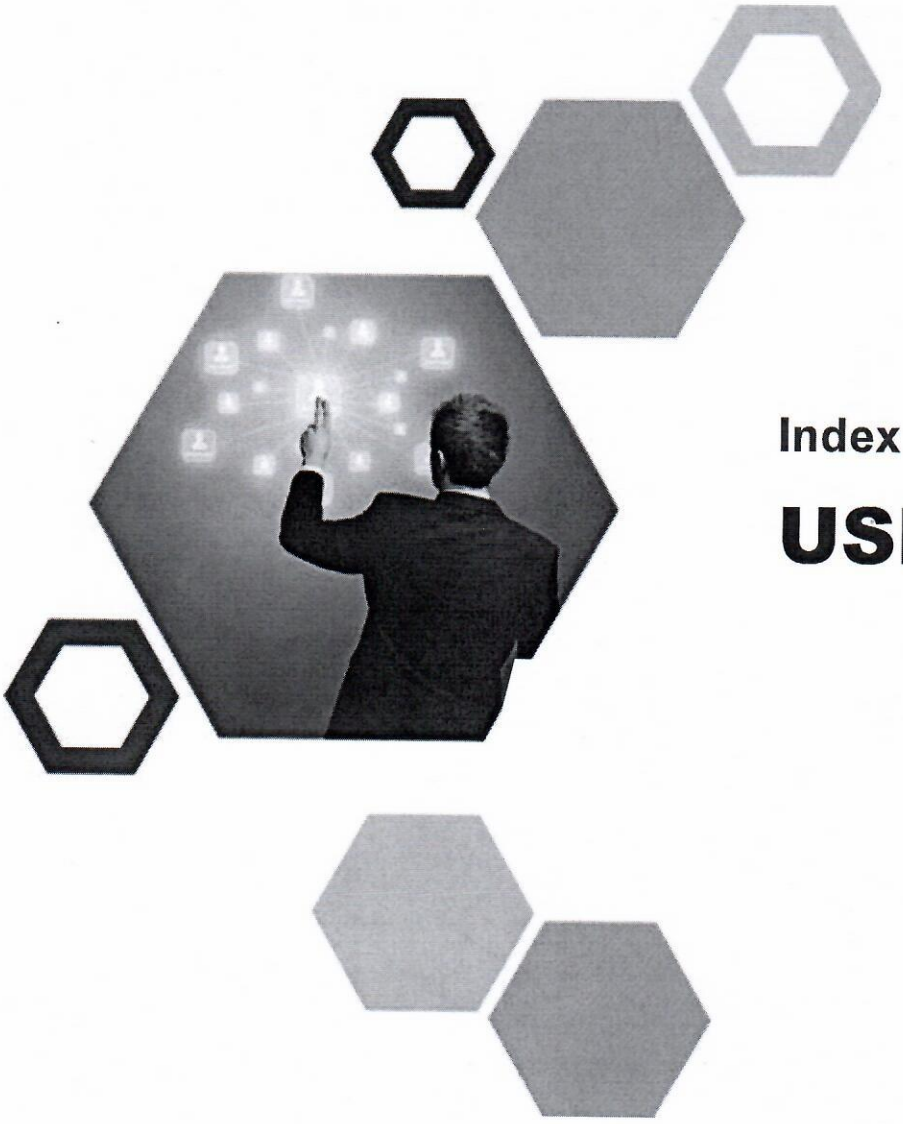
INDEX™ ENGINES
Power Over Information



Index Engines Software

USER GUIDE

Release 6.7.0

Document Version 100818

CyberSense

You can enable **CyberSense** data collection which collects the file information needed to calculate cyber analytics statistics. This option is disabled by default.

OCR

This is the default **OCR** (Optical Character Recognition) setting; you can maintain this default but change it on a per-job basis if you wish. The OCR setting only appears if you have an Index Engines OCR feature license. This "all-or-nothing" option affects all images when enabled or none at all when it is disabled. Indexing is slow when OCR is enabled since it processes all images encountered during indexing. However, see the note below:

> ✉ **NOTE:** The file server reindexing feature introduced in Release 6.6.0 lets you limit the objects processed with OCR enabled. To do this, index with OCR disabled (typically in a metadata-only mode). Then search for and select a subset of objects—those that you wish to reindex with OCR enabled—and select Reindex. Reindexing defaults to full content indexing mode. When reindexing in a federated environment, enable OCR on each engine. See the Search Guide for reindexing details.

File Filters

The indexing service uses various File Filters to process the internal contents of individual file types. To date, there are approximately 195 filters from which you can choose.

By default, all filters are enabled, i.e., files associated with the enabled filters are indexed. Disabling filters tends to increase ingestion performance while simultaneously reducing the size of index segments proportionate to the amount of indexable data associated with the disabled filters.

The *File Filters* page lets you enable the file types to be used for indexing.

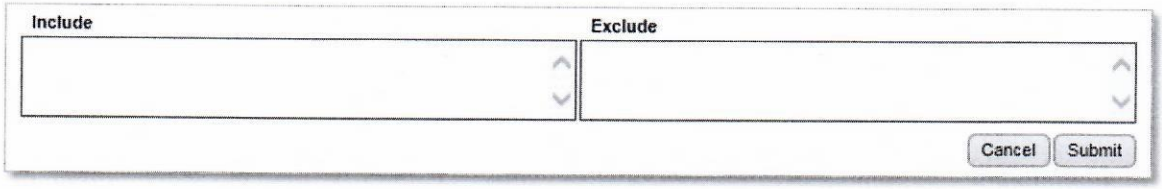
Navigate to **Administration->Home->Indexing Service->File Filters**:

CyberSense

You can run **CyberSense** analysis (assuming you enabled the *CyberSense* data collection option on the **Administration->Home->Indexing Service->Options** page). When you enable the *CyberSense* option on the Job Definition's *Advanced* menu, an email is sent containing a CSV with the statistics and a summary of any evidence of a ransomware attack. This option is disabled by default.

Include/Exclude Filters

This option lets you click inside the *Include* and *Exclude* fields to specify sub-folders and/or files in the directory that you wish to include or exclude from the indexing job. By default, all files and folders are otherwise included.



The filter format details described next can be used for NFS, CIFS, or SharePoint indexing jobs.

Filters for NDMP jobs (EMC, NetApp) have vendor-specific formats. In addition, only NetApp NDMP jobs support Include filters, while both NetApp and EMC NDMP support Exclude filters. For information on NDMP filter formats for NetApp and EMC jobs, please refer to these vendor-specific documents:

- NetApp - Data ONTAP Data Protection Tape Backup and Recovery Guide
- EMC - EMC Celerra Network Server Configuring NDMP Backups on Celerra

The Include and Exclude filters let you use wildcard patterns to specify the files and sub-folders for inclusion or exclusion during indexing. Note the following details when specifying include/exclude filters:

- If only **includes** are specified, every file is excluded except those items that match the include patterns.
- If only **excludes** are specified, every file is included except those items that match the exclude patterns.
- If **both includes and excludes** are supplied, then includes are applied first, and then the exclude patterns are applied to the included content.
- If **either** pattern matches a directory, it applies to all contents of the directory. It does not matter if the pattern ends with a path separator or not; a file or directory can match either way.
- Multiple patterns can be entered by placing each on its own line. The familiar wildcards are supported:

  * matches none or any number of characters (including path separators)