



Tanti KRISTANTI NUGRAHA

<https://tantikristanti.github.io/>

Contact:

tantikristanti@gmail.com
 <https://www.linkedin.com/in/tanti-kristanti-474520118/>

Project repositories:

<https://github.com/tantikristanti>
 <https://huggingface.co/kristanti>

Technical Skills

1. Programming languages: Python, Java
2. Machine learning and deep learning frameworks: PyTorch, TensorFlow, Keras, Hugging Face
3. Databases: PostgreSQL, MySQL
4. Vector Databases: ChromaDB, Faiss
5. Embedding models: Sentence-transformer, embeddinggemma, qwen-embedding
6. LLMs: BERT, GPT, Llama, Mistral, Deepseek, Qwen, Gemma
7. RAG and Agentic AI Platforms: LangChain, LangGraph, LlamalIndex, smolagents
8. APIs: RESTful API, FastAPI
9. Web Interfaces: Gradio, Streamlit
10. Deployment: Docker

Work Experiences

Research and Development Engineer

1. Developed a medical document extraction system.
2. Designed and implemented various systems for computational humanities projects using machine learning and deep learning approaches.
3. Developed a text mining system for extracting bibliographic and citation data from scientific documents.
4. Developed additional features for the Named Entity Recognition and Disambiguation (NERD) system.

November 2017 – April 2023

INRIA-Paris
<https://www.inria.fr/fr>

ALMAnaCH Project Team
<https://almanach.inria.fr/index-en.html>

Full-Time Lecturer

- | | |
|-------------------------------------|-------------------------------|
| - SAP Modules | - Data Structure |
| - Certified Ethical Hacking Modules | - Information Systems |
| - Software Engineering | - Knowledge Management System |
| - Project Management | - Basic Programming with C |

2008 – 2015
Faculty of Information Technology
Maranatha Christian University
<https://www.maranatha.edu/en/>
Indonesia

Education

1. Master in Language and Computer Science 2015 – 2017

Sorbonne Université, France
<https://www.sorbonne-universite.fr/>

2. Master in Informatics 2006 – 2008

Bandung Institute of Technology (ITB), Indonesia
<https://itb.ac.id/>

Languages

- French: Fluent
- English: Fluent
- Indonesian: Native speaker

Projects

1. grobid-medical-report

Development of a medical document extraction system within a collaborative project between INRIA and APHP (<https://www.aphp.fr>)

GitHub: <https://github.com/tantikristanti/grobid-medical-report>

2. DESIR-CodeSprint-TrackA-TextMining

Development of a text mining system to extract bibliographic and citation data from scientific documents for the DESIR Code Sprint within the DARIAH-EU project (<https://www.dariah.eu>).

GitHub: <https://github.com/DARIAH-ERIC/DESIR-CodeSprint-TrackA-TextMining>

Project report: <https://shs.hal.science/halshs-02415967v1/file/D4.3%20Workflow%20for%20new%20services.pdf>

3. nerdKid

Development of a Wikidata entity classification system, additional features for the Named Entity Recognition and Disambiguation (NERD) system called *entity-fishing* (<https://github.com/kemitt2/entity-fishing/tree/master>), within the H2020 HIRMEOS project.

GitHub: https://github.com/tantikristanti/NERD_KID

4. Other repositories are available at <https://github.com/tantikristanti?tab=repositories>

Shared-Task

Participated in CLEF-HIPE-2020, a shared task aimed at identifying named entities (e.g., people and places) in historical journals, <https://github.com/impresso/CLEF-HIPE-2020/>

Team: Inria-DeLFT (Team 37)

Results: - Best result for the English entity disambiguation task (#el-only-english-strict-1-literal-sense-nel-lit-micro-fuzzy-1 and #el-only-english-relaxed-1-literal-sense-nel-lit-micro-fuzzy-relaxed-1), https://github.com/impresso/CLEF-HIPE-2020/blob/master/evaluation-results/ranking_summary_final.md

Certification, Training, Workshop

[Name] – (Training Organization) (Start Date of Validity – End Date of Validity)

[GDPR Training] – (INRIA)	18/04/2023
[Scrum Way, the Way: Software Development Management with Scrum] – (ScrumWay)	03/12/2011
[SAP Module ADM100 – SAP Web AS Administration I] – (Edugate)	(12/02/2010 - 11/02/2011)
[SAP Module AC040 – Business Process in Management Accounting SAP] – (Edugate)	(08/08/2009 - 07/08/2010)
[SAP Module PLM200 – Business Process in Project Management SAP] – (Edugate)	(01/08/2009 - 31/07/2010)
[Certified Ethical Hacker] – (EC-Council)	(09/07/2009 - 08/07/2012)
[SAP Module BC400 – Introduction to the ABAP Workbench SAP] – (Edugate)	(02/02/2009 - 01/02/2010)

Research

- Development of Lecturer Portal Leveraging Text Classification Algorithm

Year: 2014 – 2015

Funded by: Maranatha Christian University.

- Classification and Clustering Analysis Model for Student and Lecturer Data in Higher Education

Year: 2013 – 2014

Funded by: Competitive grant from the Directorate General of Higher Education, Research and Technology (DIKTI), Ministry of Education and Culture, Indonesia.

Publications

Kristanti, Tanti, and Laurent Romary (2020) **DeLFT and entity-fishing: Tools for CLEF HIPE 2020 Shared Task**. CLEF 2020- Conference and Labs of the Evaluation Forum. Vol. 2696. CEUR. https://ceur-ws.org/Vol-2696/paper_266.pdf

Karén Fort, Bruno Guillaume, Nicolas Lefebvre, Laura Ramírez, Mathilde Regnault, Mary Collins, Oksana Gavrilova, Tanti Kristanti (2017) **Towards annotation through more complex corpora: the case of specialized language (Vers l'annotation par le jeu de corpus plus complexes : le cas de la langue de spécialité)**. Traitement Automatique des Langues Naturelles (TALN). <https://aclanthology.org/2017.jeptalnrecital-court.21/>

[Other publications can be found on my website, <https://tantikristanti.github.io/publications/>]

Administrative roles

- Head of Computer Laboratory**

Faculty of Information Technology

Maranatha Christian University

<https://www.maranatha.edu/en/>

Indonesia

2009 – 2015

- Head of Reviewing Board**

Jurnal Sistem Informasi (Journal of Information Systems)

2011 – 2014

ISSN 1907-1221

Maranatha Christian University

<https://www.maranatha.edu/en/>

Indonesia

Miscellaneous

Participant in Humorous Speech (in English)

15th August 2008

Maranatha Toastmasters Club

<https://www.toastmasters.org/Find-a-Club/596885-maranatha-club>

Moderator on workshops:

20th May 2010

- Thinking out of the box to protect your data

- Defend your network against hackers