

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)。

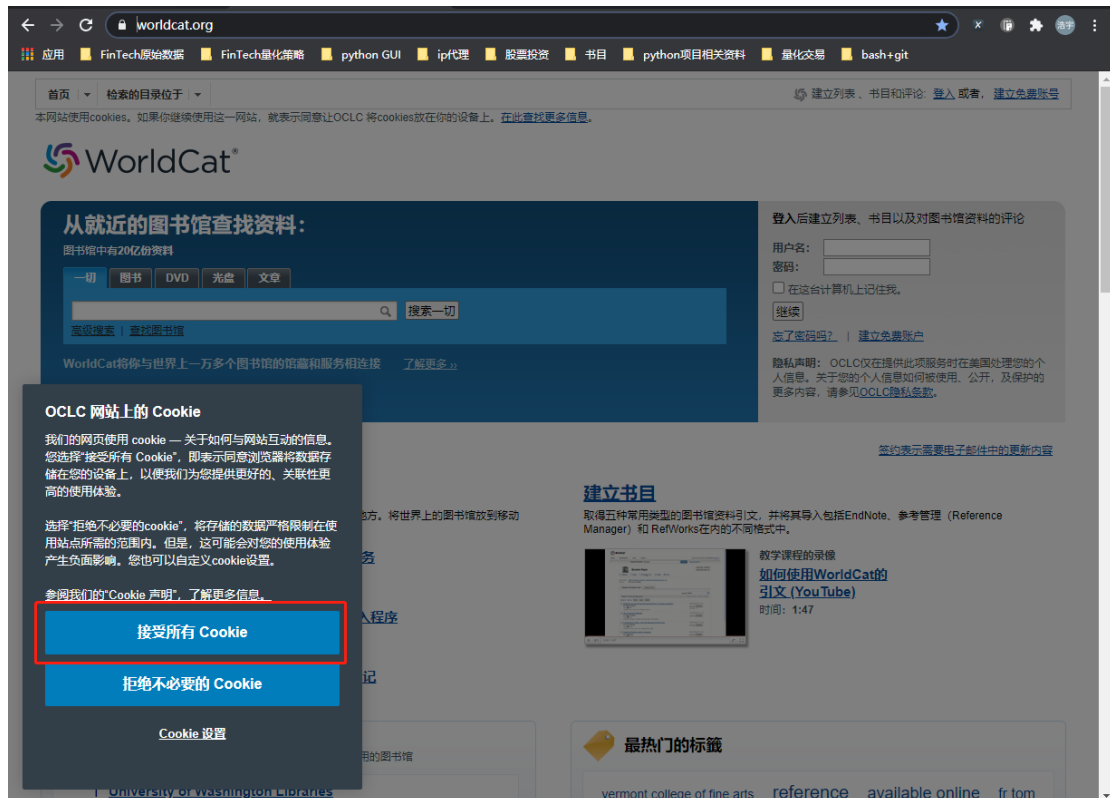


图 1

Step2.在搜索框中输入 isbn 编号, 如图 2。



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。



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。

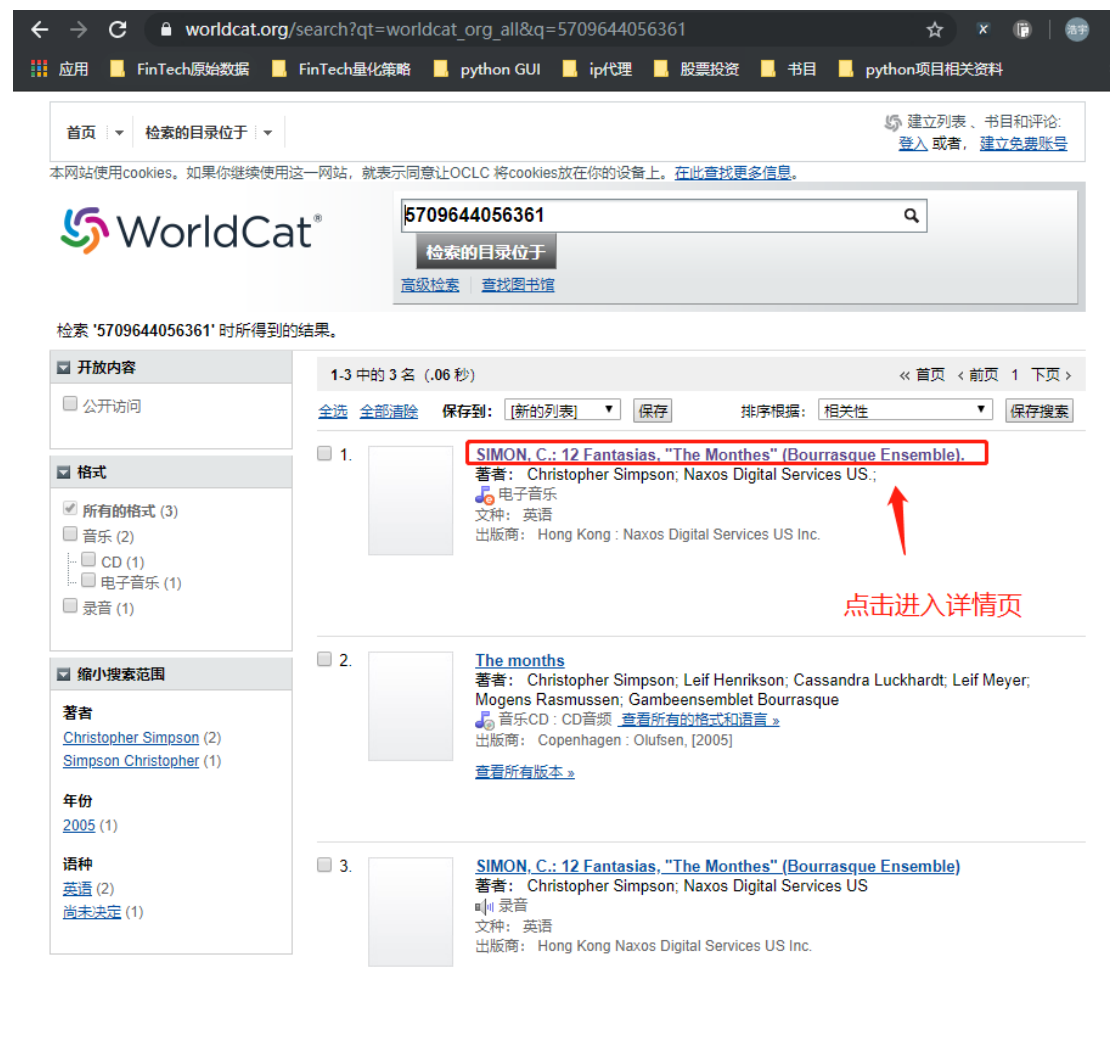


图 4(例: 5709644056361)

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.

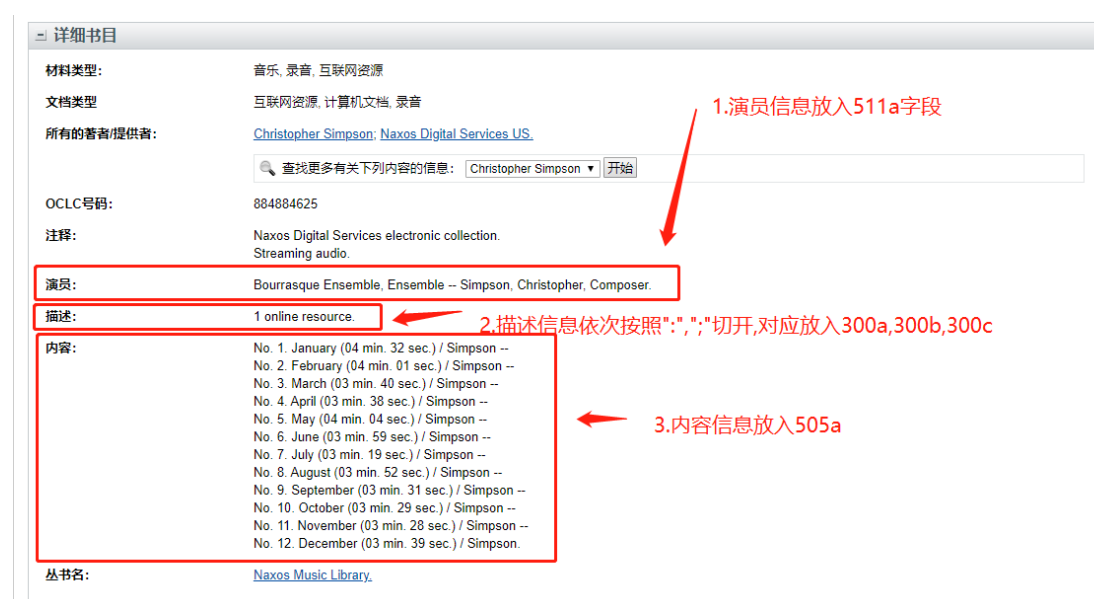


图 6

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.