hashdb Quick Reference

http://github.com/simsong/hashdb/wiki

General Usage

`hashdb <command> <options> <args>`	Run hashdb command, `-q` for quiet mode

New Database

`create [-p <hash block size>] [-m <maximum duplicates>] [-a <byte alignment>] [-t <hash truncation>] [<bloom settings>] <hashdb.hdb>`	Create a new hash database

Import/Export

`import [-r <repository name>] <hashdb.hdb> <dfxml.xml>`	Import from DFXML file into hash database
`import_tab [-r <repository name>] [-s <sector size>] <hashdb.hdb> <dfxml.xml>`	Import from tab file into hash database
`export <hashdb.hdb> <dfxml.xml>`	Export hash database to DFXML file

Database Manipulation

`add <A.hdb> <B.hdb>`	$A + B \rightarrow B$ add A into B
`add_multiple <A.hdb> <B.hdb> <C.hdb>`	$A + B \rightarrow C$ add A and B into C
`add_repository <A.hdb> <B.hdb> <repository name>`	$A_r + B \rightarrow B$ add when repository name matches
`intersect <A.hdb> <B.hdb> <C.hdb>`	$A \cap B \rightarrow C$ add when hash and source are common
`intersect_hash <A.hdb> <B.hdb> <C.hdb>`	$A \cap B \rightarrow C$ add when hashes are common
`subtract <A.hdb> <B.hdb> <C.hdb>`	$A - B \rightarrow C$ add when hash and source not common
`subtract_hash <A.hdb> <B.hdb> <C.hdb>`	$A - B \rightarrow C$ add when hashes are not common
`subtract_repository <A.hdb> <B.hdb> <repository name>`	$A_{\bar{r}} + B \rightarrow B$ add unless repository name matches
`deduplicate <A.hdb> <B.hdb>`	$A_d \rightarrow B$ add only hashes in A that are distinct

Scan Services

`scan <hashdb.hdb> <dfxml.xml>`	Scan DFXML file for matching hashes
`scan_hash <hashdb.hdb> <hash value>`	Scan for hash match
`scan_expanded [-m <number>] <hashdb.hdb> <dfxml.xml>`	Scan DFXML file for matches showing all sources
`scan_expanded_hash [-m <number>] <hashdb.hdb> <hash value>`	Scan for hash match showing all sources

Statistics

`size <hashdb.hdb>`	Print sizes of internal database tables
`sources <hashdb.hdb>`	Print source metadata
`histogram <hashdb.hdb>`	Print hash distribution
`duplicates <hashdb.hdb> <number>`	Print hashes sourced the given number of times
`hash_table <hashdb.hdb> <source_id>`	Print hashes associated with the source index
`expand_identified_blocks [-m <number>] <hashdb.hdb> <identified_blocks.txt>`	Expand to include source information for each source
`explain_identified_blocks [-m <number>] <hashdb.hdb> <identified_blocks.txt>`	Print information about less frequently observed hashes

Tuning

`rebuild_bloom [<bloom settings>] <hashdb.hdb>`	Rebuild Bloom filter

Performance Analysis

`add_random [-r <repository name>] <hashdb.hdb> <count>`	Add random hashes, log performance in `log.xml`
`scan_random <hashdb.hdb> <count>`	Scan random hashes, log performance in `log.xml`

bulk_extractor Scanner

`bulk_extractor -E hashdb -S hashdb_mode=import -o outdir1 my_media_image`	Import image
`bulk_extractor -E hashdb -S hashdb_mode=import -o outdir1 -R my_import_dir`	Import directory
`bulk_extractor -E hashdb -S hashdb_mode=scan -S hashdb_scan_path_or_socket= outdir1/hashdb.hdb -o outdir2 my_image2`	Scan image

March 23, 2015