# 以社會性標籤為基礎的擴充搜尋技術支援影音分享網站中之影片檢索

李彥賢
國立嘉義大學資訊管理學系

楊錦生*
元智大學資訊管理學系

廖國堯
天新資訊

## 摘要

WEB 2.0 的概念興起促使網路使用者從資訊接收者的角色轉變成資訊生產者，並透過適當的網路平台與其他網路使用者進行資訊分享與互動。近年來，隨著網路與資訊科技的快速發展，已使網際網路中分享的資訊媒體從過去單純的文字內容逐步演進到以影音多媒體為主流，並發展出許多影音分享網站。然而，為方便使用者在龐大的分享影音資料中進行搜尋，此類網站往往會提供影片搜尋機制。不若以文字描述為主要內容的傳統網站，影音分享網站每則影音檔案都只有少量的描繪資料可供進行關鍵字比對，容易在比對時造成字串錯配（Word Mismatch）的問題，進而影響影音檢索的效能。本研究提出以社會性標籤為基礎的擴充搜尋技術（STBQE），先藉由局部分析方法取得初步的影片搜尋結果，再試圖利用使用者分享影片時，針對影片內容所制定的描繪標籤，來進行搜尋字串擴充以及後續搜尋結果排序。本研究最後利用實證評估方式來比較現有影音分享網站的搜尋機制與本研究提出 STBQE 的差異，而結果發現本研究提出擴充搜尋技術能在特定情境之下改善現有方法的搜尋效能。

**關鍵詞**：搜尋字串擴充、影音檢索、社會性標籤、情境感知、搜尋引擎

* 本文通訊作者。電子郵件信箱：csyang@saturn.yzu.edu.tw
2011/01/11 投稿；2011/06/10 修訂；2011/07/18 接受

# A Social-Tag-Based Query Expansion Approach for Supporting Video Retrieval in Video Sharing Websites

Yen-Hsien Lee
Department of Management Information Systems, National Chiayi University

Chin-Sheng Yang*
Department of Information Management, Yuan Ze University

Guo-Yauo Liau
Future Intelligence Technology Inc.

## Abstract

The advance of Internet and information technologies has turned the information sharing on the Web from the form of textual content into that of multimedia and fostered the development of video sharing websites. To ease the access of the large number of online videos, the sharing websites usually provide web surfers the mechanism for searching their targeted videos. However, traditional searching approaches that focus on the analysis of textual contents may not effectively support the search for the online videos, which generally contain a few annotations given by the providers. As a result, keyword search has been one of the familiar approaches to the video retrieval, although it may suffer from the word mismatch problem. In this study, we intend to improve the effectiveness on retrieving the online video by proposing the Social-Tag-Based Query Expansion (STBQE) approach. To identify from the initial search results the context words relevant to the user's query, our approach can use which to expand the search and well rank the search results. Our evaluation results also suggest that the effectiveness of our proposed STBQE approach is comparable to the performance benchmark, YouTube.com, under some evaluation scenario.

**Keywords:** query expansion, video retrieval, social tagging, context aware, search engine.

* Corresponding author. Email: csyang@saturn.yzu.edu.tw
2011/01/11 received; 2011/06/10 revised; 2011/07/18 accepted

# 壹、緒論

管理大師 Peter F. Drucker（2001）認為真正推動社會進步的「Information Technology」為「Information」而不是「Technology」，僅著重技術層面的開發而忽略資訊的本質，將使資訊科技成為一具空的軀殼，而不能使社會增值（Drucker 2001）。以高互動性與個人化為核心的 WEB 2.0 概念被提出後，網路服務提供者亦逐漸接受網路使用者從過去的資訊接收者轉變為資訊生產者，藉由提供適當的網路平台讓使用者充份表達自己的意見，並與其他的使用者進行互動。著名的 WEB 2.0 技術提供者，像是 Amazon、eBay、Google、MySpace、Yahoo、YouTube 與 Wikipedia 等網路服務供應者，其提供網路使用者不需具備特殊資訊技能，便能簡單快速進行資訊分享的網路平台與工具。

近年來，隨著網路與資訊科技軟、硬體的快速發展，已使得網路中資訊分享的媒體從單純的文字內容逐步演進到影音多媒體。相較於文字內容，多媒體擁有較高的資訊豐富度，且能替代繁瑣或難以清楚表達的文字敘述，並減少文字使用量，因此較能引起網路使用者的注意。根據知名網路研究機構 comScore[1]在 2007 年 11 月針對美國網路使用者進行調查的結果顯示，超過 75%的網路使用者曾經觀看過線上影片，且每位使用者平均花費 3.25 個小時，相較於前年 1 月的調查結果成長 29%。更進一步的分析指出，在不計重複造訪人次的情況下，美國網路使用者總計在該月觀看約 9.5 億部線上影片，而其中 2.9 億的影音來自 YouTube.com，Hitwise[2]調查報告亦顯示，線上影片分享的市場正在持續大幅度地成長。

一般而言，影音分享網站多半會提供搜尋機制，方便使用者取得想要的影音檔案。然而不同於傳統以文字描述為主要內容的網站，具備眾多的文字可供搜尋引擎進行搜尋文字的比對，影音分享網站往往只能利用影音檔案的描繪資料（Meta Data），例如標題、標籤與相關描述等少量的文字描述進行檢索比對。儘管如此，由於線上影音資料數量的快速成長，以及使用者對於個人化服務的需求，影音分享網站多半提供使用者自行制定影音描繪資料的功能，由使用者依據自己的喜好來進行影音資料類別管理，讓使用者得以更快速檢索分享的影音資料。舉例來說，YouTube.com 可讓影片發佈者給定影片標題（Title）、相關類別（Type）、影片標籤（Tag）及相關內容描述（Description）等資料，除相關類別中的選項係由服務平台先行定義外，其他資料皆可由使用者依喜好自行給定，以方便日後對於個人分享影片的存取與管理。上述由使用者定義影片類別標籤的過

1 http://www.comscore.com/press/release.asp?press=2002

2 http://www.hitwise.com/news/us200804.html#body_feature

程稱為社會性標籤制定（Social Tagging）或協同式標籤制定（Collaborative Tagging），其允許使用者依照自己的喜好給予文件或檔案相關的關鍵字或標籤，讓使用者參與文件或檔案的分類過程，而非被動接受網站所提供的分類方式（Chi & Mytkowicz 2007; Passant & Laublet 2008; Rowley 1995）。

社會性標籤制定機制可結合眾人之志將線上資源進行分類標定，除直接反應使用者的喜好外，並間接反應社會潮流，更可讓使用者依據熟悉的分類方式進行資訊檢索。儘管如此，根據 Furnas et al.（1987）研究指出，不同兩個人使用相同文字描述同一個概念的機率小於 20%。一般來說，不同使用者對於相同字詞的意義往往會有不同的認定，且往往會使用許多創新且相異的詞彙來表達相同意義的概念（Golder & Huberman 2006; MacGregor & McCulloch 2006）。此外，影音分享網站提供扁平式而非階層式的社會性標籤制定，不同經驗背景的使用者在為相同影片制定標籤時，可能會擷取不同階層的概念，而形成概念階層的特異性（Specificity）（Golder & Huberman 2006; Zauder et al. 2007）。舉例來說，使用者 A 與 B 皆擁有跟「哈士奇」相關的影音，假設使用者 A 對於狗具有高度興趣及擁有豐富的相關知識，可能會採用比較特異的（Specific）標籤「哈士奇」來標註該影片；相反地，使用者 B 對於狗的喜好程度及擁有的相關知識皆不及使用者 A，則使用者 B 就可能會採用較一般化（General）的標籤「狗」來標註該影片。因此，在仍是以關鍵字比對為主要搜尋方式的影音分享網站中，若使用者在進行影片搜尋時未能給予適當的搜尋字串，則容易面臨字串錯配（Word Mismatch）的問題，影響搜尋引擎的檢索效能（Wei et al. 2007; Carpineto et al. 2002; Xu & Croft 2000）。此外，過去的研究亦指出，使用者平均只使用兩個字彙來進行資訊搜尋，過少的搜尋字串往往也會影響搜尋引擎找出合適資料的機會（Croft et al. 1995）。

為了解決影音分享網站中以使用者自訂社會性標籤作為搜尋基礎，所容易引發的字串錯配的問題，過去的研究多著重在發展自動或半自動化的技術，來擴展影音的標籤（Ballan et al. 2010; Krestel et al. 2009; Yang et al. 2011），藉由增加影音標籤的數量，以減低字串錯配的機率。然而，文獻中卻鮮少從另一個角度－增加搜尋字串的數量或稱為搜尋字串擴充（Query Expansion），來探討影音搜尋可能發生的字串錯配問題。因此，本研究將針對搜尋字串擴充技術在影音搜尋環境中的可行性與效能進行研究。在文件檢索的環境中，字串錯配一樣是影響檢索效能的一個主要問題，因此，過去有許多研究嚐試透過增加搜尋字串字數，以提高文件和搜尋字串吻合的機率，減少字串錯配所帶來的影響（Xu & Croft 1996）。研究進一步指出，即使列出一堆字串供使用者選擇用以擴充搜尋字串，使用者也很難選到正確的擴充字串（Ruthven 2003; Hoeber et al. 2005），故需透過自動化的搜尋字串擴充方式（Automatic Query Expansion），協助使用者進行搜尋字串的擴充，

提升檢索的效能（Wei et al. 2007）。考量搜尋字串擴充在傳統資訊檢索領域已被廣泛地研究，且提出了許多搜尋字串擴充的演算法，故本研究將著重在於分析目前主要的搜尋字串擴充演算法的優缺點，從中找出一個適用於影音搜尋應用的架構，並考量影音分享網站的特性，適當修改選定的演算法，再以實際驗證方式評估所提出的方法在現行的影音搜尋平台中的效能表現。

根據 Au Yeung、Ginnins 與 Shadbolt（2007）的研究指出，在某一時間點或某一影音檔內，使用者對於特定標籤幾乎不會同時用來形容不同的概念，若其伴隨著不同的標籤出現在不同的影音檔案中，則該特定標籤可能代表不同的含意，而與特定標籤共同用來描繪影片的標籤，則稱為此特定標籤的情境文字（Context Word）。據此，情境文字應可以協助評估影音所處的情境，並加以提升影音檢索的效能。因此，本研究將提出一個以社會性標籤為基礎的擴充搜尋技術（Social-Tag-Based Query Expansion，STBQE），利用情境文字的分析，來支援影音分享網站中的影片資料檢索。此外，考量影音分享網站中資料經常變動的特性，本研究採用局部回饋（Local Feedback）（Buckley et al. 1996）的搜尋字串擴充方式，將初步搜尋結果做為擴充搜尋的基礎，並利用情境文字的特性建構擴充辭典，用於擴充使用者的搜尋字串並做為搜尋結果的排序準則。

本研究後續內容編排如下，第二章將針對與本研究相關的文獻進行探討，第三章會詳述本研究提出的社會性標籤為基礎的擴充搜尋技術，第四章說明實證評估設計與重要結果，最後於第五章描述本研究結論與未來研究方向。

## 貳、文獻探討

本章的重點在探討本文的相關研究成果，因為本文提出的社會性標籤為基礎的擴充搜尋技術（Social-Tag-Based Query Expansion; STBQE）為一個以文字為基礎的資訊檢索系統，並結合了搜尋字串擴充的概念，因此本章分別就資訊檢索與搜尋字串擴充此兩大研究議題進行探討。除文字為基礎的檢索系統外，內涵式影像檢索系統（Content-based Image Retrieval）在影音或多媒體檢索的研究領域中是另一個主流的研究方向。內涵式影像檢索系統主要利用影像的低階（Low-level）、非文字（Non-Textual） 內容，例如，顏色、紋理、形狀等，來進行影像的檢索或搜尋字串擴充（Haubold et al. 2006; Rahman & Bhattacharya 2009），因本文的方法是以關鍵字為基礎的檢索方法，考量主題的相關程度與文章的篇幅限制，故未於相關文獻探討中介紹內涵式影像檢索系統及其相關搜尋字串擴充方法。有關內涵式檢索系統的相關研究，有興趣的讀者可參考 Lew 等人之研究成果（Lew et al 2006）；而內涵式影像檢索系統相關的搜尋字串擴充方法，則可參考下列文獻（Hong et al. 1998; Rahman & Bhattacharya 2009; Zhai, Liu & Shah

2006）。

## 一、資訊檢索（Information Retrieval）

數量龐大且雜亂的資訊，若未經過系統化的規劃、分類、及索引，將導致使用者無法在適當時間內取得其所需的資訊或知識（van Rijsbergen 1979）。儘管使用者仍可藉由逐一檢視的方式來搜尋所需的資訊或知識，以滿足資訊的需求，然而在面對快速成長的文件數量下，往往無法以上述人工方式檢索資料。也因此，如何快速且正確的取得使用者所需的資訊，即為資訊檢索所面臨的主要任務。從過去相關文獻來看，在建構資訊檢索系統時所採用的模型架構主要可區分成三大類，分別是布林模型（Boolean Model）、向量模型（Vector Model）及機率模型（Probabilistic Model）（Baeza-Yates & Ribeiro-Neto 1999; van Rijsbergen 1979），以下分別說明此三種模型。

### （一）布林模型

布林模型是最早被提出來的資訊檢索方法，利用布林運算子的各種組合，取得使用者所需要的資訊。使用者下達的搜尋字串，在與資料庫中的文件索引字詞，符合布林運算時才會擷取該文件，反之則予以略過（van Rijsbergen 1979）。舉例來說，假設 $Q_1$、$Q_2$、$Q_3$ 為搜尋字串，並以公式 $Q = ((Q_1\ AND\ Q_2)\ OR\ Q_3)$ 做為本次查詢基準，則檢索系統將會回傳同時具有 $Q_1$ 及 $Q_2$，或者具有 $Q_3$ 的相關文件給使用者。此種模型的主要缺點為無法對於檢索的文件進行適當排序。雖然可以經由日期或者文件某些特性進行排序，但是卻無法真正滿足使用者對於檢索結果進行相關性排序的需求（Harman 1993; Singhal 2001），也因而有向量模型及機率模型等模型架構被提出。

### （二）向量模型

向量模型以詞彙（Terms）作為文字區段的向量維度，並且將文字區段表示成一個向量（Salton et al. 1975）。一般而言，文字區段會事先排除無意義的字（Stopwords）形成一詞彙組合，詞彙組合中的每個字詞透過數學模式，計算其在此段文字中的重要程度。通常在衡量字詞重要性時主要會考量三個重要因子包括字詞頻率（Term Frenquency，TF）、文件頻率（Document Frenquency，DF）、及文件長度（Document Length）（Singhal 2001）。字詞頻率用來衡量詞彙在文字區段或文件內出現的次數，越多次代表此詞彙在文字區段或文件中具有較高的代表性。文件頻率主要在計算詞彙出現在多少文件之中，若出現在越多文字區段或文件之中，表示該詞彙為較常使用的字詞，例如：你、我、他等，文件頻率恰與字詞頻率相反，文件頻率越高的字詞其重要性越低。因此，以文件頻率為基礎的加

權方式，通常都會採用倒轉的文件頻率（Inverse Document Frequency，IDF），衡量字詞的權重（Jones 1972）。最後，當多個文字區段或文件集合中所具有的文字數量不相同時，則擁有越多文字的文字區段或者文件，通常會擁有較高的機率包含較多的詞彙及較高的詞彙重複率，而文件長度被用以進行詞彙加權的正規化（Singhal 2001）。

目前字詞重要性的衡量方式，大都以上述三項的因子加以變化而成，常見的方式包括有詞彙頻率與倒轉的文件頻率（Term Frequency×Inverse Document Frequency，TF×IDF）（Salton et al. 1983）、資訊獲利（Information Gain，IG）（Quinlan 1986; Mitchell 1997）、互訊息（Mutual Information，MI）（Fano 1961; Church & Hanks 1989）、卡方統計量（chi-square）（Schutze et al. 1995）等方式。計算所有字詞重要性後，通常會以前 $k$ 個重要程度較高的字詞做為文字區段或文件的代表字詞，並用來將文字區段或文件表示成向量，以便後續進行與搜尋字串的相似度計算。在進行文字區段或文件表達時，每個代表字詞都會被視為一個向量維度以重新表達此文件，如下列公式所示：

$$\vec{D} = (W_1, W_2, ..., W_k) \tag{1}$$

其中，$\vec{D}$代表文件的向量；$W_j$ 代表第 $j$ 個字詞在文件 $D$ 中的重要程度（權重）。在將搜尋字串和文件轉換成向量後，便可針對兩個向量進行相似度的計算。向量之間的夾角常被用來衡量兩個向量之間的相似程度，若平行則兩個向量相似度為1；若垂直則相似度為0（Singhal 2001）。其計算方式如下列公式所示：

$$Sim(\vec{Q}, \vec{D}) = \frac{\vec{Q} \cdot \vec{D}}{|\vec{Q}| \times |\vec{D}|} \tag{2}$$

（三）機率模型

根據機率排序定律（Probability Ranking Principle），文件群集應該依照文件和搜尋字串的相似機率由高往低排列，以取得較佳的資訊檢索效能（Roberson 1977）。但搜尋字串和文件的機率卻很難正確得知，因此機率模型主要以估計方式來評估文件和搜尋字串相似的機率。機率檢索一開始是由 Maron 與 Kuhns（1960）提出，雖陸續有研究提出不同的機率模型，但大部份的機率模型仍然以 Bayes 定理為基礎，推估搜尋字串和文件的相似機率（Baeza-Yates & Ribeiro-Neto 1999）。文件和搜尋字串相關和不相關的機率分別可以用$P(R|D)$ 、$P(\overline{R}|D)$表示，而文件排序則可以下列公式做為排序的基準：

$$\log\left(\frac{P(R|D)}{P(\overline{R}|D)}\right) \tag{3}$$

經由 Bayes 定理轉換可轉換成下列公式：

$$\log\left(\frac{P(R)P(D|R)}{P(\overline{R})P(\mathrm{D}|\overline{R})}\right) \tag{4}$$

因為事前機率 $P(R)$及$P(\overline{R})$，跟文件是否和搜尋字串相關互為獨立，對於排序結果並無影響，因此可將公式進一步簡化如下列公式：

$$\log\left(\frac{P(D|R)}{P(\mathrm{D}|\overline{R})}\right) \tag{5}$$

不同的機率模型對於機率預估的方式，大部份則是由簡化公式引申變化而成，而資訊檢索的效能差異，則會受到不同機率估計方式的影響。

## 二、搜尋字串擴充（Query Expansion）

搜尋字串擴充（Query Expansion）主要是利用相關文字來擴充增加使用者的搜尋字串，以解決資訊檢索時因搜尋字串過少而導致無法正確比對的問題。搜尋字串擴充的概念最早由學者 Jones（1971）提出，其透過數學模式的計算，以文字共同出現頻率為基礎將相關文字進行分群，做為擴充字串的基礎。文獻中已有許多搜尋字串擴充的方法，根據用於分析擴充字串的資訊來源之差異，主要可以分成全域分析（Global Analysis）或局部分析（Local Analysis）兩種分析方法（Attar & Fraenkel 1977; Carpineto et al. 2002; Cui et al. 2003; Dinh & Tamine 2011; Wei et al 2007; Xu & Croft 1996; 2000），以下針對文獻中一些代表性的搜尋字串擴充研究進行討論。

基於「高度相關的字串經常共同出現在相同的文件內」此一關聯假設（Association Hypothesis），全域分析在所有的文件群集中，計算字串出現的次數以及字與字之間的關係，建構一個統計式辭典，用以進行搜尋字串擴充（Jing & Croft 1994）。根據字串共同出現的情形，Jones（1971）提出將字串分成數個群集，之後找出搜尋字串涵蓋的字串群集來進行擴充。Qiu 與 Frei（1993）提出概念式（Concept-based）搜尋字串擴充技術，其特殊處在於挑選擴充字串時，擴充字串需與整個搜尋字串而非其中的單一組成字串進行相似度評估。Gauch、Wang 與 Rachakonda（1999）認為高度相關的字串有可能是互補字，並不一定會共同出現，但卻會有高度相似的情境字串（Contextual or Surrounding Terms），因此建議

估算情境字串的相似度來建構統計式辭典。考量字串可能不是獨立的況狀，Deerwester 等（1990）認為文件群集涵蓋的重要概念應該由重要的潛在結構（Latent Structure）而非所有的字串來表示，因此使用潛在語意索引（Latent Semantic Indexing）技術來找出文件群集中的重要潛在語意，之後將所有的文件與搜尋字串都轉換到 LSI 潛在語意結構中來進行檢索。上述搜尋字串擴充技術採取不同的策略來估算字串間的關係，但有一個共同點：使用文件群集中所有的文件來進行分析，因此文件群集中任兩個字串只有一個相似度估算值。然而，文件群集可能涵蓋許多的主題，兩個字串的相關程度在不同主題中可能差異很大，因此全域分析在文件群集主題太廣時，可能會有檢索效能不彰的問題（Gauch et al. 1999; Liu & Chu 2005）。例如，「哈士奇」與「臘腸」在討論狗的文章中可能是高度相關的字串，但在討論餐廳的文章中則沒有顯著關聯。

針對全域分析的缺點，局部分析不是使用整個文件群集，而是只針對原始搜尋字串所取得的部份文件進行字詞萃取，以找出具有代表性的字詞來進行字串擴充（Xu & Croft 1996）。局部分析可再細分成相關性回饋（Relevance Feedback）及局部回饋（Local Feedback）兩種方式。相關性回饋（Rocchio 1971; Salton & Buckley 1990）是一種利用使用者對於初次搜尋結果逐一進行相關性判斷，將結果區分成正、負兩種範例後，回傳給搜尋系統以修正進行再次搜尋的字串。在假設使用者可以準確判斷搜尋結果和搜尋字串有無相關性的前提下，此種方式已被證實有效（Harman 1992）。相關性回饋雖然可以提高搜尋引擎的效能，但回饋資料必須要由使用者提供，在非強迫性之下使用者通常意願較不高，故此種方法較少被採用。局部回饋（Local Feedback）（Attar & Fraenkel 1977; Khan & Khor 2004; Xu & Croft 1996; 2000）是因應使用者回饋意願不高的問題而提出，此方法假設搜尋引擎傳回的前數筆文件都是正確的，在不經過使用者判斷的情況下，直接將這些假設是正確的資料用於擴充搜尋字串。Attar 與 Fraenkel（1977）將搜尋引擎傳回的前數筆文件中的字串進行分群，再用於擴充搜尋字串。Xu 與 Croft（1996）則是找出搜尋引擎傳回的前數筆文件中出現頻率最高的字串來進行擴充。Khan 與 Khor（2004）捨棄直接將擴充字串加到原始搜尋字串中的做法，直接將每個搜尋字串視為一個額外的搜尋字串，分別進行檢索。採用與 Qiu 與 Frei（1993）相似的概念，Xu 與 Croft（2000）提出局部情境分析（Local Context Analysis）技術，其特殊處為僅分析搜尋引擎傳回前數筆文件的部分資訊（Passages）來進行搜尋字串擴充，以避免文件過長夾帶過多不相關資訊，影響搜尋字串擴充的效能。局部回饋的效能取決於搜尋引擎回傳的前數筆資料的正確性，若正確性不高，則產生的擴充字詞可能會與原始的搜尋字串相關性不高，導致字串擴充失敗（Cui et al. 2003; Huang et al. 2003; Xu & Croft 1996; 2000）。此外，搜尋字串的好壞也嚴重影響局部分析的效能（Losada 2010）。局部分析方法基本上需要花費大

量的運算時間於關聯程度的計算，因此對於即時性要求高的互動式系統（Interactive System）可能會造成運作上的困難（Gauch et al. 1999）。

不論是使用全域分析或局部分析的方式進行字串擴充，都會面臨詞意混淆，亦即同字異義與異字同義的問題，因此有研究提出以辭典為基礎（Thesaurus-Based）的方式，來解決詞意混淆問題，以改善字串擴充效能（Gong et al. 2005; Gong et al. 2010）。比較常見的辭典如 WordNet（Miller et al. 1990），其類似標準的辭典，包含字的定義及字與字之間的關係，而差異之處在於標準辭典是使用字母順序排列，WordNet 則是以概念當做排序依據（Gong et al. 2005）。此外，WordNet 亦使用階層式的概念架構，分成上位詞（Hypernym）、同義詞（Synonym）及下位詞（Hyponym）。Synset 則是 WordNet 最基本的單位，一個 Synset 代表的是一個概念；而上位詞指的是在階層概念中位於 Synset 上層的概念；下位詞則指的是位於 Synset 下層的概念。舉例來說，WordNet 3.0 對「Computer」所定義的上位詞有｛Machine, Device, Entity｝，下位詞則是｛Analog Computer, Digital Computer, Home Computer｝。儘管以事先建立的標準辭典來進行字串擴充可以解決部份詞意混淆的問題，然而當使用者可以依據自己的習慣及喜好來給予相同影片不同的描述文字，且其使用的字詞往往會隨時間經過而不斷創新時，則無法隨時更新的辭典（Thesaurus），就顯得不適用於快速變動的 WEB 2.0 環境中來進行搜尋字串擴充。

考量全域分析、局部分析與辭典為基礎等搜尋字串擴充方法的優缺點，本文認為使用局部分析來進行影音搜尋環境的搜尋字串擴充較為適當，原因在於：(1)影音分享網站內的影音資料通常涵蓋相當多的主題，因此使用全域分析無法適當處理兩個字串的相關程度在不同主題中可能差異很大的問題；(2)影音分享環境中使用者用於標註影音的標籤可能是經常變動的，例如，「小三」一詞是最近數月使用者高度使用的一個標籤，當採用辭典為基礎的擴充方法時，使用的辭典可能完全不涵蓋小三這個詞彙，若是使用全域分析，標註「小三」一詞的影音檔在整個影音資料庫中可能只占一部分，因此「小三」一詞在建構出來的全域式統計辭典可能是一個不太顯著的字串。反觀，使用局部分析方法可避免上述的兩點問題，因統計辭典建構的基礎是原始搜尋字串所取得的部份文件。此外，相關性回饋需要使用者的即時回饋，實用性較低，因此本文採用局部回饋的方式來發展所提之以社會性標籤為基礎的擴充搜尋技術（Social-Tag-Based Query Expansion，STBQE）。

## 參、社會性標籤為基礎之擴充搜尋技術

在影音分享網站的使用環境下，使用者可依喜好給予自己發佈的影片進行字詞標註，用以描繪該影片內容或對其進行個人化分類。一般來說，使用者通常會給予一個以上字詞（標籤）來共同描繪影片所呈現的內容或概念。因此，某些標籤集合往往會共同出現以表示某個特定內容或概念，而經常伴隨著某標籤 $T_i$ 共同出現以標註不同影片的相關標籤，由於其可能隱含標籤 $T_i$ 所處的情境或與標籤 $T_i$ 共同表示某個概念，在本研究中稱這些相關標籤為標籤 $T_i$ 的情境文字。基於上述標籤標註的特性，本研究提出以社會性標籤為基礎的擴充搜尋技術（Social-Tag-Based Query Expansion，STBQE）。STBQE 先試圖找出使用者所下達的搜尋字串相關的情境文字，藉由豐富搜尋字串的情境，來提高影片搜尋的效能。在現有的影音分享環境中，每部影片的發佈者通常會提供與影片內容高度相關的標題與標籤（類似文章關鍵字），因此本研究主要利用這兩種資訊來做為找尋搜尋字串相關情境文字的基礎。除此之外，本研究採用局部分析的方式來擴充搜尋字串。由於影音分享環境的資料處於經常性變動的狀態，在 WEB 2.0 中使用者往往會創造許多新用詞，例如「補教人生」、「開箱」等。如文獻所述，若使用全域分析方法則需經常進行重新分析，才能有效處理不斷創新的詞彙。局部分析的字串擴充方式只針對部份的文件進行詞彙分析，以找出具代表性的詞彙來進行字串擴充（Xu & Croft 1996）。從這個角度來看，局部分析的方式應較能因應詞彙多變的環境，本研究因此採用此方式來進行搜尋字串擴充，以期能改善影音搜尋引擎的效能。

本研究提出的 STBQE 方法主要包括情境文字辭典建構及影音檢索兩個階段（如圖 1 所示）。在情境文字辭典建構階段，首先會利用使用者的搜尋字串進行影音檔案初步檢索，並由檢索的結果中計算影音檔案標籤之間的關聯程度，以建立情境文字辭典。緊接著在影音檢索階段中，將利用前階段建立的辭典來擴充使用者的搜尋字串，並以擴充後的字串再一次進行影音檔案檢索，之後將檢索到的影音檔案依據其情境與使用者搜尋字串之間的相關程度進行排序。以下詳細說明各階段的主要工作。

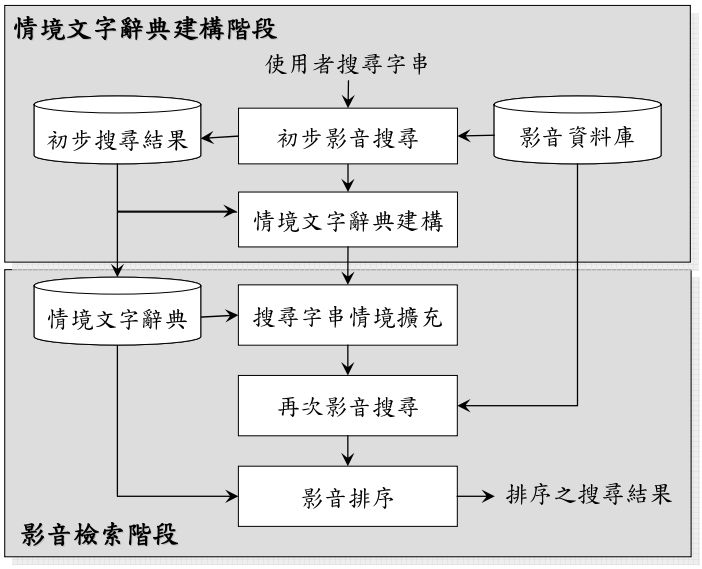


圖 1：社會性標籤為基礎的擴充搜尋技術之流程圖

## 一、情境文字辭典建構階段

情境文字辭典建構階段主要在建立情境文字辭典，作為後續搜尋字串情境擴充之依據。如圖 1 所示，此階段主要包含二個步驟，分別為初步影音搜尋及情境文字辭典建構。在初步影音搜尋步驟中，使用者所下達的搜尋字串 $Q$ 將利用關鍵字比對方式，找出所有描繪資料中（包括標題、標籤、內容描述等資料）含有搜尋字串 $Q$ 的影片集合 $V$。

緊接著，在情境文字辭典建構步驟中將分析搜尋字串 $Q$ 與影片集合 $V$ 中所出現的個別標籤 $t_i$ 的相關性。在評估搜尋字串 $Q$ 與標籤 $t_i$ 的相關性時，除考量兩者在所有影片的標籤中共同出現的頻率外，由於標題往往反應影片發佈者對於影片內容的簡短想法，而使用者通常也會參考影片標題來決定是否為所需的影片，因此本研究更進一步考量分別出現 $Q$ 與 $t_i$ 的影片標題的相似度。據此本研究提出用以衡量搜尋字串 $Q$ 與影片標籤 $t_i$ 相關性的公式如下：

$$CW_i = Freq(Q, t_i) \times Sim(\overrightarrow{TG_q}, \overrightarrow{TG_i}) \tag{6}$$

其中 $Freq(Q, t_i)$ 為標籤 $t_i$ 與搜尋字串 $Q$ 共同出現在影片集合 $V$ 中的頻率，而 $Sim(\overrightarrow{TG_q},\overrightarrow{TG_i})$ 則是影片標籤中含有 $Q$ 的影片標題集合和影片標籤中含有 $t_i$ 的影片標題集合之間的餘弦相似度（Cosine Similarity）（Singhal 2001）。本研究認為兩者標題集合的詞彙相似程度越高，則 $Q$ 和 $t_i$ 的相關程度應該越高，並藉此相似度來

調整 $Q$ 與 $t_i$ 的關聯程度。舉例說明其計算方式，假設有四部影音檔案分別為 $v_1$、$v_2$、$v_3$ 及 $v_4$，其中 $v_1$、$v_2$ 及 $v_3$ 影片的標籤中含有 $Q$（以 $TG_q$ 表示此標題集合）；而 $v_3$ 及 $v_4$ 則含有 $t_i$（以 $TG_i$ 表示此標題集合）。我們可以透過詞彙萃取的技術，來取得 $TG_q$ 和 $TG_i$ 各自包含的所有詞彙，再行比較 $TG_q$ 和 $TG_i$ 兩者擁有詞彙的相似程度。所得的相似度越高，則表示 $Q$ 和 $t_i$ 兩者共同出現的影片越多或是非共同出現的影片中字彙相似度高。

由於影片的標題通常是由句子組成，需要先進行字詞截取後方能進行比對。又標題中可能會夾雜多種語言，因此本研究先將非中文的字詞直接斷出，再將剩下的中文部份以教育部的語料庫進行詞彙比對，斷出有意義的中文字彙。緊接著將以聯集方式找出 $TG_q$ 與 $TG_i$，並將兩者以向量方式表示，並以過去研究經常採用且表現較佳的 TF×IDF 衡量方式來表達每個字彙在向量中的權重（Billhardt et al. 2002; Larsen & Aone 1999; Roussinov & Chen 1999; Wong & Yao 1992）。最後，在獲得每個字彙 $t_i$ 與 $Q$ 的相關程度 $CW_i$ 值後，進行由大到小排序，並取前 $k_c$ 個重要字彙成為情境文字集合 $C$，以做為後續情境文字擴充及影片搜尋結果排序的依據。

## 二、影音檢索階段

影音檢索階段主要包含三個步驟（如圖 1），分別是搜尋字串情境擴充、再次影音搜尋、及影音排序。在搜尋字串情境擴充階段，本研究採用前一階段建構的情境文字辭典，做為擴充搜尋字串的候選詞彙來源。由於情境文字與搜尋字串 $Q$ 具有高度相關的字彙且其具備多樣性，應有助於搜尋出更多與搜尋字串相關的影片。儘管如此，由於情境文字集合 $C$ 中的字彙數量眾多（前 $k_c$ 個 $CW_i$ 值高的字彙），為避免過度擴充，本研究進一步選取情境文字辭典中前 $k_m$ 個重要字彙做為搜尋字串的擴充字彙，用以進行影片再次搜尋。

在再次影音搜尋步驟中，本研究利用擴充後的情境文字以取得更多與搜尋字串相關的影片。由於擴充的情境文字所表達的概念可能涵蓋不同階層，有可能是搜尋字串的同義字、或是相關字、亦或是概念階層中上下層關係的字彙。基於此，本研究以 OR 的方式來進行再次搜尋，亦即逐一利用各個擴充的情境文字進行影音檢索，以取得與不同擴充情境文字相關的影音檔案，再將各別搜尋結果的聯集來做為最後搜尋結果。本研究期望可以此方式解決使用者在文字使用上的特異性，以找出不同概念階層且可能與使用者搜尋字串相關的影音檔案。

在取得重新搜尋的結果後 STBQE 將執行影片排序的步驟，亦即將最有可能符合使用者需求的影音給予較高的顯示順位。本研究藉由比對影片的描繪資訊與情境文字集合 $C$ 的相似度來進行影片的排序，描繪資訊中出現越多搜尋字串與其

情境文字，則越可能為使用者所要搜尋的影片，應該獲得較高的順位，反之則應降低其顯示順位。此外除考量情境文字出現的數量外，亦須考慮個別情境文字的重要性以及描繪資訊本身資料的長度。出現高重要性文字的影片應有相對較高的順位；而在出現相同數量與相同權重字彙的情況下，描繪資訊較短的影片也應具有相對較高的順位。舉例來說，假設影音 $v_1$ 與 $v_2$ 的描繪資訊分別擁有 10 個與 20 個字彙，且 $v_1$ 與 $v_2$ 的描繪資訊中擁有符合搜尋字串情境文字的相同字彙，則此時 $v_1$ 的情境媒合程度應較 $v_2$ 高。再者，影片的描繪資訊如前所述大致包括影片標題、標籤、內容描述、及文字評論與回應。一般來說，相較於內容描述以及文字評論與回應的內容，影片的標題與標籤通常會與影片本身具有較高度的相關。據此，本研究初步利用標題和標籤兩種影音描繪資訊來進行搜尋結果排序，並進一步考量兩者之間權重的分配對於排序結果的影響。本研究提出影片排序的衡量公式如下：

$$CR_p = \sum_{i \in C} \left( \alpha \times \frac{CW_i \cdot H_{ip}}{HW} + \beta \times \frac{CW_i \cdot T_{ip}}{TW} \right), \forall p \in RV \tag{7}$$

其中 $CR_p$ 為影音 $p$ 的排序分數，分數越高則順位越前面；$RV$ 為影片重新搜尋後的結果；$C$ 為搜尋字串的情境文字集合；$CW_i$ 指的是情境文字 $t_i$ 對於搜尋字串 $Q$ 的權重值；$H_{ip}$、$T_{ip} \in \{0, 1\}$ 分別表示影音 $p$ 的標題及標籤中是否出現情境文字 $t_i$，若有為 1，反之為 0；$HW$、$TW$ 分別代表標題萃取出的詞彙數量及標籤數量的加權值；$\alpha$、$\beta$ 分別指標題和標籤所佔的權重，且 $\alpha + \beta = 1$。

# 肆、實證評估

## 一、評估方法

### （一）實驗設計

在 STBQE 比較基準（Baseline）的選擇上，最簡單的方法是實作一個不包含搜尋字串擴充功能的傳統資訊檢索技術，如採用空間向量模型（Vector Space Model）之資訊檢索系統（Baeza-Yates & Ribeiro-Neto 1999）。然而採用這樣的技術作為比較基準可能有失公允，因現有商用影音搜索系統多已加入額外的資訊來進行搜尋結果排序的工作，例如 Google 的搜尋結果已加入 PageRank 的指標來進行排序，且搜尋結果普遍優於傳統資訊檢索技術。因此，本文未選擇傳統資訊檢索技術作為比較基準，相反地，我們決定採用最廣為使用的商業系統（State-of-the-art）作為 STBQE 的比較基準。目前影音搜尋市場主要由 YouTube.com 所領導，在所有線上影音提供商市佔率位居第一，故本研究將以 YouTube.com 的影音

搜尋結果做為比較基準，來評估使用社會性標籤擴充搜尋字串對於影音搜尋效能的影響。

為瞭解本研究提出的 STBQE 方法在實際使用環境中的表現，本研究招募 50 位曾在一個月內使用過 YouTube.com 的使用者擔任受測者。為使受測者能夠在一致的平台與介面中進行實驗，避免因實驗環境差異所造成的影響，本研究架設一個實驗用搜尋網站（如圖 2 所示）。在實驗方面，參與實驗的受測者皆被要求根據自己當下最想要搜尋的影片，下達一個或多個中文搜尋關鍵字。透過 YouTube.com 提供的 API，該網站能取得初步搜尋結果，該結果也是本研究比較基準 YouTube 的檢索結果，之後本文提出的 STBQE 會依據該結果進行後續的情境文字辭典建構、搜尋字串擴充、再次影音搜尋與影音排序等工作。在獲得兩者的檢索結果之後，網站會將結果聯集並以隨機排序方式將影片的畫面呈現在網頁中，並讓受測者勾選哪些影片與自己所下達的搜尋關鍵字相關，而受測者的回饋即為本研究評估影音檢索效能的依據。


圖 2：本研究架設之搜尋網站介面

## （二）評估準則

資訊檢索系統效能的衡量，可以透過許多不同的評估準則，衡量系統不同面向的效能。召回率（Recall）和準確率（Precision）為資訊檢索與文字探勘較常用使的評估準則（Makhoul et al. 1999），而其他的評估準則通常為此兩項指標的變化延伸。召回率評估檢索到的相關文件涵蓋所有相關文件的比率，而準確率則是

評估所有檢索的文件中有多少是相關文件。由於在實際環境中的影音檔案數量相當龐大，無法讓使用者檢視所有檔案是否符合所需，因而傳統召回率及準確率須微幅修改方能適用於本實驗評估。本研究最後採用的評估準則包括前 $N$ 筆資料準確率（Precision）、平均準確率（AP）、及平均效益（AU），其分別說明如下。

本研究採用傳統的準確率來評估影音檢索系統取回的前 $N$ 筆影音能夠符合使用者需求的命中程度，計算方式如下所示：

$$Precision = \frac{\sum_{r=1}^{N} R(r)}{N} \tag{8}$$

$N$ 為搜尋引擎檢索回的影音數量；$r$ 代表影音的排序；$R(r)$指排序為 $r$ 的影音是否為相關的影音檔案，相關為 1，不相關為 0。

影音檢索除了考量取得正確資訊的能力外，也應該考量影音檔案的排序位置。對於搜尋字串的關聯程度越高的影音檔案，應該要放置在影音檢索列表的越前端，讓使用者較先閱讀到使用者認為最相關的影音檔案，減少觀看到不相關影音的機率。為了評估檢索排序結果的優劣，本研究採用 Voorhees 與 Harman（1999）所定義的平均準確率（Average Precision, AP）做為評估準則。AP 數值越高代表系統檢索影音的排序效能越好，其計算方式如下列公式所示：

$$AP = \frac{\sum_{r \in R} P(r)}{|R|} \tag{9}$$

$|R|$代表使用者認定相關的影音集合數量；$P(r)$指的是當排序為 $r$ 的相關影音之準確率。

舉例來說，搜尋引擎根據搜尋字串檢索出 20 份的影音檔案，其中 4 份為使用者認為與搜尋字串相關的影片，其在搜尋結果的排序分別為 1、2、4 及 10。則第 1 個影片的準確率為 1/1=1，第 2 個影片的準確率為 2/2=1，第 3 個影片的準確率為 3/4=0.75，第 4 個影片的準確率為 4/10=0.4。故平均準確率（$AP$）為 (1+1+0.75+0.4)/4=0.7875。

平均準確率將搜尋結果的排序納入考量，改善傳統準確率沒有考量排序效能的缺點。儘管如此，其衡量的是使用者獲得資訊的效益，仍未考慮檢索影音過程中獲得資訊的成本。亦即若搜尋引擎提供過多非使用者所需的檔案，則使用者需要花費額外的時間進行檢視，無形中也會增加資訊獲得的成本。因此，本研究將平均準確率的分母修改為檢索取回的檔案數量，提出考量資訊檢索成本的平均效

益（Average Utility, AU）指標，如下列公式所示：

$$AU = \frac{\sum_{r \in R} P(r)}{N} \tag{10}$$

舉例來說，搜尋引擎檢索到 20 筆認為相關的影音檔案，但只有 4 個影音檔案是使用者認為相關的，且在搜尋結果的排序分別為 1、2、4 及 10，則 $AU$ 為（1+1+0.75+0.4）/20=0.1575。

## 二、影音檢索效能分析

在使用本研究所提 STBQE 時需考量關鍵字擴充數量（$k_m$）、情境文字數量（$k_c$）、標題和標籤權重分配（$\alpha$、$\beta$）以及詞彙字數長度加權方式（*HW*、*TW*）等四項參數，本研究以實驗方式決定各參數的較佳值來進行後序的比較分析。表 1 為各參數的數值範圍及最終採用的參數值，本研究最後採用擴充字數為 3、情境文字數為 45、標題權重為 0.5、標籤權重為 0.5、詞彙長度加權方式使用 log 加權。

表 1：參數範圍及採用數值

| 參數名稱 | 變數名稱 | 參數範圍 | 調整間距 | 採用數值 |
|---|---|---|---|---|
| 關鍵字擴充數量 | $k_m$ | 0~3 | 1 | 3 |
| 情境文字數量 | $k_c$ | 5~50 | 5 | 45 |
| 標題權重 | $\alpha$ | 1~0 | 0.1 | 0.5 |
| 標籤權重 | $\beta$ | 1~0 | 0.1 | 0.5 |
| 詞彙數目加權 | *HW* 與 *TW* | 1. 沒有加權<br>2. 詞彙數量<br>3. Log（詞彙數量+1）<br>4.（詞彙數量）2 | NA | Log（詞彙數量+1） |

表 2：單一關鍵字與多個關鍵字受測者勾選資料分析

| | 受測者數量 | 平均勾選 YouTube 回傳影片數 | 平均勾選 STBQE 回傳影片數 | 平均勾選影片比例 |
|---|---|---|---|---|
| 單一關鍵字 | 32 | 7.63 | 8.56 | 45.16% |
| 多個關鍵字 | 18 | 10.39 | 6.33 | 44.99% |
| 總和 | 50 | 8.62 | 7.76 | 45.10% |

（一）效能比較分析

根據先前的研究結果顯示（Wei et al. 2007），搜尋關鍵字的多寡會顯著影響擴充搜尋的效能，故本研究先將 50 個受測者根據提供的搜尋字串中關鍵字的數量分成單一關鍵字與多個關鍵字兩群，再分別進行檢索效能的比較。表 2 數據顯示，在單一關鍵字的群集中共有 32 名受測者，其平均勾選 YouTube 回傳影片及 STBQE 回傳影片數量為 7.63 及 8.56；而多個關鍵字群集中則有 18 名受測者，其平均勾選 YouTube 回傳影片及 STBQE 回傳影片數量為 10.39 及 6.33；平均總共勾選的影片比例分別為 45.16%及 44.99%。當受測者利用單一關鍵字進行搜尋時，STBQE 所回傳的前 20 部影片中能夠提供相對較多的相關影片；反之當受測者利用多個關鍵字進行搜尋時，則受測者認為 YouTube 能提供相對較多的相關影片。圖 2(a)–(c)分別為 STBQE 與 YouTube 在取回 2~20 筆影音檔案時的 Precision、AP 與 AU 等指標的比較結果。實驗結果發現，本研究所提出之 STBQE 在 Precision 與 AU 兩項指標皆優於 YouTube 的檢索結果，但 AP 指標較 YouTube 差。Precision 與 AU 佳但 AP 差表示 STBQE 在最前幾筆影音（例如，前 2 筆）的檢索排序可能較 YouTube 差，但整體（前 20 筆）的相關比數與排序結果卻是較佳的。舉例來說，如果只觀察前 2 筆的影音排序，有兩個影音搜尋引擎皆取得 1 個相關檔案，此時有一個搜尋引擎給予相關檔案排序 1，但另一個給予排序 2，前者 AP 值為 1，後者為 0.5，AP 值的差距高達 0.5，即使第 2 個搜尋引擎在其他相關檔案的排序較第 1 個搜尋引擎佳，也很難追回 0.5 的 AP 差值。整體而言，我們認為 STBQE 在單一關鍵字的檢索效能是較 YouTube 為佳的，原因在於，單一關鍵字所能表達的檢索需求較為有限，因此使用社會性標籤來進行搜尋字串擴充能夠改善檢索的效能。

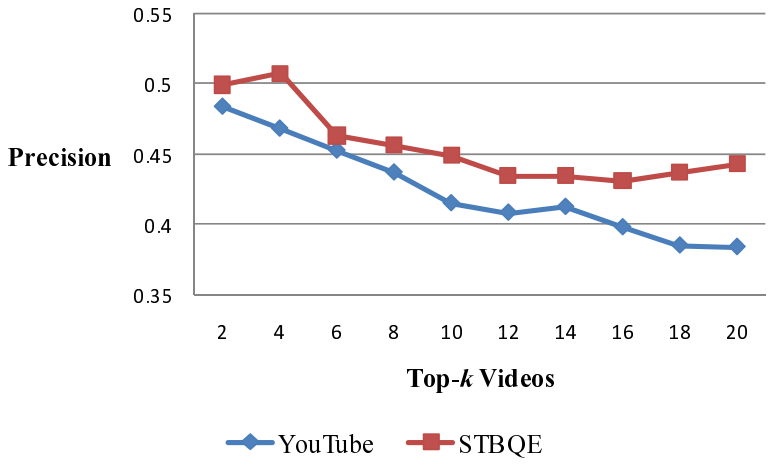


圖 3(a)：單一關鍵字 Precision 評估結果

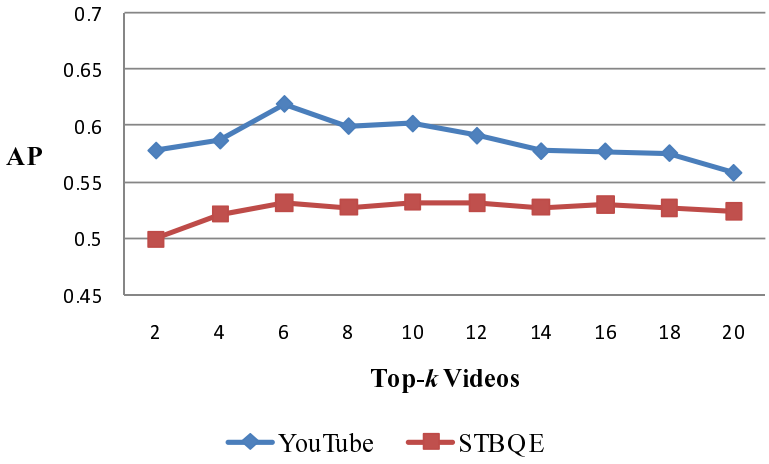


圖 3(b)：單一關鍵字 AP 評估結果

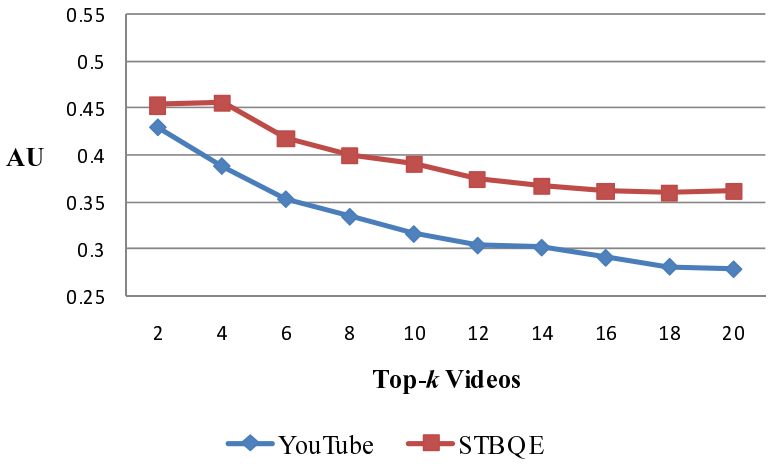


圖 3(c)：單一關鍵字 AU 評估結果

緊接著，我們分析 STBQE 與 YouTube 在多個關鍵字的檢索效能差異。如圖 4(a)–(c)所示，YouTube 在 Precision、AP 與 AU 等 3 個指標上皆顯著優於 STBQE 方法，比較結果與單一關鍵字時完全相反。本研究推測其可能的原因是兩個以上的關鍵字能夠表達的檢索需求已經相當豐富，且影音檔案所擁有的少量描繪資訊若存在兩個以上的關鍵字，但卻不會被認定為相關影音的機率相當低，因此以社會性標籤來進行搜尋字串擴充並無法提升檢索效能，甚至可能會因為過度擴充而降低搜尋效能。

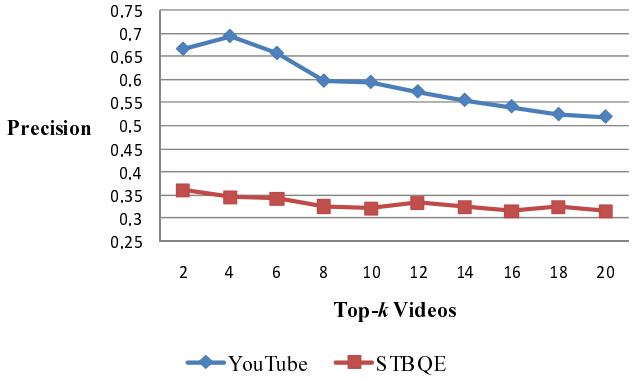


圖 4(a)：多個關鍵字 Precision 評估結果

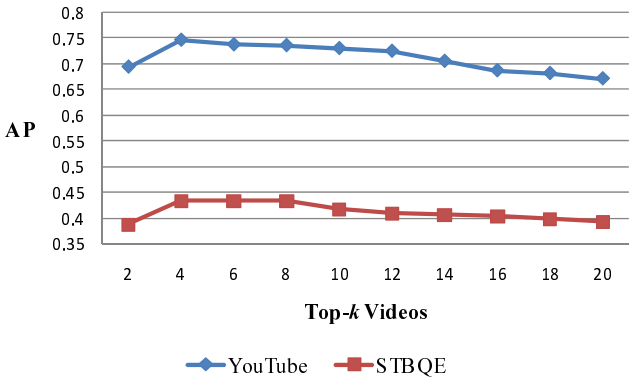


圖 4(b)：多個關鍵字 AP 評估結果

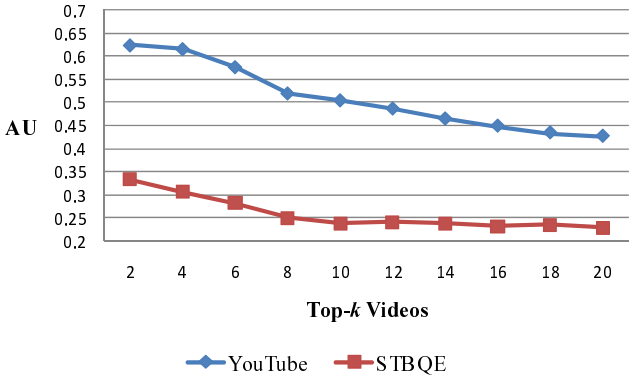


圖 4(c)：多個關鍵字 AU 評估結果

整體來說，STBQE 在單一關鍵字的檢索效能優於 YouTube，但在多個關鍵字時較 YouTube 差。造成此種現象的原因，可能是 STBQE 在擴充字串時，針對每個關鍵字各自進行擴充後再分別進行影片搜尋，而導致搜尋出的影片內容發散。而根據初步搜尋結果來建構的情境文字辭典可能也因此無法聚焦，造成後續進行排序時無法真正定位使用者所要搜尋的影片情境。此外，STBQE 在使用單一關鍵字進行搜尋的情況下獲得較好的效能。儘管 Hawking（2006）的研究指出使用者的平均搜尋字串長度為 2.3 個字（Word），但其主要是針對使用英文進行資料搜尋的環境。一般來說英文多會以複數字來表達一個概念，例如資訊檢索以英文表達則為 Information Retrieval 兩個字，相較於中文表示則只有「資訊檢索」一個關鍵字。而本研究實驗的樣本也顯示，在不限定受測者使用的搜尋對象及字串長度之下，約有三分之二的受測者會使用單一關鍵字來進行影片搜尋，這也表示本研究提出的方法能被應用在實際的搜尋環境中。

### （二）初步搜尋結果之影響分析

本研究認為初始搜尋所得到的影片與標籤的數量應該會影響檢索的效能。影片的描述標籤越少，可能暗示著影片所表達的內容單純、特定，須用或可用以描繪影片的文字數量少、差異相對較小；相反地，影片的描述標籤越多，則可能隱含著影片所表達的內容複雜，須用或可用以描繪影片的文字數量多、差異也可能相對較大。因此，若使用者所下達的搜尋字串，其初步搜尋結果中影片的描述標籤較少，可能表示使用者所要搜尋影片內容比較特定，而當使用者的搜尋想法越特定，則搜尋引擎應相對給予更精確的結果。基於此，本研究進一步藉由初步搜尋結果中影片標籤的比例來瞭解不同情境下對搜尋績效的影響。

本研究定義影音-標籤密度指標，其計算方式為每個使用者的初步搜尋結果中影片的數量除以所有共同出現的標籤數量。本研究依據計算出的影音-標籤密度由大到小排序後，再將資料均分成三群，亦即形成高、中、低密度三群，並分別觀察不同影音-標籤密度對於搜尋效能的影響。舉例來說，假設有 6 個使用者其初步搜尋結果的影音-標籤密度分別為 0.3、0.1、0.5、0.2、1.0 及 0.7，則排序後第 5、6 位使用者屬於高密度群集；第 1、3 位屬於中密度群集；第 2、4 位則屬低密度群集。高密度表示初步搜尋結果中影片描述標籤較少，影片內容可能比較特定；相反地，低密度可能表示初步搜尋結果中影片描述標籤較多，影片內容較複雜。

本研究分別分析單一關鍵字及多個關鍵字的結果，並於表 3 及表 4 分別列出兩者的相關資訊。比較單一關鍵字及多個關鍵字在高、中、低密度三群的影音-標籤密度發現，單一關鍵字的密度值遠高於多個關鍵字的密度值，表示利用單一關鍵字進行影片搜尋的使用者其所要搜尋的對象大多明確。此外，不論關鍵字數多寡，高密度的群集其影音-標籤密度都遠大於中、低密度群集，而後兩者之間的差異則相對較小。

表 3：單一關鍵字—影音及標籤數量相關資料

| | 平均初始搜尋影音數量 | 平均共同出現標籤數量 | 影音－標籤密度最大值 | 影音－標籤密度最小值 | 影音－標籤密度平均值 |
|---|---|---|---|---|---|
| 高密度 | 596.09 | 231.45 | 11.02 | 1.25 | 4.06 |
| 中密度 | 285.46 | 300.00 | 1.23 | 0.78 | 0.98 |
| 低密度 | 290.10 | 507.20 | 0.75 | 0.33 | 0.59 |
| 總計 | 393.69 | 341.19 | 11.02 | 0.33 | 1.92 |

表 4：多個關鍵字—影音及標籤數量相關資料

| | 平均初始搜尋影音數量 | 平均共同出現標籤數量 | 影音－標籤密度最大值 | 影音－標籤密度最小值 | 影音－標籤密度平均值 |
|---|---|---|---|---|---|
| 高密度 | 400.67 | 408.67 | 1.34 | 0.61 | 0.96 |
| 中密度 | 134.67 | 284.67 | 0.52 | 0.43 | 0.47 |
| 低密度 | 144.00 | 421.00 | 0.41 | 0.13 | 0.31 |
| 總計 | 226.44 | 371.44 | 1.34 | 0.13 | 0.58 |

表 5 顯示在不同影音標籤密度下使用單一關鍵字與多個關鍵字進行搜尋之受測者勾選相關影片的狀況。數據顯示在單一關鍵字群集中，除高影音標標籤密度的情況呈現差異不大的情況外，其餘兩者的情況下都是 STBQE 被受測者認為回傳的相關影片較多；反之在多關鍵字的情況下則仍是以 YouTube 的相關影片數量較多。然而不論搜尋關鍵字數量，兩種方式在低影音標籤密度的情況下，其搜尋效能皆明顯比另兩種情況來得差。推論可能的原因為，低影音標籤密度的影片其描繪標籤的用字可能較多且發散，而較不容易找到使用者認為相關程度較高的影片。

表 5：不同影音標籤密度下單一關鍵字與多個關鍵字受測者勾選資料分析

| | 單一關鍵字 | | 多個關鍵字 | |
|---|---|---|---|---|
| | 平均勾選 YouTube 回傳影片數 | 平均勾選 STBQE 回傳影片數 | 平均勾選 YouTube 回傳影片數 | 平均勾選 STBQE 回傳影片數 |
| 高密度 | 8.45 | 8.36 | 12 | 6.33 |
| 中密度 | 8.36 | 10.27 | 11.83 | 7.83 |
| 低密度 | 5.9 | 6.9 | 7.33 | 4.83 |

圖 5(a)呈現搜尋字串為單一關鍵字時，不同程度的標籤密度下 Precision 的表現。在低密度的情形下，YouTube.com 的搜尋效能明顯隨著影片排序增加而急遽下降，這可能是由於低密度群集中影片的描述標籤較多差異較大，導致單純使用關鍵字搜尋方式的效能會受到明顯影響。相對地，STBQE 透過情境文字擴充及比對方式來搜尋並排序影片，因此在影片-標籤密度低時，其搜尋準確度大致維持穩定。在高密度群集的表現上，STBQE 的數據較為出色，影片-標籤密度高代表搜尋出的影片內容較明確且所有的影片大致上呈現相同的概念。儘管如此，影音可能由不同面向來表達概念，利用情境文字來進行排序，某些程度上可以反應出目前較多使用者上傳及觀看的影音種類。而在標籤密度為中等時，YouTube 在 $k$ 小於 14 時優於本研究的 STBQE，但 $k$ 大於 14 時呈現相反結果。

圖 5(b)為單一關鍵字在不同密度時，AP 數值的變化情形。在高密度群集中仍是以 STBQE 的效果為佳。此外，值得注意的是 YouTube 在低密度的獲得的 Precision 不佳，但是卻擁有高的 AP 值，表示 YouTube 找到的相關影音較少，但在排序上給予較高的排序，因而第 6 筆資料之後呈現 AP 數值逐漸下滑的現象。而 STBQE 在低密度的數據表上現，則是持續的有找到相關的影音，因此 AP 值才會維持在一定的水準。最後，圖 5(c)為 AU 的數值，結果呈現 STBQE 在高、低密度的效能通常較 YouTube 好，且 YouTube 在低密度效能快速下降的趨勢。

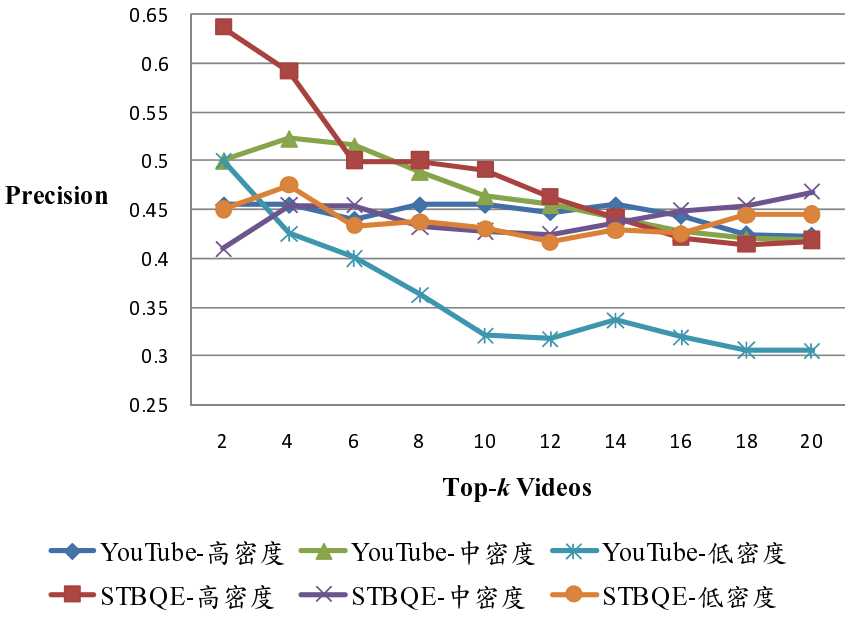


圖 5(a)：影音標籤密度對單一關鍵字 Precision 的影響

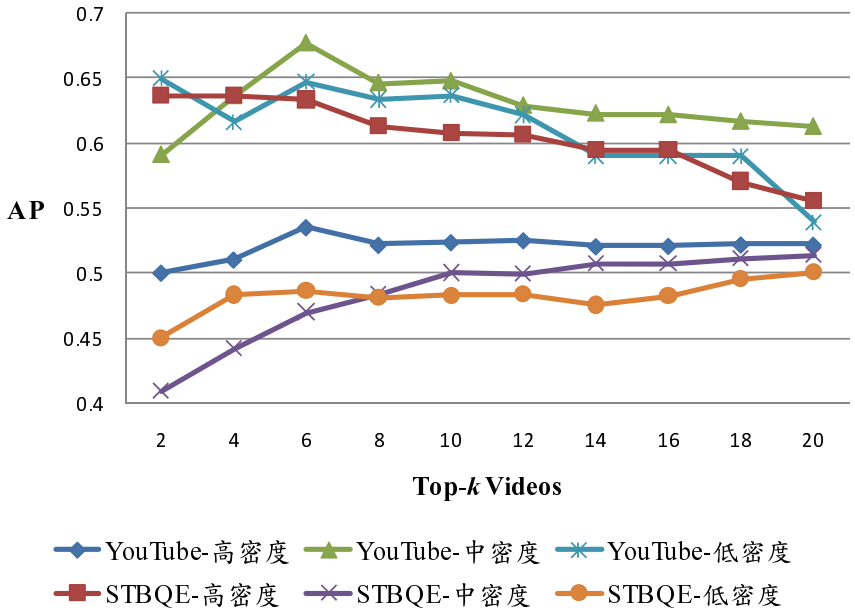


圖 5(b)：影音標籤密度對單一關鍵字 AP 的影響

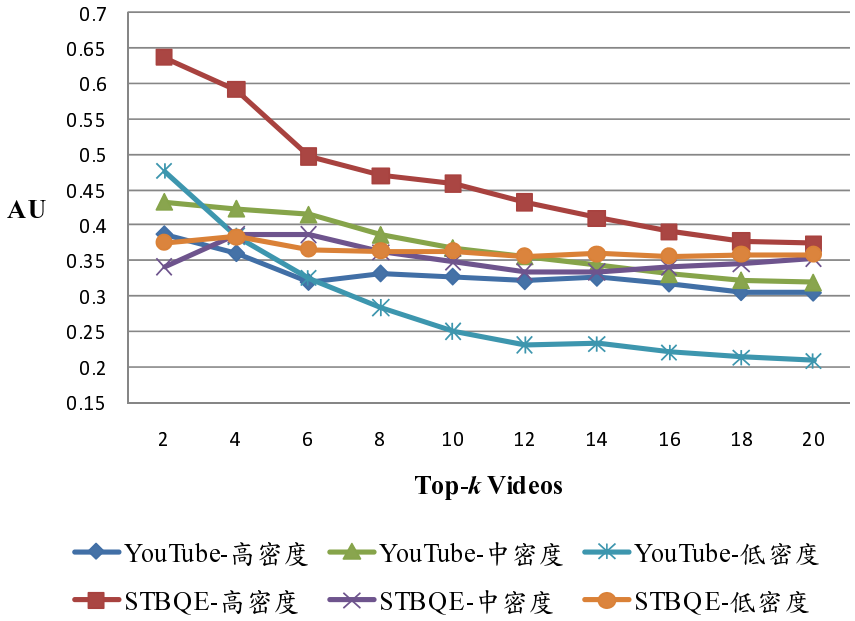


圖 5(c)：影音標籤密度對單一關鍵字 AU 的影響

接著觀察多個關鍵字在不同影音-標籤密度下檢索效能的差異，從圖 6(a)–(c)中可以得知，YouTube 在高、中、低密度的情形下，Precision、AP 與 AU 等效能指標皆明顯優於本研究提出的 STBQE，顯示當利用多個關鍵字進行搜尋時，利用社會性標籤來擴充原始搜尋字串並無法改善檢索效能，這可能是因為本研究分別對各個關鍵字進行擴充，導致過度擴充而使得主題更加模糊。反過來看，單純以關鍵字比對的方式反而能夠更清楚鎖定目標。此外，我們也發現在提供多關鍵的情形下，YouTube 在高密度群集的檢索效能反而呈現快速下降的趨勢，此結果恰好與單一關鍵字時獲得相反的結果，其可能的原因是過多的關鍵字反而在對描述標籤較少、概念較明確的影片進行搜尋比對時造成額外的干擾。

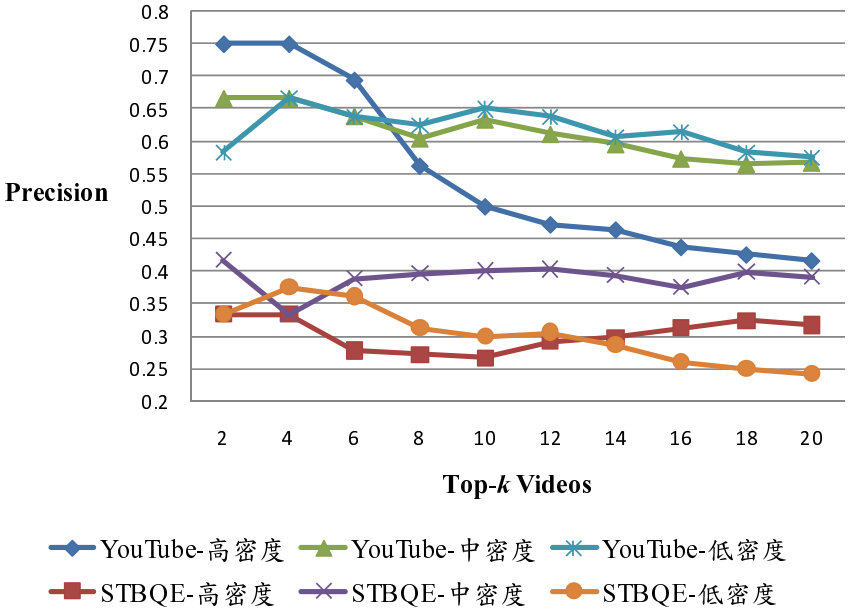


圖 6(a)：影音標籤密度對多個關鍵字 Precision 的影響

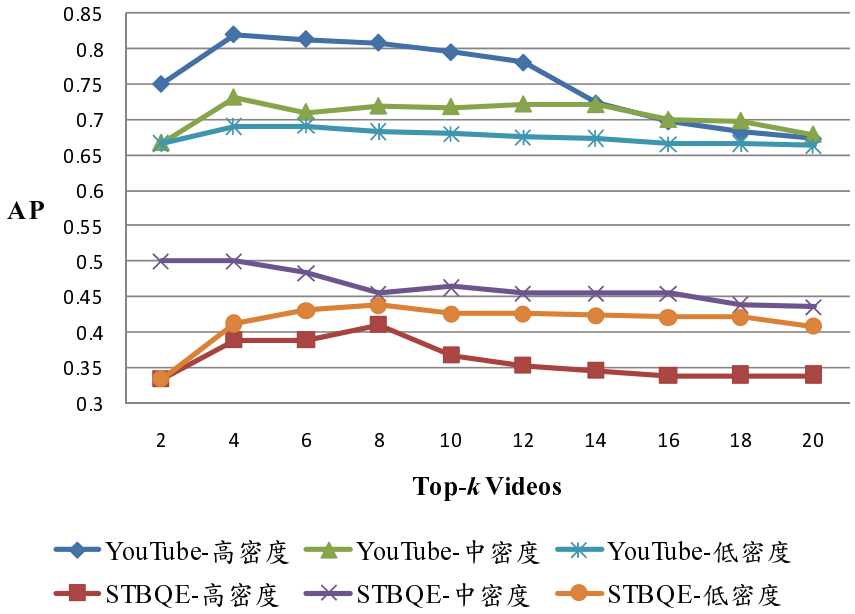


圖 6(b)：影音標籤密度對多個關鍵字 AP 的影響

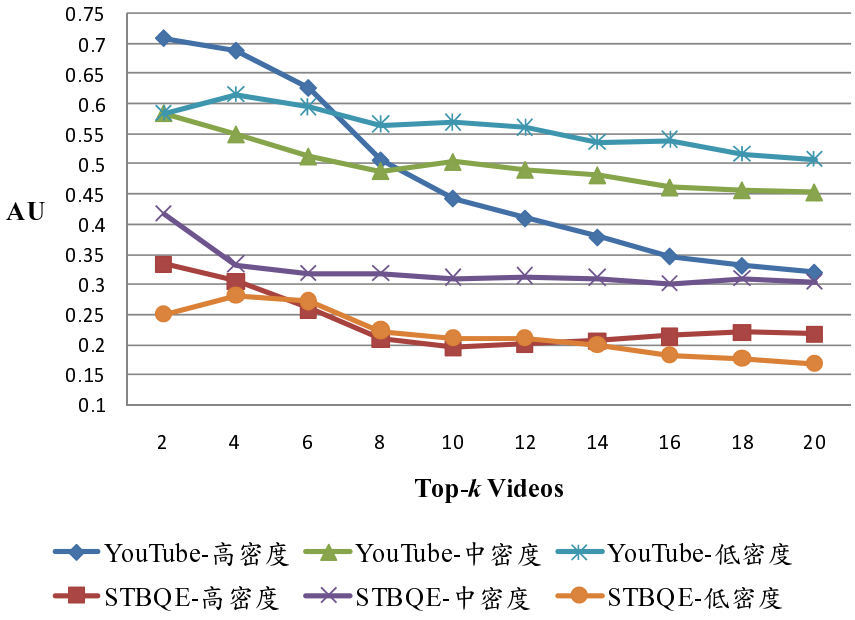


圖 6(c)：影音標籤密度對多個關鍵字 AU 的影響

從上述的數據可知，在單一關鍵字的表現上，本研究提出的 STBQE 優於 YouTube 的影音檢索。在低密度的數據可以觀察到，本研究可以有效解決使用字詞的差異，提升影音搜尋效能；而高密度的表現上，情境比對的方式可以較正確地進行排序。此外，在多個關鍵字的數據上，YouTube 則較本研究提出的 STBQE 有較佳的搜尋效能，這可能是由於 STBQE 採用聯集方式擴充搜尋範圍，反而導致模糊了搜尋結果。

## 伍、結論與未來研究方向

為解決以關鍵字為基礎的影音搜尋方式所造成字串錯配的問題，本研究提出以社會性標籤為基礎的擴充搜尋技術（STBQE），以設法擴充使用者搜尋字串的情境，提高影片搜尋效能。本研究以 YouTube 搜尋引擎做為比較基準，並分析在不同關鍵字數與不同的影音－標籤密度下，兩種影音檢索系統的效能。實證評估結果顯示，本文所提 STBQE 在搜尋字串為單一關鍵字時獲得較優於 YouTube 的檢索效能，然而在搜尋字串為多個關鍵字組成時，檢索效能不如 YouTube。此結果表示若使用者可以清楚描述所要搜尋的影片內容，則傳統關鍵字搜尋方式便能夠達成相當不錯的效能；相對地，若僅用單一關鍵字進行搜尋，則本研究提出之 STBQE 利用擴充情境文字的方式來擴充使用者搜尋字串，能有助於改善利用單一關鍵字搜尋時資訊需求不明確的問題。另一方面，由於 STBQE 針對搜尋字串中的每個關鍵字分別進行擴充及搜尋，因此在面對使用者利用多個關鍵字進行搜尋時，反而因過度擴充而讓資訊發散，進行降低檢索效能。此外本研究進一步發現，在單一關鍵字搜尋且初步搜尋結果中的影片-標籤密度高時，利用 STBQE 來進行擴充並再次搜尋，能獲得較佳的搜尋結果

本研究貢獻可從技術與實務應用兩方面來說明。在技術方面，針對影音搜尋環境可供搜尋資訊過少的問題，先前研究多著重在自動或半自動標籤建議的研究上，鮮少探討搜尋字串擴充技術在此環境下的效能，因此本研究提出以社會性標籤為基礎的擴充搜尋方式，並以實證研究方式評估搜尋字串擴充技術在實際的影音搜尋環境（YouTube）中的效能。本研究所提技術利用少量的影片標籤資訊來進行搜尋字串擴充，將影片搜尋與排序方式從傳統以「點」為主的關鍵字比對，轉變成以「面」為主的情境比對，而本研究的實證評估結果也顯示，STBQE 在搜尋字串為單一關鍵字時普遍優於單純關鍵字比對的 YouTube。從宏觀的角度來看，本研究提出以局部回饋分析，且利用社會性標籤來建構字串擴充辭典的方式，確實能用於具有字詞快速變動特性的 WEB 2.0 環境中，並有助於提升影音搜尋的效能。在實務應用方面，一般來說在中文影片搜尋環境中多半以使用單一關鍵字搜尋居多，雖目前並無相關研究統計，但以本研究的實驗結果來說，即有約

三分之二的受測者只用單一關鍵字進行搜尋，又如結果所示 STBQE 在單一關鍵字時能夠獲得相對較佳的效能，因此在實務上應有一定的實用價值。此外，本研究所僅利用影片本身的描繪資訊來進行字串擴充，其可架構在原有的利用關鍵字比對的搜尋引擎之上，不需進行大規模的系統修正即可應用。再者其雖需兩次搜尋，但在現有的分散式搜尋架構下應能使執行時間在合理的範圍內，而適用於線上即時搜尋回應的環境。

本研究仍有下列的研究限制與未來研究方向：(1)本文目前僅針對中文進行處理，但實務上一個影音檔案的標籤可能其他語言的文字，因此未來需延伸STBQE，使其能考量不同語言影片標籤的差異。(2)本文僅利用情境文字相似度來排序檢索得的影音檔案，但在 WEB 2.0 快速變動的環境下，額外的因素，例如影片發佈時間或是使用者對於新影片的偏好等，都會是影響使用者判斷影音檔案是否相關的依據，未來研究應將 WEB 2.0 環境下的重要因素納入排序時的考量。(3)STBQE 僅利用影音檔案的標籤來進行情境文字辭典的建構，但影音檔的其他描述文字，尤其是標題，可能也包含了重要的情境資訊，因此未來在建構情境文字辭典也應將這些資訊納入評估。(4)本研究著重於分析目前主要的搜尋字串擴充演算法的優缺點，並加以考量影音分享網站的特性來適當修改選定的演算法，進而提出一個適用於影音搜尋應用的演算法。因此，在實際驗證方面，本研究針對所提的概念在實際的影音搜尋環境中與現行搜尋引擎進行搜尋效能的比較，而結果也顯示本研究提出的字串擴充方法可在某些情境下提升影片搜尋的效能。基於此，在未來的研究上，我們將進一步考慮比較不同字串擴充演算法之間的優劣，以其能分析出一個更佳的應用於影音搜尋環境的字串擴充方法。

## 致謝

本研究經行政院國家科學委員會專題研究計畫部份經費補助（NSC97-2410-H-415-016、NSC 98-2410-H-155-008），作者並感謝匿名審查委員給予的寶貴建議與意見，使本研究得以更加完善，謹此致謝。

## 參考文獻

Attar, R. and Fraenkel, A.S. (1977), 'Local feedback in full-text retrieval systems', *Journal of the ACM*, Vol. 24, No. 3, pp. 397-417.

Au Yeung, C.M., Gibbins, N. and Shadbolt, N. (2007), 'Understanding the semantics of ambiguous tags in folksonomies', *Proceedings of the International Workshop on Emergent Semantics and Ontology Evolution*, Busan, South Korea, November 12, pp. 108-121.

Baeza-Yates, R. and Ribeiro-Neto, B. (1999), *Modern Information Retrieval*, Addison Wesley/ACM Press, New York, NY.

Ballan, L., Bertini, M., Bimbo, A.D., Meoni, M. and Serra, G. (2010), 'Tag suggestion and localization in user-generated videos based on social knowledge', *Proceedings of Second ACM SIGMM Workshop on Social Media*, Firenze, Italy, October 25-29, pp. 3-8.

Billhardt, H., Borrajo, D. and Maojo, V. (2002), 'A context vector model for information retrieval', *Journal of the American Society for Information Science and Technology*, Vol. 53, No. 3, pp. 236-249.

Buckley, C., Singhal, A., Mitra, M. and Salton, G. (1996), 'New retrieval approaches using SMART: TREC 4', *Proceedings of the TREC 4 Conference*, Gaithersburg, MD, November 1-3, pp. 25-48.

Carpineto, C., Romano, G. and Giannini, V. (2002), 'Improving retrieval feedback with multiple term-ranking function combination', *ACM Transactions on Information Systems*, Vol. 20, No. 3, pp. 259-290.

Chi, E.H. and Mytkowicz, T. (2007), 'Understanding navigability of social tagging systems', *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, California, April 28-May 3.

Church, K.W. and Hanks, P. (1989), 'Word association norms, mutual information and lexicography', *Proceedings of the 27th Annual Meeting of the ACL*, Vancouver, Canada, June 26-29, pp. 76-83.

Croft, W.B., Cook, R. and Wilder, D. (1995), 'Providing government information on the Internet: experiences with THOMAS', *Proceedings of the Second Annual Conference on the Theory and Practice of Digital Libraries*, Austin, Texas, June 11-13, pp. 19-24.

Cui, H., Wen, J.R., Nie, J.Y. and Ma, W.Y. (2003), "Query expansion by mining user logs', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 15, No. 4, pp. 829-839.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R.A. (1990), 'Indexing by latent semantic analysis', *Journal of the American Society for Information Science*, Vol. 41, No. 6, pp. 391-407.

Dinh, D. and Tamine, L. (2011), "Combining global and local semantic contexts for improving biomedical information retrieval', *Proceedings of the 33rd European Conference on Information Retrieval*, Dublin, Ireland, April 18-21, pp. 375-386.

Drucker, P.F. (2001), *Management Challenges for the 21st Century*, Harper Business,

New York.

Fano, R. (1961), *Transmission of Information: A Statistical Theory of Communications*, MIT Press, Cambridge, MA.

Furnas, G.W., Landauer, T.K., Gomez, L.M. and Dumais, S.T. (1987), 'The vocabulary problem in human-system communication', *Communications of the ACM* , Vol. 30, No. 11, pp. 964-971.

Gauch, S., Wang, J. and Rachakonda, S.M. (1999), 'A corpus analysis approach for automatic query expansion and its extension to multiple databases', *ACM Transactions on Information Systems*, Vol. 17, No. 3, pp. 250-269.

Golder, S.A. and Huberman, B.A. (2006), 'Usage patterns of collaborative tagging systems', *Journal of Information Science*, Vol. 32, No. 2, pp. 198-208.

Gong, Z., Cheang, C.W., and Leong, H.U. (2005), 'Web query expansion by WordNet', *Proceedings of the 16th International Conference on Database and Expert System Applications*, Copenhagen, Denmark, August 22-26, pp. 166-175.

Gong, Z., Muyeba, M. and Guo, J. (2010), 'Business information query expansion through semantic network', *Enterprise Information Systems*, Vol. 4, No. 1, pp. 1-22.

Harman, D. (1992), 'Relevance feedback revisited', *Proceedings of the 15th Annual International Conference on Research and Development in Information Retrieval*, Copenhagen, Denmark, June 21-24, pp. 1-10.

Harman, D. (1993), 'Overview of the first text retrieval conference (TREC-1)', *Proceedings of the First Text Retrieval Conference (TREC-1)*, Gaithersburg, Maryland, November 4-6, pp. 1-20.

Haubold, A., Natsev, A. and Naphade, M.R. (2006), 'Semantic multimedia retrieval using lexical query expansion and model-based reranking', *Proceedings of the 2006 IEEE International Conference on Multimedia and Expo*, Toronto, Canada, July 9-12 pp. 1761-1764.

Hawking, D. (2006), 'Web search engines: Part 2', *IEEE Computer*, Vol. 39, No. 8, pp. 88-90.

Hoeber, O., Yang, X.D. and Yao, Y. (2005), 'Conceptual query expansion', *Proceedings of the Third International Atlantic Web Intelligence Conference,* Lodz, Poland, June 6-9, pp. 190-196.

Hong, Z., Syin, C. and Lai, K.F. (1998), 'Query expansion by text and image features in image retrieval', *Journal of Visual Communication and Image Representation*, Vol. 9, No. 4, pp. 287-299.

Huang, C.K., Chien, L.F. and Oyang, Y.J. (2003), 'Relevant term suggestion in

interactive Web search based on contextual information in query session logs', *Journal of the American Society for Information Science and Technology*, Vol. 54, No. 7, pp. 638-649.

Jing, Y. and Croft, W.B. (1994), 'An association thesaurus for information retrieval', *Proceedings of the RIAO'94*, New York, October 11-13, pp. 146-160.

Jones, K.S. (1971), *Automatic Keyword Classification for Information Retrieval*, Buterworth, London, UK.

Jones, K.S. (1972), 'A statistical interpretation of term specificity and its application in retrieval', *Journal of Documentation*, Vol. 28, No. 1, pp. 11-21.

Khan, M.S. and Khor, S. (2004), 'Enhanced Web document retrieval using automatic query expansion', *Journal of the American Society for Information Science and Technology*, Vol. 55, No. 1, pp. 29-40.

Krestel, R., Fankhauser, P. and Nejdl, W. (2009), 'Latent dirichlet allocation for tag recommendation', *Proceedings of the Third ACM Conference on Recommender Systems (RecSys'09)*, New York, October 22-25, pp. 61-68.

Larsen, B. and Aone, C. (1999), 'Fast and effective text mining using linear-time document clustering', *Proceedings of the Fifteenth International Conference on Knowledge Discovery and Data Mining*, San Diego, CA, August 15-18, pp.16-22.

Lew, M. S., Sebe, N., Djeraba, C. and Jain, R. (2006), 'Content-based multimedia information retrieval: state of the art and challenges', *ACM Transactions on Multimedia Computing, Communications, and Applications*, Vol. 2, No. 1, pp. 1-19.

Liu, Z. and Chu, W.W. (2005), 'Knowledge-based query expansion to support scenario-specific retrieval of medical free text', *Proceedings of the ACM Symposium on Applied Computing*, Santa Fe, New Mexico, March 13-17, pp. 1076-1083.

Losada, D.E. (2010), 'Statistical query expansion for sentence retrieval and its effects on weak and strong queries', *Information Retrieval*, Vol. 13, No. 5, pp. 485-506.

MacGregor, G. and McCulloch. E. (2006), 'Collaborative tagging as a knowledge organisation and resource discovery tool', *Library View*, Vol. 55, No. 5, pp. 291-300.

Makhoul, J., Kubala, F., Schwartz, R. and Ralph, W. (1999), 'Performance measures for information extraction', *Proceedings of DARPA Broadcast News Workshop*, pp. 249-252.

Maron, M.E. and Kuhns, J.L. (1960), 'On relevance, probabilistic indexing and information retrieval', *Journal of the ACM*, Vol. 7, No. 3, pp. 216-244.

Miller, G.A., Beckwith, R., Felbaum, C., Gross, D. and Miller, K. (1990), 'Introduction

to WordNet: an on-line lexical database', *International Journal of Lexicography*, Vol. 3, No. 4, pp. 235-244.

Mitchell, T. (1997), *Machine Learning*, McGraw Hill, Boston, MA.

Passant, A. and Laublet, P. (2008), 'Meaning of a tag: a collaborative approach to bridge the gap between tagging and linked data', *Proceedings of the Workshop on Linked Data on the Web*, Beijing, China.

Qiu, Y. and Frei, H.P. (1993), 'Concept based query expansion', *Proceedings of the 16th Annual International Conference on Research and Development in Information Retrieval*, Pittsburgh, PA, June 27-July 1, pp. 160-169.

Quinlan, J.R. (1986), 'Induction of decision tree', *Machine Learning*, Vol. 1, pp. 81-106.

Rahman, M.M. and Bhattacharya, P. (2009), 'Image retrieval with automatic query expansion based on local analysis in a semantical concept feature space', *Proceeding of the ACM International Conference on Image and Video Retrieval*, Island of Santorini, Greece, July 8-10.

Roberson, S.E. (1997), 'The probability ranking principle in IR', J*ournal of Documentation*, Vol. 33, No. 4, pp. 294-304.

Rocchio, J.J. (1971), 'Relevance feedback in information retrieval', in Salton, G. (Ed.), *The SMART Retrieval System*, Prentice-Hall, Upper Saddle River, NJ, pp. 313-323.

Roussinov, D. and Chen, H. (1999), 'Document clustering for electronic meetings: an experimental comparison of two techniques', *Decision Support Systems*, Vol. 27, No. 1, pp. 67-79.

Rowley, J.E. and Farrow, J. (1995), *Organizing Knowledge: An Introduction to Managing Access to Information*, Gower, Brookfield, VT.

Ruthven, I. (2003), 'Re-examining the potential effectiveness of interactive query expansion', *Proceedings of the 26th Annual International Conference on Research and Development in Information Retrieval*, Toronto, Canada, July 28-August 1, pp. 213-220.

Salton, G. and Buckley, C. (1990), 'Improving retrieval performance by relevance feedback', *Journal of the American Society for Information Science*, Vol. 41, No. 4, pp. 288-297.

Salton, G., Fox, E.A. and Wu, H. (1983), 'Extended boolean information retrieval', *Communications of the ACM*, Vol. 26, No. 11, pp. 1022-1036.

Salton, G., Wong, A. and Yang, C.S. (1975), 'A vector space model for automatic indexing', *Communications of the ACM*, Vol. 18, No. 11, pp. 613-620.

Schutze, H., Hull, D.A. and Pedersen, J.O. (1995), 'A comparison of classifiers and

document representations for the routing problem', *Proceedings of the 18th Annual International Conference on Research and Development in Information Retrieval*, Seattle, WA, July 9-13, pp. 229-237.

Singhal, A. (2001), 'Modern information retrieval: a brief overview', *IEEE Data Engineering Bulletin*, Vol. 24, No. 4, pp. 35-43.

van Rijsbergen, C.J. (1979), I*nformation Retrieval*, Butterworths, London, UK.

Voorhees, E.M. and Harman, D. (1999), 'Overview of the seventh text retrieval conference (TREC-7)', *Proceedings of the Seventh Text REtrieval Conference (TREC-7)*, Gaithersburg, MA, November 09-11, pp. 1-23.

Wei, C.P, Hu, P., Tai, C.H., Huang, C.N. and Yang, C.S. (2007), 'Managing word mismatch problems in information retrieval: a topic-based query expansion approach', *Journal of Management Information Systems*, Vol. 24, No. 3, pp. 269-295.

Wong, S.K. and Yao, Y.Y. (1992), 'An information-theoretic measure of term specificity', *Journal of the American Society for Information Science*, Vol. 43, No. 1, pp. 54-61.

Xu, J. and Croft, W.B. (1996), 'Query expansion using local and global document analysis', *Proceedings of the 19th Annual International Conference on Research and Development in Information Retrieval*, Zurich, Switzerland, August 18-22, pp. 4-11.

Xu, J. and Croft, W.B. (2000), 'Improving the effectiveness of information retrieval with local context analysis', *ACM Transactions on Information Systems*, Vol. 18, No. 1, pp. 79-112.

Yang, Y., Huang, Z., Shen, H.T. and Zhou, X. (2011), "Mining multi-tag association for image tagging', *World Wide Web*, Vol. 14, No. 2, pp. 133-156.

Zauder, K., Lazic, J.L. and Zorica, M.B. (2007), 'Collaborative tagging supported knowledge discovery', *Proceedings of the 29th International Conference on Information Technology Interfaces*, Cavtat/Dubrovnik, Croatia, June 25-28, pp. 437-442.

Zhai, Y., Liu, J. and Shah, M. (2006), 'Automatic query expansion for news video retrieval', *Proceedings of the IEEE International Conference on Multimedia and Expo*, Toronto, Canada, July 9-12, pp. 965-968.