilla Graduate Student Charlotte Lin Undergraduate Student Philip Patchin Undergraduate Student Steve Rumble Undergraduate Alumnus Mahadev Satyanarayanan Faculty, Carnegie-Mellon University Adin Scannell Graduate Student Bianca Schroeder Faculty Joe Whitney Graduate Student Securing interactions between devices that do not know each other a priori is an important challenge for future ubiquitous computing applications. Our system, Amigo, uses physical proximity as an intuitive way to provide authentication. We exploit devices' shared radio environment at a particular location and time in order to determine proximity. We are also exploring leveraging multiple co-operating devices in order to improve accuracy and provide richer location information. These techniques do not require any additional hardware to be present on the devices beyond the radios, do not require any user involvement and are not vulnerable to eavesdropping.



PROJECT TEAM

Eyal de Lara Faculty (delara@cs.toronto.edu) Andre Kalamandeen Graduate Student Anthony LaMarca Intel Research Seattle Adin Scannell Graduate Student Anmol Sheth Intel Research Seattle

19 ILoveSketch: As-natural-as-possible sketching system for creating 3D curve models

ILoveSketch is a 3D curve sketching system that captures some of the affordances of pen and paper for professional designers, allowing them to iterate directly on concept 3D curve models. The system coherently integrates existing techniques of sketchbased interaction with a number of novel and enhanced features. Novel contributions of the system include automatic view rotation to improve curve sketchability, an axis widget for sketch surface selection, and implicitly inferred changes between sketching techniques. We also improve on a number of existing ideas such

as a virtual sketchbook, simplified 2D and 3D view navigation, multi-stroke NURBS curve creation, and a cohesive gesture vocabulary.



A Granular Dynamics Solver

Granular dynamics is the dynamics of a large set of small particles/ grains. Convincing simulation of natural granular phenomena (e.g. dust, sand or powders) is a challenging mathematical and computational problem. Our observation is that the more realistically the collection of grains approaches its steady state, the more natural the simulation appears. This work examines the limitations and merits of using a molecular dynamics approach in a general granular dynamics solver. The solver is qualified by its ability to form stable piles under various formation conditions.



PROJECT TEAM

Christina Christara Faculty (ccc@cs.toronto.edu) Jeff Lait SideFX Software Tina Yee Graduate Student

20 Buffer Sizing in Internet Routers

Internet routers require buffers to hold packets during times of congestion. The widely used rule-of-thumb for determining the size of these buffers results in very large buffers. There has been recent suggestions of significantly reducing the size of buffers to few dozen of packets under certain conditions. In this work, we are investigating the gap between large and tiny buffers and offering a new approach to estimate the required buffer size in an existing network.



PROJECT TEAM Yashar Ganjali Faculty (yganjali@cs.toronto.edu) Monia Ghobadi Graduate Student Geoff Salmon Graduate Student

PROJECT TEAM Seok-Hyung Bae Postdoctoral Fellow (shbae@dgp.toronto.edu) Ravin Balakrishnan Faculty Karan Singh Faculty

21 Making User-Submitted Reviews More Useful by Using Novel Extraction and Visualization

Online review websites such as Yelp.com and TripAdvisor.com are exploding in popularity. Tens of millions of people regularly visit these sites and write reviews about hotels where they've stayed or restaurants where they've eaten. However, with more and more content to read we've found that people rarely read all of the reviews about a venue and rely on subjective average ratings as well as the small number of reviews they do read to make a decision. We're studying ways of summarizing review text in such a way to increase the usefulness of online reviews. We're looking beyond more traditional summarizing techniques and exploring tag clouds, noun adjective pairs, and word sentiment to create a qualitative visualization of words that accurately represent the opinions about a venue. Preliminary research we performed on how users currently use review websites motivates this project and further studies are planned to fine tune our technique.

2 JSCOOP: A High-Level Concurrency Framework for Java

JSCOOP is a minimal extension to the sequential objectoriented programming model for concurrency. The extension consists of one keyword (separate) that avoids explicit thread declarations, synchronized blocks, explicit waits, and eliminates data races and atomicity violations by construction through a set of compiler rules. It attempts to guarantee fairness via use of a global scheduler. We are developing a new implementation of SCOOP for Java, called JSCOOP. JSCOOP introduces a new set of annotations modeled after SCOOP keywords, as well as several core library classes which provide the support necessary to implement the SCOOP semantics. A prototype Eclipse plug-in

allows for the creation of JSCOOP projects. The fact that JSCOOP programs are annotated Java programs means we can exploit existing Java formal methods tools, particularly JML and Java Pathfinder.



PROJECT TEAM

Michael Novati Graduate Student (michael.novati@utoronto.ca) Khai Truong Faculty Koji Yatani Graduate Student

Video Browsing by Direct Manipulation

We present a method for browsing videos by directly dragging their content. This method brings the benefits of direct manipulation to an activity typically mediated by widgets. We support this new type of interactivity by: 1) automatically extracting motion data from videos; and 2) a new technique called relative flow dragging that lets users control video playback by moving objects of interest along their visual trajectory. We show that this method can outperform the traditional seeker bar in video browsing tasks that focus on visual content rather than time.



PROJECT TEAM

Marsha Chechik Faculty Jonathan Ostroff Faculty, York University Faraz Torshizi Graduate Student (faraz@cs.toronto.edu)

Precise Packet Generator

We need realistic network traffic for testing network devices as well as various network algorithms, and protocols. However, generating realistic traffic in the context of large networks -especially the Internet -- is extremely difficult since in any testing environment a limited number of computers and network devices are available. The objective of this project is to emulate the traffic generated by a large number of senders using commodity computers and devices. The Precise Packet Generator ensures accurate order and timing of packet transmissions using specialized network hardware, and provides researchers the flexibility to perform realistic tests exploring a wide range of network traffic conditions.

PROJECT TEAM

Ravin Balakrishnan Faculty (ravin@dgp.toronto.edu) Jacky Bibliovic Graduate Student Pierre Dragicevic Graduate Alumnus, INRIA Derek Nowrouzezahrai Graduate Student Gonzalo Ramos Graduate Alumnus, Microsoft Live Labs Karan Singh Faculty In this project we address the problem of detecting and preventing Spam over Internet Telephony (SPIT). As price of Voice Over IP (VoIP) is going down, SPIT is expected to become a serious problem in the near future. It has the potential to become even a bigger problem than email spam, because recipients will be disturbed by each received SPIT call. We are exploring properties of SPIT calls from conversations to detect them and use these properties to prevent new SPIT calls before interrupting users.



Online discussion groups provide a rich source of information for better understanding opinions and trends in products, services, politics, etc. We study the community extraction problem within the context of networks of blogs and forums. We assume the existence of a small set of known seed nodes that accurately represent good and bad examples of sites within a community (e.g. the community of people interested in hybrid SUVs). We study the use of content information (beyond explicit link information) in the identification of the relevant community. Our approach lends itself to a new and insightful ranking scheme for members of the extracted community and an efficient algorithm for inflating/deflating the extracted community. We introduce a testable measure of quality that can be used to complement human evaluation. Using a considerably large commercial data set of blog and forum sites, we provide experimental evidence to demonstrate the utility, efficiency, and stability of our methods.

PROJECT TEAM

Hossein Bokharaei Graduate Student (hossein@cs.toronto.edu) Yashar Ganjali Faculty Alireza Sahraei Graduate Student

What's on the Grapevine?

27

Ever wondered what the world is talking about? Grapevine is an online system that conducts large scale data analysis on the social media collective, extracting and distilling information in real time from millions of blog posts, thousands of tweets, news items, etc. It automatically identifies stories, emerging events, popular or surprising entities narrowed down to any demographic of interest (e.g., "What are male teens talking about in Toronto?" or "What were financiers in Texas blogging

about during the previous week?"). Stories can be interactively explored in a variety of ways, such as modifying their scope, obtaining related content and examining their temporal evolution.



PROJECT TEAM

Allan Borodin Faculty (bor@cs.toronto.edu) Hyun Chul Lee Graduate Alumnus, Thoora/Rogers Leslie Goldsmith Affinity Systems

28 Network Measurements Using OpenFlow

We present a system for performing network measurements to support a wide-range of traffic engineering tasks in an OpenFlow enabled network. OpenFlow is added to networking devices to provide a standardized hook to allow researchers experiment with production networks without having vendors expose internals of their devices. We show how OpenFlow can make complicated traffic engineering tasks much simpler to do. Specifically, we show how we can derive the traffic matrix of a network in real time using direct measurements and also show how it can be leveraged to run a more robust network.

PROJECT TEAM

Albert Angel Graduate Student Nick Koudas Faculty (koudas@cs.toronto.edu) Nikos Sarkas Graduate Student Divesh Srivastava AT&T Labs PROJECT TEAM Yashar Ganjali Faculty Monia Ghobadi Graduate Student Amin Tootoonchian Graduate Student (amin@cs.toronto.edu) In today's computer-dependent society, we often rely on computers to control safety critical, and mission critical tasks. Ensuring the correctness of these computer systems has become an important challenge. Model checking is a technique to check whether a given hardware or software system satisfies properties of interest. Compared to other techniques, e.g., testing and deductive methods, model checking provides a full coverage of system behaviors and can be performed fully automatically.

The bottleneck of model checking is the state-explosion problem that occurs in the analysis of systems with a large or infinite number of states. To address this problem, we have investigated abstraction approaches for model checking that can prove and disprove properties over a smaller abstract model with equal effectiveness. Our results include abstract model checking of symmetric systems, a symbolic software model checker - Yasm, and analysis of partial modeling formalism used for abstraction, which increases the scope of systems where model checking can be effectively applied. "Keywords" are commonly used to summarize documents. Most people conduct searches on the web by typing in a few "keywords" that best describe what they are looking for. Information retrieval theory suggests using the TFIDF and/or BM25 functions to determine the keywords for a given document. These offer high performance but in some situations fail to pick out keywords that could help us understand the document even better. We use a Restricted Boltzmann Machine to improve on these methods. Our method is able to obtain, for example, keywords that are not actually present in the document.

PROJECT TEAM

Marsha Chechik Faculty Arie Gurfinkel SEI, CMU Ou Wei Graduate Student (owei@cs.toronto.edu)

31 Parallel Building Blocks

Parallel Building Blocks (PBB) is an open source C++ library we have developed to distill parallelism to its essence. It extends the open source Threading Building Blocks project from Intel (a high-performance C++ library for multi-core systems). Programs using PBB can run on multi-core computers and clusters, and utilize GPU cards. The significance of the PBB library is that programs can be adapted to any type of parallel architecture with minimal changes to source code.



PROJECT TEAM Geoff Hinton Faculty Nikola Karamanov Undergraduate Student (n.karamanov@utoronto.ca)

32 Learning Foreign Language Vocabulary with Contextual Translation

Students learning foreign language vocabulary primarily learn new words by rote memorization. Memorization, although effective, is a tedious method that does not effectively convey the context in which words are used. We have developed a prototype that teaches vocabulary in-context by transforming a student's everyday web browsing experience into a language learning environment. Our prototype selectively translates parts of every web page into the foreign language being learned such that the student reading the page can learn vocabulary using contextual hints provided by the untranslated words. We believe that our protoype will prove effective as a language learning tool for students at all levels.

PROJECT TEAM

Scott Briggs Graduate Student, Environmental Engineering Adrien Guillon Undergraduate Student (aj.guillon@utoronto.ca) Giovanni Grasselli Assistant Professor, Associate Director of the Lassonde Institute, Department of Civil Engineering Igor Jurisica Faculty, Computer Science and Senior Scientist, Ontario Cancer Institute, University Health Network Greg Wilson Faculty

PROJECT TEAM Khai Truong Faculty Andrew Trusty Graduate Student (atrusty@cs.toronto.edu)

33

Customizing the Composition of Actions, Programs, and Web Services with User Preferences

The goals and preferences of individuals, customers, and organizations all play a significant role in human decision making. As such, it comes as no surprise that preferences also play a significant role in AI automated planning, providing a means of customizing the generation of plans to ensure they are of high quality. Most recently our work has focused on the use of preference-based planning techniques for composition and customization of software components. Increasingly, corporations are providing services within and between organizations by publishing programs on corporate intranets or on the World Wide Web. Many of these programs represent component software that can be composed together either manually or automatically to provide value-added service. To address the task of automated software composition, we have developed means of composing software through the use of composition templates or scripts that sketch the user's firm objectives, specifying user's customizing preferences, and attempting to generate an optimal composition with respect to this specification.

PROJECT TEAM

Jorge Baier Graduate Student Sheila McIlraith Faculty Shirin Sohrabi Graduate Student (shirin@cs.toronto.edu)

35

Recovering Related Artifacts in Software Projects' History: a Comparison of Information Retrieval Based Methods

The average software project is made up of a rich history of artifacts or change events: discussions in mailing lists or forums, bug reports, source code revisions, documentation or wiki page edits, patches, and feature or support requests. Although each of these artifacts are independent, many events may be logically related. For example, an email in a project's mailing list archives may discuss a related bug report. Similarly, the bug report may be fixed by a source code revision in version control repository, and this may cause a change in the wiki pages too.

We present a system to automatically cluster related events into logical groupings, or find the artifacts related to a target artifact. In addition, we present a visualization tool to help project stakeholders explore these artifacts, which also let them to further expand their search to retrieve more related artifacts in a tree based fashion. We provide a comparison of three Information Retrieval Based Methods: Vector Space Model (VSM), Latent Semantic Indexing (LSI), and a novel approach based on approximating the most distinguishing words or phrases considering the properties of software repositories.

Using Statistical Information to Solve Logical Constraints

Boolean Satisfiability (SAT) and Constraint Satisfaction (CSP) problems are abstract logical systems that form the core of complexity studies in computer science. At the same time, they serve to represent a variety of real-world problems involving constrained systems, like planning and scheduling, software and systems verification, and even new applications in bioinformatics. In real applications, then, problems can exhibit hidden structure that allow us to overcome negative theoretical runtime bounds that apply to SAT and CSP in general. In this work, we apply statistical techniques to finds such structure, at times exponentially increasing the efficiency of solvers for these problems.

PROJECT TEAM

Chris Beck Faculty, Mechanical & Industrial Engineering Eric Hsu Graduate Student (eihsu@cs.toronto.edu) Sheila McIlraith Faculty Christian Muise Graduate Student

36 A User Interface with Semantic Tactile Feedback for Mobile Touch-screen Devices

One of the interaction problems with mobile touch-screen devices is that they do not provide tactile feedback to the users. Thus, users are required to look at the screen to interact with the devices. In this project, we explore a tactile feedback system which offers semantic information about objects users are touching on the screen. Multiple actuators are embedded in a mobile touch-screen device, and the system can generate different patterns of vibration, such as from right to left, and from top to bottom. These different vibration patterns will enable users to understand which object they are touching even without looking at the screen.

