

MathML の表現形式から意味形式への変換およびその応用

理学専攻・情報科学コース 2040634 内橋 夏実 (指導教員: 浅本 紀子)

1 はじめに

近年, Web ページにおける数式表記の重要性の高まりに伴い, 表現方法も変化している. 以前は数式を記述した画像ファイルを Web ページに埋め込む方法等が利用されていたが, 現在では数式を Web ページ上に直接記述する方法が多く利用されている. 本研究では WWW 上の数式表現の標準技術として W3C[1] から勧告されている数式記述用 XML 言語, MathML に着目した. MathML は HTML に埋め込むことで HTML 文書の一部として記述する. MathML には, 数式の見た目の位置情報を記述する表現形式と, 数式の意味構造を持ち厳密に数式の意味を表現する意味形式の二種類の記述方法が存在する. MathML で数式が記述されている Web ページの多くは表現形式が用いられているが, 表現形式は数式の視覚的情報にすぎないため, 数式処理システムへの直接の応用が困難である.

本研究では, Web ページ上の数式を数式処理システムへの応用がし易い状態にすることを目的として, 表現形式から意味形式へ変換するツールの開発を試みた. 変換が可能になることで数式データの利活用が高まり, 数式検索の実用化など様々なことが技術展望として考えられる. 本研究では, 意味形式への変換の応用例として, 数学のオンライン小テストを自動生成するツールも開発した.

2 意味形式への変換

2.1 準備

表現形式は記述したい数式の見た目が正しく表現できれば問題ないため, 同じ数式でも複数の記述方法が存在する. しかし, 意味形式への変換を考える上ではそれが課題となる.

本研究と同じく表現形式から意味形式への変換ツールの開発を扱った先行研究 [2] では, 変換前の表現形式記述で数式情報が欠落していた場合, 最も一般的に考えられる意味を表す意味形式へ変換している. また別の既存研究で, 表現形式記述で欠落した数式情報をユーザの指示通りに付加し, 一意で正規的な表現形式への変換を行う GUI ツールの開発が存在する. これを八巻澄奈が Web アプリケーション化したものを開発した [3]. 本研究では, このツールによって一意な書き方に記述しなおされた表現形式記述を意味形式に変換する. これによってユーザの解釈する意味を記述する意味形式に変換することができる.

2.2 変換手法

① MathML 記述箇所の抽出

渡された HTML ファイルの中の MathML 記述部分を抽出し, それぞれを新たなファイルに書き出す. さらにそのファイルを `xml.dom.minidom` で整形する.

② リストの作成・整理

MathML ファイルを一行ずつ読み込み, 表現形式記述で用いられるタグを利用してリストにその行の内容を

保存する (図 1).

対応する開始タグのみの行を表すリストから終了タグのみの行を表すリストまでを一つにまとめ, 整理する (図 2).

<pre> <math> <msup> <mi>x</mi> <mn>2</mn> </msup> <mo>-</mo> <mn>9</mn> <mo>=</mo> <mn>0</mn> </math> </pre>	list0	list1	list2	list3	list4
	<math>	<msup>	<mi>	<mn>	None
	None	None	x	2	None
	None	None	</mi>	</mn>	</msup>
	</msup>	<mo>	<mn>	<mo>	None
	<mo>-</mo>	-	9	=	0
	<mn>9</mn>	</mn>	</mn>	</mn>	None
	<mo>=</mo>	</mo>	</mo>	</mo>	None
	<mn>0</mn>	</mn>	</mn>	</mn>	</math>
	</math>	</math>	</math>	</math>	</math>

図 1: リストの作成

list1	list5	list6	list7	list8
<msup>	<mo>	<mn>	<mo>	<mn>
(list2,list3)	-	9	=	0
</msup>	</mo>	</mn>	</mo>	</mn>
x^2	-	9	=	0

図 2: リストの整理

③ 構文木の作成

演算子の結合度に基づき「等号・不等号, 加減, 乗除, ⁡, 項」の優先順序で, 構文木の根となるリストを図 2 の右側から探索する. 決定した根の右側と左側で, 同じように探索し左右の子ノードを決定する. これを繰り返して構文木を作成する.

④ 意味形式の書き出し

行きがけ順に構文木の各ノードを訪れ, 対応する意味形式を出力する.

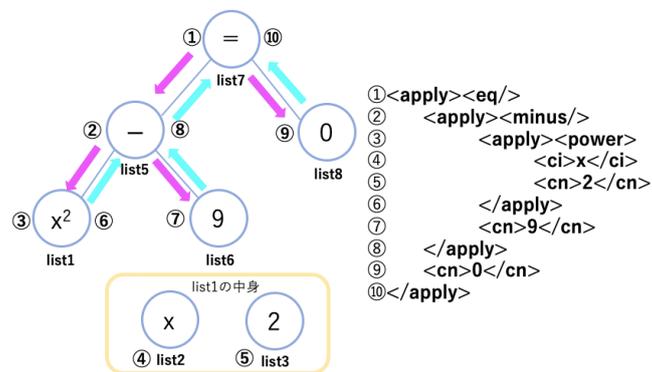


図 3: 構文木

2.3 実行結果

本コンバータツールは Python3 で実装する. 「意味形式変換ツール」ディレクトリでコマンドを実行すると (実行コマンド: `python3 makefile.py` 変換元ファイル

名), 同ディレクトリの中に `out_folder` ディレクトリが作成され, その中に表現形式で記述された部分が意味形式に置き換わったファイルが出力される.

3 オンラインテスト自動生成ツール

Moodle: オンラインでの課題提出・評価, 小テスト等を利用できるオープンソースの e ラーニングプラットフォームであり, プラグインを追加することで機能を拡張できる.

STACK: 数式による回答が可能なオンラインテスト・評価システムのことで, 数式の入力にはフリーの数式処理システムである Maxima の数式入力記法に従う. STACK3.0 以降では Moodle のプラグインとして動作する.

3.1 目的

Moodle の小テスト作成には, 問題作成者が Moodle の使い方に慣れるまでにある程度の時間を要するという課題がある. 本研究では, MathML で数式が記述された Web ページから誰もが簡単にソースコードを入力できる状況なら, 問題作成者が HTML ファイルで小テストを作成し, それを Moodle の小テストに変換することで課題を克服できると考えた.

Moodle には小テストで使用する問題をインポートおよびエクスポートするための Moodle 固有の XML フォーマットが存在する. そこで, HTML ファイルの数学小テストを Moodle 用 XML ファイルに変換するツールの開発を試みた.

3.2 意味形式の利用場所

STACK によるオンラインテストにおける問題文の数式部分は, 図 4 の「問題テキスト」二行目のように, TeX で書かれている. しかし, 受験者が入力した回答の正誤判定やフィードバックのために, 数式処理可能な Maxima でもこの数式を「問題変数」として登録しなければならない.

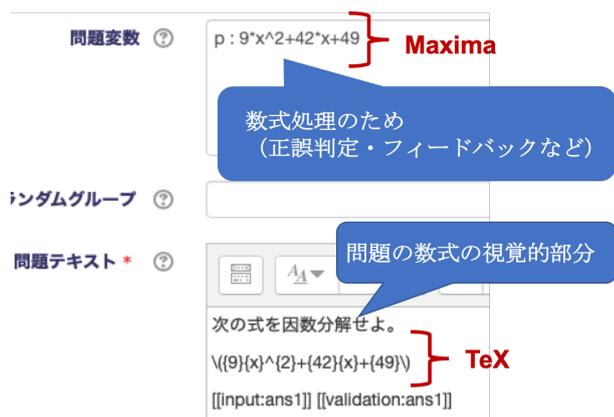


図 4: STACK の小テスト

したがって数学の問題が記述された HTML ファイルを Moodle 用 XML ファイルに変換する場合, MathML の表現形式で記述された数式を TeX と Maxima にそれぞれ変換しなければならないが, 表現形式から TeX や Maxima に変換することは手間のかかる作業である. そこで, MathML 表現形式記述を意味形式に変換し,

意味形式記述から TeX と Maxima に自動で変換する. 意味形式は意味構造を持つため, 他言語への変換は容易に可能となる.

3.3 対象とする問題

今回対象とする数学の問題は 一元多次方程式/一次方程式/展開/因数分解/微分/積分の 6 種類で, 中学・高校の代数学を範囲としている. HTML ファイルに記述されている文章や数式からどの問題を作成するか判断することはできないため, 問題作成者側が MathML 開始タグの `class` 属性を利用して指定するか, GUI で問題の種類を選択し指定する.

3.4 実装例

本ツールも Python3 で実装する (実行コマンド: `python3 makefile.py 変換元ファイル名`). コマンドを入力し実行すると, Moodle 用 XML ファイル「`quiz.xml`」が作成される. お茶の水女子大学 2019 年度版 Moodle の問題バンクに `quiz.xml` をインポートし, それをプレビューした様子が図である.

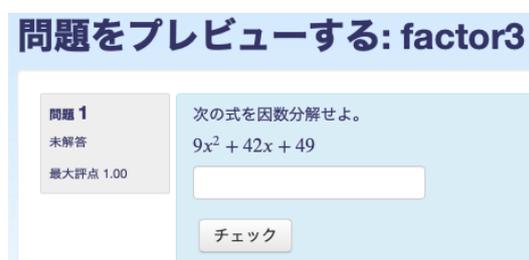


図 5: インポートした問題

4 まとめ

MathML の表現形式を意味形式へ変換するコンバータツールを開発した. さらにこの応用例として, 数式部分が表現形式で記述された HTML ファイルから, Moodle 上の STACK を利用した数学オンラインテストを自動生成するツールを開発した.

現在, 四則演算・分数・対数・平方根・累乗根・関数を含む数式の変換は精度が高いが, これ以外の数式の変換や精度の向上が課題として挙げられる. また, 一意な表現形式への変換ツールと統合させ, ウェブアプリケーションの形で実装できるようにしてユーザビリティを高めたい.

参考文献

- [1] W3C <https://www.w3.org> (2021/12/30/閲覧)
- [2] MathML における表現形式から意味形式へのコンバータツールの開発, 荒川玲佳, お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻 情報科学コース 修士論文 (2020)
- [3] MathML における一意な表現形式記述への変換ツールの開発, 八巻澄奈, お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻 情報科学コース 修士論文 (2022)