NGUYEN Nhu Khoa

khoann2209@gmail.com | linkedin.com/in/nguyen-nhukhoa
https://www.researchgate.net/profile/Nhu-Khoa-Nguyen

EXPERIENCE

NLP Researcher — Dec. 2019 – Dec. 2022

La Banque Postale - Asset Management — *Paris, France*

- **Text Mining and Processing to discover potential emerging trends in news articles.**
- As part of the *CIFRE PhD.* Program.
- Evaluate data structure and quality to ensure the best results from Text Mining.
- Developing method for exploring change of context in words/phrases over time to assess potential emergence of new trends.

Research Engineer — May. 2019 – Nov. 2019

Laboratoire Informatique, Image et Interaction (L3i) — *La Rochelle, France*

- Maintain DAnIEL system (Detail: https://daniel.greyc.fr/public/index.php?a=AboutDAnIEL).
- Generate and inject artificial noise into text corpus to imitate naturally degraded documents.
- experiment and evaluate DAnIEL robustness with generated noisy documents to simulate scenario that use historical newspapers.
- Part of NewsEye project (Detail: https://www.newseye.eu/).

Developer — Jan. 2017 – Dec. 2017

Institut de recherche pour le développement (IRD) Vietnam — *Hanoi, Vietnam*

- **GAMA Project.**
- Research and Develop a mean for running GAMA simulation distributively in a cluster.
- Purpose is to handle large simulations that cannot be handled by a single computer.

SKILLS

Languages: Vietnamese (Native), English (IELTS 7.5), French (DELF A2).
General Programming Languages: Python, C/C++, Java.
Machine Learning/A.I.: Classification (SVM), Clustering (k-Means, DBSCAN, HDBSCAN), Topic Modeling (LDA), Transformers language model
Data Analysis Tools: pandas, numpy, matplotlib, seaborn.
Machine Learning Tools: scikit-learn, pytorch.
NLP Tools: spacy, gensim, BERT transformer models, Word2Vec.
Parallel Programming: Basic knowledge with CUDA with Thrust API, OpenMP and OpenMPI, Parallel Programming Pattern.
Developer Tools: Git, Docker, Linux, Vim.

EDUCATION

Ph.D. in Natural Language Processing (NLP) — Dec. 2020 – Feb. 2023

La Rochelle University — *La Rochelle, France*

- Thesis: *Emerging Trend Detection in News Article*
- Supervisors: Prof. Antoine DOUCET, Thierry DELAHAUT
- Industrial Agreements for Training through Research (CIFRE) program funded by the French Ministry of Higher Education, Research and Innovations

Msc in Information and Communication Technology Sept. 2016 – Sept. 2018

Rennes University *Rennes, France*

- specialized in Infrastructure for High-Performance Computing

Msc in Information and Communication Technology Sept. 2016 – Sept. 2018

University of Science and Technology Hanoi (USTH) *Hanoi, Vietnam*

- specialized in Infrastructure for High-Performance Computing

PUBLICATIONS

FinNLP workshop - Empirical Methods in Natural Language Processing conference Dec. 2022

Contextualizing Emerging Trends in Financial News Articles

Symposium on Information and Communication Technology conference Dec. 2022

Utilizing Keywords Evolution in Context for Emerging Trend Detection in Scientific Publications

International Journal on Digital Libraries 23 Sep. 2022

Assessing the impact of OCR noise on multilingual event detection over digitised documents

WWW '21: The Web Conference 2021 Apr. 2021

L3i_LBPAM at the FinSim-2 task: Learning Financial Semantic Similarities with Siamese Transformers

19th International Conference of the Italian Association for Artificial Intelligence Nov. 2020

Impact Analysis of Document Digitization on Event Extraction