NGUYEN TRUONG SON

(August 2018)

Email: nguyen.son@jaist.ac.jp ; ntson@fit.hcmus.edu.vn; ntson2002@gmail.com
Phone: (+81) 080 3044 6868

EDUCATION

10/2015 – 09/2018 **PhD in Computer Science**
Japan Advanced Institute of Science And Technology (JAIST) – Japan

Title of research: Structure Analysis and Textual Entailment Recognition in Legal Texts
Keywords: Recognizing textual entailment, Natural Language Inference, Legal Text Analysis, Legal Text Processing, Deep learning
Supervisor: Associate Prof. Nguyen Le Minh

10/2007 – 09/2010 **Master of Computer Science**
Faculty of Information Technology, University of Science, Vietnam National University at Ho Chi Minh City (VNU-HCM)

Thesis area: Text mining, Information Retrieval, Machine learning
Thesis title: Exploit the information on social networks for disease outbreaks monitoring.
Thesis grade: 8.6/10
Supervisor: Associate Prof. Ho Bao Quoc

09/2002-09/2006 **Bachelor of Informatics**
Faculty of Information Technology, University of Science, VNU-HCM
Thesis area: Constraint programming
Thesis title: A study on university's scheduling algorithms and build an illustrative system
Thesis grade: 9.13/10
GPA: 8.25/10
Supervisor: D.r. Nguyen Tan Tran Minh Khang

AWARDS AND CERTIFICATES

- 911-project scholarship of Vietnamese Government for Phd program (2015)
- Best Paper Award at RIVF-2015: The 11th IEEE-RIVF International Conference on Computing and Communication Technologies (2015)

- Internship Certificate at Japan Advanced Institute of Science And Technology, Japan (2015)
- Internship Certificate at National Informatics Institute, Japan (2010)
- Excellent Student Award of the University of Science, VNU-HCM (2006)
- Annual scholarships for excellent students (2002-2006)

WORKING EXPERIENCES

11/2009-present **Lecturer**
Faculty of Information Technology, University of Science, VNU-HCM

Main courses:
- Introduction to Database
- Database and Web programming
- Information System: Design and Analysis.
- Database Management Systems

01/2010 – 07/2010 **Research Internship**
National Informatics Institute, Tokyo, Japan

Research title: Web-based health surveillance and text mining

01/2007 – 11/2009 **Teaching Assistant**
Faculty of Information Technology, University of Science, VNU-HCM

RESEARCH INTEREST

- I am interested in Natural Language Processing problems/applications such as Recognizing textual entailments, Information Retrieval, Questions Answering, Information Extraction, Document Classification, Named Entity Recognition, Shallow Discourse Parsing, Chatbot.
- I am eager to apply and Machine learning, Deep learning, Data Science to solve problems in Natural Language Processing field and real-life problems.

RESEARCH EXPERIENCE

Below are main researches which I have conducted:

Structure Analysis in legal texts

- Implement some conventional models (Conditional Random Fields) and propose several deep learning-based models based on Recurrent Neural Networks to analyze structures of legal documents which are written in Vietnamese, Japanese and English.
- Develop Restful API and demonstrated web-based applications.

Legal Information Retrieval and Textual Entailment Recognition in legal texts

- Develop an information retrieval component which find relevant articles for a legal statement.
- Propose deep learning-based methods for recognizing whether or not a legal statement is entailed from legal articles.

- Develop a web-based application which can answer whether or not a legal statement is correct.
- Language: English
- This research is a part of the CREST project (https://www.jst.go.jp/kisoken/crest/en/)

Develop a Question Answering System for Company Regulatory Documents

- Develop components for a question answering system that can answer questions related to company regulations in Japanese: data processing, question analysis (using Deep learning), document retrieval (using Elasticsearch), candidate extractions and answer extractions.
- Develop a web-based application which can retrieve relevant documents and highlight candidate answers.
- Language: Japanese
- This research is a project between Nguyen'Lab and TIS company (https://www.tis.co.jp/)

Vietnamese Named Entity Recognition

- Propose deep learning-based models for Named Entity Recognitions. The proposed models allow incorporating engineering features into deep learning models.
- Models: Long Short-Term memory (LSTM), Bidirectional LSTM, BiLSTM-CRF

Shallow Discourse Parsing

- Develop methods that can detect discourse components in texts which are discourse connectives (e.g. but, however, …) and arguments.
- Methods: Conditional Random Fields
- Language: English

Web-based health surveillance and text mining

- Develop a surveillance system which can detect unusual signs of user's behavior on social networks. The system can (1) identify the behavior of user's message on social network using text classification techniques such as SVM, Naïve Bayes, Decision Tree (2) identifying the location (3) calculating the unusual level and (4) showing the result on the map (Google maps)
- Language: English

Others (which I am not a primary person)

- Extract entities and relation between proteins and entities in biomedical documents
- Building a Bilingual (English-Vietnamese) Corpus for English-Vietnamese translation task
- Develop cross language information retrieval systems

PROFESSIONAL AND RESEARCH SKILLS

Programming language: C++ (7 years), JAVA (2 years), C#(7 years), PHP (5 years) , Python (2 year)

Web technologies:

- Server-side technology: PHP, ASP.NET, JSP, node.js
- Client-side: HTML, JavaScript, AJAX, CSS, jQuery, Single page applications
- Restful API programming

Databases: MySQL, SQL Server

Research skills:

- Ability to develop deep learning-based models based on TensorFlow, Keras, Theano and other libraries which can solve various problems in Natural Language Processing such as document analysis, document classification, question answering, information extraction, …
- Ability to use and develop applications based on machine learning, text mining, information retrieval frameworks an such as Weka, GATE, Elasticsearch, Lucene, …
- Ability to learn new things (new technologies, new programming languages, new database management systems, …), independent working, problem-solving, critical thinking.

LANGUAGE SKILLS

- Vietnamese: Native
- English:
  - Good at Speaking, Writing and Listening
  - Certificates: TOEFL PBT 520, TOEIC 830

RESEARCH PUBLICATIONS

Jounrals:

- Son Nguyen Truong, Nguyen Le Minh, Ken Satoh, Tojo Satoshi and Akira Shimazu. “Recurrent neural network-based models for recognizing requisite and effectuation parts in legal texts”, Artificial Intelligence and Law, 2018.
- Collier, Nigel, Nguyen Truong Son, and Ngoc Mai Nguyen. "OMG U got flu? Analysis of shared health messages for bio-surveillance.", *Journal of Biomedical Semantics* 2011, **2**(Suppl 5):S9, doi:10.1186/2041-1480-2-S5-S9

Conference and workshops:

- Truong-Son Nguyen, Le Minh Nguyen, and Ken Satoh. “Improving entailment recognition in legal texts using corpus generation”, SCIDOCA 2017
- Truong-Son Nguyen, Le-Minh Nguyen, Akira Shimazu and Kiyoaki Shirai. "Structural Paraphrasing in Japanese Legal Texts", JURISIN 2017
- Son Nguyen Truong, Nguyen Le Minh, Ken Satoh, Tojo Satoshi and Akira Shimazu. "Single and multiple layer BI-LSTM-CRF for recognizing requisite and effectuation parts in legal texts", In Proceedings ASAIL 2017， 2017/06
- Nguyen Truong Son, Phan Viet Anh, Nguyen Le Minh. "Recognizing entailments in legal texts using sentence encoding-based and decomposable attention models", COLIEE 2017, In Proceedings COLIEE 2017， 2017

- Nguyen Truong Son, Nguyen Le Minh. "Nested named entity recognition using multilayer recurrent neural networks", PACLING 2017, August 16 - 18, 2017, Yangon, Myanmar
- Nguyen Truong Son, Phan Viet Anh, Trieu Hai Long, Nguyen Thanh Huy, Chau Ngoc Phuong, Phan Trung Tin, Nguyen Le Minh "Legal Information Extraction/Entailment Using SVM-Ranking and Tree-based Convolutional Neural Network", JURISIN 2016.
- Son, Nguyen Truong, Le Minh Nguyen, Ho Bao Quoc, and Akira Shimazu. "Recognizing logical parts in legal texts using neural architectures." In Knowledge and Systems Engineering (KSE), 2016 Eighth International Conference on, pp. 252-257. IEEE, 2016.
- Son, Nguyen Truong, and Nguyen Le Minh. "SDP-JAIST: A Shallow Discourse Parsing system@ CoNLL 2016 Shared Task." ACL 2016 (2016): 143
- Son Nguyen, Quoc Ho and Minh Nguyen. "JAIST: A two-phase machine learning approach for identifying discourse relations in newswire texts". CONLL 2015 Shared task, 2015 , Beijing, China.
- Nguyen Truong Son, Nguyen Thi Phuong Duyen, Ho Bao Quoc, Nguyen Le Minh. "Recognizing logical parts in Vietnamese Legal Texts using Conditional Random Fields". The 11th IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF2015), 2015 , CanTho, Vietnam.
- Lai Dac Viet, Nguyen Truong Son, and Minh Le Nguyen. "Deletion-based sentence compression using Bi-enc-dec LSTM", In Proceedings PACLING 2017, 2017/08
- Bui Dac Thinh, Nguyen Truong Son, and Ho Bao Quoc, "Towards a Conceptual Search for Vietnamese Legal Text",13th International Conference on Computer Information Systems and Industrial Management Applications, 2014
- Quang Le Minh, Son Nguyen Truong, and Quoc Ho Bao. 2011. A pattern approach for biomedical event annotation. In *Proceedings of the BioNLP Shared Task 2011 Workshop* (BioNLP Shared Task '11). Association for Computational Linguistics, Stroudsburg, PA, USA, 149-150.

REFERENCES

- Associate Professor Minh Le Nguyen at School of Information Science, Japan Advanced Institute of Science and Technology (JAIST), Japan.
    - Webpage: https://www.jaist.ac.jp/profiles/info_e.php?profile_id=525
    - Email: nguyenml@jaist.ac.jp
- Associate Professor Ho Bao Quoc at Department of Information Science, Faculty of Information Technology, University of Science – VNU-HCMC, Ho Chi Minh city, Vietnam.
    - Email: hbquoc@fit.hcmus.edu.vn