

NCALS のはん（汎）用文書型定義の開発の方針と課題

5 S - 9

若鳥陸夫(CALS 技術研究組合: NCALS),

中村洋, 下市徹, 田中道夫, 北村和彦(NCALS), 西川一紀(沖電), 横浜秀夫(富士ゼロックス), 滝田善克(CRC 総研),
中村秀之(NTT データ), 伊藤和夫(NTT データ), 伊申亮二(三菱電), 秦務(松下電), 豊本康雄(東芝), 古沢嘉隆(東芝),
山本篤史(日本 IBM), 小林茂(日本ユニシス), 村上晴夫(日立), 木戸祥夫(富士通), 広川久夫(富士電), 田中洋一(凸版)

1.はじめに(forward)

NCALS では,印刷への SGML 応用ではなく,技術文書の電子化及び格納・検索・交換・提示の構造を文書記述言語 SGML¹⁾ による文書型定義(DTD)で定義した。その際,ISO9000 対応文書化,日本語の共通化及び利用者対機械間の界面の分離を試みた。現在の計算機環境では,金物,操作系,応用系などによって,文字集合,図形・映像の符号化方式の変化があり,共通文書として利用しにくい面があった。NCALS では,将来の日本語文書は,交換・処理の対象として等しいことが利用者の潜在要件であると認識し,処理対象の日本語符号及び図形・映像の符号化を選択した(以降,NCALS 共通符号と呼ぶ)。この NCALS 共通符号を用いることによって,PC,UNIX 機,はん用計算機などの種別を問わず共通な文書として取り扱う世界を共通化できると提案し,それを実証するつもりである。^{2),3)}

2.日本語技術文書の型定義の開発の方針(concept)

2.1 日本語符号の構成の要件

(1)基本的な文字集合の要件

NCALS 共通符号を構成する文字符号は,JIS X 0202(情報交換用符号の拡張法)によって,次の(A)~(C)とした。

(A)英数字記号(ISO/IEC 646-IRV)を G0 集合へ指示し,使用中の表の図形左側へ呼び出す。

(B)JIS X 0208-1990(情報交換用漢字符号)を G1 集合へ指示し,使用中の表の右側へ呼び出す。JIS 情報交換用漢字符号表から次の文字を除いた。

(a)2 バイト英数字記号(SP,A..Z,1..9 など)。ただし,二重引用符及びチルド記号を除く。

(C)JIS X 0212(情報交換用漢字符号-補助漢字)を G3 集合へ指示し,SS3 によって図形左側へ1文字だけ呼び出す。

備考 1 バイト片仮名文字は,使用しない。フォントの違いによって,符号を分けない。

(2)乱順序で処理できる文字符号

(A)通信及びファイル読み出し(状態遷移による順序処理)だけでなく,文書作成編集系の乱位置指定による処理対象にできること。

(B)すべての文字符号値が一意であること。

(3)SGML 宣言できる構成

(A)ISO/IEC 2022 に適合した符号拡張方式であること。

(B)状態遷移を使用しないで,符号値が一意の文字に割り当てられること。

2.2 内容記法の要件

NCALS-DTD では,次の(1)~(8)の内容記法を導入する。

(1)数式のための TeX 及び eqn 表現

(2)絵のための CGM 符号化方式

(3)図形のためのラスタ図形符号化方式

(4)大型図面の分割のためのタイル状ラスタ図形符号化方式

(5)3次元図形のための IGES(第1.1版拡張)

(6)写真及び彩色画のための JPEG(第1.1版拡張)

(7)動画のための MPEG(第1.1版拡張)

(8)音響のための ISO/IEC 8613-9(音響内容体系中の符号化方式) (第1.1版拡張)⁴⁾

A Study of Development of Document Type Definition(DTD) for Technical Documents in NCALS
Rick Wakatori,etal.

Nippon CALS Partnership

c/o Time 24 bldg.

2-45 Aomi,Koutou-ku,Tokyo, 135-73, Japan

3. 道具の作成及び事例の投入実験(results)

NCALS-DTD に適合した電子文書の有効性を実証するために、3.1～3.3の事項を実施している。

3.1 道具の開発

- (1) シフトジス対 NCALS 共通符号間の変換
- (2) 1 バイト片仮名と 2 バイト片仮名との変換
- (3) 2 バイト英数字記号と 1 バイト英数字記号との変換
- (4) 標準外字削除

備考 SGML 文書では、文字参照機構があり、予め定義した文字の呼び名で定義文字を参照できる余地がある。

3.2 現存する文書作成系からの変換道具の開発

- (1) ワードプロセッサ固有形式から NCALS-DTD 適合文書への変換
- (2) ワードプロセッサ固有形式から HTML 変換
- (3) 文面への HTML 構文の自動付与

3.3 実在文書の論理分割と投入

実在する文書を NCALS-DTD にしたがって論理分割する試み及びその結果を電子文書化する実験を行なっている。その結果、次の(1)及び(2)が発見されている。

(1) 論理構造の枠組みから逸脱した部分

論理構造によって誘導する処理系が出現すれば、この逸脱は減少すると推定される。その理由は、文書作成規範が頭にあるが強制する手段がないために、その場の思い付きなどで、自由な論理構造を生成してしまい勝ちだからである。例えば、2次元の表が細部では3次元になっていたり、番号付き段落によって解説が始まったりすることがある。したがって、これらの形式違反は、原文を修正することにした。更に従来の文書は、紙を提示媒体として考えていたため、電子文書の提示面にはそぐわない構成が多々発見されている。これは、電子化作業のための変換であって、部分的には、そのまま表現した静的文書と電子化形式である動的文書とを併記し、比較することも行なうつもりである。

(2) 論理構造への追加が自然な部分

初期の NCALS-DTD では、表現できないが、日本語技術文書の論理構造として、許容してもよい変化が原文書に存在する場合がある。これは、NCALS-DTD を改版することによって、吸収している。

4. 今後の課題(future works)

NCALS の技術文書だけを対象とした文書型定義の開発及びその実証実験は、次の 4.1～4.4 の課題を解決して、はじめて、実用になる。

- 4.1 日本語文書の共通符号としての流通性の向上
- 4.2 基本論理構造と応用系の利用者界面との一意的な対応の成立の実証
- 4.3 利用者が使える水準に利用者界面の向上
- 4.4 系統だった運用性の確認

5. まとめ(summary)

NCALS-DTD の開発では、“将来の日本語文書は、共通の符号体系で表現され、操作系の境界を超越して、作成・格納・伝達・提示の処理ができるのがよい。”という無謀とも思える理想をかかげて活動してきた。幸いにして、多くの計算機製造者から派遣された研究者及び技術者の前向きな応援を得て基盤を整えつつある。しかし、UNIX系でEUCを使用している操作系は小変更であるが、シフトジスを採用しているUNIX、Windowsなどは基本部分への対応を必要とするため、更に多くの方々の賛同と協力をお願いする。

参 考 文 献

- 1) JIS X 4151“文書記述言語 SGML”, 日本規格協会, 1992
- 2) 若鳥: “NCALS-DTD の紹介”, パソコンリテラシ, パーソナルコンピュータユーザ利用技術協会, 1996年3月
- 3) 若鳥: “NCALS の近況-第3 G”, CIF だより, CALS 推進協議会, 1996年8月
- 4) INSTAC: “JIS X 4109 原案”, 日本規格協会, 1996年3月