# Research Showcase

<table>
<thead>
<tr>
<th>Paper ID &amp; Title</th>
<th>Author(s)</th>
<th>Affiliation</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>03: Effective testing: A combinatorial approach for improving test efficiency</td>
<td>Abdul Rauf</td>
<td>IBM</td>
<td>5</td>
</tr>
<tr>
<td>19: Twit-Digest: A Web based Tool to Analyse and Visualize Twitter in Real-time</td>
<td>Aditi Gupta, Akshit Chhabra, Ponnurangam Kumaraguru</td>
<td>IIIT Delhi, DTU, IIIT Delhi</td>
<td>6</td>
</tr>
<tr>
<td>37: Transducer Models of Sliding Window ARQ Protocols for Noisy Channels</td>
<td>Jay Thakkar, Aditya Kanade</td>
<td>IISc</td>
<td>7</td>
</tr>
<tr>
<td>42: Designing Green Database Servers</td>
<td>Chhabi Sachan</td>
<td>IISc</td>
<td>8</td>
</tr>
<tr>
<td>44: The CODD Metadata Processor</td>
<td>Nilavalgan I, Deepali Nemade</td>
<td>IISc</td>
<td>9</td>
</tr>
<tr>
<td>47: Computer Vision based Approach for Indian Sign Language Character Detection</td>
<td>Ajith J, Niranjan M, Sai preethy M.S</td>
<td>Amrita School of Engineering</td>
<td>10</td>
</tr>
<tr>
<td>48: Smart Farming : A Step Towards Techno-Savvy Agriculture</td>
<td>Dr. Mydhili Nair, Padmashree Bhat, Priyam Borkotoky, Shahswathi N, Suhas K, Nishant Kulkarni</td>
<td>MSRIT, MSRIT, MSRIT, MSRIT, IBM</td>
<td>11</td>
</tr>
<tr>
<td>57: Semantic Based Peer To Peer Service Discovery in Cloud</td>
<td>Apsara Karen S, Aiswariya J, Saswati Mukherjee</td>
<td>CoE Guindy, Anna University</td>
<td>14</td>
</tr>
<tr>
<td>65: Discovering Coverage Patterns for Banner Advertisement Placement</td>
<td>Gowtham Srinivas Parupalli, Krishna Reddy P, Bhargav S</td>
<td>IIIT Hyderabad</td>
<td>15</td>
</tr>
<tr>
<td>66: Detecting MS Initiated Signaling DDoS Attacks in 3G/4G Wireless Networks</td>
<td>Aman Gupta, Tanmay Verma, Soshant Bali, Sanjit Kaul</td>
<td>IIIT Delhi, IIIT Delhi, IIIT Delhi, IIIT Delhi</td>
<td>16</td>
</tr>
<tr>
<td>71: Resolution Problems in Community Detection in various Social Networks</td>
<td>Samridhi Khurana, Natwar Modani</td>
<td>Indian Institute of Technology (BHU), Varanasi</td>
<td>17</td>
</tr>
<tr>
<td>73: Data Understanding using Semi-Supervised Clustering</td>
<td>Rashmi Dobarial, Priya Jain, Ashish Mahabal, Vasudha Bhatnagar</td>
<td>Agnity India, Aricent Group India, Caltech, University of Delhi</td>
<td>18</td>
</tr>
<tr>
<td>75: Energy Efficient Data Center Networks : A SDN based approach</td>
<td>Dharmesh Kakadia, Vasudeva Varma</td>
<td>IIIT Hyderabad</td>
<td>19</td>
</tr>
<tr>
<td>76: Hosting Keyword Search Engine inside RDBMS</td>
<td>Amit Mehta</td>
<td>IISc</td>
<td>20</td>
</tr>
<tr>
<td>Session</td>
<td>Title</td>
<td>Authors</td>
<td>Institutions</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>80:</td>
<td>ROSHNI: Recreating Optical Sight by virtually Navigating Indoors</td>
<td>M Balakrishnan, Ankit Kumar, Ashwini Choudhary, Devesh Singh, Dhruv Jain, Dhruv Gupta, Himanshu Meenia, Kartik Singh, Manas Sahu, Mohit Taak, Naman Gupta, Nitin Kamra</td>
<td>IIT Delhi</td>
</tr>
<tr>
<td>83:</td>
<td>Efficient Constrained Shortest path Estimators</td>
<td>Ankita Likhyani, Srikanta Bedathur</td>
<td>IIIT Delhi</td>
</tr>
<tr>
<td>85:</td>
<td>Is Government a Friend or Foe? Privacy in Open Government Data</td>
<td>Srishti Gupta, Mayank Gupta, Ponnurangam Kumaraguru</td>
<td>IIT Delhi, Delhi College of Engineering</td>
</tr>
<tr>
<td>93:</td>
<td>Enhancing Query-by-Object Approach for information Requirement</td>
<td>Ammar Yasir, Kumaraswamy Mittapally, P Krishna Reddy</td>
<td>IIIT Hyderabad</td>
</tr>
<tr>
<td></td>
<td>Elicitation in Large Databases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>94:</td>
<td>A Framework for Corruption Assessment in Public Services</td>
<td>Nidhi Rajshree, Nirmat V. Desai, Biplov Srivastava, Anshu Jain, Anuradha Bhamidipaty, Kamal Bhattacharya</td>
<td>IBM Research, India, IBM Research, India</td>
</tr>
<tr>
<td>98:</td>
<td>Mechanism Design for Resource Critical Crowdsourcing</td>
<td>Swaprama Nath, Pankaj Dayama, Dinesh Garg, Narahari Yadati, James Zou</td>
<td>PhD Student, Computer Science, IBM Research India, IBM Research India</td>
</tr>
<tr>
<td>100:</td>
<td>MODA: A Middleware for Policy-aware Adaptive Mobile Data Collection</td>
<td>Vinay Kolar, Hemant Kowshik, Palanivel Kodeswaran, Ravindranath Kokku</td>
<td>IBM Research India, IBM Research India, IBM Research India</td>
</tr>
<tr>
<td>101:</td>
<td>Eliciting High Quality Feedback from Crowdsourced Tree Networks</td>
<td>Ratul Ray, Rohith D Vallam, Yadati Narahari</td>
<td>IISc Bangalore, IISc Bangalore, IISc Bangalore</td>
</tr>
<tr>
<td></td>
<td>using Scoring Rules</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RTL-SDR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>107:</td>
<td>Efficiently Scripting Change-Resilient Tests</td>
<td>Nimit Singhania, Pranavadatta Dn</td>
<td>IBM Research, India, IBM Research, India</td>
</tr>
<tr>
<td>108:</td>
<td>Preserving Date and Time Stamps for Incident Handling in Android</td>
<td>Robin Verma, Gaurav Gupta</td>
<td>IIT-Delhi, IIT-Delhi</td>
</tr>
<tr>
<td></td>
<td>Smartphones</td>
<td></td>
<td></td>
</tr>
<tr>
<td>113:</td>
<td>ForeCAST: Social-web assisted Predictive Multicast</td>
<td>Giridhari Venkatadri, Hemant Kowshik</td>
<td>IIT Madras, IBM Research India</td>
</tr>
<tr>
<td>115:</td>
<td>Achieving skill evolution through challenging task assignments</td>
<td>Chandrashekar L. Gargi B. Dasgupta, Nirmat V. Desai</td>
<td>IISc Bangalore, IBM Research, India</td>
</tr>
<tr>
<td>116:</td>
<td>Topic Expansion using the Term Co-occurrence Analysis</td>
<td>Sumant Kulkarni, Srinath Srinivasa</td>
<td>IIT-Bangalore, IIT-Bangalore</td>
</tr>
<tr>
<td>119:</td>
<td>Optimal Incentive Strategies for Product Marketing on Social Networks</td>
<td>Pankaj Dayama, Aditya Kamik, Y Narahari</td>
<td>IBM Research, India, General Motors, IISc Bangalore, IISc Bangalore</td>
</tr>
</tbody>
</table>

2
<table>
<thead>
<tr>
<th>Video Title</th>
<th>Author(s)</th>
<th>University</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChaMAILeon: Simplified email sharing like never before!</td>
<td>Prateek Dewan, Mayank Gupta, Sheethal Shreedhar, Prof. Ponnurangam K. (&quot;PK&quot;)</td>
<td>IIIT-Delhi, DCE, NITK Surathkal</td>
<td>42</td>
</tr>
<tr>
<td>Smart Farming</td>
<td>Priyam B, Padmashree Bhat, Suhas K, Shashwathi Reddy</td>
<td>MSRIT Bangalore, MSRIT Bangalore</td>
<td>11</td>
</tr>
<tr>
<td>The CODD Metadata Processor</td>
<td>Deepali Nemade, I.Nilavalagan</td>
<td>IISc</td>
<td>09</td>
</tr>
<tr>
<td>WHOIS</td>
<td>Himanshu Shekhar, Sudhanshu Shekhar, Sherjil Ozair, Dhruv Jain</td>
<td>IIT Delhi, IIT Delhi</td>
<td>43</td>
</tr>
<tr>
<td>Indian Robot</td>
<td>MD.ATHIQ UR RAZA AHMED. M MANIKANDHAN.A. REENA JESICCA GEORGE BAILRAJ</td>
<td>SAEC, Anna University</td>
<td>44</td>
</tr>
</tbody>
</table>
Effective testing: A combinatorial approach for improving test
efficiency

Abdul Rauf EM (abdurauf@in.ibm.com)
IBM India Pvt Ltd

In this paper we are discussing how we introduced a combinatorial approach for improving test efficiency in an embedded environment using simulation testing.

Suppose you got an assignment for doing the simulation analysis of embedded software using simulation testing (Eg: Electronic power assisted steering system). The requirement we need to test is that to run multiple releases of this software on a simulator (Micro controller simulator) for specified number of kilometers (eg: 2, 50,000KM) and make sure that the system running without any issues. For simulating the situation the test team need to identify various inputs like sensing options (Like Angular range, angular resolution, angular accuracy, response time, rotation speed, speed accuracy, torque range etc), power consideration and other control aspects. Each parameter could have numerous values. As a result, there were tens of thousands of potential input variations, each producing different results, and making exhaustive testing of the simulation nearly impossible. In such a situation test team will face difficulties in selecting the suitable test combination for getting test completeness and coverage.

Manual process of selecting the input combination is mind numbing and error prone. Random selection and exploratory approaches do not provide an easy way of assessing the effectiveness of our test coverage and usually lead to redundant tests. In this situation, we need some way to systematically produce a subset of all possible tests that effectively tests important combinations; has a high probability of exposing functional, multi-modal defects and risks and provides confidence in the test coverage. Testing combinations of input variables that affect a common output is important because several industry and academic studies conclude that variable interactions are often problematic and a common cause of multi-modal bugs in software. But, it is usually not feasible to test exhaustively all possible combinations of input variables. So how do we derive a suitable set of test cases when it’s just not feasible to use all the possible combinations and variations of test parameters? The paper addresses one solution through combinatorial test approach.
Twit-Digest: A Web based Tool to Analyze and Visualize Twitter in Real-time
Aditi Gupta, Akshit Chhabra, Ponnurangam Kumaraguru

Internet and online social media have revolutionized the face and medium for people to access information and news about current events. Unlike traditional news media, online social media such as Twitter is a bidirectional media, in which common people also have a direct platform to share information and their opinions about the news events. Twitter content provides a vast resource of unmonitored and unstructured but rich information and content about events. Large volume of content generated on Twitter, makes manual monitoring and analyzing of data expensive, in terms of time and resources required. Also, anonymity of source and presence of malicious elements makes raw information consumption from Twitter challenging.

The aim of this project is to develop a real-time Twitter search tool for extraction, analysis and visualization of content from Twitter, with special emphasis on security aspects. Twit-Digest, is an easy-to-use, free-for-all web based tool, developed to provide an interface over the direct Twitter stream. The tool extracts data (tweets and user information) from Twitter and presents them to the user in real-time, along with various analysis outputs, like spam / phishing detection, credibility assessment, social network analysis and query expansion. Analyses are aimed at providing the user with quick inferences and big-picture about the activity about an event or query. Security analysts, organizations, and professional agencies that want to monitor content of its interests on Twitter can effectively use the tool.

The web-based tool has been designed using Python-Django web framework with a MySql database in the backend and calls Twitter Streaming API in the background for data. Twit-Digest is a LIVE web based tool [http://precog.iiitd.edu.in/Twit-Digest], available free-for-all, that provides users with a safer and richer Twitter based experience. Some of the features provided by Twit-Digest are:

- Credibility and spamming check for all tweets
- Geographical analysis of location of tweets and users
- Social network analysis of the reply and retweet graphs of Twitter
- Query expansion of the query given by the user to Twit-Digest
- Extraction of popular URLs

(a) Sample geographical distribution of the tweets and users corresponding to a query.
(b) The network graph, with users as nodes and edges representing the retweets and replies by one user to another.
(c) For each tweet the credibility score is computed and the tweet is marked as spam (green tick) or not spam (red cross).
Transducer Models of Sliding Window ARQ Protocols for Noisy Channels
Aditya Kanade, Jay Thakkar

In the communication system, the sender wants to transmit messages to the receiver over a transmission channel that can be either noiseless or noisy. Clearly, it is necessary that the message be transferred successfully without any error, in either case. The well-known Sliding Window ARQ protocols ensure reliable communication over unreliable communication channels. Existing approaches to verify these protocols assume channels to be perfect (reliable) or lossy, but noise-free. However, in the real world, communication channels between the sender and the receiver can be noisy. In this work, our aim is to formally model such protocols and verify their correctness, without the assumption that channels are noise-free. To verify the correctness of a given sliding window protocol, we model the sender, the receiver and the specification of the protocol as Streaming String Transducers (SSTs). A streaming string transducer makes a single left to right pass over an input string to produce an output string. In sliding window protocols, the sender reads a message, encodes the message into a frame with a sequence number, message content and CRC (Cyclic Redundancy Check) and transmits the frame. The sender then waits for an acknowledgement before transmitting the next frame to ensure reliable communication. Sender retransmits the same frame on receiving negative acknowledgement. The basic idea here is that the receiver positively acknowledges every correctly received frame, i.e., frames with correct CRC bits, and negatively acknowledges every corrupted frame. Messages and acknowledgement streams can be thought as input/output strings over certain input alphabets. Hence, we use SSTs to model such protocols as it makes a single pass over input string and it uses string variables to buffer the messages. We also model the CRC generator (for sender) and checker CRC (for receiver) using SST for a fixed generator polynomial. Next, a sequential composition of the sender SST and the receiver SST is performed. Since SSTs are closed under sequential composition, their composition is also an SST. The verification problem is then to show the functional equivalence of the composed SST and the specification SST. In this work, we are able to formally model the sender, the receiver and the specification of some protocols using SST for noisy channel and verify their correctness.
DESIGNING GREEN DATABASE SERVERS  
Chhabi Sachan, Database Systems Lab, SERC/CSA, IISc  
chhabi@dsl.serc.iisc.ernet.in

Data centres consume large amounts of energy. Since databases form one of the main energy sinks in a typical data centre, building energy-aware database systems has become an active research topic recently. Modern database engines typically choose query execution plans with the objective of minimizing estimated query execution time. Since now our aim is to build a query optimizer that minimizes energy consumed (or minimizes the maximum power extracted) by the query engine, the first task that needs to be accomplished is to develop the models to estimate the power or the energy consumed by Database Servers with highest possible accuracy. After this the steps can be taken to minimize the energy consumed by the query engine. These will also help in identifying plans with attractive tradeoffs between energy consumption and time-efficiency or peak power consumption and time-efficiency. Here we focus on taking the first step towards making the "green" database systems, i.e. developing models to estimate energy and peak power consumed in database servers. In the following sections we describe the methodology adopted to model energy/power of a Query Engine.

MODELING ENERGY: For modelling the energy consumption in DBMS an operator-based approach is followed, i.e. the energy model corresponding to each DB operator like scan, sort, join etc. is developed separately by first determining the factors affecting energy consumed by each operator, like relation size, tuple width, selectivity, index key etc. , the values of which are collected by executing the queries that involves only these operators. The collected values are then fed it to least squares regression model as training data, that outputs a corresponding function. The testing of the models developed is done against the TPC-H benchmark queries.

MODELING PEAK POWER: The same method could not be used for modelling peak power of DB operations, as peak power are non-additive in nature unlike average power or energy consumed which can be added to get total energy consumed in a query execution plan. Also, here the parallelism of operators has to be taken into account, as they represent the maximum aggregate consumption of concurrent operations, i.e. source of peaks or short term burst in power consumed. So a different approach which is based on dividing a query into pipelines is adopted to model the power consumption in it.
The effective design and testing of database engines and applications is predicted on the ability to easily evaluate a variety of alternative scenarios that exercise different segments of the codebase, or profiles module behavior over a range of parameters. A limiting factor is that the time and space overheads incurred in creating and maintaining these databases may render it infeasible to model the desired scenarios. In Database Management system there exists a class of important functionalities, such as query plan generators, system monitoring tools and schema advisory modules for which the inputs comprise solely of metadata, derived from underlying database. So we present a graphical tool, called CODD, that supports the ab-initio creation of metadata along with other functionalities such as metadata retention, scaling and porting across database engines. CODD provides following modes of operations, which covers the construction of alternative scenarios and various features of CODD.

1. **Metadata Construction:** Metadata construction lets the users to directly create or edit the statistical meta-data ensuring the validity of entered metadata without requiring presence of any prior data instance. In CODD, Histograms can be visualized graphically and can be altered to get the desired geometry by simply reshaping the bucket boundaries with the mouse. Further, CODD incorporates a graph-based model of the structures and dependencies of metadata values, implementing a topological-sort based checking algorithm to ensure that the metadata values are both legal(valid range and type) and consistent(compatible with the other metadata values).

2. **Metadata Retention:** CODD supports dropping of some or all of the raw data after extracting metadata from a database instance, permitting reclamation of the storage space without affecting metadata characteristics.

3. **Metadata Transfer:** It supports automatic porting of the statistical metadata across database engines, to an extent possible, based on the predefined metadata statistics mapping between them, thereby facilitating comparative studies of systems as well as early assessments of the impact of data migration.

4. **Metadata Scaling:** For the purpose of testing the database engine on scaled version of the original database scaled metadata is required. CODD natively supports space-scaling models along the lines of TPC-H and TPC-DS benchmarks. In addition, it also provides a novel time-based scaling model, where the metadata is scaled with respect to the execution time of the query workload.

In a nutshell, CODD is an easy-to-use graphical tool for the automated creation, verification, retention, porting and scaling of database metadata configurations. An extension to CODD is to construct alternative scenarios for testing execution modules that requires data in evaluating scenarios.

For More Details: [http://dsl.serc.iisc.ernet.in/projects/CODD/](http://dsl.serc.iisc.ernet.in/projects/CODD/)
COMPUTER VISION BASED APPROACH FOR INDIAN SIGN LANGUAGE CHARACTER RECOGNITION
Amrita School of Engineering, Coimbatore

INTRODUCTION

Computer vision has developed from being an abstract concept to a complete and ever expanding field. Gesture Recognition is one application that has recently attained a lot of focus. This project attempts to use computer vision algorithms to detect Indian Sign Language gestures. Initially we attempt to detect the English character set (A-Z). This project is significant because deaf and dumb people find it very hard to communicate with others. Also, it is not right to expect everyone to know sign language. Further sign language is highly regional. The main difference to other similar systems is that our system does not require any additional hardware.

DISTANCE TRANSFORM APPROACH

The input, which is the gesture, is taken in through a webcam and converted into frames. Each frame is taken for processing separately. Colour based segmentation using the Hue Saturation and Intensity Colour model is done to segment out the hand region from the background. Hue denotes the colour, saturation denotes the depth of the colour and intensity denotes brightness. On this segmented image, distance transform is performed. Distance transform calculates the distance of each interior pixel to the closest boundary pixel.

Distance transform is used to identify the centroid of the hand. Then a structural element (a disc) used to separate out the finger region from the hand. From the finger region the finger tips are identified. This is done by identifying the major axis of each finger. The major axis is where the distance between the current hand pixel and closest boundary pixel is maximum. The end points of the major axis are taken and a line is drawn between them.

During the initial training phase which lasts for 50 frames, the length of each finger and the distance of each open finger to every other finger is found and an upper triangular matrix is obtained. This upper triangular matrix and the length of each finger obtained is used to identify which of the fingers are half open, full open or full closed. Based on these features the recognition is done during the actual sign language recognition phase.
Agriculture is the backbone of Indian economy, with two-thirds of our population depending on farming and its agro-products for livelihood. There are many problems affecting the farmers, leading to large number of farmer suicides across India. Some of these problems are in our control to solve through timely expert advice such as what fertilizers and pesticides to apply - when and how, what crops to be grown along with the main crops or on rotation basis to increase yield, side-businesses that can be taken up etc. Over the last two decades, there has been vast amount of research addressing problems specific to the Indian farming sector, but the suggested best practices and outcomes of this research has remained largely with the scientific community. They have not been put to practice by the Indian farmers. This video proposes a set of farmer friendly services which use technology to bridge the existing wide gap between the expertise of agro-scientists and the transfer of this knowledge in a personalized way to the Indian farmers, so it can be put to use effectively. The video also presents the benefit of taking the expert advice of agricultural experts by the farmers of a sapota farm in Sulikunte, a farming village in rural Bangalore.

All these services are personalized, meaning, the information appearing on the user interface, are relevant to the farmer who uses the application. Personalization of the farmer services is achieved by capturing the location of the farm, so that data relevant to that location can be given, rather than general information. The various personalized services provided in the proposed solution are described in the subsequent sub-sections.

Registration Service: This service enables new farmers to register into the system. The details captured are - farmer’s land location, soil type, crops grown, amount invested for seeds and fertilizers in last three years, the crop yield, income or loss incurred for each crop etc.

Yield Prediction Service: This predicts the yield a farmer can expect. A ‘Regression Model’, is used where data is utilized to identify relationships among variables.

Market and Storage Locator Service: This service is used to locate the local and export markets, near the farmland, for the farmer to sell and store his produce hygienically.

Expert Advice Service: This service aims to bridge the gap between the vast amount of knowledge with agricultural scientists and the percolation of this know-how to the farmers.

Crop Monitoring Service: In this service, the agricultural expert gets to view the various stages of crop cultivation, starting from preparation of the cultivable land, through pictures uploaded from the mobile application provided by the system.
Database Management Systems (DBMS) are popularly used for storing and managing enterprise data. Modern DBMS typically store data in the form of relations, and query the information using the declarative Structured Query Language (SQL). The number of candidate strategies, called “plans”, by which the query can be processed, is exponential in the number of relations. Hence DBMSs incorporate a “query optimizer” module, which explores this exponential space using dynamic programming techniques, to find the best plan. Query optimizers are critical for efficiently processing complex queries that characterize current decision support applications.

In recent times, a fresh perspective on the behavior of query optimizers has been introduced through the concept of “plan diagrams”, which are color coded pictorial enumerations of the optimizer’s plan choices over a query parameter space. In this poster, we investigate the plan diagram coloring paradigm, which will immensely help in the analysis of plan diagrams. To determine how different any two specific plans are, we need to drill down into the plans. This process is cumbersome when more than a handful of plans have to be analyzed. We alleviate this problem by semantically coloring plan diagrams. Specifically, we color such that the differences in color between any pair of plans reflects the differences in their plan tree structures. For instance, if the biggest and second biggest plans are very similar, they would both be assigned close shades of the same color, say red. With this new approach to coloring, the plan diagram itself provides a first-cut reflection of the plan-tree differences without having to go through the details of every plan.

The challenges here include designing a quantitative metric for assessing plan differences, and developing transformation techniques for accurately representing these differences in the three-dimensional color model. To assign differences to every pair of plans, we adopt the strategy from our previous work, where plan differences are quantified to be in the interval (0, 1], using the Jaccard metric. For the transformation techniques, we adapt Kruskal’s Iterative Steepest Descent multidimensional scaling method, and test its representational quality by coloring a rich diversity of plan diagrams on benchmark database environments over a suite of commercial database engines. Our experimental results indicate that the vast majority of plan distances can be represented with satisfactory visual accuracy. Given this, we found that, in many plan diagrams, more than half the space is colored with shades of the same color, implying that large areas of plan diagrams are occupied by structurally similar plans.

This feature has been incorporated into the Picasso database query optimizer visualizer tool.
Exploring a Hybrid Approach to Multi-Objective Software Architecture Decision-Making
V.Smrithi Rekha
Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham
v_smrithirekha@cb.amrita.edu

In recent times, Software Architecture is considered to be a set of architectural design decisions (ADDs) and has evolved from a simple component based representation to a combination of Architectural Design Decisions, Rationale and formal representations. ADDs act as a bridge between requirements and design and represent decisions that architects take considering requirements, stakeholder concerns, objectives, criteria and alternatives available. Decisions are often made in the midst of severe constraints and rapidly changing requirements. Owing to the complexity of this process, the support of models and automated tools that can help architects make optimal decisions is necessary. In this paper, a hybrid approach to multi-objective software architecture design decisions has been explored.

Our approach has 2 important phases. In Phase 1, from the requirements specification, through a process of consensus and brainstorming, architects identify problems/questions/issues that need to be solved inorder to design the system. We shall call these “issues”. For each issue under consideration, the architects list several possible alternatives. Once the alternatives are listed the architects score each alternative against the criteria. The score indicates how much each alternative satisfies particular criteria. Also, the criteria are assigned weights depending on their relevance and priority.

In Phase 2, the issues are modeled as chromosomes. Then each criterion to be met is modeled in terms of an objective function, who’s score can be derives from the normalized scores of Phase 1. A genetic algorithm is then applied where the chromosome undergoes a process of mutation and crossover until the population with the best score is obtained. The final fitness value may be calculated as a single aggregate function by using weights.

This approach largely automates the decision-making process. However, it still requires manual intervention in ranking alternatives and criteria and requires architects to have domain knowledge. We propose to expand this work to include formal approaches for scoring the alternatives inorder to make it more accurate.
Semantic Based Peer To Peer Service Discovery in Cloud
Apsara Karen, S., Aiswariya, J., Dr. Saswati Mukherjee
College of Engineering, Guindy, Anna University

We propose to construct a structured overlay based on chord to form multi-ring architecture to support service discovery query applicable in cloud environment. Nodes are arranged in the ring overlay based on the node Identifier. The node ID can be separated into keyword part and random id part to differentiate nodes hosting same set of services. The keyword part of the node identifier can be obtained by referring Global Industry Classification Standard (GICS). In GICS, each company/service is assigned to a sub-industry, and to a corresponding industry, industry group and sector, according to the definition of its principal business activity. Unique id is assigned to each sub-industry, industry, industry-group and sector categories.

The overlay is constructed using the following steps. i) A ring is constructed by the order of the node ID. ii) Another ring is constructed from the nodes which have the same sector. iii) Under each Sector ring, Industry Group sub-rings are constructed which will be equal to number of Industry Groups under the sector. Nodes belonging to same industry group are placed in one of the industry group sub-rings. The same process is repeated to construct Industry sub-rings under Industry group rings and Sub-Industry sub-rings under Industry ring.

When a user issues a service discovery query, a query-id is obtained from the query by using the same rules as generating node IDs. The random part of the node-id becomes the service-id of the service request in the query id. The query id is used to find the Sector ring whose id matches with the Sector id given in the query. After reaching the Sector ring, the query id is again used to find the Industry Group ring whose id matches with Industry-Group of the service requested. The same process is repeated for Industry and Sub-industry. Once we reach Sub-Industry ring, we need to find the exact service that matches the service request by sending queries to the nodes placed in the sub-industry ring. If the name of the service requested in the query in not available in GICS, then Ontology and Semantics are used to find equivalent service in GICS. Thus semantic P2P searching techniques can be applied for efficient resource discovery in cloud computing environments.
Discovering Coverage Patterns for Banner Advertisement Placement

P. Gowtham Srinivas         P. Krishna Reddy                    S. Bhargav
IIIT-Hyderabad                           IIIT-Hyderabad                           IIIT-Hyderabad
gowtham.srinivas@research.iiit.ac.in     pkreddy@iiit.ac.in           bhargav.spg08@research.iiit.ac.in

We propose a model of coverage patterns and a methodology to extract coverage patterns from transactional databases. We have discussed how the coverage patterns are useful by considering the problem of banner advertisements placement in e-commerce web sites. Normally, advertiser expects that the banner advertisement should be displayed to a certain percentage of web site visitors. On the other hand, to generate more revenue for a given web site, the publisher has to meet the coverage demands of several advertisers by providing appropriate sets of web pages. Given web pages of a web site, a coverage pattern is a set of pages visited by a certain percentage of visitors. The coverage patterns discovered from click-stream data could help the publisher in meeting the demands of several advertisers.

In the proposed model of coverage patterns, each pattern is associated with the parameters coverage support and overlap ratio. A naive approach to find the complete set of coverage patterns for a transaction dataset leads to combinatorial explosion and high computational cost. The search space can be reduced if the coverage pattern satisfies downward closure property. However the coverage support measure does not satisfy downward closure property. In the proposed approach, the sorted closure property of overlap ratio measure is used for minimizing the search space. By combining both coverage support and overlap ratio, we have proposed an iterative item set generation and test algorithm similar to the Apriori algorithm.

We have conducted experiments on one synthetic and three real world datasets. It has been shown that the proposed model and methodology can effectively discover coverage patterns. We also demonstrated that, given a coverage support, the proposed approach provides flexibility to the publisher to meet the demands of multiple advertisers by extracting coverage patterns with distinct data items.
Detecting MS Initiated Signaling DDoS Attacks in 3G/4G Wireless Networks
Aman Gupta, Tanmay Verma, Soshant Bali, Sanjit Kaul
IIIT-D, New Delhi, India

The architecture of present day cellular data networks is hierarchical in nature. As a result, large number of base stations (BS) depend on a small number of core network elements, like core network gateways and radio network controllers (RNC), for essential services like Internet connectivity. A core network gateway exchanges several signaling messages with a large number of mobile stations (MS), in addition to handling all Internet traffic destined to (or arriving from) these MS. Internet traffic is handled by ASIC chips (fast path) whereas signaling messages are handled by gateway's CPU (slow path), which is also the central manager that controls other gateway functions, like master-slave state synchronization, programming the network processor forwarding table, slave health monitoring, and etc. A temporary increase in signaling message volume can overload the gateway CPU and paralyse it. This can delay essential time-critical management tasks leading to a temporary Internet outage in a large geographic region. A mobile botnet can exploit this network vulnerability by launching a distributed signaling attack on one or more of core network elements, causing a large number of subscribers to experience service degradation.

In this work, we propose a detector that examines a subset of IP packets transmitted by a MS to determine whether it is infected and can participate in an attack. The detector uses features extracted from IP packet headers to determine if the packets are unusually similar (or unusually dissimilar) to each other. If packet similarity appears very uncharacteristic of a normal MS, then the MS that generated this traffic is classified as an attacker. Service providers can install this detector anywhere in the data path, i.e., for example at the MS, BS, and the gateway, to detect and quarantine infected terminals.

The detector was trained using one week of IP packet traces generated by 62 different smartphones. These traces were used to form labelled training samples of a normal MS. For an infected MS, since our search for an existing signaling attack application did not yield any results, we generated our own attacks. The detector was trained and tested using 7 different attacks. Preliminary results indicate that most types of signaling attacks are detected with a probability of about 0.9 with a false alarm probability of about 0.1.
Resolution Problems in Community Detection in Social Networks

Samridhi Khurana
Indian Institute of Technology (BHU), Varanasi
Samridhi.khurana.apm09@iith.ac.in

Natwar Modani
IBM India Research Lab
Namodani@in.ibm.com

Community: A community in a social network is formed by individuals, who within a group interact with each other more frequently than with the individuals outside the group. A community can therefore be taken to be a module with dense network connections within itself, but sparser connections with other groups.

Community Detection in Social Networks: Modularity\(^{(1)}\) is often used in optimization algorithms for detecting community structures in social networks. The modularity measure compares the number of links inside a module with the expected value of the same for a randomized graph having the same degree sequence.

For a given network with \(n\) nodes and \(m\) links, where each node \(i\) has a node degree \(k_i\), Modularity can be expressed as:

\[ Q = \frac{1}{2m} \sum_{i,j} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \]

Problems lying in the Probability Estimate Term:

- Considering a graph of two nodes, i.e. both \(k_i\) and \(k_j\) being 1, \(P_{ij}\) should be 1. The proposed expression however gives a value of \(\frac{1}{2}\).
- If \(m = k_i + k_j\), then \(i\) and \(j\) have to be connected, which would mean that \(P_{ij}\) is 1, which is not satisfied by the expression.
- \(P_{ij}\) should not increase with the increase in the number of nodes \(n\) in the network, which again is not incorporated in the proposed expression.

Resolution Limit of Modularity:

Modularity is actually not consistent with its optimization which may favor network partitions with groups of modules combined into larger communities.

Ring of Clique Problem:
Consider a ring of \(n\) cliques, each of which has \(m\) nodes. Ideally, all the cliques should be identified as separate communities. The modularity measure, instead, tends to merge up two communities together.

\(Q_{\text{single}} > Q_{\text{pairs}}\) is satisfied only when \(m(m - 1) + 2 > n\).

Our Approach: Applying Recursive BGLL Algorithm

**Step 1:** The communities are identified using BGLL Algorithm\(^{(1)}\). The found communities will have a hierarchical, nested and overlapping structure. As a fact, the smaller communities would be merged with some of the bigger ones.

**Step 2:** The large communities found in Step 1 are considered, and BGLL algorithm is applied on them. This step is repeated until modularity is optimized and all the communities detected are not bigger than the specified threshold.
Data Understanding using Semi-Supervised Clustering
Rashmi Dobarial (Agnity India), Priya Jain (Aricent Group India), Ashish Mahabal (Caltech), Vasudha Bhatnagar (University of Delhi)

Abstract

In the era of E-science, most scientific endeavors depend on intense data analysis to understand the underlying physical phenomenon. Predictive modeling is one of the popular machine learning tasks undertaken in such endeavors. Labeled data used for training the predictive model reflects understanding of the domain. In this paper we introduce data understanding as a computational problem and propose a solution for enhancing domain understanding based on semi supervised clustering.

The proposed DU-SSC (Data Understanding using Semi-Supervised Clustering) algorithm is efficient, scalable, incremental, takes resolution (only user parameter) and performs single scan of data. Algorithm discovers unobserved heterogeneity among classes in labeled data, with the intent to improve understanding of domain. Given labeled (training) data is discretized at user specified resolution and placed into a imaginary grid of unit volume, representing the entire data space. Regions of data with similar classes (micro-distributions) are identified along with outliers, based on automatically computed threshold. The discovery process is based on grouping similar instances in data space, while taking into account the degree of influence each attribute exercises on the class label. Maximal Information Coefficient measure is used during similarity computations for this purpose.

The study is supported by experiments and a detailed account of understanding gained is presented for three selected UCI data sets. General observations on ten UCI datasets are presented, along with experiments that demonstrate use of discovered knowledge for improved classification.
Energy Efficient Data Center Networks - A SDN based approach
Dharmesh Kakadia and Vasudeva Varma, SIEL, IIIT-Hyderabad

The success of cloud computing is motivated by the ability to deliver reliable services while operating at very large scale. While the use of commodity hardware at scale, allows operators to amortize capital investments, the operational cost (OPEX) of a cloud-scale infrastructure is an obstacle for sustainable growth of next generation systems. The energy cost is a key component of OPEX and network components account for nearly 15\% of the amortized cost. Our goal is to minimize the energy consumption of the data center network without adversely affecting the network performance. Current networks are designed for peak load and are not power proportional. Previous optimization based solutions do not consider the utilization history and can lead to instability. We propose a SDN based approach to dynamically decide, the minimal set of network switches required to meet the demand on network. Software Defined Networking (SDN) refers to the networking architecture where control and data plane functionalities are separated, and control plane is implemented in software. OpenFlow is a leading, open interface for SDN architectures, which gives the centralized abstraction for remotely controlling the forwarding tables in network devices.

Our algorithm uses OpenFlow counters to collect traffic information on switches. It identifies all the switches, which has the utilization below the administrator controlled threshold \( \theta \), observed over time period \( \tau \). From each of these switches, we try to migrate the flows to other switches, incrementally using OpenFlow controller. If we are able to consolidate all the flows from the switch, it can be powered off or put to low-power state. While consolidating, we keep some flow capacity of switches (20\% during our experiments) reserved to take care of sudden surge in traffic.

We measured the effectiveness of the consolidation algorithm using mininet simulator on synthesized traffic. The performance-penalty due to consolidation was measured in terms of maximum delay experienced by the flow packets. During our experiments, the delay variance changed to 30 \( \mu \) s from 22 \( \mu \) s and the average delay increased to 99 \( \mu \) s from 84 \( \mu \) s, after applying consolidation.

Our algorithm was able to achieve nearly linear power consumption behavior with traffic load on the network, maintaining the delay characteristics within small bound. We have demonstrated that simple techniques can result into significant energy savings without compromising on the performance of the network.
Hosting Keyword Search Engine inside RDBMS

Amit Mehta (IISc)

Keyword search (KWS) is a widely accepted mechanism for querying in Information Retrieval (IR) systems and Internet search engines on the Web. They offer convenient keyword based search interfaces. But searching in relational database systems the user needs to learn SQL and to know the schema of the underlying data even to pose simple searches. KWS should give an easy interface to RDBMS, which does not require the knowledge of schema information of the published database. Most of the works on KWS engines use main memory data structures to perform required computations to get good performance on execution time and RDBMS is used as storage repository only. Our work focuses on effective utilization of RDBMS technologies to process computations involved in providing KWS interface. By this we can get additional benefits from RDBMS back-end technologies to handle large databases and to have persistent KWS indexes. We are also trying to make Database engine aware of the kind of indexes available for keyword search because currently these indexes are used as only relation tables and not a part of execution plan tree of structured queries. Also to get semantically stronger ranked results efficiently, we are trying to make processing of top-$k$ queries more efficient. Two prominent database models used by KWS engines are schema graph based and data graph based KWS models. We have chosen Labrador KWS engine [1] as a representative of schema based KWS model and built DBLabrador, which is functionally similar to Labrador, uses RDBMS to perform all computations and uses additional keyword index. Labrador needs to build keyword index every time before it is functional while DBLabrador gives persistent keyword indexes by storing keyword indexes as relational tables. DBLabrador keep cell-granularity term frequency information along with column-level term frequency information and does extra computation to materialize weight of each term in cell-granularity level. Experiments with various real data sets show that DBLabrador’s performance in populating keyword index is comparable with Labrador’s performance and DBLabrador’s performance in getting answer tuples and their ordering is better when full-text index is not available on published attribute compared to Labrador and its performance is almost equal to Labrador when full-text index is present on attribute as both use same SQL query to generate answer tuples. In data graph based KWS model, we have taken (PBKSC) [2] work, which is based on distinct root semantic answer model. We introduced an alternative keyword index, Node-Node, instead of Node-Keyword index to reduce the storage space consumed by the keyword index. By using properties of Node-Node index issues related to storage space of keyword index can be effectively solved by compromising with query search time. Compared to Node-Keyword index, Node-Node index uses less storage space for text based databases and gives comparable performance. Also Node-Node index uses RDBMS back-end technologies in self join procedure, which allows it be operated with small threshold path weight to get same quality of answer produced by high threshold path weight Node-Keyword index, and in the process to produce CTS answer model, where Node-Keyword index approach depends on main memory procedures.

**ROSHNI: Recreating Optical SigHt by virtually Navigating Indoors**
Ashwini Choudhary, Devesh Singh, Dhruv Jain, Himanshu Meenia, Kartik Singh, Manas Sahu, Mohit Taak, Naman Gupta, Nitin Kamra, M. Balakrishnan

Assistive Technologies Group
Department of Computer Science and Engineering
Indian Institute of Technology (IIT), Delhi
Email: dhruv.jain.iit@gmail.com

**Introduction:** Among the many challenges faced by the visually challenged persons are the constraints of independent mobility and navigation in an unfamiliar indoor environment. Finding the path to some desired location including public utilities inside the building can be a daunting task. GPS based navigation that is now getting widely used does not work in an indoor environment. Although several attempts have been made at making such indoor navigation systems, none of them have found wide acceptance in a developing country scenario. We present the design of a cell-phone based indoor mobility system for blind persons that can help them navigate independently in an unfamiliar building without any external sighted assistance. Our solution requires minimal additional building infrastructure and can be easily retrofitted onto existing buildings. Further, it is easy to use, low-cost and can be attached to the user's waist or cane.

**Approach:** We have developed a portable and self-contained system fabricated using the infra-red sensor-suite and inertial unit. The system consists of 3 modules (i) a waist-worn user module; (ii) network of wall-mounted units spaced at distance intervals; and (ii) a mobile application. By pressing keys on his/her mobile unit, the user can obtain acoustic directions to any desired location on the map from his/her current position.

**Experiment Results:** The system was installed on multiple floors of a university building and evaluated by 6 visually impaired users in real-life settings using a standardized protocol. Questionnaire based interviews were conducted before and after trials to obtain their feedback.

During questionnaire based interviews, majority of users expressed that indoor navigation is a day-to-day problem and reported seeking help from persons nearby is difficult (especially in female users). Often, they get lost inside the building which causes them extreme inconvenience as well as a delay in reaching their destination. During the experimental trials, the users successfully navigated on different paths covering multiple floors. Feedback from users shows that the device is easy to learn and can provide a sense of independence to them.
Efficient Constrained Shortest Path Estimators
Ankita Likhyani  Srikanta Bedathur
Indraprastha Institute of Information Technology, New Delhi
{ankita1120,bedathur}@iiitd.ac.in

Shortest path computation has been studied extensively in graph theory and computer networks, but the processing of voluminous data sets require efficient shortest path algorithms for disk-resident data. In applications arising in massive online social networks, biological networks, and knowledge graphs it is often required to find many, if not all, shortest length paths between two given nodes. In this work, we generalize this further by aiming to determine all shortest length paths between two given nodes which also satisfy a user-provided constraint on the set of edge labels involved in the path. It seems quite straightforward to incorporate such label constraints over an online shortest path algorithm such as the Dijkstra’s algorithm. But social networks and knowledge networks that we consider in our work lack near-planarity, hierarchical structure, and low-degree nodes that are critical for the efficiency of algorithms. This calls for algorithms with moderate pre computation cost and index size, yet still fast query time.

There exist landmark-based methods for point-to-point distance estimation in [1] and label constraint reachability methods in [2] for constraint set reachability between two nodes in very large networks. In this work, we propose a solution which extends the above two ideas by maintaining an indexing structure which stores distances from all nodes to landmarks and minimal sufficient path label set between two nodes. During query processing this index framework is used to determine all the shortest length paths between two given nodes which also satisfy the user-provided constraints on the labels of nodes/edges involved in the path. Currently, the reachability problem has been addressed on a graph size of around 100,000 nodes, so here we also aim at building an efficient solution which is scalable to graphs containing millions of nodes.

References:
Is Government a Friend or a Foe? Privacy in Open Government Data

The awareness and sense of privacy has increased in the minds of people over past few years. Earlier, people were not very restrictive in sharing their information, but now they are more cautious in sharing their personal information with strangers. For e.g. on the social networking sites like Facebook, people try to manage their account settings such that minimum information can be viewed by a person not in his / her friend list. With such mind set, it is difficult to embrace the fact that a lot of information is available publicly on the web. The information portals in form of the e-governance websites run by Delhi Government provide access to such personal information. This is privacy invasive and can be harmful to citizens.

This project aims to identify (in terms of what information can be extracted) such information and analyse possible privacy and related concerns that may arise out of public availability and easy accessibility of such data. The purpose is to showcase that such large amount of data is available openly and to demonstrate to the government and citizens that their privacy is at stake and can be misused for unlawful activity. We believe this will increase the awareness among Indian citizens about privacy and also help the government agencies make informed decisions about what data to make public.

The next stop was to spread awareness among the general public about the possible misuse of their personal data without their notice and consent. For this, we developed a system which could process the query furnishing the details of the users. [http://precog.iiitd.edu.in/profiling/]. It takes little user information as an input and gives all possible data about him / her as the output as shown in Figure1.
Enhancing the Query-by-Object Approach using Schema Summarization

Information Requirement Elicitation (IRE) recommends a framework for developing interactive interfaces, which allows users to access database systems without having prior knowledge of a query language. An approach called ‘Query-by-Object’ (QBO) has been proposed in the literature for IRE by exploiting simple calculator like operations. The QBO approach provides a web-based interface for building a query using multiple user level steps. In this approach, the user communicates with a database through a high level interface. The initial intent of the user is captured via selection of objects from an object menu. The user navigates to select granularity of these objects and operators to operate between the selected objects. The user’s actions are kept track in a query-bag, visible to the user at all stages. Finally, a SQL equivalent query is formulated and is executed at DBMS server. The QBO approach uses a database to store the objects and entities. However, for large databases, the QBO approach does not scale well. Large number of tables in the schema makes it harder for the user to locate his information of interest. Secondly, with large number of tables, the number of pairwise operations between tables also increase.

In this paper, we propose an enhanced QBO approach called Query-by-Topics (QBT), for designing calculator like user interfaces for large databases. In the proposed approach, first we identify semantically correlated elements in the database schema, representing what users perceive as a single unit of information (topics). We discuss clustering based approaches for detecting topical structures from a database by utilizing the database schema, the data stored in the database and the database documentation. Secondly, instead of defining operations between each pair of tables, we define operations between topics and within topics, reducing the number of pairs for which operators have to be defined. We also design a prototype system based on the QBT approach. The prototype is based on client-server architecture. Users interact with the system by means of a web-based user interface, analogous to the interface of a traditional calculator. The user interface allows for objects selection, operator selection and also displays query results. The back-end consists of a system which processes the inputs given by user and generates an SQL query which is executed on a relational database server (MySQL). To analyze the effectiveness of the proposed approach, we conducted a usability study. The usability study consists of an ease of use survey on a real database using real users. We developed two prototypes, one based on the QBO approach and the other based on the QBT approach. We asked the users to explore the database and pose queries from their day-to-day requirements using both the prototypes. After the session, they filled out a questionnaire, rating the prototypes. The results from the usability study are encouraging.
A Framework for Corruption Assessment in Public Services
Nidhi Rajshree, Nirmit Desai, Biplav Srivastava, Anshu Jain, Anuradha Bhamidipaty, Kamal Bhattacharya
Email: {nidhi.rajshree, nirmit.desai, anshu.jain, abhamidi, kambhatt, sbiplav}@in.ibm.com
IBM Research, India

Corruption ails most of the human delivered services especially public services leading to negative impact to the growth of economies. It undermines the fairness, stability and efficiency of these services, corrodes the image of the public agencies and projects the state as predatory and unfair. Corruption does not necessarily occur by intention, i.e., through a design of a corruptible public service to be exploited by malicious individuals. Corruption may simply emerge as service providers and consumers misuse fragments of service designs for personal gain. We believe that business process-centric approaches to service design can aid in the detection of corruptible process designs that may potentially lead to misuse. In this paper, we introduce three patterns of corruptible process designs: monopoly, lack of SLAs, and human discretion. The patterns focus on petty corruption that involves smaller sums of money and pervades the low level bureaucrats. This form of corruption covers the majority of cases reported by the service consumers. We also present a meta-model that allows expression of corrupt aspects of service designs. We study several scenarios of corruption in multiple domains of public services and generalize them into distinct patterns of corruptibility. The general patterns are valuable because they represent common root causes of corruption and hence enable a study of general remediation one might apply to mitigate corruption regardless of the domain. In future, such patterns can be formalized in a suitable vocabulary such that a static analysis against public service processes can be carried out to automatically detect corruptible aspects of processes as well as suggest remediation. Whether or not these patterns are complete with respect to a universe of corruptible elements is a topic for future research. Both the patterns and the meta-model can augment the well-known service design languages such as BPMN and SBVR. We believe that public services can benefit from such an approach that avoids corruptibility of processes at design time, hence making them more robust and less prone to corrupt practices. This does not make it impossible to undermine the system, but it may create the necessary transparency required to establish the appropriate trust relationship between the service providers and their customers.
Mechanism Design for Resource Critical Crowdsourcing

Swaprava Nath¹, Pankaj Dayama², Dinesh Garg², Y. Narahari¹, and James Zou³
¹Indian Institute of Science, Bangalore, swaprava@gmail.com, hari@csa.iisc.ernet.in
²IBM India Research Lab, pankajdayama@in.ibm.com, garg.dinesh@in.ibm.com
³Harvard School of Engineering and Applied Sciences, jzou@fas.harvard.edu

Abstract

Crowdsourcing over social networks has recently emerged as an active tool for complex task execution.

In this paper, we address the problem faced by a planner to incentivize agents in the network to execute a task and also help in recruiting other agents for this purpose. We study this mechanism design problem under two natural resource optimization settings: (1) cost critical tasks, where the planner's goal is to minimize the total cost, and (2) time critical tasks, where the goal is to minimize the total time elapsed before the task is executed. We define a set of fairness properties that should be ideally satisfied by a crowdsourcing mechanism. We prove that no mechanism can satisfy all these properties simultaneously. We relax some of these properties and define their approximate counterparts. Under appropriate approximate fairness criteria, we obtain a non-trivial family of payment mechanisms. Moreover, we provide precise characterizations of cost critical and time critical mechanisms.
MODA: A Middleware for Policy-aware Adaptive Mobile Data Collection

Hemant Kowshik, Palanivel Kodeswaran, Ravindranath Kokku and Vinay Kolar
IBM Research, India.

{hkowshik, palanivel.kodeswaran, ravkokku, vinkolar}@in.ibm.com

Abstract

With the widespread adoption of smartphones, ubiquitous and continuous personalized sensing is now a reality. Applications like activity monitoring, traffic sensing and citizen science use sensor data to determine context and intent of a mobile user. In the current ecosystem, each application operates alone, building a userbase from scratch, sensing the environment and learning user behavior models. This results in unnecessary sensing and communication overheads that prevent the effective scaling of such applications. In this paper, we propose an efficient and configurable middleware that supports datacentric queries on real-time mobile data. Such a middleware has the potential to unlock the power of the mobile sensor web, while causing minimal impact on users, and delivering maximum value to application developers and user alike.

The platform is configurable since it allows each user to specify policies regarding the acceptable communication footprint, battery usage and attribute privacy. The platform is efficient since it eliminates redundant communication and sensing, caused by applications with overlap in their sensor requirements. Further, one can dynamically adapt the frequency of updates of each sensor so as to save energy and bandwidth. Finally, the platform automatically identifies correlations between users and builds models of user behavior online. These models can be used to further compress data in subsequent application queries. These models can be used to provide applications a richer semantic description of user context and behavior. On the other hand, they can be used to package raw data in a privacy-preserving fashion to applications.

Figure 1: System Architecture of MODA
Eliciting Honest, High Quality Feedback from Crowdsourced Tree Networks using Scoring Rules
Ratul Ray, Rohith D Vallam, Y. Narahari
Indian Institute of Science, Bangalore

Abstract

Eliciting accurate information on any object (may be a product, person, or service) by collecting feedback from a crowd of people is an important and interesting problem in web-based platforms such as social networks. Peer prediction networks represent one of the known efforts in that direction. We generalize the peer prediction model to the natural setting of a tree network that gets formed when a query that originates at a (root) node propagates to different nodes in the social network or crowd in a hierarchical fashion. The feedback received from the nodes in the tree must be honest and accurate, and the feedback also must be aggregated in a hierarchical fashion to generate a high quality answer at the root level. To motivate the nodes to put in sufficient effort and report truthfully, it is required to incentivize them appropriately. We investigate this problem by proposing a generic hierarchical framework using scoring rules to incentivize the nodes to elicit truthful reports from the nodes. We consider two different scenarios based on whether or not prior probabilities are common knowledge of the nodes in the tree and discuss the way our approach can be employed in the two settings. Through simulation experiments, we validate our findings and study the relationship between the budget of the mechanism designer budget and the quality of answer generated at the root node. In the context of this problem, we also compare the performance of three well known scoring rules: logarithmic, quadratic, and spherical.
LOW COST DIGITAL TRANSCEIVER DESIGN FOR SOFTWARE DEFINED RADIO USING RTL-SDR

Abirami M, Akhil Manikkoth, Sruthi M B
Department of CEN
Amrita Vishwa Vidyapeetham
Coimbatore
akhil001@gmail.com

Gandhiraj R
Assistant professor
Department of ECE
Amrita Vishwa Vidyapeetham
Coimbatore
r_gandhiraj@cb.amrita.edu

Dr Soman K P
Professor and HOD
Department of CEN
Amrita Vishwa Vidyapeetham
Coimbatore
kp_soman@amrita.edu

Abstract

The field of wireless communication has become the hottest area and Software Defined Radio (SDR) is revolutionizing it. By bringing much functionality as software, SDR reduces the cost of hardware maintenance and up-gradation. Open source hardware such as USRP (Universal Software Radio Peripheral) and software called GNU Radio-Companion are commonly used to do experiments in SDR. Since the cost of USRP is high, a low cost set up is needed which is affordable by the student community. In this paper a low cost alternative to USRP is proposed using RTL-SDR which is only used for reception. A $20 revolution from OSMO SDR has introduced a hardware called RTL-SDR Realtek RTL2832U which is the cheapest one. The DVB-T (Digital Video Broadcast Terrestrial) dongle proved to be efficient for SDR purposes as the chip is able to transmit raw I/Q samples to the host. The operating frequency range of RTL-SDR is from 64 to 1800 MHz, with sample rate of 3.2 MS/S. For transmitting purpose, a mixer circuit can be used to map the base band signal to the band that can be received by RTL-SDR on the other end on Linux / Windows platform. The mixer gives unlimited access to the frequency range 64 to 1700 MHz for reception of all modes. An oscillator block with an output frequency of 106.25 MHz allows simple frequency translation from low frequency to higher frequency. Mixer NE602AN is a double balanced mixer and oscillator which is intended for high performance, low power communication systems. Real time implementation of mixer circuit along with RTL-SDR is possible in both Linux and Windows operating system. In Linux, the platform for reception is GNU Radio-Companion by using RTL-SDR. In windows; open source software like HDSDR, SDR# is available for reception using RTL-SDR. The cost for total transceiver system can be less than USD 100 which is 10 times less than the existing one.
Efficiently Scripting Change-Resilient Tests
Nimit Singhania, Pranavadutta DN
IBM Research, India

Abstract

In industrial practice, test cases often start out as steps described in natural language that are intended to be executed by a human. Since tests are executed repeatedly, they go through an automation process, in which they are converted to automated test scripts (or programs) that perform the test steps mechanically. Conventional test-automation techniques can be time-consuming, can require specialized skills, and can produce fragile scripts. To address these limitations, we present a tool, called ATA, for automating the test-automation task. Using a novel combination of natural language processing, backtracking exploration, runtime interpretation, and learning, ATA can significantly improve tester productivity in automating manual tests. ATA also produces change-resilient scripts, which automatically adapt themselves in the presence of certain types of user-interface changes.
Preserving Date and Time Stamps for Incident Handling in Android Smartphones
Robin Verma (IIIT-Delhi), Gaurav Gupta (IIIT-Delhi)

Present day smartphones act as mobile offices, entertainment hubs and a social tool - all packed into one compact device. Over the last decade, the processing and data storage capability of a typical smartphone has evolved to the level of a typical personal computer or a laptop. Important data that are commonly found in most smartphones include call logs, contact list, text messages, emails, photos, videos, web-history, application data, eBooks, and maps. With this much personal data to manage, employees at corporations prefer to carry their own device (often called Bring Your Own Device, BYOD) to their workplace, instead of getting another phone from their employers. Employers can save cost of device whereas for employees this proves to be convenient way of managing their personal and professional data on a single device. Although this trend is catching up in corporations around the world, tracking and controlling access to private and corporate networks is still a major challenge. The biggest threat in BYOD model is the security of data on the device, which is not only valuable for the employee but also equally or more valuable for the employer. People with malicious intent often either attempt to get access to this valuable data or try to tamper it. In both the cases metadata information, including Modification, Access, Change and/or Creation Date and Time Stamps (MAC DTS), of files which were accessed or tampered invariably gets changed. MAC DTS of digital data present in smartphones could be very crucial and most fundamental evidence in almost all such cases, thus establishing authenticity of available MAC DTS is of prime importance for forensics investigators. Malicious tampering of MAC DTS mainly consists of changing date and time stamps and/or contents of the files or both. Commercial tools including Cellebrite UFED System, FTK Mobile Phone Examiner and other mobile phone imaging and analysis tools provide solutions to recover data. However most of them prove to be inadequate when trying to establish the authenticity of MAC DTS.

The research work aims to detect such malicious actions by capturing the date and time stamps along with location details and snapshot of changes in data items in a secure place outside the smartphone. The system generated MAC DTS values are captured with the help of a Loadable Kernel Module (LKM) which hooks on to the system calls to get these values. A secure place outside the phone can be a cloud or a local server in the Enterprise. The cloud snapshot of authentic MAC DTS values can be used later to check the authenticity of MAC DTS values of questioned files on the phone.
ForeCAST: Social-web assisted Channel-aware Predictive Multicast over Wireless Networks
Giridhari Venkatadri (IIT Madras) and Hemant Kowshik (IBM Research)

Consumption of video on mobile devices is becoming increasingly popular and is expected to explode over the next few years, placing severe bandwidth constraints especially on the wireless last-hop. This bandwidth crunch would mean severe degradation in the quality of experience of users, most marked at peak load times. Though a number of videos, for example Youtube videos, are watched in an on-demand fashion, video watching is becoming increasingly social (i.e. people watch videos based on activity in social networks) and hence more predictable. In this paper, we propose a novel predictive content delivery system that exploits this predictability, using user-specific video uptake probabilities generated from activity on social networks to proactively and opportunistically multicast content so as to provide bandwidth savings and enhance quality of user experience. The system proactively pushes trending videos over spare bandwidth and opportunistically tries to multicast content over actual requests for the content. We propose a novel objective (Expected Bandwidth Savings) required for such a system. Using a threshold channel model we formulate the problem of choosing multicast groups for each video so as to maximize expected bandwidth savings as a Multiple-Choice Knapsack Problem (MCKP) for which efficient approximations are available. We also discuss how to handle limitations in terms of memory space and battery of the mobile. Simulations show promising savings in bandwidth and reduction of load during peak load times. Also, we consider an alternate content-delivery mechanism (Broadcast using Fountain Codes) and propose a greedy yet optimal algorithm to schedule broadcasts so as to maximize expected bandwidth savings.
Global service delivery heavily relies on effective cross-enterprise collaboration. Service systems in this domain are human-centric in many ways. The service outcomes produced by a service provider are experienced by the service consumer. The service provider typically leverages service workers to produce service outcomes that satisfy the consumer. The basic functional unit of a service system is a human worker who delivers work either individually or by working with a team of co-workers. In this context, a key human factor that needs investigation is the skill of the service workers. Not only are skills essential in producing satisfactory service outcomes but they also evolve as the service workers accumulate experience. The outcome of a worker’s task and his performance is typically assumed to be a function of his current skills. However it has been shown that the team’s performance can progressively improve with skill evolution of its workers. An understanding of skills and how they are gained and lost by workers can help answer key questions about management of human resources. There may be multiple methods to evolve the skills of a team: (a) class room training, where people’s times are blocked for a certain duration (b) shadowing, where a low-skill worker is expected to closely follow the work of a highly skilled person or (c) on-the-job training, where people pick up skills while actually doing the work. Specifically, this paper explores “on-the-job training”, where skill evolution of workers is achieved by challenging them with work assignment that requires skills beyond their current capabilities. This serves a two-fold purpose: (1) it helps human workers achieve personal satisfaction as they learn something new and (2) it helps in more effective work completion for the team. However, a challenge with on-the-job-training is managing the risk of task failure and eventual SLA violation. In this early exposition, we present a reinforcement learning-based technique for learning the work assignment policy that achieves a state of target skills for service workers while incurring minimum costs in terms of SLA violations. This technique is being applied to real-life service systems at a leading IT services provider. We have found that relative to a cost function, our method can recommend the fastest training plan without compromising SLA performance.
Topic Expansion using Term Co-occurrence Analysis
Sumant Kulkarni and Srinath Srinivasa, IIIT-Bangalore

*Topic expansion* for a given topic \( t \) is the process of unfolding \( t \) into a set of related concepts that are collectively “about” \( t \). Each term in a language may represent one or more senses or sub-senses. Hence, when we speak of a term, we speak of multiple topics (senses) the term refers to. When we utter the term *Java*, we might either be talking about *Java Island* or *Java Programming language*. On the other hand, when we mention *Himalaya*, we might be talking about *Himalayan Mountains* or *Himalayan Vegetation*. Likewise, each term in a language refers to many different senses. Sometimes these different senses are highly unrelated (like *Java*) and sometimes are much related (like *Himalaya*). Each of these senses represents a *topic* and not just the meaning. This is a departure from classical Word Sense Disambiguation. We expand these *topics* in terms of other related concepts. For example, in the first case, we can expand the topic referring to *Java Island*, as \(<\text{Java}, \text{East java}, \text{Central java}, \text{Yogyakarta}, \text{Javanese}>\). Similarly, we expand the same term for the sense *Java Programming Language* as \(<\text{Java}, \text{Programming Language}, \text{C++}, \text{Programming}, \text{Perl}, \text{Platform}>\). The expansion of a topic \( t \) consists of terms that lexically *co-occur* with \( t \). The *co-occurrence* patterns of other terms with *Java* ensure that it acquires the meaning of *Java Island* in the first case, and *Java Programming Language* in the second case. Topic expansion using lexical co-occurrence analysis, as explained above, is the objective of our work. We first separate the senses of a term and then unfold each of the topics (senses) using an ordered set of related concepts based on their co-occurrence patterns.

The co-occurrence algorithm based on 3-layer cognitive model\[^3\] considers each co-occurring term of a topic \( t \) as a separate dimension, and creates topic expansion clusters for each of them. These clusters represent all the major senses \( t \) acquires in the corpus. Based on the degree of overlap, we merge clusters with similar senses to form a single cluster for each of the distinct senses. We optionally drop extraneous clusters that are noisy and do not represent any distinct sense. Now, we are left with topic expansion clusters representing different senses of \( t \). We order the terms within each cluster according to their relevance to the topic. We hypothesize that the tightness of the coupling between a given sense of \( t \) and any given related term in the underlying corpus, represents the relevance of the given term in explaining that sense of \( t \). This relevance is calculated as a function of the “exclusivity” of the term in the context of that sense of \( t \). We find that our approach to topic expansion performs better than LDA and graph based Word Sense Disambiguation\[^1\]. Comparison of results between topic expansion and LDA\[^2\] can be found at: [http://bit.ly/QJMqdq](http://bit.ly/QJMqdq).

Optimal Incentive Strategies for Product Marketing on Social Networks

Pankaj Dayama\(^1\), Aditya Karnik\(^2\), and Y. Narahari\(^3\)

\(^1\)IBM India Research Lab, pankajdayama@in.ibm.com
\(^2\)General Motors Global R&D, aditya.karnik@gm.com
\(^3\)Indian Institute of Science, Bangalore, hari@csa.iisc.ernet.in

ABSTRACT

We consider the problem of devising incentive strategies for viral marketing of a product. In particular, we assume that the seller can influence penetration of the product by offering two incentive programs: a) direct incentives to potential buyers \((\text{influence})\) and b) referral rewards for customers who influence potential buyers to make the purchase \((\text{exploit connections})\). The problem is to determine the optimal timing of these programs over a finite time horizon. In contrast to algorithmic perspective popular in the literature, we take a mean-field approach and formulate the problem as a continuous-time deterministic optimal control problem. We show that the optimal strategy for the seller has a simple structure and can take both forms, namely, \textit{influence-and-exploit} and \textit{exploit-and-influence}. We also show that in some cases it may optimal for the seller to deploy incentive programs mostly for low degree nodes. We support our theoretical results through numerical studies and provide practical insights by analyzing various scenarios.
Resource Allocation in the Presence of Strategic Users with Near Budget Balance  
Thirumulanathan D and Rajesh Sundaresan

We consider the problem of allocating a single divisible good to a number of strategic users. Each user has a scalar type that represents how much he values the divisible good, and this type is his private information. Users are strategic and could potentially misrepresent their types to maximize their rewards. We design a mechanism that extracts this scalar type truthfully from every user, allocates the good efficiently, and, in addition, achieves “near” budget balance. To satisfy these properties, we consider a family of VCG mechanisms with linear rebates, where the users are asked to pay according to the Clarke pivotal mechanism for the portion of the good they receive, and the collected payments are provided back to the users as rebates. Now we additionally require the sum of rebates not to be more than the sum of payments received, as otherwise the auctioneer will be forced to pay from his pocket. We also want the utility of every user to be non-negative, as otherwise the user would pull out of the allocation process rather than participate. We model this problem as a type of mathematical programming problem called an uncertain convex program, and provide an upper bound on the number of samples sufficient for approximate optimality.

Situations such as allocation of a single divisible good to strategic users with all the aforesaid properties arise in any setting where there is a public good to be distributed, and the auctioneer is not interested in maximizing revenue. Qualitative examples include disbursement of funds for projects within a parent organization, and allocation of carbon credits across countries. Our results are applicable in all such settings where the auctioneer is not interested in maximizing his revenue, but only desires efficiency.

REFERENCES
Web search has become ubiquitous and with the availability of huge volumes of data on web and effective retrieval systems, user can now obtain good-quality information relevant to his need easily. But, unfortunately, in many cases, the required information is scattered over multiple sources. As a result, user may find it difficult to gather information from the sources for the queried topic and extract relevant and diverse information to form a digest of the topic. Also, user may not have enough time to read all the documents and therefore, some information may be left unread. Although, Wikipedia facilitates the availability of comprehensive content about a lot of entities it is not exhaustive. Less popular entities and events still remain unrepresented in it.

Entity-centric summarization assists the user by presenting a single document containing relevant information about the entity searched. Sauper et. al. [1] used high-level structure of human-generated text to automatically create domain specific templates and learned Topic-specific extractors for content selection. Filippova et. al. [2] obtained company-specific summaries from financial news on basis of relatedness to the company symbol, overall importance and novelty. However, above approaches are limited to domain-specific queries.

In this work, we propose a framework consisting of the following three key modules:

1) **Related Entity Finding:** Every entity is generally associated with various other entities. For Example, entity “Manmohan Singh” is associated with “India”, “Prime Minister”, “Sonia Gandhi”, etc. This module finds the entities associated with the input entity.

2) **Support Sentence Identification:** It identifies the sentences describing the relationship between found entity pairs.

3) **Diversity in Ranking:** It eliminates the redundant information in the support sentences. Overall, our approach is based on identifying sentences depicting all the important entity pair relationships, followed by, ranking and diversifying these sentences to form a summary of the topic.

**References**


Learning to Propagate Rare Labels for Fraud Detection
Deepesh Bharani\(^1\), Dinesh Garg\(^2\), Rakesh Pimplikar\(^2\), Gyana Parija\(^2\)

The presence of fraudulent users is inevitable in any e-Commerce, m-Commerce, and online social networking applications. Such users pose a serious threat to these services and detecting them is a non-trivial challenging task from the perspective of both modeling as well as computational requirements. Almost any outlier detection technique, which is based on just user centric features, would fail here because it is very easy for such users to tweak the values of the features in a way to circumvent the filter. On the other hand, the users inherently carry out transactions among themselves in these applications. These transaction data are less noisy and contain valuable information regarding users’ behavior. Therefore, it would be apt to leverage these data for the purpose of identifying fraudulent users. In this paper, we address this problem by proposing an innovative label propagation method over the network of users where users are connected based on the transactions they carry out. We deal with two main challenges in our work – (1) Labeled dataset is usually very small, say of the order of 0.1\%, and (2) Number of fraudulent users in the system is very less, say of the order of 0.5\%, which leads to highly skewed class distribution scenario for the label propagation. The initial set of experiments performed on KDD’99 cup dataset give very encouraging results.

\(^{1}\) IIT Delhi, New Delhi. deepeshbharani.iitd@gmail.com

\(^{2}\) IBM Research, New Delhi, India. \{garg.dinesh, rakesh.pimplikar, gyana.parija\}@in.ibm.com

38
Towards Efficient Named-Entity Rule Induction for Customizability

Generic rule-based systems for Information Extraction (IE) have been shown to work reasonably well out-of-the-box, and achieve state-of-the-art accuracy with further domain customization. However, it is generally recognized that manually building and customizing rules is a complex and labor intensive process. In this paper, we discuss an approach that facilitates the process of building customizable rules for Named-Entity Recognition (NER) tasks via rule induction, in the Annotation Query Language (AQL). Given a set of basic features and an annotated document collection, our goal is to generate an initial set of rules with reasonable accuracy, that are interpretable and thus can be easily refined by a human developer. We present an efficient rule induction process, modeled on a four-stage manual rule development process and present initial promising results with our system. We also propose a simple notion of extractor complexity as a first step to quantify the interpretability of an extractor, and study the effect of induction bias and customization of basic features on the accuracy and complexity of induced rules. We demonstrate through experiments that the induced rules have good accuracy and low complexity according to our complexity measure.

Authors:

Ajay Nagesh, IIT Bombay, ajaynagesh@cse.iitb.ac.in
Ganesh Ramakrishnan, IIT Bombay, ganesh@cse.iitb.ac.in
Laura Chiticariu, IBM Research – Almaden, chiti@us.ibm.com
Rajasekar Krishnamurthy, IBM Research – Almaden, rajase@us.ibm.com
Ankush Dharkar, SASTRA University, ankushdharkar@cse.sastra.edu
Pushpak Bhattacharyya, IIT Bombay, pb@cse.iitb.ac.in

Presenter: Ajay Nagesh

Note: This work was published in EMNLP-CoNLL 2012, Jeju, Korea
Traffic Congestion Surveillance by Neighborhood based Spatial Anomalous Region Detection

By

Ranjana Rajendran + Aditya D Telang ‡ Deepak S Padmanabhan ‡ Prasad M Deshpande ‡

University of California Santa Cruz + IBM Indian Research Lab ‡

Abstract: Traffic congestion detection methods involve installation of expensive infrastructure sensors to monitor traffic conditions or cooperative techniques involving vehicle-to-vehicle communication which apart from requiring some pre-designed synchronization protocol between the vehicles are unreliable because it is prone to all issues as are any data communication network. GPS, which collects the speed and direction of motion of its host, provides a reliable and easy method to explore location traces of moving objects. In this paper we intend to give an algorithm to detect locations of congestion of object movement using real time location traces obtained by GPS devices. The use cases of this algorithm will involve detection of congestion as in a traffic jam, rally of people or swarming of moving objects, which involve noticeable decrease in speed of movement. We will also compare and contrast our algorithm with density based clustering algorithms such as DBSCAN and OPTICS and spatial scan statistics, which have been traditionally applied in this scenario. Our method, which will basically rely on variance measure coefficients to identify near to homogeneous regions in space and time relative to its local neighborhood, can be applied to various other use cases such as tornado detection using time varying atmospheric pressure distribution and other climatic and biological regional anomaly detection. This work in progress will extend spatial anomaly detection methods to the temporal domain with real time applications.
Building A Low Cost Low Power Wireless Network To Enable Voice Communication In Developing Regions
Vijay Gabale, Jeet Patani, Rupesh Mehta, Ramakrishnan Kalyanaraman, Bhaskaran Raman, Kameswari Chebrolu
CSE Department, IIT Bombay, India

In this work, we describe our experiences in building a low cost and low power wireless mesh network to provide telephony services in rural regions of the developing world. The novelty of our work lies in its use of IEEE 802.15.4 technology. 802.15.4 is constrained by its computation, communication and storage requirements and hence is a low cost and a low power technology. Hence it was originally designed for a completely different application space of non-real-time, low data rate embedded wireless sensing. However, we use it to design and prototype a telephony system, which we term as Lo3 (Low cost, Low power, Local voice). Lo3 primarily provides two use cases; (1) local and broadcast voice within the wireless mesh network, and (2) remote voice to a phone in the outside world. A Lo3 network can cost as less as $2K, and can last for several days without power “off the grid”. Thus it has several fold less cost and power requirements in comparison to technologies such as OpenBTS, WiFi-based mesh networks or WiMAX networks. Apart from cost and power benefits, it has advantages in providing required capacity and coverage in villages in comparison to technologies such as Femto Cell. Thus, we believe that, Lo3 is an ideal choice to meet the communication need of rural regions.

To realized, Lo3, in this work, we developed a full-fledged, TDMA-based light-weight MAC protocol to enable voice communication over bandwidth-constrained 802.15.4 radio. We also developed 802.15.4 based handset and gateway nodes indegenously to establish real-time voice calls. To enhance the efficiency of our system, we incorporated several optimizations such as a delay-constrained voice call scheduler and a cost-approximate topology formation alogrithm. We test deployed a full-flended Lo3 system in a village near Mumbai, India for 18 hours over 3 days. We established voice calls with an end-to-end latency of less than 120ms, with an average packet loss of less than 2%, and a MOS of 3.6 which is considered as good in practice. The users too gave a positive response to our system. We also tested Lo3 within our department where it can be used as a wireless intercom service. To our knowledge, Lo3 is the first system to enable such a voice communication system using 802.15.4 technology, and show its effectiveness in operational settings. We believe that our prototype can also be extended to stream real-time and stored video over 802.15.4-based mesh networks.
While passwords, by definition, are meant to be secret, recent trends in the Internet usage have witnessed an increasing number of people sharing their email passwords for both personal and professional purposes. In the recent past, the practice of sharing passwords among teenagers has caught special attention in leading news media all over the world. News agencies like New York Times, China Daily, Zee News India, Times of India and many more have reported how sharing passwords has become a new way of building trust and intimacy amongst teenagers and couples. The Pew Internet and American Life Project found that 30 percent of teenagers who were regularly online had shared a password. The survey of 770 teenagers ages 12 to 17, found that girls were almost twice as likely as boys to share. And in more than two dozen interviews, parents, students and counselors said that the practice had become widespread.

While some individuals are comfortable with sharing their passwords with their best friends and partners, other individuals who do not wish to share their passwords often land up on a rough side. People even suffer a break up only because they did not share their passwords. Having to share their email password with someone not only raises concerns for an individual's privacy, but also makes them vulnerable to serious risks. Someone knowing your password could change it, delete your account before you even come to know, or even worse, send unwanted emails to your boss, partner or anyone you can think of! As sharing passwords increases the chances of your passwords being compromised, leading websites like Google strongly advise their users not to share their passwords with anyone.

To cater to this conflict of usability versus security and privacy, we introduce ChaMAILeon, an experimental service, which allows users to share their email passwords while maintaining their privacy and not compromising their security. ChaMAILeon allows users to create multiple passwords for their account, and associate a different level of access with each such password. This allows users to share this new password, and grant limited access to their emails to someone else; but giving them only a partial view of their email content. ChaMAILeon provides blocking or allowing emails from specific individuals / groups and also provides allowing emails containing only certain keywords. All these specifications and preferences can be set by the users to suit their needs.

Visit the live system at http://precog.iiitd.edu.in/chaMAILeon
WHOIS: Visualizing Universe of You
Himanshu Shekhar, Sudhanshu Shekhar, Sherjil Ozair and Dhruv Jain
Indian Institute of Technology, Delhi

Facebook, since 2005, has been the important source of connection between ourselves and the vast circle of our friends near or far. Through various activities like posts, photo/video shares, likes, statuses, comments, etc. everyone of us generate a large amount of data. The current Timeline on facebook does not extract relevant information or does any sort of ranking of the activities/posts/information related to a person whose timeline is, thereby showing the data generated by the user in a simple chronological order.

Now, considering the fact that, we generally have friends in hundreds and even many have in thousands, using the current Timeline to keep track of every friend by browsing through his profile and within each profile, various links/albums/posts is very difficult. So we feel that, given we have so much of data being generated by every of our friend, until one has some mechanism of extracting relevant information out of it, it is not possible for us to remain connected with so long list of friends, thus contradicting the very ideology of a social network.

We propose use of data mining and relevance scores based machine learning methods for ranking traffic of a particular of interest. The information on facebook associated with a user can be broadly divided to textual information and visual information. For visual information like images, we have developed a learning method based on inputs from various images like face occupancy in an image, overall color profile of the image to train a machine learning model which then can be used to predict the visual goodness of an image in the user profile. Also, ranking of textual information is done using various signals such as no. of likes, no. of comments, tags, etc. We also learn a model for automatically generating the cluster of most active friends on your profile, thereby giving a way to determine what is called close friends on facebook.

As, it is always easy to digest large data visually, so using the relevant information extracted, we generate, on the fly, an interactive video of approximate 2 minutes summarizing his/her bio, friends, interests, articles, likes, photos and his feed, in a unique 3d visual experience using impress.js.

So, WHOIS helps in again connecting our fragmented social network where we remain centered around a few people inspite of having so many friends. This will provide a much better visualization of the universe around us in a very easy, timeless and creative manner thus paving way for the next evolution of facebook timeline.
Everyone thinks of changing the world, but no one thinks of changing himself but a Robot can change everything. The Indian Robot is a Revelation for a corruption free INDIA. With Robotic programming, there can be control of traffic, Home security, preserving the nature and many more where man power is needed to make polluted INDIA a crystal clear one from various rascality, especially from bribes and corruptions. Our system is proposed to ensure the security of road traffic and Houses. The existing system has the programmed control where our system is programmable with User and Manual Mode. With the Manual mode the government has the control where the violation of traffic rules can be restricted and with User mode, any individual can monitor the security of his locked Home. The RF transmitter transmits the signals that the web camera captures. The micro controller is a microprocessor with memory unit that save the memory being handled to the database. The PIR sensor is used for Intruder Detection and LDR sensor is used for Fire Detection .The Human Body emits Infrared waves of wavelength ranging from 8 to 12 micrometers, any other higher wavelength range will be considered as an intruder. Whenever any Human being comes in the vicinity the IR system gives the Signal using the Gum broad embed. Here in the home security module, the intruders are being detected by using Facial recognition Techniques. Visual Basic code is programmed to compare the images and those which do not match is listed as intruder. In traffic Control , a 360° Motion camera records the images and transmits using the transceiver to the relevant database. In case of any traffic violation, registration plates of vehicles will be captured and sent to the RTO and will be intimated to the nearest police station to track the vehicle. In fire detection and diffusion, LDR is used to sense the fire. Normally LDR senses all types of Lights, but in our system, LDR should sense only blue and yellow lights and rejects sunlight and other luminaries. For the Robot motion, simple moving wheels are used. Based on the power supply the motion will be faster or slower. The Indian Robot will be your second Blood relation to look after your house and control the bribe notices from traffic cops and be corruption eradicator. These are only fewer applications with our robot, there can be any number of applications to make India a Developed, Corruption free Nation.