

4H-5

チェック・訂正問題を対象とした
段階的知識ベース生成支援方式

山本 佳代子, 渡辺 敏
NTT情報通信網研究所

1 はじめに

エキスパートシステムの特徴の一つとして知識の追加・修正の容易性が挙げられる。しかし、専門家の持つ概念的な知識のレベルと現状ツールによって提供される知識表現レベルとの間には大きなギャップ(以下、知識表現ギャップと呼ぶ)があり、専門家による知識の追加・修正を困難としている。知識表現ギャップを解消し、知識の追加・修正を容易とするためには、ツール知識表現の高度化、或いは知識表現ギャップを埋める支援機構が必須と思われる。

筆者らは、「専門用語の持つ概念的背景を整理した辞書が有れば、日本語表現から既存のツール知識表現への変換が可能ではないか」との着想を得、具体化した。

本報告では、専門用語の意味付けを主体に、知識整理過程を段階的に実施すると共に、形式化した日本語で記述した知識から既存ツールの知識表現に変換することを骨子とする、段階的知識ベース生成支援方式を提案する。

2 知識表現ギャップ

知識表現ギャップの存在は、専門家の持つ概念的な知識を計算機で処理するため、モデル化及び形式化を実施していることによる。すなわち、仮想的な世界の枠組みをつくることから、この世界に関する約束を全て理解しなければ、知識の与え方が分からないことになる。このギャップを解消する作業は実際のES開発では、知識処理技術や開発経験を基にKEが実施している。

知識表現ギャップをKEに依らずに解決するためには、知識表現レベルの高いツールが必要で有るが、現技術レベルでは難しい。知識表現ギャップを段階的に埋めていく支援機構を設けることが妥当と考えられる。

3 基本的な考え方

ESの開発経験から、専門用語が単に概念の整理や体系化に有効なだけでなく、知識表現においても重要な意味を持つため、知識表現ギャップを埋めるための支援機構に利用できるのではないかと考えた。専門用語の利用法を以下に示す。

- ①利用に先立ち、専門家の持つ概念的知識体系を専門用語に着目して整理し、専門用語をキーとする辞書を作成する。辞書の構成は標準用語、用語間の相関、ツール知識表現との対応、体系化情報、等からなる。
- ②用語間の相関、体系化情報等を利用して、概念的な知識から曖昧性、矛盾等を取り除き、一意に解釈可能な表現とする。
- ③専門用語とツール知識表現との対応関係を組み合わせ、最終的に、日本語表現と既存ツールで提供される知識表現とを対応付ける。

上記アプローチにより、専門家による知識の追加・修正および辞書による知識の共有化が可能になる。

4 段階的知識ベース生成支援方式の提案

4.1 概要

知識獲得の流れと本手法の適用範囲及び作業手順を図1に示す。本手法は、「概念化」までが完了し、知識表現が既知であることを前提に、「定式化」および「実現」の各作業段階において、専門家を順次支援し、既存ツールの知識表現形式に変換し、最終的に知識ベースを生成する。

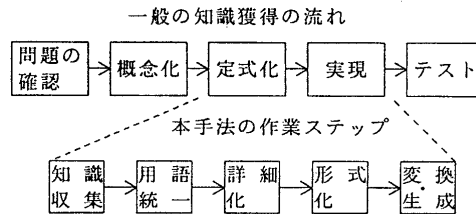


図1 知識獲得の流れ

本手法の各作業段階の実施目的を表1、処理イメージを図2に示す。

表1 各作業段階の実施目的

工程	目的
知識収集	知識ベースに投入する知識を収集する。
用語統一	曖昧性がなく的確に概念を表すことができる専門用語を整理、体系化する。
詳細化	抽象度の高い知識表現をツールの知識表現レベルまで、詳細化する。
形式化	曖昧性及び多様性を持つ日本語表現を一意に解釈できるように限定した表現に置き換える。
変換・生成	形式化された日本語表現からツールの知識表現に変換後、知識ベースを生成する。

4.2 各段階における支援方式

段階的知識ベース生成支援方式の各段階の支援内容を以下に示す。

(1)知識収集

通常の知識収集作業とし、特に支援しない。ただし、機械的な用語抽出を可能とするため、定型フォーマット上に、日本語記述することが前提となる。

(2)用語統一

専門家が作成した日本語記述知識を入力として、用語抽出を行い、命名規則に従った標準的な名称に変換する作業を支援する。この過程で、新しく現れた用語については、属性、使用法の定義、上位/下位概念の明確化、類似語等の分類・整理を本ツールと対話形式で実施し、辞書に登録する。

最終的にツールの知識表現と結び付けるために必要な計算機上へのインプリメント情報を合わせて定義する。そのため、辞書は概念を表す論理辞書とインプリメント情報を与える物理辞書から構成する。尚、物理辞書はKEが作成し、必要な機能をオブジェクトとして与える。用語辞書の作成に関しては、データ資源管理⁽¹⁾の概念を参考とした。表2に辞書構成を示す。

表2 辞書の構成

構成要素	意味
標準名	用語の標準的な名称*
品詞区分	用語の品詞(修飾語/主要語/区分語)
上位概念	他用語との階層関係
下位概念	
類似語	標準名に統一される類似用語
省略語	標準名の省略形
属性	桁数、形式、編集形式等に関する情報
値範囲	用語が取り得る値の範囲
使用法	用語の使用法、内容に関する情報

*命名規則: [修飾語...] + 主要語 + 区分語

(3)詳細化

辞書に示される用語関係(上位/下位概念)を利用して、上位概念を示す用語を下位概念の用語の組み合わせに繰り返し置き換えていくことにより、ツールの知識表現レベルまで詳細化する過程を支援する。

専門家が計算機に支援されて作業するため、曖昧性が省かれと共に、均質な詳細化が保証される。

(4)形式化

自由に書かれた日本語を一意に解釈が可能な構文則の

範囲で記述し直す。本過程では、基本構文則の提示や、用語の使用法に関する誤りの検査等を支援する。

(5)変換・生成

形式化された知識表現を解析し、ツールの知識表現に変換すると共に、知識ベースを生成する。変換にあたっては、論理的な辞書と物理的な辞書との対応関係を用いることにより実現する。

4.3 評価

本提案はルールによる知識表現が用いられる分野に適用可能と考えているが、まず、知識が比較的単純で、かつ多いデータのチェック、及び訂正業務を対象に適用する予定である。ここでは、既開発のESに適用した場合を想定し、評価した。

①方式に対する評価

実際に、段階的な過程をふまえた知識整理を机上でシミュレーションした結果、専門家による知識追加・修正が可能であるとの見通しを得た。

②生産性に対する評価

オーダ票のチェック業務のルール約100個に対し机上で確認した結果、知識表現を形式化した日本語レベルまで上げたことで、既存ツールによる知識表現と比較し、記述量が約1/2となり、記述性に優れ、生産性が向上することを確認した。

また、KEを介した知識の整理および変換作業をなくすことにより、知識の誤解防止効果もあると予想される。

5. おわりに

知識表現ギャップを解消し、専門家による知識の追加・修正を容易とする段階的知識ベース生成支援方式を提案すると共に、机上評価結果について述べた。現在システム化を進めており、インプリメント法及び定量的評価については別途報告する。

参考文献

- (1) Durell, W. R.: Data Administration, pp. 13-139, McGraw-Hill(1985).

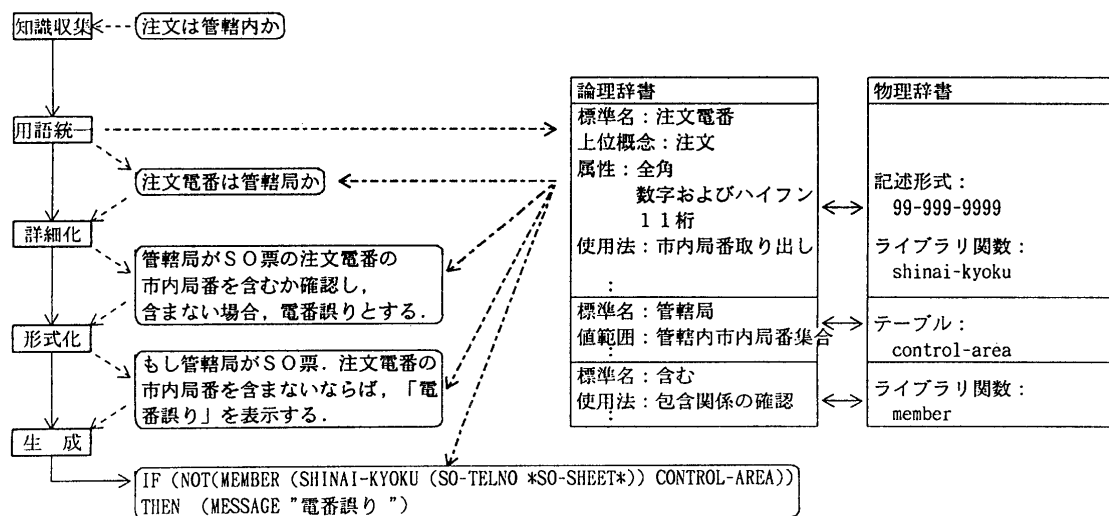


図2 処理イメージ