

CURRICULUM VITAE

PERSONAL DETAIL

Full name **DO THI NGOC DIEP**

Female

Date of birth Sep 26th, 1982

Place of birth Hanoi - Vietnam

Nationality Vietnamese



University Lecturer and Researcher

Address International Research Institute MICA, Hanoi University of Science and Technology

Building B1-1004, No1 Dai Co Viet, Hanoi, Vietnam

Office: (+84) 43 868 30 87

Mobile: (+84) 912606900

Email ngoc-diep.do@mica.edu.vn

diepdn@gmail.com

Homepage: <http://www.mica.edu.vn/perso/Do-Thi-Ngoc-Diep>

EDUCATION

December 2011 Degree of Doctor of Philosophy from Grenoble University, France

- Specialty: Information
- Subject: Extraction of parallel corpus for machine translation from/to an under-resourced language
- Thesis in French

2007 – 2011 PhD student in Grenoble University

- MSTII Doctoral School (Mathématique, Sciences et Technologies de l'Information, Informatique)
- Specialty: Information
- Supervisor: Grenoble Informatics Laboratory LIG
- Co-supervisor: International Research Institute MICA

September 2007 Degree of Master of Science from Hanoi University of Science and Technology

- Specialty: Information Processing and Communication
- Subject: Modeling and use of vocalic gestures in Automatic Speech Recognition
- Thesis in English

2005 - 2007 Master Student in Hanoi University of Science and Technology

- School of Information and Communication Technology
- Information Processing and Communication Department

2000 - 2005 Student in Hanoi University of Science and Technology

- School of Information and Communication Technology

- Software Engineering Department
- Graduation thesis: “E-learning and an application of SCORM 2004 to the Learning Content Management System”
- Grade: Excellent
- Overall result of the whole course (5 years): 8.55 (on a 10.00 scale)

1997 - 2000 Baccalaureate from Viet Duc Senior High School, Hanoi

- Grade: Excellent

Interested domains:

- Natural language processing
 - Machine translation for under-resourced languages.
 - Mining and extraction bilingual data for machine translation task.

Languages: English and French

- Certification:
 - English: IELTS 5.5, TOEFL 527 (since 2002)
 - French: DELF A1, B1- en cours d’acquisition (since 2008)

CERTIFICATE AND AWARD

- | | |
|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2004 | Get the Fundamental Information Technology Engineer Skill Standards Certificate of Vietnamese Ministry of Science and Technology (equivalent to the JITEC’s Fundamental Information Technology Engineer Skill Standards) |
| 2005 | Receive the Vietnam Post and Telecoms Ministry’s certificate of merit for Distinguished Information Technology Student of the year 2003-2004 |
| 2007-2010 | Ph.D. grant from Embassy of France in Vietnam |

RESEARCH ACTIVITIES

- Participate in research projects at School of Information and Communication Technology, Hanoi University of Science and Technology (2004-2005):
 - Analyze the development of library management software for the Department of Software Technology.
 - Research on e-learning, e-learning standard, standard SCORM 2004.
 - Establish study information management system for an e-learning system.
- Participate in research projects at MICA Institute (2005-2007):
 - Develop Camera System Controller software for SIAM center room at MICA.
 - Research for the development of an e-learning application on PDA.
 - June to August 2006, research in CLIPS laboratory (INPGrenoble, France). Subject: Shot Boundary Detection for video, using Support Vector Machine.
Tutor: Dr. Georges Quénot.

- Research in the field of natural language processing (machine translation) at LIG laboratory - Grenoble (2007 - 2011)
- Research in the field of natural language processing (text and speech processing, machine translation) at Institute MICA (2007 - until now)
 - Leader of a local project at Hanoi University of Science and Technology number T2012-175 on “Research and develop a parallel text corpus for Vietnamese and other under-resourced languages” (2012).
 - Leader of project VIVAVU-phase 1 “Vietnamese Voice Assistant for Vision-impaired User”. Project of Ministry of Information and Communications of Vietnam (2013).
 - Participate to project VIVAVU-phase 2 “Vietnamese Voice Assistant for Vision-impaired User”. Project of Ministry of Information and Communications of Vietnam, leader Pr. TRAN Do Dat (2014).
 - Leader of a 2-year international project APRIL “Avancées dans le Prototypage Rapide pour l’Informatisation des Langues”. Partners: Lacito – France and MICA (2013-2014).
 - Researcher in project KC03-2nd phase, National project on “Intelligent Human-Machine Interaction System using Vietnamese voice control in the noisy environment”, leader Dr. NGUYEN Viet Son (2012-2015).
 - Leader of a local project at Hanoi University of Science and Technology number T2015-094 on “Research and develop a speech translation system on smartphone, applied to French-Vietnamese language pair” (2015).
 - Participate to a local project at MICA institute on “Building a large corpus for researching and developing of speech synthesis system and speech recognition system in Vietnamese language”, leader Dr. MAC Dang Khoa (2014-2015)
 - Participate to an international project DOREMIFA « Données des Recherches linguistiques de Michel Ferlus en Asie du sud-est », Partners: Lacito – France and MICA, leader Dr. Alexis MICHAUD (2014-2016).
 - Internship from March to April 2015, in Multilingual Translation laboratory, Universal Communication Research Institute of NICT, Japan. Subject: Collecting parallel text data for building a Vietnamese-Japanese machine translation system. Tutor: Dr. Masao Utiyama.

RESEARCH SUBJECTS FOR STUDENT

Finished subjects:

1. *Building and deploying a Web page translation system for Vietnamese-English and Vietnamese-French.* Final year student project.
 Student name: Antoine DAVID
 Institute: GRENOBLE INP ESISAR, France.
 Internship from 01/04/2012 to 31/08/2012 at MICA.
2. *Improving language model for Khmer Automatic Speech Recognition.* Final year student project.
 Student name: KAING Hour
 Institute: Institute of Technology of Cambodia.
 Main tutor: Dr. SAM Sethserey.
 Internship from 04/03/2013 to 31/05/2013 at MICA.
3. *Building a speech translation system on mobile phone.* Forth year student project.

Student names: Rodolphe FREBY, Paul LABAT
Institute: Polytech Grenoble, France.
Internship from 22/05/2014 to 13/08/2014 at MICA

Available subjects:

1. *Mining a parallel corpus using in multilingual machine translation system*. Master student project.
2. *Using synonym in machine translation system from/to Vietnamese*. Master student project.
3. *Building a subjective evaluation system for machine translation system*. Final year student project.
4. *Machine Translation for unwritten languages*. PhD student project.

PUBLICATIONS

1. Mạc Đăng Khoa, **Đỗ Thị Ngọc Diệp**, Trần Phương Thúy, *BKLAS – Hệ thống hỗ trợ việc học tập, áp dụng chuẩn SCORM 2004*, Hội nghị sinh viên nghiên cứu khoa học lần thứ 22, Khoa công nghệ thông tin, Trường Đại học Bách Khoa Hà Nội, năm 2005.
2. **Đỗ Thị Ngọc Diệp**, Mạc Đăng Khoa, Nguyễn Ngọc Bình, *Ứng dụng chuẩn SCORM 2004 trong xây dựng hệ thống e-learning*, Kỷ yếu hội thảo khoa học quốc gia lần thứ hai, Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin, TP Hồ Chí Minh, FAIR 2005.
3. Viet-Bac Le, Laurent Besacier, Sopheap Seng, Brigitte Bigi, **Thi-Ngoc-Diep Do**, *Recent Advances in Automatic Speech Recognition for Vietnamese*, 1er International Workshop on Spoken Languages Technologies for Under-resourced languages, Ha Noi, SLTU 2008.
4. **Thi-Ngoc-Diep Do**, Viet-Bac Le, Brigitte Bigi, Laurent Besacier, Eric Castelli, *Mining a comparable text corpus for a Vietnamese - French statistical machine translation system*, Workshop on Statistical Machine Translation, Athens – Greece, WMT 2009.
5. **Thi-Ngoc-Diep Do**, Viet-Bac Le, Brigitte Bigi, Laurent Besacier, Eric Castelli. *Exploitation d'un corpus bilingue comparable pour la création d'un système de traduction probabiliste Vietnamien - Français*, Conférence sur le Traitement Automatique des Langues Naturelles, Senlis – France, TALN 2009.
6. **Thi-Ngoc-Diep Do**, Laurent Besacier, Eric Castelli, *Unsupervised SMT for a low-resourced language pair*, International Workshop on Spoken Languages Technologies for Under-resourced languages, Penang – Malaysia, SLTU 2010.
7. **Thi-Ngoc-Diep Do**, Laurent Besacier, Eric Castelli, *A Fully Unsupervised Approach for Mining Parallel Data from Comparable Corpora*, Conference of the European Association for Machine Translation, Saint-Raphaël – France, EAMT 2010
8. **Thi-Ngoc-Diep Do**, Laurent Besacier, Eric Castelli, *Apprentissage non supervisé pour la traduction automatique : application à un couple de langues peu doté*, Conférence sur le Traitement Automatique des Langues Naturelles, Montréal – Canada, TALN 2010
9. **Thi-Ngoc-Diep Do**, Laurent Besacier, Eric Castelli, *Improved Vietnamese-French Parallel Corpus Mining Using English Language*, International Workshop on Spoken Language Translation, Paris – France, IWSLT 2010.
10. Laurent Besacier, Haithem Afli, **Do Thi Ngoc Diep**, Hervé Blanchon, Marion Potet, *LIG Statistical Machine Translation Systems for IWSLT 2010*, International Workshop on Spoken Language Translation, Paris – France, IWSLT 2010.
11. **Thi-Ngoc-Diep Do**, Eric Castelli, Laurent Besacier, *Mining Parallel Data from Comparable Corpora via Triangulation*, International Conference on Asian Language Processing, Penang – Malaysia, IALP 2011.
12. Emmanuel FERREIRA, Pascal NOCERA, Maria GOUDI and **Ngoc Diep DO THI**. *YAST : A scalable ASR toolkit especially designed for under-resourced languages*. IALP 2012

13. **Thi Ngoc Diep DO**, Duy Binh NGUYEN, Dang Khoa MAC, Do Dat TRAN. *Machine Translation Approach for Vietnamese Diacritic Restoration*. International Conference on Asian Language Processing (**IALP 2013**), Urumqi, China, August 17-19, 2013
14. **Do Thi-Ngoc-Diep**, Michaud Alexis, Castelli Eric. *Towards the automatic processing of yongning na (sino-tibetan): developing a 'light' acoustic model of the target language and testing 'heavyweight' models from five national languages*, International Workshop on Spoken Language Technologies For Under-Resourced Languages (SLTU'14), Russia, 14-16 May 2014.
15. **Thi-Ngoc-Diep DO**, Masao UTIYAMA, Eiichiro SUMITA, *Machine Translation from Japanese and French to Vietnamese, the Difference among Language Families*, IALP 2015.