多篇文件自動摘要系統

沈健誠，張俊盛

清華大學資訊工程研究所

mr884354@cs.nthu.edu.tw, jschang@cs.nthu.edu.tw

摘要

目前大部分的摘要系統為單篇文章摘要系統，雖然能提示個別文章的要點，卻無法把性質相近的文章集合成摘要。能否夠發展一個多篇文章摘要系統，將敘述相同事件的文章統合成一篇摘要？如此一來，兩三個句子就能把文章的文意清楚而簡潔的表達出來，讓使用者能在一分鐘之內，明瞭這幾篇文章是否符合資訊需求，以縮短其蒐集的時間，更有效率的吸收網路上的大量資訊。

我們的目標在於發展一個多篇文章摘要系統，系統所產生的摘要能滿足以下兩個條件：指示性簡單摘要，和查詢主題相關，能因應使用者的查詢而有所改變。

為了達成此目標，我們將探討句子的指示性和查詢主題相關性，並選出重要性高而且相互獨立的句子，然後將不重要的小句刪除，以得到最終摘要。

我們針對 NTCIR 的 248 篇文章和 50 個查詢標題作實驗，所得到的摘要縮減比率為 95%以上。整體而言，產生的摘要都能指示出幾篇相關新聞以及查詢主題的要旨。

1.簡介

1.1 研究動機與目的

隨著網際網路的蓬勃發展，網路資訊（如電子報，政府公告，企業概況報導）的數量與日遽增，常常會有不同文章描述相同事件的情況，造成讀者無謂的時間浪費。摘要系統能夠將文章濃縮成精華片段，讀者只要閱讀摘要，就能瞭解此篇文章所敘述的事件和目的。因此，好的摘要系統能夠縮短使用者的閱讀時間，讓使用者在短時間內閱讀更多文章，吸收更多有用的資訊，精確地得到所需的情報。

目前大部分的摘要系統，針對單篇文章提供摘要，無法把性質相近的文章集合成摘要，讀者仍然需要一篇一篇的篩選。如果能夠發展一個多篇文章摘要系統，將敘述相同事件的文章統合成一篇摘要，在兩、三個句子之內，把文章的文意清楚而簡潔的表達出來（提示性摘要），讓使用者能在一分鐘之內，明瞭這幾篇文章是否符合資訊需求，以縮短其蒐集的時間，更有效率的吸收網路上的大量資訊。

因此，我們的目標在於發展一個多篇文章摘要系統，它所產生的摘要能滿足以下兩個條件：

1. 數篇文章的綜合摘要，而且是最簡潔的提示性摘要，最好能夠用兩三句就將這幾篇文章的主題事件描述出來。在目前的工商業社會中，資訊成長的速度遠超過想像，所以記者都會把新聞都寫得儘量簡短，只描述單一事件，並把文章的重點集中在某一段；往往只要簡短的一兩句，就能將文章的意義表達清楚。本系統的目的，便是從文章中找出足以代表整篇文章的句子，經過修飾後即成為真正的摘要。
2. 和查詢主題（topic）相關，能因應使用者的查詢而有所改變，以符合使用者的真正需求。

1.2 摘要如何產生

目前的摘要產生方式有兩種：摘述法（Extraction）與重述法（Abstraction）。摘述法就是將文章中的關鍵句抽取出來，將其組合成為摘要；重述法則是做摘要的人將文章的要義以另外的文句寫下。目前大部分的摘要系統多採用摘述的方法，因為它比較簡單，只要句子取的好，就能達到指示文章的目標。

摘要的產生方向也有兩種：通用型（Generic）與查詢主題相關型（Topic-Related）。對通用型摘要而言，不管使用者所在意的問題為何，相同文章一律產生相同的摘要；它比較固定，而且作業上比較簡單。而查詢主題相關摘要，是由使用者的查詢（query 或 topic）顯示出的使用者關心的部分所構成。

由以上敘述可得知，查詢主題相關摘要較符合使用者的需求，而利用摘錄方法所產生的摘要能夠兼顧效率與準確性。

1.3 多篇文件摘要系統的相關研究

摘要系統的研究，已行之有年；隨著網際網路的蓬勃發展，其重要性日益增高，但大多數的摘要系統是為單篇文件而設。多篇文件摘要系統的研究在最近一兩年來，逐漸引起大家的重視。McKeown 和 Radev(1998)在它們的系統中除了地名，人名，組織名的辨識，還用新聞摘要語料庫，學習摘要的產生規則，最後利用文章合成技術（Text Generation）來產生摘要。

Mani 和 Bloedorn（1999）以分析，重組，合成三步驟來產生摘要，並利用 WordNet 來計算句子間的關係，進一步尋找出各篇文章間的相同和相異處；此外，他們還利用 Spreading Activation 演算法將與查詢主題相關的句子放入摘要中，進而產生查詢主題導向（Topic-Oriented）的摘要。

Chen 和 Huang（1999）將中國時報，中央日報，中時晚報及工商時報網站上的新聞，以 Complete-Link clustering（Salton, 1989）的方法，依事件分類，然後計算句子間的相似度，產生重點式與瀏覽式摘要。

有鑑於中文的 WordNet 尚未發展完全，目前也沒有中文新聞摘要的語料庫，所以我們希望能在沒有任何訓練及參考資料的情形下，以統計方式找出最合適的句子，並進一步將句子中不重要的部分剔除，完成最精簡的摘要。

2. 摘要的生成

2.1 摘要的條件

好的摘要應該簡潔、清楚，對文章有強烈的代表性及提示性。讀者光憑摘要，就能瞭解本文所描述的事件，輕易分辨這些文章是否為其所需。此外，摘要系統必須能從文章中挑出讀者想知道的部分；換句話說，摘要應該隨使用者的查詢需求而有所調整。

大部分的摘要系統，都是從文章中挑出關鍵句，將其合成摘要。選到適當的句子，摘要就成功了一半。評量句子的方法，大多為 tfidf 的變形，或是句子間的詞彙鏈結（Lexical Chain）。多篇文章摘要的來源文章，彼此之間都有一定程度的相似，因此本系統以句子間的相似度為主。

2.2 摘要產生步驟

大多數的摘要，都是以選句為基礎，本系統亦不例外；在選句之前，要先做好斷句的工作；除此之外，句子間的字彙鏈結是以句中的動詞和名詞為主，所以必須有良好的斷詞和詞性標注（Part-of-Speech Tagging）。為句子計算分數時，必須考量其提示性和查詢主題相關度。在選取句子的部分，我們希望摘要中的句子不但有高度的代表性，而且彼此獨立。最後，

我們將探討如何進一步把摘要中不重要的小句刪除，讓摘要更簡潔。

2.3 斷句斷詞

斷句方面，以句號（。），分號（；），驚嘆號（！）和問號（？）做為斷句的準則；換言之，一個完整的句子是以上述四個標點符號作結尾。而句子中的逗號（，）則為小句的分隔符號，在句長縮短的步驟中，我們考慮將不重要的小句剔除，進一步減少摘要的長度。

斷詞方面，則以 Chen（2000）所發展的斷詞工具，對句子作斷詞和詞性標注，以名詞和動詞，作為句子評分的主要依據。

2.4 句子評分（Sentence Scoring）

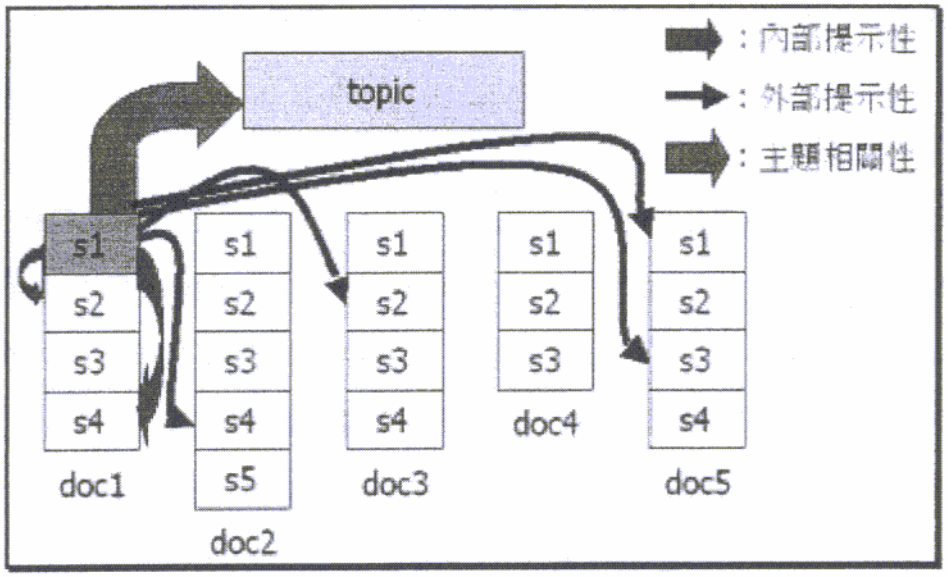
哪些句子才是關鍵句？很明顯的，為了回答這個問題，我們必須對句子作評分，決定何者為關鍵句；而關鍵句的考量有兩個方向：

1. 在此篇文章中的代表性：此句子是否足以代表本篇文章的意義。
2. 和查詢主題（topic）的相關聯程度：此句子是否和查詢主題有關係。

假設有一句子 S，其分數以兩方面來考慮：

A. 提示性（Indicativeness）：S 對其他句子的提示程度；對 S 所在文章內的其他句子的提示程度，稱為內部提示性，對其他相關文章內的句子提示程度稱為外部提示性。

B. 主題相關性（Topic Relevance）：句子與查詢主題的關聯程度。



圖一. 句子的提示性與查詢主題相關性

2.4.1 句子的提示性

要評斷一個句子提示其他句子的程度，最簡單的辦法便是計算兩句子的詞重複的數目。在大部分的情況下，句中的動詞和名詞（尤其是專有名詞）比較能夠代表整個句子的主要意義，因此在本系統中，我們以兩個字（含）以上的動詞與名詞作為計算相似度的基本單位；也就是說，如果兩個句子有一個以上的相同動詞或名詞，我們便說這兩個句子間有某種程度的相互提示性。

在本系統中，我們還將提示性分為內部提示性與外部提示性。內部提示性指的是此句對於同一文章內其他句子的提示程度，外部提示性則是此句對於其他文章內句子的提示程度。至於提示性的分數計算，我們採用三種方法：

2.4.1.1 方法 A1：

- 句子間的提示性 I_{ij}：兩句子 S_i 與 S_j 之間的相關係數，若 S_i 和 S_j 在同一篇文章內，稱為內部提示性；若 S_i 和 S_j 在同一查詢主題（topic）

的不同文章內，則稱之為外部提示性

$$I_{ij} = \frac{V_{ij} + N_{ij}}{V_i + N_i}$$

V_{ij}：S_j 和 S_i 內相同的動詞個數（重複出現不計）

N_{ij}：S_j 和 S_i 內相同的名詞個數（重複出現不計）

V_i：S_i 的動詞總數

N_i：S_i 的名詞總數

在此公式之下，如果兩句子內同時出現的動詞與名詞數越高，兩者之間便有越高的提示性。為了正規化（normalize），因此必須以 S_i 的動詞術語名詞數總和當分母；S_i 中和 S_j 相同的重要字詞比例，即為 S_i 對 S_j 的提示性。

- 句子 S_i 的提示性分數：假設 S_i 對 m 個句子（S_j，j＝1～m）有提示性，則 S_i 的分數為 I_{ij} 的總和。提示性又分為外部提示性（OI_i）和內部提示性（II_i），其計算方式如下：

$$OI_i = \sum_{j=1}^{m} I_{ij}$$，S_i和S_j在不同文章內

$$II_i = \sum_{j=1}^{m} I_{ij}$$，S_i和S_j和在同一篇文章內

例如：

S_1 ＝『為配合司法院大法官會議四四五號解釋，並進一步保障人民集會遊行自由，內政部在會商相關部會後，已研議完成「集會遊行法」部分條文修正草案，除刪除現行集遊法第四條「集會、遊行不得主張共產主義或分裂國土」的條文，同時也以更為具體、明確的文字，完備集會、遊行之申請採取許可制的相關規定。』

S_2＝『為配合司法院大法官會議解釋，並進一步保障人民集會遊行自由，內政部在會商相關部會後，已經研議完成「集會遊行法」部分條文修正草案。』

兩句共有 20 個相同動詞與名詞（其中"集會"在 S_1 出現四次，在實驗一中算成一個），而 S_1 的動詞和名詞總個數為 36，所以

$$I_{12} = 20 / 36 = 0.5556$$

2.4.1.2 方法 B1

- 句子間的提示性 I_{ij}：S_i 和 S_j 為相異兩句

$$I_{ij} = \frac{N_{ij}}{\sqrt{N_i N_j}} + \frac{V_{ij}}{\sqrt{V_i V_j}}$$

N_{ij}，V_{ij}：同時出現在 Si 和 Sj 中的名詞（動詞）總數

N_i，N_j：出現在 S_i（S_j）中的名詞數目

V_i，V_j：出現在 S_i（S_j）中的動詞數目

- 句子 S_i 的提示性：同實驗一，假設 S_i 對 m 個句子（S_j，j=1~m）有提示性，則 S_i 的分數為 I_{ij} 的總和

以 S_1 和 S_2 為例，S_1 和 S_2 之間的提示性為

$$I_{12} = \frac{8}{\sqrt{16*8}} + \frac{12}{\sqrt{20*12}} = 0.7071 + 1.0607 = 1.7678$$

2.4.2 查詢主題相關性（Topic Relevance）

本系統的目的是『產生與查詢主題相關的摘要』，因此除了考量句子本身對於文章的代表性，還要考慮句子跟查詢主題之間的相關性。我們以 NTCIR-2 查詢主題 Concepts 欄位中的詞為關鍵詞，並採用資訊檢索（Information Retrieval）的方式，來對句子的查詢主題相關性做評分。查詢主題的範例如下：

ID	SECTION	CONTENT
1	Number	CIRB010TopicZH001
1	Title	集會遊行法與言論自由
1	Question	查詢集會遊行法中有關主張共產主義或分裂國土規定之修正與討論。
1	Narrative	相關文件內容應敘述集會遊行法原本對主張共產主義或分裂國土之限制，其是否符合憲法中對言論自由等基本人權的保障，大法官對此議題的相關解釋，學者專家的討論與看法，以及集會遊行法條文的修改現況。
1	Concepts	集會遊行法、集會遊行、集遊法、憲法、言論自由、保障、共產主義、分裂國土、大法官會議、立法、修正條文。

表一. 查詢主題（topic）範例

在此步驟，我們參考資訊檢索的方式，採用了四個方法。

2.4.2.1 方法 A2

帶入如下公式

$$R_i = \frac{\#\ of\ (Terms\ in\ S_i\ \mathrm{I}\ Conecpt\ terms\ in\ topic)}{\#\ of\ (Concept\ terms\ in\ topic)} \times C$$

一個句子中出現了越多查詢主題內的關鍵詞，它和查詢主題的相關程度就越高。為了和句子的提示性分數平衡（I_i的值大多在 0～10 之間，R_i原始值在 0～1 之間），所以必須乘以常數 C。（本系統中，C的預設值為 10），例如：

S_2 =『爲配合司法院大法官會議解釋，並進一步保障人民集會遊行自由，內政部在會商相關部會後，已經研議完成「集會遊行法」部分條文修正草案。』

上述句子中包含了『集會遊行』,『集會遊行法』,『保障』,『大法官會議』四個<topic 1>的關鍵詞，而<topic 1>共有 11 個關鍵詞，因此其主題相關性的評分為 $R_2 = \frac{4}{11} * 10 = 3.6364$

2.4.2.2 方法 B2

柏克萊大學在 TREC-2（Text Retrieval Evaluation Conference）中提出一種機率統計式的文件評分方法，在 TREC-5（中文查詢）與 NTCIR-2（中文與英文查詢）都有不錯的效果。在本方法中，引用此公式來表達句子 S_i 與查詢主題間的關連程度，其公式如下：

$$
\begin{aligned}
R_i &= \log O(R \mid S_i, Q) + K \\
&\approx \log \frac{P(R \mid S_i, Q)}{P(\overline{R} \mid S_i, Q)} + K \\
&\approx -3.51 + 37.4 * X_1 + 0.330 * X_2 \\
&\quad + (-0.1937) * X_3 + 0.929 * X_4 + K
\end{aligned}
$$

$P(R \mid S_i, Q)$：S_i 和查詢（*Query*）相關的機率

$P(\overline{R} \mid S_i, Q)$：$S_i$ 和查詢（*Query*）不相關的機率

K：爲了使分數大於零，所加上的常數，在本篇預設值爲5

X_1, X_2, X_3, X_4：此公式的四個參數，計算方法如下

$$
\begin{aligned}
X_1 &= \frac{1}{\sqrt{N}+1} \sum_{i=1}^{N} \frac{qtf_i}{ql+35} \\
X_2 &= \frac{1}{\sqrt{N}+1} \sum_{i=1}^{N} \log \frac{dtf_i}{dl+80} \\
X_3 &= \frac{1}{\sqrt{N}+1} \sum_{i=1}^{N} \frac{ctf_i}{cl} \\
X_4 &= N
\end{aligned}
$$

N :< 文件 > 和 < 查詢 > 中相符的字詞數

qtf : 字詞在 < 查詢 > 中出現的頻率

ql :< 查詢 > 所包含的字詞總數

dtf : 字詞在 < 單一文件 > 中出現的頻率

dl :< 單一文件 > 所包含的字詞總數

ctf : 字詞在 < 所有文件 > 中出現的頻率

cl :< 所有文件 > 包含的字詞總數

由於此公式適用於較長的整篇文件，因此用於長度較短的句子上時，會有分數小於零的情況發生。為了和句子的提示性分數平衡，必須加

上一個常數 K（在本系統中，K 的預設值為 5）。

2.4.3 句子的分數：一個句子的重要性包含了內部提示性，外部提示性，與查詢主題相關性，計算公式如下：

$$score_i = r_{out} * OI_i + r_{in} * II_i + (1 - r_{out} - r_{in}) * R_i$$

OI_i：外部提示性

II_i：內部提示性

R_i：查詢主題相關性

r_{out}：外部提示性分數所佔的比例，介於 0～1 之間

r_{in}：內部提示性分數所佔的比例，介於 0～1 之間

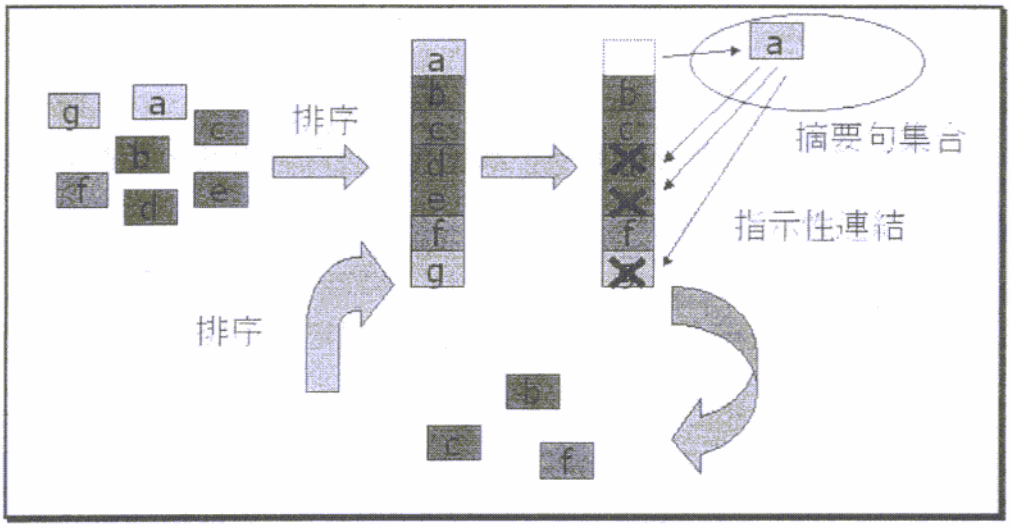
在本系統中，我們給予內部提示性，外部提示性與查詢主題相關度相同的重要性，因此 $r_{out} = r_{in} = 1/3$。

2.5 關鍵句選取（Key Sentence Selection）

良好的摘要句應該兼顧『提示性』和『相互獨立』的原則；被選入的句子本身能夠對其他句子有強烈的提示性，足以代表其他句子和文章，而句子相互之間的提示性越少越好，避免重複敘述的情況發生。為達此目的，使用如下的方法：

1. 將同一群組內不同文章的所有句子，依照句子的分數由大排到小，形成一個句子名單（list）
2. 把分數最大的句子 a 從名單中取出，放入摘要句集合中
3. 與 a 有一定程度相互提示性（大於某個下限值）的句子，從名單中移除
4. 重複步驟 1，將剩下的句子重新排列選取

如此一來，便可以選出具有高提示性又相互獨立的句子。



圖二. 選句方法示意圖

2.6 句子縮短（Sentence Reduction）

為了進一步縮短摘要長度，減少讀者閱讀摘要的時間，我們必須將句子中無關緊要的部分刪除。通常一個句子內包含了數個小句，以逗號(，)分隔之。如果某小句符合下列兩條件，則將其從摘要中剔除：

1. 毫無標題性：標題通常能代表文章的主要文意；如果以某字詞為開頭的小句常常出現在標題中，即具有標題性。若某小句的開頭為字詞 W，則其標題性計算公式如下

$$T_W = \frac{\text{標題內的小句以W開頭的次數}}{\text{所有文章（包含文章標題）內的小句以W開頭的次數}}$$

在此我們以 NTCIR-2 的所有文章（共十三萬篇）為訓練資料，將所有小句的標題性算出，當小句的 $T_w \approx 0$ 時，表示毫無標題性，可視為不重要的小句。

2. 與查詢主題不相關：如果小句內不包含任何一個查詢主題的關鍵詞（Concepts 欄位內的字詞），則認為與查詢主題不相關。

換言之，如果一個小句的標題性趨近於 0，且不包含任何一個查詢

主題關鍵詞，則將其從摘要中刪除之。除此之外，某些格式的小句可以直接刪去：

1. 括弧內的字：如【xxxxx】、（xxxxx）等文字。大部分的新聞文件，會在開頭標示『【媒體/地名/記者名/報導】』（例如『【記者賴廷恆台北報導】』）。此類句子跟文章的意義完全無關，可以直接刪除。而小括號內的文字往往做為前面名詞或動詞的註釋（例如『來自全世界各地對「金學」（即指對金庸小說的研究）學有專精的學者專家』）或是將相對時間修正為絕對時間（例如『馬友友昨（八）日展開忙碌的行程』），一般而言，容易造成重複敘述的情況，違反了摘要的精簡原則，必須剔除。
2. 三個字以下的小句：如『畢竟，但是，因此，其次，基本上，事實上，不過，此外，另外，其實……』等小句，往往是連接詞或語氣詞，作為前後句的接續之用，沒有實質的意義。雖然刪除之後偶而會造成文句的不通順，但是對於摘要的『提示性』毫無影響，因此我們可以放心的刪除之。四個字以上的小句可能會包含『某人說』（如『卡特說』），『某單位表示』，或內含動詞名詞的小句，因此予以保留。

例如：第六篇～第十篇（和<查詢主題 2>相關）原本的摘要如下：

> 既然台灣已經是一個主權獨立的國家，何以民進黨還要追求獨立呢？林義雄並爲維持現狀下一定義就是不發生戰爭，不刺激中共打台灣，如果改國號會引發中共的攻擊，美國的反對，就是破壞現狀，當然會慎重處理。在年底立委選舉時，林義雄認爲，各政黨都不可能迴避統獨議題，否則是不負責任，至於此議題對民進黨的利弊，林義雄表示難以判斷，但他相信，維護台灣安全的政黨，是能得到人民的支持，這方面民進黨比其他政黨更易得到人民的肯定。

上述摘要文章中，以黑體字標注的這幾個小句，不包含查詢主題的

關鍵詞，而且以『就是』、『當然』、『但』、『否則』為開頭的小句完全沒有在標題出現過，所以這幾個小句被視為不重要，可刪除之。經過句子刪除後，產生的摘要如下：

> 既然台灣已經是一個主權獨立的國家，何以民進黨還要追求獨立呢？林義雄並爲維持現狀下一定義就是不發生戰爭，不刺激中共打台灣，如果改國號會引發中共的攻擊，美國的反對。在年底立委選舉時，林義雄認爲，各政黨都不可能迴避統獨議題，至於此議題對民進黨的利弊，林義雄表示難以判斷，維護台灣安全的政黨，是能得到人民的支持，這方面民進黨比其他政黨更易得到人民的肯定。

經過小句刪除作業後的摘要比較簡短，而且大部分的情況下，文章尚屬流利，也不影響指示原文的效果。

3. 實驗資料與結果

3.1 實驗資料

本系統的目的為產生查詢主題相關連（topic-related）的多篇文件摘要，所以實驗資料必須包含文章與查詢主題（topic）的詳細描述。本系統專門針對中文查詢與文章，因此我們使用 NTCIR-2 的中文資料，針對每個查詢主題（topic），從大會公佈的標準答案中取出五篇文章；第 30 和 45 個查詢主題的標準答案只有四篇，所以實驗用的文章總篇數為 248 篇。

3.2 比較式評估

本文以人工評估，比較幾種計分方法所產生的摘要優劣及長度。比較項目包含了提示性的計分方式，查詢主題相關性的計分方式，關鍵詞是否分割成覆疊性雙連字，以及小句刪除之後的滿意度。

3.2.1 提示性計算方式的差異

我們在計算提示性時，採用了兩種不同計分方式。方法 A1 和方法

B1 的比較結果如下：

	較好的查詢主題（topic）編號	50 篇摘要總長度（字數）
方法 A1 較佳	1,3,4,6,9,10,11,12,14,17,20,24,26,30,31,32,33,34,37,39,42,48,50（共 23 篇）	（方法 A1）8073
方法 B1 較佳	2,5,7,8,13,16,18,19,23,25,27,28,35,38,40,41,43,44,49（共 19 篇）	（方法 B1）8582
相同	15,21,22,29,36,45,46,47（共 8 篇）	

表二 提示性的計算方式比較

在大部分的情況下，方法 A1 所選出來的句子較短，而且提示性和查詢主題相關性較高。

以 **topic 34**—威而剛的副作用為例，方法 A1 和方法 B1 所產生的摘要各為：

<方法 **A1**>：輝瑞並警告說，因胸痛而服用含有磷酸鹽藥物的病人，若再服用威而鋼甚至會致命。報導指出，目前已有三十名男子因爲服用陽痿治療藥「威而鋼」不幸死亡，另外還有七十人發生嚴重副作用。輝瑞藥廠同時強調，全世界已有至少一百萬人服用過「威而鋼」，其中大多數是中年男性，該廠將與衛生部密切配合，就極少數死亡案例進行調查。

<方法 **B1**>：食品藥物管理局及輝瑞藥廠都表示，他們正在調查六名使用者死亡的原因，食品藥物管理局的聲明中說：「我們仍然相信這種藥物對於它的病症及病人安全有效」。報導指出，目前已有三十名男子因爲服用陽痿治療藥「威而鋼」不幸死亡，另外還有七十人發生嚴重副作用。

在上述例子中，兩個摘要都有提到查詢主題的主要意義，但是方法 A1 所產生的摘要的字數較少，是比較好的摘要。

雖然在某些情況下，方法 B1 的摘要較好，但是所用的字數卻比較多，例如 **topic 7**—卡特訪台：

<方法 **A1**>：二十年來，美國和臺灣、美國和北京及臺灣和北京的關係，都有很大的進步。呂秀蓮認爲，卡特應爲台灣民主運動的受挫負責，並要求卡特就此對台灣人民道歉。卡特並認爲每個國家都有自己的問題，美國不應該負這個責任。

<方法 **B1**>：卡特在昨日的離華記者會上仍然重申他當初決定與台灣斷交並沒有錯誤，他表示他在來台之前心理有些害怕，但是他認爲要讓台灣的民眾了解爲什麼他當初要做這個決定，現在雖然有許多人對於斷交的決定持不同的看法，但是卡特仍認爲這個決定增進了美、中、台三方的關係。卡特說，中共當初承諾，要以和平方式解決兩岸問題，他不認爲中共未來會對台灣採取軍事行動，他並重申，兩岸問題應由兩岸以和平方式解決，外人不適合，也不應該介入。

上述兩摘要所描述的都是卡特訪台的相關事件，雖然著重之處不同。方法 B1 所產生的摘要篇幅較大，因此較方法 A1 的摘要詳盡。

由表二及上述例子可得知，方法 A1 和方法 B1 所形成的摘要滿意度相差不大（23：19），但是方法 A1 產生出的摘要篇幅縮得較短（8073：8582 個字）；因此，方法 A1 的摘要較合乎我們的需要。

3.2.2 查詢主題的相關性比較

接下來，我們針對兩種查詢主題相關性的評分方法做比較（關鍵詞直接使用，未拆散成覆疊性雙連字）。

	較好的查詢主題（topic）編號	50 篇摘要總長度（字數）
方法 A2 較佳	1,10,12,13,14,18,20,23,24,25,32,34,38,42,43,44,47,48,49,50（共 20 篇）	（方法 A2）6530
方法 B2 較佳	4,16,17,28,29,30,35,37,41（共 9 篇）	(方法 B2) 6902
相同	2,3,5,6,7,8,11,15,19,21,22,26,27,31,33,36,39,40,45,46（共 21 篇）	

表三 查詢主題相關性計算方式的比較

從評估結果中發現，方法 A2 的滿意度遠比方法 B2（利用 Berkeley 的 IR 公式）來得好（20:9），可能因為 Berkeley 的 IR 公式，是專門處理字數較多的整篇文章，對於長度較短的句子而言，容易造成分數低落（小於 0）的情況發生；雖然最後的分數加上一個常數 K，但這種齊頭式平等的加分法，會造成分數的不平衡，真正和查詢主題相關的句子，因而無法被突顯出來，較長的句子可能會比較佔便宜。由滿意度的比較和篇幅長短的比較，我們認為方法 A2 形成的摘要較符合我們的需求。

3.2.3 句子縮短的影響

接下來我們針對句子縮短前後的摘要，比較其字數與句子縮短後的接受程度。

	較好的查詢主題（topic）編號	50 篇摘要總長度（字數）
縮短前較佳	10,21,30,31,47（共 5 篇）	（縮短前）8073
縮短後較佳	3,4,6,8,9,11,12,13,14,15,16,17,18,19,20,24,25,26,30,33,34,35,36,38,39,40,44,45,48,50（共 30 篇）	（縮短後）6902
相同	1,2,5,7,22,23,27,28,29,37,41,42,43,46,49（共 15 篇）	

表四 句子縮短前後的比較

在五十個查詢主題的摘要中，有五個被認為是縮短之後結果變差的。雖然被刪除的小句不多，但通常刪去之後會造成整個句子的不連貫，使句子本身變得不清楚；這種情況通常出現在句子前後有邏輯論證的關係時，例如 topic 10—庫藏股制度，原本的摘要為：

庫藏股制度究竟是利多，還是利空？未來，大股東可以決定動用公司資本公積及保留盈餘進場買回自家股票，最高可達一○%，形成變相減資； 不過，由於子公司買賣母公司股票完全不必規範，使得這套防弊措施能否奏效尚待觀

察。

第三句的重點在於後半句，因此在後半句被刪除的情況下，第三句的前半段就變得較不清楚，而且和第二句有重複現象，喪失了本身的獨立性和提示性。第二句則是刪除了前半句形成的後果，使得一般人無法瞭解此事件會形成什麼樣的影響，和事件本身的重要性。

句子縮短後的摘要，達到了 14.5%的縮減率，和九成的滿意度，效果相當良好。未來我們希望將刪除的範圍從小句擴大到字詞，並做一些小句間的關連分析，希望在追求縮減率的同時，也能將句子的原意完整的保留下來。

3.2.4 斷詞 vs. 雙連字

另外，我們來比較另外一種查詢主題相關性評分方式的差異；對於查詢主題關鍵詞，未經處理而直接使用，與拆成覆疊性雙連字再套入查詢主題相關性的計算方法，有什麼樣的不同？哪一種方式比較好？

我們針對查詢主題相關性的計算方法：使用關鍵詞，和將關鍵詞拆成覆疊性雙連字（Overlapping Bigram）做比較，其結果如下：

	較好的查詢主題（topic）編號
斷詞（方法 A2）	1,3,7,11,22,26,34,35,39,45,46,49
拆成雙連字（方法 A2）	4,9,12,14,17,18,19,20,24,33,38,44,47
相同	2,5,6,8,10,13,15,16,21,23,25,27,28,29,30,31,32,36,37,40,42,43,48,50

表五 關鍵詞是否拆成覆疊性雙連字的比較

在此我們發現，有一半以上的摘要是相同的，其他互有優劣的摘要在選取的三句中往往有一兩句相同，即使完全不相同者，兩邊的滿意度

都算可以接受。以 **topic 17**—流浪狗問題爲例：

<關鍵詞>：動物保護法從去年十月底立法實施以來，已撥出五千三百萬元經費給各縣市政府改善收容所設備，而且也提供巴比妥酸鹽給收容所，進行流浪犬安樂死的人道處理。對於臺南縣政府發生溺斃、電殛流浪犬收容所作法，農委會依動物保護法的規定，中央主管機關並不能開出罰單，只能要求地方政府對行爲人開出五萬元的罰單。由農委會及省農林廳補助六百萬元興建的「花蓮縣吉安鄉流浪犬中途之家」，第一期硬體工程已經完工。

<雙連字>：動物保護法從去年十月底立法實施以來，已撥出五千三百萬元經費給各縣市政府改善收容所設備，而且也提供巴比妥酸鹽給收容所，進行流浪犬安樂死的人道處理。對於臺南縣政府發生溺斃、電殛流浪犬收容所作法，農委會依動物保護法的規定，中央主管機關並不能開出罰單，只能要求地方政府對行爲人開出五萬元的罰單。行政院農業委員會決定全面整頓流浪犬處理問題並將興建現代化的流浪犬中途之家，同時統一採購流浪犬安樂死用麻醉劑，提供給地方政府人道處理流浪犬之用。

因此，我們可以得到一個結論：不論關鍵詞有無被拆成覆疊性雙連字，結果都差不多。在 NTCIR-2 的探討會議中有提到：因為關鍵詞整理得太好，所以直接用關鍵詞來做資訊檢索（Information Retrieval）的查詢反而比利用其他欄位（如 Title，Question，Narrative 等，請參閱表一）的其他方法好。在此也是相同的情形，所以在查詢主題有很多關鍵詞的情況下，似乎不需要多此一舉，拆成覆疊性雙連字。不過除了 NTCIR-2 以外，網路搜尋引擎收到的查詢主題往往只有兩三個關鍵詞；此時，將系統找到的關鍵詞拆成雙連字，可能會有一定的改進。也可避免斷詞的錯誤所造成的負面影響。

4. 結論

4.1 主題相關的多篇摘要

本篇論文提出了一個多篇文件自動摘要系統，在此系統中我們利用簡單的斷詞斷句工具，以 NTCIR-2 的文章和查詢主題為實驗對象，針對句子的提示性和查詢主題相關性作分析，選出富有提示性又相互獨立的句子，將句子內不重要的小句刪除後，產生提示性摘要。

本篇摘要在提示性的計算上，採用了方法 A1；在查詢主題相關性的計算方面，採用方法 A2；不重要的小句也已經刪除。原文共有 197423 個字，五十篇摘要的總字數為 6902 個字，平均每篇 138 字；壓縮比為 3.5 %。

小句的刪除大致上不影響『提示性摘要』的需求，只有少數情況下，如 **topic 32**—腸病毒的摘要，效果不太理想。刪除後，摘要縮得太短。為此我們設定了字數下限，以保持足夠的指示性。當摘要的字數小於某一個值時（例如：60 字），則停止刪除小句的動作，以免刪除過多小句，喪失摘要的原意。如此摘要便能維持較適宜的狀況。

4.2 未來研究方向

本系統在詞彙的比對上，只使用了最簡單的斷詞和句子間的詞彙比對。然而文章內的句子寫法千變萬化，不同的語詞，不同的語法，可能形成相同的意念。在黃聖傑（1999）的多篇摘要研究中，試圖利用同義詞來處理這些現象，卻發現效果不大。這是因為很多新的專有名詞不會出現在辭典中，偏偏這些新詞往往是文章的重心所在。如果要解決這問題，可能需要作更深入的語意分析。

另外，目前的摘要作法還存在一個很大的問題：各摘要的篇幅差距頗大（52～349 字之間，標準差為 63.8），可能的原因為文章中句子的長

短不一，因此我們必須做長句分割，在語句停頓的地方將逗號改成句號（You Yu-Ling）。此外，也希望將句子縮短的步驟作得更徹底，除了小句之外，不重要的詞也可以一併刪除；更進一步。將長詞換成同等意義的短詞（例如：行政院長→閣揆，清華大學→清大）。

本文的研究，主要是針對新聞為主，因為一篇新聞往往只描述一個事件，有利於摘要的形成。我們希望除了新聞外，也能將這些方法推廣到其他文體（例如技術報告），產生具有指示性的多篇的摘要。

參考文獻

1. Chinatsu Aone, Mary Ellen Okurowski, James Gorlinsky. 1998. Trainable, Scalable Summarization Using Robust NLP and Machine Learning. In 36th Annual Meeting of the COLING-ACL, pp. 62-66.
2. Mark Wasson. 1998. Using Leading Text for News Summaries: Evaluation Results and Implications for Commercial Summarization Applications. In 36th Annual Meeting of the COLING-ACL, pp. 1364-1368.
3. Regina Barzilay, Karthleen R. McKeown and Elhadad. 1999. Information Fusion in the Context of Multi-Document Summarization. In 37th Annual Meeting of the ACL, pp. 550-557.
4. Adam Berger, Vibhu O. Mittal. 2000. Query-Relevant Summarizations using FAQs. In 38th Annual Meeting of the ACL, pp. 294-301.
5. Hongyan Jing. 2000. Sentence Reduction for Automatic Text Summarization. In Proceedings of the 6th ANLP / 1st NAACL, Section 1, pp. 310-315.
6. Hongyan Jing and Kathleen R. McKeown. 2000. Cut and Paste Based Text Summarization. In Proceedings of the 6th ANLP / 1st NAACL, Section 2, pp. 178-185.
7. Inderjeet Mani, Eric Bloedorn. 1999. Summarizing Similarities and Differences Among Related Documents. In Information Retrieval, Vol. 1, pp. 35-67.
8. Weiquan Liu and Joe Zhou. 2000. Building a Chinese text summarizer with phrasal chunks and domain knowledge. In ROCLING XIII, pp. 87-96.
9. D. R. Radev and K. R. McKeown. 1998. Generating Natural Language Summaries from Multiple On-line Sources. In Computational Linguistics,

Vol. 24, No. 3, pp. 469-500.

10. Yu-Jin Chen. 2000. Scalable Summarization for Chinese Text. National Tsing-Hua University, master thesis.
11. Yu-Ling You. 2000. Toward Defining Discourse Unit in Chinese Discourse. In Language researching and teaching. (in press)
12. 黃聖傑, 1999. 多文件自動方法摘要研究. 台灣大學資訊工程研究所碩士論文，台北.
13. 楊允言，謝清俊，陳淑美，陳克健. 1992. 中文文件自動分類之研究. 中華民國八十二年第六屆計算語言學研討會論文集，pp.217-233.
14. 楊允言，張俊盛，陳克健. 1993. 文件自動分類及其相似性排序. 清華大學資訊科學研究所碩士論文，新竹.