OLD DOMINION UNIVERSITY

What's Grad School All About?

Dr. Michele C. Weigle

CS 800 - Research Methods

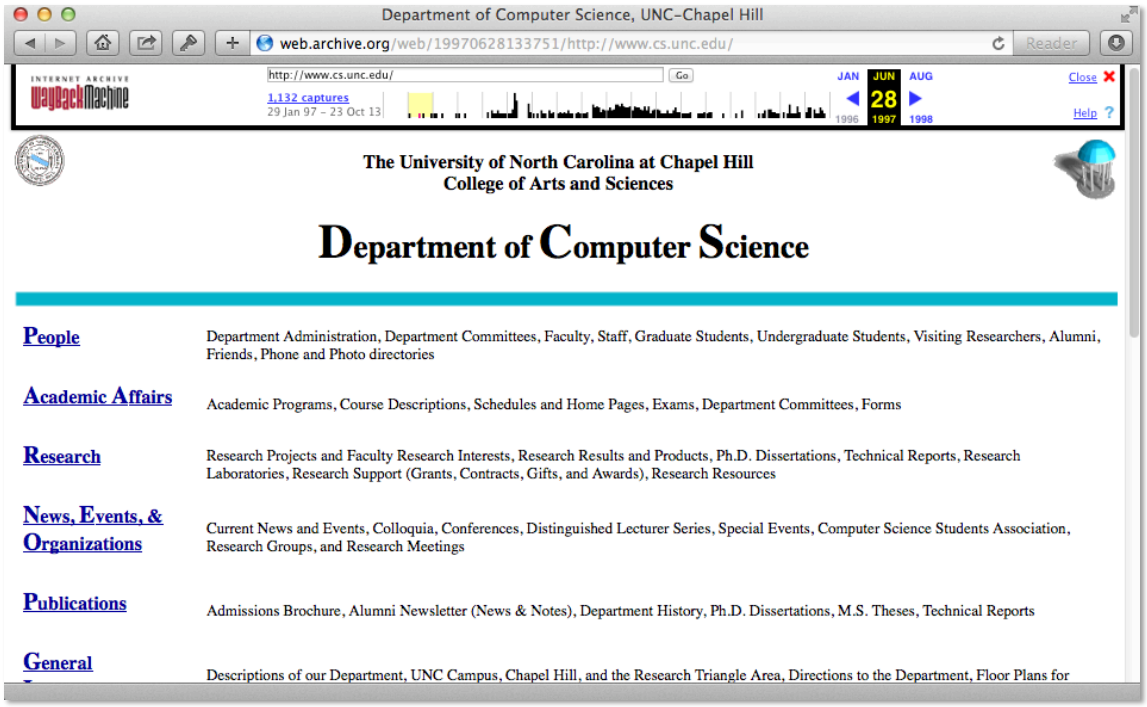
Week 1

January 14-16, 2020

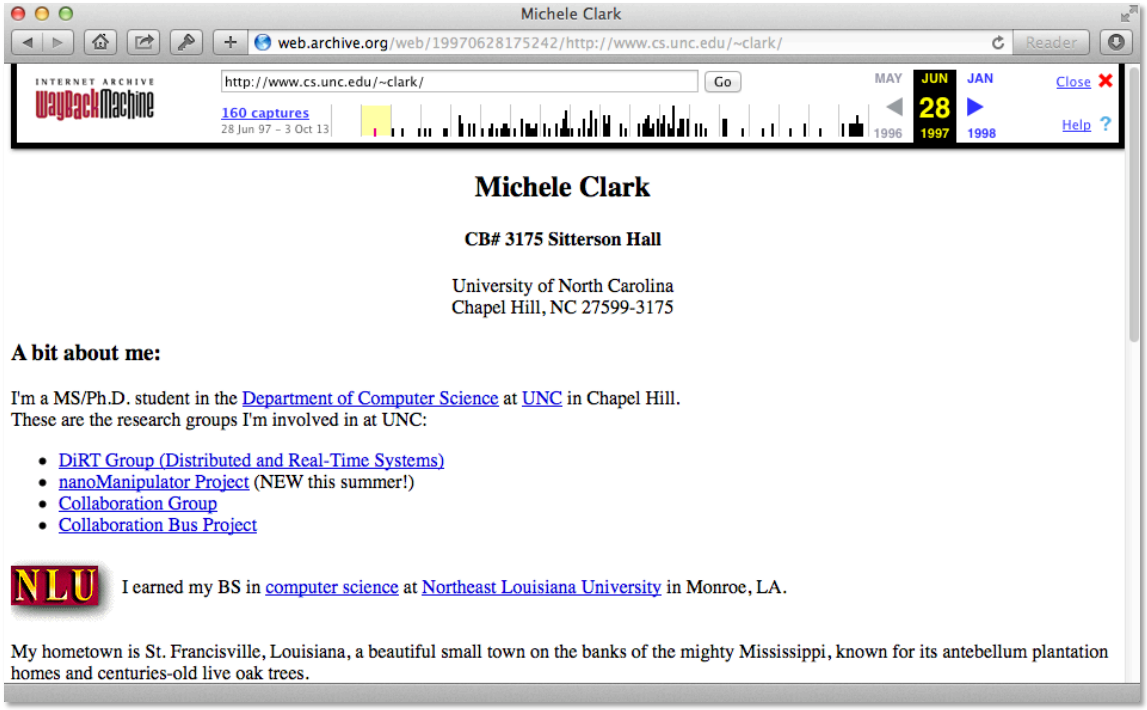
@weiglemc http://www.cs.odu.edu/~mweigle/

My Journey Through Grad School

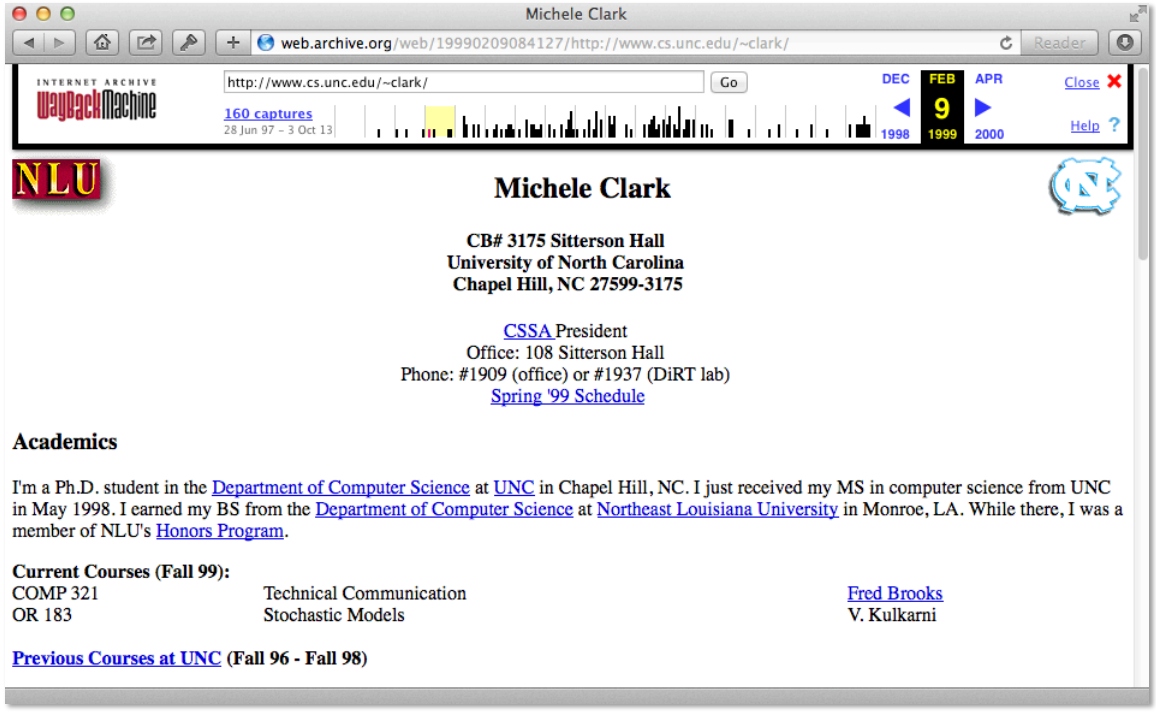
UNC-CS



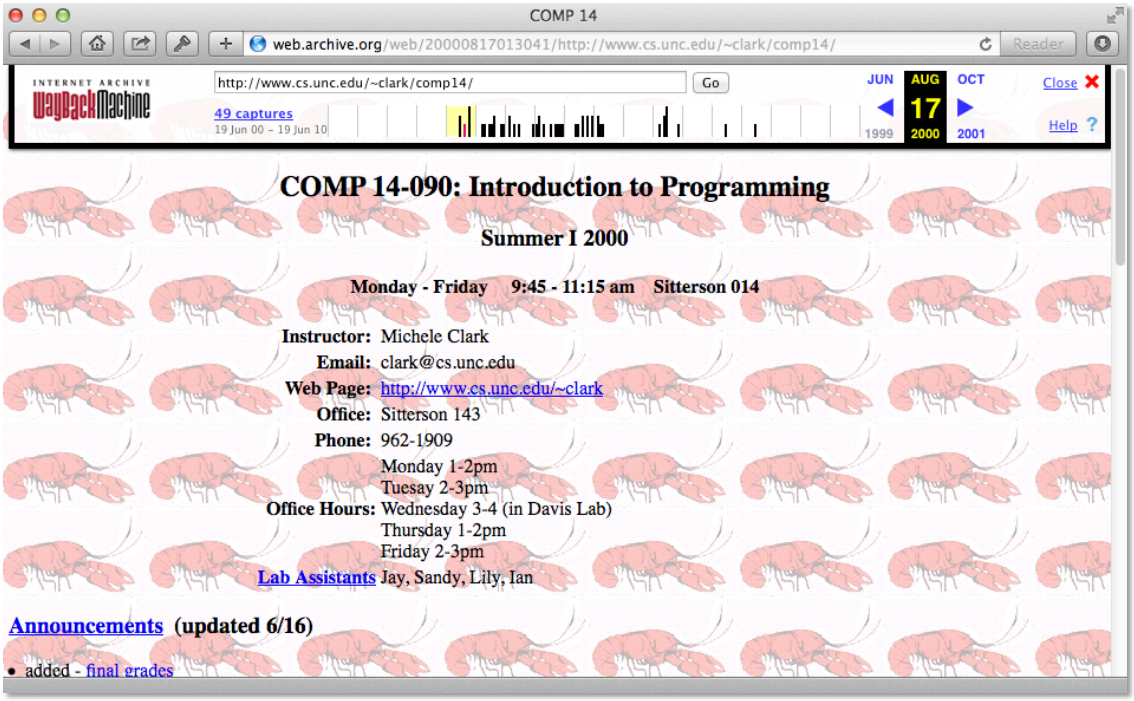
My CS Homepage - 1997



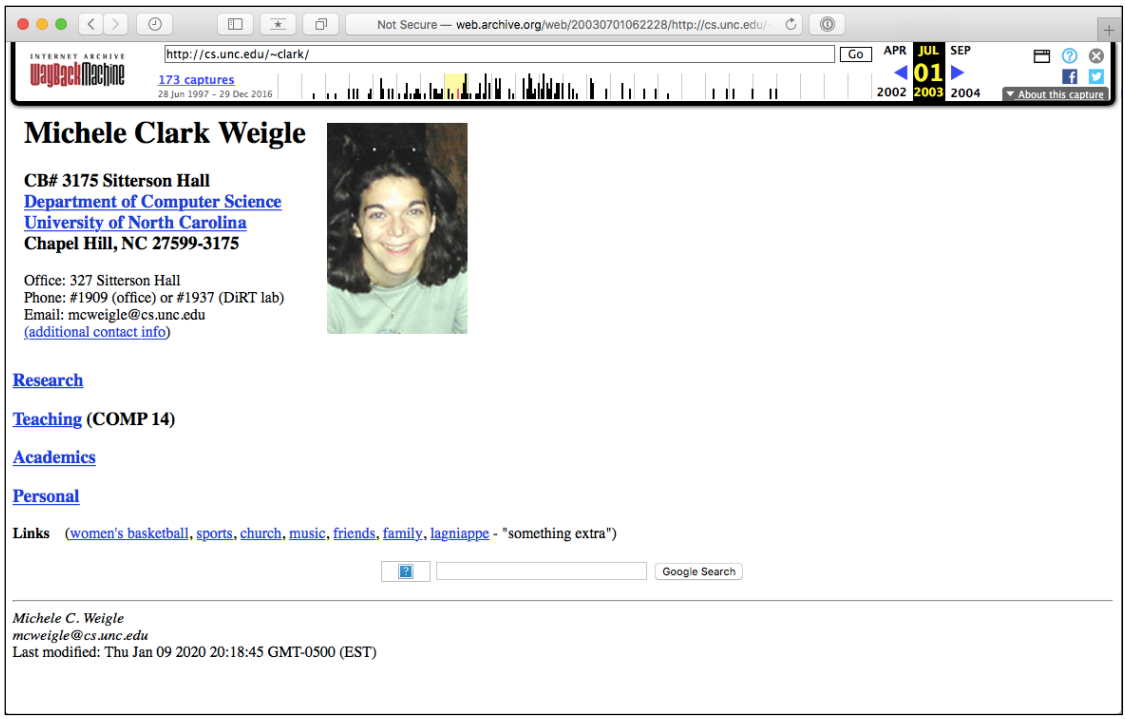
CS Student Assoc Pres - 1999



Teaching - 2000



Last Snapshot Before Grad - 2003



Grad School

- I loved grad school
  - and I love it even more now that I can look back fondly on it
- No "real-life" responsibilities
  - single, no kids, no house
- Developed life-long friendships
- Met people from all over the world
- Worked with really smart people
- Worked on interesting projects
- Got to code!
- Attended interesting (sometimes!) classes
- Got to travel (for free) - Italy, Sweden, Lake Tahoe, Norfolk

Outline

- Difference between MS and PhD
  - PhD vs. MS requirements at ODU
- Life as a PhD Student
- What is research?
- What is a dissertation?
- Surviving as a PhD Student

Acknowledgements: Some slides and material are courtesy Dr. Michael Nelson, Dr. Justin Brunelle, Dr. Kris Cooper, and Dr. Tracy Camp - gathered from course materials and personal conversations.

MS vs. PhD

MS vs. PhD – Requirements at ODU

- MS
  - 34 hours of coursework (11 courses)
  - project (only 10 courses) or thesis (only 8 courses)
    - project - usually a software development project
    - thesis - requires some novel research contribution
  - usually partial tuition waiver
  - ~2 years
- PhD (after MS)
  - 48 credit hours
  - 24 hours of coursework (8 courses)
    - 4 "real" courses (not seminar or special topics) from 3 different research areas
    - CS 800 - *this course*
  - 24 hours of dissertation credits
  - full tuition waiver
  - variable (often 3-5 years after MS)

MS vs PhD - Jobs

- A masters degree equips you to do high level, complex design and potentially lead software engineering teams.
- A PhD degree equips you to do original research and potentially lead R&D teams.
- "With a Ph.D. you will have a better chance of spending the rest of your life doing what *you* want to do, instead of what *someone else* wants you to do."
  - William Lipscomb, a Nobel Prize winner in chemistry

https://www.quora.com/What-are-the-main-differences-between-a-Masters-and-a-PhD-in-computer-science/answer/Vijay-Chidambaram
Vijay Chidambaram, computer science grad student

BS vs. MS vs. PhD

- BS - you are given the questions and the answers
- MS - you are given the questions and mostly you get to find the answers
- PhD – you must come up with the questions *and* the answers
- It sure is a lot simpler when you are given the questions ***and*** the answers.

Dr. Kris Cooper (my undergraduate advisor)

Remember your Bachelor's?

- You are given a course schedule
- Instructors pose questions, you give answers, you're told if you're right or wrong
- You showed up and did what you were told for ~4 years and got a degree in return
- You are certified to be competent in your topic area (e.g., CS)

And you may have/are working on a MS

- Increased depth and/or specialization
  - You have a few core courses
  - You probably selected about half of your courses
- This is your introduction to research
  - You are asked a question and you need to determine the answer *and* whether or not you are correct
    Answers become less definite!
- You are certified to understand the complexities of your field, potentially with a specialization

What does a PhD look like?

- No emphasis on courses (you only take a few)
- No one has the questions OR the answers!
  - You embark on original research
  - You define the questions, the answers, and have to prove that you are correct
  - Wide-open, not concrete
- "You know more and more about less and less"
  - You are highly specialized
- You will work extremely closely with your advisor (so make sure you work well together!)

A PhD is All About Questions

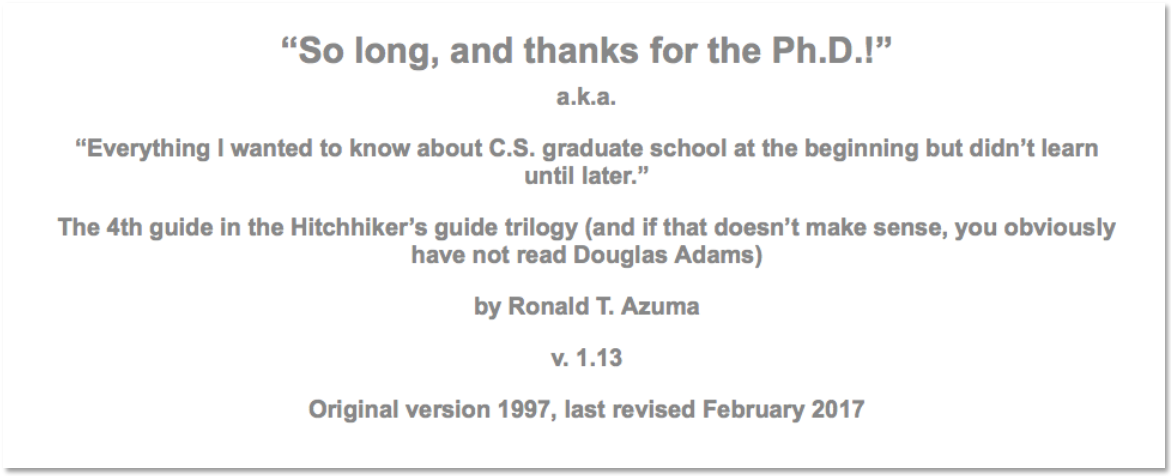
- "A PhD teaches you how to ask the right questions" http://blog.skanev.org/2013/03/why-not-do-phd-in-computer-science.html

- Your PhD advisor's job is to ask questions about your work
  - why does the output/graph/result look like this?
  - what would happen if you ran the experiment/analysis another way instead?

Life as a PhD Student

So long, and thanks for the PhD!

Ron Azuma's classic article (PhD, UNC 1995)
http://www.cs.unc.edu/~azuma/hitch4.html



"So long, and thanks for the Ph.D.!"

a.k.a.

"Everything I wanted to know about C.S. graduate school at the beginning but didn't learn until later."

The 4th guide in the Hitchhiker's guide trilogy (and if that doesn't make sense, you obviously have not read Douglas Adams)

by Ronald T. Azuma

v. 1.13

Original version 1997, last revised February 2017

What is a PhD?

- A PhD program is very different from getting a Bachelor's degree, and you must treat it as a strange type of job.
  - *Initiative*, *tenacity*, *flexibility*, *interpersonal skills*, *organizational skills*, and *communication skills* are all critical and not things that universities typically test for in selecting incoming students.
- A PhD is a means to an end: employment in a startup, commercial business, government or industrial research lab, or academia.

http://www.cs.unc.edu/~azuma/hitch4.html

Where do GRAs Come From?

- Academia is a business, and "graduate student" is a job title.
- Faculty write grant proposals to external agencies (NSF, NEH, IMLS, …).
  - fund GRA stipends, travel, small amounts of faculty summer support
  - *without grant funding, there is no GRA funding*
- These agencies expect concrete deliverables (software, publications, etc.).
- *My chances of future funding are largely based on your performance.*

http://www.cs.unc.edu/~azuma/hitch4.html

Treat Your GRA Like a Job

- You must prove to your professors that you are capable of
  - getting the work done,
  - being a team player,
  - communicating your results, and
  - most of the other characteristics needed to do well in regular jobs.

http://www.cs.unc.edu/~azuma/hitch4.html

Skills Gained Through PhD

- **Ability to work independently**
- **Critical thought**
  - A PhD candidate learns to critically examine the thoughts of others and pick out the pros and cons.
- **Perseverance**
- **Ability to work with poorly defined goals**
  - One of the bigger hurdles of the PhD is that there is no clear cut goal.
  - No one can exactly say these are the things you need to do every day.
  - Research as such involves going back and forth, exploring blind alleys and so forth.
- **Effective communication**

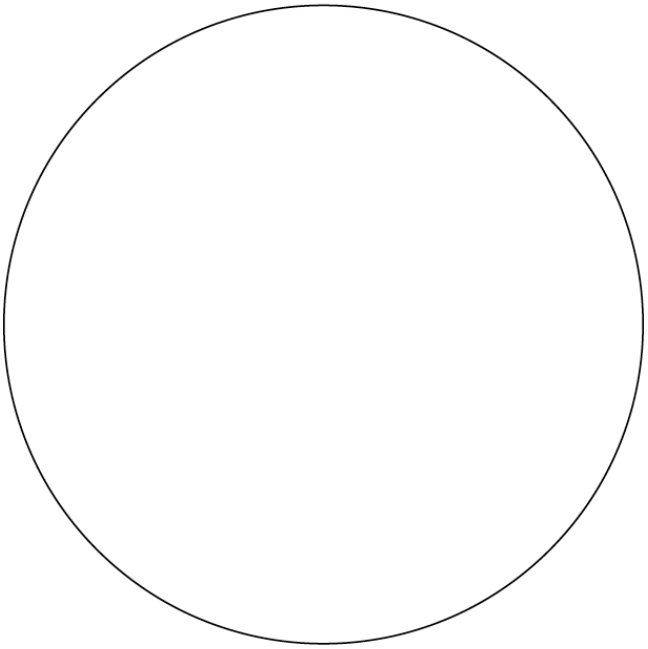
https://www.quora.com/What-are-the-main-differences-between-a-Masters-and-a-PhD-in-computer-science/answer/Vijay-Chidambaram
Vijay Chidambaram, computer science grad student

What is Research?

Matt Might (http://matt.might.net/), a professor in Computer Science at the University of Utah, created "The Illustrated Guide to a Ph.D." to explain what a Ph.D. is to new and aspiring graduate students.

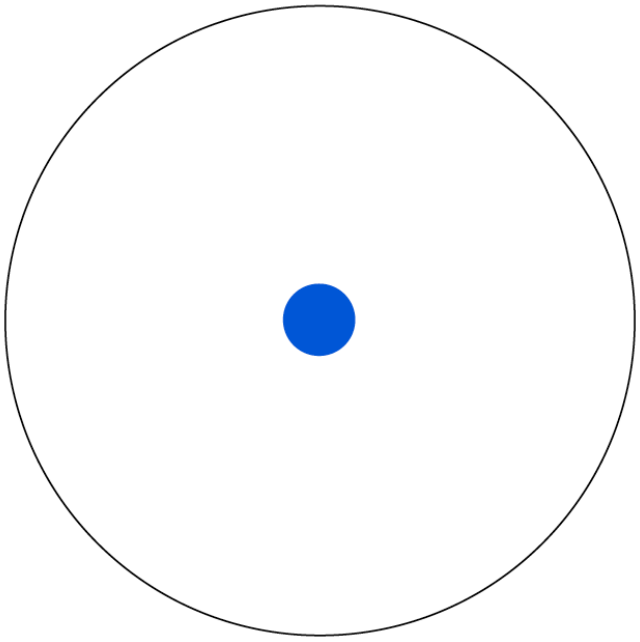
[Matt has licensed the guide for sharing with special terms under the Creative Commons license.] *http://matt.might.net/articles/phd-school-in-pictures/*

Imagine a circle that contains all human knowledge



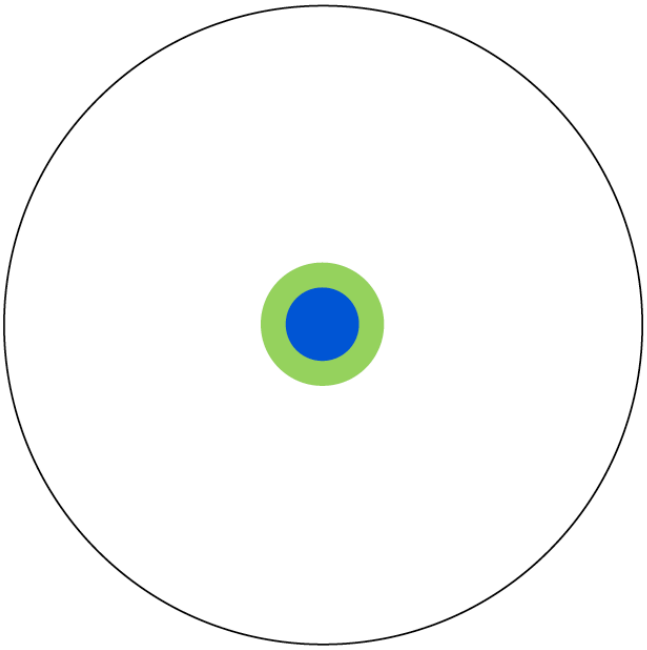
http://matt.might.net/articles/phd-school-in-pictures/

By the time you finish elementary school, you know a little



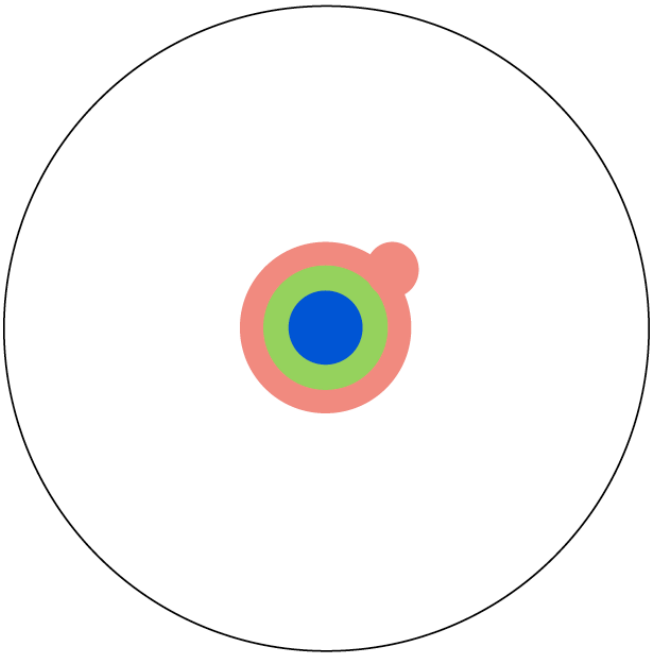
http://matt.might.net/articles/phd-school-in-pictures/

By the time you finish high school, you know a bit more



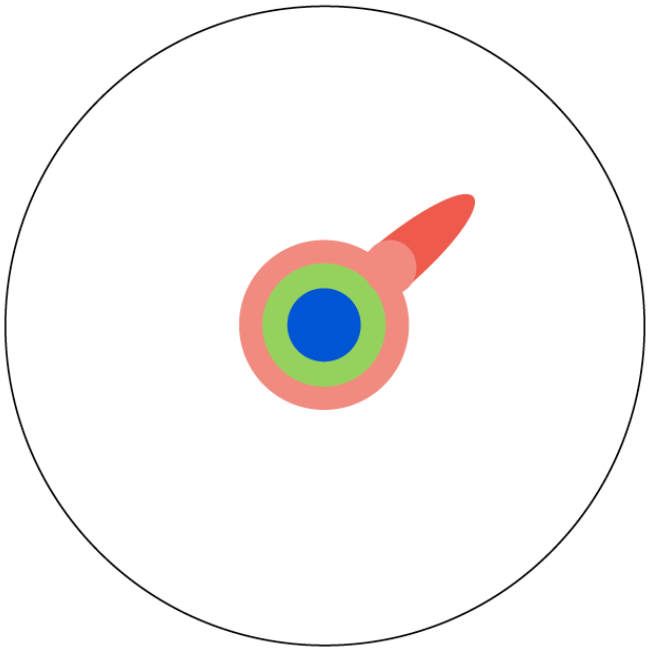
http://matt.might.net/articles/phd-school-in-pictures/

With a bachelor's degree, you gain a specialty



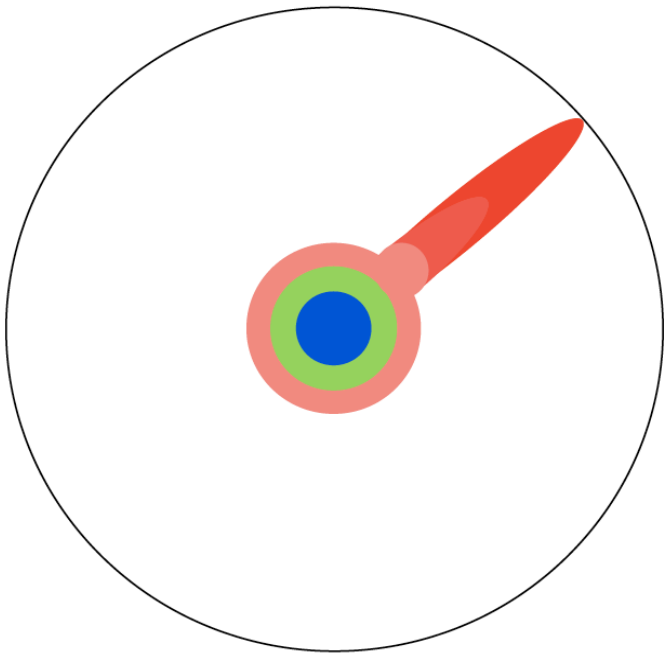
http://matt.might.net/articles/phd-school-in-pictures/

A master's degree deepens that specialty



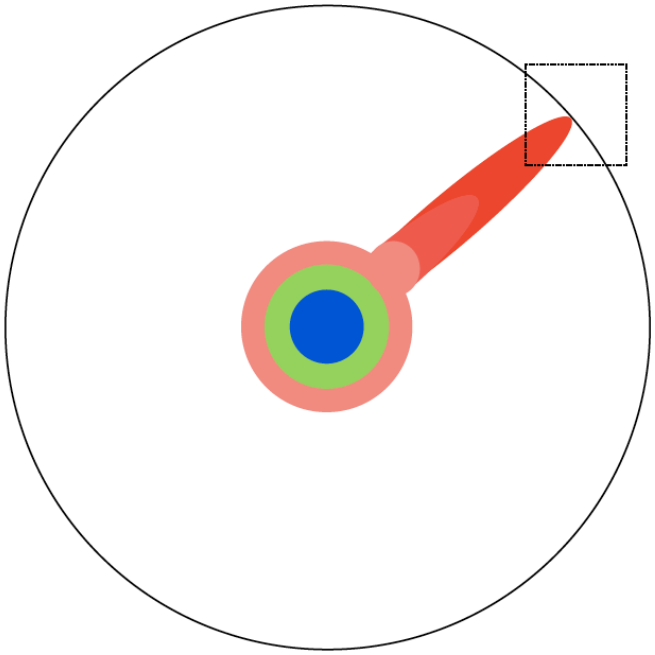
http://matt.might.net/articles/phd-school-in-pictures/

Reading research papers takes you to the edge of human knowledge



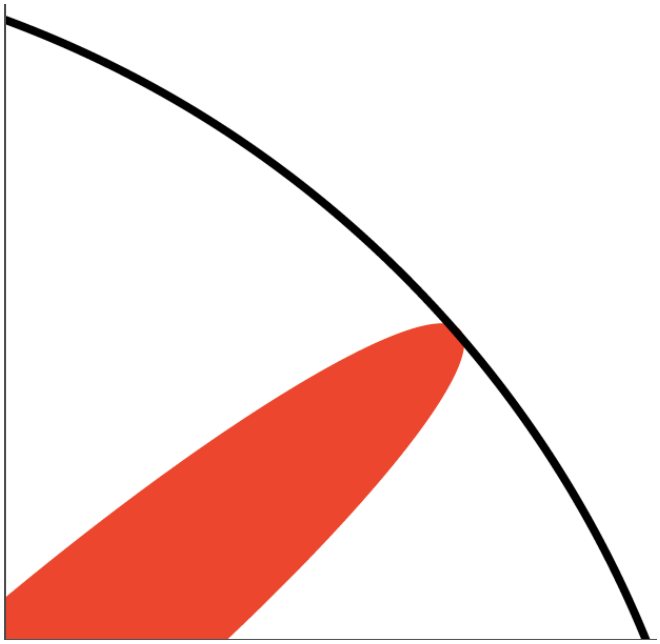
http://matt.might.net/articles/phd-school-in-pictures/

Once you're at the boundary, you focus



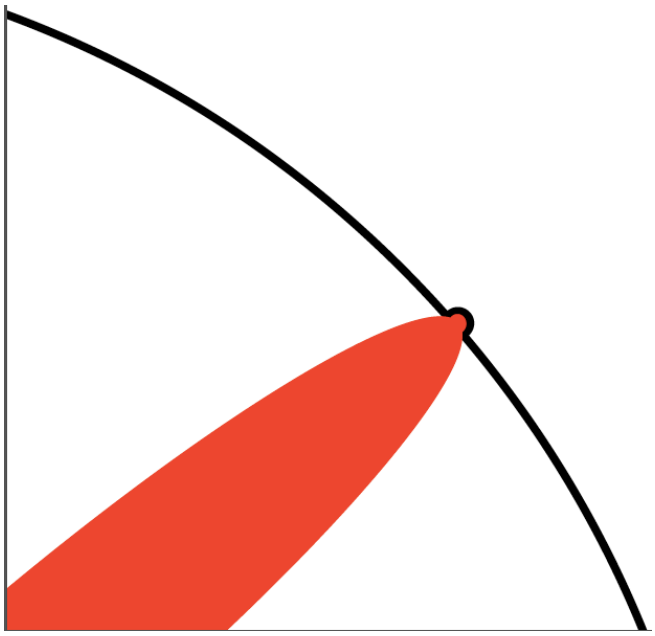
http://matt.might.net/articles/phd-school-in-pictures/

You push at the boundary for a few years



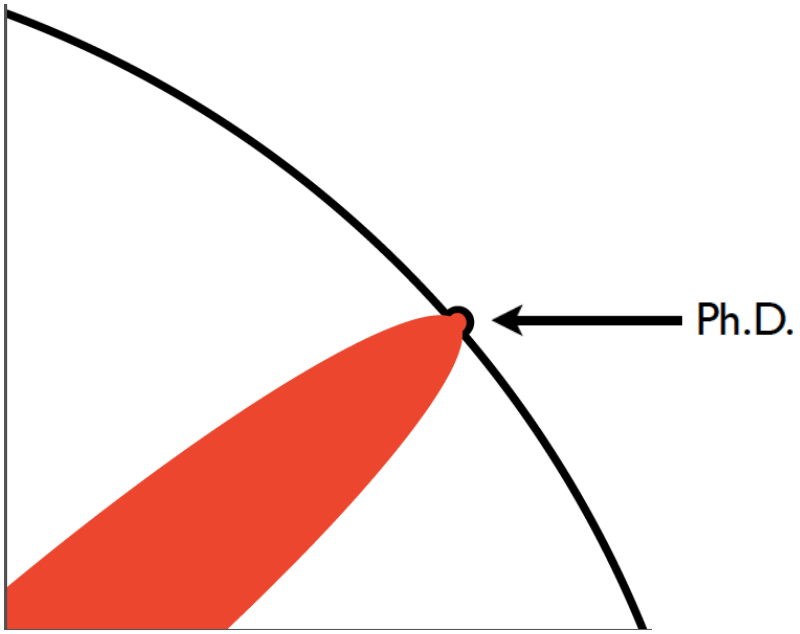
http://matt.might.net/articles/phd-school-in-pictures/

Until one day, the boundary gives way



http://matt.might.net/articles/phd-school-in-pictures/

And, that dent you've made is called a Ph.D.

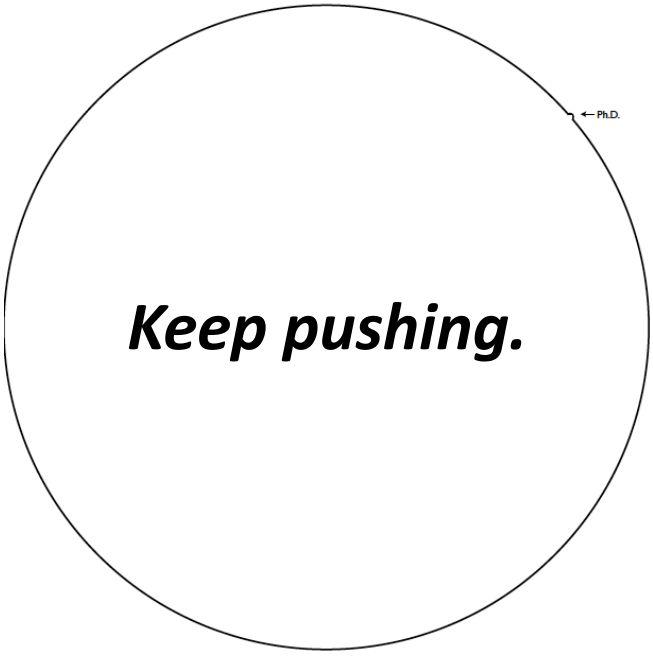


http://matt.might.net/articles/phd-school-in-pictures/

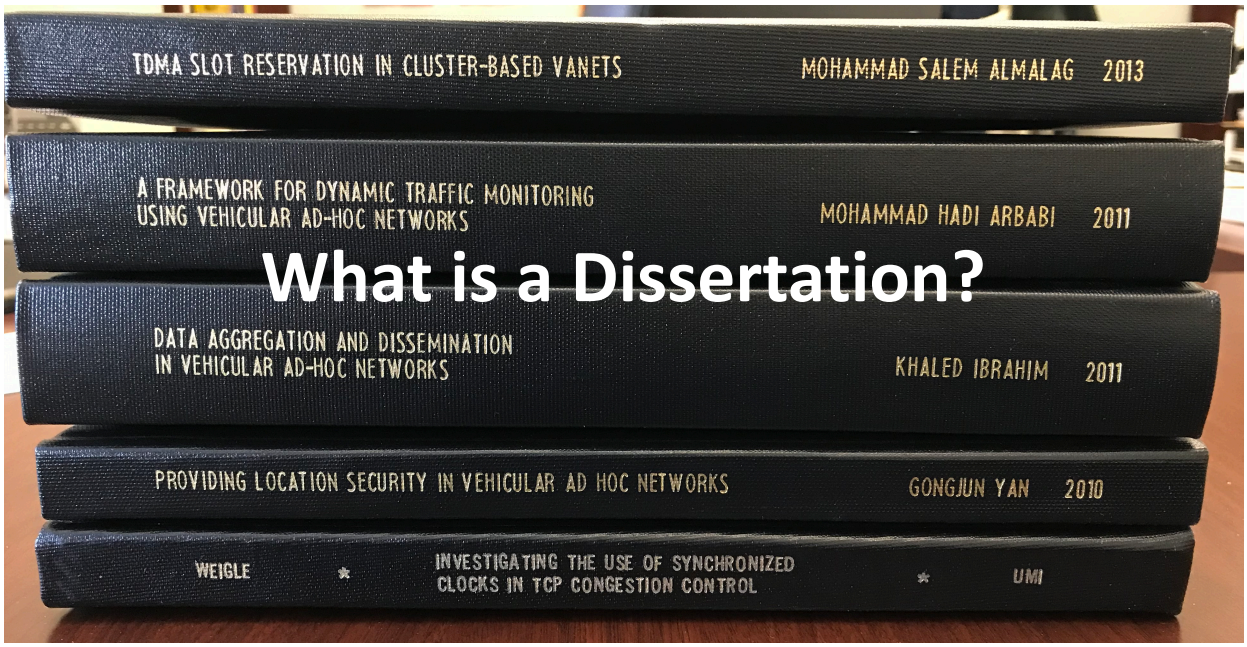
Of course, the world looks different to you now

http://matt.might.net/articles/phd-school-in-pictures/

So, don't forget the bigger picture



http://matt.might.net/articles/phd-school-in-pictures/

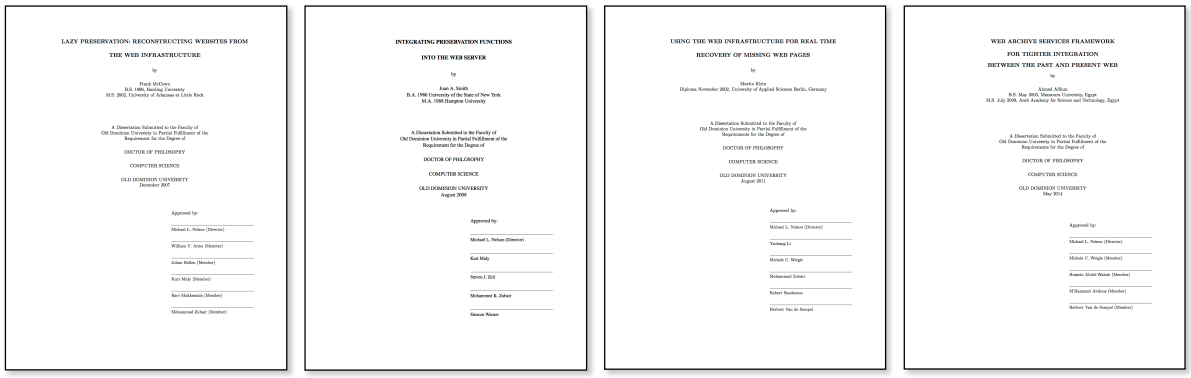


What is a Dissertation?

Before you write a dissertation, you should read one

- HW10 - read a dissertation and give a presentation on it
  - preferably a dissertation from an ODU graduate from your research group (if possible)
  - start talking with your advisor about suggestions now

ODU CS dissertations since 1989:
https://digitalcommons.odu.edu/computerscience_etds/



LAZY PRESERVATION: RECONSTRUCTING WEBSITES FROM THE WEB INFRASTRUCTURE

by

Frank McCown
B.S. 1995, Harding University
M.S. 2002, University of Arkansas at Little Rock

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirement for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTER SCIENCE

OLD DOMINION UNIVERSITY
December 2007

Approved by:

Michael L. Nelson (Director)

William Y. Arms (Member)

Johan Bollen (Member)

Kurt Maly (Member)

Ravi Mukkamala (Member)

Mohammad Zubair (Member)

INTEGRATING PRESERVATION FUNCTIONS INTO THE WEB SERVER

by

Joan A. Smith
B.A. 1988 University of the State of New York
M.A. 1990 Hampton University

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirement for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTER SCIENCE

OLD DOMINION UNIVERSITY
August 2008

Approved by:

Michael L. Nelson (Director)

Kurt Maly

Steven J. Zeil

Mohammad K. Zubair

Simeon Warner

USING THE WEB INFRASTRUCTURE FOR REAL TIME RECOVERY OF MISSING WEB PAGES

by

Martin Klein
Diploma November 2002, University of Applied Science Berlin, Germany

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirement for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTER SCIENCE

OLD DOMINION UNIVERSITY
August 2011

Approved by:

Michael L. Nelson (Director)

Yaohang Li

Michele C. Weigle

Mohammad Zubair

Robert Sanderson

Herbert Van de Sompel

WEB ARCHIVE SERVICES FRAMEWORK FOR TIGHTER INTEGRATION BETWEEN THE PAST AND PRESENT WEB

by

Ahmed AlSum
B.S. May 2003, Mansoura University, Egypt
M.S. July 2008, Arab Academy for Science and Technology, Egypt

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTER SCIENCE

OLD DOMINION UNIVERSITY
May 2014

Approved by:

Michael L. Nelson (Director)

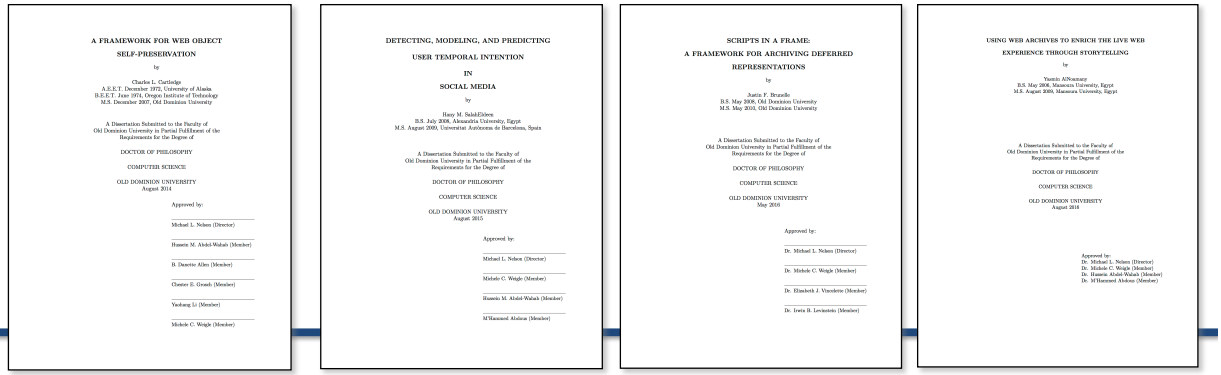
Michele C. Weigle (Member)

Hussein Abdel-Wahab (Member)

M'Hamed Abdoua (Member)

Herbert Van de Sompel (Member)

All dissertations, theses, and candidacy proposals have the same structure…



A FRAMEWORK FOR WEB OBJECT SELF-PRESERVATION

by

Charles L. Cartledge
A.E.E.T. December 1972, University of Alaska
B.S.E.T. June 1974, Oregon Institute of Technology
M.S. December 2007, Old Dominion University

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTER SCIENCE

OLD DOMINION UNIVERSITY
August 2014

Approved by:

Michael L. Nelson (Director)

Hussein M. Abdel-Wahab (Member)

B. Danette Allen (Member)

Chester E. Grosch (Member)

Yaohang Li (Member)

Michele C. Weigle (Member)

DETECTING, MODELING, AND PREDICTING USER TEMPORAL INTENTION IN SOCIAL MEDIA

by

Hany M. SalahEldeen
B.S. July 2008, Alexandria University, Egypt
M.S. August 2009, Universitat Autònoma de Barcelona, Spain

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTER SCIENCE

OLD DOMINION UNIVERSITY
August 2015

Approved by:

Michael L. Nelson (Director)

Michele C. Weigle (Member)

Hussein M. Abdel-Wahab (Member)

M'Hamed Abdoua (Member)

SCRIPTS IN A FRAME: A FRAMEWORK FOR ARCHIVING DEFERRED REPRESENTATIONS

by

Justin F. Brunelle
B.S. May 2006, Old Dominion University
M.S. May 2010, Old Dominion University

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTER SCIENCE

OLD DOMINION UNIVERSITY
May 2016

Approved by:

Dr. Michael L. Nelson (Director)

Dr. Michele C. Weigle (Member)

Dr. Elizabeth J. Vicecelette (Member)

Dr. Irwin B. Levinstein (Member)

USING WEB ARCHIVES TO ENRICH THE LIVE WEB EXPERIENCE THROUGH STORYTELLING

by

Yasmin AlNoamany
B.S. May 2006, Mansoura University, Egypt
M.S. August 2009, Mansoura University, Egypt

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTER SCIENCE

OLD DOMINION UNIVERSITY
August 2018

Approved by:
Dr. Michael L. Nelson (Director)
Dr. Michele C. Weigle (Member)
Dr. Hussein Abdel-Wahab (Member)
Dr. M'Hamed Abdoua (Member)

vi

TABLE OF CONTENTS

Page

LIST OF TABLES xi

LIST OF FIGURES xviii

Chapter

1. INTRODUCTION 1
   - 1.1 YOU ARE NOT A GADGET 2
   - 1.2 #SOPABLACKOUT: A CASE STUDY 9
     - 1.2.1 CRAIGSLIST 9
     - 1.2.2 WIKIPEDIA 13
   - 1.3 MODERN WEB ARCHIVING 16
   - 1.4 RESEARCH QUESTIONS 19
   - 1.5 DISSERTATION ROADMAP 21

2. BACKGROUND 23
   - 2.1 WEB ARCHITECTURE 23
   - 2.2 FROM HTML TO RICH INTERNET APPLICATIONS 27
     - 2.2.1 HTML 27
     - 2.2.2 REST PRINCIPLES 28
     - 2.2.3 JAVASCRIPT AND AJAX 29
     - 2.2.4 PROGRESSION OF WEB LANGUAGES 35
   - 2.3 TRENDING TOWARD THE DYNAMIC WEB 39
   - 2.4 HASHBANG URIS 41
   - 2.5 TRENDS IN WEB ARCHIVING 43
     - 2.5.1 MEMENTO 44
     - 2.5.2 ARCHIVING MORE THAN JUST THE DESKTOP WEB 45
       - The Personalized Web 46
       - The Mobile Web 47
       - The Linked Data Web 49
   - 2.6 DAVID HOCKNEY'S JOINERS 55
   - 2.7 SUMMARY 57

3. RELATED WORK 58
   - 3.1 ARCHIVAL EFFORTS AND TOOLS 58
   - 3.2 ARCHIVAL CHALLENGES: THE DEEP, HIDDEN, AND EPHEMERAL WEB 66
   - 3.3 WHAT MAKES A REPRESENTATION? 69
   - 3.4 MEASURING ARCHIVE QUALITY 70
   - 3.5 CACHING JAVASCRIPT-DEPENDENT REPRESENTATIONS 73

Examples, motivation, "why should we care?"

What basic technology do I need to understand the rest of the dissertation/thesis/proposal?

What have others done about this? (they provided the *correct* solution to some *slightly different* problem)

3.6 MONITORING THE CLIENT FOR DEBUGGING AND SECURITY 75
3.6.1 DETECTING MALWARE 76
3.6.2 DEBUGGING AND ERROR TRACKING 77
3.6.3 USER INTERACTIONS 78
3.7 SESSION REPLAY AND SHARING 79
3.8 SUMMARY 82

4. UNDERSTANDING THE CHALLENGES OF ARCHIVING JAVASCRIPT 83
4.1 REVISITING THE #SOPABLACKOUT: TECHNICAL DETAILS 83
4.1.1 CRAIGSLIST: REVISITED 83
4.1.2 WIKIPEDIA: REVISITED 85
4.2 JAVASCRIPT AND MEMENTOS OF MEMENTOS 89
4.3 ZOMBIES IN THE ARCHIVES 97
4.4 MULTIPLE REPRESENTATIONS 101
4.5 SUMMARY 108

5. MEASURING JAVASCRIPT IN THE ARCHIVES 110
5.1 MOTIVATING EXAMPLE 110
5.2 EXPERIMENT DESIGN 116
5.2.1 DATASETS 118
Twitter 118
Archive-It 119
Collection Differences
5.3 ARCHIVING THE RESOURCES
5.4 RESOURCE METRICS
5.4.1 URI COMPLEXITY
5.4.2 CONTENT COMPLEXITY
5.5 REQUESTS FROM HTML AND JAVASCRIPT
5.6 RESOURCE SET ANALYSIS
5.6.1 URI COMPLEXITY
5.6.2 CONTENT COMPLEXITY 125
5.7 ARCHIVING EXPERIMENT DISCUSSION 127
5.8 MISSING EMBEDDED RESOURCES 127
5.9 LEAKAGE 128
5.10 IMPACT OF ACCESSIBILITY 132
5.11 MEASURING ARCHIVABILITY IN THE INTERNET ARCHIVE 134
5.12 CHALLENGES IN MEASURING PAST PERFORMANCE 135
5.13 THE IMPACT OF JAVASCRIPT ON MEMENTO COMPLETENESS 140
5.14 SUMMARY OF FINDINGS 146
5.15 CONTRIBUTION TO RESEARCH QUESTION 1 147

6. MEASURING ARCHIVE QUALITY 149
6.1 MOTIVATING EXAMPLES 150
6.2 USERS' PERCEPTION OF DAMAGE 152
6.3 EVALUATING ORGANIC DAMAGE 157

Starting here, you're pretty much only talking about *your* work, contributions, etc.

Some flexibility in presentation and granularity: preliminary work that informs your framework / model / whatever …

Introduction to the framework / …, Then aspects of implementation / exploration / etc.

Chapters often roughly correspond to papers you've published (but chapter order doesn't have to follow chronological order of papers, and chapters are not necessarily 1-1 with pubs)

TPDL 2013
D-Lib 2013
IJDL 2016

6.4 CALCULATING MEMENTO DAMAGE 159
6.4.1 DEFINING D_M 159
6.4.2 WEIGHTING EMBEDDED RESOURCES 160
6.4.3 IMAGE DAMAGE CALCULATION 161
6.4.4 STYLE SHEET DAMAGE CALCULATION 161
6.4.5 THE D_M ALGORITHM 163
6.4.6 LIMITATIONS OF D_M CALCULATION
6.5 DAMAGE IN THE ARCHIVES
6.5.1 TURKER ASSESSMENT OF D_M
6.5.2 MEASURING THE INTERNET ARCHIVE 167
6.5.3 MEASURING WEBCITE 170
6.6 IMPACT OF JAVASCRIPT ON DAMAGE 175
6.6.1 CRAWLING DEFERRED REPRESENTATIONS 175
6.7 MEASURING ARCHIVE.IS 178
6.8 CONTRIBUTION TO RESEARCH QUESTION 2 180

7. A TWO-TIERED APPROACH FOR CRAWLING DEFERRED REPRESENTATIONS 184
7.1 MOTIVATING EXAMPLES 185
7.2 A FRAMEWORK FOR TWO-TIERED CRAWLING 188
7.2.1 SINGLE-TIER CRAWLING FRAMEWORK 188
7.2.2 TWO-TIERED CRAWLING APPROACH 189
7.3 COMPARING CRAWLS 191
7.3.1 CRAWL TIME BY URI 192
7.3.2 URI DISCOVERY AND FRONTIER SIZE 194
7.3.3 FRONTIER PROPERTIES AND DEDUPLICATION 197
7.3.4 DEFERRED VS. NONDEFERRED CRAWLS 200
7.4 CLASSIFYING REPRESENTATIONS 200
7.5 PERFORMANCE OF TWO-TIERED CRAWLING 204
7.6 MAPPING AND IDENTIFYING DESCENDANTS 207
7.7 DESCENDANT EQUIVALENCY 210
7.8 DESCENDANT CRAWLING APPROACH 211
7.9 EDGE CASES 216
7.10 DESCENDANT STATES 217
7.10.1 DATASET DIFFERENCES 217
7.10.2 TRAVERSING PATHS 220
7.10.3 IMPACT ON CRAWL TIME 221
7.11 ARCHIVAL COVERAGE 222
7.12 STORING DESCENDANTS 226
7.13 CONTRIBUTION TO RESEARCH QUESTION 3 230

8. FUTURE WORK, CONTRIBUTIONS, AND CONCLUSIONS 232
8.1 RESEARCH QUESTIONS REVISITED 233
8.2 FUTURE WORK 235
8.3 CONTRIBUTIONS 236
8.4 CONCLUSIONS 237

IJDL 2015 (first as JCDL 2014)

iPres 2015
JCDL 2017 (first as arXiv 2016)

Background or Related Work?

- Look at what you're citing
- Some guidelines
  - Background: textbooks, protocols, RFCs, surveys, older publications, etc.
  - Related Work: conference papers, journal articles, more recent publications
- Your advisor may have different ideas on this, they win
  - some combine background and related work chapters
  - some combine introduction and background chapters

Surviving as a PhD Student

Surviving the PhD

- Read Ron Azuma's guide and David Patterson's "How to Have a Bad Career in Research/Academia", https://people.eecs.berkeley.edu/~pattrsn/talks/BadCareer.pdf
- Perseverance
  - it can be slow, it can get boring, some days you just have to get through it
- Initiative
  - your advisor will rarely bug you each day to make sure you're working, must set your own goals
- Curiosity
  - PhD students are usually ready to graduate once they start asking their own questions about their data and research
- Coffee
  - *my PhD students told me to add this one*

A PhD is Not About Courses

- Most of what you learn in a Ph.D. program comes outside of classes:
  - from doing research on your own and in collaboration with your advisor
  - attending conferences
  - discussions with your fellow students

http://www.cs.unc.edu/~azuma/hitch4.html

Ph.D. Students Must Break Away From Undergraduate Mentality

- Grades don't matter much anymore
  - main form of evaluation is research progress (i.e., publications)
- There is no one who can tell you exactly what to do, so own your research
  - secret: *your advisors don't know all the answers!*
- When you graduate, you will be the world's expert on your dissertation topic

https://cacm.acm.org/magazines/2013/7/165494-ph-d-students-must-break-away-from-undergraduate-mentality/fulltext

Critical Skills Needed

- Initiative
- Tenacity
- Flexibility
- Interpersonal skills
- Organizational skills
- Communication skills

http://www.cs.unc.edu/~azuma/hitch4.html

Initiative

- The dissertation represents a focused, personal research effort where you take the lead on your own, unique project.
- Ph.D. students must show *initiative* to successfully complete the dissertation.
- If you never do any tasks except those that your professor specifically tells you to do, then you need to work on initiative.

http://www.cs.unc.edu/~azuma/hitch4.html

Tenacity

- No one can tell you in advance exactly how long the dissertation will take, so it's hard to see where the "end of the road" lies.
- You will encounter unexpected problems and obstacles that can add months or years to the project.
- If you are not *tenacious* about working on the dissertation, you won't finish.
- The best way to finish the dissertation is to do something every day that gets you closer to being done.

http://www.cs.unc.edu/~azuma/hitch4.html

Flexibility

- *Flexibility* means
  - taking advantage of opportunities and synergies,
  - working around problems
  - being willing to change plans as required

http://www.cs.unc.edu/~azuma/hitch4.html

Interpersonal Skills

- You need to build and maintain interpersonal relationships with your advisor, your committee, your research and support staff and your fellow students.
- Cultivating interpersonal relationships is mostly about treating people with respect and determining their different working styles.
  - Give credit where credit is due.
  - Acknowledge and thank them for their help.
  - Return favors – be a team player.
  - Respect their expertise, advice and time.
  - Apologize if you are at fault.
  - Realize that different people work in different ways and are motivated by different things

http://www.cs.unc.edu/~azuma/hitch4.html

Organizational Skills

- You will have lots of responsibilities (classes, GRA, GTA, dissertation research, publications)
- You must be *well-organized* and learn to prioritize to make sure the important things get done

Additional time management advice from Randy Pausch (former prof at UVa, CMU)
http://web.archive.org/web/20070223065627/www.cs.virginia.edu/helpnet/Time/time.html

http://www.cs.unc.edu/~azuma/hitch4.html

Communication Skills

- You will write (a lot)
- You will present your ideas (a lot)
- It will be so much better (and more efficient) for all of us if we spend more time talking about research ideas than about organization, grammar, and typos
- ***I cannot over-emphasize how important this is***
- This is so important that we'll have a whole set of slides/discussion a bit later in the semester

http://www.cs.unc.edu/~azuma/hitch4.html

Counseling Services

- Graduate school is full of highs and lows
- Don’t be afraid to ask for help if you need it
- ODU’s Office of Counseling Services provides comprehensive mental health services to enrolled students
  - https://www.odu.edu/counselingservices

Failure

- All research involves risk. If your project can't fail, it's development, not research.
- Failure just means that you had the courage to do something difficult
- People who seem to consistently succeed, in fact fail as often as anyone else.
  - They often have several projects going on all at once, only a few of which pan out.
  - The projects that do succeed have usually failed repeatedly, and many wrong approaches went into the final success.

How to do Research At the MIT AI Lab, https://dspace.mit.edu/handle/1721.1/41487

Research is Hard

- It is easy to burn out on it.
- Research always takes much, much longer than it seems it ought to.
  - you can't set a hard time limit on your PhD and expect to be successful
- If you are thinking about your research in background mode all the time, ideas will just pop out.
- Successful people are generally less *brilliant* than they are *persistent.*
- Explaining your work to others will help you keep in mind just how hard it is to understand what now seems trivial to you.

How to do Research At the MIT AI Lab, https://dspace.mit.edu/handle/1721.1/41487

Working with your advisor

- Computer science is often an apprenticeship model.
- You and your advisor plan out a plan, set of questions to explore
- You investigate and come back at the next meeting with answers and additional questions.
  - graduate student is one who writes programs, runs experiments, directly analyzes data
- If you bring experiment results, be prepared to explain any outliers or corner cases.
  - don’t present results that you haven’t thought about

Side Effect of Apprenticeship Model

- The success of a faculty member often depends upon the success of their students
- Thus, professors are highly motivated to get good students

All advisors vary

- Everyone does this relationship differently.
- We'll have some faculty come and talk about their research group styles.
- It's important that you learn early on what the expectations of your advisor are and get an idea of their working style.
  - ask your advisor directly
  - talk to senior students in the group
- One example: https://www.cs.ubc.ca/~tmm/policy.txt

Finding a Research Topic

- Generally students start on a small project that can lead into a larger body of work
- Sometimes new students assist senior students on their work.
- Ask questions. Be curious. Read papers.

Finding a Research Topic

- Good thesis topic expresses a personal vision
- Must be passionate about it - *nothing less will keep you going*
- Hardest part is figuring out how to cut your problem down to a solvable size while keeping it big enough to be interesting
- Future work sections of papers can be good sources for thesis topics

How to do Research At the MIT AI Lab, https://dspace.mit.edu/handle/1721.1/41487

Once you have a topic…

- You should be able to answer the question “What's the thesis of your thesis?"
  - what are you trying to show?

- You should have answers of varying length
  - one sentence
  - one paragraph
  - five minutes

How to do Research At the MIT AI Lab, https://dspace.mit.edu/handle/1721.1/41487

Importance of Lab Notebooks

- Either physical or online
  - WS-DL uses a shared wiki

- You will forget the details of code, experiment, or algorithm if you don’t write it down
  - don’t expect to just be able to read your code 6-9 months down the road when it’s time to write up your results for a paper

- Summarize interesting papers you’ve read

- Write down ideas as they come
  - while the information is clear in your head

Online Executable Notebooks

- Allows you to include text (explanations) along with runnable code
- Jupyter (Python, R, …)
  - requires a server for online interactivity, see https://mybinder.org
    - ex: https://hub.gke.mybinder.org/user/jupyterlab-jupyterlab-demo-adu5i3bi/lab
  - Python Overview - https://nbviewer.jupyter.org/github/phelps-sg/python-bigdata/blob/master/src/main/ipynb/intro-python.ipynb
- Observable (JavaScript)
  - no server needed, editable/executable in web browser
  - 5 min intro - https://observablehq.com/@observablehq/five-minute-introduction
  - force-directed layout - https://observablehq.com/@d3/force-directed-graph
  - bar chart race explained - https://observablehq.com/@d3/bar-chart-race-explained
- RMarkdown (R)
  - not editable online, produced via RStudio
  - intro example - http://www.math.mcgill.ca/yyang/regression/RMarkdown/example.html
  - analyze rainfall data - https://rpubs.com/JDONOGHUE16363611/565013

Research Tools

- GitHub
  - not only is it version control, but it's a place where you can show off your projects to potential employers
  - ***learn Git and use it***

- WS-DL has a group GitHub organization - https://github.com/oduwsdl
  - https://github.com/oduwsdl/MementoEmbed
  - https://github.com/oduwsdl/sumgram

Responsible Conduct of Research

- Required research ethics training for all graduate students at ODU
  - will be your HW1

- Ethics includes avoiding cheating and plagiarism
  - ODU CS page on Academic Integrity, https://graduate.cs.odu.edu/resources/academic-integrity/
  - "Cheating: The List Of Things I Never Want To Hear Again", https://www.cs.ubc.ca/~tmm/courses/cheat.html

OLD DOMINION UNIVERSITY

What's Grad School All About?

Dr. Michele C. Weigle

CS 800 - Research Methods

Week 1

January 14-16, 2020