Web Scrape Design

using Requests and Beautiful Soup

scraping webpages

inputs

keywords

list of source URLs

outputs

**article title**

date

source (publisher, author)

keyword tags

**URL**

location

steps

**access webpage** (via source URL)

if there’s good content

parse content for keywords

access webpage

if there’s good content

parse content for keywords to confirm it’s relevant

save title, URL

access webpage (source URL)

import requests module

make a request

check that request is good

if good

return content

else

log error