時序性醫療資訊的精簡表達式與決策運用

A Concise Representation of Temporal Medical Data and its Application to Decision Support

許瑞麟
Ruei-Lin Shiu
國立陽明大學衛生資訊與決策研究所
canis200285@gmail.com

陳恭
Kung Chen
國立政治大學資訊科學系
chenk@cs.nccu.edu.tw

王大為
Da-Wei Wang
中央研究院資訊科學所
wdw@iis.sinica.edu.tw

摘要

時間在醫療診斷、預測或治療等決策行為都扮演重要的角色。許多的臨床指引以時序性資料為決策判別的重要依據，如何簡潔但精準地表達時序性的指引規則，以順利落實於臨床決策支援系統中遂成為重要的研究課題。本研究提出時序性資料描述與查詢語言—TempoCom (Temporal Comprehension)，以更簡潔的方式在原始的關聯式資料庫上查詢時序性醫學資料。並延伸臨床作業指引開發工具——CPGen (Clinical Practice Guideline Application Generator)，使其可以快速開發建制時序性臨床指引決策支援系統。

關鍵字：臨床作業指引、時序性推理、決策支援系統

Abstract

Time plays a major role in clinical diagnosis, prediction, and therapy planning. Time-oriented clinical data are a major decision making factor for many clinical guidelines. An important research problem is how to model guideline concisely and clearly so that temporal constraints can be implemented in guideline-based decision support systems in a straightforward manner. The contributions of this paper are twofold: (1) to develop a temporal query language for medical domain-- TempoCom (Temporal Comprehension), which provides comprehensive yet concise expressions for querying temporal medical data and (2) to extend the clinical practice guideline development tool-- CPGen (Clinical Practice Guideline application Generator) with the capability of supporting temporal decision rules expressed in TempoCom. With the extended CPGen, one can develop temporal guideline-based decision support systems for clinical treatment rapidly.

Keywords: *Clinical practice guideline, Temporal reasoning, Decision support system.*

1、緒論

1.1、背景

臨床作業指引(Clinical Practice Guideline)的應用除了協助臨床醫生可以有系統的診斷與治療疾病，還可以使臨床教育、醫療行為審核或評估以系統化的方式進行[8]。若能將這些指引的規則嵌入到臨床決策支援系統中，並與醫令系統(COPE)或是醫療資訊系統(HIS)結合，當可大幅增加臨床指引在日常醫療行為的接受度和應用性，提供合乎臨床指引的監督和建議，並進而提升醫療品質[7]。因此，許多學者投入研究如何將臨床指引電腦化，設計各種便於開發基於臨床指引的決策支援系統的工具[4]。

1.2、動機

過去關於時序性資料的推理在醫學方面的研究相當多[8]，除了關於時序性邏輯(temporal logic)方面的基礎理論研究外，也有不少關於發展時序性資料庫(temporal database)，並將其應用在臨床決策支援上的研究[4]。但在實務上，大多數的 HIS 並非採用具有特殊時序資料處理能力、以時間為導向的資料庫；而只使用傳統的關聯式資料庫，佐以時戳作為的時序性資料的標記。關於病患資料的查詢，也多半是使用標準的資料查詢語言－SQL，對於時序性資料的查詢上並無特殊機制[3]。例如病患的過去一年間是否至少一次血清肌

酐酸濃度(Creatinine level，以下簡寫為 Cre)大於 1.4，我們必須以下列的 SQL 語句進行查詢並看結果值來判斷：

```
SELECT COUNT(Table.Cre)
FROM Table
Where Table.P_NO='00002574' and
        Table.Cre >1.4 and
        datediff('d', Table.Date,now())<= 365
GROUP BY Table.P_NO;
```

其中 P_NO 為病患的病例編號，Table.DATE 為檢驗時間，Table.Cre 為 Cre 的檢驗值。我們認為這樣的規則表示方式過於瑣碎繁複，不利系統開發人員與醫療專業人員就指引規則的電腦化進行溝通。針對時序性資料與規則的電腦化，我們需要有更高抽象化的表示方法，以方便歸納整理這類規則進行電腦化，並進而得以透過臨床決策支援系來實現它們。

1.3、研究目的

本研究探討如何以更簡潔抽象的方式來表達常見的有時序性的指引規則，這個表達方式要能滿足以下需求：

(1) 簡潔性：將時間觀念整合到醫學資料內，在表達具時序性指引規則時可以直接表示資料或其時戳，而不需要以額外的式子來指定資料與其時戳之間的關聯，以方便系統開發人員與醫療專業人員溝通。

(2) 實務性：能在既有的關聯式資料庫上實現查詢時序性資料與運算，而且要兼顧效能考量。

(3) 可執行性：能方便電腦化，有系統的將指引規則嵌入臨床決策支援系統內。

(4) 延伸性：能作為進一步發展複雜的時序性醫學推論的基礎。

此外，我們認為要使時序性臨床指引規則可以透過臨床決策支援系統導入醫療行為之中，初期的發展應著重於建立醫療專業人員對於電腦化臨床指引規則的認同。若能以比較快速的方式，開發出可支援一般常用之臨床指引的時序性臨床決策支援系統，比較可能有效的取得醫療專業人員的支持與回饋，進而得以改善我們的方法與工具。相對而言，複雜的指引規則或是龐大開發工具反而容易因為導入過程冗長繁複，可能失敗並導致醫療專業人員對這類系統的不信任。因此本研究並不企圖支援所有的時序性臨床指引的開發，而是選擇實際上已使用於醫療行為的指引規則來電腦化，提倡透過臨床決策支援開發工具，以快速開發迅速回應的方式來推動落實臨床指引的電腦化工作。

2、技術背景

本研究所設計的時序性醫療資料與指引規則表達式可以透過適當的解析而電腦化，整合於臨床決策支援系統中。目前我們是採用臨床決策支援系統開發工具 CPGen[5] (Clinical Practice Guideline Application Generator)來實現，故本章節將簡要介紹這個工具。

CPGen 是一個架構在醫學知識系統開發工具 Protégé [6]上的臨床作業指引建模與執行的工具，圖 1 勾勒出 CPGen 的執行流程，它利用 Protégé 的本體論編輯功能與友善的使用者介面為基礎，以決策圖(decision graph)的模式來建立臨床作業指引的電腦模型，然後由特製嵌入 Protégé 的規則轉換器將決策圖翻譯成 XML 格式的虛擬程式碼。此外，CPGen 提供了專屬的執行環境來解譯虛擬程式碼，可以在臨床時結合 HIS，取得病患資料以實現決策支援的功能。

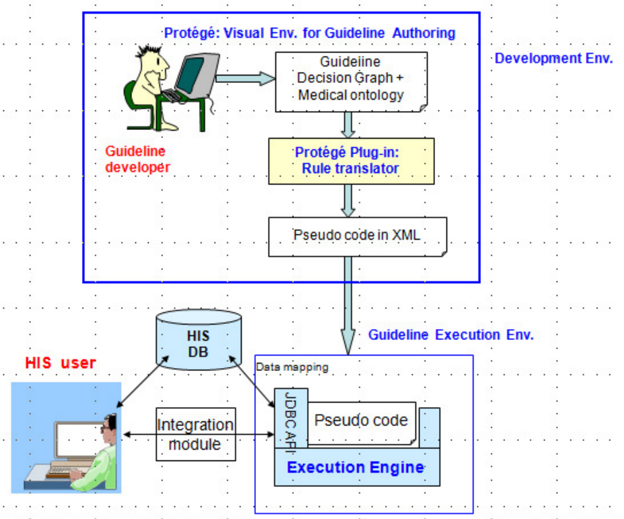

圖 1 CPGen 工具執行流程概觀

CPGen 的開發環境以 Protégé 為基礎，分成四各主要部份；本體論編輯模組、決策圖建立模組、醫療資料對應模組與規則轉譯模組。本體論編輯模組負責建立指引模型中所需要用到的醫學變數的相關資訊，例如前面提到的血清肌酐酸濃度(Cre)就是一個典型的醫學變數；此外建模會用到一些病患基本資料與其他運算函式也會建在本體論中。決策圖建立模組則是 CPGen 的核心，它將指引中的各項邏輯規則以視覺化的方式建成一個決策圖。圖 2 展示一個簡單的決策圖，圖中節

點可分成三類：開始節點(start node)、資訊節點(information node)和終止節點(terminal node)。開始節點是來標示一個決策程序的開始，使用綠色表示；資訊節點收集和計算病患在臨床決策邏輯上所會使用之資料，用深藍色表示；終止節點根據臨床指引提出建議，用灰色表示。決策圖中只可以有一個開始節點，但可以有很多資訊節點和終止節點。連結節點的邊(edge)附有使用布林條件式表達的指引規則，例如：LDL_c ≧160；只有當連結邊上的條件成立時才能從一節點進行到另一節點，決策圖和決策樹概念上相似，最大的差別在決策圖節點可以被不同的邊共享。

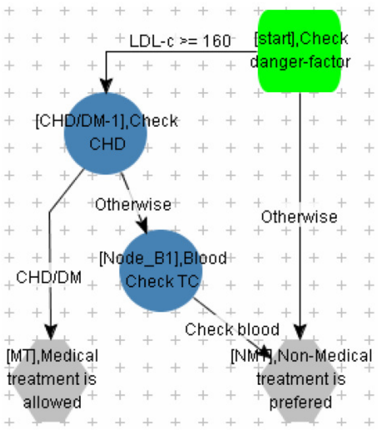

圖 2 決策圖實例

決策圖完成後，必須利用資料對應模組將指引規則內所用到一些變數關聯到的 HIS 資料庫表格，接者就可以用 plug-in 開發的規則轉譯模組把決策圖轉譯成含有資料擷取功能符合 XML 格式的虛擬程式碼。

CPGen 的執行環境分成三部份：資料對應與擷取模組、執行引擎與整合模組。執行引擎藉由資料擷取模組從 HIS 資料庫透過取得病患資料，以直譯的方法執行代表指引規則的虛擬程式碼，以提供 HIS 使用者決策支援的功能。整合模組則負責執行引擎與 HIS 的介接，在取得病患識別碼後，以適當方式驅動執行引擎，執行其功能。

3、時序性醫療資訊表示語言——TempoCom

本研究對具時序性的醫療資訊採用一個簡單且直覺的表達方式：附有時戳的資料序列。例如，一個病患歷次的血清肌酐酸濃度值可以簡單的表達如下：

$$Cre = [cre_1, cre_2, \ldots cre_n]$$

其中 Cre 變數代表病患所有的血清肌酐酸濃度值，是一個依時間排列的序列，而個別的 cre_i 值都有兩個屬性：資料值與時戳，以 cre_i.value 和 cre_i.time 表示之。

其次，在時序性的指引規則中，我們常需要集體性地操作這些時序性資料，為方便表達這些需求，我們參考了函數式語言中常用的序列抽象表示法：List Comprehension (LC) [9]，將其應用在時序性資料上。例如，某病患過去檢驗紀錄中，所有大 1.4 的 Cre 值可用下列 LC 方式簡潔地表示出來：

$$[\, x \mid x \leftarrow Cre,\ x.value>1.4 \,] \tag{1}$$

我們以"[expr | generator/filter …]"的方式建構資料序列，分隔符號"|"的左邊是建構出的序列的各個數值的運算式(公式 1 中為簡單變數 X)，右邊則是一一列舉出此序列的建構公式，這些公式有兩種：生成器(generator)與過濾器(filter)。生成器的形式為"變數←序列"，代表這個變數的值是取自某個序列(類似迴圈變數)。例如，公式 1 中"x←Cre"表示變數 x 的值是由 cre_1 到 cre_n。過濾器則是一個布林算式，用來篩選生成器產生出來的值。例如，公式 1 中"x.val>1.4"表示只有大於 1.4 的 x 才會留下來，因此公式 1 精簡地表達了某病患所有大於 1.4 的 cre 值所建構的序列。

我們稱這個時序性醫療資料的描述語言為 TempoCom (Temporal Comprehension)，除了序列建構公式外，TempoCom 也提供了一些關於時間值的算術邏輯運算與操作序列的函數來協助表達關於序列資料的指引規則。以第一節中 Cre 紀錄查詢為例，要表示病患過去一年間是否至少有一次的 Cre 值大於 1.4，我們可以用下列 TempoCom 公式表達[1]：

$$size(\, [\, x \mid x \leftarrow Cre{\downarrow}1Years,\ x.value>1.4 \,]\,) >= 1 \tag{2}$$

其中函數 size(list)計算序列 list 的元素個數，時間限縮運算"↓"限制序列的範圍，"Cre↓1Years"將給定的 Cre 序列限縮在一年內的紀錄值，常數 1Years 表示一年(kYears 表示 k 年，kDays 表示 k 天，以此類推)。此外，因為 TempoCom 的執行模式是以個別病患為主，所以目標病患的病歷號不必特標註出來。將公式 2 與第一節中的 SQL 語句對比，我們認為很明顯可以的看出 TempoCom 在表達時序性的指引規則時，要比 SQL 語句來的簡潔扼要。完整的 TempoCom 語法與函數庫請

[1] 符號" $x \leftarrow Cre{\downarrow}1Years$" 實作時轉成" $x \leftarrow Cre$, now()-x.time < 1Years" 實現，now()回傳現在時間。

參見[1]，以下是更多的 TempoCom 的公式範例。

例1. 病患的 Cre 檢驗紀錄中，最後一筆是否大於 1.4：

last([x | x←Cre]) > 1.4

例2. 病患某個 I 檢驗值在過去一年的紀錄中，是否有上升大於 5 的情況：

not is_empty ([x | x←diff(I ↓ 1Years), x.value > 5)])

其中 is_empty(list)回傳 True 如果序列 list 的元素個數為零，diff(list)則將 list 中的元素由左至右兩兩相減，例如 diff([1,3,6])==[2, 3]。

例 3：每當病患的 X 值大於 5 時，Y 的值就會小於 1：

size([(x, y) | x←X, x.value > 5, y←Y,
x.time=y.time, y,value<1]) > 0

4、系統架構與實作

本章節介紹如何擴充 CPGen 的開發與執行環境，納入 TempoCom 公式，以提供具有時序性指引規則的決策支援環境。

4.1、整合 TempoCom 至 CPGen 開發環境

首先，為了要表達時序性資料，我們將 CPGen 的資料型態予以擴充，增加序列型態；接著在本體論變數裡，針對決策邏輯上所會使用到的時序性資料，我們就可以賦予它們序列型態。其次，因 CPGen 的開發環境中，決策圖由含有規則的邊和節點組成，所以我們將規則輸入的部份擴充，讓使用者可以在規則輸入時，選擇以 TempoCom 公式來輸入規則的表達式，圖 3 展示用 TempoCom 表示式查詢病患時序性資料的決策圖。

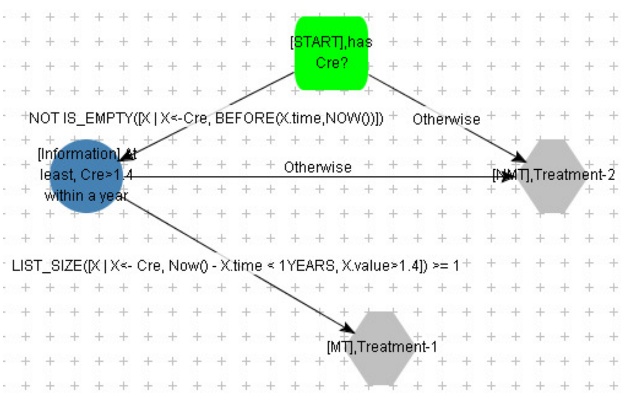

圖 3 CPGen 內以 TempoCom 表示的決策圖

CPGen 開發環境中的規則轉譯模組也必須擴充以處理 TempoCom 的規則式。這部份除了進行語彙與語法分析外，也需要對規則式中使用的變數做基本的查核，像是只有序列變數可用來生成序列元素值。這些查核可以透過本體論中提供的變數與操作函數的型態資訊完成。轉譯模組中最重要的是序列建構式(LC)的翻譯，這部份我們是依據函數式語言的翻譯技術[9]，採用以下的轉換函數(TE)完成：

$$TE[\![[E \mid v \leftarrow L, Q]]\!] \equiv \text{flatMap } (\lambda TE(v).\ TE[\![[E \mid Q]]\!])\ TE[\![L]\!]$$

$$TE[\![[E \mid B, Q]]\!] \equiv \text{IF } TE[\![B]\!]\ TE[\![[E \mid Q]]\!]\ \text{NIL}$$

$$TE[\![[E \mid]]\!] \equiv \text{CONS } TE[\![E]\!]\ \text{NIL}$$

這裡 TE 轉換函數會去比較方括弧[| |]中的 TempoCom 樣版，如果符合，會依樣版對應的函數型表示式進行轉遞迴換。TempoCom 式子中用到的變數，轉譯模組會根據 CPGen 開伐環境知識庫中本體論與資料庫對應表的資訊編譯成資料擷取指令；最後，整個決策圖為轉譯成符合 XML 格式的虛擬程式碼，供後端執行引擎執行。原來 CPGen 的虛擬程式碼必須增加處理序列資料的指令，其他部分則不必修改。圖 4 上半部展示翻譯模組與 CPGen 開發環境其它部份的關係。

4.2、TempoCom 的執行流程與架構

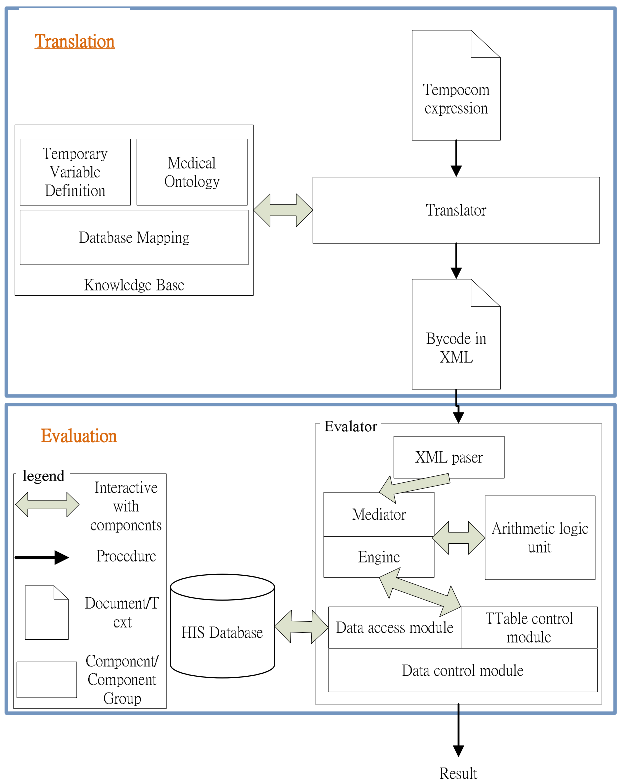

圖 4 TempoCom 的執行流程與架構概觀

如圖 4 所示，含 TempoCom 的決策算式的執行流程可

以分成兩部份，分別是轉譯階段(Translation)和求值階段(Evaluation)，以 XML 程式碼文件為兩者之間的橋梁。

求值階段的目標是把 XML 表示的 TempoCom 程式碼加以解析，求出決策規則的最終值。當 CPGen 整合模組接受 HIS 的驅動，取得病患識別碼後，就會啟動執行引擎的求值過程。整個過程可概分成四個步驟：初始化執行引擎、建立稱之為 TTable 的病患時序性檢驗資料總表、查詢 TTable 設定各醫學資料項目的內容值和解析決策圖中的指引規則。這個 TTable 相當重要，它包含病患過去此臨床作業指引要用到的檢驗資料值，它以二維方式儲存資料，橫軸是依檢驗項目區分，一個檢驗項目存一列；縱軸則是依時間展開，一個時點一行。如下表所示：

	t1	t2	t3	t4	…	tn
Cre						
TC						
…						

求值過程中所用到的序列值都存放在此 TTable 內。

求值器(evaluator)由以下七個元件組成：XML Parser、Mediator、Engine、Arithmetic logic unit、Data control module、Data access module 和 TTable control module。XML Parser 負責對 XML 剖解，抓取 XML 文件中所記錄之資料，初始化其他元件;Mediator 為各元件之中介者，負責協調各元件發出之需求與運作；Engine 負責求值流程；Arithmetic logic unit 負責求值計算；Data control module 負責存取控制求值階段中的所有靜態資料；Data access module 為資料庫存取介面和 TTable control module 負責存取所有 TTable 的內容值。

XML 轉譯階段會讀取 Knowledge Base 內的本體論，以取得變數資訊，與 Knowledge 的 Database Mapping 裡的有關特定領域的醫學專業名詞在資料庫的儲存資料表與欄位的對應，並在 XML 內加以記錄。

求值階段，會先從結合記錄在 Knowledge 的 Database Mapping 裡的資料庫對應資訊，從資料庫裡載入目標病患在 TempoCom 內欲查詢的醫學項目的所有數據，建立 TTable。時序運算與查詢會針對在記憶體內的 TTable 做出相關的運算，回傳查詢運算結果。在實作方面使用了 Model-View-Controller (MVC) pattern。model 元件記錄了所有靜態資訊，對應求值器的元件是 Data control module 和 TTable control module；view 元件會顯示病患資訊和求值結果，可依使用者需求而實作之元件，為著重求值本身，在流程說明與元件示意中不包含此元件與求值器其他元件之互動情況，焦點放在核心元件間的互動；而 controller 元件協調 model 元件和 view 元件實作求值流程，對應求值器的元件是 Mediator、Engine、Arithmetic logic unit 和 Data access module。

5、實驗評估

為了瞭解 CPGen 結合 TempoCom 後，對於據時序性資訊判斷的臨床準則的建模能力與執行效能，我們以顯影劑注射時的腎功能檢定為實作的臨床指引[2]，對已實行的顯影劑注射，做回溯性研究。完整的決策圖因篇幅限制，請參考[1]。以某醫學中心實行電腦斷層掃描(CT)和靜脈腎盂造影(intravenous pyelography, IVP)而注射顯影劑的記錄為系統的輸入。資料期間為 2003 年 1 月 1 日到 2005 年 12 月 31 日，共 26126 位病患，41381 次造影記錄。針對 41381 筆造影記錄中，排除同天同一病患施行不同部份之電腦斷層掃描資料後，共 35507 筆。採用之 Cre 實驗室資料為 2001 年 1 月 1 日至 2006 年 6 月 18 日的，共 193257 筆，排除資料遺失，包含檢驗數值遺失、檢驗日期遺失和兩者皆遺失後，共 193228 筆。

表 1 評估結果

步驟	**平均**（ms.）	**標準差**（ms.）	**最大值**（ms.）	**最小值**（ms.）
初始化	0.039468	0.044238	3.29441	0.0085
建立 TTable	281.8328	55.984597	7207.79	22.216
查詢 TTable	0.037388	0.081934	5.88511	0.00915
指引規則解析	0.391856	0.880299	64.876	0.03457
總合	**282.301**	**56.0183575**	**7207.89**	**222.387**

針對系統執行的四個求值步驟的回復時間和總回復時間分別評估結果，如表 1。各部份的平均時間均在 300

毫秒以下，而除了建立 TTable 的標準差為 56 毫秒，其餘皆在 1 毫秒以下。總回復時間平均為 282 毫秒，但標準差受建立 TTable 之步驟影響，為 56 毫秒。不過實際解析指引規則的時間都很短，而 TTable 的建立是屬於系統啟動階段，使用者通常比較能夠接受。

6、討論

在醫療領域內有許多時序性的臨床指引存，從簡單的單一檢驗數值的時序判別，到困難的化學治療的醫療計畫。為了減少人為疏失，時序性臨床決策支援系統是必需的。如何發展這方面系統的研究，在醫療資訊領域內一直不斷地進行。但是不可否認的，瞭解時序性臨床指引的醫生和實作決策支援系統的資訊人員之間，由於專業訓練的不同，確實存在鴻溝，使得時序性的臨床決策支援系統通常需要較長的開發周期。本研究使用臨床作業指引開發工具——CPGen 結合針對臨床指引實作的時序性查詢語言——TempoCom 實作可開發時序性臨床作業指引的開發工具。利用 CPGen 提供的視覺化介面，使醫生與資訊人員可以在開發階段有更有效的溝通。提供一個輕量級的工具，可供使用者依據內含時序性判斷的臨床指引，開發時序性臨床決策支援系統。其所產生的時序性臨床決策支援系統，雖然不像大型的時序性臨床指引決策支援系統，如 EON，可以支援到需強大時序運算的醫療計畫或臨床指引，但針對大多只需一般時序運算的臨床指引，本研究可以提供使用者快速開發所需的臨床決策支援系統的環境，加快時序性臨床決策支援系統之導入。此外，TempoCom 是以一般醫療資訊系統所使用的關聯式資料庫為基礎實作出來的查詢語言，可在不影響原始資料庫架構之下提供時序性的查詢，滿足簡潔性、實務性、可執行性等需求。在延伸性方面，我們將歸納分析更多醫學常用的時序性規則樣版，如病患在手術後血壓回升大於正常值的時間是否超過24小時或有連續 3 次檢驗，血糖值皆大於標準值等，設計更豐富的時序性高階函數，讓 TempoCom 可以表達更複雜的時序性規則。最後，未來我們將進一步結合圖型化介面的設計開發，提高 TempoCom 的親和力，使醫療人員可以更方便的視覺化方式輸入指引規則。

致謝

本文部份內容是第二作者於2007年在中央研究院資訊所進行短期訪問時完成。

參考文獻

[1] 許瑞麟,"時序性醫療資訊的精簡表達式與決策運用",國立陽明大學碩士論文,2007.

[2] 蔡宗佑,"以『臨床警示及決策支援系統』降低顯影劑誘發腎病變的機率",臺北醫學大學碩士論文,2006.

[3] Daniel J. Nicrin, Isaac S. Kohane, "Temporal expressiveness in querying a time-stamp-based clinical database", *JAMIA*, 2007; 7:152-163.

[4] Juan Carlos Augusto, "Temporal reasoning for decision support in medicine", *Artificial Intelligence in Medicine* (2005) 33, 1-24.

[5] Kung Chen, Chung-Hsin Chen, Shin-Yuan Chang, Chiehfen Chen, and Yu-Chuan Li, "A Rapid Development Environment for Synthesizing Guideline-Based Decision Support Systems", in the Processing of the First International Multi-symposiums on Computer and Computational Sciences (IMSCCS|06), Vol. 2, pp. 548-554, June 2006, Hangzhou, PRC.

[6] Natalya Fridman Noy, Ray W. Fergerson, Mark A. Musen, "The knowledge model of Protégé-2000: combining interoperability and flexibility", In Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000), Juan-les-Pins, France, 2000.

[7] Paul A. de Clercq, Johannes A. Blom, Hendrikus H.M. Korsten, and Arie Hasman, "Approaches for creating computer-interpretable guideline that facilitate decision support", Ar*tificial Intelligence in Medicine* (2004) 31, 1-27.

[8] Paolo Terenziani, Gianpaolo Molino, and Mauro Torchio, "A modular approach for representing and executing clinical guidelines", *Artificial Intelligence in Medicine* (2001) 23, 249-276.

[9] Simon Peyton Jones. The implementation of functional programming languages. Prentice Hall, 1987.