Text Technologies - Practical 1 - Web Crawler - Statistics

Ruaridh Thomson - s0786036

Total time to crawl (seconds):   38.371s

Average time per page (seconds):   0.047s

Unique URLs seen: 811 (including initial seed url)

Total visited and parsed: 807

Number of errors: 4 HTTP (404) errors

The crawler does not 'visit' the pages with 404 errors, hence: (total visited) = (number seen) - (number of errors).

Total number of links processed: 9852
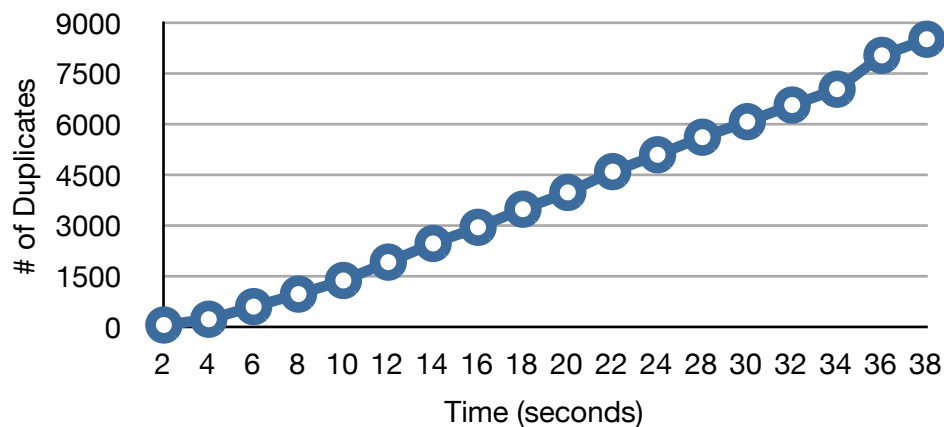
Average processed per page: 12.21

Number of links blocked by robots.txt: 510

Number of duplicate links found: 8518

Average duplicates per page: 10.56

Duplicates seen against time: