A F N L P

N E W S L E T T E R

(2016Q2 issue, April 2016)

1. President's Message

To increase the awareness of NLP research in Asia, the AFNLP newsletter starting from this issue will publish special report on the latest NLP development in different Asian country's/region's. Thanks to Virach Sornlertlamvanich, Sirindhorn International Institute of Technology, Thammasta University, Thailand; and Chai Wutiwiwatchai and Thepchai Supnithi, National Electronics and Computer Technology Center (NECTEC), Thailand, the second issue of 2016 covers Thailand's research outlook on NLP and speech processing (see Section 10).

2. Asian Linguistic Resources

Key-Sun Choi (Korea) has kindly volunteered to lead the AFNLP Asian Linguistic Resource Initiative. He keeps in contact with the LREC Steering Committee. He reported that Nicoletta, Calzolari, Chair of LREC visited Japan and met with Hitoshi Isahara and Jun-ichi Tsujii to discuss the organization of the next LREC in Asia.

Further, Key-Sun will discuss with Koiti Hasida (Japan), Chair, AFNLP's Asian Language Resources Committee (ALRC) and the AFNLP voluntary liaison person for the ISO/TC37/SC4 working group, for the possibility of AFNLP using LREC Linguistic Resource (LR) Map facilities.

For the TC37/SC4 liaison task, there is no news from the ISO Central Secretariat after they asked about the legal seat of AFNLP. The corresponding person, Merce, has mentioned nothing after the last December. Key-Sun will follow-up.

3. Publicity and Communications

AFNLP has compiled an Asian NLP Expert List with more than 130 entries in 2015Q1. The list was shared with ACL-IJCNLP-2015, NLPCC2015, BigComp2016 and NLPCC2016. It is available to members upon request. Please send email to the President if you are interested. We will initiate an update exercise for the list soon. The target is to get the name, expertise and associated contact information of 20 experts from each country/region.

Based on input from members of the executive committee, AFNLP has compiled a list of 2016 NLP-related conferences, which will be held in or related to Asia, see below:

Conference	Submission Date	Notification Date	Conference Date	Location
LREC 2016	Oct 15, 2015	Feb 1, 2016	May 23-28, 2016	Portoroz, Slovenia
NAACL 2016	Jan 6, 2016	Mar 2, 2016	Jun 12-17, 2016	San Diego, USA
SIGIR 2016	Jan 21 2016	Mar 31, 2016	Jul 17-21, 2016	Pisa, Italy
ACL 2016	Mar 18, 2016	May 24, 2016	Aug 7-12, 2016	Berlin, Germany
EMNLP 2016	Jun 3, 2016	Jul 29, 2016	Sep 1-5, 2016	Austin, TX, USA
ICHAI2016	Apr 30 2016	Jun 24, 2016	Oct 4-7, 2016	Singapore
PACLIC2016	Jun 30, 2016	Jul 31 2016	Oct 28-30, 2016	Seoul, Korea
IALP 2016	Jul 20, 2016	Aug 20, 2016	Nov 21-23, 2016	Tainan, Taiwan
NLPCC2016	Jul 22, 2016	Sep 17, 2016	Dec 2-6, 2016	Kunming, China
COLING 2016	Jun 15, 2016	Sep 19, 2016	Dec 11-16, 2016	Osaka, Japan
MT Summit XVI	TBD	TBD	Sep 18-22, 2017	Nagoya, Japan
IJCNLP2017	TBD	TBD	Nov 20-24, 2017	Taipei, Taiwan
Conferences sponsored, supported or hosted by AFNLP				

4. Scholarship and Awards

AFNLP has awarded the following papers for two AFNLP Asian Language Best Paper Awards in the 4th Natural Language Processing and Chinese Computing (NLPCC2015), October 8-11, 2015, Nanchang, China. (see http://tcci.ccf.org.cn/conference/2015/index.html)

1. *A Dynamical Glyph Generation Method of Xiangxi Folk Hmong Characters and Its Implementation Approach.* Liping Mo (Ji Shou University) and Kaiqing Zhou (Universiti Teknologi Malaysia),
2. *Uyghur Text Automatic Segmentation Method Based on Inter-Word Association Degree Measuring.* Turdi Tohti, Winira Musajan and Askar Hamdulla (Xinjiang University)

AFNLP has supported the AFNLP Asian Language Best Paper Award in the 3rd International Conference on Big Data and Grid Computing, January 18-20, Hong Kong. (see http://conf2016.bigcomputing.org/main/)

AFNLP has committed three more best paper awards to the following conferences:

a. ACL-2016 Student Research Workshop, August 7-12, 2016
b. IALP-2016, Tainan, Taiwan, November 21-23, 2016
c. COLING-2016, Osaka, Japan, December 11-16, 2016

\

5. IJCNLP 2017

The 8th International Joint Conference on Natural Language Processing (IJCNLP 2017), organized by The Association for Computational Linguistics and Chinese Language Processing (ACLCLP) and hosted by The Asian Federation of Natural Language Processing (AFNLP) Associations, will be held in Taipei, Taiwan on November 20-24, 2017. IJCNLP 2017 will feature full papers, short papers, demonstrations, tutorials, and workshops.

The Pre-Conference Workshops and Tutorials: November 20, 2017; Main Conferences: November 21-23, 2017 and Post-Conference Workshops: November 24, 2017.

Local Organizing Co-Chairs: Yuen-Hsien Tseng, National Taiwan Normal University; Hsin-Hsi Chen, National Taiwan University; Chung-Hsien Wu, National Cheng Kung University.

6. Call For Bids to Host ACL 2018

ACL invites proposals to host the conference in the Asia-Australasia-Oceania region in July or August 2018 and is now seeking preliminary draft proposals from prospective bidders in the Asia-Pacific region. The selected location for ACL 2018 will be announced at the ACL 2016 conference (Berlin, Germany, August 7-12, 2016). See the following link for details: http://www.aclweb.org/portal/content/final-call-bids-acl-2018

Notice that the following Asian NLP researchers are involved in the ACL2018 Conference Coordination Committee:

1. Pascale Fung, HKUST, Hong Kong (Program Chair ACL2013)
2. Haizhou LI, I^2R, Singapore (General Chair, ACL2012; Local Arrangement Chair, ACL-IJCNLP09)
3. Yuji Matsumoto, NIST, Japan (General Chair, ACL-IJCNLP 2015)
4. Jong Park, KAIST, Korea (Local Arrangement Chair, ACL 2012; PC Chair IJCNLP13)

7. Standardization Meeting on NLP Annotation in Asia

In March 30-31 and April 1, 2016, a standardization meeting for "Natural Language Processing Annotation Knowledge Base Framework" was held in KAIST (Daejeon, Korea). It was attended by Piek Vossen (Holland), President, Global WordNet Association; Koiti Hasida (Japan), Chair of AFNLP ALRC; and Key-Sun Choi (Korea), AFNLP Lead in Asian Linguistic Resource Initiative. The objective of the meeting is to recommend a specification for coordinating and using the different open NLP annotation standards and tools.

8. COLING2016: Call For Workshops

Key-Sun Choi, Past President, AFNLP, is the Workshop Co-Chair of COLING2016. AFNLP Members are invited to submit proposals, see Call-For-Workshop proposals (http://coling2016.anlp.jp/call-for-workshops.html)

9. COLING2016: Call For Papers

The Association for Natural Language Processing in Japan is hosting COLING2016, which will be held on December 11-16 in Osaka, Japan. AFNLP members are invited to submit papers. Dues are July 15 for full papers and August 26 for demo papers. See more details at http://coling2016.anlp.jp.

10. 2016Q2 Special Issue

Thailand Natural Language Processing and Speech Processing Research Outlook 2016

Virach Sornlertlamvanich[1], Chai Wutiwiwatchai[2] and Thepchai Supnithi[2]
[1]Sirindhorn International Institute of Technology, Thammasta University, Thailand
[2]National Electronics and Computer Technology Center (NECTEC), Thailand
virach@siit.tu.ac.th, chai.wutiwiwatchai@nectec.or.th, thepchai.supnithi@nectec.or.th

In a long range of NLP research in Thailand, Natural Language Processing and Speech Processing communities are working very closely since beginning. There is no doubt that why both communities can share their resources in a most efficient way. Corpora and tools are created to serve for both research needs. NLP research kickoff in Thailand was significantly seen after the two big projects of Machine Translation, namely ARIANE (1981) and CICC Multi-lingual Machine Translation (1987-1994). Many research groups were established in universities and research institutes since then. Currently, the achievements are also fruitful implemented in commercial solutions such as electronic dictionary, document management system, speech interface system, opinion mining, and more and more growth in the needs for Big Data analytics.

This report is partly collected from the known communities, which cannot thoroughly represent all the activities in countrywide. It is expected to reflect the basic continuous movement of the efforts in the recent days.

I. Research Communities

Groups of researchers are formed up in universities and research institutes to make contributions in the scope of their interests. By browsing the following tables, it is apparent that NLP and Speech Processing are already reaching into many groups of research interests.

NLP related communities:

Organizations	Scope of Interests
National Electronics and Computer Technology Center (NECTEC)	Thai NLP, Machine Translation, Sentiment Analysis
Linguistics, Chulalongkorn University	Corpus Development
Computer Science, Thammasat University	Thai NLP, Machine Translation
SIIT, Thammasat University	Thai NLP, Text Summarization, NER, Sentiment Analysis, Social Text Analysis, Asian Wordnet

Kasetsart University	Sentiment Analysis
Chiang Mai University	Machine Translation

Speech Processing related communities:

Organizations	Scope of Interests
National Electronics and Computer Technology Center (NECTEC)	Thai LVCSR, Thai/English TTS, Thai pronunciation assessment for medical treatment
Computer Engineering, Chulalongkorn University	Thai TTS, Technologies for people with disabilities
Electrical Engineering, Thammasat University	Speech enhancement, Speech analysis and processing for hearing impaired and speech training fields
Faculty of Liberal Arts, Thammasat University	Speech analysis and processing for hearing impaired and speech training fields
Electrical Engineering, King Mongkut University of Technology Thonburi	Speech and audio signal processing applications
Computer Engineering, Prince of Songkla University	ASR for the southern Thai dialect language
Computer Science, Konkaen University	TTS for the north-eastern Thai dialect language

II. Research Activities

The activities are grouped by their recent significant efforts. Not all research interests in the above tables are elaborated.

(1) Thai NLP and Text Processing

The fundamental of Thai NLP is essential in the recent growth of the Big Data. After releasing the first 5 million words annotated corpus for BEST contest, Thai word segmentation was significantly improved. In SNLP 2016, there are two papers related to word segmentation. The first paper combines dictionary with current CRF method to show the improved accuracy. Since word segmentation is domain dependent task, the second paper attempts to develop a domain-adaptation technique for Thai word segmentation. A CRF based Thai POS tagging is developed and obtains around 97% accuracy. A mathematical model for parameter estimation method for a generative model for PGLR parsing is introduced for Thai parser. Named Entity Recognition (NER) and Semantic Relation Extraction are yet another crucial issues for representing the essential information of the text

for understanding the relation between texts. Machine learning and deep learning approaches are gaining their potential in improving the performance in many research topics.

(2) Machine Translation

Machine Translation is one of the long histories of Thai NLP research. Currently, there are two main works we are focusing on. One is a hierarchical based machine translation for Chinese ↔ Thai. Another is the collaborative project with ASEAN organizations, ASEAN MT project. This project starts from collecting 20 thousands sentences of travel domain corpus from ten languages and develop a statistical machine translation for all language pairs. The current BLEU score is around 20-40. To support this research, a translation management system is developed for bilingual or multilingual corpus development.

(3) Sentiment and Social Text Analysis

Sentiment analysis is expected in modern business, especially to grasp the trend of customer favorite. It is applied to social media text to understand the community response or understanding the customer requirement from the feedback. Two main analysis modules, intention and sentiment, are based on classification algorithm to automatically assign appropriate intention and sentiment class labels for a given text

(4) Text Summarization

Recently there have been an increasing number of online news articles published in any language on public news web sites, social media and web portals. There is a work that applies Thai Elementary Discourse Unit (TEDU) and then proposes a three-stage method of Thai multi-document summarization, i.e., unit segmentation, unit-graph formulation, and unit selection for summarization.

(5) To comprehensively improve the performance Thai LVCSR, NECTEC joined with Computer Engineering, Chulalongkorn University to conduct a recent project named "Thai automatic speech recognition service". Under this project, speech and text corpora in various domains have been enlarged. In summary, the acoustic model has been trained by over 200 hours of speech mainly from read, broadcast, and mobile collections. The language model has been trained by over 70 million words in 12 genres. A hybrid language modeling approach proposed in INTERSPEECH 2013 has been deployed to reduce the size of vocabulary and the out-of-vocabulary rate with no degradation of the recognition accuracy. Evaluated on 3,000 open-domain utterances, the system currently achieves 23% word accuracy. More speech and text data are being collected using social media prompts. Currently, over 100,000 utterances from around 200 speakers with gender balanced and environment distributed have been collected and used to improve the system ability to recognize spoken-style utterances.

(6) The Faculty of Medicine, Khonkaen University has joined NECTEC to pursue a 3-year project called Speech assessment and training tools for cleft-palate patients (2015-2017). This project aims to research and develop tools to assist dentists and speech therapists to assess and train their cleft-palate patients. Besides word sets designed individually in several medical centers, a standard phone balanced set is expected to be systematically designed. Having such standard set used in medical center partners, semi-automatic speech assessment, and training tools are then expected to be integrated in the treatment procedure, with an ultimate goal in enhancing the treatment efficiency over the previous traditional procedure.

III. Events

Recently, a large group of researchers in language, speech, and artificial intelligence have joined setting up a new association, called Artificial Intelligence Association of Thailand (AIAT). PRICAI (Pacific-rim international conference on AI) 2016 is scheduled to be organized in Phuket during August 22-26, 2016. Consecutively, International Joint Conference on Artificial Intelligence (IJCAI) will also be held Thailand in 2019. For the NLP local events, the 12th Symposium of NLP (SNLP), will be in the last quarter of year 2017.

Important Links:
AFNLP website: http://www.afnlp.org;
Webmaster: Minghui Dong <mhdong@i2r.a-star.edu.sg>
Newsletter Editor: Min Zhang <minzhang@suda.edu.cn>