

연세대학교 포탈 수강편람

데이터베이스 구축 및 서비스 개선



산업정보관리론(IIE2102-01)

백상현 2013147051

김지환 2013147049

문승현 2015199049

한재현 2014272026

작성일: 2019-12-16

목차

1. Requirement Analysis

- 1-1. 기존 시스템 문제점
- 1-2. 사용자 요구사항
- 1-3. 프로젝트 목표 설정

2. 데이터 베이스

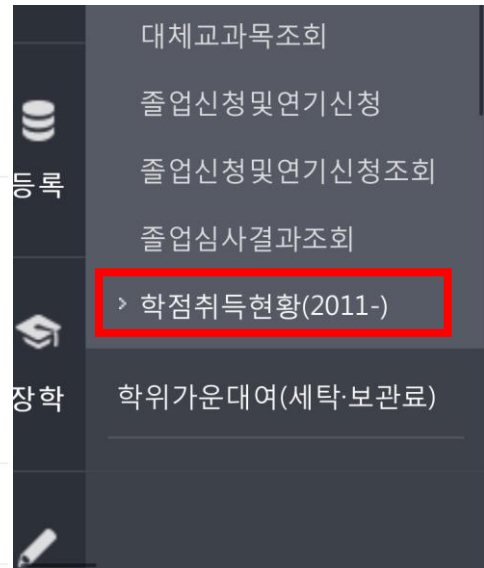
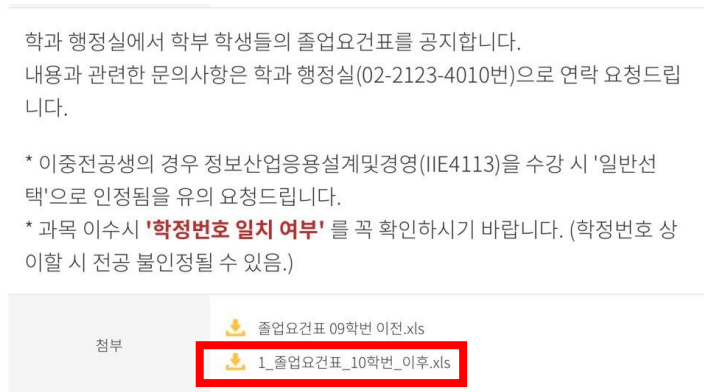
- 2-1. 데이터 베이스 추정
- 2-2. 데이터 수집
- 2-3. 데이터 전처리
- 2-4. 데이터 베이스 재설계
- 2-5. 데이터 입력

3. GUI 서비스 구현

4. 결과 분석

- 4-1. 개선된 사항
- 4-2. 한계점

각 학생은 졸업요건 및 선수과목을 동일한 ‘수강편람’ 화면 안에서 확인할 수 없다. 본인의 졸업요건을 확인하기 위해서는 1)본인의 전공 홈페이지에 들어가거나 2)개인정보가 담긴 ‘학사관리’에 들어가 테이블로 구성된 단순한 구조의 졸업요건표와 자신의 이수과목을 일일이 대조해 확인하거나, 포털에서 학점 취득 현황을 통해 확인해야 한다. 이처럼 졸업요건표 및 선수과목 여부를 수강편람을 확인하면서 수업을 미리 계획하는 과정에서 직관적으로 파악하기 힘든 어려움이 있다.



1. 산업공학과 홈페이지의 졸업요건표

2. '학사관리'의 학점취득현황

이렇듯 수강편람을 조회하는 과정부터 해당 과목이 사용자의 선수과목과 졸업요건에 어떤 영향을 미치는지 확인하는 일련의 과정 곳곳에 문제점이 있다. 사용자는 이에 대해 불편함을 겪는 것으로 확인되었고 요구사항의 도출이 필요했다.

1-2) 사용자 요구사항

이번 프로젝트를 진행하면서 팀의 구성원들이 모두 학부생으로서, 해당 서비스를 충분히 이용해보았다는 점, 아이디어 고도화 당시 2019년 겨울 계절학기에 대한 수강신청이 진행되고 있었다는 점 등을 통해 요구사항에 대해 깊게 고려하였다. 또한 팀 구성원 뿐만 아니라 다른 학우들의 다양한 의견까지 수렴할 수 있었다. 이러한 과정을 통해 연세대학교 사이트(연세포탈)를 통해 수강편람을 이용하는 사용자들의 요구사항을 크게 5가지로 선정할 수 있었다.

- 1) 이수하고자 하는 과목에 대한 자세한 검색이 어렵다.
- 2) 건물 간 위치 및 거리, 소요시간 등 위치 정보에 대한 필요성과 약어 위주의 건물명에 대한 직관적인 정보 제공의 필요.
- 3) 교원에 대한 정보와 교원들이 진행하는 수업의 평가를 따로 찾기 힘들다.
- 4) 해당 과목의 마일리지를 확인하기 번거롭다.
- 5) 개개인의 졸업요건과 수강 현황을 동시에 비교하기 어렵다.

각각의 정보들은 '수강편람' (이수하고자 하는 과목에 대한 검색), 구글맵스, 네이버지도(건물 간 위치 및 거리, 소요시간, 건물명), 에브리타임(교원 정보, 수업 평가), '학사관리' (개개인의 졸업요건 확인)에서 확인할 수 있는 정보들이었다. 이러한 정보를 한 화면에서 제공함으로써 사용자의 불편사항을 최소화하는 새로운 인터페이스를 제공할 수 있을 것으로 기대되었다.

1-3) 프로젝트 목표 설정

위 요구사항으로 얻게 된 요구사항 중 본 프로젝트의 목표를 크게 세 가지로 추려서 설정하였다. 이에 따라 본 프로젝트에서는 다른 학교의 벤치마킹 사례들을 참고해 편람에 수업 내용 및 각 과목에 대한 평가를 반영하고, 강의실 및 건물 위치를 안내하도록 개선하며, 개인별 졸업요건에 대해 확인이 가능하도록 설계할 것을 구상하였다. 먼저 데이터베이스를 구축한 이후 하

나의 화면에서 관련 정보를 최대한 종합적으로 제공할 수 있는 인터페이스를 구축하는 것을 계획했다.

2. 데이터 베이스

2-1) 데이터베이스 추정

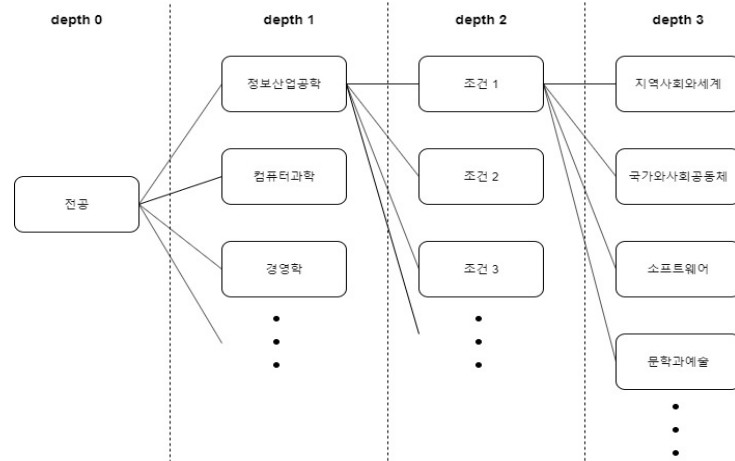
본 프로젝트를 진행하기 앞서 본 팀 구성원들은 수강편람을 위한 정보의 Entity와 attribute를 다음과 같이 정의하였다.

SUBJECT		SUBJECT/BUILDING		STUDENT		PROFESSOR	
OpenYear	년도	OpenYear	년도	StudentID	학번	ProfName	교수이름
OpenSem	학기	OpenSem	학기	Major	전공	Department	소속
G1	구분1	G3	구분3 (개설전공)	IfGraduating	졸업신청여부	BuildingName	연구실 건물
G2	구분2 (중별)	SubjectID	학정번호	SchoolYear	학년	OfficeNumber	연구실 호수
G3	구분3 (개설전공)	BuildingName	건물이름	FINISHEDSUBJECT	StudentID	Phone	연락처
SubjectID	학정번호	RoomNum	건물호수			Email	이메일
SubjectName	과목명	SUBJECT/PROFESSOR				OfficeHour	면담시간
Credit	학점	SUBJECT/PROFESSOR		OpenYear	년도	BUILDING	
SubjectTime	강의시간	OpenYear	년도	OpenSem	학기	BuildingName	건물이름
FullCapa	정원	OpenSem	학기	G3	구분3 (개설전공)	Coordinate	건물좌표
MajorCapa	전공자 정원	G3	구분3 (개설전공)	SubjectID	학정번호	AbbreviatedAs	약어
1Capa	1학년 정원	SubjectID	학정번호	Grade	성적	EVERYTIMEREVIEW	
2Capa	2학년 정원	ProfName	교수이름	WEEKLYSYLLABUS		ReviewID	강의평가 ID
3Capa	3학년 정원	Department	소속	OpenYear	년도	SubjectID	학정번호
4Capa	4학년 정원	APPLIEDSUBJECT		OpenSem	학기	SubjectName	과목명
MaxMileage	최대마일리지	StudentID	학번	G3	구분3 (개설전공)	ProfName	교수이름
IfExchangeP	교환학생 가능여부	OpenYear	년도	SubjectID	학정번호	Department	소속
SyllUploadDate	실라버스 최초등록일	OpenSem	학기	Week	주	Stars	별점
SyllLastUpdate	실라버스 최종수정일	G3	구분3 (개설전공)	DatePeriod	기간	ReviewText	리뷰
SubjectIsFor	수강대상	SubjectID	학정번호	Content	수업내용	CLASSPROPERTY	
SubjectGoal	수업목표 및 개요	MileageBet	신청마일리지	Event	교재범위,과제물	OpenYear	년도
Prerequisite	선수과목	IfEnrolled	수강성공여부	Remarks	비고	OpenSem	학기
SubjectMethod	강좌운영방식					G3	구분3 (개설전공)
SubjectGP	성적평가방법					SubjectID	학정번호
TextBook	교재및참고문헌					Property	유의사항
InfoProg	교수정보						
InfoTA	조교정보						
EngSyll	영문수업계획내용						

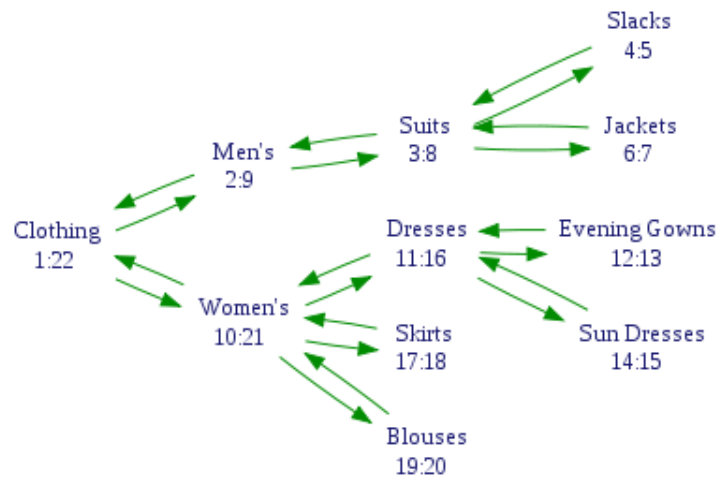
과목, 건물, 교수, 학생, 평점, 졸업요건 등 총 6가지의 영역별 entity를 설정하였다.

1) 과목 Entity

학년도학기 선택 2019학년도 2학기 ▼															
※ 마일리지 수강신청결과															
학정번호-분반-실 습	과목명	학점	담당교수	강의시간	강의실	정원	참여 인원	강원자 잔여 (2강공포함)	학년별평점				교환학생 가능여부	Max Mileage	마일리지
									1	2	3	4		최소값	최대값 평균값
IE2102-01-00	산업정보관리론	3	김우주	월7.8.4.6	공D407	50	116	40 (N)	0	0	0	0	X	36	1 36 21.16
* 정렬: 전공자정렬, 학년별평점은 수강신청기간동안 적용된 값이며, 추가수강신청 및 수강변경 기간에 적용되는 학년별 평점 중의 값(전여석)은 과목 개설학과와 동일하게 따라 반영될 수 있으므로 개설과목에서 검색하여 확인하시기 바랍니다.															
마일리지 순위 결과															
순위	마일리지	강원자/특수전공자 (전공자정원포함(여 부))	신청 과목수	졸업신청	초수강여부	총이수학점/ 졸업이수학점	직전학기이수학점/ 학기당수강학점	학년	수강여부	비고					
1	36	Y (N)	6	N	Y	0.8968	0.8888	4	O						
2	36	Y (N)	6	N	Y	0.8074	0.7894	4	O						
3	36	Y (N)	6	N	Y	0.7500	0.9736	4	O						
4	36	Y (N)	6	N	Y	0.7481	0.8684	3	O						
5	36	Y (N)	6	N	Y	0.7333	0.8888	3	O						
6	36	Y (Y)	6	N	Y	0.6071	0.8157	3	O						
7	36	Y (Y)	6	N	Y	0.4929	0.9736	3	O						
8	36	Y (N)	5	N	Y	0.6740	0.8888	3	O						
9	36	Y (N)	4	N	Y	0.8948	0.9736	4	O						
10	36	Y (Y)	3	Y	Y	0.9214	0.8947	4	O						
11	36	N (N)	6	Y	Y	0.9126	1.0000	4	O						
12	36	N (N)	6	N	Y	0.6888	0.7894	3	O						
13	36	N (N)	6	N	Y	0.5333	0.8157	3	O						
14	36	N (N)	6	N	Y	0.4037	0.8157	2	X						
15	36	N (N)	4	N	Y	0.7825	0.7368	4	X						
16	35	Y (N)	6	N	Y	0.7936	0.8686	4	X						
17	34	Y (Y)	6	N	Y	0.3857	0.8684	2	O						
18	34	Y (Y)	5	N	Y	0.8857	0.5789	4	O						



Nested set model은 leaf에 해당하는 항목에 독자적인 숫자를 좌표형식으로 부여하여 relational database에 좌표 값을 넣는 방식이다. 이를 통해 min/max cardinality문제, 정규형 및 RDB형태를 만족하는 형태로 논리적 설계를 완료하였다.



위의 6가지 entity를 바탕으로 ER 다이어그램을 구축하였다.

2-2) 데이터 수집

구상한 데이터 베이스에 입력할 데이터를 얻기 위해 연세대학교 포탈 수강편람 웹페이지에 대한 크롤링을 실시했다.

크롤링을 진행하기 앞서 여러가지 크롤링 수단을 검토해보았다. 우선 어떤 프로그래밍 언어를 활용하여 크롤링을 진행할지 검토했다. 모든 팀원들에게 파이썬 언어가 가장 익숙했기 때문에 파이썬 언어를 활용하기로 했다. 다음으로 어떤 패키지를 사용하여 크롤링을 진행할지 검토해보았다. 파이썬에서 주로 사용되는 크롤링 패키지는 Request와 Selenium이다. Request의 경우 자바스크립트 객체에 접근하는 것이 까다롭고 연세대학교 포탈 수강 편람에서 많은 버튼들이 자바 스크립트로 구성되어 있어 request를 사용하지 않기로 결정했다. 결과적으로 웹 브라우저를 통해 크롤링을 진행하는 Selenium 패키지에 크롬 드라이버를 연동하여 크롤링을 진행했다.

[연세대학교 포탈 수강편람 크롤링]

Selenium은 특성상 크롤링 진행 속도가 느리다. 실험을 해본 결과 1개 학기에 대한 크롤링은 1일 정도 소요된다. 연세대학교 포탈 수강편람에는 2003년~2019년에 대한 수업이 공개되어 있다. 이를 모두 크롤링하기에는 시간적 제약이 컸다. 그러므로 크롤링 연도 범위를 2019년도로 축소했

고 크롤링할 수업 범위도 문과대학, 상경대학, 경영대학, 이과대학, 공과대학, 생활과학대학, 교육과학대학, 언더우드국제대학, 약학대학, 글로벌인재대학, 교양기초(2019~), 대학교양(2019~), 기초교육(2019~), 공통기초(10~18), 필수교양(10~18), 선택교양(10~18), 학부기초(~2009), 학부필수(~2009), 계열기초(~2009), 학부선택(~2009), 생명시스템대학, 신과대학, 사회과학대학, 그리고 음악대학으로 한정했다.

우리는 연세대학교 포털 수강편람에 공개된 정보를 활용하여 구상한 데이터베이스를 구축하기 위해 다음 페이지의 정보에 접근했다:

1. 수강편람 메인 페이지

수강편람 메인 페이지에서 개설 과목에 대한 검색을 진행할 수 있고, 검색 결과로 특정 영역에 해당하는 혹은 특정 검색어에 해당하는 수업을 표시해준다. 수업 별로 수업의 중별, 학정번호, 학점, 교과목명, 담당교수, 강의시간, 강의실, 유의사항, 국외 교환학생 가능여부를 출력해준다.

2. 각 수업의 수업계획서 (보라색 버튼)

메인 페이지에서 검색된 수업의 수업계획서는 ‘수업계획서 조회’ 버튼에서 제공된다. 이 페이지에서는 수업계획서 뿐만 아니라 담당교수에 대한 정보도 기재되어 있다.

3. 각 수업의 마일리지 수강신청 결과 (주황색 버튼)

메인 페이지에서 검색된 수업에 몇 명의 학생이 얼마의 마일리지를 이용해 해당 수업에 수강신청을 했는지에 대한 정보는 ‘마일리지 수강신청 결과’ 버튼에서 제공된다.

개설전공

학부

기초교육(2019학번~)

전체

전체학점

하반기

2019

1학기

동일교과목 조회

교과목 개요 조회

수업계획서 조회

마일리지 수강신청 결과

영문교과목명 조회

결과내 영어강의만 보기

전체검색

학점번호

하당학기 영어강의 전체검색

중별	학정번호-분반(-실습)	학점	교과목명	담당교수	강의시간	강의실	유의사항	국외교환학...
기교	UCI2001-01-00	1	사회봉사(세브란스병원 자원봉사)	이숙자			⑤(c)	O
기교	UCI2001-02-00	1	사회봉사(자유선택 자원봉사활동)	남석인			16(c)	O
기교	UCI2001-03-00	1	사회봉사(희망원정대-선 봉사후 학점인정)	전종재			16(c)	O
기교	UCI2001-04-00	1	사회봉사(동행프로젝트)	홍원표			20(c)	O
기교	UCI2001-05-00	1	사회봉사(한국어,한국문화를 통한 국제교류프로그램A)	김현철			(c)	O
기교	UCI2001-06-00	1	사회봉사(한국어,한국문화를 통한 국제교류프로그램B)	김성조			(c)	O
기교	UCI2001-07-00	1	사회봉사(함께 나누는 세상A)	박진배			24(c)	O
기교	UCI2001-08-00	1	사회봉사(함께 나누는 세상B)	김형관			폐강	O
기교	UCI2001-09-00	1	사회봉사(함께 나누는 세상C)	한인철			24(c)	O
기교	UCI2001-10-00	1	사회봉사(시각장애인 회랑나눔교실)	나정은			(c)	O
기교	UCI2001-11-00	1	사회봉사(외국인 신입생을 위한 Peer-Mentoring Program)	유성운			26(c)	O
기교	UCI2001-12-00	1	사회봉사(드림 스테디 멘토링 가람1004 봉사단)	정종훈			폐강	O
기교	UCI2001-13-00	1	사회봉사(UIC 융합과학공학부 튜터링 프로그램)	김종학			폐강	O
기교	UCI2001-14-00	1	사회봉사(드림 스테디 멘토링 에델마을 봉사단)	정종훈			(c)	O
기교	UCI2001-15-00	1	사회봉사(평화 만들기-위안부 사안과 관련하여)	정미현			(c)	O

1-15 of 23

1. 수강편람 메인 페이지

<div>수업계획서 조회</div> <div>UC2001-09 (2019학년도 1학기)</div> <div>연세대학교 YONSEI UNIVERSITY</div> <div>최초등록일 2019-02-12 17:57:56 최종수정일 2019-02-13 15:57:22</div> <div>교과목명 사회봉사참여나눔 세상C 학점 1</div> <div>강의실 강의시간</div> <div>담당교수 한인철 담당교수소속 연합신학대학원</div> <div>연우실 주스채를 교목실 내 교목실강실 연락처 02-2123-2005</div> <div>e-mail 및 면담시간 hichul@yonsei.ac.kr</div>				주	기간	수업내용	교재명,교재물	비고
<div>수강대상</div> <div>한한년 누구나</div> <div>자녀 학기 영도로 활동한 학생 적극 환영</div> <div>수업료 및 개요</div> <div>함께나눔세상을 복한의 어린이들과 남한와 청소년들이 처한 어려움을 함께 나누기 위해 2009년 9월에 설립된 단체입니다.</div> <div>‘나를 찾아 떠나는 여행’은 대한민국 청소년들의 꿈과 희망을 찾기 위한 활동한 활동이며, 2009년 1기를 시작으로 16기 까지 활동하였습니다.</div> <div>· 연도 1명과 한의 2명이 한팀이 되어, 각 활동한 주어진 활동한 관습합니다.</div> <div>· 연도 기간별로 연도 그중을 조직하고, 연도장을 선방하여 활동한 교류합니다.</div> <div>· 문 활동한 십삼 자강기적으로 사회복지공공모금회에서 후원하며, 함께 나누는 세상에서 운영합니다.</div> <div>활동지역 및 모집인원</div> <div>· 사회봉사 과목(ABC)각 1명 나을 찾아떠나는여행 연도활동지도</div> <div>· 지역별 모집인원(연도기준)</div> <div>서대문 지역 종교대학교</div> <div>상계동 지역(자원중학교)</div> <div>서동지역 전체(자이아루서터 활동)</div> <div>· 연도 신청서 작성시 활동 희망 지역을 표기하세요 (필수)</div> <div>전공과목(비수학)</div>				1	2019-03-04 2019-03-10	멘토 오리엔테이션 (필수 참석) 3/9 (토) 15:00 - 18:00 주스채를 통일한을 8 101호	나떠여 10주 활동 교육 멘토링 교육 (상설과 대항/정소년의 이해) 봉사기관 배정 봉사기관별 멘토장선발 간식제공	(3.4.) 개강 (3.7. - 3.11.) 수강신청 확인 및 변경
				2	2019-03-11 2019-03-17	멘토멘티 반대식 3/16 (토) 11:00-13:00 멘토 멘티 첫 만남과 1주차 활동 ‘나 이런 사람이야!’	[1주차 실내활동] - 함께나눔세상 소개 나를 찾아떠나는 여행 소개 - 멘토/멘티 짝 만나기 자기소개, 활동 기대 나누기 10주 만남 일정 세우기	(3.7. - 3.11.) 수강신청 확인 및 변경
							[2주차 야외활동] 수포주(야외활동)	

2. 각 수업의 수업계획서 (보라색 버튼)

<div>학년도 학기 선택</div> <div>2019학년도 1학기 ▼</div> <div>※ 마일리지 수강신청결과</div> <div> <div>학점번호-문만-실습</div> <div>과목명</div> <div>학점</div> <div>담당교수</div> <div>강의시간</div> <div>강의실</div> <div>정원</div> <div>참여인원</div> <div>전공자 정원 (2전공포함)</div> <div> <div>1</div> <div>2</div> <div>3</div> <div>4</div> </div> <div>교한학생 가능여부</div> <div>Max Mileage</div> <div>마일리지</div> <div>최소값</div> <div>최대값</div> <div>평균값</div> </div> <div>KOR1001-01-00 한국어문학의이해 3 박애경 조태린 박무영,조갑석 금2,3,4 i진A302 50 20 0 (N) 0 0 0 0 O 36 1 15 3.75</div> <div> <div>* 정원,전공자정원 학년별정원은 수강신청기간동안 적용된 값이며, 추가수강신청 및 수강변경 기간에 적용되는 학년별 정원 등의 값(잔여석)은 과목 개설학과의 결정에 따라 변경될 수 있으므로 개설과목에서 검색하여 확인하시기 바랍니다.</div> <div>마일리지 순위 결과</div> <div> <div>순위</div> <div>마일리지</div> <div>전공자/복수전공자 (전공자정원포함여부)</div> <div>신청과목수</div> <div>출입신청</div> <div>초수강여부</div> <div>총이수학점/출입이수학점</div> <div>직전학기이수학점/학기당수강학점</div> <div>학년</div> <div>수강여부</div> <div>비고</div> </div> <div> <div>1</div> <div>15</div> <div>Y (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.5259</div> <div>0.8421</div> <div>3</div> <div>O</div> <div></div> </div> <div> <div>2</div> <div>15</div> <div>N (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.5000</div> <div>0.8881</div> <div>3</div> <div>O</div> <div></div> </div> <div> <div>3</div> <div>10</div> <div>Y (Y)</div> <div>4</div> <div>N</div> <div>Y</div> <div>0.7500</div> <div>1.0000</div> <div>4</div> <div>O</div> <div></div> </div> <div> <div>4</div> <div>6</div> <div>N (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.4087</div> <div>0.6944</div> <div>2</div> <div>O</div> <div></div> </div> <div> <div>5</div> <div>4</div> <div>Y (N)</div> <div>4</div> <div>N</div> <div>Y</div> <div>0.8222</div> <div>0.6842</div> <div>4</div> <div>O</div> <div></div> </div> <div> <div>6</div> <div>3</div> <div>Y (Y)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.9259</div> <div>0.7894</div> <div>4</div> <div>O</div> <div></div> </div> <div> <div>7</div> <div>3</div> <div>Y (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.6547</div> <div>1.0000</div> <div>3</div> <div>O</div> <div></div> </div> <div> <div>8</div> <div>3</div> <div>N (N)</div> <div>5</div> <div>N</div> <div>Y</div> <div>0.6296</div> <div>0.8157</div> <div>4</div> <div>O</div> <div></div> </div> <div> <div>9</div> <div>2</div> <div>Y (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.8333</div> <div>1.0000</div> <div>4</div> <div>O</div> <div></div> </div> <div> <div>10</div> <div>2</div> <div>Y (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.6444</div> <div>0.8694</div> <div>3</div> <div>O</div> <div></div> </div> <div> <div>11</div> <div>2</div> <div>N (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.4803</div> <div>1.0000</div> <div>2</div> <div>O</div> <div></div> </div> <div> <div>12</div> <div>2</div> <div>N (N)</div> <div>5</div> <div>N</div> <div>Y</div> <div>0.4259</div> <div>0.7894</div> <div>2</div> <div>O</div> <div></div> </div> <div> <div>13</div> <div>1</div> <div>Y (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>1.0000</div> <div>0.6866</div> <div>4</div> <div>O</div> <div></div> </div> <div> <div>14</div> <div>1</div> <div>Y (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.8074</div> <div>0.9473</div> <div>4</div> <div>O</div> <div></div> </div> <div> <div>15</div> <div>1</div> <div>Y (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.7757</div> <div>0.9722</div> <div>4</div> <div>O</div> <div></div> </div> <div> <div>16</div> <div>1</div> <div>Y (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.6587</div> <div>1.0000</div> <div>3</div> <div>O</div> <div></div> </div> <div> <div>17</div> <div>1</div> <div>Y (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.5555</div> <div>0.9736</div> <div>3</div> <div>O</div> <div></div> </div> <div> <div>18</div> <div>1</div> <div>N (N)</div> <div>6</div> <div>Y</div> <div>Y</div> <div>1.0000</div> <div>0.0555</div> <div>4</div> <div>O</div> <div></div> </div> <div> <div>19</div> <div>1</div> <div>N (N)</div> <div>6</div> <div>N</div> <div>Y</div> <div>0.9444</div> <div>0.8333</div> <div>4</div> <div>O</div> <div></div> </div> <div> <div>20</div> <div>1</div> <div>N (N)</div> <div>5</div> <div>N</div> <div>Y</div> <div>0.5000</div> <div>0.8750</div> <div>3</div> <div>O</div> <div></div> </div> </div>
--

3. 각 수업의 마일리지 수강신청 결과 (주황색 버튼)

위에서 설명했듯이 각 과목에 대해서 3개의 페이지에 접근했고, 다음과 같이 5개의 csv 파일을 생성했다:

1. CLASSPROPERTY.csv

이 테이블에는 크롤링된 과목이 가지는 수업 특성(영어강의, 절대평가 등)이 담겨있다. 과목을 식별하는 PK column들과 특성 column으로 구성되어 있다. 1개의 과목이 2개 이상의 특성을 가지는 경우에 한 cell에 2개이상의 특성이 입력된다. 추후 데이터 전처리를 통해 이 문제를 해결했다.

In [13]: raw_CLASSPROPERTY.head()

Out[13]:

	OpenYear	OpenSem	G1	G2	G3	SubjectId	Property
0	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-01-00	㉔㉕㉖㉗(c)	
1	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-02-00	㉔㉕㉖㉗(c)	
2	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-03-00	㉔㉕(c)	
3	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-04-00	㉔㉕(c)	
4	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-05-00	㉔㉕(c)	

CLASSPROPERTY.csv 샘플

총 수집된 row 개수: 6045

Note: 크롤링을 완료하기 전에는 몇 개의 column으로 수업을 정확히 식별할 수 있는지 확인할 수 없기 때문에 우선적으로 필요 이상의 수업 식별 column을 기록했다.

2. SUBJECT.csv

이 테이블에는 과목 당 하나씩 존재하는 값들을 기록했다.

```
In [11]: SUBJECT.head(2)
```

```
Out[11]:
```

OpenYear	OpenSem	G1	G2	G3	SubjectID	SubjectName	Credit	SubjectTime	FullCapa	참여인원	MajorCapa	1capa	2capa	3capa	4capa	IfExchangeP	MaxMi
2019	1학기	학부	문과대학	인문대학	HUM3005-01-00	윤동주와연세 인문정신	3	화8,9,10	100	3	0 (N)	0	0	0	0	0	O
2019	1학기	학부	문과대학	인문대학	HUM4001-01-00	인문학자양성 세미나	3	목7,8,9	12	3	0 (N)	0	0	0	0	0	O

SUBJECT.csv 샘플

총 수집된 row 개수: 5703

3. SUBJECT_BUILDING.csv

이 테이블에는 수업의 강의실에 대한 정보를 기록했다.

In [19]: raw_SUBJECT_BUILDING.head()

Out[19]:

	OpenYear	OpenSem	G1	G2	G3	SubjectId	BuildingName
0	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-01-00	I층301	
1	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-02-00	I층301	
2	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-03-00	I층301	
3	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-04-00	I층301	
4	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-05-00	I층301	

SUBJECT_BUILDING.csv 샘플

Note: 크롤러 작동 구조상 1개의 과목이 2개 이상의 강의실을 가지는 경우에 1개의

cell에 2개의 강의실이 기록된다. 이에 대해 전처리를 진행했다. 또한, 프로젝트의 목표에 따르면 건물이름과 강의실 호수가 별도로 기록되어야 하지만 이 역시 분리가 안되어 있다. 이 문제점에 대해서도 전처리를 진행했다.

총 수집된 row 개수: 6045

4. SUBJECT_PROFESSOR.csv

이 테이블에는 수업의 담당교수에 대한 정보를 기록했다.

```
In [21]: raw_SUBJECT_PROFESSOR.head()
```

```
Out[21]:
```

	OpenYear	OpenSem	G1	G2	G3	SubjectID	ProfName	Department	OfficeNumber	Phone	Email
0	2019	1학기	학부	교양기초(2019~)	채플	YCA1001-01-00	곽호철	연합신학대학원	Christine Chapel	032-749-2800	chapeli@yonsei.ac.kr
1	2019	1학기	학부	교양기초(2019~)	채플	YCA1001-02-00	곽호철	연합신학대학원	Christine Chapel	032-749-2800	chapeli@yonsei.ac.kr
2	2019	1학기	학부	교양기초(2019~)	채플	YCA1001-03-00	곽호철	연합신학대학원	크리스틴 채플 105	032-749-2800	chapeli@yonsei.ac.kr
3	2019	1학기	학부	교양기초(2019~)	채플	YCA1001-04-00	곽호철	연합신학대학원	크리스틴 채플 105	032-749-2800	chapeli@yonsei.ac.kr
4	2019	1학기	학부	교양기초(2019~)	채플	YCA1001-05-00	정용한	연합신학대학원			NaN

SUBJECT_PROFESSOR.csv 샘플

Note: 이 테이블은 단순히 모든 과목의 수업계획서에 기록된 값을 가져온 것이다. 그러므로, 같은 담당교수를 가지는 과목이 오타로 인해 다른 값의 정보를 가지는 경우가 존재한다. 이는 추후 전처리를 통해 최대한의 정리 및 통합을 진행했다.

과목과 교수 사이에는 Multivalued Dependency가 존재한다. 1개의 과목을 담당하는 교수가 여러 명인 경우가 존재하고, 1명의 교수가 여러 과목을 담당하는 경우도 빈번하다. 연세대학교 포탈 수강편람에서 이런 경우 교수 정보를 담는 테이블을 별도로 표시하기 때문에 Multivalued Attribute에 알맞은 형식으로 SUBJECT_PROFESSOR.csv를 write할 수 있었다.

담당교수	조태섭	담당교수소속	문과대학 사학
연구실	외술관 405호	연락처	010-3772-3385, 02-2123-2378
e-mail 및 면담시간	chots@yonsei.ac.kr	화요일, 목요일 오후 기타 연락하에...	
담당교수	최민규	담당교수소속	문과대학 사학
연구실	외술관 519호	연락처	010-6476-1794
e-mail 및 면담시간	mingyui@daum.net		

한 과목을 담당하는 교수가 여러 명인 경우 크롤링하기 용이한 형태이다

총 수집된 row 개수: 6870

5. WEEKLYSYLLABUS.csv

```
In [23]: raw_WEEKLYSYLLABUS.head()
```

```
Out[23]:
```

	OpenYear	OpenSem	G1	G2	G3	SubjectID	Week	DatePeriod	Content	Event	Remarks
0	2019	1학기	학부	교양기초(2019~)	채플	YCA1001-01-00	1	2019-03-04 2019-03-10	TBA	(3.4.) Spring semester classes begin	(3.7. - ...)
1	2019	1학기	학부	교양기초(2019~)	채플	YCA1001-01-00	2	2019-03-11 2019-03-17	TBA	(3.7. - 3.11.) Course add and drop period	
2	2019	1학기	학부	교양기초(2019~)	채플	YCA1001-01-00	3	2019-03-18 2019-03-24	TBA		
3	2019	1학기	학부	교양기초(2019~)	채플	YCA1001-01-00	4	2019-03-25 2019-03-31	TBA		
4	2019	1학기	학부	교양기초(2019~)	채플	YCA1001-01-00	5	2019-04-01 2019-04-07	TBA	(4.3. - 4.5.) Course withdrawal period	

WEEKLYSYLLABUS.csv 샘플

이 테이블에는 수업의 주별 수업계획에 대한 정보를 기록했다.

총 수집된 row 개수: 92266

6. APPLIEDSUBJECTS.csv

이 테이블에는 특정 과목에 신청된 마일리지의 정보를 담았다.

In [4]:

mileage.head(2)

Out [4]:

연도	학기	구분1	구분2	구분3	학점번호- 분반-실습	과목명	학점	담당교수	강의시간	마일리지	전공자/복수전공자(전공자정원포함여부)	신청과목수	졸업신청	조수강여부	총이수학점/ 졸업이수학점	직전학기이수학점/ 학기당수강학점	학년	수강여부	비고
0	2019	1	학부	교양기초 (2019~)	재품	YCA1001-01-00	재품 (1)	0.0	과목명 화2 ...	2	Y (Y)	6	N	Y	0.4603	0.8421	3	O	
1	2019	1	학부	교양기초 (2019~)	재품	YCA1001-01-00	재품 (1)	0.0	과목명 화2 ...	2	Y (Y)	5	Y	Y	0.8730	0.8333	4	O	

2 rows x 34 columns

APPLIEDSUBJECTS.csv 샘플

총 수집된 row 개수: 131883

[에브리타임 크롤링]

다음으로 과목별 리뷰 데이터베이스를 구축하기 위해 에브리타임에 대해 크롤링을 진행했다. 에브리타임은 전국 400개 대학에 대해 학교별 익명 커뮤니티 및 시간표 서비스를 제공하는 웹서비스이다. 또한 에브리타임은 대학생들이 시간표를 만드는데 가장 많이 쓰는 어플리케이션 서비스 중 하나로, 방대한 사용자들을 기반으로 많은 양의 수업 리뷰를 보유하고 있다.

에브리타임 크롤링을 진행하는데 다음과 같은 문제점이 있었다:

1. 기술적 문제

에브리타임 크롤링을 진행하며 일정 시간 이상 자동으로 웹페이지를 탐색하면 로그아웃되는 현상이 발생했다. 이는 에브리타임에 자동화 소프트웨어를 사용하는 접속자를 식별하고 방해하는 기능이라고 추측된다. 꾸준히 로그인을 하며 크롤링을 진행했지만, 방대한 양의 에브리타임 교과목 리뷰를 수집하기에는 시간 측면의 무리가 있었다.

2. 법적 문제

중간 발표에서 김우주 교수님이 사기업에서 운영하는 데이터를 사용하는데 법적 문제가 존재할 가능성이 높다는 피드백을 하셨다. 이를 고려하여 에브리타임 이용약관을 확인했다. 에브리타임 이용약관 제9조(저작권의 귀속)에 의하면 에브리타임에 게시된 모든 정보의 복제권 및 전송권을 포함한 모든 권리는 에브리타임이 가지고 회사의 승인 없이 데이터베이스의 전부 또는 일부를 복제·배포·방송 또는 전송할 수 없다고 명시되어 있다.


앞서 설명한 법적 문제 때문에 회사의 동의없이 데이터 수집을 계속 진행할 수 없었다. 그러므로, 회사의 동의를 받고자 에브리타임社에 이메일을 보내 데이터 수집을 허가해줄 것을 요청했지만 거절당했다.

<p>제9조(저작권의 귀속)</p> <p>1. 회사는 유용하고 편리한 서비스를 제공하기 위해, 2009년부터 서비스 및 서비스 내부의 기능(시간표, 강의평가, 책방 서비스 등의 체계와 다양한 기능(학점계산기, 검색 기능 등)을 직접 설계 및 운영하고 있는 데이터베이스 제작자에 해당합니다. 회사는 저작권법에 따라 데이터베이스 제작자는 복제권 및 전송권을 포함한 데이터베이스 전부에 대한 권리를 가지고 있으며, 이는 법률에 따라 보호를 받는 대상입니다. 그러므로 이용자는 데이터베이스 제작자인 회사의 승인 없이 데이터베이스의 전부 또는 일부를 복제·배포·상승 또는 전송할 수 없습니다.</p> <p>2. 회사가 작성한 게시물에 대한 권리는 회사에게 귀속되며, 회원이 작성한 게시물에 대한 권리는 회원에게 귀속됩니다. 단, 회사는 서비스의 운영, 확장, 홍보 등의 필요한 목적으로 회원의 저작물을 합리적이고 필요한 범위 내에서 별도의 허락 없이 수정하여 무상으로 사용하거나 제3자에게 제공할 수 있습니다. 이 경우, 회원의 개인정보는 제공하지 않습니다.</p> <p>3. 회사는 전항 이외의 방법으로 회원의 게시물을 이용할 경우, 서비스 내부 알림 수단과 회원의 연락처를 이용하여 회원의 동의를 받아야 합니다.</p>	<p>[에브리타임] 문의하신 내용에 답변 드립니다. 가* ★</p> <p>2019년 12월 5일 (목) 오전 11:24</p> <p>안녕하세요. 에브리타임입니다. 문의하신 내용에 답변 드립니다.</p> <p>항상 에브리타임에 관심가져주시고, 좋은 제안 주신 것 감사드립니다.</p> <p>아쉽게도, 현재 크롤러 및 API 지원 등은 허가/전달이 어렵습니다. 전달해주신 내용은 검토해보도록 하겠으며, 추후 지원할 수 있는 부분이 있다면 지원할 수 있도록 노력하겠습니다.</p> <p>앞으로도 더욱 발전하는 에브리타임이 될 수 있도록 노력하겠습니다.</p> <p>감사합니다.</p> <p>* 본 메일은 발신 전용 메일이므로 답장을 받을 수 없는 메일입니다.</p> <p>* 본 메일은 문의 시 입력한 이메일 주소로 발송된 메일입니다. 만약 메일이 잘못 발송되었다면 즉시 삭제해 주시기 바랍니다.</p> <p>Copyright © 에브리타임. All Rights Reserved.</p>
--	---

법적 제약이 존재하는 상태에서 크롤링을 계속한다면 서비스 이용정지 등의 페널티를 부여받을 수 있고 법정 대응의 가능성도 존재하여 크롤링된 데이터를 사용하지 않기로 결정했다.

[지도 데이터 크롤링]

A	B
1 BuildingName	Coordinate
2 001 연세암병원	37.561234, 126.938754
3 002 의과대학(본관 신관 도서관)	37.560800, 126.940213
4 003 제중관	37.561999, 126.939365
5 004 종합관/장례식장	37.561437, 126.940839
6 005 세브란스병원 본관	37.562266, 126.941068
7 006 재활병원	37.563063, 126.940395
8 007 어린이병원	37.562548, 126.939907
9 008 심장혈관병원	37.562703, 126.939459
10 009 광혜관	37.562376, 126.939073
11 010 안-이비인후과병원	37.561585, 126.938298
12 011 치과대학병원	37.562965, 126.941744
13 012 치과대학	37.563399, 126.942356
14 013 에비슨의생명연구센터	37.563939, 126.942576
15 014 간호대학	37.563620, 126.941911
16 015 임상의학연구센터	37.560478, 126.939443
17 016 동문회관	37.562745, 126.942454
18 017 러들교수동	37.563612, 126.941338
19 101 정문	37.560414, 126.936607
20 102 공학원	37.560890, 126.935513
21 103 포스코브리지	37.561222, 126.935674
22 104 야구장	37.561162, 126.934075



연세대학교 홈페이지에서 연세대학교 빌딩 목록을 얻었고 각 건물의 위치 좌표를 구글 지도를 이용해 획득했다.

2-3) 데이터 전처리

SUBJECT 테이블을 제외한 모든 테이블은 데이터베이스에 INSERT하기 전 전처리를 할 필요가 있었다. 요구되는 전처리 과정을 파이썬 코드를 이용해 진행했다. 수집된 데이터의 양이 적지 않기 때문에 시간 효율성을 위해 pandas 패키지에서 제공하는 pandas.DataFrame() 객체와 pandas.Series.apply() 메소드를 적극 활용했다.

[CLASSPROPERTY 전처리]

크롤링은 총 4대의 컴퓨터에서 진행했기 때문에 동일한 테이블이 4개씩 생성되었다. 4대의 컴퓨터에 분산되어 있는 csv파일들을 한 곳에 모아 하나의 단일 테이블로 통합했다.

데이터 수집 부분에서 언급했듯이 크롤러의 작동 구조상 1개의 과목이 2개 이상의 특성을 가지고 있을 때 해당 과목의 Property Attribute에 2개 이상의 특성이 기록된다. 이는 Multivalued Attribute으로 여러 개의 row으로 분리할 필요가 있다. 해당 작업을 파이썬 코드를 통해 진행했다.

	OpenYear	OpenSem	G1	G2	G3	SubjectId	Property
전	0	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-01-00	㉠㉡㉢㉣
	OpenYear	OpenSem	G1	G2	G3	SubjectId	Property
	0	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-01-00	㉠
	1	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-01-00	㉡
	2	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-01-00	㉢
	3	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-01-00	㉣
후	4	2019	1학기	학부 교양기초(2019~)	채플	YCA1001-01-00	(c)

전처리 결과

[SUBJECT_BUILDING 전처리]

CLASSPROPERTY 테이블과 동일하게 여러 대의 PC에서 수집된 SUBJECT_BUILDING 테이블들을 하나의 단일 테이블로 통합했다.

SUBJECT_BUILDING 테이블 역시 multivalued attribute이 존재한다. 강의실과 실습실이 다른 경우 1개의 과목이 2개 이상의 강의실을 가지는 경우가 발생한다. 파이썬 코드를 이용해 이런 row들을 여러 개의 row으로 분리시켰다.

SUBJECT_BUILDING테이블의 다른 문제점은 강의 건물과 강의실이 약자로 간략하게 1개의 attribute으로 기록된다는 것이다. 이 문제는 정규 표현식을 이용하여 해결했다.

하지만 극복할 수 없는 문제가 있었다. 연세대학교 포털 수강편람 시스템에서 제공하는 강의실 정보는 체계적인 데이터베이스 쿼리를 통해 제공되는 것이 아님을 전처리를 진행하며 확인했다. 이런 판단의 근거는 같은 건물에 대해 다른 표기법이 다수 존재한다는 것이다. 예를 들어 체육관 농구코트에서 진행되는 수업은 3가지가 넘는 방법으로 표기된다(체육관 농구코트, 체 육관 농구장 등). 이런 현상은 체계적인 데이터베이스 쿼리가 이루어진다면 일어날 수 없다. 강의실에 대한 표기법이 통일되어 있지 않고 수집된 SUBJECT_BUILDING테이블의 row수가 6000개가 넘는 상황에서 건물과 건물/과목 테이블을 정제하고 데이터베이스에 포함시키기에는 시간적 제한이 있다고 판단했다. 이런 이유에서 부득이하게 건물과 관련된 정보를 데이터베이스에서 제외하기로 결정했다.

[WEEKLYSYLLABUS 전처리]

한 학기는 16주로 구성 되어있기 때문에 WEEKLYSYLLABUS 테이블을 데이터베이스에 적합한 형태로 크롤링해 올 수 있었다. 그러므로, WEEKLYSYLLABUS 테이블은 특별한 전처리 과정없이 여러 개의 PC에서 생성된 테이블을 통일시키는 작업만 진행했다.

[SUBJECT_PROFESSOR 전처리]

다른 테이블과 동일하게 4대의 PC에서 수집된 테이블들을 하나의 SUBJECT_PROFESSOR 테이블로 통합했다. SUBJECT_PROFESSOR테이블의 정보는 수업계획서에서 수집했다. 각 수업의 수업계획서 내용은 해당 수업의 담당교수가 직접 입력하는 것으로 추정된다. 그 이유는 교수 이메일, 교수 전화번호 등 정보들이 통일된 양식없이 입력되어 있기 때문이다. 이런 현상은 데이터베이스 호환을 통해 자동으로 해당 정보들이 입력된다면 발생할 수 없다.

담당교수	곽호철	담당교수	곽호철
연구실	Christine Chapel	연구실	Christine Chapel
e-mail 및 면담시간	chapeli@yonsei.ac.kr	e-mail 및 면담시간	chapli@yonsei.ac.kr
담당교수	김옥경	담당교수소속	
연구실		연락처	
e-mail 및 면담시간	verokim22@hanmail.net/ 면담은 사전 약속 후		
담당교수	홍석민	담당교수소속	답
연구실	백N302		
e-mail 및 면담시간	ssmhong@yonsei.ac.kr // 월 2,5,6교시		

다양한 형태로 입력되어 있는 Email 값

이렇게 통일되지 않은 양식에서 교수진의 이메일 정보를 추출하기 위해 re모듈의 정규표현식 `([a-zA-Z0-9_+.]+@[a-zA-Z0-9-]+\.[a-zA-Z0-9-]+)`을 사용했다. 또한 대소문자 구분이 없는 이메일이 대문자로 입력되어 있는 경우가 존재해 추출된 이메일을 소문자로 통일했다.

[PROFESSOR 테이블 생성]

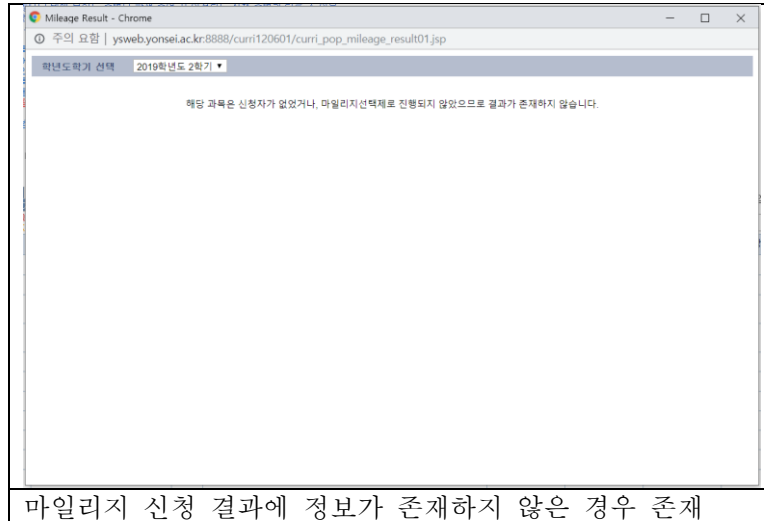
PROFESSOR 테이블을 생성하기 위해 SUBJECT_PROFESSOR에서 ProfName, Department, OfficeNumber, Phone, Email column들을 분리시켰다. 연세대학교의 경우 교수 1인당 여러 개의 수업을 가리키는 경우가 많기 때문에 단순히 SUBJECT_PROFESSOR에서 분리된 테이블에는 많은 양의 중복값이 존재한다. 그렇기 때문에 이 테이블에서 모든 PK (ProfName, Department, Email)의 조합이 유니크 할 수 있도록 `pd.DataFrame.drop_duplicates(subset=[])` 메소드를 적용했다.

[Foreign Key Integrity 확인]

앞서 전처리 과정에 대해 설명한 테이블들은 PROFESSOR 테이블을 제외하고 모두 SUBJECT테이블의 PK(OpenYear, OpenSem, G2, G3, SubjectID)를 포함하고 있다.

이론적으로 크롤러의 탐색 범위에 포함된 모든 과목에 대해 오류없이 크롤링이 진행된다면, 기타 테이블들에 존재하는 SUBJECT PK(OpenYear, OpenSem, G2, G3, SubjectID)값은 SUBJECT에 존재해야 한다.

하지만, 크롤러의 작동 특성상 그렇지 않은 경우가 있다. 이 문제의 원인은 크롤러의 SUBJECT 테이블의 write row 코드가 해당 과목의 마일리지 신청 결과버튼을 조회하는 코드 블록에 있기 때문이다. 교과 과목 중 일부 폐강되어 마일리지 신청 결과의 내용이 없는 경우가 존재한다. 이때 크롤러의 마일리지 신청 결과버튼을 조회하는 부분의 코드 블록이 try, except 예외 처리문으로 인해 실행되지 않아 SUBJECT 테이블 row는 write되지 않지만 해당 과목의 다른 테이블 row는 write되는 현상이 간혹 발생한다. 이 경우 다른 테이블에 존재하는 SUBJECT PK(OpenYear, OpenSem, G2, G3, SubjectID)값이 SUBJECT 테이블에 존재하지 않아 Foreign Key Constraint가 위반되는 경우가 발생한다. Foreign Key Constraint의 위배를 방지하기 위해 파이썬 코드로 기타 테이블에서 SUBJECT테이블에 존재하지 않는 SUBJECT PK(OpenYear, OpenSem, G2, G3, SubjectID) 값을 가지는 row를 삭제하는 전처리 과정을 진행했다.



[MAJOR 테이블 생성]

서비스 시연을 위해 간략하게 3개 전공(정보산업공학, 전자전기공학, 건설환경공학)에 대한 정보를 조사해 입력했다.

[STUDENT 테이블 생성]

학생들의 신상 정보는 민감한 개인정보이다. 그런 이유에서 연세대학교에서는 학생과 관련된 데이터를 제공하고 있지 않다. STUDENT테이블은 본 프로젝트가 목표로 하는 서비스를 구축하는데 필수적이다. 또한, 마일리지 신청 내역인 APPLIEDSUBJECT테이블에 마일리지 신청자의 PK(Student ID)가 포함되기 때문에 STUDENT테이블을 생성할 필요가 있다. numpy.random 객체를 활용하여 STUDENT테이블을 생성했다.

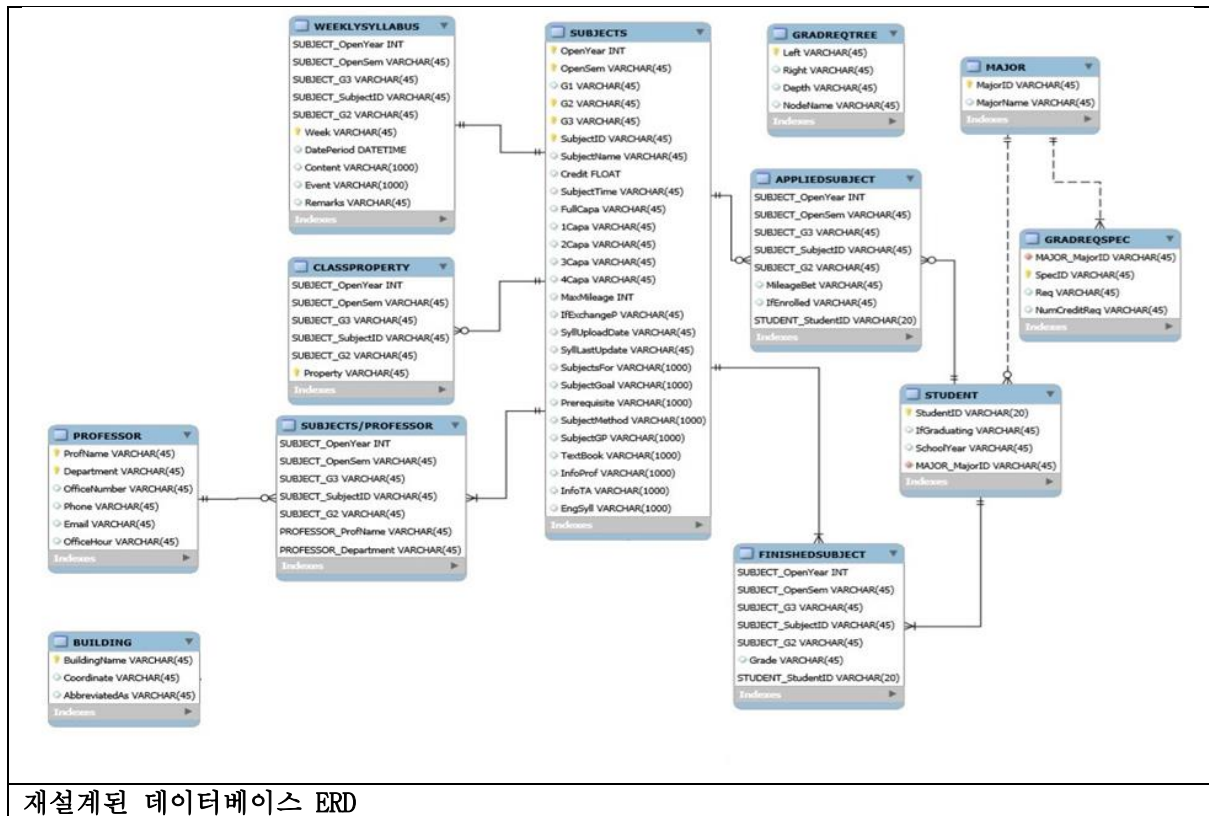
[APPLIEDSUBJECT 전처리]

크롤러가 수집한 APPLIEDSUBJECT 테이블은 특정 과목에 어떤 마일리지 신청이 있는지에 대한 정보를 담고 있다. APPLIEDSUBJECT 테이블에서는 마일리지 신청자에 대한 정보가 누락되어 있다. 데이터베이스 구축을 위해 랜덤하게 STUDENT테이블에 존재하는 학생들을 마일리지 신청 내역에 연결했다.

2-4) 데이터 베이스 재설계

데이터 전처리(2-3) 부분에서 발견된 한계점들은 데이터 베이스 추정(2-1)단계에서 제안한 데이터베이스 디자인을 그대로 가져갈 수 없음을 의미한다. 최대한 우리가 고안한 서비스 구축에 지장이 없도록 데이터베이스의 디자인을 수정했다.

수정된 결과는 다음 그림과 같다. 건물 데이터를 확보할 수 없었기 때문에 SUBJECT_BUILDING 테이블 그리고 BUILDING 테이블을 삭제했다. 에브리타임 데이터를 사용하는데 법적 문제가 존재하기 때문에 EVERYTIMEREVIEW 테이블을 삭제했다.



재설계된 데이터베이스 ERD

2-5) 데이터 입력

시간 효율성을 위해 데이터 전처리에 pandas 패키지를 이용했기 때문에 모든 데이터를 `pandas.DataFrame()` 형식으로 보유하고 있었다. 또한, 데이터의 양이 비교적 크기 때문에 `pymysql`과 `for`문을 이용해 `INSERT INTO` 쿼리를 진행하는 것은 효율적이지 않은 선택이었다. 그러므로, `pandas.DataFrame()`과 간편하게 연동되고 대량의 `INSERT INTO`문을 구현하기 간편한 `sqlalchemy.create_engine` 객체와 `pandas.DataFrame.to_sql()` 메소드를 활용하여 앞서 정제한 데이터를 mysql 데이터베이스에 입력했다.

추가적으로 프로젝트 시연을 위해 `FINISHEDSUBJECT` 테이블에 4개의 row를 입력했다. 그리고 정보산업공학과에 한해 nested set 형태의 `GRADREQTREE`를 작성하여 입력했다.

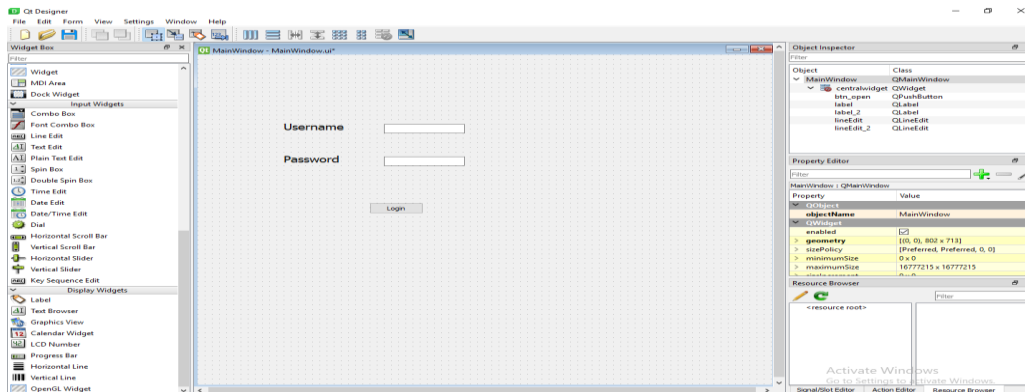
데이터베이스에 INSERT된 row entity 개수:

SUBJECTS	5703	PROFESSOR	2487
SUBJECTS_PROFESSOR	6512	CLASSPROPERTY	10598
MAJOR	3	STUDENT	131974
APPLIEDSUBJECT	131883	GRADREQTREE	12
WEEKLYSYLLABUS	88314	FINISHEDSUBJECT	4

3. GUI 서비스 구현

[tk-inter와 PyQt5 비교분석]

GUI 서비스를 제공하기 위해서는 프로그램의 선정이 필요하다. 제한된 시간에 조작이 보다 효율적이고 편리한 프로그램이 PyQt5이라고 생각되었다. PyQt5의 장점은 Designer에 있다. Designer는 각종 버튼과 툴들을 삽입하여 보다 직관적이게 GUI를 구축할 수 있게 해주며, 이렇게 구축된 PyQt5파일을 Python파일로 변환하여 자동적으로 python코드들로 변환해준다. 이러한 pyqt5의 장점은 본 프로젝트를 pyQt5로 수행하기에 충분한 명분이 되었다.



[User Requirement analysis]

현재 많은 학생들은 연세대학교 수강편람 시스템에 불편함을 겪고 있다. 졸업을 일년 앞둔 대학교 3학년인 학생이 있다. 그는 지난 3년간 100학점 이상을 이수해왔다. 그가 들은 과목들은 각 과목마다 학정번호를 가지고 있다. 졸업을 위해 그는 각 과목들의 과목명과 그 과목이 전공인지, 교양인지, 졸업을 위해 몇 학점을 들어왔는지, 더 들어야 하는지 알아야 했다. 하지만 졸업요건표에 나타나 있는 과목명과 수강편람의 과목명이 다른 과목이 있어 학정번호를 일일이 비교해 수강신청을 계획하기도 했다. 그 과목의 평가방법과 개설연도, 교수님의 수업방식도 고려해야 했다. 이러한 평가를 알아보기 위해 시간표 작성 앱인 '에브리타임'을 매 학기 별도로 반영하였다. 그는 3학년 들어와서 전공 또는 필수교양 위주로 듣는다. 전공 또는 필수교양을 적어도 한 과목씩 듣는다. 그가 들은 과목의 영역이 어디인지, 무슨 영역을 더 들어야 하는지 찾아야 한다. 현재 수강신청 사이트에 개설되는 과목들이 학기, 전공, 학점수에 따라 저장되어 있고, 별도로 학정번호, 교과목명, 교수명, 강의시간 등의 검색을 통해 찾을 수 있다.

위 사례는 5가지의 시사점을 제공한다. 첫 번째로는 상세검색이 까다롭다는 것이다. 범위를 모두 지정했음에도 불구하고 입력란에 수업명을 입력하여 검색하는 순간, 전에 설정하였던 범위는 무용지물로 변한다. 두 번째로는 건물간 위치와 거리의 줄임말을 파악하기 힘들다는 점이다. 수강 편람 시스템에서는 건물명을 모두 약어로 표시하고 있기 때문에 건물 정식 명칭 및 위치를 파악하기 힘들다. 추가적으로 교수님 및 수업평가를 따로 찾기 힘들며 마일리지 확인 또한 확인하기 어렵다. 무엇보다 가장 큰 문제라고 생각되는 것은 연세대학교 수강편람 시스템에 졸업요건표가 연동되어 있지 않다는 점이다. 학생들은 매년 수강신청 시즌이 되면 졸업요건표와 수강신청편람을 대조하느라 바쁘다. 이것은 매우 비효율적인 구조이며 번거로운 일일 것이다.

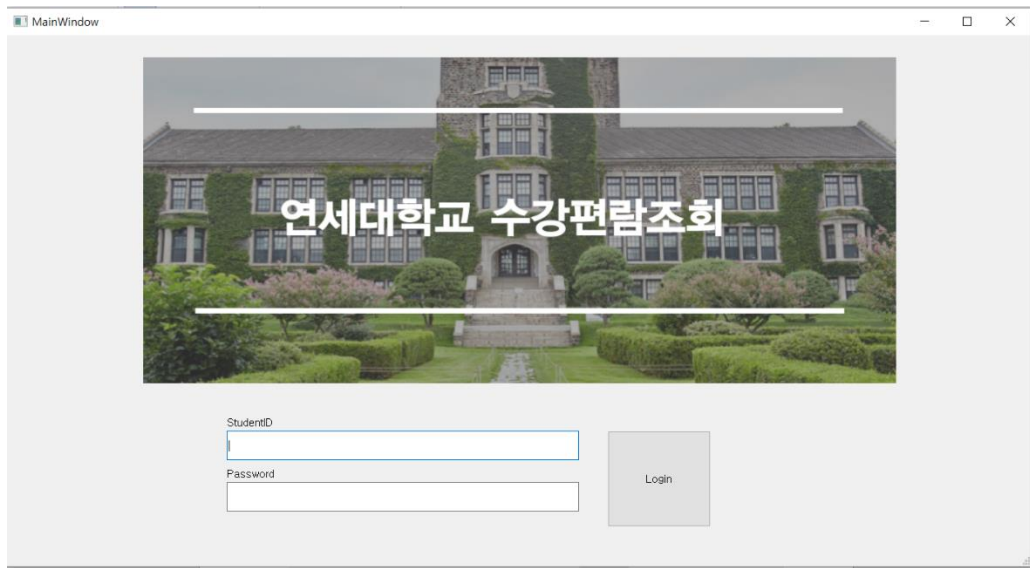
이에 우리는 본 프로젝트를 통해 위와 같은 불편 사항을 개선하고자 한다. 하지만 현실적인 구현 가능 여부에 있어서 몇가지 제약이 존재한다. 에브리타임 수업 평점 및 평가의 적용이 힘들다. 외부 DB에대한 법률적인 크롤링 제약이 존재해서이다. 두 번째로는 앞서 언급했듯이 연세대학교 포탈 건물 데이터의 표기방식의 한계로 인해 건물 데이터를 크롤링하는데 어려움이 있다. 마지막으로 학생들, 교수님들의 개인정보 접근이 불가능하다.

따라서 결과적으로 우리는 위와 같은 한계점을 배제하고 다음과 같은 기능과 과정들로 GUI를 구

현재 보았다.

1. 로그인 창

본 프로젝트에서 가장 먼저 거쳐야 되는 단계는 로그인 단계이다. 로그인을 하게 되면 메인 페이지인 수강편람 페이지에 접근하게 된다. 이렇게 설계한 이유는 학생들이 로그인을 하게 되면 이에 대응되는 정보들을 메인 페이지에 불러오기 위함이다. 가장 중요한 부분은 역시 졸업요건 현황표가 되겠다. 매 학생마다 들은 수업과 들어야 할 수업들이 다르므로 로그인 과정을 통해 학생들의 정보를 식별해 줘야한다. 또한 학생이 해당 시스템, 즉 해당 학교 학생이 아닌 경우 로그인이 안되는 코드도 추가하였다. 따라서 해당 학교 학생이면 로그인 버튼을 누르는 순간 메인 페이지로 넘어가게 되겠고 학생의 정보 또한 메인 페이지에 표시될 것이다.



2. 메인 페이지:

메인 페이지는 수강편람을 수행하는 핵심 페이지다. 옵션 및 검색창이 상단에 위치하고 있으며, 우측에는 졸업요건표가 배치되어 있다. 중앙에는 편람하고자 하는 검색결과를 나타내는 수업 정보(SUBJECTS)들이 표시된다. 오른쪽 상단에는 캠퍼스 지도와 더불어 로그아웃 창 버튼이 고정되어 있다. 메인 페이지에 이러한 구성을 한 이유는 보다 직관적이고 핵심 내용과 필요한 내용만을 담으려고 UX/UI적으로 노력했다. 수강 편람하는데 있어서 가장 중요한 요소만을 적재적소 배치해 두고 핵심요소에 비해 상대적으로 덜 중요하거나 한 화면에 담기 힘든 정보들은 버튼 형식으로 클릭하면 pop-up하는 형식으로 창을 만들어 언제든지 키고 닫을 수 있게 설계하였다.

The screenshot shows the 'MainWindow' application. At the top, there are dropdown menus for '2019', '1학기', '전체', and '전체', followed by a 'Search' button, a 'Campus Map' button, and a 'Logout' button. Below these is a 'Search Result' table with columns: '개설연도', '개설학기', '영역1', '영역2', '영역3', '학점번호', '과목명', and '학점'. The table is currently empty. To the right of the search result table is a '졸업요건' (Graduation Requirements) table with columns: '졸업요건', '만족 여부', and '부족 학점'. The '졸업요건' table contains 11 rows of requirements, each with a checkbox, a status 'X', and a credit value.

개설연도	개설학기	영역1	영역2	영역3	학점번호	과목명	학점
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							

졸업요건	만족 여부	부족 학점
1 필수교양5영역	X	15
2 논리와수리	X	12
3 자연과우주	X	6
4 생명과환경	X	6
5 전공기초	X	6
6 채플	X	2
7 전공선택	X	61
8 전체이수학점	X	140
9 글쓰기	O	0
10 대학영어	X	2
11 기독교의이해	O	0
12		
13		

3. 옵션창 및 검색창

This screenshot is similar to the previous one, but the '전체' dropdown menu is open, showing a list of options: '전체', '인간이해와역사', '과학과기술', '사회와가치', '문화과예술', '외국어와세계이해', and '건강과스포츠'. The '졸업요건' table is also visible on the right.

개설연도	개설학기	영역1	영역2	영역3	학점번호	과목명	학점
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							

졸업요건	만족 여부	부족 학점
1 필수교양5영역	X	15
2 논리와수리	X	12
3 자연과우주	X	6
4 생명과환경	X	6
5 전공기초	X	6
6 채플	X	2
7 전공선택	X	61
8 전체이수학점	X	140
9 글쓰기	O	0
10 대학영어	X	2
11 기독교의이해	O	0
12		
13		

옵션창과 검색창은 mysql의 query로 원하는 범위를 설정해주어 결과값을 불러온다. 대략적으로 각 옵션창들은 query상 WHERE 옵션1=(), 옵션2=() 인 형태로 설정하였으며 여기에 Search input값 = ()까지 일치하게 코드를 설계하였다. 해당 쿼리는 아래와 같다.

```

sql = 'SELECT OpenYear,OpenSem,G1,G2,G3,SubjectID,SubjectName,Credit,SubjectTime,FullCapi
SubjectName FROM SUBJECTS '
semester = str(self.comboBox_5.currentText())
gu2 = str(self.comboBox_2.currentText())
gu3 = str(self.comboBox_3.currentText())
sql += "WHERE OpenSem = '{0}' ".format(semester) #AND G2 = '{0}' AND G3 =
'{0}' ".format(semester,gu2,gu3)
if gu2 != '전체':
    sql += " AND G2 = '{0}' ".format(gu2)
    if gu3 != '전체':
        sql += " AND G3 = '{0}' ".format(gu3)
#print(sql)

if self.lineEdit.text() != "":
    sql = sql + " AND SubjectID LIKE '%{0}%' ".format(self.lineEdit.text()) + "UNION ALL " + sql + "
AND SubjectName LIKE '%{0}%' ".format(self.lineEdit.text())
sql += ';'

```

4. 결과 분석

4-1) 개선된 사항

개선된 사항에서는 기존 연세포탈 수강편람 시스템보다 더 효율적이고 나은 기능들을 소개하겠다. 본 프로젝트에서 개선된 사항으로는 총 5가지이며, 아래와 같다.

1. 편리한 Search Engine (옵션창(drop down)과 입력창의 연동):

기존 연세포탈 시스템에서는 Drop down의 제약조건과 입력창은 따로 개별적으로 작동한다. 예를 들어, ‘학부필수’ 영역을 선택하고 입력창에 ‘영화’를 입력하면 선택된 ‘학부필수’ 내 ‘영화’ 단어와 연관된 수업이 output으로 도출된 것이 아니라 기존 설정했던 제약 조건들이 다 날아간다. 따라서 우리는 입력창과 Drop down 박스를 연동시킴으로써 보다 효율적인 검색 창을 구축하였다.

The screenshot shows a web application interface with a search bar and a results table. The search criteria are set to '2019', '1학기', '학부필수(~2009)', and '전체'. The search term is '영화'. The results table shows 3 items, all with '학부' as the category and '학부필수(~2009)' as the subject. The right sidebar shows a list of courses with checkboxes for selection.

개설연도	개설학기	영역1	영역2	영역3	학점번호	과목명	학점	시간
1 2019	1학기	학부	학부필수(~2009)	문화의이해	UCE1105-01-00	영화의이해	3.0	화2,3,4
2 2019	1학기	학부	학부필수(~2009)	문화의이해	UCE1105-02-00	영화의이해	3.0	수7,금7
3 2019	1학기	학부	학부필수(~2009)	문화의이해	YCD1601-01-00	영화로보는현대...	3.0	화8,9,10

출입요건	만족 여부	부족 학점
1 필수교양5영역	X	15
2 논리와수리	X	12
3 자연과우주	X	6
4 생명과환경	X	6
5 전공기초	X	6
6 채플	X	2
7 전공선택	X	61
8 전체이수학점	X	140
9 글쓰기	O	0
10 대학영어	X	2
11 기독교의이해	O	0
12		
13		

위 그림을 보면, 옵션을 ‘학부필수’로 설정하고 입력창에 ‘영화’를 검색하면 정확히 해당 영역 ‘학부필수’ 안에 단어 ‘영화’와 관련된 과목들만 불러오기를 수행한다. ‘선택교양’이라는 영역안에 ‘영화로 배우는 중국어’라는 수업도 있다.

기존에 연세포탈 수강편람 시스템에서 검색할 경우, 모든 ‘영화’ 관련 과목들이 출력 된다. 하지만 본 GUI에서는 이러한 부분을 개선하였다. 이렇게 원하는 범위를 설정해준 방법으로는 앞서 설명했던 mysql query로 진행하였다.

2. 수강편람 시스템과 졸업요건 현황표의 연동:

MainWindow

2019 1학기 학부필수(~2009) 전체 Search Campus Map Logout

Search Result

	개설연도	개설학기	영역1	영역2	영역3	학점번호	과목명	학점	학점
1	2019	1학기	학부	학부필수(~2009)	문화의이해	UCE1102-01-00	미술사	3.0	수
2	2019	1학기	학부	학부필수(~2009)	문화의이해	UCE1105-01-00	영화의이해	3.0	화
3	2019	1학기	학부	학부필수(~2009)	문화의이해	UCE1105-02-00	영화의이해	3.0	수
4	2019	1학기	학부	학부필수(~2009)	문화의이해	UCE1107-01-00	음악사	3.0	화
5	2019	1학기	학부	학부필수(~2009)	문화의이해	UCE1108-01-00	디자인과문화	3.0	수
6	2019	1학기	학부	학부필수(~2009)	문화의이해	YCD1001-01-00	오래된미래, 르...	3.0	수
7	2019	1학기	학부	학부필수(~2009)	문화의이해	YCD1101-01-00	문학환경과결사...	3.0	화
8	2019	1학기	학부	학부필수(~2009)	문화의이해	YCD1201-01-00	한국문학과문화...	3.0	화
9	2019	1학기	학부	학부필수(~2009)	문화의이해	YCD1601-01-00	영화로보는현대...	3.0	화
10	2019	1학기	학부	학부필수(~2009)	문화의이해	YCD1602-01-00	대중음악의이해	3.0	수
11	2019	1학기	학부	학부필수(~2009)	문화의이해	YCE1255-01-00	유럽도시문화공...	3.0	Nc
12	2019	1학기	학부	학부필수(~2009)	문화의이해	UCD1101-01-00	현대사회와사회...	3.0	수
13	2019	1학기	학부	학부필수(~2009)	문화의이해	UCD1102-01-00	현대사회와심리...	3.0	화

졸업요건

졸업요건	만족 여부	부족 학점
1 필수교양5영역	X	15
2 논리와수리	X	12
3 자연과우주	X	6
4 생명과학경	X	6
5 전공기초	X	6
6 재충	X	2
7 전공선택	X	61
8 전체이수학점	X	140
9 글쓰기	O	0
10 대학영어	X	2
11 기독교의이해	O	0
12		
13		

빨간색으로 표시된 영역은 학생의 졸업요건 현황표이다. 기존에는 수강편람 시스템과 졸업요건표가 별개 데이터로 존재하여 졸업요건을 확인하는데 있어서 매번 엑셀 파일을 여는 행위를 반복해야 했지만, 이와 같이 기능을 추가할 경우 수강편람과 동시에 옆에 있는 졸업요건표를 참조하여 본인이 이수해야 할 항목, 이미 이수 완료한 항목들을 졸업요건과 비교하여 수강편람을 하게 해준다. 위와 같이 졸업요건, 해당 영역 만족 여부, 부족학점을 표시하게 되고 해당 영역에서 이수한 과목들을 보기 위해서는 해당 row를 클릭하면 popup창 나타나면서 정보를 제공해준다. 예로 졸업요건에서 기독교와 이해를 클릭하면 아래 그림이 표시된다. 3학점짜리인 기독교와 세계문화 수업을 이수하였고 학점은 A라는 정보가 기입되어 있다.

MainWindow

기독교의이해 수강 과목

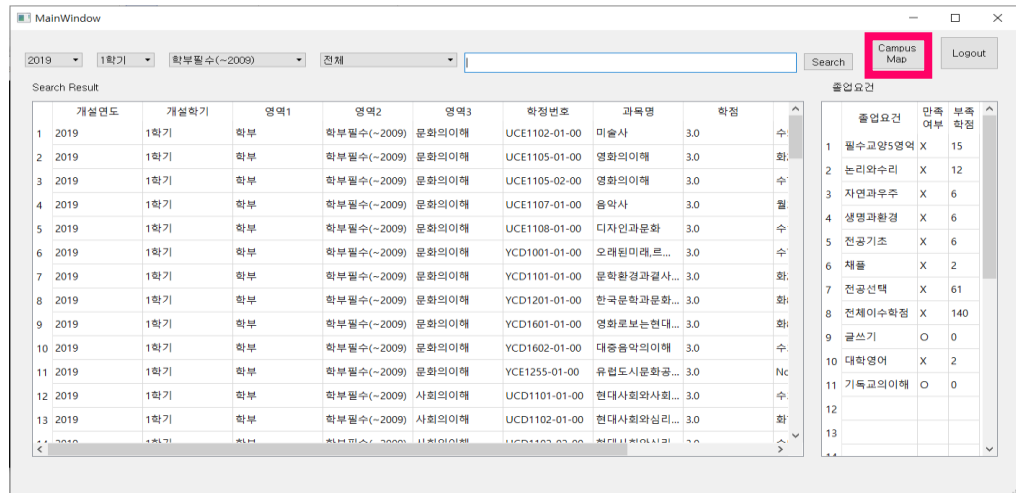
	학점번호	과목명	학점	성적
1	YCA1101-12-00	기독교와세계문...	3.0	A
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				

밑 그림은 졸업요건의 두번째 예로, 혹여 학생이 F를 맞았을 경우 졸업요건표 학점 이수란에 우리는 F를 맞은 수업의 학점이 합산된 결과로 나타나지 않은 것을 확인할 수 있다. 이것은 졸업요건을 알맞게 계산하기 위해 설계한 것이다. 졸업요건표는 종합적으로 졸업요건, 이수과목, 이수해야할 영역/과목 등 매우 유용한 정보를 제공한다.

	학점번호	과목명	학점	성적
1	YCC1001-16-00	대학기본영어 I	2.0	A
2	YCC1001-17-00	대학기본영어 I	2.0	F
3				
4				
5				
6				
7				
8				
9				
10				
11				

3. 수강편람 시스템과 캠퍼스 지도의 연동:

빨간색으로 표시된 영역은 학생의 캠퍼스 지도 버튼이다. 기존 시스템에서는 수강편람 시스템과 캠퍼스지도는 별개의 데이터로 존재하여 건물명에 대응되는 위치를 확인하는데 있어서 매번 연세대학교 홈페이지를 참고해야 했지만, 이와 같이 기능 추가로 수강편람과 동시에 건물 위치를 확인할 수 있다. 하지만, 기존 연세 수강편람 시스템 상의 ‘건물명’의 표기방식의 불일치로 인해 우리는 본 프로젝트에서 지도-건물명 연동 시각화 서비스는 제공하지 못하였다.



4. Dynamic Tab Value 기능:

기존에 명칭이 제대로 명시 되어있지도 않은 버튼을 선택하여 과목별 기본정보 및 수업계획서, 마일리지 정보를 확인할 수 있었지만, Dynamic Tab Value 기능은 과목에 대해서 직관적이고 편리하게 정보(기본정보, 수업계획서, 마일리지 정보) 관람을 가능하게 한다.

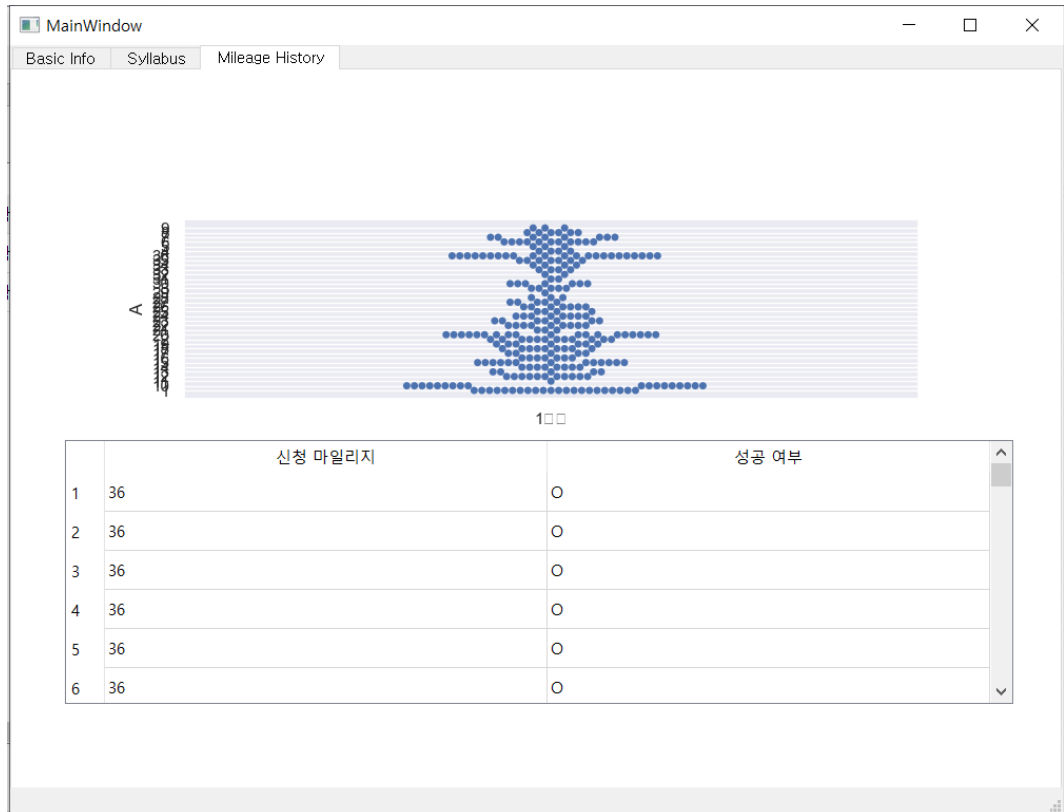


밑 그림은 Syllabus 정보를 보여준다. 수강 편람에 있어서 가장 기본이 되는 정보이며 해당 과목이 어떠한 과정으로 이루어져 있는지, 교재와 범위는 어떻게 되는지 주차별로 소개하고 있다. 중간고사와 기말고사 기간까지 명시되어 있어 편람 하는데 매우 중요한 정보라고 볼 수 있겠다.

MainWindow					
Basic Info Syllabus Mileage History					
Syllabus					
주	기간	수업내용	교재범위,과제물	비고	
1	1	2019-03-04 2019-03-10	* 강의 소개 및 수업 관련 정보 설명 제1강 장르, 분류, 스타일 - 장르영화란? - 타입과 스타일 제2강 영화의 용어 및 정의 - 영화를 가리키는 다양한 용어	* 특강과 영화 상영은 수업 상황에 따라 변동되며, 그에 따른 강의 내용은 순차적으로 밀려서 진행됩니다. * 영화 감상(여건에 따라 변동될 수 있습니다) * 루이스 자네티 <영화의 이해>	(3.4.) 개강 (3.7. - 3.11.) 수강신청 확인 및 변경
2	2	2019-03-11 2019-03-17	제3강 종합예술로서의 영화 - 제7예술설 카누도 제4강 영화 속의 거리 개념 - 클로즈업 - 롱 쇼트 - 낮설게 하기 - 패스빈더 - 뉴 저먼 시네마(사조)	* 임정택 <세계영화사 강의> * 박현섭 역 <영화기호학> * 토마스 사츠 <할리우드 장르의 구조> 참조 * 이형식 역 <영화에 대해 생각하기> 명인문화사.	(3.7. - 3.11.) 수강신청 확인 및 변경
3	3	2019-03-18 2019-03-24	제5강 영화 장면의 단위 및 중요 개념 - 미리 알면 이해하기 쉬운 용어들 - 프레임, 쇼트, 테이크, 컷, - 씬, 시퀀스 - 미장센, 몽타주 - 롱테이크, 딥 포커스 - 작가 제6강 영화 사조(참고자료용) 영화의 역사에서 가관하게 하소	*구회영 <영화에 대하여 알고 싶은 두세 가지 것들> 참조	

5. Visualizing Mileage Information:

기존에 알아보기 힘든 테이블로 구성된 과거 마일리지 배팅 정보를 마일리지 시각화 기능을 추가함으로써 그래프와 테이블로 간편화 된 마일리지 정보를 확인 가능하게 한다.



4-2) 한계점

본 프로젝트의 한계점은 아래와 같다.

1. 에브리타임 수업 평점 및 평가 적용 한계:

외부 데이터베이스를 크롤링하는 것은 법적으로 제한된다. 우리가 크롤링하려는 집단 ‘에브리타임’에서도 크롤링 거부의 입장을 밝혀 부득이 하게 수업평점과 평가의 적용은 무산되었다.

2. 건물-지도 서비스 구축 한계:

연세대학교 포탈서비스 건물 데이터 표기방식은 하나로 통일된 워딩을 쓰는 것이 아니라 다양하게 표시되고 있다. 따라서 건물정보를 크롤링 하는데 있어서 어려움이 많다. 더 나아가서 제한적인 지도 크기 안에서 직관적이게 건물 정보를 동적으로 표현해주는 UX/UI적인 제약 또한 존재한다.

3. 개인정보 (학생, 학번 정보) 접근 한계:

개인정보는 매우 민감한 사안이며 기본적으로 접근이 제한되어 있다. 따라서 우리는 이러한 한계로 가상의 학생 데이터베이스를 구축하였으며, 마일리지 정보도 이로 인해 소량만이 결과값으로 출력되는 한계점을 보여준다.