

## 산학캡스톤 프로젝트 제안서

신청 기업명	(주) 와이즈넛			사업장 위치	경기도 성남시 분당구	
담당자	소 속	성장기술연구소	성 명	양재석	직 위	차장
	연락처	010-8933-6924	e-mail	jsyang@wisenut.co.kr		
관련분야	<div> <div> <div>■ 빅데이터 처리</div> <div>■ 인공지능</div> <div>□ 임베디드SW</div> <div>□ 기타 (</div> </div> <div> <div>□ 애플리케이션SW</div> <div>□ 컴퓨터 보안</div> <div>□ 멀티미디어콘텐츠(AR/VR, 게임 등)</div> <div>)</div> </div> </div>					
프로젝트 명	다차원 벡터 활용을 통한 유사 문서 검색 시스템 개발					
프로젝트 개요	<p>문서 유사도 검색(Text similarity search)는 사용자가 입력한 임의의 텍스트에 대해서 이 텍스트와 가장 유사한 순서대로 문서들을 추출해서 정렬하는 검색 형태를 지칭한다.</p> <p>일반적으로 널리 사용되는 키워드 검색은 Sparse Embedding 방식으로 문서를 찾아내는 것에 비해, 본 프로젝트에서는 Dense Embedding 방식의 알고리즘을 활용하여 유사 문서를 검색하고자 한다.</p> <p>Sparse Embedding 방식의 검색에서는 키워드가 정확하게 매칭되는 문서를 추출해오는 특성으로 인해 의미론적인 유사성을 반영되지 않기 때문에, 정확한 쿼리를 사용해야만 원하는 문서를 가져올 수 있다.</p> <p>반면 Dense Embedding 방식의 유사도 검색 방식에서는 정확히 매칭되는 키워드가 없어도 원하는 문서를 가져올 수 있는 장점이 있다.</p> <p>본 프로젝트에서 상기한 Dense Embedding 머신 알고리즘을 기반으로 하는 최신 검색 시스템을 개발하는 것을 목표로 하고 있다.</p>					
추진 배경	□ 배경/ 필요성 - 산업현장 문제점 기술					
	<p>최근 ChatGPT가 화두가 되며 생성형 AI가 많은 관심을 이끌어 냈다. 이와 함께 검색엔진 영역에서는 질문에 대한 올바른 검색 결과를 가져오는 방법과 관련된 주제가 큰 기술 트렌드로 자리매김하게 되었는데, 본 프로젝트를 통해 학생들은 최신 검색 기술과 다양한 AI 모델링 기법을 활용/접목하여 유사문</p>					

	서 검출 시스템, 표절 문서 검출 시스템, Q&A 시스템 등이 어떻게 만들어지는지 확인해보고 이를 적용함으로써 AI 관련 지식과 시야를 전반적으로 넓혀 줄 수 있을 것으로 예상된다.
<b>개발 목표</b>	<input type="checkbox"/> 개발 목표 – 개발하고자 하는 시스템의 목표
	<ul style="list-style-type: none"> <li>- AI 기반 빅데이터 학습/분석 모듈</li> <li>- 해당 모듈을 실제 검색엔진에서 사용할 수 있도록 플러그인화</li> </ul>
	<input type="checkbox"/> 기업체 지원 가능 사항 (산업체 멘토, 개발 도구, 개발용 샘플 등)
	<ul style="list-style-type: none"> <li>- 멘토</li> <li>- 개발/시연을 위한 샘플 데이터</li> <li>- 개발/테스트 서버</li> <li>- 소스 형상 관리 서버</li> </ul>
<b>결과물</b>	<input type="checkbox"/> 최종 기대 결과물 (졸업작품으로 전시할 소프트웨어 시현물)
	<ul style="list-style-type: none"> <li>- 웹 페이지를 통한 데이터 분석 방법 시연</li> <li>- 웹 페이지를 통한 대화형 검색엔진 결과물 시연</li> </ul>
<b>인턴십 연계</b>	<input checked="" type="checkbox"/> 여름방학 인턴십 <input type="checkbox"/> 겨울방학 인턴십 <input type="checkbox"/> 인턴십 안함

※ 학생들이 본 제안서를 읽고 본인에게 적합한 과제를 선택할 수 있도록 작성 요망

※ 추진배경 및 개발목표 추가 기술이 필요한 경우 별지 작성 및 첨부 요망