



Main Menu

[IWSLT 2012](#)

[What's New?](#)

[Important Dates](#)

[Workshop Registration](#)

[Evaluation Campaign](#)

[TED Task](#)

[OLYMPICS Task](#)

[Evaluation Server](#)

[Submission](#)

[Registration](#)

[Keynote Talks](#)

[Program](#)

[Venue](#)

[About Hong Kong](#)

[HKUST iPhone App](#)

[Accommodation](#)

[Organizers](#)

[Acknowledgments](#)

[Related Events](#)

[Contact](#)

Login Form

User Name

Password

Remember Me

[Log in](#)

[Forgot your password?](#)

[Forgot your username?](#)

[Create an account](#)

Call for Participation

The International Workshop on Spoken Language Translation (IWSLT) is a yearly scientific workshop, associated with an open evaluation campaign on spoken language translation, where both scientific papers and system descriptions are presented. The 9th International Workshop on Spoken Language Translation will take place in Hong Kong on December 6-7, 2012. The workshop will be held just before the 24th International Conference on Computational Linguistics (COLING) on December 8-15, 2012 in Mumbai, India.



Evaluation Campaign

The IWSLT offers challenging research tasks and open benchmarks to develop and test speech recognition and machine translation technology. In this edition, the IWSLT will offer ASR, SLT and MT tracks around the large-vocabulary TED talks domain, as well as one MT track on the small-vocabulary Olympics domain. For all tracks, the organizers will supply participants with training and development data. The official training data are already released. Participants interested in the TED talk tasks can access the training and development data sets [here](#). Participants of the OLYMPICS task have to sign an end-user agreement in order to [get access to the data sets](#). For details on the OLYMPICS task, check [here](#).

Scientific Papers

The IWSLT invites submissions of scientific papers to be published in the workshop proceedings and presented in dedicated technical sessions of the workshop, either in oral or poster form. The workshop welcomes high quality contributions covering theoretical and practical issues in the fields of automatic speech recognition and machine translation, that are applied to spoken language translation. Possible topics include, but are not limited to:

- Speech and text MT
- Integration of ASR and MT
- MT and SLT approaches
- MT and SLT evaluation
- Language resources for MT and SLT
- Open source software for MT and SLT
- Adaptation in MT
- Simultaneous speech translation
- Speech translation of lectures
- Efficiency in MT
- Stream-based algorithms for MT
- Multilingual ASR and TTS
- Rich transcription of speech for MT
- Translation of on-verbal events

Submitted manuscripts will be peer-reviewed. Authors of accepted papers are requested to present their paper at the workshop.



Main Menu

IWSLT 2012

What's New?

Important Dates

Workshop Registration

Evaluation Campaign

TED Task

OLYMPICS Task

Evaluation Server

Submission

Registration

Keynote Talks

Program

Venue

About Hong Kong

HKUST iPhone App

Accommodation

Organizers

Acknowledgments

Related Events

Contact

Login Form

User Name

Password

Remember Me

Log in

[Forgot your password?](#)

[Forgot your username?](#)

[Create an account](#)

Organizers

Steering Committee

Marcello Federico (Fondazione Bruno Kessler, Italy)

Eiichiro Sumita (National Institute of Information and Communications Technology, Japan)

Alex Waibel (Carnegie Mellon University, USA & Karlsruhe Institute of Technology, Germany)

Workshop Chairs

Dekai Wu (The Hong Kong University of Science & Technology, Hong Kong)

Eiichiro Sumita (National Institute of Information and Communications Technology, Japan)

Evaluation Chairs

Marcello Federico (Fondazione Bruno Kessler, Italy)

Tiejun Zhao (Harbin Institute of Technology, China)

Michael Paul (National Institute of Information and Communications Technology, Japan)

Technical Program Chairs

Chengqing Zong (National Laboratory of Pattern Recognition, Chinese Academy of Sciences, China)

Chiori Hori (National Institute of Information and Communications Technology, Japan)

Local Arrangement

Dekai Wu (The Hong Kong University of Science & Technology, Hong Kong)

Program Committee

Alexandre Allauzen (LIMSI-CNRS, France)

Loic Barrault (LIUM, France)

Laurent Besacier (LIG, France)

Francisco Casacuberta (ITI-UPV, Spain)

Mauro Cettolo (FBK, Italy)

Boxing Chen (NRC-IIT, Canada)

Xiaodong He (Microsoft Research, China)

Teresa Herrmann (KIT, Germany)

Kevin Kilgour (KIT, Germany)

Katrin Kirchhoff (University of Washington, USA)

Florian Kraft (KIT, Germany)

Patrik Lambert (University of Le Mans, France)

Yves Lepage (Waseda University, Japan)

Qun Liu (ICT, China)

Shigeki Matsuda (NICT, Japan)

Arne Mauser (RWTH, Germany)

Mohammed Mediani (KIT, Germany)

Jan Niehues (KIT, Germany)

Wade Shen (MIT/LL, USA)

Xiaodong Shi (Xiamen University, China)

Hwee Tou Ng (NUS, Singapore)

Isabel Trancoso (INESC-ID, Portugal)

Haifeng Wang (Baidu, China)
Taro Watanabe (NICT, Japan)
Min Zhang (I2R, Singapore)
Ming Zhou (Microsoft Research, China)

YOU ARE HERE: [HOME](#) [▶ ORGANIZERS](#)

[TOP](#)

HKUST Human Language Technology Center



Main Menu

[IWSLT 2012](#)

[What's New?](#)

[Important Dates](#)

[Workshop Registration](#)

[Evaluation Campaign](#)

[TED Task](#)

[OLYMPICS Task](#)

[Evaluation Server](#)

[Submission](#)

[Registration](#)

[Keynote Talks](#)

[Program](#)

[Venue](#)

[About Hong Kong](#)

[HKUST iPhone App](#)

[Accommodation](#)

[Organizers](#)

[Acknowledgments](#)

[Related Events](#)

[Contact](#)

Login Form

User Name

Password

Remember Me

[Log in](#)

[Forgot your password?](#)

[Forgot your username?](#)

[Create an account](#)

Acknowledgments

Supporting organizations:



National Institute of Communication and Information Technology, Japan



Fondazione Bruno Kessler - IRST, Italy



Karlsruhe Institute of Technology, Germany



Center for the Evaluation of Language and Communication Technologies, Italy



HKUST Human Language Technology Center, Hong Kong



Harbin Institute of Technology, China

Supporting projects:



EU-Bridge Project

IWSLT 2012 is endorsed by:

[Special Interest Group for Machine Translation \(SIG MT\)](#)

[European Association for Machine Translation \(EAMT\)](#)



Main Menu

IWSLT 2012

What's New?

Important Dates

Workshop Registration

Evaluation Campaign

TED Task

OLYMPICS Task

Evaluation Server

Submission

Registration

Keynote Talks

Program

Venue

About Hong Kong

HKUST iPhone App

Accommodation

Organizers

Acknowledgments

Related Events

Contact

Evaluation Campaign

IWSLT proposes challenging research tasks and an open experimental infrastructure for the scientific community working on spoken and written language translation. The IWSLT 2012 Evaluation Campaign includes the following tasks:

TED Task

ASR track : automatic transcription of talks from audio to text (in English)

SLT track: speech translation of talks from audio (or ASR output) to text (from English to French)

MT track : text translation of talks for two language pairs plus ten optional language pairs:

OLYMPICS Task

HIT track : text translation of the sentences taken from the Olympics domain (from Chinese sentences into English)

Participants of the OLYMPICS task have to sign an end-user agreement in order to [get access to the data sets](#).

The organizations which register for a task and receive the data have a commitment to submit runs for the test data set, attend the evaluation campaign and give a presentation at the workshop. Failure to show up at the workshop or to submit runs for the test data set will jeopardize your site for future IWSLT evaluations.

Login Form

User Name

Password

Remember Me

Log in

[Forgot your password?](#)

[Forgot your username?](#)

[Create an account](#)



Main Menu

IWSLT 2012

What's New?

Important Dates

Workshop Registration

[Evaluation Campaign](#)

[TED Task](#)

OLYMPICS Task

Evaluation Server

Submission

Registration

Keynote Talks

Program

Venue

About Hong Kong

HKUST iPhone App

Accommodation

Organizers

Acknowledgments

Related Events

Contact

Login Form

User Name

Password

Remember Me

[Log in](#)

[Forgot your password?](#)

[Forgot your username?](#)

[Create an account](#)

TED Task

IWSLT proposes challenging research tasks and an open experimental infrastructure for the scientific community working on spoken and written language translation. The IWSLT 2012 Evaluation Campaign includes the TED Task, that is the translation of TED Talks, a collection of public speeches on a variety of topics. Three tracks are proposed addressing different research tasks:

[ASR track](#) : automatic transcription of talks from audio to text (in English)

[SLT track](#): speech translation of talks from audio (or ASR output) to text (from English to French)

[MT track](#) : text translation of talks for two language pairs plus ten optional language pairs:

official: from English to French and from Arabic to English

optional: from German, Dutch, Polish, Portuguese-Brazil, Romanian, Russian, Slovak, Slovenian, Turkish and Chinese to English

Main challenges of the proposed tracks are:

Open domain ASR, clean transcription of spontaneous speech, detection and removal of non-words, and talk style and topic adaptation.

Open domain SLT, translation of speech or ASR output into true-case punctuated text, and talk style and topic adaptation.

Open domain MT between distant languages, and talk style and topic adaptation.

Training of MT systems and language models for ASR is constrained to data supplied by the organizers. As for ASR acoustic modeling no training data are distributed, participants are allowed to use any publicly available data recorded before 31 December 2010.

[ASR Track](#) (English)

Input format: pre-segmented SPHERE files

Output format: CTM (description and example [here](#), at section ".ctm"), no case, no punctuation

Coding: UTF-8

[Submission guidelines](#)

Evaluation: WER

Training data for language modeling:

in-domain: [monolingual English text](#)

out-of-domain: [supplied monolingual texts](#)

Development data:

dev2010:

[audio+reference transcripts, 177MB](#)

[automatic transcripts v1, 191kB](#)

test2010:

[audio+reference transcripts, 306MB](#)

[automatic transcripts v1, 291kB](#)

[Scoring Package](#)

Test data:

tst2011, tst2012:

[audio + sentence segmentation, 273MB](#)

[SLT Track](#) (English to French)

Input format: pre-segmented SPHERE files or supplied ASR outputs (transcript, SLF lattice)

Output format: NIST XML format, true case with punctuation ([example](#))

Coding: UTF-8

[Submission guidelines](#)

Evaluation: BLEU and subjective ranking

Training data:

in-domain: from the [WIT³](#) website

out-of-domain: [supplied monolingual and parallel texts](#)

Development data:

dev2010:

[audio+reference transcripts, 177MB](#)

[lattices by LIUM, 2.5MB](#)

[automatic transcripts v1, 191kB](#)

test2010:

[audio+reference transcripts, 306MB](#)

[lattices by LIUM, 4.0MB](#)

[automatic transcripts v1, 291kB](#)

Test data:

tst2011, tst2012:

[automatic transcripts, 1.8MB](#)

MT Track (English to French, Arabic to English + 10 additional language pairs)

Input format: NIST XML format, true case with punctuation

Output format: NIST XML format, true case with punctuation ([example](#))

Coding: UTF-8

[Submission guidelines](#)

Evaluation: BLEU and subjective ranking

Training data:

in-domain: from the [WIT³](#) website

out-of-domain: [supplied monolingual and parallel texts](#)

Development data: from the [WIT³](#) website

Test data: from the [WIT³](#) website

Test References

The references of tst2011 set for ASR, SLT and MT tracks are available to participants [here](#).

-

Registration

To participate in one or more tracks please fill in the [Registration Form](#) and send it by e-mail to iwslt2012.ted AT gmail DOT com

Participants are requested to submit a system paper describing their work and to present it at the workshop.

Acknowledgments

TED talks data are copyright of TED Conference LLC and distributed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 license

Google Books ngrams are copyright of Google Inc. and distributed under a Creative Commons Attribution 3.0 license

WMT12 data are kindly supplied by the NAACL 2012 7th Workshop on Statistical Machine Translation

MultiUN data are kindly supplied by the EuroMatrixPlus project

Supplied out-of-domain monolingual and parallel training data

monolingual

[News Commentary v7](#) from WMT 2012 (en, fr), 20MB

[News Crawl](#) from WMT 2012 (en, fr), 7GB

[Europarl v7](#) (en, fr), 215MB

LDC2011T07 [English Gigaword Fifth Edition](#)

LDC2009T28 [French Gigaword Second Edition](#)

[Google books grams](#) (en, fr, de, ru, zh)

parallel

[News Commentary v7](#) from WMT 2012 (en-fr, de-en), 32MB

[Europarl v7](#) (en-fr, de-en, nl-en, pl-en, pt-en, ro-en), 900MB

WMT 2012 [UN French-English](#) (en-fr), 1.1GB

[MultiUN](#) (ar-en, en-fr, de-en, ru-en, zh-en), 800MB-1.7GB

WMT 2012 [10⁹ French-English](#) (en-fr), 2.3GB

SETimes parallel corpus: [ro-en](#) (65MB), [tr-en](#) (62MB)

TED Task Evaluation Team

Marcello Federico (FBK, Italy), Evaluation Chair

Michael Paul (NICT, Japan), Evaluation Chair

Mauro Cettolo (FBK, Italy), MT data processing (surname at fbk dot eu)

Sebastian Stueker (KIT, Germany), ASR data processing

Luisa Bentivogli (FBK, Italy), Subjective evaluation

Giovanni Moretti (CELCT, Italy), Subjective evaluation

YOU ARE HERE: [HOME](#) ▶ [EVALUATION CAMPAIGN](#) ▶ [TED TASK](#)

[TOP](#)

HKUST Human Language Technology Center



Main Menu

-
- IWSLT 2012
-
- What's New?
-
- Important Dates
-
- Workshop Registration
-
- Evaluation Campaign**
-
- TED Task
-
- OLYMPICS Task**
-
- Evaluation Server
-
- Submission
-
- Registration
-
- Keynote Talks
-
- Program
-
- Venue
-
- About Hong Kong
-
- HKUST iPhone App
-
- Accommodation
-
- Organizers
-
- Acknowledgments
-
- Related Events
-
- Contact

Login Form

User Name

Password

Remember Me

Log in

Forgot your password?
 Forgot your username?
 Create an account

OLYMPICS Task

The OLYMPICS task is carried out using parts of the [HIT Olympic Trilingual Corpus \(HIT\)](#), a multilingual corpus that covers 5 domains (*traveling, dining, sports, traffic and business*) that are closely related to the Beijing 2008 Olympic Games. The HIT corpus contains around 2.8 million words in total.

Moreover, the *Basic Travel Expression Corpus (BTEC)*, a multilingual speech corpus containing tourism-related sentences, is provided as an additional training corpus. The BTEC corpus consists of 20k sentences including the evaluation data sets of previous IWSLT evaluation campaigns.

The monolingual and bilingual language resources that should be used to train the translation engines for the primary runs are limited to the supplied HIT and BTEC corpora. This includes all supplied development sets, i.e., you are free to use these data sets as you wish for tuning model parameters or as training bitext, etc. All other languages resources besides the ones for the given translation task, such as any additional dictionaries, word lists, bitext corpora such as the ones provided by LDC, should be treated as "additional language resources".

MT track: text translation of Chinese sentences into English

official: training data limited to supplied data only.

optional: usage of additional language resources (please provide a detailed description of the resources at run submission time).

Registration:

Participants of the OLYMPICS task have to sign an [end-user agreement](#) in order to [get access to the data sets](#). In addition, participants of the OLYMPICS Task are requested to fill-in and send the following form to 'iwslt2012.oly AT gmail DOT com'. A single organization can register more times as different participants. Registration is possible until the [testdata run submission deadline](#).

```

--- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< ---

```

Registration Form IWSLT 2012 Evaluation Campaign - OLYMPICS Task

Organization Name (long, e.g. Fondazione Bruno Kessler):
 Organization Name (short, e.g. FBK):
 Country:
 Contact Person (e.g. John Smith):
 E-mail address:
 Track: + MT (Chinese to English): YES

I understand that participants in this evaluation are requested to submit a system paper describing their work and to present it at the workshop.

```

--- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< --- 8< ---

```

Training Corpus:

parallel text
 each line consists of three fields divided by the character '\'

```

Format: <SENTENCE_ID>\<PARAPHRASE_ID>\<MT_TRAINING_SENTENCE>

TRAIN_00001\01\This is the first training sentence
TRAIN_00002\01\This is the second training sentence.
...

```

Data sets:

HIT: 50K Chinese-English sentence pairs randomly selected from the HIT corpus
 BTEC: 20K Chinese-English sentence pairs randomly selected from the BTEC corpus

Corpus specifications:

coding: UTF-8
text is case-sensitive and includes punctuation

Develop Corpus:

parallel text
each line consists of three fields divided by the character '\'

```
Format: <SENTENCE_ID>\<PARAPHRASE_ID>\<MT_DEVELOP_SENTENCE>

DEV_001\01\1st reference translation for 1st input.
DEV_001\02\2nd reference translation for 1st input
...
DEV_002\01\1st reference translation for 2nd input
DEV_002\02\2nd reference translation for 2nd input
...
```

Data sets:

HIT: 2 develop sets, 1000 Chinese-English sentence pairs each, single reference
BTEC: 8 develop sets, 500 Chinese-English sentence pairs each, multiple references

Corpus specifications:

coding: UTF-8
text is case-sensitive and includes punctuation

Evaluation Corpus:

source language (Chinese) text
each line consists of three fields divided by the character '\'

```
Format: <SENTENCE_ID>\<PARAPHRASE_ID>\<MT_EVAL_SENTENCE>

HIT_TST_IWSLT12_0001\01\1st source language input sentence.
HIT_TST_IWSLT12_0002\01\2nd source language input sentence.
...
```

Data sets:

HIT: 1 evaluation set, 1000 Chinese sentences, MT engine input

Corpus specifications:

coding: UTF-8
text is case-sensitive and includes punctuation

Context Annotations:

for each sentence of the HIT corpus, context information on the *type of text* (dialogue, samples, explanation), *scene* (airplane, airport, restaurant, water/winter sports, etc.), *topic* (ask about traffic conditions, bargain over a price, front desk customer service, etc.), and the *speaker* (customer, clerk, passenger, receptionist, travel agent, etc.) is provided. Each line of the INFO files consists of three fields divided by the character '\' where the context annotations are divided by the character '|'.

```
Format: <SENTENCE_ID>\<PARAPHRASE_ID>\<TEXTTYPE>|<SCENE>|<TOPIC>|<SPEAKER>

TRAIN_00004\01\dialogue|train|when you are feeling ill|conductor
TRAIN_00005\01\dialogue|train|when you are feeling ill|Foreign guest
...
```

Submission Guidelines:

Each participant has to submit at least one run for the OLYMPICS translation task s/he registered for
multiple run submissions are allowed, but participants must explicitly indicate one PRIMARY run for each track. All other run submissions are treated as CONTRASTIVE runs. In case that none of the runs is marked as PRIMARY, the latest submission (according to the file time-stamp) for the respective track will be used as the PRIMARY run.
Runs have to be submitted as a gzipped TAR archive (format see below) and e-mailed to **iwslt2012.oly AT gmail DOT com**.

TAR archive file structure:

```
<UserID> = user ID of participant
<Set> = testset_IWSLT12
<Task> = MT_ZhEn
```

Example:

```
nict/testset_IWSLT12.MT_ZhEn.nict.primary.txt
/testset_IWSLT12.MT_ZhEn.nict.contrastive1.txt
```

Run submission file format:

```
Format: <SENTENCE_ID>\01<TRANSLATED_SENTENCE>

HIT_TST_IWSLT12_0001\01\This is the first translation
HIT_TST_IWSLT12_0002\01\This is the second translation.
...
```

Re-submitting your runs is allowed as far as the mails arrive BEFORE the submission deadline. In case that multiple TAR archives are submitted by the same participant, only runs of the most recent submission mail will be used for the IWSLT 2012 evaluation and previous mails will be ignored.

Download:

The references of the *testset_IWSLT12* evaluation data set are available to participants [here](#).



Main Menu

IWSLT 2012

What's New?

Important Dates

Workshop Registration

Evaluation Campaign

TED Task

OLYMPICS Task

Evaluation Server

Submission

Registration

Keynote Talks

Program

Venue

About Hong Kong

HKUST iPhone App

Accommodation

Organizers

Acknowledgments

Related Events

Contact

Login Form

User Name

Password

Remember Me

Log in

[Forgot your password?](#)

[Forgot your username?](#)

[Create an account](#)

Venue

IWSLT 2012 will be held in rooms 4018 (near lifts 2 and 3) and 5620 (near lifts 31 and 32) on the scenic campus of HKUST, [The Hong Kong University of Science and Technology](#).

Most of the academic rooms are in the same huge main academic building. Navigation around campus is typically given in terms of the nearest lift numbers. The web-based [HKUST Path Advisor](#) can be extremely useful in locating specific lifts and rooms.

For iPhone users, we recommend the [m.HKUST iPhone app](#) which contains the campus map and useful contacts in HKUST.

Wi-fi will be available in the conference rooms and at central locations of the HKUST campus for those belonging to the worldwide [eduroam](#) consortium.



How to get to HKUST

Public transportation in Hong Kong is very efficient, and getting to and from HKUST by [taxi](#), the [MTR](#) metro, or [bus](#) is easy, as is accessible by car. The [map](#) shows you the main routes connecting Kowloon and Hong Kong island to our campus.

The main gate of the campus features a large and easily recognizable sign with the university's name written in both English and Chinese. Past the main gate, a gently sloping driveway leads to the entrance piazza with a large red sundial at its center.

Click on image to enlarge the HKUST location map.



Remember me

[Log in](#)

[Forgot your password?](#)

[Forgot your username?](#)

[Create an account](#)

LOCATION MAP
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY



MTR Stations with bus or green minibus service to HKUST
提供往科大巴士或綠色專線小巴服務的港鐵車站

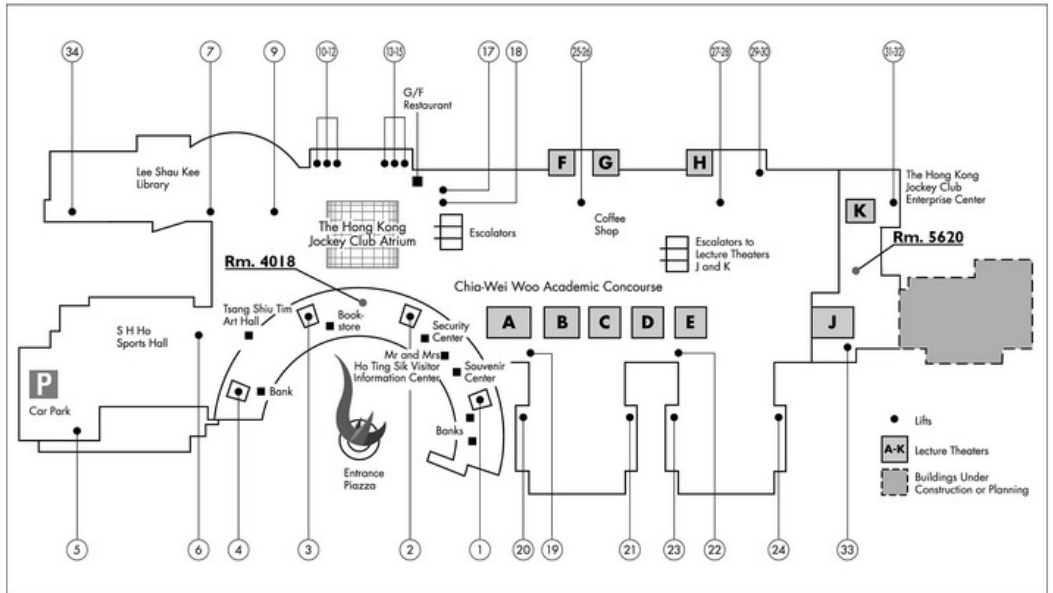
Diamond Hill 鑽石山:	91, 91M
Choi Hung 彩虹:	91, 91M, 11, 11S
Ngau Tau Kok 牛頭角:	104
Tiu Keng Leng 調景嶺:	792M
Hang Hau 坑口:	11, 11M, 11S
Po Lam 寶琳:	91M

Transportation from airport to HKUST:
For passengers with bulky luggage, taking a taxi to HKUST direct is recommended. Those with simple luggage may take Airport Bus A22 to Lam Tin, and change for taxi to HKUST.

Bus Routes 巴士路線
 Green Minibus Routes 綠色專線小巴路線

Click on image to enlarge the HKUST lift map.

Academic Building Directory
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY



MTR

The Mass Transit Railway (MTR) is an underground subway (metro) system serving Hong Kong island, Kowloon, and the airport.

The MTR is the easiest way for visitors to get around Hong Kong. Tickets can be purchased from machines in the stations on a trip-by-trip basis or you can purchase multi-use Octopus cards at MTR ticket windows.

The Choi Hung and Hang Hau MTR stations are the closest to the University, and you can board a bus or taxi (details below) to the main entrance of HKUST.

There is a train every 2 minutes. Travelling from North Point station to Hang Hau station takes approximately 20 minutes.

Taxis

Taxis provide the simplest way to get to the University, and are easily available at any MTR station, from hotels and most areas of Hong Kong. From the airport, you can take a taxi at the airport taxi stand and expect to pay approximately HK\$380 (about US\$49) and take approximately 40 minutes for the journey. Note that only HK dollars will be accepted by your driver.

Print a copy of this page and the map and bring it with you in case you need to show the route or University name and address in Chinese to the taxi driver.

The taxi drop-off point on campus is near the red sundial. For on-campus housing destinations, ask the security guard at the main gate to give the taxi driver directions.

Buses / Minibuses

From Choi Hung MTR station:	91 or 91M (MTR Exit C2; travel time: ~20 mins) 11 or 11S minibus (MTR Exit C1; travel time: ~15 mins)
From Hang Hau MTR station:	91M (MTR Exit B2; travel time: ~10 mins) 11, 11M or 11S minibus (MTR Exit B1; travel time: ~10 mins)
From Po Lam MTR station:	91M (MTR Exit A2; travel time: ~25 mins) 12 minibus (MTR Exit B1; travel time: ~40 mins)
From Ngau Tau Kok MTR station:	104 minibus (MTR Exit A; travel time: ~30 mins)
From Tiu Keng Leng MTR station:	792M (MTR Exit A1; travel time: ~30 mins)
From Sai Kung:	792M (travel time: ~20 mins) 12 minibus (travel time: ~15 mins)

Driving (General directions)

From Hong Kong Island : Take the Eastern Harbour Crossing towards Lam Tin. Continue straight after the roundabout, and get in the right lane towards the Tseung Kwan O Tunnel, following signs for Hang Hau. At the roundabout for Hang Hau, take the Ying Yip Road exit, following the road past the Clearwater Bay Film Studio onto Clearwater Bay Road, and follow signs for HKUST.

From Kowloon : Take Prince Edward Road towards Kwun Tong; after Choi Hung Estates, stay left and follow signs to Clearwater Bay. Take Clearwater Bay road and follow signs to HKUST.

From Shatin : Take Cairns or Lion Rock Tunnel, staying left onto Lung Cheung Road. Follow signs to Sai Kung. Take Clearwater Bay road, following signs for HKUST.



Main Menu

- [IWSLT 2012](#)
- [What's New?](#)
- [Important Dates](#)
- [Workshop Registration](#)
- [Evaluation Campaign](#)
- [TED Task](#)
- [OLYMPICS Task](#)
- [Evaluation Server](#)
- [Submission](#)
- [Registration](#)
- [Keynote Talks](#)
- [Program](#)**
- [Venue](#)
- [About Hong Kong](#)
- [HKUST iPhone App](#)
- [Accommodation](#)
- [Organizers](#)
- [Acknowledgments](#)
- [Related Events](#)
- [Contact](#)

Login Form

User Name

Password

Remember Me

[Log in](#)

[Forgot your password?](#)

[Forgot your username?](#)

[Create an account](#)

Program

Full IWSLT 2012 proceedings in printable form [\[pdf\]](#)

Thursday, December 6th, 2012

08:30-09:15h	Workshop Registration
09:15-09:30h	Welcome remarks — room 4018, near lift 2/3 Eiichiro Sumita, Dekai Wu
09:30-10:15h	Keynote speech 1 (<i>chair: Michael Paul</i>) — room 4018, near lift 2/3 "Toward Universal Network-based Speech Translation" [abstract] [slides] Dr. Chai Wutiwivatchai, National Electronics and Computer Technology Center (NECTEC)
10:15-10:35h	Coffee Break
10:35-12:00h	Oral Session - Evaluation Campaign I (<i>chair: Graham Neubig</i>) — room 4018, near lift 2/3
10:35-11:20h	"Overview of the IWSLT 2012 Evaluation Campaign" [pdf] [slides] Marcello Federico, Mauro Cettolo, Luisa Bentivogli, Michael Paul, Sebastian Stüker
11:20-11:40h	"The NICT ASR System for IWSLT2012" [pdf] [slides] Hitoshi Yamamoto, Youzheng Wu, Chien-Lin Huang, Xugang Lu, Paul R. Dixon, Shigeki Matsuda, Chiori Hori, Hideki Kashioka
11:40-12:00h	"The KIT Translation systems for IWSLT 2012" [pdf] [slides] Mohammed Mediani, Yuqi Zhang, Thanh-Le Ha, Jan Niehues, Eunah Cho, Teresa Herrmann, Rainer Kärgey, Alexander Waibel
12:00-14:00h	Lunch Break
14:00-16:00h	Oral Session - Evaluation Campaign II (<i>chair: Christian Saam</i>) — room 5620, near lift 31/32
14:00-14:20h	"The UEDIN Systems for the IWSLT 2012 Evaluation" [pdf] [slides] Eva Hasler, Peter Bell, Arnab Ghoshal, Barry Haddow, Philipp Koehn, Philipp McInnes, Steve Renals, Pawel Swietojanski
14:20-14:40h	"The NAIST Machine Translation System for IWSLT2012" [pdf] [slides] Graham Neubig, Kevin Duh, Masaya Ogushi, Takatomo Kano, Tetsuo Kiso, Sakriani Sakti, Tomoki Toda, Satoshi Nakamura
14:40-15:00h	"FBK's Machine Translation Systems for IWSLT 2012's TED Lectures" [pdf] [slides] Nicholas Ruiz, Arianna Bisazza, Roldano Cattoni, Marcello Federico
15:00-15:20h	Coffee Break
15:20-15:40h	"The RWTH Aachen Speech Recognition and Machine Translation System for IWSLT 2012" [pdf] [slides] Stephan Peitz, Saab Mansour, Markus Freitag, Minwei Feng, Matthias Huck, Joern Wuebker, Malte Nuhn, Markus Nußbaum-Thom, Hermann Ney

15:40-16:00h	"The HIT- LTRC Machine Translation System for IWSLT 2012" [pdf] [slides] Xiaoning Zhu, Yiming Cui, Conghui Zhu, Tiejun Zhao, Hailong Cao
16:00-17:30h	Poster Session I (<i>chair: Michael Paul</i>) — room 5620, near lift 31/32 <ul style="list-style-type: none"> • "FBK @ IWSLT 2012 - ASR track" [pdf] Daniele Falavigna, Roberto Gretter, Fabio Brugnara, Diego Giuliani • "The 2012 KIT and KIT-NAIST English ASR Systems for the IWSLT Evaluation" [pdf] Christian Saam, Christian Mohr, Kelvin Kilgour, Michael Heck, Matthias Sperber, Keigo Kubo, Sebastian Stüker, Sakriani Sakti, Graham Neubig, Tomoki Toda, Satoshi Nakamura, Alexander Waibel • "The KIT-NAIST (Contrastive) English ASR System for IWSLT 2012" [pdf] Michael Heck, Keigo Kubo, Matthias Sperber, Sakriani Sakti, Sebastian Stüker, Christian Saam, Kelvin Kilgour, Christian Mohr, Graham Neubig, Tomoki Toda, Satoshi Nakamura, Alexander Waibel • "EBMT System of Kyoto University in OLYMPICS Task at IWSLT 2012" [pdf] Chengui Chu, Toshiaki Nakazawa, Sadao Kurohashi • "The LIG English to French Machine Translation System for IWSLT 2012" [pdf] Laurent Besancier, Benjamin Lecouteux, Marwen Azouzi, Quan Luong Ngoc • "The MIT-LL/AFRL IWSLT-2012 MT System" [pdf] Jennifer Drexler, Wade Shen, Timothy Anderson, Brian Ore, Ray Slyh, Eric Hansen, Terry Gleason • "Minimum Bayes-Risk Decoding Extended with Two Methods: NAIST-NICT at IWSLT 2012" [pdf] Hiroaki Shimizu, Masao Utiyama, Eiichiro Sumita, Satoshi Nakamura • "The NICT Translation System for IWSLT 2012" [pdf] Andrew Finch, Ohnmar Htun, Eiichiro Sumita • "TED Polish-to-English translation system for the IWSLT 2012" [pdf] Krzysztof Marasek • "Forest-to-String Translation using Binarized Dependency Forest for IWSLT 2012 OLYMPICS Task" [pdf] Hwidong Na, Jong-Hyeok Lee • "Romanian to English Automatic MT Experiments at IWSLT12" [pdf] Stefan Dumitrescu, Radu Ion, Dan Ștefănescu, Tiberiu Boroș, Dan Tufis • "The TÜBİTAK Statistical Machine Translation System for IWSLT 2012" [pdf] Coskun Mermer, Hamza Kaya, İlknur Durgar El-Kahlout, Mehmet Ugur Dogan <p>+ 7 poster presentations of the papers presented in the oral sessions</p>
18:00h	Social Event — lion dance, double decker bus to Tsimshatsui, traditional sailing junk to Lamma island, Chinese seafood banquet on Lamma

Friday, December 7th, 2012

09:30-10:15h	Keynote speech 2 (<i>chair: Chiori Hori</i>) — room 5620, near lift 31/32 "Who Can Understand Your Speech Better — Deep Neural Network or Gaussian Mixture Model?" [abstract] [slides] Dr. Dong Yu, Microsoft Research
10:15-10:40h	Coffee Break
10:40-12:00h	Oral Session - Technical Papers I: ASR (<i>chair: Chiori Hori</i>) — room 5620, near lift 31/32
10:40-11:00h	"Active Error Detection and Resolution for Speech-to-Speech Translation" [pdf] [slides] Rohit Prasad, Rohit Kumar, Sankaranarayanan Ananthakrishnan, Wei Chen, Sanjika Hewavitharana, Matthew Roy, Frederick Choi, Aaron Challenner, Enoch Kan, Arvind Neelakantan, Prem Natarajan
11:00-11:20h	"A Method for Translation of Paralinguistic Information" [pdf] [slides] Takatomo Kano, Sakriani Sakti, Shinnosuke Takamichi, Graham Neubig, Tomoki Toda, Satoshi Nakamura
11:20-11:40h	"Continuous Space Language Models using Restricted Boltzmann Machines" [pdf] [slides] Jan Niehues, Alexander Waibel

11:40-12:00h	"Focusing Language Models For Automatic Speech Recognition" [pdf] [slides] Daniele Falavigna, Roberto Gretter
12:00-13:30h	Lunch Break
13:30-14:15h	Keynote speech 3 (<i>chair: Eiichiro Sumita</i>) — room 5620, near lift 31/32 "Head Finalization: Translation from SVO to SOV" [abstract] [slides] Prof. Hideki Isozaki, Okayama Prefectural University
14:15-15:35h	Oral Session - Technical Papers II: MT (<i>chair: Dekai Wu</i>) — room 5620, near lift 31/32
14:15-14:35h	"Simulating Human Judgment in Machine Translation Evaluation Campaigns" [pdf] [slides] Philipp Koehn
14:35-14:55h	"Semi-supervised Transliteration Mining from Parallel and Comparable Corpora" [pdf] [slides] Walid Aransa, Hoger Schwenk, Loic Barrault
14:55-15:15h	"A Simple and Effective Weighted Phrase Extraction for Machine Translation Adaptation" [pdf] [slides] Saab Mansour, Hermann Ney
15:15-15:35h	"Applications of Data Selection via Cross-Entropy Difference for Real-World Statistical Machine Translation" [pdf] [slides] Amittai Axelrod, QingJun Li, William D. Lewis
15:35-16:00h	Coffee Break
16:00-17:30h	Poster Session II (<i>chair: Philipp Koehn</i>) — room 5620, near lift 31/32 <ul style="list-style-type: none"> • "A Universal Approach to Translating Numerical and Time Expressions" [pdf] Mei Tum, Yu Zhou, Chengqing Zong • "Evaluation of Interactive User Corrections for Lecture Transcription" [pdf] Henrich Kolkhorst, Jevin Kilgour, Sebastian Stüker, Alexander Waibel • "Factored Recurrent Neural Network Language Model in TED Lecture Transcription" [pdf] Youzheng Wu, Hitoshi Yamamoto, Xugang Lu, Shigeki Matsuda, Chiori Hori, Hideki Kashioka • "Incremental Adaptation Using Translation Information and Post-Editing Analysis" [pdf] Frederic Blain, Holger Schwenk, Jean Senellart • "Interactive-Predictive Speech-Enabled Computer-Assisted Translation" [pdf] Shahram Khadivi, Zeinab Vakil • "MDI Adaptation for the Lazy: Avoiding Normalization in LM Adaptation for Lecture Translation" [pdf] Nick Ruiz, Marcello Federico • "Segmentation and Punctuation Prediction in Speech Language Translation Using a Monolingual Translation System" [pdf] Eunah Cho, Jan Niehues, Alexander Waibel • "Sequence Labeling-based Reordering Model for Phrase-based SMT" [pdf] Minwei Feng, Jan-Thorsten Peter, Hermann Ney • "Sparse Lexicalised Features and Topic Adaptation for SMT" [pdf] Eva Hasler, Barry Haddow, Philipp Koehn • "Spoken Language Translation Using Automatically Transcribed Text in Training" [pdf] Stephan Peitz, Simon Wiesler, Markus Nußbaum-Thom, Hermann Ney • "Towards a Better Understanding of Statistical Post-Editon Usefulness" [pdf] Marion Potet, Laurent Besacier, Herve Blanchon, Marwen Azouzi • "Towards Contextual Adaptation for Any-text Translation" [pdf] Li Gong, Aurelien Max, Francois Yvon
17:30-18:00h	Closing Ceremony - Best Paper Awards — room 5620, near lift 31/32