

OODBによる医療情報の意味論的情報表現*

5C-2

福田典夫, 津本周作, 田中博†

東京医科歯科大学難治疾患研究所情報医学研究部門医薬情報‡

1. はじめに

近年の診療情報システムの発展に伴い、患者診療情報の電子化とその蓄積は急速に進んでいる。しかし、現状では、医学概念・診療行為の意味論的考察に基づいた医療情報のスキーマが確立していないため、従来のDBスキーマでの単純なデータ蓄積が行われており、これらの診療データから事後、臨床医学的知識を抽出したり、今後の診療に役立つ症例データベースに編成し直すことはきわめて困難である。

我々は現在、診療の行為がどのような基本構造になっているか、あるいは医療の知識や情報にどんな基本カテゴリーが存在してそれぞれが診療のどの部面で働くか等、これらの枠組みを与える「診療のスキーマ」を意味論的処理から進展したオブジェクト指向分析 [2] に基づいて構成しつつあり、この診療・医学オブジェクト構造を用いて、多様なメディアからなる医療情報を変換し診療DBとして編成する方法を検討中である。我々は現在開発中のシステムを OSIRIS (Object-oriented Semantic Information Retrieval and Inference System:開発中) [1] と呼んでいる。

本大会では、この OSIRIS プロジェクトの中で、既に開発した自然言語で記載された入院患者の病歴要約(「退院時サマリー」)をその時間関係を基本にしながら病態のオブジェクト構造に併せて event 系列に要約するシステム APRIL(Automated analyzer for Patient Records based on Interval Logic)[3] について報告する。

2. 診療情報のオブジェクト構造

医療情報のオントロジーは次の3つからなっていると考えられる。

(1) **Concept ontology**: 診療の対象は疾患・病態である。この医学知識の構成特に疾患の階層的構成や症状・検査値との関連をオブジェクト構造にしたものを指す。

(2) **Task ontology**: 診療行為のオブジェクト構造を指す。診療行為がいくつかの自律的単位からなると考えるなら、それらの単位がオブジェクトである。例えば、問診、観察、検査オーダー、処方、疾病別診療プロトコール、治療などはそれぞれオブジェクトこれらが一緒になって診療というタスクを構成している。「診療のオブジェクト構造分析」が医学的言表の意味を抽出する基本枠組みとなる。

(3) **Event ontology**: 具体的には診療は、疾患に対する時間的な病態-診療行為 > Event 系列として現象し、医療情報は Event-Sequence として一次的には格納される。この Event-Sequence は患者を上位オブジェクトとして時間の granularity に従って階層化される。末端の instance は、診療日の具体的な病態-行為 > である。それらが抽象されたものが中間オブジェクトである。

さて以上の3つの Object-Ontology 間は、インスタンス結合やメッセージ通信によって連結される。医療情報は患者の主訴、理学所見、臨床検査、放射線検査などコード数値、自然語テキスト、画像、グラフなど多岐にわたる。ここでは、以上の情報構造に合わせてオブジェクト指向DB言語 Ontos を用いてこの構造を実現した。ここではその中で、医師が叙述した自然語テキストから時間要約の Event 列を抽出して EVENT オブジェクトに格納する知的 Interface 部分 APRIL を以下に述べる。

3. APRIL

APRIL は自然言語で構成された病歴から、時間情報の一階述語論理による記述にいったん抽出、domain knowledge を利用し、オブジェクトの形式に再構成する。APRIL は簡単な自然言語処理を行うための解析部と解析に必要な

*Semantic Representation of Medical Information based on OODB

†Norio Fukuda, Shusaku Tsumoto and Hiroshi Tanaka

‡Medical Research Institute, Tokyo Medical and Dental University 1-5-45 Yushima, Bunkyo-ku, Tokyo 113, Japan

知識を加えた辞書部とにより構成されている。このプログラムのアプローチは、自然言語としてきわめて制限された形をとっている病歴から時間的な知識を取り出してくるもので、いわゆる semantic analysis をオブジェクト指向的に捉らえ直したものである。

ここでは、次の様な病歴の自然言語としての特徴を利用している: 1) 統語論的性質: 病歴に含まれる構文は、ほとんど名詞句+動詞という形で構成される。動詞自体も名詞を修飾するものか、名詞で表現できるものを名詞化したものであり、名詞とそれを修飾する句によって症状を表現する傾向がある。2) 意味論的性質: 症状の経過を伝えるという目的があるため、記述はほぼ時系列的に書き下されている。記述の無矛盾性は強く要請されているため、文脈上大きな矛盾を引起す文は現れてこない。したがって、名詞句及び動詞の抽出及びその意味の retrieval が解析部の本質であり、我々の開発したプログラム APRIL では、EVENT オブジェクトを次のような手順に沿って順次、文を解析しながら生成していく。

1) 構成素への分解: システムは1つの文を得た時、それを時間を表現する句、名詞句、動詞句の3つの構成句に分解する。厳密には、時間を表現する句は動詞の修飾句に含まれるが、我々の目的上、時間的な要素が重視されるため、特別に別の構成句と仮定している。

2) 構成素の評価: 構成素に分解した後、次のような手順で評価する。

(i) 時間の表現句の評価: 絶対的な値として与えられている場合、そのまま値が評価されるが、文で相対的な時間関係が記載されている場合は、基準となる年月日を参照し、それが明示的に記述されない場合は、参照できるまで拘束条件としてその決定を保留しておく(拘束充足法)。そして、最終的に病歴をすべて解析した後、時間に関する情報をまとめて矛盾がないかを調べる。(ii) 名詞句の評価: 名詞句で表現される症状に対して、症候のオブジェクトが駆動される。このオブジェクトからの情報を参照し、その症状の特徴及びとそれに関連する病態が検索され、さらに類似の症候の存在について検索を進める。例えば、「歩行時のふらつき」であれば、それに関連する病態は「下肢を使用する運動が拙劣となる」であり、これと類似の病態が「上肢を使用する運動が拙劣となる」というものである。これらの記述を利用して、病歴上、以前解析した文で同様の病態を表現したものがいないかを調べ、そのような記述があった場合は、それらの文の解析結果を示すオブジェクトと、この名詞句によるオブジェクトとをリンクさせる。

(iii) 動詞句の評価: 動詞句が、名詞句の発生、推移を表現する場合は、それをこの句の評価として与える。例えば、「起こる」という動詞句は、「発生した」という意味をもつので、この文の時間表現句で表される時間に発生したと評価され、この「発生」に関するオブジェクトが検索され、名詞句によるオブジェクトの属性が与えられる。動詞句が症状を表現している場合は、評価は名詞句の場合に準じる。

3) 評価の assertion: それぞれの句を評価した後、各オブジェクト間のリンクを生成し、時間論理に基づいて、このリンクによるオブジェクト間の構造についても検索する。例えば、症候が慢性に進行している、急性か、発作性か等の構造を抽出し、これをオブジェクトとして生成する。

4. おわりに

以上、診療情報のオブジェクト構造とそれを抽出する知的 interface プログラム APRIL について報告した。この方法の有効性については、紙面の関係上論じることができなかったが、本大会にて供覧することとしたい。また、現在、他の診療情報の記述についても検討、開発中であり、今後その成果について報告していく予定である。

参考文献

- [1] 田中博. オブジェクトデータベースと医療情報. 第12回医療情報学連合大会論文集, 1992.
- [2] Code, P. and Yourdon, E. *Object-Oriented Analysis. 2nd Edition* Prentice Hall, NJ, 1991.
- [3] Tsumoto, S., Tanaka, H. and Satomura, Y.: Extraction of Temporal Knowledge from Patient Records. *Proc. of MEDINFO92*, 1992.