

プロファイリングによる共通テキストの効率的な作成

内海 淳

Email: utsumi@hirosaki-u.ac.jp

弘前大学人文学部

◎Key Words 共通テキスト, XML, プロファイリング

1. はじめに

授業において複数の教員が共通のテキストを作成して利用する場合がある。個々の授業担当の教員は、各クラスの受講者のレベルや所属などに応じて、その共通テキストの一部を取捨選択して授業を行うことになる。しかし、科目によってはカバーすべき範囲が広い場合や、学習者のレベルが多様である場合がある。そのため、テキストの分量が膨大になり、その構成が複雑になってしまい、テキストの作成や修正や変更などの維持管理が難しくなってしまう。それだけでなく、そのテキストを使用する学習者にとっても、自分が必要とする部分を見つけ出しにくくなり、効率的な学習を妨げる場合がある。

本発表では、DocBook や DITA などの XML 文書作成規格におけるプロファイリングの手法を用いる事により、こういった大規模で複雑な共通テキスト作成や利用を効率的に行うことができることを示す。具体的な事例として、大学における、情報および英語の科目の共通テキスト作成のケースを取り上げる。

2. XML 文書作成規格とは

2.1 DocBook

本発表で取り上げる XML 文書作成規格とは XML (eXtensible Markup Language) に基づいて定義された文書作成のためのマークアップ言語を指す。XML 文書作成規格には、DocBook や DITA (Darwin Information Typing Architecture) などがあるが、本発表では DocBook を中心に取り上げる。

DocBook は、1991 年から Norman Walsh らによって開発された XML 文書作成規格であり、コンピュータのハードウェアやソフトウェアのマニュアルといった技術文書の作成のために開発されたものである。しかし、最近では、技術系の文書に限らず、より広範囲の文書の作成にも利用されてきている。現在はバージョン 5 となる DocBook5 が公開されており、Walsh (2010)⁽¹⁾ に DocBook5 の詳細な仕様が解説されている。また、DocBook から XSLT を使って様々な表示形式への変換に関して詳しく解説しているものとして、Stayton (2007)⁽²⁾ がある。DocBook は OASIS (Organization for the Advancement of Structured Information Standard) の下で管理されており、誰でも自由に無償で利用することができる。

2.2 タグと属性

DocBook は html などの他の XML 文書と同様に、タグを使って文書を記述する。章、節、パラグラフなどの構成要素をタグでくくって次のように記述する。

```
<para>大学の教育システムは、大きく分けて、
<emphasis role="bold"> 専門教育 </emphasis> と
<emphasis role="bold"> 教養教育 </emphasis> から成り
立っています。</para>
```

上の例の para および emphasis が構成要素名であり、emphasis の開始タグ内の role="bold" の部分は、その emphasis 部分の属性 (attribute) を指定している。上記の例では、強調要素 emphasis の強調の程度としてボールドの値が設定されている。

2.3 プロファイリング

上述の属性の特定の値を出力に際に指定することによってターゲットごとに異なる出力結果を生成することが可能になる。例えば、全ての要素を日本語と英語で書き、日本語の部分には属性 lang の値を"ja"、英語の部分には"en"の値を指定しておき、出力の際に lang="ja" として指定すれば日本語のみの文書、lang="en" と指定すれば英語のみの文書を生成することができる。このような仕組みをプロファイリング (profiling) による出力と呼ぶ。

3. 「情報」の場合

3.1 多様なプラットフォームへの対応

弘前大学では、教養科目の「情報」の授業のため、複数の教員が作成した共通テキスト⁽³⁾ を使用している。このテキストの主な役割は、弘前大学内で利用できるコンピュータシステムおよびオフィススイートなどのアプリケーションソフトの基本的な使用方法の解説である。この「情報」のテキストは LibreOffice などのオープンソースのソフトウェアを主に解説している。

しかし、弘前大学内には、Windows, Mac OS X, Linux の三種の OS の教育用コンピュータが設置されている。これらのコンピュータ上にはオープンソースの LibreOffice だけでなく MS Office も搭載されているため、授業では MS Office の使用方法にも対応しなければならない。そのため、MS Office などの使用方法については、授業担当の教員が共通テキストを拡張・編集した教材

を作成し使用している。

このような状況を改善するためには、授業担当教員が使用する OS 及びオフィススイートの組み合わせを全てカバーする必要がある。そのためには、少なくとも 8 通りの場合に合わせたテキストを用意しなければならない。しかし、それを 1 冊の書籍としてまとめた場合、分量が膨大になり、その構成が複雑になってしまうため、学習者、特に「情報」の科目が想定する初学者には利用しにくいものになる。

3.2 プラットホームごとの出力

DocBook は、元々、コンピュータなどのマニュアル作成のために開発された規格であるため、os などコンピュータに関連した属性が用意されている。上述の問題は、属性 os の値として、linux, macosx, windows を設定し、属性 condition の値として libreoffice と msoffice を設定してテキストを記述すれば解決できる。例えば、LibreOffice と MS Office の違いは次のように記述する。

```
<para os="macosx;windows" >
  <phrase condition=" libreoffice">Calc</phrase>
  <phrase condition=" msoffice">Excel</phrase>
  のデフォルトのファイル形式は
  <phrase condition=" libreoffice">
    “~.ods” </phrase>
  <phrase condition=" msoffice” “~.xlsx” </phrase>
  という拡張子…</para>
```

ここで示した os と condition の値を組み合わせて、それぞれのプラットフォームごとに出力させることにより、8 通りのテキストを生成することができる。

4. 「英語」の場合

4.1 多様な学習者への対応

弘前大学では、英語の授業内容が教員ごとに異なり、評価基準もまちまちであることから、共通テキストの作成・導入が検討されている。しかし、英語の場合、学習者の習熟度の差が大きく、また学部・学科等の求める語彙レベルなども異なっている。このような中で共通テキストを作成する場合には、まず、習熟度のレベルごとに基本的な学習内容を決定しなければならない。その上で、学部・学科ごとに必要とされている語彙レベルの違いは、例文等を学習者の所属などに応じて使い分けることなどにより対応することが考えられる。

このような内容を 1 冊の書籍としてまとめた場合は、上述の情報の場合と同様で、分量が膨大になる上、個々の学習者に自分の必要する部分がわかりにくくなってしまふ。

4.2 学習者に最適化された出力

DocBook には学習者の習熟度に対応した userlevel という属性が用意されていて、その初期値として、beginner と advanced が設定されている。この

userlevel ごとに基本的なテキストの本文を記述しておくことができる。次に、学部・学科ごとに異なる語彙レベルを反映した例文を、属性 audience に所属ごとの値を指定しながら記述する。例えば、学部ごとに異なる例文を指定する場合の記述は以下のようなになる。

```
<para audience="igaku"> The survey was
sponsored by CDC. </para>
<para audience="jinbun;kyoiku"> The meeting
was sponsored by government. </para>
```

このような形で記述された共通テキストのソースファイルに対して、授業を担当する教員は、自分が担当するクラスの習熟度レベルと学生の所属を出力の際に指定することにより、そのクラスに最適化されたテキストを生成することができる。

5. テキストの電子化

ここまで示した共通テキストの作成方法は、紙媒体での印刷を前提としたものではない。もちろん、紙への印刷は可能であるが、いわゆる「多品種少量生産」の方式のため、紙での印刷の場合はコストが高くなる。紙での印刷の場合にはコストが高くなる。

この XML 文書作成規格に基づいた共通テキスト作成は、電子化したテキストの利用を前提としている。テキストを電子化することで、修正・改訂などが素早く行え、共通テキストを統一的に管理することが可能になる。

しかし、この方式の最大の利点は、教員や学習者というテキストのユーザに対して、そのユーザに最適化されたテキストを提供できるという点にある。ユーザに最適化される点は、上述の、プラットフォーム、習熟度、所属だけでなく、html, ePub, pdf などのフォーマットに関しても、ユーザの希望する形式で出力することができる。

6. おわりに

本稿では、DocBook などの XML 文書作成規格におけるプロファイリングの手法を用いる事により、大規模で複雑な共通テキスト作成や利用を効率的に行うことができることを示した。

参考文献

- (1) Norman Walsh : DocBook 5, O'Reilly, 2010. (<http://www.oasis-open.org/docbook/documentation/reference/html/docbook.html> にオンライン版が公開されている)
- (2) Bob Stayton : DocBook XSL: The Complete Guide, 4th ed., Sagehill Enterprises, 2007. (<http://www.sagehill.net/docbookxsl/> にオンライン版が公開されている)
- (3) 内海淳, 葛西真寿, 小山智史, 佐藤友暁, 鈴木裕史, 丹波澄雄, 松谷秀哉, 水田智史, 『オープンソースソフトウェアによる情報リテラシー 第2版』, 共立出版, 2013