

1B-3

『三才児』における表からの知識生成

稲葉 光行, 藤井 則夫  
富士通株式会社 システム本部

1. はじめに

エキスパートシステム (ES) の知識ベース構築作業においては、知識表現と知識獲得が大きな制約となっている。知識表現における制約の原因の一つは、エンドユーザが情報間の階層関係や因果関係を知識処理に適した形で整理できない点である。また知識獲得に関しては、システムへの情報入力が基本的に手作業であり、作業に掛かる工数や、入力された情報の信頼性の点で一部制約がある。

一方、ワードプロセッサやデータベースの普及により、日常の業務における様々な情報が電子化されている。それらの中には、ESの知識ベースにふさわしい多くの貴重なノウハウも含まれている。

そこで、エンドユーザが日常行っている情報管理の形態に近い方法で知識を整理すると共に、そこからプログラムによって自動的に知識を生成することができれば、これらの電子化された知識やノウハウを有効に活用でき、知識獲得や知識表現の制約も軽減される。

このような考えから、まず知識ベースにふさわしいノウハウや知識が含まれている表形式のデータに着目し、そこから当社のコンサルテーションモデル構築ツール「三才児」<sup>1)</sup>のフレーム型知識ベースを生成するツールを開発した。以下に、ここで用いた知識生成の方法とツールの概要を述べる。

2. 表による知識表現

現在、知識獲得の負担を軽減するために、ワークシートに従って知識を整理する方法<sup>2)</sup>や、計算機との対話を通して知識ベースを構築する試み<sup>3)</sup>が行われている。しかし大量の情報を入力する場合には、やはりユーザの負担が大きい。

そのため我々は、多くの業務で既に大量の情報が蓄積されている表形式の文書データから、直接知識ベースを生成する方法を取った(図1に例を示す)。そして、知識ベースに変換可能な表の雛形と最低限の規約を示し、利用者がそれを参考に、自

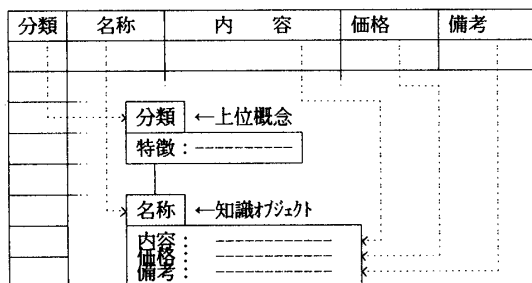


図1 表と知識ベースとの対応例

分の業務に応じた知識ベースを構築できるようにした。以下、我々が利用者に示した表の雛形と規約の概要を述べる。

2.1 既存文書における表の分類

既存の文書の中で、知識が整理されている表の形式を分類するため、社内で作成されたワークシート、会議の議事録、そして情報処理学会及び人工知能学会の全国大会論文集から、ランダムに50個の表のサンプルを収集した。そして以下に示す3つのパターンで、サンプルの表全体の約80%カバーできることが判明した。ここではこの3つを、クロス型、階層型、チェック型と呼ぶことにする。図2はこの3パターンの形態を示したものである。

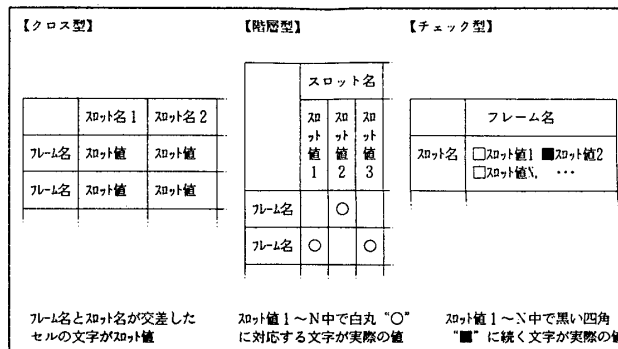


図2 表の3パターン

クロス型は、単純な2次元配列としてデータが整理された、最も標準的な表の形式である。この形式では、表の上部に並んでいる見出しと、表の側面に並んでいる見出しとが交差した部分(ここではセルと呼ぶ)の文字を、フレームのスロット値として取り出すことができる。

階層型とは、クロス型の形式に加えて、表の上部または側面にある見出し部分が、実際の値とそのカテゴリに階層化されている表である。この表の場合、階層化された見出しの下位の単語中で、上部見出しと側面見出しが交差したセル中の“○”等に対応する単語(複数個でもよい)がスロット値となる。スロット名は階層化された見出し部の上位から取り出せる。

チェック型は、基本的にはクロス型であるが、セルの中の文字または数値が複数個あり、さらにそれらの前に、実際に値となるものをチェックするための欄がある表である。このチェック欄の中に、何らかの記号が入っているもの(複数個選択されていてもよい)がスロット値となる。

2.2 表のパターンにおける制約

上記の3つのパターンでカバーできなかった表としては次の

ようなものがあった。

- ①セルの区切り記号を特定できないもの
- ②枠外の注釈等で、表中の記号の意味を示しているもの
- ③図形やイメージを含むもの

①については、人間であれば簡単に判別できるかもしれないが、プログラムで自動的に処理することは困難である。②については、記号の意味を判定するための辞書を持った個別プログラムで対処するか、手作業で表を修正する必要がある。③についても、図形データがツールや機種によって異なるため、汎用化は困難である。

②や③の表の場合、利用者が手作業で修正する必要が出てくる。この負担を少しでも軽減するため、修正作業のログをファイルに保存できるエディタと知識生成ツールを組み合わせ、以前の手続きを利用することで、2回目以降は自動的な修正を可能にするという方法を検討している。

### 3. 知識生成ツールの概要

#### 3.1 知識生成ツールの構成

既存文書中の表を分析し、上記の3パターンを抽出した後、これらの表から知識ベースを生成するツールを作成した。以下にこのツールの概要を述べる。

図3に示すように、知識生成ツールは「ユーザ応答定義部」「情報抽出部」及び「フレーム生成部」の三つに分かれる。

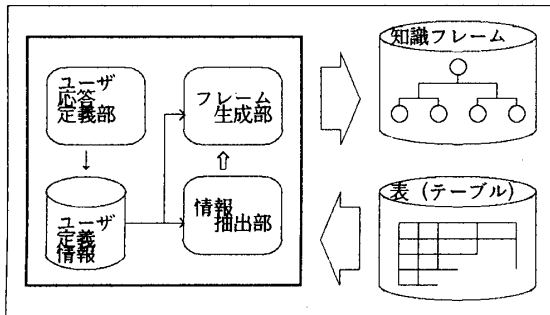


図3 知識生成ツールの構成

#### 3.1 ユーザ応答定義部

知識生成ツールでは、表中の単語の意味的な階層関係や、表の形式に対して解釈をせず、それらの情報はユーザが指定するという方法を取っている。この情報は、知識生成ツールのユーザ応答定義部との会話を行うことで容易に作成できる。この時ユーザは、管理対象(オブジェクト名)と属性(スロット)及び上位概念(クラス名)に該当する欄を指定すればとりあえず知識ベースを生成できる。また、特定の欄を無視したり、スロット名を変更することも可能である。図4は、ユーザ応答定義部によって生成されたユーザ定義情報の例である。

表の方向	縦	横
見出し	人部	の表
1	無	名
2	ク	ス
3	ラ	ク
4	ス	ス
5	ス	ス
6	ス	ス
7	ス	ス
8	ス	ス
9	ス	ス
10	ス	ス
11	ス	ス
12	ス	ス
13	ス	ス
14	ス	ス
15	ス	ス
16	ス	ス
17	ス	ス
18	ス	ス
19	ス	ス
20	ス	ス
21	ス	ス
22	ス	ス
23	ス	ス
24	ス	ス
25	ス	ス
26	ス	ス
27	ス	ス
28	ス	ス
29	ス	ス
30	ス	ス
31	ス	ス
32	ス	ス
33	ス	ス
34	ス	ス
35	ス	ス
36	ス	ス
37	ス	ス
38	ス	ス
39	ス	ス
40	ス	ス
41	ス	ス
42	ス	ス
43	ス	ス
44	ス	ス
45	ス	ス
46	ス	ス
47	ス	ス
48	ス	ス
49	ス	ス
50	ス	ス
51	ス	ス
52	ス	ス
53	ス	ス
54	ス	ス
55	ス	ス
56	ス	ス
57	ス	ス
58	ス	ス
59	ス	ス
60	ス	ス
61	ス	ス
62	ス	ス
63	ス	ス
64	ス	ス
65	ス	ス
66	ス	ス
67	ス	ス
68	ス	ス
69	ス	ス
70	ス	ス
71	ス	ス
72	ス	ス
73	ス	ス
74	ス	ス
75	ス	ス
76	ス	ス
77	ス	ス
78	ス	ス
79	ス	ス
80	ス	ス
81	ス	ス
82	ス	ス
83	ス	ス
84	ス	ス
85	ス	ス
86	ス	ス
87	ス	ス
88	ス	ス
89	ス	ス
90	ス	ス
91	ス	ス
92	ス	ス
93	ス	ス
94	ス	ス
95	ス	ス
96	ス	ス
97	ス	ス
98	ス	ス
99	ス	ス
100	ス	ス

図4 ユーザ定義情報の例

#### 3.2 情報抽出部

知識生成ツールは、テキストデータの形になっている表をディスクから読み込んだ後、第2章で述べた3つのパターンの表のセルからデータを切り出し、文字列のリストに変換する。この際、セルの範囲を特定するための区切り記号(デフォルト値は罫線)をユーザ定義情報から参照し、複数行に渡っているデータの連結処理等を行う。

#### 3.3 フレーム生成部

情報抽出部によって生成されたリスト構造のデータは、フレーム生成部の中で、ユーザ定義情報に従った知識フレームに変換される。知識階層の深さについては、ユーザによる定義の形式をできるだけ簡略化するために、2階層、つまり知識オブジェクトと、クラスのみを指定を許している。ただし、クラスの欄を複数個指定することもでき、マルチインヘリタンスが可能となっている。

また「三才児」自体では階層の数が制限されていないため、「三才児」に知識を取り込んだ後、マウス編集が可能なブラウザによってさらに深い階層化ができるようになっている。

### 4. 評価

このツールの効果を調べるため、ワープロ文書として作成された社内のハンドブックの表から、まず手作業で「三才児」に知識を入力し、その後ワープロ文書の形で保存されている同じ表のデータから、ツールによる知識ベース生成を行った。

結果として、知識ベースの作成時間を10分の1以下に短縮することができた。また、既存の文書データがそのまま取り出されるため、知識ベース中のデータをチェックする手間も省けた。これらのデータからツールの効果をまとめると、以下のようになる。

- ①既存の文書データ中の表をそのまま知識ベース化できる。
  - ②知識入力時間を大幅に短縮できる
  - ③入力ミスの可能性が減り、知識の信頼性が向上する。
- また、難形と制約条件に従った表を作成すれば、そのまま知識ベースに変換できるため、
- ④エンドユーザでも知識ベースを容易に構築できる
- ようになったと言えるだろう。

### 5. おわりに

今回開発したツールは、「三才児」の知識ベースのみをターゲットとしている。しかし「三才児」の知識フレームは、他のESの知識フレームにも変換しやすいものであるため、今回の方法をもとに、今後は他ツールのフレーム生成、さらにプロダクションルールの生成等も考えて行きたい。

#### 【参考文献】

- 1) 藤井, 銀林: "三才児をモデルにした知識自動獲得システムの応用例と試行結果", 情報処理学会大35回全国大会, 1987
- 2) 松本, 太田: "エキスパートシステム構築方法論-ES/SDEM-の開発", 人工知能学会全国大会(第1回), 1987
- 3) 藤堂, 松本: "エキスパートシステム構築アドバイザー/SAKAS", 人工知能学会全国大会(第2回), 1988