

電子出版における XSLT スタイルシートの部品化

戸澤 侑子[†] 金澤 典子 塚本 享治[†]

東京工科大学メディア学部メディア学科[†]

1 はじめに

電子出版においてデータ作成の電子化は大きく二種類に分けられる。一つは DTP を用いる形式、もう一つは組版ソフトを用いる形式である。しかし、これら既存の方式には二つの問題が存在する。一つはコンテンツとレイアウトの情報が混在しているため、コンテンツまたはレイアウト情報のみの再利用が困難なこと、もう一つは組版ルールをカスタマイズすることが困難といった問題である。そこで、再利用性の高い XML を使い、コンテンツとレイアウトの情報を分けて記述することで、第一の問題点を、組版ルールを記述した XSLT を部品化することで第二の問題点の解決を試みた。

2 想定する電子出版システム

本研究では、科学技術文書のための電子出版システムを作成した。ユーザーは表、図、グラフ、数式、索引、引用文献などの要素をもつ科学技術文書とレイアウトを指定するための設定ファイルを XML で記述する。XML 化された文書や設定ファイル、図はデータベースに蓄積される。XML ファイルは XSLT プロセッサや Java プログラムを通して、様々な組版ルールが適用され、XSL-FO [1] に変換される。その後、組版エンジンによって、PDF ファイルへと変換される。

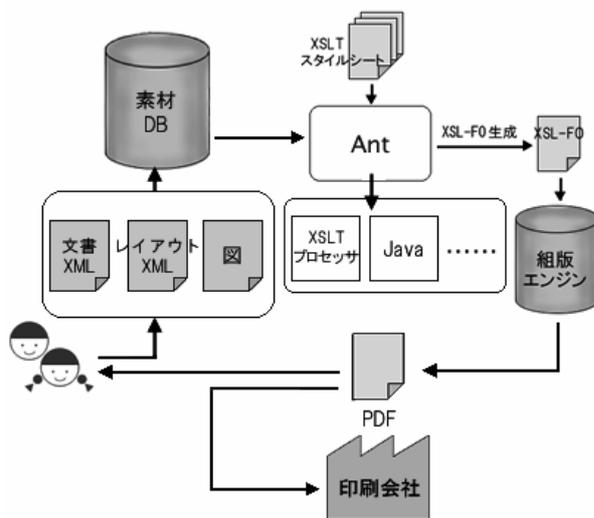


図 1: 想定する電子出版システム

3 組版ルールの部品化

3.1 部品化の必要性

出版物を作る際には、文字と図版類を組み合わせてページを作り上げていくことが必要となる。

出版物の元となる XML ファイルに対し、一つの XSL スタイルシートを適用するのでは、一つの出版物ごとに一つのスタイルシートが必要になってしまう。また、一つのスタイルシートに様々なルールが記述されている場合、カスタマイズすることが困難である。これらの問題を解決するために、組版ルールの部品化が必要となる。

3.2 部品化のコンセプト

まず、書籍がどのような構造であるかを分析した。書籍は書籍の情報といくつかのページに大きく分けられる。

それぞれのページは一つ以上の文章の固まりをもっている。その文章の固まりの中には、ヘッダ、段落、テキスト、外部ファイルデータ、箇条書き、脚注、文章のひと固まりが存在する。これらを整理したものが図 2 である。

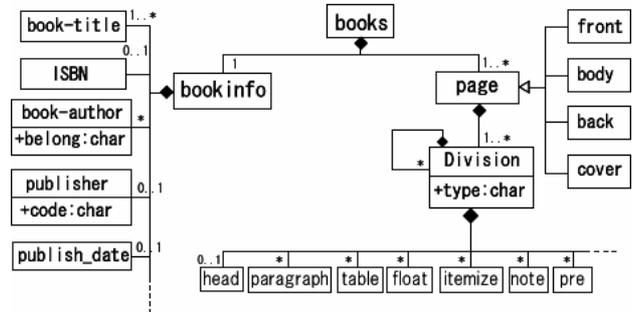


図 2: 書籍の構造

次に様々な適用される組版ルールをどのように部品化し、どのような関係性を持たせるかが重要である。

出版物には句読点の組み方や禁則処理、図に対するテキストの回りこみなどの様々なルールが必要である。必要なルールは基本的なものだけでも 20 個以上存在 [3] する。これらのルールを必要に応じてプラグイン化し、使用する。

出版物は複数の「台紙」から成り立っている。台紙にはマージンや版面が存在する。台紙には、とびら、中とびら、前書、目次、本文、あとがき、付録などが存在する。科学技術文書では、索引、引用文献なども必要となる。

台紙は複数の「構成要素」から成り立っている。科学技術文書はタイトル、サブタイトル、著者情報、見出し、テキスト、出版社名、出版社年月日、ノンブル、脚注、見出し、テキスト、キーワード、柱などから構成されて

いる。科学技術文書では図、表、グラフ、数式の要素も必要となる。これらを部品としてページを作成してゆく。

3.3 表・グラフ・数式の部品化

科学技術文書の中で使用される表、グラフ、数式も XML 形式で記述する。それらを出版物の中で使用する際は表や図に変換しなければならない。本文とは異なる変換が行われるため、異なるネームスペースを用い、区別する。

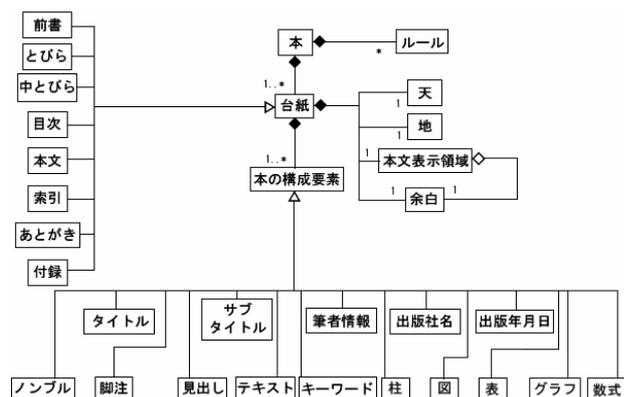


図 3: 部品の構造

4 テンプレートの部品化

4.1 組版ルール

構成要素の部品として、タイトル、サブタイトル、著者情報、見出し、テキスト、出版社名、出版社年月日、ノンブル、脚注、見出し、テキスト、キーワード、柱、図、表、グラフ、数式を、禁則処理の部品として、行末禁則、行頭禁則、行末調整を作成した。

部品はそれぞれ、継承関係、集約関係をもっている。それらの関係を表す方法として<xsl:include>を使用する方法と<xsl:import>を使用する方法がある。外部ファイルを読み込む点では二つとも変わりはないが、前者は外部ファイルを置き換えることを目的としている。後者は、ただ置き換えるのではなく、優先順位をもたせることができる。後者を使用して外部ファイルを読み込んだ場合、テンプレートは呼び出される側よりも呼び出す側の方が、前に記述された方よりも後に記述された方が、それぞれ優先順位は高くなる。優先順位の低いテンプレートを完全に打ち消すのではなく、現存するテンプレートの機能に「追加」したいという場合、<xsl:apply-template>要素と<xsl:apply-import>を使用することで、元のテンプレートの機能にほかのテンプレートの機能を入れ子で適用することが可能となる。本研究では、継承関係を表すには<xsl:import>を、集約関係を表すには<xsl:include>を使用した。

4.2 テンプレート

部品を使用する場合は、出版物の種類ごとに必要なル

ールを組み合わせ、テンプレートを作成する。テンプレートの形で管理することにより、カスタマイズを行いやすくなった。

例えばレポートであれば、台紙は本文のみでよい。台紙には、タイトル、サブタイトル、著者情報、見出し、テキスト、脚注、表、図、ノンブル、グラフ、数式の構成要素部品を使い、組み立てる。

5 変換システム

XML 化された科学技術文書は、本文、XML 化された表・グラフ、画像の参照先、SVG、MathML 形式で記述された数式から成り立っている。それらは、それぞれ XSL や Java を用い、変換が行われる。本文はナンバリングが行われ、XML 化された表・グラフ・数式からは SVG 形式の図が生成される。(図 4 参照)

この変換にはビルドツールである Ant を用いる。この際、変換元のファイルが新しい場合のみ、変換を行う。Ant の Uptodate タスクのプロパティ名を指定しておくことにより、これが可能になった。

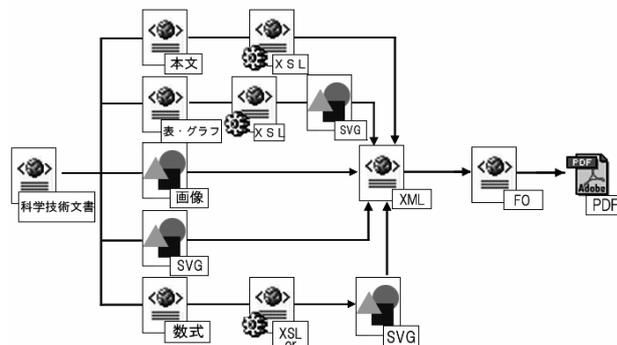


図 4: 変換プロセス

6 おわりに

本稿では、XSL で記述された組版ルールの部品化について提案し、部品化の方式について述べた。部品化を行うことにより、ある一部分に変更が生じた場合でも、その部分だけを変更し、変更がない部分を再利用することが可能となった。

今後の課題は、電子出版技術として必要なトンボや画像の CMYK 変換、フォントの組み込みなどである。

参考文献

- [1] Dave Pawson, “XSL-FO”, O'Reilly & Associates (2002)
- [2] “PREMedia Part16”, 印刷出版研究所, (2001)
- [3] “文字の組み方ルールブック ヨコ組編”, 日本エディタースクール, 2002
- [4] 塚本、花園、菊池、畑、鎌田, XML を用いた技術文書の電子出版, 情報処理学会第 64 回 (平成 14 年) 全国大会講演論文集(3)[3Y-03], 2001