

## 연수 제안서

연구 분야	머신러닝 기반, 소재 빅데이터 구축 및 활용
연구 과제명	
연수 제안 업무	머신러닝 기반, 소재 빅데이터 구축 및 활용
<p>(연수 내용)</p> <p>연구의 배경:</p> <p>소재 분야에서 방대한 양의 실험 및 계산 데이터가 문헌을 통해서 매일 보고되고 있다. 연구자들은 매일 실험 활동을 통해 데이터를 꾸준히 생산한다. 그러나, 문헌들을 보고하는 방식이나 연구노트 작성방식은 모두 제각각이기 때문에 일관된 형태의 데이터베이스 구축 및 활용이 매우 어렵다. 본 연수과정에서는 우선적으로 촉매 소재분야에 집중한다. 촉매의 합성 조건, 구조분석, 활성 성능정보 등 관련된 핵심 정보들 (text, image, graph등의 다양한 포맷)을 머신러닝을 활용하여 추출하고, 궁극적으로 소재 빅데이터를 구축하고 활용하고자 한다.</p> <p>연수 내용:</p> <p>연수의 내용은 크게 2가지로 나뉜다.</p> <ul style="list-style-type: none"><li>- 첫째, 문헌으로부터 데이터 추출이다. 본 연구팀은 이미 촉매분야 논문을 xml 데이터포맷으로 약 100,000편을 수집한 상태이다. 촉매의 합성조건, 구조분석, 활성 성능정보 등 핵심 정보들 (text, graph, image 포맷)을 머신러닝 기법으로 추출/활용하고자 한다.</li><li>- 둘째, KiRI Note(KIST 자체개발한 I-Pad 기반 전자연구노트)로부터 데이터 추출이다. KiRI Note는 기존의 제지 기반의 연구노트 개념을 극복하고, 통상적 연구활동을 아이패드 기반으로 진행한다. KiRI Note에는 raw data들이 비정형화된 형태로 저장되어 있는데, 이로부터 중요 정보들을 추출하여, 궁극적 형태의 소재 데이터베이스 구축 및 활용의 기틀을 마련하고자 한다.</li></ul> <p>- 연수기간 : 2021년 11월 1일 - 2022년 10월 31일 (추후 평가를 거쳐, 연장 가능)</p> <p>※ 연구 정보의 기밀 유지</p>	
소속 부 서 : 계산과학연구센터 연수 책임자 : 김동훈	