

kakaoenterprise

[AI기술] 자연어처리 데이터 구축 전문가 영입

Portfolio

Profile

| 남궁 영

y.namgoong@kakao.com

+82. 10. 5710. 2914

<https://github.com/aei0109>

학력

석사 컴퓨터공학, 한국해양대학교(부산)
2018. 03. ~ 2020. 02.

학사 컴퓨터정보학, 고려대학교(세종)
2010. 03. ~ 2015. 02.
이중전공: 영어영문학

경력

아르바이트 한국과학기술정보연구원(KISTI)
2020.10. ~ 2020.12.

자격

정보 정보처리기사(한국산업인력공단)
2014.05.30.

어학 KBS한국어능력시험(한국방송공사), 3+급
2019.05.19.

| 데이터 구축 관련 프로젝트 개요

Index	Year	Projects	Agency
1	2018.06.01. ~ 2019.05.31.	▪ 한국어 처리를 위한 기초 분석 도구 및 서비스 개발	미래창조과학부
2	2018.03.15. ~ 2018.11.30.	▪ 자동번역을 위한 다국어 개체명 인식의 성능 개선에 관한 연구	한국전자통신연구원
3	2019.07.31. ~ 2020.01.30.	▪ 의미역분석 말뭉치 구축	국립국어원
4	2019.08.06. ~ 2020.02.06.	▪ 구문분석 말뭉치 구축	국립국어원
5	2020.01.01. ~ 2020.02.29.	▪ 문장 문법성 판단을 위한 기초 자료 구축	국립국어원
6	2021.01.01. ~ 2021.01.31.	▪ 과학기술 기계학습 데이터 구축	한국과학기술정보 연구원
7	2020.10.22. ~ 2020.12.18.	▪ 화재 관련 이미지 데이터 구축	한국과학기술정보 연구원

프로젝트 과제 명세

1

미래창조과학부 >

한국어 처리를 위한 기초 분석 도구 및 서비스 개발

2018.06.01. ~ 2019.05.31.

수행업무

- 한국어 기저구 정의 및 표지 표준화 기술보고서 제작
- 한국어 기저구 말뭉치 구축 및 검수

데이터양

- 신문기사, 문학 등 문어체 및 구어체 대상
- 13,113문장, 17개 범주

작업도구

- 실험실 자체 언어정보 부착 시스템(LIAS)

활용범위

- 심층학습을 이용한 한국어 기저구 인식모델 학습 및 평가에 활용

(Bi-LSTM/CRF 모델을 이용한 한국어 구문
음, 한국정보과학회 학술대회논문집,
pp.631-633, 2019.)

결과물

- 한국어 기저구 표지 정의
- 한국어 기저구인식 말뭉치

2

한국전자통신연구원 >

자동번역을 위한 다국어 개체명인식의 성능개선에 관한 연구

2018.03.15. ~ 2018.11.30.

수행업무

- 한국어 개체명 말뭉치 구축

데이터양

- 신문기사, 문학 등 문어체 및 구어체 대상
- 22,135 문장, 9개 범주

작업도구

- 실험실 자체 언어정보 부착 시스템(LIAS)

활용범위

- 심층학습을 이용한 개체명 인식모델 학습 및 평가에 활용

결과물

- 한국어 개체명인식 말뭉치

3

국립국어원 >

의미역분석 말뭉치 구축

2019.07.31. ~ 2020.01.30.

수행업무

- 한국어 의미역 말뭉치 구축
- 필수역을 중심으로 의미역표지 부착

작업량

- 35,000 어절, 3,162문장
- 프레임셋: K-Propbank, ETRI frameset, U-Propbank

작업도구

- 구문&의미역 분석 말뭉치 통합도구(딥네츰릴 제공)

4

국립국어원 >

구문분석 말뭉치 구축

2019.08.06. ~ 2020.02.06.

수행업무

- 한국어 의존구문분석 말뭉치 구축
- 어절 단위로 지배소와 의존소 관계 설정 및 표지 부착

작업량

- 35,000 어절, 3,162문장
- 구문표지범주 9개, 기능표지범주 6개

작업도구

- 구문&의미역 분석 말뭉치 통합도구(딥네츄럴 제공)

5

국립국어원 >

문장 문법성 판단을 위한 기초 자료 구축

2020.01.01. ~ 2020.02.29.

수행업무

- 유사문장 자료 구축
- 도치, 어휘/조사변환, 전성, 구문분해 등의 방법으로 유사문장 생성

작업량

- 신문기사, 문학 등 문어체 및 구어체 대상
- 2,250 문장

작업도구

- 제공되는 엑셀작업 템플릿

6

한국과학기술정보연구원 >

과학기술 기계학습 데이터 구축

2021.01.01. ~ 2021.01.31.

수행업무

- 논문 본문 텍스트 데이터 구축 및 검수

작업량

- Journal 및 conference 논문 대상
- 약 650건

작업도구

- [KISTI 기계학습 데이터 구축 시스템](#)

7

한국과학기술정보연구원 >

화재 관련 이미지 데이터 구축

2020.10.22. ~ 2020.12.18.

수행업무

- 영상으로부터 화재 이미지 캡처 후 화재 여부 레이블링

작업량

- 1,040 장, 2개 범주
- YOLO v3를 활용한 화재인식모델 학습 및 평가에 활용

작업도구

- Capturing: Python3, openCV2
- Labeling: LabelMe

Publications and Skills

Publications

Thesis

Y. Namgoong, "Korean dependency parsing reflected chunking",
M.S. thesis, Korea Maritime and Ocean University, Busan, 2020. (Advisor. Jae-Hoon Kim)

| Domestic Journal

Young Namgoong, Jae-Hoon Kim* et al., **Defining Chunks and Chunking using Its Corpus and Bi-LSTM/CRFs in Korean**, *Journal of KIISE*, vol.47, no.6, pp.587-595, Jun., 2020.

| Domestic Conference

Young Namgoong, Jae-Hoon Kim* et al., **Building Korean Dependency Treebanks Reflected Chunking**, *Proc. of the 31st HCLT*, pp. 133-138, Oct., 2019.

Young Namgoong, Jae-Hoon Kim* et al., **Korean Chunking using Bi-LSTM/CRF**, *Proc. of the KCC*, pp. 631-633, Jun., 2019.

Young Namgoong, Jae-Hoon Kim* et al., **Defining Chunks for Parsing in Korean**, *Proc. of the 30th HCLT*, pp. 409-412, Oct., 2018.

Young Namgoong, Jae-Hoon Kim* et al., **Chunking auxiliary verb and particle equivalent phrases through NLTK**, *Proc. of the 42nd KOSME Spring Conference*, pp 265-266, Apr., 2018.

| Technical Document

Young Namgoong, Jae-Hoon Kim*, **Definition of Korean chunk labels**, KMOU-NLP-2018-002, Oct., 2018.

Skills

| Languages / Frameworks

Python, C++, HTML, CSS / Tensorflow, PyTorch, Flask, NLTK etc.

| Editors / VCS

Pycharm, Visual Studio Code etc. / Git

| OA

MS PPT / Word / Excel, Hancom office, Adobe Photoshop etc.

More Publications in NLP

(https://github.com/aei0109/CV_Publications)

NLP & Deep leaning projects

(<https://github.com/aei0109>)

[AI tech]

NLP Data construction expert

Name	Young Namgoong
Date	2020. 05. 01.
Research interests	NLP, Chunking, Corpus Construction, Deep Learning, BCI
Contact	y.namgoong@kakao.com
GitHub	https://github.com/aei0109