

Homework #3  
406.424 Internet Applications  
Spring, 2010

Due Date: April 15, 2010

1. (40 points) Crawl your twitter homepage using Heritrix based on following conditions:

- Profile: basic\_seed\_sites
- Parameters
  - . TooManyHopsDecideRule > max-hops: 2
  - . TooManyPathSegmentsDecideRule > max-path-depth: 5
  - . Other parameters: default value

2. (30 points) Crawl your twitter homepage by changing the value of TooManyHopsDecideRule to 3 and using the same conditions for the other parameters given in Problem 1.

3. (30 points) Crawl your twitter homepage by changing the value of profile to 'broad\_but\_shallow' and using the same conditions for the other parameters given in Problem 1.

You're asked to submit the result report for all problems, including crawl-report, hosts-report and mimetype-report from Heritrix.

***Submit your answers in MS-Word or PDF format via email to the instructor cc TA.***