

bin/nutch updatedb

Updatedb is an alias for org.apache.nutch.crawl.[CrawlDb](#)

This class takes the output of the fetcher fetcher and updates the crawldb accordingly.

Usage:

```
bin/nutch updatedb <crawldb> (-dir <segments> | <seg1> <seg2> ...) [-force] [-normalize] [-filter] [-noAdditions]
```

<crawldb>: This is the path to the crawldb directory we wish to update.

-dir <segments>: This should be the path to the parent directory containing all, if several, segments to update from.

<seg1> <seg2> ...: Here we would pass a comprehensive list of paths to individual segments to update from.

[-force]: This argument will force an update even if the crawldb appears to be locked. [blocked URL](#) : CAUTION: advised [blocked URL](#)

[-normalize]: This argument uses any current URLNormalizer's on urls in crawldb and segment (usually not needed).

[-filter]: Pass this argument to use any current URLFilters on urls in the crawldb and segment. This can provide better quality results in certain applications.

[-noAdditions]: If pass this parameter the updatedb command will only update already existing URLs, and will not add any newly discovered URLs during a fetch.

[CommandLineOptions](#)