

Od książki do e-książki

Dlaczego nie kocham DTP

Piotr Bolek

EuroBachTeX 2011

Zadanie



W czym problem?

Co to jest książka?

W czym problem?

A co to jest e-książka?

W czym problem?

(e-)książka to:

Treść



Forma

W czym problem?

(e-)książka to:



W czym problem?

(e-)książka to:



Co to jest e-książka?

- Czy plik PDF jest e-książką?

Co to jest e-książka?

- Czy plik PDF jest e-książką?
- EPUB – standard e-książki (?) – czy tylko forma prezentacji i sposób dystrybucji treści (jakich?)

Co to jest e-książka?

- Czy plik PDF jest e-książką?
- EPUB – standard e-książki (?) – czy tylko forma prezentacji i sposób dystrybucji treści (jakich?)
- Jeśli PDF, to nie e-książka, to jak z PDF-a zrobić e-książkę?

EPUB

- Oficjalny standard IDPF (International Digital Publishing Forum) od 2007 r.
- Standard formatowania (XHTML + CSS)
- Standard opisu struktury publikacji w pliku .epub (opf, ncx)
- Fizyczny sposób zapisu w postaci paczki zip.

EPUB – przykładowa struktura

Nazwa	Rozmiar	Typ
[-] META-INF	1 element	Katalog
[-] container.xml	268 bajtów	Dokument XML
[-] OEBPS	16 elementów	Katalog
[-] img	3 elementy	Katalog
cover.jpg	40,5 KB	JPEG Image
line.png	159 bajtów	Obraz PNG
woblink.png	4,1 KB	Obraz PNG
style.css	4,9 KB	Arkusz stylów CSS
Aa epubfontsrff.ttf	142,6 KB	Czcionka TrueType
Aa epubfontsrfb.ttf	137,8 KB	Czcionka TrueType
Aa epubfontsrffi.ttf	140,8 KB	Czcionka TrueType
Aa epubfontsrffi.ttf	135,1 KB	Czcionka TrueType
bk01-toc.html	6,5 KB	Dokument HTML
ch01.html	1,7 KB	Dokument HTML
ch02.html	3,5 KB	Dokument HTML
ch03.html	8,1 KB	Dokument HTML
ch04.html	13,5 KB	Dokument HTML
cover.html	442 bajty	Dokument HTML
index.html	1,2 KB	Dokument HTML
redakcyjna.html	2,0 KB	Dokument HTML
toc.ncx	12,6 KB	Dokument HTML
content.opf	7,5 KB	typ application/x-wine-extension-opf
mimetype	20 bajtów	Zwykły dokument tekstowy

EPUB – słabości

- Dobry dla książek beletrystycznych (dużo tekstu, proste formatowanie i struktura)
- Niedoskonały dla książek fachowych (technicznych, naukowych) bądź bogato ilustrowanych
- Brak obsługi wzorów jako wzorów (wzory jako obrazki)
- Problemy z adresowaniem treści wewnątrz publikacji wynikające z technikaliów (paczka zip z opisem XML)
- DRM – problemy z przenośnością

(za wikipedią)

Od książki do e-książki

- Na wejściu: pliki z tradycyjnej ścieżki wydawniczej (PDF, Indd, Quark, Doc, ...)
- Po drodze: ???
- Na wyjściu: e-książka w formacie EPUB

Od książki do e-książki

- Na wejściu: pliki z tradycyjnej ścieżki wydawniczej (PDF, Indd, Quark, Doc, ...)
- Po drodze: **dużo problemów do rozwiązania**
- Na wyjściu: e-książka w formacie EPUB

Narzędzia i metody

- Jakich narzędzi używać? Calibre, Sigil, epubtools, Indesign – słabo...

Narzędzia i metody

- Jakich narzędzi używać? Calibre, Sigil, epubtools, Indesign – słabo...
- **Wyzwania:** wydobycie tekstu, ilustracji i elementów formatowania i struktury (ile się da) z materiału wejściowego

Narzędzia i metody

- Jakich narzędzi używać? Calibre, Sigil, epubtools, Indesign – słabo...
- **Wyzwania:** wydobycie tekstu, ilustracji i elementów formatowania i struktury (ile się da) z materiału wejściowego
- **Organizacja pracy:** konwersja, korekta

Narzędzia i metody

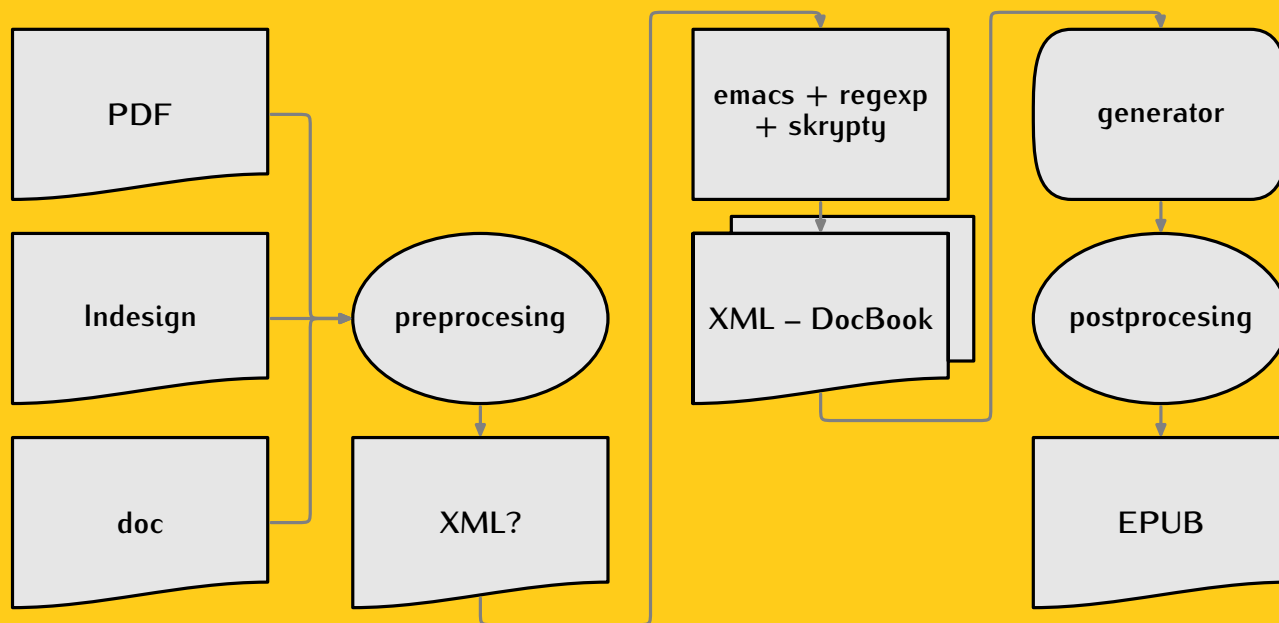
- Jakich narzędzi używać? Calibre, Sigil, epubtools, Indesign – słabo...
- **Wyzwania:** wydobyć tekst, ilustracji i elementów formatowania i struktury (ile się da) z materiału wejściowego
- **Organizacja pracy:** konwersja, korekta
- **A jednak XML!** – docbook + szablony XSLT → EPUB



```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE book PUBLIC "
<book lang="pl">
<part><title>CZĘŚĆ PIERWSZA
<chapter><title>1 <?break?> Rozmowa z
<para role="space">
```



Jak?



Preprocessing

- PDF
 - Wykrycie początków akapitów
 - Wykrycie i konwersja kodowania polskich znaków
 - „Dehyphenacja”
 - Usunięcie śmieci (paginy, informacje systemowe)
- INDD → EPUB
 - Konwersja tagów HTML na XML (docbook) i czyszczenie śmieci
 - Analiza stylów CSS → wyróżnienia w tekście (!)
 - Przetwarzanie przypisów

Postprocessing

- Poprawki błędów generatora
- Strona redakcyjna i tytułowa
- Poprawki w spisie treści
- Wersja demo (fragment)
- Cięcie tekstu na kawałki (jeśli trzeba)

Problemy – PDF

- Kodowanie znaków (!)
- Łamanie wyrazów, w szczególności na granicy stron
- Akapity i wcięcia
- Paginy i żywe paginy
- Kursywy, pogrubienia i zmiany fontów
- Kapitaliki i wersaliki
- Przypisy, ilustracje i tabele
- Wyróżnienia przez spacjowanie

Problemy – INDD

- Przypisy
- Wyróżnienia różnorakie
- Kolejność tekstu (wątki, marginalia, ramki, szpalty)
- Śmietnik po DTP

Przykłady z życia