We are pleased and proud to announce that the School of Computer Sciences USM had successfully organised the Second International Workshop on Spoken Language Technologies for Under Resource Languages (SLTU'10) from 3 to 5 May 2010 at Dewan Persidangan Universiti (DPU), Universiti Sains Malaysia. Check out SLTU'10 website: http://www.mica.edu.vn/sltu-2010/index.php for post-workshop info.

The SLTU meeting is a technical workshop focusing on spoken language processing for under-resourced languages, such as Vietnamese, Khmer, Laos, Chinese dialects, Thai, African, etc. In Malaysian context, these languages are Malay dialects, native languages from Sabah and Sarawak, orang asli and others. The areas of research involved automatic speech recognition, machine translation, speech synthesis and others. This year, we received papers from United States, Germany, France, Vietnam, Korea, Malaysia and others.

SLTU'10 workshop managed to gather researchers working on ASR, synthesis and translation for more than 20 under-resourced languages among others include portability issues, multilingual spoken language processing, fast resources acquisition. We were delighted to find so much interest in the workshop that was organised for the first time in Malaysia.

In overall, 22 papers were selected for oral communications. The authors and co-authors originated from 17 countries (Bangladesh, Burma, Cambodia, Ethiopia, France, Germany, Hungary, India, Korea, Malaysia, Pakistan, Russia, Singapore, Spain, South-Africa, Vietnam, USA). There was an increase in the quality of the papers as compared to SLTU'08 and the scientific committee had compiled a two-day scientific programme that covered the areas of automatic translation, speech recognition and synthesis, as well as collection of language resources. In addition, three prestigious speakers were invited to give keynotes: Haizou Li (I2R, Singapore), Ruhi Sarikaya (IBM, US) and Alex Waibel (CMU, US). We also received grants and sponsorship from ISCA, AFCP, CNRS, Grenoble-INP, USM's Vice Chancellor, and National library of Malaysia. The support notably by ISCA and AFCP allows us to give grants to fund five International PhD students to participate in the workshop. The students were from Bangladesh, Hungary, Korea, France, and Vietnam.

The future SLTU'12 workshop will probably be in Africa.
First of all, welcome to the Second International Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU’10). This event is co-organized by Laboratoire d'Informatique de Grenoble (LIG) from France and International Research Center for Multimedia, Information, Communication and Applications (MICA) from Vietnam. We would like to thank both of the organizations for helping us to realize this workshop. This year USM has the honour to host this important event especially in the domain of spoken language technologies.

The SLTU meeting is a technical workshop focusing on spoken language processing for under-resourced languages. This event is especially meaningful to us, as Malaysia is home to many languages, and a lot of them are rather rare. Even though Malay is the national language, communities in Malaysia are free to practise other languages. Even Malay exists in different flavour or dialects at different regions of Malaysia. The Chinese community in Malaysia on the other hand speak different Chinese languages such as Hokkien, Cantonese, Hakka, Teochew etc. Many of the languages have evolved and deviated from their origin. For example the Hokkien spoken in Penang has many vocabularies borrowed from Malay. At the east of Malaysia- Sabah is home to over 50 indigenous languages. These languages have very little language resource and not well studied. Most of them do not have a writing system. Some are at the brink of extinction. An example is the language spoken by Orang Kanaq, which is one of 19 indigenous groups in peninsular Malaysia. The Orang Kanaq language has only less than a hundred speakers. Therefore, the work in this workshop is very much relevant to us in Malaysia, and also aligns with the goal of the university which is to create a sustainable future.

In School of Computer Sciences, Universiti Sains Malaysia, there are 3 research clusters: namely the service, architecture and knowledge clusters. We have a language engineering group which is working in the area language technology such as automatic speech recognition, machine translation, speech synthesis and others. In particular, we are interested in the area of Malay automatic speech recognition and Malay speech synthesis as there is a big demand for these applications. At the national level, the development in this area will encourage the usage and development of Malay. Other areas of interest are in speech search technology, which is dubbed as the next frontier for search engine, where
we make use of the power of automatic speech recognition and grid to allow user to query about multimedia files. We also hope to explore those local speech phenomena such as code switching and have a deeper study on dialects. We hope this workshop will open up more opportunities of collaboration in the area of speech technologies and other related areas. Last but not least, we wish everyone has a fruitful workshop and enjoyable stay in Malaysia.
Foreword by Local Workshop Chair of SLTU'10

After the first SLTU workshop was held in Hanoi, Vietnam in 2008, we have the honor to organize the second Spoken Languages Technologies for Under-resourced Languages workshop with the help and support from Laboratoire d'Informatique de Grenoble (LIG), France and Multimedia, Information, Communication and Applications (MICA) Research Center, Vietnam.

It took us around a year to prepare for this event, starting from the proposal to the realizing of it. However it seems like only yesterday, as I recalled the time when Tan Tien Ping asked me to be the local chair of this event. I accepted the offer because I felt that SLTU has a very good reason d’être. Today a lot of minorities and aborigines are losing their native mother tongue. They do not have the financial and political resources to promote their languages in schools or via the TV stations. It is a shame to lose out this human heritage. One of the ways to support the existence of these languages is through the advancement of today’s ICT. This kind of effort is very much needed as it enables the preservation of human heritage.

We have successfully gathered experts from different countries to share their knowledge and expertise on Natural Language Processing and Computational Linguistics especially in spoken languages. I would like to welcome you to our beautiful university, USM and certainly to Penang which boasts of its beautiful sandy beaches and rich cultural heritage embedded within the island. Penang was recently declared a UNESCO World Heritage Site. It is also known to be the “food paradise” of Malaysia. Do discover Penang that has a rich multicultural history, and our country, Malaysia that is truly Asia.

I would like to express my sincere thanks and appreciation to the dedicated work and cooperation given by the local committee, as well as the technical and professional support from LIG, MICA, the scientific committee and the financial support from our sponsors. Without this joint effort, SLTU 2010 workshop would not have materialized today.

Thank you. “Terima kasih”.
Foreword by SLTU’10 Chair

After the 1st workshop held in Vietnam in 2008, we are very pleased to welcome participants to this 2nd international workshop on Spoken Languages Technologies for Under-resourced Languages (SLTU), in Penang, Malaysia.

For 2010, we managed to gather researchers working on ASR, synthesis and translation for more than 20 under-resourced languages (portability issues, multilingual spoken language processing, fast resources acquisition, etc.). We are delighted to find so much interest in this SLTU 2010 workshop.

Overall, 22 papers have been selected for oral communications. The authors and co-authors originated from 17 countries (Bangladesh, Burma, Cambodia, Ethiopia, France, Germany, Hungary, India, Korea, Malaysia, Pakistan, Russia, Singapore, Spain, South-Africa, Vietnam, USA). We observed an increase in the quality of the papers compared to 2008 and the scientific committee has compiled a two-day scientific programme that covers areas of automatic translation, speech recognition and synthesis, as well as collection of language resources.

In addition, three prestigious speakers have been invited to give keynotes: Haizou Li, Ruhi Sarikaya and Alex Waibel. We wish to give special thanks to them. We would also like to thank all the members of the scientific committee for their help in making this event a scientifically recognized workshop. Special thanks also to the Universiti Sains Malaysia which accepted the responsibility to host the workshop and to all the local organizers.

Finally, we would like to extend our thanks to ISCA, AFCP, CNRS and Grenoble-INP for their strong support (notably, the support of ISCA and AFCP which allowed us to give grants to fund the venue of five PhD students).

Thanks and have an excellent workshop!
SLTU’10 Organising Committees

Workshop Chair
Laurent Besacier (LIG)

Workshop Co-chair
Eric Castelli (MICA)

Local Workshop Chair
Chan Huah Yong (USM)

Scientific Committee
Tanja Schultz, CMU, USA
Ruhi Sarikaya, IBM, USA
Haizhou Li, A-star, Singapore
Lori Lamel, LIMSI, France
Jean-Paul Haton, LORIA, France
Pascal Nocera, LIA, France
Bali Ranaivo, Multimedia University, Malaysia
Alvin W. Yeo, UNIMAS, Malaysia
Ch’ng Eng Siong, NTU, Singapore
Tan Tien Ping, USM, Malaysia
Eric Castelli, MICA, Vietnam
Geneviève Caelen-Haumont, MICA, Vietnam
Vincent Berment, LIG, France
Laurent Besacier, LIG, France

Local Organising Committee
Dennis Wong Chin Phang, USM, Malaysia
Faten Damanhouri, USM, Malaysia
Norlia Mustaffa, USM, Malaysia
Norliza Hani Md. Ghazali, USM, Malaysia
Nour Azimah Zulkapli, USM, Malaysia
Sabrina Tiun Abdullah, USM, Malaysia
Siti Khaotijah Mohammad, USM, Malaysia
Tan Ewe Hoe, USM, Malaysia
Tan Tien Ping, USM, Malaysia
Marilyn Lim Chien Hui, USM, Malaysia
Ruslan Ahmad, USM, Malaysia
Badriyah Che May, USM, Malaysia
Ramli Yahaya, USM, Malaysia
Syed Mohamad Syed Sahil, USM, Malaysia
Acknowledgement

The organising of SLTU Workshop 2010 is a joint effort of many individuals and organisations. They come from within and outside Universiti Sains Malaysia. Their combined technical, administrative and financial support as well as advisory and knowledge-based skills has enabled the Organising Committee to proceed with the various tasks smoothly and systematically. Hence, the Organiser of the 2nd International Workshop on Spoken Languages Technologies for Under-Resourced Languages would like to extend its heartfelt thanks and appreciation to the following and whoever concerned, for their collaboration and support in making this workshop a success:

Vice Chancellor, USM
Deputy Vice Chancellor, Academic and International Affairs, USM
Deputy Vice Chancellor, Students Affairs and Development, USM,
Deputy Vice Chancellor, Research and Innovation, USM
School of Computer Sciences, USM
Centre for Instructional Technology and Multimedia (PTPM), USM
Centre for Knowledge, Communication and Technology (PPKT), USM
Development Department, USM
Security Department, USM
Registry Department, USM
Busary Department, USM
Public Relation Unit, USM
Multimedia, Information, Communication and Applications (MICA) Research Center, Vietnam
Laboratoire d'Informatique de Grenoble (LIG), France

~Thank You~
Sponsors

Association Francophone de la Communication Parlée

Speech Processing Asian Network

Grenoble INP

International Speech Communication Association

Centre National de la Recherche Scientifique

National Library of Malaysia
**WORKSHOP CHAIR**
- Laurent Besacier, LIG, France
- Eric Castelli, Mica, Vietnam
- Chan Huah Yong, USM, Malaysia

**LOCAL ORGANISERS**
- Dennis Wong Chin Phang, USM, Malaysia
- Faten Damanhoori, USM, Malaysia
- Norlia Mustaffa, USM, Malaysia
- Norliza Hani Md. Ghazali, USM, Malaysia
- Nour Azimah Zulkapli, USM, Malaysia
- Sabrina Tiun Abdullah, USM, Malaysia
- Siti Khaotijah Mohammad, USM, Malaysia
- Tan Ewe Hoe, USM, Malaysia
- Tan Tien Ping, USM, Malaysia

**SCIENTIFIC COMMITTEE**
- Tanja Schultz, CMU, USA
- Ruhi Sarikaya, IBM, USA
- Haizhou Li, A-star, Singapore
- Lori Lamel, LIMSI, France
- Jean-Paul Haton, LORIA, France
- Pascal Nocera, LIA, France
- Bali Ranaivo, MMU, Malaysia
- Alvin W. Yeo, UNIMAS, Malaysia
- Ch'ng Eng Siong, NTU, Singapore
- Tan Tien Ping, USM, Malaysia
- Eric Castelli, MICA, Vietnam
- Vincent Berment, LIG, France
- Laurent Besacier, LIG, France

**ONLINE REGISTRATION**
http://www.mica.edu.vn/sltu-2010

---

**IMPORTANT DATES**

Paper submission due: 15 December, 2009
Notification of acceptance: 15 February, 2010
Registration due: 1 March, 2010
Workshop dates: 3-5 May, 2010

**SUBMISSION INFORMATION**

Upload your submission in PDF or MSWord format at: [http://www.mica.edu.vn/sltu-2010](http://www.mica.edu.vn/sltu-2010)
Submission should be at most 8000 words. We suggest to authors to use IEEE Computer Society Proceedings Manuscript Formatting Guidelines:

MSWord:
ftp://pubftp.computer.org/press/outgoing/proceedings/instruct.doc

Latex:
ftp://pubftp.computer.org/Press/Outgoing/proceedings/IEEE_CS_Latex.zip

---

**SLTU 2010**
School of Computer Sciences,
Universiti Sains Malaysia,
11800 Penang, Malaysia
Tel: +604-653 4386
Fax: +604-656 3244
mailto:sltu2010@cs.usm.my

*Picture is courtesy of Sabah Tourism Malaysia*
The SLTU is an abbreviation for “Spoken Languages Technologies for Under-Resourced Languages” that represents a group of researchers that are interested in natural language processing (NLP) and computational linguistics (CL) particularly in spoken languages where their data resources are rather limited such as the native language of Sabah and Sarawak, Vietnamese language, Cambodian language, languages from Africa region and other parts of the world. Apart from that, researchers in this field are also interested in foreign dialects. Therefore, Malay dialects and Malaysian English spoken by local people have also been included. More details are available on http://www.mica.edu.vn/sltu-2010/

OBJECTIVES

SLTU research workshop will focus on spoken language processing for under-resourced languages. The main goals of the workshop are as follows:

- To attract the interest of whomever who is involved in spoken language processing for under-resourced languages, in the computational and theoretical point of view.
- To preserve and encourage the research in language studies with limited resources in Malaysia such as the Malay and Chinese dialects, aborigines languages, native languages in Sabah (example: Kadazan) and Sarawak (example: Iban) in a more scientific way.

CONFERENCE TOPICS

All areas related to the processing – computational and theoretical point of view – of under-resourced languages:

- Language resources development, acquisition, and representation: dictionary, language model, text and speech corpora, parallel corpora, grammars
- Automatic speech recognition of low-resourced languages, as well as indigenous languages, dialects, etc.
- Multilingual spoken language processing (recognition, synthesis)
- Machine translation

SPONSORED BY

To introduce the SLTU group of researchers and their research work to all the NLP community, with the hope of enhancing international research excellence and cooperation, fast resources acquisition (speech, text, lexicons, parallel corpora)