

# アクセシブル電子書籍制作・編集・閲覧システムの 多言語対応について

鈴木 昌和

(九州大学マス・フォア・インダストリ研究所)

KEY WORDS: マルチメディア DAISY, 制作・編集・閲覧システム, 多言語対応

○山口 雄仁

(日本大学短期大学部)

## 1. はじめに

デジタル教科書など電子書籍のアクセシビリティについて、わが国では国際標準規格である DAISY(Digital Accessible Information System)/アクセシブル EPUB3 や UD ブラウザー、音声埋め込み PDF など、全盲・重度弱視・読字障害など障害者の特性に応じたさまざまな形式が提案され、実際にそれらアクセシブル電子書籍が広く提供されるようになった。我々のグループでも、理数系文書用 OCR ソフトウェアやアクセシブル数式エディタなどの独自技術を活用し、墨字印刷ないし PDF の書籍からマルチメディア DAISY/アクセシブル EPUB3 形式の電子書籍を作成するための、制作・編集・閲覧用システムを開発してきた[1]。

## 2. システムの他言語対応の現状

我々の制作・編集・閲覧システムは海外での利用を念頭に、当初から日本語版と英語版の2本立てで開発を進めてきた。しかし特に理数系電子書籍のアクセシビリティは英語圏以外の諸国でも重要な課題であり、我々のシステムの他言語版を求める声は少なくなかった。そのため海外からの相談を受けたケースのうち、当該言語に十分な知識を持つ協力者が得られた場合について、2013年頃から徐々にではあるが対応を進めてきた。それらは主に拡張ラテン文字など特殊な文字を利用する欧州諸言語に関するもので、たとえばフランス語版・イタリア語版・トルコ語版・チェコ語版などがこれまでに試作されている。

現在わが国では学齢期の視覚障害者の数はかなり少なくなっているが、発展途上国では依然として数多くの視覚障害のある児童・生徒がいる。これは主に医療体制の不備などに起因するもので、そうした諸国ではそれら児童・生徒向けにアクセシブル電子書籍に対する需要が確実に増している。インドやバングラデシュ、東南アジア諸国でも、主に DAISY によるアクセシブルな電子書籍の制作が行われているが、ただ理数系書籍については数式など特殊表記の取り扱いが容易ではないため、初等・中等教育用教材の提供すらまだほぼ手つかずの状況にある。

我々のグループではこうした点に鑑み、昨年から DAISY コンソーシアム・インド、インド工科大学デリー校の研究協力者とともにインド仕様のアクセシブル電子書籍制作・編集・閲覧システム開発に着手したが、インドの場合文字体系そのものが全く異なるなどこれまでの欧州諸言語ではなかった困難な課題が多くあり、残念ながら実際のシステム試作にまでは至っていない。

こうした中で本年度、NPO 支援技術開発機構(ATDO)では、ベトナムにおいてマルチメディア DAISY によるアクセシブル理数系教科書制作を進めることとなった。ベトナムではフランス統治時代にアルファベットが全面導入され、テキストはすべて（一部特殊なアクセント符号付きの）アルファベットで表記されるため、インドに比べ対応がはるかに容易であると推測される。我々のグループではこのプロジェクトに協力し、アクセシブル電子書籍制作・編集・閲覧システムのベトナム語対応に取り組むこととなった。

## 3. 科学文書用 OCR ソフトウェアのベトナム語対応

我々の理数系文書用 OCR ソフトウェア：インフティ・リーダーは、墨字印刷ないし PDF の理数系文書を文字認識し、さまざまなアクセシブル形式に変換できる。特に最新版のインフティ・リーダー3では、文字情報が埋め込まれた e-born PDF について、PDF 解析ソフトウェアを活用して直接それら文字情報を利用するようになったため、日本語・英語以外の言語でも特別に当該言語用の OCR エンジンを使わなくても、テキスト領域で十分高い認識精度が期待できる。ただ数式や図・表など特殊表記が混在する理数系文書では、文字認識を行う前にテキスト領域・数式領域・図領域などを正しく切り分ける必要がある。この前処理を正確に行うにはその言語で頻出する短い単語(文字列)のリストが必要で、それがないとテキスト部分を数式と見なすことによる誤認識が数多く発生する。ATDO では今回ベトナム人視覚障害者を含む開発協力者を組織しており、それらの協力を得てベトナム語の理数系科目教科書のサンプルデータを元に必要な基本語リストを制作することになっている。さらにベトナムの理数系教科書に特徴的なレイアウトなどについても検討を進め、効率よくレイアウト解析を行う手法を確立する予定である。

## 4. 電子書籍編集・閲覧ソフトウェアのベトナム語対応

我々のマルチメディア DAISY 編集・閲覧ソフトウェア：チャティ・インフティでは、インフティ・リーダーの認識結果を直接読み込み、それに必要な修正・編集を加えた上で、マルチメディア DAISY/アクセシブル EPUB3 や、通常のブラウザーで閲覧できるチャティ・ボックスなどの形式で出力できる。チャティ・インフティでは数式部分の読み上げ方法を定義するファイルと、メニューやダイアログの表示内容を定義するファイルがソフトウェア本体とは独立になっており、それら2つのファイルをベトナム語用に編集すれば、原理的にはソフトウェアはベトナム語版になる。ただ、チャティ・インフティは音声合成エンジンとしてマイクロソフトのスピーチ API, Ver.5 (SAPI5)を利用するため、ベトナム語で音質のよい SAPI5 音声を利用できるかどうかは鍵であり、今回は、ベトナム語音声合成の専門家にもプロジェクトに加わって頂くことになっている。

## 5. まとめ

近年我が国はベトナムなど東南アジア諸国から数多くの留学生を受け入れており、それら留学生の中にも視覚障害・読字障害のある学生が含まれていると推察される。我々の今回の取り組みは、そうした学生を日本国内でサポートする際にも活用できるものと期待される。

(謝辞)

本研究は ATDO と協力し、一般財団法人：日本国際協力システム(JICS)の JICS NGO 支援事業の助成を得て実施されるものである。

(参考)

[1] <http://www.sciaccess.net/>

(SUZUKI Masakazu, YAMAGUCHI Katsuhito)