YONGSIN PARK

yongsin.nlp@gmail.com ◇ https://yongsinp.github.io

PROFESSIONAL & RESEARCH EXPERIENCE

NLP Research Lab, NHN diquest, Inc. Seoul, South Korea
Senior Researcher, Foundational Technologies Team *Mar. 2022 - Jun. 2023*
Associate Researcher, Foundational Technologies Team *Feb. 2019 - Feb. 2022*

· Worked on language analyzers for efficient, flexible, and robust text processing, as well as information retrieval and extraction systems to help users obtain information and gain insight from large volumes of text.

Intelligent Systems Lab, Dong-A University Busan, South Korea
Graduate Researcher *Sep. 2017 - Feb. 2019*
Undergraduate Researcher *Dec. 2015 - Aug. 2017*

· Worked on language analysis for a more accurate understanding of text and user intent, and machine reading comprehension for improved information retrieval.

EDUCATION

Master of Science in Computational Linguistics Seattle, USA
University of Washington *Sep. 2024 - Now*

Master of Science in Computer Engineering Busan, South Korea
Dong-A University *Sep. 2017 - Feb. 2019*

Bachelor of Engineering in Computer Engineering Busan, South Korea
Dong-A University *Mar. 2014 - Aug. 2017*

PROJECTS

Language Analysis

· **Danchoo:** A collection of deep learning (DL) models for part-of-speech (POS) tagging, named entity recognition (NER), dependency parsing (DP), and semantic role labeling (SRL), boosting text analysis accuracy
· **JIANA / Cotton:** Rule-based & statistical POS taggers, including one optimized for search engines
· **PLOT / DL-PLOT:** A rule-based NER system and a DL-based hybrid NER model for customizability and robustness
· **DISA:** A rule-based classifier for speech act (SA) and sentiment identification
· **SA Classification Using Sentence Types and Modalities:** A rule-based & DL-based SA classifier
· **Spelling & Spacing Correctors:** DL-based autocorrect models with statistical post-processing for accuracy
· **TimeWise NER:** Designed and built a dataset for extracting time entities from text

Information Retrieval & Extraction

· **Entity Linking for Expanding WiseWordNet:** Extracted named entities and their associated data from Wikipedia, to match entities identified in text with corresponding knowledge base records
· **English/Korean Machine Reading Comprehension (MRC):** DL-based MRC models for question answering
· **Summarization for Legal Consultations:** A DL and RL-based abstractive summarization for legal conversations
· **University Student Support System:** A recommendation system for course and job opportunity matching
· **Content Recommendation System:** A system to suggest trending and related content based on user behavior

Others

· **DL Server:** Built a platform for training, deploying, and managing machine learning (ML) models
· **AI Criminal Investigation Support:** Tool for extracting/masking personal data and automating sentence completion

SKILLS

Programming Languages
Python, Java, C/C++, and Bash

Comfortable with
TensorFlow, PyTorch, NLTK, spaCy, Regular Expressions, Git/GitFlow, Linux, Docker, SQL, and Databases