

自然言語医療情報システムの改良について

西村 靖司, 小田 誠雄 (福工短大)
白石 正人 (福教大), 横田 将生 (福岡工大)

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処理が可能なシステムの構築を目指している。これまでに検索および統計処理機能を備えた自然言語医療情報システムのプロトタイプが完成している。しかし、このシステムでは単純な質問文しか処理できなかつたり、カルテに添えられた図を取り扱うことができないなどの問題点があった。今回、これらの問題に対処するため、システムの質問処理部およびユーザインターフェース部の改良を行ったので、その概要について報告する。

On Improvement of Natural Language Medical Information System

Yasushi NISHIMURA, Seio ODA (Fukuoka Junior College of Technology)
Masato SHIRAIISHI (Fukuoka University of Education)
Masao YOKOTA (Fukuoka Institute of Technology)

The authors have been accumulating medical information in clinical records and aiming at construction of a system which can have an access to the database by natural language. We have already completed a prototype of natural language medical information system with retrieval and statistical processing functions. However, there were some such problems that it was hard to process complicated questions and have figures affixed to the texts with this system. This time, to solve these problems, we have improved the part of question processing and the part of the user-interface. In this paper, we report the outline of the improvement.

1 はじめに

これまでに我々は退院サマリ (医療用カルテの一種) の文章をコンピュータに蓄積し、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開発してきている^[1]。しかし、このシステムでは単純な質問文しか処理できなかつたり、カルテに添えられた図を取り扱うことができないなどの問題点があった。今回は、これらの問題に対処するため、質問処理部を形態素解析サーバと連携処理させるよう改良を行った。また、ユーザインターフェース部に WWW(World Wide Web)システムを採用し、図や医療画像なども取り扱えるようになった。本稿ではこれら2つの改良点について報告する。

2 取り扱う医療データ

本システムでは、医療データとして退院サマリ文と医師質問文の2種類を取り扱っている。

2.1 退院サマリ文

検索対象とな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から文章を収集し、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 7.444, 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, HolterECGにてもarrythmiaを認めなかった。

図1 退院サマリ文の例

2.2 医師質問文

医師が退院サマリからどのような情報を得ようとしているかを実際に調査し、自然言語質問文として167例の質問文を収集した。表1はこの質問文を種類別に分類し、その頻度と例をまとめたものである。これらの例から推察されるように、医師が通常の診療業務で個別に退院サマリを利用する以外に、退院サマリ全体を対象として何らかの情報を得ようとする場合、その内容は一次的な情報に対しての単なる事実検索よりも、統計的処理を施した二次的情報を主に期待していることがわかる。

表1 主な質問文の種類と例

種別	頻度	質問例
確認	6	〇〇氏は以前から肝機能障害があったのでしょうか？
誰	10	去年の4月に土曜日の夜ICUに入院した患者の名前は何でしたか？
何時	4	急性心筋梗塞で当科入院した患者の死亡率はいつ頃が高いのでしょうか？
期間	10	CCU入院患者の平均入院期間は？
比率	35	これまで入院したHOCMの患者さんの男女比は？
最大・最小値	3	当科で心カテを行なった最年長者は？
平均	4	最近3年間に入院したIHDの人のcholesterolの平均値は？
人数	11	心カテ中にcomplicateが起こった患者は何人ですか？
頻度順	19	当科にSBEで入院した患者の起炎菌は何が多いのでしょうか？
種類・分類	4	〇〇氏の治療中に出現した薬による副作用の種類は？
相関関係	9	Anticoagulationをしている患者のトロンボテストの値とEmbolic episodeの危険率との関係は？
比較	5	AMIの患者で抗狭心症薬を投与していた群としなかった群とではpostinfarction anginaの発生率に有意差はあるのでしょうか？
検査値表示	5	〇〇さんの前回入院時のHb, Htは？
一覧表示	24	これまで入院したARの患者さんのetiologyはどうなっていますか？
サマリ表示	2	〇〇さんが以前入院した時の診断、経過、治療方針を教えてください？

3 システムの概要

本システムの構成を図2に示す。本システムでは、医療データベース生成部で退院サマリから必要な情報をあらかじめ抽出し、検索処理部で使用するための医療データベースを生成している。

ユーザインターフェース部では、WWWクライアントから送られた質問文がWWWサーバを経由し、質問処理部へ送られる。質問処理部では、送られてきた質問文の形態素解析を行い、続いて質問文解析処理が行われ、キーワードが抽出される。このキーワードから同義語辞書を検索した後、得られた同義語を使って医療データベースに対する条件検索を行い、マッチする退院番号と属性値をリストとして取り出す。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返される。

なお、質問処理部の形態素解析処理については、別プロセスで動作している形態素解析サーバと連携処理を行っている。

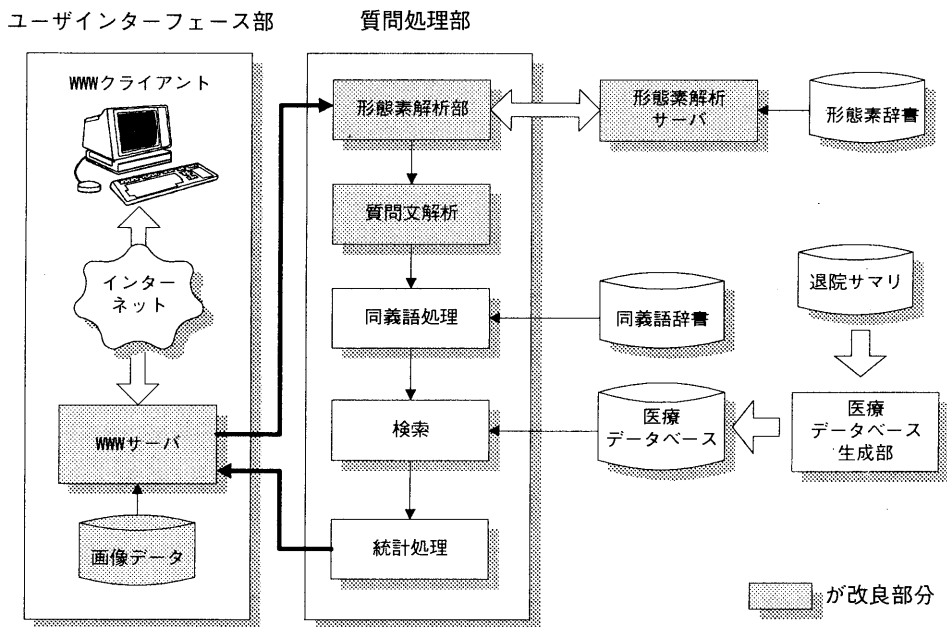


図2 システム構成図

4 システムの改良部分

4.1 ユーザインターフェース部

ユーザインターフェース部は質問文の入力と検索結果を表示する部分で、質問処理部とはサーバ・クライアント通信を行っている。従来、このユーザインターフェース部で取り扱える情報はテキストデータのみであった。このためサマリに記入されていた患部の様子や心電図の波形等の医師

¹ ICOT から公開されている形態素解析プログラムをサーバプログラム用に修正したもの。

の手書きの図は表示することができなかった。しかし、これらの図は退院サマリ文の内容と密接に関連しているため、より実用的なシステムとするためには、システムで画像データを取り扱えることが重要だといえる。今回、このユーザインターフェース部をNEXTSTEPという特殊なシステムからインターネットのWWWへと移行すると共に、サマリ中の図をイメージスキャナを使ってWWWサーバへと蓄積した。これにより、インターネットに接続された端末であれば、どこからでも画像を含む医療情報を検索・表示することができるようになった²。

退院サマリをWWWで利用できるようにするため、次の様な作業を行った。

- ① サマリのHTML(Hyper Text Markup Language)化
- ② サマリ文中への画像データの挿入

このうち①については変換プログラムを作成することでデータを自動変換した。現在のところ画像情報として蓄積しているものは、サマリに描かれた手書きの図のみであるが、この他にも、心電図の波形やCTやMRIといった医療機器を使って撮影された画像等ともリンクさせることも可能である。

4.2 質問処理部

4.2.1 処理対象となる医療データベース

本システムでは、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、あらかじめ医療データベースを作成している。これは退院サマリから検索の対象となるような属性値を持つ名詞(例えば血圧や体温)や動詞を退院番号と共に退院サマリの全文から抽出したものである。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, 4986, 4994, 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, 5499, 5511, 5511, 5511, 5511, 5515,
:, ( - ), ( + ), ( + ) , , , , , , , ( - ), ( - ) , , , , , , , ( + ) , , ( - ), ( - )
( - ), ( - ), ( - ) , , ( + ), ( - ) , , ( + ) , , ( + ) , , , , , ( - ) , , , , , ( - )
, ( + ) , , , , , ( - ) , , , , ( - ) , , ( + ) , , , , , , , ( - ) , , , , , ( - ), ( - )
```

² 勿論、実際の医療カルテを取り扱う際には、セキュリティ等に関して十分な対策を施す必要がある。

4.2.2 受理できる質問文

現在、本システムで受理可能な質問文の規則は以下の通りである。

- <規則1> 【質問文】 → 【条件部】，【主題部】，【処理部】
- <規則2> 【条件部】 → { [疾患・症状等], [の, で] }
- <規則3> 【主題部】 → 患者(さん), [の, で]
- <規則4> 【処理部】 → {処理方法}
- 【処理部】 → 属性, {処理方法} ()は省略可, { }は繰り返し可
- 【処理部】 → 属性, 属性値条件, {処理方法} []はいずれかの選択を表す

この規則を医師質問文に適用すると、例えば「急性心筋梗塞の患者さんの一覧は」の場合であれば、「急性心筋梗塞の」が条件部、「患者さんの」が主題部、「一覧は」が処理部（処理方法）に対応する。また、「急性心筋梗塞の患者で年齢が40才以上の比率は」の場合では、「急性心筋梗塞の」が条件部、「患者で」が主題部、「年齢が」が属性、「40才以上の」が属性値条件、「比率は」が処理方法に対応する。

処理方法については現在のところ表2に示すものが使用できる。また「急性心筋梗塞の患者さんのCPKの最大値と平均は」といったような複数種類の指定も可能となっている。

表2 処理方法の種類

処理方法	処理内容
平均値	属性値の平均値を表示する
最大値	属性値の最大値を求め表示する
最小値	属性値の最小値を求め表示する
一覧	KWIC のリストを表示する
サマリ	該当する患者の退院サマリを表示する
比率	比率を計算し表示する
人数	人数をかぞえ表示する
氏名	患者氏名を表示する
種類	属性値の種類を一覧表示する
値	属性値を表示する
有無	検索結果の真偽を表示する

4.2.3 質問文の処理方法

4.2.1の医療データベースに対する質問文を処理するには、質問文の中からキーワードとなる単語を抽出する必要がある。従来のシステムでは質問文に対して直接パターンマッチングすることで単語を抽出していたため、表現が少しでも違うとマッチングに失敗していた。今回、改良したシステムでは、入力された質問文を、まず形態素解析し、次に格助詞をマーカとして利用することで単語を抽出している。この方法により、表現の違いによる影響をあまり受けずに質問文が処理できるようになった。

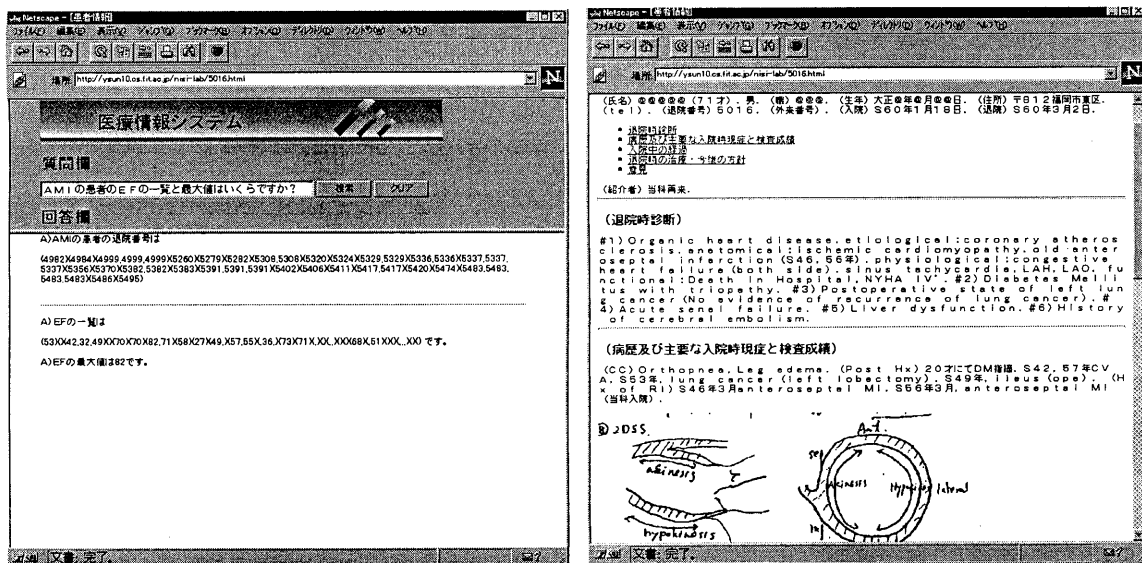
例えば「AMIの患者のEFの一覧と最大値はいくらですか?」という質問文が入力された場合、この質問文を形態素解析サーバを使って解析した結果は、

AMI, 10-の, 424-患者, 10-の, 424-EF, 10-の, 424-一覧, 16-と, 425-最大値, 10-は, 439-いくら, 10-です, 383-か, 442-?, 2

となる。ここで、各単語に付加された番号は品詞の種別を表している（AMIの場合であれば10が名詞を意味している）。この結果を質問文解析部に入力することにより、条件部の「AMI(の)」, 属性の「EF(の)」, 処理方法の「一覧」と「最大値」が取り出され、検索および統計処理を実行することができる。

5 処理結果

本システムによる質問文の処理結果の例を図 3(a)に、図を含む退院サマリの表示例を図 3(b)に示す。



(a)

(b)

図3 処理結果の例

6 まとめ

質問処理部を形態素解析サーバと連携処理させることで、質問文の表現の違いによる影響を受けにくいシステムが実現できた。また、ユーザインターフェース部にWWWシステムを採用することで、サマリ中の図や医療画像などが取り扱えるようになり、実用性が高まったといえる。今後の課題としては、受理可能な質問文の種類を増やして行くことがあげられる。

参考文献

- [1] 西村他：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, 情報処理学会研究報告 95-NL-106,(1995)
- [2] 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)
- [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] ICOT: ICOT フリーソフトウェア No.33 「形態素辞書」(1993)