COMP6331 Computer Networks
Assignment 2 - Web Crawler
u6743886 Jinpei Chen

Report:

<table>
<tr><th>Item</th><th>What to report</th><th>Results</th></tr>
<tr><td>1</td><td>The total number of distinct URLs found on the site (including any errors and redirects)</td><td>46 (including off-site URLs)
43 (excluding off-site URLs)</td></tr>
<tr><td>2</td><td>The number of html pages</td><td>32</td></tr>
<tr><td>3</td><td>The number of non-html objects on the site (e.g. images)</td><td>9</td></tr>
<tr><td>4</td><td>The smallest html pages, and size</td><td>comp3310.ddns.net:7880/B/23.html,
size: 1443</td></tr>
<tr><td>5</td><td>The largest html pages, and size</td><td>comp3310.ddns.net:7880/C/307.html,
size: 7261</td></tr>
<tr><td>6</td><td>The oldest modified page, and its date/timestamps</td><td>comp3310.ddns.net:7880A/10.html,
modified time: Tue, 01 Jan 2019 05:05:00 GMT</td></tr>
<tr><td>7</td><td>The most-recently modified page, and its date/timestamps</td><td>comp3310.ddns.net:7880C/30.html,
modified time: Sun, 05 May 2019 01:01:00 GMT</td></tr>
<tr><td>8</td><td>A list of invalid URLs (not) found (404) (including off site invalid URLs)</td><td>comp3310.ddns.net:7880/F/10.html
comp3310.ddns.net:7880/F/20.html
comp3310.ddns.net:7880/F/30.html
comp3311.ddns.net:7880/B/207.html
comp3310.ddns.net/A/19.html
comp3310.ddns.net/B/29.html
comp3310.ddns.net/C/39.html</td></tr>
</table>

<table>
<tr>
<td>9</td>
<td>A list of on-site redirected URLs found (30x) and where they redirect to</td>
<td>
<table>
<tr><th>Source</th><th>Redirect to</th></tr>
<tr><td>comp3310.ddns.net:7880/C/3A.html</td><td>comp3310.ddns.net/A/19.html</td></tr>
<tr><td>comp3310.ddns.net:7880/A/1A.html</td><td>comp3310.ddns.net/B/29.html</td></tr>
<tr><td>comp3310.ddns.net:7880/B/2A.html</td><td>comp3310.ddns.net/C/39.html</td></tr>
</table>
</td>
</tr>
<tr>
<td>10</td>
<td>A list of off-site URLs found (either 30x redirects or html references), and whether those sites are valid</td>
<td>
<table>
<tr><th>Off-site URLs</th><th>Flag</th></tr>
<tr><td>comp3310.ddns.net/A/19.html</td><td>Invalid</td></tr>
<tr><td>comp3310.ddns.net/B/29.html</td><td>Invalid</td></tr>
<tr><td>comp3310.ddns.net/C/39.html</td><td>Invalid</td></tr>
<tr><td>comp3311.ddns.net:7880/B/207.html</td><td>Invalid</td></tr>
<tr><td>www.canberratimes.com.au</td><td>Valid</td></tr>
</table>
</td>
</tr>
</table>