

Project Proposal
Jinyi Gao
Uniquename: gaojinyi

Introduction

My project is about crawling and scraping data from multiple pages of career part of Microsoft, storing the data into a tree and presenting the data by Flask.

Data Source

Data will be gained by crawling and scraping all kinds of job information on the career part of Microsoft official website.

Data Access and Storage

Raw data (JSON or HTML) will be cached into a JSON or HTML files and the processing part will directly load data from the caching part instead of refetching from the website.

Data Processing

Tree will be used to fit all the data. The original root of the trees will be all jobs and different kinds of jobs will appear as children of the original root. In addition, level next to the different kinds of jobs will be the detailed information of the jobs (basic description, place, requirement, etc.).

Data Presentation

Flask will be used to present all data and do interaction with users.