

情報処理用語の J I S 規格における関連語検索

5 V-4

寺田欣泰 林 俊文 石川佳治 植村俊亮
奈良先端科学技術大学院大学 情報科学研究科

1. はじめに

日本工業規格 (JIS) 情報処理用語には、見出し語とその定義文、番号などが見出し語ごとに独立に記述されており、見出し語相互間の関係は表示されていない。このようなテキストデータを知識の体系として機械的にかつ効率良く利用するためには、見出し語間の意味的な関係の表現、即ち、広義のシソーラスが必要である。

本稿では、情報処理用語の JIS 規格を実験材料として、見出し語の関連語を自動的に選定する手法と、見出し語間の意味的関連の自動的体系化の方法を論じる。

2. 用語の特徴

2.1 単語の特性

関連語を選定する上で重要となるのは、各々の単語が持つ意味や特徴であり、それらはその単語の品詞や日本語文法上の特徴から得ることができる。用語のほとんどが複合語であるため、それを単語に分解し、その特性を以下のように分類した。

カテゴリ 1: それ自身で一つの用語に成り得る単語であり、複合語である用語の末尾に現れ語幹的な振舞いをする単語。

(例) システム、言語、計算機、など

カテゴリ 2(a): その単語だけでは一つの用語には成り得ない単語であり、複合語の最初に現れ修飾語的な振舞いをする単語。

(例) 超、特殊、初期、疑似、など

カテゴリ 2(b): 数字、数詞。

カテゴリ 2(c): 反義、否定を意味する接頭辞。

(例) 非、否定、不、など

カテゴリ 3: 対義の単語を持つ単語。

(例) 直接 ↔ 間接、アナログ ↔ デジタル、など

カテゴリ 4: その他の普通・固有名詞。

ただし、いくつかの単語は複数の特性を持つことがある。例えば、“アナログ”はカテゴリ 3 と 4、“直接”はカテゴリ 2(a) と 3 を持つ。

2.2 説明文の構成

JIS 規格における用語の説明文の構成は、最初にその用語の定義が記述されており、その後ろに本質的な内容や例・注意が記述されている。それらを以下のように区別・定義する。

定義文: 説明文の最初の一文、すなわち“……。”、または“……であって、”までの文。

付属文: 説明文において、“(注)……。”、または“(例)……。”といった文。

具体的内容文: 定義文と付属文以外の文。

3. 関連語の選定

上記に述べた用語や説明文の特性をもとに、以下の評価を行ない、関連語を選定する。また同時に、得られた関連語に意味的關係を与える。

3.1 見出し語からの関連語選定

まず、関連語選定の起点となる用語の見出し語に注目し、見出し語だけで特定できる関連語を選定する。見出し語が複合語の場合、その見出し語を単語に分解し、各単語にカテゴリを与える。そして、以下の選定評価を行ない、該当するものを関連語として抽出する。

選定 (1.1) 見出し語にカテゴリ 1 が現れる場合、

(i) そのカテゴリ 1 の単語が登録語。

⇒ 上位概念 (上位語)

(例) 制御文字 → 文字

(ii) そのカテゴリ 1 の単語を取り除いた登録語。

⇒ 上位概念

(例) 図形文字 → 図形

選定 (1.2) 見出し語の最初にカテゴリ 2(c) が現れる場合、それを取り除いた登録語。

⇒ 反義語

Retrieval Method for Glossary of Related Terms used in Information Processing

Yoshiyasu TERADA, Chun-Wen LIN, Yoshiharu ISHIKAWA and Shunsuke UEMURA

Graduate School of Information Science,

Nara Institute of Science and Technology (NAIST)

選定 (1.3) 見出し語にカテゴリ 3 が現れる場合、それ
を対義の単語に変換した登録語。

⇒ 対義語

(例) アナログデータ ↔ デジタルデータ

選定 (1.4) 見出し語にカテゴリ 2(b) が現れる場合、そ
れ以外の部分が一致する登録語。

⇒ 類義語

選定 (1.5) 部分一致語の中で、

(i) 見出し語の前にカテゴリ 2(a) または 4。

⇒ 下位概念 (下位語)

(例) 制御文字 → 伝送制御文字、書式制御文字

(ii) 見出し語の前にカテゴリ 2(c)。

⇒ 反義語

(iii) 三つ以上の単語から成る見出し語の場合、連
接する末尾二つ以上の単語が一致する登録
語。⇒ 類義語

(例) 伝送制御文字 → 書式制御文字、装置制御文字

3.2 説明文からの関連語選定

次に、見出し語の説明文に対して、その説明文
を定義文・具体的内容文・付属文に区別し、それ
らから登録語と機能表現 (見出し語との関係を規
定する表現) 以外の意味のない語を除去した上で
関連語を選定する。

表 1 機能表現

| | |
|-------|------------------------------------|
| (a) | {。 であって、} |
| (b-1) | {の「一つ」の「一種」の「一分野」 「の「方法」に関する用語} |
| (b-2) | {の「こと」であるもの} |
| (c-1) | {「によって」に従って} |
| (c-2) | {「において」における「に対する」 「に関する」に適用される} |
| (c-3) | {「を処理する」を表す「を制御する」} |
| (c-4) | {「を用いる」をもつ「を含む」から成る} |

{ | } は、{ } 中のいずれの語でもよいことを意味する。

選定 (2.1) 定義文において、その文末に現れる登録語。

“登録語+(a)” ⇒ 上位概念 (上位語)

選定 (2.2) 定義文において、その文末に以下の機能表
現が現れる場合。

(i) “登録語+(b-1)+(a)” ⇒ 上位概念 (上位語)

(ii) “登録語+(b-2)+(a)” ⇒ 同義語

選定 (2.3) 定義文、具体的内容文において、以下の機
能表現が現れる場合。

(i) “登録語+(c-1)” ⇒ 上位概念

(ii) “登録語+(c-2)” ⇒ 上位概念 (使用分野)

(iii) “登録語+(c-3)” ⇒ 下位概念

(iv) “登録語+(c-4)” ⇒ 下位概念 (構成要素)

3.3 見出し語と説明文からの関連語選定

最後に、見出し語からの選定と説明文からの選
定の両方を考慮した関連語選定を行なう。

選定 (3.1) 見出し語の部分一致語と選定 (2) で選定さ
れなかった登録語と一致するもの。

選定 (3.2) その見出し語を説明文中に含む登録語に対
して、選定 (2) を適用する。

3.4 連鎖検索

上記の関連語選定の結果を用語間の階層構造・
意味的構造をふまえて図式的に表現すると図 1 の
ようになる。

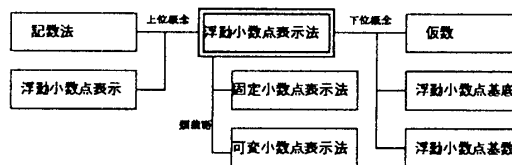


図 1 関連語選定の例

利用者はその中から用語を選択し、連鎖的に関
連語選定を行なうことによって、ある用語・概念・
分野に特化した用語の検索が行える。

4. 問題点と今後の課題

(1) 説明文中の登録語の判別。

(2) 検索手法。

(3) あいまいな質問に対する検索。

参考文献

- [1] 黒橋 慎夫, 長尾 真, 佐藤 理史, 村上 雅彦: 専門用語辞典
の自動的ハイパーテキスト化の方法, 人工知能学会誌,
Vol.7, No.2, March 1992
- [2] Yasushi OGAWA et al: A New Indexing and Text
Ranking Method for Japanese Text Databases using
Simple-Word Compounds as Keywords, DATABASE
SYSTEMS FOR ADVANCED APPLICATIONS '93,
p.197-p.204
- [3] 電子情報通信学会誌, Vol.7, No.9, 9/87, p.878-p.908