

MathML における一意な表現形式記述への自動変換

渡辺裕美*1・浅本紀子*2

Email:watanabe.hiromi@is.ocha.ac.jp

*1: 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

*2: お茶の水女子大学大学院基幹研究院自然科学系

◎Key Words MathML, 表現形式, 意味形式

1. はじめに

近年, Web ページにおける数式の表現方法はインターネット利用の多様化に伴い, その表現の仕方も変化してきている。かつては, LaTeX で作成した文書から数式部分を画像で切り出し, Web ページ貼り込む方法が多かった。しかし, 画像の張り込みによる数式の表示は, フォントや文字サイズの変更や, 異なるウィンドウサイズへの対応が出来ない。そこで, MathJax⁽¹⁾によって Web ページ内に LaTeX 表式を使って数式を表現する方法が利用されるようになった。これにより, 文章内のあらゆる箇所に数式を自由に表現できるようになった。しかし, いずれの方法も数式をただ Web 上に記号か文字として表現しているだけに過ぎず, 数式の意味までは考慮されていない。

そこで, 注目したのが MathML⁽²⁾である。MathML とは, XML アプリケーションの 1 つで, 数式を記述するためのマークアップ言語である。HTML の文書内に埋め込んで, HTML 文書として扱われる。MathML の特徴として, コンピュータによる数式の意味認識において有利となるよう設計されたという点が挙げられる。人間が MathML を直接書いたり編集したりすることは意図されていないため, 数式を 1 つ表現するにしても, 文字数数の多さや使用するタグの多さから, MathML が使用される機会は少なかったが, 電子書籍の普及や教材のデジタル化に向けて, 数式の表現が重要視されるようになったことで, MathML の使用機会が増加しつつある。そこで本研究では, MathML の普及に向けて MathML の可能性を拡大させるシステムを提案する。

2. 背景

2.1 MathML における表記の仕方

MathML はコンピュータによる数式の意味認識において有利となるよう設計された言語ではあるが, その表記の仕方には 2 種類あり, 書き方は 3 種類ある。表記の仕方は, 表現に特化した表現形式記述, 意味に特化した意味形式記述の 2 種類である。そして書き方は, 前記 2 種に加えて, 表現形式記述と意味形式記述を混合して描く書き方がある。図 1 に書き方の例を示す。混合で書く場合, 表現形式記述の中に意味形式記述を書く際は, 意味形式記述を書き始める箇所の前に `<annotation-xml encoding="MathML-Content">` を加え, `</annotation-xml>` を意味形式記述の書き終わりの箇所に加えると良い。意味形式記述の中に表現形式記述を組み込む場合は, 何も制約はなくそのまま書いて良い。

表現形式記述は, 数式を表現する際文字の大きさや位置関係を指定してレイアウト構造を定めるものであって, 概念は LaTeX と同様であるが, LaTeX より数式の意味に関して厳密に書くことができる。意味形式記述は, オペレーターや引数などの数式要素を定義してから数式の意味を厳密に表現するものであり, Web ページ上で実際に意味を確実に取ることができる表記の仕方は, 意味形式記述である。

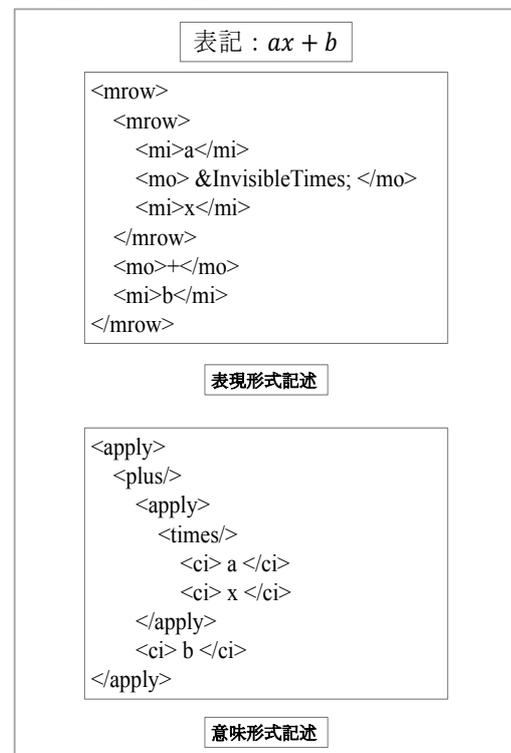


図 1 MathML の記述例

2.2 意味形式記述における問題点

意味形式記述を利用することによって, 確かにコンピュータが数式の意味を正しく読み取ることができるが, ここにはいくつかの問題点がある。1 つ目は, 意味形式記述がすべての数式を表記できないことだ。意味形式記述には, 数式を表記するためのタグが存在しないものが 1 部ある。2 つ目は, 意味に特化した記述方式なので数式が見づらい可能性がある。3 つ目は, そもそも意味形式記述で書かれた数式を正確に表示できる無料のブラウザが今現在存在しないことである。以上の問題から, 意味形式記述は今のところ Web ページではほぼ使用されていない。

2.3 表現形式記述の有用性

意味形式記述では、最大の問題点として主要なブラウザで表示することができないという点があるが、表現形式記述は Firefox, Safari ではサポートされ、MathJax と併用して使用することによって Chrome, Internet Explorer でも表示することが可能である。実際、Wikipedia などでは表現形式記述を使用して書かれている。また、表現形式記述であっても、数式の意味を読み取れる書き方が存在する。数式の見目の細部までこだわって表記することもできる。

2.4 表現形式記述の問題点

表現形式記述は数式を自由に表記できるが、自由度が高いために1つの数式であっても、複数の表記の仕方が存在するものがある。その中には、省略して書かれたことで見目が歪になっている書き方や、数式の意味を読み取れる部分が省かれた書き方が存在する。例を図2に示す。

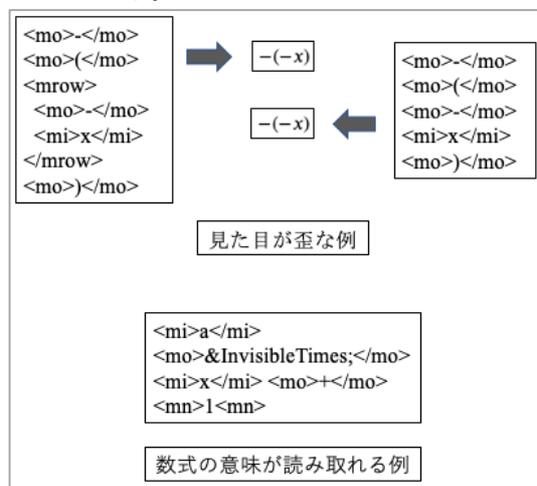


図2 表現形式記述の問題点の例

2.5 関連研究

関連研究として、LaTeX から MathML の意味形式記述に変換する研究⁽³⁾、MathML の表現形式記述から意味形式記述へ変換する研究⁽⁴⁾がある。

3. 本システム

3.1 本システムの概要

本研究では、表現形式記述で数式を正しく表現し、また、表現形式記述から意味形式記述への変換の支援を行うためのシステムを開発している。具体的には、表現形式で表記された数式に対して、その書き方が数式の見目を正確に表しているか、コンピュータで意味を介するために必要な部分が省略されていないかを、ユーザーに判断させて一意な正規的な書き方に変換する。

3.2 実装

図3に現在変換可能な数式の例を示す。

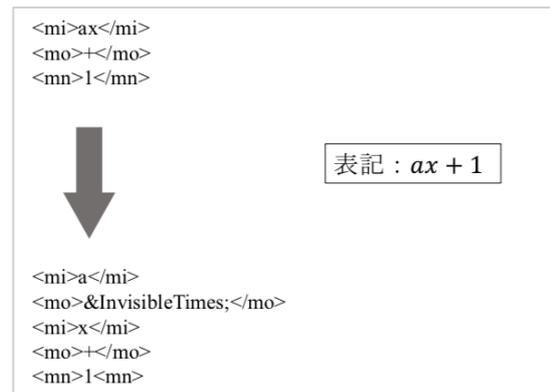


図3 変換可能な数式の例

ax が $a \times x$ か ax という変数かをユーザーに確認し、 $a \times x$ であった場合は図3の矢印先のように変換を行う。 $\⁢$ は見えない演算子といい、数式の見目に変化はないが、変数 a と変数 x の掛け算という意味を数式に持たせることができる。

3.3 今後の実装

現在、変数同士の掛け算に対してコンピュータで意味を取るように変換する段階だが、今後は、数式に意味を持たせることのできる見えない演算子が他にも存在するので、それらに対しても付け加えられるように実装していく。そして、数式の見目を綺麗に表示することができる実装を行う。また、ユーザーを介して数式の変換を行うので、ユーザーインターフェースの開発を行っていく予定だ。

4. おわりに

電子書籍の普及や教材のデジタル化に向けて、数式の細かい表現や数式の意味まで表現できる MathML は今後対応されるブラウザが増え、使用される機会が増えていくであろう。本研究では、MathML を使用するための足がかりとなるように、表現形式記述の柔軟性に着目して、その可能性を拡大するシステムの開発を行った。今後、MathML のバージョンもアップグレードして、表現できる数式の幅が広がっていくことに期待したい。

参考文献

- (1) <https://www.mathjax.org> (2019/6/12閲覧).
- (2) <https://www.w3.org/TR/MathML3/> (2019/6/14閲覧).
- (3) 渡辺千晶, 視覚障害者学習支援のための MathML 変換と利活用, 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 (2017).
- (4) 荒川玲佳, MathML における表現形式から意味形式へのコンバーターツールの開発, 2019 PC カンファレンス 全国大会 (2019/8/予定).
- (5) 道廣勇司: "MathML 数式組版入門 Ver1.1", アンテナハウス株式会社 CAS 電子出版 (2017).
- (6) <https://reference.wolfram.com/language/XML/tutorial/MathML.html.ja?source=footer> (2019/6/13閲覧).
- (7) <https://www.antenna.co.jp/AHF/> (2019/6/12閲覧).
- (8) 奥村晴彦・黒木裕介: "[改訂第6版] LATEX_ε 美文書作成入門", 株式会社技術評論社 (2013).