

Podstawowy warsztat informatyka — lista 5

Zadanie 1 (1 punkt). Pod adresem `http://www.ii.uni.wroc.pl/~piotrek/pwi` znajdują się pliki, które będą potrzebne w kolejnych zadaniach. Ściągnij je wszystkie na dysk jednym wywołaniem programu `wget`. Sprawdź programem `file` czy pliki używają kodowania UTF-8. Jeśli nie, to załóż, że jest to WINDOWS-1250 lub ISO-8859-2 (nawet jeśli `file` nie rozpoznaje tego kodowania poprawnie). Następnie wykorzystaj program `iconv` do przekonwertowania wszystkich plików na UTF-8 (oczywiście oprócz tych, które już są w tym formacie).

Jesteśmy gotowi do scalenia wszystkich plików w jeden. Przemyśl, w jakiej kolejności należy poskładać te pliki, aby wpisy były chronologicznie uporządkowane. Scal te pliki używając podstawowych poleceń linuxowych (`cat`, `>>`, `...`). Plik wynikowy nazwij `dane`.

Zadanie 2 (1 punkt). Dane z pliku `dane` były zbierane w kolejnych latach różnymi sposobami, z których niektóre były lepsze, inne gorsze. Chcemy mieć pewność, że w tym pliku każda linia oznacza dokładnie jedną wypowiedź. Używając programu `grep` napisz polecenie, które usunie linie, które nie są w formacie wypowiedzi, tzn. nie są postaci „[(jakiś ciąg znaków)] nazwa użytkownika :”. *Rada: Nie zapisuj od razu do pliku, z którego czytasz, bo inaczej błędnym zapytaniem możesz zniszczyć swoje dane. Lepiej utwórz plik pomocniczy, a gdy uznasz, że jest wystarczająco dobry, użyj poleceń `rm` i `mv`.*

Zadanie 3 (1 punkt). Znajdź w pliku `dane` te wypowiedzi, które zawierają co najmniej trzy razy jakiś trzyliterowy ciąg znaków.

Zadanie 4 (1 punkt). Znajdź w pliku `dane` wszystkie adresy e-mail (nie wypisuj całych wypowiedzi, tylko te adresy).