

再加工を考慮した表のデータ構造に関する一考察

2W-3

清末悌之、児島治彦、秋山照雄  
NTTヒューマンインタフェース研究所

1. まえがき

図表、テキストなど種々のメディアで構成された文書を認識した後<sup>(1)</sup>加工を行う場合には、各メディアが相互に関連しあうことに留意する必要がある。ここではまず、構造の明確な表の加工に焦点をあてる。文書加工の一環として表の操作を考えた時の問題点について考察し、これを解決するために階層性を持つデータ構造で表現する手法を提案する。このデータ構造を用いれば、表に対する加工が容易になる。

2. 全体の構成と既存技術の問題点

本発表の内容は図1のように位置付けることができる。このようなシステムを構築する場合、表の加工に関する既存技術の問題点は「表全体を拡大・縮小する場合には、文字とけい線を一括してイメージとして扱って処理するしかない」といった、ユーザが望む操作を自由に行えないことにある。これは、システム側がユーザの要求する加工処理を容易に実現できるようなデータ構造を提供していないことが原因であると思われる。これを解決するための表のデータ構造を提案する。

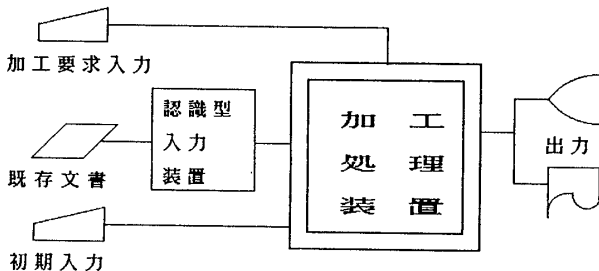
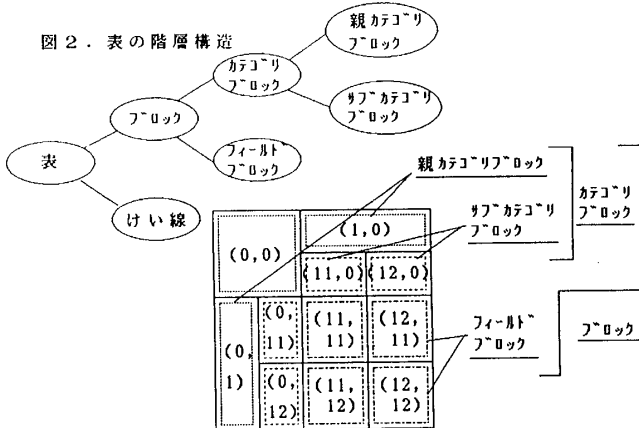


図1. システム構成図

3. 加工処理の基本要素

表はけい線とそれによって囲まれる領域(我々はこれをブロックと呼ぶ)から構成されているが、主としてブロック内のデータを重視する表現形態であると思われる。表とはまずはじめにデータが存在し、これを見やすくするために見出しを設定しブロックをこれに添って並べ、けい線を境界に引くといった表現形式である。従って、表に対してユーザが要求する加工処理はブロック内のデータに対する操作が主であり、ユーザの望む表の加工処理をシステム側で容易に実現するためには、これがダイレクトに扱えるデータ構造が望ましい。この観点から、我々は加工処理の基本要素をブロックとし、これをもとにデータ構造を決定する。



3. 用語の定義とブロックのインデックス

4. データ構造としてのインデックス管理構造

表を構成する各要素の階層構造を図2に示す。ここで用いる各用語の定義は図3で示す。後述する要素間の拘束関係はこの階層構造に基づいており、拘束関係はユーザの意図を処理に反映させるための知識であるといえる。従って、ユーザの要求する加工処理を容易に実現するためにはブロックを管理するデータ構造中に、階層性が表現されていなければならない。我々はこのためのデータ構造としてインデックス管理構造を提案する。ブロックに対するインデックスの付与例を図3に示す。インデックスとは、具体的にはこの図に示すような各ブロックに付与された(index<sub>1</sub>, index<sub>2</sub>)のようなベクトル表現である。このインデックスから①ブロックの隣接関係②カテゴリ/フィールドの区別③親カテゴリ/サブカテゴリの親子関係④カテゴリの種類(行方向か列方向か)等が判別できる。カテゴリの種類がn種存在するようなn次元テーブルへの拡張を行うには、インデックスの数を(index<sub>1</sub>, index<sub>2</sub>, ..., index<sub>n</sub>)と増やすことで対処できる。

5. 拘束関係に基づく加工処理の制御

表中の各要素間には、種々の拘束関係が存在する(表1)。我々はこれらの拘束関係を処理とは分離して管理し、加工処理が行われるごとに拘束関係が満足されているかどうかをチェックする制御方法を取ることにした。このため、より柔軟な表現を可能にするための、拘束関係の変更や削除に際しての管理が容易に行える。実際の加工処理は、次のように行う。①ユーザが加工処理を、表の中の要素もしくは表自体に対して要求し、最初の処理が行われる。②拘束関係が満たされているかどうかをチェックし、これを満たすように他の要素の加工処理が自動的に行われる。満たされれば処理を終了する。③②の加工処理は、最初の加工処理と同等のものとして扱われ全ての拘束関係が満たされるまで②を行う。

6. まとめ

マルチメディア文書中の表の加工処理を容易に行えるようにするためのデータ構造と制御方法を提案した。

参考文献 (1)第37回情処全大、児島ほか、「複雑な構造を持つ表の認識に関する基礎検討」

表1. 表の拘束関係

拘束要素	表示位置	表	ブロック	親カテゴリブロック	親カテゴリブロック	サブカテゴリブロック	フィールドブロック	フィールドブロック
表	①		①					
ブロック		①	①	③			③	②
親カテゴリブロック			①	③	③	③	②	
サブカテゴリブロック					①		②	
フィールドブロック					③			
フィールドブロック			①					
フィールドブロック			③					

拘束の種類  
①位置に関する拘束関係  
②大きさに関する拘束関係  
③要素の種類に関する拘束関係

"A Study about Data Structure of Table on Reconstruction Process"

Yasuyuki KIYOSUE, Haruhiko KOJIMA, Teruo AKIYAMA  
NTT Human Interface Laboratories