

搜尋引擎回覆款目品質探討： 以檢索失誤率為例

Performance Analysis of Search Engines Using Retrieval Error Ratio

吳政睿

Cheng-Juei Wu

輔仁大學圖書資訊系專任教授

Prof. of Department of Library & Information Science, Fu-Jen University

【摘要 Abstract】

本文以 2002 到 2004 年間 3 年的實驗數據，使用檢索失誤率為基準，衡量搜尋引擎回覆款目所提供之（目錄）資料品質。實驗分析顯示多種搜尋引擎平均的（總）檢索失誤率為 28.3%，其中會造成使用者點選不需要款目（因而浪費時間或金錢）的型 I 檢索失誤率為 16.5%；會造成使用者錯失點選需要款目（因而損失資訊）的型 II 檢索失誤率為 11.8%。若以題目類型來分析，生活類型题目的型 I 和型 II 檢索失誤率相差不大，但是學術類型题目的型 I 檢索失誤率幾乎是型 II 檢索失誤率的兩倍，這似乎暗示題目類型對檢索失誤的種類而言是一個重要的影響因素。最後以個別搜尋引擎來看，（總）檢索失誤率最優前三名為 MSN、PC Home、和 GAIS；型 I 檢索失誤率最優前三名為 PC Home、Openfind、和 MSN；型 II 檢索失誤率最優前三名為 MSN、WiseNut、和 Google，因此綜合三個檢索失誤率來看，以 MSN 和 PC Home 的表現最佳。

Based on the experimental data of 2002, 2003, and 2004, the performance of various search engines is analyzed and measured by the Retrieval Error Ratio. Firstly, the experimental results show that, on average, the Retrieval Error Ratio is 28.3%, which includes 16.5% of Type I Retrieval Error Ratio and 11.8% of Type II Retrieval Error Ratio, for the eleven search engines including Google, GAIS, and YAHOO. Secondly, the types I Retrieval Error Ratio is slightly higher than the Type II Retrieval Error Ratio for non-academic search terms. On the other hand, there are almost twice as many percentages of Types I Retrieval Error Ratio as that of Types II Retrieval Error Ratio for academic search terms. Lastly, MSN, PC Home, and GAIS are the three best search engines in terms of the Retrieval Error Ratio. For the Type I Retrieval Error Ratio, the three best search engines are PC Home, Openfind, and MSN, while MSN, WiseNut, and Google are the three best search engines in terms of the Types II Retrieval Error Ratio.

【關鍵詞 Keyword】

檢索失誤率；元資料；檢索；搜尋引擎

Retrieval Error Ratio; Metadata; Information Retrieval; Search Engine

壹、前言

1990年代在人類資訊處理上最顯著的里程碑，首推 World-Wide Web（全球資訊網，簡稱 WWW）的盛行，網際網路和 WWW 的迅速興起和結合，對資訊傳播的方式產生了重大的衝擊。WWW 盛行後，為因應檢索網頁內容的需要而有搜尋引擎的產生，搜尋引擎運作的方式，基本上是屬於全文檢索，主要是透過自動抓取程式在網際網路上抓取網頁，然後以自動拆字（或詞）作索引的方式來建立其資料庫，做為檢索的基礎，這種操作方式的特點是高運作效率和一網打盡，因此有高回收率與低精確率的特性。

元資料（Metadata）最常見的英文定義是“Data about Data”（註1），可直譯為描述資料的資料，主要是描述資料屬性的資訊，用來支持如指示儲存位置、資源尋找、文件紀錄、評價、過濾等的功能（註2）。廣義的描述性元資料包括有（搜尋引擎）全文索引、無欄位名詞集（如關鍵詞或主題）、基本欄位架構（如無修飾詞的都柏林核心集）、修飾詞欄位架構（如有修飾詞的都柏林核心集）、和複雜結構（如 MARC 和 TEI）等（註3）。

為了驗證元資料的實際效用，在「都柏林核心集對減低檢索失誤率的實務研討」一文中（註4），作者使用一個新的衡量標準——檢索失誤率（Retrieval Error Ratio，簡稱 RER）

（註5），用來評估檢索系統的效能，檢索失誤率是來比較和評估檢索系統所提供之（目錄）資訊做為使用者判斷基準的精確程度。

實驗結果顯示，七個不同的搜尋引擎在全部 140 筆回覆款目中，總共有 16 筆第 I 型失誤和 11 筆第 II 型失誤，共計 29 筆失誤，檢索失誤率（RER）是 20.7%（註6）。

隨後在「Google 網頁排序能力和回覆款目品質探討：以檢索失誤率為例」一文中（註7），針對 Google 所提供回覆款目資訊之檢索失誤率的研究發現，Google 的檢索失誤率在 19.5%-21.5%之間（註8），顯示 Google 檢索系統所提供回覆款目的品質，以協助使用者判斷的角度而言，與上述其他七個不同搜尋引擎的效能大致相同。

貳、檢索失誤率

以檢索過程而言，一般檢索系統（包括人工目錄卡片、圖書館自動化系統、和搜尋引擎等）所提供的目錄或回覆款目資料，其主要功能之一是協助使用者判斷資料是否為所需，做為取得原文或者忽略跳過的依據。如果取得原文的過程耗時費力，或者須要付費，這使得行動之前的判斷非常重要，反而是評估檢索系統效能的重要依據，能協助使用者做出正確判斷的檢索系統，其目錄或回覆款目資料的品質才是最佳的。

基於以上的認知，作者創造檢索失

誤率 (RER) 為衡量標準，來比較各種檢索系統其目錄或回覆款目資料的品質。檢索失誤率是計算失誤筆數佔總筆數的百分比，公式如下 (註 9)：

$$\text{檢索失誤率} = \frac{\text{失誤筆數}}{\text{總筆數}} \times 100\%$$

例如：總筆數為 20，失誤筆數為 4 時，檢索失誤率 (RER) 為 20%。

由於失誤型態的不同，檢索失誤率 (RER) 可分為以下兩種 (註 10)：

一、第 I 型失誤：

在閱讀檢索系統所提供的資訊時，使用者認為需要，但在調閱原文後，判定非為其所需要的資料。換言之，使用者浪費了時間或金錢。

二、第 II 型失誤：

在閱讀檢索系統提供的資訊時，使用者認為不是他或她需要的資料，但在事後調閱原文時，發現是需要的資料。換言之，使用者損失了資訊。

參、實驗設計

以選修作者所開設的研究所課程「元資料概論」的研究生為實驗者，實驗過程如下：研究生二人為一組，一人扮演讀者，一人扮演參考館員，由扮演讀者的研究生出一個題目，然後扮演參考館員的研究生自行選擇一個搜尋引擎來搜尋，參考館員從搜尋引擎所回覆的款目中挑選 20 筆 (從回覆款目的第一

筆挑選起，但是扣除純介紹性的網站首頁等不適合的款目)，將這 20 筆的回覆款目印出。然後扮演參考館員的研究生將印出的資料 (即搜尋引擎所提供的資訊)，拿給扮演讀者的研究生勾選，選出他或她認為需要的文件 (或款目) 有那些，並且記錄讀者勾選的結果，組成第一組數據 (註 11)。

然後扮演參考館員的研究生再將這 20 筆款目的原始文件一一下載取回後，印出拿給扮演讀者的研究生勾選，選出他或她認為需要的文件有那些，並且記錄讀者勾選的結果，組成第二組數據 (註 12)。

將上述兩組數據對比後，即可分別找出第 I 型和第 II 型失誤，加總後即為 (總) 檢索失誤率 (RER)。

肆、實驗結果整理與分析

由於作者所開設的研究所課程「元資料概論」是在每個學年度的上學期，因此實驗進行的時間約在每年 10-12 月期間，以下是最近三年 (西元 2002-2004 年) 來的實驗結果整理與分析。

首先，表一是實驗時間、參與人數、和題目類型的概況表格，為使不同搜尋引擎間的比較能有較公平的基準，同次參與實驗者所選的題目類型是相同的，大致上分為偏學術類型或生活類型二類。同時由於參與的人是研究所選修課同學，故每年參與人數會有變動。

表一：實驗人數和題目類型概況。

時間（西元）	參與人數	題目類型
2002 年底	5	學術
2003 年底	7	學術
2004 年底	12	生活

其次，表二、表三、和表四分別為 2002 到 2004 年的實驗結果整理（每一題目選 20 篇文件）。有關型 I 失誤和型 II 失誤的意義，請參見上面「檢索失誤率」一節中的介紹，此外，該節中也有檢索失誤率的計算公式。

表二：西元 2002 年的實驗結果整理。

題 目	搜尋引擎	型 I 失誤	型 II 失誤	失誤總計	檢索失誤率
TEI	Alta Vista	5	0	5	25%
TEI Header	Lycos	4	2	6	30%
EAD	Yahoo	4	1	5	25%
DOI	Alltheweb	4	3	7	35%
SICI	WiseNut	4	1	5	25%

註：Alltheweb 的網址為 <<http://www.alltheweb.com>>，WiseNut 的網址為 <<http://www.wisenut.com>>。

表三：西元 2003 年的實驗結果整理。

題 目	搜尋引擎	型 I 失誤	型 II 失誤	失誤總計	檢索失誤率
文件自動分類	Openfind	4	1	5	25%
兒童資訊素養	PC Home	2	2	4	20%
人本心理學	Yahoo! 奇摩	2	3	5	25%
血癌	GAIS	4	3	7	35%
李賀詩	Google	2	2	4	20%
資訊過濾	MSN	3	0	3	15%
兔唇治療	新浪網	4	4	8	40%

表四：西元 2004 年的實驗結果整理。

題 目	搜尋引擎	型 I 失誤	型 II 失誤	失誤總計	檢索失誤率
Herbal Tea	Alta Vista	5	6	11	55%
水族飼養	Google	0	2	2	10%
歌劇	MSN	3	0	3	15%
宗教音樂	Openfind	3	0	3	15%
生態旅遊	Openfind	3	7	10	50%
怪物+漫畫+浦澤直樹 +介紹	Openfind	3	1	4	20%
北台灣小吃	GAIS	2	3	5	25%
拉丁有氧	GAIS	3	0	3	15%
瑜珈	Google	5	3	8	40%
北台灣溫泉旅遊	Google	7	0	7	35%
線上遊戲	Openfind	0	10	10	50%
普普藝術	Yahoo! 奇摩	3	3	6	30%

用年度來看，表五是 2002-2004 年實驗結果的檢索失誤率比較表格，表六是 2002-2004 年實驗結果的統計比較表格。若以統計學的平均值、標準差、和全距等概念來做進一步的分析，各年度的檢索失誤率（或平均失誤筆數）相差不大，2002-2004 年分別為 28%、25.7%、和 30%。但是失誤筆數的標準差卻相差很大，2002-2004 年分別為

0.8、1.64、和 3.03；同時以檢索失誤率的變動範圍（即全距）來看，西元 2002 年實驗為 25%-35%，西元 2003 年實驗為 15%-40%，西元 2004 年實驗為 10%-55%，所以綜合來看，2004 年的檢索失誤率變動最大，至於變動較大的原因，是受參與人數多寡、題目類型、搜尋引擎、或綜合因素的影響，則有待進一步的實驗來釐清。

表五：2002-2004 年實驗結果的檢索失誤率比較表格。

時間（西元）	資料總筆數	型 I 檢索失誤率 （總筆數）	型 II 檢索失誤率 （總筆數）	檢索失誤率 （總筆數）
2002 年底	100	21% (21)	7% (7)	28% (28)
2003 年底	140	15% (21)	10.7% (15)	25.7% (36)
2004 年底	240	15.4% (37)	14.6% (35)	30% (72)
總 計	480	16.5% (79)	11.9% (57)	28.3% (136)

表六：2002-2004 年實驗結果的統計比較表格。

時間 (西元)	平均失誤 筆數	失誤筆數 標準差
2002 年底	5.6	0.8
2003 年底	5.14	1.64
2004 年底	6	3.03

註：平均失誤筆數除以 20 後換成百分比，即為檢索失誤率。

表七是學術類型和生活類型题目的統計比較表格，初步來看，兩者的檢索失誤率以學術類型題目略低，不過相差

表七：學術類型和生活類型题目的統計比較表格。

題目類型	資料總筆數	型 I 檢索失誤率 (總筆數)	型 II 檢索失誤率 (總筆數)	檢索失誤率 (總筆數)
學術	240	17.5% (42)	9.2% (22)	26.7% (64)
生活	240	15.4% (37)	14.6% (35)	30% (72)

表八是個別搜尋引擎在檢索失誤率上效能的比較，統計彙集的方式是將這三年的實驗數據以搜尋引擎為依據來彙總，例如搜尋引擎 Alta Vista 在 2002 和 2004 年分別被使用一次，各產生 5 和 11 筆的檢索失誤總筆數，因此彙總相加後，在表八上顯示有 16 筆的檢索失誤總筆數。

由於每個搜尋引擎被使用的次數不同，為求有較公平的比較基準，檢索失誤率的計算方式為：檢索失誤總筆數除以資料總筆數後的百分比。例如搜尋引擎 Alta Vista 在表八上顯示，檢索失誤總筆數和資料總筆數分別為 16 和 40，

不大；但是若以兩者失誤筆數的標準差來看，卻相差很大，學術類型和生活類型之失誤筆數標準差分別為 1.37 和 3.03，學術類型题目的標準差低很多。另外一個值得觀察的差異是兩者型 I 和型 II 檢索失誤率的對比，生活類型题目的型 I 和型 II 檢索失誤率相差不大，但是學術類型题目的型 I 檢索失誤率幾乎是型 II 檢索失誤率的兩倍，兩者相差甚巨，這似乎暗示題目類型對檢索失誤的種類而言是一個重要的影響因素。

因此 $16/40 = 40\%$ ，其餘型 I 和型 II 檢索失誤率的計算方式亦同，都是以個別搜尋引擎之資料總筆數為分母計算求得。

表九、十、和十一分別為搜尋引擎之(總)檢索失誤率、型 I 檢索失誤率、和型 II 檢索失誤率效能的排行，可以看出(總)檢索失誤率最優前三名為 MSN、PC Home、和 GAIS；型 I 檢索失誤率最優前三名為 PC Home、Openfind、和 MSN；型 II 檢索失誤率最優前三名為 MSN、WiseNut、和 Google。因此綜合三個檢索失誤率來看，以 MSN 和 PC Home 的表現最佳。

表八：搜尋引擎效能的統計彙總表格。

搜尋引擎	資料總筆數	型 I 檢索失誤率 (總筆數)	型 II 檢索失誤率 (總筆數)	檢索失誤率 (總筆數)
Alta Vista	40	25% (10)	15% (6)	40% (16)
Alltheweb	20	20% (4)	15% (3)	35% (7)
GAIS	60	15% (9)	10% (6)	25% (15)
Google	80	17.5% (14)	8.8% (7)	26.3% (21)
Lycos	20	20% (4)	10% (2)	30% (6)
MSN	40	15% (6)	0% (0)	15% (6)
Openfind	100	13% (13)	19% (19)	32% (32)
PC Home	20	10% (2)	10% (2)	20% (4)
WiseNut	20	20% (4)	5% (1)	25% (5)
Yahoo! 奇摩	60	15% (9)	11.7% (7)	26.7% (16)
新浪網	20	20% (4)	20% (4)	40% (8)

註：檢索失誤率為失誤總筆數除以資料總筆數後換成百分比。

表九：搜尋引擎檢索失誤率的比較表格。

搜尋引擎	檢索失誤率
MSN	15%
PC Home	20%
GAIS	25%
WiseNut	25%
Google	26.3%
Yahoo! 奇摩	26.7%
Lycos	30%
Openfind	32%
Alltheweb	35%
Alta Vista	40%
新浪網	40%

表十：搜尋引擎型 I 檢索失誤率的比較表格。

搜尋引擎	型 I 檢索失誤率
PC Home	10%
Openfind	13%
MSN	15%
GAIS	15%
Yahoo! 奇摩	15%
Google	17.5%
WiseNut	20%
Lycos	20%
Alltheweb	20%
新浪網	20%
Alta Vista	25%

表十一：搜尋引擎型 II 檢索失誤率的比較表格。

搜尋引擎	型 II 檢索失誤率
MSN	0%
WiseNut	5%
Google	8.8%
PC Home	10%
GAIS	10%
Lycos	10%
Yahoo!奇摩	11.7%
Alltheweb	15%
Alta Vista	15%
Openfind	19%
新浪網	20%

伍、結語

本文延續在「都柏林核心集對減低檢索失誤率的實務研討」和「Google 網頁排序能力和回覆款目品質探討：以檢索失誤率為例」二文中，使用檢索失誤率來衡量搜尋引擎回覆款目所提供之（目錄）資料品質的策略，以 2002 到 2004 年間選修作者所開設的研究所課程「元資料概論」的研究生為實驗者，所得的 3 年實驗數據為基礎，來進一步分析搜尋引擎在此方面的效能，並與上述二文中的實驗結果來相互比較。

實驗分析顯示多種搜尋引擎平均的（總）檢索失誤率為 28.3%，其中 2002、2003、和 2004 年分別為 28%、25.7%、和 30%。對照來看，在「都柏林核心集對減低檢索失誤率的實務研討」一文中的（總）檢索失誤率為 20.7%，基本上搜尋引擎的（總）檢索失誤率一般在 20%-30%之間。

細部分析顯示，2002 到 2004 的實驗中，會造成使用者點選不需要款目（因而浪費時間或金錢）的型 I 檢索失誤率為 16.5%；會造成使用者錯失點選需要款目（因而損失資訊）的型 II 檢索失誤率為 11.8%。在「都柏林核心集對減低檢索失誤率的實務研討」一文中的型 I 和型 II 檢索失誤率分別為 12.85% 和 7.85%。對照分析來看，型 I 和型 II 檢索失誤率在本文中的實驗數據略高。

若專以 Google 來看，本文中 Google 的（總）檢索失誤率、型 I 檢索失誤率、和型 II 檢索失誤率分別為 26.3%、17.5%、8.8%；在「Google 網頁排序能力和回覆款目品質探討：以檢索失誤率為例」一文中 Google 的（總）檢索失誤率、型 I 檢索失誤率、和型 II 檢索失誤率分別為 19.5%-21.5%、16.5%-17%、3%-4.5%。相互比較發現，型 I 檢索失誤率約略相同，但是型 II 檢索失誤率在本文中的實驗數據較高（因而使檢索失誤率也較高）。

若以題目類型來分析，2002 到 2004 的實驗中，學術類型題目的（總）檢索失誤率、型 I 檢索失誤率、和型 II 檢索失誤率分別為 26.7%、17.5%、和 9.2%；生活類型題目的（總）檢索失誤率、型 I 檢索失誤率、和型 II 檢索失誤率分別為 30%、15.4%、和 14.6%。相互對比可知，生活類型題目的型 I 和型 II 檢索失誤率相差不大，但是學術類型題目的型 I 檢索失誤率幾乎是型 II 檢索失誤率的兩倍，這似乎暗示題目類型對檢索失誤的

種類而言是一個重要的影響因素。此外，雖然學術與生活二種類型題目的（總）檢索失誤率相差不多，但是學術類型和生活類型題目之失誤筆數標準差分別為 1.37 和 3.03，顯示學術類型題目的變動性較低。

對比以前的研究結果，在「都柏林核心集對減低檢索失誤率的實務研討」一文中的題目為學術類型，其（總）檢索失誤率、型 I 檢索失誤率、和型 II 檢索失誤率分別為 20.7%、12.85%、和 7.85%。對照上述的學術類型題目數據，雖然上述的實驗數據略高，但基本上結果是相互吻合的。

最後以個別搜尋引擎來看，（總）檢索失誤率最優前三名為 MSN、PC Home、和 GAIS；型 I 檢索失誤率最優前三名為 PC Home、Openfind、和 MSN；型 II 檢索失誤率最優前三名為 MSN、WiseNut、和 Google，因此綜合三個檢索失誤率來看，以 MSN 和 PC Home 的表現最佳。

附 註

- 註 1 E. P. Shelley and B. D. Johnson, "Metadata: Concepts and Models," in *Proceedings of the Third National Conference on the Management of Geoscience Information and Data* (Adelaide, Australia: Australian Mineral Foundation, 1995), pp.1-5.
- 註 2 L. Dempsey and R. Heery, "An Overview of Resource Description Issues," (March 1997), <http://www.ukoln.ac.uk/metadata/DESIRE/overview/rev_01.htm>.
- 註 3 S. Weibel, R. Iannella, and W. Cathro, "The 4th Dublin Core Metadata Workshop Report," June 1997, <<http://www.dlib.org/june97/metadata/06weibel.html>>, p.3.
- 註 4 吳政叡，「都柏林核心集對減低檢索失誤率的實務研討」圖書館學與資訊科學 24 卷 1 期（1998 年 10 月），頁 50-64。
- 註 5 吳政叡，機讀編目格式在都柏林核心集的應用探討（台北市：學生書局，1998 年 12 月），頁 210。
- 註 6 同註 4，頁 62。
- 註 7 蔡佳霖，吳政叡，「Google 網頁排序能力和回覆款目品質探討：以檢索失誤率為例」書藝 38 期（2002 年 5 月），頁 40-54。
- 註 8 同註 7，頁 47-48。
- 註 9 同註 5。
- 註 10 吳政叡，「資訊的檢索失誤率探討」，中國圖書館學會會訊 109 期（1998 年 6 月），頁 26。
- 註 11 同註 5，頁 212-213。
- 註 12 在原本的實驗設計中，從搜尋引擎回覆款目所勾選的結果，不但是第一組數據，也成為對照組；此外尚有由 Dublin Core 所勾選結果組成的第二組數據（實驗組），和原始文件勾選結果組成的第三組數據。這裡為因應本文的題目，將實驗設計過程略為改寫，省略 Dublin Core 的數據，以免造成讀者的混淆。