

**CS6113**

**Translation Technology Systems**

**ADRIAN HOGAN – 18159761**

**ARCHIT KHANNA – 18155723**

**DAVID BROMELL – 18055001**

**Role David**

I contributed to this project by initiating the “Scrapy” framework, this is a web crawler which we used as a fast and efficient way to extract HTML content from various websites, and although this helped to speed up the HTML extraction process , we had to be careful about what websites we chose to use as we needed HTML code which was content rich. I also carried out the mass validation process using the Okapi Lynx command prompt, this aided in speeding up the validation process along with enabling us to quickly verify the content of a file. In addition to the above I also aided in the translation of files using SDL Trados. Lastly, I created and managed the GitHub for this project.

**Role Adrian**

I oversaw the Memsource for the project, which I was the project manager of, I sourced translation memories for the project’s initial translation within Memsource and contributed by allocating the 7 different XLF files to different jobs for the team, which allowed for easier contributions. I also used the Okapi Rainbow framework in order to extract XLF 2.0 content from our HTML files which were extracted from the “Scrapy” framework.

**Role Archit**

I contributed to the python code for Scrapy, finding appropriate websites to extract html from and assisted in carrying out the web crawling. I also worked on the translation of the xliff files in memsource by carrying out the analysis, pre-translation using the sourced translation memories and then reviewed the translations using the Microsoft machine translation tool that is built into Memsource. After verifying the translations, I also confirmed the various segments in the files. I then exported these files as bilingual mxliff first and then only xliff. I also assisted in the translation of the files in SDL Trados and post-validating the files in Okapi Lynx.

**Errors & Bugs**

1. TBX prefix not recognised, after examining Lynx specification it appears it supports the TBX prefix, however when validating same was not prevalent. Same of will be reported an recorded out BitBucket

