

# 概念を獲得する機能メモリ

5 F-9

— ハード化を考慮した知識ベース管理方式 —

増田進二 納富一宏 内山明彦

早稲田大学理工学部

## 1. はじめに

データベースの開発においては、データの検索効率を考慮した構造設計と最適なデータの配置および格納が特に重要な問題とされる。これらの問題に対処するためには、データ構造の解析によって、データベース構造の自動的な確定と自動的なデータ抽出後の保存とを優先的に検討していかなければならない。

現在、コンピュータやワープロの普及によって、電子化されたテキストが非常に多く存在し、また生産されている。そこで我々は、これらの既存の電子化文書を利用して、そのテキスト中に含まれる様々な情報を抽出し、データベース化する方略について検討している[1][2]。大量のテキストからの高速な情報収集(知識獲得)では、自然言語処理的な解析を極力排除し、単純な文字列パターンのみを用いて全プロセスを遂行しなければならない。

ある情報表現と密接に関連する典型的な文字列パターンのうち、データベースの構築に適する構造としては、テキストに現れた用語の説明や定義、内容の分類、全体の要約、注意・禁止事項などが考えられる。これらのパターンは電子化された既存のテキストにも多く含まれている。よって、そのような文をテキストの中から抽出し、パターンマッチングを行なって、検索可能なデータ構造を作ることができれば、データベース開発に役立つはずである。

さらに近年、機能メモリが注目されている。機能メモリでは、ハードウェアレベルでデータベース処理を実現できる可能性があり、先の高速度な情報収集方略(アルゴリズム)の場

合にも、機能メモリとしての実現可能性が存在する。

本稿では、上記の機能メモリの開発に向けた、テキストからの概念獲得の手法について報告する。

## 2. 概念の獲得

文字列パターンを用いて、テキスト中から必要な文を抽出する(図1)。例えば、「A

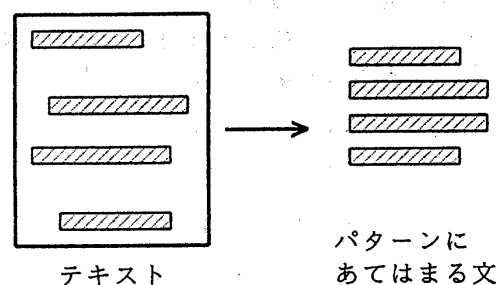


図1 文の取り出し

はBである。」という文がある。このパターンに当てはまる文は、AがBに所属しているという、AとBの所属関係を表している。

(「カナリアは鳥である。」という文は、カナリアは鳥の一種であるという意味である。) この「AはBである。」というパターンに当てはまる文から、次のようなデータの階層構造を作ることができる(図2)。

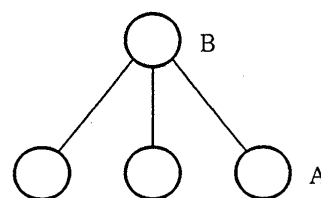


図2 階層構造

ここで、図2のような階層構造を概念と定義し、このような階層構造が作れたことを、

Functional Memory for Concept Acquisition

Shinji Masuda, Kazuhiro Notomi and Akihiko Uchiyama

School of Science and Engineering, Waseda University

概念の獲得と呼ぶことにする。また、図2のAはBに所属するものを表す場合と、Bの性質（属性）を表す場合とがある。

文を抽出するときの主なパターンを表1にまとめる。

所属を表すもの	～は～である。 ～は～に属する。 ～は～の一種である。
属性を表すもの	～は飛ぶ。 ～は泳ぐ。 ～は走る。

### 3. 概念獲得の処理手法

テキストから、概念を獲得するための処理手順をまとめると図3のようになる。

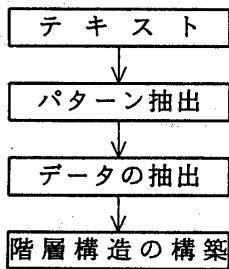


図3 概念獲得の処理手順

テキストの入力は既存の電子化されたテキストを使用する。このテキストは特にこのために作られたものでなくてよい。テキストの中から、いくつか用意されたパターンにあてはまる文を抽出する。抽出された文から、必要なデータを取り出し、そのデータから階層構造を構築する。

例として、次のような文がテキストから抽出された場合を考えてみる。

- ・鳥は動物である。
- ・猿は動物である。
- ・燕は鳥である。
- ・鳩は鳥である。
- ・動物は食物を取る。

- ・鳥は飛ぶ。
- ・猿は寝る。

これらの文から、次のようなデータ構造が構成される（図4）。図4において、“inst”は所属を、“attr”は属性を表している。

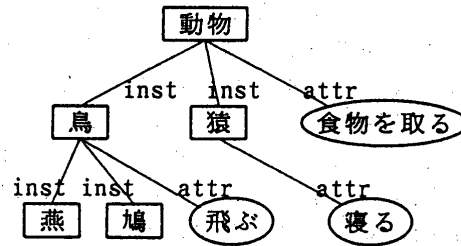


図4 データ構造

### 4. おわりに

本稿では、テキストから概念を獲得する手法について述べた。この手法の利点は、パターンとの比較とデータの切り出しという簡便な処理のみで実現できる点である。

本手法の問題点としては、次のことがあげられる。入力文がパターンに当てはまるか当てはまらないかだけを見ているため、入力文の真偽によっては信頼できないデータ構造ができてしまう。また、パターンに当てはまったとされた文が、実際は正しく当てはまらないことも生じる。

今後は、これらの問題点について検討し、本手法を取り入れた機能メモリの開発を行っていく予定である。

### 参考文献

- [1] 木子, 納富, 内山: 「自然言語解析における概念文化の手法」, 早大情科センター紀要, Vol. 11, Spring, pp. 28-31(1990).
- [2] 白石, 納富, 内山: 「概念の動的管理システム」, 情処自然言語処理研究会報告, 85-9, pp. 65-72(1991).