

科学技術分野における知見共有のためのプラットフォーム

Information Platform for Scientific Knowledge Sharing

芦野 俊宏[1], 斉藤 隆之[2]

Toshihiro Ashino[1], Takayuki Saito[2]

[1] (株)情報数理研究所, [2] 情数研

[1] IMS Lab. Inc., [2] IMSLAB

<http://www.interwall.org/~ashino/>

数値実験が一般的な手法となり、用いられるプログラムの継承、共有が重要な課題となっている現在、プログラムの実装の詳細、経験的なパラメータの設定とこれによる計算結果への影響、出力されたデータの処理、結果の解釈に至る一連の情報を共有することが研究活動を推進するうえで重要な課題となっている。本稿では、計算科学技術に関連する知識のうち、実際に計算をすることで得られる経験に基づくものとして、ドキュメントとデータ、およびそれらの間の関連づけによって記述可能な部分を「知見」と呼び、これを蓄積、共有するためのプラットフォーム開発とデータ構造標準化の試みについて述べる。

はじめに

従来、科学技術計算の分野における入力・出力情報は計算の度毎に作成されるものと見做され、出力データはグラフなどの形式でビジュアライズされ、その結果のみが論文や報告書として共有されてきた。

しかしながら、数値計算が一般的な手法となり、用いられるプログラムの継承、共有が重要な課題となっている現在、プログラムの実装の詳細、経験的なパラメータの設定とこれによる計算結果への影響、出力されたデータの処理、得られた結果の解釈に至る一連の情報を共有することが研究活動を推進するうえで重要な課題となっている。ここでは、計算科学技術に関連する知識のうち、実際に計算をすることで得られる経験に基づくものとして、ドキュメントとデータ、およびそれらの間の関連づけによって記述可能な部分を「知見」と呼び、これを蓄積、共有するための基盤となるプラットフォームについて述べる。

XMLによる知見の共有

WWWでは、HTMLがドキュメント共有の手段として用いられているが、HTMLはドキュメントの見た目を指定するという性格が強い。XMLとこれを処理するソフトウェアの充実により、計算機処理の可能な語彙を定義し、これをドキュメントに埋め込むことでドキュメントの内容を処理するための環境が整備されつつあり、従来のWebと区別してSemantic Webなどと呼ばれている。

科学技術計算の分野においても、XMLを用いた語彙を共有することによってデータやこれに関連する知見を分散したサーバ上に蓄積して関連付け、計算機を用いた共有、検索を行うことが可能となる。すでに、流体力学の分野においてはnetCDFのように計算によって出力される膨大なバイナリデータを相互に交換するための標準的なファイルフォーマットが存在する。これはタグ付のバイナリファイルであるが、タグを解釈する、あるいは変換するための枠組みは用意されていない。XMLを用いたタグはさらに汎用的なものとして、XPointerなどを用いたハイパーリンクやMathMLを用いた数式の記述などを可能とし、これらはブラウザや数式処理システムによって扱うことができる。

知見プラットフォームプロトタイプ

このような標準的なタグセットの作成とこれを利用する計算科学技術のためのプラットフォームの構成について議論する場として2000年11月、有志により、「知見プラットフォーム推進協議会」が設立された。同協議会は、分野横断的な立場から計算科学技術のためのデータ構造のありかたとその共有について議論する場である。

同協議会では、議論の叩き台となるプロトタイプとして平成11年度IPAのプロジェクトによって構築された知見プラットフォームを参照している。

このシステムにおいてデータベースは数値計算の入出力データとここから導かれた結果、書誌情報などを含む文書と計算プログラムのソースコード、マニュアルなどを含む文書の二種からなり、これらの間の関連性をハイパーリンクを用いて表現する。このことにより、他の計算例で用いられた経験的なパラメータの設定と計算への影響を調べる、他の計算例の入力パラメータを編集して用い、元となった計算例に対するリンクを保持することで関係を明らかにするといった経験の共有が可能となる。

定義されたXMLタグは、

- ・システム・データベース用タグ
 - 書誌情報
 - 文書構造

- プロトコル定義
- タグ辞書
- ・分野共通属性タグ
 - 物理定数
 - 物理量
 - データ型
 - (単位系)
- ・分野別タグ
 - 生体高分子
 - 流体解析
 - 非構造メッシュ
 - 物質材料設計

のように体系化される。このうち分野別タグは知見プラットフォーム実証実験に参加した研究グループの専門分野に対応している。実証実験においては、参加した研究グループに分散したサーバ群を配置し、データベースへの登録、結果の検索、可視化のための Java を用いたプログラムを配布した。これにより複数の研究機関にまたがって数値計算に要する知識を共有し、特に収束計算に経験的なパラメータを要するタンパク質の三次元構造計算において他者の計算事例の参照が積極的に行われ、効果的な共同作業が行われた。

今後の展望

今後、協議会においては複数の専門分野に関してその分野の情報交換を議論する分科会を組織し、そらの代表者によって、分野横断的なデータ構造の議論を行うとともに、入出力データに加えて知見としてどのような情報を付加する必要があるかを記述してゆく。一方で、データ型、バイナリデータの扱いなど、数値計算データの取扱いには必ずしも適さない面があり、このようなデータを扱うための技術的な側面、あるいは XML 関連の規格についての検討を行ってゆく予定である。