

第四课. 提取一个网页内多个元素的数据

请先看下面这一张网页截图。

颜色	插电3档调光+USB (银色)	31.90元	18456 个可售	-	0	+
	插电5档调光5档调光定时+USB线 (银色)	39.90元	49072 个可售	-	0	+
	充电5档调光5档调色定时+USB线 (银色)	49.90元	45195 个可售	-	0	+

这是本教程多次演示过的某商品详情页中的一部分，包括了商品的多个 SKU 的标题、价格和库存数据。

本课将进一步介绍 CSS 选择器的功能，用两个 CSS 定位字符串提取这个网页上所有 SKU 的标题、价格、库存组成的数据。

请读者跟着步骤试着搭建好这个流程，然后尝试用不同的操作方式反复运行，体验 CSS 选择器的这个功能，进一步掌握 CSS 定位字符串的作用。

类似的页面结构常见于搜索结果、商品目录、订单列表、用户评论等网页内容。读者可以在课后自行尝试提取类似页面内多个元素的数据，为后面学习批量采集数据打下基础。

搭建步骤详解

本课内容讲解如何用机器人流程自动化采集单个网页上多个指定元素的数据。

虽然用于演示的例子是 1688 网站的商品详情页的 SKU 数据，但读者学会后可自行开发流程采集任何网页上多个指定元素的数据。

此处以使用 1688 平台为例，选择一个商品详情页，运行机器人流程，无人工干预的提取各项 SKU 数据，所产生的最终结果类似于一张表格。（注：步骤中的图片只用来介绍方法中的主要步骤，并不连续。）

1. 第一步，运行电商记桌面版客户端的浏览器，或者安装了电商记插件的 chrome、360、QQ 等浏览器，打开“机器人流程自动化”页面 (<https://m.dianshangji.com/rpa.html>)
2. 第二步，在浏览器的地址栏输入一个商品链接 (<https://detail.1688.com/offer/587693664150.html>)。后面文中所述的“宝贝 ID”就是“587693664150”。可以看到该页面上如图所示部分，即 SKU 数据。SKU 数据包含多个数据项，每个数据项由 SKU 标题、价格和库存构成，在网页上对应一个单元，包含三个文本元素。

颜色	插电3档调光+USB (银色)	31.90元	18456 个可售	-	0	+
	插电5档调光5档调光定时+USB线 (银色)	39.90元	49072 个可售	-	0	+
	充电5档调光5档调色定时+USB线 (银色)	49.90元	45195 个可售	-	0	+

- 第三步，在“机器人流程自动化”页面中搭建第一个**浏览类**积木块，用于采集 1688 详情页。为了代表这个页面，我们会定义一个**变量**。出于演示的目的，图中将变量命名为“商品详情页”。



- 第四步，添加**浏览类**积木块，使用“CSS 选择器”，获得 1688 商品详情页的 SKU 数据项所在元素单元的 CSS 定位字符串。提取的结果是符合定位条件的一组网页元素的数据项，即一个**列表**。我们分配一个新的变量“一组 SKU 数据项”，代表这个列表。列表中的每个**数据项**本身又是一个列表，包含 3 项内容，即标题、价格和库存。



这个元素单元的 CSS 定位字符串“.table-sku tr”通过图中所示两个动作获得：1、鼠标移至包含 SKU 标题的网页元素上面，但不要点击页面；2、按下键盘的“S”键一次。这两个操作告诉 CSS 选择器，你需要的是鼠标所在位置的网页元素的 CSS 定位字符串。通过避免鼠标点击操作，网页上原有的人机交互功能，特别是与鼠标点击相关的功能，就不会运行，因此网页的结构在这两个操作步骤前后保持不变。以后运行全自动的机器人流程时，这个 CSS 定位字符串将一直是有效的（除非网站运营方更改了前端界面）。



按键盘“S”键选中SKU标题“插电3档调光+USB（银色）”后，接着按键盘“P”键两次，使选择范围扩大了两级，把标题、价格和库存都包括进来，即网页元素的一个单元。单元内有一个或多个元素包含了采集所需的数据，组合在一起成为一个数据项。



然后把鼠标移到下一个SKU标题“插电5档调光5档调光定时+USB线（银色）”，再次按键盘“S”键，此时所有SKU数据项都会被自动选中。同时，CSS定位字符串也变成“.table-sku tr”。



此时将获得的 CSS 定位字符串 “.table-sku tr” 全选后复制到机器人流程的积木块中，用于提取商品详情页的一组单元。



- 第五步，添加**数组类**积木块 B，建立一个列表，表示单元内的一组元素各自对应的 CSS 定位字符串。分配新的变量 “SKU 三个元素”，代表这个列表。由于这个列表包含三个元素的数据，因此需从**文本类**积木块中选三个空白文本块放入列表。将积木块 C 做相应的修改，即 “……单元内部的一组 SKU 三个元素 定位的……”。



- 第六步，回到 1688 商品详情页，关闭 CSS 选择器后重新运行，鼠标移到第一个 SKU 标题元素上，按键盘 “S” 选择后如图所示。



然后将鼠标移到第二个 SKU 标题元素上，再次按键盘“S”选择后如图所示。



最后将鼠标移到第三个 SKU 标题元素上，再次按键盘“S”选择后如图所示，此时获得的 CSS 定位字符串是“.name span”。

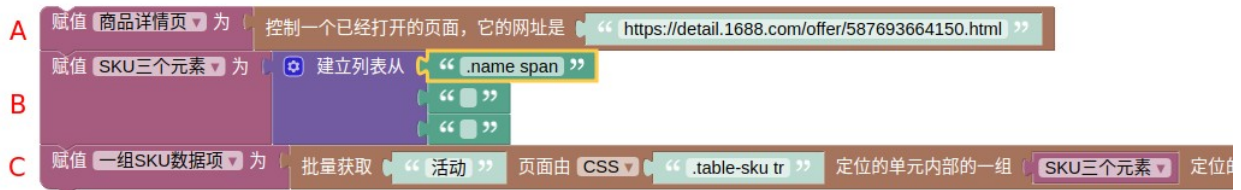


如果发现 CSS 定位字符串不是图示结果，或者没有选中所有元素，请关闭 CSS 选择器，然后点击图示下拉箭头按钮，把所有 SKU 选项都显示在网页上。再次运行 CSS 选择器，重复第六步的一组动作，**最多选三次**，可以看到这个 CSS 定位字符串

“.name span”能够把所有的 SKU 标题元素（总共八个）都选中。



7. 第七步，将 CSS 定位字符串复制粘贴到机器人流程的积木块 B 中。



8. 第八步，采取类似第六、七步的过程，最多选三次，选中总共八个价格元素，获得 SKU 价格元素的 CSS 定位字符串 “.price span em.value”，同样复制粘贴到积木块 B 中。



9. 第九步，采取类似第六、七步的过程，最多选三次，选中总共八个库存元素，获得SKU库存元素的CSS定位字符串“.count em.value”，同样复制粘贴到积木块B中。

颜色	SKU名称	价格	库存
	插电3档调光+USB (银色)	31.00元	5000个可售
	插电5档调光5档调光定时+USB线 (银色)	39.00元	5000个可售
	充电5档调光5档调色定时+USB线 (银色)	49.00元	5000个可售
	无线充5档调光5档调色定时+USB线 (银色)	59.00元	5000个可售
	插电3档调光+USB (黑色)	32.00元	5000个可售
	插电5档调光5档调光定时+USB线 (黑色)	40.00元	5000个可售
	充电5档调光5档调色定时+USB线 (黑色)	51.00元	5000个可售
	无线充5档调光5档调色定时+USB线 (黑色)	60.00元	5000个可售

点此洽

.count em.value

A 赋值 商品详情页 为 控制一个已经打开的页面，它的网址是 “ https://detail.1688.com/offer/5876 ”

B 赋值 SKU三个元素 为 建立列表从 “ .name span ” “ .price span em.value ” “ .count em.value ”

C 赋值 一组SKU数据项 为 批量获取 “ 活动 ” 页面由 CSS “ .table-sku tr ” 定位的

10. 最后一步，添加文本类积木块D，将采集到的“一组SKU数据项”显示到输出窗口。

A 赋值 商品详情页 为 控制一个已经打开的页面，它的网址是 “ https://detail.1688.com/offer/587693664150.html ”

B 赋值 SKU三个元素 为 建立列表从 “ .name span ” “ .price span em.value ” “ .count em.value ”

C 赋值 一组SKU数据项 为 批量获取 “ 活动 ” 页面由 CSS “ .table-sku tr ” 定位的单元内部的一组 “ SKU三个元素 ” 定位的元素的 “ innerText ” 属性值

D 将 一组SKU数据项 显示到输出窗口

本课的机器人流程到此已经可以运行了。你可以多次刷新 1688 详情页，并运行这个流程，验证它总是能准确采集到全部的 SKU 数据。如果你采集的数据项少于图示的八组，请检查两个 CSS 定位字符串是否有误。

```
输出窗口
导出Excel 清空
2019-09-12 10:51:43 >>>>>>>> 运行开始 >>>>>>>>
[控制一个页面] 134 =
https://detail.1688.com/offer/587693664150.html
插电3档调光+USB（银色）,31.00,5000
插电5档调光5档调光定时+USB线（银色）,39.00,5000
充电5档调光5档调色定时+USB线（银色）,49.00,5000
无线充5档调光5档调色定时+USB线（银色）,59.00,5000
插电3档调光+USB（黑色）,32.00,5000
插电5档调光5档调光定时+USB线（黑色）,40.00,5000
充电5档调光5档调色定时+USB线（黑色）,51.00,5000
无线充5档调光5档调色定时+USB线（黑色）,60.00,5000
2019-09-12 10:51:44 <<<<<<<<< 运行结束 <<<<<<<<<
```

本课总结

总结一下本课所传达的概念。

- **CSS 定位字符串**可用于采集网页上一个或多个元素的数据
- 一个**单元**是网页上包含多个元素的一个组合，用于提取对应的数据项
- 启用**键盘辅助操作**（按下“P”键）用于选择当前选中元素的上一级（父）元素

本课内容是对第三课的进一步深化，用于采集网页上类似表格结构的一组数据项。提取的结果是一个列表，且列表中每一项也是一个列表。

本课附带了一个课后练习，如果你希望继续第五课的学习，请务必用上你的鼠标和键盘，在自己的浏览器上实践几次。如果中途出错，没有关系，只要刷新一下页面就可以了，直到获得预期的结果。

课后练习

请你尝试重新搭建四个积木块，采集另一个商品详情页的 SKU 数据 (<https://detail.1688.com/offer/580299626823.html>)。与本课演示案例不同的是，该商品详情页的每一个 SKU 数据项对应的单元包含一个图片、一个价格和一个库存。

流程搭建完毕后，刷新详情页面，然后运行流程，这样反复实验多次，验证采集的一组数据项中的价格和库存数据是准确的。注意，图片元素没有文本内容，也就没有相应的“innerText”属性，因此采集到的数据是空的。没关系，下一课将介绍“src”和“data-lazy-src”元素属性，专门用于获取图片元素的网址。

在“机器人流程自动化”页面，你可以选择本地代码的“打开最近一次运行的代码”，在第三课的课后练习基础上，修改并添加所需积木块，完成本课的课后练习。

当然，我们建议你创建新的流程，从零开始搭建这四个积木块，复习 CSS 选择器的使用方法。

颜色		7.40元	296649 个可售	-	0	+
		7.40元	295610 个可售	-	0	+
		7.40元	296064 个可售	-	0	+