Erklärung der Vorgehensweise zur Optimierung der Performance

# Wie ist der grundsätzliche Ablauf des Prüfalgorithmus?

Am Beginn werden alle Pfade nach Dateien durchsucht. Dabei wird in einem *Dictionarry* die File Größe und der Pfad abgelegt, wobei die File Größe der Key ist. Haben zwei Files die gleiche Größe werden sie beide im gleichen Eintrag im *Dictionarry* abgelegt, getrennt durch ein Semikolon. Wurden alle Files erfasst werden jene mit nur einem Pfad entfernt. Danach können die Einträge mit gleicher Größe auch auf gleichen Inhalt geprüft werden. Dies geschieht in einer Parallelen foreach-Schleife. Dabei werden die jeweiligen Dateien mit gleicher Größe geöffnet und der MD5 Hash der Datei berechnet. Durch vergleichen der Hashes kann festgestellt werden ob auch der Inhalt gleich ist.

# Wann wurde wie entschieden wie viele Threads gestartet werden?

Die Anzahl der gestarteten Tasks kann entweder durch einen Übergabeparameter auf einen Maximalwert gesetzt werden. Ist dies nicht der Fall entscheidet die TPL über die Anzahl der Tasks und Threads.

# Warum wurde was wie gemacht?

Der Vergleich der Datei Größen wurde gewählt, da dadurch nicht die gesamten Files geöffnet werden müssen und nur die Meta-Daten der Dateien eingelesen werden. Somit kann sehr schnell eine grobe Filterung von nicht gleichen Dateien durchgeführt werden. Dies wird in einem Task durchgeführt, da alle Informationen in ein *Dictionarry* geschrieben werden müssen. Wäre das nicht der Fall müssten alle *Dictionarry* am Ende wieder Aufwendig miteinander verglichen und zusammengeführt werden. Und ein gemeinsamer Zugriff aller Tasks auf ein *Dictionarry* wäre auch nicht schneller.

Der Vergleich der gleich großen Dateien kann sehr gut in einer parallelen Schleife durchgeführt werden da keine Abhängigkeiten zwischen den einzelnen Dateien besteht. Um bei großen Dateien den Byte-Weisen Vergleich zu vermeiden wird ein Hash der Files zum Vergleichen verwendet. Dabei wurde das MD5 verfahren gewählt, da es eines der schnellsten Hash verfahren ist.