

bin/nutch freegen

Freegenerator is an alias for org.apache.nutch.tools.[FreeGenerator](#)

This tool generates fetchlists (segments to be fetched) from plain text files containing one URL per line. It's useful when arbitrary URL-s need to be fetched without adding them first to the crawldb, or during testing.

Usage:

```
bin/nutch freegen <inputDir> <segmentsDir> [-filter] [-normalize]
```

<inputDir>: This should be the path to the input directory containing one or more input (text) files. As with the Injector class, each text file should contain a list of URLs, one URL per line.

<segmentsDir>: The path to the desired output directory, where new segment will be created.

[-filter]: An argument to run current URLFilters on input URLs to improve the quality of the new segment(s).

[-normalize]: This argument should be passed to run URLNormalizers on input URLs prior to them being used in the process of creating new segments.

[CommandLineOptions](#)