

自然言語の文法記述との類推に基づく 構造化文書処理技術の開発

今村 誠 三菱電機(株) 情報技術総合研究所



【受賞論文】

- ・素性論理に基づく XML 文書ルール記述言語 DRDL とインターネット文書交換システムへの応用
- ・今村誠, 渡邊圭輔, 増塩智宏(三菱電機(株)), 渡部明洋(三菱電機インフォメーションシステムズ)
- ・情報処理学会論文誌, Vol.47, No.3, pp.751-764 (2006)

受賞論文の特徴である下記の2点を中心に、開発の経緯を述べる。(1)「自然言語の文法記述」と「構造化文書のルール記述」間の類推を用いて、構造化文書処理方式を提案、(2) CALS (Continuous Acquisition and Lifecycle Support) や EC (Electronic Commerce) の実用システムを開発。

文書処理との出会い～電子マニュアル～

DRDL (Document Rules Description Language) の原型は、1993年の電子マニュアル用ハイパーテキストの試作に始まる。ここで、コンテンツ変更の容易性をねらって、コンテンツ記述に SGML (Standard Generalized Markup Language) を採用し、SGML からハイパーテキストの内部構造への変換処理用の記述言語を考案した。言語処理の文法記述を SGML のルール記述と対応させると、生成規則は DTD (Document Type Definition) に相当し、意味規則は文書変換記述に相当する。後者の目的特化版を自作したわけである(拡張 DTD と自称した)。

SGML から XML へ～CALS/EC システム～

1996年には、電子部品カタログの記載内容が部品分類や属性の標準用語辞書に従うか否かを検証するシステムを開発した。拡張 DTD に、電子部品カタログ (SGML) と辞書 (SGML) とを照合するルール記述を追加した。

1997年からは、昇降機的设计仕様書を自動検証するワークフローシステムを開発する。ここで仕様書 (SGML) からチェックスクリプト付 HTML への変換が必要になり、SGML から HTML 表構造への変換記述を拡張 DTD に追加した。後に、

電子データ交換や設備管理システムにも横展開した。翌2月には XML (Extensible Markup Language) が勧告になる。

素性論理との類推による記述言語の設計

1999年には、官公庁への許認可申請を電子化するシステムの開発に参画した。ここでは、「輸出許認可申請では、輸出品目に応じて必要書類が定まる」等の XML 申請書の検証に加えて、既存の検証や変換機能を統一的に記述できるように、拡張 DTD の仕様を整理して DRDL とした。言語仕様設計時には、「自然言語の文法記述用の素性論理」と「構造化文書の検証/変換処理」間の表-1に示す類推関係を用いた。

特殊な事例かもしれないが、この一連のソフトウェア開発では、類推に基づく仕様設計は以下の3点で有用と感じた。(1) 思考を節約でき、設計時間を短縮できる。(2) 「整合性が保証された理論」と仕様とに対応関係があると、その対応関係の範囲内で仕様の整合性が保証される(仕様のバグを削減)。(3) 仕様の限界や追加すべき基本要素が分かりやすいので、拡張性に優れる。

所感

コストや期間の制約下にある企業のソフトウェア開発現場において、形式論理に基づく文書指向設計により品質や生産性を少しでも向上させたいという思いで、自分なりの工夫を積み重ねて DTD の拡張として集約してきた。このたび名誉ある賞をいただいたことへの畏れはあるが、これらの思いを評価いただいたものと大変嬉しい。また、種々の開発への参画を通じて、さまざまな機会、知識、そして励ましを受けた。(財)新世代コンピュータ開発機構、電子機器・部品 CALS、(財)ニューメディア開発協会、(財)安全保障貿易情報センター、本会研究会(デジタル・ドキュメント、情報学基礎、情報システムと社会環境)、弊社内でお世話になった皆様に感謝いたします。

(平成19年5月11日受付)

素性論理	構造化文書のルール記述	
	検証	変換
パス /	ロケーションパス	ロケーションパス
連言 ∧	連言	逐次文
含意 ⇒	含意	if 文
全称限量子 ∀	for-each 文	for-each 文
存在限量子 ∃	反復数チェック	要素の追加/削除
等式 =	等式チェック	代入
要素関係 ∈	型チェック	(対応関係なし)
包摂関係 ⊆	スキーマ上位互換	(対応関係なし)

表-1 素性論理と構造化文書の検証/変換との対応関係

今村 誠(正会員) Imamura.Makoto@bx.MitsubishiElectric.co.jp
1986年京都大学数理工学専攻修士課程修了。偏微分方程式等を学ぶ。同年三菱電機(株)入社。自然言語対話、知識表現、知的インタフェース、構造化文書処理、テキストマイニング等の研究開発に従事。