08CommandLineOptions

Command Line Options of Nutch version 0.8.x

See each entry for datails of the command arguments and options.

command	function
bin/nutch crawl	One-step crawler for intranets.
bin/nutch readdb	Read / dump crawldb.
bin/nutch readlinkdb	Read / dump linkdb.
bin/nutch inject	Inject new urls into the crawldb.
bin/nutch generate	Generate new segments to fetch.
bin/nutch convdb	Converts a crawldb from pre 0.9 format.
bin/nutch fetch	Fetch a segment's pages.
bin/nutch parse	Parse contents in one segment.
bin/nutch segread	Read data in an existing segment.
bin/nutch updatedb	Updates the crawldb from a segment.
bin/nutch invertlinks	Create or update a linkdb from a segment or segments.
bin/nutch index	Run the indexer on a segment's fetcher output.
bin/nutch merge	Merge several segment indexes.
bin/nutch mergedb	Merge several crawldb-s together. Can be used for filtering out specific content.
bin/nutch mergelinkdb	Merge several linkdb-s together. Can be used for filtering out specific content.
bin/nutch mergesegs	Merge several input segments into one or more output segments. Can be used for filtering out specific content.
bin/nutch dedup	Deletes duplicate documents in a set of segment indexes.
bin/nutch plugin	Load a plugin and run one of its classes main().
bin/nutch server	Run a search server.

Other useful commands are also available

See each entry for datails of the command arguments and options.

command	function
bin/nutch org.apache.nutch.searcher.NutchBean	Commandline interface for doing searches.
bin/nutch org.apache.nutch.net.URLFilterChecker	Utility for testing url filters.
bin/nutch org.apache.nutch.indexer.HighFreqTerms	Lists the most frequent terms in an index.