

Zadania ASU

mgr inż. Jerzy Sobczyk

dr inż. Adam Krzemienowski

5 października 2021

Projekt Porządkowanie plików

Celem projektu jest zapoznanie się z problemami występującymi przy przeszukiwaniu drzewa katalogów i porównywaniu plików. Jest to zadanie wspólne dla wszystkich uczestników kursu.

Założmy, że mamy spory zbiór dokumentów, nagrań, filmów i fotografii w katalogu **X** i jego podkatalogach. Co gorsza mamy również jedną lub więcej kopii w katalogach **Y1**, **Y2**, ... w których te same pliki są w innych miejscach drzewa lub pod innymi nazwami. Mogą się też zdarzać braki niektórych plików zarówno w katalogu **X** jak i w katalogach **Y1**, **Y2**, ... Celem jest uporządkowanie zbioru tak aby:

- w katalogu **X** znalazły się wszystkie pliki,
- zlikwidować duplikaty czyli spośród plików o identycznej zawartości zasugerować pozostawienie tylko najstarszego (nowsze daty są zapewne datami utworzenia kopii),
- zaproponować skasowanie plików pustych i tymczasowych,
- w przypadku plików o tej samej nazwie zasugerować pozostawienie nowszego,
- zasugerować ujednolicenie atrybutów np. `rw-r--r--`,
- zasugerować przemianowanie plików, których nazwy zawierają znaki mogące komplikować operowanie nimi (np. `':', '"', '.', ';', '*', '?', '$', '#', ',', '|', '\', ...`) zastępując te znaki zdefiniowanym substytutem np. `_`.

Skrypt powinien potrafić wyszukiwać:

- pliki o identycznej zawartości (choć niekoniecznie identycznej nazwie i niekoniecznie w analogicznej pozycji w drzewie),
- pliki puste,
- nowsze wersje plików o tej samej nazwie (choć niekoniecznie znajdujące się w analogicznym katalogu),

- pliki tymczasowe (*~, *.tmp, ew. inne rozszerzenia zdefiniowane przez użytkownika),
- pliki o nietypowych atrybutach np. rwxrwxrwx,
- pliki o nazwach zawierających kłopotliwe znaki,

Dla każdego znalezionej pliku skrypt powinien zaproponować odpowiednią akcję:

- przeniesienie (lub przekopiowanie) do odpowiedniego miejsca w katalogu **X**,
- skasowanie duplikatu, pliku pustego lub tymczasowego,
- zastąpienie starszej wersji nowszą,
- zastąpienie wersji nowszej starszą (w przypadku plików z identyczną zawartością),
- zmianę atrybutów,
- zmianę nazwy,
- pozostawienie bez zmian,

Skrypt nie musi wszystkich tych funkcji wykonywać jednocześnie. Jego użycie może wymagać kilkukrotnego uruchamiania, np. w jednym przebiegu wyszukujemy pliki puste, w innym duplikaty, w jeszcze innym odnajdujemy nowsze wersje, itd. Program musi być odporny na wszelkie znaki w nazwach plików i katalogów.

Parametry takie jak:

- sugerowana wartość atrybutów pliku np. rw-r-r-
- zbiór znaków kłopotliwych,
- znak będący substytutem znaków kłopotliwych,
- rozszerzenia plików uznawanych za tymczasowe,

należy wczytywać z pliku konfiguracyjnego np.: `$HOME/.clean_files`