

概念検索機能を備えた自然言語医療情報システムの試作

西村 靖司† 白石 正人†† 小田 誠雄† 横田 将生†††

†福岡工業短期大学 ††福岡教育大学 †††福岡工業大学

我々は、カルテに記述された医療情報を蓄積して、自然言語によるデータベース問い合わせが処理可能なシステムを目指している。今回、医学用語辞典の分類記号に着目し、概念表現による医療情報の検索方法について考察すると共に、概念検索機能を備えた自然言語医療情報システムを試作したので、その概要について報告する。

An Experimental Natural Language Medical Information System with Conceptual Retrieval Function

Yasushi NISHIMURA† Masato SHIRAISHI†† Seio ODA† Masao YOKOTA†††

†Fukuoka Junior College of Technology ††Fukuoka University of Education
†††Fukuoka Institute of Technology

The authors have been accumulating medical information described in the clinical records and are aiming at the construction of the system which can inquire the database in natural language. This time, we paid attention to the classification sign of a medical term dictionary. So we considered the retrieval method of medical information using conceptual representation, and developed a prototype of natural language medical information system with conceptual retrieval function. In this paper we report the outline of them.

1 はじめに

これまでに、医療を目的とした実用的患者情報システムが数多く開発されてきている。その大部分は数値あるいは符号化された定型データを対象としており、自然言語で記述された文章データは取り扱っていない。しかしながら、文章データは患者情報のより大きな部分を占めており、治療や診断により重要と考えられる。

現在、我々はカルテに記述された医療情報を蓄積して、自然言語によるデータベース問い合わせが処理可能なシステムの実現を目指している。これまでにキーワード検索機能および統計処理機能を備えた医療情報システム [1] を試作してきている。今回、このシステムに概念検索機能を組込むことにより、表現の違いによる影響を受けにくい医療情報の検索が実現できた。ここでは、概念検索の方法と自然言語医療情報システムへの応用について述べる。

2 対象となる退院サマリの内容

対象とする医療用データとして、九大病院循環器内科の退院サマリ（医療カルテの一種）から文章を収集し、現在、339名12552行の退院サマリを計算機に入力蓄積している。図1に退院サマリ文の例を示す。

(入院時現症) abdomen:liver;not palpable, no tenderness, pretibial pitting edema (+)(bil.).
 (検査)(1) blood gass;room air Po2 62.1, Pco2 38.1, pH 7.44
 4, HCO3 26.0 (2) CBC Ht 42%, WBC 8400 (3) ESR 36/68 (4) 血清生化学 GOT 191, GPT 240 (5) 血清学 CRP 1+ (6) Ccr=69.3ml/m (7) 肺機能 %VC 67%, %FEV1.0 72%.
 (入院中の経過) 入院時 orthopnea の状態で, wheezing が著明であった。胸写上, 肺血管陰影はやや増強していた。同時に黄色粘稠痰と 37°C 台の発熱を認めた。心不全と考え, lasix を投与したが, dyspnea, wheezing は完全には改善しなかった。dyspnea に対して neophylline は effective であった。気道感染ありと考え, ABPC 1.0g*2 を開始し, neblizen にて気管支拡張剤, 去痰剤を投与したところ, sputum は白色となり, 量も減少し, dyspnea, wheezing は消失した。2DSS では, MVA=1.5~2.0cm↑2, LAD=4.2cm, LV function 良好であった。doppler にて Mr, Ar は negative であった。血圧正常。胸痛の Hx なし。arrythmia の自覚なく, Holter ECG にて arrythmia を認めなかった。

図1 退院サマリ文の例

3 概念検索の方法

まず、概念検索の有用性について考察する。例えば以下の様な表現について考えてみる。

- 例1) 高血圧の患者に降圧剤を投与した。
- 例2) 高血圧の患者がレセルピンを服用した。
- 例3) 高血圧の患者にメチルドバを開始した。

これらの例ではそれぞれの表現は違っているが、基本的な概念(意味)構造は同一である。したがって、「高血圧の患者にどのような薬を投与しましたか?」という質問文に対しては、例1の様な表層的なマッチングに加え、概念レベルでのマッチングを行なうことで、例2,3も検索の対象となり、質問文に対する回答として、「降圧剤」、「レセルピン」、「メチルドバ」の3つの薬品名を得ることができる。

今回、この概念検索を実現するため、退院サマリに出現する動詞とその対象事物(～をの前のくる単語)の種別コードを組み合わせることで概念を定義した。ここで、種別コードとは表1に示す様なCD-ROM版「25万語医学用語大辞典」[2]の全ての用語に付加された分類情報のことでMeSH(Medical Subject Headings)に準拠したものである。これにより、上記の例の場合では「投与する」、「服用する」、「開始する」等の動詞とそれに対応した【薬化】という種別コードを持つ対象事物の組み合わせによって、「医療従事者が患者に薬を投与する」という概念が定義でき、概念レベルでの検索が可能となる。

表1 医学用語大辞典の種別コード(一部)

種別コード	種類	用語例
【診療】	分析、診断、治療の技術	はり麻酔法
【疾患】	病気	心弁膜炎
【症状】	病気の状態	よろめき歩行
【機器】	機器、用品	毛細管血圧計
【薬化】	薬品 および化学物質	アザプリン
【生物】	生物	アメーバ類
【材料】	医科・歯科材料	インプラント材

4 システムの全体構成

本システムは図2に示す様に Sun 4/10,2 と NeXT Station で構成されるサーバ・クライアント型のシステムで、以下の部分から構成されている。

1. 検索データベース生成部

退院サマリから必要な情報を抽出し、検索処理部で使用するためのデータベースを生成する。

2. 検索処理部 (サーバ)

ユーザインターフェース部と係をとりながら動作し、検索や統計処理を行う。さらに検索処理部は以下のサーバと通信しながら処理を行なう。

(a) 形態素解析サーバ

形態素解析プログラム `sepassp[5]` をサーバ・クライアント型のプログラムに修正したもの。

(b) CD-ROM サーバ

CD-ROM 検索プログラム `dserver[6]` を医学用語大辞典の検索用に修正したもの。

3. ユーザインターフェース部 (クライアント)

医師からの質問文の入力と検索結果を表示する部分で、検索処理部とのコミュニケーション部とユーザインターフェース部から構成される。

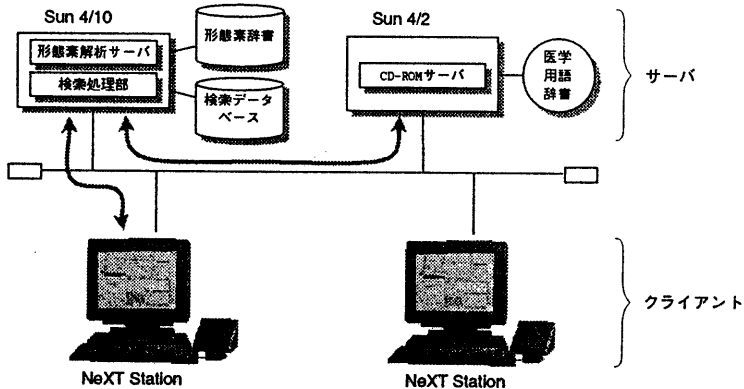


図2 システム構成

5 検索データベース生成部

検索データベース生成部は、医師からの質問に対してリアルタイムに検索結果を表示するために、退院サマリから検索の対象となりうるような属性値を持つ名詞 (例えば血圧や体温) や動詞を抽出し、あらかじめ検索のためのデータベースを作成する役割をもつ。

検索データベース生成部の流れを図3に示す。処理手順としては、まずキーワードファイルを入力として、退院サマリとのパターンマッチングを行なう。つぎにマッチングした文の形態素解析を行った後、語の分類コードを検索し、最終的に概念辞書を参照しながら検索データベースを生成する。

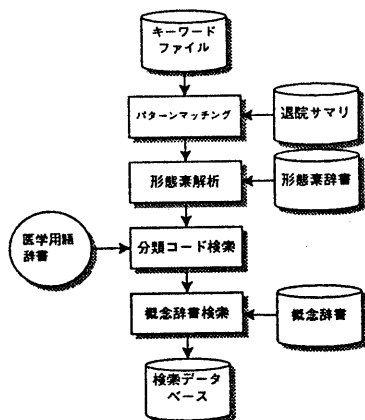


図3 検索データベース生成部

5.1 キーワードファイル

キーワードファイルは、医師が質問すると予想される事項についてのキーワードを登録したもので、キーワードごとに属性値や対象事物を抽出するための正規表現のメタ文字（メソッド）が指定できる。キーワードファイルは以下のように記述される。

<形式> {Kwd:Method}

但し、Kwd はキーワード、Method はパターンマッチングの式を表す。

なお、Method が NULL のときはデフォルトのメソッドが与えられる。属性のメソッドの場合、まず (1) でパターンマッチングを行い、これに失敗した場合には、(2) でマッチングを行う。

(1) \$line =~ /\$kwd[をにのはがと]([^-ん%]+)/;

(2) \$line =~ /\$kwd[= : はがもを]*([0-9.]+)/;

但し、\$line は入力文、\$kwd はキーワード、= はパターンマッチを行うこと、[...] はどれか1つの文字にマッチ、+ は直前のパターン要素に1回以上マッチ、* は直前のパターン要素に0回以上マッチ、(...) は括弧の内側のパターン要素を1つの要素にまとめることを示している。

5.2 概念辞書

キーワードファイル中の動詞と概念記号および種別コードを結びつけているのが概念辞書である。その形式と例を以下に示す。

<形式> {Verb:(Cpt Cls)*}

但し、Verb は動詞、Cpt は概念記号、Cls は種別コード、(...) * は繰り返しを表す。

<例>

投与する:C11【薬化】

服用する:C11【薬化】

開始する:C11【薬化】,C21【診療】

施行する:C21【診療】

5.3 検索データベース

キーワードファイルを基に、退院番号と属性値または対象事物に関する情報をメソッドに従って退院サマリの全文から抽出したものが検索データベースである。その形式と例を以下に示す。

<形式> {Kwd:Tno:Val}

但し、Kwdはキーワード、Tnoは退院番号のリスト、Valは属性値または対象事物のリスト。

<例>

```
A R :4967,4968,4968,4968,4968,4994,4994,4996,4999,5002,5002,5002,5002,5002,5005,
5011,5013,5013,5013,5013,5015,5015,5015,5015,5020,5024,5026,5026,5042,5046,
5322,5323,5329,5329,5329,5339,5341,5341,5348,5348,5348,5352,5352,5359,5377,
5383,5388,5390,5390,5392,5408,5408,5414,5414,5416,5416,5416,5416,5416,5438,
5441,5449,5449,5453,5457,5457,5457,5457,5457,5468,5476,5479,5494,5494,5499,
5499,5499,5499,5511,5511,5511,5511,5515,
:,, (-) , (+) , (+) ,,,,,, (-) , (-) ,,,,,, (+) ,, (-) , (-)
(-) , (-) , (-) ,, (+) , (-) ,, (+) ,, (+) ,,,,,, (-) ,,,,,, (-)
, (+) ,,,,,, (-) ,,,,, (-) ,, (+) ,,,,,, (-) ,,,,,, (-) , (-)
C11 :4964,4966,4966,4966,4967,4967,4967,4967,4967,4968,4971,4971,4971,4971,
4971,4972,4974,4974,4974,4975,4976,4976,4976,4976,4976,4977,4981,4981,4988,
4989,4990,4991,4991,4993,4993,4994,4998,5003,5004,5005,5005,5005,5005,5007,
5008,5008,5008,5008,5009,5010,5011,5016,5016,5019,5020,5022,5024,5024,5024,
5027,5029,5029,5029,5030,5030,5030,5032,5032,5033,5033,5033,5035,5035,
:Fluitran ltab【薬化】 ,Panaldine【薬化】 ,リオレサル
【薬化】 ,Panaldine【薬化】 ,steroid【薬化】 ,digoxin
【薬化】 ,lasix【薬化】 ,lasix【薬化】 ,去痰剤【薬化】 ,去痰剤【薬化】
,降圧剤【薬化】 ,Herbesser【薬化】 ,Nitorol【薬化】 ,Nitorol
8T【薬化】 ,Herbesser 8T【薬化】 ,錠剤【薬化】 ,RFP I
NH【薬化】 ,Mercazole【薬化】 ,利尿剤【薬化】 ,ヘルベッサ-2T【薬
化】 ,Mexiletine 600mg単独【薬化】 ,methyldigoxin
```

6 検索処理部

検索データベース生成部で生成された検索データベースを利用して実際に検索を行なうのが検索処理部である。検索処理部の処理の流れを図4に示す。検索処理部では、起動時に全ての検索データベースをメモリ上の連想配列に読み込む。クライアントのユーザインターフェース部より送られてきた質問文は、まず形態素解析された後、キーワード抽出が行なわれる。続いて、このキーワードを基に医学用語辞書と同義語辞書を検索する。ここで得られた同義語をキーワードとして検索データベースに対して条件検索を行い、マッチする退院番号と属性値をリストとして取り出す。

また、概念検索部では、質問文解析部から渡された動詞から概念辞書を検索し、対応する概念記号を得る。次に検索データベース中の概念記号を検索して、該当する対象事物のリストを取り出す。こうして得られたそれぞれのリストに対し、最後に質問文の指示に従って統計処理等が行われ、その結果がユーザインターフェース部へと送られる。

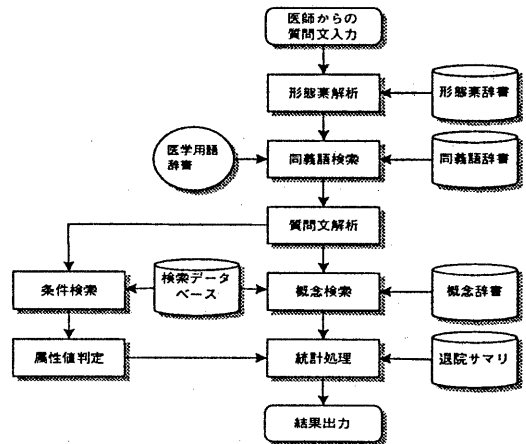


図4 検索処理部

7 検索結果

本システムを使って、実際に339名分(12552行)の退院サマリを対象に検索を行なった結果を、ユーザインターフェース部の画面と共に以下に示す。



図5 検索結果の例

8 まとめ

自然言語医療情報システムに概念検索機能を組み込んだことで、表現の違いによる影響を受けにくい情報検索が実現できたと考えている。今後の課題としては、検索データベースへの時間パラメータの取り込みや、複合概念(動詞概念が2つ以上存在するようなもの)の検索機能の実現があげられる。

参考文献

- [1] 西村他：自然言語医療データ管理システムについて，電子情報通信学会技術研究会報告(1993)
- [2] 「25万語医学用語大辞典」電子ブック，医学用電子化AI辞書研究会(1991)
- [3] 横田他：自動理解処理を目的とする退院サマリの体系的分析，情報処理学会論文誌，Vol.29，No.12(1988)
- [4] 横田：退院サマリの理解処理 - 接続した名詞の意味解釈 - ，電子情報通信学会技術研究会報告，NLC88-11(1988)
- [5] ICOT:ICOT フリーソフトウェア No.33「形態素辞書」(1993)
- [6] 西岡:CD-ROM 検索プログラム dserver(1991)