
Edited by:
Lior Shamir

ABSTRACT BOOK and AGENDA



Collab. I / Collab. II / Senior Project

Fall 2011

December 9th, December 16th, 2011, 5:45pm, S204



Math & Computer Science
Lawrence Technological University

Preface:

It is my great pleasure to welcome you to our fall 2011 final presentations. This semester we will have two sessions of presentations. Session I will take place at 5:45pm, December 9th at room S204, and session II will take place on December 16th at the same room and hour. The presentation session gives us an excellent opportunity to learn about each other's work during the semester, as well as presenting our work to others. Our session will feature a combination of academic and industry projects, and will allow us to interact, ask questions, and exchange ideas. Several of the projects that were done this semester were submitted to reputable scientific journals and will be reviewed by the leading experts in their fields. In this document you will find the agenda and the abstracts of the projects that will be discussed in the session.

As you may know, oral communication is a crucial skill in the 21st century. As the environment that surrounds us rapidly becomes richer in information, brilliant ideas are sometimes being rejected because they are not communicated and explained effectively, while average ideas are in many cases adopted because they are delivered in a proper fashion. The presentation session also gives us the opportunity to practice our technical communication skills, and express our ideas and work in a clear and condensed fashion that will be of interest to our audience.

Our session is organized such that each student(s) presents for 10 minutes, and then we will have five more minutes for questions and discussion. To save time during the session, please make sure that your presentation is ready on your computer before it is your turn to present. I would also like to remind you that submission of the course materials is due to December 20th. The submission is through Blackboard. Please consult with me if you are not sure what you need to submit.

I wish you a fruitful and enjoyable presentation session, and I look forward to attending your talks.

Sincerely,
Dr. Lior Shamir

Session I

Friday, December 9th, 2011

5:45pm-6:00pm	<p>Welcome and opening remarks <i>Lior Shamir</i></p>
6:00pm-6:15pm	<p>Increasing eastern bluebird nest box performance using satellite images <i>Sarah Svatora</i></p> <p>Abstract: Conservationists have been introducing man-made boxes in an effort to increase the bluebird population. In this study we used computer analysis of satellite images to show that the performance of the boxes used by the eastern bluebird (<i>Sialia sialis</i>) in Michigan can be improved by ~48%. The analysis is based on a correlation found between the edge directionality measured in the satellite image of the area around the box, and the preferences of the birds when selecting their nesting site. The method is based on satellite images taken from Google Earth, and can be used by conservationists to select a box placement strategy that will optimize the efficacy of the boxes deployed in a given area.</p>
5:55pm-6:10pm	<p>CloudID: Mapping Names to Numbers in Real Time on Android Devices <i>James Otto</i></p> <p>Abstract: The task of automatic author identification by text belongs in the field of natural language processing (NLP), which is considered as an interaction between computational linguistics and computer science, to be more concise artificial intelligence. In this paper we describe some linguistic principles that can lie in the base of identification. For that purpose, we will construct a minimal vocabulary, which will be common for all authors, but at the same time will help to distinguish between them. To prove our assumption several experiments were made. Following parts of speech were used to construct our vocabulary: prepositions, conjunctions, articles, adverbs and auxiliary and modal verbs. The reason that we decided to choose those particular parts of speech is because we do not want to rely on the context, which becomes possible by taking in account nouns, verbs and adjectives. Our vocabulary elements are context-free, which means that they are not overlapping with the vocabulary of each particular text. All experiments are based on the calculation of word-frequencies and computation of distances based on k-NN algorithm, which will play key role in the identification. Based on the output produced by the k-NN, we can conclude that as a part of speech conjunctions and prepositions, together allow us to determine author's name much more, than other ones.</p>
6:10pm-6:25pm	<p>CHLOE: A software tool for automatic outlier phenotype detection in microscopy image datasets <i>Saundra Manning</i></p> <p>Abstract: The recent advancements in automated microscopy and</p>

	<p>information systems allow the acquisition and storage of massive datasets of microscopy images. The rapid development of these systems has reinforced the need for computer vision and pattern recognition tools that can effectively analyze these masses of image data and turn them into knowledge. Here we describe Chloe, a software tool for automatic extraction of outlier phenotypes in microscopy image datasets. The tool is based on a comprehensive set of numerical image content descriptors reflecting image morphology, and can be used in combination with ROI detection and cell segmentation tools such as ITK. The rich feature set and the set of rules allow it to automatically detect repetitive phenotypes that are visually different from the common phenotypes in the image dataset. Experimental results suggest that Chloe can effectively detect repetitive phenotypic patterns that are different from the common phenotype. The code and software is publicly available for free download at http://vfacstaff.ltu.edu/lshamir/downloads/chloe.</p>
<p>6:25pm-6:40pm</p>	<p>Mint – A laptop inventory management system <i>Todd James</i> Abstract: Mint is a laptop inventory management system that was created to allow users to easily keep track of current laptop inventory using existing barcodes. Mint's user-friendly interface allows the user to maintain a master database that stores a list of all laptops it should keep track of along with the type and current status of each laptop. Statuses include common laptop life cycle states such as "In Collection", "In Transit", "Special Programs", and "Stolen". Users can use a barcode scanner to scan the inventory on hand and import data from Sungard HE's Banner, a laptop assignment system, to be used for comparison against the master database. The final result is a report that summarizes laptops checked out to end users by type, major, level, and department as well as laptops on hand by type, location, and status. Mint can be used by an organization to track inventory flow and make decisions regarding purchasing, sales, and special program distributions.</p>
<p>6:40pm-6:55pm</p>	<p>Identification of genetic variants to predict the survival outcome of individuals with cancer <i>Steve Jarfas</i> Abstract: Medical prognostic models predict the outcome of a treatment or disease progression; in this case, using experimental genetic variables to predict the survival outcome for individuals with cancer. The purpose of this study was to develop a R program to analyze the test data set to identify a predictive gene to determine the prognosis of a patient. The provided data set consisted of 7400 gene expressions for 180 patients along with the patient outcome. The predictive algorithm calculated a Fisher score for each gene expression. The algorithm then used the Fisher scores to produce a weighted Euclidean Distance between gene sets and compared the patient outcomes for a predetermine number of gene</p>

	<p>expressions. The gene sets were then sorted to identify and rank the top ten closest gene sets. “Majority” prognosis scores were then determined by comparing the patient outcomes. The Majority score was considered either a hit, miss, or draw based on the number of matching patient comes in the top ten closest gene expressions. Hence, 10 Majority score sets were generated per each gene expressions corresponding to comparisons for the closest gene expression, the closest two gene expressions, the closest three gene expressions, up to the closest ten gene expressions. Accordingly, the analysis required a series of runs to analyze the sample set of gene expressions, incrementing the number of gene expressions per each run according to the Fisher score ranking of the gene expression. One additional gene expression was added per analysis for the highest ranked 110 gene expressions. Ten gene expressions were added per analysis up to 2000 gene expressions. Fifty gene expressions were added per analysis up to 3400 gene expressions. Although the initial results appeared promising when considering a limited number of genes, the algorithm becomes inefficient as the number of gene expressions increases and does not meet the challenge requirement of less than 90 minute execution time.</p>
<p>6:55pm-7:10pm</p>	<p>A utility for tracking the origin supply company for part used by Ford Motor Company <i>Dhrgam AL Kafaf</i> Abstract: This Product aims to mitigate risks by providing Ford Motor Company (Purchase Department) with the ability to extract information about all the fundamental company that supplies the automotive manufacturer’s parts. This product supply Ford Motor Company with the information deep within the tiered supply based companies to provide understandings such as supplier company names, manufacturing locations, site leaders and contact data, details such as facility size, employee quantity, and logistics. By such information Ford Motor Company will have a better view of where the supply companies stand, and in case of natural disaster, or war zone Ford Motor Company can be prepared for using alternative companies for this part. The product is web base software that allow the customer to brows for the companies within the tires in different ways like, by region where they are located, by the company name and by the part number that incorporate to manufacture. And give the customer the ability to add or update new companies.</p>
<p>7:10pm-7:25pm</p>	<p>Correlation of physiological indicators with brain MRI scans using image recognition <i>Erik Schwarz</i> Abstract: The purpose of the experiment is to test what physiological indicators correlate with brain MRIs. That is, what can be learned about a person given a brain MRI scan. The image data set consisted MRI scans of 416 subjects covering the age span of 18 to 96, both men and women, and 100 subjects over the age of 60 have very mild to mild Alzheimer’s disease. The data are cross-sectional, and</p>

	<p>contained the following data about each image: age, gender, dominant hand, education level, socioeconomic status, mini-mental state examination, clinical dementia rating, estimated total intracranial volume, atlas scaling factor, and normalized whole brain volume. The process I went through involved converting each image to Tiff format using ImageJ. The next step was to take the same three slices from each image and save them into a separate folder, this was done using the program TifSplit. And finally each heading for the images had to have a text file created with the full path name to the file and the value of the heading it represents. The results were that education's best had an average Pearson correlation of .335206 on slice 75 of 16-bit normalized with 1% of features, age's best had an average Pearson correlation of .716960 on slice 50 of 16-bit with 1% of image features, and gender's best had an average Pearson correlation of .755556 on slice 50 of 8-bit normalized with %5 of the features.</p>
<p>7:25pm-7:40pm</p>	<p>Pattern Matching with Acoustics <i>Jordan Fox</i> Sounds recorded in nature are an informative data, and using sounds for ecological research can be very beneficial if they can be effectively measured and analyzed. Using a computer to find patterns, aka certain sounds like a bird chirping, would be hugely beneficial in that a computer can scan many files much quicker than any human can. The first step in searching for patterns in files is identifying sounds in a subset of files. These known patterns will be used to search other sound files for those particular patterns. The analysis is done by using signatures that numerically reflect the audio content. Several audio sound analyses are used, such as multi-scale histograms, gradient transform, mean, median, Fourier transform, and a variation of NDSI. After calculating the numerical sound signatures, the most informative sound descriptors are selected, and the values are analyzed using the Weighted Nearest Distance (WND) algorithm, which classifies the sound samples. Experimental results show that the algorithm can differentiate between the sounds of a frogs and a sound of loons recorded in nature with accuracy of 95%. sound of loons recorded in nature with accuracy of 95%.</p>
<p>7:40pm-7:55pm</p>	<p>Labyrinth: A team-based strategy game based on navigating procedurally generated mazes <i>Andrew Smith and Kyle Cupp</i> Abstract: Labyrinth is a 3-dimensional real-time multiplayer video game, written in C# using the XNA framework. The intended platform for the game is PC, with the target audience being 16 to 40 year old male fans of strategy and team play. The basis of the game is for several teams to compete against each other in accomplishing a given objective while navigating a large labyrinth. Game-play will consist of combating AI monsters and opposing players to gain experience which is then used to purchase abilities, which are in turn used to better combat the players opposition. Unique selling points for the game include procedurally generated terrain, customizable magical abilities which are based on concentration rather than depletable "mana," and skill based play in a strategy game.</p>

Session II

Friday, December 16th, 2011

5:45pm-6:00pm	<p>PHD: Static Code Analysis for PHP <i>Ron Jenkins and Caleb Davis</i></p> <p>Abstract: The purpose of this project is to develop a static code analysis tool for the PHP programming language. The tool should produce various code metrics including comment ratio, cyclomatic complexity for functions and files, and number of properties/methods per class. To achieve this, a grammar must be defined that fully describes the PHP language (there is no PHP grammar available because PHP is defined by its defacto implementation rather than by a formal specification). This grammar is matched against PHP input files and a predefined ruleset to produce a list of all the language elements present in the file. This list can then be searched to produce the desired code metric values by counting the appropriate values, such as the number of control structures and the number of class properties. This tool can be used to assist developers looking to reduce the complexity of their PHP code and find areas of their code that may benefit from refactoring, reducing the possibility of runtime errors. In the future, this tool will be expanded to include more code metrics (ie. lack of cohesion of methods, maintainability index) as well as code audit rules that represent best programming practices (ie. avoid control structures with no braces, accidental assignment in conditional expression).</p>
6:00pm-6:15pm	<p>Automatic Detection of Ring Galaxies <i>Patrick E. P. Reynolds</i></p> <p>Abstract: Out of the billions of galaxies in our universe, most galaxies can be associated with a stage on a morphological sequence call the Hubble sequence. However, some rare peculiar types galaxies are not part of that sequence, and are therefore of paramount scientific interest. One of the important types of peculiar galaxies are the Ring galaxies. In the search for ring galaxies astronomers currently have to sift through millions if not billions of images by hand just to find one. As a result, currently there are only about 132 known ring galaxies and little is known about how they acquired their unique symmetrical ring shape.</p> <p>The purpose of this project is to create a computer algorithm that can sift through and analyze thousands or even millions of galaxy images acquired by robotic telescopes and determine if the image is of a ring galaxy or not. This is done by first determining the galactic center and radius. Once these</p>

	<p>are determined a concentric banded area going out from the center of the galaxy is determined, from 40% point to the 75% point of the radius. This band is then divided up into 35 concentric circles and at each degree in the circle the pixel is converted into a value depending on its intensity. This value is then stored as part of a 35 x 360 array. Once all of the values are determined their statistical characteristics are analyzed in order to determine the morphological type of the galaxy.</p>
<p>6:15pm-6:30pm</p>	<p>Intelligent Freeway Traffic Management: Removing the Human Factor <i>Brandon See and Joel Fischer</i></p> <p>Abstract: The goal of this project is to create an artificially intelligent highway system which prevents accidents stemming from human error, and helps to eliminate congestion by organizing traffic flow. As opposed to most intelligent road systems in development, which are many-to-many, our traffic system has a one-to-many relationship. In our system, each car reports its position and speed to a server, and the server controls each car individually. Cars themselves do not have direct communication with other cars. To create this system, we created a road system, found formulae to calculate position along those lines, and randomly generated cars. We then built collision detection and put a timer system into a separate thread to help keep the system accurate. Finally, we built an AI system to intelligently control each car. The results of our project are that we have successfully created an AI system which controls cars on the freeway. Cars are randomly generated on the left of the screen, and we control them so they get to the end of the segment of highway as efficiently as possible without crashing. For visual purposes, as a car changes lanes, it turns blue. Each individual part of the project was rather simple to implement, but combining each piece together into a working program was difficult. If this project were to make it into the real world, we expect it would be merged with another AI system where the car was also able to control itself. Future goals to advance this project may be to adapt this system to allow for some cars on the road to be under human control, or to make it work with stoplights.</p>
<p>6:30pm-6:45pm</p>	<p>An Application for Automotive Industry Manufactures To Monitor Supply Chains In Disaster Situations <i>Asya Lushnikova</i></p> <p>Abstract: “Organizations face many external and internal risks that can threaten vital business operations. There risks include anything from natural disasters to man-made disruptions...”¹ Among those organizations are automakers, which might suffer from multi-million financial losses, and especially in manufacturing industry, due to delays in supplies delivery and production. To make more effective work of employees dealing with supply chains on daily basis and involved in purchasing of parts and raw materials, we constructed an two-tier application, which consists of the database with all necessary information related to company’s vendors and web</p>

	<p>interface for easy retrieval and manipulation of data. The main idea of the application is to assist purchasing department in faster look-up/ search and management of supplier data, which includes addresses, contact information (such as phone and fax numbers, managing directors' names, manufacturing brands, smelters, etc.) and Google Map links to the regions, where the companies are located, etc. The other important detail of the application is the ability to view suppliers of each part in production order of it. MySQL ver.5.6 and MySQL ver.5.1 - were both tested during software production as Database Management System for company's relational DB model. Oracle WebLogic Server ver. 10.3 allows us to deploy Java Web Application, which mainly consists of web pages, developed using JSP technology, for further data analysis. Our software provides not only user friendly interface, but as well, gives our customer ability to manage supply chains' information in a more convenient way.</p> <p>1 Balthazar, J.: Introduction. Business Continuity, Roles and Responsibilities, and Key Deliverables Defined, 2011, pp.1</p>
<p>6:45pm-7:00pm</p>	<p>Scientific Journal Website <i>Eriton Muhametaj</i> Abstract: There is an efficient way to discover and foster new research and production by building a new and simpler scientific journal websites. A newly created website will change the world of academic publication, with a goal on improving how the publication will publish to the public. Our perspective is on how to make scientific communication easier when the information is being exchanged between researchers in every field, using a novel platform. We present this new web application with an example-based approach, concentrating on the importance of deciding who and what can be published. To support this goal the projects is being developed using HTML design, UML diagrams for software specification and database model of peer-reviewed publications. The website allows creating accounts for authors, editors and two kinds of readers; with and without accounts. To make the website more efficient, the comments, rating and reviews from each reader will play a key role in the processing of scientific articles. Advanced search with the use of keywords will make it easier for the user to navigate through the dataset of articles. This platform will be used to change the way scientific work is communicated today.</p>
<p>7:00pm-7:15pm</p>	<p>Automatic Detection of Peculiar Galaxies in Digital Sky Surveys <i>Andrew Lucas</i> Abstract: There are almost an infinite number of galaxies in the Universe with more being discovered and data being continuously gathered about them. At the moment the only effective way to sort through all these data is through manually looking through pictures of the galaxies and sorting them based on the knowledge of the people analyzing them. The goal of this project was to computerize the process of sorting and then to be able</p>

	<p>to find unique galaxies out of all the data. I started with using three different already manually sorted data samples of elliptical, spiral, and edge on galaxies as the learning data for slightly modified version of Dr. Shamir's image classifier program called Wndchrm. After that classifying a set of ring galaxies acting as the test samples of unique galaxies I learned that edge on galaxies were not needed in the learning set and that more data samples of elliptical and spiral were needed. When more data samples were added better results were showing and I was starting to see a threshold that could possibly identify unique galaxies from the rest of the data. In conclusion the results are good but the end goal is still far from being achieved as more work still has to be done to help such things as having the program ignoring any artifacts that could be on data samples.</p>
<p>7:15pm-7:30pm</p>	<p>The Average Galaxy of the Universe <i>Michael Myroniuk</i> Abstract: This research is focused on the analysis of galactic and interstellar matter to study a possible cosmic parity violation. The purpose is to normalize a large amount of image data, specifically around 700,000 different images of various galaxies that have been detected by Sloan Digital Sky Survey (SDSS), and average their pixel content using image processing techniques. C code was developed to automatically download 700,000+ galaxy images from SDSS web site using the CAS server, and then average them automatically by using the standard ImageMagick utility software to generate the average galaxy representation from the image dataset. This would eventually yield one photo containing the average pixel intensity value of each photo for further analysis. Assuming cosmic parity, it is expected that the "average" galaxy image will be fully symmetric. Therefore, any asymmetry detected in the averaged image can be considered evidence that the universe is not isotropic. When the final averaged image was obtained it was displayed as a dim spherical shape with a bright center which indicates an area containing more heat. The image was afterwards analyzed using an artificial thermal image map.</p>
<p>7:30pm-7:45pm</p>	<p>ISBackfile: A document digitization system <i>Kyle Anderson and Christopher Harris</i> Abstract: Over the next decade, hospitals will be required to make the shift towards being entirely paperless. Documents will instead be managed digitally, through an Electronic Content Management (ECM) solution. In order to move paper content into their ECM system, hospitals will need a way to quickly index their backfile documents. We developed a Windows application named ISBackfile that aims to improve the way in which hospitals can import these documents into the ECM solution. OnBase by Hyland Software is the ECM product that we based this project around. Our goal was to create an application that uses touch commands and two different databases to quickly index documents based on a primary keyword value and generate a text file with this information. Once a</p>

	<p>document is opened, it is displayed in our application’s main window and each page can be rotated or moved with touch controls. Keyword information is populated from an external database. Based on one keyword, like a patient number, other keywords can be automatically populated, like first name, last name, address, etc. Documents can be indexed as different Document Types. The available document types are configurable through our application and get populated from the OnBase database. Each document type has different keywords associated with it, these are also configurable, and are also populated from the OnBase database. Each document type has a SQL Query associated with it that is used to query the customer’s external database and correctly populate keyword information. Once the document type and keyword information have all been set, our application generates a text file that contains the document type name, all of the keyword values, and a path to the image that information is associated with. OnBase then uses that text file to import the document with all of the associated keyword information. By using intuitive touch controls and automatically populated keyword values, we believe that we can greatly improve the speed at which hospitals can bring their backfile documents into OnBase.</p>
<p>7:45pm-8:00pm</p>	<p>Accelerando <i>Mike Skupmy</i> Abstract: Accelerando is a new form of role-playing game experience combining the classic dungeon crawler with the franticness of a rhythm game. Most of the game takes place in both pre-made and randomized dungeons with the accompaniment of a player-built Accelerando digital music player playlist. Each song has a unique character-enhancing perk, and more songs can be found along with other items by traversing the dungeons. There are no classes such as warriors or wizards, but rather a skill points system that allows for character customization for any preferred playing style. Once in a dungeon, players ascend through dungeon floors to get to the boss while being mindful of time spent moving along. If all songs in the Accelerando playlist are played, the player must restart the dungeon, though items can be found that allow for shuffling the playlist or repeating songs. There are three modes to dungeon crawling: movement, attack selecting, and actual attacking. For the latter, notes scroll down on one side of the screen in time with the music. When the players decides on a command such as attacking, defending, healing, or casting spells, he or she must play several measures of music by pressing keys in time to the scrolling notes hitting a line. The effectiveness of the command depends on the percentage of the notes hit. Besides this main gameplay, Accelerando will also include cutscenes that tell an light-hearted modern era story, an overworld, towns with shops for new equipment and songs, and more. Most of these additional pieces of the game will be added after the first prototype, which is meant to showcase the dungeon</p>

	gameplay.
8:00pm-8:15pm	<p>An Image Data Analysis System for Edited Photo <i>Cheng Jialiang</i></p> <p>Abstract: With the availability of digital photo editing tools, we can see many edited photos published in digital and printed media. While some photos look natural, it is difficult to know whether they were edited digitally. To identify digitally edited areas in digital photos, a tool which can detect the edited photo is developed, such that the purpose of the project is to find a method for determining whether a digital photo is fully natural, or was edited with a digital photo editing tool, and then find the edited part in the photo using image data analysis. In the paper, we introduce how to detect the edited photo and how to get the image data and analyze it. Then the final result depends on the data analysis. The method currently does not achieve perfect accuracy, and future work will be required to improve it.</p>