中華民國計算語言學學會

通訊 ACLCLP

發行人：陳信宏
主　　編：曾淑娟
執行編輯：黃　琪

The Association for Computational Linguistics and Chinese Language Processing 第十七卷第一期

E-Mail：aclclp@hp.iis.sinica.edu.tw
地 址：台北市研究院路二段128號中研院資訊所
Website : http://www.aclclp.org.tw
電 話：(02)2788-3799 ext:1502
傳 眞：(02)2788-1638
劃 撥：19166251

本 期 要 目

「中央研究院口語韻律語料庫暨工具平台」開放申請

「中央研究院口語韻律語料庫暨工具平台」(Sinica Continuous Speech Prosody Corpora & Toolkit，簡稱 COSPRO & Toolkit)，係中研院語言所語音實驗室主持人鄭秋豫教授多年語流韻律研究之心血結晶(1994-2005)。基於學術資源共享之理念，以期望能促進語音科學研究與技術能有突破性發展之初衷，釋出本語料庫與工具平台。其智慧財產權爲中央研究院所有，授權民間公司—艾爾科技公司（L Labs Inc.）發行，供國內外學術或民間機構非營利使用。

COSPRO 包含九個子語料庫（共 10.5GB），COSPRO 01-08 屬於麥克風朗讀語音，COSPRO 09 則爲麥克風自發性語音（76MB）。內容包羅了不同長度的語料，短至孤立詞組(1-4 字詞)，長至段落語篇(85-996 音節)。由 111 位語者因不同的研究目的錄製而成，其中包括 61 位女性，50 位男性。

釋出語料庫除了 wav 檔案之外，還包括了每位語者的朗讀（轉寫）文本（*.txt）、經人工調整過後的音節檔（*.adjusted / *.syl）——標記對應語音之音節邊界（子音與母音）之時間碼以及由受過訓練之標音員所標記之停延韻律標記檔（*.break）。可進行語音研究、語音合成與語者辨識等多方面應用。

另一方面，COSPRO Toolkit 係一視窗介面（Window-based）、好操作（user-friendly）的語音分析暨合成之工具平台，集合了 Adobe® Audition®, Praat© and Speech Viewer®等常見語音分析(合成)軟體之特點，其主要功能有：聲學訊號分析功能、標記口語語流功能以及重新合成語音訊號（re-synthesizing speech signals）功能。

欲知進一步申請資訊請參閱COSPRO專屬網頁：http://www.myet.com/COSPRO

學生出席國際會議獎助

獎助會議：

COLING、ACL、ACM SIGIR、ICASSP

獎助說明：

- 申請人須同時具備下列資格：
  1. 被接受論文之第一作者（指導教授不計）。
  2. 本會會員。
  3. 投稿時爲國內在學學生。
- 獎助金額：由審查委員會依地區別及論文等級審定獎助金額，每名獎助金額上限爲美金 1,000 元
- 獎助名額：每個會議補助一～二名。

申請辦法：

1. 日期：論文被接受發佈日二週內提出。
2. 手續：申請人需將論文接受函、審查意見、學生證、論文全文及申請書等相關資料郵寄至本會秘書處。

受獎助人義務：

1. 出席發表論文。
2. 論文全文必須以書面同意投稿至本會期刊。
3. 代學會攜去宣傳品及帶回相關資料。

THE 5TH INTERNATIONAL SYMPOSIUM ON CHINESE SPOKEN LANGUAGE PROCESSING

第五届中文口语语言处理国际会议

December 13-16, 2006

http://www.iscslp2006.org

CSLP 2006 UNIQUELY Singapore

Call for Papers

ISCSLP'06 will be held during December 13-16, 2006 in Singapore hosted by the Institute for Infocomm Research (I²R) and the Chinese and Oriental Languages Information Processing Society (COLIPS). This is the 8th year after its inaugural event in Singapore and we welcome ISCSLP back to her birthplace.

Singapore, popularly known as "The Garden City", is situated at the southern tip of the Malaysian Peninsula in South-East Asia and has a rich and interesting history tracking back to 1819 when the British started a trading post which later developed into an important commercial and military imperial base. Singapore is a small but prosperous cosmopolitan state diversified with 4 main ethnic groups, namely, Chinese, Malays, Indians and Eurasians. With museums exhibiting rich collections of historical information and relics, fun theme parks, zoos and night safaris, bustling shopping and dining heavens, Singapore provides an interesting stop to unwind from the daunting stress of today's society.

We invite your participation in this premier conference, where the language from ancient civilizations embraces modern computing technology. The ISCSLP'06 will feature world-renowned plenary speakers, tutorials, exhibits, and a number of lecture and poster sessions on the following topics:

- Speech Production and Perception
- Phonetics and Phonology
- Speech Analysis
- Speech Coding
- Speech Enhancement
- Speech Recognition
- Speech Synthesis
- Language Modeling and Spoken Language Understanding
- Spoken Dialog Systems
- Spoken Language Translation
- Speaker and Language Recognition
- Indexing, Retrieval and Authoring of Speech Signals
- Multi-Modal Interface including Spoken Language Processing
- Spoken Language Resources and Technology Evaluation
- Applications of Spoken Language Processing Technology
- Others

Official Language & Publication

- The official language of ISCSLP is English.
- The regular papers will be published as a volume in the Springer LNAI series.
- The poster papers will be published in a companion volume.

Paper Submission

- Authors are invited to submit original, unpublished work in English.
- Papers should be submitted via http://www.iscslp2006.org.
- All submissions should be formatted in accordance with the Springer print style at: http://www.springer.de/comp/lncs/authors.html and no more than 12 pages.
- Each submission will be reviewed by two or more reviewers.
- At least one author of each paper is required to register.

Schedule

Full paper submission by	Jun. 15, 2006
Notification of acceptance by	Jul. 25, 2006
Camera ready papers by	Aug. 15, 2006
Early registration	Nov. 10, 2006

Organizers

Sponsors

Call for papers

International Symposium on Chinese Spoken Language Processing (ISCSLP'2006),

Special Session on Robust Techniques for Organizing and Retrieving Spoken Documents

Singapore, Dec. 13-16, 2006

Multimedia data containing speech are considered "spoken documents". As the cost of storage deceases and the bandwidth of communication increases, there has been a rapid growth of multimedia information on the Internet or in online archives. There is increasing interest in efficiently and effectively processing speech content for multimedia indexing and search. Therefore, organizing and retrieving spoken documents is becoming a research area of prime importance. This special session aims to solicit papers covering all aspects of the technologies and systems in the area. The topics cover, but are not limited to, the following:

- Spoken document segmentation, clustering and transcription
- Spoken document indexing and retrieval
- Spoken document summarization and information extraction
- Spoken document understanding, organization and visualization
- Crosslingual and multilingual spoken document retrieval
- Spoken dialogue systems for accessing spoken archives
- Spoken document corpora
- Performance evaluation metrics

Please submit your papers to the special session chair Dr. Hsin-min Wang (whm@iis.sinica.edu.tw) before June 15, 2006. For details about ISCSLP'2006, please refer to http://www.iscslp2006.org/.

International Symposium on Linguistic Patterns in Spontaneous Speech

November 16-17, 2006
Centre for Academic Activities, Academia Sinica, Taipei

Organized by
Institute of Linguistics, Academia Sinica

We would like to draw your attention to the International Symposium on Linguistic Patterns in Spontaneous Speech, which will be taking place on November 16-17, 2006 in Academia Sinica, Taiwan. This symposium will focus on research results on the processing, analysis and understanding of the linguistic characteristics of spontaneous speech. And it will also provide an excellent opportunity for linguists, speech engineers, and practitioners of natural language processing and system developers to share with each other the latest thinking and results on spontaneous speech. With no restriction on the languages under investigation, we are interested in contributions concerned with phenomena in all languages. Selected papers from the symposium will be published in a book in the Language and Linguistics Monograph Series by the Institute of Linguistics, Academia Sinica. For further details about this symposium, please visit our website at http://www.lpss.sinica.edu.tw.

We are looking for papers on the following topics:

- Corpora of spontaneous speech
- Linguistic analyses on spontaneous speech
- Characteristics of disfluency in spontaneous speech
- Understanding spontaneous speech in dialogue systems
- Automatic speech recognition of spontaneous speech
- Automatic or manual labelling of spontaneous speech data and
- Any other topics related to spontaneous speech

Important dates:

Deadline for full-paper submission	July 31, 2006
Acceptance notification to authors	September 15, 2006
Deadline for camera-ready paper	October 15, 2006
Symposium dates	November 16-17, 2006

Invited speakers:

Dafydd Gibbon	Universität Bielefeld
Kikuo Maekawa	National Institute for Japanese Language
Elizabeth Shriberg	SRI and ICSI

Organization: Shu-Chuan Tseng

自然語言問答系統的製作與研究

許聞廉
中央研究院資訊科學所
hsu@iis.sinica.edu.tw

前言

早在 1961 年，Green [4]就發展了第一個問答系統 (Question Answering System)，用來回答單季美國大聯盟相關比賽問題。該系統執行於 IBM 7090 平台，以今日的觀點來看，其硬體資源相當貧乏，但由於問答的範圍狹小，系統正確率尚能令人滿意。近年來，網際網路成長快速，在資訊、流量、使用人數、以及應用領域上都有驚人的發展。截至目前爲止[5]，Google 已經索引了超過八十億個網頁資料；MSN BETA、Yahoo 也分別有四十億與二十億個網頁資料。整個網際網路總索引量則高達一百一十五億個網頁，頗爲驚人。此趨勢帶動了近幾年問答系統的研究風潮，盼能解決網路搜尋如大海撈針的困境。報導指出[2]，從 2000 年到 2005 年，網路人口成長了 1.7 倍，在前十名的語言中，中文人口成長率爲 284.8%，高達一億兩千萬，遠遠超過英文人口的成長率。此數據顯示了中文處理的重要性，這也是本實驗室致力於相關研究的動機之一。

從問答系統外部的行爲上來看，其與目前主流資訊檢索技術有兩點不同：首先是查詢方式爲完整而口語化的問句，再則其回傳的答案爲高精準度網頁結果或明確的字串。以 Ask Jeevs [1]這個搜尋引擎爲例，使用者不需要思考應該使用甚麼樣的關鍵詞才能夠得到理想的答案，只需要用口語化的方式直接提問如「請問誰是美國總統？」即可。而系統在瞭解使用者問句後，會非常清楚地回答「布希是美國總統」。當這種系統的精準度夠高時，使用者不再需要像以往一樣費心去一一檢視搜尋引擎回傳的網頁，對於資訊檢索的效率與資訊的普及都有很大幫助。從系統內部來看，問答系統使用了大量有別於傳統資訊檢索系統自然語言處理技術，如自然語言剖析 (Natural Language Parsing)、問題分類 (Question Classification)、專名辨識 (Named Entity Recognition)等等。少數系統[7]甚至會使用複雜的邏輯推理機制，來搜尋答案。在系統所使用的資料上，除了傳統資訊檢索會使用到的資料外(如字典)，問答系統還會使用本體論 (Ontology) 等語意資料，或者利用網頁來增加資料的豐富性。

問答系統的分類

我們可以從知識領域、答案來源等角度來替問答系統做分類。從知識領域來看，可分爲「封閉領域」以及「開放領域」兩類系統。封閉領域系統專注於回答特定領域的問題，如醫藥或特定公司等。由於問題領域受限，系統有比較大的發揮空間，可以導入如專屬本體論等知識，或將答案來源全部轉換成結構性資料，來有效提升系統的表現。開放領域系統則天文地理無所不問，系統中所有知識與元件都必須儘量做到與領域不相關，難度也相對地提高。

若根據答案來源來區分，可分爲「資料庫問答」、「常問問題問答」、「新聞問答」、「網際網路問答」等系統。資料庫是最常見的結構化資料儲存媒介。雖然透過操控 SQL 語言便能夠有效率地存取資料，但有些系統試圖提供更直覺的自然語言查詢介面，希望能進一步降低學習門檻。1970 年代的 LUNAR 系統[15]算是早期成功的案例，其正確答題率可以達到百分之七十，可回答月球隕石相關資料。微軟的 English Query[6]則是近期的一個商業產品。English Query 在剖析完英文問句後，會根據底層資料庫結構，自動產生出相對應的 SQL 查詢。雖然有這些成功系統案例，但資料庫問答系統似乎很難被大眾所接受，其中一個因素可能是因爲對於結構化資料來說，結構化的查詢介面在查詢上更爲方便。常問問題(Frequently Asked Questions, FAQs)是公司或者長期經營領域中常見的重要資源。一份 FAQ 資料包含了一個問句以及相對應的答案描述。FAQ 問答系統的主要責任在比對使用者問句與現有 FAQ 問句的相似度，此與其他問答系統著重在答案語料中擷取答案的作法不同。另一種重要的系統爲新聞問答系統。今日新聞媒體都已經數位化了，每日累積所產生的新聞資訊量是相當可觀的，加上新聞的內容廣泛豐富，作爲開放領域問答系統的答案來源是

最適合不過的。這樣的特性使得此類系統的評估較爲容易，因此稍後提到的國際評估會議都是採用此類系統作爲評估對象。最後一類的是網際網路問答系統，這些系統利用搜尋引擎回傳的結果網頁，從中擷取答案。主要挑戰在於如何處理網路眾多異質性的資料，以及高雜訊網頁過濾等問題。

問題類型

問答系統接受的是自然語言問句，爲了有效控制研究變因，多會訂定可接受的問題類型來引導研究範圍。最基本的類型爲「仿真陳述問答」(Factoid Question Answering)，此類系統根據答案語料所述資訊，取出一小段字串作爲答案。由於答案的正確與否是根據答案語料的內容來決定，在現實生活中不一定爲真，故稱爲仿真陳述問答。有些系統把問答範圍進一步縮小，限定在人、地、組織等明確的專有名詞上。若此類系統有能力回答如「請列舉美國歷屆總統」這種清單型的問句，則稱爲「清單問答」(List Question Answering)；若能回答定義問題，則稱爲「定義問答」(Definition Question Answering)；依此類推還能定義出其他類型的問題。除了這些與問句資訊內容有關的類型外，最近評鑑會議引進如「時間限制問題」(Temporally Restricted Questions)與「序列問題」(Series of Questions)等複雜的問題類型。時間限制型的問題會在問句中明確指出答案的時間範圍限制，比如說以「民國九十年時的國民黨主席是誰」這問句來說，系統必須有根據答案語料結構化資料，或上下文來推論正確答案的能力。序列問題則把問答系統未來的應用定位在互動式的系統上。經過來回多次問答的方式來滿足使用者的資訊需求。瞭解這些問題類型分類，有助於研究範圍的界定，同時在分析比較上也較有依據。

國際性評估會議

截至目前爲止，世界主要語言都有問答系統發表在文獻上，甚至還有少數跨語言的案例。在過去問答系統的研究中，所有研究都是在各自的假設下進行，加上系統複雜度高，不同單位的研究成果很難拿來做客觀的評估與比較。除此之外，這類系統的評估是非常消耗人力的，事前的準備包含要產生足夠多且合適的問題題目，同時每一題可能出現的答案都必須以人工方式從比賽語料中挑選出來。以上種種對問答系統的研究發展都產生許多不確定因素。有鑑於此，由單一組織舉辦、多個研究單位共同參與的問答系統比賽應運而生。

英文問答系統早在1999年就開始由TREC (Text REtreival Conference) 會議主辦進行這類型的比賽；日文的比賽於 2003 年由日本國立情報學研究所 NII 的 NTCIR 會議 (NTCIR Workshop) 所主辦；歐洲同樣於 2003 年由 CLEF (Cross Language Evaluation Forum) 會議主辦歐洲語言的比賽。根據 2004 年的報告[14]，目前最佳英文問答系統的水準已經可以達到 70%左右的正確率。也就是說，在一百個自然語言問句中，有七十題系統可以直接回答精準而正確的答案。此最佳英文系統由 Language Computer Corporation 所發展，邏輯推理能力爲其致勝關鍵。在日文系統方面，正確率稍微低了些，但也有 51%。日本電信電話公司(NTT)[9]是目前成績最好的團隊。歐洲方面，QA@CLEF 在規模上相當大，參與比賽的語言高達九種，加上跨語言問答的項目，比賽內容最爲豐富。其中法文、葡萄牙文等語言系統於 2005 年[13]都已經可以達到六成多的正確率。相較於其他語言，中文雖然是世界上第二大語言，但中文問答系統比賽直到 2005 年才開始由日本 NTCIR 會議所主辦。包含臺灣的中研院[10]以及成功大學[11]共九個團隊參加評鑑，經過半年多的努力後，最後只有五個團隊送出結果。其中最佳的正確率爲中研院的 45.5%。

閱讀這些評鑑會議數據時必須注意評鑑方式間的差異。TREC 會議主要的評鑑項目有「仿真陳述」、「列舉」、以及「定義」問題，各類型又有其特定的評鑑標準。而 CLEF 看似與 TREC 的「仿真陳述」類型相同，但最近特別強調「時間限制問題」，使得問題更有挑戰性。而 NTCIR 的 2005 年的日文題目則全爲「序列問題」。就算題目類型相同，評鑑方式仍可能不同。TREC 使用三位評鑑者來評估每一結果，而 CLEF 依照語言的不同，使用一或兩位來評鑑每一題。2005 新引進的 NTCIR 中文問答則使用了兩位評鑑者。評鑑標準最大的差異在於是否有考慮「文章支持度」的問題，TREC、CLEF 以及 NTCIR 的中文問答都會考慮答案所在的文章是否「支持」該答案爲真，若證據不明確，就算答案字串正確，該題仍會被視爲是錯誤的。早期 NTCIR 日文問答則沒有考慮文

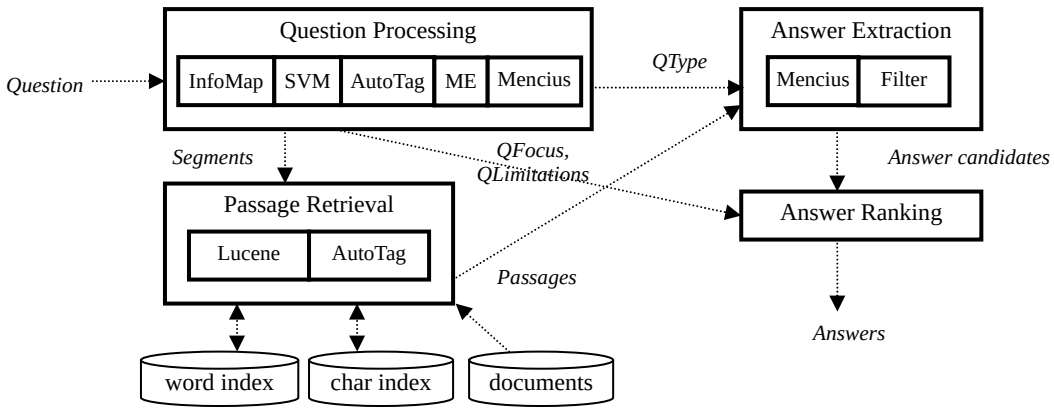
章支持度的問題。根據 TREC 的評鑑結果，有考慮跟沒考慮文章支持度的評鑑結果差距可達十幾個百分比之多。

本實驗室研究狀況

中央研究院資訊科學研究所智慧型代理人實驗室多年前即開始從事問答系統研究，第一個開發的實驗系統爲「中研院答詢系統」 (http://qa.iis.sinica.edu.tw/)。此系統爲一「常問問題」問答系統，允許使用者使用如「請問如何到中研院？」之類口語化的問句提問，理解問句後如果有找到合適的中研院網頁，會直接將該頁面顯示出來；如果沒有合適網頁的話，會將問題導到一般搜尋引擎去。系統使用了 InfoMap[8]知識表示方式，將常問問題拆解整合成領域知識的樹狀結構，用來輔助使用者問句與常問問句間之比對。

另外一個系統爲開放領域的仿真陳述問答系統，名爲 Academia Sinica Question Answering System (ASQA)，此系統參加了此次 NTCIR 中文問答系統評鑑。爲了參加這次的比賽，我們參考過去開發「中研院答詢系統」的經驗，加上更多語言分析與機器學習技術，最後完成的系統包含「問題分析」、「文句擷取」、「答案抽取」、「答案排序」等四大模組。自然語言問句在經過「問題分析」模組取得重要關鍵字後，透過「文句擷取」模組取得高度相關的句子。「答案抽取」模組負責從這些文句中抽取辨識出符合問句類型的候選答案，最後再利用「答案排序」模組將最合適的答案輸出。

在「問題分析」中，問句的類型、斷詞結果、問題焦點、問題限制等資訊都是分析的重點。問題類型的分類包含人、地、組織、物、時間、數字等六大類，這六大類又可再細分成六十二小類。問題分類的處理採 InfoMap 與 Support Vector Machine(SVM)並用的方式來處理。InfoMap 負責提供問句知識與句型比對，而 SVM 則是根據事先標記好的訓練資料來決定問句類型。斷詞部分則是採用資訊所詞庫小組 (CKIP) 所提供的中文斷詞程式[3]來處理。問題焦點與限制是指問句中用來指稱答案類型以及相關答案限制的字串。這些字串在之後的「答案排序」階段扮演著相當重要的角色。我們將問題焦點的辨識視爲一個序列標記問題，人工標記部分語料後，利用 Maximum Entropy Model 輔以一些 Heuristic 的規則來進行辨識。「文句擷取」不同於一般資訊檢索系統，著重於以完整句子作爲檢索對象。這些句子雖然比較短，但相對於以逗點分隔的短句來說，其中包含的資訊尚稱完整，用來作爲問答系統檢索單位相當合適。我們利用 Lucene 文件檢索引擎，根據新聞語料建立了兩個索引結構，這兩個索引分別以詞與字爲單位，以便同時兼顧斷詞結果正確，以及因不正確而導致檢索失敗的狀況。檢索結果前幾名的句子會被送到「答案抽取」模組進行答案抽取。「答案抽取」步驟中最主要的程序爲專有名詞辨識，目前所處理的問題類型中，大多數類型所需要的答案都是專有名詞。爲此，我們開發了一套專有名詞辨識(NER, Named Entity Recognition) 系統—「莊子」(Mencius)[12]，用在辨識人、地、組織名稱上。Mencius 也是採用



圖一:ASQA 系統架構圖

Maximum Entropy (ME) 機器學習模型，整合了許多與專有名詞相關的知識與模版，整體識別正確率約在八成左右，對於答案的抽取有相當大的助益。「答案排序」爲最後關鍵步驟，所有候選答案都在此做信心強度的分數計算與排序。分數的計算必須考慮答案在文句中的位置，同時也要考慮到該答案與「問題焦點」、「問題限制」間的關係。最後從這些關係，我們可以得到一個加權分數，用來做最後的排序，並輸出最高分者爲最後答案。

結語

完成問答系統所需的技術相當繁雜，從領域知識的蒐集、現成模組的整合到機器學習演算法甚至是邏輯推演、完整文句解析，每一項都是一個挑戰。但問答系統也提供一個邁向智慧型語言系統的一個窗口，它比傳統資訊檢索系統複雜，但又比對話系統來的單純，可以用來檢測相關技術在智慧型語言系統的表現。許多研究都是從這樣的角度來發展問答系統，如 LCC 利用問答系統驗證其邏輯推理機制；NTT[9]利用問答系統驗證其資訊檢索系統等。TREC 在資訊檢索方面經過多年經營，使得近年資訊檢索表現大幅提昇。我們對問答系統也有相同的期望，希望能早日完成真實可用的系統。

參考文獻

1. "Ask Jeeves," http://www.ask.com/
2. "Internet Users By Language, Internet World Stats," http://www.internetworldstats.com/stats7.htm
3. "Autotag, CKIP, Academia Sinica," http://ckipsvr.iis.sinica.edu.tw/
4. Green, B., Wolf, A., Chomsky, C., and Laughery, K. "BASEBALL: an automatic question answerer," in: *Readings in natural language processing*, Morgan Kaufmann Publishers Inc., 1986, pp. 545-549.
5. Gulli, A., and Signorini, A. "The Indexable Web is More than 11.5 billion pages," Poster proceedings of the 14th international conference on World Wide Web, 2004.
6. Gunderloy, M., and Sneath, T. *SQL Server Developer's Guide to OLAP with Analysis Services* Sybex, 2001.
7. Harrabagiu, S., Moldovan, D., Clark, C., Bowden, M., Williams, J., and Bensley, J. "Answer Mining by Combining Extraction Techniques with Abductive Reasoning," Proceedings of TREC, 2003.
8. Hsu, W.-L., Wu, S.-H., and Chen, Y.-S. "Event Identification Based on the Information Map - INFOMAP," IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE), 2001.
9. Isozaki, H. "NTT's Question Answering System for NTCIR QAC2," NTCIR Workshop, 2004.
10. Lee, C.-W., Shih, C.-W., Day, M.-Y., Tsai, T.-H., Jiang, T.-J., Wu, C.-W., Sung, C.-L., Chen, Y.-R., Wu, S.-H., and Hsu, W.-L. "ASQA: Academia Sinica Question Answering System for NTCIR-5 CLQA," NTCIR Workshop, 2005.
11. Lin, S.-J., Shia, M.-S., Lin, K.-H., Lin, J.-H., Yu, S., and Lu, W.-H. "Improving Answer Ranking Using Cohesion between Answer and Keywords," NTCIR Workshop, 2005.
12. Tsai, T.-H., Wu, S.-H., Lee, C.-W., Shih, C.-W., and Hsu, W.-L. "Mencius: A Chinese Named Entity Recognizer Using Maximum Entropy-based Hybrid Model," *Computational Linguistics & Chinese Language Processing* (9) 2004, pp 65-82.
13. Vallin, A., Giampiccolo, D., Aunimo, L., Ayache, C., Osenova, P., Peñas, A., Rijke, M.d., Sacaleanu, B., Santos, D., and Sutcliffe, R. "Overview of the CLEF 2005 Multilingual Question Answering Track," Cross Language Evaluation Forum (CLEF), 2005.
14. Voorhees, E.M. "Overview of the TREC 2004 Question Answering Track," Text REtreival Conference (TREC), 2004.
15. Woods, W.A. "Progress in Natural Language Understanding - an application to lunar geology," American Federation of Information Processing Societies, 1973, pp. 441-450.