A Message from the Computer Science Department Chair

Welcome to UWORCS'06. Once again this year the Department is pleased to run UWORCS (University of Western Ontario Research in Computer Science).

UWORCS is an opportunity for the Department to highlight the research of its graduate students and it is a wonderful opportunity for all faculty, staff and students to learn a little about the research being done by their colleagues. It is also a chance for our students to gain experience in preparing presentations and speaking in a conference-like environment.

We have over 30 students presenting their research this year and once again UWORCS will have parallel sessions. We have a diverse range of topics being covered this year, including artificial intelligence, databases, distributed and grid systems, human-computer interaction, imaging, software engineering, symbolic computation, and theory.

I would like to thank Dr. Mahmoud El-Sakka, Kehinde Oladosu and Ben Stephenson for their efforts at organizing the conference, the solicitation of abstracts, preparation of these proceedings, and other details associated with the conference organization. Preparing for the conference takes time and effort. They have done an excellent job and I congratulate them on their efforts.

These proceedings provide a useful summary of the work of the students. I hope you will read through some or all of the abstracts, especially ones where you were not able to attend the presentation. I think you will find the work to be of high quality, diverse and interesting.

Michael Bauer
Department Chair
Computer Science Department

A Message from the Computer Science Graduate Chair

Welcome to the new edition of the University of Western Ontario Research in Computer Science conference, UWORCS'06.

This year, we received 32 abstracts and we will have three parallel technical sessions in the morning and another three in the afternoon.

The collection of topics in UWORCS'06 is a good representation of the most recent research activities at the Computer Science Department. Due to the diversity and quality of research being presented at the conference, the Computer Science Department has decided to increase the number of awarded prizes from five to six, one in each technical session. I would like to thank the volunteer judges for their time and effort in evaluating the presented research.

Similar to previous years, it has been an interesting and enjoyable experience to see the pieces of this conference brought together. I wish to express my thanks and appreciation to Ben Stephenson and Kehinde Oladosu who put a lot of time and effort towards organizing this conference.

I am sure that the UWORCS'06 conference will be exciting and I hope that you will enjoy it. I look forward to seeing you all there.

Mahmoud El-Sakka
Graduate Chair
Computer Science Department

A Message from the UWORCS'06 Conference Chairs

It is our delight to welcome everyone to UWORCS'06 conference held annually by the department of Computer Science at the University of Western Ontario.

The importance of this conference to graduate students and the faculty members within the department cannot be over emphasized. It serves as the focal point of most of the research work carried out within the department annually, and as a day where all graduate students and faculty members can exchange ideas.

This year, we are proud to have a total of 32 talks during the conference. These talks include interesting topics reflecting the various research areas of the department. It is also delightful to note that the number of entries this year continues the upward trend in the number of entries received annually.

We will like to thank all the judges and session chairs who have volunteered to adjudicate the talks and moderate the sessions. A special thank you goes to Dr. Mahmoud El-Sakka and other members of the GEC for their efforts during the preparation of this conference.

Once again, you are welcome to UWORCS'06.

Kenny Oladosu
Conference Co-Chair
Computer Science Department

Ben Stephenson
Conference Co-Chair
Computer Science Department

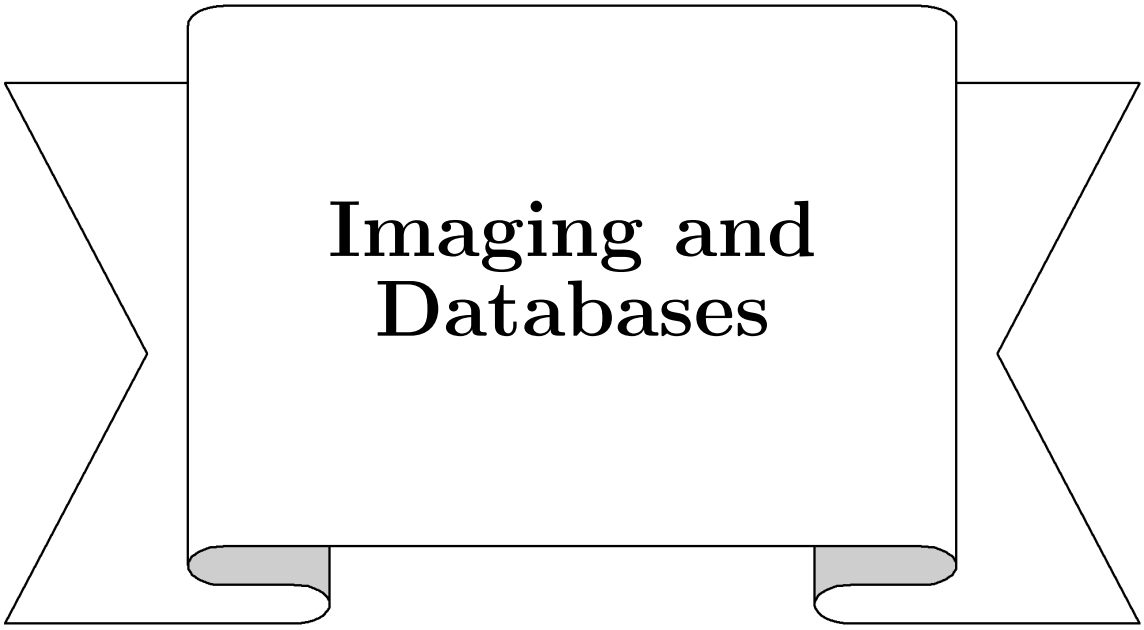
Session Schedule

Time	Session	Location
9:00am – 9:30am	Opening Session	MC320
9:50am – 11:30am 9:30am – 11:30am 9:30am – 11:10am	Imaging and Databases Software Engineering Symbolic Computation	MC300 MC316 MC320
11:30pm – 12:30pm	Lunch Break	
12:30pm – 2:10pm 12:30pm – 2:30pm 12:30pm – 2:10pm	Artificial Intelligence and Human Computer Interaction Distributed and Grid Systems Theory of Computation	MC300 MC316 MC320
2:30pm – 3:00pm	Break	
3:00pm – 3:15pm	Closing Session	MC320

Contents

Contents

Contents



Imaging and Databases

Time	Title	Location
9:50am-10:10am	Skeleton-Based Hook Echo Detection in Doppler Radar Precipitation Density Imagery	MC300
10:10am-10:30am	Carotid Artery Ultrasound Image Segmentation Using Fuzzy Region Growing	
10:30am-10:50am	Semiautomatic Segmentation with Compact Shape Prior	
10:50am-11:10am	Delegation in the Administrative Role Graph Model	
11:10am-11:30am	An Efficient Algorithm for Identifying the Most Contributory Substring	

Skeleton-Based Hook Echo Detection in Doppler Radar Precipitation Density Imagery

Hongkai Wang, John Barron and Bob Mercer

Doppler radar has been successfully used in weather forecasting and severe storm prediction for many years. Meteorologists can manually detect tornadoes in Doppler imagery. Due to the large size of Doppler datasets, it would be helpful if algorithms could be designed to assist with analyzing these datasets automatically, efficiently and accurately. In this thesis, we identify tornadoes from Doppler radar imagery by detecting hook echoes. A hook echo is an important signature of tornadoes in radar reflectivity imagery. The medial axis representation, also known as a skeleton, is employed as the shape descriptor of storms. Using four properties of hook echoes, curvature, orientation, thickness variation and boundary proximity, hook echoes are modelled. We apply our hook echo detection algorithm on the radar datasets collected from one tornado outbreak in central Oklahoma, May 1999.

Carotid Artery Ultrasound Image Segmentation Using Fuzzy Region Growing

Amr R. Abdel-Dayem and Mahmoud R. El-Sakka

We propose a new scheme for extracting the contour of the carotid artery using ultrasound images. Starting from a user defined seed point within the artery, the scheme uses the fuzzy region growing algorithm to create a fuzzy connectedness map for the image. Then, the fuzzy connectedness map is thresholded using a threshold selection mechanism to segment the area inside the artery. Experimental results demonstrated the efficiency of the proposed scheme in segmenting carotid artery ultrasound images, and it is insensitive to the seed point location, as long as it is located inside the artery.

Semiautomatic Segmentation with Compact Shape Prior

Piali Das

We present a semiautomatic segmentation algorithm, that can segment an object of interest from its background. Though not fully automated, our approach requires minimum guidance from the user, who just has to select a single seed pixel inside the object of interest. To obtain a reliable and robust segmentation with such low user interaction, we have to make several assumptions. Our main assumption is that the object to be segmented is of compact shape, which can be relaxed a bit to segment objects of more general shapes. We base our work on the powerful Graph Cut segmentation algorithm of Boykov and Jolly. In order to make the Graph Cut approach suitable for our low user interaction framework, we address several well-known issues in their work. First we show how to incorporate the compact shape prior as a hard constraint. The additional benefit of incorporating the compact shape prior is that a parmeter biasing segmentation towards larger objects can be introduced into the Graph Cut framework. This is a very important parameter, as it helps to counteract the well known bias of Graph Cut based segmentation algorithm to shorter segmentation boundaries. Segmentation results are quite sensetive to the choice of the bias parameter, and so another contribution of our paper is that we show how to select the bias parameter automatically. We demonstrate the effectiveness of our method on the challenging industrial application of transistor gate segmentation in an integrated chip, for which it produces highly accurate results in real-time.

Delegation in the Administrative Role Graph Model

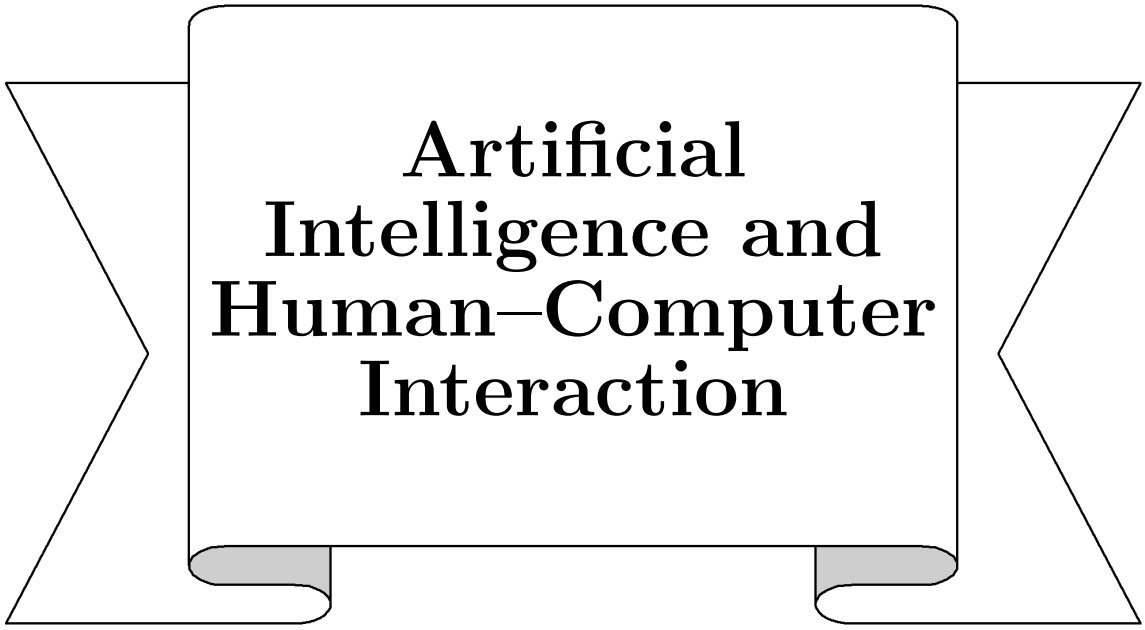
He Wang and Sylvia Osborn

Delegation has received lot of attention in the literature. In this paper, we present a delegation role graph model that is based on our decentralized administrative role graph model. We use the combination of user/group assignment and user-role assignment to support user to user, permission to user and role to role delegation. A powerful source-dependent revocation algorithm is described. We separate our delegation model into a static model and a dynamic model, then define and discuss the static model and its operations. We provide detailed partial revocation operations and algorithms. We also give the details about the changes of role hierarchy, user/group structure and RBAC operations that are affected by delegation.

An Efficient Algorithm for Identifying the Most Contributory Substring

Ben Stephenson

Detecting repeated portions of strings has important applications to many areas of study including data compression and computational biology. This talk defines and presents a solution for the Most Contributory Substring Problem. The goal of this problem is to identify the single substring that represents the largest proportion of the characters within a set of strings. We show that a solution to the problem can be achieved with an $O(n \log k)$ running time (where n is the total number of characters in all of the input strings and k is the number of input strings) when overlapping occurrences of the most contributory substring are permitted. Furthermore, we present an extended algorithm that does not permit occurrences of the most contributory substring to overlap. The expected running time of the extended algorithm is $O(n \log n \log(j + k))$ where j is the size of the alphabet and k is the number of strings in the input set.



Time	Title	Location
12:30pm-12:50pm	Test Strategies for cost-sensitive learning	
12:50pm-1:10pm	Can Machine Learning Challenge the Efficient Market Hypothesis?	
1:10pm-1:30pm	High Accuracy and Low Storage Prefetching	MC300
1:30pm-1:50pm	On the use of Classifier Committees for Automated Text Categorization Tasks	
1:50pm-2:10pm	Towards a framework of navigation interaction in computer-based learning environments	

Test Strategies for cost-sensitive learning

Shengli Sheng and Charles X. Ling

In medical diagnosis doctors must often determine what medical tests (e.g., X-ray, blood tests) should be ordered for a patient to minimize the total cost of medical tests and misdiagnosis. In this paper, we design cost-sensitive machine learning algorithms to model this learning and diagnosis process. Medical tests are like attributes in machine learning whose values may be obtained at a cost (attribute cost), and misdiagnoses are like misclassifications which may also incur a cost (misclassification cost). We first propose a lazy decision tree learning algorithm that minimizes the sum of attribute costs and misclassification costs. Then we design several novel test strategies (that can request to obtain values of unknown attributes at a cost (similar to doctors ordering of medical tests at a cost) in order to minimize the total cost for test examples (new patients). These test strategies correspond to different situations in real-world diagnoses. We empirically evaluate these test strategies, and show that they are effective and outperform previous methods. Our results can be readily applied to real-world diagnosis tasks. A case study on heart disease is given throughout the paper.

Can Machine Learning Challenge the Efficient Market Hypothesis?

Jun Yan, John Nuttall and Charles X. Ling

The Efficient Market Hypothesis (EMH) has several forms. A commonly believed weak form of the EMH hypothesizes that the future stock price is completely unpredictable given the past trading history of the stock. Recent research suggests that there exists a predictable component in past trading information. However, data snooping is used, or the profit after the trading costs is not evident. With the rapid development of powerful computers and effective machine learning algorithms, we can now build adaptive models from a tremendous amount of raw data points of past stock trading history to possibly extract weak regularities for predicting future stock price. We describe an adaptive stock portfolio selection method based on machine learning, and show that it produces a clear profit improvement compared to other approaches. After taking into account reasonable trading costs, our model can still make a sizable profit over the wide range period of 1978 to 2004. The results seem to seriously challenge the weak form of EMH.

High Accuracy and Low Storage Prefetching

Qinghui Liu

Prefetching is one of the most effective techniques to reduce user-perceived Web latency. Web prefetching predicts the future Web requests of a user, fetches the corresponding objects from the Web, and stores them in a local cache. If the predictions are correct and the prefetched objects are requested in the near future, such requests will be served from the cache; therefore, the user will not need to wait for the object to be transferred from remote Web locations.

The Prediction-by-Partial-Matching (PPM) algorithm is widely recognized as one of the best known prefetching algorithm. It predicts users' future requests based on their previous Web visiting behavior. PPM stores every user's visiting history in a data structure called the History Structure. As more data is stored in it, the size of the History Structure can grow rapidly to an impractical size.

Many researchers have proposed different variation of the PPM algorithm to reduce the size of the History Structure, while minimizing the loss of prediction accuracy. Compared with the existing methods, our History Structure Cache PPM (HSCPPM) technique is strongly competitive in the following two regards. First, it is the first method employing a limited-size cache to contain the History Structure. This allows users to preset the maximum size of the History Structure and efficiently manage precious memory resources. With our cache management system, only the information that is most useful for predictions is kept in the cache.

Second, our technique is the first one that improves the prediction method of the PPM algorithm. This improvement is achieved by including a new factor which affects the prediction performance when predicting users' requests.

In addition, we have also developed a technique to upper bound the maximum prediction accuracy of any history-based prefetching technique. Our experiments show that the performance of the HSCPPM technique is very close to the upper bound.

On the use of Classifier Committees for Automated Text Categorization Tasks

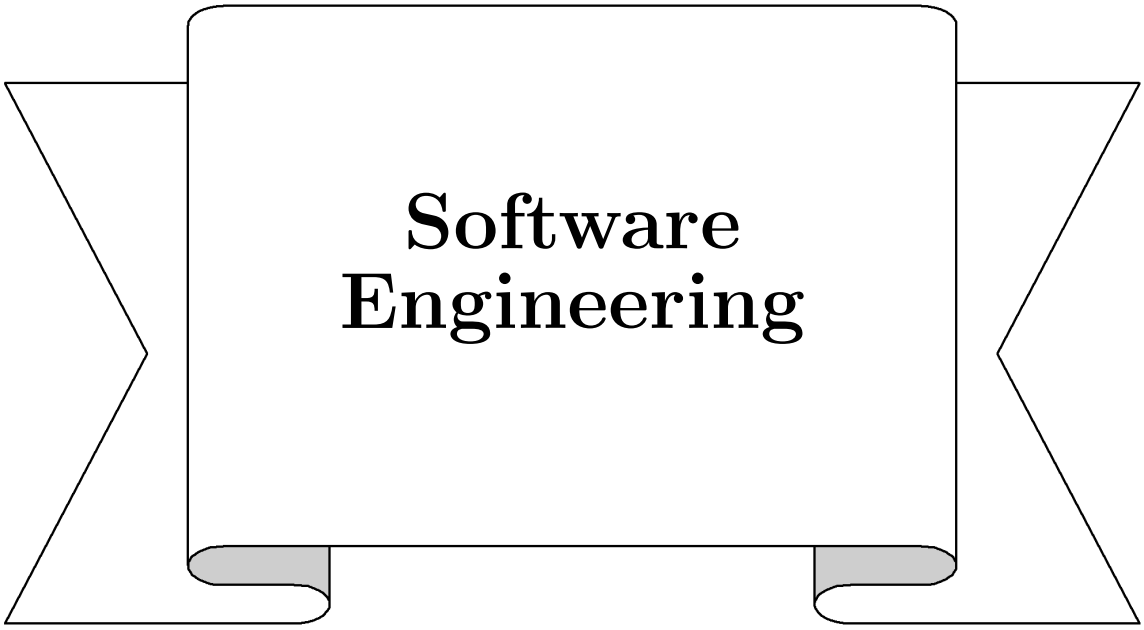
Jeff Taylor

Automated text categorization refers to the use of machine learning techniques to apply predefined categorical labels to natural language texts. A classifier committee is a machine learning technique in which a number of different classifiers are applied to the same problem, and their outputs are combined in an appropriate way. Classifier committees have been used to perform automated text categorization with mixed results. To date, no comprehensive studies have been done about the different ways that the individual members of a committee can be built for an automated text categorization task. To that end, we built several different classifiers by using different indexing approaches and inductive methods. These classifiers were then used in several different classifier committees and tested on the Reuters corpus, a well-known corpus in text categorization tasks. Here, we present the results of our experiments.

Towards a framework of navigation interaction in computer-based learning environments

Hai-Ning Liang and Kamran Sedig

Computer-based learning environments (CBLEs) can assist learners in achieving different learning goals. Learners are often required to interact with the visual representations used within CBLEs. One common, fundamental form of interaction is navigation. Navigation allows learners explore and learn the properties of an information space and its embedded objects. There are three models of information navigation that can be used in CBLEs: Spatial, semantic, and social. A CBLE can provide these three types of navigation to support learners exploration of its information space. An information space is mapped on and displayed in the interface using visual representations. Based on the overall representational framework used, a CBLE can be categorized as hypermedia, virtual reality, or visual abstraction. Navigation in each category can support different types of learning. An analysis of how each model of navigation interaction is used within each type of CBLE is provided. This analysis represents the initial step for creating a framework that can assist designers of CBLEs in deciding which navigational model (or combination of models) is most appropriate to support the exploration of information spaces used within CBLEs.



Time	Title	Location
9:30am-9:50am	An empirical study on comparing mutants and faults	MC316
9:50am-10:10am	Semantic Agreement Service	
10:10am-10:30am	Using WS Level Agreements to Differentiate Web Service Offerings	
10:30am-10:50am	The Impact of Requirements Knowledge and Experience on Software Architecting: An Empirical Study	
10:50am-11:10am	The Role of Software Architecture in Decision Making During Requirements Engineering	
11:10am-11:30am	A User-Centered Approach to Improving System Testing	

An empirical study on comparing mutants and faults

Akbar Siami Namin

One of the main challenges in software testing research is the cost-effectiveness study of different software testing techniques. Experimental assessment is one approach to address this issue. As a result, in order to fulfill an empirical investigation, we need some subject programs with real faults and a large enough number of test cases. However, in reality, there are not too many choices of subject programs for experiments. The Mutation technique has been studied for generating a mutant version of a program, in which some pre-defined operators are used to simulate a fault in the code. So, in this way, researchers are able to generate a large collection of mutants behaving as faulty code. Nevertheless, we do not know if the statistical analysis and results achieved in this way are valid. In other words, the main challenge is: Are the mutants good representers of real faults? In order to address this latter problem, we have designed an empirical study on a widely used program with real faults and a large enough test pool. We are interested in investigating the cost-effectiveness ratios for both mutants and faults in terms of different coverage measurements as well as random generation of test suites. The statistical analysis shows that mutants can be treated as real faulty codes of a program. Also, we provide some evidence to address the cost-effectiveness issues of using different coverage criteria for generating test suites, and we compare them with the random generation case.

Semantic Agreement Service

Qian Zhao

For greater efficiency and lower costs, computer scientists and engineers have been pursuing automation of versatile process management. Naturally, it is widely expected that highly structured agreements should be able to be executed and enforced automatically. In the progress of building an executable license agreement framework, we have noticed that different agreement processing and automation have something in common, which leads us to propose a generic architecture that is able to provide various automated agreement services leveraging ontology, rule and agent technologies.

Using WS Level Agreements to Differentiate Web Service Offerings

Halina Kaminski, Khalid Shredil, Nazim Madhavji and Mark Perry

The advent of Service Oriented Architecture (SOA) paradigm and increasing use of Web Services (WS) implies that the future will see a large number of services transferred between providers and consumers, using many applications or agents working on behalf of humans. Discovering and using the services is the easy part. Negotiating and selecting the best services from amongst the plethora of similar ones, depending on their cost and quality, is the challenging issue. However, existing WS-I standards neither cater to provision of Service Level Agreements (SLAs), nor their exchange between parties. These standards are confined merely to WS description (WSDL). Once WS are discovered and selected, SLAs are merely used to monitor service compliance. We propose a novel method that allows service-providers to dynamically generate the SLAs, and then transfer them to clients for selection amongst competitive service providers. The clients use Application to Application (A2A) communication to choose the best service provider at run time, and then bind to it to available services. Our method complies with all WS-I standards, and hence does not require any modifications to the UDDI or WSDL. Instead of using the SLA as just a contractual document for compliance monitoring of the service, we also use it as a means of service selection. We demonstrate and validate our method using a prototype developed in laboratory settings, which uses multiple 'Weather Service Providers' to obtain various indicators for weather forecasting. Our paper describes a novel method that allows service-providers to dynamically generate the SLAs. The consumer uses these SLAs for service selection. The method complies with all WS-I standards, and hence does not require any modifications to the UDDI or WSDL. The transfer of a SLA between parties is packaged as a service itself, which can then be published on a UDDI registry. We demonstrate and validate our method using a prototype developed in a laboratory setting, which uses multiple 'Weather Service Providers' to obtain various indicators for weather forecasting. Our paper suggests only a simple approach in dynamically selecting WS using WSLAs under the A2A paradigm. We expect that more advanced approaches would have to be used as the nature of services becomes mission critical. The field of WS is only in its infancy, and as new standards are developed, they would facilitate this process. Artificial Intelligence (AI) and other techniques would be needed, possibly involving software mobile agents, for negotiation amongst parties. The clients would also like to monitor the QoS parameters for compliance, such as the response time, to ensure that other WSLA obligations are fulfilled. Despite all these extensions, the foundation framework involving the WS protocol stack would remain the same, because the very success of the WS framework lies not in any new technological advancement but in world wide acceptance to standards.

The Impact of Requirements Knowledge and Experience on Software Architecting: An Empirical Study

Remo Ferrari and Nazim Madhavji

While the relationship between Requirements Engineering and Software Architecture (SA) has been studied increasingly in the past five years in terms of methods, tools, development models, and paradigms, that in terms of the human agents conducting these processes has barely been explored. This presentation describes the impact of requirements knowledge and experience (RKE) on SA tasks. Specifically, it describes an exploratory, empirical study involving a number of architecting teams, some with requirements background and others without, all architecting from the same set of requirements. The overall results of this study suggest that architects with RKE perform better than those without, and specific areas of architecting are identified where these differences originate. We discuss the implications of the findings on the areas of training, education and technology.

The Role of Software Architecture in Decision Making During Requirements Engineering

James Miller

The presence of an existing software architecture within a software project (i.e., one that is in an evolutionary stage) is recognized in the requirements literature as being an important influence on the decisions being made in the requirements engineering process. In spite of this recognition it is not known what the impact of this influence is. Based on an exploratory control study, this paper explores how software architecture influences decision making during requirements engineering. We will not only examine how the properties of requirements (e.g., cost, priority, implementability, etc.) are influenced but also how the decisions which produced them are affected when an existing architecture is present. The results demonstrate how requirement engineering decisions can be enabled, constrained or otherwise influenced in a significant way by software architecture, and we will discuss what the implications of this might be on projects in different contexts (e.g., software as infrastructure or software as a commercial product)

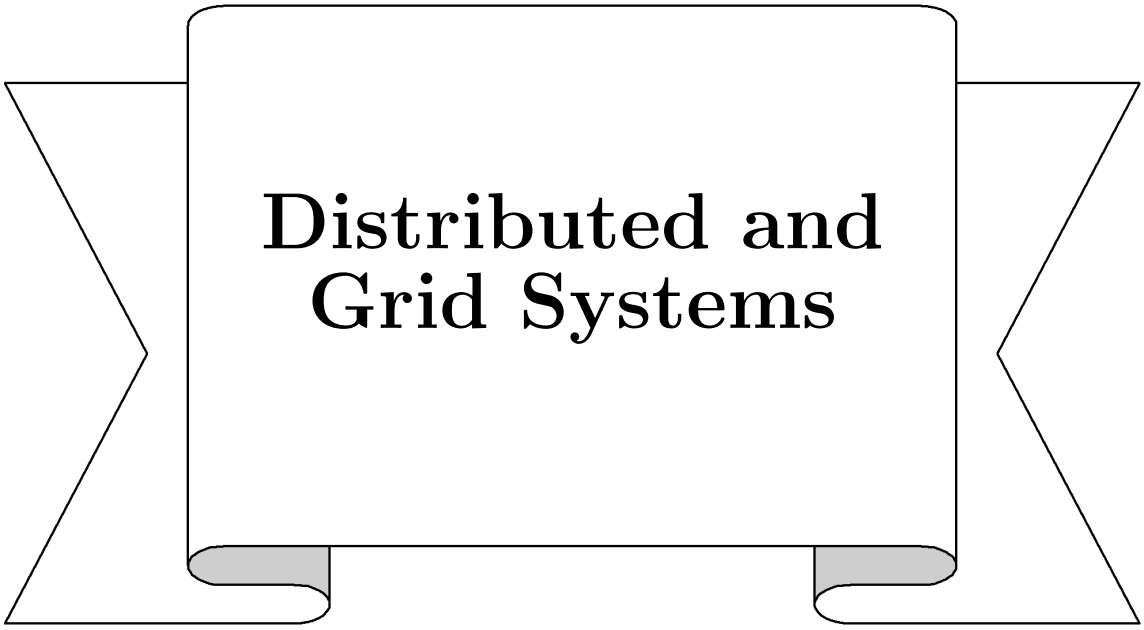
A User-Centered Approach to Improving System Testing

Andriy Miranskyy

Despite extensive efforts made to eliminate all bugs pre-release, large scale software systems can fail in many unexpected ways. Such software defects found in the field are expensive to repair.

One way to prevent such field-defects is to improve the system-testing processes. We approach this problem by logging the program's execution trace at the level of function calls. Traces can be captured during field use of the system (called user-traces) and in-house during system testing (called test-traces). Knowing which of the test-traces are "closest" to the fault user-trace could give a clue as to the conditions under which the system is failing and a lead into the reasons for such failure. This can then help in improving test scenarios which are used for system testing, thereby, helping to reduce field defects.

This field and test-trace comparison must be done automatically, but is complicated by the size of each individual trace [O(10^4 Mb)] and by the number of test-traces [O(10^7)]. In order to deal with this complexity, we devise "fingerprints" for traces so that similar ones can be readily identified. To do this, the trace structure is analyzed using an entropy measure. Based on the findings, we create trace "fingerprints" using l-word entropy, and Fourier and Wavelet Packet Transforms. In this presentation, we will describe the status of our current research and how we intend to validate it.



Time	Title	Location
12:30pm-12:50pm	Communication Factors for Jobs Across Multiple HPC Clusters	MC316
12:50pm-1:10pm	Avoiding TCP Packet Drops Using SmoothTCP	
1:10pm-1:30pm	Policy-based Autonomic Management of an Apache Web Server	
1:30pm-1:50pm	Towards Automating the Adaptation of Management Systems to Changes in Policies	
1:50pm-2:10pm	A Policy-Based Framework for Managing Data Centers	
2:10pm-2:30pm	Developing Autonomic Feedback Control for Heterogeneous Systems Using Cascaded Controllers	

Communication Factors for Jobs Across Multiple HPC Clusters

Jinhui Qin and Michael Bauer

High Performance Computing (HPC) clusters continue to be popular as a means of satisfying ever-increasing computing demands for many research institutes and commercial organizations. To more effectively use such clusters for even larger computations, users are looking to interconnect multiple HPC clusters, creating a grid, as a means of improving turn-around times and better use compute resources. To effectively use such grids, it may be desirable to split and co-allocate jobs requiring many processes across multiple clusters. While splitting a very large job across multiple clusters in an attractive possibility, the benefit, in terms of improving turn-around time, ultimately depends on the communication patterns between processes, workload on the communication links, and the maximum bandwidth of the links. Some researchers have approached this problem by assuming fixed slowdown ratios to represent and study the influence on job execution if a job was split across clusters. However, it is difficult to estimate such a slowdown ratio in advance, and it may not be uniform when there is a choice of multiple clusters with different communication links. Moreover, the slowdown ratio may actually change dynamically based on the jobs communication patterns, workload on the network links, and the maximum bandwidth of the network links, etc. The objective of this work is to understand the impact of communication on multi-processor jobs in order to develop scheduling strategies and job allocation algorithms for multi-cluster grids which can accommodate communication factors. In this presentation we report on initial investigations of some co-allocation strategies. This evaluation is based on a simulator that has been implemented and validated experimentally across two HPC clusters.

Avoiding TCP Packet Drops Using SmoothTCP

Elvis Vieira and Michael Bauer

Packet drops have a great impact in the RTT variation and the throughput experienced by some TCP applications. However, using packet drops is the primary way TCP uses to discover the bandwidth available in order to proceed with its packet transmissions. A different approach is used by SmoothTCP-q, where ICMP-SQ messages are sent to the SmoothTCP-q sender every time a threshold is reached in the queue size of the router. In this mechanism, SmoothTCP-q tries to avoid packet drops since it is not necessary to overload the router queue to discover the network bandwidth. Consequently, it is import to have some way to determine the queue threshold and evaluate its effect on SmoothTCP-q. This paper briefly describes SmoothTCP-q and presents a simple model for this relationship that can be used to evaluate whether or not there will be packet drops in a SmoothTCP-q connection.

Policy-based Autonomic Management of an Apache Web Server

Raphael Bahati, Michael Bauer, Elvis Vieira, Chang-Won Ahn, and O.K. Baek

Web-based servers, services and applications are becoming key elements in the way many organizations deliver their services and provide support. Ensuring that expected performance and behavioral objectives are met is therefore critical. These objectives may be defined internally within an organization that manages its own services or specified as service level agreements when the services are managed by third-party organizations. Policies can be used to specify expected behaviors of the systems, applications and services in the cluster and can provide the kinds of directives which an autonomic management system can and should rely on. We look at how such techniques can be applied to manage the Apache Web Server. In particular, an architecture for autonomic management of the Apache web server is described and a means of specifying policies is presented. A prototype autonomic management system is described and its performance in managing Apache under different scenarios is illustrated.

Towards Automating the Adaptation of Management Systems to Changes in Policies

Abdelnasser H. Ouda, Hanan Lutfiyya, and Michael Bauer

The goal of distributed systems management is to provide reliable, secure and efficient utilization of the network, processors and devices that comprise those systems. The management system makes use of management agents to monitor attributes that characterize system and application behavior e.g., length of a user session, CPU load, web server throughput. The monitored information is analyzed and events are generated in the case of undesired behavior e.g., CPU load is too high, session length exceeds maximum allowed. The events and other monitored information are further analyzed to the appropriate action. Policies are used to define the appropriate action based on the analysis. For example, if a session length exceeds 3 minutes then e-mail the administrator. Systems and policies governing the behavior of the system and its constituents change dynamically. For example the policy about session length may change to "if a session length exceeds 3 minutes for user nasser then terminate the session". Essential a policy is an event-triggered, action-condition rule. An event triggers the evaluation of a rule of the form "if condition then actions". An event is generated as the result of some condition of the state of the system being true. The state of the system is characterized by a set of attributes. A subject carries out action on or for specified targets if a specific event has occurred and a specific condition has been satisfied. The system monitoring and action execution are carried out by management agents. If the set of effective policies is dynamic then the set of management agents and the conditions that the agents are monitoring must also be dynamic. This work explores this problem.

A Policy-Based Framework for Managing Data Centers

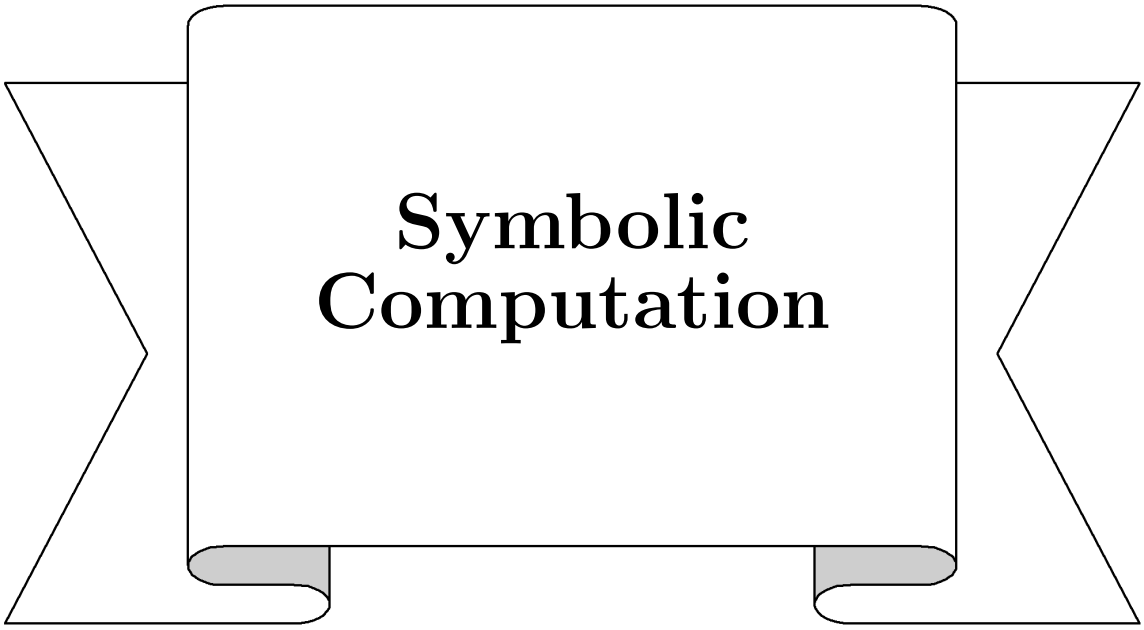
Bradley Simmons, Hanan Lutfiyya, Mircea Avram, and Paul Chen

A data center is defined as a set of computing resources that is owned by an organization and shared among multiple applications from different client organizations. Applications often have non-functional run time requirements for different classes of users. These are referred to as service level objectives (SLO) and are considered part of a service level agreement (SLA). Allocating resources to an application should be dynamic and based on specific run-time conditions. It should be possible to change these conditions without recoding. This talk describes a framework, a prototype based on this framework and an application of the prototype that dynamically allocates resources and allows the application environment to adapt in the case of having insufficient resources.

Developing Autonomic Feedback Control for Heterogeneous Systems Using Cascaded Controllers

Wael Hosny Fouad Aly and Hanan Lutfiyya

The motivation of the work is based on the static use of reference point values in dynamic feedback control systems that are based on control theoretic techniques. Reference point values are threshold values that values of attributes characterizing system behavior can be compared to. The result of the comparison is used to adjust the tuning parameter of the system being controlled. If the entity being controlled is a server then the best choice for the reference value may depend on many factors such as the required processing times, the machine speed that the server is executing on, .etc. This paper illustrates a novel approach to automatically change the reference value for different installations of servers. Comparisons are made between the approach proposed in this paper and other dynamic feedback control approaches. Results show that the proposed approach outperforms the other dynamic feedback control approaches. That is, the number of violations was reduced by about 14% and the number of processed requests was increased by about 18%.



Time	Title	Location
9:50am-10:10am	Implementation Techniques for Fast Polynomial Arithmetic in a High-level Programming Environment I: The generic case	MC320
10:10am-10:30am	Implementation Techniques for Fast Polynomial Arithmetic in a High-level Programming Environment II: The non-generic case	
10:30am-10:50am	Parallel Triangular Decompositions	
10:50am-11:10am	Complex pattern matching over sequences in Common Lisp	
11:10am-11:30am	Complexity and Regularity	

Implementation Techniques for Fast Polynomial Arithmetic in a High-level Programming Environment I: The generic case

Akpodigha Filatei

To solve large mathematical problems in science and engineering, there is an increasing need for both faster implementations of the well-studied classical algorithms, and, new faster algorithms in computer algebra. The theoretical asymptotic complexity of these algorithms are mostly known but achieving comparable practical complexity (efficiency) is a challenge.

Our mission is to implement algorithms for fast polynomial computations over a generic coefficient ring. Implementations of fast algorithms in low-level languages have been widely investigated but the same is not true for implementations in high-level languages. Considering that it is desirable for mathematicians to reuse and build on top of implementations in high-level languages rather than in low-level languages, we use a high-level programming environment for our implementation: Aldor. We also write generic code that can be useful in a wider variety of applications.

Our main focus is on fast algorithms for which implementation techniques remain to be investigated, such as the Fast Extended Euclidean Algorithm (FEEA). Our main application is fast algorithms for GCDs over fields and field extensions which are major tools for polynomial system solving.

Implementation Techniques for Fast Polynomial Arithmetic in a High-level Programming Environment II: The non-generic case

Xin Li

In the last decade, several software for performing symbolic computations have put a great deal of effort in providing outstanding performances, including successful implementation of asymptotically fast arithmetic. Today, it is common practice to assume that a new algorithm, say for Hensel lifting techniques, can rely on asymptotically fast polynomial multiplication.

The goal of this work is to provide fast algorithms with efficient implementation for univariate/multivariate polynomials in a high-level language, namely AXIOM. On the contrary of A. Filatei's work, we focuss on special coefficient rings, prime fields, since they play a central role in computer algebra.

We mix high-level generic code (AXIOM code), middle-level code (Lisp) and low-level machine dependent code in a transparent way for the high-level end-user. High performance is achieved by selecting suitable data structure, using fast integer and floating point arithmetic, understand restrictions of compiler, understand memory performance and processor's architecture. Our implementation is based on Intel-compatible processor, running on Linux. Our specialized implementation in AXIOM leads to comparable performances and often outperforms those of the most famous comparable software, MAGME and NTL.

Parallel Triangular Decompositions

Yuzhen Xie

Since the discovery of Greobner bases, the algorithmic advances in Commutative Algebra have made possible to tackle many classical problems in Algebraic Geometry that were previously out of reach. However, algorithmic progress is still desirable, for instance when solving symbolically a large system of algebraic (non-linear) equations is needed.

For such a system, in particular if its solution set consists of geometric components of different nature (surfaces, curves, points) it is necessary to combine Greobner bases with decomposition techniques, such as triangular decompositions.

The efficiency of this approach depends on its ability to detect and exploit geometrical information during the solving process. Its implementation, which naturally involves symbolic parallel computations, is another challenging topic.

Complex pattern matching over sequences in Common Lisp

Geoff Wozniak

The usefulness of regular expressions for matching patterns over sequences of characters is evident by its presence in a myriad of programs and libraries. While "string" regular expressions are very useful, they suffer from some limitations:

- they only match against strings;
- they only predicate matches;
- backreferences are not scoped and are not symbolic, often making their use clumsy;
- writing patterns over multiple strings that share information is a complex and tedious endeavour;
- the specification language is often different from the source language, a disconnect that makes creating new regular expressions from old ones difficult.

In the study of biological processes as formal models (and vice versa), examining a sequence for patterns is a common task. However, common regular expressions are ill-suited for the task due to the vices listed above. In this work, I will demonstrate a pattern matcher that addresses these issues; that is, a pattern matcher that matches over any sequence, generates all possible matches, uses lexically scoped, symbolic backreferences, allows for dependent patterns with constraint matching, and whose specification is in the same language as the source language, namely, Common Lisp.

Complexity and Regularity

Sorin Constantinescu

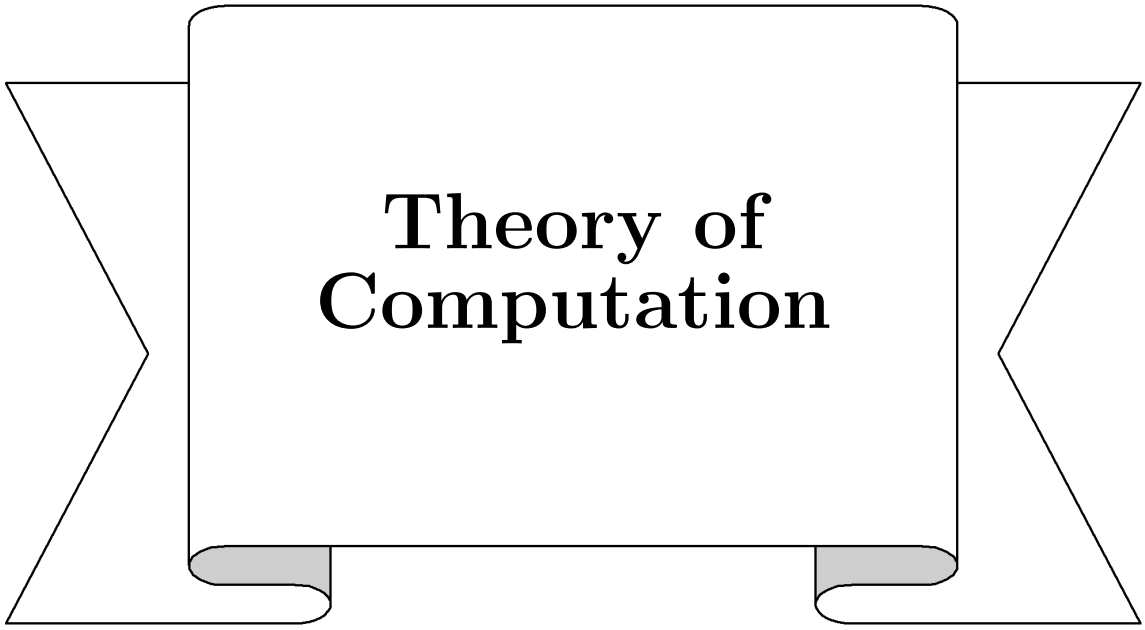
String complexity is a concept that lies at the core of numerous topics in Computer Science with exciting applications in other disciplines such as Biology and Medical Science. Its purpose is to capture the amount of information that is contained in a given string of characters. For instance, is the string 01010101 more or less complex than 00001111? Which of the two strings holds more information?

One of the earliest and most fundamental works in this area is due to Kolmogorov who proposed, as a measure of the complexity of a string, the length of the shortest program that will produce the specified string. This approach ideally captures the idea of complexity from an algorithmic perspective. It contrasts the idea of regularity and structure with the one of randomness: highly structured strings will have low complexity while random ones will most likely have high complexity.

One important observation is that this measure of complexity indicates how much a string (or, in the context of practical computing, a file, which can be regarded as a binary string) can be compressed: the ideal compressor for a given file is the shortest program that outputs that file.

The major disadvantage of Kolmogorov's complexity is the fact that it is not calculable by any means. To correct this shortcoming, various computable measures that try to capture the same idea of complexity have been introduced. Naturally, none of them performs as well as Kolmogorov's, assigning high complexity to some highly regular sequences, depending on the nature of the actual measure. In this paper, I am focussing on two popular measures: the Factor Complexity and the Lempel Ziv Complexity, the latter being very closely related to the extremely widespread Lempel Ziv compression algorithm.

The topic of this paper is how these measures behave on some classes of strings, notably infinite strings of high regularity. This gives an idea about the robustness of a complexity measure since those strings will have a low Kolmogorov complexity while also featuring an unbounded length, making them ideal test subjects for measure comparison.



Time	Title	Location
12:30pm-12:50pm	Introduction to Process Traces	MC320
12:50pm-1:10pm	State complexity of combined operations	
1:10pm-1:30pm	The Church-Turing Thesis and the Continuum View of Computability	
1:30pm-1:50pm	An Infinite Hierarchy of Languages Induced by Depth Synchronization	
1:50pm-2:10pm	XML Schema of Glycan and its Application in Glycan Sequencing	

Introduction to Process Traces

Qing Zhao

Mazurkiewicz's theory of traces is one of the popular theories that describe the behavior of concurrent systems. In this talk, we argue that this theory is not totally adequate for describing concurrent processes. We introduce process systems and ptraces, as well as P-expressions, which we consider can describe concurrent processes more adequately. In addition, we introduce an extension to P-expressions by adding an Option operation. We show the applications as well as the properties of the operations. We also show that adding the new operation makes P-expressions more powerful and more convenient to describe concurrent systems faithfully.

State complexity of combined operations

Yuan Gao

There have been a lot of papers in the area of state complexity since 1990s. However, in all of those papers, state complexity is considered for only individual operations, e.g.,intersection, union, catenation and star. But in practice, not only individual operations but also combinations of operations are often required to be performed on FA. The state complexity of a combination of operations are not necessarily equal to the combination of the state complxity of the individual operations. Here we review some recent results on state complexity of several basic operations on regular languages.We also show the proof that the state complexity of star of union operation is much better than the direct combination of state complexity of each of them.

The Church-Turing Thesis and the Continuum View of Computability

Maia Hoeberechts

In 1936, Alan Turing described a calculating machine which since has come to be known as the Turing Machine. Turing's original idea was that his machine would model the type of computation that a human with pen and paper was capable of doing. Other early models intended to capture the process of computation are Church's lambda calculus and recursive function theory. These models and many others developed in subsequent years were all shown to be capable of calculating the same set of functions.

The properties of Turing Machines and Turing-equivalent computation models have been extensively studied and the Church-Turing Thesis, which essentially states that Turing Machines capture our intuitive notion of computation, is commonly accepted. The idea that there are fundamental properties shared by every "reasonable" computing model is a compelling one. Over the last 70 years, the Turing Machine has taken a central position as the paradigmatic model of a computer. Most computer scientists take for granted that the computing capabilities of a Turing Machine represent those of any past, present or future computer.

The primary benefit of focusing on Turing-equivalent models is the way in which it allows us to define the concept of computability. The "computable languages" are normally understood to be those languages accepted by a Turing Machine. Computable functions are conventionally defined as those functions which can be computed using a Turing-equivalent device. However, it is well known that there are machine models which are more powerful than a Turing Machine, such as some neural networks or Turing Machines using real numbers as input. Given these more powerful models, why would we limit what we call "computable" to that which is computable by a Turing Machine?

In this talk, I will review some of the history and debate surrounding the Church-Turing Thesis and the concept of computability. I will then introduce a new notion of computability according to which the machine in use determines what is computable — computability is no longer an absolute classification. I will call this theory, which allows machines with more or less computing power than Turing Machines to also define classes of "computable" functions, the "continuum view" of computability. Finally, I will discuss how conventional computability theory fits within this new framework.

An Infinite Hierarchy of Languages Induced by Depth Synchronization

Franziska Biegler

Synchronized context-free (SCF) grammars are an extension of context-free grammars in which so-called synchronization symbols can be attached to the nonterminals. In order for a derivation tree to contribute to the language, the situation sequences along the paths in the tree have to be either all equal or all in prefix relation. SCF grammars are known to generate the well studied family of ET0L languages (extended interaction-less Lindenmayer systems with tables). In ET0L systems, however, there is no distinction between synchronized and non-synchronized derivation steps.

The depth synchronization function of an SCF grammar maps each natural number n to the maximal length of the situation sequence required to generate a word of length at most n. The definitions can be extended to SCF languages to provide characterizations of the complexity of the grammars for the language. It is immediate from previous results, that all SCF languages that are not context-free require at least logarithmic and at most linear depth synchronization. The family of SCF languages with depth synchronization below logarithmic is equivalent to the family of context-free languages.

The existence of SCF languages with logarithmic depth synchronization measure is obvious with $\{a^{2^n} \mid n \in \mathbb{N}\}$ being one example.

In this presentation we show that the language $L_0 = \{w\$w \mid w \in \{0, 1, \#\}\}$ requires linear depth synchronization, i.e. there does not exist any SCF grammar generating L_0 with sub-linear depth. This shows that the previously known upper bound is tight.

We also provide examples of languages with depth-synchronization measures in $\Theta(n^{\frac{1}{k}})$ for each $k \geq 2$, which gives rise to a strict infinite hierarchy within the family of SCF (=ET0L) languages.

The question whether there are SCF languages, the depth synchronization measure of which is neither logarithmic, nor linear, nor in $\Theta(n^{\frac{1}{k}})$ for some $k \geq 2$ remains an open problem.

XML Schema of Glycan and its Application in Glycan Sequencing

Baozhen Shan

Glycosylation is one of the most common post-translational modifications in which carbohydrates are attached to cell surface and extracellular matrix proteins as well as to lipids. The carbohydrates of glycoproteins and glycolipids are commonly referred as glycans. The glycan moieties cover a range of diverse biological functions. As a natural extension of proteomics, glycomics provides a better understanding of glycoproteins, glycosylation process, and its role in the protein function.

Glycan structure sequencing, which is to determine the primary structure of a glycan using tandem mass spectrometry, remains one of most important tasks in glycomics. Most of current software tools use de novo sequencing technique, which is to determine structure without the aid of a known glycan database. De novo sequencing is powerful to determine novel structures that are not in the database. However, in many cases, de novo method lacks accuracy due to errors in experiment. Considering structural conservation, comparing with known structures in the database provides an efficient solution to this problem.

The use of glycomics databases has become indispensable for the daily work of biochemists and biologists. However, the resources for glycomics are very poor as compared with those for proteomics and genomics. Exchange of information of glycan structures by web is not as efficient as DNAs/RNAs or proteins. One major reason is that unlike DNAs/RNAs and proteins where sequences are linear and data type is string which can describe their primary structures, carbohydrates are characterized by their two dimensional sequence, linkage and stereochemistry. So far, There is no efficient and well-accepted language to describe carbohydrate sequences.

The eXtensible Markup Language (XML) is a language that describes text and tags. The availability of development tools (XPath, XQuery, XML Parser) promotes in popularity of XML usage. We noticed that a well-formed tree-structure of XML is very suitable for describing carbohydrate sequence structures. In this presentation, we give the XML schema of glycan, which describes the primary structure of glycan clearly. With XML database of glycan, structural query is very efficient, which is used for glycan sequencing. We first use de novo sequencing technique to obtain a list of structure candidates. Then, the structure candidates are refined by query the XML database of glycan.