

07CommandLineOptions

Command Line Options of bin/nutch

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

command	function
bin/nutch_admin	Web page and link database administration, including creation
bin/nutch_analyze	Adjust database link-analysis scoring
bin/nutch_crawl	Perform complete crawling and indexing of a set of root urls
bin/nutch_datanode	NDFS data node
bin/nutch_dedup	Deletes duplicate documents in a set of segment indexes
bin/nutch_fetch	Fetch a segment's pages
bin/nutch_fetchlist	Print the fetchlist of a segment
bin/nutch_generate	Generate new segments to fetch
bin/nutch_index	Run the indexer on a segment's fetcher output
bin/nutch_inject	Inject new urls into the web page and link database
bin/nutch_merge	Merge several segment indexes
bin/nutch_mergesegs	Merges multiple segments & removes duplicates
bin/nutch_namenode	NDFS name node
bin/nutch_ndfs	NDFS administrative access
bin/nutch_parse	Parse contents in one segment
bin/nutch_prune	Prunes existing Nutch indexes of unwanted content
bin/nutch_readdb	Read data from the web page and link db
bin/nutch_segread	Read data in an existing segment
bin/nutch_segslice	Divide data from one segment into several segments
bin/nutch_server	Run a search server of IPC connections
bin/nutch_updatedb	Updates the web page and link db from the segment fetcher output