

# XML パーサを使った論文書式自動校閲ツール Word-Lint の開発

松谷栄輝 花川典子

阪南大学大学院 企業情報研究科

論文やレポート課題は高等教育機関において必要不可欠な教育手法である。しかし、論文やレポートの添削には教員の多くの時間が必要であり、特に論文の形式や書式のチェックは単純作業あるにもかかわらず、細かく手作業でチェックするケースが多い。そこで、論文の形式や書式のチェックを自動的にチェックするツールを開発した。これによって、単純作業の形式チェックの労力を削減し、教員は本来の専門的知識や論理的思考を評価する時間を確保することができる。

## A lint tool for theses formatted on MS-Word using a XML parser.

HIDEKI MATSUTANI NORIKO HANAKAWA

HANNAN UNIVERSITY GRADUATE SCHOOL OF CORPORATE INFORMATION

### 1. はじめに

大学等の高等教育機関において、論文やレポートによる課題は必要不可欠な教育手法である。学生に論文やレポートの作成を課すことにより、知識の獲得、データ分析能力、文章構成力、文章作成能力を養うことが出来る。

しかし、論文やレポート課題には、添削に多大な時間と労力が必要という問題点がある。添削には、専門的記述内容や論理展開のチェックと、形式や書式のチェックというふたつのチェックに分けることが出来る。内容や論理の展開の添削には、各分野に精通した教員の専門知識が必要であり、人間の知的判断が必要不可欠である。一方、形式や書式の添削は文体を統一したり、段落の先頭の字下げや句読点を統一したりする等、専門知識が不要な単純作業である。本論文ではこの単純作業を自動化して、知的内容のチェックを充実させることのできるツールを提案する。

### 2. 関連研究

現在、実現している添削作業に関するツールやサービスの代表的なものは、MS-Word の文章校正機能やテンプレート機能、また外部の専門家による添削・校閲サービスを利用する方法である。MS-Word の文章校正機能では、スペルチェックや表記の揺らぎの自動チェック等があるが、形式や書式までは確認できない。テンプレート機能では、スタイルの設定や書式の追加等、テンプレート機能を熟知していても、書式や形式の統一に扱うには限界がある。また、外部の専門家によるサービスは教員の学生指導という本質的な部分が欠落する可能性がある。

また、笠原らはソフトウェア製品のユーザマニュアル自動評価の研究として、マニュアルの校閲作業における文書推敲支援ツールの実適用評価をおこなった[1]。この研究では目視による校閲の品質向上のため、校閲作業支援ツールを開発し実適用評価をした。校閲作業支援ツールにより校

閲総工数の 69%を削減できたと報告された。校閲に着目したことは、我々のツールと同様であるが、本ツールはマニュアルに限定されていない、一般性の高い日本語校閲を目指している。

### 3. Word-Lint の紹介

本項では前述の問題を解決するために開発した Word-Lint について説明する。Word-Lint は MS-Word2007/2010 形式の文書の内部構造(XML)を解析し、自動的に論文・レポートの形式や書式を添削するツールである。ここでは MS-Word2007/2010 形式の文書の概要および Word-Lint の要件について説明する。

#### 3.1 MS-Word2007/2010 形式の文書の概要

MS-Word 2007/2010 形式の文書は Microsoft 社製の Office スイートである MS-Word 2007 および MS-Word2010 で保存される電子文書の保存形式である。この保存形式は従来の Word 2003 までのバイナリ保存形式と異なり、XML と呼ばれるタグを用いて文書を表現する技術と ZIP と呼ばれるファイル圧縮技術を組み合わせで実現されている[2]。これにより、従来までは他のアプリケーション製品で互換性を得られなかった MS-Word 形式が XML をサポートするアプリケーションであれば作業することが可能となり、ユーザ自身もデータを抽出したり、別の目的に使用したりできるようになった。図 1 に Office (2007) Open XML ファイル (MS-Word 2007 文書)を展開したものを示す。このように、内容はプレーンテキストの XML 文書であるため、ユーザ自身の手によって様々な操作が可能となっている。

#### 3.2 MS-Word2007/2010 形式の文書の解析

Word-Lint の機能の実現には、MS-Word2007/2010 形式の文書を解析する必要がある。MS-Word2007/2010 形式の文書は図 1 のような構造を持っている。これを解析することにより、文書内の問題となるフォントやテキストを探し出したり警告としてコメントを挿入したりする。

```

- <w:r w:rsidR="00CC11D2">
+ <w:rPr>
<w:t>利用者が特定の非標準的な書式を適用している。</w:t>
</w:r>
- <w:r w:rsidR="00A373BC" w:rsidRPr="00A373BC">
- <w:rPr>
<w:rFonts w:ascii="HG丸ゴシック M-PRO" w:eastAsia="HG丸ゴシック M-PRO" w:hAnsi="HG丸ゴシック M-PRO" w:cs="Tahoma" />
<w:color w:val="000000" />
<w:sz w:val="22" />
<w:szCs w:val="21" />
</w:rPr>
<w:t xml:space="preserve">利用者が特定の非標準的な書式を適用している。</w:t>

```

図 1 MS-Word2007 文書の構造

たとえば、本文中の文字列は、「<w:t>」というタグで定義されている。そのため、「<w:t>」タグ内の文字列を読むことで本文中のテキストを抽出できる。また、フォントの情報は「<w:rFonts>」タグで定義されている。図 1 で確認できるように、「<w:t>」から「</w:t>」の間に本文が書かれており、「<w:rFonts>」には「w:ascii」と「w:eastAsia」属性として「HG 丸ゴシック M-PRO」というフォントが設定されていることがわかる。「w:ascii」と「w:eastAsia」属性はそれぞれ範囲内の英字フォントと日本語フォントを設定している。これらの XML タグや属性等を確認することで、書式の統一や「である調」の統一等を確認することが出来る。

### 3.3 Word-Lint の要件

Word-Lint は論文やレポートの形式や書式を添削する。そのための要件として次の 3 つを挙げた。一つ目は書式チェック、二つ目は日本語ルールのチェック、三つ目は論文形式のチェックである。

書式チェックは、フォント・文字サイズ等の文字に関する情報の統一チェックを行う。表題・副題や章・節・項の書式や表記の違い、本文の書式の統一の確認を行う。日本語ルールチェックは、文章における日本語ルールの欠如や表記ゆれを確認する。段落の先頭の字下げ、句読点の統一、文体の統一の確認を行う。論文形式のチェックは、表題・日付・所属・氏名の有無、目次の有無、参考文献の有無、謝辞の有無、文献番号への参照の確認等を行う。これらの設定は別途設定ファイルまたは設定画面を設けることで添削者が柔軟に対応できるようにする。

### 4. 適用例

Word-Lint の試作品の適用例を図 2 に示す。Word-Lint は試作品であるため、すべての要件を実装していない。日本語ルールチェックのみ実装している。その機能は、段落の先頭の字下げ、句読点の統一、文体の統一の確認である。検出された項目は図 3 のように警告として既存の MS-Word 文書に書き込まれコメントとして表示される。図 3 中の①は、段落の字下げが無い点についての警告である。通常、段落の最初は字下げが必要だが、この段落では字下げが行われていないために警告を表示している。②は文体についての警告が表示されている。「である調」で統一しな

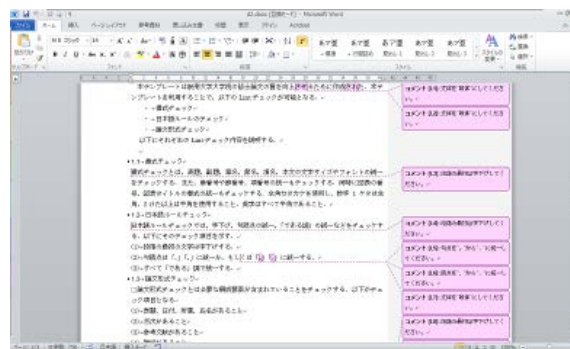


図 2 Word-Lint の適用例

- ① 同プロジェクト コメント [L7]: 段落の最初は字下げして...
- ② ます コメント [L6]: 文体を"常体"にしてください...
- ③ いなかった コメント [L3]: 句点を"。"から"。"に統一...

図 3 コメント部分の拡大

ければならないにも関わらず、「ですます調」で記述されている部分があるところのように「常体」にするよう警告が表示される。③は句点に対する警告である。句点を「.(ピリオド)」で統一する必要がある場合に、「。(マル)」が使用されていた場合に「.」で統一するように警告が挿入される。現時点ではこれら三つの項目についてのみ検出を行っているが、手作業と比較して一瞬で形式や書式のチェックを終えることが出来る。

### 5. まとめ

Word-Lint は文書の書式、論文の形式、日本語のルールを自動で校閲するツールである。このツールを使用することにより、論文・レポートの添削者(教育担当者)は、論文の内容についてより多くの時間を割くことが出来るようになる。その結果、課題を提出する学生に対し、ひとりひとりに質の高いフィードバックを与えることが可能となる。これにより、Word-Lint は教育効果の向上が狙えるツールである。

今後の課題としては、まず試作品の Word-Lint に対して三つの要件を実装し、それぞれの検証を行う。次に、開発設計やリファクタリングを経て、添削者に使いやすいユーザーインターフェイスを持つツールとして開発を行う予定である。

### 参考文献

1) 笠原健成, 小林栄一, 荒井真人, 絹川博之: マニュアルの校閲作業における文書推敲支援ツールの実適用評価, 情報処理学会論文誌, Vol.42, No.5, pp.1242-1253 (2001)  
 2) [http://msdn.microsoft.com/ja-jp/library/aa338205\(v=office.12\).aspx](http://msdn.microsoft.com/ja-jp/library/aa338205(v=office.12).aspx)