

Program Log

```
1 C:\Python27\python.exe C:/Users/PrestonSSD2/PycharmProjects/WebCrawler/crawler.py
2 Crawling...
3   index.html ... indexed
4     SMU-CSE-LOGO.gif ... image
5     mailto:fmoore@lyle.smu.edu ... external
6     CSE5337_syllabus.pdf ... skipped
7     CSE7337_syllabus.pdf ... skipped
8     schedule.htm ... new
9     https://smu.instructure.com ... external
10    dontgohere/badfile1.html ... new
11    SMU-CSE-LOGO.gif ... image
12    dontgohere/badfile1.html ... restricted
13    schedule.htm ... indexed
14      index.htm ... new
15      index_duplicate.htm ... new
16      does_not_exist.htm ... broken
17      misc/count_letters.txt ... new
18      misc/count_letters_duplicate.txt ... new
19      misc/google-and-or-example.jpg ... image
20      misc/tokenizing_exercise.pdf ... skipped
21      misc/algorithm-1-7.pdf ... skipped
22      misc/jaccard_example.pdf ... skipped
23      misc/mercator-article.pdf ... skipped
24      misc/levenshtein.html ... new
25      http://www.gedpage.com/soundex.html ... external
26      misc/porter_stemmer_example.html ... new
27      misc/word-morphing-puzzle.pdf ... skipped
28      misc/urlexample1.htm ... new
29      this_aint_gonna_work.htm ... broken
30      misc/example-2017-03-02.xlsx ... excel
31      misc/examl.html ... new
32      misc/permuttermindex-example.jpg ... image
33      misc/excel-examples.html ... new
34      misc/activity-mar28.pdf ... skipped
35      misc/useragent.html ... new
36      misc/sim_example.xlsx ... excel
37      misc/roadmap-04-06-2017.jpg ... image
38      misc/text/index.php ... skipped
39      misc/bayes.html ... new
40      http://www.smu.edu/EnrollmentServices/Registrar/Enrollment/FinalExamSchedule/Spring2017 ...
external
41      misc/bayes.html ... indexed
42      misc/bayes.js ... skipped
43      misc/useragent.html ... indexed
44      misc/excel-examples.html ... indexed
45      misc/page113.xlsx ... excel
46      misc/excel-icon.jpg ... image
47      misc/page113-ex2.xlsx ... excel
48      misc/excel-icon.jpg ... image
49      misc/examl.html ... indexed
50      misc/urlexample1.htm ... indexed
51      ... skipped
52      index.htm ... new
53      index.htm ... skipped
54      ... skipped
55      http://lyle.smu.edu/~fmoore ... external
56      index.htm ... duplicate
57      misc/porter_stemmer_example.html ... indexed
58      http://tartarus.org/~martin/PorterStemmer/ ... external
59      http://en.wikipedia.org/wiki/Document_classification ... external
60      http://en.wikipedia.org/wiki/Stop_words ... external
61      http://en.wikipedia.org/wiki/Tf*idf ... external
62      http://lucene.apache.org/core/ ... external
63      http://search.carrot2.org/stable/search ... external
64      http://9ol.es/porter_js_demo.html ... external
65      misc/levenshtein.html ... indexed
66      misc/count_letters_duplicate.txt ... indexed
67      misc/count_letters.txt ... duplicate
68      index_duplicate.htm ... duplicate
69      index.htm ... already visited
70 Crawling finished
71
72 Process finished with exit code 0
73
```