

XML を用いた構造データ交換方法に関する一考察

梅崎 利矢 村山 隆彦

NTT 情報流通プラットフォーム研究所

umezaki.toshiya@lab.ntt.co.jp, murayama.takahiko@lab.ntt.co.jp

5W-4

1. はじめに

近年、電子商取引市場ではサプライチェーンや業界の枠を越えたデータ交換を可能にする技術が求められており、XML を用いた帳票の交換技術が提案されている。しかし、新規のサプライチェーンやマーケットプレイスを構築するには、あらかじめ複数の企業や業界で議論をし、共通に使える帳票形式を決める作業や異なる形式間で必要な変換ルールを作成する作業等の前準備が必要である。このため、実際のシステム構築以前に時間や労力がかかっている。

本研究では上記問題に対し、XML Repository^[1]等に蓄積された既存の変換ルールを再利用して必要な新規変換ルールを自動生成する方式を提案する。この技術により、XML をベースとした帳票流通の自動化を支援でき、サプライチェーンや業界の枠を越えたデータ交換が可能となる。

2. 帳票流通モデルにおける問題点

図 1 のように、既定サプライチェーンや既に取り引のある企業間には変換ルールが存在する。しかし新規の取引等では必要な変換ルールが存在しない。このように変換ルールが欠落した部分で発生する流通経路の断絶が問題となる。

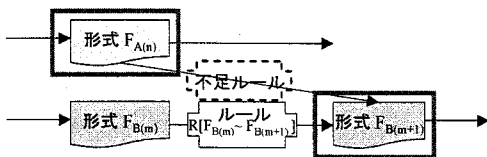


図 1: 帳票流通の例

3. 変換ルール自動生成問題

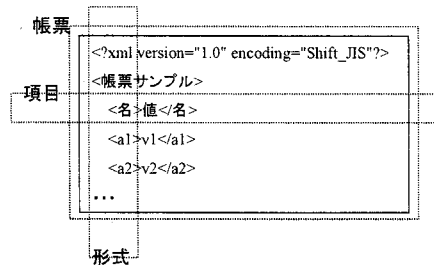
本研究における言葉の定義と、必要な定式化を行う。名称または記号に : が続き、その右辺に名称または記号の意味を記述する。

- #: 集合の要素の個数を示す
- $n, m, l \in \mathbb{N}$ (\mathbb{N} は自然数)
- 名, 値: 文字列の集合に属する
- 項目: [名, 値] の組

- 形式 F_n : 名の順序を示したもの。名は一意に決まるものとし、同じ名が複数回出現する場合の順序は問わない
- 項目数 $I[F_n]$: 形式 F_n に含まれる名の数。同じ名が複数回出現する場合、繰り返し回数はカウントしない
- 帳票: 形式に従った項目の集合
- 転記: ある帳票に含まれる値を異なる形式の帳票内にある同義の名を持つ項目に代入すること。特に、前者を転記元、後者を転記先、と呼ぶ
- 変換: 転記
- ルール $R[F_n \sim F_m]$: 形式 F_n から形式 F_m への転記に用いる、名の対応関係の全部、および転記元または転記先の形式を一部または全部記述したもの

以上の関係を XML で表現した例を図 2 に示す。変換ルールには XSLT^[2], eCo-Field^[3]等を用いる。

- 対応項目 $S[F_n \sim F_m]$: 任意の 2 形式 F_n, F_m 間において、意味的に同じであり、1:1 で結び付けられる項目の集合
- 新規項目: 形式 F_n にも出現する項目、および形式 F_n 以降にも出現する項目。この項目は、変換ルール $R[F_{n-1} \sim F_n]$ を手動で生成した場合でも転記不可能である
- 新規項目数 $N[F_n]$: 形式 F_n に含まれる新規項目の数



ルール = eCo-Field ルール, XSLT 等

図 2: XML 帳票の例

A study in structural data exchange process using XML
Toshiya UMEZAKI, Takahiko MURAYAMA
NTT Information Sharing Platform Laboratories

4. 定式化

図 1 において、形式 $F_{A(n)}$ は企業 A 系列のサ

サプライチェーン上にある n 番目の形式、形式 $F_{B(m)}$ は企業 B 系列のサプライチェーン上にある m 番目の形式を示す。

形式 $F_{B(m)}$ を転記元とし、形式 $F_{B(m+1)}$ を転記先とする変換ルール $R[F_{B(m)} \sim F_{B(m+1)}]$ には、既に形式 $F_{B(m+1)}$ の情報（繰り返し条件等）が含まれている。このため、各帳票に含まれる項目数と対応関係に着目することによって、以下に示す比率を求めることが可能である。

適合率:

$$A[F_{A(n)} \sim F_{B(m+1)}] = \frac{\#(S[F_{B(m)} \sim F_{B(m+1)}] \cap S[F_{A(n)} \sim F_{B(m)}])}{I[F_{B(m+1)}] - N[F_{B(m+1)}]}$$

$$(0 \leq A \leq 1)$$

転記先 $F_{B(m+1)}$ の項目のうち、正しく値が埋まった比率を適合率と呼ぶ。この時、変換ルールに記述されている転記元の項目名及び形式は、与えられた対応付けによって置き換えられるものとする。

変換ルール自動生成法の目標は、適合率 A を 1 に近づけることであると言え換えることができる。

転記率:

$$T[F_{A(n)} \sim F_{B(m)}] = \frac{\#(S[F_{A(n)} \sim F_{B(m)}])}{I[F_{B(m)}] - N[F_{B(m)}]}$$

$$(0 \leq T \leq 1)$$

一般的に、形式 $F_{A(n)}$ と形式 $F_{B(m)}$ 間で、項目をどの程度対応付けられるかの比率を転記率と呼ぶ。

結合率:

$$C[F_{B(m)} \sim F_{B(m+1)}] = T[F_{B(m)} \sim F_{B(m+1)}]$$

$$(0 \leq C \leq 1)$$

変換ルール $R[F_{B(m)} \sim F_{B(m+1)}]$ をもともと規定されている転記元 $F_{B(m)}$ に適用した場合、転記先 $F_{B(m+1)}$ の項目を正しい値で埋めることができる比率を結合率 $C[F_{B(m)} \sim F_{B(m+1)}]$ と呼ぶ。結合率は、既存サプライチェーンにおける帳票間のつながりの強さを示すものと考えてよい。

類似率:

$$P[F_{A(n)} \sim F_{B(m)}] = T[F_{A(n)} \sim F_{B(m)}]$$

$$(0 \leq P \leq 1)$$

形式 $F_{A(n)}$ と形式 $F_{B(m)}$ 間で、項目をどの程度結びつけることができるかの比率を類似率と呼ぶ。

5. 変換ルールの作成アルゴリズム

変換ルールの作成アルゴリズムを図 3 に示す。

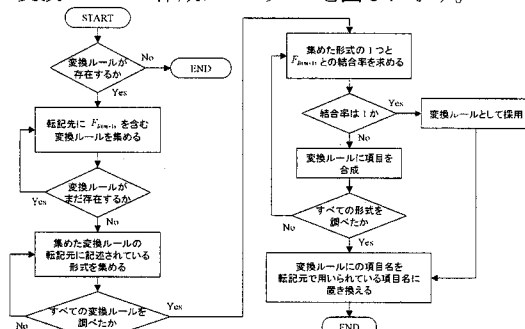


図 3: 変換ルール生成アルゴリズム

6. 例題を用いた実験と考察

代表的な宅配便業者 4 社の伝票を用いて、変換ルール生成アルゴリズムの検証実験を行った。結果、20 件中 18 件が適合率 1 になった。適合率 1 は、転記できる項目数が手動で生成した変換ルールと同じであり、これ以上転記できる項目がない、完全な変換ルールであることを意味している。これにより、前項の手法が有効なものであると言える。

7. おわりに

今回、結合率が 1 となる強力なサプライチェーン間の変換ルールがあれば、転記元 $F_{A(n)}$ と転記先 $F_{B(m+1)}$ の直前 $F_{B(m)}$ との項目対応関係を与えるだけで、 $F_{A(n)}$ から $F_{B(m+1)}$ へ転記可能な変換ルールを自動生成できることがわかった。この手法により、XML ベースの帳票流通自動化を支援でき、サプライチェーンや業界の枠を越えたデータ交換が可能となる。

今後は、実用レベルまで高めるために、複数の変換ルールを合成する手法における変換ルールの効率的な検索方法探求及びさらなる適合率の向上、構造を考慮した場合の検証が課題である。

参考文献

- [1] 梅崎・村山・山本: リポジトリを利用した企業間情報交換方式の一考察, 信学技報 Vol.100, No.206, pp.71-76
- [2] XSLT, <http://www.w3.org/TR/xslt/>
- [3] 横関・村山・山本: 制約条件に基づく XML 変換言語に関する一考察, 信学技報 Vol.100, No.441, pp.57-64