

「ロタス123」文書ファイル・「一太郎」文書ファイルの集計用4GLツール（第1版）

6J-6

岡田 尚
（株）東洋情報システム

1. はじめに

近年、一般業務において、PC上での簡易ソフト（表計算ソフト・ワープロソフト）による文書編集が日常化している。そして、管理業務では、簡易ソフトで作成された文書ファイルのデータ集計・分析作業を簡易ソフトで行うことの複雑さ・不十分さに直面している。

そこで、本稿では、文書file処理用4GLツール（第1版）を開発して実際の業務に使用した結果、及び、文書fileを対象としたデータ処理支援環境の検討について報告する。

2. 4GLツールの開発目的と機能

従来の4GLは、データベース（特に、RDB）処理用簡易言語として利用されているが、今回開発する4GLは

対象	データベース	文書file
論理構造	論理構造そのものがあらかじめ定義されている。	文書のレイアウト・タイトル・記入文字列などの物理構造に、論理構造が隠されている（論理構造としては、明示されていない）。
データ型とその値	データ型は、あらかじめ直接決められていて、その値の意味も一意に決まる。	論理構造が明示されていない上、文書のレイアウトや用途にデータの内容が依存するため、データ型は曖昧である。また、データ値は、自然言語による表現が一般的であり、意味を抽出することが難しい。

表1. データ処理の観点での相違点

NO	機能の説明
1	文書のレイアウトと論理構造（4GL言語表現）との対応づけをする機能を持つ。
2	データ値の不十分さ（NULL値の混在）やデータ型の曖昧さ（文字列と数値の混在：多相型）を考慮したデータ処理機能を持つ。
3	データ値（コマンド文・自然言語文など）を解析して論理データを抽出する機能を持つ。
4	論理データのより非手続的な論理演算式でデータ処理を記述できる機能を持つ。
5	処理結果などの論理データから、PC上での簡易ソフト（表計算ソフト・ワープロソフト）で再利用できるデータ（データ表記文・コマンド文・自然言語文など）を生成出力する機能を持つ。

表2. 文書file処理用4GLが満たすべき機能一覧

A Study of 4GL Tool for Document File
by "LOTUS-123" and "ICHTARO"
Takashi Okada
TOYO INFORMATION SYSTEMS CO, LTD.

文書file処理用簡易言語として利用されることを目的としている。処理対象が、データベースである場合と文書fileである場合のデータ処理の観点での相違点を表1に示す。また、文書file処理用4GLが開発目標にしている満たすべき機能を表2に示す。

3. 言語仕様・インタプリタ実行機能の特徴と性能

上述の機能を、「ロタス123」（表計算ソフト）を使用した計数データ処理の実現に限定して、言語仕様とインタプリタ実行機能を具体化した。すなわち、4GLインタプリタが「ロタス123」で作成された文書file内の一部（または、全部）のデータを4GLプログラムで定義した論理テーブルに対応付け、論理テーブルの演算（関係演算

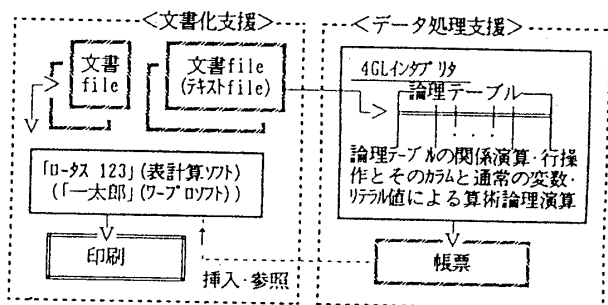


図1. 文書file処理用4GLの利用形態

NO	満たすべき機能	言語仕様・インタプリタ機能
1	文書レイアウトと論理構造との対応づけ	文書file（テキスト）に対し、行桁の引数指定で論理テーブルとの対応づけを行う関数が言語仕様として提供されている。
2	データ値の不十分さ（NULL値の混在）・データ型の混在	実行時のオプション指定で、NULL値の変数・テーブルのカラムを無視する指定がある。（実行エラーとしない指定）
3	データ値の解析による論理データの抽出	インタプリタは、ELL(2)型の構文解析手法を用い、カタログによる構文文法定義ができる。よって、4GL言語仕様と同様にしてデータ値の解析用の構文文法を定義しておけば実行可能だが、現在、使用していない。
4	論理データの非手続的な論理演算表現での処理記述	論理テーブルの関係演算（選択など）と、論理テーブルのカラム値更新操作で、対象行を選択条件式で指定できる。
5	処理結果の簡易ソフトでの再利用可能なデータ生成出力	「ロタス123」のデータ表記は、文字列データを「」でくくる仕様であり、実行時のオプション指定で実現している。

表3. 4GLの言語仕様・インタプリタ機能の特徴

・代入・算術論理演算)を行い、結果を帳票出力するものである(図1)。今回開発した4GLの言語仕様・インタプリタ機能と文書file処理用4GLが満たすべき機能(表2)との対応を表3に示す。

図2で示すように、このインタプリタは、構文解析に解析速度の遅いELL(2)型構文解析手法を用い、実行制御では実行用関数が解析木データを直接照会する方法を用いている。このため、解析速度・実行制御速度が共に遅く、また、メモリ使用量も大きくなった。このような方法を用いた目的は、

『構文文法(文脈自由文法)と解析木の処理関数のみの定義だけで、容易に新しい言語文法のインタプリタを作成(自動生成)できる。』

機能の実現のためである。その理由を表4に示す。

NO	おもな理由
1	文書file内の解析対象が多様(帳票・マトリクス・データ列・自然言語・図形など)で、且つ、物理構造と論理構造との対応付けが簡易なプログラムや作成文書fileのフォーマットに依存するため、言語仕様・実行制御仕様を一意に決定するのが難しい。
2	言語仕様・実行制御仕様が、データ処理業務の運用上のオーバーヘッドや処理結果の品質に大きな影響を与える。
3	文書fileを対象としたデータ処理は、PC上でのトランザクション処理(バッチ処理)であるため、インタプリタのデータ処理時間は、あまり重要ではない。

表4. 言語処理系の構築機能が必要な理由

その他、このインタプリタには、4GLプログラムの開発用として、構文解析のみに限定するモードや実行単位(解析木データ単位)に実行内容を表示するモードなどがある。

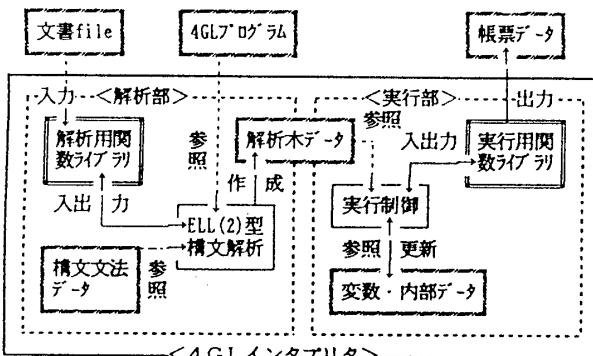


図2. インタプリタ内部処理の概略図

4. 4GLツール(第1版)の利用効果とその問題点

この4GLツール(第1版)により、「ロータス123」の操作・機能に制約されない計数処理の自動化を実現できた。このツールを文書file内表データに対する計数処理を中心とした業務に利用した時の効果と問題点を表5に示す。

NO	説明
利用効果	① 文書fileのフォーマットに対する論理テーブルとデータ処理を論理表現で明示的に指定されていて、処理内容が明確に分かる。 ② 一度4GLプログラムを作成しておけば、複数の文書fileに対するデータ処理が何度でも実行できる。また、その処理は個々にトランザクション処理(バッチ処理)である。 ③ 論理テーブルに対して、リンクが介在するような複雑な集計・分析が可能で、リンク記述の再利用ができる。 ④ 「ロータス123」上の初期値のデータ型(数値・文字列・NULL)を、プログラム属性として明確化でき、キャスト(行・桁)が介在しないテーブル(論理テーブル)を作成・処理できる。
問題点	① 構文解析に時間がかかり過ぎる。 ② 文書fileとのインタフェースが関数引数での行NO・桁NO指定であり定義に時間がかかる。 ③ 「ロータス123」の場合、文字列と数値の混在する同一ファイルのデータが、よく存在する。型の混合が許されていないため、エラー値が付きまとう。(多相型の変数をサポートしていない) ④ リンク記述や処理データ量が比較的多くなると、メモリオーバーが起こる。

表5. ツール(第1版)の利用効果と問題点

5. 4GLツール(第1版)の改良

言語仕様とその実行制御仕様、及び、4GLインタプリタの実行性能をふまえて、改善すべき点を表6に示す。

NO	改良すべき点
1	ELL(2)型解析処理の高速化(パス選択時の字句2度読み回避など)
2	動的メモリ管理を導入する。
3	表2の5つの機能項目を考慮し、言語仕様の問題点を改良する
4	簡易なプログラムに適した論理構造の適応方法とその解析方法を検討し、より複雑な論理構造の抽出を考える。

表6. 4GLツール(第1版)の改良すべき点

6. おわりに

本稿では、「ロータス123」などの簡易プログラムと連携のとれた文書file処理用4GLツールの機能の検討と開発運用している4GLツール(第1版)の評価を行い、その利用効果と問題点を述べた。この中で、表2に示す文書file処理用4GLツールが持つべき5つの機能は、十分高等な機能である。これらを実現する手段として、最善な言語仕様を確立していく方法と、言語処理系をユーザが構築(または、介入)できる機能を充実していく方法とがある。筆者は後者をとるが、どちらがよいかは、インド・コンピュータ環境の問題である。

今後は、分散DBMS環境(SYBASEなど)下での文書業務における4GLツール(文書file処理用)の機能と利用形態を検討していきたい。

[参考文献]

- [1]大須賀節雄:「データベースと知識ベース」,オーム社, P73-
- [2]J. C. Cleaveland:「データ型序説」,共立, P189-P194
- [3]情報処理学会:「情報処理ハンドブック」, P920-P923
- [4]「LOTUS 123 ユーザーズガイド」