

Changxu Duan

NLP Researcher

Contact

-  Darmstadt, Hessen, Germany
-  linkedin.com/in/changxu-duan/
-  <https://fireblossom.github.io/>
-  changxu.duan@foxmail.com

Education

M.Sc. Computational Linguistics
University of Stuttgart
2018 - 2021

B.Sc. Computer Sciences
Henan Normal University
2014 - 2018

Skills

- Python | Java | C | C++
- Machine Learning
- Deep Learning
- Natural Language Processing
- Algorithms
- DevOps
- Large Language Model
- PyTorch | Huggingface
- Spacy | NLTK | WordNet
- SPARQL | SQL
- ArangoDB | MySQL
- Data Structure
- Unstructured Data
- Time series Data
- Data Annotation

Profile

I am an enthusiastic researcher bridging the realms of language and AI. With a strong penchant for knowledge mining, I excel in deriving structured insights from unstructured data. My expertise extends to devops for large language models (LLM), showcasing my commitment to harnessing cutting-edge AI solutions.

Work Experience

Research Assistant

Technische Universität Darmstadt 08.2021 - Present

BMBF-funded project InsightsNet, emphasizing academic publications knowledge mining.

- Developed an annotation process pipeline for parsing these publications.
- Established cross-disciplinary connections based on the extracted content.

Machine Learning Intern

Robert Bosch GmbH 04.2020 - 10.2020

Crafted time series models with RNNs, GANs, and VAEs for bicycle sensor predictions.

- Optimized data preprocessing, model tuning, stayed abreast of ML and time series trends.
- Partnered with cross-functional teams for insights and model refinement.

Projects & Tools

Master Thesis

Master Thesis Student at Robert Bosch GmbH 10.2020 - 04.2021

Investigated event-centered emotion analysis applicability in Semi-supervised Learning.

- Designed a method using similarity to select unlabeled data, manage clustering attributes.
- Identified key attributes of SSL Emotion Analysis tasks for performance prediction.

RainbowLatex Annotator

<https://github.com/InsightsNet/texannotate>

Transforms LaTeX codes into detailed PDF layouts, specifically for academic publications.

- LaTeX Compilation: transformed LaTeX to PDF with dockerized TexLive2023.
- LaTeX Annotation: applied color-labels to LaTeX elements for layout extraction.
- Data Extraction: extracted text & figures from PDFs, identifying both types and positions.

SPARQL Autocompletion

<https://github.com/Fireblossom/sparql-auto-completion>

A language server tailored for enhancing SPARQL query functionalities.

- Syntax Highlighting: improved the syntax visualization of VSCode SPARQL extension.
- Advanced Features: auto-completion for prefixes & IntelliSense for classes and properties.