

統計メタデータ記述の 技術動向とXML文書の利用

研究開発部試験作成支援研究部門 大 津 起 夫

大学入試センター試験において試験内容、実施方法などの改良を行うためには、過去の試験の実施によって得られたデータに含まれる情報を十分に把握し利用することが必要である。しかし取り扱うべきデータが膨大であり、採点基準など多くの付加的な情報が分析において必要とされる。また、センター試験業務の他に、研究開発部では各種関連分野の調査が積極的に行われているが、そこで得られたデータは個別の研究者によって管理されており、分析結果は紙媒体の報告書中に表現されている。今後、これらのデータと資料が蓄積されて大量となり、研究者の入れ替わりなどが生じると、次第に過去の業務と研究によって得られた情報の全体像を把握することが難しくなる可能性がある。

この問題に対応するためには、各種のデータと資料の管理情報を統一された形式で電子化し検索可能とする必要

がある。このような技術基盤への要求は、企業や教育研究機関において共通に生じている。データベース技術はこのような要求に応える一つの手段であったが、従来は定型的な情報のみを主たる管理対象としていた。現在では、インターネット上のWebの普及を背景に、マークアップ言語を利用する環境が整備され、各種のソフトウェアツールが無償または低コストで利用できる。このため、XMLで記述された情報をデータ管理に用いようとする試みが広がりつつある。本稿では、統計データ（特に調査データ）のコードブック（実施要領、調査票、コード表などの記録）の記述を中心に、XMLによる統計メタデータ記述の研究と利用の動向を紹介した。

マークアップ言語は複雑な入れ子の構造を持っており、文書を読み込みこの構造を理解するソフトウェアを自作するのは容易ではない。ここでは論理

型記号処理言語 (Prologと呼ばれる計算機言語) 上のXMLパーザを用いたXML文書の利用と生成の方法を加えて紹介した。この手法は、処理効率の面で幾分性能が悪く、また現在のところ広く普及はしていないが、論理型言語の持つパターンマッチング機能を用いることにより、極めて柔軟なXML文書の操作が可能であり。XSLを用いた処理と比較しても多くの利点を持つ。特に、多様な構造を持つ文書を扱う研究や業務目的に有益である。