

数式埋め込みコンテンツと標準化 -現状理解と標準化プロセスについて-

長坂耕作

KOSAKU NAGASAKA

神戸大学発達科学部

FACULTY OF HUMAN DEVELOPMENT, KOBE UNIVERSITY*

1 はじめに

近年のXML技術の進歩により情報の集約、相互運用などが可能になっている。これに伴い、数学コンテンツを表現するMathML規格もW3Cから勧告されており、少なくないソフトウェアでMathMLを用いて数学記号や数式の表現が可能となっている。このことは、どの種の文書にも必ず含まれる数や数式などの数学オブジェクトを、画像や特殊なフォーマット(TeXなど)で記述せずとも、その他の文章と同じレベルで扱えることを意味している。実際に、数学知識データベースの開発や専門家でない人も容易に扱える数学掲示板など、必要であるにも関わらず、その特殊性(通常のテキストで表現できない点)から実現できていなかった方面での利用が進んでいる。

これらMathMLとその他のXML技術により数学コンテンツに関連する問題は解決済みのように思えるが、それは間違いである。MathMLやその他のXML技術は、数学に関連する最終成果物(データのみを扱うビジネス文書、数式を集めただけの公式集、小説のように読むことだけが目的の文書など)における数学オブジェクトの表現に多大な寄与をしたが、その生成過程に直接的には寄与していない。例えば、次のような場合、MathMLとその関連技術では表現できず、独自規格などを利用する必要がある。

- MathMLによって記述された数学オブジェクトの正当性の確認
- MathMLによって記述された数学定理(命題)の自動証明
- 技術レポートなどで、初期データを変更した場合に計算結果を再計算
- 教科書などにおいて、数値やデータを変更してグラフを再描画
- 専門家が利用した計算ソフトの結果を、別のソフトで検証すること
- 専門家がその研究のためにアルゴリズムや実験を統一的に扱うこと
- ネットワーク分散型の数値数式融合計算システムを協調動作させること
- その他の、数学オブジェクトが静的でなく動的に扱われる文書

*nagasaka@main.h.kobe-u.ac.jp

これらの状況を鑑みると、XMLなどのICT技術による情報の流通・共有化が進む中で、非常に重要な数式埋め込みコンテンツが流通・共有化の波に乗り遅れていることがわかる。さらに、現存する勧告であるMathMLについて、日本からワーキンググループに参加しているものはなく、この分野の日本の国際貢献度は非常に少ない。このような状況を数式処理に携わる研究者は変えていく必要があるのではないだろうか。本稿では、ひとつの可能性としての包括的な規格や、その重要性などの提案をしたい。

2 現状理解と標準化プロセス

2.1 既存の規格など

数式埋め込みコンテンツに直接的に関連する規格として、MathML, OpenMath, OpenXMの3つが研究開発されている(一部終了)。さらに間接的に関与すると考えられる規格・概念として、(株)ジャストシステムによるxfy技術の中のVocabulary Connection Descriptor(VCD)があげられる。

MathML W3C勧告の1つであり、XMLを利用して、数学オブジェクトを表現するマークアップ言語である。現在は規格のメンテナンスが主な活動となっている。

OpenMath 2004年まで約10年に渡りXMLによる数学オブジェクトの表現について研究していたプロジェクトであり、多大な影響をMathMLに与えている上に、数学オブジェクトの構造を正しく表現する点でMathMLを越えている。OpenMathに関連するものとして、現在は、ActiveMathやWebALTなどのプロジェクトが進行中である。

OpenXM 国内で研究が進められている規格で「Open message eXchange for Mathematics」の略である。OpenXMの対象は数式処理システム間における数学オブジェクトの交換であり、国産システムのRisa/Asirを中心に実装が進んでいる。

VCD XML文書におけるXSLTの役割(異なるXMLへの変換やスタイル定義等)を包括するような概念で、XMLをXSLT等により変換したオブジェクトが変更された際に、どのようにして元のXMLデータを変更するかを指定することを可能にする。

これらを組み合わせることで次のようなことが可能である(主なものを例示)。

- OpenMathによる記述で数学的な正確さを保ったまま、XSLTやVCDにより、表示部においてはMathMLを使うことで、相互に関連する数値等の単純な数学オブジェクトを有機的に結合し、静的なコンテンツにおいて動的な変化を可能とするコンテンツの作成。
- MathematicaやMapleのような数学オブジェクトを含む多様なコンテンツ作成を可能にするシステムを使いながら、OpenXMによる通信を行うことで、汎用的ではない特定分野向けの特種なソフトウェアで一部の演算を行うようなコンテンツの作成。

しかしながら、同時に次のような制限が課せられる(主なものを例示)。

- VCDでデータを有機的に結合する場合、その結合を実際に行う部分は別途準備しなければならない。特に、結合部において高度な演算が必要となる場合、実際に計算を行うソフトウェアとの通信部分を個別に実装する必要があり、XMLを中心とする標準規格だけで記述することは出来ない(例えば、Mathematicaで計算させる場合はMathLinkによる通信を、Mapleで計算させる場合はMapletによる通信を実装する必要があり、方法も異なる)。

- OpenXM は通信規約に過ぎず、作成したコンテンツはやはり同じソフトウェアでなければ利用することが出来ない。
- 数学的に非常に単純なアルゴリズムであっても、その記述通りにコンテンツ内で計算を実行させることは出来ない（個別のシステムごとに実装を行えば別であるが、流通や共有は困難である）。

これらの制限は、動的に変化するような生きたコンテンツを制作するための国際標準が存在しないために生じている。数式埋め込みコンテンツには、数学オブジェクトの表現だけでなく、それ以外のオブジェクト（グラフィックス、アニメーション、プログラミング言語としての文法など）とそれらの有機的な結合が必要なためである。これを実現する規格は存在しない。一方で、数式埋め込みコンテンツを作成し利用するために開発されているプラットフォーム（世界的なシェアを考慮し、数値数式融合計算システムに限定して解説を行う）においては、次のようなコンテンツを作成し、同一プラットフォーム利用者内で流通し、利用することが可能になっている。繰り返しになるが、プラットフォームを越えては利用することが出来ない。

- 十分な印刷品質を持つ純粋数学や工学的分野などの学術論文を作成すること
- 小学校、中学校、高校、大学における動的な教材を作成すること（式や数値を変更した結果を確認したり、グラフがリアルタイムに変更されたりする）
- ネットワーク越しに複数の同一プラットフォームを接続し、ネットワーク分散型の並列計算を実行すること（OpenXM を利用することで、OpenXM 対応のプラットフォームに関しては相互に接続させることは可能）
- 特定の条件を満たす問題（数学に限らない）を自動生成させること
- 特定の条件を与えて、命題の自動証明を行わせること
- 実験データとともに記載されているアルゴリズムを実際に動かすこと
- その他、計算が必要となるありとあらゆる分野で求められていることを、統一されたファイル（文書）として取りまとめ、必要に応じて実行・計算を行うこと（統計的処理、可視化処理、シミュレーション、スペルチェック、アニメーション、タイプセット、画像処理、多種多様なデータの入出力、WBT、データベース処理など）

2.2 標準化すべき規格など

数式埋め込みコンテンツとインターフェイスの仕様を考えた場合、次のような視点から複数の規格を定める必要があると思われる。

- 数式やグラフィックスなど様々なコンテンツとその有機的結合の表現に関するもの。例えば、MathML や OpenMath を内包する XML スキーマの定義（XSD で記述）など。名前を付けるならば、ECASML: Extended Computer Algebra System ML だろうか。
- 動的コンテンツ部を計算処理するエンジン部に係る通信に関するもの。例えば、数値数式融合計算システムにおけるエージェント間の通信に対応するもの。既存の MathLink(Mathematica のもの) や Maplet(Maple のもの) などの一般化。名前を付けるならば、ECOAP: ECAS Object Access Protocol だろうか。

- ユーザインターフェイスが備えるべき機能に関するもの。例えば、動的な表現やユーザ操作などの要件定義など。名前を付けるならば、GCUIS: GemC User Interface Standard だろうか。

そして、これらの集合体をもって、数式埋め込みコンテンツに係る仕様とすれば、この枠組みを GemC: Generic Embedded Mathematics Content と称することを提案する (gem は、英語で珠玉、優れたもの、という意味を持つ)。

以下、これらの規格についてどのようなものにすべきかを述べる。

2.2.1 GCUIS (ユーザインターフェイスの要件規定)

既存の主なシステムを包括する形で、数式埋め込みコンテンツに必要なユーザインターフェイスの要件を考慮すべきであろう。これまでの多種に渡る既存の知見を反映し、将来のニーズにも幅広く対応可能なものとするために、次のような視点から検討することを提案する。

- ICT 技術を用いた数理情報教育に求められている課題とその対策
- 複合した XML 文書を統一的に扱う xfy 技術
- 世界シェアを二分する Mathematica に組み込まれている多様な機能
- 工学的応用で抜きん出ている Maple に組み込まれている多様な機能
- システム間通信において先駆的な役割を担った OpenXM の長所と短所
- 自動証明など Mathematical Knowledge Management に関する知見

これらは、編集機能、国際化、入出力、グラフィックス (3D 操作含む)、アニメーションなどを網羅し、言語や数学オブジェクトとの同レベルでのフロントエンド構造の統合も行う (Mozilla における XUL のようなもの) べきだろう。

2.2.2 ECASML (コンテンツ表現のためのマークアップ)

前述の要件規定 GCUIS が定めれば、既存の規格である MathML や OpenMath 等を内包する形で、コンテンツ表現のためのマークアップについて検討が可能となる。その際には、情報収集や意見調整のために、W3C において MathML の保守を担当している Math IG (数学利害グループ) へ参加している研究者と連携し、既存技術との整合性を可能な限り取った上で、検討を進めるべきだろう。

2.2.3 ECOAP (コンテンツの通信に係る規定)

実際に、ECASML で記載されたコンテンツに動的な変化が発生したら、その変化に応じてコンテンツを処理する必要が出てくる。この処理はユーザインターフェイス部分ではなく、別途それ専用のプログラム (エンジン部またはカーネルと呼称) を利用して行うのが良いだろう。ADL などにおける SCORM 規格のように、コンテンツ部と実際のエンジン部は分離は自然なことである。これによって、複数の計算エンジンを利用したネットワーク分散型の並列計算なども対象にすることが可能である。これらフロントエンド部とカーネル部をネットワーク分散型のエージェントと捉え、エージェント間の通信プロトコルをコンテンツの通信に係る規定として策定することを提案する。なお、汎用性を考慮すると XML のメッセージ交換を主とし、SOAP などをベースにすることが一般的であるが、実際の計算システムでは割り込みや起動などの機能を有しなければならないため、XML 技術だけでは実現できないと思われる。

2.3 標準化プロセスなど

数式埋め込みコンテンツを数式処理におけるキラーアプリと考えるのであれば、世界を牽引するほどの優れた規格を生み出さなければならないと共に、それらを国際標準化を行うことも重要である。一言に国際標準化と言っても多種多様な団体があり、また標準化しようとする内容によっても適する団体は変わってくる。前述の ECASML のように、MathML や OpenMath を内包する形で、コンテンツ表現に関する規格を考えるのであれば W3C が、ECOAP のように通信に係る規格であれば IETF が、標準化を提案する先として適切であろう。

仮に、W3C において国際標準化に取り組む場合についてのプロセスを紹介しておく (MathML と密接な関わりを持つ規格と仮定する)。W3C が勧告案を出すためには、1) 当該トピックへの関心が高まる、2) 当該トピックに関する活動提案がディレクターにより行われる、3) 活動提案により設立された WG、IG、CG などによりテクニカルレポート等が作成される、4) 成熟し必要性が認められれば勧告として公表される、というステップが踏まれる。今後も MathML のメンテナンスを行っている Math IG が引き続き活動を行うことを仮定すると、MathML に関する国際標準化プロセスは、次のようになると考えられる。以下において、AC とは諮問委員会、AC 代表とは W3C に所属する組織から派遣される AC へ参加する研究者のことを指す。

1. W3C への参加
2. AC 代表の Math IG への参加
3. Math IG で、その規格の必要性のコンセンサスを得る
4. その規格に関する IG の活動ノート作成
5. ディレクターによる活動提案の作成
6. AC によるレビュー
7. その規格に関する WG や IG の設立
8. テクニカルレポートの作成
9. W3C による勧告

個々のステップには比較的長い時間が必要で、例えば、MathML 第 2 版は、1999 年 12 月に草案 (これは、IG などのテクニカルレポートよりも進んでいる状態) が公開されてから W3C 勧告になるまでに、約 1 年 3 ヶ月かかっている。

一方、IETF において国際標準化に取り組む場合のプロセスはまったく異なる。

1. IETF への規格の提案を行う
2. Internet-Drafts として公開される (または差し戻され修正後に再提案)
3. Standards Track に移動するに十分な根拠を与える実装の公開
4. Proposed Standard として公開される
5. Draft Standard として公開される
6. Standard として公開される

IETFによるRFCの策定においては、具体的な実装の有無、異なる実装間における互換性の確保などが要求される。そもそも、Internet-Draftsは自由に提案できるため、多種多様なものが日々50件近く提案されているが、Standardとなる規格は相対的に数少ない。

参 考 文 献

- [IETF] <http://ietf.org/>
- [Maple] <http://www.maplesoft.com/>
- [MATLAB] <http://www.mathworks.com/>
- [Mathematica] <http://www.wolfram.co.jp/>
- [MathML] <http://www.w3c.org/Math/>
- [MuPAD] <http://www.mupad.de/>
- [OpenMath] <http://www.openmath.org/>
- [OpenXM] <http://www.openxm.org/>
- [Risa/Asir] <http://www.math.kobe-u.ac.jp/Asir/>
- [W3C] <http://www.w3.org/>
- [xfy] <http://www.xfytec.com/>