

# Ailegis, Coding Assignment

Thank you for your interest in joining us!!

**Goal:**  
The purpose of this assignment is to assess your scraping capabilities.  
Due to a lack of creativity, the goal will be to scrape a section of the ETHZ website.

**Task:**  
The website is: <https://ethz.ch/en/doctorate.html>  
Scan all possible subpages of the doctorate path.  
Create a single .json file to store all of the results.  
The .json format should include an array of all the subsections of the website.  
Each subsection should have:

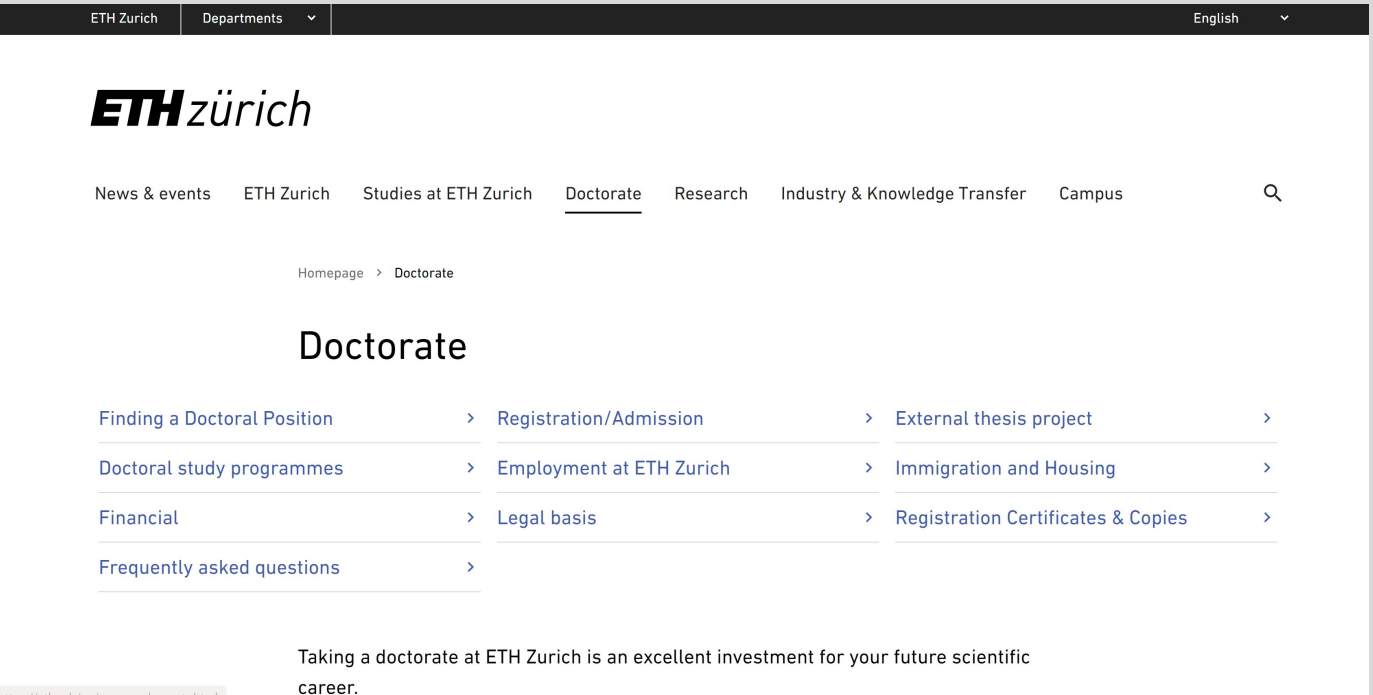
- Title ( a string )
- URL ( a string ) (if possible the url should map to the specific subsection with #identifiers)
- Content ( a string )
- An array with all the attached files (file name and URL , both strings )

Note that some sections of the web-page are not relevant, such sections should not be in the content string.

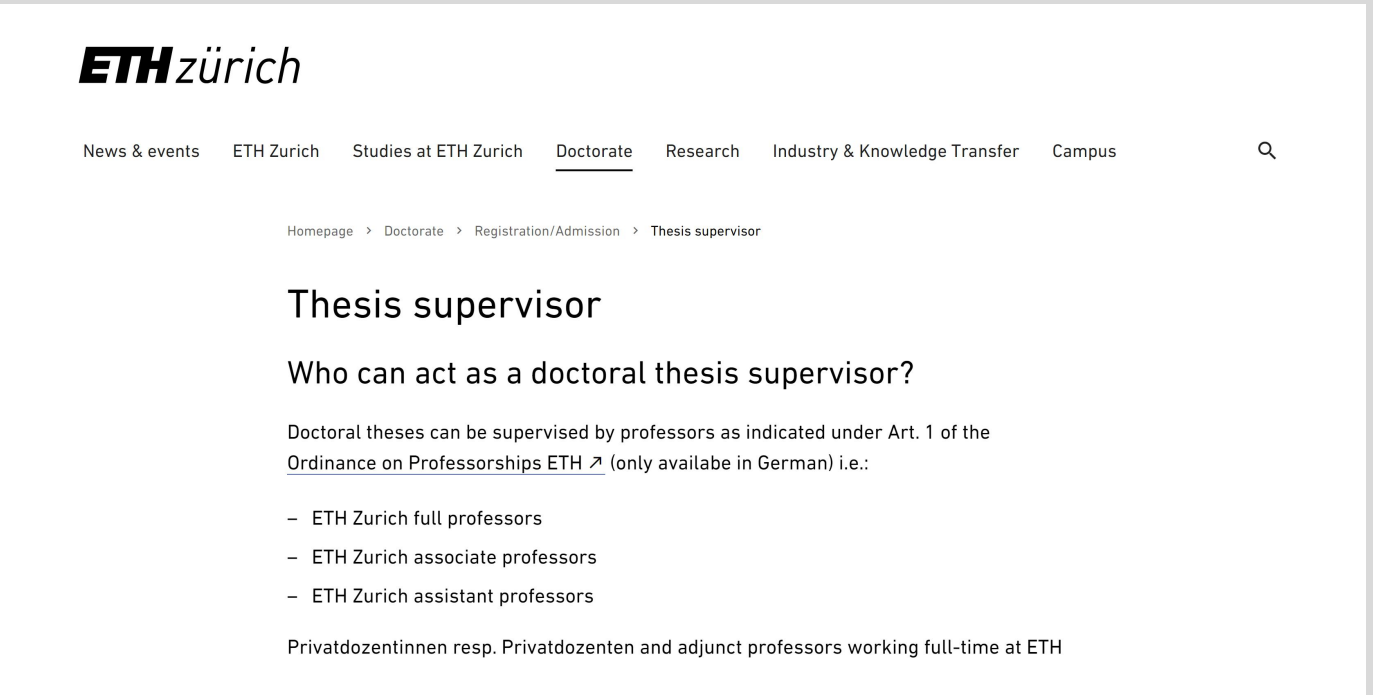
**Bonus:**  
Scrape for all available languages (have a .json file per language).  
Scrape the content of the files attached.  
Any other feature you think could be useful, or show off your skills.

**Constraints:**  
Use Python or TypeScript.

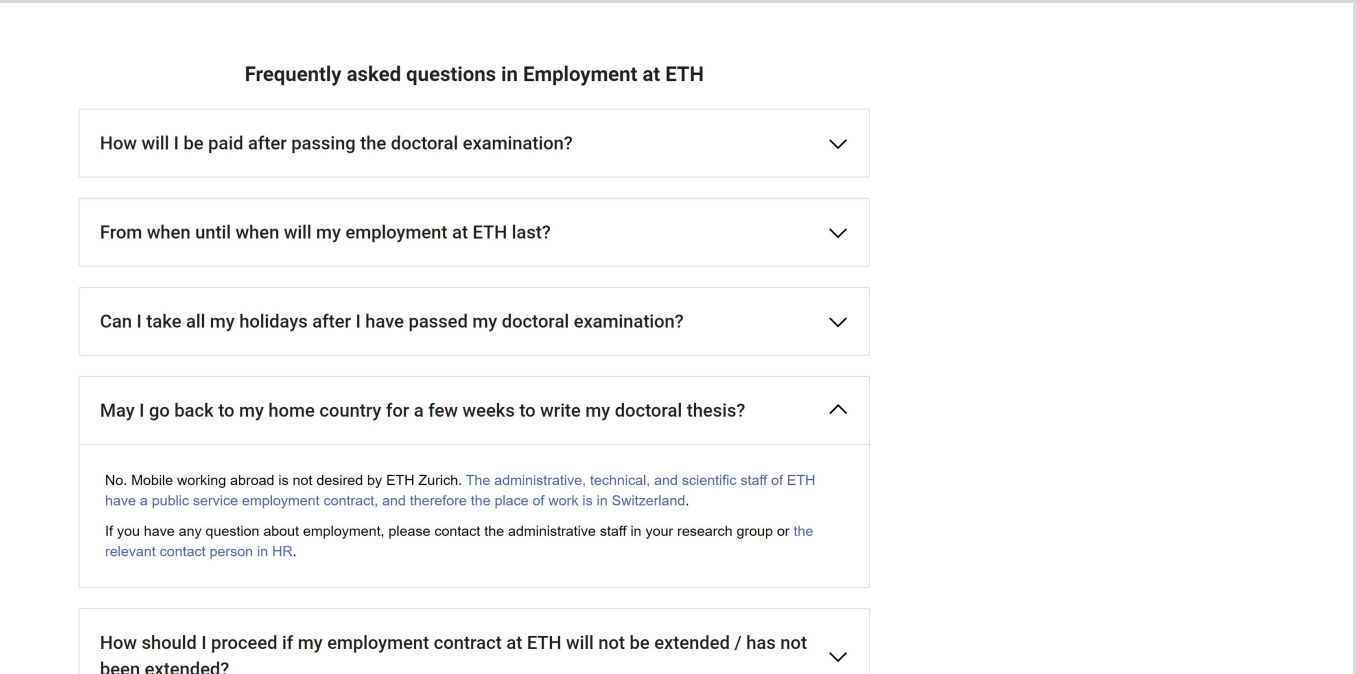
**Delivery:**  
Create a GitHub repository with all the code used under src/ and the scraped data under results/.  
Write a maximum of 2 sentences on what you would have done if you had more time.



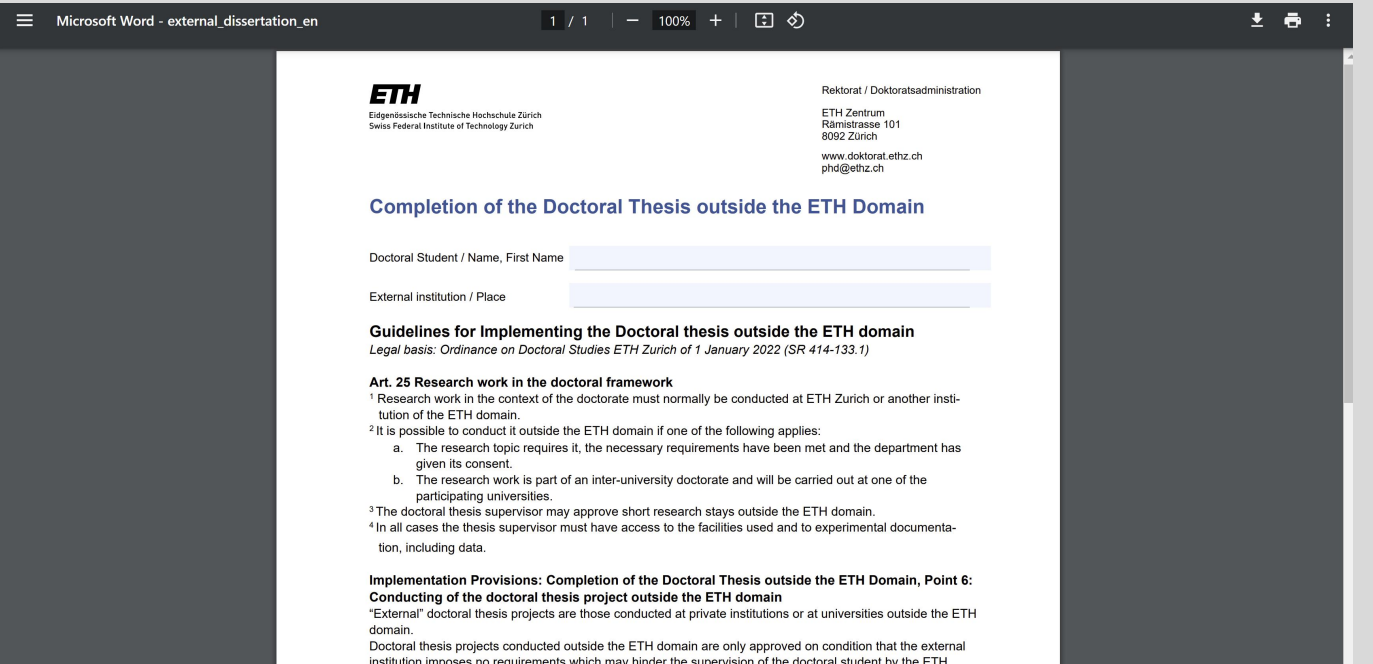
Main page



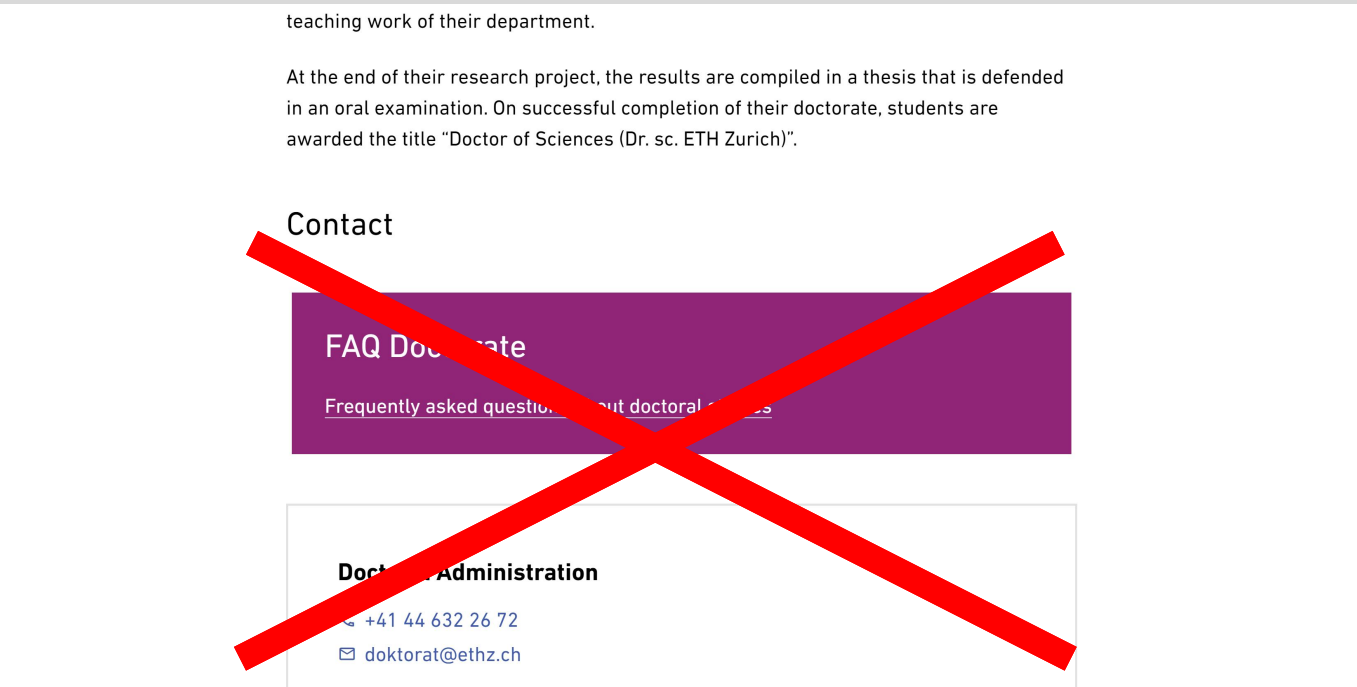
Example of Subpage



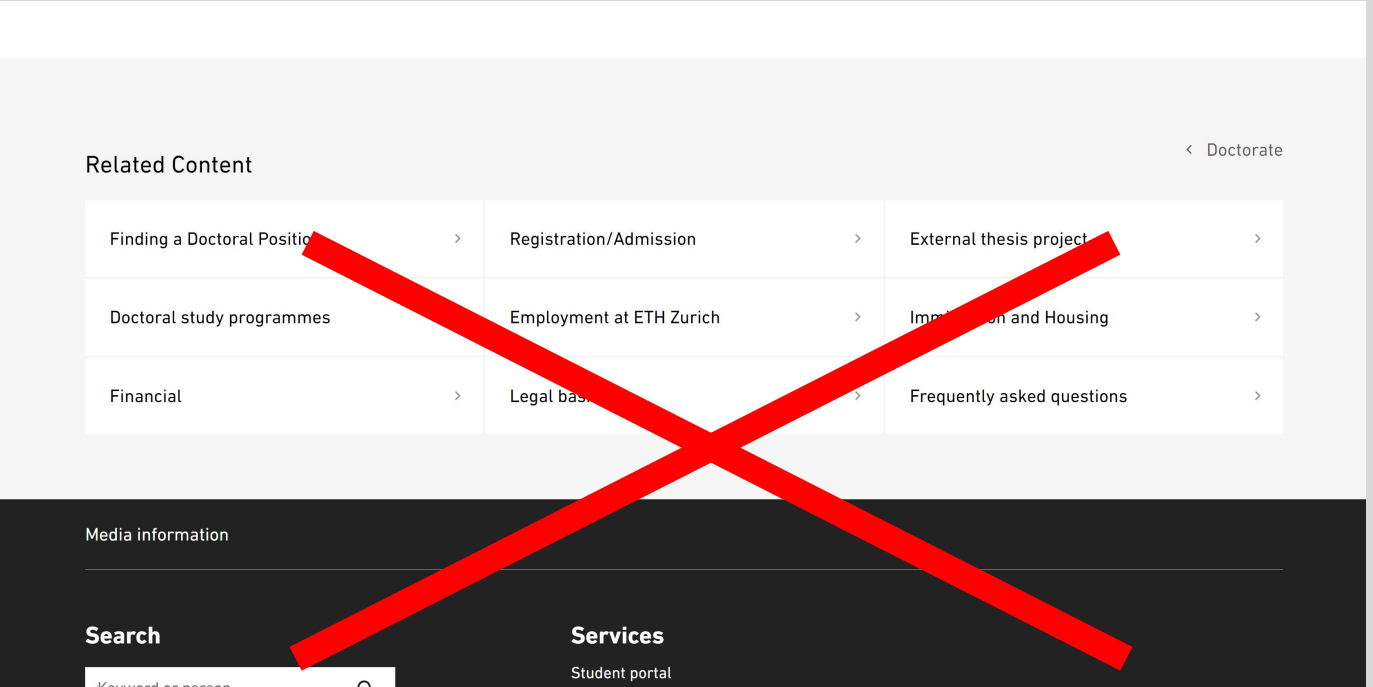
Example of Subpage



Example of pdf attached



Example of sections that do not need to be scraped



Example of sections that do not need to be scraped