## *[Ninjabot]*

#### *[Derwin Lim Chin Wee]*

#### *[2427221L]*

## Proposal

### Motivation

*[Clearly motivate the purpose of your project; why someone would care about what you are doing]*

Web User Interface (UI) widget components offer the opportunity to do a specific task in the Web. However, the identification of techniques of the

. Therefore the project motivation is to attempt to back trace from the origin to attempt to identify widget components.

### Aims

*[Clearly state what the project is intended to do. This should be something which is measurable; it should be possible to tell if you succeeded]*

Design a Web robot (bot) to automatically scrape the Web page, so that Web UI components can be identified using Machine Learning techniques. This bot should be configurable to scrape pages that intentionally orchestrated content for specific user-agent in a discrete manner.

## Progress

*[Briefly state your progress so far, as a bulleted list]*

1. Research the purposes and intent for WCAG 1.0 and 2.0
2. Research Web UI Design patterns
3. Researching on JavaScript jQuery to inject into

## Problems and risks

### Problems

*[What problems have you had so far, that have held up the project?]*

1. Replicating Web pages as near identical as possible.
2. Attempts to bypass Browser Cross-Origin Resource Sharing (CORS) to replicate the scrape site as close as possible.
3. Understanding the Browser Architecture to inject JavaScript to redirect the required resource from website
4. Understanding the ECMAScript Architecture to retrace the possibility of Web Widget Component existing

### Risks

*[What problems do you foresee in the future and how will you mitigate them?]*

* Many widget component to identify. **Mitigation:** will narrow down to one to two widgets.
* Web page Sign-in authentication. **Mitigation:** Any form of Sign-in authentication will be avoided in this project.
* Unclear how to evaluate success of Widget Identification. **Mitigation**: will do background research to investigate how success of Widget Identification has been performed in the research literature review.
* Retracing of code and ensuring the accuracy of the Identified component. No clear mitigation available.

## Plan

*[Time plan, in roughly weekly to monthly blocks, up until submission week]*

Semester 2

* Week 1-2: Research tracing JavaScript to identify Web Widgets.
* Week 2-3: Integrate WAI-Aria requirements to project once Widget is identified.
* Week 3-5: Integrate tracing with Machine learning.
* Week 5-6: Merge the Integration with scraper.
* Week 6: Research on how to evaluate the system performance.
* Week 7-9: Machine Learning Training on Identification of Web Widget components
* Week 9: Evaluate test experiments run
* Week 8-10: Write up