

УДК 004.12.1

Є.М. Грабовський

Харківський національний економічний університет імені Семена Кузнеця, Харків

АВТОМАТИЗАЦІЯ ПРОЦЕСУ СТВОРЕННЯ ЕЛЕКТРОННИХ КНИГ У ФОРМАТІ EPUB

У статті наведено методику автоматизації процесу створення електронних книг у форматі epub. Розглянуто специфіку розробки прототипу електронної книги формату epub, верстання якої ідентичне оригінал-макету у форматі InDesign. Особливу увагу наділено механізмом усунення недоліків неякісної конвертації даних з одного формату в інший, а також загальним особливостям розробки алгоритму конвертації документів InDesign у формат epub і побудови параметричної моделі для палітри символів.

Ключові слова: електронна книга, автоматизація, формат.

Вступ

Постановка проблеми. Процес узгодження параметрів продукції між замовником і поліграфічним підприємством дуже трудомісткий, багатоаспектний, потребує, в силу своєї трудомісткості, наявності відповідного методичного забезпечення. В даний час виробництво друкованої продукції все більше інтегрується в єдиний простір медіа-індустрії. Складовою частиною цього процесу є інтеграція друкованих та електронних видань. Основною частиною більшості електронних видань, також як і книг, випущених типографським способом, є текстові фрагменти. Деякі видання взагалі цілком текстові. Якщо таке видання набрано на комп'ютері в будь-якому текстовому редакторі, то його можна назвати електронним виданням. У зв'язку з суттєвою кількістю книг у форматі Adobe InDesign «*.indd» і актуальністю використання електронних книг у сучасному житті необхідність автоматизації процесу конвертації з формату «*.indd» у формат epub «*.epub» надзвичайно актуальна і необхідна.

Аналіз останніх досліджень і публікацій. У результаті дослідження бібліографії з тематики статті [1-5] було виявлено, що у спеціалізованих літературних джерелах досить повно відображені основні положення теорії цифрових зображень, розглянуто специфіку створення електронних видань на основі використання HTML, XHTML і CSS. Але при цьому у літературі недостатньо висвітлені питання автоматизації процесу створення електронних книг у форматі epub.

Метою даної статті є аналіз специфіки автоматизації процесу створення електронних книг у форматі epub.

Виклад основного матеріалу дослідження

Конвертація даних є дуже необхідним нині засобом розповсюдження інформації, яка може бути представлена майже у будь-якому форматі. Процес

конвертації представляє собою перетворення даних з одного формату в інший. Зазвичай із збереженням основного логічно-структурного змісту інформації. Але у процесі конвертації з одного формату в інший існує ряд суттєвих недоліків. Конвертація може бути з втратою інформації або без втрати інформації.

Очевидно, що усунення недоліків неякісної конвертації з формату «*.indd» у формат «*.epub» можливо шляхом автоматизації процесу збору параметрів документу «*.indd» та формування CSS файлу, який є необхідним елементом у формуванні структури формату epub. Нажаль на сьогоднішній день не існує універсального інструментарію здатного виконати досконалий збір інформації з документу Adobe InDesign.

Аналізуючи причини даної ситуації, було встановлено, що складність створення подібного інструментарію, обумовлена слабкою структурованістю вже існуючої функції експорту документу «*.indd» у формат «*.epub». Недоліком такої конвертації є значна втрата інформації документу та втрата ідентичності верстки документу.

Саме тому головною метою даного проекту є отримання якісної електронної книги у форматі epub (з максимальною наближеністю до оригіналу) створеної на основі існуючого шаблону у видавничому програмному забезпеченні Adobe InDesign «*.indd».

У результаті досконального вивчення і дослідження властивостей видавничого програмного забезпечення Adobe InDesign було виявлено, що для більш коректного експорту документів InDesign у формат epub необхідно вдосконалення експорту параметрів документів, стилів символів, об'єктів, абзаців та таблиць. Саме ці стилі втрачають свої параметри при конвертації у найбільшій мірі. На основі цього факту необхідно приділити особливу увагу, при розробці прототипу саме цим елементам верстання та розробити алгоритм експорту документів з формату InDesign «*.indd» формат epub «*.epub».

Тож, розробка алгоритму конвертації документів InDesign у формат epub відбувається із фокусу-

ванням перш за все на такі параметри, як параметри документів, стилі символів, об'єктів, абзаців та таблиць. Виходячи з цього було розроблено параметричну модель для палітри символів.

Character (символ) призначений для управління атрибутами текстових символів, такими як гарнітура і розмір кегля. Форматування, що виконується за допомогою даної опції, застосовується до атрибутів окремих символів або набірних знаків абзацу.

Параметрична модель символу включає перелік таких параметрів, як: шрифт; капіталізація; стиль шрифту; розмір шрифту; інтерліньяж; мова; колір заливки; підкреслення; колір підкреслення; відтінок підкреслення; тип підкреслення; ширина підкреслення; закреслення; колір закреслення; відтінок закреслення; тип закреслення.

Такий перелік параметрів є достатнім для описання властивостей символу та можливим для реалізації засобами CSS.

Object (об'єкт) призначений для роботи з фреймами. Фрейми – це контейнери, куди поміщається графіка або текст. Фрейми можуть також використовуватися як графічні форми. Існує три типи фреймів, які ви можете створювати порожні, графічні і текстові.

Параметрична модель об'єкту включає перелік таких параметрів, як: стиль абзацу; параметри нижнього лівого кута; параметри радіусу нижнього лівого кута; параметри нижнього правого кута; параметри радіусу нижнього правого кута; параметри ефектів вмісту; параметри прозорості вмісту; колір заливки; відтінок заливки; параметри прозорості заливки; параметри підгону; праве закінчення лінії; ліве закінчення лінії; вирівнювання обведення; колір лінії; відтінок лінії; тип лінії; ширина лінії; параметри обтінання; параметри відступу; параметри верхнього лівого кута; параметри верхнього правого кута; параметри прозорості.

Такий перелік параметрів є достатнім для описання властивостей об'єкту та можливим для реалізації засобами CSS.

Параметрична модель параграфа включає перелік таких параметрів, як: шрифт; стиль шрифту; розмір шрифту; інтерліньяж; абзацний відступ; підкреслення; закреслення; трекінг; виключка; вирівнювання по базовим лініям; відступ з лівого краю; відступ з правого краю; відступ до абзацу; відступ після абзацу; напрям тексту; буквиця; кількість символів буквиці; кількість строк буквиці; колір заливки; відтінок заливки.

Такий перелік параметрів є достатнім для описання властивостей абзацу та можливим для реалізації засобами CSS [14].

Параметрична модель таблиці включає перелік таких параметрів: тип верхньої межі таблиці; колір верхньої межі таблиці; тип нижньої межі таблиці;

товщина нижньої межі таблиці; колір нижньої межі таблиці; тип лівої межі таблиці; товщина лівої межі таблиці; колір лівої межі таблиці; тип правої межі таблиці; товщина правої межі таблиці; колір правої межі таблиці; відступ до таблиці; відступ після таблиці.

Такий перелік параметрів є достатнім для описання властивостей таблиці із додаванням параметрів абзацу та можливим для реалізації засобами CSS.

Для розробки прототипу експорту необхідно розробити метод на основі якого буде реалізовано процес конвертації з формату InDesign «*.indd» формат ерив «*.erub». Важливо пам'ятати, що першим кроком роботи програми конвертації є збір вже існуючих стилів у документі InDesign, що у свою чергу можуть розділятися на підстили, згідно з приміною до тексту ієрархією.

При конвертації документа InDesign можна піти двома шляхами:

залишити структуру стилів в існуючому вигляді, і перенести їх у CSS, створивши відповідні стилям класи;

розробити власний метод обробки існуючих стилів, з метою оптимізації їх структури.

Для розробки прототипу був обраний другий варіант конвертації. При розробці методу структуризації стилів були враховані такі поняття:

- 1) узагальнення;
- 2) спадкування;
- 3) ієрархія;
- 4) деревовидна структура.

Принцип формування батьківського стилю приведений у табл. 1.

Таблиця 1

Принцип формування батьківського стилю

Стиль	Шрифт	Кегль	Підкреслення	Мова	Вирівнювання
1	Calibri	12	Ні	Російська	По ширині
2	Calibri	14	Так	Російська	По ширині
3	Calibri	14	Ні	Російська	По ширині
4	Cambria	16	Ні	Російська	По центру
5	Arial	14	Ні	Російська	По ширині
6	Arial	14	Ні	Російська	По ширині

Таким чином, перш за все необхідно виявити батьківський стиль «Батьківський стиль» буде формуватися по принципу найбільш повторюваних значень параметрів стилів, тобто той параметр, що зустрічається у більшості випадках автоматично стає батьківським стилем.

Таблиця 2

Результат формування батьківського стилю

Батьківський стиль			
Calibri	14	Hi	Російська

Тож виходячи з даних табл. 2 можна зробити висновок, що батьківський стиль формується по принципу вибору параметрів стилю, що зустрічаються частіше. Всі інші параметри стають дочірніми, тобто наслідують лише деякі параметри батьківського класу, а ті що залишилися є зміненими або взагалі новими (своїми власними).

Таким чином ми отримуємо деревовидну ієрархічну модель існуючих класів (рис. 1), у якій є батьківським стилем, а всі останні є його дочірні стилі. Дочірній стиль може брати лише кілька параметрів батьківського стилю, а все останні будуть його власними параметрами, або зміненими параметрами батьківського стилю.

Збір усіх стилів документу буде реалізовано методом зчитування параметрів з XML-файлів, які знаходяться у zip-контейнері формату IDML.

Для реалізації зчитування параметрів стилів формуємо метод `IdmlReadService` за допомогою мови програмування C#, у якому реалізувались зчитування кольору документу (`Color Process`), стилю оформлення документу (`Styles Process`), контенту (`Content Process`).

Формат IDML по своїй структурі подібний структурі формату електронних книг `epub`. Провівши аналогію, було виявлено що збір даних саме з формату IDML є найбільш коректним для його подальшого конвертування у формат `epub`.

Можна зробити висновок, що для більш коректного збору параметрів стилів документу формат IDML є найбільш зручним форматом, який можна



Рис. 1. Деревовидна модель стилів

використати як шаблон для формування CSS файлу. Сформувавши CSS файл, у якому ми зберігаємо усю необхідну інформацію про документ, ми переходимо до формування структури `epub` формату та автоматизації її створення.

При формуванні CSS файлу враховується також і те, що стилі що використовуються два рази і більше формується як класи, а стилі, що зустрічаються одноразово формуються як ідентифікатори.

Висновки

Таким чином в результаті роботи прототипу отримуємо документ `epub`, верстання якого ідентичне оригінал-макету у форматі `InDesign`. Для інтеграції розробленого модулю конвертації в веб-середовище був розроблений веб-сервіс, що дозволяє оптимізувати роботу онлайн системи. Науковий результат запропонованого рішення представлений у формі методики автоматизації процесу створення електронних книг у форматі `epub`. Практичним результатом є документ `epub`, верстання якого ідентичне оригінал-макету у форматі `InDesign`. Подальшим напрямком даного дослідження може стати розроблення методики контролю якості електронних книг у форматі `epub`.

Список літератури

1. Jeffrey Richter. CLR via C#. Third Edition / J. Richter. – L.: Ms Press, 2010. – 896 p.
2. Учебник CSS [Електронний ресурс]. – Режим доступу : <http://www.wisdomweb.ru/CSS/css-first.php>.
3. EPUB-книга [Електронний ресурс]. – Режим доступу : <http://epubbook.ru/html..>
4. InDesign / Об'єктна модель [Електронний ресурс]. – Режим доступу : http://pt.scribd.com/nacho_pr/d/86888837/24-Object-references.
5. HTML / Entities [Електронний ресурс] – Режим доступу : <http://www.htmlhelp.com/reference/html40/entities>.

Надійшла до редколегії 4.03.2016

Рецензент: д-р техн. наук, проф. Ф.В. Новіков, Харківський національний економічний університет імені Семена Кузнеця, Харків.

АВТОМАТИЗАЦИЯ ПРОЦЕССА СОЗДАНИЯ ЭЛЕКТРОННЫХ КНИГ В ФОРМАТЕ EPUB

Е.Н. Грабовский

В статье приведена методика автоматизации процесса создания электронных книг в формате `epub`. Рассмотрена специфика разработки прототипа электронной книги формата `epub`, верстка которой идентично оригинал-макету в формате `InDesign`. Особое внимание надлено механизму устранения недостатков некачественной конвертации данных из одного формата в другой, а также общим особенностям разработки алгоритма конвертации документов `InDesign` в формат `epub` и построения параметрической модели для палитры символов.

Ключевые слова: электронная книга, автоматизация, формат.

AUTOMATING THE PROCESS OF CREATING E-BOOKS IN THE EPUB FORMAT

Е.М. Hrabovskyi

In the article the method of automating the process of creating e-book format `epub`. The specificity develop a prototype e-book format `epub`, which is identical DTP layout format `InDesign`. Special attention endowed corrective mechanism defective converting data from one format to another, and the general features of the development of algorithms convert `InDesign` documents into a format `epub` and building parametric models for Character Palette.

Keywords: eBook, automation, format.