

Core - Bug # 56726

Status:	New	Priority:	Should have
Author:	Frans Saris	Category:	File Abstraction Layer (FAL)
Created:	2014-03-10	Assigned To:	
Updated:	2014-12-05	Due date:	
TYPO3 Version:	6.2		
PHP Version:			
Complexity:	easy		
Is Regression:	No		
Sprint Focus:			
Subject:	MetaDataExtraction isn't triggerd after file is uploaded		
Description			
<p>Currently the metaDataExtraction is only called through scheduler task. So when a editor uploads a new file he has to wait until the scheduler task is triggered again.</p> <p>It would be a great improvement if it is directly called for every new uploaded file.</p>			
Related issues:			
duplicates Core - Bug # 57408: Call of the meta extractor services for local ...		New	2014-03-28
duplicated by Core - Task # 57546: Call ExtractionService on new files and n...		Accepted	2014-04-02

History

#1 - 2014-03-10 15:13 - Steffen Ritter

- Status changed from New to Needs Feedback

metadata vs. indexing

metadata extraction always should be asynchronously because it could be very heavy.

#2 - 2014-03-10 15:16 - Frans Saris

- Category set to File Abstraction Layer (FAL)

I know it could be heavy but I guess for 1 file at a time it should be not a problem.

Heavy services could provide a own check in canProcess() that they only be executed in cli context etc.

#3 - 2014-03-11 11:05 - Steffen Ritter

Frans Saris wrote:

| Heavy services could provide a own check in canProcess() that they only be executed in cli context etc.

no - that's exactly why this "processing" has been detached form indexing process (despite it was in the old indexer)

#4 - 2014-07-10 14:55 - Alexander Opitz

Hi,

what's the state of this issue?

#5 - 2014-07-10 15:17 - Xavier Perseguers

It was done on purpose, so this should not be changed.

If you really want to index right away, EXT:[extractor](#) lets you do that.

#6 - 2014-07-13 20:23 - Fabien Udriot

+1 for metadata extraction upon upload. The actual situation is not satisfying, IMO -> Users do not want to wait until the next cron run.

If there is the fear to overload the system, a threshold (number of files on upload) could be added where to disable the metadata extraction. However, I believe on the majority of cases that won't be a problem.

#7 - 2014-07-14 09:38 - Xavier Perseguers

Just to be complete here, automatic metadata extraction is not only a problem of overloading the system but it slows down the upload itself a lot in case you are relying on binaries, such as tika (Java-based). Test for yourself, you'll see.

#8 - 2014-07-17 16:37 - Fabien Udriot

(By overloading the system, I meant slowing down the upload <-- just for the sake of clarity.)

By far not everyone has Tika deployed which is reserved to some advance set-up. Furthermore PHP based metadata extraction, is quite fast to my experience.

Could we make it as an opt-out option: by default indexing after upload which can be disabled by some configuration. This would be a compromise. Again, as a User it looks unsatisfying to have to wait for the next cron cycle to get the metadata.

#9 - 2014-07-17 20:12 - Ingo Renner

FWIW: Tika can also be run in server mode, which then saves the start up time of the JVM and the making it a lot faster. It's just that EXT:tika does not support server mode (yet).

#10 - 2014-07-29 21:58 - Frans Saris

Maybe we can add a checkbox to the storage settings to enable auto metadata extraction for that storage?

#11 - 2014-12-05 14:44 - Alexander Opitz

- Status changed from Needs Feedback to New