**世界猫**

输入文件:csv(utf-8) 输出文件:csv(utf-8)

Step1.先打开网页: https://www.worldcat.org/，如图1。根据图示1单击”接受所有Cookie”按钮(记得写需求之前需要先清除浏览记录,否则类似按钮不会出现)。

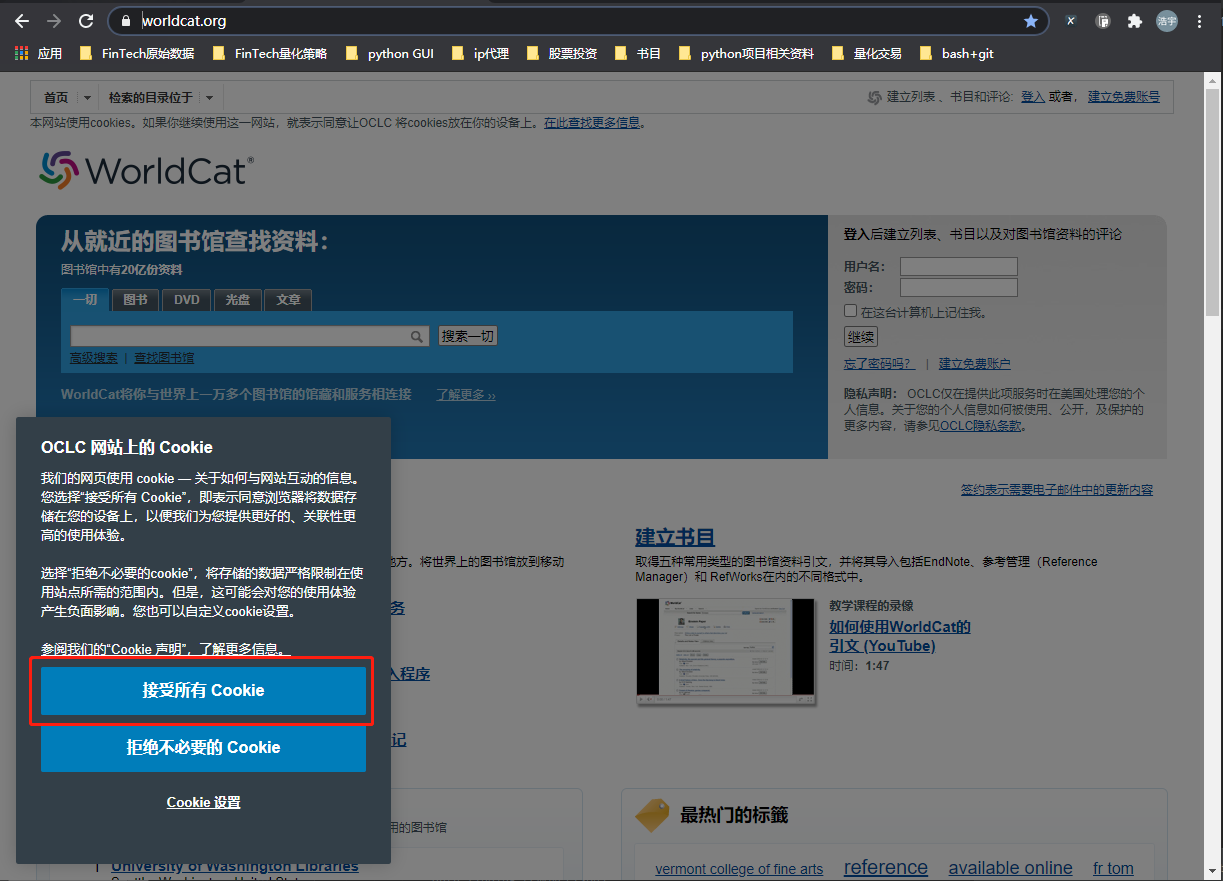


图1

Step2.在搜索框中输入isbn编号，如图2。



图2

Step3.如果出现图3的情况，则说明该待搜索信息没有任何结果。那么在该编号前添加一个0,返回Step2重新输入带搜索信息再次搜索,直到重复三次失败则放弃该编号。



图3

Step4.如果出现图4(例: 5709644056361)的情况，点击第一个图片或者右侧粗黑字体以进入详情页。

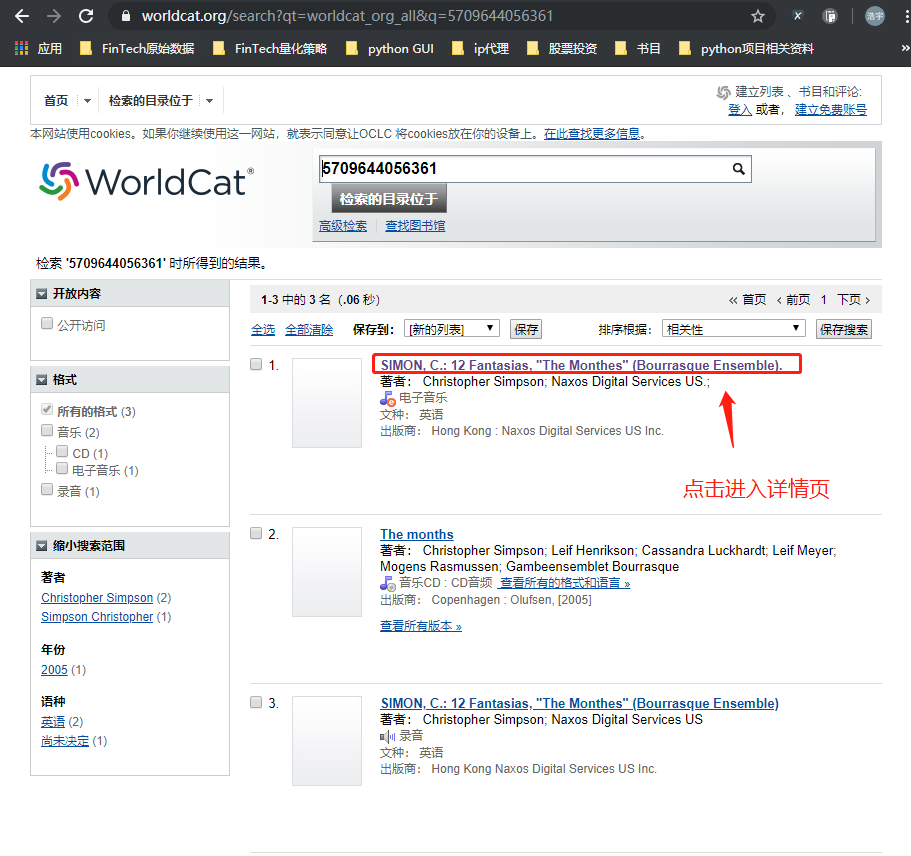


图4(例: 5709644056361)

Step5.依次抓取字段1.

* 如图5,题目中”:”前内容录入245a,之后内容录入245b.如无冒号全录入245a.
* 如图5,著者信息放入245c.
* 如图5,出版商信息把冒号前内容放入260a,冒号到其之后第一个数字/逗号之前放入260b,四位数字年份放入260c.
* 如图5,丛书信息中”. “换成” ; “放入490a.
* 如图5,主题内容中间用” ; “隔开,放入650a.



图5

Step6.依次抓取字段2.

* 如图6,下方内容中的演员信息放入511a字段.
* 如图6,下方内容中的描述信息按照冒号与分号切开,分别对应放入300a,300b,300c字段.
* 如图6,把内容部分字样衔接,放入505a字段.



图6

最终结果示意:

024a: 5709644056361

028a: ClassCd 663

245a: SIMON, C.

245b: 12 Fantasias, "The Monthes" (Bourrasque Ensemble).

245c: Christopher Simpson

100a: Simpson, Christopher

260a: Hong Kong

260b: Naxos Digital Services US Inc.

260c:

650a: Classical Music.

490a: Naxos Music Library.

511a: Bourrasque Ensemble, Ensemble -- Simpson, Christopher, Composer.

300a: 1 online resource.

505a: No. 1. January (04 min. 32 sec.) / Simpson -- No. 2. February (04 min. 01 sec.) / Simpson -- No. 3. March (03 min. 40 sec.) / Simpson -- No. 4. April (03 min. 38 sec.) / Simpson -- No. 5. May (04 min. 04 sec.) / Simpson -- No. 6. June (03 min. 59 sec.) / Simpson -- No. 7. July (03 min. 19 sec.) / Simpson -- No. 8. August (03 min. 52 sec.) / Simpson -- No. 9. September (03 min. 31 sec.) / Simpson -- No. 10. October (03 min. 29 sec.) / Simpson -- No. 11. November (03 min. 28 sec.) / Simpson -- No. 12. December (03 min. 39 sec.) / Simpson.