

4J-8 知的検索における 自然言語インタフェースの一手法

堀内 正人、 仮山 周一、 土金 由幸、 村岡 輝雄
日本ビクター(株) 総合技術研究所 第八研究室

1. はじめに

近年の計算機環境の目ざましい発展、とりわけ、外部記憶装置の大容量化・マルチメディア化は、情報検索のあり方にも大きな影響を及ぼしつつあるが、検索対象データの規模の拡大や多様性に対応するために、知識を積極的に利用する方法が考案されている[1]。本稿では、検索対象データをより抽象的なレベルで扱うことができ、かつ、柔軟な検索を可能とする自然言語インタフェースの一手法を提案し、これを実現させるために必要な知識の枠組みについて、筆者らが進めている解剖学画像検索システム[2][3]を例に考察する。

2. 事項検索と名詞句モデル

事項検索(fact retrieval)は、問い合わせに対する答として検索結果をダイレクトに表示するものであり[4]、マルチメディア・データベースの場合、文書・音声・画像・数値などのデータが、その検索対象になり得る。事項検索に於ける問い合わせでは、最終的に期待する答は対象データそのものであり、自然言語インタフェースを実現させる上では、データそのものに直結するところの名詞句の解析機能を強化する必要がある[5]。

本手法に於いては、情報検索用の自然言語インタフェースに用いられる名詞を、図1に示す区分に従って分類している。図1に於ける各分類(obj_abs, obj_con, ...)は、原則的には、語彙そのものの内的性質としてア・プリオリに定義できるものである。検索対象データは

obj_con 中の名詞の属性の一つとして扱うので、自然言語の世界での語義の階層性 (= 検索対象データの抽象レベル) に捕らわれることなく検索を行うことができる。

図1の名詞分類を利用することによって、問い合わせ文に出現する名詞間の修飾関係の中で、情報検索にとって全く意味の無いものを早期に除外することが可能となる。

3. 名詞句モデルの生成

本手法の中心的な考え方の一つである、問い合わせに於ける「名詞句モデル」とは、次のようなものである。

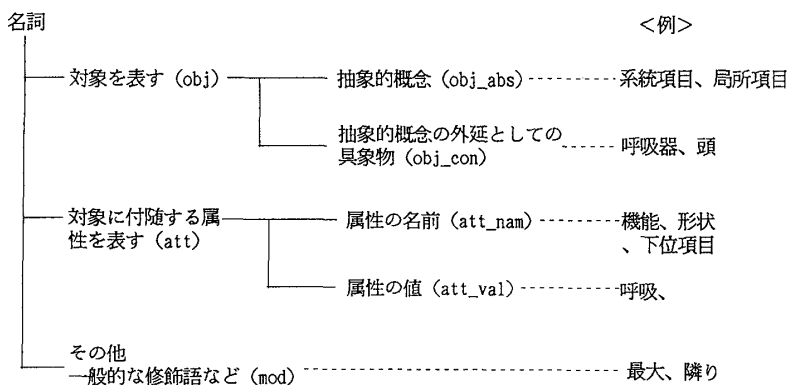
検索対象を表す名詞句は、

- a) その名詞句のレベルで機能するもの
- b) その名詞句の指示対象のレベルで機能するもの

の2つに大別されるが[6]、a)においては特定化された検索対象項目そのものを、また、b)においては名詞句に示された性質を充足する集合及びその要素を、それぞれ問い合わせの対象にしているものとする。「問い合わせとして意味を成さない名詞句」とは、検索対象項目が特定化されなかったり、集合の要素が数え上げられなかったりする名詞句である。

問い合わせ文中の名詞句を解析することによって、名詞句モデルとして、a)についてはインスタンス・オブジェクトが、また、b)については示された性質を持つ集合変数が、それぞれ生成される。

問い合わせ文の解析に基づいた名詞句モデルの生成は、



obj_abs	の	obj_con
att	の	obj_con
att_con	の	att_abs
x	の	att_con
x	の	x

図2 問い合わせとして意味を成さない名詞句の例

x	の	obj_abs	→	obj_con	の集合
x	の	att_nam	→	att_val	の集合

図3 名詞句の複合による意味属性の変化

図1 情報検索に用いる名詞の分類

語彙機能文法 (LFG) [7] [6] の枠組みを利用して行う。

その基本戦略は、次の2項に示す通りである。

(1) 各名詞の語彙記述 (f-description) の中に、名詞句モデルの生成に必要とされる意味マーカを、属性の一つとして付加する。これを利用して、名詞間の修飾関係をチェックしながら f-構造をビルドアップしていく。図1に記したような意味マーカを利用することによって、例えば、「名詞句 a」+「の」+「名詞句 b」の形の名詞句であっても、図2に示すような組合せの場合には、問い合わせ文として意味をなさないものとして処理する。また、解析結果の複合名詞句についての意味属性は、図3のように変更される。

(2) f-構造を、統語構造のゆらぎに強い名詞句モデルに変換する操作を行う。f-構造は、統語構造の情報をそのまま保持する訳ではないが、表層文に於ける表現の多様性の影響を少なからず受ける。ここでは、異なる統語構造を持つ名詞句から、それらの属性を利用して同一の名詞句モデルを生成することにより、表層表現の多様性を吸収する。その方法は、現在のところヒューリスティクスに頼るものである。

4. 問い合わせ文における省略表現・指示詞の解析

本手法のもう一つの特長は、問い合わせ文及び検索結果についての名詞句モデルを対話制御部に保持することによって、問い合わせ文における省略表現・指示詞の解析に柔軟に対応できる点にある。この場合、参照される名詞句モデルが、問い合わせ文中／検索結果のどちらであるかを判別する処理が重要となる。本手法では、省略表現は問い合わせ文を、また、指示表現は検索結果を、それぞれ参照することを基本戦略とするが、特に省略表現は多分に曖昧性を持っており、“でしゃばらないインタフェース”を実現させる上では注意する必要がある。省略表現・指示詞を含む問い合わせ文の解析例を図4に示す。

```

nalana : 【 頭の下位項目 】          /* 質問 */
engine : 頭の下位概念を表示します    /* 推論エンジンの返答 */
engine : 【 部位 】
engine : 【 [前頭部, 頭頂部, 側頭部, 後頭部, 耳介部, 乳突部] 】
nalana : 【 顔について 】
engine : 顔の下位概念を表示しますか ( y/n ) ?
nalana : 【 y 】
engine : 顔の下位概念を表示します
engine : 【 部位 】
engine : 【 [眼窩部, 眼窩下部, 鼻部, 口部, 頬部, 頬骨部, 耳下腺咬筋部, オトガイ部] 】
nalana : 【 その中で呼吸に関係するもの 】
engine : 呼吸に関する部位を表示します
engine : 【 部位 】
engine : 【 [鼻部, 口部, 頬部, 頬骨部] 】

```

図4 省略表現・指示詞を含む問い合わせ文の解析例

5. おわりに

本稿では、知的情報検索における自然言語インタフェース実現の一手法について、解剖学画像データを対象とした事項検索システムを例として論じた。本手法の特長は、事項検索における問い合わせ文中の名詞句の重要性に着目した点にあり、語彙機能文法の枠組みを利用した「名詞句モデル」を生成し、これらを対話制御部のワーキングメモリに保持することにより、問い合わせ文の省略表現・指示詞の解析に対応することが可能である。

本手法に於いては、検索対象データに依存した語彙レベルの知識と、f-構造→名詞句モデル変換、及び対話制御のための、ヒューリスティックな知識という、相異なる2つのレベルの知識を利用している。特に前者の知識の場合、語彙記述に埋め込まれる形で使われるため、知識のモジュラリティを高めるためには工夫が必要となる。また、本手法で対象とした解剖学に関する知識は、極めて静的なものである。知的検索のためのユーザ・フレンドリな自然言語インタフェースを実現するために、更に検討を加えていきたい。

参考文献

- [1] 藤澤他：“概念ネットワークを用いた知的ファイリングシステム”、電子情報通信学会研究会、0586-48 など
- [2] 土金他：“解剖学を対象とした知的検索技術の開発”、情報処理学会第36回全国大会論文集
- [3] 土金他：“知的検索に於ける知識構造間のクオリファイについて”、情報処理学会第37回全国大会論文集
- [4] 伊藤：“情報検索”、ソフトウェア講座19、昭晃堂
- [5] 藤崎：“日本語によるデータベース照会”、高橋編「日本語情報処理」、近代科学社
- [6] Bresnan, J ed.：“The Mental Representation of Grammatical Relations”、MIT Press
- [7] 淵監修：“自然言語の基礎理論”、知識情報処理シリーズ4、共立出版
- [8] Mellish, C. S., 田中訳：“コンピュータのための自然言語意味理解の基礎”、サイエンス社