

東 海 大 學

工業工程與經營資訊學系

高階醫務工程與管理碩士在職專班

碩士論文



建構智慧化疾病分類編碼輔助系統-
以一般外科為例

研 究 生：張維書

指 導 教 授：謝宛霖 博士

中 華 民 國 一 〇 八 年 六 月

**The Development of Intelligent Support Systems for
Coding International Classification of Diseases–
The Case Study of General Surgery Medical Record**

By
Wei-Shu Chang

Advisor : Prof. Wan-Lin Hsieh

A Thesis

Submitted to Tunghai University
in Partial Fulfillment of the Requirements
for the Degree of Master of Health Administration

June 2019
Taichung, Taiwan

建構智慧化疾病分類編碼輔助系統-以一般外科為例

學生：張維書

指導教授：謝宛霖 教授

東海大學工業工程與經營資訊學系高階醫務工程與管理碩士在職專班

摘 要

自 1995 年全全民健康保險開辦至今，國際疾病分類代碼(International Classification of Disease)為各醫療院所向健保署申請醫療費用給付之重要依據，疾病分類代碼直接影響 DRGs(Diagnosis-related Groups)歸屬及費用給付，疾病分類代碼的正確性和適當性越高，才能確保病患所做的醫療處置都能向健保署申請合理的給付，疾病分類師編列代碼在醫療院所扮演著相當重要角色，然而，疾病分類代碼的編排方式，具有其分類架構及分類邏輯，關鍵字的判定及查找正確的代碼，成為疾病分類師編碼的第一重要步驟。現今疾病與處置編碼的工作，已有電子工具書可輔助疾病分類師使用關鍵字協助編碼，然而，在使用此電子工具書前，仍須自行從主要診斷或主要處置描述中，判定其關鍵字。對於資淺的疾病分類師，仍需花費大量的時間與容易產生錯誤。

本研究運用「文字探勘」協助建構主要診斷與主要處置之關鍵字判定系統。首先，整理歸納個案醫院中 2016 與 2017 年所有主要診斷與主要處置中，每個案例中協助判斷其關鍵字之判定字以及所對應之關鍵字，用以建立「主要診斷」及「主要處置」關鍵字判定系統之知識庫。此外，並整理出判定字搜尋的規則，作為關鍵字判定系統的設計邏輯。本研究以 Python 進行編碼，並以 PyInstaller 將此判定系統包裝成執行檔，透過 2018 年之病歷案例來驗證系統之準確性。

本研究最後證實此「主要診斷與主要處置之關鍵字判定系統」，可以有效地降低疾病分類師進行「主要診斷」及「主要處置」的關鍵字判斷，主要診斷個案，資淺疾病分類師必須耗費 1620 秒，資深疾病分類師必須耗費 930 秒，使用關鍵字判定系統只須耗費 450 秒；而對於主要處置個案，資淺疾病分類師必須耗費 2070 秒，資深疾病分類師必須耗費 1200 秒，使用關鍵字判定系統只須耗費 540 秒。

關鍵字詞：ICD-10-CM/PCS、文字探勘、關鍵字

The Development of Intelligent Support Systems for Coding International Classification of Diseases— The Case Study of General Surgery Medical Record

Student : Wei-Shu Chang

Advisor : Prof. Wan-Lin Hsieh

Master Program for Health Administration
Department of Industrial Engineering and Enterprise Information
Tunghai University

ABSTRACT

Since the inception of the universal health Insurance in 1995, the International Classification code has become an important issue for medical institutions to apply to the health insurance Department for medical expenses. Disease classification code can directly affects DRGs attribution and expense payment. Disease classifiers play an important role on the medical institutions to make sure classification codes are organized based on their architecture and logic. Taiwan Government and some companies develop electronic tool book to assist disease classifiers while searching for correct disease codes. However, disease classifiers should be able to identify key words of diagnosis statement and treatment statement before using the electronic tool book. Therefore, junior disease classifiers still need to spend lots of time and it is easy to cause human errors.

This study uses "text mining" to assist in the construction of keyword determination system for principal diagnosis and principal procedure. First of all,summary of case hospitals in 2016 and 2017 of all principal diagnoses and principal procedure, each case assists in judging the keyword's decision word and the corresponding keyword, knowledge base for establishing "principal diagnosis" and "principal procedure" keyword determination system. In addition,and sort out the rules for judging word search,as the design logic of keyword determination system. This study adopts Python for coding and PyInstaller for packing the system into an execution file. Then, the execution file is used to verify the accuracy of the system through a 2018-year medical record case.

This study finally confirms that this "keyword determination system for principal diagnosis and principal procedure", can effectively reduce the disease classifier to carry out "principal diagnosis" and "principal procedure" keyword judgment,principal diagnostic case,disease division must take 1620 seconds, deep disease division must take 930 seconds, use the word to determine that the system takes only 450 seconds, and for the principal procedure, disease division must take 2070 seconds, deep disease division must take 1200 seconds, use the word to determine that the system takes only 540 seconds .

Keywords : ICD-10-CM/PCS, Text Mining, Keyword

誌謝

論文完成了，真的有股莫名的感動與開心，對於唸在職專班的我，在工作與唸書中不斷地努力，終於完成研究論文的撰寫。

首先，要向指導教授—謝宛霖老師說～謝謝老師細心的指導，老師您辛苦了！老師總是很樂意的與我討論論文，不論是和老師約的時間、假日或是半夜，老師不但非常細心且耐心地指導我寫論文，一次又一次的討論論文內容，啟發了我寫論文的智慧；論文一點一滴的完成了，由衷地感謝老師；謝謝口試委員—黃欽印主任及李怡慶老師，於 108 年 05 月 24 日口試當天指導學生加強撰寫論文的重點內容，使得我的論文內容更臻完善。

謝謝我的好同學—許雅鈴、陳清華的互相鼓勵、加油、打氣，我常常提問班上最聰明的「雅鈴同學」問題，雅鈴總是回我～我教妳，我的問題又迎刃而解了，「清華同學」則是我們的開心果，有清華在就有歡樂的笑聲，非常的開心在美麗的校園裡學習知識，有著好同學的陪伴，真好！

感謝辛苦的媽媽～又要工作，又要唸書的我，終於完成研究論文了。

維書 謹誌於東海大學醫務工程與管理碩士在職專班

2019 年 6 月

目錄

摘要.....	i
ABSTRACT.....	ii
誌謝.....	iii
目錄.....	iv
表目錄.....	vi
圖目錄.....	vii
第一章 緒論	1
1.1 研究背景與動機.....	1
1.2 研究目的.....	4
1.3 研究限制.....	4
第二章 文獻探討	6
2.1 疾病分類的目的與發展.....	6
2.1.1 疾病分類目的.....	6
2.1.2 國際疾病分類.....	7
2.1.3 我國疾病分類發展現況.....	13
2.2 疾病分類的編碼、複雜度及重要性.....	14
2.2.1 疾病分類的編碼步驟.....	14
2.2.2 疾病分類編碼的複雜度.....	14
2.2.3 疾病分類編碼的重要性.....	15
2.3 資料探勘與文字探勘.....	17
2.3.1 資料探勘的定義與目的.....	18
2.3.2 資料探勘的流程、演算法及模型.....	19
2.3.3 文字探勘的定義.....	21
2.3.4 文字探勘的系統結構.....	22
2.4 專家系統.....	23
2.4.1 專家系統的構造.....	24
第三章 研究方法	25
3.1 研究流程.....	25
3.2 研究工具.....	27
3.3 病歷資料處理.....	27
3.4 建置疾病分類知識庫.....	28
3.5 設計關鍵字之判定系統.....	30
3.5.1 主要診斷關鍵字之判定系統.....	30
3.5.2 主要處置關鍵字之判定系統.....	32
第四章 系統驗證	34
4.1 主要診斷系統驗證.....	34
4.1.1 107年個案－主要診斷系統驗證.....	34
4.1.2 主要診斷關鍵字判定系統之效率測試.....	38

4.2 主要處置系統驗證.....	39
4.2.1 107 年個案-主要處置系統驗證.....	39
4.2.2 主要處置關鍵字判定系統之效率測試.....	43
第五章 結論與建議	45
5.1 研究成果.....	45
5.2 未來研究方向.....	47
參考文獻	48
中文部分.....	48
英文部分.....	50
附錄一 ICD-10-CM 工具書介紹	52
附錄二 ICD-10-PCS 工具書介紹	54

表目錄

表 2.1	ICD-9 與 ICD-10 之差異對照比較.....	9
表 2.2	ICD-9-CM 與 ICD-10-CM (診斷代碼) 編碼差異對照比較.....	10
表 2.3	ICD-9-CM 與 ICD-10-PCS (處置代碼) 編碼差異比較.....	12
表 2.4	一般外科常見診斷代碼舉例	16
表 2.5	一般外科常見處置代碼舉例	17
表 2.6	處置代碼第 1 碼包含之主要章節	55
表 2.7	處置代碼第 1 碼包含之輔助章節	55
表 3.1	105-106 年主要診斷知識庫	29
表 3.2	105-106 年主要處置知識庫	30
表 4.1	105-106 年主要診斷資料庫	34
表 4.2	107 年主要診斷資料庫	35
表 4.3	無法找到主要診斷描述判定字個案	36
表 4.4	重複主要診斷描述判定字個案	37
表 4.5	107 年找到的主要診斷描述判定字及關鍵字	37
表 4.6	資淺及資深疾病分類師使用關鍵字判定系統前後耗時比較表	38
表 4.7	105-106 年主要處置資料庫	39
表 4.8	107 年主要處置資料庫	40
表 4.9	無法找到主要處置描述判定字個案	42
表 4.10	重複主要處置描述判定字個案	42
表 4.11	107 年找到的主要處置描述判定字及關鍵字	43
表 4.12	資淺及資深疾病分類師使用關鍵字判定系統前後耗時比較表	44

圖目錄

圖 1.1 疾病分類編碼進化圖	3
圖 2.1 資料探勘流程圖	19
圖 2.2 文字探勘的簡易輸入—輸出模型	22
圖 2.3 用於使用者輸入和輸出的反覆運算迴圈	23
圖 2.4 專家系統圖	24
圖 3.1 本研究之研究流程	26
圖 3.2 程式語言 PYTHON 之執行檔	27
圖 3.3 主要診斷—關鍵字之系統流程圖	31
圖 3.4 主要處置—關鍵字之系統流程圖	33
圖 4.1 程式語言 PYTHON 之執行檔	36
圖 4.2 程式語言 PYTHON 之執行檔	41
圖 5.1 主要診斷—使用關鍵字判定系統前後耗時比較圖	46
圖 5.2 主要處置—使用關鍵字判定系統前後耗時比較圖	47

第一章 緒論

1.1 研究背景與動機

自 1995 年全民健康保險開辦以來，各醫療院所向健保署申請醫療費用，必須以國際疾病分類代碼(International Classification of Disease)為申請醫療費用之給付依據，2016 年起健保署以 ICD-10-CM/PCS(International Classification of Disease and Related Health Problems, Tenth Revision, Clinical Modification/Procedure Coding System)為申報依據，衛生福利部中央健康保險署（以下簡稱健保署）自 2010 年導入住院 Tw-DRGs（Taiwan Diagnosis Related Groups 台灣版診斷關聯群，以下簡稱 Tw-DRGs）支付制度，疾病分類代碼直接影響 DRGs（Diagnosis-related Groups 診斷關聯群，以下簡稱 DRG）歸屬及給付費用，疾病分類代碼的正確性和適當性越高，才能確保病患所做的醫療處置都能向健保署申請合理的給付，對於醫療院所的大宗醫療收入影響相當大(Debry, C. A., Lapane, K. L., Feldman, H. A., & Carleton, R. A, 2001；Stavem K., Bjerke G., Kjelsberg FN., Ruud EA., & Saxrud SO, 2002；Baker, 2002；黃瑞典等，2010)。

醫療科臨床人員將病人的「出院診斷」書寫於出院病歷摘要(Discharge Note)，疾病分類師以出院病歷摘要之「出院診斷」為編碼依據，並詳細閱讀住院病歷摘要(Admission Note)、病程記錄(Progression Note)、手術記錄(Operation Note)…等相關病歷資料（李瑞華，1992），將醫療科臨床人員施予病人的手術、處置、檢查…等各項治療處置，依據國際疾病分類編碼原則，從多個診斷及多個處置中，依據病人的主訴及相關症狀，判斷主要診斷及次要診斷；依據與主要診斷最相關之處置，判斷主要處置及次要處置，查找診斷及處置關鍵字，依序編列主要診斷、次要診斷、主要處置、次要處置，本研究一般外科個案，並無出現多個診斷，因此無法進行主要診斷之判斷。但是卻有多個處置須進行主要處置判斷的情況。然而，因為時間關係，本研究並未針對此部分進行系統之建立，屬此研究之缺陷。再判斷完主要診斷與主要處置之後，需進行代碼的編列，僅能依靠疾病分類師人工作業編列代碼，不僅耗費大量時間，也耗費相當的人力；疾病分類師編列正確編碼，出院個案歸屬正確且合理的 DRG，醫療院所向健保署申請給付費用，「診斷與處置代碼」成為健保署給付醫療院所醫療費用之基本條件及重要關鍵。

在早期疾病分類師編碼工作中，疾病分類師必須倚賴紙本工具書來協助進行編碼的工具（疾病分類編碼 1.0）。近年來，由於資訊技術的進步，健保署及私人公司分別建立「疾病分類電子工具書」，以協助疾病分類師可以更快速地編列診斷及處置代碼（疾病分類編碼 2.0）。然而，疾病分類師必須閱讀疾病診斷及手術處置的描述，自行判定診斷及處置的關鍵字，才能使用疾病分類電子書工具書，查找診斷及處置代碼，對於本研究個案醫院為例，新進人員或是經驗不足的疾病分類師，自行判定診斷及處置的關鍵字，所耗費的時間是三年以上資深疾病分類師的一點五倍以上，以研究個案醫院為例，以隨機抽樣法選取一般外科 30 個案，資淺疾病分類師查閱主要診斷的關鍵字，平均每個案耗費 54 秒，主要處置的關鍵字，平均每個案耗費 69 秒；資深疾病分類師查閱主要處置的關鍵字，平均每個案耗費 31 秒，主要處置的關鍵字，平均每個案耗費 40 秒，再加上疾病分類師人才短缺、離職率高（范碧玉等，2013），醫療院所必須不斷地耗費相當多的時間，進行新進人員的培訓及教育訓練。

目前醫療院所聘任之疾病分類師，以三年以下之資淺疾病分類師居多，原本由人腦來判定主要診斷及主要處置的關鍵字，會因編碼經驗不足，導致疾病分類師編碼工時增加，且容易因為人為疏失造成編碼錯誤，導致影響編碼品質，不僅提高醫療院所的人事成本，更直接影響醫療院所的醫療收入（黃于珊、施驊真，2004）。因此，本研究希望透過智慧化技術的導入，建立一套診斷及處置之智慧化關鍵字判定系統，期望能透過此專家系統用以協助疾病分類師能更精準地判斷主要診斷及主要處置的關鍵字，將原本由人腦進行主要診斷及主要處置關鍵字判斷的作業，改由智慧化系統協助判斷，使疾病分類編碼工作邁入到新的階段（疾病分類編碼 3.0），疾病分類編碼進化圖，如圖 1.1。

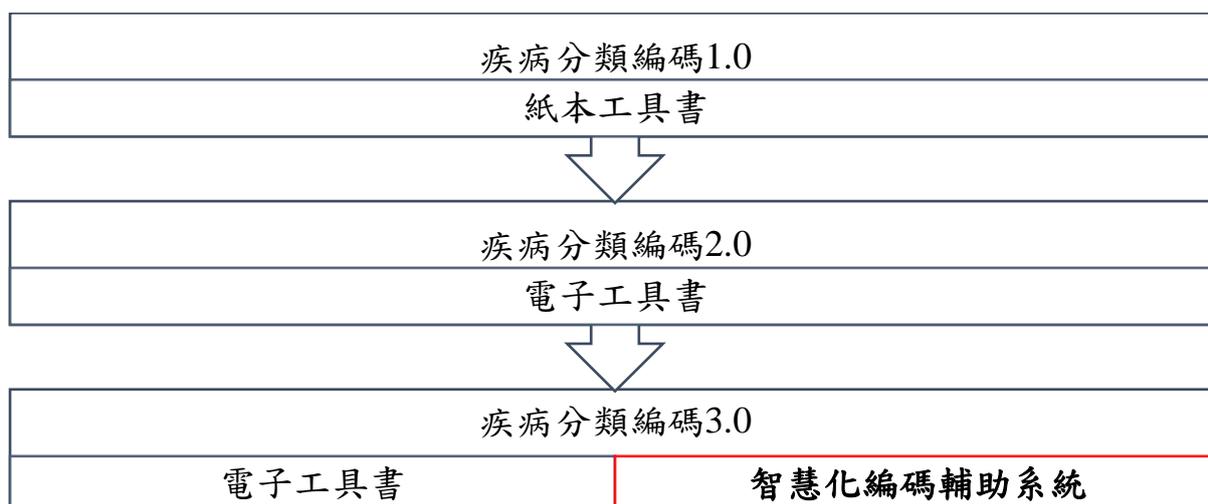


圖 1.1 疾病分類編碼進化圖

資料來源：本研究整理

資料探勘(Data Mining)技術可用於在非常大量文字或數據資料裡，挖掘資料模型與資料之間具有關聯的規則，其過程必須是自動的或半自動的，發現的模型必須是有意義的，因為它將會是產生具有貢獻的價值，然而，挖掘出的模型與關聯規則能夠事先預測，有效的幫助使用者探討資料與資料之間的相關性、關聯性，因此被推崇為有效的解決現今資訊流通迅速的時代，所面臨的資料繁複、不具有規則性、資訊貧乏困境中的一種有效方式(Witten et al., 2016；張芸濤、龔玲，2007)。

1.2 研究目的

本研究將建立一套專家系統，將目前已有的主要診斷及主要處置關鍵字，結合資深疾病分類師在判斷其關鍵字的依據，整理歸納成一知識庫(database)，並建立一套關鍵字判定系統，使得新進人員在即使沒有任何編碼相關經驗的情況下，也可透過這套專家系統，輸入主要診斷及主要處置的描述，快速準確地找到診斷及處置的關鍵字，以減少資淺疾病分類師在經驗不足的情況下，發生診斷及處置的關鍵字判定錯誤的情況。如此，除了可降低工作的時間之外，也可降低因人為因素所造成的編碼錯誤率，進而增加診斷及處置代碼的準確性及適當性。

本研究目的如下：

- 1.歸納整理診斷描述之判定字與其對應之疾病病灶關鍵字的知識庫。
- 2.歸納整理手術(處置)描述之判定字與其對應之治療目的關鍵字的知識庫。
- 3.建立關鍵字判定系統，以提升疾病分類師編碼之正確性及適當性。

1.3 研究限制

疾病分類代碼的產生，必須由疾病分類師查閱醫師書寫於病人的出院病歷摘要之出院診斷及病歷相關內容，從多個診斷及多個處置中，依據病人的主訴及相關症狀，判斷主要診斷及次要診斷；依據與主要診斷最相關之處置，判斷主要處置及次要處置，查找診斷及處置關鍵字，依序編列主要診斷、次要診斷、主要處置、次要處置，編列診斷及處置代碼完成後，正確歸屬 DRG(Diagnosis related Group 診斷關聯群，以下簡稱 DRG)，以作為醫療院所向健保署申請給付醫療費用的依據；本研究一般外科個案，並無出現多個診斷可為主要診斷之情況；但因研究時間不足，本研究並無針對多個處置進行主要處置判斷之系統建立。

影響 DRG 歸屬的六個變項因素為 1.診斷，2.手術或處置，3.轉歸，病人出院時的狀態，例如：改門診治療、一般自動出院、死亡、轉院…等，4.合併症(Comorbidity)或併發症(Complication)，5.年齡，6.性別(健保署，2019)；醫療院所向健保署申報之診斷及處置代碼數，最多申報診斷 20 個代碼及處

置 20 個代碼，診斷代碼包含當次住院相關診斷及病人所患有的慢性疾病等，處置代碼包含當次住院之治療手術處置及其他相關之診斷性處置，診斷及處置代碼繁複且包含範圍廣，因此萃取住院病人最重要的「主要診斷」及「主要處置」為研究題材。

本研究以中部地區某區域教學醫院 105-107 年「一般外科」已申報個案為例，僅以「主要診斷」及「主要處置」為研究資料，所建立的知識庫可能會有遺漏的案例。若可以進一步收集到醫學中心等級的醫院個案，將可使知識庫更完整。

第二章 文獻探討

2.1 疾病分類的目的與發展

2.1.1 疾病分類目的

醫療院所所稱的診斷與處置編碼(coding)，就是將門診及住院的病人疾病診斷(diagnosis)與手術處置(procedure)轉譯(translate)成為數字或代碼，成為診斷及處置代碼，可作為統計上可運用的重要參考資料。疾病分類編碼資料可提供醫師及臨床科醫療人員查詢、檢索，作為統計分析、研究、臨床診療及教學之用；疾病分類代碼也是醫療院所向健保署申請病人醫療費用給付的重要依據；另外，疾病分類統計資料，可提供醫療院所及衛生主管機關施政及決策參考，也進一步是醫院經營管理與醫療品質評估的指標(范碧玉、賴淑芬，2014)。

疾病分類的標準一般可由各專科醫學會、醫療衛生機構或保險機構自行訂定，也可由國家統一訂定。WHO(World Health Organization 世界衛生組織，以下簡稱 WHO)為統合全世界的疾病及死亡統計，編撰了有關疾病、傷害等的國際分類標準(International Classification of Disease, ICD，以下簡稱 ICD)供全世界使用，目前 ICD 出版到第 11 版(ICD-11)(World Health Organization, 2018)。

疾病分類通常由受過專業訓練的疾病分類師，依醫師的病歷記錄(Medical Record)、病歷摘要(Discharge Note)、住院病歷摘要(Admission Note)、手術記錄(Operation Note)、病程記錄(Progression Note)、醫囑記錄及檢查(驗)報告、會診記錄等，詳細閱讀相關的診斷與處置資料，查找出正確及適當的診斷及處置分類代碼，將代碼輸入電腦儲存，以作為統計、分析整合與運用(范碧玉、賴淑芬，2014)。

然而，疾病分類目的可包含下列九點：(一)提供健保署給付醫療院所費用之重要依據；(二)提供醫學研究及分析；(三)醫療品質之依據；(四)作為醫療資源管理控制與編列預算之工具；(五)財政計劃之參考；(六)醫療資源審核；(七)提供醫療院所之間的成本比較；(八)教學評鑑之參考依據(九)醫療技術儀器取得(曾卉珍、徐均宏，2016)。

因此，若醫療院所之疾病分類師執行編碼作業流程順暢且完善，編列之代碼具有正確性及適當性，充分運用豐富的疾病分類編碼知識，將使得前線臨床醫療人員對於病人的診斷及施予的手術（處置），得以於疾病分類代碼完整呈現，以利於各方面之研究、分析、運用及費用給付等，由此可見，疾病分類師扮演著相當重要的角色（古思明等，2010）。

2.1.2 國際疾病分類

WHO 將 ICD(International Classification of Disease)譯為國際疾病分類，是一套國際診斷及處置標準代碼，其為國際上有系統的將各種疾病傷害、相關問題及手術處置，予以結構性及系統性的分類，以作為臨床醫療科人員統計分析與研究等用途。

ICD 是一種“變項－主軸(variable-axis)”的分類。其結構由威廉·法爾(William Farr)在早期國際上討論分類結構之際所提出者，他的構想是「對於流行病學用途的、具有實用性的統計資料，以下列方式的歸類分群」(Farr, 1855)：

- 流行性疾病(Epidemic diseases)
- 一般疾病或體質(General diseases or constitutional)
- 依解剖部位歸類的局部疾病(Local diseases arranged by site)
- 發展性疾病(Developmental diseases)
- 傷害(Injuries)

WHO為統一全世界的疾病、傷害及死因，發展出ICD(International Classification of Diseases and Injuries)國際疾病分類系統，1975年ICD(International Classification of Diseases and Injuries)第9版(ICD-9)公布實施，美國為使ICD-9更適用於其國內的臨床與研究之用，於1977年將它修正為ICD-9-CM(International Classification of Diseases, Ninth Revision, Clinical Modification國際疾病分類臨床修訂第九版，以下簡稱ICD-9-CM)。自1995年我國全民健康保險開辦後，各層級的醫療院所健保病人之「診斷」與「處置」分類，規定必須以ICD-9-CM編碼，且為醫療院所向健保署申報醫療費用之給付依據。。

ICD-9-CM 在美國自西元 1970 年開始使用，已使用超過 30 年，純數字

編碼方式是多年前設計的，編碼組合及設計已不合乎需求，其醫療技術資料的分類亦無法符合現今的醫療科技及醫療儀器，醫師對於病人確診的診斷，手術與處置也不能精準的呈現，且代碼已超出早期所設計的用途與目的，診斷與處置代碼已不符合現況；新增代碼也無空間可編列，無法呈現新增疾病及新的醫療手術處置，不但影響診斷及手術、處置的統計分析，亦影響病人醫療照護品質及健保費用申報的正確性。

1993年WHO公布ICD-10(International Statistical Classification of Disease and Related Health Problems, Tenth Revision.國際疾病與有關健康問題統計分類第十版，以下簡稱ICD-10)，美國很快也將它修正為ICD-10-CM(International Statistical Classification of Disease and Related Health Problems, Tenth Revision, Clinical Modification.國際疾病與有關健康問題統計分類第十版臨床修訂，以下簡稱ICD-10-CM)，其中並將ICD-9-CM第3冊處置分類部分編修為ICD-10-PCS(International Statistical Classification of Disease and Related Health Problems, Tenth Revision, Procedure Coding System.國際疾病與有關健康問題統計分類第十版處置代碼系統，以下簡稱ICD-10-PCS)，ICD-10於1990年代定稿，自1999年開始，美國生命統計機構負責的死因報告（死亡證明書）的編碼，已採用ICD-10；美國考量國際醫學統計資料分享與其國內需求之因素，於2015年10月1日起，施行ICD-10-CM編碼系統。

基本的ICD是由前三項字元組成「類目碼」。在ICD-10以數字與英文字母混合配置代碼(alphanumeric code)，即第一個字元是「英文字母」，第二、三、四個字元為「數字」；取代了先前各修訂版僅有「數字」之代碼系統；第四碼位於小數點之後。因此「診斷代碼」由A00.0到Z99.9。其中字母“U”則為保留碼，以提供未來新增的疾病的擴充使用。

ICD-10有多種之子系統，其中以ICD-10-CM和ICD-10-PCS兩者為主。ICD-10-CM用於急診、門診及住院病人之疾病、診斷分類；ICD-10-PCS用於急診、門診及住院病人之手術、處置分類。ICD-9與ICD-10之差異比較，如表2.1。

表 2.1 ICD-9 與 ICD-10 之差異對照比較

項目 \ 版本	ICD-9	ICD-10
分類名稱	國際疾病分類	疾病與有關健康問題國際疾病分統計分類
章節	17 章 001-999	21 章 A00-Z99, “U” 碼除外 U00-U49: 保留於新增的疾病及無法確定病因的疾病 U50-U99: 保留於研究用途
補充分類	兩種補充分類 1. 傷害中毒外因 E800-E999 2. 影響健康狀態與接觸醫療服務的因素 V01-V82	成為正式章節 1. 第 20 章: 疾病與死亡外因 V01-Y98 2. 第 21 章: 影響健康狀態與接觸醫療服務的因素 Z00-Z99
冊數	1. 第一冊-詳細列表說明 2. 第二冊-疾病英文字母排列之索引	1. 第一冊-詳細列表說明 2. 第二冊-使用指引手冊 3. 疾病英文字母排列之索引
疾病診斷筆數	13,000	69,000
處置代碼筆數	3,700	86,000

資料來源：本研究整理

由表2.1來看，ICD-10與ICD-9比較，除了將以往ICD-9兩種補充分類放入正式分類之外，診斷代碼總數(total codes)與處置代碼總數(total codes)明顯倍增，疾病的分類更加詳細，例如：側性、細部位的區分。ICD-10-CM 和ICD-9不同之處，有非常多代碼呈現疾病的病況，診斷代碼第一個字元為英文字母，第二個字元至第七個字元為數字或字母，可擴充之空間加大、「V」和「E」碼不再以補充碼使用，感覺器官異常從「神經系統」分離出來、外傷則改以「解剖部位」來分類，而非以「外傷類別」來分類、部份手術後併發症已移至身體系統之特定處置章節，增加側性、虛擬位置以利擴充等。ICD-9與ICD-10之編碼方式差異對照比較，如表2.2。

表 2.2 ICD-9-CM 與 ICD-10-CM (診斷代碼) 編碼差異對照比較

版本 項目	ICD-9-CM	ICD-10-CM
字元數	3-5 個字元 新增代碼空間有限	3-7 個字元 新增代碼空間較有彈性
代碼內容	第一個字元是數字， 僅 E (傷害中毒外因) 及 V (影響健康狀態與接觸醫療服務的因素字母) 代碼第一個字元是字母， 第 2-5 字元是數字	第一個字元是字母， 第 2-7 字元是字母或數值
代碼格式	3-5 碼 病因、解剖部位、嚴重度 ↑ □□□.□□ ↓ 類目碼	3-7 碼 病因、解剖部位、嚴重度 ↑ □□□.□□□ □ ↓ ↓ 類目碼 擴充碼

項目 \ 版本	ICD-9-CM	ICD-10-CM
代碼舉例	540.2 急性闌尾炎合併腹膜炎 865.09 脾臟撕裂傷 550.90 腹股溝疝氣	K35.3 急性闌尾炎合併局部腹膜炎 S36.039A 脾臟撕裂傷 K40.90 腹股溝疝氣

資料來源：本研究整理

手術處置的部分，ICD-10-PCS取代ICD-9-CM的處置碼，ICD-10-PCS與ICD-9-CM處置代碼架構截然不同，ICD-9-CM 處置碼僅有3-4位數字，ICD-10-PCS則由7個字元組成一個處置代碼，包含數字及字母，其代碼可完整呈現臨床醫療上的手術處置方式及使用的醫療器材，使得臨床醫療資料的具多軸性架構，且具有價值最大化，具有完整性、擴充性、容易使用、不具重覆性的優點，詳述臨床醫療技術的細節，且能完整呈現醫療科技的先進技術，顯示現今醫療科技之進步與發達。ICD-9-CM之處置代碼與ICD-10-PCS之編碼差異對照比較，如表2.3。

表 2.3 ICD-9-CM 與 ICD-10-PCS (處置代碼) 編碼差異比較

版本 項目	ICD-9-CM	ICD-10-CM
字元數	3-4碼	7碼
代碼內容	數字	字母或數字
代碼格式	<p style="text-align: center;">次分類碼</p> <p style="text-align: center;">↑</p> <p style="text-align: center;">□□.□□</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">類目碼</p>	<p style="text-align: center;">Root Operation Qualifier</p> <p style="text-align: center;">手術方式 修飾詞</p> <p style="text-align: center;">Section Approach</p> <p style="text-align: center;">章節 手術途徑</p> <p style="text-align: center;">↑ ↑ ↑ ↑</p> <p style="text-align: center;">□ □ □ □ □ □ □</p> <p style="text-align: center;">↓ ↓ ↓</p> <p style="text-align: center;">Body System Device</p> <p style="text-align: center;">身體系統 裝置物</p> <p style="text-align: center;">Body Part</p> <p style="text-align: center;">身體部位</p>
代碼舉例	<p>47.09 闌尾切除術</p> <p>85.23 乳房部分切除術</p> <p>06.2 單側甲狀腺葉切除術</p> <p>50.22 肝部分切除術</p>	<p>0DTJ0ZZ 開放性闌尾全部切除術</p> <p>0HBT0ZZ 開放性右側乳房部分切除術</p> <p>0GTG0ZZ 開放性左側甲狀腺葉全部切除術</p> <p>0FB14ZZ 腹腔鏡肝臟右葉部分切除術</p>

資料來源：本研究整理

2.1.3 我國疾病分類發展現況

我國衛生福利部中央健康保險署自西元 1980 年起，要求各醫療院所之疾病分類(Nosology/Taxonomy)，以 WHO 於 70 年代出版的 ICD-9 為分類與命名的統一標準。自 1991 年 4 月起，勞保、公保也陸續要求各醫療院所的「住院病人」，必須以診斷與處置代碼為醫療費用申請之重要依據，逐漸改以美國版本的 ICD-9-CM。

全民健康保險自 1995 年實施後，健康保險局規定各醫療院所的「住院病人」、「門診」與「急診」的醫療費用申報，必須以 ICD-9-CM 編碼，ICD-9-CM 成為我國疾病分類與命名的黃金標準。

國內疾病分類版本發展沿革，自 1995 年全民健保實施後，我國疾病與手術處置的分類系統，採用美國 1992 年版 ICD-9-CM，2004 年改採用美國 2001 年版 ICD-9-CM，2009 年衛生福利部統計處使用 ICD-10 統計死因分類，2010 年改採用美國 2014 年版本 ICD-9-CM 為分類基礎，導入住院 Tw-DRGs(Taiwan, Diagnosis-Related Groups 台灣版診斷關聯群，以下簡稱 Tw-DRGs)支付制度。

由於美國已公告自 2014 年 10 月改用 ICD-10-CM/PCS (請參閱附錄 ICD-10-CM 與附錄二 ICD-10-PCS 之工具書介紹)，將不再持續維護 ICD-9-CM (以往每年持續不斷地更新)，屆時台灣將面臨無新版本可用的困境，若繼續使用 ICD-9-CM 編診斷及處置代碼，將導致無法與世界各國家接軌，也可能阻斷台灣與世界各國進行醫療相關統計資料比較、對照及研究的道路。

因此，健保署自 2010 年啟動 ICD-10-CM/PCS 導入計畫，2016 年 1 月 1 日起，疾病分類編碼採用美國 2014 年版 ICD-10-CM/PCS 為編碼依據，且為各層級醫療院所向健保署申請醫療費用給付之重要申報依據。

2.2 疾病分類的編碼、複雜度及重要性

2.2.1 疾病分類的編碼步驟

疾病分類師執行編列診斷及處置代碼時，首先，從診斷及手術處置中找出「關鍵字」，亦即最能呈現該疾病的診斷病灶，該手術處置的治療目的，查找出代碼後，詳細閱讀代碼表列說明，以確切說明代碼的正確性及適當性。

完整的疾病分類編碼包含五大步驟：1.詳閱全本病歷，查找診斷與處置相關之重要資料，例如：出院病歷(Discharge Summary)、住院摘要(Admission Summary)、病程記錄(Progress Note)、手術記錄(Operation Record)、病理報告(Pathology Report)、檢驗報告(Laboratory Report)、醫囑單(Physician Order)、麻醉記錄(Anesthesia Report)、護理記錄(Nurses Note)、急診病歷(Emergency Room Record)等。2.查找關鍵字(Key word)。3.詳閱關鍵字的註解說明、修飾詞或交互性之參考文字，編列適當之代碼。4.對照 ICD-10-CM/PCS 電子書中診斷及處置代碼表列說明，確認代碼之正確性及適當性。5.確認代碼與醫師書寫之出院診斷或手術處置是否相符(李瑞華，1992；范碧玉、賴淑芬，2014)。

2.2.2 疾病分類編碼的複雜度

依據 AHIMA(American Health Information Management Association 美國健康資訊管理協會，以下簡稱 AHIMA)統計，ICD-10-CM 診斷代碼筆數約有 69,000 個，ICD-9-CM 診斷代碼筆數約 13,000 個，ICD-10-CM 與 ICD-9-CM 的診斷代碼，相較之下，增加 5.3 倍之多。而依據 AHIMA 統計，ICD-10-PCS 處置代碼筆數約 86,000 個，ICD-9-CM 處置代碼筆數約 3,700 個，ICD-10-PCS 與 ICD-9-CM 的處置代碼，相較之下，增加 23.2 倍之多。

ICD-10-CM/ICD-10-PCS較ICD-9-CM的編碼系統、編碼長度、定義、內容...等，有大幅度之變動，對於資深疾病分類師（亦即有ICD-9-CM編碼基礎）而言，除了必須拋棄既有ICD-9編碼規則與分類邏輯，重新學習ICD-10之編碼定義、編碼指引，以及編碼指引的改變。

診斷代碼中的側性，例如：左側、右側、雙側等，以及身體細部位，必須於診斷描述中加以確認及辨識；對於手術處置代碼中的手術方式(Root Operation)、身體部位(Body Part)、手術途徑(Approach)、是否有使用裝置物(Device)、修飾詞(Qualifier)...等，必須加強了解臨床專業知識，才能具有足夠的專業能力解讀臨床醫療科人員所書寫的病歷記錄。

對於資淺疾病分類師（亦即無ICD-9-CM編碼基礎）而言，因為ICD-10之編碼規則與定義細緻，診斷及處置代碼筆數鉅增，其分類架構精細且複雜，編碼困難度鉅增，須熟稔臨床醫學知識，加強醫學術語、解剖學、生理學等專業醫療知識(Majerowicz,2011；李玲美，2016)，學習疾病分類編碼更是一項艱難的技術與挑戰。

范碧玉等(2002)於「醫療院所申報健保住診費用之疾病分類編碼適當性分析」研究表示，單一主要診斷及主要處置編碼完全一致之比率為65.0%，所有疾病診斷與處置編碼一致率為46.6%。編碼不一致之主要原因則以「編列錯誤代碼」、「忽略可編之次要診斷」及「因資訊不足，疾病分類師無法編列詳細部位代碼」為最多；楊美雪等(2007)於「疾病分類編碼一致性與其DRGs分派之影響--以膽結石處置病例為例」研究表示，處置代碼編列錯誤是影響外科DRGs歸屬之主要編碼不一致原因；各醫療院所持續不斷地要求疾病分類師編碼品質一致率，然而，因為ICD-10-CM/PCS編碼困難，且需要至少花費三年以上的學習及教育訓練（依據研究者編碼經驗及培訓新進疾病分類師之經驗），才能維持一定的編碼品質，由此可見，學習疾病分類的入門，查找診斷與處置之「關鍵字」，是創造編碼實力的重要關鍵。

2.2.3 疾病分類編碼的重要性

疾病編碼之依據來自於醫師、臨床醫療人員及各類醫事人員之病歷記錄，醫師與醫事相關人員對於ICD-10也必須有基本的認知，促使臨床醫療人員書寫之病歷記錄達到ICD-10-CM編碼需求（一般外科常見診斷代碼舉例，如表2.4），使得疾病分類師能夠依據病歷資料內容，編列正確且適當的診斷代碼。

透過更詳細的編碼，疾病分類手術處置代碼可以與健保申報醫令進行交叉比對，以確認手術處置代碼符合健保申報醫令，而不同的手術方式

(Root Operation)、手術途徑(Approach)、裝置物(Device)代碼(一般外科常見處置代碼舉例,如表2.5),也有可能影響Tw-DRGs的歸屬,甚至連動影響住院個案的權重(Related Weight),影響醫療院所向健保署申請醫療費用給付,更直接影響醫療院所的大宗醫療收入(Sabanovic et al., 2001; Reid et al., 2005; 魏秀美, 2004; 黃瑞典等, 2010)。

李瑞華(1992)提及疾病分類師應具備足夠的理解、閱讀病歷資料內容的能力,正確的判斷主要診斷、次要診斷、主要處置、次要處置的順序。目前已施行第二階段Tw-DRG,未來將全面擴大實行Tw-DRGs,分類代碼的適當性及正確性,對於醫療院所的醫療給付費用影響甚鉅,疾病分類師應檢視病歷之重要參考資料,以編列完整及正確的「診斷」與「處置」代碼。

表 2.4 一般外科常見診斷代碼舉例

英文診斷名稱	中文診斷名稱	診斷代碼
Acute appendicitis with localized peritonitis	急性闌尾炎合併局部腹膜炎	K35.3
R't inguinal hernia with obstruction	右側腹股溝疝氣合併阻塞	K40.30
Gall stones with chronic cholecystitis	膽囊結石合併慢性膽囊炎	K80.10
Small bowel obstruction	小腸阻塞	K56.60
Right thyroid nodular goiter	右側甲狀腺結節	E04.9
Left breast abscess	左側乳房膿瘍	N61
英文診斷名稱	中文診斷名稱	診斷代碼
Bilateral breast cancer	雙側乳癌	C50.912
Acute peritonitis	急性腹膜炎	K65.0

資料來源：本研究整理

表 2.5 一般外科常見處置代碼舉例

英文處置名稱	中文處置名稱	處置代碼
Bilateral total mastectomy	開放性雙側乳房全部切除術	0HBV0ZZ
laparoscopic cholecystectomy	腹腔鏡膽囊切除術	0FT44ZZ
left total thyroidectomy	左側甲狀腺全部切除術	0GTG0ZZ
Hernia repair with mesh	疝氣成形術合併使用彌補物	0YU60JZ
Laparoscopic appendectomy	腹腔鏡闌尾切除術	0DTJ4ZZ
subtotal gastrectomy	開放性次全胃切除術	0DB60ZZ
segment excision of ileum	開放性迴腸部分切除術	0DBB0ZZ
laparoscopic partial hepatectomy,s5-7	腹腔鏡右側肝葉部分切除術	0FB14ZZ

資料來源：本研究整理

2.3 資料探勘與文字探勘

資料探勘(Data Mining)的執行方法，僅適用於具有結構性的相關表格資料，無法應用於未具有結構性的資料；文字探勘(Text Mining)則結合了資料探勘技術、自然語言處理(Natural Language Processing)、資訊萃取(Information Retrieval)，進而從文字資料中挖掘其核心價值，包括資料之間的關聯程度(Association Degree)、重要的說法(Phrase)及關鍵的術語(Term)。

資料探勘(Data Mining)與文字探勘(Text Mining)的目的，都是從大量的文字資料裡，有效率地整理、歸納、分析出資料中的重要資訊。文字探勘必須再運用特別的資料處理程序，以及遵循規則上的歸納步驟。

2.3.1 資料探勘的定義與目的

資料探勘(Data Mining)是收集大量的資料，挖掘出有意義、實用性的資訊，探討出正確、具有規則性，且有用的隱含知識，且具有潛在性的邏輯與規則，最終可以讓人們了解的寶貴知識與實用資訊，是知識發現的重要技術，可作為資訊技術的發展。其主要的意涵是 Mining From Data，從資料中挖掘金礦，資料探勘的另外一個常見的同義詞KDD(Knowledge Discovery in Databases 資料庫知識發掘，以下簡稱KDD)-資料庫中的知識挖掘，換句話說，在KDD中進行知識發現的過程是為資料探勘，然而，資料探勘是KDD中的一個非常重要的探索過程(Han et al., 2011；張芸濤、龔玲，2007)。

Chen et al. (1996) 提到資料探勘可以從資料庫中探索出潛在有用的、重要的、事先未知的資訊，而Berry與Linoff (2004) 則表示資料探勘是以大量的資料，進行半自動或自動的分析方法，探索出具有意義及價值的關聯法則或規則。Daniel與Chantal(2014)進一步指出資料探勘是利用統計、數學、模型識別技術篩選資料庫中的大量資料，發現有意義的新關聯、模型和趨勢的過程。

因此，資料探勘(Data Mining)就是要挖掘「主要診斷」中不能靠直覺發現的疾病病灶，與「主要處置」中非直覺連想出來的治療目的，資料探勘所挖掘的資訊越有意義，就可能越有價值；「主要診斷」的描述中，隱含著疾病的病灶，「主要處置」的描述中，潛在地說明治療疾病的目的，挖掘的知識的過程中，往往不容易經由簡易的歸納與分析就能獲得，往往隱藏於資料裡，透過個案資料的整理、歸納、比較分析，應用專門處理大量資料的資料探勘工具—程式語言 Python，挖掘出「關鍵字」潛在規則與邏輯，進而產出的疾病分類知識庫，使未來在各「主要診斷」與「主要處置」中，可以透過系統直接產生「關鍵字」，對於資淺的疾病分類師而言，執行疾病分類編碼時，具有實際效用及效益。

張芸濤及龔玲於2007年「資料探勘原理與技術」一書中提到，資料探勘(Data Mining)並無明確的說法或定義，在不同的研究領域或其他不同的文獻，也有另有其定義，如Jonn提到資料探勘是在資料裡挖掘，具有實際效益模型的過程，Ferruzza提到資料探勘是用在知識挖掘過程裡，發現資料之間的未知關聯和模型的方式或技巧，Zekulin則表示資料探勘是一個從大型資

料庫中擷取讓人了解且可以實行的資訊；Lodhi(2012)則表示資料探勘是使用智慧的方法，從資料中找尋模型；本研究應用「資料探勘」的概念，搜集中部某區域教學醫院105-106年一般外科個案的「主要診斷」及「主要處置」，整理及歸納其診斷與處置描述中之關鍵字判定字，萃取出編列代碼重要的關鍵字。

2.3.2 資料探勘的流程、演算法及模型

Fayyad(1996)與Fayyad et al.(1996)將知識挖掘的流程分為資料選擇(Selection)、資料前置處理(Preprocessing)、資料轉換(Transformation)、資料探勘(Data Mining)、解釋與評估(Interpretation/Evaluation)等步驟，如圖2.1。

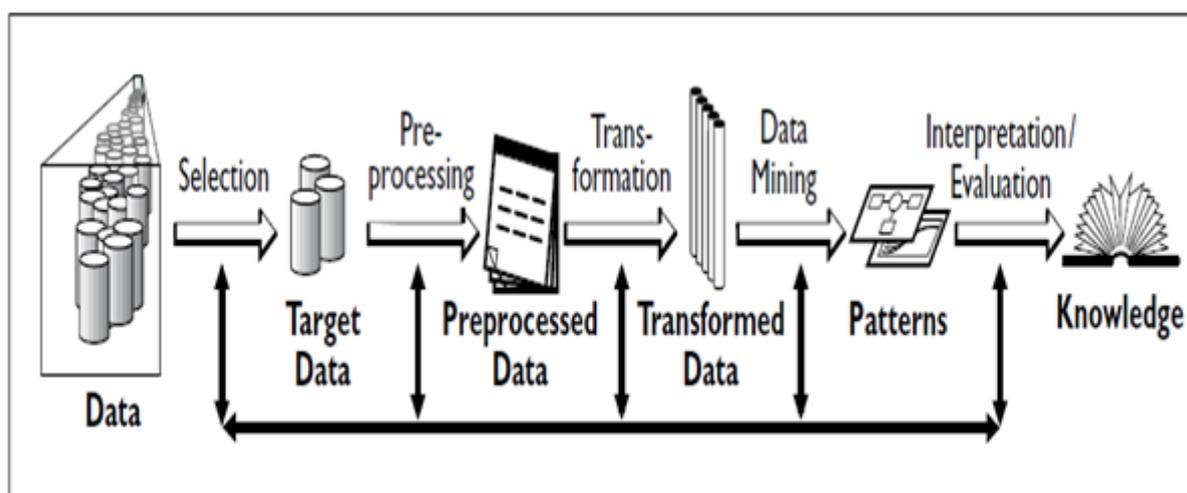


圖 2.1 資料探勘流程圖 (Fayyad, 1996:29)

資料探勘演算法主要由三個元件組成，包含 1.模型(The model)，2.優先準則(The preference criterion)，3. 搜尋演算法(The search algorithm)(Fayyad et al. ,1996)，以下針對各項做詳細介紹：

1.模型(The model)

是指有兩個或兩個以上相關的因素組合而成

2.優先準則(The preference criterion)

依據給予的資料，優先以一種模型或一組特徵為基準，標準通常是某種模式的適配度檢定。

3.搜尋演算法(The search algorithm)

演算法用於查找特定模型和特徵，給予資料或模型(或一系列的模型)和優先準則。

資料探勘常見的功能模型包括：1.分類模型(Classification)，2.群聚模型(Clustering)，3.關聯分析(Association Analysis)，4.時序性分析(Sequence analysis)，5.摘要(Summarization)等，(Fayyad et al., 1996; Han、Kamber, 2000; Khan et al., 2008)。以下針對各項進行詳細說明：

1.分類模型(Classification)

將資料項目分類為幾個預定義的分類類別。

2.群聚模型(Clustering)

將資料項目預設為幾個類別項目(或群聚)，其中必須確定資料來源，與預定義的分類不同。通過基於相似性度量或概率密度模型找到數據項的自然分組來定義群集。群聚是定義資料挖礦群組或機率密度模式。

3.關聯分析(Association Analysis)

另一種說法是購物籃分析(Basket Analysis)，可從眾多交易中(多個購物籃中)，探討與解析購物籃中，購物品項之間的關係，探究出哪幾項產品被顧客同時購買的頻率高於其他產品，探討出來結果，不但可以幫助產品的業者規劃促銷方針及策略，也可以運用在客戶社群及市場行銷。

4.時序性分析(Sequence analysis)

模式序列包含(資料具有時序性的依賴，例如時間序列分析)，目標是模擬生成序列的過程的狀態，探討時間的偏差和趨勢的關聯。

5.摘要(Summarization)

對於所有領域，以數據資料提供簡要的描述，以一個簡單的舉例來表達意旨和標準偏差。精緻的功能包含摘要規則，多個變項視覺化技術和功能變項之間的相關性。摘要功能常用於互動式的探索性資料分析和報表的產生。

2.3.3 文字探勘的定義

謝邦昌(2017)提及文字探勘(Text Mining)是跨領域的資料運用，透過資料探勘技術、資訊檢索技術、程式語言的資料處理方式，使得未經整理及建檔的文字資料，經由電腦程式判定、分析，執行知識歸納、資料檢索及自動分類。文字探勘的主要技術，主要來自於資料探勘，從個案分析和文字資料之交互比對，查找出難以發現的訊息，呈現出經驗價值與未知的技巧(Hale, 2005)。

文字探勘的定義為「整理、歸類及分析龐大的資料的程序過程，用以提供特定使用者的新穎資訊，且發現資料與資料之間的相互關聯」；然而，探勘的處理過程中，必須依賴電腦程式語言和模型的應用，才能處理龐大的資料，「自然語言處理」(natural language processing)必然獲得跨越各項專業領域的應用價值(Sullivan, 2001；Blake, 2011；陳世榮，2015)。

文字探勘彙整各項傳統的資訊檢索功能，包括自動檢索摘要、資料自動歸類、全部文章的檢索及最具實用性的「關鍵字」萃取。

2.3.4 文字探勘的系統結構

文字探勘系統接收輸入的原始資料，產生各種類型的輸出資料，例如：模式、關聯、趨勢，（如圖 2.2）；然而，以人為中心的知識挖掘，則產生一個較複雜的輸入-輸出範例，（如圖 2.3）(Feldman 與 Sanger, 2007)。

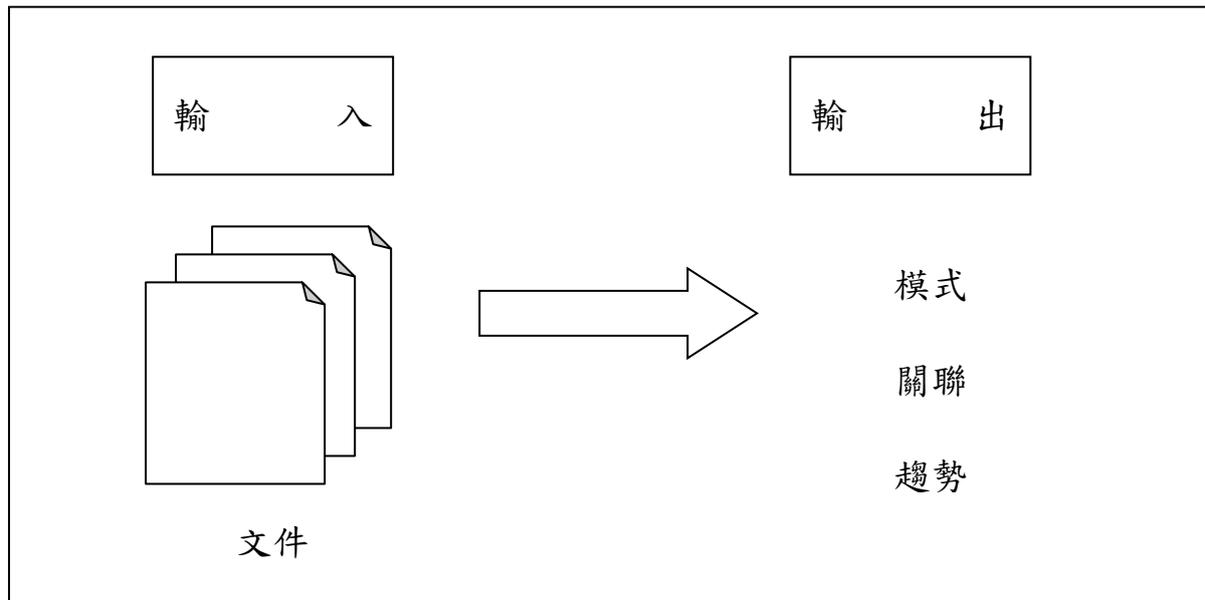


圖 2.2 文字探勘的簡易輸入—輸出模型

資料來源：譯自 Feldman 與 Sanger(2007)

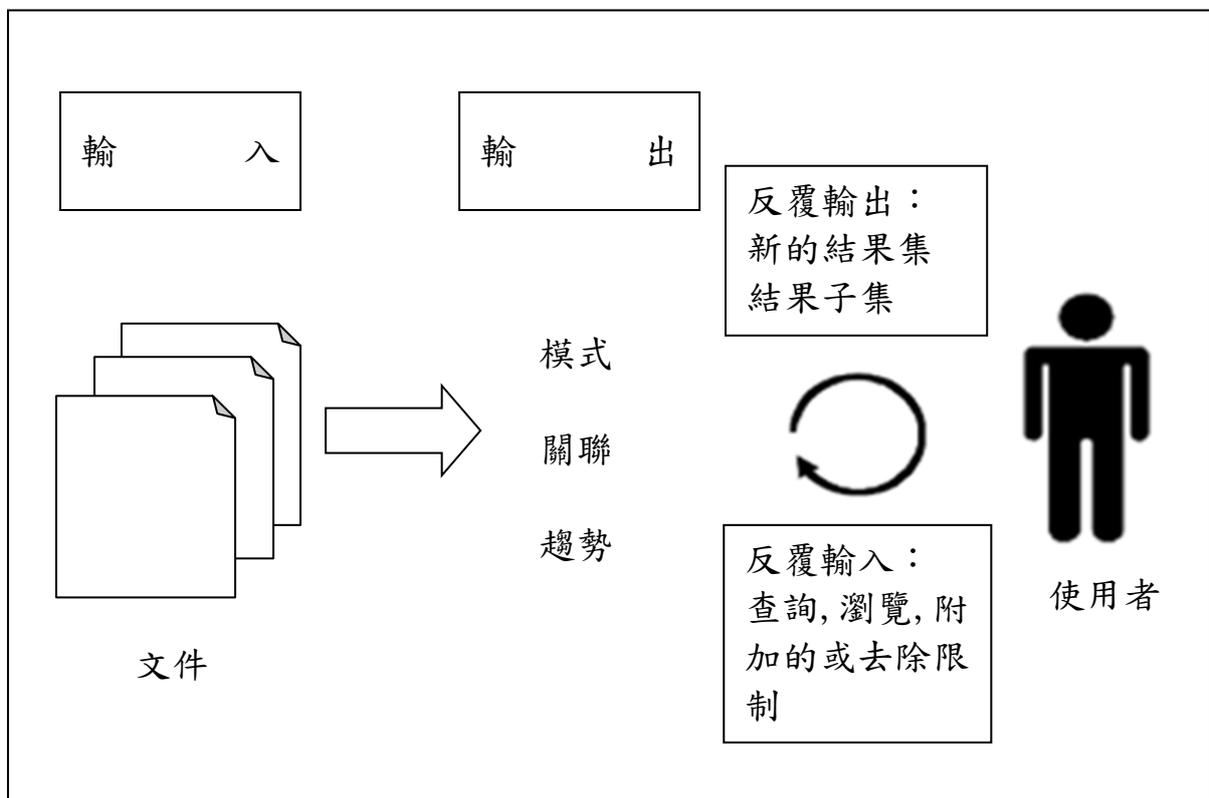


圖 2.3 用於使用者輸入和輸出的反覆運算迴圈

資料來源：譯自 Feldman 與 Sanger(2007)

2.4 專家系統

早期人工智慧(Artificial Intelligence, AI)的子系統—專家系統(Expert System)是在某些特定的專業領域中，萃取專家經年累月且寶貴的經驗知識之程式系統或軟體。「專家」受過專業訓練，並且累積多年的經驗，在特定的領域中具有專業知識(Expertise)，亦即一般人沒有的專業知識與技能。專家系統可以透過電腦系統，經由特定的規則及知識，建立特定領域，例如：疾病分類，專家的經驗知識庫(Knowledge Base)，執行如同專家的專業知識及解決所面臨的問題能力，讓程式系統執行專家所具有的思考及想法，進一步再進行自動推理(Automated Reasoning)。

專家系統被定義為具備智慧化的電腦系統，以專家知識為依據，運用推理方式的概念設計軟體的結構。然而，它被建立於決策支援輔助的一項工具，並給予使用者參考的依據，但它是強化「知識庫」的系統，運用專家解決問題的模式與技巧，儲存大量的知識來解答使用者所面臨的問題，

也是決策者用以解決知識問題的軟體(Ariena et al., 1991; Travis et al., 1992; Goodell et al., 1993; Boyd and Sun, 1994)

費根鮑姆等人於 1968 年研製成功第一個專家系統 DENDEL，專家系統突飛猛進的發展，因而被廣大的應用於不同的專業領域，例如：醫療、教學、汽車、農業等，也為使用者帶來明顯的經濟效益，如今，專家系統成為人工智慧領域中最受關注、最具有專業價值的系統(Jabbar et al., 2015)。

專家系統包含「知識庫」及「推理機」，也稱為知識系統(Knowledge-based system)。一般而言，專家系統具有三個重要的因素：1.某項專業領域專家知識、2.執行專家的思考和想法、3.具有專家等級的程度。

2.4.1 專家系統的構造

使用者將事實或其他資訊提供專家系統，專家系統處理後，即獲得專家的解答（如圖 2.4），專家系統包含知識庫(Knowledge base)及推理引擎(Inference engine)。知識庫是系統的核心，主要儲存專業領域的專業知識；推理引擎則依使用者提供之事實描述，透過知識庫中的知識，經由推理步驟解決問題。

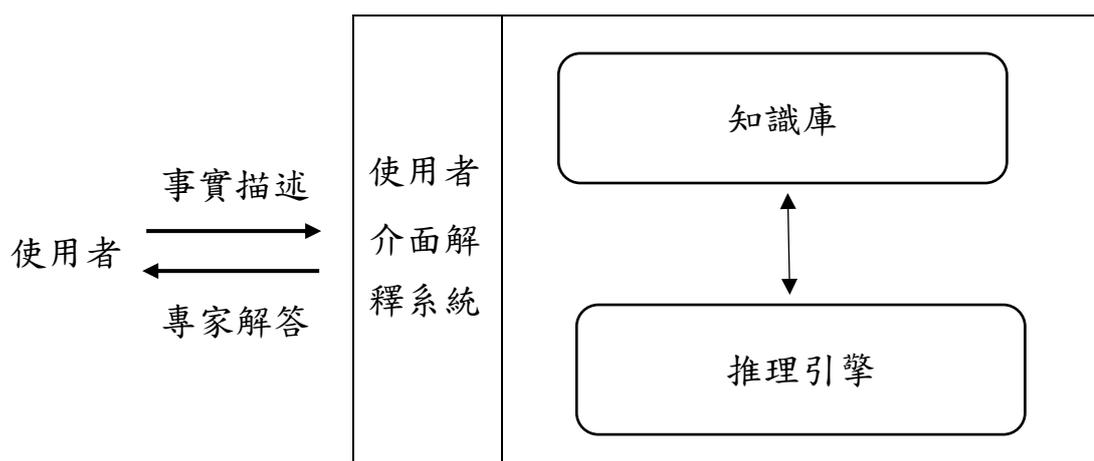


圖 2.4 專家系統圖

資料來源：譯自 Tseng 與 Hwang (2007)

第三章 研究方法

本研究取得中部某區域教學醫院 105-107 年一般外科住院病人之「主要診斷」及「主要處置」病歷資料，針對 105-106 年一般外科住院病人之「主要診斷」及「主要處置」，應用文字探勘進行個案資料整理與建檔，建立關鍵字判定知識庫，其內容包含個案之主要診斷及主要處置之判定字以及其所對應之關鍵字。與設計「判定字」文字探勘之判定流程系統的規則。最後，以 107 年一般外科住院病人之「主要診斷」及「主要處置」病歷資料加以驗證系統流程之準確性，透過程式語言(Python)進行編碼，再利用 PyInstaller 將所撰寫之語言包裝成執行檔，以協助疾病分類師進行「主要診斷」及「主要處置」的關鍵字判斷。

3.1 研究流程

本研究針對一般外科住院病人之「主要診斷」及「主要處置」加以整理、歸納、分析，探索出主要診斷與疾病病灶關鍵字、主要處置與治療目的關鍵字之間的關聯規則。將一般外科的診斷及其主要治療的手術、處置，透過資深疾病分類師的經驗與知識，建立一套判定字與關鍵字對應之知識庫。應用「文字探勘」設計關鍵字判定之邏輯與規則，以建立一套用於主要診斷與主要處置關鍵字判定系統。最後，再透過系統驗證，以確立系統之準確性，作為具有實用性及價值性的「專家系統」。本論文研究流程如圖 3.1。

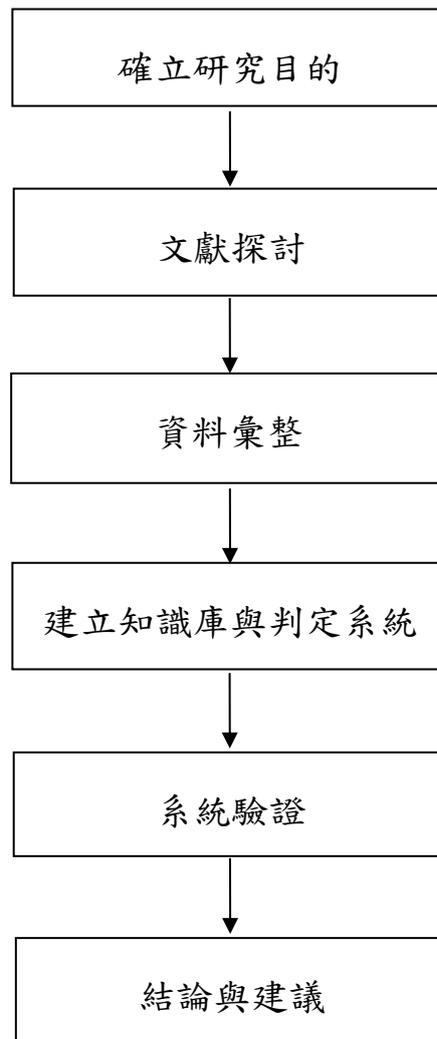


圖 3.1 本研究之研究流程

3.2 研究工具

首先，本研究先使用 Microsoft Excel 整理歸納 105-106 年所有的主要診斷與主要處置之個案，並且根據研究者之經驗知識，整理各案例之判定字以及所對應的關鍵字。接著，使用程式語言是 Python（如圖 3.2），將所設計之文字探勘規則撰寫成系統，為了讓使用者有簡易的操作介面，再利用 PyInstaller 將所撰寫之系統包裝成執行檔。

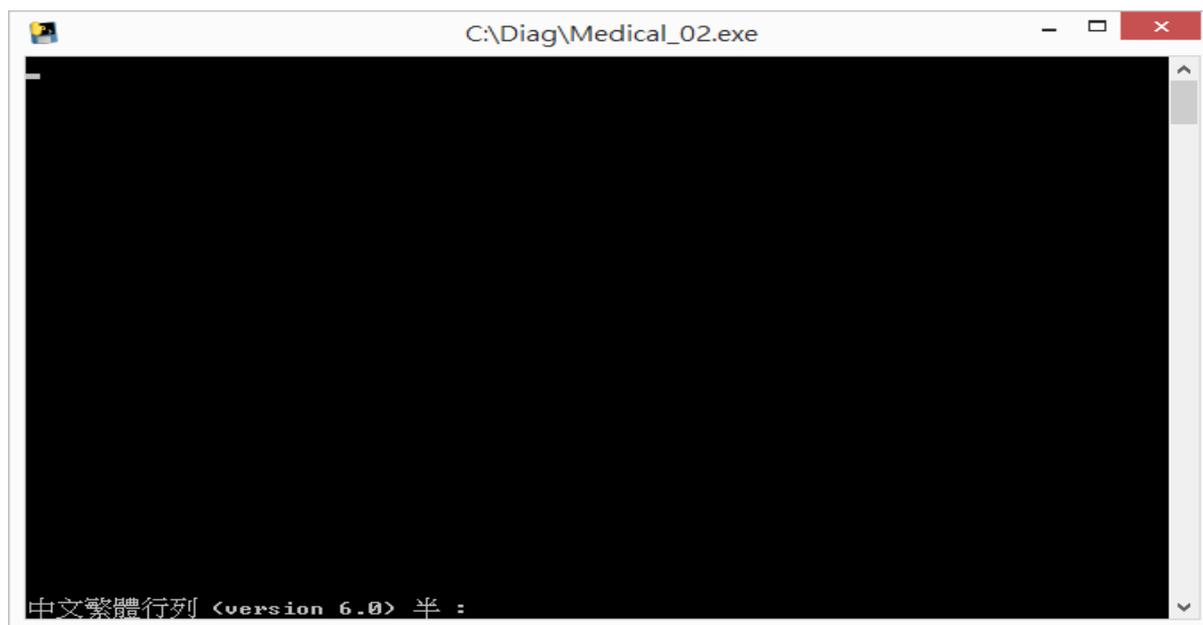


圖 3.2 程式語言 Python 之執行檔

3.3 病歷資料處理

「主要診斷」的定義是個案被確定診斷為引起病人住院治療的主要原因，引起病人此次住院之主要原因有兩個診斷時，則可以選取醫療資源耗用較高之診斷；「主要處置」是與主要診斷最相關之治療性手術處置，如果有兩個以上之手術處置與主要診斷都有相關性時，則以治療性的手術處置為主要處置，如果有兩個以上的手術處置與主要診斷都有相關性時，且都是治療性的手術處置，則以複雜度較高、醫療資源耗用較多者為主要處置（健保署，2015）；當疾病分類師執行編碼的個案有兩個或兩個以上的主要診斷或主要處置時，經由疾病分類師與該個案之主治醫師共同判斷，並取得共識。

「關鍵字」為疾病分類師執行診斷及處置編碼(Coding)時，必須查找診斷及手術處置描述中，診斷的疾病病灶，以及手術處置的治療目的，進而查找相關的描述，例如：身體的部位（Liver 肝臟、Gallbladder 膽囊、Appendix 闌尾…等）、身體的側性（Left 左側、Right 右側、Bilateral 雙側…等）。

「主要診斷描述判定字」為查找關鍵字的相關判定字眼，例如：Acute appendicitis with peritonitis 急性闌尾炎合併腹膜炎，判定字可為「appendicitis」及「peritonitis」，但依據疾病分類編碼規則，主要判定字應為「appendicitis」，相關的附帶疾病為「peritonitis」，則主要的關鍵字為「appendicitis」；「主要處置描述判定字」為查找關鍵字的相關判定字眼，例如：Laparoscopic partial hepatectomy 腹腔鏡部分肝臟切除術，主要判定應為「hepatectomy」肝臟切除術，「partial hepatectomy」表示部分肝切除術，編碼時應依據主要處置的詳細描述，查找合適的關鍵字，為「Excision」部分切除術。

依據個案的病歷記錄，將個案的「主要診斷」及「主要處置」匯出於 Microsoft Excel，分列為四項欄位：1.「個案」，2.「主要診斷／主要處置」，3.「關鍵字」，4.「主要診斷／主要處置描述判定字」，將病歷記錄整理及建檔；然而，本研究醫院一般外科個案，並無兩個或兩個以上之診斷均可為「主要診斷」之情況，但因研究時間不足，本研究並無針對多個處置進行「主要處置」判斷之系統的建立，而是由研究者判斷主要處置。因此，本研究之關鍵字判定系統，並無針對多個診斷進行「主要診斷」之判斷或多個處置進行「主要處置」之判斷的步驟。

3.4 建置疾病分類知識庫

針對 105 年出院診斷—主要診斷，共 1,326 筆主要診斷，個案有執行手術處置之主要處置，共 1,258 筆主要處置，其中 68 個案未執行手術處置之原因可能包含：1、以抗生素治療疾病，2、因病人要求至他院尋求第二意見，3、因病人之疾病狀況不良，不適合進行手術處置，4、藥物治療已改善病況，不願意進行手術處置；106 年出院診斷—主要診斷，共 1,389 筆主要診斷，個案有執行手術處置之主要處置，共 1,345 筆個案，其中 44 個案未執行手術處置之原因可能包含：1、以抗生素治療疾病，2、因病人要求至他

院尋求第二意見，3、因病人之疾病狀況不良，不適合進行手術處置，4、藥物治療已改善病況，不願意進行手術處置；105-106 年主要診斷，共 2,715 筆（如表 3.1），105-106 年主要處置，共 2,603 筆（如表 3.2），其中 112 個案未執行手術處置之原因可能包含：1、以抗生素治療疾病，2、因病人要求至他院尋求第二意見，3、因病人之疾病狀況不良，不適合進行手術處置，4、藥物治療已改善病況，不願意進行手術處置，進行知識庫整理及建檔，105-106 年主要診斷共建立 156 組主要診斷描述判定字及關鍵字，主要處置共建立 81 組主要處置描述判定字及關鍵字。

表 3.1 105-106 年主要診斷知識庫

	A	B	C	D
1	個案	出院診斷-主要診斷	關鍵字	主診斷描述判定字
2	1	ileus	ileus	ileus
3	2	R't breast IDC,cStage IV(pHx: pSTAGE:III A,T2 N2 M0)	neoplasm	IDC
4	3	gall stones with suspected cholecystitis	calculus	stones
5	4	GB stone	calculus	stone
6	5	gall stone with cholecystitis	calculus	stone
7	6	Splenic laceration - AAST gade II	laceration	laceration
8	7	R't breast microcalcification, UIQ, 12-1/3, nature	microcalcifications	microcalcification
9	8	left breast mastitis	mastitis	mastitis
10	9	GB stone with cholecystitis	calculus	stone
11	10	HCC, cT1N0M0 - stage I , pT2N0M0 stage II	carcinoma	HCC
12	11	Gallbladder stones with chronic cholecystitis	calculus	stones
13	12	Acute suppurative appendicitis	appendicitis	appendicitis
14	13	Right breast Ductal carcinoma in situ with apocrine differentiation	neoplasm	carcinoma

表 3.2 105-106 年主要處置知識庫

	A	B	C	D
1	個案	主要處置	關鍵字	主處置描述判定字
2	1	Small intestine series	fluoroscopy	series
3	2	partial mastectomy	excision	partial mastectomy
4	3	laparoscopic cholecystectomy	resection	cholecystectomy
5	4	laparoscopic cholecystectomy	resection	cholecystectomy
6	5	laparoscopic cholecystectomy	resection	cholecystectomy
7	6	splenectomy	resection	splenectomy
8	7	partial mastectomy	excision	partial mastectomy
9	8	debridement	excision	debridement
10	9	laparoscopic cholecystectomy	resection	cholecystectomy
11	10	partial hepatectomy	excision	partial hepatectomy
12	11	laparoscopic cholecystectomy	resection	cholecystectomy
13	12	Laparoscopic appendectomy	resection	appendectomy
14	13	partial mastectomy	excision	partial mastectomy

3.5 設計關鍵字之判定系統

3.5.1 主要診斷關鍵字之判定系統

疾病分類師查找主要診斷—關鍵字之編碼步驟包含：1.疾病分類師判定個案之主要診斷、2.輸入主要診斷描述、3.判斷是否僅有一個主要診斷描述判定字：是→主要診斷關鍵字，例如：主要診斷—Right inguinal indirect hernia（右側腹股溝直接型疝氣），主要診斷描述判定字：hernia（疝氣），查找的關鍵字—hernia（疝氣），關鍵字：hernia（疝氣），4.判斷是否僅有一個主要診斷描述判定字：否→以主診斷描述 with 之前的判定字為主診斷描述的判定字，例如：主要診斷—gall stone with acute cholecystitis（膽囊結石合併急性膽囊炎），主要診斷描述判定字：stone（結石）、cholecystitis（膽囊炎），以主診斷描述 with 之前的判定字為主要診斷描述的判定字，查找的關鍵字—calculus（結石），關鍵字：calculus（結石），疾病分類師查閱主要診斷—關鍵字之系統流程圖，如圖 3.3。

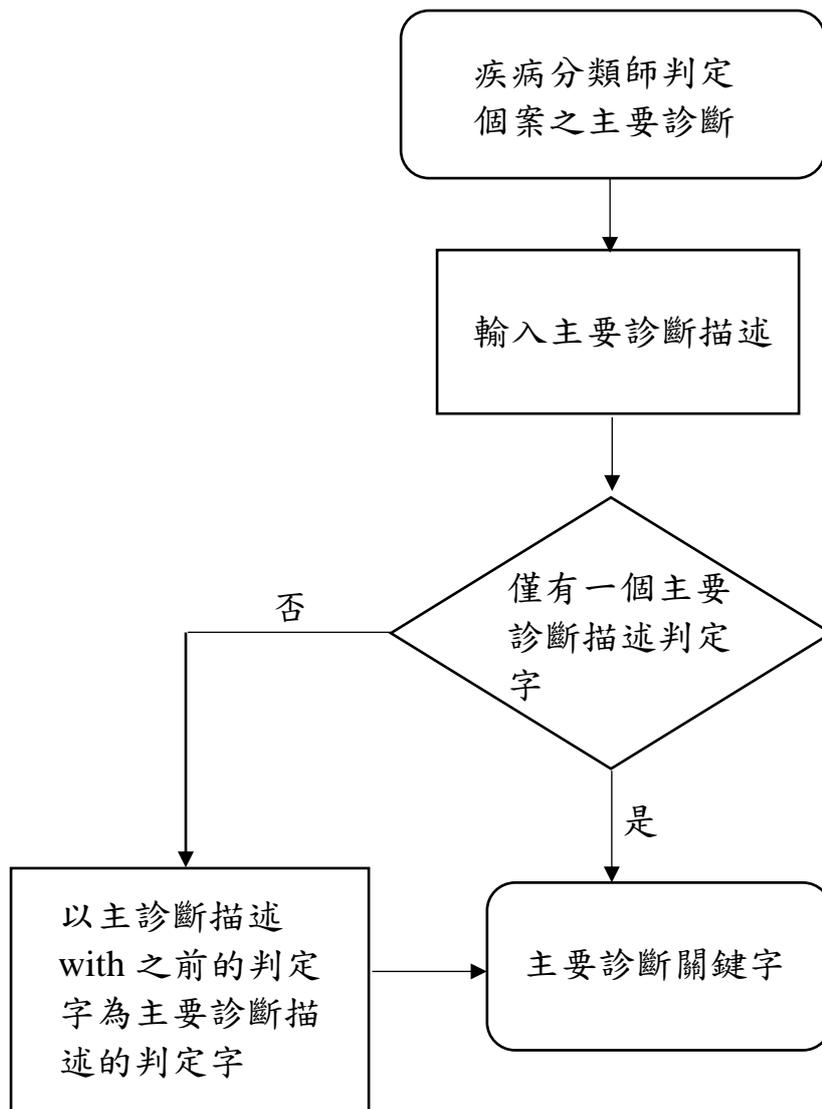


圖 3.3 主要診斷—關鍵字之系統流程圖

資料來源：本研究整理

3.5.2 主要處置關鍵字之判定系統

疾病分類師查閱「主要處置」－關鍵字之編碼步驟包含：1.疾病分類師判定個案之主要處置、2.輸入主要處置描述、3.以「主要處置描述判定字」為主要處置判定字、4.主要處置－關鍵字。

疾病分類師依據「手術方式(Root operation)之內、外科處置編碼指引」，由 ICD-10-PCS 字母索引工具書查找關鍵字。手術處置之關鍵字為治療疾病的目的，如：部分切除(Excision)，全部切除(Resection)，強化(Supplement)，繞道(Bypass)，修補(Repair)，引流(Drainage)，摘除(Extirpation)，鬆解(Release)，擴張(Dilation)，復位(Reposition)，視查(inspection)，放射(Beam radiation)...等。

例如(1)：主要處置－partial mastectomy (部分乳房切除術)，主處置描述判定字：partial mastectomy (部分乳房切除術)，查找的關鍵字－excision (部分切除術)，關鍵字：excision (部分切除術)；例如(2)：主要處置－laparoscopic cholecystectomy (腹腔鏡膽囊切除術)，主處置描述判定字：cholecystectomy (膽囊全部切除術)，查找的關鍵字－resection (全部切除術)，關鍵字：resection (全部切除術)；例如(3)：主要處置－hernia repair with mesh (疝氣修補術合併使用彌補物)，主處置描述判定字：hernia repair with mesh (疝氣修補術合併使用彌補物)，使用「彌補物」為置入合成物質或生物性物質，用來增強、強化身體部位全部、局部的功能，查找的關鍵字－supplement (強化)，關鍵字：supplement (強化)；例如(4)：主要處置－left thyroid near total lobectomy (左側甲狀腺幾近全部切除術)，主處置描述判定字：near total lobectomy (幾近全部切除術，意指”部分切除術”，”非”全部切除術)，關鍵字：excision (部分切除術)；疾病分類師查找「主要處置－關鍵字」之系統流程圖，如圖 3.4。

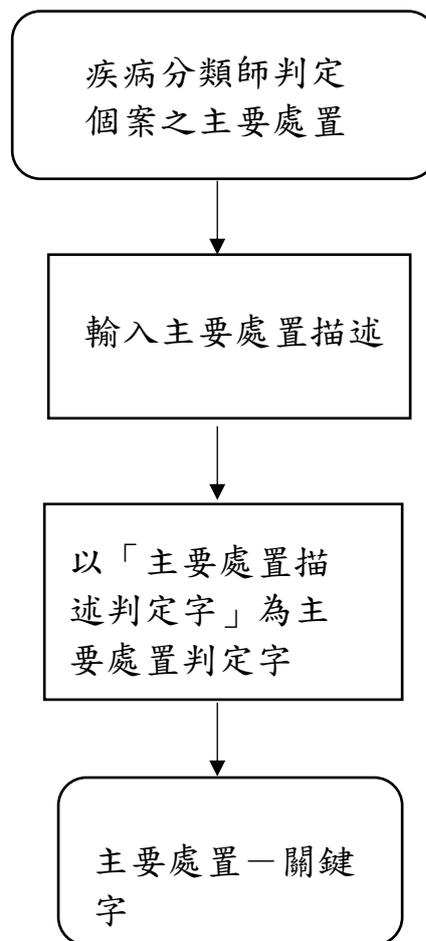


圖 3.4 主要處置－關鍵字之系統流程圖

資料來源：本研究整理

第四章 系統驗證

4.1 主要診斷系統驗證

本研究以中部某區域教學醫院,107 年一般外科出院診斷之「主要診斷」1,076 個案,以及「主要處置」1,033 個案,其中 43 個案未執行手術處置之原因可能包含:1、以抗生素治療疾病,2、因病人要求至他院尋求第二意見,3、因病人之疾病狀況不良,不適合進行手術處置,4、藥物治療已改善病況,不願意進行手術處置,5、病人決定不開刀,改以門診追蹤治療,進行案例系統驗證。

4.1.1 107 年個案—主要診斷系統驗證

將 107 年一般外科—主要診斷 1,076 個案,進行驗證步驟:1.建立資料庫 Input_data_base.xlsx,資料夾名稱:data_base-主要診斷,建立之檔案放置於電腦 C 槽,檔案名稱爲 Diag,資料庫(database)包含 105 及 106 年全部個案的主要診斷描述判定字及關鍵字,如表 4.1。

表 4.1 105-106 年主要診斷資料庫

	A	B
1	主診斷描述判定字_database	關鍵字_database
2	hernia	hernia
3	CA	neoplasm
4	stones	calculus
5	stone	calculus
6	cancer	neoplasm
7	appendicitis	appendicitis
8	carcinoma	carcinoma
9	hyperplasia	hyperplasia
10	abscess	abscess
11	goiter	goiter
12	Gallstones	calculus
13	hyperparathyroidism	hyperparathyroidism
14	fibrocystic	fibrocystic
15	diverticulosis	diverticulosis

將 107 年一般外科—主要診斷 1,076 個案，進行驗證步驟：2.建立輸入資料 Input_Diag.xlsx，資料夾名稱：107 年主要診斷，建立之檔案放置於電腦 C 槽，檔案名稱為 Diag，包含 107 年所有案例的「出院診斷-主要診斷」，如表 4.2。

表 4.2 107 年主要診斷資料庫

	A	B
1	個案	出院診斷-主要診斷
2	1	Gallstones with chronic cholecystitis
3	2	L't large incarcerated indirect inguinal hernia
4	3	Acute appendicitis with ruptured, diffuse peritonitis
5	4	Gallstones with chronic cholecystitis
6	5	Left axillary lymphadenopathy
7	6	Cholelithiasis with chronic cholecystitis and acute
8	7	ileus
9	8	R't breast microcalcification, UOQ, 10/2
10	9	HCC with rupture 5.8x6.8 cm in S2 .cT4N0M0 StageIIIB
11	10	HCC with ruptured
12	11	HCC,cT4N0M0,stage:IIIB,pT3N0M0,stage:IIIA
13	12	duodenum Gastric type polyp
14	13	gastric outlet obstruction

將 107 年一般外科—主要診斷 1,076 個案，進行驗證步驟：3 按下 exe. (如圖 4.1)，等待約 1 分鐘左右，程式產生三個檔案包含：1.無法找到主要診斷描述判定字的個案，Output_loss.csv (如表 4.3)，若同一個案有找到主要診斷描述判定字的個案，則於 A 欄呈現個案，B 欄呈現 FALSE；2.重複的主要診斷描述判定字的個案，Output_repeat.csv (如表 4.4)，若同一個案有找到重複的主要診斷描述判定字的個案，則於 A 欄呈現個案，B 欄呈現找到幾個主要診斷描述判定字；3.最後結果為 107 年找到的主要診斷描述判定字及關鍵字，Output_Search_result.csv (如表 4.5)。

107 年一般外科—主要診斷共找到 1,076 個案，證實主要診斷關鍵字判定系統建立之規則 100% 準確。

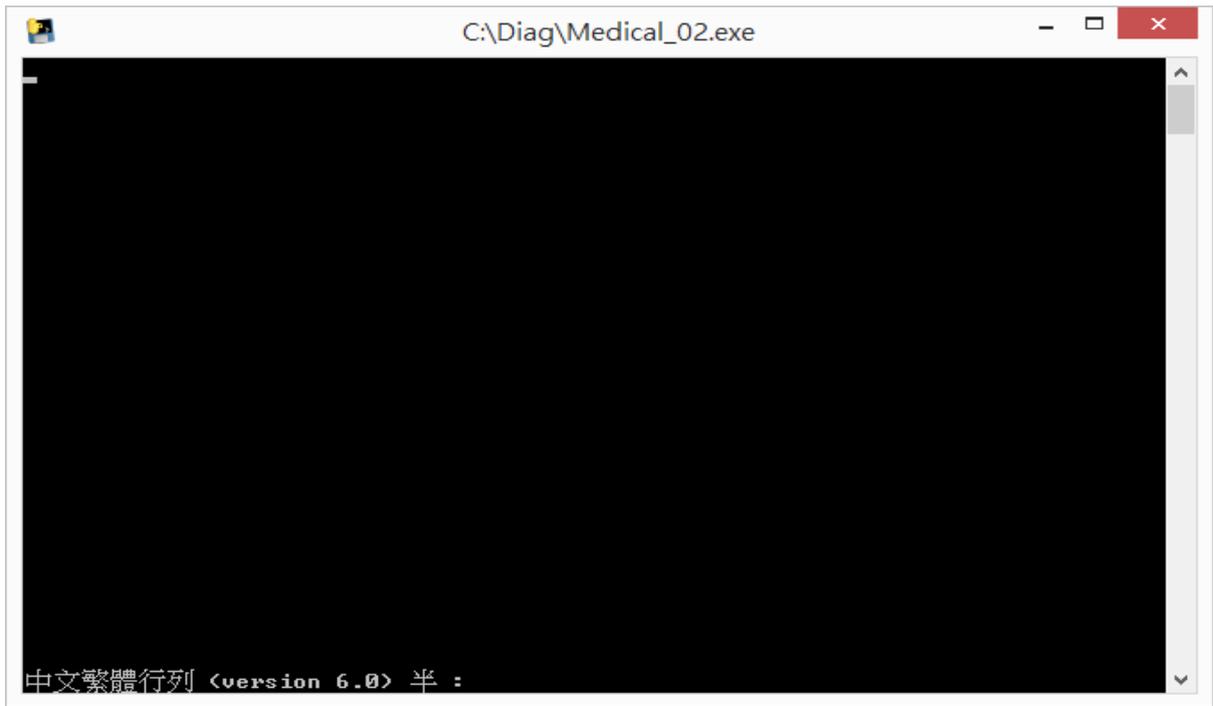


圖 4.1 程式語言 Python 之執行檔

表 4.3 無法找到主要診斷描述判定字個案

	A	B	C
1	Case	CHECK	
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			

表 4.4 重複主要診斷描述判定字個案

	A	B	C
1	Case	repeat	
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			

表 4.5 107 年找到的主要診斷描述判定字及關鍵字

	A	B	C	D
1	Case	Diag	Key_Diag	Key_DB_word
2	1	ileus	ileus	ileus
3	2	r't breast idc,cstage iv(phx: pstage:4 a ,t2 n2 m0)	idc	neoplasm
4	3	gall stones with suspected cholecystitis	stones	calculus
5	4	gb stone	stone	calculus
6	5	gall stone with cholecystitis	stone	calculus
7	6	splenic laceration - aast gade ii	laceration	laceration
8	7	r't breast microcalcification, uiq, 12-1/3, nature	microcalcification	microcalcifications
9	8	left breast mastitis	mastitis	mastitis
10	9	gb stone with cholecystitis	stone	calculus
11	10	hcc, ct1n0m0 - stage i . pt2n0m0 stage ii	hcc	carcinoma
12	11	gallbladder stones with chronic cholecystitis	stones	calculus
13	12	acute suppurative appendicitis	appendicitis	appendicitis
14	13	right breast ductal carcinoma in situ with apocrine	carcinoma	carcinoma
15	14	right breast fibroadenoma	fibroadenoma	fibroadenoma
16	15	gall stones,cbd stones with chronic cholecystitis ar	stones	calculus

4.1.2 主要診斷關鍵字判定系統之效率測試

以研究個案醫院，每一天平均每位疾病分類師，必須編碼約 30 個案，以隨機抽樣法抽取 30 個案，進行編碼耗時計算；原編碼方式為進入病歷資料查詢系統，逐一輸入個案病歷號碼，查閱出院病摘之出院診斷—主要診斷，疾病分類師判斷主要診斷，以疾病分類師之經驗知識判定關鍵字，2 位資淺疾病分類師平均耗時 1620 秒（27 分 00 秒），3 位資深疾病分類師平均耗時 930 秒（15 分 30 秒），如表 4.6。

以本研究之「智慧化編碼輔助系統」協助疾病分類師編碼，依研究個案醫院，每一天平均每位疾病分類師，必須編碼約 30 個案，以隨機抽樣法抽取 30 個案，進行編碼耗時計算；將個案之出院診斷匯出於 Microsoft Excel，依研究個案醫院 5 位疾病分類師及該個案之主治醫師共同判斷，並取得共識之「主要診斷」，複製至 Input_Diag.xlsx 之出院診斷—主要診斷，按下執行檔 exe.，2 位資淺疾病分類師及 3 位資深疾病分類師平均編碼耗時 450 秒（7 分 30 秒），如表 4.6。

表 4.6 資淺及資深疾病分類師使用關鍵字判定系統前後耗時比較表

疾病分類師／ 關鍵字判定 系統	資淺疾病分類 師／未使用關 鍵字判定系統	資深疾病分類 師／未使用關 鍵字判定系統	資淺及資深疾 病分類師／使 用關鍵字判定 系統
關鍵字			
主要診斷	1620 秒	930 秒	450 秒

4.2 主要處置系統驗證

4.2.1 107 年個案-主要處置系統驗證

將 107 年一般外科—主要診斷 1,033 個案，進行驗證步驟：1.建立資料庫 Input_data_base.xlsx，資料夾名稱：data_base-主要處置，建立之檔案放置於電腦 C 槽，檔案名稱為 Diag，資料庫(database)包含 105 及 106 年全部個案的主要處置描述判定字及關鍵字，如表 4.7。

表 4.7 105-106 年主要處置資料庫

	A	B
1	主要處置描述判定字_database	關鍵字_database
2	series	fluoroscopy
3	partial mastectomy	excision
4	cholecystectomy	resection
5	splenectomy	resection
6	debridement	excision
7	partial hepatectomy	excision
8	appendectomy	resection
9	hemicolecotomy	resection
10	hernioplasty with mesh	supplement
11	excision	excision
12	repair	repair
13	total parathyroidectomy	resection

將 107 年一般外科—主要處置 1,033 個案，進行驗證步驟：2.建立輸入資料 Input_Diag.xlsx，資料夾名稱：107 年主要處置，包含 107 年個案有執行醫療手術處置的主要處置，如表 4.8。

表 4.8 107 年主要處置資料庫

	A	B
1	個案	主要處置
2	1	laparoscopic cholecystectomy
3	2	hernioplasty with mesh
4	3	Laparoscopic appendectomy
5	4	laparoscopic cholecystectomy
6	5	excision of axillary LN tumor
7	6	laparoscopic cholecystectomy
8	7	Small intestine series
9	8	partial mastectomy
10	9	laparoscopic partial hepatectomy
11	10	laparoscopic partial hepatectomy
12	11	laparoscopic partial hepatectomy
13	12	excision of duodenal tumor

將 107 年一般外科—主要處置 1,033 個案，進行驗證步驟：3 按下 exe. (如圖 4.2)，等待約 1 分鐘，程式產生三個檔案包含，1.無法找到主要處置描述判定字，Output_loss.csv (如表 4.9)，若同一個案無法找到主要處置描述判定字，則於 A 欄呈現個案，B 欄呈現 FALSE；2.找到重複的主要處置描述判定字，Output_repeat.csv，(如表 4.10)，若同一個案找到重複的主要處置描述判定字，則於 A 欄呈現個案，B 欄呈現找到幾個重複的主要處置描述判定字；3.最後結果，107 年找到的主要處置描述判定字及關鍵字，Output_Search_result.csv (如表 4.11)。

107 年一般外科—主要處置共找到 1,033 個案，證實主要處置關鍵字判定系統建立之規則 100% 準確。



圖 4.2 程式語言 Python 之執行檔

表 4.9 無法找到主要處置描述判定字個案

	A	B	C
1	Case	CHECK	
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			

表 4.10 重複主要處置描述判定字個案

	A	B	C
1	Case	repeat	
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			

表 4.11 107 年找到的主要處置描述判定字及關鍵字

	A	B	C	D
1	Case	Diag	Key_Diag	Key_DB_word
2	1	laparoscopic cholecystectomy	cholecystectomy	resection
3	2	hernioplasty with mesh	hernioplasty with mesh	supplement
4	3	laparoscopic appendectomy	appendectomy	resection
5	4	laparoscopic cholecystectomy	cholecystectomy	resection
6	5	excision of axillary ln tumor	excision	excision
7	6	laparoscopic cholecystectomy	cholecystectomy	resection
8	7	small intestine series	series	fluoroscopy
9	8	partial mastectomy	partial mastectomy	excision
10	9	laparoscopic partial hepatectomy	partial hepatectomy	excision
11	10	laparoscopic partial hepatectomy	partial hepatectomy	excision
12	11	laparoscopic partial hepatectomy	partial hepatectomy	excision
13	12	excision of duodenal tumor	excision	excision
14	13	laparoscopic subtotal gastrectomy	total gastrectomy	resection

4.2.2 主要處置關鍵字判定系統之效率測試

依研究個案醫院，每一天平均每位疾病分類師，必須編碼約 30 個案，以隨機抽樣法抽取 30 個案，進行編碼耗時計算；原編碼方式為進入病歷資訊查詢系統，逐一輸入個案病歷號碼，查閱個案之相關手術處置，疾病分類師判斷主要處置，以疾病分類師之經驗知識判定關鍵字，2 位資淺疾病分類師平均耗時 2070 秒（34 分 30 秒），3 位資深疾病分類師平均耗時 1200 秒（20 分 00 秒），如表 4.12。

以本研究之「智慧化編碼輔助系統」協助疾病分類師編碼，依研究個案醫院，每一天平均每位疾病分類師，必須編碼約 30 個案，以隨機抽樣法抽取 30 個案，進行編碼耗時計算；將個案之手術處置匯出於 Microsoft Excel，以研究個案醫院之 5 位疾病分類師及該個案之主治醫師共同判斷，並取得共識之「主要處置」，複製至 Input_Diag.xlsx 之 107 年主要處置資料夾後，按下執行檔 exe.，2 位資淺疾病分類師及 3 位資深疾病分類師平均耗時 540 秒（9 分 00 秒），如表 4.12。

表 4.12 資淺及資深疾病分類師使用關鍵字判定系統前後耗時比較表

疾病分類師 / 關鍵字 判定系 統 關鍵字	資淺疾病分類師 / 未使用關鍵字 判定系統	資深疾病分類師 / 未使用關鍵字 判定系統	資淺及資深疾病 分類師 / 使用關 鍵字判定系統
主要處置	2070	1200	540

第五章 結論與建議

全民健康保險自 1995 年開辦至今，國際疾病分類代碼(International Classification of Disease)為各醫療院所向健保署申請醫療費用給付之重要依據，疾病分類代碼直接影響 DRG 歸屬及費用給付，疾病分類代碼的正確性和適當性越高，才能確保病患所做的醫療處置都能向健保署申請合理的給付，由此可見，疾病分類師編列代碼在醫療院所扮演著相當重要角色，然而，疾病分類代碼的編排方式，具有其分類架構及分類邏輯，關鍵字之判定及查找正確關鍵字，成為疾病分類師編碼的第一重要步驟。

經由文獻探討中發現，建置疾病分類知識庫的過程中，主要診斷與疾病病灶存在著資料與資料之間的關聯，主要處置與疾病治療目的存在著既有的準則，精準判定正確的關鍵字，進而查找正確又精確的代碼，是值得探討且具貢獻價值的問題。

疾病分類師編列疾病代碼必須仰賴多年的編碼技巧及編碼經驗，才能熟稔各種疾病診斷及手術處置之編碼細節及注意事項，準確的判斷疾病診斷與手術處置之「關鍵字」，本研究透過資深疾病分類師十二年的編碼經驗建立知識庫，並應用「文字探勘」協助建構一套主要診斷與主要處置之關鍵字判定系統，用以協助疾病分類師能更精準與快速地判斷主要診斷及主要處置之關鍵字，使疾病分類編碼工作邁入到新的階段(疾病分類編碼 3.0)。

5.1 研究成果

本研究以中部某區域教學醫院，以 105-106 年一般外科住院病人之「主要診斷」及「主要處置」病歷資料，應用文字探勘進行個案資料整理與建檔，建立關鍵字判定知識庫，其內容包含個案之主要診斷及主要處置之判定字以及其所對應之關鍵字。與設計「判定字」文字探勘之判定流程系統的規則。

最後，以 107 年一般外科住院病人之「主要診斷」及「主要處置」病歷資料加以驗證系統流程之準確性，透過程式語言(Python)進行編碼，再利用 PyInstaller 將所撰寫之語言包裝成執行檔，使用執行檔後，程式產生三個檔案，沒有出現無法找到主要診斷描述判定字的個案，Output_loss.csv；沒有出現找到重複的主要診斷描述判定字的個案，Output_repeat.csv；最後

結果，107 年病歷資料找到的主要診斷及主要處置描述判定字及關鍵字，Output_Search_result.csv；驗證結果證實此「智慧化疾病分類輔助系統」，有效地協助疾病分類師進行「主要診斷」及「主要處置」的關鍵字判斷。

本研究最後證實此「主要診斷與主要處置之關鍵字判定系統」，可以有效地降低疾病分類師進行「主要診斷」及「主要處置」的關鍵字判斷之耗時時間，以隨機抽樣法抽取 30 個案，資淺疾病分類師尚未使用「疾病分類編碼輔助系統」前，查找主要診斷個案需花費 1620 秒，資深疾病分類師查閱主要診斷個案需花費 930 秒，資淺及資深疾病分類師使用「疾病分類編碼輔助系統」，查找主要診斷個案需花費 450 秒（如圖 5.1）；而對於主要處置個案，資淺疾病分類師尚未使用「疾病分類編碼輔助系統」前，查找主要診斷個案需花費 2070 秒，資深疾病分類師查閱主要診斷個案需花費 1200 秒，資淺及資深疾病分類師使用「疾病分類編碼輔助系統」，查找主要診斷個案需花費 540 秒（如圖 5.2），有效減少編碼工時，進而有效降低因人為疏失所導致之編碼錯誤。

本研究建置之「關鍵字判定系知識庫」為其導入編列正確代碼之重要入門，若醫療院所疾病分類師人力不足時，即使非為疾病分類師，也能夠依編碼操作流程，依據「主要診斷與主要處置之關鍵字判定系統」判定關鍵字，以正確之診斷及處置之關鍵字，編列疾病分類代碼，以補足疾病分類師人力不足的缺口。

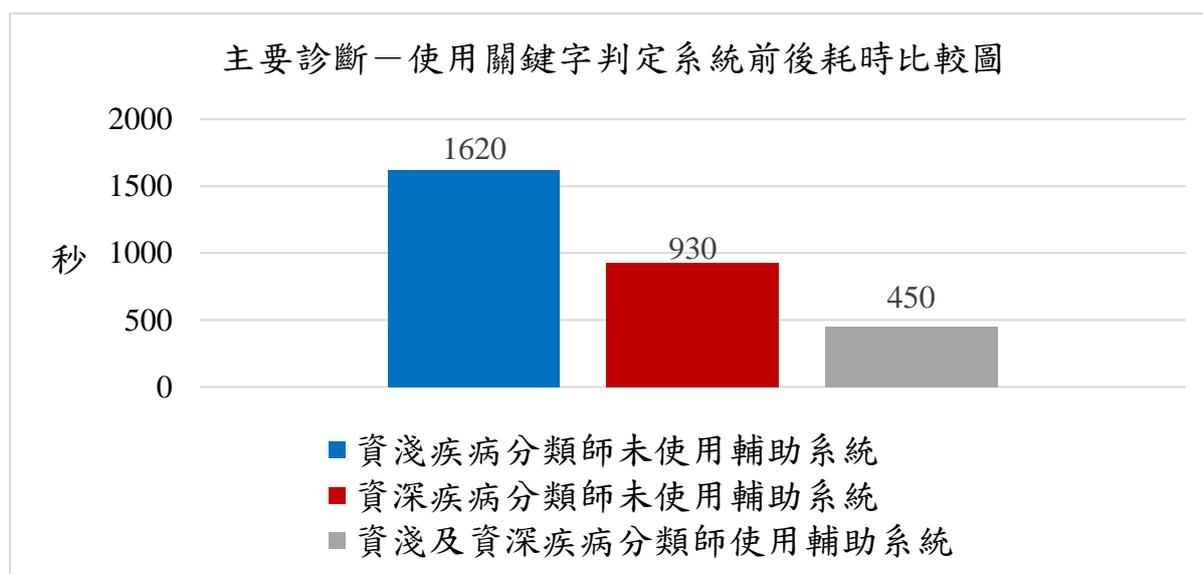


圖 5.1 主要診斷—使用關鍵字判定系統前後耗時比較圖

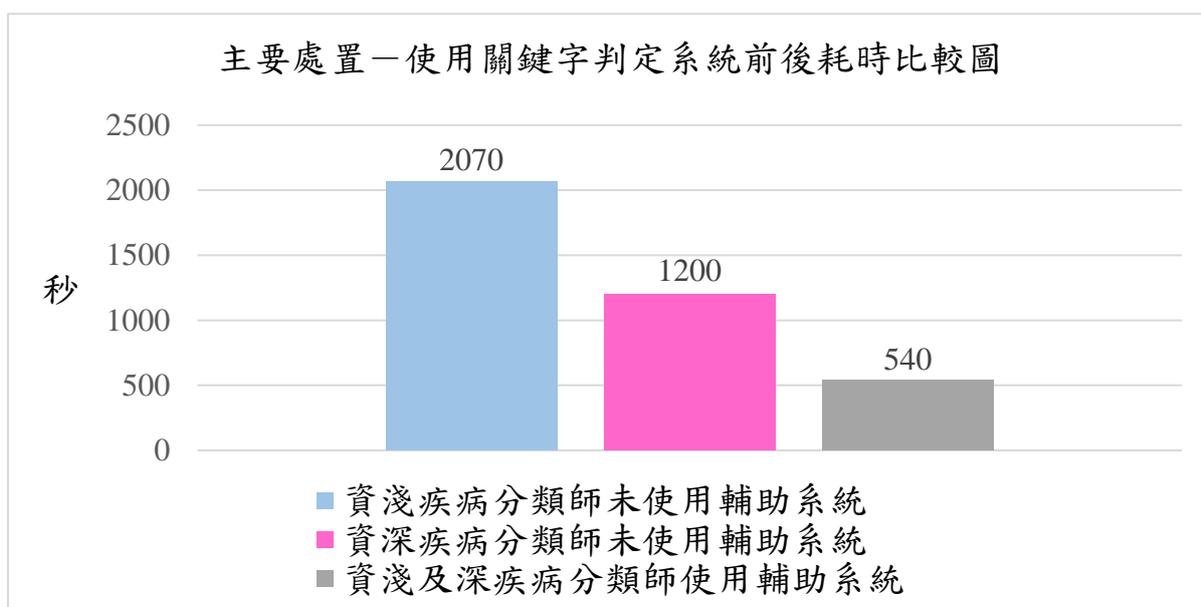


圖 5.2 主要處置—使用關鍵字判定系統前後耗時比較圖

5.2 未來研究方向

本研究提出之「關鍵字判定系統」，得以精準且快速地判定「主要診斷」及「主要處置」的關鍵字，經由系統驗證證實判定系統百分之百準確，期望未來能夠建置疾病嚴重度較高的科別，例如：心臟外科、整型外科、骨科等極具複雜度的科別；具有兩個或兩個以上等多重診斷和處置，均可為「主要診斷」及「主要處置」多重選擇之判定，加入本研究中，並進行「主要診斷」及「主要處置」智慧化的判斷，以突破疾病分類編碼的複雜度及困難度，增加編碼的熟悉度，避免因人為疏失，導致編列錯誤代碼，進而影響疾病分類編碼品質。

參考文獻

中文部分

1. 古思明、林俊榮、林政彥 (2010)。我國疾病分類人員角色認知調查研究。《病歷管理期刊》，9(2)，84-103。
2. 李玲美 (2016)。台灣北部某區域醫院針對 ICD-10-CM/PCS 實施的管理分享。《病歷管理期刊》，15(1)，42-51。
3. 李瑞華 (1992)。疾病分類人員於閱讀病歷時之注意事項。《醫院》，25(1)，58-60。
4. 范碧玉、黃麗秋、鄭茉莉、簡雅芬 (2002)。醫療院所申報健保住診費用之疾病分類編碼適當性分析。《病歷管理期刊》，2(2)，34-54。
5. 范碧玉、賴淑芬、張晏晏、薛德興、莊孟嫻 (2013)。我國疾病分類採用 ICD-10-CM/PCS 的重要因素及推動策略探討。《病歷管理期刊》，12(2)，47-68。
6. 范碧玉、賴淑芬 (2014)。ICD-10-CM/PCS 分類規則彙編。1-36。
7. 吳瑞堯、周駿賢 (2011)。運用資料探勘技術於六大死因慢性疾病之研究。《資訊管理學報》，18(1)，187-211。
8. 陳世榮 (2015)。社會科學研究中的文字探勘應用：以文意為基礎的文件分類及其問題。《人文及社會科學集刊》，27(4)，684。
9. 張芸濤、龔玲 (2007)。資料探勘原理與技術。臺北市：五南文化。
10. 梁定澎 (2006)。決策支援系統與企業智慧 = Decision support systems and business intelligence。臺北市：知識達。
11. 黃于珊、施驊真 (2004)。疾病分類編碼品質與效率。《病歷管理期刊》，4(2)，58-66。
12. 黃瑞典、柯成國、藍守仁、嚴雅音 (2010)。疾病分類編碼品質對住院醫療給付影響之研究。《病歷管理期刊》，10(1)，12-26。
13. 曾卉珍、徐均宏 (2016)。ICD-9 轉換為 ICD-10 對疾病分類人員造成的工作壓力之探討-以北部某區域教學醫院為例。《醫院》，49(3)，28-38。
14. 楊美雪、蘇秀玉、施驊真、陳寶如、范春寶 (2007)。疾病分類編碼一致性與其 DRGs 分派之影響-以膽結石處置病例為例。《病歷管理期刊》，6(2)，53-64。
15. 鄭凱今 (2011)。應用文字探勘技術於疾病分類碼之研究-以腸胃科病歷摘要為例(碩士論文)。國立雲林科技大學，雲林縣。
16. 楊正銘 (2004)。以文字探勘技術應用於疾病分類之輔助系統-以出入院病歷摘要為例(碩士論文)。台北醫學大學，臺北市。
17. 魏秀美 (2004)。醫療機構健保申報品質之研究-以北部某醫學中心申報與疾病分類一致性為例(碩士論文)。陽明大學，臺北市。
18. 謝和昆 (2009)。疾病分類支援系統之建構-以神經外科與放射科為例(碩士論文)。

東海大學，台中市。

19. 謝邦昌 (2017)。Text Mining 文本探勘。臺北市：元華文創。
20. 衛生福利部中央健康保險署 (2014)。推動及應用國際疾病分類第十版 ICD-10-CM/PCS 於臨床疾病分類計劃—ICD-10-CM/PCS 疾病分類教育訓練教案。
https://www.nhi.gov.tw/Content_List.aspx?n=3C0BD93EC042CC67&topn=D39E2B72B0BDFA15。取得日期 2019 年 1 月 22 日。
21. 衛生福利部中央健康保險署 (2015)。全民健康保險住院診斷關聯群 (Tw-DRG) 案件審查注意事項—通則。
https://www.nhi.gov.tw/Content_List.aspx?n=F075AD6CF8E194E0&topn=D39E2B72B0BDFA15。取得日期 2019 年 3 月 31 日。
22. 衛生福利部中央健康保險署 (2015)。DRG 支付制度-支付通則全民健康保險住院診斷關聯群 (Tw-DRG) 支付通則。
http://www.nhi.gov.tw/webdata/webdata.aspx?menu=23&menu_id=819&webdata_id=937&WD_ID=819。取得日期 2019 年 1 月 26 日。

英文部分

1. Ariena, H. C., D. A. Neher and T. R. Weicht. (1991). Teaching computer-based diagnosis of plant disease. *Plant Dis.* 75(3), 320-322.
2. Baker, J. J. (2002). Medicare Payment System for Hospital Inpatients, Diagnosis-Related Groups. *Journal of health Care Finance*, 28(3), 1-13.
3. Berry, M. A., Linoff, G. S. (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. Wiley Publishing, Inc. ISBN 0-471-47064-3.
4. Blake, C. (2011). Text mining. *Annual review of information science and technology*, 45(1), 121-155.
5. Boyd, D. W. and M. K. Sun. (1994). Prototyping an expert system for diagnosis of potato diseases. *Computers and Electronics in Agriculture*. 10(3), 259-267.
6. Chen, M. S., Han, J., Yu, P. S. (1996). *Data Mining: An Overview from a Database Perspective*. *IEEE transactions on knowledge and data engineering*, 8(6), 866-883.
7. Daniel, T. L., Chantal D. L., (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, Inc., Hoboken, New Jersey.
8. Debry, C. A., Lapane, K. L., Feldman, H. A., Carleton, R. A. (2001). Possible Effect of DRGs on the Classification of Stroke. *Stroke*, 32, 1487-1491.
9. Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From Data Mining to knowledge Discovery in Databases. *AI magazine*. 17(3), 37-51.
10. Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *AI magazine*, 39(11), 27-34.
11. Feldman, R., Sanger J. (2007). *The text mining handbook: Advanced Approaches in Analyzing Unstructured Data*. 13-14.
12. Goodell, P. B., J. F. Strand and M. Ostergard. (1993). Delivering expert systems to agriculture: Experience with CALEX/Cotton. *AI Appl.* 7, 14-20.
13. Jabbar, H. K., Khan, R. Z. (2015). Survey on development of expert system in the areas of Medical, Education, Automobile and Agriculture. 2015 2nd International Conference on Computing for Sustainable Global Development.
14. Han, J., Pei, J., Kamber, M. (2011). *Data mining: concepts and techniques* (3th ed.), Morgan Kaufmann, ISBN 978-0-12-381479-1.
15. Hale, R. (2005). Text Mining: Getting more value from literature resources, *Published in Drug Discovery Today*, 10(6), 377-379.
16. Journal of American Health Information Management Association. (2009). *Transitioning to ICD-10-CM/PCS-an academic timeline*. 80, 59-64.
17. Lodhi, O. S. (2012). *Most commonly used techniques in data mining*. *International Journal of Advances in Engineering Research*. 4(3).

- Available from : <http://www.ijaer.com>. [Accessed 5 Mar 2019]
18. Majerowicz, A. (2011). *Developing an ICD-10-CM/PCS Coder Training Strategy*. Journal of AHIMA.
Available from : <http://library.ahima.org/doc?oid=103803#.XMURP8J7IMs>
[Accessed 1 Mar 2019]
 19. Reid, B., Allan, C., McIntosh, J. (2005). Investigation of leukemia and lymphoma AR-DRGs at a Sydney teaching hospital, *Health Information Management*, 34(2), 34-39.
 20. Sabanovic, Z., Masic, I. (2001). Computerized information system support in continuous quality improvement in hospital care, *Med Arh*, 55(2), 113-116.
 21. Stavem K., Bjerke G., Kjelsberg FN., Ruud EA., Saxrud SO. (2002). Coding of diagnosis in chronic obstructive pulmonary disease-economic consequences, *Tidsskr Nor Laegeforen*, 122(23), 2290-2293.
 22. Sullivan, D. (2001). *Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing, and Sales*. John Wiley & Sons, Inc. New York.
 23. Travis, J. W., E. Rajotte, R, R. Bankert, K. D. Hickey, L.A. Hull, V. Eby, P. H. Heinemann, R. Crassweller, J. McClure, T. Bowser and D. Laughland. (1992). A working description of the Penn State apple orchard consultant, an expert system. *Plant Dis*. 76(6), 545-554.
 24. World Health Organization (2018, June 18). *History of the development of the ICD*.
Available from : <https://www.who.int/classifications/icd/en>
[Accessed 10 Mar 2019]
 25. Witten, I. H., Frank, E., Hall, M. A., Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (9th ed.), Morgan Kaufmann, ISBN 978-0-12-804291-5.

附錄一 ICD-10-CM 工具書介紹

ICD-10 臨床修訂版(ICD-10-CM)，包含第一冊及第二冊，共二冊，其內容如下列詳細說明（健保署，2014）：第一冊：疾病代碼列表說明(Tabular List of diseases)，第二冊：字母索引(Alphabetic Index of diseases)。

第一冊：疾病分類代碼列表說明架構(Tabular List of diseases)，（1）疾病分類代碼之各疾病列表大多數章節以「解剖學」為分類主軸，少部份的章節以「病因學」（如腫瘤）或依其他標準（如外部引起之原因）為分類主軸，共分成二十一章。

第一章：感染症和寄生蟲(A00-B99)

第二章：腫瘤(C00-D49)

第三章：血液和造血器官及涉及免疫機轉的疾患 (D50-D89)

第四章：內分泌、營養和代謝疾病(E00-E89)

第五章：精神與行為疾患(F01-F99)

第六章：神經系統與感覺器官的疾病 (G00-G99)

第七章：眼睛及其附屬器官的疾病 (H00-H59)

第八章：耳與乳突之疾病(H60-H95)

第九章：循環系統疾病(I00-I99)

第十章：呼吸系統的疾病(J00-J99)

第十一章：消化系統的疾病(K00-K95)

第十二章：皮膚及皮下組織的疾病 (L00-L99)

第十三章：肌肉骨骼系統及結締組織疾病(M00-M99)

第十四章：生殖泌尿系統疾病(N00-N99)

第十五章：妊娠、生產與產褥期合併症(O00-O9A)

第十六章：源於週產期的指引(P00-P96)

第十七章：先天性畸形、變形與染色體異常(Q00-Q99)

第十八章：症狀、癥候、與臨床和實驗室的異常發現(R00-R99)

第十九章：傷害中毒與其它外因造成的特定影響 (S00-T88)

第二十章：導致罹病或致死之外因(V00-Y99)

第二十一章：影響健康狀況及健康服務(Z00-Z99)

ICD-10-CM 代碼的第一碼是大寫英文字母，每一個字母代表一特定的章節。字母 D 和 H 除外，字母 D 分別屬於工具書第二章（腫瘤）和第三章（血液、造血器官及免疫機轉有關的疾病）兩個章節。字母 H 則分屬工具書第七章（眼睛及附屬器官的疾病）和第八章（耳及乳突疾病）兩個章節。工具書內的代碼仍有很多代碼尚未被使用，為考量將來的代碼修正或者擴充需求（例如，新興的疾病）。

第二冊 字母索引(Alphabetic Index of diseases) ，ICD-10-CM內的字母索引分為三個部分：第一部份為第一章到十九章及二十一章，但「腫瘤一覽表」及「藥品及化學物質一覽表」除外。2. 第二部份為「腫瘤一覽表」及「中毒及藥品副作用的藥品」及「化學物質一覽表」。3. 第三部分為疾病外因碼索引及歸類於工具書第二十章之字詞。

附錄二 ICD-10-PCS 工具書介紹

ICD-10-PCS具有邏輯性、一致性架構的分類系統。個別字母與數字稱為數值(values)，依順序填滿分類代碼的七個位置，稱之為碼(characters)。每一個分類代碼都有七位碼，每位碼均代表處置的治療目的。

第1位碼 (Character 1)：章節 (Section)，包含內、外科相關的處置，共九個章節 (如表2.6)，由B到D及第F到H章節，構成了ICD-10-PCS的輔助章節，共有六個章節 (如表2.7)；第2位碼 (Character 2)：身體系統 (Body System)，用來定義執行處置的解剖區域或生理系統；第3位碼 (Character 3)：手術方式 (Root Operation)，用來定義處置的治療目的；第4位碼 (Character 4)：身體部位 (Body Part)，用來定義執行處置的身體部位或特殊的解剖部位；第5位碼 (Character 5)：手術途徑 (Approach)，用來定義到達手術部位的途徑或所使用的技術；第6位碼 (Character 6)：裝置物 (Device)，用來定義執行處置後，留置於身體內的裝置物；第7位碼 (Character 7)：修飾詞 (Qualifier)，用來定義處置的附加特性，例如，Diagnostic (診斷性)。

表 2.6 處置代碼第 1 碼包含之主要章節

章節數值	章節名稱
1	Obstertrics (產科相關處置)
2	Placement (外在裝置物之置放)
3	Administration (輸液治療)
4	Measurement and Monitoring (測量或監測)
5	Extracorporeal Assistance and Performance (體外輔助器械之執行)
6	Extracorporeal Therapies (體外療法)
7	Osteopathic (整骨術)
8	Other Procedures (其他處置)
9	Chiropractic (按脊術)

表 2.7 處置代碼第 1 碼包含之輔助章節

章節數值	章節名稱
B	Imaging (影像)
C	Nuclear Medicine (核子醫學)
D	Radiation Oncology (放射腫瘤學)
F	Physical Rehabilitation and Diagnostic Audiology (復健及診斷性聽力學)
G	Mental Health (心理衛生)
H	Substance Abuse Treatment (藥物濫用治療)