

分散知識ベースシステムにおける 個人用シソーラスの対応付け方式

3M-2

大場雅博, 伊藤英則
(財) 新世代コンピュータ技術開発機構

1. はじめに

第5世代コンピュータシステム研究開発プロジェクトでは、研究の一環として、推論マシンや知識ベースマシンがLANで結合された分散知識ベースシステムを対象として、知識ベースマシン分散モデルの研究開発を行なっている[1]。このような対象では、個人(或いは研究グループ)は、自分の仕事に必要なデータや知識を、自分の用語を用いて記述し、それらを自分の観点や目的にもとづいて分類整理する。推論マシンや知識ベースマシン上に独自に作成した、このような種類のデータや知識の集まりをパーソナル知識ベース(PKB)と呼ぶ。PKBは個別に維持管理されるのが通常である。しかし、それらを全体としてまとめて、仕事に活用したいという要求がある。本研究は、その実現に必要な分散制御技術や知識管理方式などの技術開発をおこなうことを狙いとしている。既に、PKB群を共有化するには各PKB中の用語の意味を相互理解することが重要と考え、そのための基本的枠組みを提案した[2][3]。本講演では、PKB間で用語同志の対応をとるための基本的考え方を報告する。

2. パーソナル知識ベースの共有化のための基本的考え方

PKB群を相互に使用するためには、次のような点が要求される。

- (1) 利用者は、自分の情報要求にとってどのPKBが有用であるかを判断するために、他者のPKBの内容や意味を理解できねばならない。
- (2) 利用者は、自分自身の語彙を用いて検索キーワードを記述できることが好ましい。

本研究では次のことを仮定する。各PKB作成者は、自分の概念体系や観点、問題意識などを表現しているパーソナルシソーラス(PT)を所有している。PKB内に格納されている知識やデータの各々に対し、その内容説明のために、自分のPT内の用語をいくつか選んで付与している。

用語の相互理解の問題は、利用者のPTと提供者のPTを照合しあい、意味的類似性を考慮して各PT内の用語同志を対応付けることにより解決できる。この方式をPTマッチングと呼ぶ(図1)。PTマッチングにより次のことが可能となる。

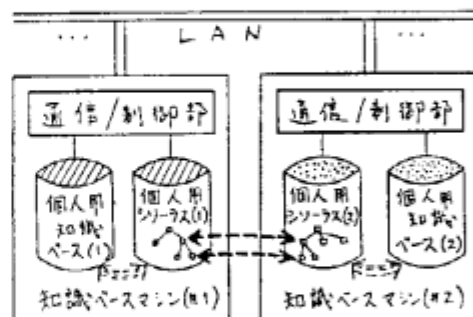


図1. 共有化のためのメカニズム

- (1) 利用者の問合せ中のキーワードが、提供者のPKBにアクセスするために、意味の点で適したキーワードが割り付けられる。
- (2) 提供者のPKBの格納物に付与されている分類語は、利用者のPT内の適切な分類語に変換される。利用者は自分自身の用語で、提供者のPKBの内容を理解することが出来る。

3. PTの特徴とPTマッチングのための要求機能

ここでは、利用者の検索要求に対し、何等かのやり方でその要求に適した分野の知識やデータを格納しているサイト群が選ばれているとし、同じ分野の分類語を有しているPT群を対象とする。A分野のPTの例を図2(PT1)と図3(PT2)に示す。図中、()でくくった用語は分類語間の関係を表現する関係語を示す。

このようなPT群は次のような特徴を持っている。

- (1) 異なるPTに、類似語が出現している。基本的な術語を核とした複合語が出現している。用語間で文字列(字面)が類似していたり、複合語同志が共通の単語を有しているからといって、それらの用語が意味的に類似しているとは限らない。
- (2) 同義語が出現している。
- (3) 同一の用語群に対して、人が違えば異なる分類構造を与えることがある。例えば、用語の上位/下位関係が逆転することもある。

上記の特徴により、PTマッチングのためには、次のような機能が要求される。

(1) 単語(ストリング)マッチング

同義語は、PKB作成者が入力するものとする。マッチングのための基本機能は、単一の単語及び複合語、同義語を取扱える単語のマッチング機能である。

(2) 意味的類似性の判定

単語マッチングを行うと、完全一致或は部分一致する用

Personal Thesaurus Matching Method for
Distributed Personal Knowledge Base System
by Masahiro OBA, Hidenori ITOU (ICOT)

人工知能<1>
 (課題) 人工知能構築環境<2>
 (実装) 人工知能本言語<3>
 ニキスパートシステム構築ツール<4>
 (評価) 知識表現<5>
 (例) フレーム型<6>/近接型<7>/ルール型<8>
 知識ベース<9>
 (格別) 事実知識<10>/ルール知識<11>/メタ知識<12>
 知識ベース管理<13>
 推論エンジン<14>
 利用者インタフェース<15>
 人工知能システム構築方法<16>
 (手順) 知識の獲得<17>
 知識の分析・整理<18>
 (目的) 無矛盾性<19>
 知識ベース構造設計<20>
 知識ベースの作成<21>

図2. PT1

語が多数出現する。このような用語に対し、意味的類似性を判定することにより、妥当な対応付けを絞り込めることが好ましい。

4. 意味的類似性

別々のPTに出現する用語の意味的類似性を判定するためには、その用語だけに注目しているだけでは不十分で、PT内の多くの用語を用い、多くのPT群を用いることが重要である。次に、意味的類似性を判定、評価する基準を述べる。

(1) 用語群のグループ性

別々のPTの2つの用語群グループにおいて、用語同志がお互いにマッチングする時、意味的に類似していると思われ。この時、分類構造は必ずしも同一である必要はない。この種の類似性を、グループ的類似性と呼ぶ。

(2) 用語の定義

グループ的類似性をより厳密に捉えたものであり、グループ的類似性の条件に対し、さらに、分類構造の同一さを追加した条件を満たす時、意味的に類似していると思われ。この種の類似性を定義的類似性と呼ぶ。

グループ的類似性と定義的類似性とは、分類構造が類似している場合に有用である。

(3) 用語間の性質の継承

1つのPTの中で、1つの用語は、直接の上位/下位にある用語だけではなく、隔れたところにある用語とも、幾つかの関係語を介して関連している。この関連を取り扱うために、各関係語に対し、性質の継承のモードを考える。性質継承モードとは、関係語をはさんだ上位概念と下位概念が、各々もう一方の性質を有しているかどうかを示した特性であり、①継承する、②継承するかもしれない、③継承しない、の3モードある。このモードは、PKB作成者により各関係語に付与される。直接の上位/下位ではない2つの分類語の間の性質継承モードは、1つの分類語からもう1つの分類語の方へ関係語を順にたどっていくことで、その種類と程度を推論することが出来る。それぞれ別々のPT上にある2組の用語群(仮に2つずつの用語とする)は、

AI<31>
 (課題) 実用化課題<32>
 (実装) 総合AIシステム構築環境<33>
 (構成要素) 知識表現機能<34>
 (例) 扁平的表現法<35>
 (例) フレーム型<36>
 論理型<37>
 ルール型<38>
 メタ知識<39>
 推論機能<40>
 ユーザインタフェース<41>
 基盤的研究課題<42>
 (項目) 知識ベースの構造<43>
 (例) 知識獲得支援システム<44>
 知識ベース設計手法<45>
 知識ベースの表現・管理<46>
 (例) 知識ベースの一貫性<47>

図3. PT2

各組間で用語同志が対応付けられ、かつ、各組内での用語間の性質継承モードが互いに類似している時意味的に類似であると見做す。この類似性を性質継承的類似性と呼ぶ。

5. PTマッチングの事例

事例を用いてPTマッチングのラフなイメージを与える。次のような状況を想定する：利用者はサイト1上にPT1を、提供者はサイト2上にPT2を持っており、利用者は「『述語型』『知識表現』と『メタ知識』による『無矛盾性』検出を取り扱っている『AIシステム構築方法』に関する論文」を探している。利用者は同義語として「人工知能」=「AI」や「無矛盾性」=「一貫性」等を与える。以下、〈X〉はPT1とPT2中の分類語の番号を示し、〈X〉〜〈Y〉は分類語〈X〉と〈Y〉とが対応付けられることを示すものとする。PTマッチングの結果として、〈16〉〜〈43〉、〈17〉〜〈44〉、〈19〉〜〈47〉、〈20〉〜〈45〉がグループ的類似性として、〈4〉〜〈33〉、〈5〉〜〈34〉、〈14〉〜〈40〉、〈15〉〜〈41〉、〈5〉〜〈35〉、〈6〉〜〈36〉、〈7〉〜〈37〉が定義的類似性として、〈4〉〜〈33〉、〈12〉〜〈39〉が性質継承的類似性として判定される。この結果、サイト2は〈35〉、〈37〉、〈39〉、〈47〉の用語をキーワードとしてPKBの検索を行う。

6. おわりに

個人用シソーラス対応付けのために、用語同志の意味的類似性の判定基準の基礎検討を行なった。今後は、この対応付けを通信/制御により実行する方式の詳細を詰めていく予定である。

参考文献

- [1] 伊藤 他：KBMS PHI(1) - 分散知識ベースシステムのシステム構成方式 -、情報処理第32回全国大会
- [2] 大場 他：分散知識ベースの協調に関する一考察、情報処理第34回全国大会
- [3] H. Oba, et al.: Basic Method for Mutual Utilization of Distributed Personal Knowledge Bases, IFIP CONFERENCE ON DISTRIBUTED PROCESSING, October 1987 (to appear)