

## DU-도전학기제 참가신청서

성명				학번		
단과대학				학과(전공)		
휴대전화 번호				이메일		
보호자 성명				보호자 연락처		
도전학기 지도교수						
도전학기 과제명	(한글) 글에서 중요 단어를 추출하여 해당하는 이미지 연결하기 (영문) Extract important words from the article and connect with similar image					
도전 기간	2018-1학기	도전 영역	<input type="checkbox"/> 일반선택영역 <input checked="" type="checkbox"/> 전공영역			
도전학기 과제 내용 요약	해당하는 글의 주요 단어를 분석해 아주 밀접한 이미지 또는 그림을 연결해주어 상상하기 힘들거나 윤곽조차 잡히지 않는 내용들에 대해 이해를 돕는다.					
대학 재학 중 주요 교내외 활동	기관명	활동기간	활동내용			
2017-1 교내외 장학금 수혜내역	구분(교내/교외)	장학금명		장학금액(원)		
상기와 같이 도전학기제에 지원합니다. 2018년 2월 5일 신청인 :						

## DU-도전학기제 도전계획서

성명		학번	
단과대학		학과(전공)	
도전학기 과제명	(한글) 글에서 중요 단어를 추출하여 해당하는 이미지 연결하기 (영문) Extract important words from the article and connect with similar image		
신청학점 및 교과구분	전 공: 3 학점	예상 소요 예산	책값 30만원
	일반선택: 학점	소속 학과장	
학과장 의견	학생의 도전과제는 자연어처리나 기계학습등 이용해 주어진 글에서 중요 단어를 추출, 관련 이미지를 보여주는 과제인 과제의 내용도 진용 교과목의 내용과 밀접한 관련이 있다고 사료됨		

### 1. 도전의 배경

교과 외 활동을 하면서 원하는 전공 공부를 할 좋은 기회.

### 2. 도전 및 도전 과제의 목표

소설, 문학과 같은 글에서 그림 또는 이미지는 극히 일부에 불과하거나 아예 없는 경우가 많다. 이는 독자의 상상력을 증진시키며 오히려 긍정적인 효과를 낼 수도 있지만 자신만의 상상이기에 닫힌 그림밖에 떠오르지 않는다. 또한 글에 관련된 지식이 미리 없다면 상상조차도 할 수 없는 애매한 표현들이 많기에 이에 글을 알맞은 단위로 끊어 프로그램으로 주요 단어를 조합해 이와 관련된 이미지를 보여주면서 상상을 도와준다.

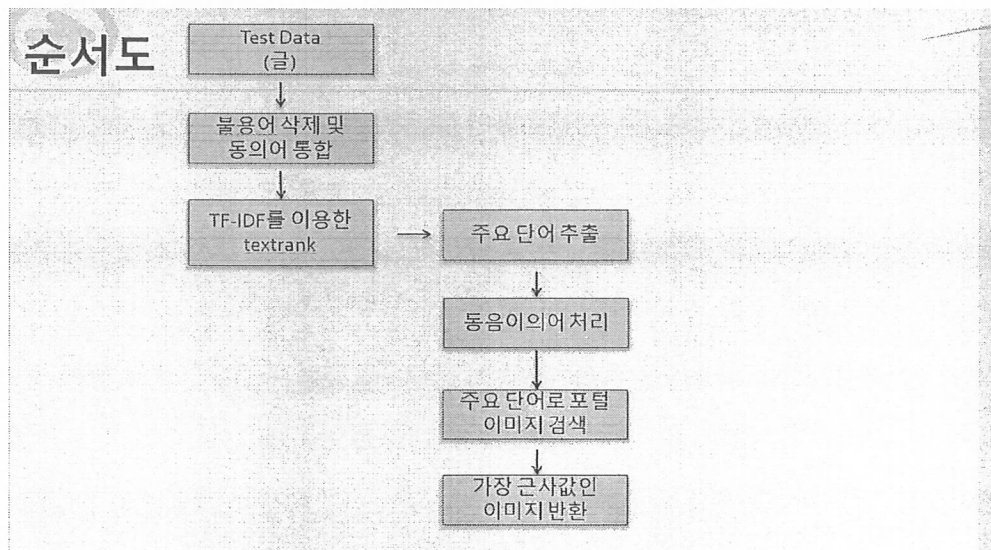
그리고 자연어처리 및 기계학습의 이해도를 높이고 이와 관련된 제대로 된 프로그램을 만든다.

### 3. 도전 과제 내용

참고문헌 : CNN을 이용한 소셜 이미지 자동 태깅(저자 : 장현웅, 조수선 정보 과학회논문지 2016.01)

: LSI를 이용한 차원 축소 클러스터 기반 키워드 연관망 자동 구축 기법(저자 : 유한목, 김한준, 장재영 정보과학회논문지 2017.11)

글에서 중요 단어들을 추출하여 이 단어들에 해당하는 연관된 이미지를 가져오는 것이 목적이며



우선 간단하게 프로그래밍 하는 틀을 잡으면

1. 글이 주어지면 이 글의 불용어 삭제 및 동의어 들을 하나로 통합한다.
2. TF-IDF 기반 TextRank 알고리즘을 이용해 단어의 TextScore를 내림차순으로 정리한다.
3. 정리한 것 중 상위 주요 단어 몇 개를 색인어(주요 Topic)으로 추출한다.
4. 추출된 주요단어가 동음이의어 일 경우 어떤 의미인지 결정한다.
5. 이전 단계의 색인어를 이용해 포털(구글 or 인스타그램)에서 검색 후 이 검색된 상위 이미지들을 가져온다.

위의 기초 프로그램이 완성되었을 시 성능들을 높이기 위한 방법들을 구상해야 한다.

예를 들면 주요 단어를 추출하기 위한 textrank의 방식이나 주요단어를 이용해서 이미지를 가져올 시 하나의 단어만으로 가져올지 여러 단어를 복합적으로 포함한 이미지를 가져올지 아니면 몇 개의 단어에 또 다른 색인 방법을 추가할지 이 과정이 주요 도전이 될 것이다.

테스트 데이터를 위한 글들을 크롤링 하며

불용어 삭제 및 동의어통합을 위한 한국어 사전 라이브러리를 가져오고 프로그램 자체적으로 인스타그램 또는 구글 이미지를 크롤링 해올 수 있어야 한다. 오류는 수도 없이 많을 것이고 성능이 안나오거나 알고리즘 자체가 잘못됐으면 수정해야 할 것이다.

#### 4. 도전 과제 추진일정

3월 2일 도전학기제 시작.

3월 2일~(1주차)

앞으로의 학업에 필요한 지식들에 관련된 책 목록을 살펴보고 구매한다.

3월 5일~(2주차)

textrank를 매기는 방식에 대해 고민하고 이와 관련된 논문 또는 기술을 찾아보며 직접 실습하며 실험한다.

3월 12일~(3주차)

textrank 로 매겨진 단어들을 어떤 규칙(위에서 3개) 또는 수식을 통해 가져올지 고찰하며 실습한다.

3월 19일~(4주차)

위의 3주차의 새로운 규칙 또는 수식을 통해 추출된 단어로 이미지를 어디서 가져올 것인가에 대해(구글, 인스타그램) 고찰하며 실습한다.

3월 26일~(5주차)

어디서 가져올까에 대해 결론이 났다면 어떻게 그 이미지를 가져올지에 대해(단순하게 단어로 검색해서 상위 이미지만 크롤링?) 집중 고찰한다.

4월 2일~(6주차)

기본 데이터인 글에서 불용어들을 어떻게 처리할 것인가에 대해 고찰한다.

4월 9일~(7주차)

가져온 이미지가 믿을만한 이미지인가 이를 어떻게 검증할 것인가에 대해(문맥상 동물 사자인데 가져온 이미지는 사자탈, 또는 검색 결과에 맞지 않는 잘못된 사진) 집중적으로 고찰한다.

4월 16일~(8주차)

4월 중순으로써 중간보고서 기간이다. 이에 따라 현재까지의 학습 내용 및 결론에 대해 고민하고 지도교수의 면담, 지도를 통해 도전 프로젝트의 수행 점검 및 이후 방향 설정한다.

4월 23일~(9주차)

가져온 이미지 하나만을 연결할 것인가 여러 이미지를 조합할 것인가에 대해서(단순히 어딘가에서 가져온 이미지 그 하나만을 띄울지 아니면 가공 또는 편집하는 어떤 과정을 거칠지) 집중적으로 고찰한다.

4월 30일~(10주차)

글에서 문단 단위로 연관된 이미지를 연결할 것인가 소제목 하나 단위로 할 것인가 사용자가 원하는 단위를 설정 할 수 있게 해야 하는 것인가? 각 설정에 따라 프로그램 성능에 영향을 끼치거나 처리 곤란한 점이 있는가?에 대해 고찰한다.

5월 7일~(11주차)

추출된 단어가 동음이의어라면(사람“눈”, 하늘에서 내리는 하얀“눈” 어떻게 구분하여 색인어으로써 사용할 것인가에 대해 고찰한다.

5월 14일~(12주차)

이 프로그램에 들어갈 글은 어떤 종류인가(뉴스, 소설, 일기등) 또한 그 종류에 따라 성능에 영향을 끼치는가에 대해 고찰한다.

5월 21일~(13주차)

이미 완성된 이미지를 가져오는 것이 아니라 어설프터라도 색만 새로 조합하거나 새롭게 이미지를 창조할 수 있는가에 대해 고민한다.

5월 28일~(14주차)

이 프로그램을 어디서 시연 할 것인가.(웹? 핸드폰?) 이를 구현할 수 있는가? 할 수 있다면 구현한다.

6월 4일~(15주차)

이 프로그램에서 쓰일 규칙, 별도 데이터 등 데이터베이스를 구현한다.

6월 11일~(16주차)

현재까지의 과정 중에서 부족했거나 하던 와중 급격하게 변화해야 할 내용이 있는지에 대해 고찰하고 구현한다.

6월 18일~(17주차)

6월 중순 결과 보고서 기간이다. 지도교수의 면담, 지도를 통한 도전 프로젝트 수행에 대한 최종 보고를 한다.

6월 21일 도전학기제 종료

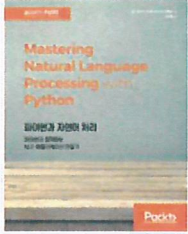
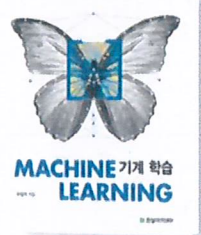

일정 중에 결과가 나온다 싶으면 논문을 제출하여 이에 대한 결과를 얻는다.

## 5. 예상 소요 예산 상세 내역

책값 30만원

## 6. 도전과 관련하여 제출할 수 있는 결과물

1. 각 주차마다 해당하는 목표를 완성했다면 그 프로그램이 돌아가는 모습 자체를 컴퓨터 내부 동영상으로 짧게 찍어 제출 가능.
2. 성능을 개선한다면 이전 데이터와 대조하며 개선 수치 및 방법론 제출
3. 논문 제출
4. 최종 결과물 프로그램

	책 이름	구매 목적	가격
	파이썬과 자연어 처리	현재 목표하는 프로그램의 주요 기능으로써 문자열을 자신이 원하는 방식으로 데이터화하고 불필요한 데이터를 제거하는 등 자연어 처리에 관련하여 빠른 이해를 돕는다.	25,200 원
	기계 학습	비전공자도 볼 수 있을 정도의 기초적인 기계학습에 대한 지식을 확립한다.	32,000 원
	파이썬을 이용한 머신러닝, 딥러닝 실전 개발 입문	실질적인 파이썬 예제 코드로써 다양한 과정 및 결과물을 직접 도출해낸다.	27,000 원
필요시 추후 구매 예정		프로그램의 완성도를 높이기 위한 구매	



## 개인정보 수집 및 활용 동의서

소속 :

학번 :

성명 :

위 본인은 대구대학교 DU-도전학기 참여와 관련한 개인정보를 국고사업 및 각종 평가 실적, 학교 홍보 등의 자료로 활용하는데 동의합니다.

개인정보 수집 및 이용 항목	
성명, 소속, 학번, 연락처, e-mail, 도전과제 및 결과물	<input checked="" type="checkbox"/> 동의함 <input type="checkbox"/> 동의하지 않음

\*위 목적이외의 다른 용도로는 활용을 하지 않습니다.

2018년 2월 5일

성명



## DU-도전학기제 서약서

소속 :

학번 :

성명 :

위 본인은 DU-도전학기 과제 수행과 관련하여 아래의 내용에 대하여 서약합니다.

1. 2018년 3월부터 7월까지 DU 도전학기 활동을 성실히 수행할 것을 약속하며, 과제 수행 중 휴학 및 자퇴를 할 경우 지원금액 전액을 반환할 것이다.
2. 교내 프로그램 및 타 국고사업과 동일 또는 유사한 과제로 중복지원하지 않을 것을 약속하며, 이를 위반할 경우 DU-도전학기 이수학점 취소 및 지원금액 전액을 반환할 것이다.

2018년 2월 5일

서약자