

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。

- 一、依據向量空間模式 (Vector-Space Model) 建構的資訊檢索系統，是將每一份文件表達為對應的文件向量，請說明如何建構文件的向量，解釋文件向量分量 (Component of Vector) 的意義，並討論如何產生每一個分量的數值。(25分)
- 二、許多資訊檢索系統在建構索引時，排除了停用詞 (Stop Words)，但是某些資訊檢索系統，並不排除停用詞。請先說明何謂停用詞，繼之討論排除停用詞與不排除停用詞的利弊得失。(25分)
- 三、在大數據 (Big Data，亦稱為巨量資料) 的風潮下，圖書館界亦關心研究資料的典藏、管理及使用，也就是近年備受關注的資料度用 (Data Curation) 議題。圖書館若是提供資料度用的服務，則必須建構資料度用系統，並提供檢索功能，請討論研究資料的檢索與文件的檢索，在功能開發上，有何不同的考量。(25分)
- 四、有一些資訊檢索系統提供同義處理、同音處理、近似處理或是詞幹處理 (Stemming) 後的檢索等，我們可以將前述的處理都視為是建構等價群集 (Equivalence Class)，例如相同意義的詞彙形成一個等價群集；發音相同的詞彙形成一個等價群集。請討論提供前述處理功能的檢索系統，對於檢索績效 (Performance) 的正面與負面的影響。(25分)