

將AI科技導入國家考試

南臺科技大學 資工系
鄭淑真



Southern Taiwan University of Science and Technology



Intelligent System Lab. (iLab)



國家考試現有的需求

- 題庫難易度只能試後分析無法試前預測
- 如何檢視難易度分配比例是否洽當?
- 如何精進題庫試題品質?
- 如何篩檢並更新題庫電子試題，淘汰不合時宜的試題?
- 由專家針對題庫試題審題是否與考古題相似，僅能搜尋相同文字之考題，無法找出核心概念雷同之考題
- 由專家針對題庫試題審題是否偏離科目重點，工作繁重
- 如何檢視歷年題庫試題之核心概念?
- 共同科目之試題對應至不同等級考試之難易度難以訂定



科技部107年度AI專案研究計畫

----人工智慧試題難易度預測分析系統

- 蒐集考古題試題參數(難度、鑑別度)進行分析，並與新試題進行比對，**評估新試題可用性，提供教師作為出題參考。**
 - **處理難易度的方式：**分析每題核心關鍵詞，計算關鍵詞之間的相關性，再計算題目之間的相似度，以考古題當中相似度最高的題目難易度，預測新題目難易度。
 - **訓練數據量：**107年度考選部所提供的初步資料均已進行處理，共處理**1,620道資訊類型考題、1,025題申論題型、建立詞庫5,156個詞、766,938個文字，切割出19,403項資訊類詞語。**
 - **網路資源量：**借由爬蟲工具蒐集類科網路文章
 - **預測資料的正確：**目前關鍵詞數量眾多，初步估計使用隨機抽樣關鍵詞檢視正確率，經由專家判斷達正確率**90%**。



Southern Taiwan University of Science and Technology



Intelligent System Lab. (iLab)



研究成果demo畫面



Southern Taiwan University of Science and Technology



Intelligent System Lab. (*iLab*)



選擇科目

- 原住民四-類科(電子)-科目(計概)
- 地方四-類科(電子)-科目(計概)
- 外交四-類科(外交行政人員資訊組)-科目(資訊安全與網路管理概要)
- 外交民航原住民四-類科(電子)-科目(計概)
- 普考-類科(電子)-科目(計概)
- 警察鐵路員級四-類科(電子)-科目(計概)
- 警察鐵路退除役員級四-類科(電子)-科目(計概)
- 鐵路員級-類科(資訊處理)-科目(計算機概要)
- 關務身心四-類科(資訊處理)-科目(計算機概要)
- 關務身心四-類科(電子)-科目(計概)
- 高普考四-類科(電子)-科目(計概)

1040615-警察鐵路退除役員級四-類科(電子)-科目(計概)-33

33 下列關於網路處理封包碰撞的敘述，何者錯誤？(A)交換器 (switch) 可以減少碰撞的機會 (B)IEEE 802.11 採用碰撞避免 (collision avoidance) 的方式來減少碰撞 (C)乙太網路會進行碰撞偵測 (collision detection) ，若發現碰撞之後則立即重傳封包 (D)乙太網路使用碰撞偵測的方式，其效能與網路線的長度有關 【1040615 - 員級四 - 計概 - 電子】 C 8 - 3 。 關鍵字：封包碰撞，交換器 (switch) ， IEEE 802.11 ， 碰撞避免 (collision avoidance) ， 乙太網路，碰撞偵測 (collision detection) ； 難易度： 3

考題難易度： 3 碰撞偵測 乙太網路 封包

1040615-警察鐵路退除役員級四-類科(電子)-科目(計概)-34

34 下列有關TCP/IP 網路架構的應用層 (application layer) 的敘述，何者錯誤？(A)應用層負責提供服務給網路使用者 (user) (B)應用層提供網域名稱 (domain name) 與IP 位址 (address) 之對映服務 (mapping service) (C)應用層提供FTP (file transfer protocol) 服務 (D)應用層傳送資料時以封包 (packets) 為資料單位 (data unit) 【1040615 - 員級四 - 計概 - 電子】 D 8 - 1 。 關鍵字：TCP/IP 網路架構，應用層 (application layer) ； 難易度： 2

考題難易度： 2 應用層 服務 封包

1040615-警察鐵路退除役員級四-類科(電子)-科目(計概)-36

106070_6808-鐵路員級-類科(資訊處理)-科目(計算機概要)-14

指定科目，可在題庫中搜尋出含有該關鍵詞的考題(資訊類)

選擇科目

第一次-護理師-內外科護理學

第一次-護理師-基本護理學

第一次-護理師-產兒科護理學

第一次-類科(護理師)-科目(內外科護理學)

第一次-類科(護理師)-科目(基礎醫學)

第一次-類科(醫師(二))-科目(醫學(三))

第一次-類科(醫師(二))-科目(醫學(五))

第一次-類科(醫師(二))-科目(醫學(四))

第二次-護理師-內外科護理學

第二次-類科(護理師)-科目(內外科護理學)

第二次-類科(護理師)-科目(基本護理學)

第二次-類科(護理師)-科目(基礎醫學)

第二次-類科(醫師(一))-科目(醫學(二))

第二次-類科(醫師(二))-科目(醫學(六))

106020_3302-第一次-類科(醫師(二))-科目(醫學(五))-8

8.進行全靜脈營養 (total parenteral nutrition, TPN) 時，下列與中央靜脈導管相關之併發症的敘述，何者錯誤？ A.導管內血栓 (thrombus) B.氣胸 (pneumothorax) C.臂叢神經損傷 (brachial plexus injury) D.尿路感染 (urinary tract infection)

考題難易度：易

中央靜脈

靜脈導管

氣胸

血栓

106100_2301-第二次-類科(醫師(一))-科目(醫學(二))-82

82.在續發性高血液凝固狀態 (hypercoagulable states) 的因素中，下列何者發生血栓的風險相對性最低？ A.長期臥床 B.心肌梗塞 C.心房顫動 D.抽菸

考題難易度：中

血液凝固

血栓

心房

106100_2301-第二次-類科(醫師(一))-科目(醫學(二))-83

106030_3106-第一次-護理師-內外科護理學-52

指定科目，可在題庫中搜尋出含有該關鍵詞的考題(醫護類)

搜尋項目: 主記憶體 其他相關: 存取 | 快取記憶體 | 暫存器 |

列出所搜尋之關鍵詞相關的一系列關鍵詞(資訊類)

存取

102080_6810-員級鐵路人員-類科(電子)-科目(計概)-13

102080_6810-員級鐵路人員-類科(電子)-科目(計概)-37

102090_5405-普考-類科(電子)-科目(計概)-3

102190_5438-地方四-類科(電子)-科目(計概)-27

102190_5438-地方四-類科(電子)-科目(計概)-5

103050_5142-關務身心四-類科(電子)-科目(計概)-31

103070_6812-員級鐵路人員-類科(電子)-科目(計概)-5

103080_5417-普考-類科(電子)-科目(計概)-32

103180_5436-地方四-類科(電子)-科目(計概)-8

1040419-關務身心四-類科(電子)-科目(計概)-1

1040419-關務身心四-類科(電子)-科目(計概)-30

30 有關C++中class 對於data member 與member function 存取權限的敘述，下列何者錯誤？(A)提供public、protected 與private 三種存取權限 (B)若某class A 的data member 為public，則在同一程式的其他函式都可以讀寫該public data member (C)class 的constructor 的存取權限必須為public (D)若某class A 的data member 為private，則所有不屬於class A 的函式都不可以讀寫該private data member 【1040419 - 身心四 - 計概 - 電子類】D 4 - 6，關鍵字：C++，class，data member，member function；類別成員存取權限 (Accessibility) 難易度：3

考題難易度：3 函式 存取

1040419-關務身心四-類科(電子)-科目(計概)-7

7 下列何者是循序存取的輸出入設備？(A)硬碟機 (B)磁帶機 (C)軟碟機 (D)CD 光碟機 【1040419 - 身心四 - 計概 - 電子類】B 5 - 5，關鍵字：循序存取，輸出入設備，硬碟機，磁帶機，軟碟機，CD 光碟機；難易度：1

考題難易度：1 軟碟機 光碟機 循序 硬碟 設備

搜尋項目: 靜脈 其他相關: 下腔靜脈 | 肺動脈 | 動脈 | 呼吸 |

列出所搜尋之關鍵詞相關的一系列關鍵詞(醫護類)

下腔靜脈

104030_1109-第一次-類科(護理師)-科目(基礎醫學)-13

104030_4109-第一次-類科(護理師)-科目(產兒科護理學)-16

106030_1106-第一次-護理師-基礎醫學-26

26 胎血循環中，靜脈導管連接臍靜脈與下列何者？
 門靜脈 上腔靜脈 下腔靜脈 肝臟

考題難易度：適中

下腔靜脈

臍靜脈

肺動脈

104030_2102-第一次-類科(醫師(二))-科目(醫學(四))-32

104030_3109-第一次-類科(護理師)-科目(內外科護理學)-29

104090_3302-第二次-類科(醫師(二))-科目(醫學(五))-3

106020_2302-第一次-類科(醫師(二))-科目(醫學(四))-13

106080_3302-第二次-類科(醫師(二))-科目(醫學(五))-15

106100_1301-第二次-類科(醫師(一))-科目(醫學(一))-20

20. 當以食指穿入心包橫膈 (transverse pericardial sinus) ， 那些血管會位於食指之前？
A. 肺動脈與下腔靜脈 B. 主動脈與上腔靜脈 C. 主動脈與肺動脈 D. 肺靜脈與上腔靜脈

考題難易度：中

主動脈

肺動脈

上腔靜脈

106-高普考四-類科(電子)-科目(計概)

information

1060711-高普考四-類科(電子)-科目(計概)-4	+
1060711-高普考四-類科(電子)-科目(計概)-5	+
1060711-高普考四-類科(電子)-科目(計概)-7	-

7 關於資料鏈結層 (Data Link Layer) 的敘述，下列何者正確？ (A)TCP/IP 協定無定義此層 (B)通常再細分為媒體存取控制及實體介面兩個子層 (C)無提供訊框包裝 (framing) (D)OSI 參考模型第一層 【1060711 - 普考四 - 計概 - 電子】 A 8 - 1， 關鍵字：資料鏈結層 (Data Link Layer)，TCP/IP，媒體存取控制，實體介面子層，訊框包裝 (framing)，OSI 參考模型； 難易度：2

考題難易度：2
預測難易度：2
關鍵字：協定 媒體存取控制 訊框 資料鏈結層 實體
相似題目

1060711-高普考四-類科(電子)-科目(計概)-12	-
-------------------------------	---

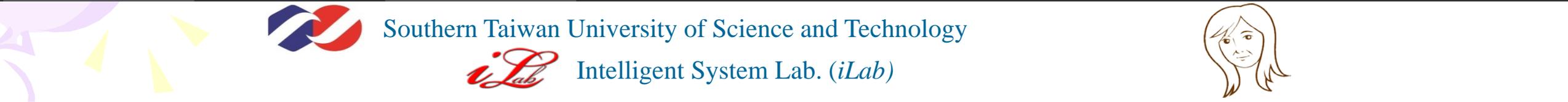
12 某嵌入式處理器僅具有加法器 (adder) 與移位器 (shifter)，而不具備乘法器，欲執行運算 $F = A * 14$ ，下列運算方式何者正確？ (A) $F = A \ll 4$ (B) $F = (A \ll 4) + (A \ll 1)$ (C) $F = (A \ll 4) - (A \ll 1)$ (D) $F = (A \ll 5) - (A \ll 2)$ 【1060711 - 普考四 - 計概 - 電子】 C 4 - 4， 關鍵字：加法器 (adder)，移位器 (shifter)； 難易度：3

考題難易度：3
預測難易度：2
關鍵字：運算 加法器 嵌入式
相似題目

列出與所選的題目相似度高的一系列考題

共含有12筆相似題目，相似題目包含：

- 12 $AB + \bar{A}B$ 相當於對 A 與 B 進行那一種邏輯運算？ NAND NOR XOR Equivalence
【關鍵字】：運算
【相似程度】：99.999%
- 15 有關 NOR 閘之敘述，下列何者錯誤？ 所有輸入皆為 0 時，輸出才為 1 先執行 NOT 運算再做 OR 運算 可將數個 NOR 閘連接起來設計成一個 AND 閘 任何邏輯布林電路都可以只用 NOR 閘就設計出來
【關鍵字】：運算,連接,布林,輸入
【相似程度】：90.015%
- 21 下列關於作業系統的敘述，何者錯誤？ 介於電腦硬體與應用程式之間的軟體 負責控制電腦的硬體及分配電腦的資源 Microsoft Excel 為常見的作業系統之一 提供執行應用軟體的環境
【關鍵字】：作業系統,硬體,軟體,應用程式,環境
【相似程度】：70.018%
- 1 某處理器以 8 位元(bit)加法器進行無號整數(unsigned integer)運算，下列運算中，何者將產生溢位(overflow)？ (A)12+25 (B)37+128 (C)206+57 (D)128+127 【1060911 - 原民四 - 計概 - 電子】 C 關鍵字：無號整數(unsigned integer)，溢位(overflow)； 整數表示法 難易度：1
【關鍵字】：運算,溢位,加法器,整數,處理器
【相似程度】：82.219%
- 22 下列何者不是使用程序(procedure)的優勢？ (A)將功能模組化 (B)抽離複雜運算中的部分運算，以利於理解程式 (C)減少程式中重複出現的程式碼片段 (D)加快程式執行速度 【1060911 - 原民四 - 計概 - 電子】 D 4 - 3， 關鍵字：程序(procedure)，模組化； 結構化程式設計 難易度：2
【關鍵字】：運算,模組,程式碼,程序
【相似程度】：84.366%



人工智慧技術---文字探勘



Southern Taiwan University of Science and Technology



Intelligent System Lab. (iLab)



使用AI文字探勘技術

- 經由AI技術針對考古題文字進行斷詞
- 自動計算每個關鍵詞之重要性

1 假設虛擬記憶體的虛擬位址 (virtual address) 是 32-bit，一個分頁 (page) 的大小是 4KB，則該虛擬記憶體共有幾個分頁？
(A)220 220
(B)232
(C)212
(D)218
Ans:A

虛擬記憶體 虛擬記憶體 virtual address bit
page KB virtual-address

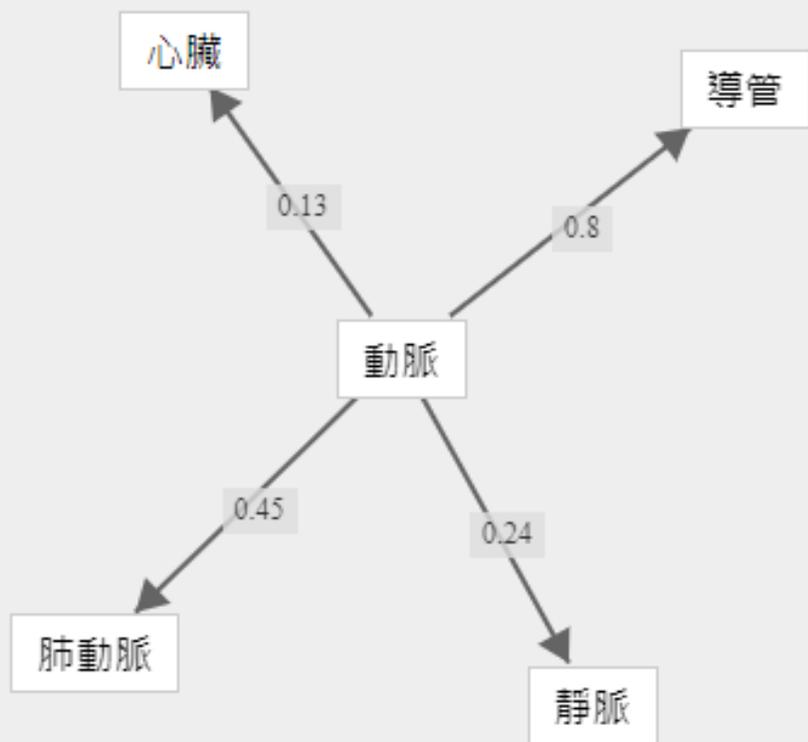
表一、題目關鍵詞之TF-IDF數據

Words	TF Value	IDF Value	TF-IDF Value
虛擬記憶體	0.036363636	5.446359543	0.198049438
virtual-address	0.018181818	7.004504161	0.127354621
kb	0.018181818	6.193573944	0.112610435
page	0.018181818	5.618209799	0.102149269
bit	0.018181818	5.299756068	0.096359201
virtual	0.018181818	5.171922697	0.094034958
address	0.018181818	4.956811317	0.090123842

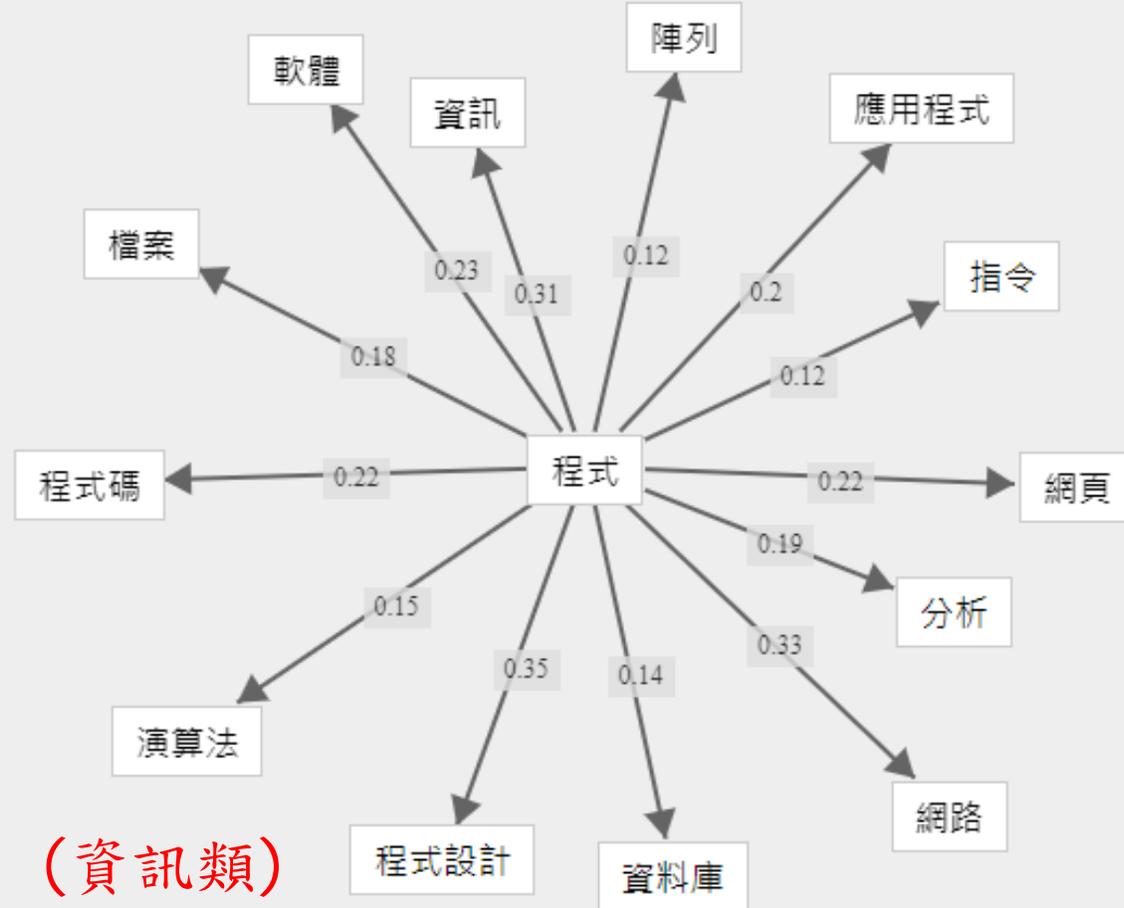


使用AI文字探勘技術

- 自動計算關鍵詞之間的相關性(數字代表相關性高低)



(醫護類)



(資訊類)

AI 文字探勘結果之應用

- 建立每道題庫試題之核心關鍵詞
- 彙整歷年題庫試題之核心概念
- 挖掘新舊資訊差異與變更，進而篩檢並更新題庫電子試題
- 題庫試題審題(是否偏離科目重點)、入闈協助決定試題
- 當年度命題與考古題之相似度分析
- 共同科目之試題相似度分析



試題難易度分析

- 參考試題反應理論，分析每道題目的關鍵詞難易度等級
- 利用試題關聯法則分析相同難易度的關鍵詞
- 考古題未曾出現之關鍵詞可由專家設定難易度等級



試題難易度預測

- 利用當年度命題與考古題之相似度分析結果預測難易度
- 可預測當年度命題難易度
- 題庫試題審題(難易度分配是否恰當)、入闈協助決定試題
- 精進題庫試題品質
- 共同科目之試題對應至不同等級考試之難易度分析



現有題庫之分析限制

- 由於每道考題所涵蓋之關鍵詞並不多，相同關鍵詞出現率亦不算高，端賴考古題與題庫之文字進行探勘實有不足
- 可結合網路文章，利用程式自動進行爬文，藉以建立更多的關鍵詞相關性分析
- 或利用教科書之專業術語，藉以建立更多的關鍵詞相關性分析



如何驗證正確率

- 試題相似度之分析結果可請專家進行驗證比對
- 題庫難易度分析之正確率可使用IRT測驗統計結果對照檢驗
- 建立每道題庫試題之核心關鍵詞結果可請專家進行驗證比對
- 彙整歷年題庫試題之核心概念結果可請專家進行驗證比對



結語

AI技術應用於國家考試

- 可進行題庫難易度分析與預測
- 分析題庫之難易度分佈，精進題庫試題品質
- 可挖掘新舊資訊差異與變更，進而篩檢並更新題庫電子試題，淘汰不合時宜的試題
- 分析試題核心概念，輔助專家決定試題審題是否偏離科目重點或是否與考古題相似
- 可檢視歷年題庫試題之核心概念
- 共同科目之試題對應至不同等級考試之難易度分析

