



2014
Teaching Excellence Award
Finalist Portfolio

Huzefa Rangwala
Computer Science



Table of Contents

Teaching Portfolio

Huzefa Rangwala

- | | | |
|-------------------------------------------------------------------------------------|---|---------------------------------------------|
|  | 1 | Introduction |
|  | 2 | Table of Contents |
|  | 3 | Teaching Statement |
|  | 4 | Teaching Vita |
|  | 5 | Reflections and Evidence |
|  | 6 | Evidence of Teaching Effectiveness & Impact |

Section 1: Cover Letter

Huzefa Rangwala, Ph.D.

I am honored and excited to be the 2014 George Mason University Teaching Finalist. It is gratifying to know that my peers, students and past winners of this valuable award find my activities as a faculty member in the excellent category.

My preference for a faculty position has been driven by a desire to teach, to nurture and inspire students. I am committed to fostering the synergy between teaching and research by providing an environment for all students to develop intellectually and professionally. This philosophy has helped me navigate my life as a tenure-track faculty in the Engineering School where it is necessary to achieve excellence in research for acquiring tenure. As a personal goal I wanted to achieve excellence in both teaching and research.

The teaching portfolio provides a window into my journey from a student to a faculty member; always experimenting with new ideas, always learning and importantly always inspiring my students and peers. This portfolio includes examples of the different activities I try out in my classes. It documents how I have engaged graduate, undergraduate and high school students in activities within and outside the classroom that require them to think creatively, discover and develop innovative computing approaches for tackling grand challenges in big data analytics and genome sciences. It also showcases my desire to bring interesting pedagogical approaches to engineering classes and share them with all my colleagues and peers.

As part of developing this portfolio, I requested anonymous endorsement and testimonials letters. I was fascinated by the high response rate. A Computer Science Professor I was most excited by an endorsement letter that I received from an undergraduate student that said "*the biggest advantage of taking Professor Rangwala's course was in learning how to teach*". I am thrilled to know that students understand the material I teach. I am more thrilled to know that some of them would like to teach like me. One of my doctoral student (graduated) shared that "*Huzefa has been a great teacher, mentor, and an example of how to approach academic work and life.*" My tenure case is under review. I am happy and proud to know that the promotion and tenure committees at the departmental and engineering school level have voted on my tenure application as excellent in both categories of research and teaching.

I cherish and embrace all my responsibilities as a faculty member. The driving force behind my passion for research, student mentoring, teaching and service to my community is the philosophy that seeks improved ideas and ways to empower my students, collaborators and peers with the ability to better society. I am looking forward to the next several years as a member of the Mason faculty; because Mason has helped me innovate, think ahead and become better.

I hope you enjoy reading this portfolio as much as I have enjoyed putting it together.

-Huzefa Rangwala

TEACHING STATEMENT

[..] one of the biggest advantages of taking Professor Rangwala's course was in learning how to teach. [...] His lively and engaging presentations were a big takeaway, not to mention his friendliness and accessibility towards his students. He also demanded a lot of his students. He expected us to learn and participate in class in a reasonable manner that truly pushed us. These are takeaways that I have kept in my notebook as foundations for what I aim to achieve as a computer science professor" – Computer Science, Mason, 2013.

Since joining George Mason University (Mason) in 2008, I have embraced the entrepreneurial and innovative spirit that defines Mason, consistently making contributions in my role as an academic that exemplify what it means to be an outstanding educator, researcher, mentor and collaborator. I have a strong passion for teaching and mentorship, and my record will demonstrate how I engage graduate, undergraduate and high school students in activities within and outside the classroom that require them to think creatively, discover and develop innovative computing approaches for tackling grand challenges in big data analytics and genome sciences.

My philosophy in teaching and mentorship comes from a continuous process of learning for myself; and a deep sense of gratitude towards my teachers and mentors who have assisted me in my thought process to think independently and achieve a deep understanding of my field of study.

My classroom experiences as a student and an instructor, have led me to envision an ideal lecturer. I believe the best teacher does not have a single trait but adapts to classroom situations and student reactions. As a teacher, I see myself as a motivator, moving students toward critical and independent thinking. I also see myself as a knowledge sharer, helping students understand and study computer science principles and concepts along with practical examples. A reward for my efforts is students being able to formulate and form a strong understanding of the subject matter, seek higher learning, and thus advance the spectrum of the computing field in foundations, principles and applications.

Within all my courses, I implement classroom activities that force students to think critically and participate unequivocally. I design assignments and projects that model real world problems and require students to overcome steep challenges with hard work, creativity and collaboration. These sophisticated and forward-looking pedagogies are, in part, a result of having attending a Preparing Future Faculty program as a graduate student (further demonstrating my commitment to teaching and learning). Despite the greater commitment required for students who take my classes, the course evaluation scores are outstanding, averaging 4.66 out of 5. In comparison, the average evaluation score for the Engineering School was 4.0.

Within my graduate data mining classes, I have implemented several different competition-style assignments. The spirit of competition and "learning by doing" approach pays rich dividends in helping student understanding and mastery of the material. Students are provided with tasks like analyzing a chemical molecule to estimate its toxicity or identifying the positive or negative nature of reviews/comments posted on Amazon.com. Completing these tasks involves creative thinking, programming and engineering to achieve the most optimal solution. I introduce new material every week that helps students advance and improve their working solution. During the 4-week competition, students work in teams to submit their solutions to a server that ranks each of them. A student reported, *"The class project we did was a fantastic example of the difference your teaching makes: you want people to learn concepts by asking them to apply them to a context of their own choosing. This truly engages people."*

Another example of effective use of engaging classroom techniques is noted by a student, "I remember one "experimental" teaching technique that worked out pretty well. The purpose of the exercise was to teach four different multiple sequence alignment algorithms (topic). Each of these algorithms were pretty difficult to understand, and it wasn't reasonable for a student to fully understand each on his or her own within the span of two classes. To get these ideas across more efficiently, Huzefa broke the class into four groups and assigned each group a paper on a different multiple sequence alignment as homework. In the next class, each group discussed their assigned paper. Then groups took turns explaining their papers to the rest of the class. I learned a lot from these group presentations because I was able to hear a description of each method from a classmate in terms that were easier to understand than the standard textbook explanation. Also, I was able to get a much more detailed look at the method in the paper I was assigned because I both discussed the paper with my group and had to distill it down to a few core points that I then had to help present. In general, Huzefa's combination of enthusiasm and constant willingness to try out new ideas to help students learn better impressed me as characteristics of a great teacher"

I routinely evaluate my classes, in order to adapt them for improvement through classroom assessment techniques. Sometimes I request students to report "the muddiest point" for a lecture, which is the portion of the lecture that students found the most confusing or challenging. Armed with this information, I can begin the next class addressing the problem by repeating the discussion or crafting an in-class exercise.

In Fall 2011, I was selected from early engineering faculty across the United States to present at the National Academy of Engineering sponsored Frontiers of Engineering Education Symposium. At this event, I shared these pedagogical practices developed for his engineering and interdisciplinary classes. I also regularly share my best teaching practices at the *Innovations in Teaching and Learning Conference* organized by the Center for Teaching and Faculty Excellence and with my departmental colleagues.

Beyond the formal classroom, I am committed to fostering the synergy between teaching and research by providing an environment for all students to develop intellectually and professionally. I have supervised and mentored ten graduate, nine undergraduate and five high school students from various disciplines. My mentoring approach differs, depending on the level of the student. My approach to working with graduate students revolves around collaboration and learning from one another. I provide enough latitude to my students so that they pursue their own interests but also acts as a guide for helping them achieve their milestones. On the other hand mentoring undergraduate and high school students is more successful when there are well-defined projects that students can complete during a semester period. These research projects are designed to expose students to interdisciplinary research, challenge their intellect and creativity, and provide them with the satisfaction resulting from creating something novel and useful.

I am especially passionate about undergraduate student scholarship and entrepreneurship. With support of the Mason Student as Scholars program and in collaboration with Professors Wang and Allbeck in the Computer Science Department, I am implementing a project wherein inquiry-based learning approaches and undergraduate research are synthesized within the Computer Science curriculum. Specifically, we designed a three-course sequence that first helps students in understanding the "discovery" process of research by following successful research projects completed by departmental faculty or industry partners. The second phase involves "inquiry" where students learn how to ask relevant and critical questions. This involves training the next generation of scientists and industry professionals by exposing them to contemporary grand computing solutions like Google's search algorithm. The final phase allows students to implement a complete research product (be it a software or experiment) and present the results at a student seminar. The successful implementation of these curriculum improvements will result in local industry collaborations where students will serve as clients for sponsored projects. Such initiatives will assist in training the next "doers", "innovators" and "entrepreneurs" who will drive

successful economies for the future.

I have also participated as one of the founding members of the Mason bioengineering department and developed a computational track within the undergraduate bioengineering program. This involved designing a curriculum that advanced the needs of bioengineering from a computing standpoint. I also stepped to the forefront to help create a new graduate program in data analytics and engineering. This program deals with the management, archival, retrieval and analysis of the behemoth quantities of complex data available from commercial and scientific systems. Dr. Rangwala was instrumental in formulating the course sequence and will be offering a new course related to the use of several computers for speeding up the analysis process.

To support research initiatives for undergraduates I teamed with the undergraduate research assessment committee to develop new approaches for measuring the impact of undergraduate research efforts in classrooms due to direct investments by Mason. I volunteered to present introductory research at several across the University events organized for orienteering students and exposing them to their first view of research practices. Within my own home department, I participate on undergraduate studies related committee and recently, developed a web-based resource to match students with mentors within Mason.

I seek to inspire students, provide them with resources for access to research at the university and enable opportunities that may not be available otherwise. Every summer, I invite students from area high schools to participate in a summer research internship. As part of the internship, students design and implement a research project that they are encouraged to present at local and national science fair competitions. In summer 2011, I mentored a high school sophomore, Eric Tao from the local Thomas Jefferson High School of Science & Technology. Eric participated in the National Siemens Science Competition and was a regional finalist. In Fall 2013, Eric joined MIT for his undergraduate education and shared *"Through working with Huzefa, I learned how to think about and approach a problem, how to tackle barriers in my work, and how to create a well-written and comprehensive research paper. I could not have gained this invaluable knowledge elsewhere."*

AVERAGE RATINGS OF TEACHING EFFECTIVENESS AT GMU

Dimensions of Teaching	Average Ratings Across Courses	Engineering School Average (Spring 2013)	Dimensions of Teaching	Average Ratings Across Courses	Engineering School Average (Spring 2013)
Teaching Rating	4.66	4.26	Course Ratings	4.38	4.03
Respect	4.78	4.69	Accessibility	4.81	4.55
Syllabus	4.69	4.48	Assignments	4.66	4.31
Organization	4.65	4.35	Readings	4.45	3.99
Explanation	4.58	4.26	Returned	4.69	4.37
Feedback	4.49	4.13	Material	4.70	4.54
Involvement	4.80	4.27	Stimulating	4.62	4.18

FEEDBACK FROM GRADUATE STUDENT MENTEES

I took Huzefa's computational biology class. Here, he also showed his dedication to teaching by trying new teaching methods and always bringing positive energy to the classroom. The class was started at 7:20 and ended 10pm, and most people - including me - showed up pretty tired. Every evening, I was surprised at how energetic Huzefa was and how, at some point, this energy usually seemed to transfer over to the class. —

[Dr. Rangwala] mentored me for my Masters for over a year, during which, I had the opportunity to work very closely with him. [...] From day one, Huzefa encouraged me to think out of the box and keep myself up to date with state of the art research [...] The relationship that Huzefa nourishes with all his mentees is significantly different from what I commonly knew. Having worked in the Internet industry for the past three plus years, I know now that the way he works has a lot in common with how the co-founders of successful startups work. Thus, the value created is at par with success stories that the digital press glorifies, except that our deliverables were appreciated publications in prominent journals. [...] Beyond successful deliverables, Huzefa cultivated in myself and other mentees the foundational research skills that are invaluable and applicable across all scientific domains. At my workplace, I constantly find myself utilizing this skill set to divide and conquer the most complex problems competently. I now consider myself equipped with an innate desire, confidence, and skill set to venture out and re-apply with perseverance the very same techniques in the industry. His persistent encouragement pushed me to keep adding value, bit by bit. In sum, he made research enthralling and challenging for us through his constant dedication and unparalleled mentorship skills.

Huzefa's passion and vision for data mining and machine learning has been a great source of inspiration for me during my PhD career, and will always be. I admire him for giving me the freedom and encouragement to pursue my own interests, as well as providing me with priceless academic and professional advices. He always offered me effective feedback and suggestions about innovative ideas and how to reach my goals. I always feel more hopeful and passionate after having a meeting with him. [...] he encourages you to learn about new technologies and ideas, recommended me to attend research conferences and meet different people to get their insights and thoughts. Now I am a Data Scientist in a professional industry and I feel proud that I can utilize all the experiences, which I learnt under Huzefa's supervision. Having Huzefa as my mentor, it was a great graduate school experience and I am confident that I can make a positive difference in the world with all the things I have learned from him. -2

FEEDBACK FROM STUDENTS IN CLASS

I took CS 465: Computer Systems Architecture with Dr. Rangwala in the Fall 2012 semester at George Mason University. Dr. Rangwala is a professor that cares deeply about his students. This became clear to me early in the semester – Dr. Rangwala offered to lend me an extra copy of the textbook that he had upon learning that I did not have much money left over to pay for textbooks... Dr. Rangwala is also a fantastic teacher. He explained the material well and often engaged the class using in-class exercises and discussion questions. These greatly helped me in understanding the material [...] class and assembly projects instilled in me a great appreciation for high-level abstractions and programming languages. Dr. Rangwala is a great teacher and I will never forget his kindness and generosity –

Your ability to integrate real-world problems into both your teaching and the assignments was refreshing. All too often instructors cling to ivory-tower theory with a disdain for application. You teach the theory necessary to understand a solution, but never pursue it for its own sake. The focus is entirely on solving tough, meaningful problems for people. Our discussion of PCA is a perfect example. We talked about how it works (by finding orthogonal eigenvectors that maximize the variance) and why it works (more variance = more information; addresses the "curse of dimensionality") and continued to discuss when it is used for actual problems. –**Anonymous Student 1.**

Huzefa, you taught me about data mining this past semester with an excitement for teaching that I haven't seen yet at George Mason, and this was my fourth semester. I was thrilled to be a part of a class that was both informative and fun. I found myself asking if that was possible several

times during the semester. I'm a part-time graduate student and I'm at the age where learning something is more important to me than my grades in a class. I know that good grades don't mean much if you don't remember what you learned when the class is over! The class project we did was a fantastic example of the difference your teaching makes: you want people to learn concepts by asking them to apply them to a context of their own choosing. This truly engages people -- how much more exciting to work on a large project if the work is something one is passionate about. Not only does it capture one's attention, but the concepts learned while working through a problem important to oneself are not quickly lost, I believe. Thank you for a great semester and for engaging us with enthusiasm and encouragement. It's obvious how much you care about teaching. —**Anonymous Student 2.**

Overall, your teaching in class has been amazing. Instead of focusing on what will be on the exam, you teach the class in a way that encourages independent thinking. Most professors teach what is needed to pass the class, but you teach what is important about the subject and tell us to think beyond the confinement of the classroom walls. You managed to make complex things very simple such as how the processor executes machine instructions. You taught how various parts of the computer work and how you can apply that knowledge to various scenarios instead of teaching us to just memorize..... It was an absolute pleasure having you as my professor because I have been able to think beyond my graduation as an undergraduate. I now have goals and ambitions of what I want to do in the computer science field. —**Anonymous Student 3.**

Huzefa Rangwala

George Mason University
Department of Computer Science
4400 University Dr.
Fairfax, VA 22030

Phone: (703) 993-3826
Email: rangwala@cs.gmu.edu
Homepage: cs.gmu.edu/~hrangwal

Education

- 2008 **Ph.D.** in Computer Science, University of Minnesota, Twin Cities
Thesis: Comparative Modeling using Machine Learning. Advisor: Prof. George Karypis.
GPA: 3.972
- 2005 **M.S.** in Computer Science, University of Minnesota, Twin Cities
Thesis: Greedy Window based Protein Sequence Alignments. Advisor: Prof. George Karypis.
GPA: 3.963
- 2003 **B.E.** in Computer Engineering, V.J.T.I, Mumbai University, Mumbai
GPA: 4.0, First Class with Honor Distinction

University Teaching Experience

August 2008-Present: *Assistant Professor.* Dept. of Computer Science, George Mason University, Fairfax, VA, USA.

August 2008-Present: *Affiliate Appointment.* Dept. of Bioinformatics & Computational Biology, George Mason University, Prince William, VA, USA.

August 2011-Present: *Affiliate Appointment.* Dept. of Bioengineering, George Mason University, Fairfax, VA, USA.

September 2006-December 2006: *Instructor.* Dept. of Computer Science, University of Minnesota, Minneapolis, MN, USA.

Honors/Awards (Related to Teaching/Mentorship)

- May 2013** Volgenau School of Engineering (VSE) Outstanding Teacher Award.
- April 2013** George Mason University's Rising Star nominee for Outstanding Faculty Award, sponsored by State Council of Higher Education for Virginia (SCHEV).
- March 2013** NSF Early Career Award (2013–2018).
- January 2013** Featured on Center For Faculty & Teaching Excellence Newsletter, "Spotlight on Innovation in Teaching Faculty".
- May 2012** Outstanding Teaching Faculty Award, Computer Science, George Mason University.
- November 2011** Selected Participant at the National Academy of Engineering (NAE) organized "Frontiers in Engineering Education" symposium.

Pedagogical Development/Curriculum Development

Jigsaws and Competitions: Interactive Ways of Learning (Teaching Table) at the Innovations in Teaching and Learning Conference organized by the Center for Faculty and Teaching Excellence, Sept 25, 2012.

Discussed Active Learning at the Teaching Symposium organized by the Computer Science Department, George Mason University, April 25, 2012.

Jigsaws and Competitions: Learning in Bioengineering and Computer Science Classes at the 2011 Frontiers of Engineering Education Symposium organized by the National Academy of Engineering (NAE). Meeting held in Irvine, CA between Nov 13 to 16, 2011.

Preparing Future Faculty Program: Participant at the University of Minnesota (2007).

Co-developed new undergraduate Bioinformatics concentration for the Applied Computer Science program. (BS ACS BNF), Fall 2009.

Assisted in developing the B.S. in Bioengineering Program, Fall 2009.

New Classes Developed - CS 611: Computational Methods for Genomics and CS 795: Biological Data Mining.

OSCAR: Students as Scholars Track B: Curriculum Development and Revision. PI: Pearl Wang, Co-PIs: Jan Allbeck and **Huzefa Rangwala**, \$21,500, 08/01/2012-07/30/2014.

Teaching Record

Assistant Professor, Department of Computer Science, George Mason University.

Class	Semester	Enrol.	Instructor Eval.	Course Eval.
			() shows Dept. Mean	
INFS 755: <i>Data Mining</i>	Fall 2008	25	4.35 (4.31)	4.10 (4.04)
CS 795: <i>Biological Sequence Analysis</i>	Spring 2009	25	4.38 (4.36)	4.06 (4.11)
CS 795: <i>Biological Data Mining</i>	Fall 2009	13	4.73 (4.25)	4.45 (4.03)
CS 795: <i>Biological Sequence Analysis</i>	Spring 2010	18	4.91 (4.31)	4.55 (4.06)
CS 465: <i>Computer Systems Architecture</i>	Fall 2010	46	4.71 (4.28)	4.26 (4.08)
CS 750: <i>Data Mining</i>	Spring 2011	18	4.72 (4.28)	4.67 (4.09)
INFS 755: <i>Data Mining</i>	Fall 2011	34	4.79 (4.25)	4.52 (4.08)
CS 465: <i>Computer Systems Architecture</i>	Fall 2012	48	4.55 (4.27)	4.10 (4.08)
CS 659: <i>Data Mining</i>	Spring 2013	32	4.81 (4.29)	4.69 (4.13)

Instructor, Department of Computer Science, University of Minnesota, Twin Cities.

Csci 5481: *Computational Techniques for Genomics*, Fall 2006.

Teaching Assistant, Department of Computer Science, University of Minnesota, Twin Cities.

Csci 5481: *Computational Techniques for Genomics*, Fall 2005, 2004.

Student Mentoring

Ph.D. advisees

1. **Mr. Nuttachat Wisittipanit**, Ph.D. in BINF, (Co-Advised with Prof. Gillevet), Spring 2009-Spring 2012. Graduation Date: 03/2012. Thesis: *Machine Learning Approach for Profiling Human Microbiome*.

2. Mr. Zeehasham Rasheed, *Ph.D. in CS*, Spring 2010-Spring 2013. Graduation Date: 04/2013.
Thesis: Data Mining Framework for Metagenome Analysis.
3. Mr. Samuel Blasiak, *Ph.D. in CS*, Summer 2010-Present. Expected Graduation Date: 08/2013.
Thesis: Latent Variable Models of Sequences for Classification and Discovery.
4. Mr. Anveshi Charuvaka, *Ph.D. in CS*, Fall 2009-Present.
5. Ms. Tanwistha Saha, *Ph.D. in CS*, (Co-Advised with Prof. Domeniconi), Summer 2009-Present.
6. Mr. Xin Guan, *Ph.D. in CS*, Spring 2013-Present.

M.S. Thesis Advisees

1. Mr. Azad Naik, *M.S. in CS*, Fall 2012-Spring 2013. *Thesis: Using Multi-Task Learning for Large-Scale Document Classification.*
2. Mr. Syed Mahmood, *M.S. in CS*, Fall 2009-Fall 2012. *Thesis: GPU-based Eulerian Assembly of Genomes.*
3. Mr. Ammar Naqvi, *M.S. in BINF*, (Co-Advised with Prof. Gillevet), Spring 2009-Fall 2009. *Thesis: Understanding and analyzing the human microbiome: taxonomy identification and potential interactions.*
4. Mr. Salman Jamali, *M.S. in CS*, Spring 2009-Summer 2009. *Thesis:: Comment Mining, Popularity Prediction and Social Network Analysis.*

M.S. Project Advisees

1. Ms. Na-Young Shin, *M.S. in BINF*, Fall 2010.
2. Mr. Ronak Shah, *M.S. in BINF*, Spring 2010.
3. Mr. Shanmuga Chiripiralla, *M.S. in BINF*, Spring 2010.
4. Ms. Rithika Ganni, *M.S. in BINF*, Spring 2010.
5. Ms. Meher Boorgula, *M.S. in BINF*, Spring 2010.

Undergraduate Mentees

1. Ms. Anisha Kolla, *B.S. in CS*, Summer 2013-Present. (Sponsor: USRP, OSCAR, GMU).
2. Mr. Liam Dannaher, *B.S. in CS*, Summer 2013. (Sponsor: NSF REU).
3. Mr. Hozaifah Zafar, *B.S. in CS*, Spring 2013-Present. (Sponsor: USRP, OSCAR, GMU).
4. Ms. Emily Eastlake, *B.S. in BENG*, Summer 2012. (Sponsor: SURE Program, Bioengineering, GMU).
5. Mr. Liban Hassan, *B.S. in ECE*, Summer 2011. (Sponsor: NSF LSAMP Program, GMU).
6. Ms. Amanda Zouzolou, *B.S. in BENG*, Summer 2011. (Sponsor: SURE Program, Bioengineering, GMU).
7. Mr. Roderick Tolbert, *B.S. in CS*, Summer 2011. (Sponsor: NSF REU to Rangwala).
8. Mr. Charles Sweet, *B.S. in CS*, Summer 2010-Fall 2012. (Sponsor: UAP Program, GMU).
9. Mr. Minh Bui, *B.S. in CS*, Summer 2010. (Sponsor: UAP Program, GMU).

High School Mentees

1. **Mr. Ashwin Sekar** *Poolesville Maryland High School, Poolesville, MD*, Summer 2013.
2. **Mr. Eric Tao**, *Thomas Jefferson High School, Alexandria, VA*, Summer 2010-Fall 2012.
3. **Mr. Abbas Idris**, *Governor's School, Prince William, VA*, Fall 2011-Spring 2012.

Student Committee Membership (Besides Advisees)

Student	Program	Advisor	Grad. Date
Mr. Cheol Young Park	Ph.D. in Information Technology	Kathryn Laskey	-
Mr. Gautum Singh	Ph.D. in Computer Science	Jana Kosecka	-
Mr. Daniel Saxton	Ph.D. in Statistics	Anand Vidyashankar	-
Mr. Joshua Church	Ph.D. in Information Technology	Ami Motro	-
Mr. Keith Sullivan	Ph.D. in Computer Science	Sean Luke	-
Ms. Nada Baset	Ph.D. in Computer Science	Harry Wechsler	Fall 2012.
Ms. Andmorgan Fisher	Ph.D. in Environmental Sciences	Patrick Gillevet	Fall 2012.
Mr. Sean Smith	Ph.D. in Bioinformatics	Patrick Gillevet	Fall 2011.
Mr. Pu Wang	Ph.D. in Computer Science	Carlotta Domeniconi	Spring 2011
Mr. Andrew Heekin	Ph.D. in Bioinformatics	Patrick Gillevet	Summer 2011.
Mr. Keenan Amundsen	Ph.D. in Bioinformatics	Don Seto	Fall 2009.

Service (Teaching Related)*University and Departmental Service*

- Students as Scholars Assessment and Program Design Subcommittee, George Mason University, Fall 2012-Present.
- Undergraduate Studies Committee, Computer Science Department, George Mason University, Fall 2012-Present.
- Guest Speaker for New Faculty, Center for Teaching and Learning Services, George Mason University, Fall 2009.
- Presentation Teaching Recipes at the Mini-Teaching Symposium organized by the Computer Science Department, George Mason University.

Section 5: Reflections and Evidence

Huzefa Rangwala, Ph.D.

Assistant Professor of Computer Science, George Mason University

I have been at George Mason University for five years as a tenure track faculty in the Computer Science Department, Volgenau School of Engineering. I have thoroughly enjoyed all the different roles played by a faculty member as an educator, mentor, researcher and a colleague. Be it the process of teaching my first class at the University of Minnesota as a graduate student, or advising my first student towards completion of his thesis, or receiving the NSF Career award and mentoring a high school student to compete in the Siemens Science Finals Competition, I have learned and improved. The tenure process, I contend is a performance where one intertwines and embodies these different faculty member roles. I have cherished and enjoyed each one of them and I am excited to be the 2014 Teaching Finalist as it validates my performance as an educator.

I have segmented this section of my teaching portfolio into the following themes that showcase my growth as a teacher, take you into one of my classrooms and show my passion for mentorship and sharing of materials with my colleagues.

- 5.I Computer Science Classes can be “Active”.
- 5.II Solving “Real World Problems” in Class.
- 5.III Inspiring New Generation of Researchers.

5.I: Computer Science Classes can be “Active”

My journey towards a Computer Science Professor

The television set, the cassette player, the speaker and power adapters were all connected with an entanglement of colorful wires to my first Casio pocket computer. I was 12 at that time, and I remember pushing some keys and hearing some beeps and tunes from the speaker and seeing squiggly digital lines on the TV set. I recollect that my wonder and fascination at that time were no less than the gurgling smile my 11-month old son shares with me when I show him the screen of my iPad today. My relatively early introduction to computers was courtesy of my late mother. My mother has been a role model and inspiration for the person, educator and researcher I am today. Besides providing us the latest computers that she could afford, she instilled in my brother and I an appreciation for science, discovery, engineering marvels and an eye towards assisting and aiding others. She was a pharmacist working on drug development at Pfizer wanting to help cure diseases affecting the largest of populations. Even today her quest for betterment of society through work continues to be a guide map for me, whether it is standing in front of a class full of inquisitive freshmen or solving problems in collaborative interdisciplinary teams of graduate, undergraduate and high school students. If not for her tireless efforts and support, I would never have attended the University of Minnesota (UMN). She used her own retirement and healthcare savings to afford me the opportunity to pursue a graduate degree.

My experiences as a graduate student at the UMN transformed me in several positive ways. When I arrived in Minnesota, the plan was to graduate with a masters degree in computer science and earn a well paid salary as a software engineer at one of the bay area companies. This was a career path that would be followed by most of my graduating class from top-notch undergraduate universities in India and would also allow me to give back to my mother. However, with a mind full of curiosity and a willingness to explore I enrolled in a new “Computational Genomics” class my first semester. This particular course was related to unlocking the genetic secrets within our living cells using the power of computing. It was within the first few weeks of attending lectures that I realized the potential of computer science as a discipline and its societal impact. For the first time I realized that computing could help save lives, engineer better fuels and even feed the hungry. I wanted to play a role in these interdisciplinary endeavors and I wanted to learn more. My thirst for scholarship and discovery was kindled and I transferred to the Ph.D. program with a focus on bioinformatics that involves the development of algorithms for problems in biology.

With a strong technical undergraduate education (with little room for studying anything besides computer engineering), bioinformatics provided me the ideal path for exploring new ideas. It was overwhelming to be the only graduate student in a 120-capacity filled Genetics 101 class. I also studied Plant Biology, Genome Technologies and subjects that helped me understand the problems and challenges in designing innovative, elegant, and useful computational solutions. This philosophy persists with me while developing new research ideas and assisting students in formulating their own projects.

During graduate school I signed up for the Preparing Future Faculty Program. Looking back this was one of the best decisions I have taken as a student. We learned about different styles of teaching in higher education. The class discussed several active learning approaches like write-pair-share activities, collaborative group activities and discussed methods for getting feedback from students. There was focus on preparing day one for class and producing a syllabus. The class consisted of students from various disciplines and the discussion would several times veer towards the lack of use of these approaches

within engineering and hard science disciplines. There was a sense that these techniques would be hard to implement within a “Computer Science” classroom. I had seen a few percentage of my professors use some of these ideas. I was itching to get a faculty job and put these ideas into practice. I did not want “Computer Science” students to miss out on the fun. *I hypothesized that computer science classes could be “active”.*

Theory to Practice

I have had the opportunity to experiment different learning approaches in a total of eight graduate classes and three undergraduate classes at Mason. I also had the unique opportunity to be the instructor for a graduate class at University of Minnesota, CS 5481: Computational Methods for Genomics. One of my graduate class, CS 795: Special Topics in Computer Science: Biological Data Mining was a research-based seminar class. Below, I will showcase evidence of different approaches and assessment techniques used in these different classes.

Write/Think-Pair-Share: I have implemented the write pair share activity consistently in all my classes. The activity involves three steps. The activity begins by asking students a specific question. The type of question can vary from problem-solving exercise with one correct answer or questions that can have no unique solution, and could potentially force students to think about multiple approaches/ideas. Students are requested to write their approach and solution on a piece of provided paper. After a few minutes, students are requested to share their answers with their neighbor. Finally, as the instructor I request solutions from pairs that have disagreed or student pairs that have the same answer. The merits of this activity are well documented in the literature but I have noticed over the years that it engages students. It allows students to form bonds with their peers. It increases the chance of involving every student in the class. *To improve this activity I noticed that walking around the classroom, while student participants are writing or sharing their solutions, forces the non-participating students to get involved.*

Since, Fall 2010 I have renamed this activity as “[Write-Pair-Share With your Favorite Neighbor](#)”. This served two purposes. It allowed me to inject humor in day-to-day activities and forced students to participate. No student walked from their neighbor to find another one. As evidence for this activity I present two examples of questions that were used in the undergraduate Computer Systems Architecture Class. They are presented in [Appendix 5.I \(WPS\)](#).

As part of my commitment to designing effective and engaging learning experiences, I actively evaluates my classes in order to adapt them for improvement through classroom assessment techniques. Sometimes I request students to report “the muddiest point” for a lecture, which is the portion of the lecture that students found the most confusing or challenging. Armed with this information, I begins next class addressing the problem by repeating the discussion or crafting an in-class exercise.

Importance of Day One: As a graduate student in the Preparing Future Faculty Program, I read “McKeachie’s Teaching Tips, Twelfth Edition”. The book was certainly an eye opener about several different issues but two things are still fresh in my mind: (i) Day One and (ii) Designing Syllabus. In my past experience as a student, I noticed that almost every instructor I encountered would tell us about the administrative policies, grading policies and structure of the class. Although, important *I plan my first day lecture with the following objectives: (i) First, I want students in the class to get a broad overview of the class topics that would be covered and why they were important. (ii) Second, I wanted students to get a feel for me as an instructor and set forth my expectations of them. (iii) Third, I wanted to discuss some of*

the administrative policies but hopefully, when discussing these policies students would be motivated to stay enrolled in the class.

Currently, I am 33 years of age and it was essential for me to establish control of my class, given that students in my graduate class were older than me. With the years coming ahead, much to my chagrin this problem will disappear. Here are some of the first day impressions that I aspire my students to have of me, that I am focused on student learning, that I care about their performance, that I am approachable, that I am interactive, that I am strict on deadlines associated with class assignments and that coming to class is fun and helpful.

My first semester at Mason (Fall 2008), I walked into my graduate “Data Mining” class. It was the large auditorium class in Innovation Hall. I had surveyed the classroom a couple of days ago to feel a little bit nervous about the platform style podium. On day one, I walked in wearing a miner’s hat (Picture in [Appendix 5.1 \(Day One\)](#)). Students burst out laughing. I was successful in connecting with them. I got them interested and I felt less nervous. In Spring 2013, when I taught “Data Mining” again, I began with an example of how data mining was useful during the 2012 US Presidential Elections. I discussed how the President Obama’s data mining team had figured out by using demographics and other factors, which voters could be targeted for persuasion to vote for them and which voters could be persuaded for grass root based campaigning and fund raising. I present class slides as evidence in [Appendix 5.1 \(Day One\)](#).

The undergraduate class that I have taught, CS 465: Computer Systems Architecture is a senior required class. It is also a class that students find extremely challenging and are generally not interested in it. My goals for day one is to provide students the reason to be in this class (even though, it is required). I focus on the fact that for being a well rounded computer scientist/software engineer and working in the industry it was fundamental to understand how a computer’s hardware and lowest software layer looked like and understanding the same would allow them to build fast and efficient programs. After, all performance of computer systems was always a metric that everyone in the real world cared about. I also present them a vision for the semester that would involve doing real world assignments geared towards building better computer systems. I present as evidence, class slides in [Appendix 5.1 \(Day One\)](#).

I also share examples of class syllabus for my classes that essentially lists the different administrative policies, grading criterion, class topics and a schedule in [Appendix 5.1 \(Day One\)](#). Clearly listing the schedule of assignments, grading policies and topics helps keep the class structured and gives the student the needed information to plan ahead.

Jigsaw Activity: Jigsaw is a collaborative learning activity where students learn a topic of interest by splitting the task. The topic or class module is split among students, so that groups of students are responsible for a particular sub-topic. These students are expected to read, and study the sub-topic at sufficient depth so as to call themselves as experts of the particular topic. During class students first convene in expert groups, discuss the sub-topic among them. After strengthening their ideas about the particular topic students are asked to gather in mixed groups such that every group has at least one expert from all the sub-topics. Collectively each mixed group discusses the entire topic and learning from others. The instructor can have a few minutes of summary activity at the end of the two stages. In terms of preparation, the instructor needs to help the discussion within different groups by asking pertinent questions to individual groups or to the class as a whole.

Traditionally, jigsaws have worked well for classes where discussion material is subjective, and the value of it in engineering disciplines is not completely known. Motivated by the peer review model, I have

designed and implemented the jigsaw activity in all my graduate level classes.

As an example, I set up the jigsaw exercise to allow the class to participate in the discussion and review of several state-of-the-art clustering algorithms in a graduate data mining classes. Two weeks before the class session, students were assigned to read and thoroughly understand one of the five chosen research articles, so as to be considered as experts of one clustering method. During class time, experts belonging to the same group were asked to convene and discuss among themselves the concepts learned. Students were then mixed into groups, such that each group had an article representative to discuss and contrast among themselves the strengths and weaknesses of each of the papers. I specifically, developed questions that mixed groups would answer regarding use of the different algorithms in different test case scenarios.

I used class room survey at the end of the assignment, and found that all 12 responding students (out of 18 class students) liked the jigsaw activity and would recommend use in future classes. Even though, the activity was used once in a semester, students recommended use of this learning strategy twice (50% of respondents) and three times (50% of respondents). 75% of the respondents understood the other papers due to their peers in the mixed groups. A visible advantage of using this activity was that students were eager to participate and read their assigned papers. I present as evidence two samples of this activity in [Appendix 5.1 \(Jigsaws\)](#). Once, this was for the Data Mining Class and the second sample is for Biological Sequence Analysis Class.

The pioneering use of jigsaws for engaging the classroom was noted by a student, *"I remember one experimental teaching technique that worked out pretty well. The purpose of the exercise was to teach four different multiple sequence alignment algorithms (topic). Each of these algorithms were pretty difficult to understand, and it wasn't reasonable for a student to fully understand each on his or her own within the span of two classes. To get these ideas across more efficiently, Huzeфа broke the class into four groups and assigned each group a paper on a different multiple sequence alignment as homework. In the next class, each group discussed their assigned paper. Then groups took turns explaining their papers to the rest of the class. I learned a lot from these group presentations because I was able to hear a description of each method from a classmate in terms that were easier to understand than the standard textbook explanation. Also, I was able to get a much more detailed look at the method in the paper I was assigned because I both discussed the paper with my group and had to distill it down to a few core points that I then had to help present. In general, Huzeфа's combination of enthusiasm and constant willingness to try out new ideas to help students learn better impressed me as characteristics of a great teacher".*

Sharing of Pedagogical Materials I am also eager for my colleagues and peers to try this technique within their classes. As such, in Fall 2011, I was selected from early engineering faculty across the United States to present these ideas at the National Academy of Engineering sponsored Frontiers of Engineering Education Symposium. At this event, I shared the pedagogical practices that I developed for my engineering and interdisciplinary classes. I also regularly share my best teaching practices at the Innovations in Teaching and Learning Conference organized by the Center for Teaching and Faculty Excellence and with his departmental colleagues. In [Appendix 5.1 \(Pedagogy\)](#) I am also presenting the poster I presented at the Frontiers of Education Engineering Symposium and slides on teaching ideas that I presented to my departmental colleagues. I am also presenting a letter of reference from the President of National Academy of Engineering, Dr. Rivest sent to the former Dean of Engineering School, Dr. Griffiths. I am also including a newsletter written by the Center For Teaching and Faculty Excellence about me.

A) A processor has a 32 byte memory and an 8 byte direct-mapped cache. Table 0 shows the current state of the cache. Write hit or miss under the each address in the memory reference sequence below. Show the new state of the cache for each miss in a new table, label the table with the address, and circle the change:

Addr	10011	00001	00110	01010	01110	11001	00001	11100	10100
H/M									

0. Initial state

Index	V	Tag	Data
000	N		
001	Y	00	Mem(00001)
010	N		
011	Y	11	Mem(11011)
100	Y	10	Mem(10100)
101	Y	01	Mem(01101)
110	Y	00	Mem(00110)
111	N		

1.

Index	V	Tag	Data
000			
001			
010			
011			
100			
101			
110			
111			

2.

Index	V	Tag	Data
000			
001			
010			
011			
100			
101			
110			
111			

3.

Index	V	Tag	Data
000			
001			
010			
011			
100			
101			
110			
111			

4.

Index	V	Tag	Data
000			
001			
010			
011			
100			
101			
110			
111			

5.

Index	V	Tag	Data
000			
001			
010			
011			
100			
101			
110			
111			

6.

Index	V	Tag	Data
000			
001			
010			
011			
100			
101			
110			
111			

7.

Index	V	Tag	Data
000			
001			
010			
011			
100			
101			
110			
111			

How many hits and how many misses ?

B) A processor has a 32 byte memory and an 4-way set associative byte cache. Table 0 shows the current state of the cache. Write hit or miss under the each address in the memory reference sequence below. Show the new state of the cache for each miss in a new table, label the table with the address, and circle the change:

Addr	10011	00001	00110	01010	01110	11001	00001	11100	10100
H/M									

0.

Set	Tag	Data	Tag	Data	Tag	Data	Tag	Data
0			0011	Mem(00110)			1010	Mem(10100)
1	0000	Mem(00001)	1101	Mem(11011)	0110	Mem(01101)		

1.

Set	Tag	Data	Tag	Data	Tag	Data	Tag	Data
0								
1								

2.

Set	Tag	Data	Tag	Data	Tag	Data	Tag	Data
0								
1								

3.

Set	Tag	Data	Tag	Data	Tag	Data	Tag	Data
0								
1								

4.

Set	Tag	Data	Tag	Data	Tag	Data	Tag	Data
0								
1								

5.

Set	Tag	Data	Tag	Data	Tag	Data	Tag	Data
0								
1								

6.

Set	Tag	Data	Tag	Data	Tag	Data	Tag	Data
0								
1								

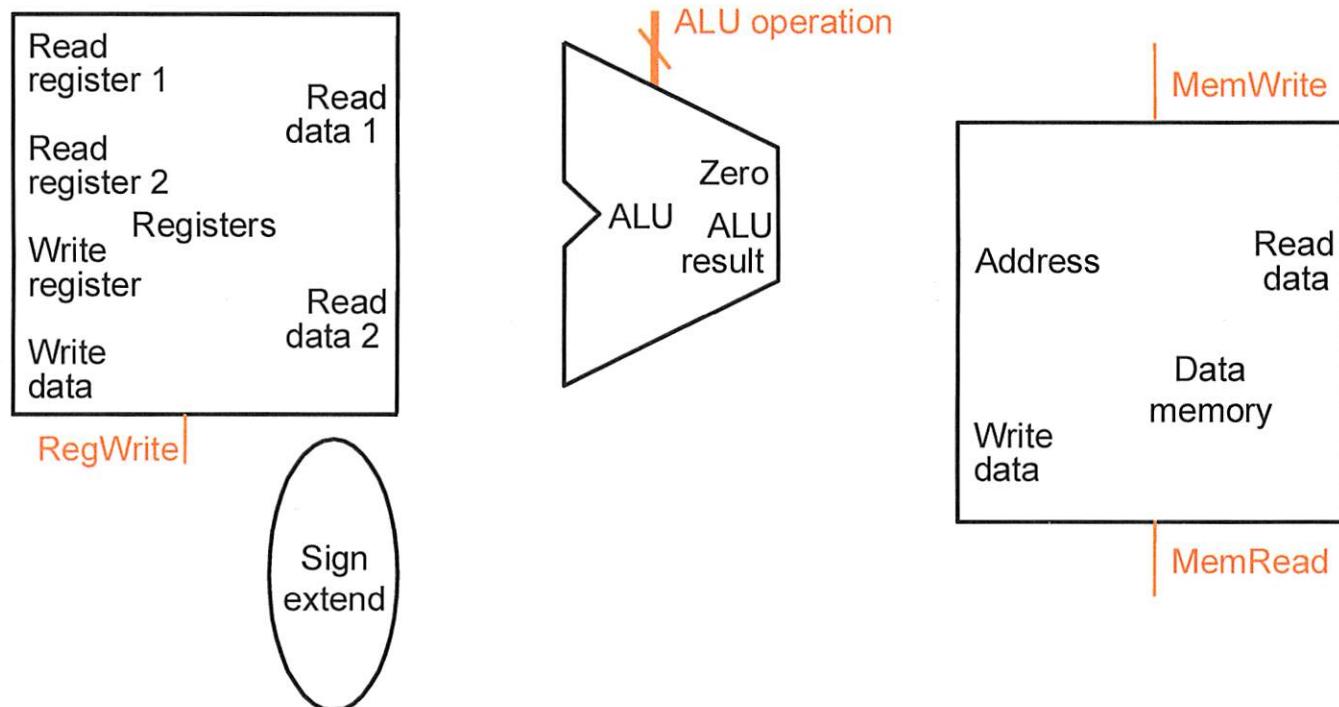
7.

Set	Tag	Data	Tag	Data	Tag	Data	Tag	Data
0								
1								

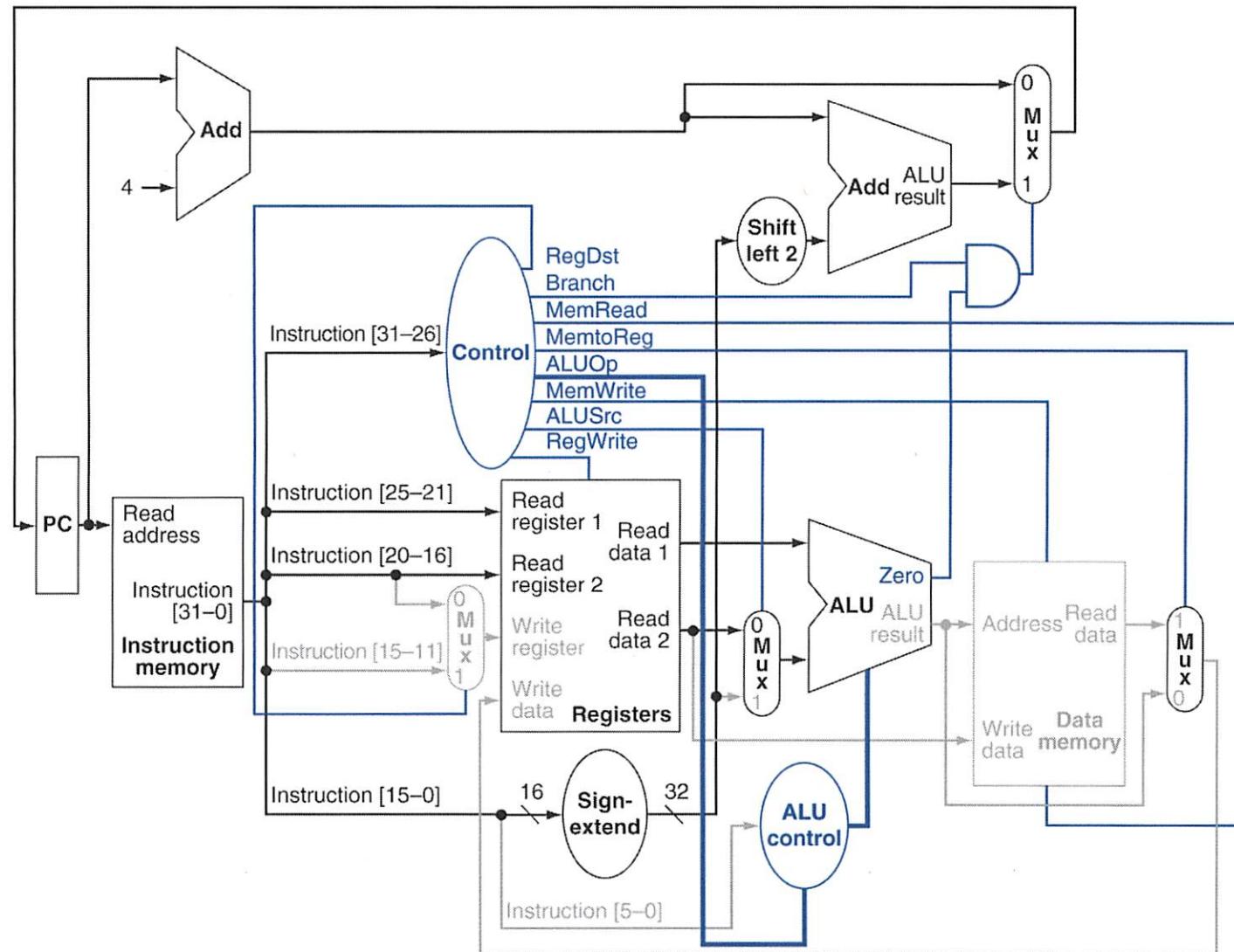
How many hits and how many misses ?

WPS Activity

Question



Solution



FALL 2008 [INFS 755]

HUZEFA IN A MINER'S HAT



Introduction to Data Mining

CS 659

Data Mining
(Spring 2013)

Huzefa Rangwala
Assistant Professor,
Computer Science
George Mason University

Email: rangwala@cs.gmu.edu

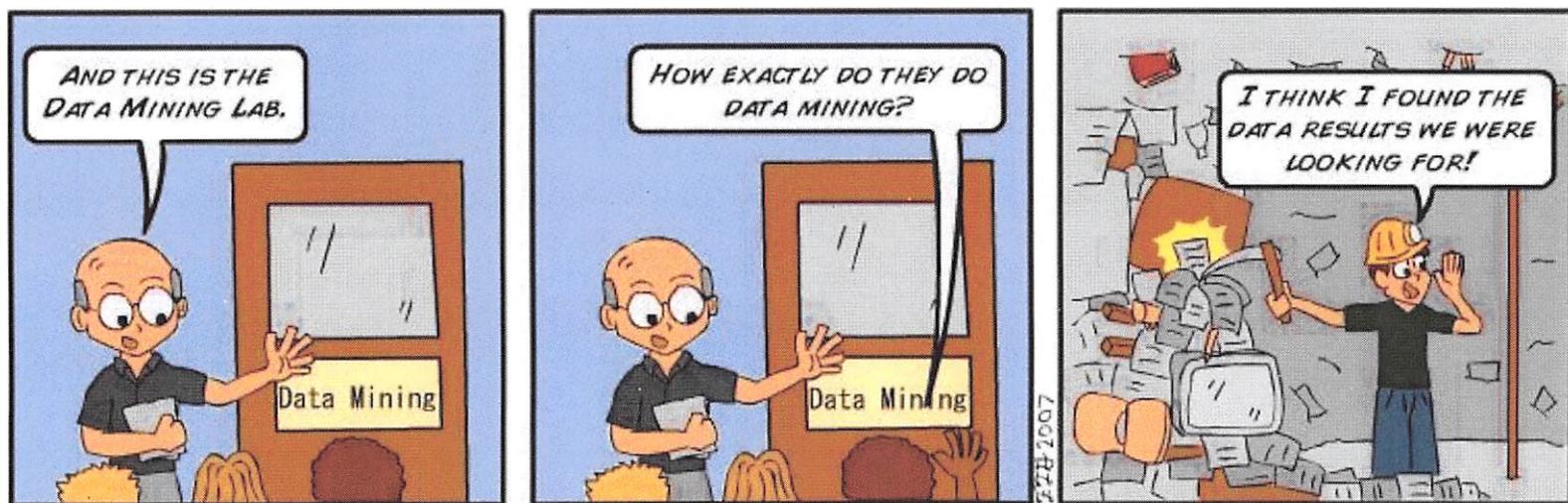
Website: www.cs.gmu.edu/~hrangwal

Slides are adapted from the available book slides developed by Tan, Steinbach and Kumar

Roadmap for Today

- Welcome & Introduction
 - Survey (Show of hands)
- Introduction to Data Mining
 - Examples, Motivation, Definition, Methods
- Administrative/ Class Policies & Syllabus
 - Grading, Assignments, Exams, Policies
- 10-15 minute break.
- Data
 - Lets begin!

What do you think of data mining?



- Please could you write down examples that you know of or have heard of on the provided index card.
- Also write down your own definition.

Election 2012 Data Mining

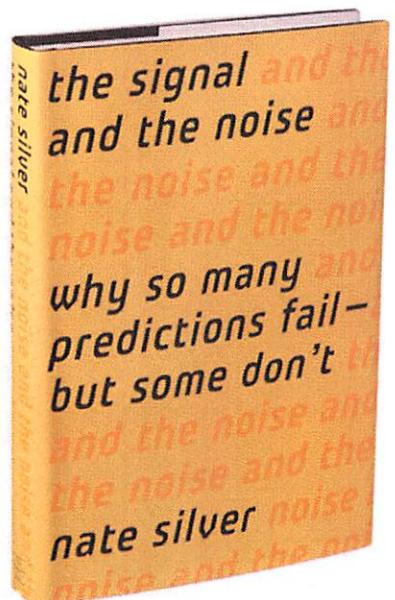
Inside the Secret World of the Data Crunchers Who Helped Obama Win

Read more:

<http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/#ixzz2luhEmNcB>

Mining Truth From Data Babel --- Nate Silver

http://www.nytimes.com/2012/10/24/books/nate-silvers-signal-and-the-noise-examines-predictions.html?_r=0



Data Deluge

<http://www.economist.com/node/15579717>



Brett Ryder

Large-scale Data is Everywhere!

- There has been enormous data growth in both commercial and scientific databases due to advances in data generation and collection technologies
- New mantra
 - Gather whatever data you can whenever and wherever possible.
- Expectations
 - Gathered data will have value either for the purpose collected or for a purpose not envisioned.



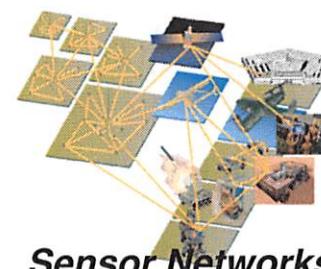
Homeland Security



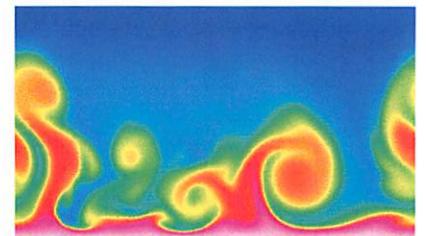
Business Data



Geo-spatial data



Sensor Networks



Computational Simulations

Why Data Mining? Commercial Viewpoint

- Lots of data is being collected and warehoused
 - Web data
 - Yahoo has 2PB web data
 - Facebook has 400M active users
 - purchases at department/grocery stores, e-commerce
 - Amazon records 2M items/day
 - Bank/Credit Card transactions
- Computers have become cheaper and more powerful
- Competitive Pressure is Strong
 - Provide better, customized services for an edge (e.g. in Customer Relationship Management)

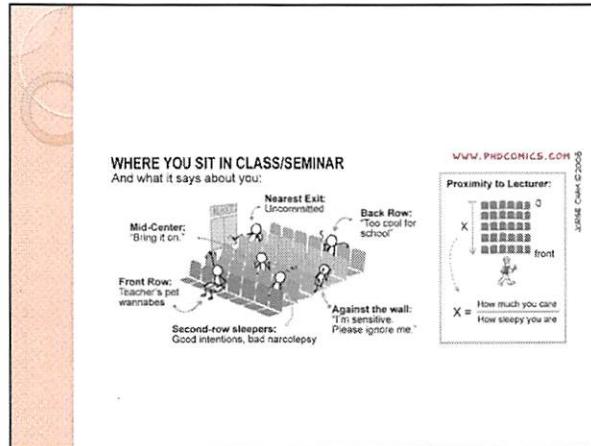


DAY ONE SLIDES

Computer Systems Architecture
Fall 2013

Huzefa Rangwala, PhD
<http://www.cs.gmu.edu/~hrangwal>

*Slides adapted from Computer Organization and Design by Patterson and Henessey



Today ..

- Welcome
- Syllabus
 - Grading
 - Policies
 - Things you want to know and care about.
- Why should you be in this class ?
- What is computer architecture ?
- Demos

• **SYLLABUS**

Location and Office Hours

- Class Time: TThu – 10:30 am to 11:45 am
- Class Location: Arts Building, 2026
- Instructor: Huzeфа Rangwala
 - Office Hours: Tuesday (2:00-4:00 pm) in Engineering Building #4423.
 - Email: rangwala@cs.gmu.edu
- TA: Azad Naik
- Office Hours: W-F(2:00-3:00 pm) in Engineering Building #4457
 - Email: anaik3@gmu.edu

Grading

- | | |
|-----------------------|-------------------|
| • HW 0 | 0% |
| • HW 1 | 10% (Use of MIPS) |
| • HW 2 | 10% (Use of MIPS) |
| • HW 3 | 15% (Use of MIPS) |
| • HW 4 | 15% (Cache/MIPS) |
| • Mid-Term | 20% (Closed Book) |
| • Final Exam | 25% (Closed Book) |
| • Class Participation | 5% |
| • Quizzes (Extra) | 5% |

Class Website and Announcements

- Keep Track of
 - <http://www.cs.gmu.edu/~hrangwal/cs465F2013>
- Specific Readings will be posted after each lecture. Note, the book is detailed and I will make clear what you are responsible for exams/assignments and also what is extra or “if-interested” reading.

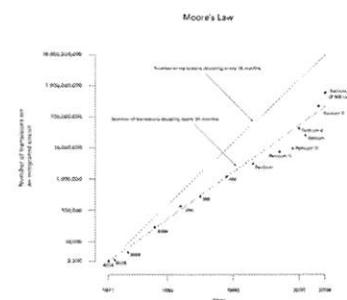
Write-Pair-Share Activity*

- What does computer architecture mean for you ?
 - Think about this individually and write 2-3 bullets. (2 minutes)
 - Meet your closest neighbor and discuss this with him/her (3 minutes)

Why? Besides it being required

- CS majors

Moore's Law

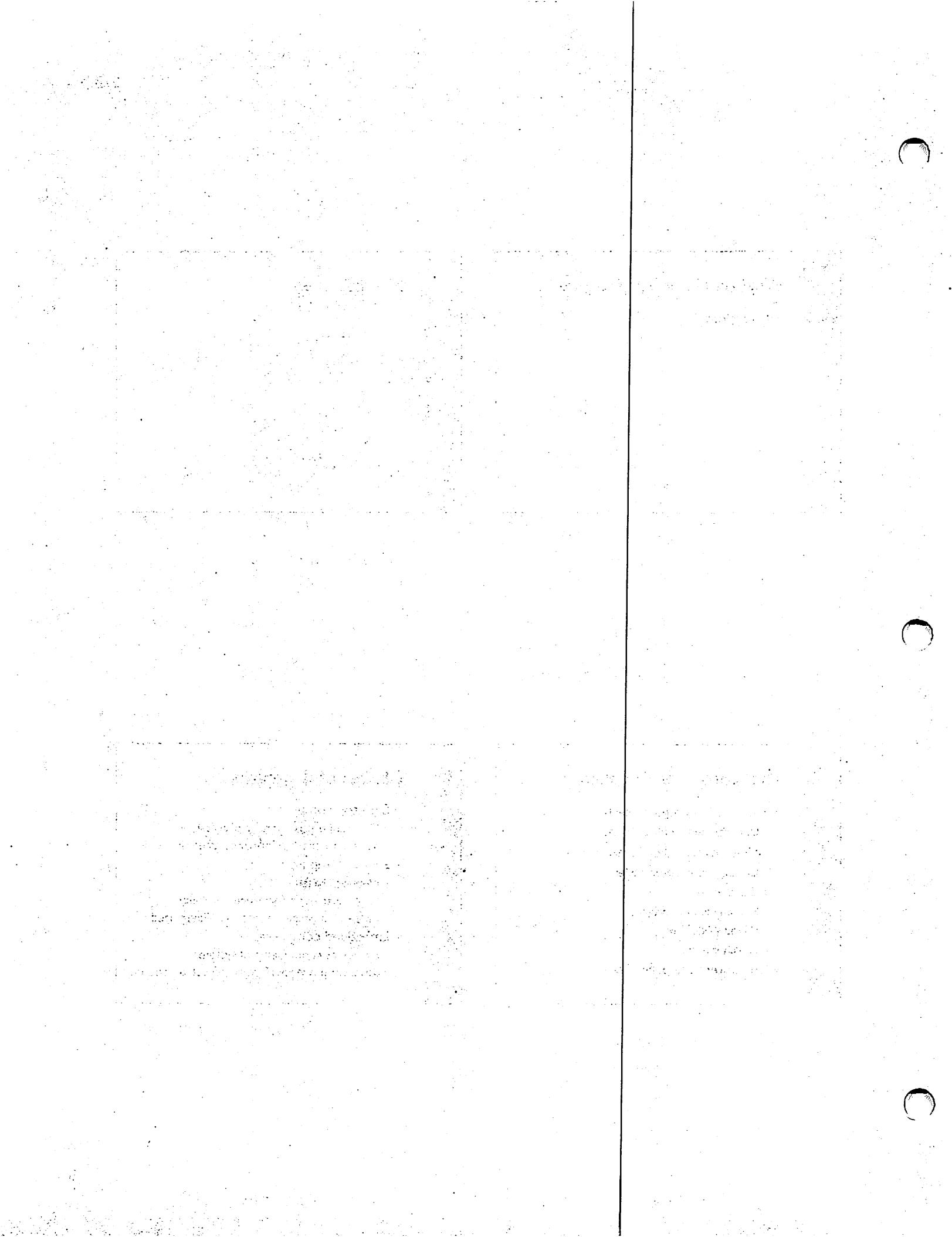


The Computer Revolution

- Progress in computer technology
 - Underpinned by Moore's Law
- Makes novel applications feasible
 - Computers in automobiles
 - Cell phones
 - Human genome project
 - World Wide Web
 - Search Engines
- Computers are pervasive

Classes of Computers

- Desktop computers
 - General purpose, variety of software
 - Subject to cost/performance tradeoff
- Server computers
 - Network based
 - High capacity, performance, reliability
 - Range from small servers to building sized
- Embedded computers
 - Hidden as components of systems
 - Stringent power/performance/cost constraints



Published on *MLBio+ Laboratory* (<http://www.cs.gmu.edu/~hrangwal>)

[Home](#) > CS 465: Computer Systems and Architecture (Fall 2012)

SYLLABUS

CS 465: Computer Systems and Architecture (Fall 2012)

Class Information

Instructor: Huzefa Rangwala [1], Room #4423 Engineering Building,
rangwala@cs.gmu.edu [2]

Time & Location: Planet Hall 206, Tue-Thu 12:00 (noon) - 1:15pm

Text Book: Computer Organization and Design by Patterson and Henessey, Revised Fourth Edition. 2011 [Book's companion website](#) [3]

Teaching Assistant: Azad Naik (anaik3@gmu.edu [4]), Engineering Building # 4456

Office Hours: Instructor: Tue 2:30-4:30 pm

TA: Fri: 3-5 pm or by email appointment, #4457

Please note the syllabus is subject to change to enrich the student's learning experience :). Feel free to email rangwala@cs.gmu.edu [2] for questions, concerns, or even say hi

About the Course

Course Description

This course provides an introduction to the fundamental concepts in computer architecture. Topics include: Basic system components, Performance measurements, Instructions and their representation, Number representation, Implementation of Arithmetic operations, Processor organization, Pipelining, The memory hierarchy

Course Prerequisites

(Computer Systems Programming (CS 367)) / Digital Electronics (ECE 301) OR ECE 303. Students not satisfying the prerequisites will be dropped from the class.

Course Format

Lectures will be given by the instructor. Besides material from the textbook, topics not discussed in the book may also be covered. Grading will be based on homework assignments and exams. Homework assignments will require some programming. Exams and homework assignments must be done on an individual basis unless stated. Any deviation from this policy will be considered a violation of the GMU Honor Code.

Course Outcomes

As an outcome of taking this class, a student will be able to

- Be able to explain the organization of the classical von Neumann machine and its major functional components
- Be able to compare performance of simple system configurations and understand the performance implications of architectural choices
- Be able to show how instructions are represented at both the machine level and in the context of a symbolic assembler; be able to understand small MIPS programs and write MIPS assembly program segments
- Be able to use different formats to represent numerical data and convert numerical data from one format to another
- Be able to explain how an instruction is executed and the concept of datapaths and control
- Be able to explain basic instruction level parallelism using pipelining and the major hazards that may occur
- Be able to explain the effect of memory latency on running time; be able to describe the use of memory hierarchy to reduce the effective memory latency, in particular, the role of cache and virtual memory; be able to understand the principles of memory management
- Be able to explain the basic I/O implementation and data transfers

Detailed Class Schedule with Topics Covered and Assignment due dates will be made [available here](#) [5]

Assignments/Exams Grading

HW 0	0%
HW 1	10%
HW 2	10%
HW 3	15%
HW 4	15%
Mid-Term	20%
Final Exam	25%
Class Participation	5%
Surprise Quizzes (Extra)	5%

Grade Distribution

Grade Score Range

A+	>97
A	92-97
A-	88-92
B+	84-88
B	80-84
B-	76-80

C+	72-76
C	68-72
C-	64-68
F	< 64

Policies:

Attendance

Attendance is not compulsory but highly recommended for doing well in the class. This class has lots of active learning exercises, and they will be a lot of fun. Extra credit i.e., surprise quizzes require classroom attendance and 5% grade is devoted to active discussions in the classroom.

Email Communication:

The best form of communication to the TA and instructor is via email. Please ensure you put "[CS 465]" in the subject header for a timely and quick response.

Assignment Solutions and Exam Solutions

Exam and HW solutions cannot be made available online to prevent future classes from plagiarism. Copies will be made available on the day the graded assignments are returned.

Assignment Submission

Please ensure that the assignments are submitted on-time, before class begins (hard-copy and soft-copy). No late submissions.

Make-Up Exams & Incompletes

Make up exams and incompletes will not be given for this class.

Late to Class

Please arrive to class on-time. Coming late to class is disruptive to your fellow class mates.

Academic Honesty and GMU Honor Code

Please visit the [University's Academic Honesty Page](#) [6] and [GMU Honor Code](#) [7].

Disability Statement

If you have a documented learning disability or other condition that may affect academic performance you should: 1) make sure this documentation is on file with the Office of Disability Services (SUB I, Rm. 222; 993-2474; [www.gmu.edu/student/drc](#) [8]) to determine the accommodations you need; and 2) talk with me to discuss your accommodation needs.

Source URL: <http://www.cs.gmu.edu/~hrangwal/cs465-fall2012>

Links:

- [1] <http://www.cs.gmu.edu/~hrangwal>
- [2] <mailto:rangwala@cs.gmu.edu>
- [3] <http://www.elsevierdirect.com/v2/companion.jsp?ISBN=9780123747501>
- [4] <mailto:anaik3@gmu.edu>
- [5] <http://www.cs.gmu.edu/~hrangwal/node/404>
- [6] http://www.gmu.edu/cte/Teaching/Getting_Started/Designing_Syllabus/academic_honesty.html
- [7] <http://honorcode.gmu.edu/>
- [8] <http://www.gmu.edu/student/drc>

Published on *MLBio+ Laboratory* (<http://www.cs.gmu.edu/~hrangwal>)

[Home](#) > CS 659: Theory and Applications of Data Mining

SYLLABUS

CS 659: Theory and Applications of Data Mining

Class Information

Class/Sec:	CS 659 (002): Theory and Applications of Data Mining
Instructor:	Huzaifa Rangwala [1], Room #4423 EB, rangwala@cs.gmu.edu [2]
Class Time & Location:	Planetary Hall 127, R 4:30pm - 7:10pm
Text Book:	Pang-Ning Tan, Michael Steinbach, and Vipin Kumar <i>Introduction to Data Mining</i> , Addison Wesley, 2006. Book's companion website [3]
Teaching Assistant:	Tanwistha Saha (tsaha@gmu.edu [4])
Office Hours:	Instructor: R 2:00-4:00 pm, TA: Mondays 3-5 pm, Engineering 5321

Please note the syllabus is subject to change to enrich the student's learning experience :). Feel free to email rangwala@cs.gmu.edu [2] for questions, concerns, or even say hi.

If you have taken CS 750 (or INFS 755), then you will not receive credit for CS 659

About the Course

Course Description

Over the past decade there has been an exponential increase in the amount of data. This has lead to development of techniques to discover useful and interesting information from the large collections of data. This course aims to provide a overview of the key data mining methods and techniques like classification, clustering, and association rule mining. The course will also provide interesting application examples of data mining, especially in the field of bioinformatics, social network analysis and "big data".

Course Prerequisites

CS 580 + Programming experience is expected. Students should be familiar with basic probability and statistics concepts, and linear algebra. Please expect some programming in the assignments and class projects. If you are not sure about the pre-reqs send me an email.

Course Format

Lectures will be given by the instructor. Besides material from the textbook, topics not discussed in the book may also be covered. Research papers and handouts of material not covered in the book will be made available. Grading will be based on homework assignments, exams, and a project. Homework assignments will require programming.

Course Outcomes

As an outcome of taking this class, a student will be able to

- Understand the various classification, clustering, association rule-mining algorithms.
- Apply the data mining techniques learned to real world applications.
- Read research papers pertaining to data mining and cloud computing

Schedule

01.24.2013	Welcome, Introduction to Data Mining (Chapter 1)	
01.31.2013	Data (Chapter 2)	[HW1 out]
02.07.2013	Classification (Chapter 4 & 5)	
02.14.2013	Classification (Contd)	[HW1 due]/[HW2 out]
02.21.2013	Clustering (Chapters 7 & 8)	
02.28.2013	Clustering (Contd.)	[HW2 due]/[HW3 out]
03.07.2013	Clustering (contd)/Exam I [15%]	
03.14.2013	Spring Break	
03.21.2013	Association Rule Mining I.	[Proposal Due]
03.28.2013	Association Rule Mining II	[HW3 due]
04.04.2013	Recommender Systems/Social Media Analysis	
04.11.2013	Exam II [15%]/ Bioinformatics Application	
04.18.2013	Massive Data Mining	
04.25.2013	Project Presentations [Could be posters]	
05.02.2013	Project Presentations [Could be posters]	
05.10.2013	Project Report Due Via Email (5:00pm EST)	

Assignments/Exams

Deadline	Type	% Weight
02.14.2013	HW 1	10
02.28.2013	HW 2	10
03.28.2013	HW 3	15
03.07.2013	Exam I	15
04.11.2013	Exam II	15
-	Class Participation	0
	Project	35
03.21.2013	Project Proposal (2 pages)	5
	Project Presentations	10

Grade Distribution

Grade Score Range

A+	>98
A	94-98
A-	90-94
B+	86-90
B	82-86
B-	78-82
C+	74-78
C	70-74
C-	66-70
F	< 66

Policies:

Attendance

Attendance is not compulsory but highly recommended for doing well in the class. This class has lots of active learning exercises, and they will be a lot of fun.

Assignment Submission

Please ensure that the assignments are submitted on-time. No late submissions are allowed.

Make-Up Exams & Incompletes

Make up exams and incompletes will not be given for this class.

Academic Honesty and GMU Honor Code

Please visit the [University's Academic Honesty Page](#) [5] and [GMU Honor Code](#) [6].

Disability Statement

If you have a documented learning disability or other condition that may affect academic performance you should: 1) make sure this documentation is on file with the Office of Disability Services (SUB I, Rm. 222; 993-2474; [www.gmu.edu/student/drc](#) [7]) to determine the accommodations you need; and 2) talk with me to discuss your accommodation needs.

Source URL: <http://www.cs.gmu.edu/~hrangwal/cs659Spring2013>

Links:

- [1] <http://www.cs.gmu.edu/~hrangwal>
- [2] <mailto:rangwala@cs.gmu.edu>
- [3] <http://www-users.cs.umn.edu/~kumar/dmbook/index.php>
- [4] <mailto:tsaha@gmu.edu>
- [5] <http://academicintegrity.gmu.edu/honorcode/>
- [6] <http://honorcode.gmu.edu/>
- [7] <http://www.gmu.edu/student/drc>

BIRCH: An Efficient Data Clustering Method for Very Large Databases

Tian Zhang

Computer Sciences Dept.

Univ. of Wisconsin-Madison

zhang@cs.wisc.edu

Raghu Ramakrishnan

Computer Sciences Dept.

Univ. of Wisconsin-Madison

raghu@cs.wisc.edu

Miron Livny*

Computer Sciences Dept.

Univ. of Wisconsin-Madison

miron@cs.wisc.edu

Abstract

Finding useful patterns in large datasets has attracted considerable interest recently, and one of the most widely studied problems in this area is the identification of *clusters*, or densely populated regions, in a multi-dimensional dataset. Prior work does not adequately address the problem of large datasets and minimization of I/O costs.

This paper presents a data clustering method named *BIRCH* (Balanced Iterative Reducing and Clustering using Hierarchies), and demonstrates that it is especially suitable for very large databases. *BIRCH* incrementally and dynamically clusters incoming multi-dimensional metric data points to try to produce the best quality clustering with the available resources (i.e., available memory and time constraints). *BIRCH* can typically find a good clustering with a single scan of the data, and improve the quality further with a few additional scans. *BIRCH* is also the first clustering algorithm proposed in the database area to handle "noise" (data points that are not part of the underlying pattern) effectively.

We evaluate *BIRCH*'s time/space efficiency, data input order sensitivity, and clustering quality through several experiments. We also present a performance comparisons of *BIRCH* versus *CLARANS*, a clustering method proposed recently for large datasets, and show that *BIRCH* is consistently superior.

1 Introduction

In this paper, we examine *data clustering*, which is a particular kind of data mining problem. Given a large set of multi-dimensional data points, the data space is usually not uniformly occupied. *Data clustering* identifies the sparse and the crowded places, and hence discovers the overall distribution patterns of the dataset. Besides, the derived clusters can be visualized more efficiently and effectively than the original dataset [Lee81, DJ80].

*This research has been supported by NSF Grant IRI-9057562 and NASA Grant 144-EC78.

Permission to make digital/hard copy of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copying is by permission of ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee.

SIGMOD '96 8/96 Montreal, Canada
© 1996 ACM 0-89791-794-4/96/0006...\$3.50

Generally, there are two types of attributes involved in the data to be clustered: *metric* and *nonmetric*¹. In this paper, we consider metric attributes, as in most of the Statistics literature, where the clustering problem is formalized as follows: *Given the desired number of clusters K and a dataset of N points, and a distance-based measurement function (e.g., the weighted total/average distance between pairs of points in clusters), we are asked to find a partition of the dataset that minimizes the value of the measurement function.* This is a nonconvex discrete [KR90] optimization problem. Due to an abundance of local minima, there is typically no way to find a global minimal solution without trying all possible partitions.

We adopt the problem definition used in Statistics, but with an additional, database-oriented constraint: *The amount of memory available is limited (typically, much smaller than the data set size) and we want to minimize the time required for I/O.* A related point is that it is desirable to be able to take into account the amount of time that a user is willing to wait for the results of the clustering algorithm.

We present a clustering method named *BIRCH* and demonstrate that it is especially suitable for very large databases. Its I/O cost is linear in the size of the dataset: a *single scan* of the dataset yields a good clustering, and one or more additional passes can (optionally) be used to improve the quality further.

By evaluating *BIRCH*'s time/space efficiency, data input order sensitivity, and clustering quality, and comparing with other existing algorithms through experiments, we argue that *BIRCH* is the best available clustering method for very large databases. *BIRCH*'s architecture also offers opportunities for parallelism, and for interactive or dynamic performance tuning based on knowledge about the dataset, gained over the course of the execution. Finally, *BIRCH* is the first clustering al-

¹Informally, a *metric* attribute is an attribute whose values satisfy the requirements of Euclidian space, i.e., self identity (for any X , $X = X'$) and triangular inequality (there exists a distance definition such that for any X_1, X_2, X_3 , $d(X_1, X_2) + d(X_2, X_3) \geq d(X_1, X_3)$).

Abstract

Jigsaws and competitions are two active learning approaches that are tailored for classes in inter-disciplinary and computer science classes. Jigsaw is a collaborative learning activity where students learn a topic of interest by splitting the task. Competition allows students to engineer solutions towards building a predictive model for exercises in drug activity prediction, protein sequence classification. The question arises on how to best assess whether these activities are working ? The competition framework encourages students and focuses on learning-by-doing but can be challenging and requires hard work.

Objectives

- Learning-by-doing
- Active Learning Approaches.
- Collaborative Thinking and Engineering New Solutions.
- Introduce several concepts that differ slightly from underlying principles.
- Motivate and engage using spirit of competition.

Background:

Traditionally, jigsaws have worked well for classes where discussion material is subjective, and the value of it in engineering disciplines is not completely known. Motivated by the peer review model, I have designed and implemented the jigsaw activity in four graduate level classes (twice in bioinformatics research classes and twice in data mining classes).

The classes I teach have a predictive modeling component. As a student, having participated in blind protein structure prediction competitions and data mining competitions like KDD Cup, I have implemented these form of competitions in my bioinformatics & data mining classes.



Jigsaws & Competitions

Huzefa Rangwala, Ph.D.

Computer Science, George Mason University

Email: rangwala@cs.gmu.edu

<http://www.cs.gmu.edu/~hrangw/>



Implementation of Jigsaw

Phase I:

Two weeks before class assign Students in expert groups.

Every student in a colored "expert" Group is responsible for reading about The same topic.



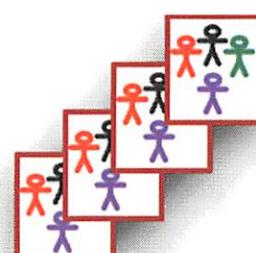
EXPERT GROUPS

Phase II (In-Class):

Students discuss the topic/paper they Were responsible for in expert groups.

Instructor can motivate discussion by asking questions.
E.g., Run-time of algorithms

Could ask students to prepare slides.



MIXED GROUPS

Phase III (In-Class):

Students meet in mixed groups.

Explain each other different algorithms/approaches/papers.

Reason about which approaches work for what cases.

Try to rank different approaches.

Discuss with instructor.

Execution of Jigsaw:

Two Data Mining Classes

Introduced papers related to hierarchical clustering. Focus on algorithmic details, efficiency and accuracy of these algorithms.

Performed Assessment using web-survey.

12 of 18 students responded.

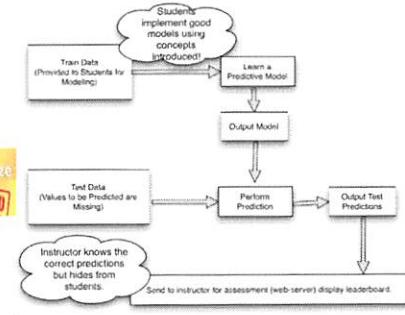
100% of respondents liked the activity.

75% of respondents understood papers due to mixed groups.

50% of respondents would like to have the activity 2 times.

50% of respondents would like to have the activity 3 times.

Also used the jigsaw approach in two bioinformatics classes to introduce multiple sequence alignment algorithms.



Acknowledgements:

IIS 0905117

2011 Frontiers of Engineering Education Symposium

Irvine, California
November 13 - 16

Sponsored by:

The National Academy of Engineering and
The O'Donnell Foundation





NATIONAL ACADEMY OF ENGINEERING
OF THE NATIONAL ACADEMIES



Frontiers of Engineering Education

500 Fifth Street, NW
Washington, DC 20001

January 31, 2012

Dr. Lloyd J. Griffiths
Dean, Volgenau School of Engineering
George Mason University
MS 4A3
4400 University Drive
Fairfax, VA 22030-4444

Dear Dean Griffiths:

We were delighted that Dr. Huzefa Rangwala of the Computer Science Department was able to participate in the National Academy of Engineering's Third Frontiers of Engineering Education Symposium and share some of his educational research ideas with an enthusiastic group of like-minded engineering educators. One of the goals of the symposium is to share ideas with colleagues at the home institution. To this extent, it would be extremely valuable if you could meet with Dr. Rangwala to learn of his experience of the symposium and how what he gained might be best transferred to the Volgenau School of Engineering at George Mason University. Certainly one way would be for him to give a lecture or seminar to the faculty highlighting that experience.

We hope that our program has helped better prepare Dr. Rangwala to be a resource for innovation in engineering education at your institution and even on the national scene. We would also appreciate any feedback you might have regarding his experience at FOEE. We hope that this program will evolve into an important force for continuous improvement in U.S. engineering education. We started it in part as a response to the fact that only 4.5 percent of U.S. university graduates are engineers and that half of all students who declare an intent to major in engineering do not graduate with an engineering degree.

We also look forward to your nominating other faculty for the 2012 FOEE; an announcement to that effect will be forthcoming once the final dates have been set. For your information, the most recent symposium, held November 13 – 16, 2011, in Irvine, CA, brought together 65 innovative engineering educators for two and a half days of discussion, reflection, and collaboration. The aim of the FOEE Symposia is to strengthen the engineering and innovation capacity and capability of the nation by catalyzing a vibrant community of *emerging* engineering education leaders like Dr. Rangwala. We do this by providing a symposium that recognizes accomplishment, facilitates learning, broadens collaboration, and promotes dissemination of best practices in engineering education. The attendees participate in panel presentations, small group discussions with other attendees, mentoring sessions with more experienced faculty, and workshops. In addition, each attendee presents a poster that highlights an engineering education research project. Dr. Rangwala's poster focused on Jigsaws and Competitions.

Dr. Griffiths
January 31, 2012
Page 2

We hope that the FOEE attendees will strengthen their professional capacity for engineering education innovation by:

- identifying and understanding how to apply identified best practices in engineering education;
- developing new ideas to advance their innovations in engineering education;
- developing an understanding that engineering educational innovation should be guided by the evolving evidence-based body of knowledge on engineering learning, in part established through research in engineering education;
- establishing long-lasting professional relationships with those attending the symposiums, and through those relationships establish new or broadened networks with other educational innovators; and
- becoming agents of change to help advance the U.S. capacity for engineering education innovation

We encourage you to support Dr. Rangwala and view him as an important asset for institutional change. One of our goals is to maintain the momentum gained during this year's symposium toward sustainable innovation in engineering education.

Nominations for the 2012 FOEE Symposium will be announced in early April. We encourage you to nominate one or more faculty members for this year's meeting.

Sincerely,



Charles M. Vest
President
National Academy of Engineering



Larry J. Shuman
University of Pittsburgh
FOEE Symposium Chair

CMV/LJS/etc



Center for Teaching & Faculty Excellence

PROMOTE • SUPPORT • CELEBRATE



[Home](#) [Professional Development](#) [Teaching](#) [Awards](#) [About Us](#) [Contact Us](#) [Site Map](#)

Faculty Spotlight on Innovation: Huzefa Rangwala

[Home](#) » Faculty Spotlight on Innovation: Huzefa Rangwala

Originally published in "Notes of Excellence" newsletter, issue 007 – January 2013.



Huzefa Rangwala is an Assistant Professor of Computer Science in the Volgenau School of Engineering. Since arriving at Mason in 2008, he has taught and developed a variety of graduate courses on data mining, computer architecture, and biological sequence analysis. A rising star in his discipline, he won his department's Outstanding Teaching Award in 2012, is the principal investigator for Mason's NSF-funded Machine Learning in Biomedical Informatics (MLBio+) Laboratory, and co-wrote an OSCAR-funded proposal to revise and integrate scholarship into the Computer Science undergraduate curriculum.

What is the most innovative thing you do with your students and/or your classes? Why do you think it is effective?

I design class assignments to supplement the understanding of the subject material. Good assignments are meant to evoke questions that were not raised in class. I believe that a student's performance on assignments is a reflection of the instructor's teaching performance. I have always found that the "learning by doing" idea pays rich dividends in understanding important concepts or principles. I developed assignments where the students develop state-of-the-art algorithms explained in class and compare the performance to available public or commercial versions. The classes I teach have a predictive modeling component. I have participated in blind protein structure prediction competitions (CASP) and data mining competitions like KDD Cup. Motivated by this model, I have implemented these forms of competitions in my bioinformatics and current data mining classes.

Specifically, in fall 2011, INFS 755 students had to train a predictive model to determine if a drug molecule was active or inactive. As part of this competition, the true values were hidden, and the developed models had to make a prediction and submit their results using a web service hosted on Kaggle. An automated program evaluated the results using the true values and provided a ranking of the students based on the predictive performance. To improve the model's predictive performance, students researched and engineered innovative solutions. The spirit of competition and using extensions of techniques learned in class encouraged students to go beyond the typical classroom.

What do you do that creates a strong learning environment for your students?

My lectures focus on interactive and student-centered learning using a wide range of active learning techniques. To train myself with the best active learning strategies, I participated in a semester-long course geared towards preparation of future faculty as a student. One activity I cherish came from middle school teaching.

Jigsaw is a collaborative learning activity. The topic/class module is split into several sub-parts, and a student group is responsible for reading and understanding a sub-part (called "topic"). Students are expected to read and study the assigned topics at sufficient depth to become experts. During class, students convene in expert groups and discuss the topics amongst themselves. After strengthening their ideas about the particular topic, students are asked to gather in mixed groups so that every group has at least one expert from all the topics. Collectively, each mixed group discusses the entire class module and learning from others. The instructor can have a few minutes of summary activity at the end of the two stages. In terms of preparation, the instructor needs to guide the discussion within different groups by asking challenging questions to individual groups or to the entire class. Jigsaw works well for classes when discussion material is subjective, and the value of it in engineering disciplines is not completely known.

What's one tip that you would offer to faculty new to teaching at Mason?

Students love to be engaged. Especially in a two-hour-plus class, it is important that student voices are heard. There are several easy exercises that I can prescribe: 1) write-pair-share activities, 2) quizzes, 3) class room assessments, and 4) jigsaws.

What's the most challenging thing for you in your teaching, and how do you address this challenge?

The biggest challenge is getting feedback about what's working for students' learning during the semester. While lecturing, I elicit responses from students by pausing at several points during the lecture to ask questions about the material presented. I use the write-pair-share activity and classroom assessment techniques (CAT) like surveys and Muddiest Point. Use of CATs allows me to adapt my class through the semester for a richer learning experience.

Copyright © 2012 George Mason University

Copyright © 2010 - This Site · All rights reserved. Conforms to W3C Standard XHTML & CSS

PRESENTED TO CS
DEPARTMENTAL COLLEAGUES

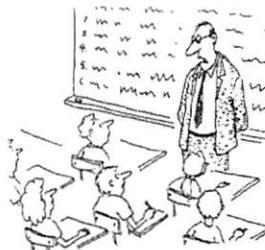
Teaching Recipes*

By Huzefa Rangwala



"I run, I work out, I don't smoke, I only drink in moderation. If I weren't so smug about it, I'd be damn near perfect."

Engage your students!



"I expect you all to be independent, innovative, critical thinkers who will do exactly as I say!"

Objectives ... (Using index cards)

What is your top objective for your class ?

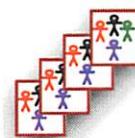
Objectives as an instructor ..

- How to achieve your objectives as an instructor?
 - Want students to listen to me.
 - Want students to participate.
 - Want students to read the book.
 - Want them to do the assignments before time.
 - **Want them to learn the material.**
 - Want them to have a good time doing it. (Put myself in their shoes)

In my classes ...

- Write-Pair-Share/Think
 - Write-Pair-Share
 - Think-Pair-Share
 - **(with your favorite Neighbor)

- Completion Exercises (Example: Handout)
 - Discuss in class
 - Maybe use as WPS.



JIGSAW ACTIVITY Example (Clustering Algorithms)

INFS 755 (Fall 2011)

CS 750 (Spring 2011)

INFS 755 (Fall 2008)

CS 795 (Spring 2009)

Reading List

- A. DBSCAN: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise; Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu (KDD96)
- B. CURE: An Efficient Clustering Algorithm for Large Databases; Sudipto Guha, Rajeev Rastogi, Kyuseok Shim (ACM 1998)
- C. Chameleon: Hierarchical Clustering Using Dynamic Modeling; George Karypis, Sam Han, Vipin Kumar (IEEE 1999)
- D. CLUTO: A Comparison of Document Clustering Techniques; Michael Steinbach, George Karypis, Vipin Kumar (SIAM 2000)
- E. BIRCH: An Efficient Data Clustering Method for Very Large Databases; Tan Zhang, Raghu Ramakrishnan, Miron Livny (SIGMOD 1996)

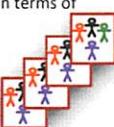
Expert Groups - [15 minutes]

- Discuss the algorithm
- Basic Idea, Flowchart, Steps ?
- Compare it to k-means
- Compare its methodology to other algorithms.
- Why is your algorithm good ?
 - Novel Features ?
- Evaluation
 - Strategy ?
 - Benchmarks ?
 - Did it compare to other methods ?
 - How did it do ?
- Complexity
 - Space
 - Run Time



Form Mixed Expert Groups based on your Number ID - 30 minutes

- Play the DEVIL's advocate. Your algorithm is the best :-)
 - Describe features of your clustering algorithm to other members of your group
 - Do you see similarities between methods ? Think about them
 - Do you see differences or how these algorithms evolved ?
 - Do you see why your algorithm is the best ?
 - What kind of evaluation methods did your algorithm use ?
 - Try to come up with a ranking of the algorithms in terms of complexity and accuracy.
 - Birch vs CURE ?
 - CHAMELEON vs CURE vs BIRCH ?
 - DBSCAN vs k-Means vs Chameleon ?



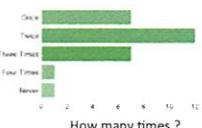
Wrap-Up Discussion [10 - 15 minutes]

- What did we learn ?
 - Differences
 - Similarities
 - Various benchmarks and evaluation methods.
 - Feel free to comment.
- Can we think of novel ways of designing a new clustering solution?
- Thank You!

What the students say ..



Use it Again ?



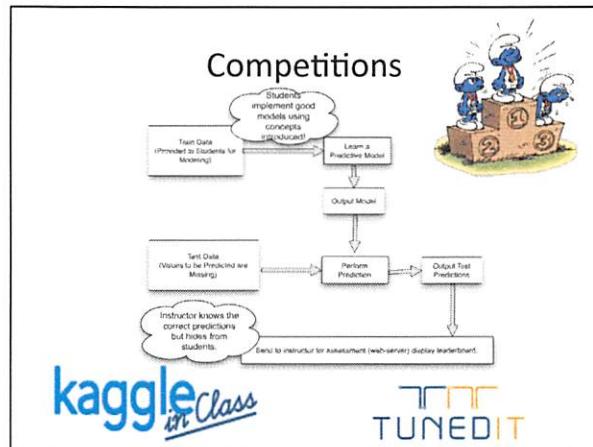
How many times ?

 “ This is the first time I participated in such an in-class activity. It worked for me as I learned through interaction with peers and also learned something in a very different way. It also helped me know my classmates.”



“ Still not everyone may read.”

In class competitions



Implementation

In Biological Data Mining Class (Spring 2010)

- Predict Solenoid Proteins.
- Used Email to receive predicted Solutions.

In Data Mining Class (Fall 2011)

- Predict activity of drug molecules. [Related to NSF funded Project 0905117]
- Involves feature engineering, choosing right models.
- Using kaggle-in-class.

At the end of competition.

- Discussion on different approaches taken by students.
- What worked & what did not.
- Grading is done on effort.
- Extra Credit for students high on leader board.

In Fall 2012 ...

- SpimBot (Mips based)
- CS 465.

A screenshot of a computer screen showing a simulation titled 'Map'. The interface includes a toolbar at the top and a main window displaying a grid-based map with several small robot icons. One robot is labeled 'robot1' and another is labeled 'robot2'.

Project-based Learning

Project-based learning.

- Data Mining Class (Easy)
 - Lots of real world datasets.
 - Project Resources online
 - Twitter Feed
 - Lots of competition
- Emphasize Real World Challenges and Datasets.



Rubric for Projects

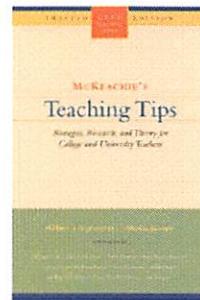
- A great way to assess.
 - Sets student expectations.
 - Sets your expectations.
 - Objective manner of “subjective” material.
- See handout for example.
 - You can create one that suits your class.
 - Several resources available online.
- Presentations, Poster Day, Software Demos
(Joao, Sanjeev)

How to Assess ?

CATs

- Muddiest Point.
- Question for an exam.
- Pop Quizzes
- Summary of Papers via email.
- Peer Review Papers.
 - Discuss reviews in class.

Closing Thoughts .. (Thank you)



Jigsaws and Competitions – Thanks to



5.II: Solving “Real-World Problems” in Class

Within my graduate classes, I have developed and implemented different competition-style assignments.

The spirit of competition and “learning by doing” approach pays rich dividends in helping student understanding and mastery of the material. Students are provided with tasks like analyzing a chemical molecule to estimate its toxicity or identifying the positive or negative nature of reviews/comments posted on Amazon.com. Completing these tasks involves creative thinking, programming and engineering to achieve the most optimal solution. I introduce new material every week that helps students advance and improve their working solution. During a 4-week competition, students work in teams to submit their solutions to a server that ranks each of them.

As a graduate student, my thesis was focused on development of a computational method for protein structure prediction. The community had setup a blind biennial competition called CASP¹ having participated in blind protein structure prediction competitions and data mining competitions like KDD Cup, I have implemented these form of competitions in my bioinformatics class. A student reported, *“The class project we did was a fantastic example of the difference your teaching makes: you want people to learn concepts by asking them to apply them to a context of their own choosing. This truly engages people.”*

In the Biological Data Mining Class (Fall 2009) as part of the final project I required students to train a predictive models to distinguish a specific class of proteins called “solenoids” using the available protein sequence information. As part of this competition, the truth values were hidden from the students and they had to make a prediction (guess) and submit their results to me (via email). I evaluated the results using the truth values and provides a ranking of the class students based on the predictive performance. The concepts introduced in class allow the students to build base line predictive models, but to improve performance, students had to research, think critically and come up with innovative solutions. I present this assignment as evidence in [Appendix 5.II \(Competitions\)](#).

Since, then I have automated the process of submission and evaluation of prediction results by using a server called Kaggle. In short, Kaggle is a San Francisco based start-up that allows corporations to setup predictive modeling competitions. This was inspired by the famous Netflix Prize Competition that provided the general public data from recommendations made by users on Netflix.com in the aspiration to build a better recommendation engine. A recommendation algorithm predicts given a user what movie the user may like the best. I wanted students in my data mining class to be involved in solving such real world problems. I wanted them to use data mining to solve problems, and compete on Kaggle.com. From Kaggle’s website “Kaggle is the world’s largest community of data scientists. They compete with each other to solve complex data science problems, and the top competitors are invited to work on the most interesting and sensitive business problems from some of the world’s biggest companies through Masters competitions”. Kaggle has been used by Facebook as a recruiting tool. See [Appendix 5.II \(Competitions\)](#) for description of this project.

In Fall 2011 and Spring 2013, when I taught data mining I setup prediction competitions on Kaggle-in-class. Both these competition descriptions, leader-board results and student efforts are shown in [Appendix 5.II \(Competitions\)](#). Specifically, in Fall 2011 students were to develop a data mining model to predict the toxicity of molecule. Overall 19 teams participated and there were a total of 192 submissions, averaging of 10 submissions per team. In Spring 2013 students were to develop a model for predicting

¹<http://predictioncenter.org>

ratings for text reviews extracted from Amazon.com. 38 teams comprising of students from both sections of the class offering participated and there were a total of 273 submissions. It is exciting to have an average of 7-10 submissions per team for a given assignment. I am extremely proud that I could engage students in such a competition where they participated with enthusiasm and rigor. The exciting part was to know that equipped with this knowledge students participated in other Kaggle competitions that forecast weather and estimated influenza outbreaks.

All graduate classes that I have taught also include a project component. Doing a project lets students take their learned theories and put it into use. I enforce that projects should be chosen by students, but should involve real world applications. Over the past, students have implemented range of projects ranging from learning user's music preferences, estimating crime probability in cities and estimating who could be potential donors for non-profits. I assist them by setting milestones that include project proposal, presentation and final report. I provide them with grading rubrics for project report and presentation as it helps students appreciate the best practices for presenting and reporting. I provide detailed feedback on the project proposal that is generally 1.5 months before semester completion. Finally, to foster collaborative learning and group sharing I have boot-camp like sessions towards the end of class where students can discuss their project ideas, progress, bottlenecks and potential solutions forward. Students also vote for the best project presentations. As evidence I have provided the project proposal template, project report template and rubrics in [Appendix 5.II \(Projects\)](#). I am also including an email titled "CS 659 Feedback" that discusses the Kaggle competition and other project based assignments.

Resources Setup After the first few offerings of my classes at Mason, I realized the need to have students get familiar with the resources that would be required for the different assignments that focused on "learning by doing". In my data mining classes students needed to install a package called "Weka". Not having it done before the first programming assignment was a cause of frustration for several students because they would be spending time doing the same. The same problem was identified in my undergraduate Computer Architecture class and in my current Spring 2014 Parallel Computing Class.

As such, I have made several changes to my class offerings. I always release a 0-credit Homework "0" in the first week of class. Within the data mining class it request students to install the required programming package and test it with a simple provided test case. In the undergraduate class (Architecture) I require students to install an assembly program simulator called MIPS and test it with a provided simple starter program. In the parallel computing class offered currently, we needed to access a cluster of machines and have assignments on different programming platforms. A detailed Homework 0 was designed to test the different types of computer resources and programming paradigms with sample test cases. These are also called "Hello World" programs in computing. I am including three different Homework 0 assignments for the different classes as evidence in [Appendix 5.II \(Resources\)](#).

To allow for student collaboration and discussion when not in class, I use discussion boards and forums. Previously, I implemented a message board system using the PHP. However, I find that using Piazza (a new classroom discussion board) is helpful and allows students/instructors to interact and communicate towards learning of materials in a quick and efficient manner. As evidence I am providing the statistics report generated by Piazza in [Appendix 5.II \(Resources\)](#).

Assignments

[COMPETITION STYLE ASSIGNMENT]

CS 795 : BIOLOGICAL DATA MINING

By *rangwala*

Created 08/24/2009 - 12:33

Assignment 1: Develop a sequence-based classifier for detecting proteins with a solenoid-based repeat.

Due date: 10/13/2009: 11:59:59 PM EST

Collaboration Allowed: Groups of 2 (max)

Background Paper: REPETITA: detection and discrimination of the periodicity of protein solenoid repeats by discrete Fourier transform^[1] by Marsella et. al

Description of the Assignment:

The purpose of this assignment is to build a binary classifier that will be able to discriminate between protein sequences having a solenoid repeat structure and protein sequences not having a repeat structure. You are free to choose any classification method (kNN, SVMs, decision trees, HMMs, CRFs, Bayes, Ensemble) that you think is suitable for the purpose, and can try one or several features (Amino Acid, AAIndex, PSI-BLAST, Secondary Structure) that can be derived using protein sequences and choose the appropriate representations.

The data-set including the training and test sequences for solenoids as well as non solenoids can be downloaded from [here](#).^[2]

Deliverables:

A detailed report (max 4 pages) explaining the method you followed, the features you used, and the experiments performed. Your results should include the precision, recall, the Q_2 accuracy, and ROC score for the different benchmarking done. Please zip your source code along-with instructions on how to run your code. Submit the entire package via email to *rangwala @ cs dot gmu dot edu* with the Subject "CS 795 Assignment 1".

Grading (100 points)

1. 60 points is for the design of your method, your idea, and explanation of the same in the report
 2. 40 points is dependent on the implementation and the results produced
 3. 10 points extra credit for achieving the highest ROC score in class.
-

Resources:

You can chose to implement your own code for this purpose or use any software that you may find interesting. Examples include

1. BioWeka
2. Weka
3. SVM Light
4. PyML
5. aasim (written by me)
6. ProSAT (written by me)

Source URL: <http://www.cs.gmu.edu/~hrangwal/node/108>

Links:

- [1] <http://bioinformatics.oxfordjournals.org/cgi/content/full/25/12/i289>
- [2] <http://protein.bio.unipd.it/repetita/help.html>



Facebook Recruiting III - Keyword Extraction [1 Page]

Finished

Friday, August 30, 2013

Jobs • 367 teams

Friday, December 20, 2013

Dashboard

- [Home](#)
- [Data](#)
- [Make a submission](#)

Information

- [Description](#)
- [Evaluation](#)
- [Rules](#)
- [Prizes](#)

Forum

- [Leaderboard](#)
- [Public](#)
- [Private](#)

My Team

My Submissions

Leaderboard

1. Owen
2. Ruslan Mavlyutov
3. E. G. Ortiz-García
4. Sergey
5. Roberto-UCIIIM
6. Alexander Kuznetsov
7. Data.Science
8. omnar
9. Silogram
10. Philip Wahrlich

Forum (81 topics)

Interviews
35 minutes ago

Interviews
6 days ago

Team anonymity and points
14 days ago

Share your approach?
18 days ago

Resume Collection

[Competition Details](#) » [Get the Data](#) » [Make a submission](#)

Identify keywords and tags from millions of text questions

Looking for a data science position at Facebook? After two successful prior Kaggle competitions, Facebook continues their mission to identify the best data scientists and software engineers that Kaggle has to offer. In this third installment, they seek candidates who have experience text mining large amounts of data.



This competition tests your text skills on a large dataset from the Stack Exchange sites. The task is to predict the tags (a.k.a. keywords, topics, summaries), given only the question text and its title. The dataset contains content from disparate stack exchange sites, containing a mix of both technical and non-technical questions.

Positions are available in Menlo Park, Seattle, New York City, and London; candidates must have, or be eligible to obtain, authorization to work in the US or UK.

Please note: you must compete as an individual in recruiting competitions. You may only use the data provided to make your predictions. Crawling stack exchange sites to look up answers is not permitted. Facebook will review the code of the top participants before deciding whether to offer an interview.

This competition counts towards rankings & achievements. If you wish to be considered for an interview at Facebook, check the box "Allow host to contact me" when you make your first entry.

Acknowledgements

We thank Stack Exchange (and its users) for generously releasing the source dataset through its [Creative Commons Data Dumps](#). All data is licensed under the cc-by-sa license.

COMPETITION EXAMPLE 1 [2 PAGES]

Drug Activity Prediction • limited

Finished

Friday, September 30, 2011

Knowledge • 19 teams

Monday, November 7, 2011

Dashboard

[Home](#)
[Data](#)

Information

[Description](#)
[Evaluation](#)
[Rules](#)
[Submission Instructions](#)
[Prizes](#)

Forum

[Leaderboard](#)
[Public](#)
[Private](#)

Leaderboard

1. Uma & Venkat
2. Paul and John
3. X.Guan
4. B. Shepard
5. INFSjoe
6. Dusan Puletic
7. Sahil & Manal
8. Neeraj.Zach
9. arezou
10. Data Miners

Forum (0 topics)

teams

players

entries

This competition is private-entry. You can view it but not participate.

Develop predictive models that can determine given a particular compound whether it is active (1) or not (0).

Drugs are typically small organic molecules that achieve their desired activity by binding to a target site on a receptor. The first step in the discovery of a new drug is usually to identify and isolate the receptor to which it should bind, followed by testing many small molecules for their ability to bind to the target site. This leaves researchers with the task of determining what separates the active (binding) compounds from the inactive (non-binding) ones. Such a determination can then be used in the design of new compounds that not only bind, but also have all the other properties required for a drug (solubility, oral absorption, lack of side effects, appropriate duration of action, toxicity, etc.).

The goal of this competition is to allow you to develop predictive models that can determine given a particular compound whether it is active (1) or not (0). As such, the goal would be develop the best binary classification model.

A molecule can be represented by 100000 binary features which represent their topological shapes and other characteristics important for binding.

Caveats:

+ Remember not all features will be good for predicting activity. Think of feature selection, engineering, reduction (anything that works)

+ The dataset has an imbalanced distribution i.e., within the training set there are only 78 actives (+1) and 722 inactives (0). No information is provided for the test set regarding the distribution.

+ Use your data mining knowledge till now, wisely to optimize your results.

Started: 9:49 pm, Friday 30 September 2011 UTC

Ended: 11:59 pm, Monday 7 November 2011 UTC(38 total days)

COMPETITION EXAMPLE 2 [5 PAGES]



Product Review Classification

Finished

Wednesday, February 27, 2013

Knowledge • 38 teams

Sunday, April 28, 2013

Dashboard

Home

Data

Information

Description

Evaluation

Rules

Prizes

Forum

Leaderboard

Public

Private

My Team

Administration

Show public leaderboard

Show private leaderboard

Benchmarks

View all submissions

Download submissions

Manage invitations

Manage admin users

View download log

Wizard

Leaderboard

1. r2d2

2. MeDals

3. cs659_dataminers

4. United659

5. 1204

6. Swami

7. ssydorko

8. gmuMiners

9. HarMony

10. sambravo88

Forum (0 topics)

This competition is private-entry. You have been invited to participate.

The goal of the competition is to predict the ratings (1.0-5.0) for a given text review of a product from Amazon.com's website.

Users have opinions and they express it. Be it books, movies or even professors, end user can express opinion about them. Sometimes the reviews are accompanied by a rating.

In this competition, given a text review we would like to automatically predict its 5-star rating. (5-highest and 1-lowest).

You are provided the following data files that you should use.

A training file **train.dat** is available that includes **5744** reviews in a sparse term-frequency format. The class label is given by the last column and is either 1.0, 2.0, 4.0 or 5.0.

A test file **test.dat** is provided to you that includes **5275** reviews in the same format as the training file. There are no labels in the last column. In fact, your goal is to predict the same.

An additional **unlabeled.txt** file is provided to you with **2131** reviews but there are no ratings. You may want to use these reviews to help your classification models. Using unlabeled examples within your classification models is called semi-supervised classification. This may or may not help your overall classification results.

Goal: For each row (example) in the test file predict one of the ratings (one per row). Ratings can be 1, 2, 4 or 5. There should be 5275 predicted ratings. Note, it should be integer ratings.

Caveats.

1. The words (features) that exist across the training, testing and unlabeled text files may not necessarily be the same i.e., you may have to figure out a way to represent the mappings.
2. You are free to use WEKA but this is not in WEKA-supported format. If you want to use Weka, you need to write a program to convert the data.

[teams](#)

[players](#)

[ENTRIES](#)

3. Other options include using packages like MALLET, ORANGE, SVM LIGHT or write your own algorithm.
4. Not all features may be important. Think of feature selection, reduction, engineering.
5. Use Data Mining wisely and correctly.

Started: 8:46 pm, Wednesday 27 February 2013 UTC

Ended: 12:00 am, Sunday 28 April 2013 UTC(59 total days)

[Edit this page »](#)



Product Review Classification limited

Finished

Wednesday, February 27, 2013

Knowledge • 38 teams

Sunday, April 28, 2013

Dashboard

Home

Data

Information

Description

Evaluation

Rules

Prizes

Forum

Leaderboard

Public

Private

My Team

Administration

Show public leaderboard

Show private leaderboard

Benchmarks

View all submissions

Download submissions

Manage invitations

Manage admin users

View download log

Wizard

Leaderboard

1. r2d2
2. MeDals
3. cs659_dataminers
4. United659
5. 1204
6. Swami
7. ssydorko
8. gmuMiners
9. HarMony
10. sambravo88

Forum (0 topics)

This competition is private-entry. You have been invited to participate.

Evaluation

Evaluation will be based on multi-class classification accuracy i.e., number of examples incorrectly predicted divided by total number of test examples.

Per day you are allowed 2 submissions.

[Edit this page »](#)



Product Review Classification

D limited

Finished

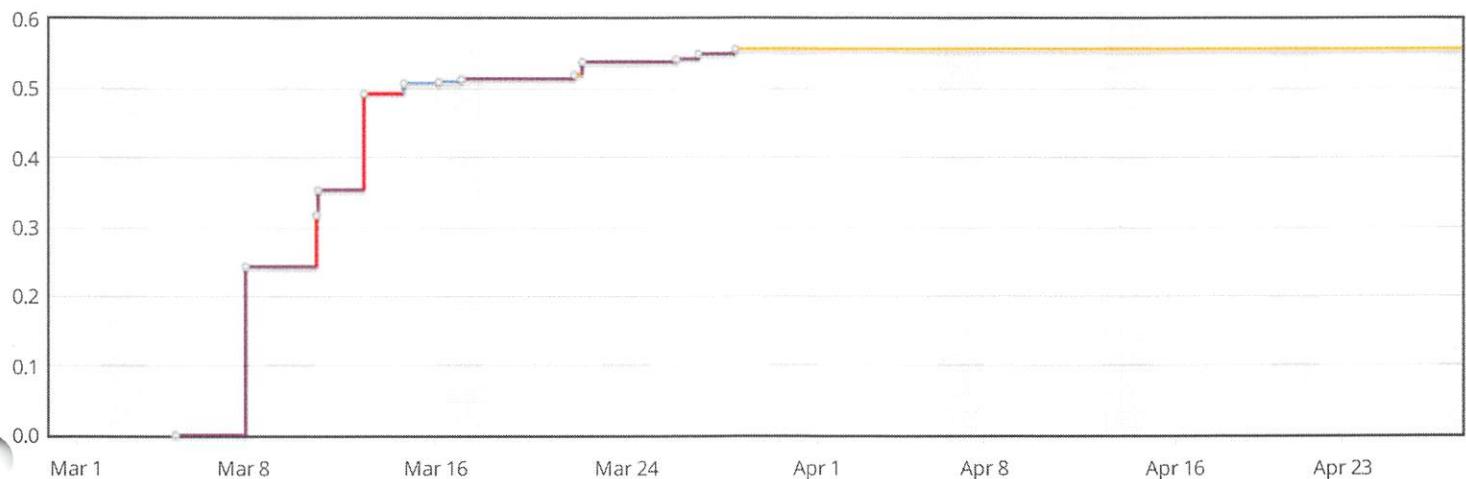
Wednesday, February 27, 2013

Knowledge • 38 teams

Sunday, April 28, 2013

Dashboard

Public Leaderboard - Product Review Classification



This leaderboard is calculated on approximately 75% of the test data.
The final results will be based on the other 25%, so the final standings may be different.

See someone using multiple accounts?
[Let us know.](#)

#	Δ1w	Team Name * in the money	Score	Entries	Last Submission UTC (Best – Last Submission)
1			0.55707	2	Thu, 28 Mar 2013 13:47:40
2			0.54973	22	Tue, 23 Apr 2013 03:15:52 (-27.1d)
3			0.54720	2	Thu, 28 Mar 2013 13:25:43
4			0.53328	18	Wed, 24 Apr 2013 23:05:02 (-26.9d)
5			0.53202	7	Tue, 09 Apr 2013 14:05:55
6			0.53176	12	Thu, 28 Mar 2013 20:25:54 (-2.8d)
7			0.53176	12	Thu, 28 Mar 2013 20:29:45 (-5.6h)
8			0.53151	6	Thu, 28 Mar 2013 15:37:44 (-36.6h)
9			0.52746	15	Sat, 27 Apr 2013 23:59:37 (-23.9d)
10			0.52544	12	Thu, 28 Mar 2013 16:06:20
11			0.52442	1	Thu, 28 Mar 2013 16:01:38
12			0.52341	2	Thu, 28 Mar 2013 16:04:19

13	0.52316	20	Thu, 28 Mar 2013 00:48:16
14	0.51911	4	Thu, 28 Mar 2013 06:48:47
15	0.51506	19	Thu, 28 Mar 2013 03:49:09 (-2d)
16	0.51455	3	Wed, 27 Mar 2013 03:42:15 (-9.3h)
17	0.51050	3	Fri, 29 Mar 2013 00:08:07
18	0.49835	19	Thu, 28 Mar 2013 20:12:53 (-7.8d)
19	0.49253	11	Wed, 27 Mar 2013 04:36:45
20	0.48924	2	Thu, 28 Mar 2013 08:08:14
21	0.48747	7	Thu, 28 Mar 2013 04:04:37 (-2.1d)
22	0.48722	6	Thu, 28 Mar 2013 00:18:07
23	0.48697	9	Tue, 26 Mar 2013 21:17:15 (-22.2h)
24	0.48646	5	Tue, 26 Mar 2013 06:24:12
25	0.48494	14	Thu, 28 Mar 2013 11:46:17 (-2.7d)
26	0.47836	3	Thu, 28 Mar 2013 08:04:38 (-4.6h)
27	0.46824	6	Wed, 27 Mar 2013 19:46:29 (-46.8h)
28	0.45457	3	Fri, 29 Mar 2013 04:13:25
29	0.45128	2	Thu, 28 Mar 2013 06:53:03 (-0.1h)
30	0.44773	3	Thu, 28 Mar 2013 20:12:09 (-44.3h)
31	0.44470	3	Thu, 28 Mar 2013 17:16:01 (-38.7h)
32	0.44065	5	Thu, 28 Mar 2013 02:41:03
33	0.42420	4	Fri, 29 Mar 2013 00:11:40
34	0.40268	1	Wed, 03 Apr 2013 13:32:26
35	0.32523	2	Thu, 28 Mar 2013 02:38:57
36	0.24677	1	Tue, 12 Mar 2013 19:25:47
37	0.20628	2	Wed, 27 Mar 2013 02:04:49 (-0h)
38	0.00000	5	Sat, 30 Mar 2013 02:36:02 (-30.7h)

[Download raw data](#)

CS 659 Data Mining (Spring 2013)
Project Proposal Guidelines
Due Date: 03/21/2013 by email & in-class

I urge you to develop your project ideas by talking with your advisor & me. If you do not know what to do I can suggest some projects related to temporal data mining, motif mining, approximate nearest neighbors, association rule mining with applications in recommender systems, medical informatics, bioinformatics and social media analysis.

Please follow these guidelines for a good project proposal.

Project Guidelines

1. Provide a brief introduction with a background (if needed) of the project you intend to do. Motivate the problem, and convince me why its important to address this problem.
2. Provide clearly the objectives of your project. What will be the outcome of the project. Please explain to me how the objectives if addressed will solve the problem.
3. Methods: Please explain what will your methodology be. What kind of tools will you use/develop ? Where will your data be downloaded, derived or even generated from?
4. Literature Search: Also perform a search to see if there exists any other competitive methods that were solving the same problem. How will you compare your work to the same ?
5. Milestone/Timeline: Please provide some milestones, as to when you would expect to finish the various tasks of the project.

Document Guidelines

1. Maximum Page Limit is 2 pages (single spaced, at most 10 point, Times New Roman Font).
2. Please state the title of the project identifying the team members.
3. Maximum Team Size Two, but identify what every individual will be doing in the proposal.

TITLE FOR YOUR PROJECT

Huzefa Rangwala

November 8, 2008

Abstract

Please write a 250 word abstract for your project. An abstract should summarize the project, and highlight the key points or results achieved in your project.

1 Introduction

Use this space to introduce the topic to the reader. Provide a background, if needed. Also give an outline of what the coming chapters will be like.

2 Problem Statement

Succinctly describe your problem. This should be brief but clear.

2.1 Notations

Introduce any notations that you use in your report here.

3 Literature Review

Summarize the literature search you have performed with respect to your project. I can figure very easily how much reading you have done. This is critical to success of the project.

4 Methods and Techniques

Describe in detail the methods and techniques used for your particular project. Any method specific details should go in here

5 Discussion and Results

5.1 Datasets

Discuss which datasets you used, where you got them from. Statistics about the same (Chapter 2 in your book is all about data.)

5.2 Evaluation Metrics

What kind of metrics did you used to evaluate yourself ?

5.3 Experimental Results

Describe the experiments performed, results with tables or charts along with explanations. Just putting a nice figure without explaining in text will get you no points.

Analyze the successes and failures for your project. Describe all the alternate strategies you tried.

6 Conclusion

Provide conclusions for your project. A summary of highs and the lows can go here. What you learned from this exercise ?

6.1 Directions for Future Work

What would you do next with this project ? How would you extend it ?

References

Provide all the references here.

Project Report Grading Rubric

By *rangwala*

Created 11/21/2011 - 20:25

Component	Excellent	Good	Fair
Content (30 points)	The deliverable did a fantastic job on all the components - problem definition, background, motivation, literature search, experimental evaluation and conclusions.	The deliverable lacked in one or two of the components defined for criterion "excellent".	The deliverable did a poor job on more than 2 of the components defined in criterion "excellent".
Contribution (30 points)	The deliverable explained core ideas of the project in an excellent manner & demonstrated novel concepts and ideas in relation to previous work.	The deliverable demonstrated adequate understanding of the problem and research approach, and discussed the novelty aspects.	The project was not novel and uninteresting from a research perspective. Either the problem was ill-defined or the methods required no thought.
Subject Knowledge (30 points)	The deliverable demonstrated knowledge of the course content by integrating major and minor concepts into the project. The deliverable also demonstrated evidence of extensive research effort and a depth of thinking about the topic.	The deliverable demonstrated knowledge of the course content by integrating major concepts into the project. The deliverable also demonstrated evidence of limited research effort and/or initial of thinking about the topic.	The deliverable did not demonstrate knowledge of the course content, evidence of the research effort or depth of thinking about the topic
Composition (10 points)	Excellent presentation & grammar. Excellent technical vocabulary. Fit in 8 pages and well-formatted.	Good presentation & grammar. Good technical vocabulary. Fit in 8 pages and well-formatted.	Poor presentation & grammar. Poor technical vocabulary. Did not fit within 8 pages and poorly formatted.

Presentation Rubric

By *rangwala*

Created 11/21/2011 - 20:31

Component	Excellent	Good	Fair
Content & Creativity (35 points)	The presentation contained an abundance of material which clearly related to the main goals of the project. External research was used to justify arguments or solutions. The presentation of the material was original and presented in a creative way that held audience attention	The presentation contained sufficient materials but did not justify arguments well or did not present external research. Slides did not present information in a creative manner	The presentation lacked clarity and did not include sufficient details. There was a lack of external research. Slides were a bulleted list of text.
Coherence & Organization (35 points)	Motivation, problem definition and project execution was well explained. Transitions were done seamlessly and there were no logical errors	Motivation, problem definition and project execution were adequately explained. Transitions were done well and there were a few logical errors	Motivation, problem definition and execution statements were missing. Transitions were poorly performed.
Presentation Skills, Participation (30 points)	All team members presented. Did not go above the time limit. Enthusiasm and confidence was exuded. Good eye contact, good speech volume and captured the audience.	All team members presented. Did not go above the time limit. Enthusiasm and confidence was mild with ok eye contact, ok speech volume and captured 80% of the audience.	All members did not present. Went above the time limit. Did not show enthusiasm for presenting.

Source URL: <http://www.cs.gmu.edu/~hrangwal/node/366>



Huzefa Rangwala <rhuzeafa@gmail.com>

CS 659 Feedback

[REDACTED] Wed, May 15, 2013 at 10:56 AM

Hello Professor,

I am writing down to you to let you know my experience about the Data Mining course that I took under you in the last semester.

I felt the course content was perfectly structured and covered. It was conveyed also in a very easy to understand manner. The way you have been giving real life examples where Data Mining can be and is being applied kind of gave a kick start to my thinking process. After seeing all the project presentations by the end of the semester I have seen data mining being applied to various interesting domains which I had never thought about.

Also to mention about your assignments. They were absolutely the perfect food for thought problems and best way to understand the various concepts in practice. The 3rd assignment idea was just awesome and the competitive vibe on the leader board made us push our limits and try and better ourselves. As a result we did well in the assignment and guess what that earned me a job.

I mentioned about the kaggle competition on my resume and the course project and the company that hired me as an intern picked me up because of these data mining projects that I did during the semester as a part of your course work. I will now be working on the data mining project they are about to start with. And I think its all thanks to you.

Hope to take some other course too under you if it fits into my program requirements.

Hope you have a great summer ahead. Good Day.



CS465 (HW0)

HW0: Assembly Programming and Installing MARS MIPS Simulator [0%]

The purpose of this HW assignment is to prepare you for the upcoming assignments by learning about assembly programs and setting up the MARS MIPS simulator on your preferred computing environment.

MARS is a MIPS Assembler and Runtime Simulator developed by Volmar from Missouri State University. Please download the assembler from [here](#).^[1] MARS is written in Java and is provided as a "jar" package. Be sure you have java installed on your preferred computing environment.

`java -jar Mars_xxx.jar` should load up MARS [where "xxx" is the version number.]

Part 1. Write your first MIPS Program.

- First open MARS simulator using command line option as shown above.
- Create a new assembly program
- Type in Editor the following code

```
#Author: Huzefa Rangwala
#Date: 08.29.2012

.data
silly_str: .asciiz "Hello World, Fall 2012 is going to be a fun time."

.text

main:
li $v0, 4
la $a0, silly_str
syscall

# To exit off main
li $v0, 10
syscall
```

- Save the code as "hello_world.asm"
- Assemble the Code
- Run the Code
- Does it show the desired output ? Yes ? Congratulations!

Part 2. Tutorial By MARS Authors using fibonacci.asm

- Download Tutorial: <http://courses.missouristate.edu/KenVollmar/MARS/CCSC-CP%20material/MARS...> [2]
- Download Fibonacci.asm from [here](#): [3]
- Follow Steps 1 to 11 of Tutorial in Part 1.

There is no deliverable and the assignment is not graded. However, I encourage you to try this and let [REDACTED] me (hrangwal@gmu.edu [5]) or the class (Piazza) know if you have issues completing this assignment.

Source URL: <http://www.cs.gmu.edu/~hrangwal/node/412>

Links:

- [1] <http://courses.missouristate.edu/KenVollmar/MARS/index.htm>
- [2] <http://courses.missouristate.edu/KenVollmar/MARS/CCSC-CP material/MARS Tutorial.doc>
- [3] <http://courses.missouristate.edu/KenVollmar/MARS/CCSC-CP material/Fibonacci.asm>
- [5] <mailto:hrangwal@gmu.edu>

HW 0 (CS 635)

HW 0: Hello World in MPI, pThreads, OpenMP and CUDA using VSE Resources.

The purpose of this HW assignment is to prepare you for the upcoming assignments by learning about different parallel programming platform and resources that are available to us.

There is 0% credit for this assignment and no deliverables are due. It is expected that after completion of this assignment you will be ready for HW1-5. If you have trouble completing this assignment, please use the discussion forum on Piazza.

There are 4 parts to this assignment. They will involve setting up the parallel programming environment, logging into appropriate resources, compiling and executing the "Hello World" program for MPI, pThreads, OpenMP and CUDA (GPUS).

Description of resources, access information and starter hello world programs have been made available at the [Resources Page](#)

1. Run Hello World Program using MPI on Medusa Cluster.
2. Run Hello World Program sing OpenMP on Medusa and tesla (tesla1.vsnet.gmu.edu).
3. Run Hello World Program using pThreads on Medusa and tesla.
4. Run Hello World Program using CUDA on tesla

Source URL: <http://www.cs.gmu.edu/~hrangwal/node/489>

Links:

[1] <http://www.cs.gmu.edu/~hrangwal/node/486>

Parallel Computing Resources

For the programming assignments we will use the cluster [medusa.vsnet.gmu.edu](#) [1]. This will allow us to do MPI/OpenMP and pThread assignments.

For GPU/CUDA assignments we will use [tesla1.vsnet.gmu.edu](#) and [tesla2.gmu.edu](#)

You should be able to access these machines using your GMU netid. Remember passwords may need to be reset. Here are some general instructions on [how to ssh into these machines](#).

[2]

You also need to do passwordless login setup. This can be done by using [ssh-keygen as follows](#). [3] Essentially first login into medusa and then follow the steps to login into other compute nodes and tesla1/tesla2.

The following are some general instructions on how to get your MPI/OpenMP/pthreads/Cuda codes compiled and running:

MPI on Medusa

1. Log into the head node [medusa.vsnet.gmu.edu](#)
2. Run "module load rocks-openmpi"
3. Using your favorite text editor create a file named "hosts" in your \$HOME directory.
The file should contain the following list of hostnames:

```
compute-0-0 slots=4
compute-0-1 slots=4
compute-0-2 slots=4
compute-0-3 slots=4
compute-0-4 slots=4
compute-0-5 slots=4
compute-0-6 slots=4
compute-0-7 slots=4
compute-0-8 slots=4
compute-0-9 slots=4
compute-0-10 slots=4
compute-0-11 slots=4
compute-0-12 slots=4
compute-0-13 slots=4
compute-0-14 slots=4
compute-0-15 slots=4
compute-0-16 slots=4
compute-0-17 slots=4
```

compute-0-18 slots=4
compute-0-19 slots=4

4. To test MPI create a file called "hello_mpi.c" and copy

```
// Author: Huzefa Rangwala
// Date: 11.26.2013

#include < mpi.h > // remember to include the mpi header
#include < stdio.h >

int main(int argc, char *argv[])
{
    int num_procs, myid, name_len;
    char proc_name[MPI_MAX_PROCESSOR_NAME];

    // Initialize MPI
    MPI_Init(&argc, &argv);

    // Obtain the number of processes
    MPI_Comm_size(MPI_COMM_WORLD, &num_procs);

    // Obtain the process id
    MPI_Comm_rank(MPI_COMM_WORLD, &myid);

    // Obtain name of machine process is executing on
    MPI_Get_processor_name(proc_name, &name_len);

    printf("Hello World from processor %d out of %d, executing on %s\n", myid,
    num_procs, proc_name);

    // Last call to MPI (REQUIRED)
    MPI_Finalize();

    return 0;
}
```

5. Compile the program: mpicc -o hello_mpi.ex hello_mpi.c
6. Run your compiled program: mpirun --hostfile ~/hosts -np 4 ./hello_mpi.ex
- here "4" is the number of processors

- [OpenMP on Medusa/Tesla](#)
- [pThreads on medusa](#)
- [CUDA on Tesla](#)

Source URL: <http://www.cs.gmu.edu/~hrangwal/node/486>

Links:

- [1] <http://labs.vse.gmu.edu/index.php/Systems/MedusaCluster>
- [2] <http://labs.vse.gmu.edu/index.php/FAQ/SSH>
- [3] <http://www.thegeekstuff.com/2008/11/3-steps-to-perform-ssh-login-without-password-using-ssh-keygen-ssh-copy-id/>
- [4] <http://computer-graphics.se/hello-world-for-cuda.html>



Share this

Piazza Report for:

CS 659 (002)

In total, students asked:

30 Questions

Either students, instructors, or both responded to
100% of Questions

Counting all posts, responses, edits, followups, and comments, there were:
335 Contributions

The average response time was:

125 Minutes

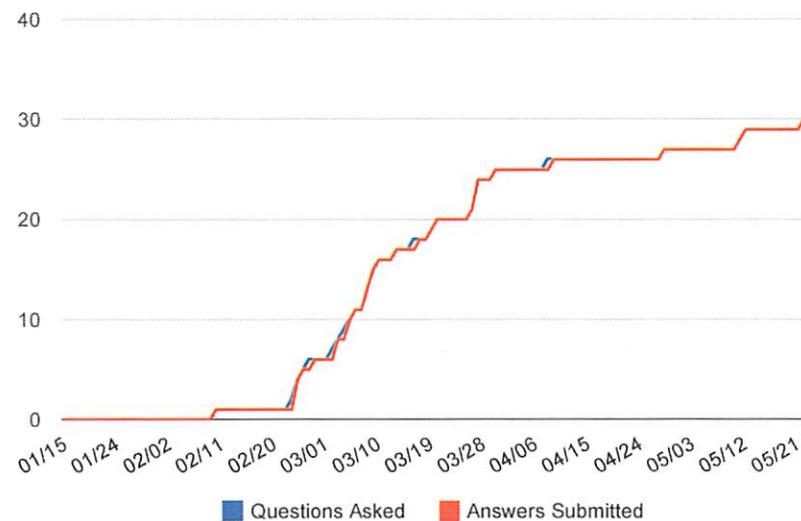
Top Contributors



42 students were enrolled...

...and **76%** of them made at least one contribution (**32** in total).

Questions & Answers



100% of questions received
instructors' responses
(**30** in total).

27% of questions received
students' responses
(**8** in total).

And **38%** of those were
endorsed by an instructor
(**3** in total)!

Top Content

33
views

Presentation Sign-Up and

Dear All, Two class periods will be devoted for project ...

- 0 Answerers
- 20 Followup Discussions
- 3 Followup Replies

34
views

HW3 submission format

To make sure my format was correct, I did a ...

- 1 Answerers
- 3 Followup Discussions
- 7 Followup Replies

36
views

Mid-Term 2 coming up on

This will be a closed * exam. We will try ...

- 0 Answerers
- 4 Followup Discussions
- 5 Followup Replies

Share this

Piazza Report for:

CS 465

In total, students asked:

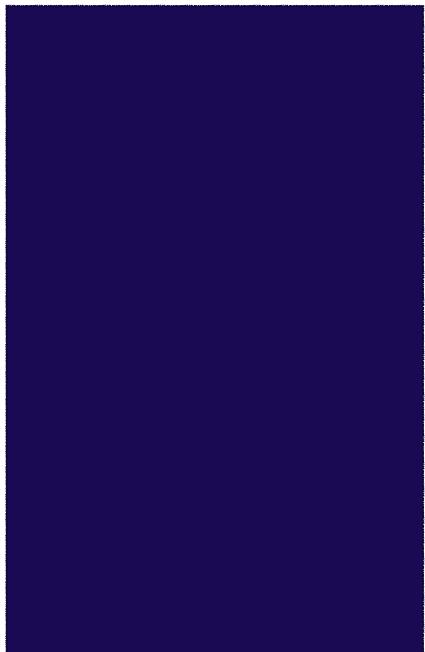
40 Questions

Either students, instructors, or both responded to
100% of Questions

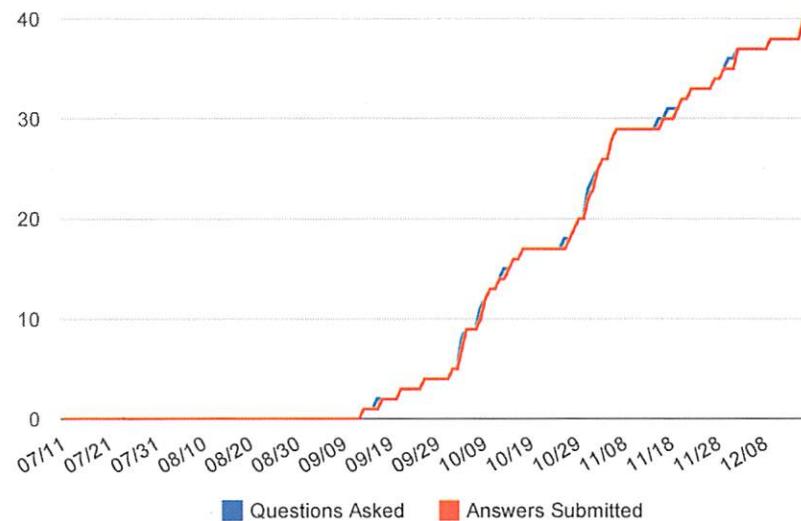
Counting all posts, responses, edits, followups, and comments, there were:
269 Contributions

The average response time was:

77 Minutes



Questions & Answers



43 students were enrolled...

...and **56%** of them made at least one contribution (**24** in total).

75% of questions received instructors' responses (**30** in total).

40% of questions received students' responses (**16** in total).

And **38%** of those were endorsed by an instructor (**6** in total)!

Top Content

Reminder: Final Exam

Final Exam will be on Tuesday Dec 17, 2013 from ...

- 0 Answerers
- 4 Followup Discussions
- 14 Followup Replies

HW4 is online

[http://www.cs.gmu.edu/~hrangwal/node/...](http://www.cs.gmu.edu/~hrangwal/node/)
Due Date: 12.03.2013 Teams: Max 2

- 0 Answerers
- 3 Followup Discussions
- 6 Followup Replies

HW 1 is online

HW 1 is uploaded.
[http://www.cs.gmu.edu/~hrangwal/node/...](http://www.cs.gmu.edu/~hrangwal/node/)

- 0 Answerers
- 4 Followup Discussions
- 4 Followup Replies

5.III: Inspiring New Generation of Researchers

The same desire to better society that guided my mother's work has influenced my interest in helping students of all levels and from all backgrounds. Being a professor at an entrepreneurial and student-focused institution like Mason aligns well with my career goals and vision. I aspire to engage, inspire and help students be involved in the multiple facets of scholarly pursuits. Involving students from the Mason or local high schools within my research program allows them to go beyond the classroom, find their passion, and take ownership towards solving pertinent real world problems. Over the years, I have mentored students from different backgrounds, different disciplines and of different ages. Not only has this been incredibly rewarding, but it has also helped me think about explaining complex research concepts to a broader audience, improving my ability to collaborate and lead diverse interdisciplinary teams.

Introducing Research Ideas in Undergraduate Classes

In Fall 2010, I taught my first undergraduate class. I was excited to teach CS 465: Computer Systems Architecture. It is a class about understanding computer system performance by unveiling the several layers underneath an application-level software stack.

After mulling about the course content and assignments, I was committed to introducing students an exercise that would allow them to experience the role of a researcher. Students would analyze the performance of different cache systems under different performance parameters. They would compare different options and comment on how the cache design choices would affect computer system performance under different conditions. Looking at the reports received as part of this assignment, I think students understand the importance of computer system assessment and how to execute it in a fair manner. This was the first time such an assignment was used in a core senior class. I present the assignment as well as a student sample in the [Appendix 5.III \(Class\)](#).

"Learning by doing" was an important component for this class. As such, I wanted students to build computer programs that were written in assembly language. For the reviewer, assembly language is the lowest level of computer program and programming in this language is very challenging than for example programming in languages like Java or Python (called high-level languages). I have developed several different assignments that involve programming in the assembly language and I provide samples of two of these assignments used in Fall 2013. One relates to computing the validity of a phone number and the second one relates to a puzzle that involves computing some chess moves. Both, these assignments involve deep critical thinking, programming effort and understanding of computer science concepts that students may have learned in their previous semesters but never implemented. I am also including my ABET report that documents my changes made to the CS 465 class in Fall 2010 in [Appendix 5.III \(Class\)](#).

Incorporating Research Within the Curriculum

Mason is an entrepreneurial university. My colleagues at Mason are inspiring. OSCAR, the office of Student as Scholars is one example of an initiative that I have not seen at other universities. The goal of OSCAR office to involve every undergraduate at Mason with a research/scholarly experience resonates with me. I want to be able to influence the students I engage in creating something new, something useful, something that they can present and be proud of.

Being passionate about undergraduate student scholarship and entrepreneurship I am excited to note that with the support of the Mason Student as Scholars program and in collaboration with Professors

Wang and Allbeck in the Computer Science Department, we are leading a project wherein inquiry-based learning approaches and undergraduate research are synthesized within the Computer Science curriculum. Our analysis of the curriculum identified that students within the program lacked or had minimal understanding of what was meant by scholarly research and how to create scholarly products. This was concerning given that Computing as a field is always rapidly innovating. Specifically, we are designing and implementing a three-course sequence that first helps students in understanding the “discovery” process of research by following successful research projects completed by departmental faculty or industry partners. The second phase involves “inquiry” where students learn how to ask relevant and critical questions. This involves training the next generation of scientists and industry professionals by exposing them to contemporary grand computing solutions like Google’s search algorithm. The final phase allows students to implement a complete research product (be it a software or experiment) and present the results at a student seminar. The successful implementation of these curriculum improvements will result in local industry collaborations where students will serve as clients for sponsored projects. Such initiatives will assist in training the next “doers”, “innovators” and “entrepreneurs” who will drive successful economies for the future. As evidence, I am presenting the analysis of the undergraduate curriculum in Computer Science that was done by our collaborative team in [Appendix 5.III \(Curriculum\)](#)

Mentoring Graduate, Undergraduate and High-School Students

Beyond the formal classroom, I am committed to fostering the synergy between teaching and research by providing an environment for all students to develop intellectually and professionally. In the past five years, I have supervised and mentored ten graduate, nine undergraduate and five high school students from various disciplines. I have already advised three Ph.D. students towards successful completion of their doctoral dissertations.

My mentoring approach differs, depending on the level of the student. My approach to working with graduate students centers around collaboration and learning from one another. I consider all students on par with me, and build a relationship that encourages listening viewpoints, identifying research challenges and evaluating strengths/weakness of proposed approaches. I provide enough latitude to my students so they pursue their own interests but also act as a guide for helping them achieve their milestones. My research interests lie at the intersection of Dr. Rangwala’s primary research interests lie at the intersection of data analytics (information science) and biology. An underlying descriptive feature transcending my scholarly pursuits is the ability to understand the challenges and problems in an interdisciplinary domain and contribute with innovative state-of-the-art computing solutions. Present-day biological science involves integration and analysis of data obtained from different biotechnological machines (like genome sequencers and automated robotic laboratories) to validate key hypotheses towards an understanding of living systems, disease and development of cures. I have developed computational systems that analyze these large, heterogeneous and complex biological datasets to advance critical scientific questions. I have published 35 peer-reviewed conference proceedings, 17 journal articles, an edited book, 4 book chapters, and 4 refereed workshop articles (at least 70% of them co-authored with my advising students). All of my research is published in premier journals and conferences in the fields of data mining and bioinformatics with highly selective acceptance rates. I have been successful in receiving funding to support my interdisciplinary research efforts and mentoring students, securing a total of \$4.44 million dollars of external funding. I also received the National Science Foundation’s (NSF) CAREER award. This award is NSF’s most coveted and prestigious award that supports a small number of junior faculty who exemplify the role

of teacher-scholars through outstanding research, excellent education and the integration of education and research for creating broader impact in the society and lives of students.

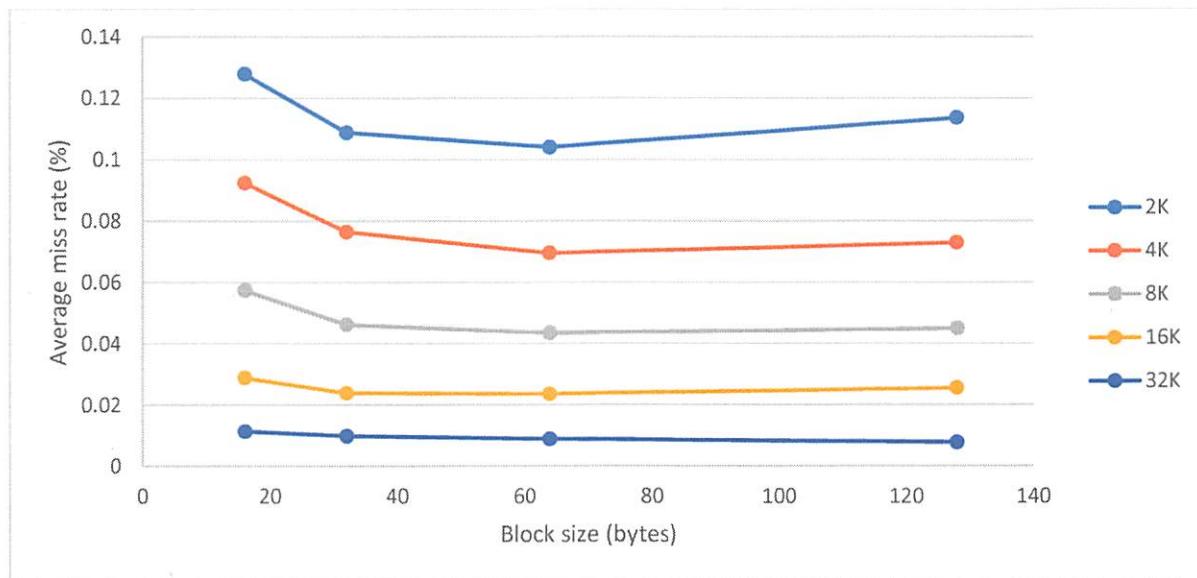
On the other hand, I have learned that undergraduate and high school students are more successful when there are well-defined projects that students can complete during a semester period. These research projects are designed to expose students to interdisciplinary research, challenge their intellect and creativity, and provide them with the satisfaction resulting from creating something novel and useful. I have co-authored two high-impact journal articles with my undergraduate mentees. I have also received NSF sponsored funding for undergraduates as supplements to my main research awards.

I also nurture the aspirations of young scientist researchers in Northern Virginia region. I have engaged diverse and minority students by presenting my research at NSF Louis Stokes Alliances for Minority Participation (LSAMP) event at Mason. I seek to inspire students, provide them with resources for access to research at the university and enable opportunities that may not be available otherwise. I invite students from area high schools to participate in a summer research internship. As part of the internship, students design and implement a research project that they are encouraged to present at local and national science fair competitions. In Summer 2011, I mentored a high school sophomore, Eric Tao from the local Thomas Jefferson High School of Science and Technology. Eric participated in the National Siemens Science Competition and was a regional finalist. In Fall 2013, Eric joined MIT for his undergraduate education and shared "Through working with Huzefa, I learned how to think about and approach a problem, how to tackle barriers in my work, and how to create a well-written and comprehensive research paper. I could not have gained this invaluable knowledge elsewhere." As evidence I present some of these testimonials about my mentorship record in Section 6 of this teaching portfolio. In the [Appendix 5.III \(Mentoring\)](#), I am including a news article from Fairfax TJ High regarding Eric Tao's performance. I am including a list of students that I have mentored at George Mason University and a list of grants that have supported these students over the years.

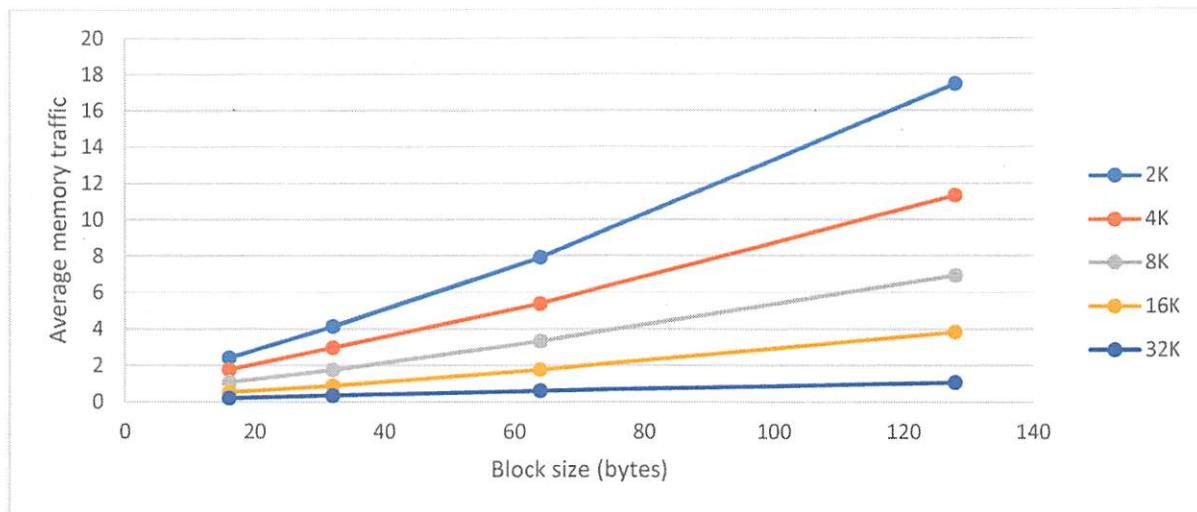
I cherish and embrace all my responsibilities as a faculty member. The driving force behind my passion for research, student mentoring, teaching and service to the community is my mother who is with me forever. I am always seeking improved ideas and ways to empower my students, collaborators and peers with the ability to better society.

Part 2: Cache Simulation Using DineroIV [40 points]

Task 1: Evaluation of cache sizes and block sizes.



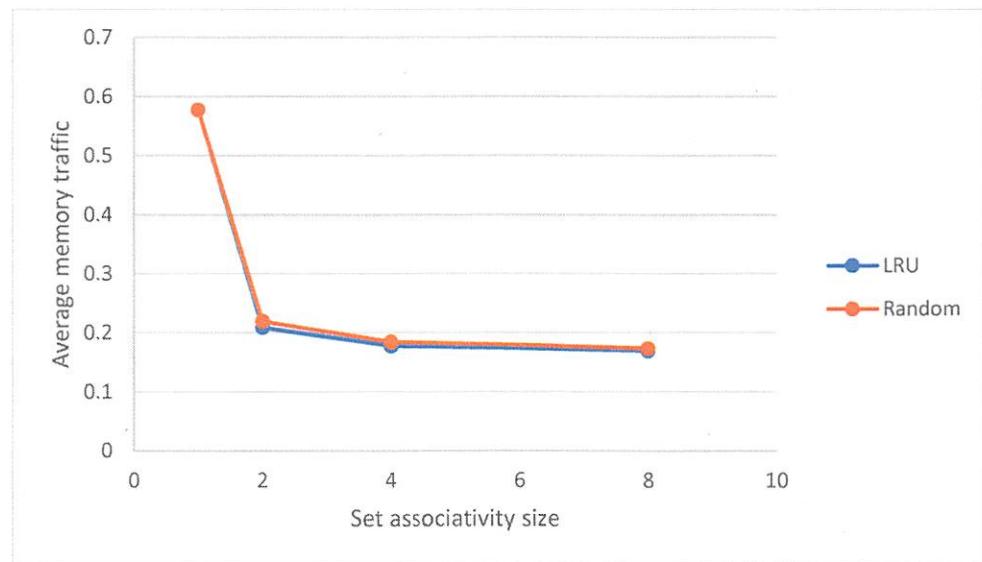
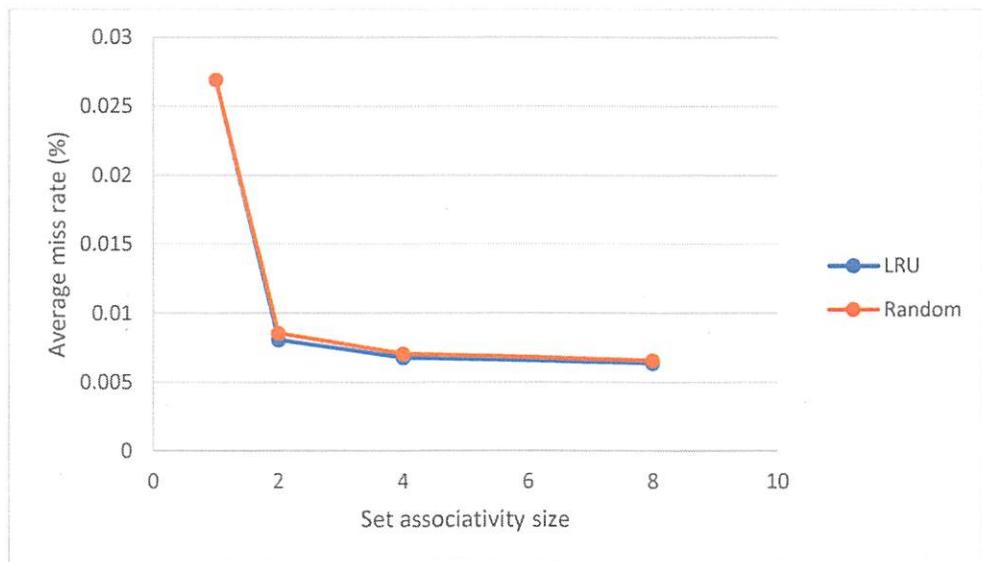
In this graph for average miss rate, the average miss rate goes down as we increase the unified cache size. The configuration of unified cache size of 32K has the smallest average miss rate overall. Additionally, larger block sizes (but not too large) only helps reduce the average miss rate if the cache size is small. For larger unified cache size like 32K, the block size has no significant effect on the average miss rate.



In this graph for average memory traffic, in most of the cache sizes configuration, as we increase the block size, the average memory traffic increases. Furthermore, the smaller the cache size, the faster the average memory traffic increases as we increase the block size. For 32K cache size, there is not much change in average memory traffic as we increase the block size.

Overall, if we assume that the cost of the different cache size is the same or we don't care about the cost, I think 32K cache size and 16 bytes block size is the best configuration since it has the lowest average miss rate and average memory traffic. We could have a higher block size configuration, but it would give the same result as 16 bytes. On the other hand, if cost of the cache is important to us, then we would choose the smallest cache size, which is 2K. The best block size for 2K would be 64 bytes if we want the lowest average miss rate or 32 bytes if we don't want to have a lot of average memory traffic.

Task 2: Evaluation of associativity and replacement policy for the cache blocks.



The trend is similar for both graphs. For both LRU and Random, as we increase the set associativity size, the average miss rate and the average memory traffic decreases. However, this only improves a lot if we go from 1 to 2 set associativity size. For higher set associativity sizes, there is not much change in average miss rate and average memory traffic. Overall, LRU and Random have trends that are similar to each other, though LRU may have lowest average miss rate and memory traffic.

Course: CS 465
Instructor: H. Rangwala
Term: Fall 10

[ABET REPORT]

1. Please indicate the percentage of students that met each of your course outcomes:

Course Outcome	Percentage			
	Strongly Met	Met	Weakly Met	Not Met
1. Be able to explain the organization of the classical von Neumann machine and its major functional components	30%	40%		30%
2. Be able to compare performance of simple system configurations and understand the performance implications of architectural choices	80%	20%		
3. Be able to show how instructions are represented at both the machine level and in the context of a symbolic assembler; be able to understand small MIPS programs and write MIPS assembly program segments	90%	10%		
4. Be able to use different formats to represent numerical data and convert numerical data from one format to another			50%	50%
5. Be able to explain how an instruction is executed and the concept of datapaths and control	80%	20%		
6. Be able to explain basic instruction level parallelism using pipelining and the major hazards that may occur	100%			
7. Be able to explain the effect of memory latency on running time; be able to describe the use of memory hierarchy to reduce the effective memory latency, in particular, the role of cache and virtual memory; be able to understand the principles of memory management	70%	30%		
8. Be able to explain the basic I/O implementation and data transfers			20%	80%

2. What basis of judgment did you use for calculating these percentages? Did you use your final exam results, a homework assignment, a term project, an exam, a specific question on an exam, etc... ? Be as specific as possible.

1. This was brought forward every two weeks, but the explanation is not elaborate in the textbook. In the final a disappointing 30% did not answer anything for the question "What is von Nuemann architecture?". In the future, additional reading can be devoted to this section.
2. HW1, Mid-Term and several class room exercises were detailed towards understanding the performance of computer systems.
3. HW1,2,3 had a programming component in MIPS assembly language. 80% of students were able to score close to maximum on the programming part. Classroom exercises, quizzes and Mid-Term focused on questions related to programming in MIPS.

4. We had a question on the Mid-Term to convert from one format to another. The result was that 50% understood how to do this efficiently. I assumed students know this, and did not give them any assignment related to this.

5. HW3 and 4, along with questions on Final and Quizzes show that students understand concepts of data path and control path.

6. HW3 and 4, along with questions on Final and Quizzes show that students understand concepts of pipelining.

7. HW4, final and a simulation assignment using Dinero (a cache simulator) allowed the concept of memory hierarchies to be illustrated.

8. Not enough time to go into details of I/O data transfer. The book does a poor job on this topic. As such, very few students answered questions about DMA transfer in the Finals.

3. Were your students adequately prepared for your course?

The students were adequately prepared for the class.

4. Please state any general observations you would like to make about your class this term.

Students enjoyed active learning exercises that I used:

- a) write-pair-share activity
- b) jigsaw activities
- c) assessment techniques (muddiest point)
- d) class quizzes

They cherished the opportunity to program in MIPS. Initially, they found that it was a step back from programming in x86 which they have been exposed to in CS367 (a pre-req). However, the class was focused on explaining the concepts of simple MIPS processor and programming in MIPS allowed one to do that.

5. What did you change in the course over previous offerings? What would you change if you were to teach this class again?

From previous course offerings:

- + Added programming assignments (learning by doing)
- + Added cache simulation assignments.

Future:

- + Skip the section on I/O and focus on parallel processors.
- + Add some more readings about von Neuman architecture.

CURRICULUM ANALYSIS

OSCAR GRANT

Appendix A - Curriculum Map of CS Foundation Courses

LEGEND					
X indicates the course currently contains this material			Indicates that this is a course we will be focusing on		
/ indicates the course partially contains this material			Indicates a course that we will be considering for future revisions toward scholarly outcomes		
O indicates the course could feasibly be altered to contain this material			Indicates a course already contributes significantly to students as scholars		

	Discovery of Scholarship										Scholarly Inquiry								Creation of Scholarship				
	Students will understand how knowledge is generated and disseminated through scholarship, and the importance of scholarship to society.					Students will understand how they can engage in the process of scholarship, by being introduced to:					Students will articulate a scholarly question; engage in the key elements of the scholarly process; and situate the concepts, practices, or results of scholarship within a broader context.								Students will create an original scholarly or creative project.		Students will communicate knowledge from an original scholarly or creative project.		
Computer Science	1. Distinguish between personal beliefs and evidence.	2. Articulate how scholarship influences society.	3. Understand epistemological or historical perspectives of a specific body of knowledge.	4. Evaluate credibility of source information.	5. Understand research methods used in a discipline.	6. Understand how knowledge is transmitted within a discipline, across disciplines, and to the public.	How Mason faculty are engaged in scholarly work	Undergraduate scholarly work at Mason	Opportunities offered by the Students as Scholars initiative.	1. Articulate and refine the question.	2. Follow ethical principles.	3a. Choose an appropriate discovery process for scholarly inquiry.	3b. Gather evidence appropriate to the question.	3c. Apply appropriate scholarly conventions during scholarly inquiry.	3d. Apply appropriate scholarly conventions when reporting or performing.	4a. Assess the validity of key assumptions and evidence.	4b. Situate the scholarly inquiry within a broader context.	1. Justify that the project intends to be engaging and novel to a particular audience.	2. Take responsibility for executing the project.	3. Present their understandings from a scholarly perspective for a specified audience.	4. Demonstrate awareness of broader implications.		
101 Preview of CS	X	X	X	/		X	X	X	X														
105 Computer Ethics and Society	X	X			X		/					X											
112 Intro to Computer Programming																							
211 Object-Oriented Programming																							
262 Intro to Low-Level Programming																							
306 Synthesis of Ethics and Law for Computing Professionals	X	X	X	X		X					X	X											
310 Data Structures																							
321 Software Requirements and Design Modeling		X	X		/	X					/								X	/			
330 Formal Methods and Models			X																				
367 Computer Systems and Programming		X																					
390 Research Methods											O	O	O	O	O	O	O	O	O	O			
465 Computer Systems Architecture	X	X	X		X						O	O	O	O	O	O	O	O	O				
483 Analysis of Algorithms		X	X				O				O	O	O	O	O	O	O	O					
490 Design Exhibition																							

Concentration courses in the ACS Game Design Major																							
325 Intro to Game Design	X	O	O		X	O	O	/	O	X	X	X	X	X	X	X	I	X	X	X	O	O	
425 Game Programming I	O	O	X	/	O	X	O	O	O	/								X	X	X	X	O	
426 Game Programming II				O	O	O	O	O	O	X	X	X	X	X	X	X		X	X	X	X	O	

Appendix B - CS 390 (New Course)

Appendix C - CS 490 (Course Revision)

Creation of Scholarship (all 4 required)									
Scholarly Inquiry (at least 3 required)									
Students will create an original scholarly or creative project	Students will communicate knowledge or creative project	Students will articulate a scholarly question; engage in the key elements of the scholarly process; and situate the concepts, practices, or results of scholarship within a broader context	Course specific Outcomes	General Education Outcomes					
CS 490 Design Exhibition		Students will articulate a scholarly question; engage in the key elements of the scholarly process; and situate the concepts, practices, or results of scholarship within a broader context		Scholarly Inquiry (at least 3 required)					
1. Justify that the project intends to be engaging and novel to a particular audience	2. Take responsibility for executing the project.	3. Present their understandings from a scholarly perspective for a specified audience.	4. Demonstrate awareness of implications for a broader scholarly inquiry.	1. Articulate and refine the question.	2. Follow ethical principles.	3a Choose an appropriate scholarly inquiry procedure for the question.	3b Gather evidence appropriate to the conventions during scholarly inquiry.	3c Apply appropriate scholarly conventions when reporting or performing.	3d Apply appropriate scholarly conventions when reporting or performing.
Explore a problem space						x	x	x	x
Write and revise a project proposal including SOW	x					x	x	x	x
Review proposed projects			x			x	x	x	x
Execute project	x					x	x	x	x
Write big describing process in progress		x	x	x		x		x	x
Present research paper	x	x	x	x		x	x	x	x
Industrial audience	x	x	x	x		x	x	x	x

tj TODAY



NOVEMBER 15, 2011

VOLUME 27, ISSUE 3

MLK Memorial unveiled

Civil Rights legacy preserved in Washington, D.C. attraction

by Sara Asad and Rabia Idrees

News Editor and Editor in Chief

Maryland resident Cliff Bright held his son up on his shoulders. The two of them made their way through crowds of visitors to stand next to the towering statue of Martin Luther King Jr. Bright smiled widely as his wife Yvanna snapped a photo.

One week after the official opening of the memorial on Independence Avenue in the District, Sunday crowds lingered in the brilliant sunlight, telling stories, reading inscriptions and feeling the history ingrained in the presence of the imposing stone structure.

This memorial represents the legacy of this country and how Martin Luther

King's movement was not just an African American movement, but one for unity in the United States," Bright said. "My son didn't get to see the struggles then, but he gets to see the legacy they left through the life he lives now."

The idea for the memorial was first considered early in 1984 when George Scaley, a member of the predominantly black Alpha Phi Alpha fraternity, proposed the idea of dedicating a national memorial to King. In 1998 former President Bill Clinton signed congressional legislation for the establishment of a memorial in Washington, D.C.

The 30-foot statue of King was carved by sculptor Lei Yixin in Hunan, China and later transported to its current position on the National Mall. The estimated total cost of the memorial is projected near \$120 million of which more than \$114 million has already been raised through private and corporate contributions.

"The memorial is so gorgeous," St. Louis resident Elaine Wyatt said. "It gives us a chance to talk about the Civil Rights Movement with our children."

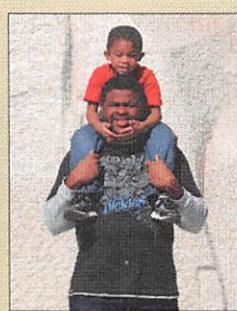
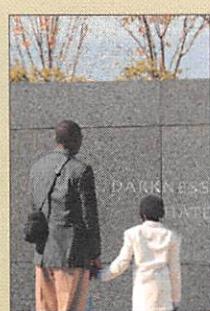
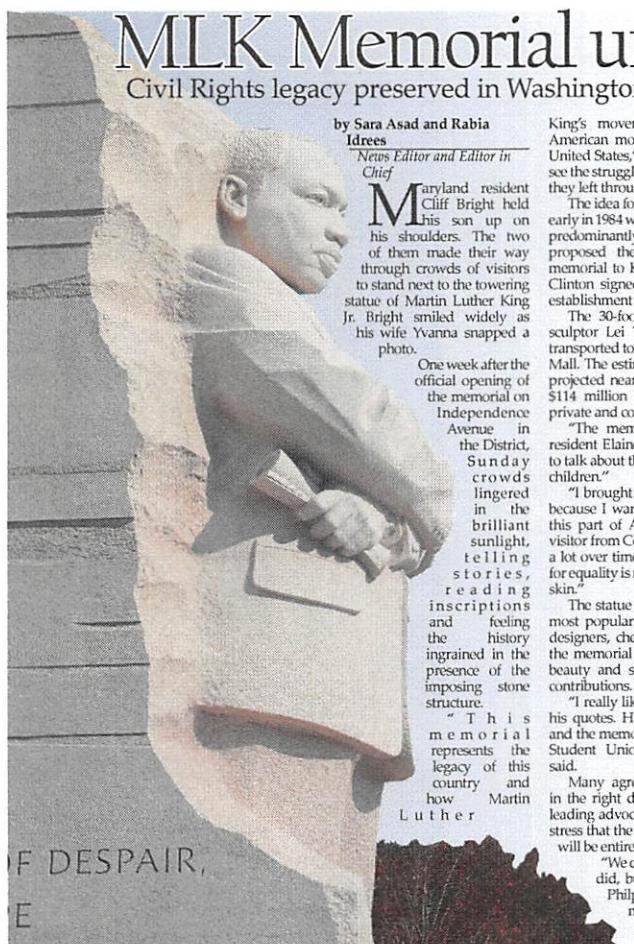
"I brought my son down to see the memorial because I wanted him to be able to understand this part of American history," Scott Rubin, a visitor from Connecticut, said. "We have changed a lot over time. The struggle King went through for equality is more important than the color of his skin."

The statue of King is surrounded by 14 of his most popular quotes and 182 cherry trees. For designers, cherry trees and quotes surrounding the memorial are important in appreciating the beauty and significance of the King's life and contributions.

"I really like how the memorial was based on his quotes. He embodied the power of speech, and the memorial really brought that out," Black Student Union president Michael Wattendorf said.

Many agree that the memorial was a step in the right direction and well-deserved by the leading advocate for civil rights. Some, however, stress that the nation still has ways to go before it will be entirely free of prejudice as King wanted.

"We didn't go through life as hard as King did, but we still have struggles," Cedric Philpot, a visitor from Miami, said. "The memorial shows that we have come along, but we aren't there yet."



Left to right: A father and his son reflect on one of the many quotes on the MLK memorial; two tourists take pictures of the memorial; Cliff Bright holds his son in front of the MLK statue.

Four regional finalists named by Siemens

by Mallika Patkar
News Editor

Four students were named regional finalists in the Siemens Competition in Math, Science and Technology on Oct. 21. Seniors Daniel Jang, William Qian and Marvin Qian as well as junior Eric Tao are among 96 regional finalists nationwide.

In addition to Jefferson's four regional finalists, seniors Pritha Bhattacharya, Rachel Chuang, Nishant Garg, Darvin Li, Karishma Popli, Rithvik Prasannappa, Sameer Singh, Yohan Sumathipala and Shriram Sundaram were named semi-finalists.

The regional finalists will go on to compete to qualify for national finals in December.

Jang was team leader for a group project that involved improving the power output of fuel cells. The project began last summer at Stony Brook University in New York.

"My project dealt with the most important part of the hydrogen fuel cell, the proton exchange membrane," Jang said. "My team was able to figure out how to modify this layer with nanoparticles to enhance the overall power output of the fuel cell."

Jang's team will compete at Carnegie Mellon University on Nov. 18 and 19.

Tao and Marvin Qian worked on their project. Their project deals with computer algorithms in the field of bioinformatics.

"Although biologists have uncovered a myriad of protein sequences, very little is known about their structures," Tao said. "We developed a way to classify proteins into solenoid or nonsolenoid categories, and locate the boundaries for the solenoid repeats."

The sequences Tao and Qian identified could be beneficial in the field of medicine.

"Proteins have a vast influence on the molecular machinery of life. Knowing the three-dimensional structure of proteins is crucial to advances in biology, as this information provides insight into how proteins operate," Asst. Professor Huzefa Rangwala from George Mason University said.

Rangwala helped the students understand structural bioinformatics.

"Siemens competition is an excellent opportunity for students to be exposed to large scale research and industrial projects within STEM fields," Rangwala said.

Tao and Qian along with William Qian participated in regional finals Nov. 5 and 6 at Georgia Tech. They did not move on.

William Qian's project dealt with computational exploration of the Earth's inner core.

"Recently there was a Science article, which dealt with the same topic and the results were, excitingly, the same as our results," Qian said.

Students will be competing for College Board administered scholarships ranging from \$1,000 to \$100,000 in individual and team categories.

"I am proud of our students for participating and sharing their research," Principal Evan Glazer said. "I like to learn about their research to see how we could cultivate the approaches they used to excel in their projects among other students at our school."



SPORTS
Winter sports previews P6



SPREAD
Jefferson's global community P8-9



HEALTH
Male body image P12



FEATURES
Observatory joint project P14

Student Mentoring

Ph.D. Advisees

1. **Mr. Nuttachat Wisittipanit**, *Ph.D. in BINF*, (Co-Advised with Prof. Gillevet), Spring 2009-Spring 2012. Graduation Date: 03/2012. *Thesis: Machine Learning Approach for Profiling Human Microbiome.*
2. **Mr. Zeehasham Rasheed**, *Ph.D. in CS*, Spring 2010-Spring 2013. Graduation Date: 04/2013. *Thesis: Data Mining Framework for Metagenome Analysis.*
3. **Mr. Samuel Blasiak**, *Ph.D. in CS*, Summer 2010-Present. Graduation Date: 08/2013. *Thesis: Latent Variable Models of Sequences for Classification and Discovery.*
4. **Mr. Anveshi Charuvaka**, *Ph.D. in CS*, Fall 2009-Present.
5. **Ms. Tanwistha Saha**, *Ph.D. in CS*, (Co-Advised with Prof. Domeniconi), Summer 2009-Present.
6. **Mr. Xin Guan**, *Ph.D. in CS*, Spring 2013-Present.

M.S. Thesis Advisees

1. **Mr. Azad Naik**, *M.S. in CS*, Fall 2012-Spring 2013. *Thesis: Using Multi-Task Learning for Large-Scale Document Classification.*
2. **Mr. Syed Mahmood**, *M.S. in CS*, Fall 2009-Fall 2012. *Thesis: GPU-based Eulerian Assembly of Genomes.*
3. **Mr. Ammar Naqvi**, *M.S. in BINF*, (Co-Advised with Prof. Gillevet), Spring 2009-Fall 2009. *Thesis: Understanding and analyzing the human microbiome: taxonomy identification and potential interactions.*
4. **Mr. Salman Jamali**, *M.S. in CS*, Spring 2009-Summer 2009. *Thesis:: Comment Mining, Popularity Prediction and Social Network Analysis.*

M.S. Project Advisees

1. **Ms. Na-Young Shin**, *M.S. in BINF*, Fall 2010.
2. **Mr. Ronak Shah**, *M.S. in BINF*, Spring 2010.
3. **Mr. Shanmuga Chiripiralla**, *M.S. in BINF*, Spring 2010.
4. **Ms. Rithika Ganni**, *M.S. in BINF*, Spring 2010.
5. **Ms. Meher Boorgula**, *M.S. in BINF*, Spring 2010.

Undergraduate Mentees

1. **Ms. Anisha Kolla**, *B.S. in CS*, Spring 2013. (Sponsor: URSP, OSCAR, GMU).
2. **Mr. Hozaifah Zafar**, *B.S. in CS*, Spring 2013-Present. (Sponsor: URSP, OSCAR, GMU).
3. **Ms. Emily Eastlake**, *B.S. in BENG*, Summer 2012. (Sponsor: SURE Program, Bioengineering, GMU).
4. **Mr. Liban Hassan**, *B.S. in ECE*, Summer 2011. (Sponsor: NSF LSAMP Program, GMU).

5. Ms. Amanda Zouzolou, *B.S. in BENG*, Summer 2011. (Sponsor: SURE Program, Bioengineering, GMU).
6. Mr. Roderick Tolbert, *B.S. in CS*, Summer 2011. (Sponsor: NSF REU to Rangwala).
7. Mr. Charles Sweet, *B.S. in CS*, Summer 2010-Fall 2012. (Sponsor: UAP Program, GMU).
8. Mr. Minh Bui, *B.S. in CS*, Summer 2010. (Sponsor: UAP Program, GMU).

High School Mentees

1. Mr. Eric Tao, *Thomas Jefferson High School, Alexandria, VA*, Summer 2010-Fall 2012.
2. Mr. Abbas Idris, *Governor's School, Prince William, VA*, Fall 2011-Spring 2012.
3. Mr. Ashwin Sekar, *Poolesville Maryland High School, MD*, Summer 2013.

Student Committee Membership (Besides Advisees)

Student	Program	Advisor	Grad. Date
Mr. Cheol Young Park	Ph.D. in Information Technology	Kathryn Laskey	-
Mr. Gautum Singh	Ph.D. in Computer Science	Jana Kosecka	-
Mr. Daniel Saxton	Ph.D. in Statistics	Anand Vidyashankar	-
Mr. Joshua Church	Ph.D. in Information Technology	Ami Motro	-
Mr. Keith Sullivan	Ph.D. in Computer Science	Sean Luke	-
Ms. Nada Baset	Ph.D. in Computer Science	Harry Wechsler	Fall 2012.
Ms. Andmorgan Fisher	Ph.D. in Environmental Sciences	Patrick Gillevet	Fall 2012.
Mr. Sean Smith	Ph.D. in Bioinformatics	Patrick Gillevet	Fall 2011.
Mr. Pu Wang	Ph.D. in Computer Science	Carlotta Domeniconi	Spring 2011
Mr. Andrew Heekin	Ph.D. in Bioinformatics	Patrick Gillevet	Summer 2011.
Mr. Keenan Amundsen	Ph.D. in Bioinformatics	Don Seto	Fall 2009.

Research Support

External Funding

NSF CAREER: *Annotating the Microbiome Using Machine Learning Methods.* IIS-1252318. PI: **Huzafe Rangwala**, \$550,000, 03/01/2013-02/28/2018.

NSF III: *Collaborative Research: Computational Methods to Advance Chemical Genetics by Bridging Chemical and Biological Spaces.* IIS-0905117. PI: **Huzafe Rangwala**, \$331,537, 09/01/2009-08/31/2013. Additional REU Supplement: \$8000

NSF CPS: *CPS: Synergy: A Novel Biomechatronic Interface Based on Wearable Dynamic Imaging Sensors.* To be Awarded. PI: Siddhartha Sikdar, Co-PI: **Huzafe Rangwala**, Jana Kosecka and Houman Homayoun. \$995,055, 09/01/2013-08/31/2017.

NIH: *Systems Biology Approach to Identifying Biomarkers for Alcoholic Liver Disease.* NIAAA-1RC2AA010405-01. PI: Ali Keshavarzian & Patrick Gillevet. Co-PI: **Huzafe Rangwala**, Robin Couch. \$1,974,161 (\$972,966 GMU portion), 10/01/09-09/30/12. This is a highly competitive ARRA GO (grand opportunities) award.

DARPA: *CyNomix: Detecting Zero-Day Malware by Generating Behavioral Cyber Genome Sequences.* DARPA-CyberGenome. PI: Angelo Stavrou, Co-PI: **Huzafe Rangwala**, \$1,527,225, 09/01/10-08/31/14.

USDA: *Bovine microRNA Transcriptome Analyses: Discovery, Tissue Specific Expression Profile and Target Gene Prediction..* PI: **Huzafe Rangwala**, \$45,000. 11/01/2011-04/30/2012.

nVidia: *GPU Computing for Assembly of Genomes.* PI: **Huzafe Rangwala**, \$3,000 worth equipment donation. 02/01/2010.

NSF: *Career Mentoring Forum and Student Travel Support for 2012 IEEE International Conference on Data Engineering.* PI: **Huzafe Rangwala**. Co-PIs: Carlotta Domeniconi and Alex Brodsky, \$24,000. 03/01/2012-03/01/2013.

nVidia: *CUDA University Research Center.* PI: Saleet Jaffri, Co-PI: **Huzafe Rangwala**, 08/30/2011-08/30/2012.

NSF II-NEW: *II-NEW: An Experimental Infrastructure for Cross-Domain Research in Wireless Computing, Cybersecurity and Data Mining.* PI: Robert Simon, Co-PIs: Daniel Barbara, Brian Mark and Angelos Stavrou. Senior Personnel: **Huzafe Rangwala**, \$547,307, 09/01/2012-08/31/2015.

Internal Funding

OSCAR: Students as Scholars Track B: Curriculum Development and Revision. PI: Pearl Wang, Co-PIs: Jan Allbeck and **Huzafe Rangwala**, \$21,500, 08/01/2012-07/30/2014.

University Seed Grant: Parallel Assembly of Genomes and Metagenomes. PI: **Huzafe Rangwala**, \$20,000, 01/01/2010-12/31/2010.

Volgenau Seed Grant: Disease State Inference from Microbiome in the Gut. PI: **Huzafe Rangwala**, \$30,233, 09/01/2009-12/30/2010.

Volgenau Seed Grant: Adverse Event Detection Mining in Health Data. PI: Jessica Lin, Co-PIs: **Huzafe Rangwala**, Carlotta Domeniconi and Daniel Barbara, \$3,500. 09/01/2009-12/30/2010.

Section 6: Evidence of Teaching Effectiveness and Impact

Huzeфа Rangwala, Ph.D.

This section contains the three required parts: (A) Summary of student ratings from the past five years, (B) testimonials ans (C) closing statement.

For Part (B) I have included the following information:

- Letters requested from my peers in the Computer Science Department, [REDACTED] and [REDACTED] [2 pages]
- A letter from my graduated doctoral student, [REDACTED]
- A letter received from N [REDACTED] was my high school mentee. [1 page]
- Four excerpts from letters received from past students who worked on a thesis or project with me. [2 pages]
- Testimonials of teaching that I collected anonymously via a Google online document. Students had the option of signing their names.

Section 6: Evidence of Teaching Effectiveness & Impact

Table 1: AVERAGE RATINGS OF TEACHING EFFECTIVENESS AT GMU

<i>Dimensions of Teaching</i>	<i>Average Ratings Across Courses</i>	<i>Engineering School Average (Spring 2013)</i>	<i>Dimensions of Teaching</i>	<i>Average Ratings Across Courses</i>	<i>Engineering School Average (Spring 2013)</i>
Teaching Rating	4.66	4.26	Course Ratings	4.38	4.03
Respect	4.78	4.69	Accessibility	4.81	4.55
Syllabus	4.69	4.48	Assignments	4.66	4.31
Organization	4.65	4.35	Readings	4.45	3.99
Explanation	4.58	4.26	Returned	4.69	4.37
Feedback	4.49	4.13	Material	4.70	4.54
Involvement	4.80	4.27	Stimulating	4.62	4.18

Table 1 shows the average ratings as reported in the course evaluations across all courses taught by me from Fall 2008 to Spring 2013. The different dimensions of teaching correspond to the questions that are surveyed in the teaching evaluations form. It also compares my teaching ratings to the average ratings of all courses offered by the Volgenau School of Engineering in Spring 2013.

Some key highlights that can be observed that in all the metrics of teaching effectiveness my ratings are higher than the average VSE ratings. In terms of “Involvement” and creating a “stimulating” environment I am excited to report a high rating.

Table 2 (next page) shows ratings for my instruction and course for the different classes taught from Fall 2008 to Fall 2013. In parenthesis “()” I show the departmental average as a point of comparison. Some interesting observations include my improved performance in the second offering of CS 795: Biological Sequence Analysis from Spring 2009 to Spring 2010. The instructor rating increased from 4.38 to 4.91. I incorporated more active learning approaches in Spring 2010 and also teamed up biology-majoring students with computer science students when performing group assignments. This improved learning for inter-disciplinary material. Consistently my teaching ratings outperform the department, engineering school and university ratings (data not shown for university).

Table 2: RATINGS FOR INDIVIDUAL CLASSES.

Class	Semester	Enrol.	Instr. Eval	Course Eval
<i>INFS 755: Data Mining*</i>	F2008	25	4.35 (4.31)	4.10 (4.04)
<i>CS 795: Biological Sequence Analysis**</i>	Sp2009	25	4.38 (4.36)	4.06 (4.11)
<i>CS 795: Biological Data Mining**</i>	F2009	13	4.73 (4.25)	4.45 (4.03)
<i>CS 795: Biological Sequence Analysis</i>	Sp2010	18	4.91 (4.31)	4.55 (4.06)
<i>CS 465: Computer Architecture*</i>	F2010	46	4.71 (4.28)	4.26 (4.09)
<i>CS 750: Data Mining*</i>	Sp2011	18	4.72 (4.28)	4.67 (4.09)
<i>INFS 755: Data Mining</i>	F2011	34	4.79 (4.27)	4.52 (4.08)
<i>CS 465: Computer Architecture</i>	F2012	48	4.55 (4.29)	4.10 (4.08)
<i>CS 659: Data Mining</i>	Sp2013	32	4.81 (4.29)	4.69 (4.13)
<i>CS 465: Computer Architecture</i>	F2013	43	4.69 (4.30)	4.56 (4.12)
<small>* Denotes – First Time Offering of Class and ** Denotes – Developing a new class that was not in the catalog.</small>				

My ratings for undergraduate class *CS 465: Computer Systems Architecture* highlighted in red. CS 465 is a core, required senior level computer science class. Different faculty members in the Computer Science Department have taught it over the past years. In Table 3 below I show the average ratings for CS 465 when I am the instructor and when I am not the instructor. I also report the standard deviation for statistical significance in parenthesis “()”. The percentage improvement in class rating when I am the instructor for CS 465 is 25%. In Section 5 I documented the changes that were implemented in CS 465.

Table 3: COMPARING RATINGS For CS 465.

Assessment Metric	My CS 465 Ratings	Other Instructor's CS 465 Ratings
Instructor Rating	4.65 (0.087)	3.64 (0.32)
Class Rating	4.30 (0.233)	3.43 (0.13)

[SCANNED STATEMENTS FROM COURSE
ASSESSMENT FORMS]

Programming assignments helped me learn the material

The book questions were somewhat ambiguous and hard to answer completely. Test questions and quizzes were far more helpful in gauging an understanding.

The professor constantly evaluated our understanding the material. The professor clearly made this course worthwhile. Unfortunately I may have to take it again if I can't get the grade I desire, depending on how heavily homework counts in .

Smart guy that obviously cares about his students.

1) Teaching style is great, class participation / involvement helped me focus.

① Instructor was great, put in lots of effort to keep us engaged in perhaps the dryest least exciting class in the curriculum.

② Prof. host is in scisack,
Horrible: I had a much easier time concentrating in our original class room in the art building.

Rengwala is one of the best CS professors I have had. He is energetic and demands participation. If he is not tenured I recommend him to remain at UI.



Department of Computer Science
The Volgenau School of Engineering

4400 University Drive, MS 4A4, Fairfax, Virginia 22030-4444, USA
+1 703-993-1660 (o), +1 703-993-1657 (fx), pammann@gmu.edu, <http://www.cs.gmu.edu/~pammann/>

February 4, 2014

Dr. Kimberly Eby
Director, Center for Teaching and Faculty Excellence
George Mason University
4400 University Drive MS 4D6
Fairfax, VA 22030

Dear Dr. Eby:

It is my pleasure to write this letter of recommendation in support of Professor Huzefa Rangwala, a finalist for a 2013-2014 George Mason University Teaching Excellence Award.

Last fall, Prof. Rangwala went through the T&P process at the CS department and VSE school levels. As part of that process I visited his classroom and reported my findings for inclusion in Prof. Rangwala's T&P dossier. In a rare move, and one that I strongly endorsed, the CS department found that Prof. Rangwala had achieved genuine excellence not only in research, but also in teaching. That finding was seconded by the VSE P&T committee. (At the time of this writing, the recommendations of the Dean and the Provost are yet to be determined.) This letter captures my report and also includes additional direct knowledge I have of Prof. Rangwala's teaching activities.

Last fall, when the CS T&P committee solicited volunteers to conduct peer-evaluations for our tenure candidates, I jumped at the opportunity to visit Prof. Rangwala's classroom. Frankly, I had an ulterior motive in volunteering to evaluate Prof. Rangwala: I figured I would learn something about teaching that I could use in my own classes.

And I did. I quote from my report under the section "Communication and Rapport":

This aspect of the class was especially well done. Prof. Rangwala varied his presentation between slides, handouts with questions, and working exercises on the whiteboard, with no extended time spent in any of these phases. Student's attention was well maintained, and answers were solicited from many students, and not just a few volunteers. Incorrect student answers were constructively and tactfully handled, and correct answers received positive feedback.

I've used in-class exercises informally for years. But I hadn't ever tried splitting them up as finely as Prof. Rangwala did. So, in my next class, I gave it a go. Success! The students liked it, and so did I, mainly because it made the potentially deadly "PowerPoint" sections much shorter in duration.

Prof. Rangwala and I followed up my classroom visit with teaching-related discussions. Prof. Rangwala lent me one of the books he has found useful in crafting his instruction (*McKeachie's Teaching Tips*,

14th Edition). Usually, senior faculty give junior faculty teaching tips; it was a pleasure to have the normal order inverted.

Prof. Rangwala obviously loves to teach. Equally importantly, he excels at it. In our research-driven environment, there are strong cultural signals that emphasize research, often at the expense of teaching. These signals are especially robust for junior faculty in the engineering school because research excellence is the *de facto* basis of T&P decisions.

Nonetheless, Prof. Rangwala has clearly resolved both to conduct world-class research and also to excel in the classroom. Prof. Rangwala shows by example that it is possible to achieve genuine excellence in both teaching and research. And he has elaborated that example while building his research reputation as a junior faculty member. While I would be reluctant to advise other junior faculty to follow his lead, I certainly am impressed at how he has achieved his goals.

In summary, I highly recommend Prof. Rangwala for a George Mason University Teaching Excellence Award. Please do not hesitate to contact me if you have further questions.





Department of Computer Science
The Volgenau School of Engineering
4400 University Drive, MS 4A5, Fairfax VA 22030, USA
Phone: 703-993-1530 Fax: 703-993-1710
<http://www.cs.gmu.edu>

Kimberly Eby
Associate Provost, Faculty Development
Director, Center for Teaching and Faculty Excellence
Provost's Office
George Mason University
4400 University Drive MSN 4D6
Fairfax, VA 22030 USA

Prof. Eby:

I am writing in strong support of Professor Huzefa Rangwala's case as finalist for the 2014 GMU excellence in teaching award. I have known Professor Rangwala since he joined the Department of Computer Science at GMU in Fall 2008. Through the years, I have been able to observe Huzefa both as a teacher in a classroom and as a mentor of graduate students. I have collaborated with Huzefa on a number of projects, we co-advised several graduate students, and we coordinated the teaching of a graduate class on data mining.

Huzefa is an unquestionably outstanding teacher and mentor. He has mastered a rich set of effective techniques for helping students to learn. In the classroom, he conveys a natural passion for teaching; he constantly interacts with the students and keeps them engaged with questions. For the data mining class we co-taught, Huzefa set up a project as a fun competition involving the analysis of Big Data through Kaggle (<http://www.kaggle.com>). He also set up the evaluation procedure to rank students' solutions. Once the students turned in their solutions, he organized the students in groups so that they could discuss each other solutions, difficulties encountered, and the pros and cons of the applied methodologies. This was an effective approach for students to learn from each other experience throughout the project. Students stayed engaged, and demonstrated enthusiasm and participation.

Huzefa is also an excellent mentor and advisor of graduate students. I have participated in many discussions with Huzefa and students we have co-advised. Huzefa is a creative and pro-active researcher. As such, he is excellent in providing research directions to students, injecting ideas in the conversation, while challenging the student to follow up with some thinking on their own. He organizes weekly lab meetings with students, and regularly provides feedback to student on how to improve their presentation skills.

In summary, I strongly believe that professor Huzefa Rangwala has demonstrated excellence in

his teaching activities.

If you have any further questions, please do not hesitate to contact me.

Sincerely,



GMU Teaching Excellence Committee
4400 University Dr.
Fairfax, VA 22030

Dear GMU Teaching Excellence Committee:

Huzefa was my advisor at GMU from the summer of 2010 to the fall of 2013 when I successfully defended my dissertation. During this time, he has shown by example how to be a great researcher, teacher, and generally good person.

I first met Huzefa when I was looking to become a research assistant during my first semester at George Mason. I talked to a number of GMU professors about their work, and Huzefa definitely stood out from the others, immediately impressing me with his enthusiasm.

After this introduction, I decided to take Huzefa's computational biology class. Here, he also showed his dedication to teaching by always bringing positive energy to the classroom. The class started at 7:20pm and ended 10pm, and most of the students arrived looking pretty tired. Every evening, I was surprised at how energetic Huzefa was and how, at some point, this energy usually seemed to transfer over to the class.

In the class, Huzefa also tested out a number of "experimental" teaching techniques, and one of these worked particularly well for me. This technique was designed to teach students four different bioinformatics algorithms. Each of these algorithms was described in a separate research paper, and the papers were difficult enough that few students would be able to fully understand all of them within the span of two classes. To get these ideas across more efficiently, Huzefa broke the class into four groups and assigned each group a different paper. In the next class, each group discussed their assigned paper. Then the groups took turns presenting their papers to the rest of the class. For me, these group presentations were effective teaching tools because I was able to hear a description of each method in more accessible terms. Also, I was able to achieve a very thorough understanding of the paper that I was assigned because my group had to distill it down to a few core points for the presentation. In general, Huzefa's combination of enthusiasm and constant willingness to try out new ideas to help students learn better impressed me as characteristics of an excellent teacher.

Most of my interaction with Huzefa has been during my dissertation research, where we have coauthored five papers together. One danger of research is that it is easy to get discouraged. I have often been excited about an research idea, only to be disappointed when it failed later on. One distinguishing characteristic of great researchers is their ability to keep working even after suffering many initial failures. Huzefa certainly helped me improve in this respect during our work together. After an idea of mine didn't work as well as expected, he was always positive about it, and this optimism encouraged me to keep working hard until I resolved the problems in my current project. Huzefa's ability to pass on enthusiasm toward research to students is a key characteristic of a great mentor.

Another reason why Huzefa has been an excellent advisor is because he allowed me flexibility in choosing my own research topics. For an academic advisor, keeping the balance between giving students specific guidance and letting them work independently can be difficult. In my case, I feel that Huzefa was able to do this very well, keeping me excited about my research, which, in turn, improved the overall quality of my work.

Huzefa has been a great teacher, mentor, and has served as an example to me of how to approach academic work and life. In all of our conversations and interactions, Huzefa has, above all, impressed me as being a thoroughly nice guy and good person. He has guided me successfully through the process of completing a PhD-level work at George Mason University, and I have definitely enjoyed my time working with him.

Two years ago, I emailed Dr. Huzefa Rangwala to seek a research opportunity. I was thrilled when he responded and invited me to his office to discuss possible research ideas. Even during my first meeting with him, his passion for research and his love for helping students were evident. I am very fortunate to have been under his guidance; he truly is an extraordinary mentor. While working with Huzefa for the past two years, never once did I feel lost or alone about the work I was doing. Huzefa was always supportive and ready to help me. At the same time, he also left significant room and flexibility for me to explore various solutions to problems and to grow on my own. Huzefa was able to skillfully balance these two equally important but conflicting roles, thus making him a highly effective mentor.

The research I conducted under the mentorship of Huzefa had a tremendous impact on my career goals. Although I had always been interested in computer science, I had never known exactly what field of computer science I wanted to study. My research with Huzefa changed that; he showed me the practical use of computer science and machine learning when applied to bioinformatics. This area computer science was extremely interesting and appealed to me. Machine learning is such a powerful way to solve problems that don't seem intrinsically computer science related. I intend to study machine learning and bioinformatics more at college.

Apart from the specific project, I also learned a great deal about conducting research in general and gained valuable experience. I saw experienced first hand how to conduct research in a professional research institution. Through working with Huzefa, I learned how to think about and approach a problem, how to tackle barriers in my work, and how to create a well-written and comprehensive research paper. I could not have gained this invaluable knowledge elsewhere.

Not only was Huzefa an excellent mentor, he has consistently been supportive and interested in my endeavors outside of research. I can feel that he cares about me personally. I consider him not only a mentor, but also a great friend. I am really thankful that I met him and had the unique and rewarding opportunity of researching under his guidance.

Excerpts of Letters from Past Student Mentees:

- 1)
- 2)
- 3)
- 4)

[Dr. Rangwala] mentored me for my Masters for over a year, during which, I had the opportunity to work very closely with him. [...] From day one, Huzefa encouraged me to think out of the box and keep myself up to date with state of the art research [...] The relationship that Huzefa nourishes with all his mentees is significantly different from what I commonly knew. Having worked in the Internet industry for the past three plus years, I know now that the way he works has a lot in common with how the co-founders of successful startups work. Thus, the value created is at par with success stories that the digital press glorifies, except that our deliverables were appreciated publications in prominent journals. [...] Beyond successful deliverables, Huzefa cultivated in myself and other mentees the foundational research skills that are invaluable and applicable across all scientific domains. At my workplace, I constantly find myself utilizing this skill set to divide and conquer the most complex problems competently. I now consider myself equipped with an innate desire, confidence, and skill set to venture out and re-apply with perseverance the very same techniques in the industry. His persistent encouragement pushed me to keep adding value, bit by bit. In sum, he made research enthralling and challenging for us through his constant dedication and unparalleled mentorship skills. –

Huzefa's passion and vision for data mining and machine learning has been a great source of inspiration for me during my PhD career, and will always be. I admire him for giving me the freedom and encouragement to pursue my own interests, as well as providing me with priceless academic and professional advices. He always offered me effective feedback and suggestions about innovative ideas and how to reach my goals. I always feel more hopeful and passionate after having a meeting with him. [...] he encourages you to learn about new technologies and ideas, recommended me to attend research conferences and meet different people to get their insights and thoughts. Now I am a Data Scientist in a professional industry and I feel proud that I can utilize all the experiences, which I learnt under Huzefa's supervision. Having Huzefa as my mentor, it was a great graduate school experience and I am confident that I can make a positive difference in the world with all the things I have learned from him.

It's an honor to write an endorsement letter for Professor Rangwala. I took a couple of courses offered in Bioinformatics and Computer Science departments with Dr. Rangwala during my master's program. I found his approach very inspiring and motivating. I also did my final masters project under his mentorship. Professor Rangwala is a positive influence on me and a lot of my peers. He helped me with innovative ideas in my research project and supported through out by always being there with sound advice. He gave me just the right amount of help and support that helped me think like a researcher

but was never left us stuck. He encourages healthy discussions and coming up with new ideas and methods to solve a problem while reading old papers and research articles for guidance. I learnt to quickly read and analyze research papers in one of his classes which helped a lot for my project. Professor Rangwala is truly committed towards his students and in guiding them in class and research work. He has strong will to drive his students in the right direction. He is very prompt in responding back in time of need which shows his dedication. He is very patient and always ready to answer any questions I had during my research. He took time for meeting with each mentee individually and going through the updates and research data in detail. Dr. Rangwala also encouraged me to write a paper on my research. His passion towards research is truly motivating. I am very grateful for his mentorship style which helped me build my confidence during and after my master's program. I learnt to work efficiently in teams and individually during his classes which is a great help now in my current job as a bioinformatics software analyst where I have to coordinate between various teams and gather data. As a successful graduate and past mentee I can strongly say Professor Rangwala is a great mentor. He can be a positive influence on the students and young researchers. I am really glad I could have a wonderful and able mentor like Professor Rangwala.

[REDACTED],
NCBI/NIH

As a Bioinformatics Masters student at George Mason University, I was privileged to have had a mentor of the caliber of Dr. Rangwala. He puts in additional effort to make his courses extremely interesting through his challenging and in depth thought provoking assignments. As part of our Data Mining course, he used to assign top rated articles and papers as readings and expected each one of us to present our individual understanding of it. Being a Professional himself, he taught us how to interpret, analyze and understand research articles and papers to extract valuable information from them. I personally feel his teaching models were a key for me to develop a keen interest towards reading from a scientific perspective. One of the many special mentorship techniques of Dr. Rangwala was to blend different methods like lecture, discussions and quiz together in order to get everyone involved in the subject. Every lecture of Dr. Rangwala had a new challenge both in terms of technicality of the course and intellect of the student. He has great mentorship qualities which was immensely helpful in clarifying all our doubts instantaneously. He was one of a kind who would go an extra mile to help students solve problems with their coursework and provide letters of support anytime they needed. I was fortunate to have had an opportunity to complete my Masters project under Dr. Rangwala's guidance. His expertise in computational techniques and problem solving abilities helped me overcome all the hurdles in my project work. He added various new dimensions to my project through his valuable suggestions. He always gives personal attention to each and every student, encourages them and provides challenging tasks to aid their individual development. I strongly support his nomination for this valuable and esteemed award.

[REDACTED]



Huzefa Rangwala <rhuzeafa@gmail.com>

Request for Endorsement Letter



Professor Rangwala is absolutely worthy of the many awards and accolades he has received. I am fortunate to have taken his course CS 365 - Computer Systems Architecture last fall. The course ranks as one of the most informative and well designed courses I have had throughout my undergraduate career, and what I've learned has provided a foundation for my continued education in topics such as a multithreaded programming, computer networking, and operating systems. Professor Rangwala's course has also built confidence in my understanding of computers and computer hardware; I will be assembling my own computer this summer, which is an undertaking I would not have attempted before taking his course.

Another one of the biggest advantages of taking Professor Rangwala's course was in learning how to teach. I am currently enrolled in a Ph.D. program with goals to become a computer science professor. Many of the notes I took in CS 365 were less about the subject material and more so about Professor Rangwala's pedagogical practices. His lively and engaging presentations were a big takeaway, not to mention his friendliness and accessibility towards his students. Professor Rangwala also demanded a lot of his students. He expected us to learn and participate in class in a reasonable manner that truly pushed us to learn. These are takeaways that I have kept in my notebook as foundations for what I aim to achieve as a computer science professor.

FEEDBACK FROM STUDENTS IN CLASS

[Google Form Used for Collecting Anonymous Feedback]

Undergraduate Class Feedback

I took CS 465: Computer Systems Architecture with Dr. Rangwala in the Fall 2012 semester at George Mason University. Dr. Rangwala is a professor that cares deeply about his students. This became clear to me early in the semester – Dr. Rangwala offered to lend me an extra copy of the textbook that he had upon learning that I did not have much money left over to pay for textbooks... Dr. Rangwala is also a fantastic teacher. He explained the material well and often engaged the class using in-class exercises and discussion questions. These greatly helped me in understanding the material [...] class and assembly projects instilled in me a great appreciation for high-level abstractions and programming languages. Dr. Rangwala is a great teacher and I will never forget his kindness and generosity - [REDACTED]

Overall, your teaching in class has been amazing. Instead of focusing on what will be on the exam, you teach the class in a way that encourages independent thinking. Most professors teach what is needed to pass the class, but you teach what is important about the subject and tell us to think beyond the confinement of the classroom walls. You managed to make complex things very simple such as how the processor executes machine instructions. You taught how various parts of the computer work and how you can apply that knowledge to various scenarios instead of teaching us to just memorize..... It was an absolute pleasure having you as my professor because I have been able to think beyond my graduation as an undergraduate. I now have goals and ambitions of what I want to do in the computer science field. – **Anonymous Student.**

Except for the processor and cache sections, I don't really enjoy the topics in this class (not your problem). However, your enthusiasm when presenting the material really kept me interested during lectures. You were always very loud and clear when presenting topics and you were obviously open to answering questions. It was refreshing to see a professor that has the patience to answer all of his students' questions with that level of respect. Providing hardcopy solutions to the homework and exams were also very helpful. Thanks for the awesome semester, and good luck with your nomination! – **Anonymous Student**

I like how you try to add a little humor in your lectures. I like how you make sure students understand the material by asking for in class participation. I learned a lot in this class and enjoyed it very much. It's a hard topic so some of us we're unsure but you took the time to go over problems for an entire week before the exam. That helps a lot and shows that you want your students to succeed. Over all a positive experience and I feel the award would be well deserved - [REDACTED]

Dr. Rangwala is one of my favorite teachers so far. I am now a senior. It was nice that Professor Rangwala was personal when addressing questions to you. By the first couple weeks, he had all his students' names memorized.

The class was fast paced, but we learned a lot of material. The programming projects were a blast, but they weren't easy. They were the type of problems that you need about the same time to think about an appropriate algorithm before you actually start.

It has also been a pleasure to have taken a course with him. For those students that did not care much about their grade and acted out or came late, he let them know that they were interrupting the class. Pretty soon everyone was attentive. This is so important to me. Plenty of teachers let those things by and the other students suffer from it. He keeps the class engaged by taking in questions from last class, quick reaction questions to specific students, and hands out in-class exercises regularly. The exercises given really helped shape my understanding of the previous classes. It felt like a smooth flow through the course. There were times when I was not as fast as the class at an exercise and needed more of an explanation. I cleared these issues up with him, briefly in office hours, or during review or while anyone had questions before the next class. Or even on Piazza. Anyway, there were no excuses not to get in touch with him and he replied promptly.

It was nice to walk into his office hours. He was eager for us to tell him what we know so he can solve our problems. Professor Rangwala's teaching style and helpfulness excelled my computer science knowledge further than before, which makes me proud and confident to soon become a graduate! – **Anonymous Student**

Dr. Huzeifa Rangwala has fantastic motivation when teaching. He purposefully samples the class such that a wide variety of students are involved in class discussions on a regular basis. His use of worksheets and open discussions helped when working problems as a class. Dr. Rangwala handles criticisms and challenges from students extremely well. So well, in fact, that he is often able to successfully turn a criticism/challenge around on the student and into a lecture contribution if the comment is appropriately out of line. Dr. Rangwala's class was easily my most looked forward to class of the semester. I would recommend any fellow student to take any one of his classes. [REDACTED]

Graduate Class Feedback

Your ability to integrate real-world problems into both your teaching and the assignments was refreshing. All too often instructors cling to ivory-tower theory with a disdain for application. You teach the theory necessary to understand a solution, but never pursue it for its own sake. The focus is entirely on solving tough, meaningful problems for people.

Our discussion of PCA is a perfect example. We talked about how it works (by finding orthogonal eigenvectors that maximize the variance) and why it works (more variance = more information; addresses the ""curse of dimensionality"") and continued to discuss when it is used for actual problems. —Anonymous Student.

Huzefa, you taught me about data mining this past semester with an excitement for teaching that I haven't seen yet at George Mason, and this was my fourth semester. I was thrilled to be a part of a class that was both informative and fun. I found myself asking if that was possible several times during the semester. I'm a part-time graduate student and I'm at the age where learning something is more important to me than my grades in a class. I know that good grades don't mean much if you don't remember what you learned when the class is over! The class project we did was a fantastic example of the difference your teaching makes: you want people to learn concepts by asking them to apply them to a context of their own choosing. This truly engages people – how much more exciting to work on a large project if the work is something one is passionate about. Not only does it capture one's attention, but the concepts learned while working through a problem important to oneself are not quickly lost, I believe. Thank you for a great semester and for engaging us with enthusiasm and encouragement. It's obvious how much you care about teaching. —Anonymous Student.

Classes are always fun with lot of interaction and has a really good way of explaining thinks. He is very clear about what the course structure is and assignments cover a lot of the subject . Had a very exiting class under him and always available for doubts. —Anonymous Student.

Section 6: Closing Statement

Huzefa Rangwala, Ph.D.

I hope that my teaching portfolio could unequivocally convey my passion for teaching, mentorship and scholarly pursuits.

From a teaching perspective, I hope it walked you into one of my classes where I constantly strive to engage my students with new and innovative learning techniques. I focus on solving real world problems that improve society; within and outside the classroom. By doing this I am constantly learning; learning new programming skills, learning new analysis methods, learning new technologies and learning new teaching approaches.

I hope the portfolio also provided evidence towards my pursuits as a research scholar. I have tried to involve everyone; at all levels (doctoral students, undergraduates and high school students) and from multiple disciplines (computer science, bioengineering, biology, digital history and environmental sciences). I have embraced the entrepreneurial spirit of colleagues at Mason and been involved with the Students as Scholars Office to make curriculum improvements that inspire all undergraduates to pursue research. I have collaborated with faculty members from Environmental Sciences and Bioengineering to win competitive research awards that expose students to cutting-edge research.

Several positive things have happened to me in the past five years at Mason. I have won several competitive research grants (totaling \$4.4 million dollars). I am also the 2013 National Science Foundation (NSF) Early Career Faculty Award Recipient. This award is NSF's most coveted and prestigious award that supports a small number of junior faculty who exemplify the role of teacher-scholars through outstanding research, excellent education and the integration of education and research for creating broader impact in the society and lives of students. I was also the winner of the 2013 Engineering School Teaching Award, the 2012 Computer Science Department Teaching Award, the 2011 Computer Science Department Junior Researcher Award and the 2014 State Council for Higher Education in Virginia (SCHEV) Outstanding Rising Star Award nominee from Mason.

I am grateful to the Mason Teaching Excellence Committee for selecting me as a finalist this year. It is a honor to know that I am making a difference at Mason in my role that faces our students and creates an improved and better learning environment. I appreciate your time in reviewing my portfolio.