中華民國計算語言學學會
通訊 ACLCLP
The Association for Computational Linguistics and Chinese Language Processing
第十七卷第四期
發行人：陳信宏
主　編：曾淑娟
執行編輯：黃　琪

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 真：(02) 2788-1638
劃 撥：19166251

本 期 要 目

ROCLING-2006 開始報名

「第十八屆自然語言與語音處理研討會」已開放線上報名(網址：http://www.speech.cm.nctu.edu.tw/Rocling2006/index.html)，報名截止日 8/25 日；會議時間：9/6～9/8；會議地點：Tutorial-中研院活動中心，Main Conference 新竹交通大學，議程請參閱第三～六頁。

博碩士論文獎申請

第六屆博碩士論文獎申請已於 7/31 日截止，本次共收到 11 篇碩士論文及 3 篇博士論文之申請，評審結果預計九月初公布於學會網頁，並將於 9/7 日之會員大會舉行頒獎儀式。

會員大會開會通知

第九屆第二次會員大會謹訂於 95 年 9 月 7 日假新竹交通大學舉行（將合併「第十八屆自然語言與語音處理研討會」同時舉行），本次大會除了學會之會務報告及提案提請通過外，並將追認通過九十四年度財務報表及九十五年度預算收支表。大會中亦將頒發博碩士論文獎及 ROCLING-2006 最佳論文獎。不克出席之會員懇請簽署委託書，以利大會進行順利，委託書請參閱第十二頁。

中央研究院中英雙語知識本體資料庫

「中央研究院中英雙語知識本體資料庫」(The Academia Sinica Bilingual Ontological Database)，網址：http://bow.sinica.edu.tw/為一涵蓋約十一萬中文詞之中英雙語知識本體資料庫。本資料庫以 IEEE 批准執行的 SUMO（Suggested Upper Merged Ontology，(http://www.ontologyportal.org）知識本體為基礎架構，並以台灣地區的語言使用為經驗基礎。提供的訊息包含中英雙語知識本體架構與概念內容，中文語言資訊與概念架構（知識本體）的連結。在使用語言與詞彙資料的基礎上，提供了知識運籌的基本架構(infrastructure)。讓不同來源的典藏知識內容，可以轉換成互通的(inter-operable) 訊息。本資料庫之英文原始內容(SUMO)由 teknowledge.com 管理並公開授權；中文原始資料由語數位典藏國家型科技計畫技術分項計畫下之語言座標計畫，(NSC94-2422-H-001-009-)開發。現經中研院授權本會發行，經費由國科會補助，中央研究院執行並持有中央研院院中英雙語知識本體資料庫。公開授權資料分別以純文字以及 XML 檔案格式儲存。

資料庫內容及範例請參閱第七～九頁，申請方式請至本會網站查詢或洽詢本會秘書處。

第三屆國際中文語言處理競賽結果

由中研院資訊所許聞廉老師主持的「智慧型代理人系統實驗室(IASL)」，參加第三屆國際中文語言處理競賽(簡稱 SIGHAN Bakeoff 2006)表現優異。

SIGHAN Bakeoff 爲計算語言學會(ACL)中文語言處理特別小組(SIGHAN)所主辦的國際中文語言處理競賽，第一屆國際中文分詞競賽起始於2003年在日本札幌舉行。今年(2006)爲第三屆國際中文語言處理競賽將於 2006 年 7 月 22-23 日在澳洲雪梨舉辦。第三屆國際中文語言處理競賽包括斷詞(WS)及專有名詞辨識(NER)二大項競賽。本屆斷詞競賽提供五組語料，包括中央研究院繁體中文語料庫(WS CKIP)、香港城市大學繁體中文語料庫(WS CityU)、微軟亞洲微究院簡體中文語料庫(WS MSRA)、賓州大學與科羅拉多大學簡體中文語料庫(WS UPUC)。此外，本屆專有名詞辨識提供三組語料，包括香港城市大學繁體中文語料庫(NER CityU)、微軟亞洲微究院簡體中文語料庫(NER MSRA)、語言資料協會簡體中文語料庫(NER LDC)。其中，參賽者可參加開放式競賽(Open Track)或封閉式競賽(Closed Track)，參加開放式競賽者，除可使用訓練語料外，並可使用任何外部語料庫、詞典、網路等語料；而參加封閉式競賽者，則僅可使用與測試語料出處相同之訓練語料，不可使用其它任何外部資料。

中研院資訊所智慧型代理人實驗室(IASL)累積多年中文自然語言處理相關系統開發經驗，在今年第一次參加國際中文語言處理競賽之香港城市大學繁體中文語料庫封閉式斷詞競賽 (WS CityU Closed Track) 即獲得 97.2%正確率，並在此分項參賽的 13 隊中獲得的第一名(微軟亞洲研究院 MSRA 第二名)。此外，IASL 在中央研究院繁體中文語料庫封閉式斷詞競賽(WS CKIP Closed Track)中獲得 95.7%正確率，並在此分項參賽的 10 隊中獲得第二名(微軟亞洲研究院 MSRA 第一名)。最後，IASL 在香港城市大學繁體中文語料庫封閉式專有名詞辨識 (NER CityU Closed Track)中，獲得 88.61%正確率，並在此分項參賽的 8 隊中獲得第二名(日本情報通信研究機構 NICT 第一名)。

「中研院口語韻律語料庫暨工具平台」產品項目及價格調整

「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，http://www.myet.com/COSPRO)，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，已於今年一月釋出本語料庫與工具平台，授權民間公司—艾爾科技（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，分別儲存於 4 張 DVD 光碟片中。

考量到一次購買大批語料庫，對於同行的研究經費是一大負擔；再者，不同的研究領域，各個研究者的研究需求不同，購買大批語料庫並不符合研究上的經濟效益。因此，爲求使 COSPRO 能達到最大的研究使用效益，同時也能兼顧 COSPRO 語料設計之核心精神，自 2006 年 7 月 1 日起，調整語料庫售價，同時針對不同研究需求，提供產品組合的搭配建議。產品項目及售價請參閱第十～十一頁。

第十八屆自然語言與語音處理研討會
(ROCLING-2006)
http://www.speech.cm.nctu.edu.tw/Rocling2006/index.html

會議日期：95 年 9 月 6 日～95 年 9 月 8 日

會議地點：**Tutorial-**台北中研院活動中心第一會議室
Main Conference-交通大學電子與資訊中心

主辦單位：交通大學電信工程系、中華民國計算語言學學會(ACLCLP)

Pre-ROCLING Tutorials Program

September 6		
Time	**Session**	**Speaker**
09:20-09:30	**Registration**	
09:30-10:40	Ontologies and NLP (1/2)	Laurent Prevot
10:40-11:00	**Coffee break**	
11:00-12:20	Ontologies and NLP (2/2)	Laurent Prevot
12:20-13:30	**Lunch break**	
13:30-14:30	New Resource in Chinese Language Processing - Fully Tagged Chinese GigaWord Corpus	Chu-Ren Huang & Keh-Jiann Chen
14:30-14:50	**Coffee break**	
14:50-16:50	Corpus Management and Processing Tools	Yuji Matsumoto

Tutorial Abstract

Ontologies and Natural Language Processing
Laurent Prevot (Academia Sinica)

This talk provides/concerns the use of ontologies in computational linguistics. It will first define ontologies and explores their diversity from foundational ontologies to domain ones. Then it will detail their applications in computational linguistics but also the use of Natural Language Processing for building and improving them. A special attention will be given to linguistic ontologies such as famous lexical resources (WordNet and FrameNet) and their combination with traditional ontologies.

Tutorial Abstract

New Resources in Chinese Language Processing -
Fully Tagged Chinese GigaWord Corpus

Chu-Ren Huang and Keh-Jiann Chen
(Academia Sinica)

The Chinese GigaWord Corpus (CGW Corpus hereafter), first released by LDC in 2003 and updated in 2005, has the following crucial characteristics:

- CGW Corpus is the largest publicly available Chinese Corpus. CGW Corpus 1.0 contains more than 1,100 million characters, while CGW Corpus 2.0 contains nearly 1,300 million characters.
- CGW Corpus is the only corpus that containing sizable data from both PRC and Taiwan. CGW Corpus 2.0 contains more than 790 million characters from Taiwan, more than 470 million characters from PRC, as well as more than 28 million characters form Singapore.

The complete CGW Corpus is easy to access since internal formatting has been unified, and each text is clearly marked and classified by topic. However, neither versions 1.0 or 2.0 were segmented or tagged. In order to turn CWG Corpus as the basic resource for more versatile Chinese language processing in the future, a fully tagged version of the CGW Corpus was prepared (Ma and Huang 2006), and will be made available shortly. This tutorial introduces the tagged CGW Corpus.

The following topics will be covered in this tutorial:

- the content and composition of CGW Corpus, including its text formatting and topic classification
- the tagset adopted (CKIP-SinicaCorpus tagset)
- the tagging methodology
- quality assurance: specific improvements and independent evaluation
- availability of the data: Licensing from LDC and browsing Chinese WordSketch

Selected References

Chinese GigaWord Corpus http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2003T09
Chinese GigaWord Corpus Second Edition
http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2005T14
Wei-yun Ma, and Chu-Ren Huang**.** 2006. Uniform and Effective Tagging of a Heterogeneous Giga-word Corpus. Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC2006). Genoa, Italy. 24-28 May, 2006.

Tutorial Abstract

Corpus Management and Processing Tools

Yuji Matsumoto (NAIST)

Annotated corpora are valuable resource for natural language processing as well as linguistic research. We have been developing a corpus management tools that store POS and syntactic dependency annotated corpora and provide various functions such as search and visualization. This tutorial introduces the system together with freely available language processing tools we are developing, such POS taggers, chunkers and dependency structure analyzers.

ROCLING-2006 Main Conference Program

September 7		
Time	**Session**	**Chair**
08:30-09:00	**Registration**	
09:00-09:10	Opening Ceremony-陳信宏教授	
09:10-10:10	Keynote Speech : Prof. Yuji Matsumoto	黃居仁教授
10:10-10:40	**Coffee Break**	
10:40-12:00	Session 1 – 語法及語意處理	高照明教授
12:00-13:20	**Lunch** **中華民國計算語言學會會員大會**	
13:20-14:20	Invited Speech – I : 簡立峰教授	陳信宏教授
14:20-14:40	**Break**	
14:40-15:40	Session 2 – 語音信號處理	廖元甫教授
15:40-16:00	**Coffee Break**	
16:00-16:50	Panel Discussion – Questions from Speech Technology Development to Linguistic and Phonetics – Dialogue and Future Directions 語音科技與語音學對談：談談語音科技開發過程中的語言語音問題	鄭秋豫博士
18:00-	**Banquet**	

September 8		
Time	**Session**	**Chair**
09:10-10:10	Invited Speech – II : 黃居仁教授	張俊盛教授
10:10-10:40	**Coffee Break**	
10:40-12:00	Session 3 – 語音及語者辨認	王新民教授
12:00-13:20	**Lunch**	
13:20-14:20	Session 4 – 資訊檢索，文件分類及語意網	盧文祥教授
14:20-14:40	**Break**	
14:40-16:00	Session 5 – 語言模型及其應用	張景新教授
16:00-16:20	**Closing**	

Session 1 語法及語意處理

1. 以語料爲基礎的中文語篇連貫關係自動標記
   鄭守益、吳典松、梁婷
2. 中文動詞名物化判斷的統計式模型設計
   馬偉雲、黃居仁
3. 大規模詞彙語意關係自動標示之初步研究 - 以中文詞網爲例
   Petr Šimon、謝舒凱、黃居仁
4. Automatic Learning of Context-Free Grammar
   Tai-Hung Chen、Chun-Han Tseng
5. Improve Parsing Performance by Self-Learning
   Yu-Ming Hsieh、Duen-Chi Yang、Keh-Jiann Chen

Session 2 語音信號處理

1. 國語雙字語詞聲調評分系統
   古鴻炎、孫世諺、張小芬
2. 一種用於網路電話之遺失封包補償方法
   古鴻炎、陳佳新
3. 基於字詞內容之適應性對話系統
   朱育德、張嘉惠
4. 一種適用於大量連續語料的語音文句校準方法
   林千翔、張嘉惠

Session 3 語音及語者辨認

1. 利用聲學與文脈分析於多語語音辨識單元之產生
   王士豪、黃建霖、吳宗憲
2. 統計圖等化法於雜訊語音辨識之進一步研究
   林士翔、葉耀明、陳柏琳
3. 應用不定長度特徵之條件隨機域於口語不流暢語流修正
   吳宗憲、吳維彥、葉瑞峰
4. 鑑別性事前資訊應用於強健性語音辨識
   丁川偉、吳柏樹、簡仁宗
5. 結合韻律與聲學訊息之強健性漢語語者驗證系統
   張文杰、陳鼎允、陳子和、曾志仁、廖元甫、莊堯棠

Section 4 資訊檢索，文件分類及語意網

1. Personalized Optimal Search in Local Query Expansion
   Shan-Mu Lin、Chuen-Min Huang
2. 以本體論爲基礎之新聞事件檢索與瀏覽
   許孟淵、黃純敏
3. 以部落格文本進行情緒分類之研究
   楊昌樺、陳信希
4. Software for minimalist experimental syntax
   James Myers

Section 5 語言模型及其應用

1. 基於特製隱藏式馬可夫模型之中文斷詞研究
   林千翔、張嘉惠
2. 以字串特徵做爲文本資料之錯誤偵測
   劉吉軒、鄭瓊瑋
3. Learning to Parse Bilingual Sentences Using Bilingual Corpus and Monolingual CFG
   Chung-Chi Huang、Jason S. Chang
4. 使用流暢性改善詞組翻譯的統計式機器翻譯
   夏敏翔、張耀升、盧文祥
5. An Evaluation of Adopting Language Model as the Checker of Preposition Usage
   Shih-Hung Wu、Chen-Yu Su

中央研究院中英雙語知識本體資料庫
The Academia Sinica Bilingual Ontological Database

1 簡介

「中央研究院中英雙語知識本體資料庫」(The Academia Sinica Bilingual Ontological Database)，網址：http://bow.sinica.edu.tw/爲一涵蓋約十一萬中文詞之中英雙語知識本體資料庫。本資料庫以IEEE批准執行的SUMO（Suggested Upper Merged Ontology，http://www.ontologyportal.org）知識本體爲基礎架構，並以台灣地區的語言使用爲經驗基礎。提供的訊息包含中英雙語知識本體架構與概念內容，中文語言資訊與概念架構（知識本體）的連結。在使用語言與詞彙資料的基礎上，提供了知識運籌的基本架構(infrastructure)。讓不同來源的典藏知識內容，可以轉換成互通的 (inter-operable) 訊息。本資料庫之英文原始內容(SUMO)由teknowledge.com 管理並公開授權；中文原始資料由語數位典藏國家型科技計畫技術分項計畫下之語言座標計畫，(NSC94-2422-H-001-009-)開發。經費由國科會補助，中央研究院執行並持有中央研院院中英雙語知識本體資料庫。公開授權資料分別以純文字以及XML檔案格式儲存。

2 資料內容說明

開放授權的資料包括下列所示的三大類訊息：

A. SUMO知識本體中文版：

依據SUMO2002版本的中文版資料，涵蓋11大類的概念，每大類又分爲二至五個類別，總共囊括3,912個概念，每項概念下附有解釋與定理，詳細資料請請參閱參考文件一。

B. 中英雙語領域分類樹：

參考「中國圖書分類法」爲基準，並參考各知識分類與實際研究經驗，提出：包含九大類的知識分類（Knowledge Content），涵蓋427個領域，並因應語言資源特性加入下列語言使用（Language Usage）的各類訊息：專名（說明文字符號的指涉）（Proper Name）、語體（說明文字符號的使用）（Genre/Strata）、各種語言／詞源（Language/Etymology）、各國地名（Country Name）。知識分類（Knowledge Content）的九大類分別是：人文學科(Humanities)、社會科學（Social Science）、形式科學（Formal Science）、自然科學（Natural Science）、醫療科學（Medical Science）、工程科學（Engineering Science）、應用產業(Production Industry）、藝術（Fine Arts）以及休閒娛樂（Recreation）。

C. 中文與SUMO知識本體對應資料庫：

由中文詞彙出發，經由以英文WordNet定義的同義詞集（synset）爲基準，對應到SUMO知識本體的概念節點。資料內容包括：

(1) 中文詞彙：109,982詞形；部份詞彙對應到一個以上英文同義詞集，以每個同義詞集對應計一次，共有149,780筆資料。
(2) 每個詞形所對應的英語WordNet同義詞集記錄碼(offset)。共對應到99,642個同義詞集。
(3) 同義詞集與SUMO知識本體的對應關係與知識本體分類（SUMO概念）間的關係：SUMO

工作小組將WordNet1.6每個同義詞集對應SUMO概念，並針對兩者之間對應的關係分成三種，分別是—上位（Hypernym）、同義（Synonymy）以及示例（Instantiation），詳細資料請請參閱參考文件四。

(4) 知識本體分類（SUMO概念英文）：SUMO概念的英文版。

(5) 知識本體分類（對應SUMO概念中文）：SUMO概念的中文版。

(6) 詞彙領域分類：針對同義詞集給與相對應於領域分類樹之的領域訊息，詳細資料請請參閱參考文件4。

3 範例

A. SUMO知識本體中文版：檔案名稱SUMO_Chinese_2002.txt
(termFormat cn Abstract "抽象的")
(termFormat cn AbstractionFn "描述函數")

B. 中英雙語領域分類樹：檔案名稱20050311domain.xls
人文學科Humanities
語言學　linguistics

C. 中文與SUMO知識本體對應資料庫：由中文詞彙出發，經由以英文WordNet定義的同義詞集（synset）爲基準，對應到SUMO知識本體的概念節點。

＃中文與SUMO知識詞網
＃每筆資料包括:
＃（1）序號（無特殊意義，主鍵爲「中文詞形」+「詞類」+「相對應的 WordNet1.6 同義詞集offset」）
＃（2）中文詞形
＃（3）詞類
＃ (4) 與WordNet 對應關係—ISA對應關係 OF相對應的 WordNet1.6 同義詞集 offset
＃（5）知識本體分類-- @ SUMO概念 @ SUMO中文

I. 純文字檔：檔案名稱Ontological_License中英雙語知識本體.txt

1　安布利亞Noun　ISA Instantiation　OF 06467131N @ LandArea @ 陸地
2　楓香樹　Noun　ISA Hypernym　OF 08619764N @ FloweringPlant @ 開花植物

II. XML檔：檔案名稱Ontological_License中英雙語知識本體.xml

```
<Record Count="1">
<ChineseLemma>安布利亞</ChineseLemma>
    <POS>Noun</POS>
    <WordNetSynsetOffset Version="1.6">06467131N</WordNetSynsetOffset>
    <SUMO>
        <SUMORelation>Instantiation</SUMORelation>
        <SUMOConcept>LandArea</SUMOConcept>
        <SUMOChi>陸地</SUMOChi>
    </SUMO>
<Record Count="2">
        <ChineseLemma>楓香樹</ChineseLemma>
```

```
        <POS>Noun</POS>
        <WordNetSynsetOffset Version="1.6">08619764N</WordNetSynsetOffset>
        <SUMO>
          <SUMORelation>Hypernym</SUMORelation>
          <SUMOConcept>FloweringPlant</SUMOConcept>
          <SUMOChi>開花植物</SUMOChi>
        </SUMO>
</Record>
```

4 參考文件

1. Niles, I., & Pease, A. (2001). “Toward a Standard Upper Ontology”. In Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001).Ogunquit, Maine, October 17-19, 2001. pp.2-9.
2. WordNet Reference Manual, http://www.cogsci.princeton.edu/~wn/doc.shtml
3. Niles, I., and Pease, A. (2003). “Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology”. Proceedings of the IEEE International Conference on Information and Knowledge Engineering. (IKE 2003). Las Vegas, Nevada, June 23-26, 2003. pp.412-416.
4. Chu-Ren Huang, Ru-Yng Chang, Shiang-Bin Lee. (2004). “Sinica BOW (Bilingual Ontological Wordnet): Integration of Bilingual WordNet and SUMO”. 4th International Conference on Language Resources and Evaluation (LREC2004). Lisbon. Portugal. 26-28 May, 2004. pp.1553-1556.
5. 張如瑩,黃居仁. 2004. 中央研究院中英雙語知識本體詞網 (Sinica BOW)：結合詞網，知識本體，與領域標記的詞彙知識庫。發表於第十六屆自然語言與語音處理研討會（ROCLING XVI）September 2-3.
6. Sinica BOW: http://BOW.sinica.edu.tw/
7. SUMO: http://www.ontologyportal.org

COSPRO 語料庫產品分項與售價一覽表

項　目	產 品	容 量（GB）	購買區別	Institution Type	License Fee
全套（4片裝）	1.COSPRO DVD No.1~4	10.5GB	國內	Non-profit Organizations	NT $140,000
				Others	NT $420,000
			國外	Non-profit Organizations	US $9,000
				Others	US $27,000
兩片裝	2. COSPRO DVD No. 1~2 建議用途：語音辨識	5.54GB	國內	Non-profit Organizations	NT $80,000
	3. COSPRO DVD No. 3~4 建議用途：語音合成	5.039GB		Others	NT $240,000
	4.COSPRO DVD No. 1&3 建議用途:韻律研究 speaker dependent	5.187GB	國外	Non-profit Organizations	US $5,000
	5. COSPRO DVD No. 2&4 建議用途：韻律研究 corpus-type dependent	5.392GB		Others	US $15,000

COSPRO 語料庫組合與 DVD 容量一覽表

DVD 編號	語料庫名稱及編號	DVD 檔案容量（GB）
1	COSPRO 01 六人次音段字調平衡語段語料庫	2.807GB
	COSPRO 04 二人次輕重音型平衡語料庫	
	COSPRO 05 二人次兩岸辭彙平衡語料庫	
2	COSPRO 07 二人次多樣(含無義字串)語料庫	2.733GB
	COSPRO 09 二人次自發朗讀對照語料庫	
	COSPRO 02 多人次詞、短句、語段語料庫	
3	COSPRO 03 七人次字句調平衡語料庫	2.38GB
4	COSPRO 06 二人次韻律句群焦點位置語料庫	2.659GB
	COSPRO 08 二人次韻律單位語料庫	

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

中 華 民 國 計 算 語 言 學 學 會

第九屆第二次會員大會委託書

開會時間：95 年 9 月 7 日中午 12:40~13:20 (星期四)
開會地點：新竹市交通大學電子與資訊中心

委 託 書

本人因故不克出席中華民國計算語言學學會第九屆第二次會員大會，茲委託本會會員＿＿＿＿＿＿＿＿＿＿＿＿代表本人出席。

此致

中華民國計算語言學學會

委託人：＿＿＿＿＿＿＿＿＿＿＿＿

受託人：＿＿＿＿＿＿＿＿＿＿＿＿

中華民國　　年　　月　　日

注意事項：

1. 每一會員僅能接受一位其他會員之委託。
2. 會員若<u>不克出席</u>，又無法找到適當受託人者，學會可代為安排，會員只需在<u>**委託人欄位上簽名**</u>即可。
3. 請將本委託書於 8/25(五)日前傳真至本會(02)2788-1638 或交由受託人於開會時至報到處繳交。
4. 本委託書僅供參考，會員若自行開具『委託書』，亦屬有效。
5. 本次會議將與「第十八屆自然語言與語音處理研討會」合併舉行。