Ph.D. Qualification Portfolio

Sayed Mohsin Reza
The University of Texas at El Paso
Department of Computer Science
Email: sreza3@miners.utep.edu

04/01/2020

TABLE OF CONTENTS

1 INTRODUCTION

This document presents my research plans to obtain a PhD at University of Texas at El Paso. The document presents my personal statement where I outline my personal characteristics and goals. The document also presents my research plans and activities.

1.1 BRIEF BIOGRAPHY

I am Sayed Mohsin Reza, a Ph.D. student in Computer Science (CS) at the University of Texas at El Paso (UTEP). I started my Ph.D. program in Fall 2019. Before joining UTEP, I was at University of Colorado Denver, USA. This portfolio is a part of my Ph.D. qualifying process. The portfolio describes the experiences that I have gained to date and my future research plans to achieve a doctoral degree.

1.2 DOCUMENT ORGANIZATION

This document is organized as follows. Section 2 presents my personal statement, section 3 presents my research statement, and finally, section 4 presents evidence to demonstrate my readiness to conduct high-quality and impactful research.

In the personal statement, I discuss my academic preparations and my professional experiences. I also discuss my planned contributions to computer science research and the way I have personally chosen to construct the path towards my doctoral degree. At the end of this section, I discuss briefly my goals after my graduation.

In my research statement, I describe my research interests focusing on investigating software evolution. I provide a set of research questions and outline the progress I had made so far.

In the evidence section, I describe my course works, communication skills to show my academic skills that fulfill my Ph.D. requirements for qualifying. Additionally, I provide my research & software development experience to show technical skills to conduct doctoral research.

2 PERSONAL STATEMENT

From my school life, I had a keen relationship with computer and information handling. In college, at first, I got some knowledge of computing and after that, the computer is now like a friend of mine. I had a very young and enthusiastic teacher Mr. Alok Kumar, who taught me the course: Introduction to Computer. His lab work and enthusiastic behavior made the computer as popular subject to me. At the undergraduate level, my interests became more focused on software development and information systems. After graduating with a master's and bachelor's in Information Technology (IT) from Jahangirnagar University, Dhaka, Bangladesh, I felt I should gain some practical experience working in a lab environment to gain first-hand knowledge of what the life of a developer entails. I took admission to UTEP.

In UTEP, I have completed almost all course works, have chosen those courses that can help me in my research. And currently, I am researching the software engineering area with Dr. Omar Badreddin, Assistant Professor, CS.

The following summarizes my research progress, my course work, and my software development experiences.

2.1 RESEARCH PROGRESS

My current research with Dr. Omar is to investigate the effectiveness of machine learning techniques in predicting software quality attributes (discussed in Section 3) throughout the software lifecycle. To do such research, we need to collect a lot of software quality attributes dataset from maintained open source repositories. To collect those repositories, I have developed ModeMine, a research repository mining tool that can mine repositories information from open source repositories. The tool is now available in the following link: http://www.smreza.com/projects/modelmine/. In addition to that, I have already worked on a research paper in analyzing handwritten code in model-driven engineering and will submit to a conference. In this paper, we have used the ModelMine tool to mine model-based repositories information. Currently, I am taking a Data Mining class with Dr. Hossain to learn more about machine learning techniques that I will apply to predict software quality attributes.

The details of my research statement are available in Section 3.

2.2 COURSE WORK

From the beginning, I tried to finish the course works under the CS Ph.D. course requirements. I have completed all core courses and breadth courses. I have carefully chosen the technical elective courses towards my research interests. This semester, I have taken CS 6390 - Special Topics Data Mining to learn about machine learning techniques that I will use in my research. Moreover, I took Graduate Research Methods where I earned knowledge about different aspects of computer science researches ideas, methodology and evaluation techniques which is currently helping in my research directly.

The detailed evidence of my coursework towards my Ph.D. degree is available in Section 4.1.

2.3 Software Development Experience

From my undergraduate level, I tried to develop my software development skills by engaging in the software engineering area and apply the knowledge to practical problems about solving different issues. In my senior level, I developed a project related to academic conferences named **"PROCONF Integrated Multi-conference System with Payment Method"**. The project is now used by several conferences in Asia and hosted in the following link: https://www.proconf.org. After completion of this project, my supervisor, Shamim Al Mamun was satisfied with my practical skills and was happy to give me a recommendation that led to my first job in a software company. Recently, I have developed a tool named **ModelMine** to mine model files from open source repositories towards my current Ph.D. research interests in CS. The tool is available on http://www.smreza.com/projects/modelmine/. In addition to these, I have also learned some data science skills that I used in one of my projects to analyze the weather, population and geological data to provide heat island effect in the USA. The project is now live at http://www.smreza.com/projects/heat-island-effect/.

The detailed evidence of software development experience is available in Section 4.4.

Last but not least, Research is essential to continue the legacy of humanity and for the betterment of the society, and a good research profession can be built on top of my research and educational experiences, as provided by the evidence in Section 4. In a Ph.D. program at Computer Science, I have a greater opportunity for research and development that can be used to help mankind which will fulfill my dream. People will find in me a person giving his best in all challenging environments and utilizing my knowledge and resources for the betterment of society.

3 RESEARCH STATEMENT

My research will investigate the effectiveness of machine learning techniques in predicting software quality attributes throughout the software lifecycle. I plan to extract features from large datasets from open sources, such as class sizes, method sizes, number and type of attributes, cyclomatic complexity, cohesion, and coupling, code smells, technical debt and others, and their histories over an extended period. My research will investigate whether we can predict how these features evolve. This will open new pathways for more effective and proactive code refactoring. Instead of waiting until issues emerge in the codebase, engineers will be empowered to proactively apply refactoring techniques to eliminate or minimize code quality concerns before they materialize.

3.1 INTRODUCTION AND BACKGROUND

Software systems continue their exponential growth in size and complexity. With this growth, software quality and maintenance has become a major concern. Not only that maintenance efforts often eclipse that of the initial software development, but it has become the reason why many software systems become unsustainable over time.

Software engineers apply many refactoring techniques to maintain a reasonable code quality as the software evolves. These refactoring techniques are typically applied reactively only after quality degrades and maintenance becomes a concern. My research will aim to identify quality concerns and codebase degradation before it materializes. This way, refactoring's can be applied proactively well before the emergence of deficiencies.

There is a significant and growing research that aims to predict software failure and software bugs using many techniques including machine learning. My research will build on these advancements and will focus on predicting code quality attributes, rather than predicting failures and bugs. One specific paper has ignited my interest in this research area authored by Liu [14]. In this paper, the authors applied a deep learning technique using several neural networks to predict the emergence of code smells. In this approach, the authors used a data set from Quality Corpus Repository [17] and extracted several code qualities features. The approach demonstrated the significant potential of machine learning in predicting code smells over time. The deep learning approach was found to be significantly superior to the most prevalent techniques and code analysis tools.

There are numerous open questions and challenges in this research area. The first big challenge is the availability of sufficiently large and appropriately labeled training datasets. Another big challenge is to identify the features that impact code quality evolution over time. And yet another challenge is to identify and evaluate the most appropriate and effective machine learning technique to predict several aspects of code quality attributes. I expect that my research will 1) create a large number of labeled data sets that can help propel research in this area, 2) identify key features that have the most impact on code quality and its evolution, and 3) contribute to identifying the effectiveness of several machine learning techniques in achieving the desired predictive code quality analysis.

3.2 Research Questions

My research on predictive code quality analysis is motivated by the following three research questions.

RQ1: How effective are machine learning techniques in predicting code quality evolution over time?

This first research question will investigate whether machine learning techniques can be effective in achieving basic prediction of code quality attributes evolution over time. For this research question, my first task is to construct sufficiently large datasets by mining open source repositories. I will apply my expertise in static code analysis and mining tools in automatically constructing and labeling the dataset. The data set that I plan to collect will include more than twenty code quality attributes (some of which were mentioned earlier) for prominent codebases, such as the Linux repository. This data will be collected for each code revision over an extended period.

RQ2: What are the key code features that have the most impact on codebase evolution and its quality attributes over time?

I am learning about several machine learning techniques in the Data Mining course with Dr. Shahriar. Throughout this course, I am continuously experimenting with several techniques on the small data set that I have already collected. The goal of this research activity is to develop a preliminary understanding of the effectiveness of these techniques on predicting the evolution of code quality attributes. The findings of this research task will inform the direction of my research.

RQ3: Can the machine learning techniques inform the formation of proactive code refactoring's that minimize or eliminate quality concerns before they materialize?

The goal of my research is to develop a proactive code refactoring approach that eliminates or minimizes code quality concerns before they materialize. For example, my research may suggest a specific refactoring not based on an existing quality concern, but rather, based on a quality concern that may emerge sometime in the future. As such, the refactoring will be much more effective in improving the codebase sustainability over the lifetime of the project. In other words, rather than waiting for code smells to emerge, this proposed approach will enable engineers to proactively refactor the code so the smell will never emerge.

3.3 Progress to Date

I have made little but important progress in this direction.

1) Preliminary data mining from major open source repositories
2) Initial experimentations with various machine learning techniques
3) Literature review on predictive code analysis techniques, approaches, and tools
4) Developed a custom-built tool to facilitate data mining of software repositories. The tool will be published as a tool paper
5) Co-authored a paper related to my research in mining open source repositories
6) Presented a seminar lecture on an IEEE Transactions paper
7) Conducted a Unity Workshop for the ACM and ACM-W student organization

4 EVIDENCE

The evidence of my ability to succeed in the Ph.D. program is demonstrated by this section. Four subsections describe my performance in Coursework, Communication Skills, Research Experience and Software Development experience.

4.1 COURSEWORK

In this subsection, I have demonstrated my coursework details to demonstrate my expertise in CS level research as well as required course fulfillment. I completed all my core course works in the University of Colorado Denver and transferred those credits to UTEP. Here is the list of completed core courses.

- CS 6315 Theory of Computation (Transferred Course - CSCI 5446 Theory of Automata)
- CS 6350 Advanced Algorithms (Transferred Course - CSCI 5451 Algorithms)
- CS 6392 Graduate Research Methods (Transferred Course - ISMG 7214 Mixed Methods Research)

Appendix B part B1 contains my transcripts and Appendix B part B2 contains my coursework and artifacts examples.

4.2 COMMUNICATION SKILLS

In this subsection, I have demonstrated my written and oral communication skills.

1.3 Written Communication Proficiency:

- Technical report [Attached in Appendix C: Communication Skills]
- Other project reports [Attached in Appendix B: Coursework]
- Papers published and in progress
    - A paper analyzing behavioral factors of turbulent exhale flows was published at the IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) in 2018. [Attached in Appendix D: Research].
    - A paper on e-Health services in Bangladesh using Low Cost and Portable Patient Monitoring System was published in the 2016 International Conference on Computer Communication and Informatics (ICCCI). [Attached in Appendix D: Research]
    - Another paper titled “Innovative approach in web application effort & cost estimation using functional measurement type” was published in the 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT). [Attached in Appendix D: Research]

Area Improvement Initiatives:

I have taken several initiatives to improve my writing skills. For example, I enrolled in a Grant Writing (RWS 5313) in Rhetoric and Writing Studies program at UTEP. I am also currently working with Dr. Maggy Smith a senior professor in RWS with my writing skills. She is helping with my revisions and identifying issues with my writing, and I am learning how to construct and organize ideas in academic writing.

In addition to that, writing needs a lot of reading; this includes reading academic papers as well as mainstream books. Therefore, I am reading several research papers to understand the writing styles,

grammar usage, and transitions that various authors use, and how they construct and present their ideas easily and understandably. To develop better writing skills, I also use some reference books [1][2][3][4] that are helpful while writing.

Besides, I regularly visit the University Writing Center at UTEP. The consultants in the writing center review my writings and provide feedback. I use that feedback to revise my writing, which is one of the best practices for improving these skills.

Finally, writing is a skill that cannot be developed overnight, but I have taken some necessary initiatives that will help to improve my writing skills in research and academia.

1.4 Oral Communication Proficiency:

Since Fall 2017, I have completed several courses that help with presenting my research work, addressing questions and concerns from the audience. All my coursework required several individual presentations and oral communications. In the graduate research method class (ISMG 7214), I presented several research papers and engaged in class discussions.

Additionally, I attended the RaCAS (Research and Creative Activities Symposium) at Denver, Colorado in 2019 & 2018 as a graduate researcher which helped me to communicate with other researchers and share knowledge with them. In 2015, I presented my graduate research [5] at an international IEEE conference in Bangladesh and in 2014, I presented my undergraduate research works [6] at an international conference at the Military Institute of Science and Technology in Dhaka, Bangladesh.

Here is the list of presentations & talks.

- Technical poster presentation of a project: *Urban Heat Island Effect in the Borderland* in Web-Based Data Integration with Geological Science Course [Attached in Appendix C: Communication Skills]
- Presented a seminar lecture on an IEEE transaction paper [Attached in Appendix C: Communication Skills]
- Presented a published paper at IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) in 2018.
- Presented our poster related to our research at RaCAS (Research and Creative Activities Symposium) at Denver, Colorado in 2019.

Area Improvement Initiatives:

To improve my oral communication, I have taken several steps that will support my success in all areas of the Ph.D. program, especially in oral communication. In research, oral communication is the most important communication skill; to disseminate research, researchers need to communicate verbally. The purpose of the research is to share the knowledge that can be achieved through oral communication.

I joined a development forum, ACM UNITY Workshop at UTEP in 2019. I also attend weekly meetings and take part in the developing game in Unity where we are asked to respond to topics about which we have no prior knowledge. This concept provides a chance to think and generate my ideas instantly and present my game ideas with them in proper ways. I came to learn different techniques for performing a perfect academic presentation that keeps the audience's attention and makes it interesting.

Consequently, when I present something, some evaluators give me feedback on my presentation both verbally and in writing, indicating the places I need to improve. The evidence of this work is attached in "Appendix C: Communication Skills."

To improve these skills, I am currently following Ted Talk videos on YouTube. In these videos, expert speakers in fields such as science, arts, physiology, and literature provide talks on different topics. I follow their presentations to learn their tactics and techniques to do better presentations on my own.

4.3 Research Experiences

After joining the computer science Ph.D. program at UTEP, I focused on research in software engineering, particularly on software design and code smells. I started working in the area of machine learning in software code smell and as an initial work, I designed a tool that can mine repositories information from the web to faster the research in this area [18]. This tool is mainly mining model-based repositories (.uml, .xmi, etc.), code files and commits details. In addition to that, several filter techniques (by programming language, repository size, etc.) are integrated with this tool to facilitate variation in search results. The results of this work are planning to submit to the MODELS'20 international conference.

Additionally, I worked in analyzing model-driven engineering behavior on codebase quality in 2019. It was group research. This study aims to investigate the hypothesis about the handwritten code quality developed in the context of Model-Driven Engineering (MDE). I analyze unique code fragments and compares their characteristics to handwritten code in repositories where code generation is not present. The study finds that handwritten code quality in the MDE context suffers from elevated Technical Debt (TD) and Code Smells (CS). We observe keyCode Smell that is particularly evident in this handwritten code. These findings imply that code generators must optimize for human comprehension, prioritize extensibility, and must facilitate integration with handwritten code elements. This result of this work is already submitted to the Journal of Software Practices and Experience.

In 2018, I published a research paper at the 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) conference which is attached in "Appendix C: Communication Skills." It was a part of my independent study with Dr. Min Choi, a professor of the CS department at the University of Colorado Denver. The research work that I conducted at that time was presenting a structural foundation for the respiratory analysis of turbulent exhale flows through the visualization of dense CO_2 density distributions. In this work, we used a thermal imaging device to target high-resolution respiratory modeling. My idea was to integrate software modeling & vision algorithms to generate sparse flow exhale models. The contributions of this work are to target the acquisition of numerous respiratory behaviors including, breathing rate, exhale strength and capacity, towards insights into lung functionality and tidal volume estimation.

Another group research on portable health monitoring systems was done in 2015-2016 and published at the 2016 International Conference on Computer Communication and Informatics (ICCCI -2016), India. This paper proposed a low cost and portable patient monitoring system for e-Health services in Bangladesh. To develop the system, we have used a raspberry pi microcomputer to collect different sensors data such as ECG signal, blood pressure signal, heartbeat signal, Situation of Oxygen in Blood (SPO2) and temperature from a patient and send these signals to a specialist doctor who is in a center or a hospital. A web-based application has been developed for both doctors and paramedics for efficient

communication with each other. The proposed system is evaluated with existing systems and found that the proposed system can be suitable for the village health care center of Bangladesh.

"Appendix D: Research**"** contains some of the evidence of my research experience.

4.4 Software Development Experiences

I worked on several software development projects in my undergraduate and graduate studies that include a conference management system, peer to peer review management system, result processing system, heat island effect map tool, model mining tool, etc. These projects were developed in different languages such as Java, C, C++, PHP, etc. Also, I worked on a special project that only deals with software design. Below I provided details of my projects that were being developed by me or with a team.

4.4.1 ModelMine: A tool to mine models from open source repositories

Website URL: http://www.smreza.com/projects/modelmine/

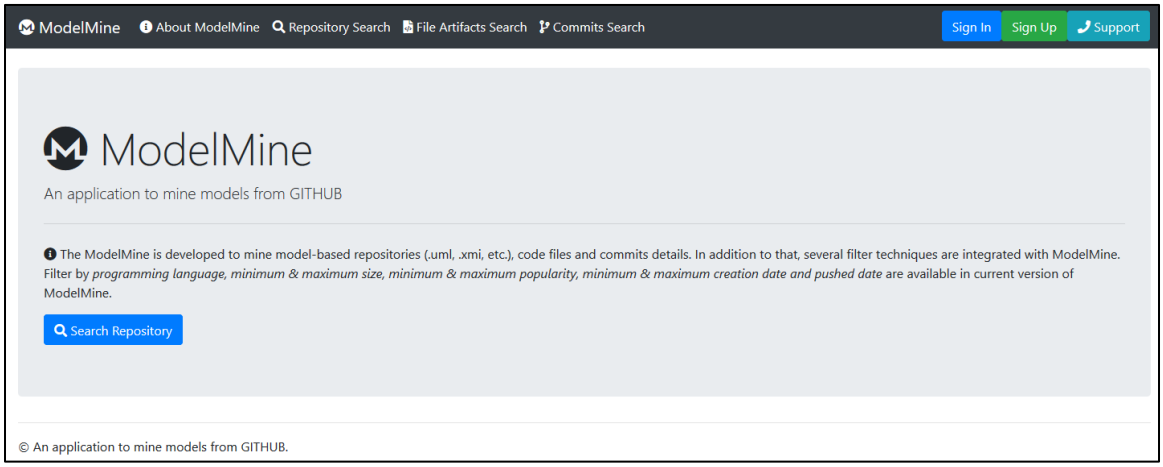


Figure 1: Homepage view of ModelMine tool

Abstract: Mining Software Repositories (MSR) has become a powerful research tool that opens a new and rich source of research data. This field enables researchers to recover data about defects, development activities, processes, patterns and more. Available mining tools are geared towards data extractions and analysis from textual artifacts.

This tool, ModelMine, a novel software repository mining tool that focuses on mining model-based artifacts. ModelMine is designed particularly to address software engineering researchers who aim to mine software repositories to uncover data about software design practices in the open-source communities. ModelMine supports features to identify repositories based on their level of modeling artifacts and querying repositories to extract models and design artifacts. It supports phase-by-phase caching of intermediate results to speed up the processing to enable the mining of many repositories.

4.4.2 PROCONF - A Conference & Peer Review Management System

Website URL: https://proconf.org/

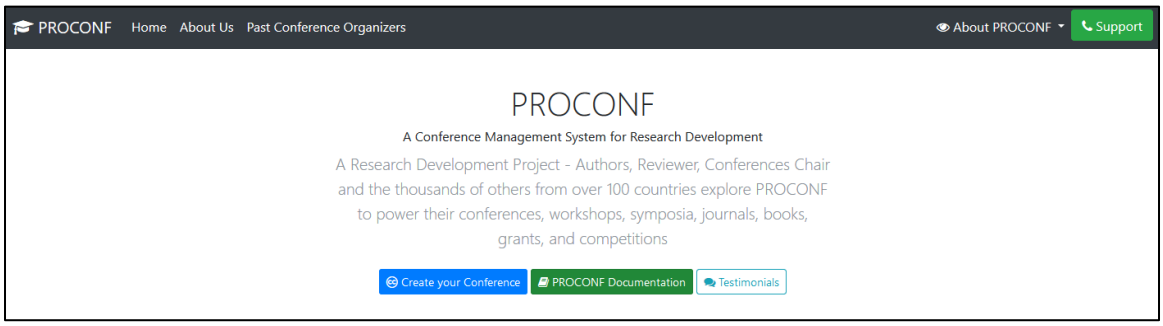


Figure 2: Homepage view of PROCONF system

Abstract: PROCONF is a web-based conference & peer to peer review management system to organize paper submission and review. PROCONF is widely used in Asian conferences, since 2014, in the scientific community, with reportedly more than 0.1 million researchers in the world. The software also provides an open access online publication service for conference proceedings. When launched, in 2014, the service was for conferences in Bangladesh only due to some business legislation, but in 2015 it was expanded to all other countries. Around 20 conferences are successfully managed & more than 80K research documents are peer-reviewed by PROCONF. It is like other event management system software such as EasyChair developed by Andrei Voronkov and OpenConf hosted by the Department of Computer Science at the University of Manchester.

4.4.3 Heat Island Effect for Borderland Area

Website URL: http://www.smreza.com/projects/heat-island-effect/

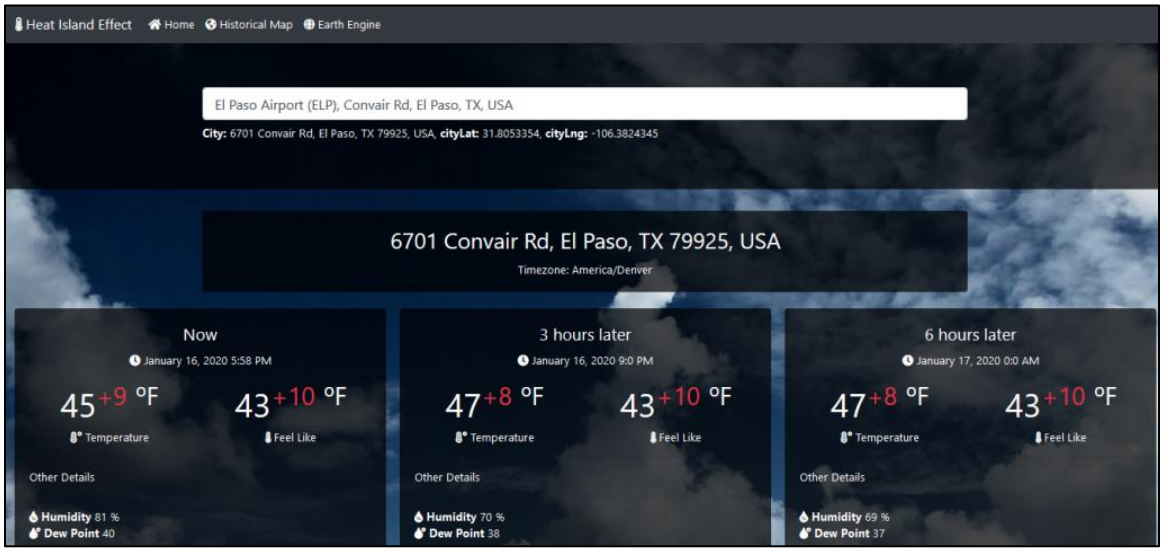


Figure 3: Homepage view of Heat Island Effect application

Abstract: An Urban Heat Island (UHI) is an urban area or metropolitan area that is significantly warmer than its surrounding rural areas due to human activities. The temperature difference is usually larger at night than during the day and is most apparent when winds are weak. UHI is most noticeable during the summer and winter [1]. The term "heat island" describes built-up areas that are hotter than nearby rural

areas. The annual mean air temperature of a city with 1 million people or more can be 1.8–5.4°F (1–3°C) warmer than its surroundings. In the evening, the difference can be as high as 22°F (12°C) [2]. Heat islands can affect communities by increasing summertime peak energy demand, air conditioning costs, air pollution and greenhouse gas emissions, heat-related illness and mortality, and water pollution [3].

Major impacts of heat island are increased energy consumption; elevated emissions of air pollutants and greenhouse gases; compromised human health and comfort; and impaired water quality [4]. In our project, the surface temperature with local metrological data is analyzed and visualized in the form of a web application where users can search their area and see the heat island effect in temperature value. The project also visualizes the historical map of heat index over the years (2002-2019). Also, the project includes the Google Earth Engine mapping system to visualize the heat index map.

4.4.4 [Software Design] Customizable Resume Maker Based on Job Requirements

Website: http://www.smreza.com/projects/hci-vita/

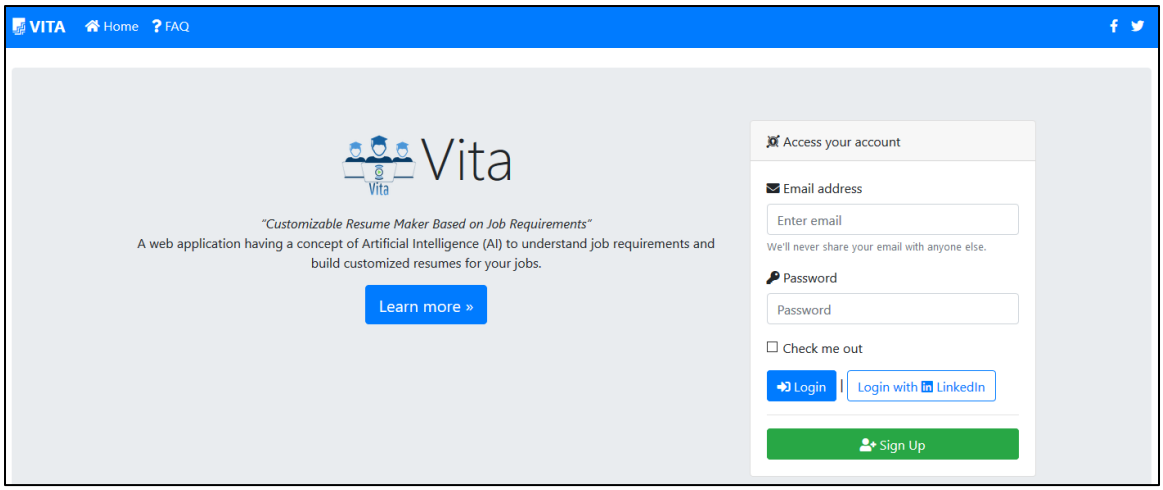


Figure 4: Homepage view of VITA software design

Abstract: Resume and CV are the documents to show a person's educations, experiences, skills, achievement, etc. Resumes and CVs are the way a recruiter can evaluate a person's qualifications for a job. By presenting well the skills and qualifications, the chance of getting the job will increase. So, the skill of presenting person skills and qualifications is very important to get hired in a company. Usually, people go to a resume reviewer to get advice and revision on their resume or CV. It will cost a person around $100 for each appointment.

The average number of people who apply for any given job is 118. Twenty percent of those applicants get an interview1. So, the users must apply for more than a job to increase the chance of hiring. On average, a person has an 8.3% probability of getting a job interview from one job application. That means it takes 10-20 applications to get one interview. And, on top of that, it takes 10-15 interviews to get one job offer1. This means a person must apply for at least 100 jobs to be sure he/she is getting the job offer. Each job has its job requirements and demands for specific skills and experience. Users can increase their chances by customizing their resume based on each job application. But it will be costly if users want to meet a resume maker to customize their resume for each job.

The purpose of building this system is to help users to prepare their resume and CV online and help them to customize their resume based on any job requirements. The system will help users to import their data easily by offering them the functionality of importing the data from LinkedIn and let them add more data manually. Then it will ask the user about the job requirements and suggest a different type of resume or CV based on that job application. Users don't need to add their data in the system in the next attempts, because their data will be stored in the system database. They just need to add a new job requirement to get a new resume or CV for that specific job. This can save time for the users, improve their resume qualification, and helps them to customize their resume for each job very fast. By reducing the time of preparing a resume, we can help the users to apply for more jobs and increase their chances to get a job faster.

APPENDICES

APPENDIX A: CURRICULUM VITA

Name: Sayed Mohsin Reza

Professional Address:

Room: 3.1202G, Department of Computer Science

University of Texas at El Paso (UTEP)

500 W University Avenue, El Paso, TX 79968

E-mail: sreza3@miners.utep.edu

Website: https://www.smreza.com

(a) Professional Preparation

University	City, State, Country	Department	Degree, Year
Jahangirnagar University	Dhaka, Bangladesh	Information Technology	B.Sc., 2014
Jahangirnagar University	Dhaka, Bangladesh	Information Technology	M.S., 2015

The semester that started Ph.D. in Computer Science at UTEP: Fall 2019

Name of your research advisor: Dr. Omar Badreddin

(b) Appointments

Duration	Job Responsibility
2019-present	Teaching Assistant, Department of CS, University of Texas at El Paso
2019	Full Stack Software Developer Intern, Pupptech Inc., Denver, Colorado
2017-2019	Research Assistant, Department of CS, University of Colorado Denver
2017-2018	Instructor, Department of CS, University of Colorado Denver
2017	Lecturer, Sonargaon University, Dhaka, Bangladesh

(c) Products

Published papers:

- Transue, S., Reza, S.M.*, Halbower, A.C. and Choi, M.H., 2018, March. Behavioral analysis of turbulent exhale flows. In 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) (pp. 42-45). IEEE.

- Rahman, S., Sharma, T., Reza, S.M.*, Rahman, M.M. and Kaiser, M.S., 2016, December. PSO-NF based vertical handoff decision for ubiquitous heterogeneous wireless network (UHWN). In 2016 International Workshop on Computational Intelligence (IWCI) (pp. 153-158). IEEE.
- Paul, M.C., Sarkar, S., Rahman, M.M., Reza, S.M.* and Kaiser, M.S., 2016, January. Low cost and portable patient monitoring system for e-Health services in Bangladesh. In 2016 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-4). IEEE.
- Reza, S.M. *, Rahman, M.M., Parvez, M.H., Kaiser, M.S., and Al Mamun, S., 2015, May. Innovative approach in web application effort & cost estimation using functional measurement type. In 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT) (pp. 1-7). IEEE.
- Reza, S.M. *, Rahman, M.M., and Al Mamun, S., 2014, April. A new approach for road networks - A vehicle XML device collaboration with big data. In 2014 International Conference on Electrical Engineering and Information & Communication Technology (pp. 1-5). IEEE.

Other products:

- Sayed Mohsin Reza* (2014, January 1). PROCONF Conference Management System (Version 1.2) [Web Application]. https://www.proconf.org.
- Sayed Mohsin Reza* (2019, December 1). ModelMine, A Tool to Mine Models from Open Source Repositories (Version 2.2) [Web Application]. https://www.smreza.com/projects/modelmine.

(d) Synergistic Activities

Activities as a Computer Science Ph.D. student at UTEP:

- I Joined 2-day Workshop: Empowering the Borderland in January 2020.
- I Joined in the ACM Unity Workshop in October 2019.
- Volunteered as Judge, Darrel Chong Student Activity Award, IEEE MGA Student Awards & Recognition Committee in September 2019.

Activities before joining CS@UTEP as a Ph.D. student (list no more than five activities):

- BHI 2018 Best Paper Award for the paper "Behavioral Analysis of Turbulent Exhale Flows" at the IEEE Conference on Biomedical and Health Informatics (BHI) 2018, https://bhi-bsn.embs.org/2018/paper-and-poster-awards/
- Joined Research and Creative Activities Symposium (RaCAS) at the University of Colorado Denver, USA in April 2019 and 2018.
- Graduate Student Member, IEEE, 2015-2019.
- Chair, IEEE Student Branch, Jahangirnagar University (Session 2015-2016).

APPENDIX B: COURSEWORK

B1: Transcript

Display Transcript

80694177 Sayed Mohsin Reza
Dec 24, 2019 02:35 pm

This is NOT an official transcript. Courses which are in progress may also be included on this transcript.

Transfer Credit Institution Credit Transcript Totals Courses in Progress

Transcript Data
STUDENT INFORMATION

Curriculum Information

Current Program
Doctor of Philosophy
College: College of Engineering
Major: Computer Science/PhD

***Transcript type:UNO is NOT Official ***

TRANSFER CREDIT ACCEPTED BY INSTITUTION -Top-

SP18: Colorado, U Of Denver

Subject	Course	Title	Grade	Credit Hours	Quality Points	R
CS	63TR	CSCI 5446 Theory of Automata	TB	3.000	0.00	
CS	63TR4	CSCI 5593 Adv Computer Archite	TB	3.000	0.00	

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Current Term:	0.000	0.000	6.000	0.000	0.00	0.00

Unofficial Transcript

F17: Colorado, U Of Denver

Subject	Course	Title	Grade	Credit Hours	Quality Points	R
CS	63TR2	CSCI 5451 Algorithms	TB	3.000	0.00	

CS	63TR7	CSCI 5565 Intro Cmptr Graphics	TB	3.000		0.00

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Current Term:	0.000	0.000	6.000	0.000	0.00	0.00

Unofficial Transcript

F18: Colorado, U Of Denver

Subject	Course	Title	Grade	Credit Hours	Quality Points	R
CS	63TR3	ISMG 7214 Mixed Methods Rsrch	TA	3.000		0.00
CS	63TR6	CSCI 7800 Rsrch in VR & AR	TA	3.000		0.00

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Current Term:	0.000	0.000	6.000	0.000	0.00	0.00

Unofficial Transcript

SP 19: Colorado, U Of Denver

Subject	Course	Title	Grade	Credit Hours	Quality Points	R
CS	63TR5	CSCI 7654 Algrthms Comm Ntwrks	TA	3.000		0.00
CS	63TR8	CSCI 7840 Indpndt Study	TB	3.000		0.00

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Current Term:	0.000	0.000	6.000	0.000	0.00	0.00

Unofficial Transcript

INSTITUTION CREDIT -Top-

Term: Fall 2019

College: College of Engineering
Academic Standing: Eligible to Re-enroll

Subject	Course	Campus	Level	Title	Grade	Credit Hours	Quality Points	R
CS	5317	Main Campus	GR	Human-Computer Interaction	B	3.000	9.00	
CS	5374	Main Campus	GR	Software Construction	A	3.000	12.00	
CS	5390	Main Campus	GR	Special Topic Computer Science	A	3.000	12.00	

Term Totals (Graduate)

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Current Term:	9.000	9.000	9.000	9.000	33.00	3.66

Cumulative:	9.000	9.000	9.000	9.000	33.00	3.66

Unofficial Transcript

TRANSCRIPT TOTALS (GRADUATE) -Top-

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Total Institution:	9.000	9.000	9.000	9.000	33.00	3.66
Total Transfer:	0.000	0.000	24.000	0.000	0.00	0.00
Overall:	9.000	9.000	33.000	9.000	33.00	3.66

Unofficial Transcript

COURSES IN PROGRESS -Top-

Term: Spring 2020

College: College of Engineering

Subject	Course	Campus	Level	Title	Credit Hours
CS	6390	Main Campus	GR	Special Topics	3.000
CS	6394	Main Campus	GR	Doctoral Research	3.000
RWS	5313	Main Campus	GR	Grant Writing	3.000

Unofficial Transcript

RELEASE: 8.7.1

B2: Example Course Artifacts

In this subsection, I demonstrate the required course details and artifacts.

CS 6315 Theory of Computation

Transferred Course - CSCI 5446 Theory of Automata

Assignment 1 – Computability, Unsolvability

Chapter 1 - Computability

The NICE Programming Language

1. Define a model of computation that does not depend on computers or programming.

Answer: Machines do not provide a complete model for computational problem-solving. Computer programming paradigms are the input for a model. Computer programming behavior is fixed and does not depend on time. It is the study of a class of concrete and well-defined phenomena of the running time of the computation on a computer or programming.

However, the Turing model is not as well matched for the natural, interactive, and continuous. With this model, we can incorporate the structured programming movement to make an easy human computation.

Turing focused upon human computation and thought about the way that people compute things by hand. The model includes reading data from a tape, writing corresponded to that data and follow the same or next instructions. It executes a specific list of instructions placed on a scratch tape.

This model doesn't need any computer or programming for computation.

2. We used floating-point numbers instead of real numbers in our programming language. Why?

Answer: Our only restriction will be that at any instant during a computation everything must be finite. That means no numbers or arrays of infinite length. Huge - yes, but not infinite! This means that the infinite decimal expansion 0.333... for one third is not allowed yet several trillion 3's following a decimal point is quite acceptable.

We should also note that even though we have a number type named real, these are not real numbers in the mathematical sense, but floating-point numbers.

```
program superexpo(x, y)
var m, n, w, x, y, z: integer;
begin
  w = 1;
  for m = 1 to y do
      begin
        z = 1;
        for n = 1 to w do z = z*x;
        w = z
      end;
  halt(z)
end
```

3. Examine the following program: What are the values of the function it computes when y equals 1, 2, and 3? Describe this function in general.

Answer:

function superexpo(x, y)	x	y	w
	x	1	x
	x	2	x^x
	x	3	x^{x^x}

4. How many multiplications are performed by the programs named expo and super Expo? (Express your answer in terms of x and y.)

Answer:

	Multiplications performed
program expo(x, y)	$y\ times$
program super expo(x, y)	$1 + x + x^x + x^{x^x} + \cdots . + x^{(y-1)x}\ times$

5. Write a program which computes the function

$$bar(x)\begin{cases} x\ if\ is\ odd\ and\ positive \\ undefined\ otherwise \end{cases}$$

Combine this with the program for the function fu(x) from the NICE language section to get an identity function program.

Answer: This halts only for odd, positive integers and computes the function described as:

$$fu(x) = x\ if\ x\ is\ odd\ and\ positive, otherwise\ undefined$$

Here is the program that computes odd and positive integers otherwise halts.

```
program fu(x)
var n, x: integer;
begin
     n = 1;
     while not x = n do n = n - 2;
     halt(x)
end
```

Turing Machines

6. What does the Turing machine of figure 2 that adds 1 to its input do when given #000 as input? What about the inputs: #bbb, #b011, and #11b10?

sweep right to end of input

	read	*write*	*move*	*goto*
I1	0	0	right	same
	1	1	right	same
	#	#	right	same
	b	*b*	left	next

change 1's to 0's on left sweep, then change 0 to 1

I2	0	1	halt	
	1	0	left	same
	#	#	right	next

input = 11...1, so sweep right printing 1000...0 (print leading 1, add 0 to end)

I3	0	1	right	next

I4	0	0	right	same
	b	0	halt	

Figure 2 - Successor Machine

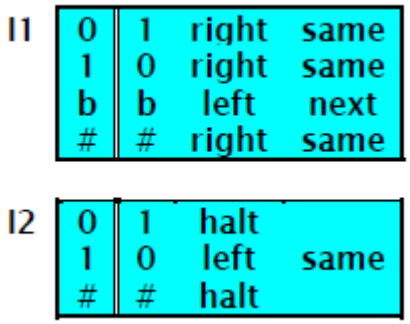
Answer:

Turing Machine – ND – Not defined

	1st					2nd				
Input	#	0	0	0	b	#	b	b	b	b
	#	0	0	0	b	#	b			
				1(halt)		#	ND			
Output	#	0	0	1		#				

	3rd						4th						
Input	#	b	0	1	1	b	#	1	1	B	1	0	b
	#	b					#	1	1	B			
	#	ND					#	0	0				
								1	0	0(halt)			
Output	#						#	1	0	0	1	0	

7. Examine the following Turing machine:
What does it do when present with the inputs #1011, #1111, and #0000? In general, what does this machine accomplish?



Answer:

	1st						2nd						3rd					
input	#	1	0	1	1	b	#	1	1	1	1	b	#	0	0	0	0	b
	#	0	1	0	0	b	#	0	0	0	0	b	#	1	1	1	1	b
					1(halt)						1(halt)		#(halt)	0	0	0	0	
output	#	0	1	0	1		#	0	0	0	1		#	0	0	0	0	

8. Design a Turing machine that subtracts 1 from its input.

Answer: To subtract 1

Instructions	Read	Write	Move Head	Next Instruction
I1	0	0	right	Same
	1	1	right	same
	#	#	right	same
	b	b	left	Next
I2	0	1	left	Same
	1	0	halt	
	#	#	halt	

9. Write down the instructions for a Turing machine which determines whether its input is zero. What happens when this machine is given a blank tape as input?

Answer:

Instructions	Read	Write	Move Head	Next Instruction
I1	0	0	Right	next
I2	0	0	Right	Same
	B	B	halt	

10. Design a Turing machine that computes the fu(x) function of the NICE language section.

Answer:

	Read	Write	Move Head	Instruction
I1	0	0	Right	Same
	1	1	Right	Same
	#	#	Right	Same
	B	B	Left	Next
I2	0	0	Halt	

A Smaller Programming Language

11. Describe the ways in which division must change after floating-point numbers have been replaced by triples of integers which denote their signs, absolute values, and decimal points.

Answer:

```
Reduce = 0
If pointy < pointx then
        While pointy < pointx do
                Begin
                        Y= y/10
                        Pointy = pointy +1
                        Reduce = reduce + 1
                End
Else
        While pointy > pointx do
                Begin
                        Y= y*10
                        Pointy = pointy -1
                        Reduce = reduce -1
                End
Z = x/y
If reduce = 0
        Pointz = reduce
If signx = singy then signz =1
Else signz =0
```

12. Define integer division. Show that division can be replaced by subtraction in much the same way that multiplication was replaced by addition.

Answer:

```
z =0
While x not = 0
        z = z + 1
        x = x -y
```

13. If we allow the predecessor operation (x = x - 1) to be included in a programming language, then subtraction is not needed. Show this.

Answer:

```
z =0
While x not = 0
        z = z + 1
        while n = 1 to y
                x = x –1
```

Equivalence of the Models

14. Design a "blank-squeezing" Turing machine. That is a machine that converts #xbby to #xby.

Answer:

	read	write	move	goto
I2	0	0	Right	Same
	1	1	Right	same
	#	#	Right	Same
	b	b	Left	Next

	read	write	move	goto
I2	b	b	left	same
	1	b	right	I3
	0	b	right	I4
	#	b	right	I5

	read	write	move	goto
I3	b	1	left	I2

	read	write	move	goto
I4	b	0	left	I2

	read	write	move	goto
I5	b	#	halt	

15. Discuss size trade-offs between Turing machines and programs that compute the same function.

Answer: Small programs can be executed in $O(\log n)$ time or $O(n)$ time which means linearly. This is the running time which we generally get from small programs and this is what is expected from bigger problems as well. It is hard to achieve $O(1)$ running time which is made possible in case of hashing. But there are worst cases too.

But in case of a turning machine, it is also the same just one need to make sure is that the machine does halt. The key insight is that for a Turing machine to halt, it cannot enter a loop. Since a Turing machine will always follow the same sequence after being a specific state, if it ever becomes that same state twice, we know the machine is caught in an infinite loop and will never finish. Therefore, the theoretical maximum number of steps it can run is the maximum number of possible different states for the machine without being the same state twice.

Chapter 2 - Unsolvability

Arithmetization

1. Design an algorithm for a Turing machine that recognizes encodings of Turing machine instructions. (Allow it to use the alphabet: {0, 1, b, s, n, ...}.) Indicate how this machine could be the basis for one which recognizes Turing machine descriptions.

Answer: Let us begin our task. If we take an instruction line such as:

0	1	Left	same

lose the nice formatting, and just run the parts together, we get the string:

$$01leftsame$$

This is still understandable since we were careful to use only certain words in our Turing machine instructions. We shall modify this a little by translating the words according to the following chart.

Character	0	1	b	same	next	left	right	halt	instruction no	string separator
Encoded	0	1	b	s	n	$\leftarrow$	$\rightarrow$	$\downarrow$	I17	·

This translation converts our previous instruction line to:

$$01 \leftarrow s$$

which is a little more succinct, but still understandable. In the same manner, we can transform an entire instruction such as:

0	1	left	same
1	1	right	I17
b	0	halt	

into three strings which we shall separate by dots and concatenate. The above instruction becomes the following string of characters.

$$01 \leftarrow s \cdot 11 \rightarrow I10001 \cdot b0 \downarrow$$

The instruction has been encoded.

2. Describe an algorithm for a Turing machine which receives the integer n as input and proceeds to write the description of the n^{th} Turing machine from the standard enumeration on its tape.

Answer: Turing machine takes input from the tapes and according to instructions it produces output. For executing any program, we need to design a Turing machine.

Let's consider the tuning machine which receives the n as input. This Turing machine can be done by encoding all the Turing machines and then ordering them according to this encoding. First, we assign some symbols to our Turing machine instructions.

Character	0	1	b	same	next	left	right	halt	instruction no	string separator
Encoded	0	1	b	s	n	←	→	↓	I17	·

Let's take some entire instruction and encode this instruction using the above symbols.

0	1	left	Same
1	1	right	I17
b	0	halt	

This above instruction can be encoded as shown below into three strings, separated by dots.

$$01 \leftarrow s \cdot 11 \rightarrow I10001 \cdot b0 \downarrow$$

So, form the above discussion there are some instructions by which the Turing machine accepts n as input and proceeds further.

Once the number is identified (in our case "n") the control will switch over to the n^{th} Turing machine and starts reading from its tape. This information can be got by a list called index which contains all listing of Turing machines which is the standard Enumeration.

Algorithm for a Turing machine which receives the integer n as input and proceeds to write the description of the n^{th} Turing machine is as follows: -

```
Case:
        0: begin
                Print a 1;
                Move one tape square left;
                Go to same instruction
                End;
        1: begin
                Print a 1;
                Move one tape square right;
                Go to next instruction I35
                End;
        Blank: begin
                print 0;
                halt end;
End case.
```

3. As we know from our study of computability, programs can be translated into Turing machines. If we check off these Turing machines (those which are transformed programs), we find that there are still some leftovers in our standard enumeration. Does this mean that there are more Turing machines than programs? Comment.

Answer: No, according to the theorem there are the same number of programs as the Turing machines the total number of Turing machines in the standard enumeration is $\aleph_0$. for every program we can derive a one-one relation with the Turing machines then we can say that bijection

exists, and the number of programs is equal to ℵ0. Each program has a separate Turing machine. And for a moment if we assume that there is some Turing machine that is not connected with any programs then there is not any necessity of such a machine. There is no meaning of any existence of such a Turing machine that cannot be associated with any program.

Properties of the Enumeration

4. Show that there are exactly as many rational numbers as there are nonnegative integers.

Answer: Let Q^t be the set of rational numbers that are nonnegative and $\aleph \cup \{0\}$, the set of nonnegative integers. Then we can describe as

$$Q^- = \{0\} \cup (\bigcup_{k=1}^{\propto} Ak)$$

Where

$$Ak = \left\{\frac{1}{k}, \frac{2}{k}, \frac{3}{k}, \frac{4}{k}, \ldots\ldots, \frac{n}{k}, \ldots\ldots\right\}, k \in \mathbb{N}$$

We construct a mapping

$$f: \mathbb{N} \cup \{0\} \to Q^+ \; by$$

$$f(0) = 0$$
$$f(k) = A - k, k \in \mathbb{N}$$

It is a one to one correspondence.

Since $\mathbb{N}$ in infinite, Q^+ the set of nonnegative rational is infinite.

5. Every Turing machine is equivalent to some other machine in our enumeration. Why? How many other machines are each machine equivalent to? How many times is each Turing-computable function represented in our enumeration? Be sure to justify your answers.

Answer: Some Turing machines seem equal, but they have the instructions in a different order. So, they are not the same. one can observe that adding additional instructions (which are never executed) to a machine changes its description but does not alter its behavior. There may be $\aleph 0$ number of Turing machine which are equivalent to some other machine in our enumeration because of $\aleph 0+\aleph 0=\aleph 0$ and $\aleph 0*\aleph 0=\aleph 0$. Each Turing machine computable function is represented $\aleph 0$ times in our enumeration. Because if there is a small change even in the order of instruction it will be considered as a new Turing computable function.

6. Prove that there exist more real numbers between zero and one than there are integers.

Answer: Consider the set

$$S = \{n \in \mathbb{R}: 0 < n < 1\}$$

Assume S is finite. Then S has a least element. Call the least element m. Then for all $p \epsilon S$, where $p \neq 1$, it is the case that 0<m<p. But consider $p = \frac{1}{2}m$. By hypothesis, $m = \frac{1}{2}m$. Since m>0, it follows that $1 < \frac{1}{2}$, which is a contradiction. Thus, S is infinite.

We find more real numbers between zero and one than there are integers.

7. Some infinite sequences of zeros and ones (such as 1010101... or 1101001000...) are not too difficult to compute. We could design Turing machines which compute them. Are there any binary sequences that cannot be computed by Turing machines? Why?

Answer: Yes, there are binary sequences that cannot be computed by Turing machines. To prove this, let us assume that all the binary sequences are computed by the Turing machines and derive a contradiction. We can use the Cantor theorem of diagonalization**.**

Let's take binary sequences $B1, B2, B3, \ldots, Bn$. Now define a binary sequence D, by choosing a 0 in the first column if B_1 has a 1 in that column and 1 if B_1 has a 0 in that column. We then choose a 0 in the *second* column if B_2 has a 1 in that column and 1 if it has a 0 and so on. The resulting binary sequence, D, cannot be in the list because it did not match with any of the binary sequences in the Turing machine, say B for some *n*. But we have just deliberately made sure that the *n*th column of D differs from B_n. this contradicts our assumed statement which clearly states that some binary sequences cannot be computed by the Turing machine. According to Theorem #3, some sets are not computable. So, we can find some strings of 0s and 1s that cannot be computable by any Turing Machine. Because they cannot be placed in one-to-one correspondence with the integers. It is a fact that there are countable Turing Machines and thus countable computable sets, but many uncountable sets.

Solvability and the Halting Problem

8. Might the following be a Turing machine? Explain your reasoning.

$$M(i) = \begin{cases} halt\ if\ M\ (n)\ halts\ for\ some\ input \\ diverge\ otherwise \end{cases}$$

Answer: Yes, this is a Turing machine. Let's consider the Turing machine which accepts a set and halt for some input, then that input is a member of a set. Thus, Turing machine halts for members of the set and provides no information about inputs that are not members. Let's take k is the set of Turing machine that halts when given their indices as input.

$$K = \{i | M_i(i) halts\} = \{i \mid i \in W_i\},$$

Consider the machine M that constructed from universal Turing machine Mu as follows: $M(i) = M_u(i, i)$.
And since it was just $Mu(i, i)$ i) we know exactly how to build it and even find its index in our standard enumeration.

Every Turing machine takes some input and gives some output according to the set of instructions. There are only two possibilities for any Turing machine is whether it will diverge or halt. Here, the above Turing machine takes some inputs and gives the output. This Turing machine halts at some input and diverged at all other inputs. So, we can say that the above is a Turing machine.

9. We showed that since the problem concerning a machine halting on its index is unsolvable, the general halting problem for Turing machines is unsolvable. Does this imply that any superset of an unsolvable problem is unsolvable? Provide proof or a counterexample.

Answer: Yes, any superset of the unsolvable problem is unsolvable too. According to theorems of halting problem states that.

- Theorem 1 states that whether Turing machine halts when gives its index is unsolvable.
- Theorem 2 states that general halting problem for arbitrary integer $i\ and\ x$, whether or not $M_i(x)$ halts are unsolvable. So, for every Turing machine, it will be unsolvable.
- Theorem 3 states that whether an arbitrary Turing machine halts for all inputs is an unsolvable problem too.

Proof: Let's consider the halting problem machine M be turning machine which is solving the halting problem. Let's take UM be the "Unsolvable Machine M_{UM}," or the one that will cause the halting problem machine to fail. The UM itself is an adaptation of a Universal Turing machine Mu, and its actions depend on the instructions of the Halting problem machine.

If machine M could compute and it will read the whatever encoded instructions and input which MUM would read after a finite number of steps then the machine M going to decision procedure as" yes" it will tell that machine M_{UM} is going to halt or either "no" means it will not halt.

Then after computation started the machine M_{UM} will behave as Mu, and read the machine M encoded instruction as its input, then its MUM code is executed on secondary, the machine M will compute and decide as yes or no.

If the machine M_{UM} process shows that machine M decides as yes, then machine MUM will halt then instructions as its input begins to diverge. Then again machine M_{UM} process shows that machine M decides No the MUM will not halt then the machine will not write the instruction it causes the contradiction the machine M.

The machine will not able to predict the decision procedure that it will halt or not halt on a finite number of steps then it fails and its contradiction.

By calculating what machine M predicts do and then doing the opposite the machine MUM can propose a Turing machine that is supposed to be able to solve the halting problem.

10. Prove that the membership problem for any finite set is solvable.

Answer: Yes, the membership problem for any finite set is solvable.

Proof: Let's consider a set of all squares of first 1000 natural number, that means set of all squares of first 1000 natural numbers is a countable finite set. It can be solvable by using the cantor theorem of diagonalization.

Let's design a Turing machine M which identifies squares of an input natural number.

The Turing machine M which accepts natural number input K and identifies its square i.e. K^2.

$$A = \{K^2 : K \in \mathbb{N} \; and \; K \leq 1000\}$$
$$i.e. A = \{1,2,4, \ldots . , K^2\}$$

Since this K^2 is Perfect Square of natural number which is present in the above finite set. Then machine M performs function f $i.e. f(K) = K^2$. Now consider the table which contains all set square numbers of finite sets which is accepted by machine M as shown below.

F(k)	1	2	3	4	…	k
1	**1**	2	3	4		
2	2	**4**	6	8		
3	3	6	**9**	12		
4	4	8	12	**16**		
…						
k	k					

The above table shows that each element up to k which belongs to N and it is in the table. According to the above table, we can tell that all members in the list got itsk^2along the diagonal which is also a member in set {1, 4, 9, 16}. And we know K is less than or equal to 1000. So, we can design a Turing machine that can accept each member of a finite set.

So, in this case, each K^{th} element is less than 1000 and in general, each member is less than or equal to any finite set endpoint. So as per the above proof, we can tell that any finite set membership problem is solvable.

11. Show that whether an arbitrary Turing machine ever executes one of its instructions is unsolvable. (This is the same as the problem of detecting unreachable code in a program).

Answer: The idea behind this problem is to show that it cannot be determined whether an instruction is executed by a Turning machine ever or not. This can be reasoned out as follows:

Consider a Turing machine M, which takes another Turing machine M_i and instruction in M_i as input and outputs whether the instruction is executed or not.

This machine (M) can be used to determine whether M_i halts or not, by checking for each instruction that leads to decision problem which accepts or reject, whether it is executed by M_i or not. This is unsolvable. So, it is undecidable to say whether an instruction is executed or not.

The problem is to decide whether a machine M_i halts on an input w. Suppose that there exists a machine M that reads machine M_i and input w and determines whether M_i halts on w. We define a machine M_s as follows:

It reads the description of M_i as input, and using M, determines whether M_i halts when its description is given as input to it. If M_i is found to halt on its description, then M_s will go into an infinite loop. If M_i does not halt on its description, then M_s halts.

Now, if M_s is given as input to itself, going by the description of M_s, M_s will halt if M_s does not halt on input M_s and M_s will not halt, if M_s halts on input M_s this is naturally impossible. So, our initial assumption about the existence of M is false. So, we cannot decide whether a machine halts on input or not. Hence when an arbitrary Turning machine is considered, it may or may not have dead code like instructions. Therefore, whether an arbitrary Turing machine ever executes an instruction is unsolvable.

Enumerable and Recursive Sets

12. Are the following sets recursive? Are they recursively enumerable? Justify your conjectures.

a) $\{ x \text{ is an even integer} \}$
b) $\{ i \mid M_i \text{ halts for all inputs} \}$
c) $\{ i \mid M_i \text{halts only for prime integers} \}$
d) $\{ i \mid M_i \text{ is not a Turing machine} \}$

Answer:

a) A Turing machine can be constructed to accept the set of even integers. Hence this set is recursively enumerable. Hence, according to the theorem if a Turing machine exists which accepts the set, then the set is recursively enumerable.
Theorem 2 states that a set is recursive if and only if both the set and its complement are recursively enumerable. Say

$$T = \{x \mid x \text{ is an even integer}\}$$
$$T' = \{x \mid x \text{ is an odd integer}\}$$

Since Turing machines can be constructed for both T and T', the sets are recursive.

b) Let us name the set as T. So,

$$T = \{i \mid M_i \text{ halts for all inputs}\}$$

If M_i halts for all the inputs, then it is one of the enumerated Turing machines. Since it is a Turning machine, it can be implied that the set A is accepted by it and hence the set T is recursively enumerable**.**
Now, T' will be the set for which M_i diverges. If we can say whether an integer x is a member of T then, we can also say if it is not a member of A by evaluating the Turing machine which doesn't halt for all the inputs. Hence there is a Turning machine for T' as well. Since T and T' both can be accepted by Turing machines, by corollary they are recursively enumerable.
Hence T is a recursive set.

c) Let us name the set as A. So,

$$A = \{i \mid M_i \text{ halts only for prime numbers}\}$$

If M_i halts only for prime numbers, then it is one of the enumerated Turing machines. Since it is a Turning machine, it can be implied that the set A is accepted by it and hence the set A is recursively enumerable**.**
Now, A' will be the set for which M_i diverges. If we can say whether an integer x is a member of A then, we can also say if it is not a member of A by evaluating the non-primes which diverges on Mi. Hence there is a Turing machine for A' as well. Since A and A' both can be accepted by Turing machines, by corollary they are recursively enumerable.
Hence A is a recursive set.

d) Let us name the set as A. So,

$$A = \{i \mid M_i \text{ is not a Turing machine}\}$$

In this case, to have the set A as recursively enumerable, we need to construct a Turing machine which can accept the set A. That is, we need a Turing machine which identifies if any M_i is a Turing machine or not. We need the Turing machine to decide if a machine halts or diverges. It is solvable to determine if a machine halt. But no machine can determine the divergence of another machine. Hence, such a machine cannot be constructed. This implies that there is no Turing

machine that can accept set A and A' hence set A is not recursively enumerable**.** By corollary of Theorem 2, since the set A is not recursive, it is not a recursive set**.**

13. Prove that if the set A is not recursively enumerable and can be reduced to the set B, then B cannot be recursively enumerable.

Answer: Suppose that B is recursively enumerable. Then, we can create a recognizer for A as follows:

Given a string to be checked for membership in A, it applies the transformation that converts the strings from A to the equivalent strings in B. Then, it uses the recognizer for B to decide whether the obtained string is in B. If so, the original string is accepted.

This recognizer cannot exist since A is not recursively enumerable. So, B cannot be recursively enumerable.

14. Show that the following sets are not recursively enumerable.

a) $\{i \mid W_i = \emptyset\}$
b) $\{i \mid W_i = all\ integers\}$

Answer:

a) As the set is empty then Turing Machine will not halt for any input. So, it diverges at every input and a set is recursively enumerable. if and only if it is accepted by a Turing Machine. But by Collorary X, the divergence of a Turing Machine cannot confirm by another Turing Machine. So, it is not recursively enumerable.

b) Here Turing Machine halts at every input but it will not diverge anywhere. So, this seems an unsolvable problem. Thus, it is not recursively enumerable.

15. Prove that every index set which is recursively enumerable is a complete set.

Answer: Any indexing of the partial recursive functions the complement $\bar{k}$ of the set

$$k = \{x \in \mathbb{N} \mid \psi_x(x) converges\}\ is\ not\ recursively\ enumerable$$

If $\bar{k}$ is recursively enumerable, then k is also recursively enumerable. A set k^k is recursive iff both k and the complement $\bar{k}$ are recursively enumerable.

The set k would be recursive a contradiction. The $\bar{k}$ and $\overline{k_0}$ are examples of sets that are not recursively enumerable.

This shows that the recursively enumerable set is not closed under complementation. However, we leave it as an exercise to prove that the recursively enumerable sets are closed under union and intersection. It is a complete set.

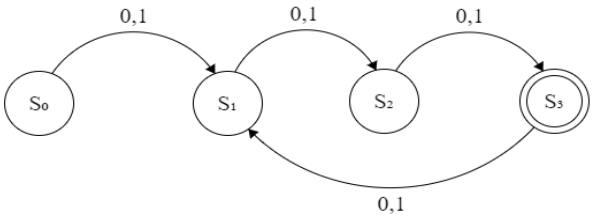
So, a recursively enumerable index set is a complete set.

Assignment 2: Finite Automata, Pushdown Automata

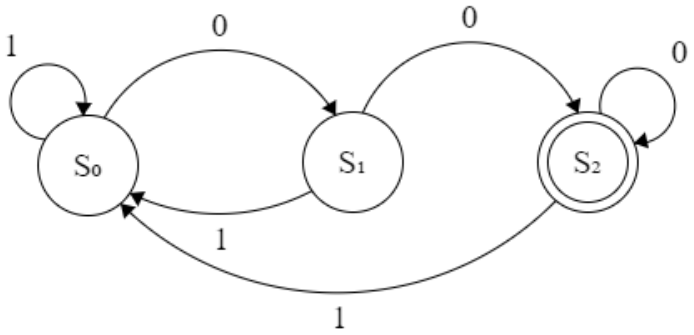
Finite Automata

1) 1. Draw the state graphs for the finite automata which accept sets of strings composed of zeros and ones which:

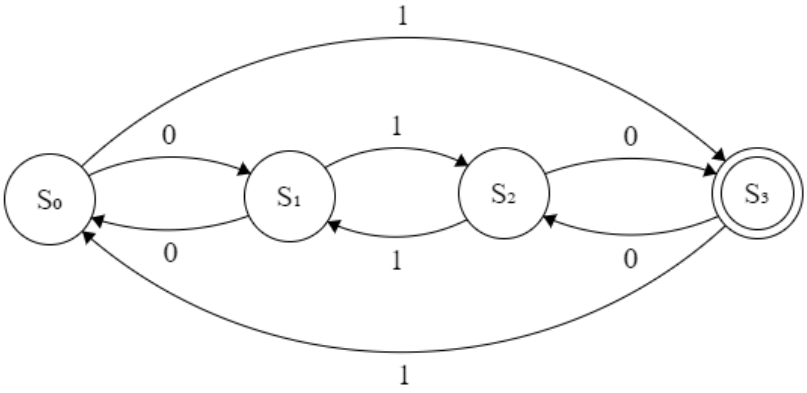
a. Are a multiple of three in length



b. End with the string 00



c. Possess runs of even numbers of 0's and odd numbers of 1's.



2) 2. Describe the sets accepted by the finite automata pictured below.

a) The first finite automata don't accept any set since there is no incoming arrow to the accept (final) state.

b) The second automata accept all strings that end with 0. As the graph shows all incoming arrows to the final state are 0s.

3) 4. Define formally and provide state tables for the finite automata that accept strings of zeros and ones which:

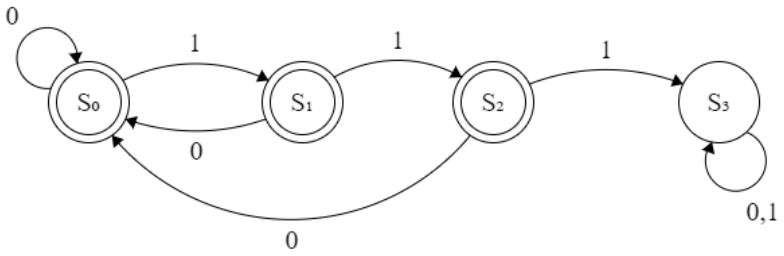
Answer: A finite automaton M is a quintuple M = (S, I, δ, s_0, F) where:

- S is a finite set (of states)
- I - finite alphabet (of input symbols)
- δ: S × I → S (next state function)
- s0 ∈ S (the starting state)
- F ⊆ S (the accepting states)

a) Never contain three adjacent ones.

Let's consider number of strings which accepted by FSA like A= {10, 100, 1000, 110101, 10100, 01011, 110010, 00011, 1010101000, 101110000….}

M = ({s_0, s_1, s_2, s_3}, {0,1}, δ, s_0, {s_0, s_1, s_2})

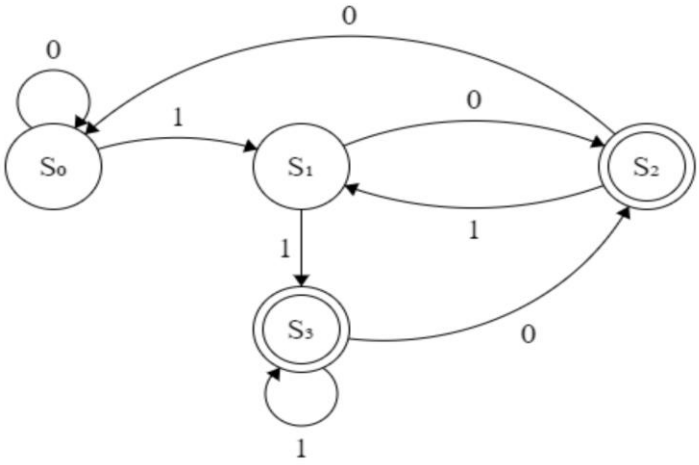


States	Input		Accept
	0	1	
S_0	S_0	S_1	yes
S_1	S_0	S_2	yes
S_2	S_0	S_3	yes
S_3	S_3	S_3	no

b) Have a one as the next to last symbol.

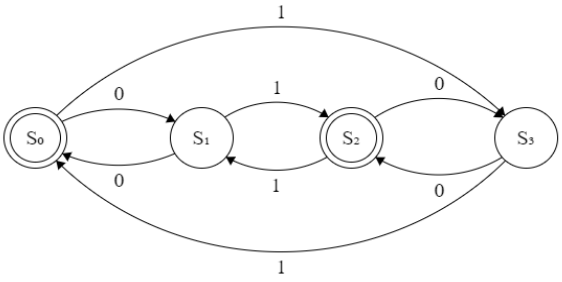
Answer: Let's us consider following string A= {0010, 100011, 00011010, 1001010, 0110, 1110001010……}

M = ({s_0, s_1, s_2, s_3}, {0,1}, δ, s_0, {s_2, s_3})



States	Input		Accept
	0	1	
S0	S0	S1	No
S1	S2	S3	No
S2	S0	S1	Yes
S3	S2	S3	Yes

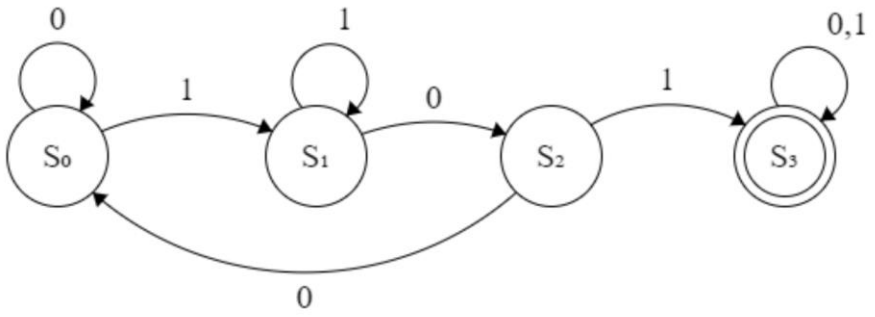
c) Contain an even number of zeroes or odd numbers of ones, not both.
$M = (\{s_0, s_1, s_2, s_3\}, \{0, 1\}, \delta, s_0, \{s_0, s_2\})$



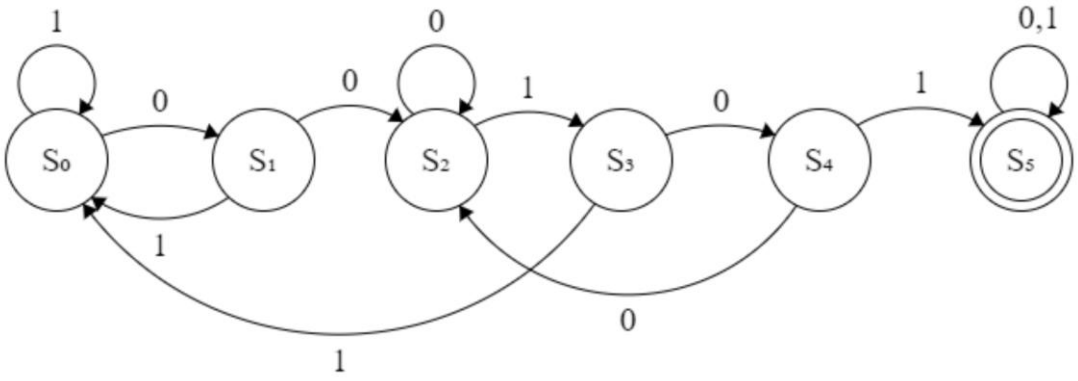
States	Input		Accept
	0	1	
S0	S1	S3	Yes
S1	S0	S2	No
S2	S3	S1	Yes
S3	S2	S0	No

4) 5. String searching and pattern matching can be done easily by finite automata. Design a machine that accepts only strings containing 101 as a substring. Now do the same (design a machine) for the substring 00101.

a) a machine that accepts only strings containing 101 as a substring



b) a machine that accepts only strings containing 00101 as a substring



Closure Properties

5) 1. Suppose that the finite automata M_i and M_k accept strings over the alphabet {0,1}. Design an automaton that accepts strings of the form x#y where x is accepted by M_i and y is accepted by M_k.

Answer: Let the finite automaton for two machines that accept strings of the form x#y.

Machine $M_i = (S, I, \delta, s_0, F)$

Machine $M_k = (Q, I, \gamma, q_0, G)$

As the strings are separated by #. We can design a machine M_{ik}, that can accept a full string and look at the operation of M_i, accepting then looking for the transition table that goes from accepting states of M_i to the start state of Machine M_k consuming # letter. Accepts states of M_k will be accepted states of M_{ik}.

$\xi, = \delta \cup \gamma \cup \{(F, \#) = q_0\}$

$S_0 = s0$ (start state of the M_{ik} is start state of M_i)

$G = G$ (final states of the M_{ik} are final states of M_k)

Machine $M_{ik} = (S \cup Q, I, \xi, S_0, G)$

6) 2. Prove that the class of sets accepted by finite automata is closed under intersection. In other words, given M_i and M_k construct the finite automaton M_m such that:

$$T(M_m) = T(M_i) \cap T(M_k)$$

Let $M_i = (S, I, \delta, s0, F)$ and $M_k = (Q, I, \gamma, q0, G)$ be two arbitrary finite automata. To prove the theorem, we must show that there is another machine (M_m) that accepts strings accepted by both M_i and M_k.

The proving process of finite automata closed under intersection is like the process of proving finite automata closed under union. multiprocessing pebble machine concept will be used here too. The only difference from the union is instead of considering combined states that contain either accept states from Mi and Mk, the accept states are only those states that have combined accept states from both.

Machine Mi = (S, I, δ, s0, F)

Machine Mk = (Q, I, γ, q0, G)

Mik = (S×Q, I, ξ, <s0, q0>, H)

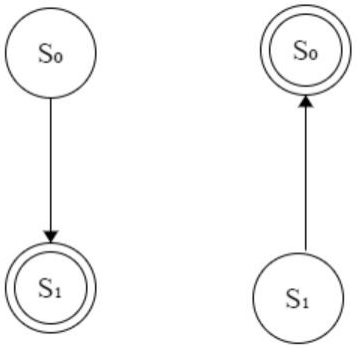
ξ is the combination of both transition function of δ and γ. ξ(<si, qi >, a) = <δ(si, a), γ(qi, a)>. H represent final states that H = (F×Q) ∩ (S×G). Thus formally, for all string x over the alphabet I. x ∈ T(Mm) iff δ*(s0,x) ∈ F and γ*(q0,x) ∈ G iff ξ*(<s0, q0 >,x) ∈ H.

7) 3. Let X^R denote the reversal of the string X. (for example, if X=010011 then X^R =110010 and so forth.) Prove that the sets accepted by finite automata are closed under string reversal by constructing for any finite automaton, a new machine that accepts the reversals of the strings accepted by the original automaton.

solution: Let the finite automaton M= (S, I, δ, S_0, F) accept the set T(M). We must now show that there is a finite automaton that accepts the reversed strings accepted by M. We could do the following and build M^1 which accepts the reversed strings of M by the below steps.

We use X^R for the reversal of string X, we know that X^R is the string written backward that's a_n a_{n-1}……. a_1 of word X=a_1, a_2,…, a_n. The reversal string M^1 is M^{1R}= {X^R | X ∈ M^1}.
The following steps describe how the finite automaton that accepts the reversal of the strings accepted by the original automata.

- Reverse all the arcs (edges) in the transition diagram M.
- Exchange start state and the accepting state.



Then the new automaton becomes M^1= (S, I, **φ,** P_0, S_0) our new machine M^1 has state set Q ∪ (Q× I) where Q is the start set of M; It has transitions of the type δ^1(q, a)= [q, a] for all q ∈ Q and a ∈ I and transition of the type δ^1([q, a],b)= δ(δ(q, b), a) for all q ∈ Q and a, b ∈ I its start state is q_0, the start state of M; and its accepting states are the accepting states of M.
Example: Consider the binary strings which are accepted by finite automata like x=010011 the x^R =110010 i.e. we can design the graph using above binary strings and create an FSA. **i.e.** saying that 0 should be the 4th from the left is reverse of saying that 0 should be 4th from the right. We can use the string reversal sets accepted by finite automata.
8) 5. The *minus* operator on sets is usually defined as:

A - B = {x | x ∈ A and x ∉ B}.

Prove that the class of sets accepted by finite automata is closed under minus.
Solution: The difference of two sets is A - B = {x | x ∈ A and x ∉ B}. We can prove this using the relation A−B= A⋂B′. We know that if a set is accepted by finite automation then its complement can be accepted by finite automation and also, the class of sets accepted by finite automata is closed under intersection, As we have seen proofs for these theorems in the class. Therefore, from these two theorems, B` is accepted and so A ∩ B` are accepted by finite automata and so A-B` is closed and is accepted by finite automata.

Regular sets and Expressions.

9) 1. What are the regular expressions for sets of strings composed of zeros and ones which?

a) Are a multiple of three in length.

Answer: The regular expression for sets of strings that are multiple of three in length is: $((0+1)\ (0+1)\ (0+1))^*$

b) End with the string 00.

Answer: The regular expression for the set of strings that end with the string 00 is: $(0+1)^*00$

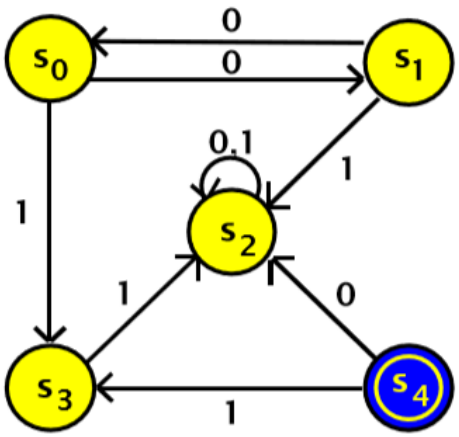
The least string that could be possible by the above regular expression is 00.

c) Possess runs (substrings) containing only even numbers of zeros and odd numbers of ones.

Answer: The regular expressions for a set of strings that contain only even numbers of zeros and odd numbers of ones are:
$(00+11+\ (01+10)\ (00+11)^*(01+10))^*(1+(01+10)(11+00)^*0)$

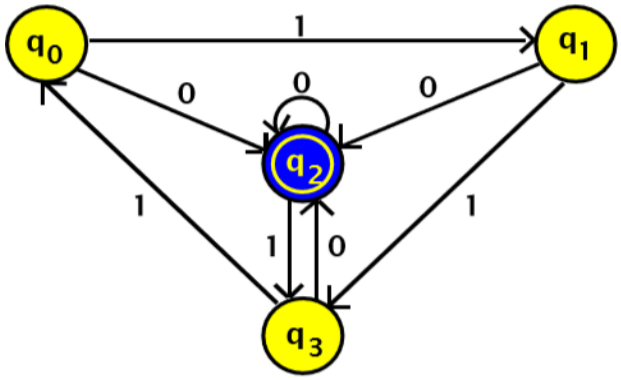
10) 2. Derive the regular expressions for the sets accepted by the finite automata whose state graphs are pictured in the second problem of the first section?

a)



Answer: There will be no regular expression as an accept state can't be reached by any strings. Thus, it's ∅.

b)



Answer: The regular expression is (0+1) * 0.

Decision problems for finite automata.

11) 2. Can the following sets of strings be accepted by finite automata? Justify your answers!

a) $\{1^n \mid n \text{ is a prime number}\}$

Solution: This is not a regular set. This language is infinite. This can be proved using Pumping Lemma. Assume X= 1111…11111

---u------v------w----

i) X= uvw

ii) Let |v|=a and |uw| = b has to be prime number.

iii) Choose a value of X which is not prime. uvkw =uvvv…vvw and |uvv..vvw| is not a prime number. At k=b.

uvvv...vvw= |uw| + b|v|

=b+b(a)

= b(1+a)

which is not a prime and is a composite number. So, FSA accepts one string which is not in the language L, thus the set of strings is not regular. The above set of the string is not a regular language. For example: - Let n=3 hence 1^3 (|uv| < 3 and |w| >0)

Let x=uvw such that |uw|<3 and |v|>0

That is u=1, v=1, w=1.

Now let $x^1=uv^2w$ that is $w^1=1^4$ where 4 is not prime.

So, L is infinite, this string is not a regular set and hence it couldn't be accepted by finite set automata.

b) $\{0^{2n}\,1^{2m} \mid n \text{ and } m \text{ are integers}\}$

Solution: This has two cases

i) When n ≠m. Yes, it could be accepted by finite set automata. It is a regular set. We can build FSA consisting of an even number of zeros (00) * followed by a concatenation of even number of 1's, (11) *.

ii) When n = m. This set is not accepted by finite automata. This is not a regular set that can be provided by pumping lemma.

Let X = 000….0011….11

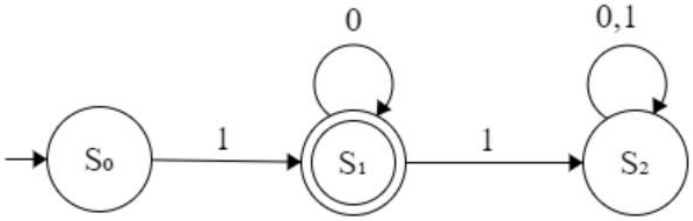
u---v---w. u, w can be empty.

i) If u is empty v starts from left i.e. Subset of zero. Set i=3. The machine accepts but not in the correct form, it

has more zeros than 1's. Thus, w is the rest.

ii) If w is empty, v = at most k 1's. Thus the string is uvv but has more 1's than 0's. Thus u and w emptiness is not possible.
iii) Let u and w are empty at a time. Then the entire string is v. this is a string accepted by machine but not the original string.
iv) If u and v both are not empty. Then the machine accepts the string but not in the correct form. Thus, the language is not acceptable by FSA.
Thus, the language is not acceptable by FSA.

c) {x | x is a binary power of two}
Solution: Yes, it is accepted by FSA. It is a regular set as we can build a finite state machine that recognizes this set. The string can be 1,10,100,1000. So, from the above, we can observe that the language is of form 10* And this language 10* can be accepted by the following finite automata.



Here S_2 is a dead state.

d) {x | the center symbol of x is a 1}
Solution: This is not a regular set. This cannot be accepted by finite automata. This can be proved by pumping lemma using the same example discussed in the class. From pumping lemma u, w could be empty.

Let X = 000….0011….11 let this represent u---v---w. u, w can be empty.

i) If u is empty v starts from left i.e. Subset of zero. Set i=3. The machine accepts but not in the correct form, it has more zeros than 1's. Thus, w is the rest.
ii) If w is empty, v = at most k 1's. Thus, the string is uvv but has more 1's than 0's. Thus, u and w emptiness is not possible.
iii) Let u and w are empty at a time. Then the entire string is v. this is a string accepted by machine but not the original string.
iv) If u and v both are not empty. Then the machine accepts the string but not in the correct form. Thus, the language is not acceptable by FSA.

12) 3. Show that the regular sets are not closed under infinite union by producing an infinite family of regular sets whose union is not regular.

Solution: Let $A_1 \cup A_2 \cup A_3$ ……is interpreting this way is not regular. Each set A has a finite set and contains only one string like $A_1 \cup A_2 \cup A_3$……A_j as {0} ∪ {000} ∪ {0000} ……. Since 0^{n2} regular sets are not closed under infinite union, by using pumping lemma we can tell that it is not a regular set. By the Pumping Lemma, we can prove that the language 0^n1^n is not regular.

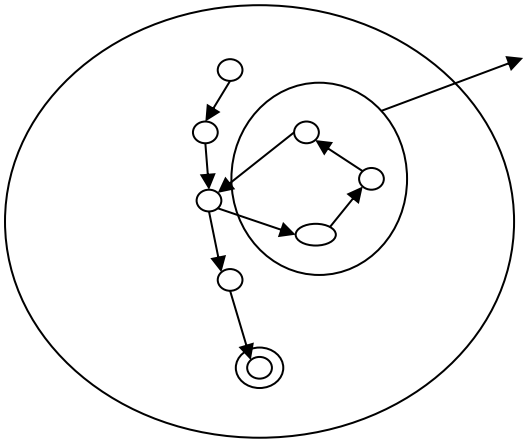
However, if regular sets were closed under infinite union, then we could build FSAs for each $0^1 1^1, 0^2 1^2, 0^3 1^3 \ldots 0^n 1^n$ and then simply union them together to produce an accepting FSA and thus a regular set. Therefore, regular sets are not closed under infinite union since this would violate the pumping lemma.

For example: - consider the sets {0}, {01}, {0011} etc. Each one is regular because it only contains one string. But the infinite union is set $\{0^i 1^i \mid i >= 0\}$ which we know is not regular. So we can say that the infinite union cannot be closed for regular sets.

13) 8. Design an algorithm to determine whether a finite automaton accepts an infinite set. Prove that your algorithm is correct.

Solution: To design an algorithm that detects whether an automaton M accepts infinitely many strings, suppose that L (M) is an infinite language. Then there exists a string w accepted by M such the length of w is greater than the number of states in M. From the proof of the pumping lemma, we see that there exists a run $s_1, s_2, ..s_m, s_{m+1}$ of M on w with the following properties:

1. s_1 is an initial state and s_{m+1} is a final state.
2. There exist i and i+k with k! =0 such that $s_i, ..s_{i+k}$ is a path and $s_i = s_{i+k}$. We call sequence s_i, s_{i+k} a loop and s_i a looping state.



Thus, if L (M) is infinite then the automaton M has the following properties. There exists a path from an initial state to a looping state and from the looping state to an accepting state.

Conversely, assume that in the transition diagram of the NFA M there exists a path from an initial state to a looping state and from the looping state to a final state. Then it is easy to see that L (M) accepts infinitely many strings. Indeed, we take a string u_1 that labels the paths from the initial state to the looping state, say q. Since q is a looping state there must exist a nonempty string v that labels a path from q to itself. Now take a string u_2 that labels a path from the looping state to the final state. Thus, we have that the strings u1vtu2 are all accepted by M, where $t \geq 0$. Therefore L (M) is an infinite language. Below is the algorithm for it.

1. Transform M into an NFA with no transitions.

2. Construct the set x= {q| q is a looping state}. This set can be constructed using the path existence (G, s, t) algorithm.

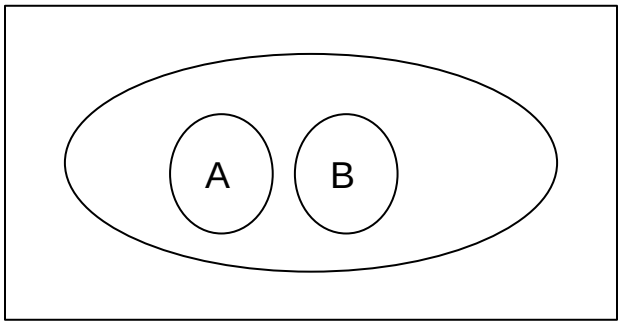
3. Check if there exist paths from an initial state to state q∈x and from q to final state. If there are such paths then output “L (M) is infinite”. Otherwise, output “L (M) is finite”.

Example: - Represent an FSA as a directed graph with labeled edges. Check if there exists a cycle in the graph, this could be done in many ways, one of the ways to check for cycles in a graph is by using a depth-first search (DFS). DFS for a connected directed graph produces a tree. If there is a back edge in this tree, then there is a cycle in the graph. (back edge means an edge from a node to itself or one of its ancestors in the tree). If a cycle was detected then, we have to find a path from the start state (start node) to any node in the detected cycle and we have to find also a path from any node in the detected cycle to the acceptance state. This can be done using any search procedure such as BFS or DFS. If these two paths were found and a cycle was detected, then our FSA accepts infinite set.

14) 9. Exhibit an algorithm that detects whether one finite automaton accepts a subset of the set accepted by another machine. Show that this procedure works.

Solution: Consider two FSA’s FA and FB if one accepts the other is accepted based on decision algorithm. We consider 3 conditions to be satisfied.

1. A is in B
2. A string is outside A but inside B.



The second condition can be of the form B-A=Bn A`. From the theorems, we know that there is an FSA for intersection and complement as we know that they are closed under intersection and complement. Thus, there exists an FSA for B-A. Now we check for the emptiness of this string using the Depth-first search (DFS) technique and check that it is empty. For the first condition, A must not be having anything outside of B. the condition here is A-B= An B`. This is an FSA from the proof above. Thus, the emptiness of the string is checked using the DFS technique and check that it is not empty. On checking the emptiness of both the strings we can conclude the algorithm that finite automata accept a subset of the set accepted by another machine.

Pushdown Automata

15) 3. Prove that acceptance by empty stack is equivalent to accepting by the final state for pushdown automata.

Answer: If L is L(M1) for PDA M and 'L' is N(M2) for some PDA M2

$M1 = (Q, \sum, l1, \delta1, q1, z1, f)$

M1 accepts 'L' by final state

$M2 = (Q2, \sum 2, l2, \delta2, q0, z0, \alpha)$
$Q2 = Q1 \cup \{q2, q1\}$
$\sum 1 = \sum 2$
$I2 = I1 \cup \{Z2\}$

M2 accepts L by empty stack

- $F2(q0, E, Z2) = \{(q, Z1, Z2)\}$
- for all q in Q,
  a in $\sum \cup \{E\}$ & $Z \in I1$
  $f1(q, a, Z) = f1(q1, a, Z)$
- for all $q \in f$ & $Z \in I2$
  $f2(q,E,Z)$ {qe, E}
- for all $Z \in I2$
  $f2(Ne, E, Z) = \{qE, E\}$

CS 6350 Advanced Algorithm

Transferred Course - CSCI 5451: Advanced Algorithms

Assignment 2 (Due 21 September 2017)

- (Closed Unit Intervals Covering Points) CLR problem 16.2-5: Describe an efficient algorithm that, given a set $\{x_1, x_2, \dots, x_n\}$ of n points on the real line, determines the smallest set of closed unit-length intervals that contains all of the given points. Prove that your algorithm is correct. Note that the word "efficient" means that complexity is no worse than polynomial-time. Be sure that your algorithm is written in readable and understandable pseudocode.

  Answer: First we sort the set of n points $\{x_1, x_2, \dots, x_n\}$ to get the set $Y = \{y_1, y_2, \dots, y_n\}$. Such that $y_1 \leq y_2, \dots \leq y_n$.

  Next, we do a linear scan on $y_1, y_2, \dots, y_n$ started from y_1. Every time while encountering y_i, for some $i \in \{1; \dots; n\}$, we put the closed interval $[y_i;\ y_i + 1]$ in our optimal solution set S, and remove all the points in Y covered by$[y_i;\ y_i + 1]$. Repeat the above procedure, finally, output S while Y becomes empty. Next show that S is an optimal solution.

  We claim that there is an optimal solution which contains the unit-length interval $[y_1;\ y_1 + 1]$. Suppose that there exists an optimal solution S$*$ such that y1 is covered by $[x';\ x' + 1] \in S*$ where $x' < 1$. Since y1 is the leftmost element of the given set, there is no other point lying in $[x';\ y1)$.

  Therefore, if we replace $[x';\ x' + 1]\ in\ S*\ by\ [y1;\ y1 + 1]$, we will get another optimal solution. This proves the claim and thus explains the greedy choice property.

  Therefore, by solving the remaining subproblem after removing all the points lying in $[y_1;\ y_1 + 1]$, that is, to find an optimal set of intervals, denoted as S', which cover the points to the right of $y_1 + 1$, we will get an optimal solution to the original problem by taking the union of $[y_1;\ y_1 + 1]$ and S'.

  The running time of our algorithm is $O(n\ log\ n + n) = O(n\ log\ n)$, where $O(n\ log\ n)$ the time for sorting is and $O(n)$ is the time for the linear scan.

- (L-Tiling) Consider C_n, a 2^n by 2^n checkerboard with the upper right square removed. An L-tiling of C_n is a tiling of C_n with L-shaped tiles (composed of three squares) with no overlaps and no square of the checkerboard left uncovered. Prove or disprove: for every $n \in Z^+$, there is an L-tiling of C_n: See Figure 1 for pictures of C_1, C_2, and C_3, and Figure 2 for pictures of all possible L-shaped tiles.

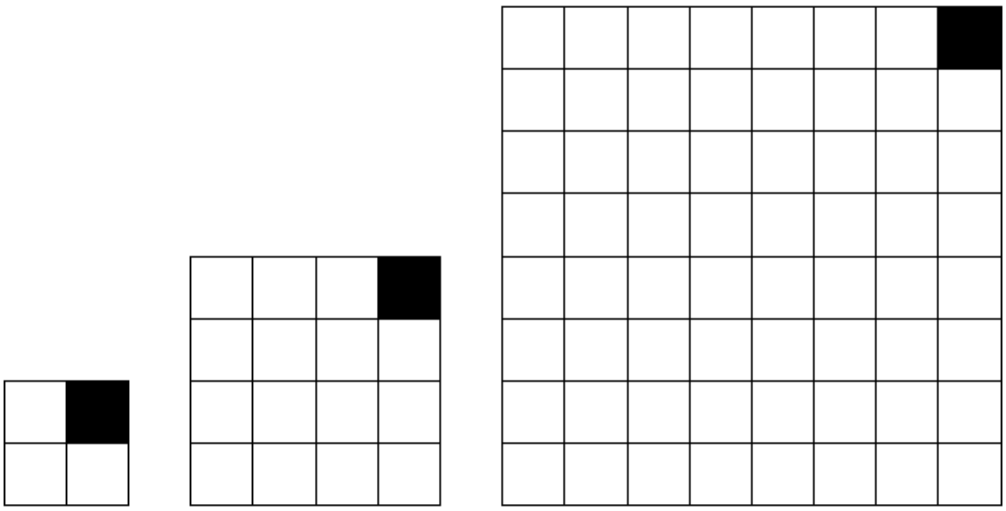


Figure 1: C_1, C_2, and C_3: the black square has been removed from the board.

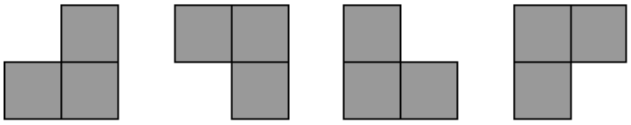


Figure 2: All possible L-tiles composed of three squares.

Answer: We proceed by induction on n. The base case is obvious; the decomposition consists of just one piece. For the induction step, let us assume that the tiling is possible for such a $2^n \times 2^n$ board and consider a $2^{n+1} \times 2^{n+1}$ board.

Start by placing a piece in the middle of the board as shown in Figure 43. The remaining surface decomposes into four $2^n \times 2^n$ boards with corner squares removed, each of which can be tiled by the induction hypothesis.

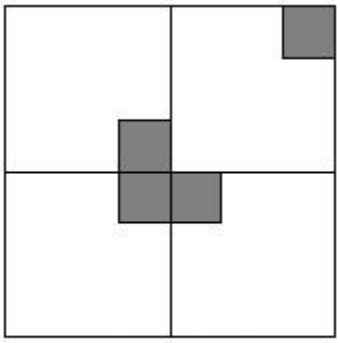
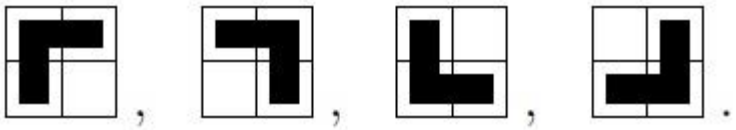


Figure 43

Hence, we are done or else consider the other solution

Basis: For $n = 1$, the four cases are  .

Induction: Let $n = 1$ and assume a $2^n \times 2^n$ chessboard with anyone square removed. It can be tiled using L-shaped pieces. Divide the $2^{n+1} \times 2^{n+1}$ chessboard into four $2^n \times 2^n$ chessboards. Choose any square to omit, and, by induction, tile the $2^n \times 2^n$ chessboard containing that square, omitting the chosen square. By induction, we may tile the remaining three $2^n \times 2\hat{}n$ chessboards so that each tiling omits the square closest to the center of the $2^{n+1} \times 2^{n+1}$ chessboard. These three omitted central squares can then be covered with an L-shaped piece, completing the desired tiling.

Assignment 3 (Due 5 October 2017)

- (Compression) Prove that there is no single compression* scheme that compresses all n-bit files.
  *A compression scheme must permit the reconstruction of the original n-bit file from the compressed version, and for a compression scheme to compress a file, the compressed version must be smaller than the original.

  Answer:
  Base Case: Prove that there is no single compression scheme that compresses all $n - bit$ files with the use of the pigeon-hole principle. First start with a function,P and a file that is length of$n$, such that the file length can be set to pn. There are 2^n files of length n, and after using program P you get $2\text{^}pn$ files of length pn.

  The important part here is to make sure the original file must map to the compressed file; which does NOT make this a one-to-one function which means no inverse can exist which means that uncompressing will NOT result in the same original file it once was.

  Claim: Using the pigeonhole principle observes that if you have n pigeons but less than n holes then at least two pigeons will share a hole. In this case, we need to show that a possible file holds 2 variations of a file.

  Reality Check: Here is an example of case n = 3. There are 8 possible files of bit length 3, because of $2\text{^}3 = 8$

  000
  001
  010
  011
  100
  101
  110
  111

  Apply program P to these 8 FILES which will now give you 66% compression, since this will give you all 8 files to have bit length 2. Here is how: $n = 3$ compress by taking $n - 1$, n now equals $n = 2$ so now we have $2\text{^}2 = 4$ files:
  00
  01
  10
  11
  We have $\frac{2}{3}$ of the bits here and took $\frac{1}{3}$ bits off so that leaves $\frac{2}{3} = 66\%$ total compression
  If you try to map the 3-bit set of files to the 2-bit set of files at least one 2-bit file will represent more than one 3-bit file by pigeonhole principle!! For example, 00 could hold 000 and 001, which if we add all bit fields together that is 8 bits way more than what we had, which is not possible. If you continue

to compress the files from here on out, you will eventually compress the file into nothing! WHICH IS IMPOSSIBLE!

Induction Hypothesis: Define B as bits such that it is the lowest limit a file can compress, such that no program can be able to compress a file below this certain limit B. Start with a small file size such as 1KB or 1024 bits to be your B. Prove that you cannot compress your file any farther.

Induction Step: Compress all files of length B + 1, there are 2^B+1 of these files, but only 2^B files can map them onto. Going back to the pigeonhole argument, at least 2 of those original files, which started differently, will have the same compressed form and will then be uncompressible!

Conclusion: We proved that the pigeon-hole argument holds in the induction step, which also proves that you cannot decompress a lower limit of bits for a file because the original file mapping would be impossible.

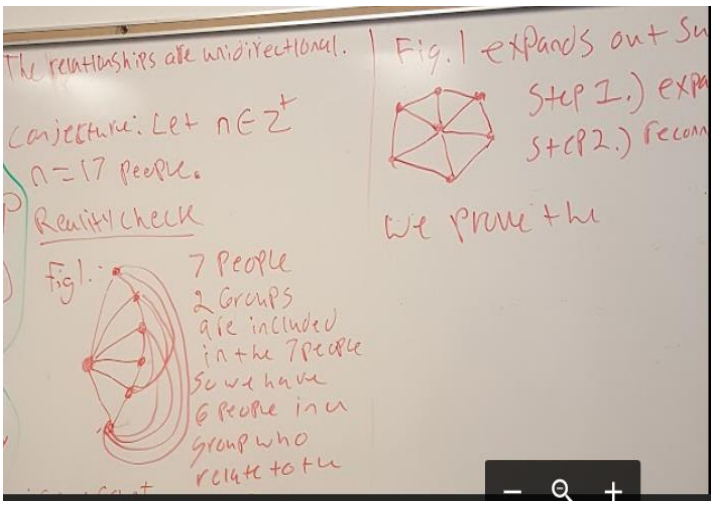
- (Party On) Suppose that in the world every pair of people either
    a) likes one another,
    b) dislikes one another, or
    c) is indifferent toward one another.

  Prove that in any gathering of 17 people, there is a group of three people all of whom satisfy one of the conditions (a), (b) or (c).

Answer:

Base Case: Use the pigeonhole principle to prove that at least one person holds a relationship with a group of 3 people and that 2 people of the group share the same relationship with the one person. Remember relationships are unidirectional and make the conjecture that n where $\mathbf{n} \in \mathbf{Z}$ is the number of total people at a party. We will start with n = 17.

Reality Check: Create 7 people, such that there is 1 person who does not belong to any of the 2 groups = 6 people (remember 1 group = 3 people). Create a connection as shown below such that all 6 people can relate to this 1 person.



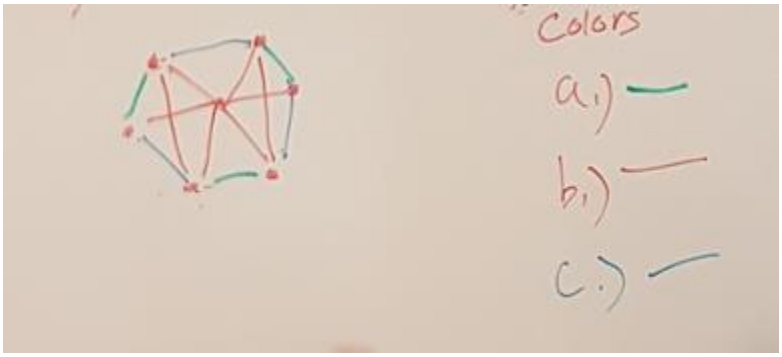
When connecting the edges so that the total number of edges is unique, we will get 21 relationships between pairs of people (edges)! Again, take the graph you have and think of it as a compressed version that can unfold out to a tetrahedron. This shows that at least one person holds a relationship with a group of 3 people and that at least 2 people of the group share the same relationship with one person.

Induction Hypothesis: First let us use the pigeonhole argument to help prove that there are always at least 3 people who share a relationship with at least one of the three properties. Prove this by use of chromatic numbering, such that every edge is unique and that there is no **monochromatic** face (triangle). We also want to prove that a bound exists such that the pigeonhole holds true. Conjecture: Let **B** be a bound that will hold the pigeonhole true such that **B** = 6 <= **B** <= 16.

Induction Step: Go back to the figure Check and now color the edges such that the edges are unique and that no adjacent edge shares the same color on the **outside parts of the graph**. Again, give the 3 properties a color:

- **A. = GREEN**
- **B. = RED**
- **C. = BLUE**

Again, the graph we are looking at is a subgraph of the given graph of 17 people



As you can see we were not able to make unique triangles such that no **monochromatic triangle** could exist. So this means that the pigeon-hole argument holds and we prove by way of contradiction (BWOC).

Conclusion: By way of contradiction we were able to disprove our claim in the hypothesis but kept our pigeon hole argument strong, such that the graph holds a bound of how many people can hold a relationship with one person outside of the group such that it will create a monochromatic triangle relationship.

Assignment 4 (Due 26 October 2017)

- (Coloring Earth/Moon Graphs) Let G be a thickness-two graph. Design an algorithm that properly colors the vertices of G with no more than 12 colors. You need not prove that your algorithm is correct, but an incorrect algorithm will receive either partial or no credit. Be sure to write your algorithm in readable and understandable pseudocode.

  Answer: Here is the algorithm.

  Input: $1 \leq v \leq 12$,
  Output: G with proper color
  Start
  Step 1: Initialize $\theta(G) = 2$
  Step 2: Check the outer vertices of the graph
  Step 3: List them from $V_i \; to \; V_n$ confirm that vertices is odd

  $List = null;$
  $for(\, int \, i = 0 \,;\, i < n \,;\, i++)$
  $\quad if(\, v[\, i \,] \,\%\, 2 \,! = 0)$
  $\quad\quad List = List + v[\, i \,];$

  Step 4: Now check for two properties

  $$\chi(G) \leq 2\,[\,Edges]$$
  $$\chi(G) \geq 2\,[\,proper\ coloring]$$

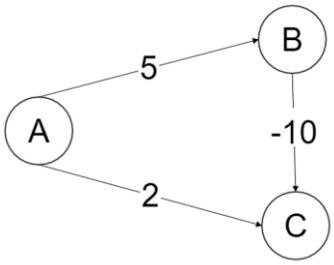
  Step 5: Now we use the GCA algorithm (HW3Q4)

  $For \; i = 1 \; to \; n \; do$
  $\quad Li = \{1 \ldots\ldots i\}$
  $For \; i = 1 \; to \; n \; do$
  $\quad Set \; Ci = the \; first \; color \; is \; Li$
  $\quad For \; j = 1 \; to \; n \; do$
  $\quad\quad if(i < j) \; and \; (Vi, Vj) \epsilon E(G) \; do$
  $\quad\quad\quad Set \; Li =: i \backslash Ci$

  Step 6: Now the graph is properly colored.
  Step 7: Go inside the graph and break it into subgraphs.
  Step 8: Count the number of vertices the highest of the $subgraph \leq 12$ & check two properties (i) and (ii) above.
  End

- (Dijkstra Thought)
  a) Give an example of a graph with negative edge weights for which Dijkstra's algorithm fails. Show a trace of the algorithm on your graph.
  b) (b) Analyze the worst-case runtime of Algorithm Dijkstra.

  Answer: Consider the graph shown below with the source as Vertex A. First try running Dijkstra's algorithm yourself on it.



When I refer to Dijkstra's algorithm in my explanation I will be talking about Dijkstra's Algorithm as implemented below.

```
1  function Dijkstra(Graph, source):
2
3      create vertex set Q
4
5      for each vertex v in Graph:             // Initialization
6          dist[v] ← INFINITY                  // Unknown distance from source to v
7          prev[v] ← UNDEFINED                 // Previous node in optimal path from source
8          add v to Q                          // All nodes initially in Q (unvisited nodes)
9
10     dist[source] ← 0                        // Distance from source to source
11
12     while Q is not empty:
13         u ← vertex in Q with min dist[u]    // Source node will be selected first
14         remove u from Q
15
16         for each neighbor v of u:           // where v is still in Q.
17             alt ← dist[u] + length(u, v)
18             if alt < dist[v]:               // A shorter path to v has been found
19                 dist[v] ← alt
20                 prev[v] ← u
21
22     return dist[], prev[]
```

So starting the values (the distance from the source to the vertex) initially assigned to each vertex are,

$$A = 0; \quad B = \infty; \quad C = \infty$$

We first extract the vertex in $\boldsymbol{Q} = [\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]$ which has the smallest value, i.e. A, after which $\boldsymbol{Q} = [\boldsymbol{B}, \boldsymbol{C}]$. Note A has a directed edge to B and C, also both of them are in Q, therefore we update both of those values,

$$A = 0; \quad B = min(-\infty, 5) = 5; \quad C = min(-\infty, 2) = 2$$

Now we extract C as (2<5), now Q = [B]. Note that C is connected to nothing, so the line16 loop doesn't run.

$$A = 0; \quad B = 5; \quad C = 2$$

Finally, we extract B, after which $Q = \phi$. Note B has a directed edge to C but C isn't present in Q, therefore, we again don't enter the for loop in line16,

$$A = 0; \quad B = 5; \quad C = 2$$

So, we end up with the distances as

$$A = 0; \quad B = 5; \quad C = 2$$

Note how this is wrong as the shortest distance from $A\ to\ C\ is\ 5 + -10 = -5$ when you go $A \to B \to C$.

So, for this graph, Dijkstra's Algorithm wrongly computes the distance from A to C.

This happens because Dijkstra's Algorithm does not try to find a shorter path to vertices which are already extracted from Q. What the line 16 loop is doing is taking the vertex u and saying "hey looks like we can go to $\boldsymbol{v}$ from source via $\boldsymbol{u}$, is that (alt or alternative) distance any better than the current $\boldsymbol{dist}[\boldsymbol{v}]$ we got? If so let's update $\boldsymbol{dist}[\boldsymbol{v}]$"

Note that in line 16 they check all neighbors v (i.e. a directed edge exists from u to v), of u which are still in Q. In line 14 they remove visited notes from Q. So, if x is a visited neighbor of u, the path $source \to x \to u$ is not even considered as a possible shorter way from source to v. This is useful if the edge weights are all positive numbers, because then we wouldn't waste our time considering paths that can't be shorter. So, I say that when running this algorithm if x is extracted from Q before y, then it's not possible to find a path - $source \to \cdots \to y \to x$ which is shorter. Let me explain this with an example,

As y has just been extracted and x had been extracted before itself, then $\boldsymbol{dist}[\boldsymbol{y}] > \boldsymbol{dist}[\boldsymbol{x}]$ because otherwise y would have been extracted before x (line 13 min distance first). And as we already assumed that the edge weights are positive, i.e. $\boldsymbol{length}(\boldsymbol{x}, \boldsymbol{y}) > \boldsymbol{0}$. So the alternative distance (alt) via y is always sure to be greater, i.e. $\boldsymbol{dist}[\boldsymbol{y}] + \boldsymbol{length}(\boldsymbol{x}, \boldsymbol{y}) > \boldsymbol{dist}[\boldsymbol{x}]$. So the value of $\boldsymbol{dist}[\boldsymbol{x}]$ would not have been updated even if y was considered as a path to x, thus we conclude that it makes sense to only consider neighbors of y which are still in Q (note the comment in line 16). But this thing hinges on our assumption of positive edge length if $\boldsymbol{length}(\boldsymbol{u}, \boldsymbol{v}) < \boldsymbol{0}$ then depending on how negative that edge is we might replace the $\boldsymbol{dist}[\boldsymbol{x}]$ after the comparison in line 18.

So any $\boldsymbol{dist}[\boldsymbol{x}]$ calculation we make will be incorrect if x is removed before all vertices v - such that x is a neighbor of v with a negative edge connecting them - is removed. Because each of those v vertices is the second last vertex on a potential "better" path from source to x, which is discarded by Dijkstra's algorithm. So in the example, I gave above, the mistake was because C was removed before B was removed. While that C was a neighbor of B with a negative edge!

Just to clarify, B and C are A's neighbors. B has a single neighbor C and C has no neighbors. $length(a, b)$ is the edge length between the vertices a and b.

Worst-case running time:

Every time the main loop executes, one vertex is extracted from the queue. Assuming that there are V vertices in the graph, the queue may contain $O(V)$ vertices. Each pop operation takes $O(log\ V)$ time assuming the heap implementation of priority queues. So, the total time required to execute the

main loop itself is $O(V\ log\ V)$. Also, we must consider the time spent in the function expand, which applies the function $handle\_edge$ to each outgoing edge. Because expansion is only called once per vertex, $handle\_edge$ is only called once per edge. It might call $push(v')$, but there can be at most V such calls during the entire execution, so the total cost of that case arm is at most $O(V\ log\ V)$. The other case arm may be called $O(E)$ times, however, and each call to $increase\_priority$ takes $O(log\ V)$ time with the heap implementation. Therefore the total run time is $O(V\ log\ V\ +\ E\ log\ V)$, which is $O(E\ log\ V)$ because V is $O(E)$ assuming a connected graph.

(There is another more complicated priority-queue implementation called a Fibonacci heap that implements $increase\_priority$ in $O(1)$ time, so that the asymptotic complexity of Dijkstra's algorithm becomes $O(V\ log\ V\ +\ E)$; however, large constant factors make Fibonacci heaps impractical for most uses.)

- (Kempe Chain Flaw) Suppose that the planar graph G in Figure 1 has been run through a four coloring algorithm using the Kempe chain argument given in class (this was Theorem 3, the awed proof of Kempe that "showed" that all planar graphs can be four colored). All vertices of G except for v have been magically colored. We are faced with coloring the final vertex v. Try to use the Kempe chain switches to color v and explain in detail what goes wrong.

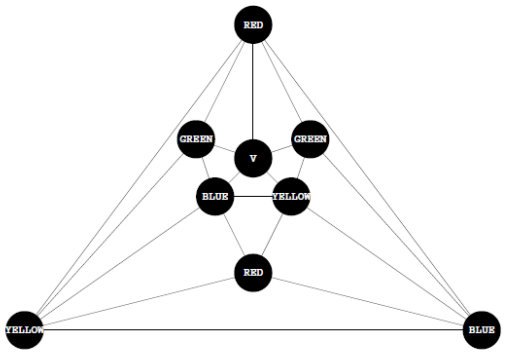
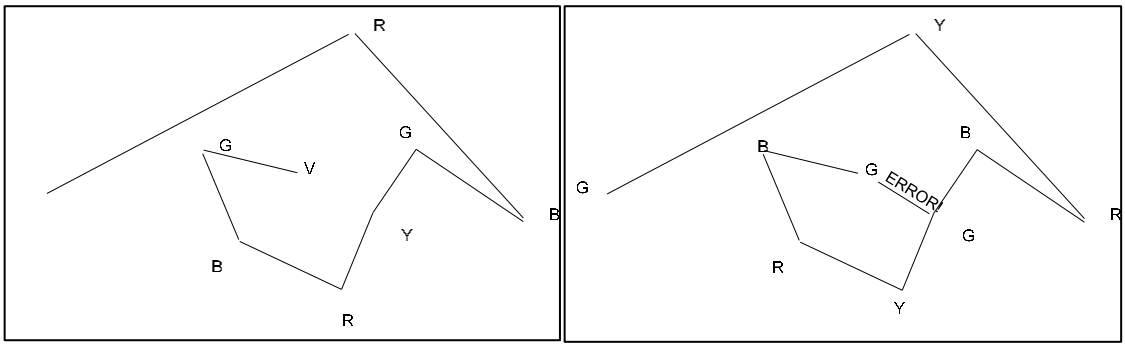


Figure 1: **Color v by using Kempe chain switches??**

Answer:
Step 1: First step to drawing the chain so it does not tangle!
Step 2: shift the colors clockwise so that V becomes G and G becomes B and so on so forth. We get an error as V or what is now G is connected to a vertex that is G as well so colors collide which fails the algorithm!



Assignment 5 (Due 9 November 2017)

- Suppose p and q are distinct odd primes and assume with have an implementation of RSA with the public key$(n;\ e)$, where $n\ =\ pq$. Let the encryption function be denoted by$E$. A block b of the RSA message is said to be fixed by E if $E(b)\ =\ b$. How many blocks are fixed by RSA when $p\ =\ 3, q\ >\ 3$, and $e\ =\ 3$? Hint: learn and use the Chinese Remainder Theorem. Show all your work.

  Answer: $gcd(e,d)\ =\ 1$ means that the numbers e and d have no common factors except for 1.
  The use of modular inverse may help us benefit here when finding the blocks with the Chinese remainder theorem.

  - Select primes $p = 11, q = 3$.
  - Calculate $n\ =\ pq\ =\ 11.3\ =\ 33$
  - Calculate $phi\ =\ (p-1)(q-1) =\ 10.2\ =\ 20$
  - Choose $e = 3$
  - Check $gcd(e, p-1)\ =\ gcd(3,10)\ =\ 1$ (i.e. 3 and 10 have no common factors except 1),
  - Check $\gcd(e, q-1) = \gcd(3,2) =\ 1$
  - Therefore $\gcd(e, phi) = \gcd\big(e, (p-1)(q-1)\big) = \gcd(3,20) =\ 1$
  - Find d from $d = \frac{phi\ n * k\ +1}{e}$
  - Check $ed - 1\ =\ 3.7\ -\ 1\ =\ 20$, which is divisible by phi.
  - $Public\ key\ =\ (n,e) =\ (33,3)$
  - $Private\ key\ =\ (n,d) =\ (33,7)$

- (RSA) The message

  19 14 3

  Was encrypted using the RSA cryptosystem with public key n = 118 and e = 39. Decrypt the message and give the corresponding plaintext message (using A = 10, B = 11, …, Z = 35, and BLANK= 99). Show all your work.

  Answer: From the question, we know that $n\ =118,\ e\ =\ 39$. To find the value of p and q we need to factorize the value of $n\ =118\ i.e.(\ 2\ *\ 59)$

  Then $p\ =\ 2\ and\ q\ =\ 59$

  $$\phi(n)\ =\ (p-1) * (q-1)$$
  $$\phi(n)\ =\ (2-1) * (59-1)$$
  $$\phi(n)\ =\ (1) * (58)$$
  $$\phi(n)\ =\ 58$$

  Now we must find out the value of d by using the equation $d\ =\ (k\ *\ \phi(n) + 1)\ /e$

  $d\ =\ (\ 2\ *\ 58\ +\ 1\ )/39$ (Where k is an integer which is greater $ZERO\ (k > 0)$)
  $d\ =\ (117)\ /39$
  $d\ =\ 3$

  The original message is encrypted using the formulae $c\ =\ M\text{^}e\ mod\ n$. We can now decrypt the Message 19 14 3 using the formulae $M\ =\ c\text{^}d\ mod\ n$.

  19:

$M = 19^3 mod\ 118$
$M = 15 = F$ ($Form\ the\ question\ we\ know\ that\ the\ value\ of\ 15 = F$)

14
$M = 14^3 mod\ 118$
$M = 30 = U$ ($Form\ the\ question\ we\ know\ that\ the\ value\ of\ 30 = U$)

3
$M = 3^3 mod\ 118$
$M = 27 = R$ ($Form\ the\ question\ we\ know\ that\ the\ value\ of\ 27 = R$)

The message after decrypting ID $F\ U\ R$. Now if we can encrypt the message using the formulae $c = M^e mod\ n$.

15:
$C = 15^{39} mod\ 118$
$C = 19$

30:
$C = 30^{39} mod\ 118$
$C = 14$

27:
$C = 27^{39} mod\ 118$
$C = 3$

- (Number Theory) Use only the following tools: pen, paper, and your friends.
  a) Prove that if $2^n - 1$ is prime then n must be prime. Show all of your work.
  b) Find the last digit of 7^{355}. Show all of your work.

  Answer:
  a) We show the equivalent statement that if n is composite, then $2\hat{}n - 1$ is also composite. Suppose n is composite. Then $n = ab$ for some integers $a, b \geq 2$. Since $2\hat{}a \equiv 1\ mod\ (2\hat{}a - 1)$, we have $2\ \hat{}n = (2\hat{}a)\hat{}\ b \equiv 1\hat{}\ b = 1\ mod\ (2\hat{}a - 1)$. Thus, $2\hat{}n - 1$ is divisible by $2\hat{}a - 1$, and since $1 < a < n$, the integer $2\hat{}a - 1$ is a proper divisor of $2\hat{}n - 1$ (i.e., strictly greater than 1 and less than n). Hence $2\hat{}n - 1$ is composite.
  b) Notice, the pattern of the last digits. They are 7, 9, 3, 1, 7, 9, 3, and 1 ... and so on. The last digit repeats in a pattern that is 4 digits long: 7, 9, 3, 1.

Digit	d^2	d^3	d^4	d^5
7	9	3	1	7

  Note that $355\ divided\ by\ 4\ is\ 88$ with a remainder of 3 so the pattern will repeat 88 times, with three extra entries at the end. These last three entries are $7, 9\ and\ 3$ so the last digit of $7^{355}\ is\ 3$.

- (Quantum Factorization) Let $N = 899$ and suppose $y = 4$ (y was chosen at random, but that does not affect how you proceed with this problem.) Use the technique given in class to factor N. Show all your work.
  Answer: Here the steps to facto N.

Step 1: Find $N\ and\ y.\ N\ =\ 899\ and\ y\ =\ 4.$

Step 2: Find the T using the equation $y^T\ mod\ N$ such that $y^T\ mod\ 899\ =1$

$$4^1 mod\ 899\ \neq 1$$
$$4^2 mod\ 899\ \neq 1$$
$$4^3 mod\ 899\ \neq 1$$
$$.$$
$$.$$
$$.$$
$$4^{70} mod\ 899\ =1$$

T=70

Step 3: Since $T\ =\ 70$ and it is an even then

$$4^{70}\ =\ \left(\frac{y^T}{2}\right)^2\ =\ x\ =\ y^T$$
$$y^T\ =\ y^{35}\ =\ x$$
$$X-1\ =\ 4^{35}\ -\ 1$$
$$X+1\ =\ 4^{35}\ +\ 1$$
$$GCD(N, X\ -\ 1) > 1$$
$$GCD(N, X\ +\ 1) > 1$$

Step 4:

$$p\ =\ GCD\left(\frac{4^T}{2}-\ 1, N\right), q\ =\ GCD\left(\frac{4^T}{2}+\ 1, N\right)$$
$$p\ =\ GCD\left(\frac{4^{70}}{2}-\ 1, 899\right), q\ =\ GCD\left(\frac{4^{70}}{2}+\ 1, 899\right)$$
$$p\ =\ GCD(4^{35}-\ 1, 899), q\ =\ GCD(4^{35}+\ 1, 899)$$
$$p\ =\ 31, q\ =\ 29$$

Therefore, the Factors of $N\ are\ 31\ and\ 29.$

- (Secret Sharing) Suppose Uzi, Kyle, Keenan, and Todd are in a room, and it is known that one of them is a spy. They are participating in a $(2, \omega)$-secret sharing scheme such that Uzi is given $(1,4)$, Kyle is given $(3,7)$, Keenan is $(5,1)$, and Todd is given $(7,2)$. Who is the spy and what is the message?

Answer: In the $(2, \omega)$-Shamir scheme being used, the secret is the constant term of a polynomial over Z_{11} of degree 1, that is, a linear polynomial. Since the graph of this polynomial is a line, we can use geometric ideas to answer the question instead of the interpolation polynomial approach. The shares of the scheme are the coordinates of points in a plane, and the three "good" ones are points on the same line. To determine which three points are on a line, we use the determinant area formula for triangles. That is, we calculate the area of the triangle determined by say, $A, B\ and\ C$. If this is zero, then the points are on a line, otherwise, they are not. For A, B, C we get

$$\det \begin{matrix} 1 & 4 & 1 \\ 3 & 7 & 1 \\ 5 & 1 & 1 \end{matrix} = 7 + 20 + 3 - 35 - 12 - 1 = -18 = 4 \bmod 11$$

Which says, since it is not zero, that the foreign agent is one of A, B or C. doing the same for $A, B\ and\ D$ gives

$$\det\begin{matrix}1 & 4 & 1\\ 3 & 7 & 1\\ 7 & 2 & 1\end{matrix} = 7 + 28 + 6 - 49 - 12 - 2 = -22 = 0 \bmod 11$$

And so, $A, B\ and\ D$ are on the same line, thus C is the foreign agent. To determine the secret, we calculate the equation of the line determined by any two of $A, B\ or\ D$ and read off the constant term. Thus, for the linear equation $y = xm + k$, using A and B we have:

$$\begin{array}{r} 4 = 1m + k \\ 7 = 3m + k \\ \hline 12 = 3m + 3k \\ 7 = 3m + k \\ \hline 5 = 2k \quad 30 = 12k \quad 8 = k\ (mod\ 11). \end{array}$$

And so, the secret is 8.

CS 6392 Graduate Research Methods

Transferred Course - ISMG 7214 Mixed Methods Research

Week 2 Summary Report, 27 August 2018

Engaging Customers in Coproduction Processes: How Value-Enhancing and Intensity-Reducing Communication Strategies Mitigate the Negative Effects of Coproduction Intensity

Till Haumann, Pascal Gunturkun, Laura Marie Schons, & Jan Wieseke

Prepared By: Sayed Mohsin Reza (Student ID # 108570451), Ph.D. Student, Computer Science & Engineering, University of Colorado Denver, USA, sayedmohsin.reza@ucdenver.edu.

Summary: The research article's objective is to show the potentially negative effects of coproduction intensity and new insights into how firms can mitigate the effects by using strategies to influence customers' perceptions of coproduction processes. The term "co-production" has a role in customers' active participation in the focal organization and independent of direct service employee involvement. Previous researches on coproduction processes have mostly investigated differences between situations of co-production, in which consumers actively participate in the production of goods and services, and situations of firm production, in which consumers have no part in the production process. The paper contributed to marketing research and management differently by investigating the consequences of coproduction intensity in real coproduction processes. They showed how marketers can effectively manage the negative effects of coproduction intensity. One of the portions of their research on relational customer goals by investigating how firms can enhance relational customer value in utilitarian consumption settings. In later, they connected the research on coproduction and inoculation mechanisms in marketing. In summary, they built a theoretical framework on equity theory to derive a conceptual model that investigates the role of perceived coproduction intensity in affecting customers' satisfaction with the co-production process.

Review: The article investigated the relationship between coproduction intensity and customers' satisfaction with the co-production process and examined how value-enhancing and intensity-reducing communication strategies affect this relationship. But they did not generalize their findings to other coproduction contexts. They did not mention any procedure how other coproduction settings will work on a different scenario. They overwrite the issue by mentioning "make a replica on other research". The paper missed referring to adequate communication strategies that companies may employ to shape customers' perceptions of coproduction processes.

Cited papers progression: One of the papers by Li-Wei Wu is about the bright side and dark side of co-production: a dyadic analysis [1]. The paper explores the links between co-production, emotional labor, employee satisfaction, value co-creation, co-production intensity, and their effects on customer satisfaction. The result shows that co-production influences deep acting, surface acting, value co-creation, and co-production intensity which have different effects on employee satisfaction and customer satisfaction. Another paper is extending this paper's research on several aspects of customer participation (CP) and how this proposed typology improves the conceptual and empirical clarity of CP research [2].

Another paper explores CE with brands for consumers exhibiting differing cultural traits and develop a set of research propositions for these individuals' cognitive, emotional, behavioral, and social CE in brand interactions [3]. They talked more about individual cultural values over coproduction. All these papers try to extend the research ideas om coproduction processes.

References

[1] Wu, L.W., 2017. The bright side and the dark side of co-production: a dyadic analysis. Management Decision, 55(3), pp.614-631.

[2] Dong, B. and Sivakumar, K., 2017. Customer participation in services: Domain, scope, and boundaries. Journal of the Academy of Marketing Science, 45(6), pp.944-965.

[3] Hollebeek, L.D., 2018. Individual-level cultural consumer engagement styles: Conceptualization, propositions, and implications. International Marketing Review, 35(1), pp.42-71.

Week 4 Summary Report, 09 September 2018

Theory Construction as Disciplined Imagination

KARL E. WEICK

Prepared By: Sayed Mohsin Reza (Student ID # 108570451), Ph.D. Student, Computer Science & Engineering, University of Colorado Denver, USA, sayedmohsin.reza@ucdenver.edu.

Summary: The paper discussed theory construction in organizational studies and how the quality of the theory affects accuracy and problem statement. Before emphasizing theory construction, the author tried to show how existing pieces of literature described theorizing activities mentioning mapping, conceptual development and speculative thought during. In his paper, the author contributed to imagination, representation, and selection to make the theorizing process to be more understandable. The paper described this theorizing process model as the components of problem statements, thought trials and selection criteria. The paper was provided with an example of marine navigation using radar to illustrate the parallel relation between theory building and the evolutionary process. In this theorizing model, the theoretical problem should be consisting of descriptions that can vary in the finiteness of detail, accuracy, and explicitness of assumptions. It means we should pay attention to the problems given by theorists as well as practitioners. Then this paper introduced thought trials as a component of the theorizing process. A theorizing process was characterized by a greater number of diverse conjectures(ranges) that produces better theory than a process characterized by a smaller number of homogeneous conjectures(classification). At last, they discussed the self-conscious selection process of theorizing. When theorists apply selection criteria to their conjectures, they ask whether the conjecture is interesting, obvious, connected, believable, beautiful, or real, in the context of the problem they are trying to solve. This paper result showed that this model will improve the quality of theoretical thinking but to do better theorists have to think better.

Review: The paper is very interesting in terms of how we can make organization studies easy by theorizing the organizational process. The paper provided a model about the theorizing process but didn't analyze the process in terms of strengths, the capacity to generate theory that is accurate, parsimonious, general, and useful. But this paper can be used to develop theoretical forms using different strategies.

Cited papers progression: One of the cited papers on theorizing from process data discussed and compared several alternative generic strategies for the analysis of process data, looking at the consequences of these strategies for emerging theories [1]. Another paper on following this research showed how bad management theories can destroy good management practices [2]. Another paper on Human Resource Management was follow up research of this article. Mainly they discuss the implications of a stronger theoretical approach to SHRM research and practice over the theorizing process [3].

References

[1] Langley, A., 1999. Strategies for theorizing from process data. Academy of Management review, 24(4), pp.691-710.

[2] Ghoshal, S., 2005. Bad management theories are destroying good management practices. *Academy of Management learning & education*, *4*(1), pp.75-91.

[3] Wright, P.M. and McMahan, G.C., 1992. Theoretical perspectives for strategic human resource management. *Journal of management*, *18*(2), pp.295-320.

Week 5 Summary Report, 17 September 2018

Plant Information Systems, Manufacturing Capabilities, and Plant Performance

Banker, R.D., Bardhan, I.R., Chang, H. and Lin, S.,

Focused on: Overall research design: Data source/s, How or what way the data is collected, How/what variables coded, what regressions/empirical method is followed, and why.

Summary: This paper is based on the theory of dynamic capabilities of how firms create value by creating relevant processes and structures to leverage their IT assets. The business value from those assets is linked to the complementary revolution in organizational process capabilities. To find the relation, they surveyed to collect information about manufacturing trends and performance metrics. They collected those datasets using a mail-in survey and smaller telephone conversion survey with manufacturing executives and plant managers Out of 1738 response, they selected 1077 plants that provided complete responses to the variables of interest. In their research, mainly they tried to measure manufacturing capabilities on a three-point scale depending on the implementation: currently, partially or extensively implemented and measure plant performance during the period of 1994 -1999. In the analysis, they need to find the consistency between proposed factors and data that were first checked by exploratory factor analyses (EFA). To validate the analysis, construct validity was established by measuring the convergent and discriminant validity of the survey items. The results are organized in three panels, representing the regressions for plant quality, time to market, and efficiency. Mainly the regression estimation of the structural equation was used to create the model. The fundamental contribution of our paper is to extend prior research by developing and empirically validating a theoretical framework that integrates the relationships between IT resources, manufacturing capabilities and plant performance using manufacturing plant data.

Identifying Words:

Validity, reliability: Mainly they used this variable to test the model constructs of their measurement.

Generalizability: They focused on plants that employ a minimum of only 100 employees limit the generalizability of their results to industries with larger manufacturing plants. They did not account for the country- or culture-specific differences in manufacturing characteristics since the scope of the survey were limited to U.S. plants.

Relevance: In turn, CSP capabilities mitigate the effects of demand uncertainty by providing stakeholders with greater visibility into process workflows. Such collaboration fosters customer relevance and improves plant performance by reducing stock-outs and improving inventory turnover.

Critical: Understanding the role of IT in enabling the development of manufacturing capabilities is critical in evaluating and rationalizing IT investments.

Week 6 Summary Report, 24 September 2018

Informating the Clan: Controlling Physicians' Costs and Outcomes

Rajib Kohli, William J. Kettinger

Method: This research moves beyond directly controlling informated workers through the legitimized managerial authority to a better understanding of how to informate autonomous professionals. This method underlies in *action research*[*] by which they introduced how hospital management attempted to employ a decision support system to directly informate the hospital's physicians. Note that it is not a *dialogical action research*[**] method.

Execution Process: The research interventions follow an action research cycle (Figure 1). They use this cycle for the two interventions.

Interventions 1: Informating, Transparency, and Agency Theory (*period 1991-1995*). In this stage, costs were diagnosed and introduce a decision support system to physicians to make the behaviors and outcome to be more transparent to hospital management.

Intervention 2: Legitimize information within clan-based concretive control (*period 1996-2001*). In this stage, researchers diagnosed information into the context of a clan to indirectly undertook actions into the clans' conservative control process.

How it is different than other research methods

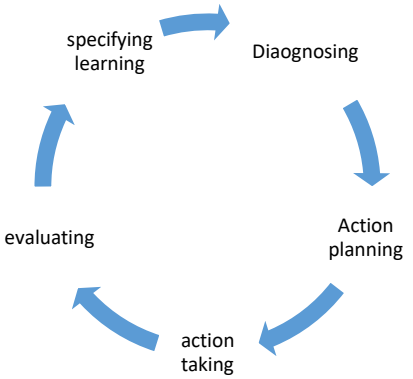


Figure 1: Action research cycle

For IS professionals, this research demonstrated that the clan information for the control change path provides an important role. To be extended, researchers can involve and try to improve the situation. The cons of this research method for any administrators is that implementing a performance monitoring system is direct control over cost and time and doesn't involve in one-to-one dialogues with the practitioner. Other methods like the epistemological process, case studies might need to one-to-one dialogues to improve the system. In case studies, we work on individual cases (scientific, philosophical or real-life problems) to make decisions considering every case.

When to follow this method research and why?

This method improves implementing a performance monitoring system to make the best option management for exercising indirect influence. If any research is about the evaluation of any action process that might cover improvement in the process directly involving researchers herself, we can use such ARC process.

Week 7 Summary Report, 1 October 2018

Survey Instruments in Information Systems

Peter R. Newsted, Sid L. Huff, Malcolm C. Munro

Prepared By: Sayed Mohsin Reza (Student ID # 108570451), Ph.D. Student, Computer Science & Engineering, University of Colorado Denver, USA, sayedmohsin.reza@ucdenver.edu .

Summary: This paper includes actual IS survey instruments and introductory information about how to start with the survey methodology. Moreover, the paper provides information about the usefulness of surveys like

- Allow researcher to give score and code
- Allow external user responses for user generalization
- Allow researcher to predict behavior
- Allow researchers to propose theories.
- Allow researcher to quantify any qualitative research

Development of Survey Instruments

They started with the development of ***Calgary Surveys Query System (CSQS)*** that can be used as a database of survey information. During their research, they tried to distribute the system with a small fee via mailed diskettes if anyone interested in IS survey research.

Later, they started working on creating a *central repository* for a wide variety of information about surveys and their instruments. To create information out of surveys, they incorporated slicing and dicing technologies in the system to predict behavior across surveys.

The "***Survey in Information Systems***" repository is available but is not available for research purposes. The surveys are still in the repository and available at http://www.misq.org/ as per a researcher website stated.

Week 8 Summary Report, 8 October 2018

The Assimilation of Software Process Innovations: An Organizational Learning Perspective

Robert G. Fichman and Chris F. Kemerer

Prepared By: Sayed Mohsin Reza (Student ID # 108570451), Ph.D. Student, Computer Science & Engineering, University of Colorado Denver, USA, sayedmohsin.reza@ucdenver.edu .

Summary: This paper is about an empirical study using data on the assimilation of the software process. The idea behind the paper is to show the importance of the hypothesized factors in organizational learning surrounding software process innovations.

Organization Learning: A researcher, Atwell draws a dominant dissimilarity between the communication of "signaling" information about the existence and potential gains from using the innovation and technical knowledge. His theory suggested that many organizations will defer adoption until knowledge barriers are sufficiently lowered. He suggested organizations that can accommodate organizational learning, by applying what they learn about new technologies in multiple client settings.

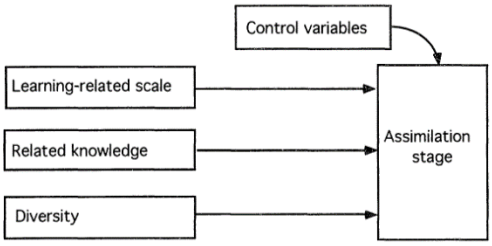


Figure 1: Model

Software Process Innovation: In the 1990's, CASE based tools and object-oriented programming technologies are major innovations in software process technology. These technologies are termed as *Software Process Innovations* (SPI).

Model: The assimilation of SPIs will be exhibited by organizations with higher learning-related scale, related knowledge, and diversity (Figure 1). A six-stage model of SPI assimilation was employed: (1) awareness, (2) interest, (3) evaluation/trial, (4) commitment, (5) limited deployment, (6) general deployment.

Hypothesis:

Learning-Related Scale	The learning-Related Scale is positively related to Object-Oriented Programming Languages Assimilation Stage.
Related Knowledge	Related Knowledge is positively related to Assimilation Stage.
Diversity of Knowledge and Activities	Diversity is positively related to Assimilation Stage.

Analysis of Biases:

1. Analysis of Potential Response Bias: 2 kinds of response bias analysis. A *logistic regression analysis* was performed using Object Oriented adoption as the dependent variable and host size, IT size, and government as the independent variables.
2. Analysis of Potential Method Bias: A paper-based follow-up survey was used to support the analysis of method bias. A one-page survey was constructed to measure the assimilation stage.

Implication: The main implication of this research for end-user organizations is that they would be well advised to view SPI assimilation as a multiyear process of organizational learning, and to carefully assess the extent to which they fit the profile of an early and sustained assimilator.

Citations:

[1] Orlikowski, W.J. and Barley, S.R., 2001. Technology and institutions: What can research on information technology and research on organizations learn from each other? MIS quarterly, 25(2), pp.145-165.W

[2] Armstrong, C.P. and Sambamurthy, V., 1999. Information technology assimilation in firms: The influence of senior leadership and IT infrastructures. Information systems research, 10(4), pp.304-327.

[3] Gallivan, M.J., 2001. Organizational adoption and assimilation of complex technological innovations: development and application of a new framework. ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 32(3), pp.51-85.

Week 9 Summary Report, 15 October 2018

Impact of Workload on Service Time and Patient Safety: An Econometric Analysis of Hospital Operations

Diwas S. Kc and Christian Terwiesch

Prepared By: Sayed Mohsin Reza (Student ID # 108570451), Ph.D. Student, Computer Science & Engineering, University of Colorado Denver, USA, sayedmohsin.reza@ucdenver.edu .

Summary: This paper presented a model of service worker productivity that includes the effect of load and overwork on service rates. To show that, they tried to find the relationship between the processing speed of service workers and system load.

In health care operations, productivity and service operation is crucial to the *quality*. The authors used operational data from a secondary data source (patient transport services and cardiothoracic surgery) to make a framework. Prior researches were done based on capacity utilization not through service rates. But this paper presented a valid framework of service operations where workers vary their service rates with the state of the system. Moreover, they researched the impact of production system design on productivity.

Hypothesis Development:

LOAD: measurable component of the level of utilization of the system's resources that relates to the unit of work.

SVCTIME: (Service Time) Measurable component of the service time taken to process a request.

OVERWORK: Measurable component of increasing function in the difference between the observed LOAD and average time.

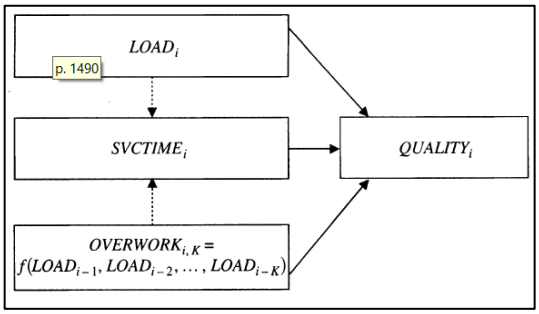


Figure 2: Effect of LOAD, Overwork, and Service Time on Quality.

Data Collection	Patient Transport Study	Cardiothoracic Surgery Study
Econometric Analysis	To capture a potentially nonlinear relationship between LOAD and SVCTIME, they created a categorical variable for LOAD for values in the ranges 0-0.3, 0.3-0.5, 0.5-0.65, 0.65-0.8, and 0.8-1 such that we had approximately similar numbers of observations within each range.	They observed the lengths of stay and quality measure of patients who pass through a single cardiothoracic surgery unit. In addition to that, they observed the admission and discharge dates for each patient, which are used to compute the patient length of stay as well as the daily census.
Results	The correlation between LOAD and OVERWORK is 0.295. Load and overwork have opposing effects, so at any given point in time, depending on the relative magnitudes of load and overwork, the net effect might be either a decrease or an increase in the service rate.	2 effects related to quality. a. Overwork leads to an increase mortality rate. b. Increased levels to early discharges, which in turn a small increase in the post discharge.

Citations:

[1] Staats, B.R. and Gino, F., 2012. Specialization and variety in repetitive tasks: Evidence from a Japanese bank. Management science, 58(6), pp.1141-1159.

[2] Singer, S.J. and Vogus, T.J., 2013. Reducing hospital errors: interventions that build safety culture. Annual review of public health, 34, pp.373-396.

[3] Anand, K.S., Paç, M.F. and Veeraraghavan, S., 2011. Quality–speed conundrum: Trade-offs in customer-intensive services. Management Science, 57(1), pp.40-56.

Week 10 Summary Report, 22 October 2018

Competition Among Virtual Communities and User Valuation: The Case of Investing-Related Communities

Bin Gu, Prabhudev Konana, Balaji Rajagopalan, and Hsuan-Wei Michelle Chen

Summary: This research is mainly based on the cause and effect of virtual communities and examine the trade-offs between information quantity and quality. They proposed a key factor that determines the direction of network external communities is posting quality from virtual communities.

Data Collection Method: They collected a set of online postings (around stratified 500,000 stocks postings) from following three large Virtual Investing-related Communities (VIC).

- Yahoo Finance, Silicon Investor, Raging Bull

Experiments: First, they analyze text messages and classify emotive content from those online postings. They developed an automated classifier based on the approach of Das and Chen (2001) and implemented a decision tree classifier to categorize messages using Foltz (1999) method. They used 5 classifiers to classify posting quality.

1. Lexicon Based Classifier (LBC): Categorize a message by matching content against a category.
2. Readability Based Classifier (RBC): Subset of messages are pushed into 3 variables: word count, mean word length, and several unique words.
3. Weighted Lexicon Classifier (WLC): To eliminate bias from the results from LBC, they used this classification to get a total number of keywords in classes.
4. Vector Distance Classifier (VDC): It classifies the messages as word vectors in D-dimensional spaces.
5. Weights Lexicon Classifier (DWLC): It is another variation of LBC but it can overcome the equal weight bias.

Causality*: *An increase in posting volume* (cause: quantity) has implications for community providers because of the *increased revenue* (effect: quality) from advertisers.

Citations:

[1] Goh, K.Y., Heng, C.S. and Lin, Z., 2013. Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. Information Systems Research, 24(1), pp.88-107.

[2] Oh, C. and Sheng, O., 2011, December. Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement. In Icis (pp. 1-19).

***Terms:**

1. *Causality:* The relationship between cause and effect and the principle that everything has a cause.

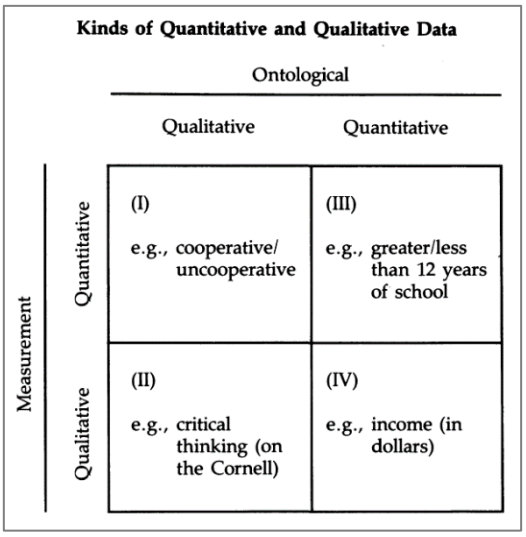
Week 12 Summary Report, 5 November 2018

Against the Quantitative-Qualitative Incompatibility Thesis or Dogmas Die Hard

KENNETH R. HOWE

Prepared By: Sayed Mohsin Reza (Student ID # 108570451), Ph.D. Student, Computer Science & Engineering, University of Colorado Denver, USA, sayedmohsin.reza@ucdenver.edu.

Summary: The paper talked about the distinction of quantitative-qualitative distinction which can be applied at different levels: data, design and analysis, interpretation of results, and epistemological paradigms.



At the level of data, they provide two distinctions – *ontological* and *measurement*.

At the levels of design, analysis, and interpretation of results, quantitative and qualitative researchers diverge in the traditions they are willing to make and how much attention they pay to "experience-near" data.

At the level of epistemological paradigms, the philosophy of science has moved on, into a "new" or "post positivistic" era.

While some researchers can be anticipated to follow both quantitative and qualitative methods, they need at least a fundamental understanding of what substitute approaches can provide and, accordingly, they should bring a collaborative attitude to research.

Moreover, a few philosophers will be made insecure as it blurs the lines between philosophical theory (epistemology) and research practice (methods). But ultimately, a philosophical perspective is valuable just to the extent that it helps shape practice and helping to shape a practice requires careful attention to just what the practice is.

Conclusion: It is high time to close the quantitative versus qualitative conversation.

Citations:

[1] Yin, R.K., 2015. Qualitative research from start to finish. Guilford Publications.
[2] Greene, J.C., 2007. Mixed methods in social inquiry (Vol. 9). John Wiley & Sons.
[3] Bogdan, R. and Biklen, S.K., 1997. Qualitative research for education. Boston, MA: Allyn & Bacon.
[4] Denzin, N.K., 2008. Collecting and interpreting qualitative materials (Vol. 3). Sage.

Findings: Mostly this paper is cited by books related to quantitative and qualitative research.

Week 13 Summary Report, 12 November 2018

Is Mixed Methods Social Inquiry a Distinctive Methodology?

Jennifer C. Greene

Summary: The paper is talked about how the mixed-methods approach embraces multiple paradigmatic traditions and has distinctive methodological components and distinctive markers of practice. The author identified 4 methodological domains to find accomplishment and what important questions remain to be engaged in these domains.

Domains	Accomplishments	Important Questions
Philosophical Assumptions and Stances	• Possibility and sensibility of mixing philosophical frameworks. • The sensibility of mixing paradigms or mental models.	• What does influence inquirers' methodological decisions in practice? • Does further engagement in this philosophical domain concern the alternative paradigm stance?
Inquiry Logics	• Purposes of hypothesis testing or explanation. • Dimensions: Interaction, status, timing, transformative intent study, and strands. • The integrated analysis involves joint and interactive mixed analysis.	• Do we use a mix of methods with offsetting biases of error and perspective? • What specific characteristics would best fulfill such purposes and design parameters? • What does mixing happen?
Guidelines for Practice	• Sample integration is one of the legitimation criteria. • Grounded visualization integrates grounded theory analysis.	• How to choose methods in each inquiry context? • What does mixing happen?
Sociopolitical Commitments	• Unsettled domain. To settle, the challenge is taken for granted. • The way of thinking also generates questions.	• Is mixed methods social inquiry a distinctive methodology?

Citations:

[3] Venkatesh, V., Brown, S.A., and Bala, H., 2013. Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. MIS Quarterly, 37(1).

[4] Denscombe, M., 2008. Communities of practice: A research paradigm for the mixed methods approach. Journal of mixed methods research, 2(3), pp.270-283.

[5] Onwuegbuzie, A.J., Johnson, R.B. and Collins, K.M., 2009. Call for mixed analysis: A philosophical framework for combining qualitative and quantitative approaches. International journal of multiple research approaches, 3(2), pp.114-139.

Week 15 Summary Report, 26 November 2018

Positioning and Presenting Design Science Research for Maximum Impact

Shirley Gregor, and Alan R. Hevner

Prepared By: Sayed Mohsin Reza (Student ID # 108570451), Ph.D. Student, Computer Science & Engineering, University of Colorado Denver, USA, sayedmohsin.reza@ucdenver.edu.

Summary: The paper addresses the issues associated with knowledge contributions in design science research in information systems.

Objective: The objective of this paper to clarify some perceived confusion in terminology and the types of contributions.

Contribution: To do this, the authors developed a DSR knowledge contribution framework and DSR publication schema.

Design Science Research Publication Schema: The following table outlines a publication pattern for a DSR study.

Table 3. Publication Schema for a Design Science Research Study

Section	Contents
1. Introduction	*Problem definition, problem significance/motivation, introduction to key concepts, research questions/objectives, scope of study, overview of methods and findings, theoretical and practical significance, structure of remainder of paper.* For DSR, the contents are similar, but the problem definition and research objectives should specify the ***goals*** that are required of the artifact to be developed.
2. Literature Review	*Prior work that is relevant to the study, including theories, empirical research studies and findings/reports from practice.* For DSR work, the prior literature surveyed should include any prior design theory/knowledge relating to the class of problems to be addressed, including artifacts that have already been developed to solve similar problems.
3. Method	*The research approach that was employed.* For DSR work, the specific DSR approach adopted should be explained with reference to existing authorities.
4. Artifact Description	A concise description of the artifact at the appropriate level of abstraction to make a new contribution to the knowledge base. This section (or sections) should occupy the major part of the paper. The format is likely to be variable but should include at least the description of the designed artifact and, perhaps, the design search process.
5. Evaluation	Evidence that the artifact is useful. The artifact is evaluated to demonstrate its worth with evidence addressing criteria such as validity, utility, quality, and efficacy.
6. Discussion	*Interpretation of the results: what the results mean and how they relate back to the objectives stated in the Introduction section. Can include: summary of what was learned, comparison with prior work, limitations, theoretical significance, practical significance,* and *areas requiring further work.* Research contributions are highlighted and the broad implications of the paper's results to research and practice are discussed.
7. Conclusions	*Concluding paragraphs that restate the important findings of the work.* Restates the main ideas in the contribution and why they are important.

Suggestions for new researchers: Researchers should be able to exercise their discretion to vary what is proposed or to suggest and implement improvements. Mainly, the proposed contribution framework and publication schema should be guidelines; they are not meant to be followed blindly or serve as a straitjacket for future work.

Citations:

[1] Vaishnavi, V.K. and Kuechler, W., 2015. Design science research methods and patterns: innovating information and communication technology. Crc Press.

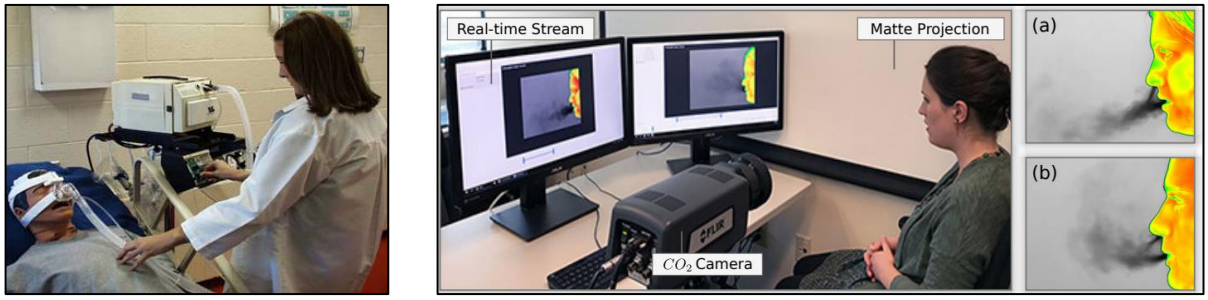
[2] Lee, A.S., Thomas, M. and Baskerville, R.L., 2015. Going back to basics in design science: from the information technology artifact to the information systems artifact. Information Systems Journal, 25(1), pp.5-21.

[3] Turber, S., Vom Brocke, J., Gassmann, O. and Fleisch, E., 2014, May. Designing business models in the era of internet of things. In International Conference on Design Science Research in Information Systems (pp. 17-31). Springer, Cham.

Research Paper Proposal

The Impact of Non-contact based Behavioral Breathing Pattern Analysis in Healthcare System

Sayed Mohsin Reza, Hawkar Oagaz, Dr. Jiban Khuntia, Dr. Min Choi



(a) Traditional contact-based respiratory monitoring system with patient discomfort

(b) Non-contact based respiratory analysis preserving patient comfort

Figure 1: Difference between contact and non-contact based respiratory analysis

Objective: Respiration monitoring in the healthcare system are getting importance for pulmonary patients with common medical problems such as asthma and chronic obstructive pulmonary disease like sleep apnea. No existing non-contact method of behavioral respiration monitoring perfectly placed in the healthcare system rather than contact methods. Both non-contact and contact-based methods have some trade-off between comfort and accuracy. Contact methods require placing sensors directly on the patient's body, provide reliable measurements, but are uncomfortable for the patient, which alters the main objective of natural breathing behaviors. Conversely, non-contact methods monitor respiration remotely and comfortably. In our previous research, we presented a structural foundation for the respiratory analysis of turbulent exhale flows through the visualization of dense CO_2 density distributions [1]. In this proposal, we will present a non-contact based respiratory monitoring healthcare system through a behavioral pattern of exhale to be useful for pulmonologists. During this research, we will use a thermal imaging camera to visualize the exhaled airflow of an individual, and record that exhale raw dataset of a patient for analysis. Breathing rate and respiratory behaviors can be extracted in real-time. Our proposed respiration monitoring technique will accurately report respiratory behaviors including, breathing rate, exhale strength and capacity, towards insights into lung functionality. We will exemplify the research by collecting real-time data from pulmonary disease patients for behavioral breathing analysis.

Possible method: As we are interested to give patient comfort during respiratory monitoring or data collection, we must use non-contact-based methods in our system. In our experiments, we will use a custom FLIR A-series thermal camera that produces 640x512 images at 30 frames per second. To make

Figure 2: Recorded breathing flow from the mouth (left), nose (center), and both nose and mouth simultaneously (right). Through our imaging process, we obtain an accurate illustration of the CO2 density distribution and flow behavior.

proper turbulent CO_2 exhale flows, we will coordinate the development of a visualization process that directly targets the CO_2 spectral band (3-5[μm]) to get breathing raw data.

Expected contributions: The primary contributions of the proposal to do respiratory analysis evolving a limited set of quantitative metrics such as breathing rate, flow analysis, and tidal volume estimates (Figure 2). Later, we will move our research to detect abnormalities within normal breathing based on the cross-sectional view with another patient dataset. We will make a robust breathing analysis system that can evaluate qualitative information like nose-mouth distribution (Figure 2), strength, and minute flow variances associated with abnormal breathing. The ability to see changes in exhale patterns in real-time can alert medical professionals to subtle breathing difficulties.

Motivation: The motivation behind such research is to make a respiratory healthcare system that allows the patient to give comfort during breathing analysis (Figure 1). Currently, the most accurate methods of respiration monitoring, such as placing ECG electrodes on the patient's body, putting thermistors in the patient's nose, having the patient wear an abdominal strain-gauge transducer, or monitoring multiple bio-physiological parameters concurrently with polysomnography, all involve place sensors directly on the patient's body (Figure 1). These direct measurements have a high rate of accuracy, but cause discomfort and alter the natural breathing of the patient. This problem is the main motivation behind this research into innovative methods that measure respiration remotely and preserve patient comfort. Our proposed non-contact methods of respiration monitoring utilize remote thermal sensors. From the perspective of our motivation, this sensor will work as a noncontact device and will give more comfort to the patient. Specially for the children and new infants, our system will able to get a normal breathing pattern.

Intellectual Merit: This system will be capable of measuring breathing rate, as well as other metrics not obtainable through other respiration monitoring methods such as nose to mouth distribution (Figure 2). Additionally, the system will allow you to get insight into respiration characteristics such as exhale strength, flow, and pattern. To make a healthcare information system, the data collection will work independently of the physical characteristics of the patient. It can be used to monitor individuals of various ages and sizes, including young children that are unable to use certain contact methods. The system will able to use the core features of live streaming as a standalone measurement for easy clinical use to visualize behavioral exhale patterns. Overall, the system will provide medical professionals with

the visualization application to comprehensively examine breathing activities and can be used for various clinical applications.

Broader Impacts of the Proposal: Preservation of the patient's behavioral exhale record can open a new arena to examine the patient's pulmonary functionalities over a long time. It will able to evaluate patient behavior from exhaling flow behavior characteristics. From exhale flow analysis, we will be able to determine the tooth anatomy of children. Moreover, CO_2 exhale imaging opens a new medical significance in the monitoring of normal versus pathological airflow from the lower and upper airways. In a bigger scenario, the system will be able to make other physical measurements like lung capacity or effect of therapy in pulmonary patients with common medical problems such as asthma and chronic obstructive pulmonary disease. We will able to record the patient's exhaled flow and later we can able to determine airflow velocity and nose-mouth distribution that can help medical professionals to find airway obstruction or those at risk for sleep apnea or sudden infant death. In special cases, the monitoring system will able to record airflow of infants sleeping prone versus supine or in car seats or soft bedding, the advancement of research into the reason and prevention of Sudden Infant Death Syndrome (SIDS) cases achieves a new trajectory in high-resolution CO_2 respiratory behavioral analysis in the healthcare system.

References

[1] Transue, S., Reza, S.M., Halbower, A.C. and Choi, M.H., 2018, March. Behavioral analysis of turbulent exhale flows. In Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on (pp. 42-45). IEEE.

[2] Schoun, B., Transue, S., Halbower, A.C. and Choi, M.H., 2018, March. Non-contact comprehensive breathing analysis using thermal thin medium. In Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on (pp. 239-242). IEEE.

Final Research Paper Report

Quantitative and Qualitative Research in Mouth Breathing: A Systematic Review*

Sayed Mohsin Reza[1], Hawkar Oagaz[1], Jiban Khuntia[2], and Min Hyung Choi[1]

***Abstract*— Mouth Breathing (MB) is the habitual respiration through the mouth instead of nose and itâs a concerning issue for healthcare professionals due to its different types of effects. The underlying cause of this disorder is an obstructed nasal airway. People of all ages may have this problem which causes snoring, dry mouth, chronic fatigues as short-term effects. Specially, children have the symptoms of large tonsils, dry, cracked lips, daytime sleepiness. This disorder has long term negative effects on the development of facial morphology and orthodontic structure. Researchers do their research based on the causes, effects as well as diagnostic test processes to find quantitative and qualitative analysis. This paper objective is to systematically review the different types of mouth breathing research papers to find the effectiveness of different diagnostic process with their analysis. To do that we collected research articles on mouth breathing from electronic databases PubMed, Wiley Library, Science Direct etc. Based on the inclusion and exclusion criteria, we concise the no of review articles to visualize the overall research scene in MB. Our study reveals that there is not yet qualitative assessment in mouth breathing research. Moreover, most quantitative researches are done using Rhinomanometry test process. Other findings of this paper is that newer clinical test are mostly done in controlled way to detect MB and recently proved as future complementary diagnosis of MB.**

I. Introduction

Breathing is a significant physiological phenomenon in human respiration system. This system includes processes like oxygen being inhaled and carbon dioxide being exhaled released through the nose or mouth. The entire process from the inhalation to exhalation is known as a respiration cycle [1]. Generally, the mouth does not contribute to respiration. When your nose is congested, the body habitually resorts to the mouth source that can provide oxygen. Also, after exercising strenuously, mouth breathing can help get oxygen to muscles faster.

Respiration process is generally done through the nose as it can filter and retains small particles in the air. In addition to that nose can add moisture to the air to prevent dryness in the lungs and warms up cold air to body temperature before going to lungs. Apart from this importance, nose produces nitric oxide to improve lungsâ ability to absorb oxygen. The properties of nitric oxide are anti fungal, antiviral and antibacterial to help the immune system to fight infections.

*This work was not supported by any organization

[1]H. Kwakernaak is with Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, 7500 AE Enschede, The Netherlands h.kwakernaak at papercept.net

[2]P. Misra is with the Department of Electrical Engineering, Wright State University, Dayton, OH 45435, USA p.misra at ieee.org

Besides the importance of nose breathing, another physical phenomenon mouth breathing has short-term negative effects as well as long-term negative effects in a healthier body. Mouth-breathing can affect breather during sleep that exhibits poor sleep quality from getting insufficient oxygen by breathing through the mouth. Generally, if a breather breath through mouth, certain sounds making is more difficult to produce that causes impaired speech.

Moreover, MB can make breatherâs mouth dry which results in halitosis. This will produce less saliva in breather mouth to clean out the germs. To give a treatment of such issues, dentists generally do some predefined tests like Mirror, Water Holding, Butterfly, Rhinomanometry test, and CO2 sensor test etc. These are contact method tests and provide an idea about breather breath through mouth or nose most. Exhale imaging test is non-contact method and reliable to visualize the exhale and categorize the exhale pattern to nose or mouth breathers. Rhinomanometry and exhale imaging test provides quantitative information of breathing. Others tests are observable by the pulmonologists or researchers.

Researchers are doing mouth breathing research on the availability of patients and clinical tests. Also they mostly focus on a specific cause or effects that has relation to mouth breathing. For that reason, a lot of research articles are found which focus on short term effects of MB. Also some researchers use observable study to produce quantitative analysis for mouth breathing research. Nowadays we have better technologies to analysis breath pattern as quantitative data as well as find long term effects of MB by doing qualitative research.

In this paper, we try to do a systematic review after analyzing research papers on mouth breathing based on the causes, effects and diagnostic process to find the relation between the data collection tools and type of researches. We will visualize different types of methods of research, their findings and concise the research ideas by categorizing the objective. Our objective is to finalize the research arena in MB with reviewing the research articles from different country, causes, effects or clinical test.

The paper starts with the methods of doing systematic review in section II. In this section, we provided the search criteria based on country, quantitative and qualitative research. Moreover, we refine the selection criteria do this

systematic review based on the causes, effects and diagnostic process. Later in Section III, we showed the analysis of reviewed papers after inclusion and exclusion of papers. Then in Section IV, we discuss the whole analysis in a descriptive way. A conclusion will be drawn in Section V.

II. Methods

A systematic literature review was started from collecting research papers from the following electronic databases: PubMed, PMC, BIR, Wiley Library, Scopus, PsycInfo, PROCONF database, and Science Direct. All the research papers were collected between September and December 2018. Before start searching research papers, we tried to find some statistics where most collaborative and non-collaborative research were done. We found an article from Forbes to get the idea of languages of research papers [21] and country rankings from Science Journal Ranking (SJR) in medical industry [22]. Figure 1 will provide a statistic of countries which are now leading in medical research and innovation during last 21 years (1996-2017).

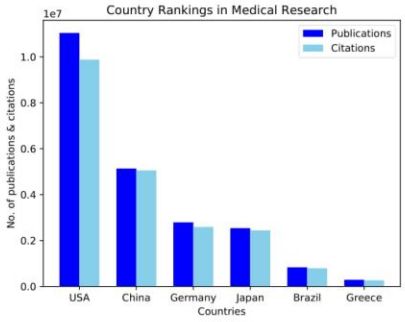


Fig. 1: Country Rankings in Medical Research

To do research articles, descriptors and crossings, we used the top leading countries languages (English, Chinese, German, Japanese). Chinese and Japanese language-based papers are hard to convert to English. For this reason, we decided not to include those research papers. We also found that research on mouth breathing is good in Brazil and Greece. So in our inclusion criteria we added those country languages. Sample crossings are shown in Table 1.

The search was conducted independently, following criteria of inclusion and exclusion. As inclusion criteria, original articles (whose manuscripts were published in any language) addressing the types of research, cause, effects and diagnostic process in the evaluation of mouth breathing were selected.

The methodological characteristics of articles were addressed according to inclusion criteria, statistical analysis and statistical comparison between selected groups (Table 2)

Crossings in English	**Crossings in German**
"Cause and Effects" and "Mouth Breathing"	"Ursache und Wirkung" und "Mundatmung"
"Diagnosis" and "Mouth Breathing"	"Diagnose" und "Mundatmung"
"Clinical test" and "Mouth breathing"	"âKlinischer Test" und "Mundatmung"
"Breathing Disorders" and "Clinical test"	"AtemstÃ¶rungen" und "Klinischer Test"
"Contact Method" and "Breathing Disorders"	"Kontaktmethode" und "AtemstÃ¶rungen"
"Non-Contact Method" and "Breathing Disorders"	"BerÃ¼hrungslose Methode" und "AtemstÃ¶rungen"

TABLE I: Search criteria for articles

Beside methodological classification, we also classify all the articles into following categories to see how the research articles target mouth breathing research.

- Obstructive MB: A resistance or complete obstruction of nasal airflow.
- Habitual MB: Person who breathes through the mouth by force as a habit whether their nasal airway is free or not.
- Anatomical MB: Breather whose upper lip does not permit complete closure without undue effort. This type of MB is due to the poor development of facial morphology.

In third category, we refine the articles by the cause and effects of mouth breathing. In general, allergies, physical obstructions, and chronic infections cause many breathers to breathe through the mouth. Apart from these generalities, we identify some other the causes in the following ways.

Deviated Nasal Septum (DNS): If nasal septum is displaced or deviated, then it allows the mouth to inhale or exhale proper amount of oxygen or carbon dioxide respectively.

Larger Nasal Polyps: Nasal polyps are soft, painless, noncancerous growths on the lining of your nasal passages or sinuses. But larger nasal polyps can lead block your nasal airways.

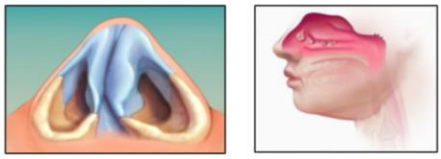


Fig. 2: a) Deviated Nasal Septum (left) b) Larger Nasal Polyps (Right)

Resistance in nasal airway affects natural breathing showed negative effects in a healthier body. This is also a concerning issue for health professionals in terms of period and age. We categorize the period effects into short-term and long-term effects.

Short-Term Effects: Generally, when a breather breath through mouth, it shows the effects on face externally by exhibiting lips slacking, short upper lip heavy lower lip and Increased mandibular plane angle. Sometimes those effects become worst if there is any occlusion of teeth. It can change the distal relation of the mandible to the maxilla. The changes in oral structure may seriously affect speech performance. Also, this change can affect breatherâs smiling. In case of children, drying of the gingiva causes saliva that accumulates debris resulting in bacterial infection.

Long-term effects: The effects of mouth breathing are not stoppable if the treatment is not done earlier. Negative effects of MB are extended timing during orthodontic treatment, altered facial morphology and sleep apnea. Inadequate airflow through the mouth imbalance the forces by the tongue which will lead to constricted maxillary arch. This changes the teeth orientation during growing up. As a result, higher chances of teeth shifting may happen and for braces treatment, it will take longer time. Moreover, it will change the facial morphology as a symptom of the narrow jaw, flatter face, and smaller chin.

In final touch, we try to refine the articles through to clinical tests in mouth breathing. In general, to evaluate any respiratory behavior, we need to qualitatively and quantitatively find respiratory metrics like respiratory rate, pulse transit time, tidal volume estimation. Generally, the researchers use the following test for their research. To see the intent of research, we categorize the articles in clinical test process (F3).

Graded Mirror (Fog) Test: In this test, a double-sided mirror is used and placed between nose and mouth. They qualitatively measure nasal or mouth breathing based on fogging on the nasal side or mouth side.

Masslerâs Water Holding Test: Breather is asked to hold the mouth full of water. During a period if breather canât retain the water, they qualitatively detect as the breather tries to breathe through the mouth.

Butterfly test (cotton test): A butterfly-shaped pied of cotton is placed over the upper lip below the nostrils. If cotton flutters down it indicate nasal breathing.

Lip Sealed Test: In this test, breatherâs mouth is completely stapled for 3 minutes to check whether breather can do nasal breathing or not. Based on the qualitative measurement, they find whether the breather is a nasal breather or mouth breather.

Fig. 3: a) Mirror (Fog) Test b) Water Holding Test(Right)

Rhinomanometry Test: A computerized test that assesses nasal airflow obstruction based on pressure and flows during breathing. To get the anatomical MB, this test is prescribed. Before the test, breathers use a decongestant spray that can open nasal airway due to obstructive MB. After a few moments, this test is conducted to check the mouth breathing is from obstructive MB or anatomical MB.

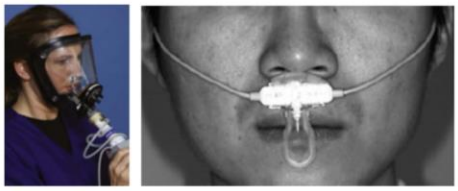


Fig. 4: a) Rhinomanometry Test (left) b) CO2 Sensor Test (Right)

Exhale Imaging Test: Exhale imaging provides a basis for evaluating high-level breathing traits over time based on a set of observable phenomena to provide metrics that are helpful for physicians as well as for patient reporting. All these monitoring techniques include momentary changes within exhale flows that inherently subsidize to the secondary level of exhaling analysis associated with breathing disorders and abnormalities. To extend respiratory analysis, our previous research provides CO2 imaging and respiratory monitory framework that enables the visualization of dense exhale flow behaviors [20]. Moreover, this test established a foundation in respiratory monitoring system that provides nose mouth exhale ratio.

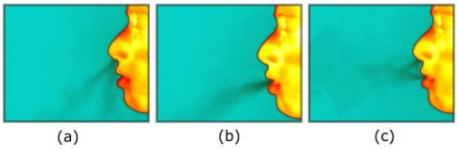


Fig. 5: Visualization of (a) only nose (b) only mouth and (c) nose-mouth exhaling

All these refining criteria allow us to review a concise number of mouth breathing research paper that has

Articles and Criteria	Bonuck et al. (2012)	Daimon et al. (2013)a	Pinto et al. (2015)	Anselmo-Lima et al. (2011)	Piperopoulou et al. (2011)
Specified inclusion criteria	Yes	Yes	Yes	Yes	Yes
Qualitative analysis	Yes	Yes	Yes	Yes	No
Quantitative Analysis	No	No	No	Yes	Yes
Randomness during research	Yes	No	Yes	Yes	No
Control group	Yes	Yes	Yes	Yes	Yes

TABLE II: Sample Methodological classification of articles

a combination of all search and selection criteria. Table 2 provides the samples of methodological classification of selected papers.

III. Results

From our flowchart (Figure 6), we can observe the number of research articles are observed after inclusion and exclusion of research articles, descriptors.

To visualize the results of such review we decided to show our study variables of selected articles: author/year, type of research, samples, methods, category and results (Table 3).

IV. Discussion

The primary contribution of the paper is to do a systematic review that allow analysis on mouth breathing research based on the country research, causes, effects and diagnosis process.

In 2012, a quantitative cluster analysis was done by Karen Bonuck et al. with the aim to visualize cluster-based sleep disordered due to mouth breathing. 5 emerged cluster has been showed in that research doing a quantitative analysis on 10441 patients.

The cause of choosing this paper is to reflect the observable study for mouth breathing research and specified inclusion criteria and randomness in the research. In our investigation, we found that another two papers using the same technique to address the randomness by adding family and patient characteristics in the paper.

The clinical test of MB directs a research to a quantitative and qualitative research. To get a reliable quantitative analysis, we choose a paper which had a clinical test based on CO2 sensor. The author had a control group to do quantitative analysis. To relate mouth breathing with habitual perception, they looked for the changes in muscle activity during chewing food. This was a unique study to check habitual MB. From the study they quantitatively showed that degree and duration of vertical force on the posterior teeth are lowered initiating mouth breathing.

Apart from anatomical and habitual MB, researchers also look for obstructive MB. Pinto et al. did a research on getting cephalometric changes between mouth breathers and nose breathers. To find that level of changes, they did the analysis using a software named cephalometric tracing module of dolphin imaging. We choose this paper to show a kind of different technology which can allow MB researchers to do qualitative analysis over a long time as well as their paper provides a cross-sectional study to get rid of the bias results. Their final conclusion is that obstructive MB leads to a different cephalometric pattern than nasal breathers.

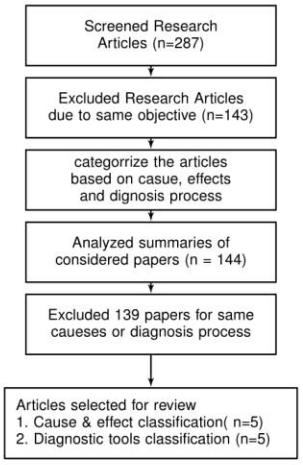


Fig. 6: Flow chart of screened, excluded and analyzed papers

Morphological and dentofacial alternations have been caused for mouth breathing. Anselmo-Lima et al. examined the influence of MB on children facial morphology. The author did that examination before and after adenoidectomy or adenotonsillectomy to compare the mouth breathing and nose breathing groups. They showed that after MB rehabilitation mandibular growth direction is normal and mandible inclination decreased which leads the facial height to be increased. This paper has been chosen after getting a very good conclusion with a qualitative analysis.

Another effects habitual snoring that is related obstructive apnea-hypopnea and nocturnal enuresis (NE). Piperopoulou et al. did a study that finds the relationship among habitual snoring, MB and daytime sleepiness with obstructive sleep apnea. In summary, relationship between habitual MB and obstructive MB. For that reason, we selected that papers

Author, Year and Country	Type of Research	Samples	Methods	Category			Results
				F1: Type of observed MB	F2: Type of MB effects	F3: Type of clinical tests	
Katherine Freeman, Karen Bonuck, 2012, USA	Quantitative cluster analysis	10,441 children from age 6, 18, 30, 42, 57, 69, and 81 months.	Avon longitudinal study of parents and children (ALSPAC)	Anatomical MB	Sleep disordered breathingâs (SDBâs)	None, observed data are collected from the study.	5 clusters emerged. Normal (50%) were asymptomatic throughout. Late snores and mouth-breathing (20%) remained asymptomatic until 4 years old. Early snores (10%) and early apnea (10%) clusters had peak symptoms at 6 and 18 months, respectively. All SDB after infancy (10%), symptoms peaked from 30 to 42 months.
N. Ikenaga, K. Yamaguchi, S. Daimon, 2013, Japan	Quantitative Analysis	45 adult volunteers with 3 women.	Monitoring of expiratory airflow from the mouth with a CO2 sensor.	Habitual MB	Changes in muscle activity	CO2 Sensor test	The total and resting duration during chewing were significantly (P ¡ 0.05) longer during mouth breathing compared with nose breathing.
Leticia P. Franco, Jorge A. Pinto, 2015, Brazil	Quantitative Analysis	26 prepubescent children (13 MB and 13 nasal breathing (NB) controls).	A comprehensive ENT evaluation and treatment of the breathing pattern.	Obstructive MB	Obstructive Sleep Apnea	ENT clinical examination	A significant difference was observed in the measurements of the SNB (P ¡ 0.036), NSGn (P ¡ 0.028) and PFH/ TAFH ratio (posterior facial height/ total anterior facial height) (P ¡ 0.012).
Sara E. M. Matta, Wilma T. Anselmo-Lima, 2011, Brazil	Quantitative and qualitative analysis	33 MB children who restored nasal breathing (NB) after surgery and 22 NB children.	Quasi-experimental study	Anatomical MB	Changes in craniofacial and dentofacial	Otorhinolaryngologic evaluation, anterior rhinoscope, Otoscope, lateral skull radiography.	28 months after the MB surgical intervention, patients still presented a dolichofacial morphologic pattern.
Afroditi V. Sakellaropoulo, Fanni Athanasiadou-Piperopoulou, 2011, Greece	Quantitative Analysis	28 males and 14 females, age ranged from 3.5 to 14.5 years old.	Otolaryngology evaluation and Polysomnographic study was performed.	Habitual MB	Habitual snoring	Polysomnographic study (PSG)	Children with OSAHS and NE had a higher arousal index (8.14+-8.05) compared with OSAHS children without NE (4.61 Â±7.95)($p = 0.19, z = â1.28$).

TABLE III: Study variables for selected papers

for review. To find such relationship, at first they performed Otolaryngological evaluation and then polysomnography. As this study is related to sleep, they took minimum of 6 hours of every patient to make sure the data quality.

These all papers results indicate that most of the paper is done in quantitative way. Also some of the papers have less subjective population, randomness and chosen methods during research. In categorization, we touched every aspect of anatomical, habitual and obstructive MB. For anatomical study we find that anterior rhinoscope, Otoscope, lateral skull radiography is important to find change in morphological structure or populate MB group for research.

In another category we touched cause and effects of MB to initiate research. We found that most researches are done on sleep disordered breathing (SDB) or Sleep apnea. But there are some other aspects that can qualitative relate mouth breathing with anatomical structure or habitual tasks. Obstructive mouth breathing has a lot of research rather than the anatomical MB or habitual MB.

The last category, clinical tests have been introduced to see the variation in mouth breathing research. If we see the trend from water holding test to exhale imaging or CO2 sensor test, we will also see the qualitative research in this area. From the perspective of test, imaging based diagnostic tools will provide quite good platform for practitioner to give treatment to the patients as well as researcher. But mostly Rhinomanometry test has been selected for their research to see obstructive breathing.

As mouth breathing is trivial problem that lead to other diseases, the researchers are mostly looked for obstructive reason in a quantitative way. Most of the mouth breathing researches are conducted quantitatively not in qualitative

way.

This is high time to do some qualitative research as we have better clinical test process and find the relationship between MB causes and effects in long term. All of these paper results indicate the need for more fine grained studies in mouth breathing research, with concise objective and quantitative methods to derive qualitative information.

V. Conclusion

In this systematic review, the comprised studies demonstrate that the use of newer clinical test can open new qualitative research arena in mouth breathing research. Although there may be a theoretical consensus regarding cause, effects and clinical tests in mouth breathing research, it has been possible to reveal scientifically and clinically the relationship among the anatomical MB, habitual MB and obstructive MB.

Besides the relationship, obstructive MB can be test with new clinical test process. The older techniques like water holding, glass based test alter the natural breathing. To keep mouth breatherâs data natural, we have to move forward to use CO2 sensor based test or exhale imaging test. It will allow researchers to collect data remotely as well as giving the patient comforts.

According to the research papers investigation, most of the time Rhinomanometry test has been used for obstructive research but it can be developed using CO2 sensor test and exhale imaging test to make controlled studies for qualitative research.

References

[1] ALâKhalidi, F.Q., Saatchi, R., Burke, D., Elphick, H. and Tan, S., 2011. Respiration rate monitoring methods: A review. Pediatric pulmonology, 46(6), pp.523-529.

[2] Transue, S., Reza, S.M., Halbower, A.C. and Choi, M.H., 2018, March. Behavioral analysis of turbulent exhale flows. In Biomedical Health Informatics (BHI), 2018 IEEE EMBS International Conference on (pp. 42-45). IEEE.

[3] Schoun, B., Transue, S., Halbower, A.C. and Choi, M.H., 2018, March. Non-contact comprehensive breathing analysis using thermal thin medium. In Biomedical Health Informatics (BHI), 2018 IEEE EMBS International Conference on (pp. 239-242). IEEE.

[4] Stausholm, K., Rosenberg-Adamsen, S., Skriver, M., Kehlet, H. and Rosenberg, J., 1995. Comparison of three devices for oxygen administration in the late postoperative period. British journal of anaesthesia, 74(5), pp.607-609.

[5] Smith, R.M. and Giannoudis, P.V., 1998. Trauma and the immune response. Journal of the Royal Society of Medicine, 91(8), pp.417-420.

[6] Courtney, S.E., Durand, D.J., Asselin, J.M., Hudak, M.L., Aschner, J.L. and Shoemaker, C.T., 2002. High-frequency oscillatory ventilation versus conventional mechanical ventilation for very-low-birth-weight infants. New England Journal of Medicine, 347(9), pp.643-652.

[7] A. R. Fekr et al. Design and evaluation of an intelligent remote tidal volume variability monitoring system in e-health applications. In IEEE J. of Biomedical and Health Informatics, 2015.

[8] P. Nguyen, S. Transue, M.-H. Choi, A. C. Halbower, and T. Vu. Wikispiro: Non-contact respiration volume monitoring during sleep. In Proceedings of the Eighth Wireless of the Students, by the Students, and for the Students Workshop, S3, pages 27â29, 2016.

[9] Vainer, B.G., 2018. A Novel High-Resolution Method for the Respiration Rate and Breathing Waveforms Remote Monitoring. Annals of biomedical engineering, 46(7), pp.960-971.

[10] Hu, M., Zhai, G., Li, D., Li, H., Liu, M., Tang, W. and Chen, Y., 2018. Influence of image resolution on the performance of remote breathing rate measurement using thermal imaging technique. Infrared Physics Technology.

[11] G. F. Lewis et al. A novel method for extracting respiration rate and relative tidal volume from infrared thermography. Psychophysiology, 48(7):877â887, 2011.

[12] Z. Zhu et al. Tracking human breath in infrared imaging. In Bioinf. and Bioeng., 2005. BIBE 2005., pages 227â231. IEEE, 2005.

[13] K. S. Tan et al. Real-time vision based respiration monitoring system. In Communication Sys. Networks and Digital Signal Processing (CSNDSP), 2010 7th Intl. Symposium on, pages 770â774. IEEE, 2010.

[14] Schoun, B., Transue, S. and Choi, M.H., 2017, November. Real-time Thermal Medium-based Breathing Analysis with Python. In Proceedings of the 7th Workshop on Python for High-Performance and Scientific Computing (p. 5). ACM.

[15] S. Transue et al. Real-time tidal volume estimation using iso-surface reconstruction. In IEEE CHASEâ16, pages 209â218. IEEE, 2016.

[16] Li, C., Lubecke, V.M., Boric-Lubecke, O. and Lin, J., 2013. A review on recent advances in Doppler radar sensors for noncontact healthcare monitoring. IEEE Transactions on microwave theory and techniques, 61(5), pp.2046-2060.

[17] A. R. Fekr et al. Design and evaluation of an intelligent remote tidal volume variability monitoring system in e-health applications. In IEEE J. of Biomedical and Health Informatics, 2015.

[18] Folke, M., Cernerud, L., EkstrÃ¶m, M. and HÃ¶k, B., 2003. Critical review of non-invasive respiratory monitoring in medical care. Medical and Biological Engineering and Computing, 41(4), pp.377-383.

[19] J. Fei, Z. Zhu, and I. Pavlidis. Imaging breathing rate in the co2 absorption band. In Engineering in Medicine and Biology, 2005.

[20] Transue, S., Reza, S.M., Halbower, A.C. and Choi, M.H., 2018, March. Behavioral analysis of turbulent exhale flows. In Biomedical Health Informatics (BHI), 2018 IEEE EMBS International Conference on (pp. 42-45). IEEE.

[21] Matthew Herper, The Most Innovative Countries In Biology And Medicine, Forbes Magazine, March 23, 2011.

[22] Scimago Journal and Country Rank, https://www.scimagojr.com/countryrank.php.

APPENDIX C: COMMUNICATION SKILLS

In this section, I describe my communication skills by describing technical writing and oral communication proficiency.

C1 Writing Communication: A technical report on project heat island effect

Urban Heat Island Effect in the Borderland

Team Name

Heat Index

(Team #4)

Team Members

Sayed Mohsin Reza

Jake Lasley

America Alvarez

Supervisors

Dr. Villanueva Rosales

Dr. Gutierrez-Jurado

Document Version

4.2

December 2019

Section 01 Introduction

An Urban Heat Island (UHI) is an urban area or metropolitan area that is significantly warmer than its surrounding rural areas due to human activities. The temperature difference is usually larger at night than during the day and is most apparent when winds are weak. UHI is most noticeable during the summer and winter [1].

The term "heat island" describes built-up areas that are hotter than nearby rural areas. The annual mean air temperature of a city with 1 million people or more can be 1.8–5.4°F (1–3°C) warmer than its surroundings. In the evening, the difference can be as high as 22°F (12°C) [2]. Heat islands can affect communities by increasing summertime peak energy demand, air conditioning costs, air pollution and greenhouse gas emissions, heat-related illness and mortality, and water pollution [3].

Major impacts of heat island are increased energy consumption; elevated emissions of air pollutants and greenhouse gases; compromised human health and comfort; and impaired water quality [4]. In our project, the surface temperature with local metrological data is analyzed and visualized in the form of a web application where users can search their area and see the heat island effect in temperature value. A sample web page view for El Paso, TX is shown in Figure 1. The project also visualizes the historical map of heat index over the years (2002-2019). Also, the project includes the Google Earth Engine mapping system to visualize the heat index map.

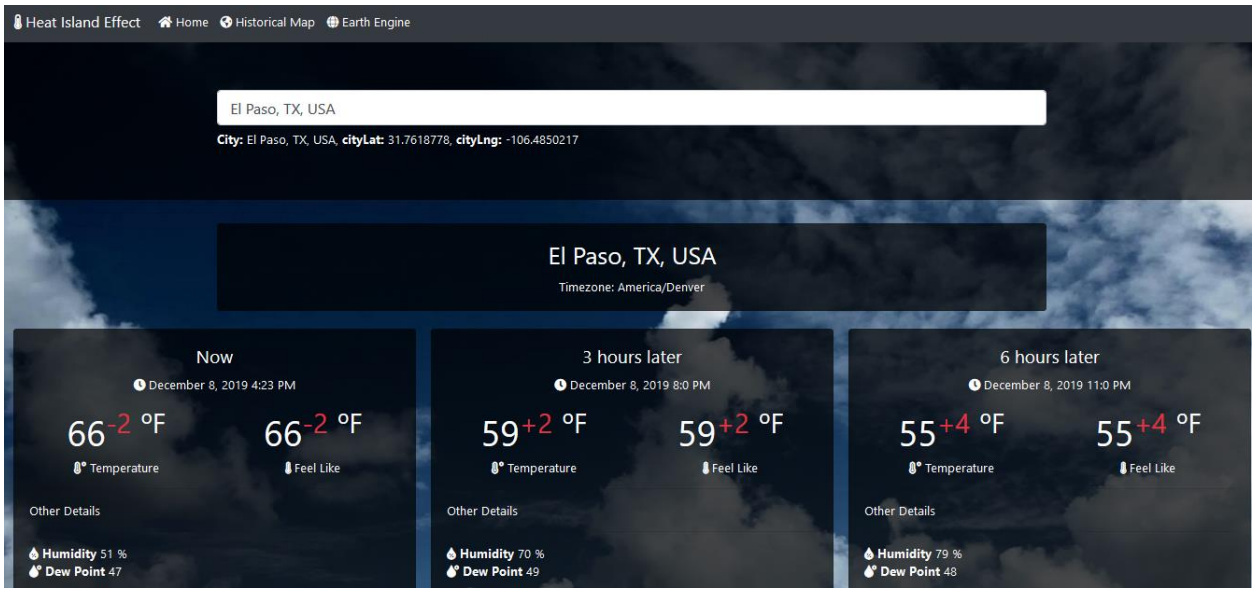


Figure 1: A web view of Heat Island Effect

Section 02 Project Scope

The scope of this project is to integrate the local metrological data and surface temperature data from the satellite to analyze the heat island effect in an area. To accomplish this, several competency questions, datasets, and the cities were first defined to make our project scope clear and comprehensive.

Purpose

The purpose of this project is to help residents living in our Borderland, the cross-border cities of El Paso, TX, USA and Ciudad Juarez, Chihuahua, Mexico, to monitor the heat conditions caused by the Urban Heat Island Effect. Additionally, the purpose of the project is to visualize the vulnerability maps within the Borderland region. Finally, we also analyze the historical vulnerability map of some selected cities.

Motivation

Higher air pollution reduced nighttime cooling, and increased temperatures as outcomes of urban heat island can adversely affect human health. Human health is negatively impacted because of increased general discomfort, exhaustion, heat-related mortality, respiratory problems, headaches, heatstroke, and heat cramps. To solve these issues, we were motivated to make a web application that visualizes the overall heat index and help people to see the current temperature & predicted urban heat index.

Significance

The heat island application will provide a place where the user will be able to find the heat island effect and predicted temperature just by searching the location. In the beginning, the project is developed for El Paso and Juarez city which can be extended in the future. Also, users will be able to see the summertime heat island map from 2002-2019.

Originality

The project is developed in both map and text-based alarming for users who want to know the effect of the heat island. The originality is that we integrated remote sensing data (surface temperature) from the MODIS Terra satellite and local weather station data (air temperature and other meteorological parameters) to predict the effect of heat island by applying machine learning methods for improved accuracy.

Usefulness

The application is useful to El Paso-Juarez residents who wish to know the effect of urban heat islands via temperature values. Residents will be able to search the area where they live or any other location of interest and see the heat island effect as a map and text. By monitoring the information, residents can take decision about what to do in dangerous heat condition.

Section 03 Competency & Research Questions

As a result of climate change, there has been an increase in the frequency, intensity, and duration of heatwave events [5]. In response to the devastating mortality and health hazards of recent heatwave events, we introduced heat island warning systems. This system is designed to prevent and reduce intense heat exposure posed by heatwaves through a web page where users can search for a designated location and see the current and forecasted temperatures along with predicted heat index.

Before the development of the system, we came up with some research questions that can be answered during the research and development of the project. The research questions are as follows:

1. Can we integrate local weather station data and satellite data to create an application for residents of the borderland to monitor dangerous heat conditions caused by the Urban Heat Island Effect?
2. Can we integrate local weather station data and satellite data to identify areas vulnerable to the urban heat island effect?
3. Can we use the integrated data to create a map of the most vulnerable areas of El Paso and Juarez?

Section 04 Concept Map

Figure 2 illustrates the concept map that was crucial in the development of our project during the beginning stages. It is important to note that the concepts directly related to the project and collection-site data are the most clearly annotated. However, concepts related to sensor data are also included.

The concept map illustrated below describes datasets and their sources & properties. Further, the process and tools that we used for this project are also shown in the concept map

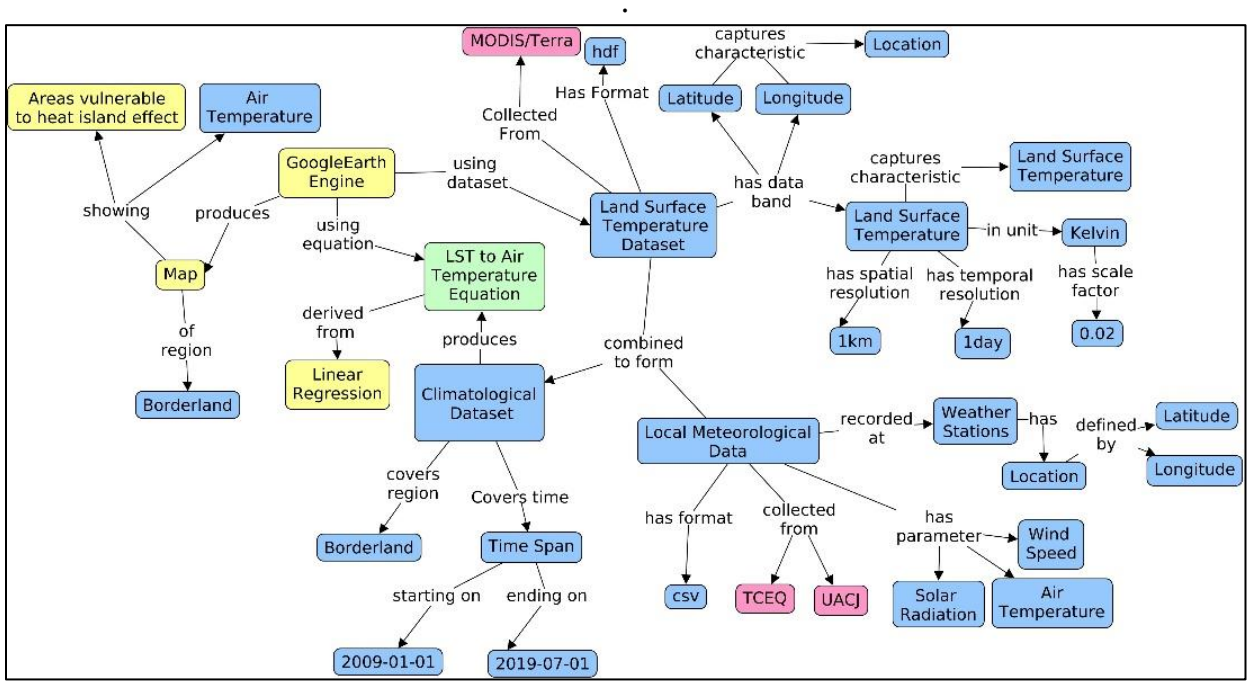


Figure 2: Concept Map of Heat Island Application

Section 05 Datasets

The section describes the data retrieval, processing, and manipulation tools that were used to collect related meteorological and satellite datasets. The full scenario with some questions related to these datasets are given below:

Table 1: Questions related data and its manipulation process

Category	Questions
Data-Related Questions	1. Where is the data?
	2. What is the format of the input data?
	3. How do you retrieve the data (i.e., API, web-service)?
	4. What is the data providing (i.e., characteristic, entity, units, temporal/spatial coverage)?
Manipulation/ Model/ Tools	5. What tool or application will you use to manipulate your data?
	6. What coverage/resolution is expected?
	7. Which algorithm/model will you be using? What are the assumptions of this model or tool? In which cases will it work and in which cases will it not?

Data Collection Areas

The data that we collected to determine the heat island effect of the El Paso-Juarez is as follows:

- Historical satellite data
  - Surface Temperature
- Observed Meteorological data
  - Air Temperature, Wind Speed, Humidity, etc.

Data Sources

We have collected those datasets from the following web resources.

- Earth Data – NASA
  - Need sign in to download the data
  - Website: https://earthdata.nasa.gov/
- Google Earth Engine
  - Need signup additionally to get the earth engine.
  - Website: https://code.earthengine.google.com/
- Dark Sky Net API
  - Need subscription to get the weather data
  - Website: https://darksky.net/
- Texas Air Monitoring Information System (TAMIS) Web Interface
  - Website: https://www17.tceq.texas.gov/tamis/index.cfm?fuseaction=report.main
- The Autonomous University of Ciudad Juarez (UACJ) Data Repository
  - Need to request direct access for Juarez weather station data

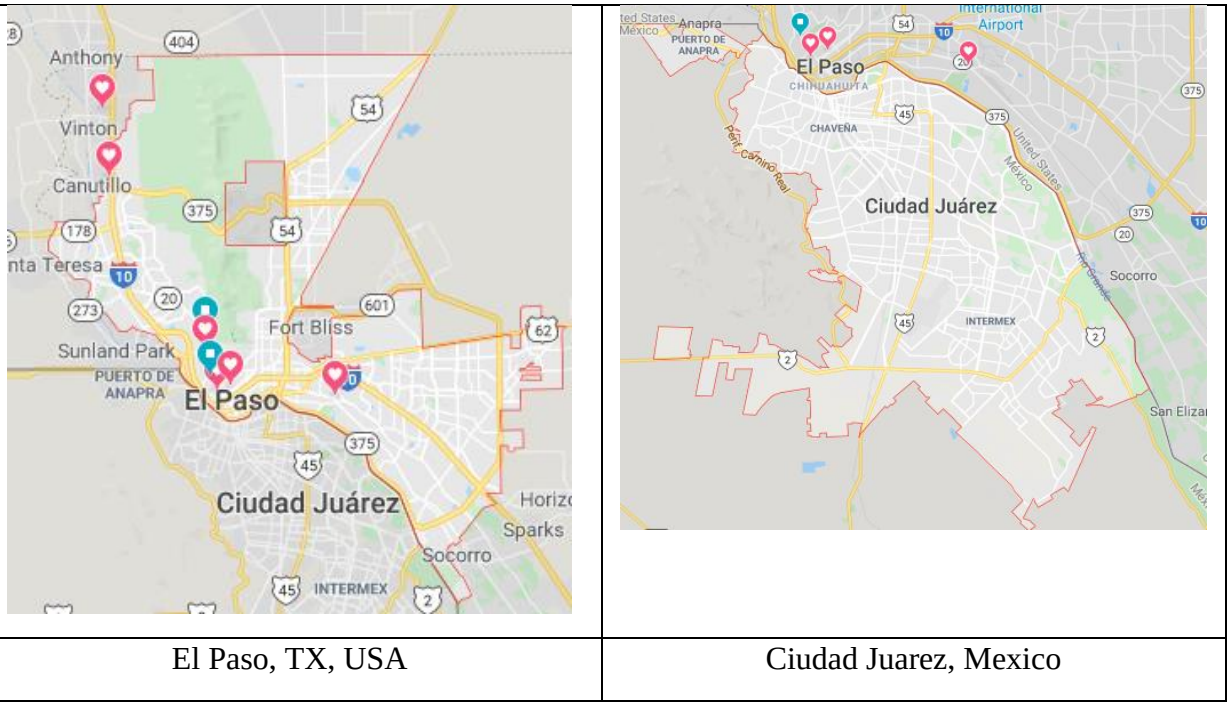
- Website: http://erecursos.uacj.mx/handle/20.500.11961/4201

This will answer question 1.

Data Coverage

We began the research with some selected areas with borderland. The areas we are working on to show heat island effect are listed below:

1. El Paso, TX, USA
    - Latitude: 31.7618778,
    - Longitude: -106.4850217
2. Ciudad Juárez, Chihuahua, Mexico
    - Latitude: 31.6903638
    - Longitude: -106.4245478



El Paso, TX, USA	Ciudad Juarez, Mexico

Figure 3: Selected cities for deriving heat index

Data Format & Resolution

To answer question 2 & 4 in Table 1, we discuss the format and resolution of each dataset.

Historical Surface Temperature: The format of the surface temperature is in HDF (image format). The details of the dataset are given below:

- Name: MOD11A1.006 Terra Land Surface Temperature and Emissivity Daily Global 1 KM.
- Image Collection ID: MODIS/006/MOD11A1
- Resolution: 1000 meters

- Cadence: 1 day
- Dataset DOI: https://doi.org/10.5067/MODIS/MOD11A1.006

Table 2: Bands Information

Name	Units	Min	Max	Scale	Description
LST_Day_1km	Kelvin	7500	65535	0.02	Daytime Land Surface Temperature

This dataset is only covered for land. A sample view of the land surface temperature is given below.

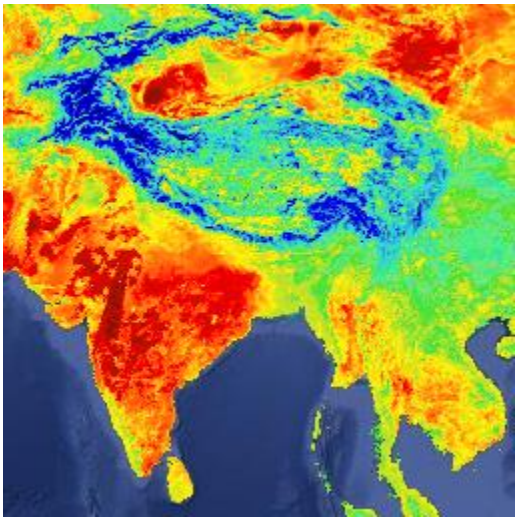


Figure 4: Sample Surface temperature Image

Historical Air Temperature: The format of the air temperature is in CSV format. The details of the dataset are given below

- Target List: Metrological Parameters (16 parameters)
- Sample Parameters: Temperature, Solar Radiation, Wind Direction, etc.
- City: El Paso, TX, USA
- Date Range: 2018-2019
- Duration: Every 1 Hour

Table 3: Sample Dataset of historical air temperature

AQS Code	Site Name	Date & Time	Parameter	Temperature
481410058	Skyline Park	2019-01-01 00:00	Temperature	35.9198
481410058	Skyline Park	2019-01-01 01:00	Temperature	36.7238
481410058	Skyline Park	2019-01-01 02:00	Temperature	35.1511
….	….	….	….	….

….	….	….	….	….
481410058	Skyline Park	2019-08-31 21:00	Temperature	32.4378
481410058	Skyline Park	2019-08-31 22:00	Temperature	31.1324
481410058	Skyline Park	2019-08-31 23:00	Temperature	31.4336

Current Weather Data: The format of the weather data us in JSON format. Sample JSON output data for El Paso weather is given below.


```
{
  "latitude": 31.7618778,
  "longitude": -106.4850217,
  "timezone": "America/Denver",
  "currently": {
    "time": 1575881567,
    "summary": "Drizzle",
    "icon": "rain",
    "nearestStormDistance": 0,
    "precipIntensity": 0.005,
    "precipIntensityError": 0.002,
    "precipProbability": 1,
    "precipType": "rain",
    "temperature": 53.27,
    "apparentTemperature": 53.27,
    "dewPoint": 47.63,
    "humidity": 0.81,
    "pressure": 1013.3,
    "windSpeed": 6.16,
    "windGust": 9.05,
    "windBearing": 281,
    "cloudCover": 0.67,
    "uvIndex": 0,
    "visibility": 10,
    "ozone": 280.8
  },
```


Figure 5: Current weather data format

Used Tools

We use several tools to collect those data and their manipulations. In this section, we discuss those used tools and their usage.

Surface Temperature: To get this surface temperature data we use Earth Data Nasa and Google Earth Engine. To do image analysis we have collected the data using the Earth Data tool.

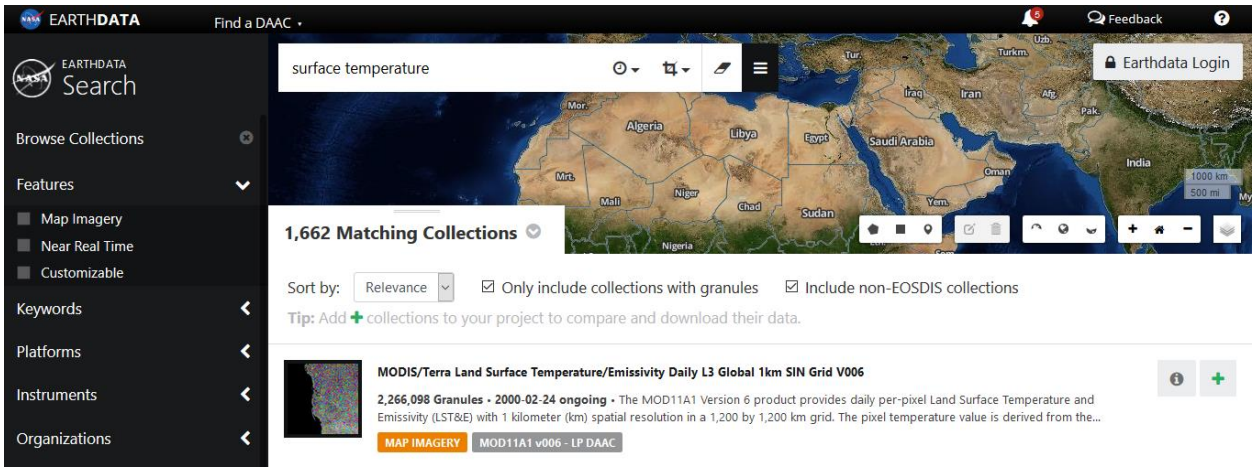


Figure 6: Earth Data Search Tool

Air Temperature Data: This dataset is collected from the Texas Commission on Environmental Quality (TCEQ) website. The website has a web interface to collect the data. They call it "Texas Air Monitoring Information System (TAMIS) Web Interface".

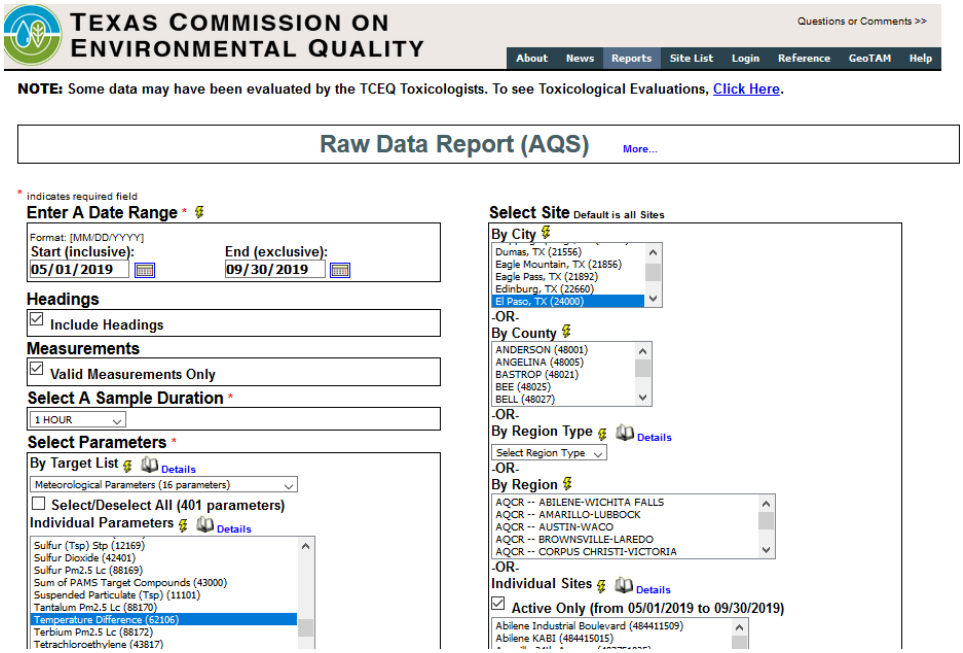


Figure 7: TAMIS Web Interface

Current Weather Data: This dataset is fetched using a web API from a weather company named Dark Sky. They provided current weather data as JSON format which is helpful to align with our project.

Section 06 Models & Interface

Our current model only considers the land surface temperature and time of the year when calculating the air temperature. There are however other, more accurate models that consider wind, net radiation, and albedo of the surface [7, 8, 9].

All the data manipulation is done using our Python code and Google Earth Engine. Python is used to filter the local weather station data and the *HDF* files, and to produce the combined climatological dataset. Our Python code is also used to perform the linear regression between the land surface temperature and the air temperature, generating an equation that calculates air temperature from land surface temperature. The code can be found in our git repository, which contains the code used, the local meteorological data, and a script to download the land surface temperature data [6].

Google Earth Engine is used to apply the equation determined by the linear regression to the land surface temperature dataset to generate both the historical and present-day vulnerability maps. To generate the historical maps, the land surface temperature data is first filtered by date to the summer months of the year. Next, the mean image of the filtered image collection is generated. This data in the image is transformed from land surface temperature to air temperatures by the equation generated from our Python code. This new image is clipped to the El Paso and Juarez region, and the air temperature values in the image are mapped to color with a palette.

Section 07 Results

In this section, we discuss the results that we produce in our project. We categorize the results in three major areas. These areas will be discussed in the following subsections.

Heat Index Visualization

This result of the heat index is visualized as text format. Residents can search location and see the current outside temperature with probable temperature increment/decrement. The increment/decrement temperature is analyzed from historical surface temperature and its relationship with air temperature.

The resulted heat index visualization can be checked online through the following link.

Link: https://www.smreza.com/projects/heat-island-effect/

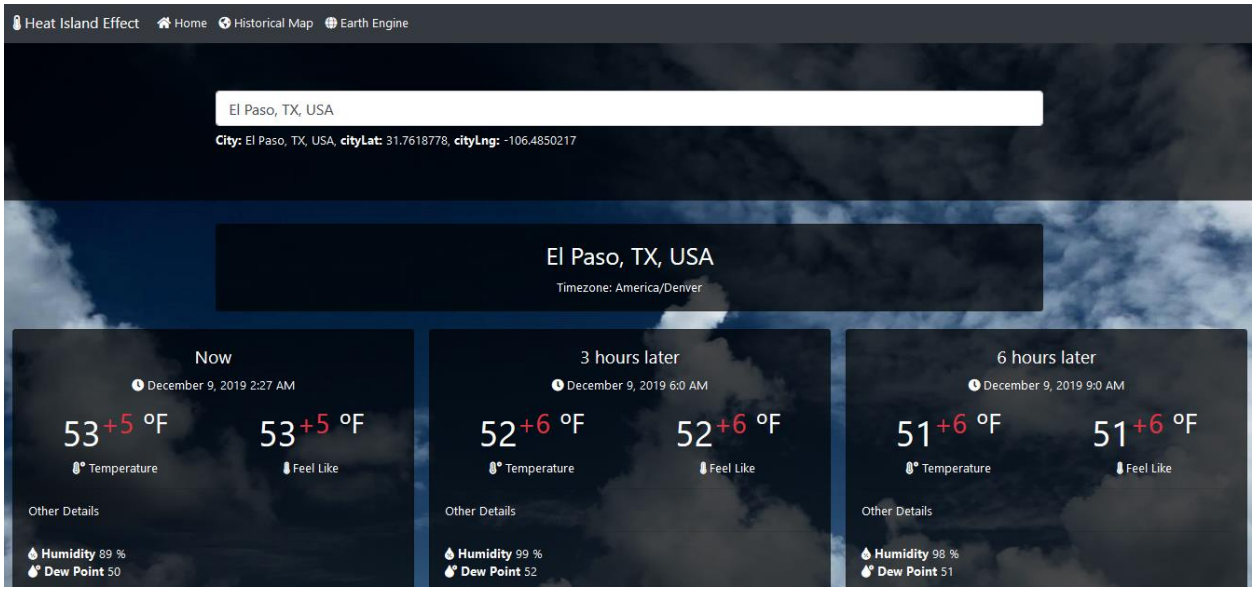


Figure 8: Heat Index Prediction by locations

Historical Heat Index Visualization

In our results, we integrated the historical predicted heat index after applying the regression model on the datasets of surface temperature and local air temperature. The resulted historical heat index map visualization can be checked online through the following link.

Link: http://www.smreza.com/projects/heat-island-effect/map.php

The results we produced are given below.

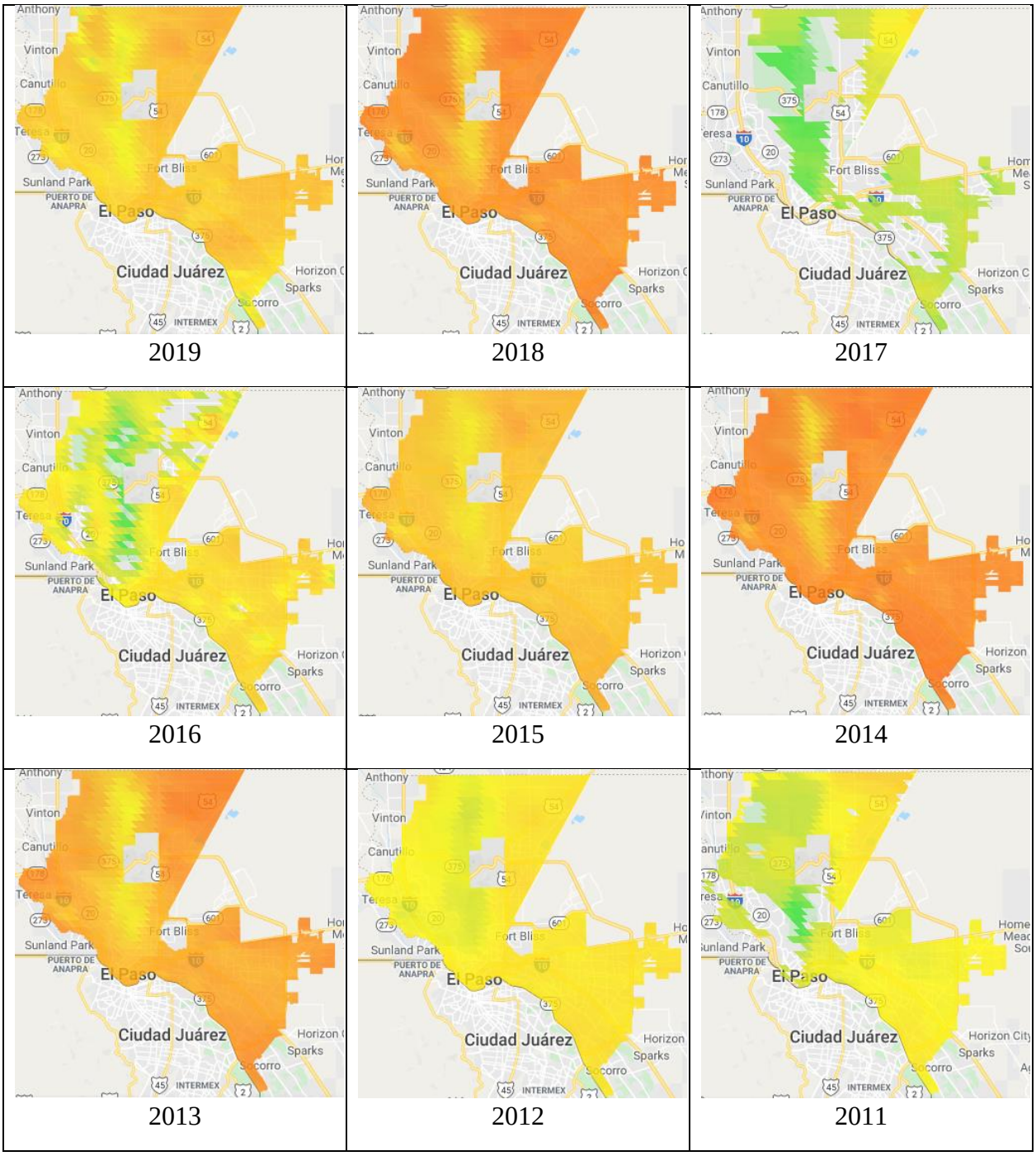


Figure 9: Historical vulnerability map in summertime (each year)

Integration of Google Earth Engine

In our results, we integrated google earth engine application that is developed in the earth engine and published as an application.

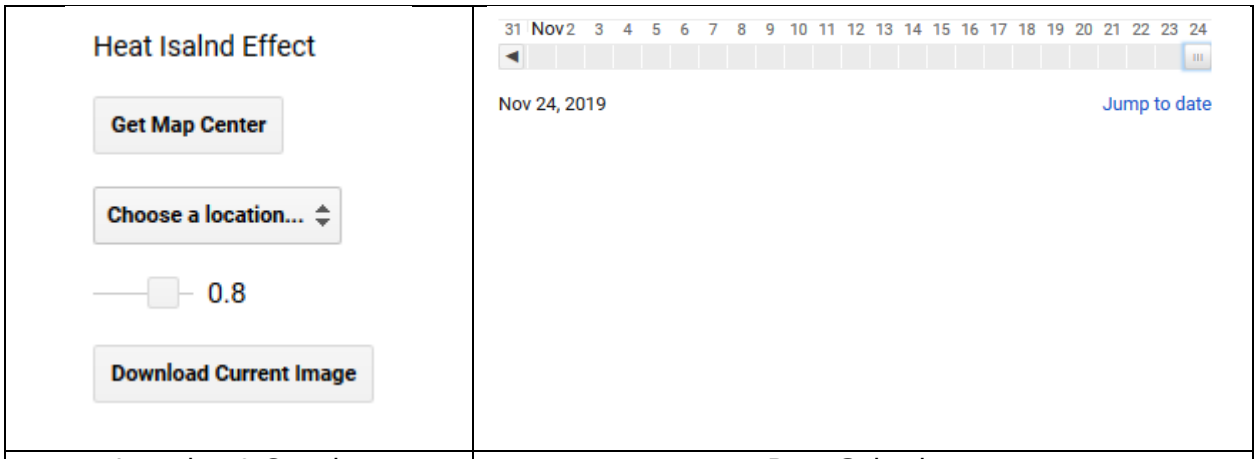


Figure 10: Widgets in our application

The resulted heat index map visualization can be checked online through the following link.

Link: https://www.smreza.com/projects/heat-island-effect/earth.php

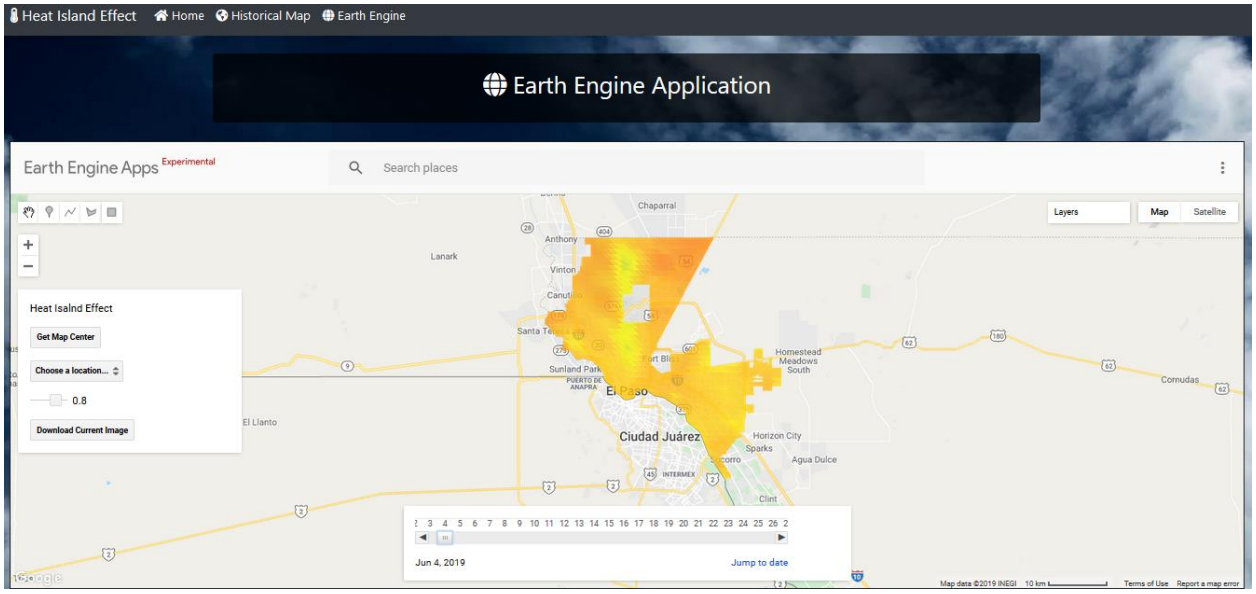


Figure 11: Integration of Google Earth Engine App in our project

Section 08 Conclusion & Future Work

Cities are often warmer than their suburbs because of a phenomenon called "the heat island effect." The way a city is designed — the building materials used, the way streets are arranged, the lack of canopy — can sequester heat. To measure the heat energy, researchers are analyzing different resource to understand the relationship to generate those heat. There are some alarming systems developed to visualize the heat island effect.

In our project, we integrate local weather station data and satellite data and make a web application for residents of the borderland to monitor dangerous heat conditions caused by the Urban Heat Island Effect. The results are visualized in both text and map. The results we produced shows that in the summertime, the temperature may grow in the selected cities in a significant way. Each day data visualization is made public to use to monitor dangerous heat conditions. The application is hosted online and can be accessed through the following link: http://www.smreza.com/projects/heat-island-effect/.

This project can be extended using different cities' data to understand more about the heat island effect. Our application framework is designed in such a way that a new developer can plug in more data for different cities, generate a predictive regression model for each city and plugin to our application.

Sections 09 Team Member's Contributions

In the development of our project, our team members did a good job of teamwork. The responsibilities are distributed among the team members. Some major contributions are given below.

Table 4: Contributions of team members

Deliverables	**Responsible**
Create a Web Application Framework	Sayed Mohsin Reza
Integrate Google Earth Engine in Web Application	Sayed Mohsin Reza
Visualize Historical Heat Index Map Images	Jake Lasley
The formula for predicting air temperature based on Surface Temperature	Jake Lasley
Temperature API integration	Sayed Mohsin Reza
Analysis of heat islands. Yield average difference in temperature between heat islands and non-heat islands for adjusting forecast	America Alvarez
Collection/Filtering/and integration of Data	Jake Lasley
Document Report	All
Poster Presentation	All

References

1. Mirzaei, P. A. (2015). Recent challenges in modeling of urban heat island. Sustainable cities and society, 19, 200-206.
2. Heat Island Effect, https://www.epa.gov/heat-islands.
3. Heaviside, C., Macintyre, H., & Vardoulakis, S. (2017). The urban heat island: implications for health in a changing environment. Current environmental health reports, 4(3), 296-305.
4. Bhargava, A., Lakmini, S., & Bhargava, S. (2017). Urban heat island effect: it's relevance in urban planning. J. Biodivers. Endanger. Species, 5, 5-187.
5. Lowe, D., Ebi, K. L., & Forsberg, B. (2011). Heatwave early warning systems and adaptation advice to reduce human health consequences of heatwaves. International journal of environmental research and public health, 8(12), 4623–4648. doi:10.3390/ijerph8124623
6. Code Repository, https://github.com/magiking/urban-heat-island
7. Liu, S., Su, H., Zhang, R., Tian, J., & Wang, W. (2016). Estimating the surface air temperature by remote sensing in northwest China using an improved advection-energy balance for air temperature model. Advances in meteorology, 11, 1-11.
8. Asaeda, T., Thanh Cu, V., & Wake, A. (1996). Atmospheric environment, 30(3), 413-427.
9. Wang, J., & Bras, R. L. (2011). A model of evapotranspiration based on the theory of maximum entropy production, Water resources research, 47, W03521, doi:10.1029/2010WR009392.

C2 Oral Communication: A technical poster presentation on project heat island effect

Urban Heat Island Effect in the Borderland

Sayed Mohsin Reza, Jake Lasley and America Alvarez*

Enlarged View Link: https://www.smreza.com/projects/heat-island-effect/poster.pdf

Urban Heat Island Effect in the Borderland

Cyber-ShARE Center of Excellence

Sayed Mohsin Reza[1], Jake Lasley[1] America Alvarez[2]

[1]Department of Computer Science, [2] Department of Geological Science

sreza3@miners.utep.edu, jlasley@miners.utep.edu, aalvarez26@miners.utep.edu

Problem Statement & Research Question

Problem Statement: Increased daytime temperatures, reduced nighttime cooling, and higher air pollution levels associated with urban heat islands can affect human health by contributing to general discomfort, respiratory difficulties, heat cramps and exhaustion, non-fatal heat stroke, and heat-related mortality.

Principal Research Question: Can we integrate local weather station data and satellite data to create an application for residents of the borderland to monitor dangerous heat conditions caused by the Urban Heat Island(UHI) Effect?

Background

The term "*heat island*" describes built up areas that are hotter than nearby rural areas. The annual mean air temperature of a city with 1 million people or more can be *1.8–5.4°F (1–3°C)* warmer than its surroundings. In the evening, the difference can be as high as *22°F (12°C)* [1, 2]. Heat islands can affect communities by increasing summertime peak energy demand, air conditioning costs, air pollution and greenhouse gas emissions, heat-related illness and water pollution [3].

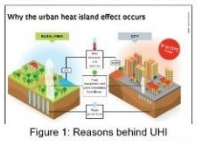


Figure 1: Reasons behind UHI

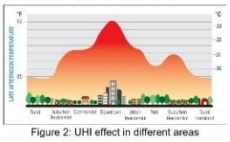


Figure 2: UHI effect in different areas

Project Scope

The scope of this project is to integrate the local metrological data and surface temperature data from satellite to analyze the heat island effect. In summary, the main purpose is to help residents living in borderland (*El Paso, TX, USA* and *Ciudad Juarez, Mexico*) to monitor the heat conditions caused by the Urban Heat Island Effect and visualize the vulnerability map of the borderland areas.

Concept Map & Data Sources

- Earth Data – NASA
- Google Earth Engine
- Dark Sky Net API
- TAMIS Web Interface
- UACJ Web Interface

Figure 3: Data Sources

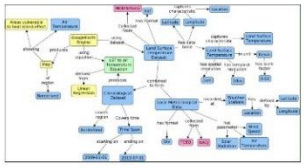
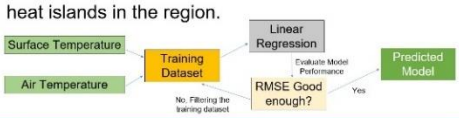


Figure 4:Concept Map

Methodology

Historical satellite data and observed meteorological data were collected and filtered for the El Paso-Juarez region during summertime months (May-September) from 2002-2019. A formula was determined based on scientific literature to visualize historical vulnerability maps of the urban heat island effect and yield a temperature difference between heat islands and non-heat islands in the region.



Results

Figure 5: Heat Index Prediction by locations

Figure 6: Google Earth Engine Application

Figure 7: Historical vulnerability map in summertime (each year)

Conclusion

Cities are often warmer than their suburb and rural areas because of the "heat island effect" phenomena. Our project serves as a monitoring system for residents of the El Paso-Juarez metroplex and aims to identify intense heat exposure. This project can be extended using data of different cities to further assess the heat island effect, especially in desert cities and arid regions.

References

[1] Mirzaei, P. A. (2015). Recent challenges in modeling of urban heat island. Sustainable cities and society, 19, 200-206.

[2] Heat Island Effect, https://www.epa.gov/heat-islands.

[3] Heaviside, C., Macintyre, H., & Vardoulakis, S. (2017). The urban heat island: implications for health in a changing environment. Current environmental health reports, 4(3), 296-305.

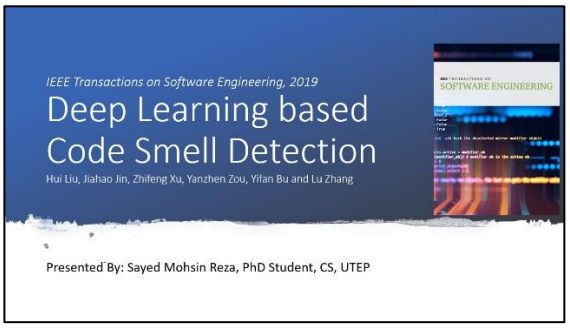
This material is based on the course project for the Web-based Data Integration with Geological Science course.

C3 Oral Communication: A seminar lecture on an IEEE transaction paper

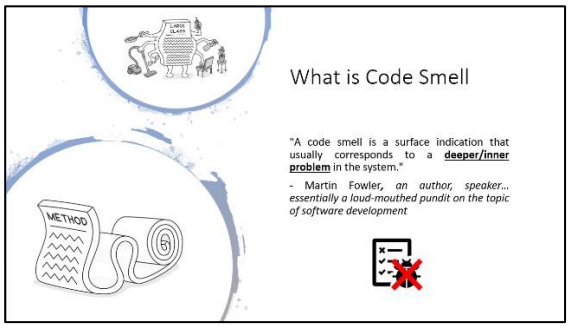
Deep Learning-based Code Smell Detection

Presented By: Sayed Mohsin Reza, PhD Student, CS, UTEP

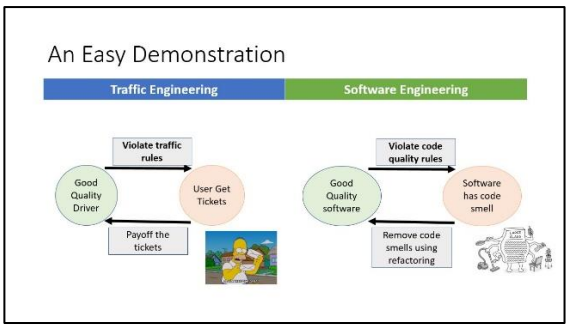
Full Presentation Link: https://www.smreza.com/web/index.php/research/code-smell-detection/



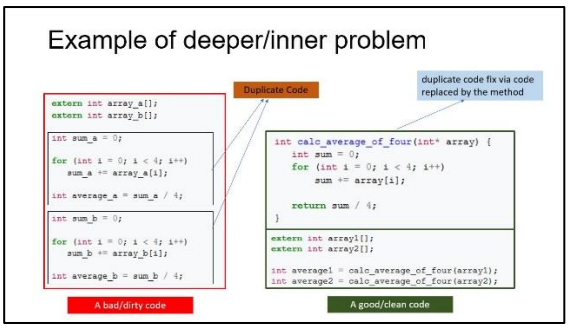
Slide 1



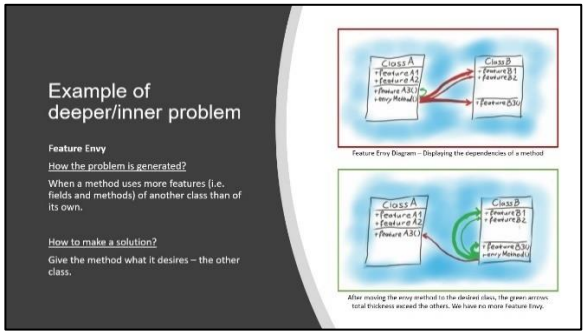
Slide 2



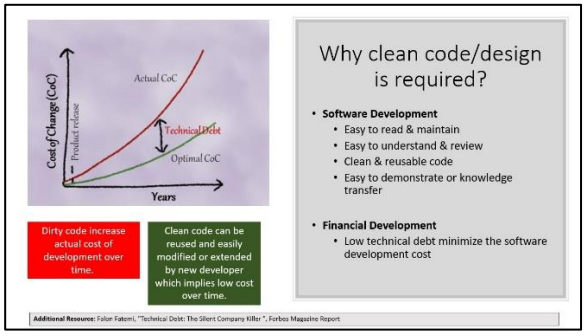
Slide 3



Slide 4



Slide 5



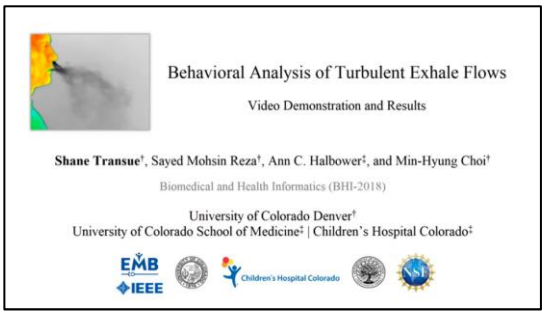
Slide 6

C4 Oral Communication: A technical oral presentation in BHI conference

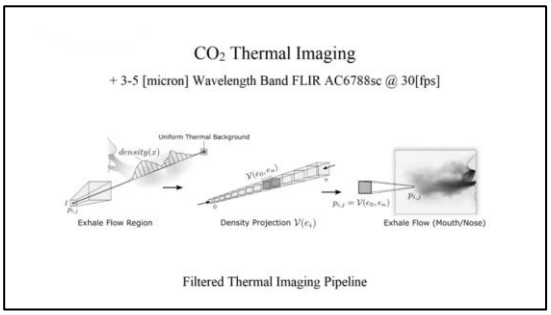
Behavioral Analysis of Turbulent Exhale Flows

Shane Transue, Sayed Mohsin Reza, Ann C. Halbower, Min-Hyung Choi*

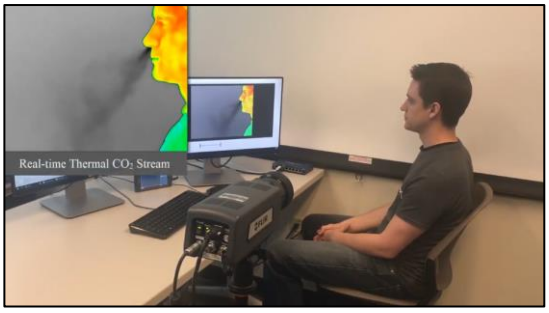
Full Video is available in https://www.youtube.com/watch?v=6052zUSVS4w



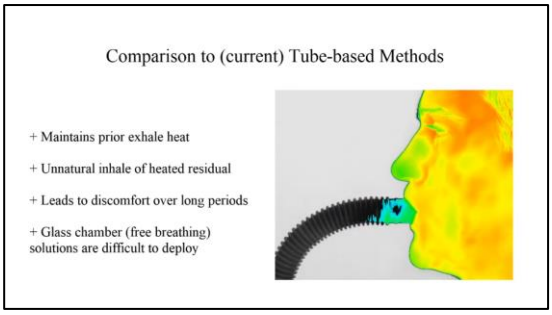
Slide 1



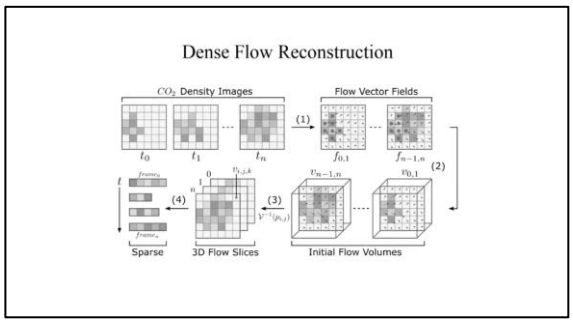
Slide 2



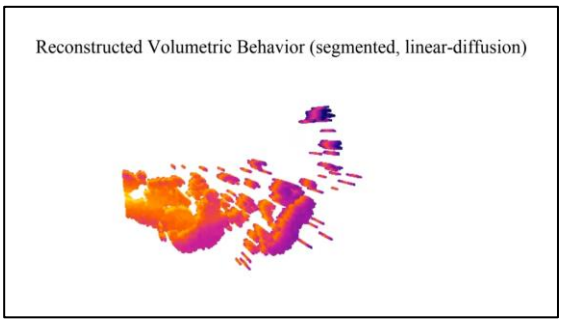
Slide 3



Slide 4



Slide 5



Slide 6

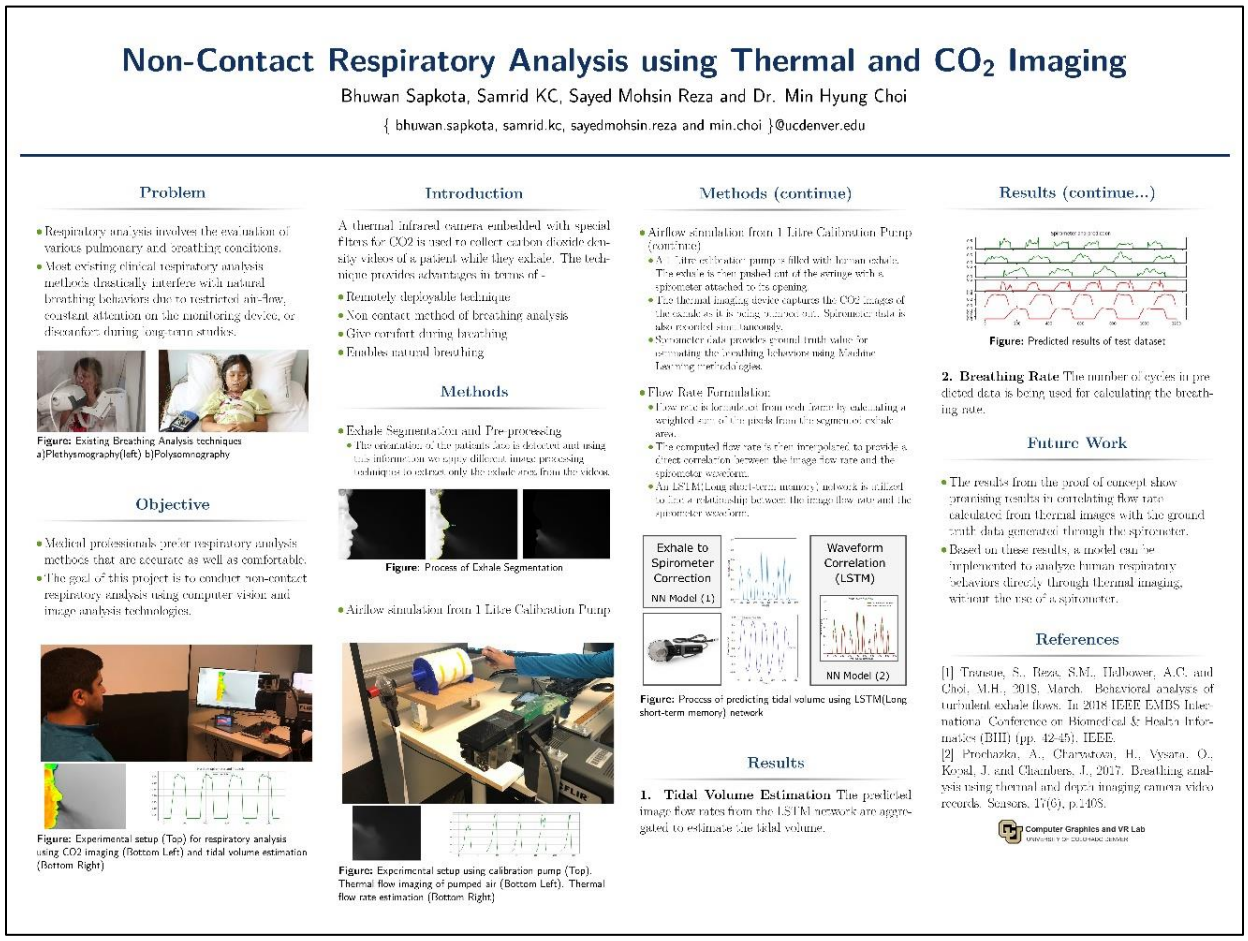
C5 Oral Communication: A poster presentation in RaCAS symposium

Non-Contact Respiratory Analysis using Thermal and CO2 Imaging

Bhuwan Sapkota, Samrid KC, Sayed Mohsin Reza and Dr. Min Hyung Choi*

Enlarged view link: https://www.smreza.com/web/index.php/rememberance/racas-2019/

Summary: I Joined the 2019 RaCAS (Research and Creative Activities Symposium) at the University of Colorado Denver, USA in April 2019 and presented our research on Non-Contact Respiratory Analysis using Thermal and CO2 Imaging.



APPENDIX D: RESEARCH

Paper published in 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)

DOI: 10.1109/BHI.2018.8333365, **Publisher:** IEEE
IEEE PDF Link: https://ieeexplore.ieee.org/abstract/document/8333365

2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)
4-7 March 2018
Las Vegas, Nevada, USA

Behavioral Analysis of Turbulent Exhale Flows

Shane Transue[†], Sayed Mohsin Reza[†], Ann C. Halbower[‡], and Min-Hyung Choi[†]

***Abstract*— Dense exhale flow through CO_2 spectral imaging introduces a pivotal trajectory within non-contact respiratory analysis that consolidates several pulmonary evaluations into a single coherent monitoring process. Due to technical limitations and the limited exploration of respiratory analysis through this non-contact technique, this method has not been fully utilized to extract high-level respiratory behaviors through turbulent exhale analysis. In this work, we present a structural foundation for respiratory analysis of turbulent exhale flows through the visualization of dense CO_2 density distributions using precisely refined thermal imaging device to target high-resolution respiratory modeling. We achieve spatial and temporal high-resolution flow reconstructions through the cooperative development of a thermal camera dedicated to respiratory analysis to drastically improve the precision of current exhale imaging methods. We then model turbulent exhale behaviors using a heuristic volumetric flow reconstruction process to generate sparse flow exhale models. Together these contributions allow us to target the acquisition of numerous respiratory behaviors including, breathing rate, exhale strength and capacity, towards insights into lung functionality and tidal volume estimation.**

I. INTRODUCTION

Accurate non-contact respiratory analysis has recently gained popularity within the domains of wireless signal processing [8] and computer vision [12] to automate and significantly broaden the class of quantitative respiratory metrics that non-contact methods can reliably address. Numerous techniques exist for both contact and non-contact respiratory analysis [4], however all of these methods *indirectly* infer breathing behaviors or utilize correlation functions for respiratory analysis. Techniques within computer vision have introduced thermal infrared cameras with spectral filters for CO_2 imaging for respiratory analysis [2], however the applicability of these techniques to comprehensive respiratory analysis is severely underdeveloped and the adoption of these methods has been very limited. This is due to three primary factors shared between most prior vision-based techniques: (1) prior objectives only emphasize simple quantitative measures such as respiratory rate [3] within limited Regions of Interest (RoI) and strength [6], limiting potential high-level behavioral analysis, (2) prior devices lack the sensitivity required to monitor subtle density variances and complex flows behaviors for identifying respiratory conditions, and (3) frame-rate limitations inhibit the ability to accurately capture

[†] This work is partially supported by the DoEd GAANN Fellowship: P200A150283 and NSF Grant: 1602428. S. Transue, S. M. Reza, and M-H. Choi are with the University of Colorado Denver Department of Computer Science and the Comcast Media and Technology Center. Emails: {shane.transue, sayedmohsin.reza, min.choi}@ucdenver.edu
[‡] A. C. Halbower is an MD of Pulmonary Medicine at the University of Colorado School of Medicine and with Children's Hospital Colorado. Email: ann.halbower@childrenscolorado.org

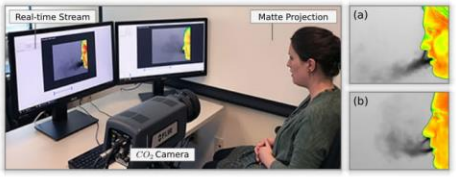


Fig. 1. Dense exhale flow analysis through optimized CO_2 imaging for illustrating unique respiratory behaviors of multiple individuals (a-b).

rapid and turbulent respiratory behaviors. To develop a device for directly analyzing turbulent CO_2 exhale flows [3], we have coordinated the development of a *hyper*-sensitive FLIR thermal camera that contains an embedded spectral filter that directly targets the CO_2 spectral band (3-5[μm]). From our requirement specification[1], the device provides raw CO_2 *count images* that contain the infrared wavelength activation counts within the CO_2 absorption band [10], [11]. Through the development these imaging methods and our *direct* measurements of breathing behavior, we introduce a new vector in vision-based clinical respiratory analysis. This includes direct flow and thermal analysis for subtle alternations in airflow related to asthma, Chronic Obstructive Pulmonary Disease (COPD), developmental conditions related to nose and mouth breathing distributions, cognitive function [13], sleep apnea, and Sudden Infant Death Syndrome (SIDS).

II. RELATED WORK

The primary evaluation criteria within respiratory analysis revolves around the collection of a limited set of quantitative metrics such as breathing rate, flow analysis, and tidal volume estimates. Extensive research has culminated numerous *contact* [4] and *non-contact* [8], [12] methods that obtain these metrics with promising levels of accuracy. However, based on these existing methods, all current respiratory evaluation is performed using *indirect* methods, that is, they infer measurements through secondary signals such as visible chest movements, vibration, pressure, acceleration, or sound [4]. Prior methods using spectral analysis for CO_2 visualization [3] have introduced a direct means of evaluating respiratory behaviors using *direct* exhale measurements for breathing rate. Similar to this form of direct analysis, we measure and model the exhale flow consisting of the visualized thermal signature of the CO_2 waveform. While prior methods only provide a breathing rate evaluation, this form of visualization is underutilized due to its ability to generate

[1]FLIR A6788sc InSb CCF, 640x512 resolution CO_2 count images @ 30-120[fps] with programmatic camera control and raw data acquisition.

978-1-5386-2405-0/18/$31.00 ©2018 IEEE

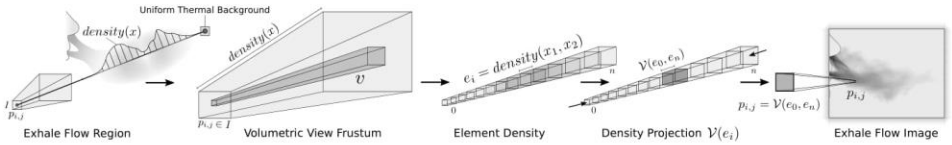


Fig. 2. Volumetric and density modeling of the CO_2 exhale region defined by the view frustum of the imaging device. The exhale flow region contains a non-linear CO_2 concentration distribution function $density$ over distance x. Each individual pixel $p_{i,j}$ represents a continuous volume through which the exhale flows. Based on the density value evaluated at each element e_i, the final value of pixel $p_{i,j}$ is the projection of all element densities within v.

numerous additional metrics such as nose/mouth distribution, velocity, dissipation, behavioral characteristics and even insight into lung efficiency in controlled environments. To enable these metrics, the result of our work significantly diverges from existing CO_2-based imaging techniques [2], [3] in both the level of analysis and the resolution of our modeling, as shown in Figure 3 (b), as compared with the prior result introduced within [3] shown in Figure 3 (a).

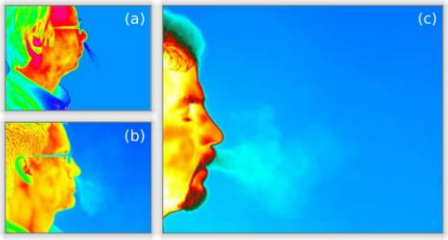


Fig. 3. Thermal respiratory CO_2 imaging: (a) results from [3], compared with our direct reconstruction in of the experimental setup in (b), with a detailed example of our exhale visualization result in (c).

Evaluating objective reconstructions of vortex behaviors within turbulent flows is an open problem within computational fluid dynamics due to reference frame dependent flow behaviors. Recent flow visualization techniques have proposed visualizations of vortex behavior through optimally local reference frame optimization [5] for reconstructing complex vortex flows. Other flow reconstruction techniques attempt to build models of complex gas flows through refraction-based Background-oriented Schlieren methods [9] or through Light-Path approximations captured through multiple imaging devices using visible light wavelength [7]. These techniques represent a divergence from traditional tracer-based method such as Particle Image Velocimetry (PIV) that requires discrete cross-correlation of discrete tracer particles which is impractical for our clinical domain.

Due to the limitations of the clinical domain, these tracer and light-based, fixed environmental constraints are impractical, therefore we introduce a heuristic-based model for analyzing turbulent respiratory exhale flows. Based on this contribution, we build a framework that: (1) provides an accurate turbulent exhale flow analysis, (2) model direct flow behaviors are required for identifying potential respiratory conditions, and (3) introduce a new methodology for identifying condition-trait signatures using clinical non-contact respiratory analysis with an aim of associating exhale flow behaviors with common pulmonary conditions and diseases.

III. Method

Exhale flow behavior modeling provides a basis for evaluating high-level respiratory characteristics based on a set of observable phenomena that is not facilitated by current monitoring techniques. This includes momentary fluctuations within exhale streams that inherently contribute to secondary flow behaviors associated with obstructed breathing, subtle changes between nose-mouth breathing distributions, lung functionality, and the ability to identify abnormal exhale CO_2 signatures. To extend respiratory analysis to include these metrics, we introduce a dense flow reconstruction process including: (1) flow estimation through dense optical flow, (2) heuristic-based flow slice extrapolation, and (3) provide a volumetric sparse scalar field representation of recorded exhale behaviors for extended monitoring periods.

A. Dense Exhale Modeling

Carbon dioxide density images obtained through our camera are characterized by the projection of volumetric densities of the observable gas flows with general infrared radiation, filtered to the spectral wavelength interval required for CO_2 imaging. To maximize clarity in this measurement, we improve the sensitivity of our recording model by adding a thin matte surface parallel to the imaging plane that contains a uniform heat distribution. Through the view frustum of the camera shown in Figure 2, we model the continuous volume v of pixel $p_{i,j}$ from this surface to the image plane I as a discrete set of n elements with an unknown density distribution function $density$ as a function of distance x. This per-element density function $V(e_i)$ is projected to pixel $p_{i,j}$ resulting in an irreversible loss of this distribution. Our model is based on heuristic approximations of the inverse $\mathcal{V}^{-1}(e_i)$ of this volumetric projection to determine the per-frame scalar density value $v_{i,j,k}$ within a sparse voxel grid:

$$v_{i,j,k} = \mathcal{V}^{-1}(p_{i,j}) = \mathcal{V}^{-1}\Big[\sum_{i=1}^{n} density(e_i)\Big] \quad \forall p \in I \quad (1)$$

Since the consolidation of this density volume to an image representation is *unrecoverable* due to the projection of the per-element densities, the reconstruction is inherently limited to an approximation of the original volume but preserves overall flow behavior. Figure 4 illustrates the resulting intensity images of the exhale obtained through this method, demonstrating multiple dense turbulent flows. Due to this result, we do not limit our analysis to a sub-image RoI, rather we consider the entire exhale region to model both behavioral characteristics and diffusion properties of exhale sequences to build a per-patient respiratory profile.

Fig. 4. Recorded turbulent exhale flow from the mouth (left), nose (center), and both nose and mouth simultaneously (right). Through our imaging process, we obtain an accurate illustration of the CO_2 density distribution and flow behavior with minimal background interference.

B. Dense Flow Reconstruction

Dense flow reconstruction from two-dimensional imaging is inherently ambiguous and cannot be directly recovered. To approximate a resulting distribution of the flow within a reconstruction we employ a four step process for estimating exhale density flow behaviors based on consecutive CO_2 image pairs over time and space, outlined in Figure 5.

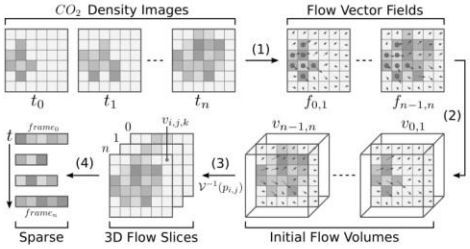


Fig. 5. Exhale flow reconstruction process. We approximate the reconstruction of the projected density volume by estimating the function $density(x)$ using heuristic approximations.

In this process we collect the set of n density images over time t, (1) compute the apparent flow through dense optical flow [1], (2) emplace these flow frames into a volumetric voxel grid as a seed slice in the middle of this volume, (3) extrapolate slice flow estimates, and (4) convert these scalar fields into sparse representations for each frame. The sparse representation is due to the dense resolution of the volumetric grid, which encodes the density value, and flow vector of each cell. We evaluate each frame independently within this single compute volume. This results in an n frame recording, each composed of a sparse 3D scalar field that approximates disjoint flow behaviors recorded in each frame.

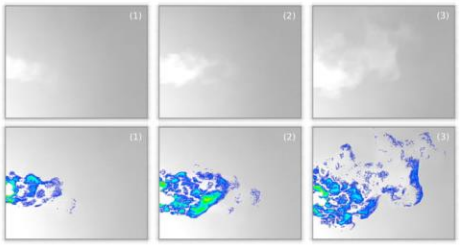


Fig. 6. Turbulent exhale optical flow vectors. The generated vector field illustrates the *apparent* flow computed through a standard dense optical flow algorithm. The (top) row illustrates the original CO_2 density images, and the (bottom) row illustrates the resulting vector norm-color-mapped flow.

IV. Experimental Results

The primary objective of flow-based respiratory analysis is to detect abnormalities within normal breathing based on the cross-sectional view of the exhale without the interference of background sources of infrared radiation. Since the objective of our reconstruction is to identify minute changes in the turbulent flow of an individual's exhale, we assume ideal posture and exhale region with a projection screen matte background. To focus the region to the exhale volume, we refine our region of interest to exclude the face, but do not limit our analysis to a region of interest. In Figure 9 we present two six-frame segmented exhale flow sequences.

A. Behavioral Analysis of Exhale Traits

The results of the proposed method present both flow visualization and approximated volumetric flow reconstruction to evaluate exhale strength and velocity shown in Figure 7, nose-mouth distribution, strength, and minute flow variances that can be associated with abnormal breathing. These contributions will greatly broaden the horizon of exhale analysis over existing frequency-based methods that utilize Fourier transforms to directly compute breathing rate.

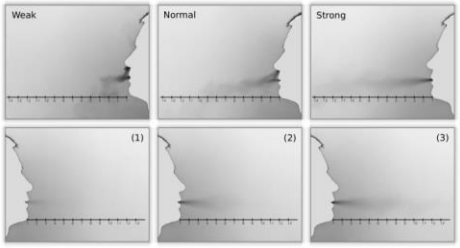


Fig. 7. Resulting videos for weak, normal, and strong exhale strengths (top row). In exhale velocity, we differentiate per frame for slow, normal, and fast velocities (bottom row), as a function of the exhale speed, CO_2 concentration, and the environmental thermal dissipation rate.

Quantitatively, we express the relative nose-mouth distribution as an estimated contribution to the exhale. For the velocity component , we express the change in exhale length over time as a result of exhaled CO_2 concentration.

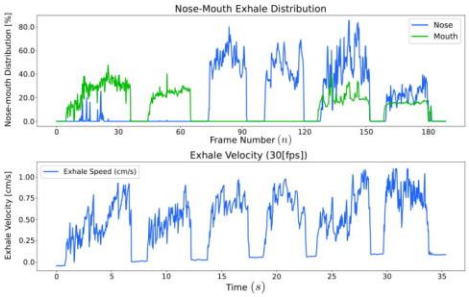


Fig. 8. Exhale nose-mouth distribution (top) for two exhales in each configuration: mouth, nose, and simultaneous oral-nasal breathing. The exhale velocity plot (bottom) demonstrates exhale characteristics for distance, strength, and the dissipation factor linked to CO_2 concentration.

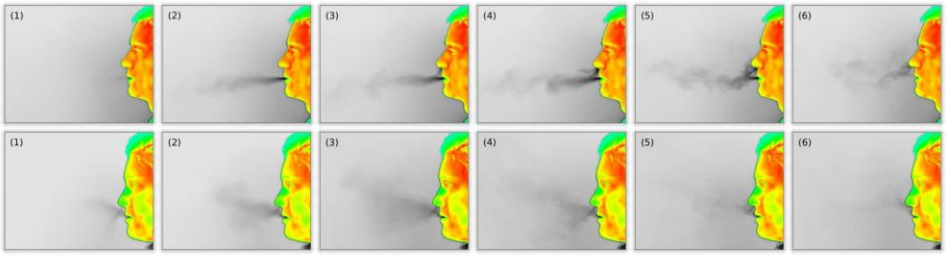


Fig. 9. Resulting CO_2 density distribution images illustrating unique respiratory patterns between individuals (top vs bottom rows). For each image sequence, one exhale period has been recorded and visualized, showing the clear separation between the nose-mouth distribution and density flow behaviors. These flow behaviors unique to each individual are subject to their own physiology and can be evaluated to identify per-individual exhale traits.

B. Flow Reconstruction

The extrapolation of our flow analysis into our 3D sparse scalar field is segmented by the optical flow vector norms. This cleanly separates the exhale densities from the residual general interference from background sources and natural environmental airflow. Using this technique, we can completely reconstruct segmented exhale behaviors. The resulting reconstruction is rendered as volumetric clouds within Figure 10.

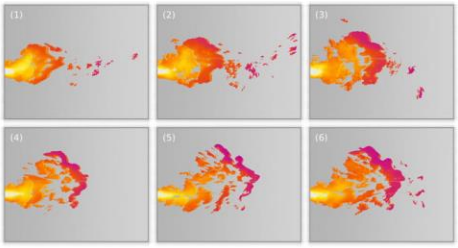


Fig. 10. Reconstructed 3D sparse exhale flow sequence rendered as a color-mapped sparse cloud (from top-left to bottom-right). Our framework incorporates interchangeable heuristics for extrapolating 3D flow fields.

V. Evaluation and Discussion

Turbulent flow analysis is inherently complex. Differentiating intersecting flows further complicates this analysis for nose/mouth separation. Accurate 3D reconstruction of volumetric flows without depth measurements or advection is also intangible. To preserve validity of our model, we focus on maintaining flow behavior characteristics, but do not formulate an exact volumetric exhale reconstruction. Rather, this form of CO_2 exhale imaging opens up a new medical significance in the monitoring of normal versus pathological airflow from the lower and upper airways. Volume measurements quantify lung capacity or effect of therapy in pulmonary patients with common medical problems such as asthma and chronic obstructive pulmonary disease without the need to utilize bulky pulmonary function machines with the required patient cooperation. Subtle alterations of airflow velocity and nose-mouth distribution can determine upper airway obstruction or those at risk for sleep apnea or sudden infant death. With the ability to monitor subtle changes in airflow in infants sleeping prone versus supine or in car seats or soft bedding, the advancement of research into the cause and prevention of SIDS cases achieves a new trajectory in high-resolution CO_2 respiratory behavioral analysis.

VI. Conclusion

In this work we have coordinated the development of a CO_2 imaging and respiratory monitoring framework that facilitates the recording and modeling of dense exhale flow behaviors. This non-contact, direct method establishes the foundation for a respiratory monitoring solution that includes new metrics for: rate, strength, nose/mouth distribution, condition trait identification, and volumetric modeling. This framework will be used to establish extensive patient studies utilizing diverse exhale quantitative metrics as future work.

References

[1] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *Intl. Journal on Comp. Vision*, 12(1):43–77, 1994.

[2] J. Fei and I. Pavlidis. Analysis of breathing air flow patterns in thermal imaging. In *Engineering in Medicine and Biology Society, 2006. EMBS'06*, pages 946–952. IEEE, 2006.

[3] J. Fei, Z. Zhu, and I. Pavlidis. Imaging breathing rate in the co_2 absorption band. In *Engineering in Medicine and Biology*, 2005.

[4] A. R. Fekr et al. Design and evaluation of an intelligent remote tidal volume variability monitoring system in e-health applications. In *IEEE J. of Biomedical and Health Informatics*, 2015.

[5] T. Günther, M. Gross, and H. Theisel. Generic objective vortices for flow visualization. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 36(4):141:1–141:11, 2017.

[6] D. Hanawa, T. Ohguchi, and K. Oguchi. Basic study on non-contact measurement of human oral breathing by using far infra-red imaging. In *Telecommunications and Signal Processing (TSP), 2016 39th International Conference on*, pages 681–684. IEEE, 2016.

[7] Y. Ji, J. Ye, and J. Yu. Reconstructing gas flows using light-path approximation. In *Computer Vision and Pattern Recognition*, 2013.

[8] P. Nguyen, S. Transue, M.-H. Choi, A. C. Halbower, and T. Vu. Wikispiro: Non-contact respiration volume monitoring during sleep. In *Proceedings of the Eighth Wireless of the Students, by the Students, and for the Students Workshop*, S3, pages 27–29, 2016.

[9] M. Raffel. Background-oriented schlieren (bos) techniques. *Experiments in Fluids*, 56(3), Mar 2015.

[10] R. Rehm et al. Dual-colour thermal imaging with inas/gasb superlattices in mid-wavelength infrared spectral range. *Electronics Letters*, 42(10):577–578, 2006.

[11] F. Rutz et al. Imaging detection of co2 using a bispectral type-ii superlattice infrared camera. In *Quant. InfraRed Thermo.*, 06 2012.

[12] S. Transue et al. Real-time tidal volume estimation using iso-surface reconstruction. In *IEEE CHASE'16*, pages 209–218, 2016.

[13] C. Zelano et al. Nasal respiration entrains human limbic oscillations and modulates cognitive function. *Neurosci*, 7(49), 2016.

Paper published in 2016 International Conference on Computer Communication and Informatics (ICCCI)

DOI: 10.1109/ICCCI.2016.7479974, **Publisher:** IEEE
IEEE PDF Link: https://ieeexplore.ieee.org/abstract/document/7479974

2016 International Conference on Computer Communication and Informatics (ICCCI -2016), Jan. 07 09, 2016, Coimbatore, INDIA

Low Cost and Portable Patient Monitoring System for e-Health Services in Bangladesh

Mithun Chandra Paul,Suman Sarkar,Md. Mahfujur Rahman,Sayed Mohsin Reza
Institute of Information Technology
Jahangirnagar University
Savar-1342,Dhaka
Email:{mithunpauliit,sumaniit02,mrrajuiit,smrezaiit}@gmail.com

Dr. M Shamim Kaiser
Associate Professor
Institute of Information Technology
Jahangirnagar University
Email:mskaiser@juniv.edu

***Abstract*—This paper proposes a low cost and portable patient monitoring system for e-Health services in Bangladesh. A raspberry pi microcomputer based system has been developed which can be used by paramedics for collecting different sensor data such as ECG signal, blood pressure signal, heart beat signal, Situation of Oxygen in Blood(SPO2), temperature and generating different signals from a patient and send these signals to specialist doctor who are in a centre or in a hospital. A web based application has been developed for both doctor and paramedics for efficient communicate with each other. It has been found that the system can be suitable for village health care centre of Bangladesh.**

***Index Terms*—ECG, e-health, raspberry pi, Situation of Oxygen in Blood(SPO2), blood pressure signal, raspbian etc.**

I. Introduction

The e-Health is an early arena in the connection of medical informatics, community health and business, mentioning to health facilities and information provided or improved through the internet and associated technologies [1]. In a wider sense, the term describes not only a technical expansion, but also to improve health care locally, county-wide, and worldwide by using information and communication technology. The e-Health covers Electronic Medical Records such as patient records, digital imaging, prescribing, internet based telemedicine and telecare services [2]. Itś also covers virtual genuineness, digital imaging, computer supported surgery, wearable and portable monitoring systems, health portals and decision making supporting for self-care.

For getting health care in the remote area is the critical fact because of poor financial infrastructure and communication problem in bangladesh. The specialist doctors are also not willing to go to the remote area. With the advent of the ICT specialised doctor can give their advises to the distant patient using e-Health system for diagnosis for proper treatment and prevention of disease. e-Health system can be beneficial for the rural patients through using telecommunication and information technologies. It helps people who leaving in a rural area to eliminate distance and can improve access to medical services that would not be often consistently available in rural communities. It is also the solution package for emergencies and critical solution [3].

e-Heath can be beneficial to patients living in isolated communities and remote regions, who can receive care from specialist doctor without having to travel to the doctors location or chamber. Recent developments in mobile collaboration technology can allow healthcare professionals in multiple locations to share information and discuss patient issues as if they were in the different place.

Bangladesh is a developing country. There are more than 68 thousand villages in Bangladesh, most of the peoples are living in these villages, the per capita income of this villagers are low compared to whose are working in urban city. Moreover the average per capita income is less that that of India and Pakistan [4] and the Road traffic communication is poor. Due to these lack of facilities, specialist doctors are not willing to go in thee villages. In Bangladesh, 70 percent people live in rural area. In rural areas about 40 percent of the population was count up as poor and 30 percent as extreme poor in 2014 by World Bank. They cańt get proper medical treatment lacks of services in the village. This paper has approached e-Health service by which rural people can easily make communication with the specialist doctor.

This paper describes a low cost portable device for patient monitoring system. Here the system has proposed to work about ECG signal, blood pressure signal, heart beat signal, Situation of Oxygen in Blood(SPO2), temperature and generating different signals using corresponding sensor to analyze the electrical activity of patient condition. Raspberry pi and a web application has also developed for deployment. The signal has been transmitted over web server to specialist doctor and displays the data in doctorś site on an organized Graphical User Interface (GUI). After getting the signal and available information, the doctor can give proper treatment for the patient.

The main goal of this paper is to introduce the difficulties related to urban areas health care. This project also provides a possible solution to overcome one of the main problems that comes collaboration of specialist doctor. Proposed model fulfil the requirements both the patient or remote paramedics and the specialist doctor. This portable device provides cheapest, fastest, home based health care services in the whole country especially elderly and remote areas.

II. Literature Review

Currently, the health care system is undergoing a cultural shift from a traditional approach to a modernized patient centered approach. In the traditional approach the health care professionals play the major role. They need to appointment the patients for essential diagnosis and instructing. There are two basic complications related with this approach. Firstly, the health care professionals must be on site of the patient all the time and secondly, the patient remains admitted in a hospital, wired to bedside biomedical instruments, for a period of time. There are various work on e-Health and research is now still going on the development of it.

Health Gear System is a real-time wearable system for observing and analyzing physiological signals. For the execution of Health Gear System using a blood oximeter to monitor the users blood oxygen level and pulse while sleeping [5]. Tracking Devices For Elderly Care System By Using GPS: At present, elderly people are surfing from dementia and Alzheimers disease across the world. Those elderly who are suffering from dementia and Alzheimers disease necessitate continuous attention. Hence it is possible to look after not only a single patient but also monitor a number of patients at the same time. Here they use RFID and GPS tracking system for the elder people to take care of them [6]. CareWell will primarily focus on the facility of care and provision to older people who have multifarious health and social care needs, are at high risk of hospital or care home admittance and require a range of high-level intermediations due to their weakness and several chronic diseases. Smart Personal Health is a European Commission provision action to endorse consciousness about problems and challenges of interoperability among personal health systems (PHS) and to other e-Health systems [7].

There have been many studies to offer a clarification to e-Health. They use GSM, satellite communication, RFID etc. Particularly, portable patient motoring system using raspberry pi is a new research arena that achieves low cost patient healthcare, which provide the potential for highly flexible medical services that are not possible with standard e-Health system.

III. Patient Monitoring System

This paper is based on client-server architecture [8]. The server application is responsible for storing data from patients and sending to the doctors sever. Doctors monitors and transmitting treatment to patient through Internet after getting those signal information from patient. Figure 1 shows the flow chart of proposed model. Proposed paper has discussed with four main components. These are:

- Sensors;
- Microcomputer;
- Database;
- Application;

A. Sensors

Proposed paper is based on low cost and portable patient monitoring system for Bangladesh. For signal acquisition form patient body, different sensors are used. The ECG Sensor model AD8232 is used for getting ECG signal [12], Pulsesensor for heart rate [13], RM2426 sensor [14] for blood pressure, AFE4400 Pulse Oximeter Shield Kit for SPO2 and DS18B20 sensor used for temperature sensor [15].

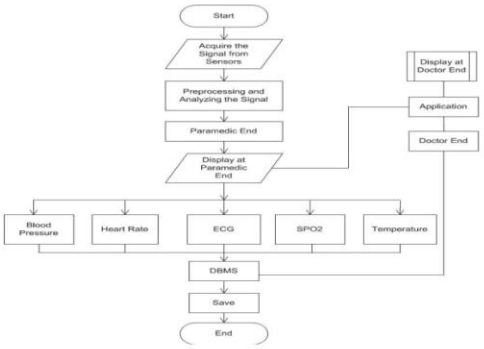


Fig. 1: Flow Chart of Proposed Model

B. Microcomputer

Raspberry Pi is a microcomputer which is used for receiving and analyzing signal from different sensors and sending those signal to specialist doctor. GPIO pins of Raspberry pi are used for receiving signal from different sensors [9]. The Raspberry Pi is a low cost, credit-card sized computer with monitor and uses a standard keyboard and mouse. It is accomplished with tiny device that assists people to explore computing. Proposed model has used raspbian operating system in raspberry pi with 16 gb memory card [10].

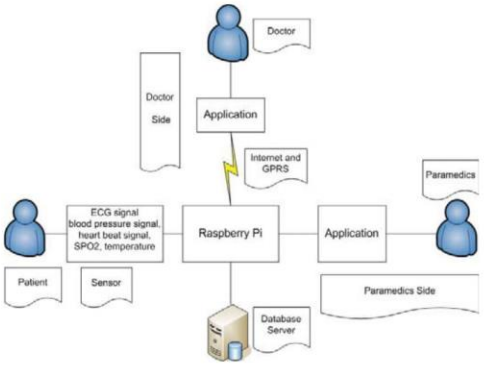


Fig. 2: Entire Scenario of the System

C. Database

A relational DBMS has been used for storing the patients and doctors information. DBMS is also used for storing the signal which acquires from different sensors. Doctors treatment are also available in the database. Paramedics and doctors can access the information whenever they will be needed.

D. Application

A web based application using ASP.Net is responsible for the client-server architecture which assists to communicate between paramedics and specialist doctors [11,17].

*Client Side:*Client-side allows interaction within paramedics and doctors. Paramedics are responsible for acquiring signal from patients body using different sensors. Those signals are transmitted over internet using GPRS to specialist doctors. A web application that is used for visualizing signals to patient or paramedics and the doctor [18,19]. The remote paramedics first access the application and input the patients information and attach the picture of both physiological and biological signal which has been acquired from the device. When paramedics will submit patients information with signal, he/she will get a patient pin or id number for access later. Here paramedics will add new patient information or edit an older patient information if needed. The specialist doctor can see all the information with signal picture. As a result, The doctor will give better advice and treatment for the patient by analyzing patient information.

Server Side: The Server is used for storing and making available incoming vital signals which originated from the clients. The server also has a database to store all the information both patient and the doctor. If the patient or doctor wants to communicate each other, they can easily get all the information from the data store. Basically, the server is developed with ASP.Net application and a relational database. The application compromises the following features: a) Patients information; b) Visualization of signals; c) Show sending patient information and treatment d) Show receiving patient information and treatment d) Prescribing Patient and chat conversation.

IV. Experimental Setup

Proposed model has used the raspberry pi microcomputer for acquiring the signal using different sensors from patient's body. A web application has been developed for sending those signals with required information to specialist doctor for getting better treatment. Here patients, paramedics and doctors are vital users for this proposed system. Figure 2 show entire scenario of the system.

Paramedics fill up all the patients information such as name, address, age, sex, phone, disease field in application. He can also attach a blood pressure signal, ECG signal, heart beat signal, Situation of Oxygen in Blood(SPO2), temperature signals which getting form sensors. After completing all the fields he will save it to the database and prepared for sending information to doctor. Doctor is responsible for analyzing the signal and giving treatment to patient(figure 3).

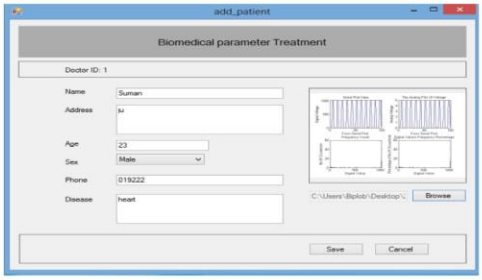


Fig. 3: Patient's Information

After filling up all patients information, then he would be able to assigning patient to a specialist doctor. He could select a corresponding doctor to a corresponding patient for getting better treatment(figure 4).

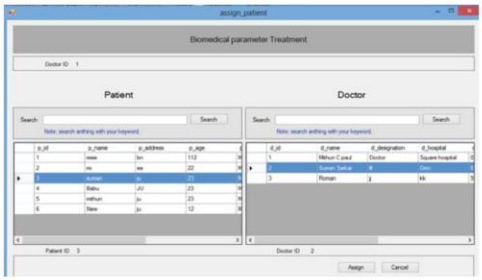


Fig. 4: Assign patient to Specialist Doctor

After seeing all the information about patient, doctor could prescribe to the patient and then paramedics could download the treatment file as .pdf , .doc or .xls(figure 5).

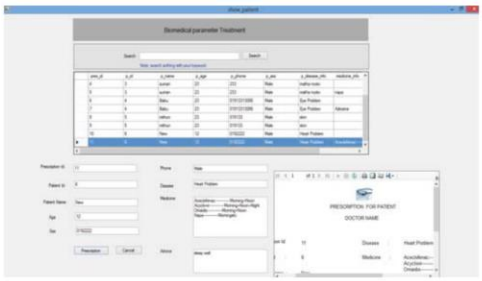


Fig. 5: Prescribing Patient by Specialist Doctor

V. Result Discussion

The system was implemented as previously described. As mentioned, both client and server side utilized ASP.NET programming language. The proposed system is meant to operate in an application by doctor and paramedics. The major task will be monitoring the patient health condition by checking on different signals and doctor could prescribe to the patient and then paramedics could download the treatment file as .pdf , .doc or .xls(figure 5).

Evaluating the result is an important issue of patient monitoring system. In order to evaluate the effectiveness , the proposed system works with different types data set. For example, ECG signal data set, blood pressure data set, heart beat data set, Situation of Oxygen in Blood(SPO2) data set, temperature data set. After getting data set proposed model preprocess the signal [16], analyze the signal and finally evaluate the result [figure-6].

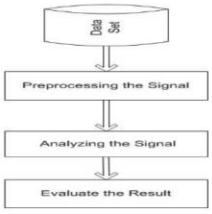


Fig. 6: Data Set and Result Evaluation

VI. Conclusion

In this paper a portable patient monitoring system for e-Health of Bangladesh has been discussed. As most of the people of Bangladesh are living in villages, and infrastructure development and imcome of these villages are low, the proposed portable patient monitoring system can be used by paramedics in the Health care center. The specialist doctors give them consultation from any part of the country using very good ICT infrastructure of the country. In future a doctor assisting system will be included along with the portable patient monitoring system. It moderates expenditures and increases the quality of life of patients. As the modern life becomes more hectic and severe diseases appear, sustained treatments become more necessary. This paper fulfills the requirements to both the patient or remote paramedics and the specialist doctor. The proposed model is more sustainable for using GPRS. This time communication take place via the wireless transmission which is more fast and easy system. If proper research and development work is initiated in this field today, then it is possible to propose a better solution to urban health care or remote areas those are far away from specialist doctor. Construction of standard prototypes and patents for home based health care makes our life more easier. e-Health care embodies a rising field in the health assistance. It moderates the need of transporting patients between hospital and patients house. e-Heath department in every central or city hospital and making it an integral part of the hospital management system reduces the difficulites of urban health care.

References

[1] Yan Xu, Jianwen Zhang, Eric Chang, Maode Lai, and Zhuowen Tu, Contexts-Constrained Multiple Instance Learning for Histopathology Image Analysis (Oral). *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*,June 2012.

[2] N M Hjelm,Benefits and drawbacks of telemedicine,*Journal of Telemedicine and Telecare*,Volume 11, November 2005.

[3] M. V. M. Figueredo, J. S. Dias, Mobile Telemedicine System for Home Care and Patient Monitoring.*Proceedings of the 26th Annual International Conference of the IEEE EMBS San Francisco, CA, USA*,September 1-5, 2004.

[4] https://www.google.com.bd/webhp?*sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=per%20capita%20income%20of%20bangladesh,1 November,2015.*

[5] Nuria Oliver, Fernando Flores-Mangas, A Real-time Wearable System for Monitoring and Analyzing Physiological Signals. *Microsoft Research, Redmond WA 98052 USA.*

[6] Michael, K, McNamee, The emerging ethics of humancentric GPS tracking and monitoring.*Proceedings of the International Conference on Mobile Business, Copenhagen, Denmark*,25-27 July 2006.

[7] David Muchangi Mugo, Dr. David Nzuki (Ph.D), "Determinants of Electronic Health in Developing Countries.*International Journal of Arts and Commerce*,Vol. 3 No. 3, April, 2014.

[8] Baker, J. , Beechwood Data Syst., Clark, NJ , Savino, S., "The role of client/server computing technology in the management of global enterprises.*International Conference on Management and Technology*,27-31 Jul 1997.

[9] Amna Abdullah, Asma Ismael, Aisha Rashid, Ali Abou-ElNour, and Mohammed Tarique,Real Time Wireless Health Monitoring Application Using Mobile Devices. *International Journal of Computer Networks & Communications (IJCNC)* ,Vol.7, No.3, May 2015.

[10] Sanjana Prasad, P.Mahalakshmi, A.John Clement Sunder, R.Swathi,Smart Surveillance Monitoring System using Raspberry PI and PIR Sensor. *International Journal on Applications of Information and Communication Engineering* ,Volume 1,Issue ,January 2015.

[11] Sayed Mohsin Reza, Md.Mahfujur Rahman, Md.Hasnat Parvez, Shamim Al Mamun,M. Shamim Kaiser, Innovative Approach in Web Application Effort & Cost Estimation using Functional Measurement Type. *International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, May 2015.

[12] Yun-Hong Noh ; Jiunn Huei Yap ; and Do-Un Jeong, Implementation of the Abnormal ECG Monitoring System Using Heartbeat Check Map Technique. *In the proceedings of International Conference on IT Convergence and Security*, December 16-18, 2013.

[13] Qingchao Gong, Guangming Li, Peng Sun and Yong Pang, Design and Implementation of Pulse Signal Detection System Based on Bluetooth Transmission*International Journal of Control and Automation*, Vol. 8, No. 7 (2015), pp. 141-148.

[14] Claudia S. Robertson, Eric L. Zager, Raj K. Narayan, Neal Handly, Alok Sharma, Daniel F. Hanley, Homero Garza, Eileen Maloney-Wilensky, Justin M. Plaum, Carolyn H. Koenig, Anne Johnson, and Timothy Morgan, "Clinical Evaluation of a Portable Near-Infrared Device for Detection of Traumatic Intracranial Hematomas," *JOURNAL OF NEUROTRAUMA*, 27:15971604 (September 2010.

[15] Haiming Jing,Design and Development of the Temperature Detection System,*International Journal of Control and Automation*, Vol. 8, No. 2 (2015), pp. 409-416.

[16] M. Shamim Kaiser, Zamshed Iqbal Chowdhury, Shamim Al Mamun, Amir Hussain, Mufti Mahmud,Solar Powered Wheel Chair For Physically Challenged People Using Surface EMG Signal, *2015 IEEE Symposium Series on Computational Intelligence*.

[17] F. Abtahi, B. Aslamy, I. Boujabir1, F. Seoane, and K. Lindecrantz,An Affordable ECG and Respiration Monitoring System Based on Raspberry PI and ADAS1000: First Step towards Homecare Applications. *16th Nordic-Baltic Conference on Biomedical Engineering*,IFMBE Proceedings 48,DOI: 10.1007/978-3-319-12967-9_2.

[18] K. Siau,Health Care Informatics. *IEEE Transactions on Information Technology in Biomedicine*, vol. 7,March,2003.

Paper published in 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT 2015)

DOI: 10.1109/ICEEICT.2015.7307462, **Publisher:** IEEE
IEEE PDF Link: https://ieeexplore.ieee.org/abstract/document/7307462

2nd Int'l Conf. on Electrical Engineering and Information & Communication Technology (ICEEICT) 2015
Jahangirnagar University, Dhaka-1342, Bangladesh, 21-23 May 2015

Innovative Approach in Web Application Effort & Cost Estimation using Functional Measurement Type

Sayed Mohsin Reza; Md. Mahfujur Rahman; Md. Hasnat Parvez; M. Shamim Kaiser; Shamim Al Mamun
Institute Of Information Technology
Jahangirnagar University,Dhaka,Bangladesh
smrezaiit@gmail.com; mrrajuiit@gmail.com; hasnatiit847@gmail.com; mskaiser@juniv.edu; sam@juniv.edu

***Abstract*—Software cost models and effort approximations support project supervisors to distribute resources, control budgets and agenda and develop modern practices, leading to projects completed on time and within financial plan. If cost and effort are determined suspicious in software projects, suitable occasions can be missed; whereas expectant predictions can be affected to some resource losing. In the context of web development, these issues are also vital, and very challenging given that web projects have short schedules and very fluidic opportunity. Since software projects are continually changed in nature, earlier projects may not necessarily cover all aspects of a new project when used as a basis for cost estimation. Preliminary software estimation models are constructed on regression analysis or mathematical sources. This paper aims to propose an approach to develop the correctness of software effort and cost estimation using the structure of data set of a web application. All the measures collected, apart from total effort, were introduced using the original web hypermedia applications to ensure that functional measurement types were precisely measured.**

***Index Terms*—Software Effort & Cost Estimation, Web Application, Functional Measurement Type etc.**

I. Introduction

The cost of software projects and the quality of products are affected by the correctness of software effort estimation. Accurate cost estimation of a software development effort is critical for good management decision making. Predicting software development effort with high precision is still a great challenge for project supervisors. On the otherhand, accurate estimation of the effort and cost of a software system is one of the vital and challenging tasks for software project management. It aids in agreement negotiations, project planning and effective distribution of resources. However, estimates at the introductory stages of the project are the most difficult to acquire because the primary source to estimate the cost comes from the requirement specification documents[1][2][3].

A number of pointers should be measured to estimate the software cost and effort.A number of pointers should be measured to estimate the software cost and effort. One of the utmost vital pointers is the size of the project.The approximation of effort and cost governed by the exact prediction of the size[4]. In general, the effort and cost approximations are challenging in the software projects. Because, software projects are often not exclusive and there has no experience or previous knowledge about them. So, prediction appears convoluted.Moreover, production of projects is not touchable so the amount of effort, cost and the amount of enhancement in the software project is very challenging.

Nowadays, web sites and web portals are more and more complex, and have to control and deliver to their visitors vast amounts of information. Unfortunately, developing web applications through PHP(Hypertext Pre-processor) is not discharged from cost, time and effort estimations, as in traditional software projects.Many estimation models are available, but currently there is no model able to effectively measure the effort of a Web application[5].

The main objective of this paper is demonstrating the abilities of the web application cost estimation methods and clustering based on sub-functional measurement, functional measurement and complexity calculation which helps project supervisor to better understanding.

II. Cost and Effort Estimation Model

Exploration on effort and cost estimation of software development has been plentiful and expanded since the end of the Seventies[6]. This arena is still very much buzzing, as shown by the numerous works existing in the literature. In this arena, experts have extensively investigated the topic, in relation to both approximation approach and investigation approach[5]. These studies were carried out in both industrial and academic contexts. A number of cost estimation methods exist and they can be classified into three main categories [5]. These categories are:

1) Expert Judgment: In expert judgment, a software project estimate the cost and effort established on chronological data and related projects to estimate software. This technique is very subjective and it absences adjustments and thus, cannot be returnable[9].

2) Algorithmic Models: This models is widespread category in the effort and cost estimation technique[8].

978-1-4673-6676-2/15/$31.00 ©2015 IEEE

These prototypes include FPSE, LM, MM, COCOMO, SLIM, SEER-SEM, SLOC and PM [7].

3) Machine Learning: Newly, this techniques are being used in aggregation or as replacements to algorithmic models. These methods include neuro-fuzzy, genetic algorithm ,neural networks, fuzzy logic and regression trees.

III. Background and nature of web applications

Web Applications Effort and Cost Estimation is a complex task due to a number of different but compelling reasons compared to traditional software engineering effort and cost estimation techniques[9].Listed a number of factors that makes web application development slightly different from traditional web application development, indicates that there is no standard to sizing web applications given that there is a wide and diverse set of technologies that can be used to develop web applications e.g. Java (Servlets, Java Beans, Applets, Java Server Pages), HTML, JavaScript, XML, XSL, PHP, ASP.NET etc.

Many different metrics have been used to estimate the size and therefore the effort required to complete a web application project. Examples include the number of web pages, the number of multimedia elements, and the number of links and so on. Additionally, there have been attempts to apply Function Point principles to sizing web applications[3][4]. This approach seeks to derive from a combination of size metrics, a number of functional requirements that would be needed in order to develop the application[23].

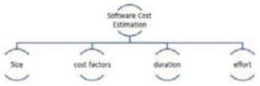


Fig. 1: Software Cost Estimation

IV. Proposed Approach

The proposed approach combines the concepts of Functional Measurement Type and Linear Algebra Rules. The functional measurement Type is used to represent the corresponding linguistic requirement variables for each interval instead of representing them as crisp interval. Thus, this should help us to derive the final prediction. Determining the corresponding Software Effort & cost estimation is performed by using association rule. The approach is described by following as explained below:

A. Dataset Description

The Proposed approach has dataset includes 9 basic requirements of projects with 5 Functional Measurement types and 4 complexity Factor of the software development effort. So, using dataset for evaluating the proposed model is based on Algorithmic model. The second attempt was to create an all requirement dataset based on one of requirement, Table 1, based on this model and algorithm.

TABLE I: Web Application Dataset Structure Development

SFMT	Description
SD	Defines the framework, MVC based development of a project.
LgD	Defines the development issue of logical term
LoD	Defines the development of conditional statement with Looping Condition
FD	Defines the logical separate code development using Multiple Parameter Development
AcD	Defines the development of form connectivity among forms
AD	Defines the issues of accessibility issue of the project
SeD	Defines the development of authorization, authentication etc.
CD	Defines the Page transfer Development of a project
RD	Defines the development of easier communication with a project

Where, SFMT means Sub Functional Measurement Type, SD means Structure Development, LgD means Logical development, LoD means Looping Development, FD means Functional Development, AcD means Action Development, AD means Additional Development, SeD means Security Development, CD Connectivity Development, RD means Readability Development.

B. Step wise development of approach

Step 1: define or specify one of your criteria for each requirement in dataset, then asserts it into several equal intervals (lengths). After that, each requirement should be partitioned into a number of equal to others intervals where the number and length of intervals should be predefined by estimator. Assuming n is the number of intervals then the length of interval metrics are calculated as follows:

DD(Metrics) = $SD(Metrics) * \frac{10}{\sqrt{3}}$
LD(Metrics) = $SD(Metrics) * \frac{10}{\sqrt{2}}$
FD(Metrics) = $SD(Metrics) * \frac{3}{\sqrt{3}}$
AD(Metrics) =$SD(Metrics) * \frac{3000}{\sqrt{2}}$

Where SD means Structural Development, DD means Decisional Development, LD means Looping Development, FD means Functional Development, AD means Advance Development.

Step 2: Cost Factor matrix development:
See Table 2

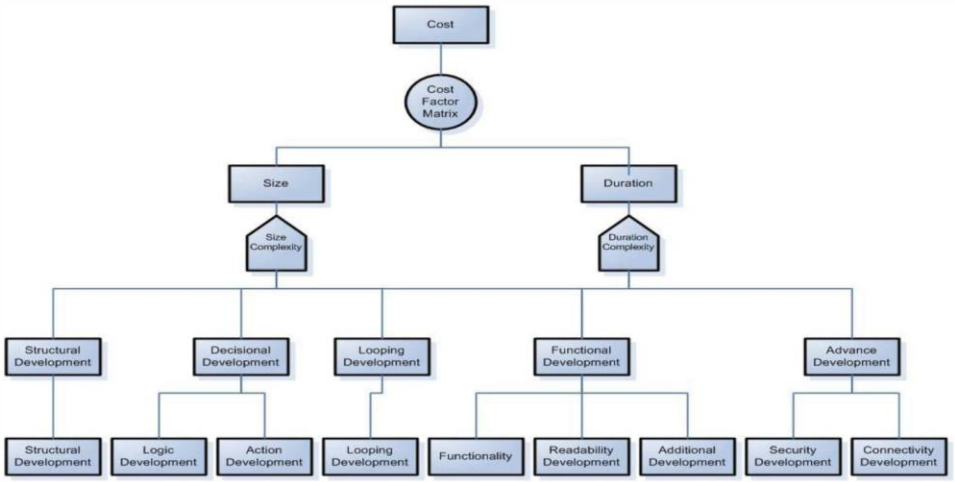


Fig. 2: Functional Measurement Tree

TABLE II: Cost Factor Matrix

	SD	DD	LD	FD	AD
SD	1	$\frac{\sqrt{3}}{10}$	$\frac{\sqrt{2}}{10}$	$\frac{\sqrt{3}}{3}$	$\frac{\sqrt{2}}{3000}$
DD	$\frac{10}{\sqrt{3}}$	1	-	-	-
LD	$\frac{10}{\sqrt{2}}$	-	1	-	-
FD	$\frac{3}{\sqrt{3}}$	-	-	1	-
AD	$\frac{3000}{\sqrt{2}}$	-	-	-	1

Step 3: Define a corresponding extra linguistic variable for each interval of requirement of Functional Measurement Type. This step has been used intensively to get the prediction cost value.

These linguistic variables can be divided into 5 categories

- Context And Peripheral Environmental Analysis
- Site Arrangement
- Graphics And Plots
- Multipleinstance Common Elements
- Reporting And Query

This variables will help the users to predict software effort estimation.

Step 4: Project management software caters to the following primary functions:

Web Application Project development: To explain a Web application project agenda, a project executive can practice the software diagram of a web application project jobs and visually define assignment communications[**?**].

Task supervision: Lets user to generate and assigned of responsibilities, goals and status intelligences.

Document allocation and teamwork: Efficiency is amplified via a principal manuscript repository retrieved by development stakeholders.

Time Constraint and communication development: Web Application Project timelines comprise planned meetings, activity periods and links that should automatically inform through all Project Executive and stakeholder schedules.

Bug and error supervision: Web Application Project management software enables bug and error reporting, observing, informing and apprising for stakeholders.

Period Analysis: Software need to have the capability of tracking period for all tasks preserve histories for third-party specialists.

Step 5: Calculate all the basic sub functional requirement to calculate Software effort and cost estimation.

$$Effort = \sum_{i=1}^{n} F_m * C_f \quad (1)$$

where,
F_m= Functional Measurement type,
C_f= Complexity factor for those sub functional measurement,
n = number of functional measurement type.

$$Cost = \sum_{i=0}^{n} C_F(F_m + L_v + P_{mc}) + M \quad (2)$$

Where,
L_v = Linguistic Variable Cost,
P_{mc}= Project Management Cost,
M = Miscellaneous cost.

C. Algorithm Development

The proposed method is a combination of analogy method and Linear algebra, in which clustering has been used to make the data as high-normal as possible. In fact, the data analysis is improved by using the smooth and consistent data sets. The consistency of a dataset is measured by the level of normality that exists among the projects[17].

Data: Data is collected from user.
Result: Summarize the Predicted Software Effort and cost estimation.
Process the data known as WEB APPLICATION DATA SET take proper decision
while *taking data from user/project managers* **do**
  read sub functional measurement value;
  if *measurement is positive value* **then**
    show other values of sub functional measurement type and calculate effort;
    if *linguistic variables and project management metrics value provide* **then**
      Calculate software cost.;
    else
      request to provide one of sub functional measurement value;
      break;
    end
  else
    request to provide one of sub functional measurement value;
  end
end

Algorithm 1: Algorithm for Predicted Software Effort & Cost.

In other words, when the relationship between the independent features and the development effort is the same among different projects, the number of outliers is decreased and the consistency of the data set is increased. This can be achieved by the clustering of projects. The estimation model is constructed in this stage, which includes two main sections as stated in the following. Efficient data needs to consider the sample sets, in which the number of projects is equal or more than the number of features. As the number of features is increased, the number of projects should be increased, and this ensures that the data analysis is completed[13].

Data: Resources are from the Project Document
Result: Summarize the Actual Software Effort and cost estimation.
start:
forall the *directories of Project* **do**
  Iteratively read the Project Directory Folder.;
  if *Check whether is it file or not* **then**
    Get the filename if it is checked as file.;
    if *Check whether file information exists or not* **then**
      Calculate software effort & cost using sub functional measurement type of that file. ;
    else
      return 0;
    end
    if *Check whether folder information exists or not* **then**
      Check that Folder to start iteratively read the directory. go to start. ;
    else
      return 0;
    end
  else
    do nothing;
  end
end

Algorithm 2: Algorithm for Actual Software Effort & Cost.

In proposed approach algorithm, the best number of effort and cost is determined based on the conditions of the dataset. Indeed, the proposed algorithm will identify on how many sub functional measurement type must be considered in the project management process to have the maximum number of Emplyee, Working hours etc, in which the number of projects is equal or more than the number of features. The maximum number of functional measurement type will ensure the high level of consistency for the existing projects. Algorithm 1 shows the procedure for finding the software effort and cost of projects[22].

V. Evaluation Result

After transferring the data, the proposed approach was conducted three main case studies to evaluate result. These cases, which used same datasets , were utilized to perform training on the parameter values. The data points and the project management cost were adopted for testing purposes. The original functional measurement parameter values are calculated in each case [14].

The parameter values of the four cases are different but in linear. This reason causes the prediction performance difference amongst the Project cases . In order to assess the prediction performance of the approach, Calculation effort from functional measurement and Cost factor with the proposed approach framework. Performance metrics were used for the analysis of each project case. Accordingly, Table

3 & Table 4 presents the results from Cases 1 to 3 project data points.

For an example, proposed model is done on three project.

- opencart
- openconf
- sibco

Result based on this model shows that in Table 3 and Table 4.

TABLE III: Result of Proposed Model

Web Application	SD	LOD	LPD	FD
opencart	213327	37168	29679	12391
openconf	41154	6329	5401	2202
sibco	48970	9481	7428	3406

Where, SD means Structural Development,LOD means Logical Development,LPD means Looping Development,FD means Functional Development.

TABLE IV: Result of Proposed Model

Web Application	ACD	AD	SD	CD	RD
opencart	34292	8664	4	26	13574
openconf	34292	8664	4	26	13574
sibco	3502	3336	0	6	10751

Where, ACD means Action Development,AD means Additional Development,SD means Security Development,CD means Connectivity Development,RD means Readability Development.

Case Study 1: This case involved the parameters of opencart projects. This project is one of furnished open source application. This open source application is formed by PHP MVC framework. There are several projects that meet this requirement. Since the proposed approach is to estimate software cost estimation, the project content was done with opencart project data points, while sevral pieces of project data and the functional data points were used for testing.

Case Study 2: This Proposed approach used the data points from an open source application named openconf projects to calibrate the software cost estimation without removing the other data points. The testing was performed with the project dataset used in the prposed approach and with the functional project data points. In comparison to Case 1, this test attempted to ascertain the prediction performance when the result involved in the results.

Case Study 3: In the previous two cases, all data points from the Sibco projects were used for . However, in Case 3, the proposed approach used part of this dataset to calibrate the this approach, and the rest of the data points, along with the project data points, were used for testing. The objective of this case was to determine the impact of the dataset size on the calibration results.

VI. Comparison Between Proposed Approach And Estimation Methods

At this division allowing existing topics, it is conceivable to associate cited estimation methods centered on benefits and drawbacks of them. This evaluation could be suitable for picking a proper technique in a particular web application project. Selecting the estimation method is completed established on abilities of approaches and state of the Web application. Table shows a comparison of declared methods for estimation. For doing comparison, the general surviving estimation methods have been nominated.

A. Comparison with Algorithmic Type

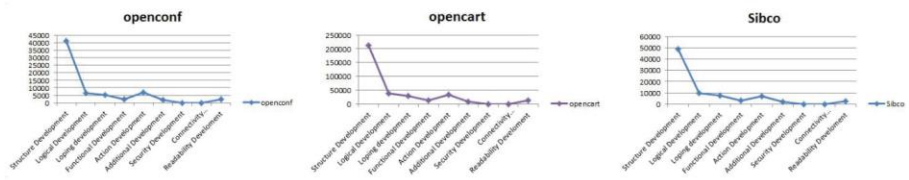
COCOMO, Function Point are Algorithmic Software Effort Estimation Method. COCOMO provides clear results, very common for all Software. For Software Effort Estimation using COCOMO is required much data, It is required historical data for any project. On the Other side, Function Point is Semantic free; Its results are superior than SLOC. in Function point, Computerization is hard to do, excellence of productivity are not considered[4][12]. In this approach, Optimization of this problem is solved. This approach provides better results for all web application. Its required only Web application structures. Historical data is not analyzed in this approach [16][21].

B. Comparison with Non-Algorithmic Type

Expert Judgment, Analogy, Neural Networks, Fuzzy Approach are Non Algorithmic model of Software Cost Estimation. Expert judgment is Fast Predicated estimation method. It has some procedure of adaption of Especial Projects. Its success depend on expert knowledge. Analogy approach Works created on definite practices, having special expert is not essential. A lots of information about past projects is required. Neural Network approaches have guideline of designing[4][21]. This approaches performance depends on large training data. Another popular approach is Fuzzy logic approach. in this approach Training is not require. Flexibility is the main issue of the approach. It is hard to custom, Keeping the degree of importance is challenging. This paper approach is Non-justified to the Non algorithmic Model. Expert Judgment is not stand here for developing the approach[16].

C. Result Comparison

The Appraisal of proposed approach was performed in 3 web applications used in Result analysis. Web application Effort & Cost Estimation was conducted using Schneiders Model, Karners model and proposed approach. Most Specialists use MMRE to calculate the error Percentage of Software Effort & Cost Estimation. MMRE is the mean of the Magnitude of Relative Error. It is very communal principle used to



(a) OpenConf (b) OpenCart (c) Sibco

Fig. 3: Case Study of Three Web Application

evaluate software cost estimation models[3][6][7]. Magnitude of Relative Error (MRE) for each surveillance can be obtained as:

$$MRE_i = \frac{|AE_i - PE_i|}{AE_i} \tag{3}$$

Where, AE means Actual Effort,PE means Predicted Effort. MMRE can be accomplished be an average of the summation of MRE over N interpretations[20][24][11].

$$MMRE = \frac{1}{N}\sum_{i=0}^{n} MRE_i \tag{4}$$

TABLE V: Comparison Between Old Models and Proposed

	OpenConf	OpenCart	Sibco	Kar	Sch
MMRE	27(%)	32(%)	29(%)	34(%)	30(%)

VII. Conclusion

This paper appraised the frame of research on effort and cost estimation models for web applications by scrutinizing the procedures that were castoff to shape approach, the datasets that were castoff and the research types engaged. This was done in the environment of espousing effort and cost estimation practices from traditional software development. Although many revisions have been accompanied by effort estimation models for web applications,There is no strong suggestion that there is a certified method or a set of verified approaches for estimating the effort and cost of web applications. All of the performances used are tailored forms of systems taken from traditional software engineering. No ominously new techniques have been projected. Moreover, there is countless discussion about what scope drivers should be used to originate approximations[15][25].

Supplementary new size metrics are being technologically advanced and tailored from current approaches Object Points, Web Objects, Data Web Points etc. Other frequently used size drivers in traditional software development such as Lines of Code (LOC) are hardly used in Web Application Effort and Cost Estimation Models[19].

Proposed case study displays that significance of the output result and the proposals comprising the premise of linear algebra. Without a doubt, each proposal has a principle. After providing a normal clarification of the principle and output of each linear algebra rule, the proposed approach proposed an details of the Association rule. Each rule can be construed as parallel assertion of software effort and cost estimation in which is informal for users to realize[18].

References

[1] Sayed Mohsin Reza,Md. Mahfujur Rahman,Shamim Al Mamun, A New Approach for Road Networks - A Vehicle XML Device collaboration with Big Data. *2nd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*,, Jahangirnagar University, Dhaka, Bangladesh, May 2015.

[2] Sayed Mohsin Reza,Md. Mahfujur Rahman,M Manjur Mahmud,Shamim Al Mamun A New Approach of Big Data collaboration for Road Traffic Networks considering Path Loss Analysis in context of Bangladesh. *JU JOURNAL OF INFORMATION TECHNOLOGY*,, VOLUME 1, JUNE 2014.

[3] Abeer Hamdy, Fuzzy Logic for Enhancing the Sensitivity of COCOMO Cost Model. *Journal of Emerging Trends in Computing and Information Sciences*,VOL. 3, NO. 9, SEP 2012.

[4] Vahid Khatibi, Dayang N. A. Jawawi, Software Cost Estimation Methods: A Review.*Journal of Emerging Trends in Computing and Information Sciences*,VOL.2 NO. 1, 2010-11.

[5] Erika Corona, Giulio Concas, Michele Marchesi, Giulio Barabino, Daniele Grechi, Effort Estimation of Web Applications through Web CMF Objects.*Seventh International Conference on Software Process and Product Measurement (IWSM-MENSURA)*, 2007.

[6] A.J. Albrecht, Measuring application development productivity. *Proceedings of the Joint SHARE, GUIDE, and IBM Application Development Symposium*.pp. 8392, Monterey, California,IBM Corporation, October 1979.

[7] Ali Bou Nassif,Luiz Fernando Capretz, Estimating Software Effort Based on Use Case Point Model Using Sugeno Fuzzy Inference System.*23rd IEEE International Conference on Tools with Artificial Intelligence*,2011.

[8] L. C. Briand and I. Wieczorek, "Resource Estimation in Software Engineering.*Encyclopedia of Software Engineering*,vol. 2, pp. 1160-1196, 2002.

[9] Lenandlar Singh, Where are we? A Review of Effort and Cost Estimation Techniques used for Web Applications. *Georgian Electronic Scientific Journal: Computer Science and Telecommunications*,2009.

[10] Emilia Mendes, Ian Watson, Chris Triggs, Nile Mosley, Steve Counsell , A Comparative Study of Cost Estimation Models for Web Hypermedia Applications. *Empirical Software Engineering*, June 2003.

[11] Ziauddin, Shahid Kamal, Shafiullah khan and Jamal Abdul Nasir, A Fuzzy Logic Based Software Cost Estimation Model. *International Journal of Software Engineering and Its Applications*, Vol. 7, No. 2, March, 2013.

[12] Vishal Sharma and Harsh Kumar Verma, Optimized Fuzzy Logic Based Framework for Effort Estimation in Software Development. *International Journal of Computer Science Issues*, Vol. 7, Issue 2, No 2, March 2010.

[13] Preeth.R,Shivakumar.N,Balaji.N , Software Effort Estimation Using Attribute Refinement based Adaptive Neuro Fuzzy Model. *PIERS Proceedings*, Czech Republic, 2007.

[14] Wei Lin Du, Danny Ho, Luiz Fernando Capretz, Improving Software Effort Estimation Using Neuro-Fuzzy Model with SEER-SEM. *Global Journal of Computer Science and Technology*,Vol. 10 Issue 12 (Ver. 1.0) October 2010.

[15] Alaa F. Sheta and Sultan Aljahdali, Software Effort Estimation Inspired by COCOMO and FP Models: A Fuzzy Logic Approach. *International Journal of Advanced Computer Science and Applications*,Vol. 4 No 11, 2013.

[16] Iman Attarzadeh and Siew Hock Ow, A Novel Algorithmic Cost Estimation Model Based on Soft Computing Technique. *Journal of Computer Science*, 2010.

[17] Ch. Satyananda Reddy, KVSVN Raju, A Concise Neural Network Model for Estimating Software Effort. *International Journal of Recent Trends in Engineering*, Issue. 1, Vol. 1, May 2009.

[18] Nabil Ibrahim El Sawalhi, A Modeling the Parametric Construction Project Cost Estimate using Fuzzy Logic. *International Journal of Emerging Technology and Advanced Engineering*, Issue. 4, Vol. 2, April 2012.

[19] Rama Sree P, Prasad Reddy P. V. G. D and Sudha K. R, Hybrid Neuro-Fuzzy Systems for Software Development Effort Estimation. *International Journal on Computer Science and Engineering (IJCSE)*, Issue. 23, Vol. 4, December 2012.

[20] Venus Marza, Amin Seyyedi, and Luiz Fernando Capretz, Estimating Development Time of Software Projects Using a Neuro Fuzzy Approach. *World Academy of Science, Engineering and Technology*, 2008.

[21] P.Ramaiah Chowdary, G.Vihari, A Concise Fuzzy Model Approach for Estimating Software Effort. *International Journal of Computer Science And Technology*, 2013.

[22] Sumeet Kaur Sehra, Yadwinder Singh Brar, and Navdeep Kaur, Soft Computing Techniques For Software Project Effort Estimation. *International Journal of Advanced Computer and Mathematical Sciences*, Vol 2, Issue 3, 2011.

[23] Sandeep Kad and Vinay Chopra, Software Development Effort Estimation Using Soft Computing. *International Journal of Machine Learning and Computing*, Vol. 2, No. 5, October 2012.

[24] Rajender Bathla, Sukhdev Singh, Mittar Vishav & Rahul Gupta, Innovative Scenario In Software Development Effort Estimation Based On A New Fuzzy Logic Model. *International Journal of Information Technology and Knowledge Management*, Vol. 2, No. 2, 2010.

[25] Gaurav Kumar and Pradeep Kumar Bhatia, Automation of Software Cost Estimation using Neural Network Technique. *International Journal of Computer Applications*, Vol. 98, No. 20, 2014.

Paper published in 2014 International Conference on Electrical Engineering and Information & Communication Technology

DOI: 10.1109/ICEEICT.2014.6919153, **Publisher:** IEEE
IEEE PDF Link: https://ieeexplore.ieee.org/abstract/document/6919153

International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT) 2014

A New Approach for Road Networks - A Vehicle XML Device collaboration with Big Data

Sayed Mohsin Reza
Institute of Information Technology
Jahangirnagar University
Savar, Dhaka.
Email: smrezaiit@gmail.com

Md. Mahfujur Rahman
Institute of Information Technology
Jahangirnagar University
Savar, Dhaka.
Email: mrrajuiit@gmail.com

Shamim Al Mamun
Institute of Information Technology
Jahangirnagar University
Savar, Dhaka.
Email: sam@juniv.edu

***Abstract*—Large amount of money is lost to manipulate traffic system or road congestion worldwide every year. By providing the travel time for a specific road, traffic congestion can be minimized. Many approaches has been taken such Automatic Vehicle Identification, Loop Detectors etc. But this methods are costly. In this paper, A system is proposed that provides traffic intensity level information to the user according to recent traffic data analysis. To minimize the traffic congestion, the paper proposal is to use big data concept in this arena. This proposal develops a structure of a simple XML device which is installed on a vehicle to trace and provide information of traffic intensity. It can estimate travel times in a road network accurately. Further, according to this system anyone can develop application for business process development and increase information transfer to the local user.**

***Index Terms*—Traffic, Big Data, Congestion, Intensity, XML device, Road Networks, travel times etc.**

I. Introduction

Every day, time is wasted in traffic congestion. This problem is suffered by millions of people. This traffic congestion wastes millions of money of every government. Every country government propose budget every year to solve this type of traffic mobbing. Escalation of traffic volume, every government is on dilemma. On the basis of this need, new dynamic travel time system with traffic intensity information should be developed [1].

Big data is castoff to increase many aspects of municipalities and countries. For example, it allows municipalities to adjust road traffic movements based on real time traffic flow information. Several cities are currently conducting big data analytics with the aim of turning themselves into Keen Cities, where the transportation infrastructure and utility routes are all combined.

Further, Analysis of Big Data refers as road traffic data turn into valuable resources. It can increase efficiencies of country's competitive advantage. For this, Data Science will help us to produce valuable information from raw data [2]. Big Data refers that -Traditional enterprise data, Machine generated or sensor data, Social data etc.

In this proposal, machine generated data will be included. The system manipulates these machine-generated data and use the observations per road segment and estimates a travel time based on the average speed.

II. Utilization of Big Data in Society

Analyzed and distilled big data with traditional enterprise data can enhance productivity, competitive stronger position and can innovate much better significant impact on society.

At first, social media sites wouldnt be present without big data. By capturing available data of a user or a member of social media, business developers can model their business and modified experience on the web. Additionally, in healthcare facilities, long term in-home monitoring device to measure progress and vital signs can be useful to ensure patient health. This type of service assistances the patient to reduce appointment to the doctor and hospital admittance.

III. Big Data Platform

Now a day, IT infrastructure is developing in such a way that produces unique requirements considering big data platforms components. To develop the paper proposal, the system should go through some big data requirements [3].

- Big data acquisition
- Big data organizing process
- Big data Analysis
- Solution prediction

IV. Road Traffic Intensity Parameter using Big Data

There are several surveys happens to find out the parameter in the concept of Big Data. In simple way, we can assume that the volume of data (Road traffic data) is one of the parameters. In fact, it is not only the parameters that controls Big Data terms. There are Volume, Velocity, Variety and Value.

Each year, data size is mounting 40 percent per year as saying by McKinsey Global Institute researchers [1]. As their say, larger quantities of data are produced by Machine Generated data. In Big data terms, Traditional data and non-traditional data formats are on contrast position.

Among these concepts, we face the challenge that which data is valuable and that valuable data should be used for analysis road traffic summarize data to ensure road traffic intensity and can potentially help to improve the traffic situation in large cities.

978-1-4799-4819-2/14/$31.00 ©2014 IEEE

V. Road Traffic Data [Big Data] Infrastructure

The continual growth of road traffic increases the need for solutions allowing monitoring and controlling traffic helping to increase road safety, improve traffic flow, and to protect the environment. So, a system which provides traffic management solutions allowing road authorities and operators to manage, monitor and maintain their roadways while giving the road users intelligent information to ensure convenient and save trip [4].

For ensure a Road Traffic Network system, we should follow some proper ways to monitor, control traffic. These are -

A. Data Acquisition

Data acquisition is the method of evaluating an electrical or somatic phenomenon such as current, voltage, pressure, temperature and sound with a computer. A Data Acquisition system consists of devices, hardware, and a computer with programmable software. Compared to traditional measurement systems, PC-based Data Acquisition systems abuse the processing power control, productivity, demonstration, and connectivity capabilities of industry-standard computers providing a more authoritative, stretchy, and cost-effective measurement clarification [5].

B. Data Organization

All data from the devices as well as manual actions of the operating staff are considered. The Traffic Management Systems enables the automated generation of traffic information. Raw traffic data and environmental data are processed to traffic information so that customers can access it via a broad spectrum of telecommunication technologies such as web services [6].

C. Data Analysis

Assembling, evaluating and processing the entire data from various devices on the roads in order to understand and monitor in real-time the prevailing traffic conditions. This analysis is the last step to provide correct information to the user.

VI. System Description

A. Introduction

In traffic system, Government of every country expends their money for organizing the traffic system. Billions of dollars are lost. People waste their times for congestion in traffic. People want to save their time but is there any way to save time? To ensure traffic intensity, we can survey road traffic data. But how the road traffic data is collected? Or in which way we will provide the users the traffic information?

Is Big Data concept applies here to ensure any possibility of road traffic Data? Data acquisition is a now challenge for us. In past, there were proposed some ways to provide this type of service.

To acquisition of road traffic data, we have to collect data using point based approach. For a short time, a control server

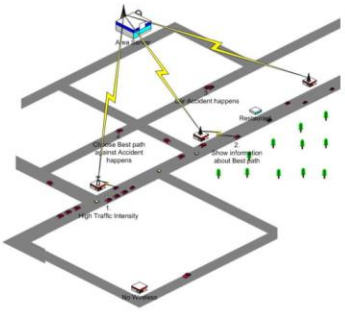


Fig. 1. Entire Scenario of the System

will summarize the accumulating data to traffic organizers data. This data will be promoted to the other area users but not these area users [7].

For time consume in this method, how much bandwidth and transfer rate will be needed? To solve this, we have to use less data information but not whole. We can use XML format to show information to users device.

B. Algorithm to Develop this System

XML vehicle device to server data handover and server to that XML device information transfer is a continuous process to develop this system. For this, we develop an algorithm which helps us to ensure the process of this system.1

Data: Data Acquisition from vehicle XML device (simple black box) send to the Server.
Result: summarize the accumulating data to traffic organizers data.
Process the data known as Road Traffic Data (Big Data) to take proper decision
while *device status* **do**
  read current vehicle information;
  if *traffic intensity level is high* **then**
    appearance possible way(if exist) to go to destination, time and appearance local information.;
    if *path changes* **then**
      go to the beginning of section.;
    else
      check black box provide service or not?;
      break;
    end
  else
    one way decision to go to destination, time and appearance local information;
  end
end

Algorithm 1: Algorithm for develop the system

C. Work flow

Road Traffic congestion system should follow a work flow diagram. As far as system data acquisition, data organization and at last data analysis is shown in this work flow diagram.

Here, a server as Google file system or Hadoop Distributed File System will work. It acquires data from Vehicle device that transmit traffic information as XML format. System Server further processes this information (Big Data or Road Traffic Data). These processing raw data will produce valuable resources for sending information to vehicle which is under around nodes.

These information contains the traffic intensity level or road accident information that helps the around user vehicle to take proper decision about his/her vehicle.

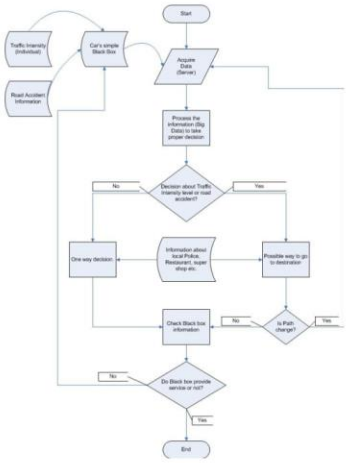


Fig. 2. Data Flow Diagram

If any vehicle changes its path then next node/station will acquire its information and send it to the server. This continuous process will produce Big data refers as Road Traffic Data.

At last, every data sync time, we send the information about device that is the device services working or not? If not then the server take an automated decision, else if it continues its continuous process.

D. Data Schema

To estimate travel times, different data source (users information) are relied on. Acquisition of data should be manipulated by a Data schema developed in the Control server. From that data, users are provided various types of information like longitude, latitude, timestamp, possible ways etc.

Here, we proposed a data schema 3 installed on the control server. This data schema will be helpful to the developer to collect several types of information which is related to the users information.

To be built a data schema, the main table data acquisition, contains data such as drivers information, connections information, source information etc. The abbreviations FK and PK are Foreign key and Primary Key respectively. By querying the main table, developer can summarize the area information which is under of Control Server.

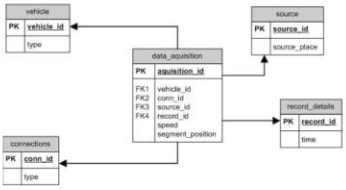


Fig. 3. Data Schema

VII. Feasibility Study

To build up the system we have to go through a feasibility study. In Active Traffic Management, there are some criteria to go through the feasibility study to provide estimated travel time and condition reports to communicate travel and traffic conditions. These are Speed Harmonization, Queue Warning, Junction Control, Hard Shoulder Running and Dynamic Rerouting.

In the system, such like that we determine some criteria related to the feasibility study.

- Data Rate Analysis
- Power Optimization
- Output Scenario

A. Data Rate Analysis

The speed with which data can be transmitted from one device to another is known as data rate. Data rates are often measured in megabits (million bits) or megabytes (million bytes) per second. These are usually abbreviated as Mbps and MBps, respectively. The system should be proposed under

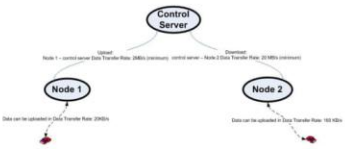


Fig. 4. Data Transfer Rate

some feasibility study. Here, some terms are needed to sign in. Data Transfer rate (DTR) is much more important to time consumption in this system.

Calculation:

Normally, a vehicle goes through Node 1 to Node 2 and the distance is d km. the vehicle takes time t s. Distance (Node 1 to Node 2) = d km At last, we can give this type of conclusion.

$$DT1 + DC1 + DP + DC2 + DT2 < t \quad (1)$$

where,

Device to node1 Data Transfer Rate = DT1
Device to node1 Data Transfer Rate = DT1
Node 1 to Control Server Data Transfer Rate = DC1
Control Server Data Process Time = DP
Control Server to node 1 Data Transfer Rate = DC2
Node 1 to device Data Transfer Rate = DT2

If it is possible then the system is feasible to development. Because if transferring data is longer before car reaching to the next station/node, then the system doesnt provide information to the users vehicle device. In the context of IEEE 802.11 technology, minimum net bit rates are 1 or 2 Mbit/s. It has some advantage like forward error correction code [8].

TABLE I
IEEE 802.11 TECHNOLOGY AND RATE

Technology	Rate
diffuse infrared operating	1 Mbit/s
frequency-hopping spread spectrum operating	1 Mbit/s or 2 Mbit/s
1 Mbit/s or 2 Mbit/s	1 Mbit/s or 2 Mbit/s

A vehicle device produce data in XML format size consists of Less than 10KB. So, we use IEEE 8-2.11 technology for simple wireless data transfer manipulation. Minimum wireless Data transfer rate - I

DT1 = 250 KB/s (25 car per second)
DT2 = 250 KB/s (25 car per second)

We proposed a High speed broadband connection between control server and Nodes. Minimum Data transfer between Nodes to control server:

DC1 = 2 MB/s (250 vehicles data)
DC2 = 2 MB/s (250 vehicles data)

Traffic data acquisition by node covers average 1 KM. So between two nodes we can easily control data acquisition. In control server, for better data manipulation and data organization we can assume some more time. But by calculating data transfer rate among the node we can say that the proposed system can easily pass the time consumption process.

B. Power Optimization

Wireless power transmission should be established on inductive power. Now a days wireless data communication is become popular and expectation is now on rise. So power distribution or measurement is considerable in sense.

Wireless power technology in consumer devices benefit greatly from the availability of a standard that is widely acceptable. To establish interoperability between wireless power chargers and power devices, we have to issue a Wireless power consortium [9].

The standard defines a power transmission of 5W. 110 kHz is Nominal operating frequency. Power control can be initialized by controlling amplitude or frequency. For better power control we have to calculate differences of output value of power transfer. Output may be voltage or current or power.

C. Output Scenario

In the proposed system, at first the XML device which is installed on that vehicle detects the local base station and show the base station name.

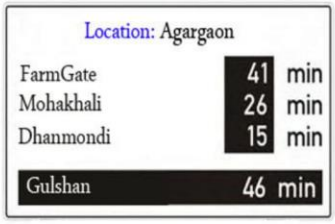


Fig. 5. Probable Output of XML Device in Vehicle

VIII. IMPLEMENTATION OF BIG DATA FOR DATA MANIPULATION

One of the challenges: data acquisitions are complete in arbitrary sense. The data which is collected from the user [vehicles] should be manipulated. After manipulation, summarize data will be transferred to the user output devices in XML format.

XML format is much easier to transfer into a device or we know XML was designed to transport and carry data. XML tags are not predefined. Self-descriptive design is the main theme of XML format.

A. Parsing XML into Data Structures

This data structure (a mixture of hashes and arrays) that relates to the arrangement and content of an XML file. For example, XML device on behalf of a configuration file. XML

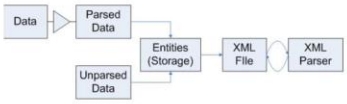


Fig. 6. Process of Parsing XML

files are created for storage elements called entities, which contain either parsed or unparsed data. Parsed data contains data.XML delivers a appliance to execute restrictions on the storage outline and logical structure.

A XML device module called an XML processor is hand-me-down to read XML files and afford access to their content and structure. It is expected that an XML processor is doing its effort in support of another module, called the server application.

IX. Tools for Data Manipulation

The proposed system produces huge data as independent computational agent like Machine generated data. It unknowingly generates the data. So we have to manipulate it to give proper credit like valuable information. To do that, we will use some of these tools for my proposed system [10].

A. Google File System [Hadoop Distributed File System]

Since the proposed system produce large data sets and we have to choose such a system that can store, stream that large data sets at higher bandwidth then a tool named Hadoop Distributed file system will be the proper answer. Hadoop provides a distributed file system to analyze and transformation of very large data sets. Its framework works as like as to develop computation among large amount of hosts [11].

In the proposed system, large data set will produce to collect traffic data which should be store then analyze that data and transform into information. In this case, large computation may be needed.

To increase the performance of application, Hadoop Distributed file system is perfect because it is patterned after UNIX file system. This file system improves the execution of program computations in parallel. These file system stores its metadata on its dedicated server and program data are stored on Data Nodes. For better reliability, the content data are located in multiple data Nodes. Every server is connected with each other using TCP based protocols [12].

In HDFS, user can access a library that exports HDFS file system interface. This system supports various operations like read, write and other operation on directory. Here, the data nodes are on different location, but user can run their application without knowing that. [11]

Traffic data acquisition and then manipulation concept also contains different location so Hadoop distributed file system will work here properly as the paper proposal. Moreover, HDFS contains some more advantage like Backup Node. It is capable of creating periodic checkpoints. This node can be viewed as Name Node (read only). [12]

To ensure security about traffic data, HDFS provides backup node for the valuable data. The data may be in read only format so its security establish as more proper format.

X. Conclusion

In this paper, proposal is developed in such a way so that server system can determine travel time approximation, this system estimate travel times for a road traffic network using a simple device which can send data and retrieve data from server. To collect traffic data, the road network, and the processing data that permits us to filter out exact Road Traffic data, a data schema has been proposed for flexible efficiencies to provide accommodations, upcoming add-ons to the travel-time system as well as vehicle and traffic facts, road circumstances.

It has been conversed how the input scenario and output scenario can be joined in preparation to analyze travel times and make available one possible approach. As a side benefit, this methodology can be used to inspect the frequencies of Road Traffic intensity level. This is used by traffic experts when analyzing road traffic data.

The principal of this system is now in a state where we can initiate to guess travel times. The following is a little list of the future work we would like to look into.

- Mobile Android Apps
- Vehicle Trace Application
- Built A Web Application to Show An Entire City Road Traffic Information

References

[1] Detection and dissipation of road traffic congestion using vehicular communication-http://ieeexplore.ieee.org/xpl/freeabs_all.jsp? arnumber=409762&abstractAccess=no&userType=inst

[2] The Awesome Ways Big Data Is Used Today To Change Our World

[3] An Oracle White Paper, June 2013: Oracle: Big Data for the Enterprise

[4] Travel-Time Estimation in Road Networks Using GPS Data, Anders Forum Jensen and Troels Villy Larsen.

[5] Data Acquisition - http://www.ni.com/data-acquisition/what-is/

[6] Feasibility of Collecting Traffic Safety Data from Law Enforcement Agencies, U.S. Department of Transportation, National Highway Traffic Safety Administration, April 2011.

[7] Road traffic congestion detection through cooperative Vehicle-to-Vehicle communications-http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5735780&abstractAccess=no&userType=inst

[8] IEEE 802.11 - http://en.wikipedia.org/wiki/IEEE_802.11

[9] Bandwidth (signal processing)- http://en.wikipedia.org/wiki/Bandwidth_%28signal_processing%29

[10] Data Science - http://searchcio.techtarget.com/definition/data-science

[11] The Hadoop Distributed File System by Robert Chansler, Hairong Kuang, Sanjay Radia, Konstantin Shvachko, and Suresh Srinivas

[12] Hadoop Distributed File System (HDFS) Architectural Documentation Use Cases - http://itm-vm.shidler.hawaii.edu/HDFS/ArchDocUseCases.html

[13] Statistical Data Analysis - https://reference.wolfram.com/mathematica/guide/Statistics.html

[14] The WEKA Data Mining Software: An Update by Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten

Paper published in Proceeding of the 2nd Workshop on Information and Communication Technology, 2013, at IUT, Bangladesh

A Study on Network Security Services with Cryptography and an Implementation of Vigenere-Multiplicative Cipher

Khandoker Abdul Rahad
Institute of Information Technology
Jahangirnagar University
Savar, Dhaka
Rahad.baten@yahoo.com

Sayed Mohsin Reza
Institute of Information Technology
Jahangirnagar University
Savar, Dhaka
smrezaiit@gmail.com

Abstract- Different types of encryption techniques are used for many years to secure the information. The purpose of this paper is to construct a new method named as Vigenere-Multiplicative cipher. Here we develop the algorithm and its algebraic description. In addition, we provide the frequency analysis of this method. Further, implementation of the method, security attacks, comparisons between given cipher and most common ciphers are briefly discussed.

Keywords: Network Security, Encryption, Vigenere Cipher, Multiplicative Cipher, Encryption Standards etc.

I. Introduction

Development in communication has brought unknown people together through information technology. We are all familiar with the internet, however, web applications out of these that most of us uses are unsafe browsing facilities. There are lots of transactions happening everyday .We have to make sure that these transactions are safe and secure. This technology is known to us as encryption. The internet has so many connections, as we do not have control over the activities such as hacking, sniffing, breach of conduct, and etc. For this we have to come up with a system that is totally impossible to break.

Cryptography is a tool that can be used to keep information confidential and to ensure its integrity and authenticity. All modern cryptographic system are based on Kirchhoff's principle of having a publicly-known algorithm and a secret key. Many cryptographic algorithms use complex transformation involving sub-situations and permutations to transform the plaintext into cipher text. Cryptographic algorithm can be divided into symmetric –key algorithms and public-key algorithms. [1]

II. Art of Communication

In Network Security, Art of communication provides two types of action: Encryption and Decryption. To secure data transfer we have to encrypt our data by scrambling our information. After scrambling the information, it is quite impossible to understand by a third party.

To retrieve the encrypted information properly, we have to decrypt the scrambling information. It enables the receiver to get the exact information. [2]

Even if we really understand what how encryption plays a part in our day-to-day life, but we use it often. As most of the businesses completely depended upon the internet for buying, selling, and transferring money over e-banking system, organizing, teleconferencing, and providing various other services these all need encryption for safe connection and privacy. [3]

Computers in the early form of development did not have networking facility. As the growth of computer networking industry, there are more software that more readily available for individuals. The advantage of the doing business over the internet is performed and, thus, to keep the unauthorized people away from hacking, network encryption has been started.

III. Model of Network Security

Model of computer network security has some basic structure as follows [3]

1) Network Security and their scenario must be included in the algorithm.
2) A proper mathematical function should be developed in the algorithm.
3) Business relevant property with authentication must be declared.
4) Shared key, Protocol of the method should be defined.

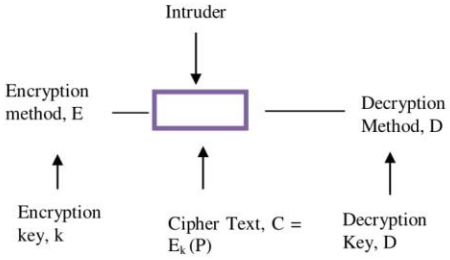


Fig. 1: The encryption model

A mathematical function of the encryption technique with shared key enables the sender to scrambles the message. The output of encryption process is then transmitted, often by a program or software. We assume that the hacker or intruder receives the complete cipher text. However, unlike the intended recipient of the information, the hacker does not know what the decryption key is and so can not decrypt the cipher text easily.[1]

IV. Classification of Encryption

Cryptography algorithms can be divided into two groups symmetric-key cryptography algorithms and asymmetric cryptography algorithms. [4]

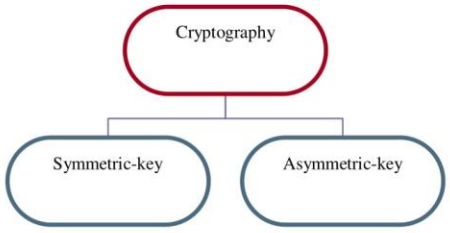


Fig. 2: categories of cryptography

A. *Symmetric-key*

Symmetric-key cryptography is the same key that is used by both encryption and decryption. It is also called single key encryption.

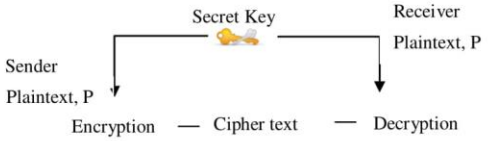


Fig. 3: Symmetric key

Symmetric key Algorithm types:

Stream cipher: plaintext are encrypted one at a time. This conversion varying digits.

For instance,

Plaintext, P = "LAPTOP"
Cipher Text, C = "MBQUPQ"
Block Cipher: plaintext with block are encrypted. Length and unvarying digits has been fixed for each block.

For instance,

Plaintext, P = "LAPTOP"
Cipher Text, C = "13 11 53 44 43 53"

B. *Asymmetric-key*

Asymmetric-key uses two different keys for encryption and decryption. One is public key and another is private key. Public key is used for the encryption and private key is used for decryption. Symmetric key encryption does not provide as much security. Hence, the importance of the asymmetric key, also known as public key encryption, is greater.

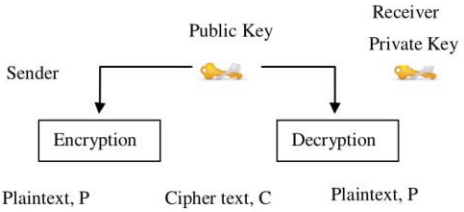


Fig. 4: Asymmetric Key

Asymmetric key has the combination of public key and private key. private key is used by internal host. At the mean time the public key is given to other computers with which it is able to communicate securely.[3]

1)RSA

The most common public key algorithm is RSA. The RSA stands for the last names of the inventors Ron Rivest, Adi Shamir, and Leonard Adleman. They developed this algorithm in 1978. Since then it is widely used. It uses two numbers e and d, as the public and private keys. There is a special relationship between e and d. [4]

RSA provides more data security that's why other algorithms does not cope the RSA. Mathematically, RSA is so strong to maintain the privacy of the message.

2) Public Key Infrastructure (PKI)

For data security in network, we have to develop the keys by using which the information should be transferred. For these purposes, a well known organization distribute, maintain the rules named as Public-key infrastructure (PKI). It relies on a set of unique keys with a mathematical function. Each pair of keys has a public key and a private key. The public key, which is published to users within the public domain, the private key is private for each entity. With public key Infrastructure (PKI) the encrypted data is transmitted to the public key of the recipient. To unlock the encrypted data or information, we use private key as like as a digital signatures. [8]

3) Certificate Authority (CA)

An entity that issues digital certificate named as Certificate authority (CA). This helps relying parties to rely upon signatures made by the private key that corresponds to the public key that is certified. In this model of trust relationships, a CA is a trusted third party that is trusted by both the owner of the certificate and the party or user relying upon the certificate. CA's are characteristic of many public key infrastructure (PKI) schemes. Certificate authority only can issue the Digital Certificate, which contains both the public and private key for encryption and decryption of the information. Depending upon the volume of the identity verification, Certificate Authority can issue Digital Certificate for different level of trust. [9]

4) Registration Authorities (RAs)

Registration Authorities(RAs) have similar functionality as like as CA. The RA can issue the temporary digital certificates. The validation of temporary digital certificates are limited. And it is not fully trustable, unless CA verifies them. A Registration Authority (RA) is a subsystem that accepts enrollment requests and authenticates them in a local context .Upon successful authentication, the RA then forwards the enrollment request to the designated Certificate Authority (CA) to generate the certificate. Based upon the type of enrollment, they can develop appropriate plugins to authenticate the request. [9]

5) Digital Certificates for authentication

Digital certificates are used to authenticate the identity of a computer or a company through Certificate Authority (CA). It can also be used to retrieve rights and authority. Most commonly, it is rapidly used in Ecommerce. To build up the trust between Customer and the authority, Digital certificate plays a vital role. Ecommerce sites have public digital certificates that anybody can view to build up the trust. [3]

V. CIPHER: ALGORITHM IN CRYPTOGRAPHY

Every Sender-Receiver pair needs their very own unique cipher for a secure communication. Millions of communication pairs can be served by one cipher. Now, question is how the cipher is created? We want to secure the communication by creating the cipher as complex as it is unique. Blaise De Vigenere, a French Mathematician tried to create a secret key stream. In a Vigenere Cipher, each letter is shifted in sequence with different shift values. This cipher consists of Caesar ciphers. [1]

Let,

Plaintext, $P = P_1P_2P_3...$
Key Stream, $K = [(K_1K_2K_3...), (K_1K_2K_3...)...]$
Encrypted Text, $C_i = (P_i + K_i) \bmod 26$
Cipher Text, $C = C_1C_2C_3...$

On the other hand, Multiplicative cipher works as the key multiple with plaintext. At the decryption time, we find the message calculating the multiple inverses with Cipher Text.

Let,

Plaintext = P
Key = k_m
Cipher Text, $C= (P * K_m) \bmod 26$

Now, we implement these two ciphers to evaluate a new combination. It works as powerful algorithm with more security and more unique. [4]

VI. Idea of Vigenere-Multiplicative Cipher

The multiplicative cipher is substituting the letter so that it should not be exposed. And the Vigenere cipher is simple enough to be a field cipher if it is used in conjunction with cipher disks. Combination of these two resources can make a more secure cipher.

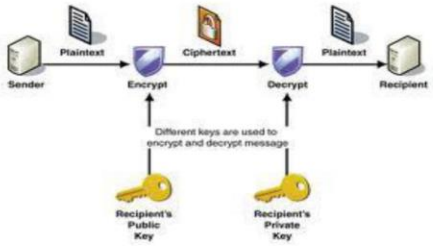


Fig. 5: The Vigenere-Multiplicative Cipher

A. Scenario:

1. Encryption Process

The Vigenere Cipher creates the ciphers by concatenating the encrypted text, and then modifying that text to Cipher Text. In the Vigenere-Multiplicative cipher, we think about those cipher text as Half Cipher Text.

This Half Cipher Text is used as next plaintext of Multiplicative Cipher Text. We assume that, after multiplicative cipher, the original cipher text is introduced.

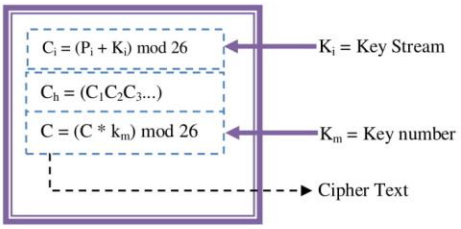


Fig. 6: Eencryption Scenario

2. Decryption Process

We have to figure out the decryption process of the proposed Vigenere-multiplicative cipher. After encryption, we have encrypted/ cipher text.

At first, we have to find out the multiplicative inverse of key number. Then, cipher text multiply with the multiplicative inverse number of the key. It produces the half plaintext as our algorithm.

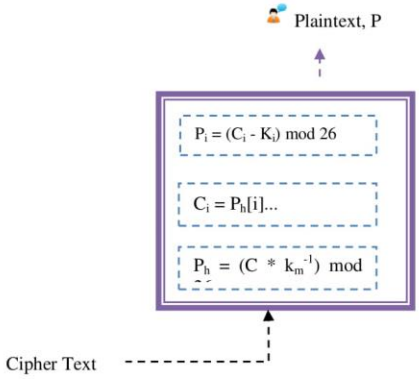


Fig. 7: decryption Scenario

B. Algebraic Description:

Vigenere-Multiplicative can also be viewed algebraically. If the letters A–Z are taken to be the numbers 0–25, and addition is performed modulus of 26, then Vigenere-Multiplicative encryption E using the key K can be written,

Encrypted Text, $C_i = E_k(M_i) = (P_i + K_i) \bmod 26$
Half Cipher Text, $C_h = C_1C_2C_3...$
Full Cipher Text, $C = (C_h * k) \bmod 26$
And decryption D using the key K_i and K_m.
Half Plain Text, $P_h = D_{km}(C) = (C * k_m^{-1}) \bmod 26$
Cipher Text, $C_i = P_h[i]$
Plain Text $P_i = (C_i - K_i) \bmod 26$

C. Example of Vigenere- Multiplicative Cipher

Encrypt the message "I am studying." using Vigenere cipher with the 6-character keyword "mobile" and a key of $K_m = 7$ using multiplicative cipher.

Solution:

Step 1: Key Stream: 'MOBILE'

Letter	Value
M	12
O	14
B	01
I	08
L	11
E	04

Step 2: Using Formula: $C_i = (P_i + K_i) \bmod 26$

Plain Text	P's values	Key Stream	HC's values	half Cipher Text
i	08	12	20	U
a	00	14	14	O
m	12	01	13	N
s	18	08	00	A
t	19	11	04	E
u	20	04	24	Y
d	03	12	15	P
y	24	14	12	M
i	08	01	09	J
n	13	08	21	V
g	06	11	17	R

Half Cipher text is " UONAEYPMJVR"

Step 3: Using Formula: $C = (C_h * K_m) \bmod 26$

half Cipher Text	HC's values	cipher text	Cipher text
U	20	10	K
O	14	20	U
N	13	13	N
A	00	00	A
E	04	02	C
Y	24	12	M
P	15	01	B
M	12	06	G
J	09	11	L
V	21	17	R
R	17	15	P

The original Cipher Text: "KUNACMBGLRP"

Step 4: using the formula $P_h = (C * K_m^{-1}) \bmod 26$.

Cipher text	C's values	P_h's values	Key Stream	P's values	Plain Text
K	10	20	12	08	I
U	20	14	14	00	A
N	13	13	01	12	M
A	00	00	08	18	S
C	02	04	11	19	T
M	12	24	04	20	U
B	01	15	12	03	D
G	06	12	14	24	Y
L	11	09	01	08	I
R	17	21	08	13	N
P	15	17	11	06	G

Here, $K_m = 7$ so $K_m^{-1} = 15$. Then use the plaintext value $P = (P_h - K_i) \bmod 26$.

The message is "I am studying".

D. Frequency Analysis

In the Vigenere-Multiplicative Cipher, there is less percentage to discover shifted plaintext frequencies. But in Vigenere Cipher, once every letter in the key is known, the cryptanalyst can simply decrypt the cipher text and reveal the plaintext.

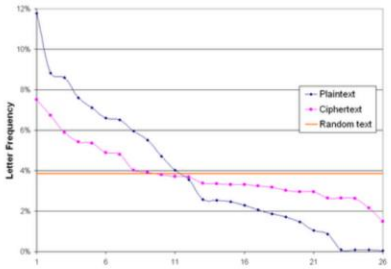


Fig. 8: Letter Frequency in Vigenere Cipher

For Example, if P is the most frequent letter in a cipher text whose plaintext is in English, one might suspect that P corresponds to E, because E is the most frequently used letter in English. But in our Vigenere-Multiplicative Cipher, There is random Number of cipher Text of Letter P.

Another weakness of the Vigenere cipher is the repeating nature of its key. If a cryptanalyst correctly guesses the key's length, then the cipher text can be treated as interwoven Caesar ciphers, which individually are easily broken.

We solve this problem by using multiplicative cipher. That's why; an intruder doesn't break the cipher text. For doing that, an intruder must know the key number then the key stream. This is quite impossible to find out two key and figure out the original message.

E. Time consumption analysis

In a network, network traffic cannot be held by the cipher. But a cipher should be helpful to do its job properly. We know, Vigenere algorithm takes a little time to scramble the message/ information. Multiplicative cipher takes a little time more than Vigenere cipher.

Individually, the algorithms take a little time. But in composition, it takes time to encrypt the data. But our

information is more valuable as we need more secure algorithm like Vigenere- Multiplicative Cipher.

F. Comparison with pros and cons

	Vigenere Cipher	Multiplicative Cipher	Vigenere-Multiplicative Cipher
Key Stream	single or multiple key	Mostly use single key	two key
Security strength	low	high	More secure
Half cipher text	No	No	Yes
Mathematical Function	Simple add function	Multiplicative inverse needed	both function needed
Time assumption	Less than multiplicative cipher	less than Vigenere-multiplicative cipher	less for same key value
Frequency analysis	sometime same cipher text for given plaintext	secure than Vigenere cipher	different cipher text for different plaintext

VII. Security Attacks [7]

When we want benefit from some resource, there should be problems to retrieve the information at risk. People benefits from the internet but there is always security attacks. The attacks may be, such as stealing the user names, passwords, credit card details, social security numbers, personal identification numbers, or any others details which can be used and have the benefits and services.

These networks and data are vulnerable to any of the following types of attacks if we do not have a security plan.

A. Sniffing: as use of Eavesdropping

An attacker always tries to gain the access of the information. If the data is in unsecured form, the information may be hacked. By interpreting the traffic, an intruder finds the data paths of the desired network. When an attacker is using eavesdropping on communications, it is known as sniffing or snooping.

B. Data Modification – intruder's action

For modification of the data, an intruder tries to find out the proper access on the network and read the information. Then the intruder can modify the data in the packet to break the confidentiality of the information.

C. IP Address Spoofing- Routing falsely

IP address of a computer is a valid entity to build up the trust between the networks and the operating systems. In the mean time, it is possible for an attacker to modify IP address to be falsely assumed. Most of the case, an intruder use special programs to construct IP packets that appear to originate from valid addresses inside the corporate internet. When the attacker gain the access to the network with a valid IP address, the attacker can modify, reroute, or delete data.

D. Password-Based Attacks

Password based attacks occurs in the operating systems and in the networks. That means our access is at risk in and out. But using proper utilization of validation we can hide our data by means of usernames and passwords.

When an attacker finds a valid user account, the intruder thinks himself as the real user. If the user has administrator-level rights, the intruder can do anything. When an intruder gains access to a network with a valid account he/she use the network falsely.

E. Denial-of-Service Attack

The denial-of-service attack prevents normal use of computer or network by valid users. The intruder tries to jam the server by creating some programs that send request to the server continuously. That's why, Normal server users find the server busy and can't gain access. These types of attack are called Denial-of Service attacks.

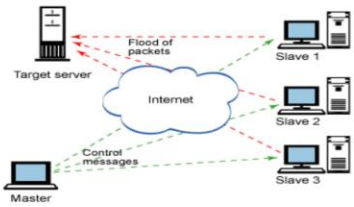


Fig. 9: Denial of service attacks

F. Man-in-the-Middle Attack

This type of attack is done by the intruder who is awake to monitor, capture the information which is transferred between the sender and the receiver. At the primary level of OSI reference model, it is easy for the man in the middle to find out the communication between the informer and receiver.

G. *Sniffer Attack*

In this scenario, the attacker monitors the data between the computer of the sender and the server. He collects data about the shopper or steals personal information, such as credit card numbers. If the packets are not encrypted, a sniffer provides a full view of the data inside the packet.

VIII. Encryption Standards

A. *Data Encryption Standards (DES)[3]*

The most commonly used encryption programs are based on the Data Encryption Standard, which is implemented in 1977 by National Bureau of Standards (NBS) .The algorithm that is used for the data is known as Data Encryption Algorithm. It is a product cipher that operates on 64-bit blocks of data, using a 56-bit key.

The possible combination for the 2^{256} is over 72,000,000,000,000,000 keys. It is considered secured, but now days the speed of the computers increased tremendously. To break this key today's computer take very short time.

B. *Advanced Encryption Standards (AES)*

For protecting electronic data, Advanced Encryption Standards (AES) is the FIPS approved cryptographic algorithm. It is a block-cipher that encrypts and decrypts the information. By encryption process, the information is converted to an unintelligible form. At receiver point, the unintelligible form is converted to original information by decryption process. The AES algorithm is capable of using cryptographic keys of 128, 192, and 256 bits to encrypt and decrypt data in blocks of 128 bits. [6]

To perform AES algorithm, rearrangement of data and replace of data plays the important role. To rearrangement of data, we have to do permutations. Several techniques are used to do these mathematical functions. Moreover, to replace of data is called substitutions.

IX. Encryption in future: Aspect of Cryptography

From a view of technology, encryption works as operational perspective with best result. For data security, we can leave our data scrambled forever. If we do this, it will be the worst case of aspect of cryptography.

A successful enterprise security deployment will be the next target in the sense of developing encryption-decryption technique. Now, we have a problem like if a key is lost, access to all of the data originally is lost. After that, our data is totally at risk at that time. Special thought about key management should be developed. If we want to secure our electronic data with shredding, we have to create backup of the keys. [5]

X. Conclusion

An intruder always tries to retrieve the valuable information so that they can use this information falsely. Hence, computer network security with cryptography provides the remedy. Our Vigenere-Multiplicative cipher is one of our try to manage the data security with confidentiality. Hence, much more advanced security measures would be more useful. So always keep our eye on network security as it is much more important.

REFERENCES

[1] Text book Andrew s. Tanenbaum, Computer networks, fourth edition, 2003.

[2] Encryption- http://searchsecurity.techtarget.com/definition/encryption .

[3] Dr. Kamaljit I. Lakhtaria " Protecting Computer Network with Encryption Technique: A Study" published in International Journal of u- and e- Service, Science and Technology Vol. 4, No. 2, June, 2011.

[4] Text book Behrouza Forouzan, Data Communications and Networking, fourth edition, 2006.

[5] The future of Encryption http://www.net-security.org/article.php?id=1113&p=3S .

[6] Announcing the ADVANCED ENCRYPTION STANDARD (AES) Federal Information Processing Standards Publication 197 November 26, 2001.

[7] Common types of network attacks- http://www.technet.microsoft.com/en-us/library/cc959354.aspx .

[8] Comparing suitable network security http://blagovision.org/comparing-suitable-network-security-keys-kerberos-and-pki/

[9] Chey Cobb, Cryptography for Dummies the PKIPrimer, John Wiley and Sons, 2004.

APPENDIX E: SOFTWARE DEVELOPMENT

E1 ModelMine: A Cloud-Based Software to Mine Models from Open Source Repositories

Website URL: http://www.smreza.com/projects/modelmine/

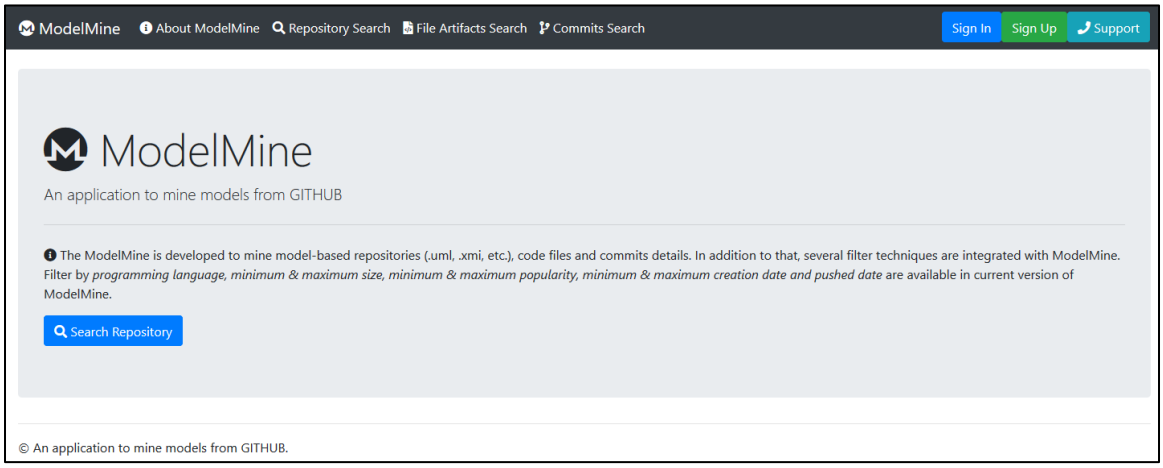


Figure 1: Homepage view of ModelMine tool

Brief Synopsis: Mining Software Repositories (MSR) has become a powerful research tool that opens a new and rich source of research data. This field enables researchers to recover data about defects, development activities, processes, patterns and more. Available mining tools are geared towards data extractions and analysis from textual artifacts.

This tool, ModelMine, a novel software repository mining tool that focuses on mining model-based artifacts. ModelMine is designed particularly to address software engineering researchers who aim to mine software repositories to uncover data about software design practices in the open-source communities. ModelMine supports features to identify repositories based on their level of modeling artifacts and querying repositories to extract models and design artifacts. It supports phase-by-phase caching of intermediate results to speed up the processing to enable the mining of many repositories.

E2 PROCONF - A Conference & Peer Review Management System

Website URL: https://proconf.org/

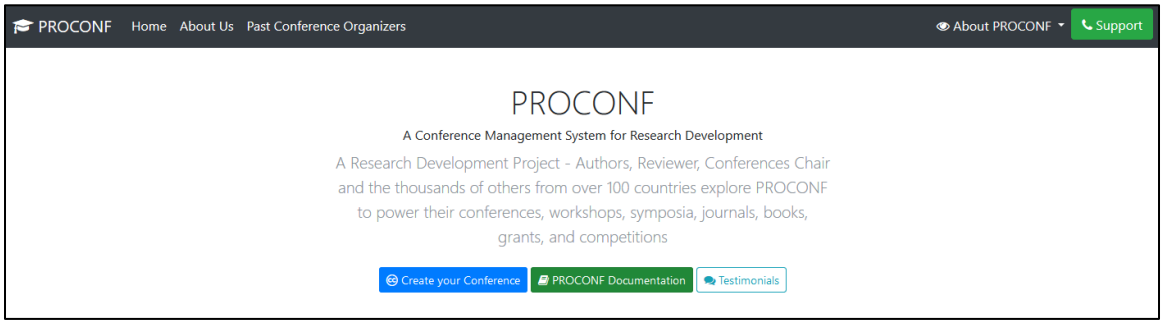


Figure 2: Homepage view of PROCONF system

Abstract: PROCONF is a web-based conference & peer to peer review management system to organize paper submission and review. PROCONF is widely used in Asian conferences, since 2014, in the scientific

community, with reportedly more than 0.1 million researchers in the world. The software also provides an open access online publication service for conference proceedings. When launched, in 2014, the service was for conferences in Bangladesh only due to some business legislation, but in 2015 it was expanded to all other countries. Around 20 conferences are successfully managed & more than 80K research documents are peer-reviewed by PROCONF. It is like other event management system software such as EasyChair developed by Andrei Voronkov and OpenConf hosted by the Department of Computer Science at the University of Manchester.

E3 Heat Island Effect for Borderland Area

Website URL: http://www.smreza.com/projects/heat-island-effect/

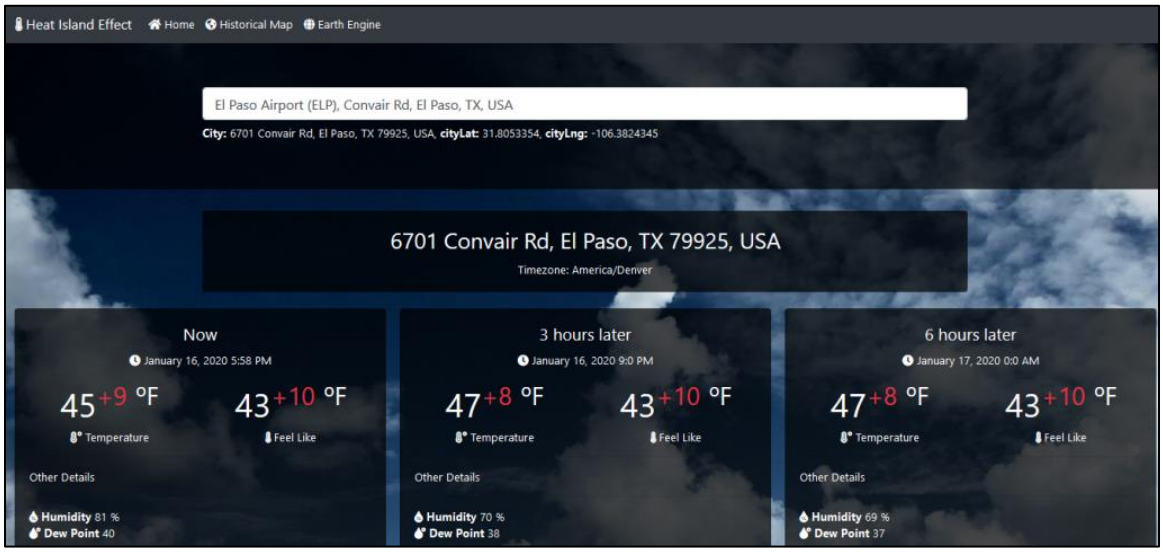


Figure 3: Homepage view of Heat Island Effect application

Abstract: An Urban Heat Island (UHI) is an urban area or metropolitan area that is significantly warmer than its surrounding rural areas due to human activities. The temperature difference is usually larger at night than during the day and is most apparent when winds are weak. UHI is most noticeable during the summer and winter [1]. The term "heat island" describes built-up areas that are hotter than nearby rural areas. The annual mean air temperature of a city with 1 million people or more can be 1.8–5.4°F (1–3°C) warmer than its surroundings. In the evening, the difference can be as high as 22°F (12°C) [2]. Heat islands can affect communities by increasing summertime peak energy demand, air conditioning costs, air pollution and greenhouse gas emissions, heat-related illness and mortality, and water pollution [3].

Major impacts of heat island are increased energy consumption; elevated emissions of air pollutants and greenhouse gases; compromised human health and comfort; and impaired water quality [4]. In our project, the surface temperature with local metrological data is analyzed and visualized in the form of a web application where users can search their area and see the heat island effect in temperature value. The project also visualizes the historical map of heat index over the years (2002-2019). Besides, the project includes the Google Earth Engine mapping system to visualize the heat index map.

E4 [Software Design] Customizable Resume Maker Based on Job Requirements
Website: http://www.smreza.com/projects/hci-vita/

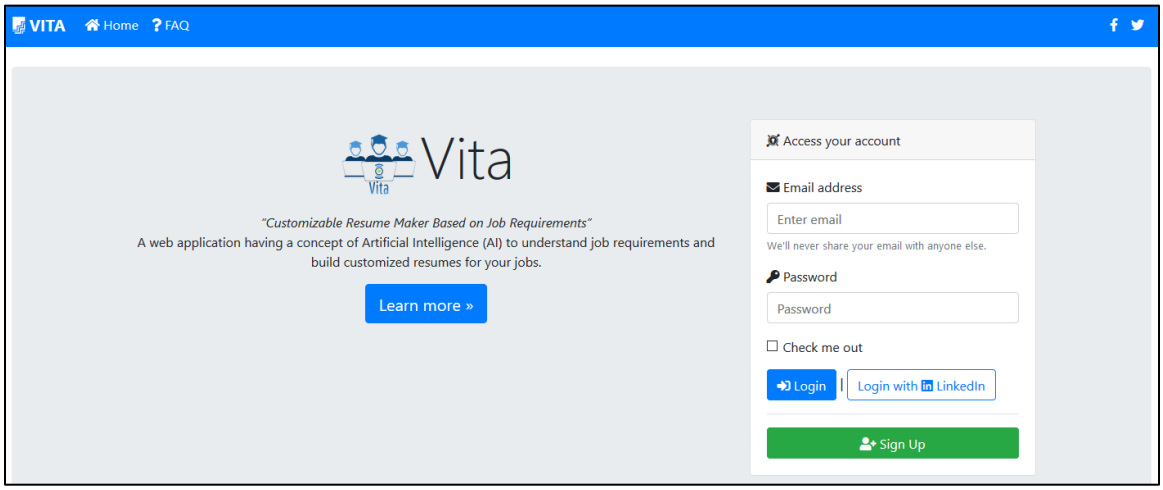


Figure 4: Homepage view of VITA software design

Abstract: Resume and CV are the documents to show a person's educations, experiences, skills, achievement, etc. Resumes and CVs are the way a recruiter can evaluate a person's qualifications for a job. By presenting well the skills and qualifications, the chance of getting the job will increase. So, the skill of presenting person skills and qualifications is very important to get hired in a company. Usually, people go to a resume reviewer to get advice and revision on their resume or CV. It will cost a person around $100 for each appointment.

The average number of people who apply for any given job is 118. Twenty percent of those applicants get an interview1. So, the users must apply for more than a job to increase the chance of hiring. On average, a person has an 8.3% probability of getting a job interview from one job application. That means it takes 10-20 applications to get one interview. And, on top of that, it takes 10-15 interviews to get one job offer1. This means a person must apply for at least 100 jobs to be sure he/she is getting the job offer. Each job has its job requirements and demands for specific skills and experience. Users can increase their chances by customizing their resume based on each job application. But it will be costly if users want to meet a resume maker to customize their resume for each job.

The purpose of building this system is to help users to prepare their resume and CV online and help them to customize their resume based on any job requirements. The system will help users to import their data easily by offering them the functionality of importing the data from LinkedIn and let them add more data manually. Then it will ask the user about the job requirements and suggest a different type of resume or CV based on that job application. Users don't need to add their data in the system in the next attempts, because their data will be stored in the system database. They just need to add a new job requirement to get a new resume or CV for that specific job. This can save time for the users, improve their resume qualification, and helps them to customize their resume for each job very fast. By reducing the time of preparing a resume, we can help the users to apply for more jobs and increase their chances to get a job faster.

REFERENCE

[1] Chris Anson (2016). A Guide to College Writing. North Carolina State University: Pearson.

[2] Mike Palmquist, Barbara Wallraff (2018). In Conversation: A writer's guidebook. Colorado State University: Bedford/St. Martin

[3] Becker, Howard S. Writing for social scientists: How to start and finish your thesis, book, or article. University of Chicago Press, 2008.

[4] Flemons, Douglas. Writing between the lines: How to compose riveting social science manuscripts. WW Norton & Company, 2001.

[5] Reza, S.M., Rahman, M.M., Parvez, M.H., Kaiser, M.S. and Al Mamun, S., 2015, May. Innovative approach in web application effort & cost estimation using functional measurement type. In 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT) (pp. 1-7). IEEE.

[6] Reza, S.M., Rahman, M.M. and Al Mamun, S., 2014, April. A new approach for road networks - A vehicle XML device collaboration with big data. In 2014 International Conference on Electrical Engineering and Information & Communication Technology (pp. 1-5). IEEE.

[7] Van Emden, E. and Moonen, L., 2002, October. Java quality assurance by detecting code smells. In Ninth Working Conference on Reverse Engineering, 2002. Proceedings. (pp. 97-106). IEEE.

[8] Song, Q., Jia, Z., Shepperd, M., Ying, S. and Liu, J., 2010. A general software defect-proneness prediction framework. IEEE transactions on software engineering, 37(3), pp.356-370.

[9] Pelayo, L. and Dick, S., 2007, June. Applying novel resampling strategies to software defect prediction. In NAFIPS 2007-2007 Annual meeting of the North American fuzzy information processing society (pp. 69-72). IEEE.

[10] F. Arcelli Fontana, M. V. Mäntylä, M. Zanoni, A. Marino, "Comparing and experimenting machine learning techniques for code smell detection", Empirical Software Engineering, vol. 21, no. 3, pp. 1143-1191, 2016.

[11] Fontana, F.A., Ferme, V., Zanoni, M. and Yamashita, A., 2015, May. Automatic metric thresholds derivation for code smell detection. In 2015 IEEE/ACM 6th International Workshop on Emerging Trends in Software Metrics (pp. 44-53). IEEE.

[12] Palomba, F., Zanoni, M., Fontana, F.A., De Lucia, A. and Oliveto, R., 2016, October. Smells like teen spirit: Improving bug prediction performance using the intensity of code smells. In 2016 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 244-255). IEEE.

[13] Fontana, F.A. and Zanoni, M., 2017. Code smell severity classification using machine learning techniques. Knowledge-Based Systems, 128, pp.43-58.

[14] Liu, H., Jin, J., Xu, Z., Bu, Y., Zou, Y. and Zhang, L., 2019. Deep learning-based code smell detection. IEEE Transactions on Software Engineering.

[15] Paiva, T., Damasceno, A., Figueiredo, E. and Sant'Anna, C., 2017. On the evaluation of code smells and detection tools. Journal of Software Engineering Research and Development, 5(1), p.7.

[16] Saarimaki, N., Baldassarre, M.T., Lenarduzzi, V. and Romano, S., 2019, August. On the Accuracy of SonarQube Technical Debt Remediation Time. In 2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA) (pp. 317-324). IEEE.

[17] Tempero, E., Anslow, C., Dietrich, J., Han, T., Li, J., Lumpe, M., Melton, H. and Noble, J., 2010, December. The Qualitas Corpus: A curated collection of Java code for empirical studies. In 2010 Asia Pacific Software Engineering Conference (pp. 336-345). IEEE.

[18] Sayed M. Reza, ModelMine: A Tool to Mine Models from Open Source Repositories, http://www.smreza.com/projects/modelmine/.
[19] Transue, S., Reza, S.M., Halbower, A.C. and Choi, M.H., 2018, March. Behavioral analysis of turbulent exhale flows. In 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) (pp. 42-45). IEEE.
[20] Rahman, S., Sharma, T., Reza, S.M., Rahman, M.M. and Kaiser, M.S., 2016, December. PSO-NF based vertical handoff decision for ubiquitous heterogeneous wireless network (UHWN). In 2016 International Workshop on Computational Intelligence (IWCI) (pp. 153-158). IEEE.
[21] Paul, M.C., Sarkar, S., Rahman, M.M., Reza, S.M. and Kaiser, M.S., 2016, January. Low cost and portable patient monitoring system for e-Health services in Bangladesh. In 2016 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-4). IEEE.
[22] Mirzaei, P. A. (2015). Recent challenges in modeling of urban heat island. Sustainable cities and society, 19, 200-206.
[23] Heat Island Effect, https://www.epa.gov/heat-islands.
[24] Heaviside, C., Macintyre, H., & Vardoulakis, S. (2017). The urban heat island: implications for health in a changing environment. Current environmental health reports, 4(3), 296-305.