xpath练习:抓取猪八戒数据

不墨迹, 直接怼.

干猪八戒.

```
import requests
from lxml import etree

url = "https://beijing.zbj.com/search/f/?type=new&kw=saas"

headers = {
    "User-Agent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_4) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/87.0.4280.88 Safari/537.36"
}
resp = requests.get(url, headers=headers)

# 丢给etree, 生成Element对象
tree = etree.HTML(resp.text)
# 拿数据吧
```

```
els = tree.xpath("//div[@class='witkey-list-grid j-service-provider-wrap ']/*/div[@class='witkey-item grid-box']")

for div in els:
    name = div.xpath("./div/div[@class='grid-top-right']/section/h4/a/text()")[0]
    intro = div.xpath("./div/div[@class='grid-top-right']/section/h4/div/div[last()]/text()")[0]
    shopdesc = "".join(div.xpath("./div/div[@class='grid-top-right']/section/h4/div/div[@class='witkey-shopdesc']//text()")).replace("\n", "")

    firstline = div.xpath("./div/div[@class='grid-top-right']/section/h4/div/div[@class='witkey-shopdesc-firstline']//text()")
    if firstline:
        shopdesc += "".join(firstline).replace("\n", "")
    fav = "".join(div.xpath("./div/div[@class='grid-top-right']/section/h4/div/div[@class='expert-tree']//text()")).replace("\n", "")
```

```
25
26
27 # 后面你想干嘛就干嘛吧. 记录数据库,写入文件,扔给全文检索工具 都可以
28
```