

데이터마이닝 인턴 모집 공고

반갑습니다.

이번 14/15 겨울방학동안 데이터마이닝센터에서는 두 가지 연구분야에 대하여 인턴 과정이 진행됩니다.

인턴 신청은 어떤 분들도 가능합니다. 각 인턴 신청자들은 두 가지 연구 주제 중 한가지를 선택하여 신청하여 주시기 바랍니다.

인턴 과정 신청 기간은 12월 14일 (일요일)까지 입니다. 신청해주시면 19일 (금요일) 전에 개별적으로 답장 드리겠습니다.

연구 분야에 관련된 내용은 언제든지 hyunjoong@dm.snu.ac.kr (연구 분야 1), hccho@dm.snu.ac.kr (연구 분야 2)로 문의주시기 바랍니다. 그 외의 인턴 관련 질문은 hyunjoong@dm.snu.ac.kr로 주시기 바랍니다.

인턴 신청 역시 각 연구 분야별로 hyunjoong@dm.snu.ac.kr 혹은 hccho@dm.snu.ac.kr로 해주시기 바랍니다.

감사합니다.

연구 분야 1. 한국어 자연어처리 및 응용

간략한 연구 내용은 다음과 같습니다.

한국어 자연어처리는 많은 텍스트 마이닝 관련 어플리케이션의 핵심 과정 중 하나입니다. 그 중 part of speech tagging은 주어진 문장으로부터 '명사, 동사, 조사, ..'와 같은 품사를 판별하는 작업으로, 키워드 추출, 의미 분석 등의 다른 여러 자연어처리 작업의 기본 과정입니다.

본 인턴 과정에서는 기존의 한국어 자연어처리 엔진이 뉴스문서에 적용하기에 적절한지 실험적으로 알아보고, 이를 이용하여 간단한 웹 어플리케이션을 만드는 작업을 수행합니다. 만약 기존의 한국어 자연어처리 엔진이 웹문서에는 부적합할 경우, 이를 보완한 연구를 수행할 수 있습니다.

본 인턴 과정에서는 한국어 자연어처리를 바탕으로 스트리밍으로 제공되는 뉴스 문서에서 현재의 트렌드를 쉽게 파악/추적할 수 있는 연관어 분석 툴을 만들어봅니다. 연관어 분석은 한 단어와 다른 단어간의 연관성을 자동으로 추출하여 관련 개념들을 쉽게 파악할 수 있도록 도와줍니다.

웹 어플리케이션은 인턴 과정 중 모두의 합의 하에 변동될 수 있습니다.
또한 인턴 과정 중 개인이 새로운 어플리케이션을 제안/진행하여도 됩니다.

본 인턴 과정에서 반드시 배울 지식은 다음과 같습니다.

1. 문서 수집 방법
2. 기존의 한국어 자연어처리 엔진의 사용 및 성능 평가

본 인턴 과정에서 어쩌면 배울 지식은 다음과 같습니다.

1. 연관어 분석을 위한 데이터마이닝 알고리즘
2. 웹문서에 적합한 한국어 자연어처리 엔진 연구 및 개발

본 인턴 과정을 위한 요구사항은 다음과 같습니다.

1. 기본적인 프로그래밍 경험
2. 토론을 즐기는 적극적 자세

인턴 과정 기간은 **12월 29일 ~ 2월 14일 (7주)**입니다.

또한 웹 어플리케이션은 진행 사항에 따라 바뀔 수 있습니다. 이번 인턴 분야 1은 {김미숙, 김한결, 김현중, 박은정} 연구원과 함께 진행됩니다.

연구 분야 2. Deep learning 패키지를 이용한 한글 인식 프로그램 개발

간략한 연구 내용은 다음과 같습니다.

현재 이미지에서 한글 캐릭터를 인식하는 연구는 많이 진행되었으나 Deep Learning 을 사용한 방법은 아직 없거나 연구중에 있습니다. 본 과정에서는 Deep Learning 을 이용하여 이미지에서 한글을 인식하는 방법을 연구하고, 나아가 실제로 추출하는 프로그램을 개발하는 것을 목표로 하고 있습니다.

본 과정을 두 부분으로 나누면, 처음 3~4 주는 이미지에서 한글을 인식하는 연구 사례와 Deep Learning 을 조사하여 공유하고, 모델 학습을 위한 데이터를 구성합니다. 그 후 3 주간은 실제로 Deep Learning 을 이용한 한글 인식 프로그램을 제작하는 것을 목표로 합니다.

본 인턴 과정에서 배울 수 있는 지식은 다음과 같습니다.

1. 기초적인 이미지 프로세싱
2. 기존의 글자 인식 알고리즘
3. Deep Learning 에 관한 기초적인 지식
4. Deep Learning 관련 프로그래밍 패키지 사용방법

본 인턴 과정을 위한 요구사항은 다음과 같습니다.

1. 기본적인 Java 프로그래밍 능력. 혹은 본인이 익숙한 다른 주력 프로그래밍 언어 지식
2. 기본적인 선형대수 지식

인턴 과정 기간은 **12월 29일 ~ 2월 14일 (7주)**입니다.

이번 인턴 분야 2은 {조현창, 안진원} 연구원과 함께 진행됩니다.