

## &lt; Back Easy Python with find and stack



▲

bw1226

★ 95

Last Edit: January 9, 2020 12:39 AM

586 VIEWS

6

- get the hostname
- use a stack to process htmlParser.getUrls, starting with the startUrl
  - if hostname is contained (find), and if not already in result set (for O(1) lookup): add to result, append to stack

```
def crawl(self, startUrl: str, htmlParser: 'HtmlParser') -> List[str]:
    hostname, grab = '', False
    # get hostname
    for index, char in enumerate(startUrl):
        if grab:
            if char == '/':
                break
            hostname += char
        if char == '/' and startUrl[index-1] == '/':
            grab = True
    hostname = 'http://' + hostname
    # process stack
    result, stack = set([startUrl]), [startUrl]
    while stack:
        q = stack.pop()
        urls = htmlParser.getUrls(q)
        for url in urls:
            if url.find(hostname) != -1:
                if url not in result:
                    result.add(url)
                    stack.append(url)
    return result
```

If you like this solution please consider giving it a star on my github. Means a lot to me.