

TR-408

個別に開発した知識ベースを分散環境下で
相互利用するための用語対応付けの一方式

大場 雅博

June, 1988

©1988, ICOT

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03) 456 3191 ~ 5
Telex ICOT J32964

Institute for New Generation Computer Technology

個別に開発した知識ベースを
分散環境下で相互利用するための
用語対応付けの一方式

大場 雅博

(財) 新世代コンピュータ技術開発機構

概 要

個人や研究チームが、自分の仕事や目的に必要な知識やデータを収集し、それらを自分の観点で分類／整理したものを、パーソナル知識ベース（PKB）と呼ぶ。LANで相互接続されたPKB群は分散PKBシステムと呼ばれ、PKBの内容に対する相互運用性が要求される。そのための基本方式として、パーソナルシソーラス（PT）照合方法が既に提案されている。PTとは、各PKB作成者が自分の使用用語（分野専門語を含む）を各自の観点や目的に従って構造化したものである。各自のPKB中の格納アイテムには、その内容説明のために、いくつかのPT内の用語が付与されている。PT照合により別々のPT内の用語同志が対応付けられ、他者PKBの格納アイテムの内容参照が可能となる。

本論文では、PT照合の実現上での重要課題である、2PT間の用語同志の意味類似度の定量的評価尺度を提案する。意味類似性は、2つのPT間、PKB間の構造的な共通性を基盤として原則的に規定される。さらに、PTやPKBに見られる微妙な個人差を考慮するために、共通性の基盤の上に構造上の相違性を取り込んで拡張される。これに基づき、4種類の意味類似度の定義式が構成される。拡張された意味類似度は、意味の曖昧さを取り扱えるという特徴をもつ。実際のPTに適用し、提案した意味類似度の有効性を確認する。最後に、各類似度の役割や特性について考察し、利用時の指針を与える。

ABSTRACT

An individual (or team) collects and classifies data and knowledge necessary for his jobs and for his purpose. This type of collection of knowledge is called a personal knowledge base (PKB). PKBs connected by a LAN are considered as a distributed knowledge base system. Inter-operability is required for mutual utilization of PKBs.

The personal thesaurus(PT) referencing approach has already been proposed as the basic method. The PT is a hierarchical system of each PKB creator's vocabulary, which reflects his objective, view, and problem field. Words in his PT are attached to each record in his PKB. PT referencing enables word matching among PTs.

This paper proposes a measurement of similarity of word meanings in two PTs, which is one important subject in realizing the method. The measurement is specified principally by the degree of equivalence and difference of structural features observed in PTs and PKBs, and is enhanced by considering the ambiguity stemming from diverse individual usage of words. Its validity is also discussed using an example.

目 次

- 1 緒 言
 - 2 対象のモデル化と用語対応付け問題の定義
 2. 1 分散 P K B システム
 2. 2 P T と P K B の定義
 2. 3 2 サイト間の用語対応付け問題
 2. 4 記法や用語の説明
 - 3 意味類似度の定義
 3. 1 基本的考え方
 3. 2 意味類似性の原則的定義
 3. 3 意味類似度定義上の留意点
 - 3.3.1 P T や P K B の個人差の反映
 - 3.3.2 相互適用時における不自然さの解消
 3. 4 意味類似性の定量的評価尺度
 - 3.4.1 文字的意味類似度
 - 3.4.2 指示的意味類似度
 - 3.4.3 構成的意味類似度
 - 3.4.4 定義的意味類似度
 - 4 事 例
 - 5 考 察
 5. 1 各意味類似度の役割
 5. 2 定義の拡張性
 5. 3 分散 P K B システム環境下での活用
 - 6 結 言
- [謝辞]
- [参考文献]

1 緒 言

CAD や研究情報管理、意志決定などの分野においては、文献情報、報告書、プログラム、プログラミングツール等の、いわゆる知識を取扱うことへの要求が生じている[5][20][4][6]。このような知識ベースの管理・利用形態には、中央集中型[5][20] や、研究者（達）が個人で（ときには、小グループで）維持・管理し（小グループ内で共用し）ているパーソナル型[4] が既に見られるが、これらの他に、パーソナル型知識ベース（以下、PKB）を公開して相互利用する形態[6] が考えられる。このような流れの一方で、エンジニアリングワークステーションやパソコンの発達に伴い、利用者相互の間で情報を互いに呼び出し活用する、相互運用性（インタオペラビリティ）の重要性が認識されている[8]。本論文では、PKB の構築されているシステム（マシン）をLAN で相互接続した分散PKB システムを対象として[1][3]、PKB群をマルチユーザ環境で相互利用するための問題を取り扱う。

分散PKB システムでは、各PKB に格納されている知識の内容は、その作成者の目的や専門分野、取り扱っている問題、関心、観点、等に応じて様々である。このようなPKB 群の相互利用においては、利用者（検索要求者）は自分自身の視点を通して他者のPKB の内容を理解することが必要である。そのためには、まず、個々のPKB の記述に使用されている用語の差異を意識することなく、用語の「意味」を相互に理解することが要求される。これを、分散PKB システムにおける用語の透明性と呼ぶことにする。

分散PKB システム全体で用語の統一をとり管理するというアプローチには限界[7] が予想される。そこで、用語の透明性に対する1つのアプローチとして、既にパーソナルソース照合方式[1][3]が提案されている。それは、各PKB に格納する知識は作成者の専門分野の用語や作成者自身の用語で自由に記述し[4]、PKB 作成者自身が自分の使用用語を各自の問題意識や観点に応じて体系化・整理した構造体（以下パーソナルソース（PT）と呼ぶ）として管理しておき、利用要求の発生時点で利用者と提供者の間でパーソナルソース同志を照合しあって用語の対応を取る、というものである。しかし、[1][3]では、考え方は示されてはいるものの、その具体的方式は明らかにされていない。その具体的実現のために、本論文では、利用者サイトと提供者サイトの2サイト間で行われる、用語の意味を考慮した用語対応付け方式[2] に焦点をあてる。

一般に、同じ表現（字面）の用語であっても、それに込められている「意味」は人や文脈によって様々に使い分けられている。逆に、言い替えや同義語は、一つ概念を表現（字面）の異なる用語で表現したものである。用語の意味に関しては、不幸なことに、普遍的に受け入れられる「意味」の定義は未だない[17]のが現状である。本論文では、用語間の意味の同一性や類似性に対する判断は、最終的には人間（利用者）が責任を持つべき

ものであると考える。システム（マシン）には人間の活動を支援する機能をもたせることにする。その一方式として本論文では、システム（マシン）が利用者サイトの P T 中の用語に対し提供者サイトの P T の中から意味的に類似している用語の候補を拾い出し、利用者自身がそれらの中から対応付けとして適したものを選択し絞り込んでいくという方式を前提とする。この方式ではマシンの機能を実現するために、意味の類似性を概略的に判定できる定量的尺度が問題となる。本論文では、哲学や言語学をはじめいろいろの分野で続けられている用語の意味に関する議論[12][13][18][19]に基づいて、P T や P K B の形態や構造上の特徴を用いたいくつかの意味類似性[2] およびその定量的評価尺度（意味的類似度）を提案する。

以下、第 2 章で対象をモデル化し、2 サイト間での用語対応付け問題を定義する。第 3 章で P T 内の用語群同志の関係や P K B 内の知識同志の関係等を用いた意味類似度を提案する。第 4 章で提案した意味類似度の適用事例を紹介し、有効性を示す。第 5 章で、各意味類似度の役割や拡張可能性について考察する。第 6 章で結論を述べる。

2. 対象のモデル化と用語対応付け問題の定義

2. 1 分散PKBシステム

分散PKBシステムは、LANで接続されたサイト群から成る（図1）。各サイトは、パーソナル知識ベースPKB、パーソナルシソーラスPT、PKBの維持・管理のための知識ベース管理システムKBMS、PT同志の照会及び他サイトのPKBの参照を分散環境下で実行する通信制御部CCから構成されている。

パーソナルシソーラスPTはそのサイトの所有者自身の用語や専門分野用語を、その人の目的や問題意識、観点に従って階層的に分類整理（構造化）したものである。PTの3つの例を図2～図4に示す。以降、それぞれのPTをPTE1、PTE2、PTE3として引用する。

尚、本論文では、PTを、

用語 用語間の階層的な直接的な上下関係

↓

↓

ノード

各用語に対応するノードを結ぶリンク

とするネットワークと見立てて表現することもある。その場合には、ノードの識別のために、用語wに対するノードをノードwと呼ぶことにする。

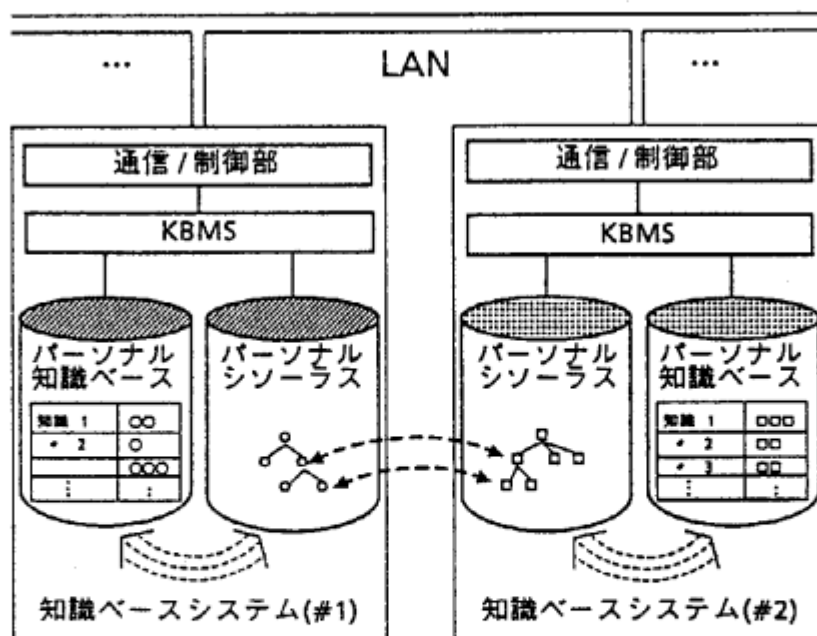


図1. 分散PKBシステムの構成

PTE1

人工知能の課題
方法論
適用領域選定方法
応用システム構築方式
統合人工知能システム構築環境
知識表現機能
基本的表現法
フレーム型
述語型
ルール型
曖昧な知識
深層知識表現
知識継承機能
メタ知識
推論機能
曖昧推論
帰納的推論
知識ベースの矛盾検出
ユーザインタフェース
性能
ワーキングエリア
AI専用マシン
知識ベース設計ツール

図 2. パーソナルシソーラス (例 1)

PTE2

人工知能
基本技術
ハードウェア
従来型
パーソナルコンピュータ
次世代型
推論マシン
知識ベースマシン
人工知能構築環境
人工知能基本言語
エキスパートシステム構築ツール
知識表現
フレーム
述語
ルール
知識ベース
事実知識
ルール知識
メタ知識
知識ベース管理
ワーキングステージ
推論エンジン
利用者インタフェース
システム構築方法
知識の獲得
知識ベースの無矛盾性
知識ベース構造設計

図 3. パーソナルシソーラス (例 2)

PTE3

知識情報処理(人工知能)
基本技術
コンピュータハード
逐次形推論マシン
並列形推論マシン
ソフトウェアシステム
AI言語
エキスパート構築ツール
メタ知識
推論・問題解決
知識ベース
知識獲得
知識表現
知識ベース管理システム
知的インタフェース
応用技術
ソフトウェア生産技術
自然言語処理
画像ビジョン
知能ロボット
教育システム

図 4. パーソナルシソーラス (例 3)

パーソナル知識ベースPKBには、そのサイトの所有者の知識が格納されている。格納されている個々の知識には、その内容や特徴を説明するために、PT内の用語がいくつか付与されている。このとき、各知識は付与されている用語により述定されているという。PT中の各用語は何個（0～複数個）の知識を述定していてもよい。また、述定のために付与されている用語は格納されている知識を検索するためのキーワードとしても活用される。

本論文では、PTやPKBの格納モデルはサイト間で透明である（サイト間での相違は不問）とする。つまり、格納モデルはサイト間で共通であるか、あるいは、たとえサイト間で異なってもその相違を意識することなく他サイトのPTやPKBをアクセスできる機能をKBMSが持っていると前提する（異種モデル間の変換については、例えば、[9][10][11]）。

2. 2 PTとPKBの定義

PTは2つ組 $\langle W, H \rangle$ で定義される。 W は用語を要素とする有限集合である。これは、そのサイトの所有者が使用する全ての用語を意味する。各用語 w は文字列として表現されている。 H は $W \times W$ の部分集合であり、PT内での用語間の関係を意味する。つまり、 $(w_a, w_b) \in H$, $w_a, w_b \in W$ のとき、用語 w_a と用語 w_b との間に階層上の直接的な上下関係がある。本論文では、 w_a を上位、 w_b を下位としておく。

PKBは $K \times W$ の部分集合として定義される。 K はサイト中に格納されている知識 k を要素とする有限集合である。個々の知識 k は構文及び意味が明確に定義された記述言語を用いて、文字列（或いは、その集合）として表現されている。PKBの要素 (k, w) は知識 k が用語 w により述定されていることを意味する。

2. 3 2サイト間の用語対応付け問題

2サイト間の用語の意味を考慮した対応付け問題は、

$$\begin{aligned} \text{(a) サイト } i \text{ 上の } PT_i \text{ と } PKB_i, \quad i = 1, 2 \\ PT_i = \langle W_i, H_i \rangle, \quad H_i \subseteq W_i \times W_i \\ PKB_i \subseteq K_i \times W_i \end{aligned}$$

$$\begin{aligned} \text{(b) } W_1 \text{ 中の用語群から } W_2 \text{ 中の用語群への写像 } \Gamma \\ \Gamma \subseteq \begin{matrix} W_1 \\ 2 \end{matrix} \times \begin{matrix} W_2 \\ 2 \end{matrix} \end{aligned}$$

(c) Γ に対する規制条件 Σ

$$\begin{aligned} & \text{用語集合 } X_1 \subseteq W_1, X_2 \subseteq W_2 \text{ に対し、} \\ & \Sigma(X_1, X_2) = \text{true} \Leftrightarrow (X_1, X_2) \in \Gamma \end{aligned}$$

により、 $\langle PT_i, PKB_i, \Gamma, \Sigma \rangle$ で定義できる。ここで、 Σ は、 W_1 中の用語群と W_2 中の用語群との間の意味類似性の条件である。 Σ として妥当な定量的評価尺度を提案することが、本論文の目的である。

以下、本論文では次の前提を置く。

(1) 用語の対応付けは 1 対 1 対応である。つまり、

$$\begin{aligned} & \Gamma \subseteq W_1 \times W_2 \\ & w_1 \in W_1, w_2 \in W_2 \text{ に対して、} \\ & \Sigma(w_1, w_2) = \text{true} \Leftrightarrow (w_1, w_2) \in \Gamma \end{aligned}$$

(2) $PT_i, i = 1, 2$ は木構造である。

PT の階層上の上下関係は、相対的に、根ノードの方を上位方向とする。

2.4 記法や用語の説明

以下、用語 w のサフィックスとして、 $1, 2, \dots$ はサイト間の区別を、 $a, b, c, \dots$ は 1 つのサイト内での区別を表すのに用いる。

(1) PT 上の距離

1 つの PT 上の用語 w_a と用語 w_b の間の距離 $dst(w_a, w_b)$ を、 PT をネットワークと見たとき、ノード w_a とノード w_b との間のリンクの最少数と定義する。

(2) 上位語、下位語

PT の根ノードの用語を特に先頭用語と呼び、 R と記す。 PT の葉ノードの用語を特に終端用語と呼び、その集合を N と記す。

$(w_a, w_b) \in H$ のとき、用語 w_a を用語 w_b の直接上位語、用語 w_b を用語 w_a の直接下位語と呼ぶ。用語 w の直接下位語の集合を $DV(w)$ と記す。

但し、 $DV(w_n | w_n \in N) = \text{empty}$ とする。

用語 w (但し $w \notin R$) の上位語とは、用語 w の直接上位語を w_u としたとき、 PT 上で用語 w_u と用語 R との通路上にある用語のことである。用語 w の上位語の集合を $U(w)$ と記す。 $U(R) = \text{empty}$ とする。

用語 w (但し $w \notin N$) の下位語とは、 PT 上で用語 w から到達可能な終端用語の集合を N^w としたとき、用語 w の全ての直接下位語と用語 $w_n^w \in N^w$ との通路上にある用語の

ことである。用語 w の下位語の集合を $V(w)$ と記す。

$V(w_n \mid w_n \in N) = \text{empty}$ とする。

(3) 兄弟関係

同一の直接上位語を持つ用語は兄弟関係にある、という。

(4) 部分木

- PTをネットワークとして見たとき、PT上の任意の部分木は、その根ノードと全ての葉ノードの集合とで一意に規定できる。用語 w のノードを根とし、用語集合 S のノード群を葉とする部分木を $IT\langle w, S \rangle$ と記す。部分木が唯一のノード w から構成されるとき、 $S = \text{empty}$ とし、これを単純部分木と言うことにする。終端ノードの用語は、単純部分木しか持たえない。

- 部分木 $IT\langle w, S \rangle$ に含まれるノードの集合を、 $\{IT\langle w, S \rangle\}$ と記すことにする。

- 任意の部分木 $IT\langle w, S \rangle$ に対し、 $\{IT\langle w, S \rangle\}$ に含まれる全用語を $Q1$ とし、 $Q1 - \{w\}$ に含まれる用語と兄弟関係にある全用語を $Q2$ としたとき、ノード w を根とし、 $Q1 \cap Q2 - \{w\}$ を要素として含む部分木を完全部分木とよび、 $IT\langle w, \Diamond S \rangle$ と記すことにする。

(5) 用語、知識の文字列（字面）

用語 w と知識 k に対し、表現している文字列（字面）そのものを指示するときには、 w や k を『 』でくくって表記することにする。

(6) 対応付け

$(w_1, w_2) \in \Gamma$ のとき用語 w_1 と w_2 は対応付けられているといい、 $w_1 \approx w_2$ と記すことにする。

3. 意味類似度の定義

用語の対応付けのためには、意味類似性の条件Ⅰを、人間にとって妥当であり、しかも、マシンが取り扱える形で規定する必要がある。本章では、意味類似性の定量的な定義方法を提案する。

3. 1 基本的考え方

本論文では、分野や問題意識が共通している人々の間では、その人達の作成したPTやPKBの形態や構造上の特徴にも共通性がある、と仮定する。例えば、専門用語やその分類の仕方、収集している知識などの共通性である。そして、このような人々のPTの間での用語対応付けを考察対象とする。

用語の意味的な類似性は上記のような形態的特徴の合同性（一致性）と関係があると考えられる。本論文では、2つのPTやPKBに共通にみられる形態上の合同性を基盤として、2つの用語の意味類似性を原則的に規定する。この規定に従って、2つのサイト間の用語同志の意味類似性の定量的評価尺度（意味類似度）は、形態上の合同性の度合いにもとづいて定義できる。しかし、基盤とした形態的共通性の中にも微妙な個人差、相違が観察される。用語間の意味類似度を定義する際には、それが融通性・柔軟性を持つように、意味類似性の原則的規定内容を十分に反映した形で個人差をも取り扱えるよう考慮することが重要となる。

一方、用語の意味は用語を使用する人や状況により様々に異なる。2つのサイトの用語同志の意味類似性をどのように取り扱い、どのように評価したいかについては、利用者が利用時に指定できる必要がある。これは、意味類似度の値の許容範囲（どの程度の厳しさ或は粗さで判定するか）を利用者が設定するという方式で取り扱うことにする。

以上から、意味類似性の規制条件Ⅱは、種々の形態上の合同度と相違度にもとづく2つの用語間の意味類似度と、その許容範囲とで規定できる。

以下、2つのPT上の用語 $w_1 \in W_1$ と用語 $w_2 \in W_2$ に対して、4種類の意味類似性の原則的定義及びそれらの実際的な定量的定義方法を提案する。

3. 2 意味類似性の原則的定義

本節では、2つのPT (PT_1, PT_2) 上の用語 $w_1 \in W_1, w_2 \in W_2$ の意味類似性をPTやPKBを用いて原則的に規定する。

(1) 文字的意思類似性

分野が共通した人々の間では、専門用語のように文字列(字面)の共通した用語が使用される。用語 w_1 と用語 w_2 が文字列上一致する場合、 w_1 と w_2 は文字列的に合同であるという(文字列合同性)。この時 w_1 と w_2 とは意味的にも類似していると考えられる。この類似性は、共通感覚[12]或は慣用法[13]に基づいたものであり[*脚注1]、文字的意思類似性と名付ける。

(2) 述定的意味類似性

本論文では、2つのPKB中の知識 $k_1 \in K_1$ と $k_2 \in K_2$ は、それぞれの文字列が同一のとき、つまり、「 k_1 」=「 k_2 」のとき、同一であると仮定する。分野や問題意識などの共通した人々の間では、同一の知識群がある用語により述定されていることがある。用語 w_1 と用語 w_2 の述定する知識群が同一である場合、それら知識群は合同であるという(知識群合同性)。この時、 w_1 と w_2 とは意味的にも類似していると考えられる。この類似性は、用語の述定機能[18]に基づいたものであり[*脚注2]、述定的意味類似性と名付ける。

.....
[*脚注1] 専門分野を同じくする人達が議論している時、使用する用語について、厳密に同じ意味で使用しているか(専門的な含みが同じか)否かは判明できないが、相互にある程度の了解は出来ている。これはその分野の人達が誰でもごくあたりまえのこととして知っている、いわば無意識的な自明性[12]に裏付けられているからだと考えられる。

[*脚注2] 話者の立場から言語の機能を捕らえると、言語は、一定の事態を叙示し、話者の意識態勢を表出し、そのことによって所期の反応を喚起する。叙示の機能については、或る対象的与件を指示する機能[14]と、それについてしかじかとして述定する機能とが区別される。
.....

(3) 構造的意味類似性

分野や問題意識などの共通した人々の間では、共通した用語群が共通した分類をされることがある。PT₁、PT₂上でw₁とw₂それぞれの近傍の用語群が同一であり、さらに、それらの分類の仕方が同一であれば、w₁とw₂は構造的に合同であるという（構造的合同性）。この時、w₁とw₂は意味的にも類似していると考ええる。この類似性は、近距離にある用語同志は相互に関連が強いことを前提としており、分類の仕方（つまり、観点や問題意識）に対する共通感覚に基づいたものである。これを構造的意味類似性と名付ける。PT上の分類の仕方のとらえ方により、次の2種類の意味類似性を導入する。尚ここで、w₁とw₂の近傍の用語群の同一性においては、本節で挙げた(1)～(3)の意味類似性のどれを満足していてもよいとする。

(3-1) 構成的意味類似性

1つのPT内では、用語wにより表わされている概念或は事物の構成要素が、その用語wの直接下位語により表わされている概念或は事物として表現されていることが多い。全ての構成要素が一致する2つの概念或いはや事物は意味的に類似であると考ええる。用語w₁と用語w₂の直接下位語群同志が全て対応付く場合、2つの用語の構成要素は合同であるといい（構成要素合同性）、このとき、w₁とw₂は意味的に類似であるとする（構成的意味類似性）。尚、直接下位語群同志の対応付けは、本節で定義するどの意味類似性のどれを満足していてもよい。

(3-2) 定義的意味類似性

1つのPT内では、用語wの直接上位語w_uはwの類概念を示し、wの直接下位語w_vはwの種差を示している[15]と見ることでよいことが多い。概念や事物は種差と類概念により定義される[19]ことから、定義の仕方が同一である2つの概念や事物は意味的に類似であると考ええる。用語wの直接上位語と直接下位語をそれぞれ用語wの上位定義要素、下位定義要素と呼び、この対を定義要素対と呼ぶ。用語w₁とw₂の各定義要素同志が対応付く場合、2つの用語の定義の仕方は合同であるといい（定義要素合同性）、このとき、w₁とw₂は意味的に類似であるとする（定義的意味類似性）。尚、直接上位語同志および直接下位語同志の対応付けは、本節で定義するどの意味類似性のどれを満足していてもよい。

3. 3 意味類似度定義上の留意点

前節で規定した意味類似性に基づいて、それらの定量的評価尺度を提案するに当たり、次の2点に注意する必要がある。

3.3.1 PTやPKB の個人差の反映

PTやPKB には、前節で述べた共通性に対して、實際上、いくつかの個人差が見られる。PT間で用語の対応付けに用いる定量的評価尺度は、融通性・柔軟性のために、それらの個人差を考慮できなければならない。そこで、意味類似性の考え方を拡張することが必要となる。

(1) PTやPKB における個人差

- ① 用語の表記においては、個人差により、用語中の文字列が部分的に異なることがある。
- ② 調査範囲や目的などの個人差により、収集されている知識は人毎に異なる。合同な用語により指示されている知識群の間には同一でない知識を含む場合がある。
- ③ 構成要素のとらえ方や構成方法に個人差があるので、用語の構成要素がその用語から少し離れた用語（直接下位語或いは下位語）である場合がある。
- ④ 用語の構成要素（直接下位語或いは下位語）同志が完全には一致しない場合がある。
- ⑤ PTの構造化の個人差により、用語の定義要素がその用語から少し離れた用語（上位語，下位語）である場合がある。
- ⑥ 観点の個人差により、定義の仕方が逆転している場合がある。

(2) 個人差への対処法 —意味類似性の考え方の拡張—

上記個人差に対しては次のように対処する。

文字的意味類似度を、2つの用語が同一の文字を共有する度合い（文字列合同度）により定義する。（上記①への対処）

指示的意味類似度を、2つの用語が同一の知識を共有する度合い（知識群合同度）により定義する。（上記②への対処）

構成的意味類似性の定義における構成要素及びそれらの対応付け条件の取扱いを次のように拡張する。（上記③、④への対処）

- 用語wの構成要素は、その用語の下位語により表されている。但し、構成要素に含まれている用語は、全て上下関係にないものとする。言い換えると、用語wの構成要素の集合をSとしたとき、Sにより部分木 $I(T < w, S >)$ を張ることができる。
- w_1 と w_2 の構成要素をそれぞれ S_1 、 S_2 としたとき、 S_1 と S_2 の要素の用語同志

で対応付く用語群に着目し、それら用語群が $\diamond S_1$, $\diamond S_2$ の全用語に占める割合や、 w_1 , w_2 からそれら用語群への距離の逆比を用いて、構成的意味類似度を定義する。
尚、ここで、用語群が $\diamond S_1$, $\diamond S_2$ の全用語に占める割合を用いることにしたのは、構成要素は兄弟関係にあるものが全部そろってはじめて意味があると仮定したからである。
また、距離の逆比を用いることにしたのは、PT内で近距離にある用語同志は相互関連が強い、という前提を反映するためである。

定義的意味類似性の定義における定義要素及びそれらの対応付け条件の取扱いを次のように拡張する。(上記④、⑤への対処)

- 用語 w の上位語と下位語が定義要素を表している。
- 定義要素の一致の必要条件を、次の(i), (ii)の少なくとも一方が成立することとする。
 - (i) 上位定義要素同志、下位定義要素同志がともに対応付く
 - (ii) 各上位定義要素が共に他方の下位定義要素と対応付く
- それら対応付く用語への w_1 , w_2 からの距離の逆比を用いて、定義的意味類似度を定義する。これは、上記と同じ理由による。

3.3.2 相互適用時における不自然さの解消

2用語間の構成的意味類似度と定義的意味類似度は、それら用語の構成要素や定義要素の間で対応付く用語に注目して定義されている。このとき、構成的意味類似度や定義的意味類似度を算出するために、さらに構成的あるいは定義的類似度を算出することが繰り返し(再帰的に)適用される、ということが生じうる。つまり、 w_1 と w_2 との構成的或いは定義的意味類似度を算出するためには、 w_1 と w_2 の構成要素や定義要素である上位語や下位語間の対応付けが必要であり、そのためにこれら上位語や下位語の構成的或いは定義的意味類似度を算出するためには、これらの上位語や下位語のさらに上位語・下位語の対応付けが必要となり、……。このとき、次のような2つの点に留意しなければならない。

(1) 構成的意味類似度と定義的意味類似度を組合せて何重にも適用すると、意味的類似度の算出対象であったもとの用語が、その用語の上位語や下位語(のさらに上位語・下位語...)の意味的類似度算出のための候補として用いられることが生じうる。このことは不自然である

例えば、図5において、 $c_1 \approx c_2$ と $w_1 \approx d_2$ から $a_1 \approx a_2$ をいい、 $a_1 \approx a_2$ と

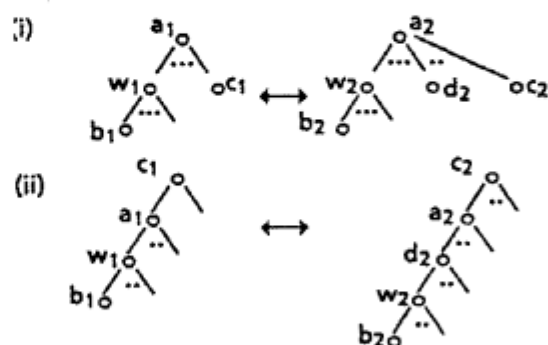


図5. 再帰的適用時の不自然さ(その1)

$b1 \approx b2$ から $w1 \approx w2$ をいうのは不自然。

そこで本論文では、ある用語同志の構成的或いは定義的意味類似度を算出するとき、その用語の上位語・下位語の対応付けにおいては、もとの用語は使ってはならない、という制限を置く。

(2) 定義の仕方の一致条件は、用語間の局所的な上下関係(逆転の場合を含む)を保障する。しかし、定義的意味類似度を何重にも適用すると、用語間の全体的な上下関係が一致しない場合が生じる。これは不自然である。例えば、図6において、 $c1 \approx c2$ と $e1 \approx e2$ から $a1 \approx a2$ 、 $d1 \approx d2$ と $f1 \approx f2$ から $b1 \approx b2$ 、 $a1 \approx a2$ と $b1 \approx b2$ から $w1 \approx w2$ をいうのは不自然。

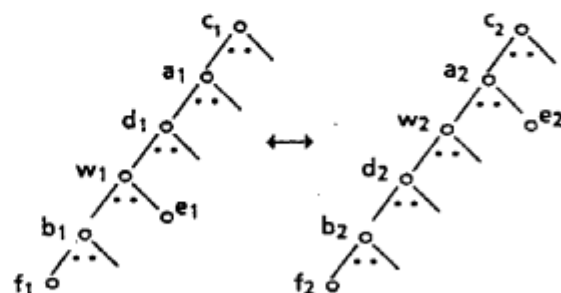


図6. 再帰的適用時の不自然さ(その2)

そこで本論文では、ある用語同志の定義的意味類似度を算出するとき、その用語の定義要素の対応付けのために用いる用語を選出する場合に、次のような制限を置く。つまり、上位定義要素に対しては、もとの用語の上位語と、その上位語の下位語(但し、もとの用語とその下位語を除く)を対象とし、下位定義要素に対しては、もとの用語の下位語を対象とする。

上記第1の制限は第2の制限に含まれる。2つの制限を統一的に解決するために、構成的及び定義的意味類似度の定義においては、対象領域とする部分木を指定することが必要となる。つまり、2用語間の構成的或いは定義的意味類似度を算出する際、もとの用語でPTを分割し、それら用語の構成要素や定義要素が含まれるサブ部分木を、構成要素間や定義要素間の類似度算出のための対象領域とする。この部分木をベース部分木と呼ぶことにし、in [部分木] for [用語] という記述で指定することにする。

3.4 意味類似性の定量的評価尺度

前節に基づき、2用語間の意味類似性の定量的評価尺度を次のように決める。

2つのPT上の用語 $w_1 \in W_1$ と $w_2 \in W_2$ とが与えられているとする。

3.4.1 文字の意味類似度

(1) 文字列合同度

2用語 w_1 と w_2 に対し、文字列合同度 $CA(『w1』, 『w2』)$ を次のように定義する。

$$CA(『w1』, 『w2』) = \frac{n1}{n1+n2+n3} \quad \dots\dots\dots ①$$

ここで n_1 : 『w1』と『w2』に共通に出現する文字の数

n_2 : 『w1』だけに出現する文字の数

n_3 : 『w2』だけに出現する文字の数

である。

(2) 文字の意味類似度

2用語 w_1, w_2 に対し、文字の意味類似度 $SA(w_1, w_2)$ を次のように定義する。

$$SA(w_1, w_2) = CA(『w_1』, 『w_2』) \quad \dots\dots\dots ②$$

文字の意味類似度の許容値を α_{SA} とすると、規制条件 Σ は、

$$\Sigma_1(w_1, w_2) = \text{true} \Leftrightarrow SA(w_1, w_2) \geq \alpha_{SA} \quad \dots\dots\dots ③$$

となる。 Σ_1 が満足されるとき、 $w_1 \approx w_2$ である。

例1 : $v1$ = フレーム、 $v2$ = フレーム型、の場合。

$n_1 = 4$ (『フレーム』), $n_2 = 0$, $n_3 = 1$ (『型』)

から、 $SA(w_1, w_2) = 0.8$

例2 : $v1$ = 人工知能築環境、 $v2$ = 統合AIシステム構築環境、の場合。

$n_1 = 4$ (『構築環境』), $n_2 = 4$ (『人工知能』),

$n_3 = 8$ (『統合AIシステム』)

から、 $SA(w_1, w_2) = 0.25$

3.4.2 述定的意味類似度

(1) 知識群合同度

2つの用語 w_1 と w_2 が述定する知識の集合をそれぞれ Tw_1, Tw_2

$$Tw_i = \{k \mid (k, w_i) \in PKB_i\} \quad i = 1, 2 \quad \dots\dots\dots ④$$

としたとき、 Tw_1 と Tw_2 の合同性の度合い (知識群合同度) $CB(Tw_1, Tw_2)$ を次のように定義する。

$$CB(Tw_1, Tw_2) = \frac{|T_c|}{|T_c| + |Tw_1 - T_c| + |Tw_2 - T_c|} \quad \dots\dots\dots ⑤$$

ここで、

$|A|$: 集合 A の要素数 (以下同様)。

$$T_c = \{k \mid [k] = [k'], k \in Tw_1, k' \in Tw_2\}$$

である。集合 T_c は Tw_1 と Tw_2 に共通に出現する同一の知識の集合を表しており、それらが与えられると一意に定まる。分母の第2項、第3項はそれぞれ Tw_1 だけ、 Tw_2 だけに出現する知識の数を表わしている。

(2) 指示的意味類似度

2用語 w_1 と w_2 の指示的意味類似度 $SB(w_1, w_2)$ を、次のように定義する。

$$SB(w_1, w_2) = CB(Tw_1, Tw_2) \quad \dots\dots\dots ⑥$$

指示的意味類似度の許容値を α_{SB} とすると、規制条件 Σ は、

$$\Sigma_2(w_1, w_2) = \text{true} \Leftrightarrow SB(w_1, w_2) \geq \alpha_{SB} \quad \dots\dots\dots ⑦$$

となる。 Σ_2 が満足されているとき、 $w_1 \approx w_2$ である。

例3：今、次のような $PT_i, PKB_i, i=4,5$ があるとする。

$w_1 \in W_4, w_2 \in W_5,$

$(ka, w_1), (kb, w_1), (kc, w_1), (kd, w_1), (ke, w_1), (kf, w_1), (kg, w_1), (kh, w_1) \in PKB_4$

$(ka, w_2), (ki, w_2), (kb, w_2), (kj, w_2), (kc, w_2), (kk, w_2), (ke, w_2) \in PKB_5$

このとき、 $T_c = \{ka, kb, kc, ke\}, Tw_1 - T_c = \{kd, kf, kg, kh\}, Tw_2 - T_c = \{ki, kj, kk\}$

から、 $SB(w_1, w_2) = 0.36$

3.4.3 構成的意味類似度

2つの用語 $w_1 \in W_1$ と $w_2 \in W_2$ に対し、それらの対応付けのためのベース部分木として部分木 $IT_1 \subseteq PT_1$ と $IT_2 \subseteq PT_2$ が与えられている。

use $[IT_1 < t_1, St_1 >]$ for $[w_1]$

use $[IT_2 < t_2, St_2 >]$ for $[w_2]$

(1) 構成要素合同度

用語 w_1, w_2 のノードを根とする、ベース部分木上の任意の部分木をそれぞれ P_{w1}, P_{w2} とする。

$$P_{w1} = IT_{w1}.Sw1 \subseteq IT_1$$

$$P_{w2} = IT_{w2}.Sw2 \subseteq IT_2$$

この部分木の組に対し、構成要素合同度 $CC(P_{w1}, P_{w2})$ を次のように定義する。

(i) P_{w1}, P_{w2} の少なくとも一方が単純部分木の場合

$$CC(P_{w1}, P_{w2}) = 0 \quad \dots\dots\dots \textcircled{8}$$

(ii) P_{w1} と P_{w2} が共に単純部分木でない場合

次のように定義する。

(a) Sw_1 から Sw_2 への写像 G を次のようにとる。

$$G = \{ (x_1, x_2) \mid x_1 \in Sw_1, x_2 \in Sw_2, \\ \exists k : \Sigma_k(x_1, x_2) = \text{true}, k = 1 \sim 4 \} \quad \dots\dots\dots \textcircled{9}$$

この写像 G は、 Sw_1 と Sw_2 が与えられると一意に定まる、多対多写像である。

但し、 $\Sigma_k(x_1, x_2)$ の判定において構成的類似度 Σ_3 或いは定義的類似度 Σ_4 を用いるときには、 x_1, x_2 のベース部分木はそれぞれ次の通りである。

$$\Sigma_3 \text{ では } \begin{aligned} & \text{use } [IT < x_i, M1 >] \text{ for } [x_i] \quad i=1,2 \\ & M1 = V(x_i) \cap St_i \end{aligned} \quad \dots\dots\dots \textcircled{10}$$

$$\Sigma_4 \text{ では } \begin{aligned} & \text{use } [IT < w_i, M2 >] \text{ for } [x_i] \quad i=1,2 \\ & M2 = V(w_i) \cap St_i \end{aligned} \quad \dots\dots\dots \textcircled{11}$$

(b) G の部分集合 G' を、 G' が 1 対 1 写像となるようにとる。 G' は一意ではない

(c) 写像対 $\langle G, G' \rangle$ に関する、 P_{w1} と P_{w2} との合同度 $PPC(P_{w1}, P_{w2})$ $\langle G, G' \rangle$ を次のように定義する。

$$PPC(P_{w1}, P_{w2}) \langle G, G' \rangle = \begin{cases} \frac{f1(G')}{2 \times |G'|} \times \frac{f1(G')}{f1(G') + f2(G')} & \dots\dots |G'| > 0 \text{ のとき} \\ 0 & \dots\dots\dots |G'| = 0 \text{ のとき} \end{cases} \quad \text{.....⑫}$$

ここで

$$f1(G') = \sum_{(x,y) \in G'} \left(\frac{1}{dst(w1,x)} + \frac{1}{dst(w2,y)} \right)$$

$$f2(G') = \sum_{x \in NX} \frac{1}{dst(w1,x)} + \sum_{y \in NY} \frac{1}{dst(w2,y)}$$

$$NX = \Diamond Sw1 - \{ x \mid (x,y) \in G' \}$$

$$NY = \Diamond Sw2 - \{ y \mid (x,y) \in G' \}$$

(d) P_{w1} と P_{w2} との構成要素合同度 $CC(P_{w1}, P_{w2})$ を次のように定義する。

$$CC(P_{w1}, P_{w2}) = \max \{ PCC(P_{w1}, P_{w2}) \langle G, G' \rangle \mid G' \subseteq G, G' \text{ は 1 対 1 の対応} \} \quad \text{.....⑬}$$

(2) 構成的意味類似度

2つの用語 w_1, w_2 に対し、与えられたベース部分木中での構成的意味類似度 $SC(w_1, w_2)$ を次のように定義する。

$$SC(w_1, w_2) = \max \{ CC(P_{w1}, P_{w2}) \mid P_{w1} \subseteq IT_1, P_{w2} \subseteq IT_2 \} \quad \text{.....⑭}$$

構成的意味類似度の許容値を α_{SC} とすると、規制条件 Σ は、

$$\Sigma_3(w_1, w_2) = \text{true} \Leftrightarrow SC(w_1, w_2) \geq \alpha_{SC} \quad \dots\dots\dots (15)$$

となる。 Σ_3 が満足されているとき、 $w_1 \approx w_2$ である。

例4： PTE1, PTE2 において、PTE1中の「基本的表現法」とPTE2中の「知識表現」の各々の下位語の中で、今仮に次の対応付けが Σ_1 を満足しているとする。

PTE1	PTE2
フレーム型	≅ フレーム
述語型	≅ 述語
ルール型	≅ ルール

このとき、CC（基本的表現法、知識表現）＝1

例5： PTE1, PTE2 において、PTE1中の「統合人工知能システム構築環境」とPTE2中の「エキスパートシステム構築ツール」「人工知能構築環境」の各々の下位語の中で、次の対応付けが $\Sigma_1 \sim \Sigma_4$ を満足しているとする。

PTE1	PTE2
基本的表現法	≅ 知識表現
メタ知識	≅ メタ知識
ユーザインタフェース	≅ 利用者インタフェース
ワーキングエリア	≅ ワーキングステージ
知識ベース設計ツール	≅ 知識ベース管理

このとき、

SC（統合人工知能システム構築環境、エキスパートシステム構築ツール）

$$= \frac{8}{10} \times \frac{8}{8+5} = 0.49$$

SC（統合人工知能システム構築環境、人工知能構築環境）

$$= \frac{5.8}{10} \times \frac{5.8}{5.8+4.2} = 0.34$$

3.4.4 定義的意味類似度

2つの用語 $w_1 \in W_1$ と $w_2 \in W_2$ に対し、それらの対応付けのためのベース部分木として部分木 $IT_1 \subseteq PT_1$ と $IT_2 \subseteq PT_2$ が与えられている。

use [$IT_1 < t_1, St_1 >$] for [w_1]
use [$IT_2 < t_2, St_2 >$] for [w_2]

(1) 定義構造合同度

2つの用語 w_1, w_2 が共に、与えられたベース領域上に定義要素を持つ場合、つまり

$w_1 \in \{IT_1 \text{ の根ノード, 葉ノード} \}$ かつ $w_2 \in \{IT_2 \text{ の根ノード, 葉ノード} \}$

のとき、各用語の上位、下位の定義要素をそれぞれ $u_{w_1}, v_{w_1}; u_{w_2}, v_{w_2}$ とし、定義要素対をそれぞれ Q_{w_1}, Q_{w_2} とする。

$$Q_{w_1} = \langle u_{w_1}, v_{w_1} \rangle$$

$$Q_{w_2} = \langle u_{w_2}, v_{w_2} \rangle$$

このとき、 Q_{w_1} と Q_{w_2} との構成要素合同度 $CC(Q_{w_1}, Q_{w_2})$ を次のように定義する。

$$CD(Q_{w_1}, Q_{w_2}) =$$

$$\frac{1}{4} \left(\sum_{x \in NX} \frac{1}{dst(w_1, x)} + \sum_{y \in NY} \frac{1}{dst(w_2, y)} \right)$$

$$NX = \{uw_1, vw_1\}, NY = \{uw_2, vw_2\}$$

……………定義要素同志が対応付くとき

0

……………上記以外の場合

……………⑩

但し、定義要素同志の対応付けの判定において構成的類似度 Σ_3 或いは定義的類似度 Σ_4 を用いるときには、各定義要素のベース部分木はそれぞれ次の通りである。

use [$IT < t_i, M3 >$] for [u_{w_i}] $i=1,2$

$$M3 = St_i - V(w_i)$$

…………… w_i の兄弟用語 $\in IT_i$ のとき

$$M3 = \{St_i - V(w_i)\} \cup d_{w_i},$$

$$(d_{w_i}, w_i) \in H_i$$

…………… w_i の兄弟用語 $\notin IT_i$ のとき

use [$IT < y_i, M4 >$] for [v_{w_i}] $i=1,2$

$$y_i \in DV(w_i) \cap U(v_i)$$

$$M4 = St_i \cap V(w_i)$$

……………⑪

(2) 定義的意味類似度

2つの用語 w_1, w_2 に対し、与えられた誘導木の中での定義的意味類似度 $SD(w_1, w_2)$ を次のように定義する。

$$SD(w_1, w_2) = \begin{cases} \max\{CD(Qw_1, Qw_2) \mid \\ u_{vi} \in U(w_1) \cap IT_i, v_{vi} \in V(w_2) \cap IT_i, i=1,2\} \\ \dots\dots\dots w_1, w_2 \text{ 共に定義要素対を持つ場合} \\ 0 \quad \dots\dots\dots w_1, w_2 \text{ の少なくとも一方が定義要素対を持たない場合} \end{cases} \quad (18)$$

定義的意味類似度の許容値を α_{SD} とすると、規制条件 Σ は、

$$\Sigma_4(w_1, w_2) = \text{true} \Leftrightarrow SD(w_1, w_2) \geq \alpha_{SD} \quad (19)$$

となる。 Σ_4 が満足されているとき、 $w_1 \approx w_2$ である。

例6： PTE1とPTE3において、PTE2中の「ハードウェア」とPTE3中の「コンピュータハード」の各々の上位語と下位語の中で、次の2組の対応付けが $\Sigma_1 \sim \Sigma_4$ を満足しているとする。

- (a) PTE2 PTE3
 基本技術 $\approx$ 基本技術
 推論マシン $\approx$ 逐次型推論マシン

このとき、 $CD(\text{ハードウェア}, \text{コンピュータハード}) = 0.88$

- (b) PTE2 PTE3
 知識情報処理 $\approx$ 人工知能
 (人工知能)
 推論マシン $\approx$ 並列型推論マシン

このとき、 $CD(\text{ハードウェア}, \text{コンピュータハード}) = 0.58$

以上から、 $SD(\text{ハードウェア}, \text{コンピュータハード}) = 0.88$

4. 事 例

PTE1とPTE2が与えられているとして、PTE1中の「総合人工知能システム構築環境」に対応する用語をPTE2から探し出す場合を考える。また、PKBに関しては、PKB1中で「総合人工知能システム構築環境」が述定している知識は $\{ka, kb, kd, kf, kg, kh, ki, kk, kl, km, kn\}$ であり、これらと同一の知識を述定するPKB2中の用語及びそれが述定する知識は次の通りとする。「推論マシン」： $\{ka, kb, kc\}$ 、「知識ベースマシン」： $\{kd, ke\}$ 、「エキスパートシステム構築ツール」： $\{kf, kg, kh, ki, kj\}$ 。

(1) 「総合人工知能システム構築環境」に対し、SAの値が大きい幾つかの用語を挙げると、次のようになる。

SA(総合人工知能システム構築環境、人工知能構築環境) = 0.57

SA(総合人工知能システム構築環境、システム構築方法) = 0.38

SA(総合人工知能システム構築環境、人工知能) = 0.29

SA(総合人工知能システム構築環境、エキスパートシステム構築ツール) = 0.26

(2) 「総合人工知能システム構築環境」に対し、SBの値が決まる用語を挙げると、次のようになる。

SB(総合人工知能システム構築環境、エキスパートシステム構築ツール) = 0.33

SB(総合人工知能システム構築環境、推論マシン) = 0.17

SB(総合人工知能システム構築環境、知識ベースマシン) = 0.08

(3) 「総合人工知能システム構築環境」に対し、SCの値が大きい幾つかの用語を挙げると、次のようになる。

SC(総合人工知能システム構築環境、エキスパートシステム構築ツール) = 0.64

SC(総合人工知能システム構築環境、システム構築方法) = 0.59

SC(総合人工知能システム構築環境、人工知能) = 0.43

SC(総合人工知能システム構築環境、人工知能構築環境) = 0.42

SC(総合人工知能システム構築環境、基本技術) = 0.34

尚、上記の計算では、構成要素の対応付けとして、図7に示した対応付けをベースとした。

(4) 「総合人工知能システム構築環境」に対し、SDの値が大きい幾つかの用語を挙げると、次のようになる。

SD(総合人工知能システム構築環境、システム構築方法) = 1.00

SD(総合人工知能システム構築環境、エキスパートシステム構築ツール) = 0.83

SD(総合人工知能システム構築環境、基本技術) = 0.83

SD（総合人工知能システム構築環境，人工知能構築環境）＝ 0.75

SD（総合人工知能システム構築環境，知識ベース）＝ 0.69

SD（総合人工知能システム構築環境，知識表現）＝ 0.65

尚、上記の計算では、上位定義要素は「人工知能の課題」 $\approx$ 「人工知能」の対応付けを、下位定義要素は図7に示した対応付けをそれぞれベースとした。

PTE1中の「総合人工知能システム構築環境」に対し、PTE2の用語の中には第3.2節で述べた原則的意味類似性を満足するものは存在しない。（PTE2中の「システム構築方法」は上記計算ではSDの値が1.0となっているが、定義要素同志が厳密に一致している訳ではない。）しかし、上記の結果から、拡張された意味類似度により、各類似度ごとに、いくつかの用語群が対応付け対象として選出されうること、つまり、対応付けの可能性の広がっていることが確認できる。

対応付け対象となりうる用語群やその中での類似値の順位は、各意味類似度ごとに少しずつ違っている。このことから、各々の意味類似度の特徴を活かした対応付けを得ることが期待できる。

また上記の用語群を見ると、全般的に無理のない用語が対象として挙っていることが分かる。

以上から、利用者自身が $\alpha_{SA} \sim \alpha_{SD}$ の値を操作して、選出されてくる用語の数を調整しながら、また、各意味類似度の特性や役割を考慮しながら、自身の判断により好ましい対応付けを指定することが可能である。

<u>PTE1</u>	<u>PTE2</u>
知識表現機能	知識表現
基本的表現法	フレーム
フレーム型	述語
述語型	ルール
ルール型	知識ベース
深層知識表現	ルール知識
メタ知識	メタ知識
知識ベースの矛盾検出	知識ベース管理
ユーザインタフェース	ワーキングステージ
ワーキングエリア	利用者インタフェース
知識ベース設計ツール	知識の獲得
	知識ベースの無矛盾性
	知識ベース構造設計

図7. 構成要素、構成要素の対応付けデータ

5. 考 察

5. 1 各意味類似度の役割

意味類似度は、先ず、2つのPTやPKBの間の共通性に着目して原則的に規定され、次に、幾く種類かの個人差を考慮するために、PT間やPKB間の相違性を取り込んで拡張された。拡張された意味類似度は、意味の曖昧性を取り扱えると言ってよい。この点は、第4章の事例検討において確認された。ところで、本論文では、第1章で述べた通り、この曖昧性は利用者の判断により解決されることを前提としている。この解決のための指針を与えるために、それぞれの意味類似性の役割を考察する。

PTを観察すると、PT中には、広義に専門用語と考えられる用語と、より一般的な用語（日常語）とが含まれている。専門用語は、その使用法という観点から見ると、つぎの2種類を区別せねばならない。第1は、正確（厳密）な定義が広く行き渡った用語。第2は、技術者や科学者が自分でよいと思うように定義し直して使用している用語[16]。この場合、字面が同じであっても意味は異なっていたり、字面は異なっているが意味が似通っていたりする。一方、日常語は、用語の表記や使用法については混乱が少なく、第1の専門用語と同類と考えてよい。第1の専門用語や日常語に対しては、文字的意味類似度が有効である。第2の専門用語に対しては、用語の使い方を考慮できる述定的及び構造的意味類似度が有効である。

述定的意味類似度と構造的意味類似度の違いは、概念の説明の仕方に関わっている。述定的意味類似度は、具体物を用いての説明に相当し、構造的意味類似度は、ある概念を別な概念で以て説明することに相当する。個人的な概念体系の中で、より低位の基本的な用語に対しては、述定的意味類似度の適用可能性が高く、より高位の用語に対しては、構造的意味類似度の適用可能性が高い。

構造的意味類似度の中で、構成的意味類似度と定義的意味類似度の違いは、概念間の関係の仕方に関わっている。構成的意味類似度は、全体／部分の関係に注目して意味を捕らえている場合に対して、定義的意味類似度は、上位／下位概念、クラス包摂に注目して意味を捕らえている場合に対して、それぞれ有効である。

しかし、本論文で提案した述定的や構造的意味類似度は、概念の多くの説明法や用語間に成り立つ多くの意味関係のほんの一部にしか過ぎない。このことは銘記すべきである。

構造的に意味類似であるためには、構成要素なり定義要素がいずれかの意味類似性を満足することが必要である。構造的意味類似度の再帰的適用が停止するためには、もとの用語からPT中の先頭用語や終端用語に至る途中の用語が文字的に意味類似であるか述定的に意味類似であるかのいずれかが必要である。このことは、用語の定義や説明においては、概念上の基本的用語や日常語を用いる、という事に相当するものと考えられる。

5. 2 定義の拡張性

意味類似度は、先ず、2つのPTやPKBの間の共通性に着目して原則的に規定され、次に、幾く種類かの個人差を考慮するために、PT間やPKB間の相違性を取り込んで拡張された。本論文で提案した意味類似度の計算式では、共通性と相違性は同等の重みで扱われている。しかし、使用者や用語が違えば、その扱いに対する考え方も違ってこよう。例えば、利用者側の用語 w_1 と提供者側の用語 w_2 との構成的意味類似性を考える場合を想定してみると、 w_1 と w_2 の構成要素をそれぞれ C_1 、 C_2 として、 C_2 と共通しない C_1 の要素は、 C_1 と共通しない C_2 の要素より重要視すべき場合が少なくない。この点に関しては、本論文の計算式は、共通項と相違項のどれを重視するか重み付け係数を追加することは容易である。

述定的意味類似度では、対象としている用語自身が述定している知識だけを考慮している。しかし、その用語の近傍の用語が述定している知識を含めても好い場合がある。これは、④式において、述定知識の集合の定義を修正することで対処できる。

5. 3 分散PKBシステム環境下での活用

本論文では、2サイト間の用語対応付けのための意味類似度を提案した。分散環境下での知識の検索（つまり、或る1つのサイトから他の全サイトを対象として、検索の精度を（大きく）損なうことなく、知識を検索する）に、この類似度を適用することも可能である。

6. 結 言

相互接続されたパーソナル知識ベース（PKB）群を相互利用するための基本方式として既にパーソナルシソーラス（PT）照合方式が提案されている。それを実現するための重要課題である、2PT間の用語同志の意味類似度の定量的評価尺度を提案した。

本論文では、分野や問題意識が共通している人々の間では、PTやPKBの形態上の特徴にも共通性があると仮定した。基盤となる4種類の共通性（用語の文字列、述定知識、構成要素、定義の仕方）にもとづき、2つの用語間の意味類似性を原則的に規定した。次に、基盤となる共通性の中に見られる微妙な個人差を考慮するために、共通性の枠内での相違性を取り込んで原則的規定を拡張した。この中の2種類の意味類似度は、無制限に再帰的に適用すると、PT上の大局的な用語関係が守られないという不自然さが生ずる。その解消のために、PTの使い方に制限を置いた。以上の考え方にもとづき、4種類の意味類似度の定義式を導いた。拡張された意味類似度は意味の曖昧性を取り扱えるという特徴を有する。

提案した意味類似度を実際のPTに適用し、納得でき、しかも無理のない用語対応付けの得られること、及び、利用者自身の判断により好ましい対応付けを選出できることを確認した。もって、本類似度の妥当性と有効性を示した。

それぞれの意味類似度の役割や特性を、用語の使用法や概念の説明法、概念間の関係等に関連させて考察した。これにより、利用者が対応付けの適合を判断する際の、一つの指針を与えた。

〔謝辞〕

本研究の機会を与えて下さった当所潤一博所長、同第3研究室室長伊藤英則博士に感謝します。

[参考文献]

- 1) 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、1987。
- 2) 大場、伊藤：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、情報処理第35回全国大会、1987。
- 3) Oba, M., and Itoh, H.: Basic Method for Mutual Utilization of Distributed Personal Knowledge Bases, in Distributed Processing (ed. by M. Barton et. al.), North-Holland (1988).
- 4) Kondo, H. and Koguchi, T.: A New Approach to Knowledge Base Management Systems, in Methodologies for Intelligent Systems (ed. by Z. Ras and M. Zemankova), Elsevier Science Publishing Co., Inc. (1987).
- 5) 持田、他：知識情報に基づく駆動型メソッドベース、データベースシステム 57-12 (1987)。
- 6) 松田孝子：研究活動におけるデータベース技術の応用と展望、情報処理, Vol.23, No.10 (1982)
- 7) 吉田将：辞書構築における諸問題、情報処理、Vol.27, No.8 (1986)。
- 8) 棟上、植村：分散処理の展望、情報処理、Vol.28, No.4 (1987)
- 9) Demurjian, S.A. and Hsiao, D.K.: The Multi-Lingual Database System, Proc. Third Int. Conf. Data Engineering, pp.44-51, 1987.
- 10) Katz, R. H. and Wong, E.:Decompiling CODASYL DML into Relational Queries, ACM Trans. on Database Systems, Vol.7, No.1, March 1982.
- 11) Rosenberg, R.L. and Landers, T.: An Overview of MULTIBASE, Distributed Data Bases, H.-J. Schneider, ed., North-Holland Publishing Company, 1982.
- 12) 中村雄二郎：術語集一気になることばー、岩波新書(黄版) 276, 岩波書店、東京 (1984)。
- 13) Hospers, John : An Introduction to PHILOSOPHICAL ANALYSIS< Second Edition, Prentice-Hall, Englewood Cliffs (1967).
- 14) Clark, H.H. & Clark, E.V. : Psychology and language-An Introduction to Psycholinguistics, Harcourt Brace Jovanovich, New York (1977).
- 15) 長尾貞監修：日本語情報処理、第7章、電子通信学会 (1984)。
- 16) Ullman, S.: Semantics- An Introduction to the Science of Meaning, Basil Blackwell and Mott Ltd.(1962).
- 17) Miller, G.A.: Language and Speech, W.H. Freeman, (1981)
- 18) 廣松渉：言語の意味と認識の問題、(田島、坂本他編、講座現代の哲学第3巻、言語の内と外)、弘文堂 (1977)。
- 19) 栗田、古在編：岩波哲学小辞典、岩波書店 (1979)。
- 20) アドバンスト・データベース・シンポジウム論文集、情報処理学会 (1981)