

自然言語処理と概念の体系化について(3)

5G-1

田中 康 仁
(姫路短期大学)

吉田 将
(九州工業大学)

1. はじめに

「自然言語処理と概念の体系化」と題して情報処理学会第37回(昭和63年後期)全国大会で発表した。さらにその続編を「自然言語処理と概念の体系化(2)」と題して情報処理学会第38回(昭和64年前期)全国大会で発表した。今回のものはその続編と考えていただきたい。概念の体系化を行うにあたって概念の属性をどのように収集するかということは重要な課題である。ここではUDCの利用を考えてみる。

2. 概念の体系化

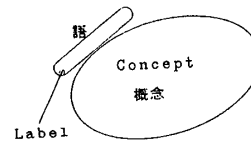
1) 概念の属性データの収集

- 概念の属性データの収集としては既存の知識、辞書から集める。例えば百科辞典から集める方法である。この方法は特定の語に対しては有効であるが全ての概念について網羅的に集めることはできない。
- 日本語の場合"の"による名詞の結合関係より集める。日本語の"の"は英語の"of"に似ている。完全に一致していないが語と語の単純な関係付けに用いられる。この方法はあらゆる語(概念)に対して有効である。集め易いし、集めながら体系化し、不足データを補うことができる。
- 国際十進分類法(UDC)や各種シソーラスを用いる方法国際十進分類法は図書館の本の整理や文献の整理等に広く利用されてきた。また、情報検索等にも重要な役割をはたしている。これは概念のシソーラス体系ができていて、国際的に認められ、同一基準で各国語(英語、フランス語、独語、日本語……)に作られている。このためマルチリンガルの概念シソーラス作成に有効である。しかし、自然言語処理のためには再検討を行い、再構築が必要である。だが何もない状態から考えるよりは、はるかに早く出来ると思われる。

3. 概念の属性

語は概念に付けられたラベルである。

これを図示すると次のようになる。さらに概念の記述内容を示す。



概念の記述

- 1) 語に関する属性
- 2) 語に関する方法属性
- 3) 概念に関する属性

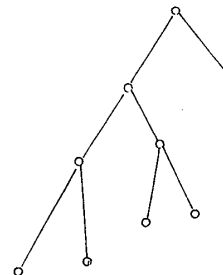
① 内包

② 外延

さらに詳細な内容は省く。

概念の属性を個々の概念ごとに全部集めようとする大変である。

しかし、概念シソーラスを使うとこの労力を軽減することができる。



概念シソーラス

概念シソーラスを利用するメリットを考えてみよう。シソーラスは次のような性質がある。

- 1) シソーラスの個々の要素は共通部分と固有部分を持っている。
- 2) 共通部分を持つことから、属性の記述についての重複をさけることができる。

この意味からも概念シソーラスの利用は重要である。

UDCのシソーラスの中には概念の内包的属性を分類項目に使っているものもある。

例

靴(くつ)

- ・ 大きさ
- ・ 機能
- ・ 色
- ・ 利用者
- ・ 材料
- ・ 形

Natural Language Processing and
Systematization of Concept (3)

Yasuhito Tanaka
Himeji College

Sho Yoshida
Kyushu Institute
Technology

UDCには靴の分類基準がいくつか述べられている。
この分類基準を利用して概念記述ができる。

4. 概念の演算

概念の機械的操作ができないと自然言語処理には使えない。それではどのような操作を考えればよいのであろうか又それは何を意味するか考えてみる。

A概念とB概念の演算

- 1) $A \cap B$ 概念間の共通部分の抽出
- 2) $A \cup B$ A概念、B概念を包含するもの
- 3) \bar{A} A以外の概念

例えば \bar{A} はどのように考えたらよいのであろうか？
例えば $\bar{男}$ を考えると多くの人々は女と答えるであろう。 $\bar{男}$ を石とか自動車と答える人は少ない。これはなぜであろうか。我々は概念シソーラスの上で無意識のうちに考えていて同一概念木の領域内で考えているにちがいない。それ故答えが女となったのである。我々の常識もこのようにして推論しているのであろう。

- 4) 上位概念を求める演算
- 5) 下位概念の1つを抽出する演算
- 6) 概念の内包的属性の1つを取り出す演算さらにその属性の内容を詳細化した内容を取り出す演算
その他の演算が考えられる。それらについては今後の研究に期待していただきたい。

次にこれらをコンピュータ上でどのように表現するかという問題がある。これについても今後の研究に期待していただきたい。

5. マルチリンガル・概念シソーラス

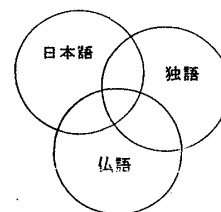
UDCはマルチリンガル・概念シソーラスを作る上からも大変便利である。

しかし、各国で文化、歴史、習慣の相違があり、同一概念体系を作ることは困難である。例を示す。酒を考える。各国では色々な酒を連想する。

例

酒 (アルコール)						
仏	英	独	中	韓	日	USSR
ワイン	スコッチウイスキー	ビール	白酒	しようちゅう	清酒	ウオッカ

酒というものの共通概念は何であろうか。



6. おわりに

今後、概念の体系化のためにいっそう研究してゆきたい。

研究にあたっては(株)情報科学技術協会中村幸雄氏の助言をいただいた。記して感謝の意を表す。

