

XSLT 化した既存のテンプレート文書を用いた複合文書の生成 Generation of Compound Document from Existing Document using XSLT

西原 裕則[†] 塚本 享治[‡]
Hironori Saibara Michiharu Tsukamoto

1. はじめに

複合文書はテキストだけでなく図表やグラフを含んだデジタル文書である。Microsoft Word や Microsoft Office Word など近年の多くのソフトは複合文書を扱う。

決められた形式で、何度も作成する必要があるデータにはテンプレートがあれば便利である。情報処理学会でも論文テンプレートが公開されている。筆者らの大学にも卒業論文のテンプレートが用意されているが、提出されたものを見ると形式が崩れていることが多い。

そこで TeX を見習い、テキストエディタを用いて論文を XML で記述し XSLT 変換を行い、Word ドキュメントを作り上げる。

2. 複合文書作成のアプローチ

筆者らの大学の論文テンプレートを使い、同じ内容の Word ドキュメントに変換していく。

まず論文テンプレートを圧縮解凍し、パッケージ化する。次に論文テンプレートの内容を記述した XML データを用意して、Word パッケージに合う形へ XSLT 変換を行う。変換された XML データをパッケージの中へ入れて再び圧縮し、Word ファイルを作成する。

論文テンプレートは表紙、概要、目次など記述する内容が決められている箇所と、本文のように決められていない箇所がある。XSLT を記述する際は、表紙、概要には論文テンプレートの XML 骨格をそのまま使用する。本文も骨格を使用するが、フォントスタイルなどの骨格のみを使用する。

この手順のうち、括弧内は手作業で行なっていく、圧縮解凍、変換、圧縮の流れはスクリプトで行う。

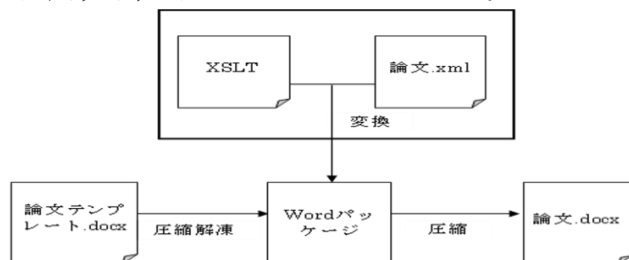


図1 変換の手順

3. 変換方法

3.1 XML 形式論文の用意

論文テンプレートの内容を、論文タイトルや著者情報、本文に分けて XML で記述していく。章、節、項、図、表、グラフは通常の論文と同様に出現する順に記述する。主に文

[†] 東京工科大学 大学院 バイオ・情報メディア研究科
Tokyo University of Technology Graduate School

章は<段落>で囲み記述していくが、見出しなどは別の言葉で囲む。表を記述する際は HTML のテーブルタグと同様に記述していく。

3.2 Word パッケージ構造

Word データを圧縮解凍すると、複数のデータが入ったフォルダが作成される。これをパッケージと言う。Word フォルダ内の document.xml がメインドキュメントであり、文章やフォントスタイルを指定するデータである。XSLT 変換を行った XML 形式論文は、この document.xml を上書きして、再び docx 形式に圧縮する。

3.3 XSLT 記述

論文テンプレートの XML 骨格をそのまま使用する表紙、概要、目次と、フォントスタイルなど一部 XML 骨格を使用する本文に分けて記述していく。

表紙、概要、目次は XML 骨格をそのまま使用するが、人の手で情報を入力する箇所は論文 XML から引用する形で記述する。

本文の章、節、項、段落などは何度も出現するため、別々に XSLT を記述していく。段落は何行にも及ぶことがあるため章節項内で繰り返すように記述する。

```

<!-- 段落の記述 -->
<xsl:template match="段落">
  <w:p>
    <w:pPr>
      <w:pStyle w:val="a2"/>
    </w:pPr>
    <w:r>
      <w:rPr>
        <w:rFonts w:hint="eastAsia"/>
      </w:rPr>
      <w:t>
        <xsl:value-of select="."/>
      </w:t>
    </w:r>
  </w:p>
</xsl:template>
<!-- 章の記述 -->
<xsl:template match="章">
  <w:pw:rsidR="00BA4A8B"
  w:rsidRDefault="00BA4A8B" w:rsidP="00747387">

```

図2 XSLT の記述例

4. 変換の詳細

スクリプトによって変換された論文 XML は図3のようになる。ここでは XSLT で記述した論文に必要な要素の変換の詳細をまとめる。

● 2.3 作成環境

作成環境で利用する Operating System について全学生に調査した結果を表 21 に示す。表中において、Operating System とその Operating System を用いると回答した人数を示す。なお、この調査では複数回答可であるので、合計人数は学生数よりも多い。

表 卒業論文作成環境 (Operating System)

OS	人数
Windows 98/SE/ME	150

次に、作成環境で用いる文書作成支援ソフトウェアについて調査した結果を図 23 に示す。

これらの調査結果から、圧倒的に Windows 上で Word を用いることが多いことがわかる。Operating System による相違は文書作成そのものには大きな影響は与えないので、ここでは Windows XP を前提とする。

また、文書作成支援ソフトウェアについても、Word のバージョンの相違がある。これは機能やダイアログの形式などに違いがあることを意味している。しかし、個々で用いる機能の多くは本質的にはバージョンの相違の影響を受けにくいと考えられることから、個々では最新版の Word (Office XP) を前提とする。

図 3 変換された論文の一部

● 表紙, 概要, 目次

表紙, 概要には論文タイトルや著者情報が入るため、論文 XML では<論文タイトル><氏名>の形で記述し、変換を行った。概要の文章は<段落>で記述するが上に<概要>を設け、差別化する。目次は章, 節, 項などの見出しが入るため<章>内の<タイトル>を引用する形で変換する。

表紙, 概要は形式が崩れることがなく変換することができた。目次も見出しを変換することができた。

論文タイトルが長い場合、タイトルの途中で改行を行い調整することができない。そのため変換後に手動で改行を行うか、論文 XML の段階で改行した記述をする必要がある。目次ページは該当する見出しのリンクやページ番号の取得が行なえていないため、今後取得方法を模索していく。

● 章, 節, 項, 段落

章, 節, 項にそれぞれフォントスタイルが決められているため、それに合わせる形で変換を行った。なお章は新しいページから始まるようにする。

全て問題なく変換することができ、フォントスタイルも合わせることができた。段落も全行変換することができた。

変換後、見出しや段落がページの切れ目直前に入ることがある。そのため改行などの調整を手動で行う必要があるが、今後任意の場所で改行を行えるような記述方法を模索していく。

● テキストボックス

テキストボックス内に文章を表示するように変換を行った。この文章は<段落>で記述している。<テキストボックス>内に入り、第 1 章のあとに表示されるようにする。

変換を行ったところ意図した場所にテキストボックスが作成され、文章の表示も成功した。

今後はテキストボックスサイズの変更や細かい位置調整の記述方法を調べていく。

● リスト, 項目

リストの書式は①, ● など複数あるため、XSLT の記述も複数用意して変換を行った。なおリスト内に表示する文章は<項目>として記述する。

変換には成功し、複数の種類のリストを作成することができた。

リストを作成することはできたが、意図した書式にならず、別の書式として変換された。Word によって割り振られている番号が異なっていると考え、同じ論文テンプレートでリストを作成し、解析を行ったところ違いがあったため、今後改善していく。

● 図表グラフ数式

表は XSLT 変換が可能であったが、図, グラフ, 数式に関しては変換が難しい。

Word データのパッケージは複数のデータで成り立っているが、データごとにリレーションシップ ID を割り振られている。リレーションシップ ID は document.xml.rels で管理している。割り振られたリレーションシップ ID は、document.xml などメインドキュメントで ID を呼び出す形で使用する。

図や数式は画像データを media フォルダ内に入れ、document.xml.rels にリレーションシップ ID を追加することでパーツとして使用可能となる。Excel で作成されたグラフは、charts フォルダ内の chart1.xml に変換されてリレーションシップ ID に登録される。

このため画像やグラフのデータを任意のフォルダ内へ移動し、リレーションシップ ID を登録し、ID を間違いなく記述する必要がある。また論文 XML を作成する際にも記述方法を考えなければならない。

● 参考文献

参考文献は表紙, 概要と同じように記述する内容が決まっているため、同じように XML 記述と変換を行う。

必要な情報を変換することができたが、本文中の文献の引用について今後考えていく必要がある。

5. おわりに

本稿では論文テンプレートを XSLT 化し、論文テンプレートと同じ内容の Word ドキュメントを作成することを目的とした。

ほとんどが自動的に変換できたが、文章構成のバランスなど細かい修正は手動で行なっていく必要がある。また図, グラフ, 数式などは複雑な作業が求められるため、今後変換方法を模索していく。

なお本論文はこの報告内容を踏まえて、スクリプトで変換して作成したものである。画像やリストの修正は、Word ファイル形式で出力するため手動で行った。

参考資料

- [1] Michael Kay 著, 佐藤 直生 監修, IDEA・C 訳, “XSLT バイブル”, (2002)
- [2] アンテナハウス株式会社 XSL Formatter グループ, “Office Open XML Formats 入門”, (2007)
- [3] “ISO/IEC 29500-1:2011”
http://www.iso.org/iso/catalogue_detail.htm?csnumber=59575