**Jsoup을 이용한 멜론 순위 차트 크롤링 설명서**

멜론 사이트의 현재 순위를 보여주는 노래의 순위, 타이틀, 가수, 앨범 출력

대상 URL : https://www.melon.com/chart/index.htm

크롤링 시 필요한 사전 설정

1. <https://jsoup.org/downloan>에서 jsoup-1.14.3.jar 파일 다운로드

2. 작업하는 프로젝트 buildpath 에서 classpath에 external JarRs로 다운로드 받은 jsoup-1.14.3.jar추가하기

1. 멜론차트 정보 크롤링 ( 순위, 곡 정보, 가수, 앨범)

1-1 크롤링 대상 URL 가져오기

텍스트이(가) 표시된 사진

자동 생성된 설명

2. 순위 차트에 있는 음악 정보가 모두 lst50클래스의 tr에 포함되어 있음.

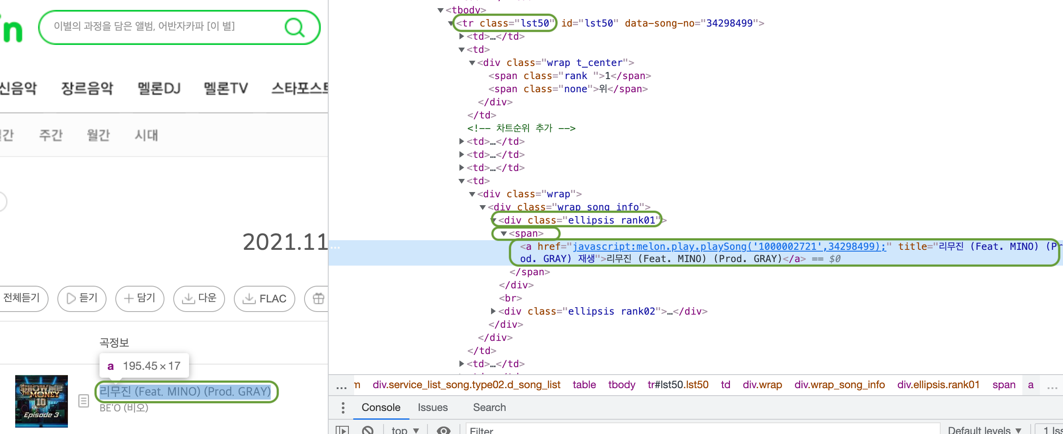
2-1 대상 순위 추출

tr.lst50 > div.t\_center.wrap > span.rank



2-2 대상 곡 정보 추출

tr.lst50 > div.rank01 > span > a



2-3 대상 가수 정보 추출

tr.lst50 > div.rank02 > span > a



2-4 대상 앨범 정보 추출

tr.lst50 > div.rank03 > a

텍스트이(가) 표시된 사진

자동 생성된 설명

[ 크롤링 데이터 출력 결과 ]

텍스트이(가) 표시된 사진

자동 생성된 설명

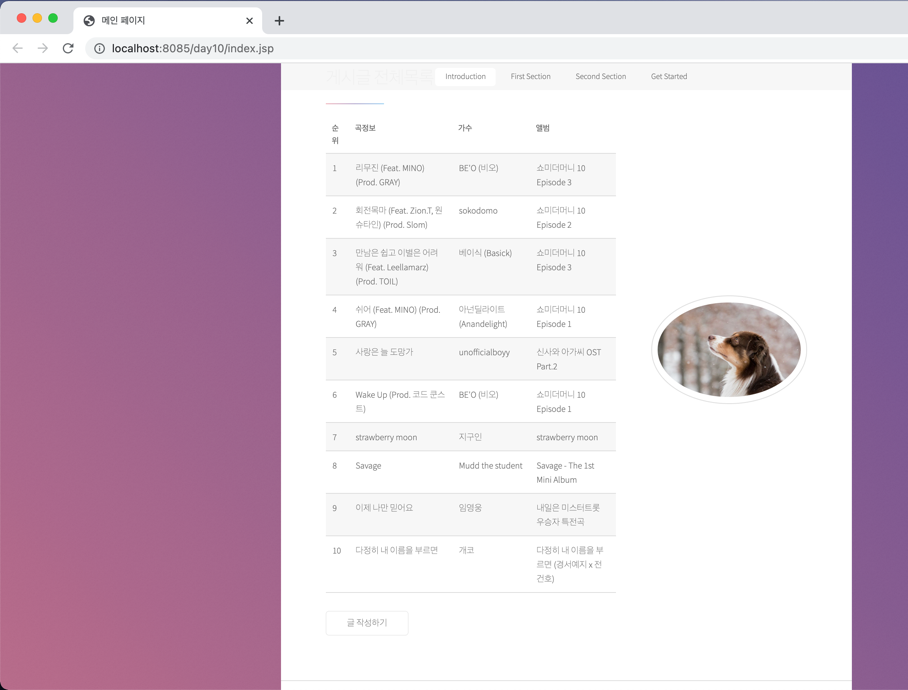
[ 코드 문제 발생 시 해결 방법]

문제발생 1) 크롤링 대상 url 가져온 후 선택자로 태그 선택했는데 결과가 나오지 않음.

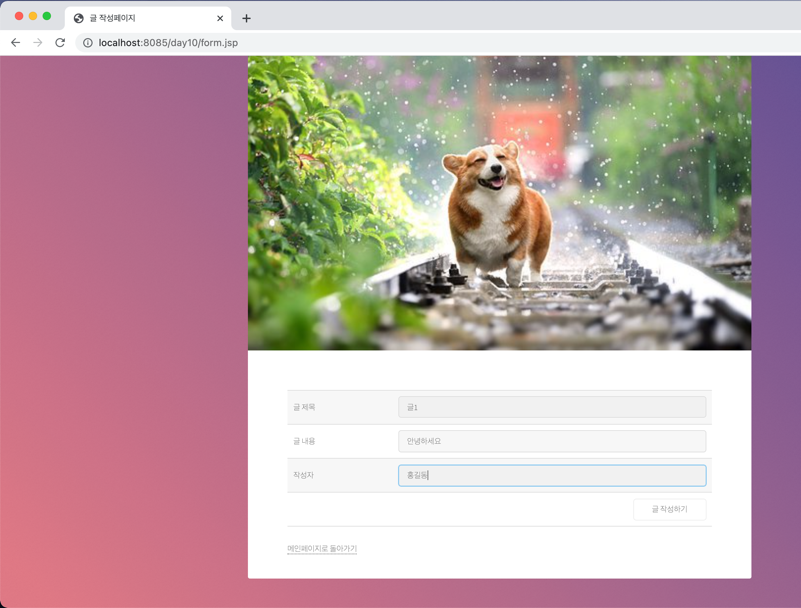
해결방법 : chrome 개발자 도구에서 본 코드랑 추출해서 가져온 코드가 정확하게 맞지 않음 -> 해당 사이트에서 크롤링 하지 못하게 막아둔 것을 알게 됨.

[ 크롤링 한 내용을 게시판에 적용한 예 ]

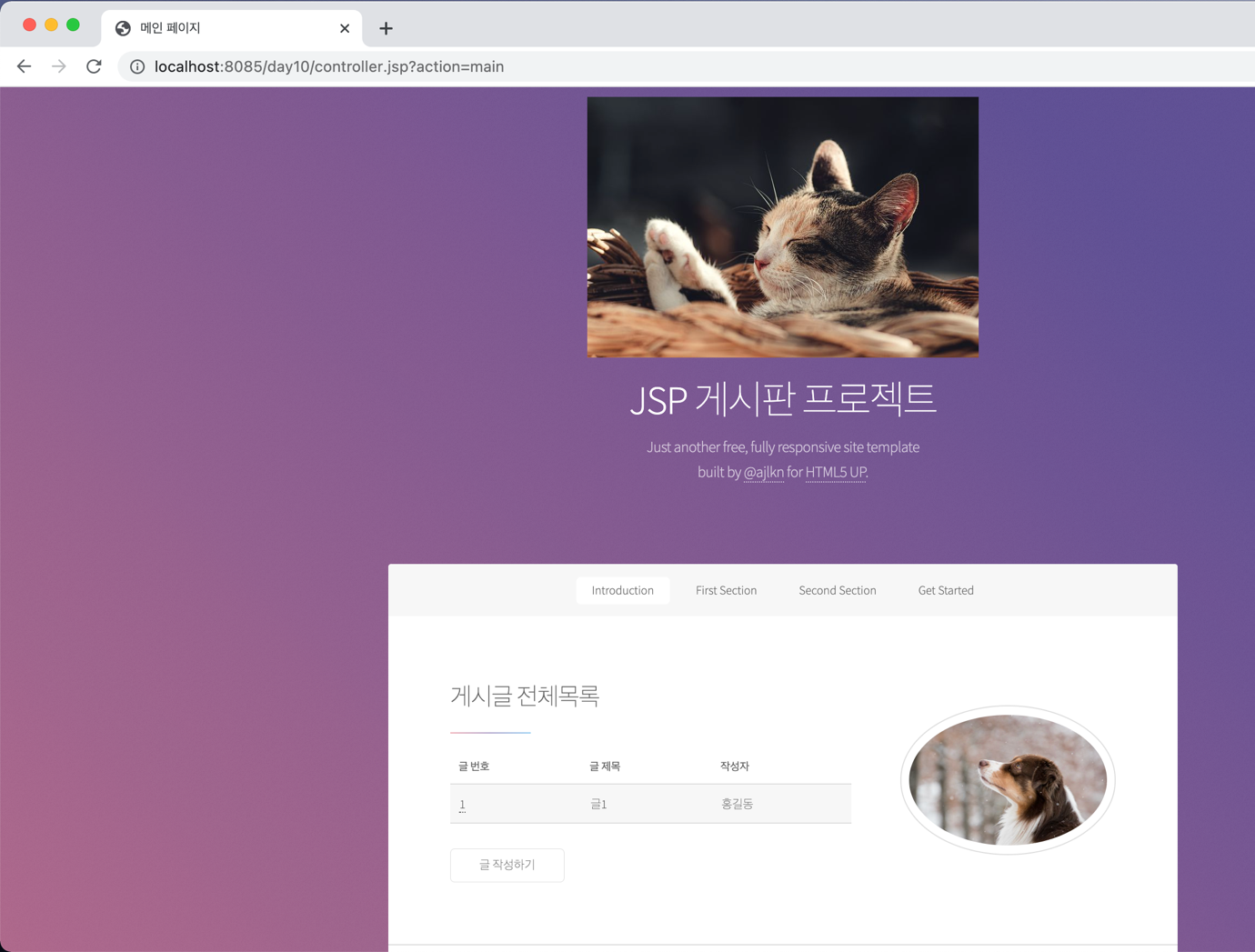
1. 게시판에 내용이 없을 경우 크롤링한 내용 출력



2. 글 작성 페이지



3. 게시판에 글이 추가되면 크롤링한 내용은 더 이상 보이지 않고 글이 추가된 게시판 출력



[ 크롤링 시 데이터의 가공 ]

텍스트이(가) 표시된 사진

자동 생성된 설명

크롤링 한 음악 순위는 50위까지이지만 데이터 양 조절을 위해 페이지에 출력하고 싶은 순위10위까지만 데이터베이스 테이블에 저장함. (cnt 변수 사용)

[ 아직 미해결 부분 ]

1. 페이지를 실행 할 때마다 음악 순위 1-10위까지 매번 데이터베이스에 들어감.

2. 페이지 정리 필요