

Systemy Operacyjne

Laboratoria nr 5

Zadanie z potokiem

Mikołaj Nowak

280082

1 Wstep

Należało wymyśleć skrypt używajacy potok w swoim działaniu. Skrypt, który stworzyłem wykorzystuje potoki by przetworzyć plik .epub na .txt.

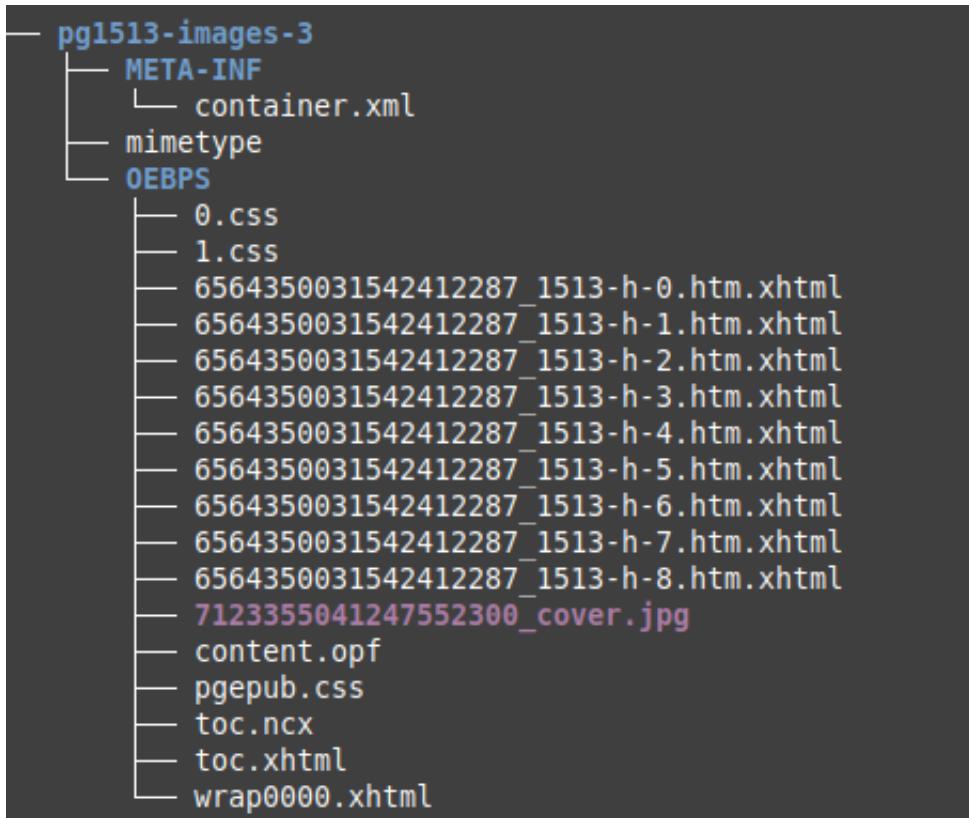
2 Kod

```
1 #!/bin/bash
2
3 #Sprawdzenie czy poprawna ilosc argumentow została przekazana
4 if [ "$#" -ne 1 ]; then
5     echo "Niepoprawna liczba elementów. Podaj: <plik> epub"
6     exit 1
7 fi
8 book="$1"
9 #sprawdzenie czy został podany plik o poprawnym formacie
10 if ! [[ $book == *.epub ]]; then
11     echo "Podany plik nie istnieje/z y format pliku"
12     exit 1
13 fi
14
15 nazwapliku=$(basename "$book" .epub)
16 unzipped_content="unzipped_$nazwapliku"
17 #rozpakowanie pliku do do katalogu
18 unzip -q "$book" -d "$unzipped_content"
19 #wyciagniecie z pliku container.xml folderu w którym bedzie znajdowac
# sie plik .opf oraz reszta potrzebnych folderow
20 pathstype=$(sed -n 's/.*full-path="\\"([^\"]*)\\"/\\".*\\\1/p'
"$unzipped_content/META-INF/container.xml")
21 extracted="$unzipped_content/$pathstype"
22 #Funkcja find wyszukuje w folderze z rozpakowanym .epub plik .opf po
# czym funkcja sed wyszukuje w tym pliku wszystkie zasoby lacznie z
# sciezками do nich, by nast pnie kazdy z nich został osobno
# przetworzony
23 find "$unzipped_content" -name "*.*.opf" -exec cat {} \; | sed -n
's/.*href="\"([^\"]*)\".*\\\1/p' | while read -r line; do
24 #sprawdzenie rozszerzenia pliku, odczytanego z pliku strukturalnego,
# by dodanie zawartosci do pliku .txt nie powodowało konfliktow
25 if [[ "$line" == *.html || "$line" == *.xhtml ]]; then
26 sed 's|<[^>]*>||g' "$extracted/$line" >> "$nazwapliku".txt #usuwa
# wszystkie tagi html'owskie i wpisuje przetworzon do pliku .txt
27 fi
28 done
29 #usuniecie rozpakowanego pliku
30 rm -r "$unzipped_content"
```

3 Test

Skrypt został przetestowany dla pliku .epub zawierajacego tekst Romeo i Julii. Plik został pobrany ze strony Projektu Gutenberg.

3.1 Plik .epub



Struktura pierwotnego pliku .epub

SCENE I. A public place.

Enter SAMPSON and GREGORY armed with swords and bucklers.

SAMPSON.

Gregory, on my word, we'll not carry coals.

GREGORY.

No, for then we should be colliers.

SAMPSON.

I mean, if we be in choler, we'll draw.

GREGORY.

Ay, while you live, draw your neck out o' the collar.

SAMPSON.

I strike quickly, being moved.

GREGORY.

But thou art not quickly moved to strike.

SAMPSON.

A dog of the house of Montague moves me.

GREGORY.

To move is to stir; and to be valiant is to stand: therefore, if thou art moved, thou runn'st away.

SAMPSON.

A dog of that house shall move me to stand.

I will take the wall of any man or maid of Montague's.

GREGORY.

That shows thee a weak slave, for the weakest goes to the wall.

Fragment pierwotnego pliku .epub

3.2 Plik .txt

ACT I

SCENE I. A public place.

Enter Sampson and
Gregory armed with swords and bucklers.

SAMPSON.

Gregory, on my word, we'll not carry coals.

GREGORY.

No, for then we should be colliers.

SAMPSON.

I mean, if we be in choler, we'll draw.

GREGORY.

Ay, while you live, draw your neck out o' the collar.

SAMPSON.

I strike quickly, being moved.

GREGORY.

But thou art not quickly moved to strike.

SAMPSON.

A dog of the house of Montague moves me.

GREGORY.

To move is to stir; and to be valiant is to stand:
therefore, if thou art moved, thou runn'st away.

SAMPSON.

A dog of that house shall move me to stand.

I will take the wall of any man or maid of Montague's.

GREGORY.

That shows thee a weak slave, for the weakest goes to the wall.

Fragment przetworzonego pliku .txt

4 Wnioski

Skrypt spełnia swoje zadanie, tworzy plik .txt z pliku .epub.

5 Bibliografia

- <http://jaroslaw.rudy.staff.iiar.pwr.wroc.pl/files/so/4-pipes.pdf>
- <https://www.gutenberg.org/ebooks/1513>
- <https://www.edrlab.org/open-standards/anatomy-of-an-epub-3-file/>