

等 別：四等考試  
類 科：圖書資訊管理  
科 目：資訊系統與資訊檢索概要  
考試時間：1小時30分

座號：\_\_\_\_\_

※注意：(一)禁止使用電子計算器。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

(三)本科目除專門名詞或數理公式外，應使用本國文字作答。

- 一、請以資料、用戶、服務三個面向，說明數位圖書館 (Digital Library) 與網頁搜尋引擎 (Web Search Engine) 之差異。(25分)
- 二、當系統沒有要求使用者登入時，如何取得使用者查詢紀錄；若系統有要求使用者登入時，如何得知其之前的查詢條件？(25分)
- 三、假設某機構有一萬筆文件資料，擬採購檢索系統，針對 A、B 兩套系統做成效評估的比較：針對查詢詞 T，已知有 5 筆相關文件，其中 A 系統回應有 10 筆資料，其第 1、3、9、10 筆為相關文件，而 B 系統回應有 6 筆資料，其第 2、3、5 筆為相關文件。就查詢詞 T 而言，計算並比較說明其檢索成效。(25分)
- 四、在有 N 篇文長近似的文件 (如 N 篇摘要，都約 300 詞) 的全文檢索系統中，索引詞彙 T 在每一篇文件都有其權重。最簡單的詞彙權重設定方式是布林 (Boolean) 邏輯方式，亦即詞彙 T 出現在文件 D，則其權重為 1，若沒出現在文件 D 則其權重為 0。另一種詞彙權重設定方式為  $TF \times IDF$ ，亦即：T 出現在 D 的次數 (或是正規化後之詞頻，TF)「乘以」T 在 N 篇文件中出現篇數的倒數 (或是正規化後之文件篇數倒數，IDF)。請說明並比較布林權重、TF、IDF 及  $TF \times IDF$  等四種權重方式，對相關文件排序的影響。(25分)