

U t i l i s pを用いた階層データ入出力プログラム

長 田 弘 康 (国鉄・鉄道技術研究所)

1. はじめに

L i s p言語の特徴として、v e c t o rなどを除けばオブジェクトをそのまま入力または出力できるのでプログラムの開発やデバックが簡単にすることがある。しかし、一般のユーザがL i s pのS式を扱うことは抵抗があるので、ならかのS式を取扱い易い形式に変換したうえで入出力のできるプログラムを作る必要がある。このようなプログラムを個々のエキスパート・システムなどのアプリケーションに対して開発をおこなっていたのではマン・パワーの問題や開発時間に問題がある。以上の問題点を解決するためS式のサブセットである階層データの定義を行い、プロトタイプ・システムの為の階層データの入出力を行うプログラムを開発をおこなった。

2. 全体の構成

この入出力プログラムでは、各ルーチンを使用する前に階層データ型のあらかじめ行い、その定義を利用して各ルーチンが動作する形式をとる。階層データ型の定義(D E F T Y P E)をしたものは、まず定義のチェック(T T E S T)をおこない、端末からのデータの入力(T R E A D)、端末へのデータの出力(T P R I N T)、修正(T E D I T)、ファイル入力したS式などの検査(D T E S T)に利用される。

端末からの入力はフル・スクリーンで行うのが望ましいが、U t i l i s pの場合は大型計算機で使用されるのでC R Tの場合もテレタイプのように使用することを前提とした。ただし、使い勝手の面から、データを入力した後でY e s / N oというような確認入力をするのがなくキー数を少なくすることを目標とした。

3. 階層データの定義

3. 1. シンタクス

階層データの定義自身はS表現によって表わし、入出力プログラムがこの定義を参照して動作する。以下に文法をB N F記法によって表わす。

```
< 型 > : : =      ( < 名 前 >   < 体 > )  
                | ( < 名 前 >   < 体 >   < 暗 黙 値 > )  
< 体 > : : =      < 簡 易 体 > | ( < 複 合 体 > ) | < リ ス ト 体 >  
< 複 合 体 > : : =  < 簡 易 体 >   | < 簡 易 体 > < 複 合 体 >  
< 簡 易 体 > : : =  ( F I X N U M )  
                | ( F I X N U M   < 整 数 下 限 >   N I L )  
                | ( F I X N U M   N I L   < 整 数 上 限 > )
```

```

      | ( F I X N U M   < 整数下限 >   < 整数上限 > )
      | ( F L O N U M )
      | ( F L O N U M   < 実数下限 >   N I L )
      | ( F L O N U M   N I L   < 実数上限 > )
      | ( F L O N U M   < 実数下限 >   < 実数上限 > )
      | ( S Y M B O L )
      | ( S Y M B O L   < 記号列 > )
      | ( S T R I N G ) | ( N I L ) | < 外部名 >
< 記号列 > : : =   < 記号 >
                | < 記号 >   < 記号列 >
< リスト体 > : : = ( L I S T   < リスト型 > )
< リスト型 > : : = < 直列リスト型 > | < 直列リスト型 > ( < 直列リスト型 > )
                | ( < 直列リスト型 > )
< 直列リスト型 > : : = < 型 > | < 型 >   < 直列リスト型 >

< 名前 >、< 外部名 >、< 記号 >           ~   L i s p の シンボル
< 整数上限 >、< 整数下限 >               ~   L i s p の 整数
< 実数上限 >、< 実数下限 >               ~   L i s p の 浮動小数点数

```

上記で『～』は左辺が右辺と同等な文法によるという意味で用いている。

3. 2. セマンテックス

データ型の定義は、S式のサブ・セットに対応しており、データ型を用て入出力されるのはLispのオブジェクトである。そこで、データ型の定義に対応するオブジェクトについて説明する。

まず、シンタックスで用いた用語の説明をする。

- (1) <型>はデータ型全体を表わす。
- (2) <名前>はデータ型を識別するのに用いる。
- (3) <体>はデータ型本体を表わす。データの階層化の方法、データの種別および、データの有効範囲を与える。後で詳しく述べる。
- (4) <暗黙値>は、<体>が<簡易体>または<複合体>のときのみ有効であり、<体>の仕様を満足するデータを置く。<暗黙値>は、データの入力時の手間を少なくする為に用いる。

次に、<体>に対応するLispのオブジェクトについて述べる。

(a) 簡易体

- (a. 1) (F I X N U M) : 整数に対応する。
- (a. 2) (F L O N U M) : 浮動小数点数に対応する。
- (a. 3) (S Y M B O L) : シンボルに対応する。
- (a. 4) (S T R I N G) : スtringに対応する。
- (a. 5) (F I X N U M < 整数下限 > N I L) : 整数下限以上の整数が対応する。
- (a. 6) (F I X N U M N I L < 整数上限 >) : 整数上限以下の整数が対応する。

- (a. 7) (FIXNUM <整数下限> <整数上限>) : 整数下限以上整数上限以下の整数が対応する。
- (a. 8) (FLONUM <実数下限> NIL) : 実数下限以上の浮動小数点数が対応する。
- (a. 9) (FLONUM NIL <実数上限>) : 実数上限以下の浮動小数点数が対応する。
- (a. 10) (FLONUM <実数下限> <実数上限>) : 実数下限以上実数上限以下の浮動小数点数が対応する。
- (a. 11) (SYMBOL <記号1> <記号2> . . . <記号n>) シンボル<記号1>, <記号2>, . . . <記号n>のいずれかのシンボルが対応する。
- (a. 12) (NIL) : オブジェクトが対応する。
- (a. 13) <外部名> : <外部名>と同じ名前を持つデータ型の体が引用される。
- (b) 複合体
 <体>が(<簡易体1> <簡易体2> . . . <簡易体n>)の場合には、<簡易体1>から<簡易体n>のいずれかに対応するオブジェクトが対応する。
- (c) リスト体
 (c. 1) (LIST <型1> <型2> . . . <型m>) :
 <型1>で定義されるオブジェクトから<型m>で定義されるオブジェクトを要素とするリストが対応する。
- (c. 2) (LIST <型1> <型2> . . . <型m>
 (<型1> <型2> . . . <型n>)) :
 <型i>で定義されるオブジェクトを e_i と書き、<型j>により定義されるオブジェクトを e_j と書くと、 e_1 から e_m と0個以上の e_1 から e_n の繰り返しからなるリストが対応する。
- (c. 3) (LIST (<型1> <型2> . . . <型n>)) :
 <型1>で定義されるオブジェクトから<型n>で定義されるオブジェクトの1個以上の繰り返しからなるリストが対応する。

入出力プログラムの関係上、データ型の定義に以下の制約条件がある。

- (ア) 再帰的な構造はできない。つまり、外部名を用いて自分自身を引用することはできない。
- (イ) リスト体(外部名によって引用される体を含めて)は、暗黙値を取ることはできない。
- (ウ) 複合体では、リスト体(外部名によって引用される体を含めて)と複合を取ることはできない。
- (エ) リスト体の要素では暗黙値を持つ要素以後では必ず各要素は暗黙値を持たなければならない。

データ定義の例： 年月日は以下のように定義される。

```
(DATE (LIST (YEAR (FIXNUM))
              (MONTH (FIXNUM 1 12))
```

(DAY (FIXNUM 1 31)))

4. データ型の定義の方法

データ型を入出力プログラムで利用する為には、あらかじめデータ構造を貯える必要がある。

DEFTYPE (Macro) name body . default

名前が<name>、体が<body>、暗黙値が<default>とするデータ型を定義する。

CLEARTYPE

データ型の定義を全て空にする。

ERASETYPE name

名前が<name>のデータ型を消去する。

以上の様なプログラムが準備されている。データ型は特定のシンボルの値としてペア・リストの形で貯えられているのにすぎない。データ型の数が増えた場合にはハッシュ化などの必要があるだろうが現状では今の方法で十分である。

なお、データ型のシンタックス及び制約条件のチェックは以下のプログラムにより行い個々の入出力ルーチンではチェックはおこなわない。

TEST (Macro) name

5. 会話的入力プログラム

定義されたデータ型を参照して端末から会話的にデータを入力するプログラムであり、データ型を参照することにより、その構造を満たすデータのみが入力される。このチェック機構があるためにアプリケーションでは処理のみを集中的に作成すればよいことになる。

TREAD name

名前<name>に対応するデータ型に従ってデータの入力を行い、そのデータを値として返す。

データ型の定義から入力時のプロンプトなどが生成されるがそれだけではデータの意味がわからなくなる場合があるので、簡易体については『?』を入力することにより、その簡易体に付随したメッセージが出力される機構を準備した。

DEFCOMMENT (Macro) name . comlists

データ型の名前<name>に対して、コメント・メッセージ<comlist>を与える。コメント・メッセージはストリングでなければならない。

前に示した年月日のデータ型を例にとり T R E A D の操作例を示す。

```
・データ型の定義
> (DEFTYPE DATE (LIST
>                     (YEAR (FIXNUM))
>                     (MONTH (FINUM 1 12))
>                     (DAY (FIXNUM 1 31))))
((DATE (LIST ???)))
・コメント・ストリングの定義
> (DEFCOMMENT MONTH "1 --- Jan"
>                     "2 --- Feb"
>                     ".....")
("1 --- Jan" "2 --- Feb" ".....")
・T R E A D によるデータ入力
> (TREAD 'DATE)
DATE :: I ::
.YEAR:<<(FIXNUM)>>=I=> 1984
.MONTH:<<(1--12)>>=I=> Feb
.MONTH:<<(1--12)>>?=I=> ?
      1 --- Jan
      2 --- Feb
      .....
.MONTH:<<(1--12)>>?=I=> 3
.DAY:<<(1--31)>>=I=> <
.MONTH:<<(1--12)>>I3I=I=> 2
.DAY:<<(1--31)>>=I=> 14
(1984 2 14)
```

図 1 T R E A D の実行例

図 1 に示す様に T R E A D プログラムでは、データ型の内容を参照してプロンプトの生成及び入力されたデータのチェックを行っている。また、T R E A D では、単にデータのを入力をするだけではなく『<』などの文字によりデータの打ちまちがいを修正する機能をもつ。『<』などの特殊文字の機能としては以下のようなものが使用できる。

(1) 暗黙値呼出し 『#』: 暗黙値が定義されている場合、『#』を入力すると暗黙値を入力したのと同じ効果がえられる。

(2) 逆もどし 『<』: 入力点を一つ前に戻す。

(3) 繰り返しの切除 『!』: リスト型の繰り返し部分を切除する。

(4) コメント・メッセージの表示 『?』

(5) リスト型の入力終了 『)』: 繰り返し部分や固定部分の途中でそれ以後の要素に暗黙値がある場合にはその値が補充される。また、繰り返し部分の初めの場合には当該リスト型の入力の終了を意味する。

(6) 入力の終了 『≠』: 『<』で入力位置を変更したときに用いる。

注: リスト型の繰り返しの追加の機能は特殊文字ではなくて『<』により繰り返し部分の先頭に戻ると自動的に追加モードになる。繰り返しの追加をスキップするには単に空行を入力すればよい。

入力の手間を省く為に空行の入力により『<』により既にデータの入力されている所では前のデータが、暗黙値の定義がある要素ではその暗黙値を入力するのと同じ意味を持つ。つまり、空行が暗黙動作に対応している。また、1 行内の多くのデータ要素の入力は特殊文字を含めて空白で区切ることにより可能である。(これは、T R E A D ルーチンにより実現した機能ではなく U t i l i s p の入力機構による。)

T R E A D プログラムは、データに関するスタックとデータ型に関するスタックの両方を持ち、入力された内容に従って2つのスタックを操作するプログラム

である。面倒なのは、リスト体をもつデータ型のときの繰り返しに関するスタック操作である。

6. 製書出力プログラム

データ型の定義に合ったデータを端末に出力するプログラムを用意している。

T P R I N T n a m e d a t a

名前<name>に対応するデータ型に従ってデータ<data>の出力を行う。

```
> (TPRINT 'DATE '(10984 2 14))  
DATE  
.YEAR:1984    .MONTH:2    .DAY:14
```

図2 T P R I N T の実行例

データの出力で問題となるのはリスト体をデータ型が持つ場合である。T P R I N Tでは、リスト型の名前及びレベルや繰り返しの数などの付帯情報を印字したのちに改行をしてリスト体に対応するデータの内容を印字するの

が基本的ルールである。リスト体の内容に印字については、もし全ての要素が基本体を持つならば、1行になるべく多くの要素について印字をし、要素のなかに基本体以外の体を持つ型がある場合には各要素の印字ごとに改行をする。

7. 会話的修正プログラム

データを端末から会話的に修正する為のエディタを準備した、データ型を参照することにより、その構造を満たすようにデータを修正する。このチェック機構があるためユーザの作成するプログラムの負荷が低減される。

T R E R E A D n a m e d a t a

名前<name>を持つデータ型に従ってデータ<data>を再入力して、変更されたデータを値として返す。

T R E R E A DプログラムはT R E A Dルーチンを変更モードにすることにより実現しているので、操作の方法についてはT R E A Dルーチンと同じである。

T E D I T n a m e d a t a

名前<name>に対応するデータ型に従ってデータ<data>を会話的に修正を行い、の修正されたデータを値として返す。

T E D I Tルーチンは、コマンド式のエディタであり、コマンドはL i s pのシンボルからなる修正を行うためのものと、整数から成る対象データを詳細化するためのものがある。

整数コマンドは、現在の修正対象がリスト体に対応しているときに有効であり、整数コマンドで示される番号の要素に修正の対象を移動する。要素と整数の対応は、コマンド『P』または『?』により情報出力させる。

シンボル・コマンドは修正の対象となっているデータ型と上位のデータ型により有効であったり、無効であったりするが、コマンド入力時のプロンプトに有効なコマンドが全て表示されるので、ユーザは、有効なコマンドがなんであるのか

```

> (SETQ DATE-1 '(1984 2 14))
(1984 2 14)
> (TEDIT 'DATE DATE-1)
DATE: in EDIT mode
COMMAND=<?/P/PP/Q>===> 2
.MONTH: in EDIT mode
COMMAND=<?/</>/¥/S/L/U/P/PP/R/Q>===> R
.MONTH:<<1--12>> [2]=R=> 5
.MONTH: in EDIT mode

?: "Help"                                <: "Move Before Element"
>: "Move Next Element"                  ¥: "Go to Toplevel"
S: "Move Start Element"                  L: "Move Last Element"
U: "Scope Level up"                      P: "Print Now Element"
PP: "Precise Print Now Element" R: "Replace Now Element"
Q: "Quit Edit"
SCOPE DATA=5

.MONTH: in EDIT mode
COMMAND=<?/</>/¥/S/L/U/P/PP/Q>===> L
.DAY: in EDIT mode
COMMAND=<?/</>/¥/S/U/P/PP/R/Q>===> R
.DAY:<<1--31>> [14]==R=> 7
.DAY: in EDIT mode
COMMAND=<?/</>/¥/S/U/P/PP/R/Q>===> ¥
DATE: in EDIT mode
COMMAND=<?/P/PP/R/Q>===> PP
DATE:
.YEAR:2 MONTH:5 DAY:7
DATE in EDIT mode
COMMAND=<?/P/PP/R/Q>===> Q
(1984 5 7)

```

図3 TEDITの使用例

を知ることができる。以下にシンボル・コマンドの種類と機能をしめす。

- | | |
|---------------------|--------------------|
| (1) 『?』: 詳細な情報の表示 | (2) 『<』: 前の要素への移動 |
| (3) 『>』: 次の要素への移動 | (4) 『!』: 消去 |
| (5) 『¥』: 最上位レベルへの移動 | (6) 『S』: 初めの要素への移動 |
| (7) 『L』: 最後の要素への移動 | (8) 『U』: レベルを1つあげる |
| (9) 『P』: 表示 | (10) 『PP』: 詳細な表示 |
| (11) 『R』: 修正 | (12) 『I』: 入力 |
| (13) 『A』: 追加 | (14) 『Q』: 終了 |

現在の T E D I T ルーチンでは、コマンドの取消し (u n d o) や要素の複写などが無いという問題点がある。

8. データ検査プログラム

ファイルなどから入力したデータを、データ型を参照することにより、その構造を満たすかを検査し、誤りがある場合にはエラー・メッセージを出力し、修正可能なものは、データを変更する。

D T E S T n a m e d a t a

名前 < n a m e > を持つデータ型を参照して、データ < d a t a > の検査を行い、修正されたデータを返す。

データの中に特殊文字『井』があるときには、データ型に定義された暗黙値を書くのと同様な動作をし、またリスト型において対応するデータの数が少ない場合は、それ以降のデータ型に暗黙値が定義されているときはデータに暗黙値が補充される。

データ検査ルーチンでは、対象となるデータとデータ型をマッチングするのが主な内容である。マッチングをする場合、データ型が基本体をもつ場合には簡単に表 1 のようにデータ型との対応できまる評価値を用いて行う。評価値が 4 の場合にはデータが受理される。評価値が 3 の場合にはデータは受理できないが、類似のデータに変換をする (表 1 では変換データとして表わしている)。評価値が 2 以下の場合にはデータが変換できない場合でありエラーとなる。

データ型がリスト体をもつ場合には、データがアトムの場合にはエラーになる。データもリストの場合にはリストの要素とリスト型の要素となる型で局所的に処理を次のように行う。

(1) 処理の対象となるリストの要素とリスト型の要素の評価値が 4 の場合にはその要素を受理して次の要素を処理する。

(2) それ以外の場合には、まず次の 3 つの評価値を計算する。

V_{21} = 現在の処理対象となっているデータと現在のデータ型の評価値。

V_{12} = 現在の処理対象となっているデータと次のデータ型の評価値。

V_{22} = 次の処理対象となるデータと次のデータ型の評価値。

(2 . 1) 評価値の最大値が V_{12} であり、それが 3 以上の場合には、現在のデータがたらないものとして \$ E R R O R \$ を受理したものとする。次の処理対象をデータに対しては変化させずに、データ型については 1 つ前に進める。

(2 . 2) 評価値の最大値が V_{21} であり、それが 3 以上の場合には現在の処理対象になっているデータが余分にあるものとして、次の処理対象をデータに対しては変化させずに、データ型については 1 つ前に進める。

以上がリストに対する基本的な処理の方法であるが、他に暗黙値の埋め込みの処理などがある。このアルゴリズムは局所的に処理を行うので、リスト体を構成する型に変化が多くなると誤りの含んだデータに対して必要以上のエラーが検出されてしまうことになる。これを回避する方法としてはリスト全要素についての大域的な判断をすることなどが考えられる。

表1 DTESTで用いるマッチングの評価値

データ型の体	データ	評価値	変更データ
(NIL)	任意	4	----
(FIXNUM)	整数 浮動少数点数 任意	4 3 1	---- 整数にしたもの ----
(FLONUM)	浮動少数点数 整数 任意	4 3 1	---- 浮動少数点数にしたもの ----
(SYMBOL)	シンボル 任意	4 1	---- ----
(STRING)	ストリング シンボル 任意	4 3 1	---- ストリングにしたもの ----
(FIXNUM imin NIL)	整数 $\geq$ imin 浮動少数点数 $\geq$ imin 数値 任意	4 3 2 1	---- 整数にしたもの 整数にしたもの ----
(FIXNUM NIL imax)	整数 $\leq$ imax 浮動少数点数 $\leq$ imax 数値 任意	4 3 2 1	---- 整数にしたもの 整数にしたもの ----
(FIXNUM imin imax)	imin $\leq$ 整数 $\leq$ imax imin $\leq$ 浮動少数点数 $\leq$ imax 数値 任意	4 3 2 1	---- 整数にしたもの 整数にしたもの ----
(FLONUM amin NIL)	amin $\leq$ 浮動少数点数 amin $\leq$ 整数 数値 任意	4 3 2 1	---- 浮動少数点数にしたもの 浮動少数点数にしたもの ----
(FLONUM NIL amax)	浮動少数点数 $\leq$ amax 整数 $\leq$ amax 数値 任意	4 3 2 1	---- 浮動少数点数にしたもの 浮動少数点数にしたもの ----
(SYMBOL . symbols)	シンボル $\in$ sinbols シンボル 任意	4 2 1	---- ---- ----

9. あとがき

エキスパート・システムなどのLispの応用プログラムを効率良く作成する為に、入出力ライブラリ・プログラムを作成した。本ライブラリを使用すること

により入出力はデータ仕様を書くだけで済むことになる。

本入出力ライブラリ・プログラムについて改善すべき点について述べる。

- (1) データ型の種類が少ない。特にプロダクション・システムのルールのような構文をもったデータは扱いにくい。
- (2) 入力が行単位なので、使い勝手が悪い。今後はフル・スクリーン又はビット・マップディスプレイでの操作が可能なものを考えていく予定である。
- (3) 複合的なデータの検査ができない。つまり、『月日』のようなデータ型について、『月』が1から12ということは検査できるが、4月31日というデータをリジェクトする機能はない。現状では、本入出力ライブラリの外部でチェックする以外に方法はない。この複合的なデータ検査機能は望ましいが(2)の問題を解決しないと入力時にエラーをプログラムが発見した場合にどこを修正するかを指定しなくてはならない。『月日』のように単純ではなくて検査が複数の要素にわたる場合は更に困難な状況がありうる。

10. 文献

- (1) Chikayama, T. : " Utilisp Manual " , MET R 81-6 , Department of mathematical engineering , University of Tokyo , 1981
- (2) 長沢孝次、長田弘康、後藤和則 : " 有機物質の赤外線吸収スペクトラム解析の知識工学的手法の応用 "、情報処理学会28回大会、1059-1060 (1984)
- (3) 長田弘康、長沢孝次、後藤和則 : " 赤外線吸収スペクトラム解析プログラムの構成方法 "、情報処理学会28回大会、1061-1062 (1984)
- (4) 長田弘康 : " Lispを使った入出力プログラム(その1)"、鉄道技術研究所速報A-84-122 (1984)