

# Honors Capstone Proposal



**Student Name: Jonathan Thomas**

**Date: 1/22/18**

1. Please list the Focus Area courses that have been approved for this capstone (Note: if you have made changes to your focus area since you received approval, please resubmit the Honors Focus Area Proposal on CTools):

Course Number	Course Title	Credits	Term	Approved?
EECS 381	Object Oriented and Advanced Programming	4	W 17	Yes
EECS 485	Web Systems	4	W 18	Yes
EECS 489	Computer Networks	4	F 18	Yes

2. Project Title: (100 characters or fewer)

## **Web page comparisons using graph algorithms on post-load DOM trees and memory heaps**

3. Introduction: Provide a brief description of your project (appx. 250 words). The first paragraph should give an introduction to the topic/problem for non-experts, while the second paragraph should provide a concise overview of the project and its purpose/goal.

Code is always in a continuous state of improvement and being rewritten. As such, source code for web pages also changes drastically over time, and it is important to know what it really means for a web page to be the same as another or the same as a previous version of itself. In order to correctly compare if two web pages are the same after a change, two main entities related to the structure of a web page are important to consider. The first is the Document Object Model (DOM) tree, which keeps track of the structure of the web page at a high level. The second is the Javascript memory heap, which is used to create and destroy objects and functions as they are used and reclaimed throughout the course of visiting a site.

In this project, I will write a correctness checker to determine how similar two web pages are from the perspective of a user visiting either page, though these sites may have differences in their actual source code. In order to determine if the sites are the same, we will use graph-diffing algorithms to compare both the post-load DOM trees as well as the memory heap allocations to ensure that the same behavior is achieved. The end goal of this project to be able to deploy this tool on real-world web pages and be able to accurately retrieve information about similarities and differences between them.

4. Project question(s) to pursue: Describe the primary objectives and scope of the project, list your research question(s), or explain the problem to be addressed.

**The primary objective of this project is to create a program that can compare the resulting post-load DOM trees and memory heaps of loading two Chrome web pages and can quantify, with respect to some metric (to be defined), how similar two web pages are.**

5. Please describe the significance of this project: How does this project advance the field or contribute to the broader community (outside of UM, your lab, etc.)? Why is it important to the community, field, or organization? Who is your audience/target, and whose goals does this project serve?

**This project will serve to evaluate the correctness of existing work being done in the field of web page optimization research, where rewriting the source code of websites is often performed in order to gain improvements in speed and memory usage. This project will allow users to determine if and how their changes have affected the actual content of the rewritten web pages.**

6. Key Skills: What knowledge, skills, or experience do you need to have in order to successfully pursue this capstone? What knowledge and skills do you hope to gain through this capstone?

**In order to execute this project successfully, I will need to use the extent of both my mathematical and programming knowledge. This problem will require a rigorous understanding of graph comparison algorithms in order to efficiently and properly perform the necessary computations on the underlying graphs. I will also need programming skills, including a deep understanding of the Javascript execution model how objects and functions are represented, as well as more general knowledge of metaprogramming and other programming language theory concepts such as reflection. Finally, I hope to gain a better understanding of web performance research and comfort with using Google Chrome Developer tooling, Node.js and other related tools.**

7. Methodology of investigation: Describe your plan for pursuing the project—what are the professional or scholarly techniques, tools, and processes you will use to achieve the project's objectives? What are the major steps for this project (list as bullets)?

**In order to proceed with this project, I will utilize the expertise of Prof. Madhyastha and Mr. Goel to provide relevant reading and resources to turn to for help, as well as a high level overview of the project milestones. This project will require research into Javascript internals as well as the current state of web performance optimization research. Major steps for this project include:**

- **Gain working knowledge of the Javascript execution model and a deeper understanding of the problem.**
- **Figure out how to dump the associated data and parse this into a data structure that can be reasoned about effectively.**

- Develop a way to compare two such pieces of data, and how to determine if two results are effectively equal.
  - Work to expand the accuracy and efficiency of these comparisons using more advanced algorithms and more relevant criteria.
  - Package the developed tools and write documentation to help future users. Knowledge transfer to rest of the team in order to help with future development.
8. Work plan with timeline (attach additional documents as needed): Include a detailed plan for your work (minimum month-to-month), including how and how often you will interact with your capstone advisor for project mentorship. Please plan your work so your final poster will be complete a minimum of 1 week in advance of Design Expo for printing, and your final report will be complete 1 week following Design Expo.

**January:** Background research and reading, along with weekly meetings to get up to speed and make faster initial progress.

**February:** Start with preliminary implementation of diff checking data. Basic DOM comparison, starting with an end-to-end comparison with two plain HTML pages and working up to testing DOM comparison on real, more complicated web pages.

**March:** Work on the JS heap memory comparison. Start by comparing simple data types, and add in ways to compare complex data and functions with progress. Incorporate more advanced techniques and criteria to determine graph equivalence.

**April:** Finish up documentation and testing. Prepare poster as well as report for Design Expo. Present and tie up loose ends before ceasing development.

**I will meet with Mr. Goel on a weekly basis due to the short nature of this project. I will meet with both Prof. Madhyastha and Mr. Goel on a biweekly basis.**

9. Outcomes (be specific): Describe the intended outcomes or baseline goal for your work (be specific). What will you create, design, build, discover, learn, solve, or address through this capstone? What knowledge, recommendations, service, or device will be created through this project? (Note: this is different from how you will present your results—this question asks what you want to achieve as your result). Include a baseline goal, and a stretch goal. How will you assess whether your project is successful?

**As a result of this work, I will build a tool to compare the contents of two web pages. A baseline goal for this project is to be able to compare the post-load DOM trees and memory heaps of two web pages and correctly determine if the two are equivalent. A stretch goal would be to provide more sophisticated analysis of the similarities and differences between two pages to handle more complicated cases, such as when objects and nodes of the DOM refer to other nodes (turning the**

tree into a cycle). Additionally, another stretch goal is finding a way to detect changes in nodes that are primarily caused by changes from other nodes and adjust to account for these propagations.

10. Team Capstone Projects: If you are completing this capstone project in a group, this section should define your individual responsibilities and contributions to the project. What are your individual goals and objectives? What are you contributing that makes this your capstone project?

N/A.

11. Additional documentation attached (if any):

N/A.

12. Capstone Completion/final products:

- ☒ Present poster at Design Expo in (select one) April or December
  - ☒ Final report uploaded to CTools
  - ☒ Final report, poster, and additional artifacts uploaded to Honors Capstone Library on Seelio
    - Additional artifacts may include photos, videos, PowerPoint, etc.
  - ☐ Other:
-

**This proposal should be signed by the student and faculty supervisor for the project, then submitted to the Honors Program for approval.**

**Student and Faculty Supervisor agreement**

Please sign below to indicate you have read and agree to the following:

- We have discussed and agreed to this Honors capstone project proposal
- We agree that this project is commensurate with Honors-level work, as outlined in the attached letter to Honors Capstone Advisors
- We confirm that this project extends beyond the requirements of the senior design experience required by the student's engineering department
- If the student's capstone project changes, the student agrees to meet with the honors academic advisor to discuss next steps

Student Name: Jonathan Thomas

Faculty Name: Prof. Harsha Madhyastha

Student Email: [jonthoma@umich.edu](mailto:jonthoma@umich.edu)

Faculty Email: [harshavm@umich.edu](mailto:harshavm@umich.edu)

Student Department: EECS

Faculty Department: EECS

Student Signature:



Faculty Signature:



INTERNAL USE ONLY

Honors Faculty Advisory Board approval: