UriState State Diagram

The URI is entered in the system (seed). The new URI is created in the global URI db, DNS resolved and replicated to the corresponding local cluster db.

IManageUri.CreateSeedUri()

The pased document was added to the index staging area and is queued for a new download schedule.

The downloaded document is being parsed into document tokens and URI-s.

Parsed

Parsing

Created

Create new URI-s

The URI should not be downloaded (HTTP 401, 404, policies, manually set, robots.txt, etc.)

The downloaded document has not changed (same hash) since the previous download.

Suspended

Scheduled

Downloading

URI-s in any state can be transitioned to Suspended status or removed from the system alltogether.

The URI is in the scheduler's queue or being processed at the moment: apply various crawler and target site policies determine the next time to download and refresh.

The URI has been activated, is in a download queue, or being downloaded at the moment

DownloadFailed