

● **학과목표**

글로벌 언어 자원과 자연언어처리에 대한 이해를 갖고 이를 바탕으로 언어처리 도구와 글로벌 빅데이터 분석 기법을 활용할 수 있는 인력을 양성하여 정부와 산업계에서 요구되는 국내외 정보 수집과 이의 분석을 통한 이슈 발굴, 동향 파악, 추세 예측 전문가 수요에 대비하고자 다음과 같은 인력을 양성한다.

교육적인 측면에서 언어 분석 시스템에 대한 이해와 운용 능력, 융합적 사고와 능력을 배양

언어인식학과 통계학 및 정보처리 공학의 융합 분야에서 혁신적인 지식을 창출하고 기술을 선도하고 적용함으로써 산업 경쟁력의 강화에 기여할 수 있는 인력 양성

연구적인 측면에서 다국어 융합 서비스를 위한 언어학, 공학, 통계학 등의 기본 지식을 활용하여 다국어 문서를 이해, 분석, 응용하는 분야에서 학제간 융합 연구를 선도적으로 활성화

세계 최고의 경쟁력을 갖춘 연구 개발을 수행함으로써 관련 분야의 새로운 경제 원동력을 창출하고 혁신적인 패러다임을 발굴하여 국가 산업 경쟁력 강화에 기여할 수 있는 연구 인력 양성

● **교수진현황**

남 지 순 언어학 박사(프랑스 파리7대학). Electronic Dictionary, Corpus Analysis, Computational Linguistics, Sentiment Analysis

송 재 목 언어학 박사(런던대학교). Language Typology, Morphology, Mongolian Language.

윤 성 우 철학 박사(프랑스 파리7대학). Philosophy, Ontology, French Philosophy.

윤 일 동 공학박사(서울대). Image Processing, Computer Vision, Medical Image Analysis

이 윤 석 공학박사(한국과학기술원). Distributed & Parallel Systems, Computer Architecture

임 승 호 공학박사(한국과학기술원). Embedded Systems, Operating Systems

조 상 영 공학박사(한국과학기술원). Computer Architecture, Embedded System

최 재 영 공학박사(한국과학기술원). Machine Learning, Pattern Recognition

EYCHENNE JULIEN 언어학 박사(프랑스 툴루즈2대학). Phonology, Phonetics, Speech Processing

● **교과목표**

학수번호	구분	교과목명	영문명	수강대상	개설 학과
G67001	필수	자연어처리	Natural Language Processing	석사	협동과정
G67002	필수	감성분석	Sentiment Analysis	석박사	협동과정
G67003	선택	딥러닝	Deep Learning	박사	협동과정
G67004	선택	추세예측론	Trend Prediction	석박사	협동과정
G67005	선택	다국어코퍼스구축론	Multi-lingual Corpus Construction	석박사	협동과정
G67006	선택	다국어코퍼스베이스	Multi-lingual Corpus Base	박사	협동과정
G67007	선택	다국어정보검색	Multi-lingual Information Retrieval	석박사	협동과정
G67008	선택	정보시각화	Information Visualization	석사	협동과정
G67009	논문연구	연구세미나1	Research Seminar	석박사	협동과정
G67010	논문연구	연구세미나2	Research Seminar	석박사	협동과정
G67011	선택	산학세미나1	Industrial Seminar	석박사	협동과정
G67012	선택	산학세미나2	Industrial Seminar	석박사	협동과정
G66004	선택	기계학습	Machine Learning	석사	컴퓨터전자시스템
G66006	선택	데이터마이닝	Data Mining	석사	컴퓨터전자시스템
G66007	선택	데이터베이스	Database	석사	컴퓨터전자시스템
G66015	선택	병렬처리	Parallel Processing	석박사	컴퓨터전자시스템
G55629	필수	분류학	Theory of Classification	석박사	통계학
G66018	선택	분산처리	Distributed Processing	석박사	컴퓨터전자시스템
	선택	수리통계학	Mathematical Statistics	석사	통계학
G55509	선택	시계열분석	Time Series Analysis	석사	통계학
G66026	선택	신경회로망	Neural Network	석사	컴퓨터전자시스템
	선택	의미론	Semantics	박사	언어인지과학
G66034	선택	인공지능	Artificial Intelligence	석사	컴퓨터전자시스템

G15659	선택	전산언어학	Computational Linguistics	석사	언어인지과학
G15666	선택	전자사전구축방법론	Machine-Readable Dictionaries	석사	언어인지과학
G66040	필수	정보검색시스템	Information Search System	석사	컴퓨터전자시스템
G15669	선택	한국어정보처리	Korean Information Processing	석박사	언어인지과학
G15101	선택	형태론	Morphology	석박사	언어인지과학
	선택	화용론	Pragmatics	박사	언어인지과학
G55502	필수	확률론	Probability Theory	석사	통계학
G55602	선택	회귀분석	RegerSSION Analysis	석박사	통계학

● 강의 교과목 내용

- **자연어처리(Natural Language Processing)** 컴퓨터를 이용하여 사람 언어의 이해, 생성 및 분석을 다루는 인공 지능 기술을 학습한다. 일상에서 사용하는 생활 언어를 형태 분석, 의미 분석, 대화 분석 등을 통하여 컴퓨터가 처리할 수 있도록 변환시키는 작업과 컴퓨터가 처리한 결과물을 활용하여 사람의 편의성에 입각하여 텍스트, 음성, 그래픽 등을 생성하는 작업에 대하여 알아보고 문서 처리, 색인 작성, 언어 번역, 질문 응답 등 많은 응용 분야의 활용을 탐색한다.
- **감성분석(Sentiment Analysis)** 일반인 작성한 다양한 웹 텍스트 문서에서 감성과 관련된 텍스트 정보를 자동으로 추출하고 이를 이용하여 특정 주제에 대한 문서의 의견을 판단하는 텍스트 마이닝(Text Mining) 기술 영역을 학습한다. 이를 이용하여 특정 분야 또는 특정 주제에 대한 웹 상의 의견 흐름 분석의 기초 데이터를 확보하는 방법을 학습한다.
- **딥러닝(Deep Learning)** 음성학과 영상 비전에서 많은 성과를 내어 주목을 받고 있는 최신 딥러닝 기법과 응용 방안에 대해 학습한다. 딥러닝의 응용으로 자연어 처리에 적용되고 있는 다양한 신경회로망의 적용 방법과 형태에 대해서 알아보고 최근의 자연어 처리에 적용된 다양한 연구의 최신 동향을 알아본다.
- **추세예측론(Trend Prediction)** 일반인 작성한 다양한 웹 텍스트 문서에서 감성과 관련된 텍스트 정보를 자동으로 추출하고 이를 이용하여 특정 주제에 대한 문서의 의견을 판단하는 텍스트 마이닝(Text Mining) 기술 영역을 학습한다. 이를 이용하여 특정 분야 또는 특정 주제에 대한 웹 상의 의견 흐름 분석의 기초 데이터를 확보하는 방법을 학습한다.
- **다국어코퍼스구축론(Multi-lingual Corpus Construction)** 단일 언어 코퍼스 구축을 확장하여 다국어 응용을 위한 다국어 코퍼스의 논리적 구축 방법 및 이론에 대해 연구한다. 유사 언어 집합 기반, 인근 지역 집합 기반, 사회적 친밀 집합 기반 등 다양한 분류 체계에 의한 코퍼스 구축에 대해 알아보고 통일 표현 체계에 기반한 구축방안을 탐색한다.
- **다국어코퍼스베이스(Multi-lingual Corpus Base)** 다국어 응용을 위한 다국어 코퍼스의 물리적 저장 방법 및 색인 검색 시스템에 대하여 연구한다. 실제 적용 가능한 다양한 병렬/분산 데이터베이스를 다국어 코퍼스베이스에 적용하는 방법을 탐색하고 적용 모델의 시뮬레이션을 통하여 최적의 방안을 찾는 방법을 알아본다.
- **다국어정보검색(Multi-lingual Information Retrieval)** 언어별 온라인 웹 상의 문서 표현 특징에 대해 알아보고 언어 특징별 국지적 웹 상의 대량 정보에 대한 조직적인 저장 및 관리와 검색 기법에 대한 고급 지식을 학습한다.
- **정보시각화(Information Visualization)** 웹 문서의 다양한 분석 결과를 정보의 통합적 관점을 원하는 사용자에게 효과적으로 제시하기 위한 정보 시각화 방법을 다룬다. 표로부터 그래프 처리 기법, 그래픽을 이용한 동적 표현 기법, 시각적 요소를 사용한 감각적 기법, 친밀한 예를 사용한 감성적 기법들에 대해 학습한다.
- **연구세미나(Research Seminar)** 다국어 언어 자산 구축, 도구, 알고리즘 및 활용에 관련된 최신 연구 동향에 대한 세미나를 진행하고 다양한 연구 테마의 아이디어를 도출하고 이의 해결 방향 및 방법을 모색한다.
- **산학세미나(Industrial Seminar)** 언어 자산을 이용한 산업계의 다양한 응용 분야에 대하여 살펴보고 실무적 접근을 위한 실제 시스템의 설계, 구현, 평가 방법에 대하여 연구한다. 이를 통해 다국어 언어 자산을 활용한 실제 산업계 활용 방안을 탐색한다.
- **기계학습(Machine Learning)** 기계학습은 다양한 분야의 응용에 적용되는 인공지능 기법으로 기존 알고리즘의 한계를 넘으며 새롭게 각광받고 있는 분야이다. linear regression, logistic regression, neural network, support vector machine, decision tree, dimension reduction, Markov chain Monte Carlo methods 등을 학습한

다.

- ◆ **데이터마이닝(Data Mining)** 대규모 데이터로부터 유용한 정보를 자동으로 발견하는 기법을 다룬다. 연관규칙탐사, 클러스터링, 분류, OLAP과 같은 데이터마이닝 기법을 강의하며, 다양한 데이터 형태에 적합한 알고리즘을 연구하고 실세계 데이터에 응용하는 능력을 습득하게 한다.
- ◆ **데이터베이스(Database)** 데이터베이스의 설계와 사용에 대한 지식을 배운다. 데이터 모델, 관계대수와 관계 해석, SQL과 같은 질의 언어, 데이터 정규화를 통한 데이터베이스 설계, 트랜잭션 처리, 물리적 데이터베이스 구조, 인덱싱, 질의 처리 방법에 대해서 강의한다. 수강생은 이러한 지식을 대용량 데이터 처리를 위한 데이터베이스 튜닝에 적용할 수 있어야 한다.
- ◆ **병렬처리(Parallel Processing)** Array processing, pipeline processing, multiprocessing 시스템 등 고속병렬 처리 컴퓨터 구성의 기본적 기술과 이론을 학습하며 GPGPU를 활용한 병렬 알고리즘과 대규모 병렬 시스템을 위한 병렬 알고리즘의 사례를 연구한다.
- ◆ **분류학(Theory of Classification)** 자료를 통한 사물의 분류는 중요한 과제이다. 선형판별분석, 로지스틱 회귀모형 뿐 아니라 나무모형, 신경망 등 패턴인식 분야에서 소개되었던 비모수적 분류방법을 학습한다. 아울러 LVQ, K-nearest neighbor 등 음성, 화상인식에서 활용되는 기계학습용 알고리즘을 살펴보고 알고리즘의 분류 정확도 측정방법을 소개한다.
- ◆ **분산처리(Distributed Processing)** 분산 환경을 제공하는 컴퓨터 시스템의 구조, 운영체제, 분산화일 시스템, 네트워크, 통신 기술 등을 학습하고 대규모 인터넷 분산 환경에서의 계산, 저장, 검색의 사례를 통하여 확장 가능한 시스템의 설계 및 구현에 대하여 연구한다.
- ◆ **수리통계학(Mathematical Statistics)** 표본분포, 순서통계량 분포, 다변량 정규분포, 추정, 가설검정 등의 고전적 추론, Bayesian, 우도비 검정에 대해서 강의한다.
- ◆ **시계열분석 (Time Series Analysis)** Box-Jenkins 방법을 중심으로 ARIMA 모형에서의 모형 식별, 추정, 검정 및 예측, 분해 시계열 분석법의 스펙트럼 분석법을 학습한다. 또한 비정상 시계열의 단위근 검정법의 기초도 다룬다.
- ◆ **신경회로망(Neural Network)** 신경세포의 기본단위인 뉴런에 대해 이해하고 뉴런에 대한 수학적 모델과 뉴런의 집합인 신경회로망을 모방한 인공신경회로망 모델을 학습하고 음성, 영상, 자연어처리 분야에 대한 응용을 통해 실제 문제에 적용하는 방법을 배운다.
- ◆ **의미론(Semantics)** 의미론 연구의 이론과 방법, 언어의 의미구조와 의미 기능을 체계적으로 연구한다.
- ◆ **인공지능(Artificial Intelligence)** 문제해결을 위한 일반적 알고리즘, 지식베이스 구성, 그리고 AI 언어에 대해 연구한다.
- ◆ **전산언어학(Computational Linguistics)** 컴퓨터를 이용하여 어떻게 자연 언어를 처리, 분석할 수 있는가에 대해 다룸과 동시에 이에 필요한 소프트웨어의 개발을 위하여 어떠한 언어학적 작업들이 필요한가에 대하여 다룬다.
- ◆ **전자사전구축방법론(Machine-Readable Dictionaries)** 사전 편찬의 방법론과 이에 따른 문제들을 살펴보고, 이렇게 완성된 사전 원고를 전산화 할 때 어떤 문제들이 있는가를 살펴보고, 이어서 이들 문제를 풀기 위한 구체적인 방법론에 대해 심도 있게 논의한다.
- ◆ **정보검색시스템(Information Search System)** 웹 상의 대량 정보에 대한 조직적인 저장 및 관리와 검색 기법에 대한 기본적인 지식을 학습한다. 정보 검색 시스템의 여러 요소 기술을 구현해보는 과제를 통하여 정보 검색 시스템을 구축하고 평가해 본다.
- ◆ **한국어정보처리(Korean Information Processing)** 한국어로 된 다양한 형태의 문서 및 텍스트 정보를 자동으로 분석하고 처리하기 위한 여러 연구 방법론과 실제 응용분야 등에 대해 연구한다.
- ◆ **형태론(Morphology)** 형태소와 단어, 품사분류의 기준과 실제 분류(명사, 대명사, 수사, 조사, 동사, 형용사, 관형사, 부사, 감탄사)와 단어의 형성(단일어, 복합어 즉 파생어, 합성어) 등을 다룬다.
- ◆ **화용론(Pragmatics)** 실제 상황에서 말하는 이와 듣는 이의 관계, 시간과 장소의 적절성, 효과적인 주제의 선택 등과 관련한 용법과 규칙을 다루며 의사소통 시의 발화에 대한 언어론이다.
- ◆ **확률론(Probability Theory)** 추상확률공간, 확률변수의 분포, 기대치, 독립성, 수렴 형태의 분류 및 관계 등

을 다룬다.

- ◆ **회귀분석 (RegerSSION Analysis)** 모수적인 방법으로 회귀모형의 추정 및 가설검정, 모형의 적합도 검사 등을 다룬다. 아울러, 이상치 및 영향력이 큰 관측치를 검출하고 잔차분석을 다루는 회귀진단에 대해서도 학습한다.