

Innerhalb von Dateien suchen

Sollen die Inhalte von Dateien aus den [Dateiarchiven](#) von der [Suchfunktion](#) durchsucht werden können, und hast Du ein Skript, welches den jeweiligen Dateiinhalt in einen Text extrahiert, dann kannst Du dieses Skript mit dem MIME-Typ verknüpfen und die Dateiinhalte werden sodann indexiert.

Sollen Dateien in den Dateiarchiven durchsucht werden, musst Du den Handler zum Extrahieren des Textes aus den Dateien angeben. Die Befehle, wie z.B. *strings* oder *pdftotext* müssen dabei auf Deinem Server installiert sein. Der jeweiligen Typ-Befehl muss im [Indexierungs-Tab](#) unter [Dateiarchive verwalten](#) definiert werden.

MIME Typ

application/vnd.oasis.opendocument.presentation

application/vnd.oasis.opendocument.spreadsheet

application/vnd.oasis.opendocument.text

Systembefehl

odt2txt %1

odt2txt %1

odt2txt %1

Ubuntu/Debian Paket mit Befehl

odt2txt

odt2txt

odt2txt

application/ms-excel

application/ms-powerpoint

application/msword

application/pdf

application/postscript

application/ps

application/rtf

xls2csv %1

catppt %1

catdoc %1

oder

strings %1

pstotext %1

oder

pdftotext %1 -

pstotext %1

pstotext %1

catdoc %1

catdoc

catdoc

catdoc

pstotext

pstotext

pstotext

catdoc

application/sgml

application/vnd.ms-excel

application/vnd.ms-powerpoint

application/x-msexcel

application/x-pdf

application/x-troff-man

text/enriched

col -b %1

oder

strings %1

xls2csv %1

catppt %1

xls2csv %1

pstotext %1

man -l %1

col -b %1

oder

strings %1

bsdmainutils

catdoc

catdoc

catdoc

pstotext

man-db

bsdmainutils

text/html

text/plain

text/richtext

text/sgml

elinks -dump -no-home %1

col -b %1

oder

strings %1

col -b %1

oder

strings %1

col -b %1

oder

strings %1

elinks

bsdmainutils

bsdmainutils

bsdmainutils

text/tab-separated-values

col -b %1

oder

strings %1

bsdmainutils

Zum Extrahieren können verschiedene Werkzeuge genutzt werden. Viele Unix Sites benutzen "strings", welches versucht Text in Dateien zu erkennen - allerdings nicht mit der Genauigkeit eines speziellen Werkzeuges für den MIME-Typ.

Stelle sicher, dass der Systembefehl seine Ausgabe auf dem Bildschirm (Standardausgabe) und nicht in eine Datei vornimmt. Probiere den Befehl vorher in einer Konsole und lies die Anleitung. Z.B. benötigst pdftotext ein nachfolgendes "-".

Ggf. muss der Tiki [Cache](#) nach der Installation eines neuen Handlers geleert werden, damit das System den Handler erkennt.

Es ist vorteilhaft, [Fileinfo](#) zu installieren, um falsch indentifizierte MIME-Typen zu vermeiden.

Verwandt:

- <http://stosberg.net/odt2txt/>