



교환교 DB 요약 프로젝트

김나현, 김소민, 황채원

목차

01 주제소개

02 어려웠던 부분

03 주제 구체화

04 진행상황



주제 소개

교환학생 파견보고서 요약하기

- 기존 파견교 리스트(엑셀형식)에서는 교환교에 대한 정보를 얻기 어렵다
- 파견보고서를 여러 개 살펴보기엔 각 파견보고서의 분량이 너무 길다
- **국제교류처의 파견보고서를 요약하여 교환교에 대한 핵심정보를 간단하게 제공하자**

파견국가	학교정보		학교정보										파견가능인원	파견가능인원	파견가능인원	Beginning Date of the Semester (for Winter/Spring)	End Date of the Semester (for Winter/Spring)	Qualification for incoming exchange students	What is the primary language of instruction?	If the primary language of instruction is not English	Is there any restriction for exchange students?
	Name of University	Country	State (City)	Website for exchange/study abroad	Academic calendar	Please indicate the support the university provides	Name	Title	Email address for student	Mailing address (Courier)	Postal Code	Phone Number	수학가능여부	1학기 파견	1년 파견						
네덜란드	Tilburg University	Tilburg Univ	The Netherlands	Tilburg	https://www.tilburguniversity.edu/en/exchange	https://www.tilburguniversity.edu/en/exchange	Rachael Vicker	Study Abroad	Studyabroad@tilburguniversity.edu	Tilburg University	5037 AE	+31 13 466 4471	영어권	1	0	January 2023	July 2023, exact date to be confirmed	Both undergraduate and postgraduate	English	Yes with NO restrictions	Some restrictions
미국	Smith College	United States	Massachusetts	https://www.smith.edu/about/exchange	https://www.smith.edu/about/exchange	The Office of International Disability	Angelo Pisano	Angelo-Study Abroad	studyabroad@smith.edu	Lewis Global Studies	1063	413-585-7598	영어권	1	0	January 26, 2023 - First	45058	Undergraduate only	English	Yes with NO restrictions	None
미국	Northwest University	USA	Washington (Kirkland)	https://eagle.northwestu.edu/exchange	http://catalog.northwestu.edu/exchange	No	Mr. Davis Gibson	International Program	davis.gibson@northwestu.edu	5520 108th Ave NE	98033	425-889-5363	영어권	2	0	January 9th, 2023	May 4th, 2023	Undergraduate Only	English	N/A	Students may not exchange
스웨덴	Karlstad University	Sweden	Karlstad	www.kau.se/international	https://www.kau.se/international	https://www.kau.se/international	Nina Kusche &	International Co	exchange@kau.se	Karlstad University	65188	+46 54 700 1000	영어권	1	0	January 16th, 2023	June 4th, 2023	Both undergraduate and postgraduate	English	Yes with NO restrictions	Some restrictions
독일	Osnabrück University	Germany	Osnabrück	https://www.uni-osnabrueck.de/en/exchange	https://www.uni-osnabrueck.de/en/exchange	Support by our	Sabine Jones	Mrs.	incoming@uni-osnabrueck.de	Universität Osnabrück	49074	+49 541 969 4500	독어권	3	0	winter October	official end winter	Both undergraduate and postgraduate	German or English	Yes with NO restrictions	Knowledge of German required
스웨덴	Lund University	Sweden	Lund	www.lunduniversity.se/en/exchange	www.lunduniversity.se/en/exchange	www.lunduniversity.se/en/exchange	Benjamin Nydén	Coordinator	benjamin.nyden@stud.lund.se	it is a post box, not a physical address	21100	462220384	스웨덴	2	0	see calendar	see calendar	Both undergraduate and postgraduate	Swedish	Yes with SOME restrictions	RESTRICTIONS
스페인	Abat Oliba CEU University	Spain	Barcelona	https://www.ceu.es/en/exchange	https://www.ceu.es/en/exchange	Please see our website	Fanny Abela	International Co	internacional@ceu.es	C/ Bellesguard, 110	8022	+34 93 254 09 00	스페인	1	0	31-01-2023	19-05-2023	Undergraduate	Spanish	Yes with NO restrictions	Can I contact you?
독일	Goethe University Frankfurt	Germany	Frankfurt am Main	https://www.goethe-universitaet-frankfurt.de/en/exchange	https://www.goethe-universitaet-frankfurt.de/en/exchange	https://www.goethe-universitaet-frankfurt.de/en/exchange	Julia Freier and	Student Mobility	incoming@uni-goethe-frankfurt.de	Global Office, C/	60323	+49 69 798 -1500	독어권	5	0	Winter term starts	Winter term ends	Both undergraduate and postgraduate	German	Yes with SOME restrictions	Course selection
스페인	Universidad Carlos III de Madrid	Spain	Madrid	https://www.uc3m.es/en/exchange	https://www.uc3m.es/en/exchange	https://www.uc3m.es/en/exchange	Ricardo Checa	Incoming Student	incoming@uc3m.es	C/ Madrid 126,	28903	34916249550	스페인	3	0	Jan 25 2023	June 30 2023	Undergraduate	Spanish, English	Yes with NO restrictions	https://www.uc3m.es/en/exchange
독일	Hochschule Düsseldorf - University of Applied Sciences	Germany	Düsseldorf (F)	https://www.hs-duesseldorf.de/en/exchange	https://www.hs-duesseldorf.de/en/exchange	HSD offers various	Sophie Haman	Student Exchange	incoming@hs-duesseldorf.de	Münsterstraße 40	40476 D	49 211 4351 815	독어권	1	0	Summer Semester	Summer Semester	Both undergraduate and postgraduate	German. The English language is also possible	Yes with NO restrictions	/
스페인	Universidad Católica San Antonio	Spain	Murcia	www.international.uc3a.es	https://www.international.uc3a.es	https://www.international.uc3a.es	Lonneke Meijer	Exchange coordinator	incoming@uc3a.es	Campus de los	30107	34968278601	스페인	1	0	1-2-2023	10-6-2023	Undergraduate	Spanish	Yes with SOME restrictions	Health insurance
독일	Europa-Universität Viadrina	Deutschland	Flensburg	https://www.uni-europa.de/en/exchange	https://www.uni-europa.de/en/exchange	Infrastructure	Lucila Morales	International Co	incoming@uni-europa.de	Auf dem Campus	24943	33059302567	독일, 영어권	5	0	Winter/Fall 1st	Winter/Fall 28th	Both undergraduate and postgraduate	German	Yes with SOME restrictions	The student must have a valid passport
프랑스	Université Catholique de Lille	France	Lille	https://www.univ-lille.fr/en/exchange	https://www.univ-lille.fr/en/exchange	Students with	Charlotte Prevost	Incoming Student	charlotte.prevost@univ-lille.fr	charlotte.prevost@univ-lille.fr	60 Boul	33059302567	불어권	6	0	Early January	Late May	Both undergraduate and postgraduate	French and English	Yes with SOME restrictions	The following information is required

어려웠던 부분

보고서마다 다른 형식, 목차

-> 소주제 별로 통합하여 요약하는 것이 매우 까다로움

데이터 문제

- > 모델을 훈련시키기에 데이터가 충분한가?
- > 훈련을 위한 데이터를 직접 라벨링을 할 것인지
- > 라벨링을 한다면 어떠한 기준으로 요약문을 생성할 것인지

모델선정

-> 카카오브레인 KoBART의 추출요약기능을 사용하기로 결정

목차	
1. 서문	1. Alma College 소개
2. 출국준비	2. 날씨
a. 거주허가증 신청	3. 준비과정
b. 기숙사 신청	1)비자
c. 비행기표 사기	2)비행기표
d. 짐싸기	3)유심구매
3. 도착 후 할 일	4)예방접종
4. 학교 생활	5)보험
a. 수강신청 및 학교 수강 시스템	6)핸드폰 정지
b. 수강한 강의 소개 및 평가	7)서류준비
5. 버디 위크 (Buddy Week)	8)코로나검사
6. 오슬로 생활	9)은행
a. 마트	10)짐싸기
b. 운동	4. 학교 안(교내 시설)
c. 여행	5. 학교 밖(식당, 편의시설)
	6. 교내활동
	7. 수강 과목
	8. 교통 및 여행
	9. 인증차별
	10. 귀국 후 소감

주제 구체화

크롤링

- 소주제마다 따로 크롤링을 하는 것은 현실적으로 어렵다
- 보고서 전체를 크롤링하는 방식 선택

문서요약

- 학교 별로 요약(multi document) VS 보고서 문서 하나에 대한 요약(single document)
 - > 보고서마다 내용과 형식이 제각각이므로 single document에 대한 요약을 선택
- Abstractive VS Extractive
 - > 필요한 정보를 선택적으로 제공하는 것이 주된 목적이므로, **추출요약**을 선택

etc

- 토픽모델링을 이용하여 파견보고서에 대한 키워드 뽑기 → 태그 달기

진행 상황

크롤링

- Selenium과 BeautifulSoup을 이용한 크롤링
- 페이지 로그인 + 임의의 게시글로 들어가기

```
[57] # 크롬드라이버 버전에 맞게 설치하시고 크롬드라이버가 위치한 주소 넣어주시면 돼요
s = Service('C:/Users/user/Desktop/EWHA/2022 summer/EURON/chromedriver_win32/chromedriver.exe')
driver = webdriver.Chrome(service = s)
driver.get("https://sso.ewha.ac.kr/SSO_IDP/swift/sso/loginForm.jsp?RSP=rwcms.ewha.ac.kr&RelayState=%2Fssologin")
Python
```

```
[58] # 로그인하기
# send_keys() 안에 아이디 비밀번호 작성해주시면 돼요
driver.find_element(By.NAME, 'login_id').send_keys('id') #아이디
driver.find_element(By.NAME, 'usr_pwd').send_keys('pw') #비밀번호
Python
```

```
[59] # 로그인 버튼 클릭
driver.find_element(By.XPATH, '/html/body/div/div[3]/div/div[2]/form/fieldset/div/div/button').submit()
driver.implicitly_wait(1)
Python
```

임의의 게시글로 들어가기

```
driver.find_element(By.XPATH, '/html/body/div/div[2]/div[2]/div[2]/div[4]/article/div[2]/div[2]/form[1]/table')
driver.implicitly_wait(1)
```

```
report_list
[66]
... Output exceeds the size limit. Open the full output data in a text editor
['Incoming Students APPLY HERE',
 '[2022년도 1학기_국문보고서]',
 '',
 '',
 '',
 '',
 '',
 '국가',
 '노르웨이',
 '파견시기',
 '2022년 1학기',
 '파견기간',
 '1학기',
 '이름',
 '홍**',
 '',
 'University of Oslo',
 '',
 '']
```


진행 상황

데이터 전처리

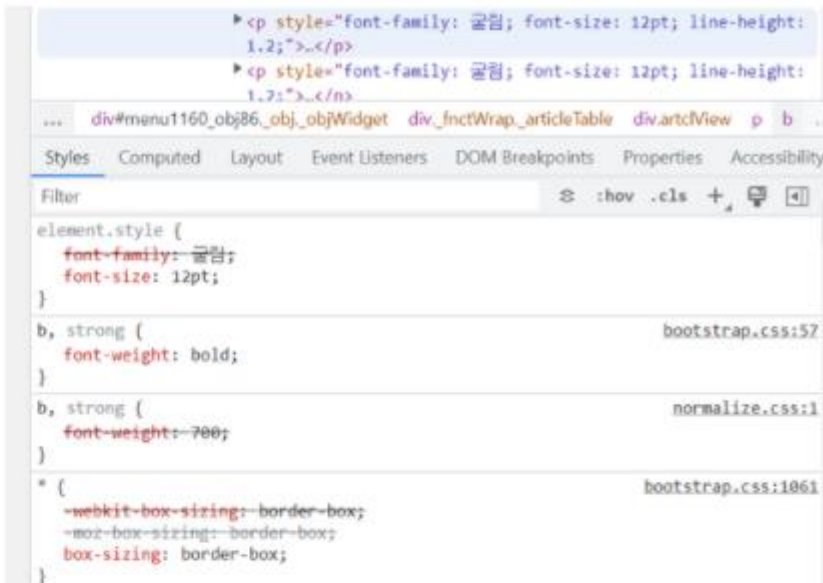
- 파견보고서의 대략적인 형식(목차 + 국문 보고서 + 영문 보고서 + 평가지)을 고려하여 국문으로 된 부분만 이용한다
- 원어로 된 고유명사는 제외할 것

- 본문 내용: P (몇몇 게시글은 본문이 'span', margin이 'p'인 경우도 있다.)

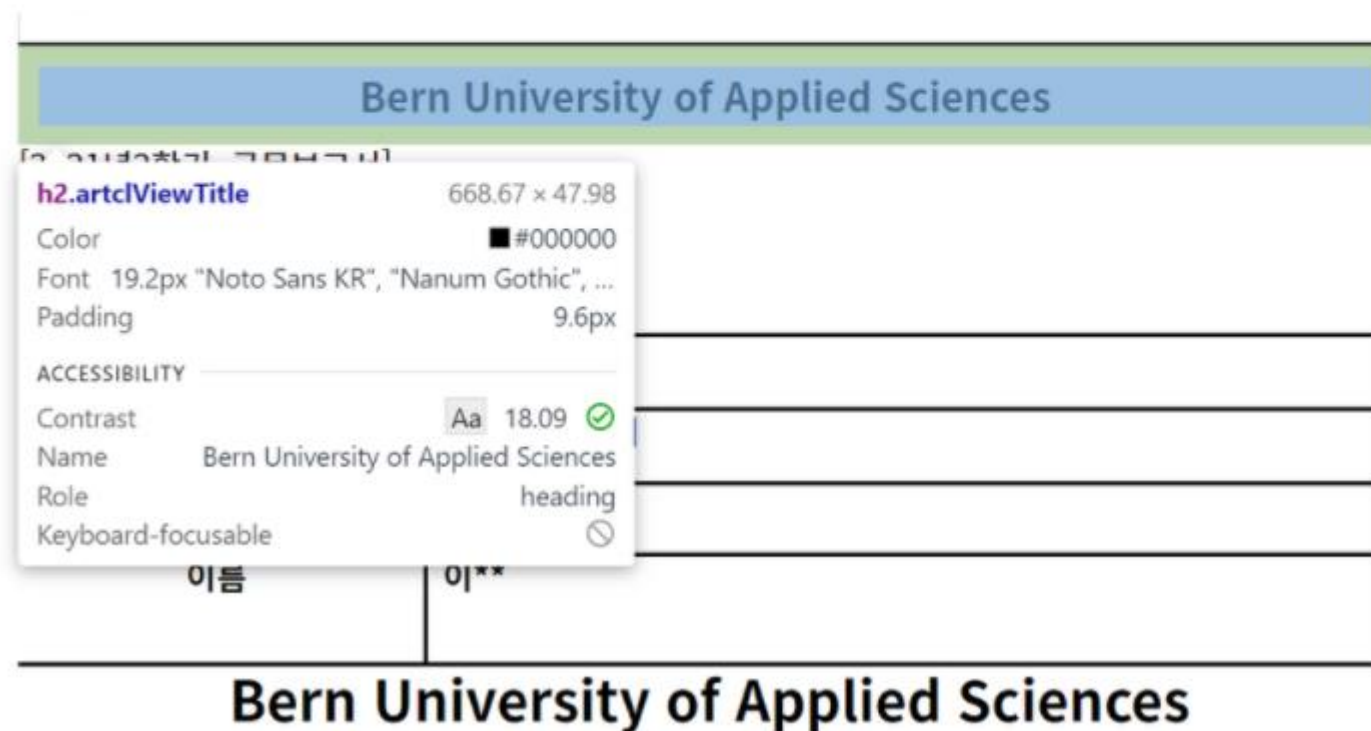
2) 베른

베른은 스위스 수도이자 베른주의 주도입니다. 스위스 중심에서 서쪽, 베른 알프스에서 북쪽으로 20km 떨어진 고원지대에 위치하고 있으며 주로 독일어를 사용하는 독일어권에 해당합니다. 한국에는 잘 알려져 있지 않지만 스위스 내에서도 큰 도시에 속하기 때문에 안내데스크를 방문하거나 장을 보는 것과 같이 일상생활 속에서는 독일어를 전혀 하지 못해도 큰 지장이 없었습니다. 약국이나 역의 티켓사무소 같은 경우는 직원들의 명찰에 어느 언어를 할 수 있는지 표시되어있고 젊은 세대는 모두 영어를 잘하는 편에 속하기 때문에 의사소통에 전혀 무리가 없었다는게 개인적으로 가장 큰 장점이었던 것 같습니다. 베른은 1848년 스위스연방의 연방시(연방의회 소재지)가 되면서 스위스의 대외적인 수도가 되었습니다. 실제로 포털사이트에 검색해보면 베른이 수도로 나오지만, 정착 스위스 현지인들은 수도라는 개념에 익숙하지 않아보이는게 신기했습니다.

구시가지(Oldtown)가 유명한데 이 때문에 구시가지(정된 바 있습니다. 구시가지는 중심으로 16세기 도시 곳곳에 산재해 있어 '분수의 도시'로도 불릴 정도고, 야마시기도합니다. 그리고 3개의 시계탑이 있는데 이 요 건물들 중 하나입니다. 베른에는 아레(Aare)강이 (Rosengarten)까지 오르면 베른 윈스터의 침탑과 함께 잘 볼 수 있어서 날이 좋으면 자주 갔던 곳입니다. 아레 강에서 수영도 할 수 있습니다. 끝으로, 꿈은 꿈공원이 있습니다. 날씨가 추워지면 실내로 들어가기 전에 가시는걸 추천드립니다.(참고: 네이버 지식백과,



- 소제목 : b
- 파견교명 : h2.artclViewTitle



- 영어 보고서 부분 : span