**构建IT系统**



**2019年秋天,**

**带回家的任务2:人气云(权重:30%)。2019年11月8日(星期五))**

***概述***

此任务要求您使用本单元(Python)、数据库查询(SQLite)和web开发(HTML)中介绍的几种技术。

***目标***

这个任务包括编写一个查询SQLite的Python程序



数据库并生成HTML文档作为其输出。因此，这是本单元主题的一个很好的例子。，不同的“计算机语言”可以一起用来建立新的IT系统。

根据对存储在数据库中的电影客户的调查，这项任务的目的是开发一个程序，创建HTML文档，显示电影演员在粉丝中的受欢迎程度。更具体地说，给定特定类别的影迷，例如，女性或年龄在30到40岁之间的客户，我们想知道存储在电影数据库中的演员在这一组中的受欢迎程度。所需的输出是一组HTML页面，每个客户组一个。每个页面必须包含演员知名度的可视化表示，其中更受欢迎的演员姓名显示在更大的字体中。

***数据库***

一个名为movie\_survey的SQLite数据库备份文件或“dump”。提供了包含用于此任务的数据的sql。它包含了微软员工关于他们最喜欢的演员和电影的调查结果。为了完成这个任务，您需要首先通过导入SQLite DB浏览器中提供的脚本来重新创建这个数据库。这个数据库中有四个表:actors、customers、actors\_movies和favorite\_actors。其中，这个任务需要两个表:customers和favorite\_actors。下面将解释这两个表的定义。

***表客户***

Table customers提供关于电影顾客或顾客的信息。表中的每一行由七个字段组成，如下所示。字段customerID是客户的标识符，是主键。字段2至7提供客户的个人信息，包括年龄、教育程度、性别、婚姻状况、子女数量和车辆数量。字段由下面的模式定义:

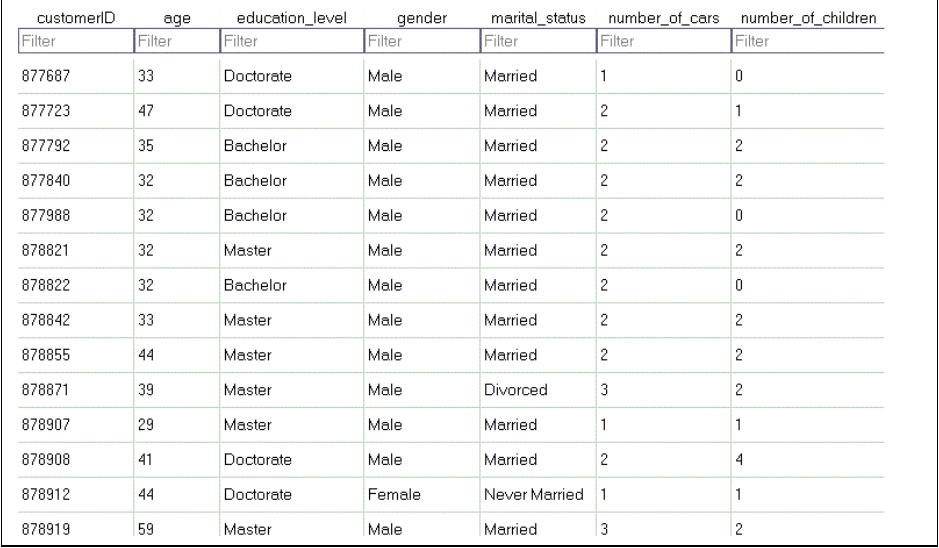
CRICOS 00213 j

第1页

**JIT104构建IT系统，2019年秋季**



customerID INT(6)，年龄INT(11)，教育程度VARCHAR(30)，性别VARCHAR(10)， marital\_status VARCHAR(20, number\_of\_cars INT(11)， number\_of\_children INT(11)，主键(customerID)



***表favorite\_actors***

Table favorite\_actors提供了影迷最喜欢的演员的信息。表中的每一行由三个字段组成，如下所示。ID是该记录的标识符，customerID是客户的标识符，而actor是该粉丝的最爱之一的actor的名称。字段由下面的模式定义:

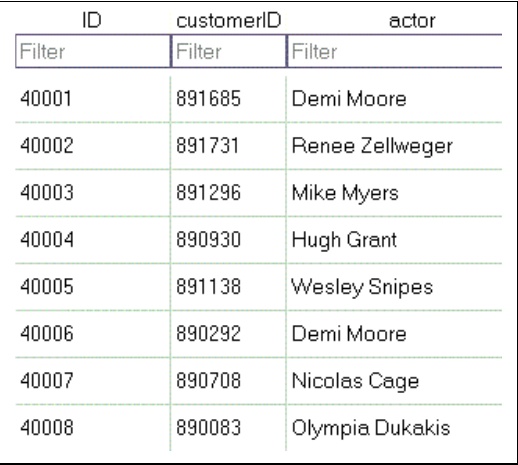
ID INT(6)， customerID INT(6)， actor VARCHAR(100)，主键(ID)

CRICOS 00213 j

第二页

**JIT104构建IT系统，2019年秋季**





请注意，这个数据库中的演员名格式是“given\_name姓氏”，如上表所示。当您按此任务中要求的顺序并显示名称时，可以将每个名称“given\_name姓氏”作为一个字符串处理，而不需要分别处理给定的名称和姓氏。(这意味着名字将根据演员的名字排序。)

***一般要求***

这个任务的目的是在视觉上展示演员在特定类型影迷中的受欢迎程度。您需要开发一个Python程序来访问上面描述的SQL数据库来生成HTML文档。程序的用户指定:•显示多少结果;和

•要展示哪类影迷的观点。

结果是一个超链接的HTML文档集合，每个电影客户类别一个。用户可以指定的不同类别的客户有:

•男粉丝;

•女粉丝;

•一定年龄范围内的粉丝，如30 - 40岁，包括;•所有粉丝。

图1显示了其中一个HTML文档的示例。它显示了演员在30 - 40岁(含30 - 40岁)电影客户中的受欢迎程度。(该页面是测试2在提供的模板文件popularity.py中生成的页面之一。)一个演员的受欢迎程度是由喜欢这个演员的顾客数量决定的。

他们的想法是用更大的字体显示更受欢迎的演员。例如，汤姆·汉克斯比桑德拉·布洛克更受欢迎，因为汤姆·汉克斯在这个年龄段有420名粉丝

CRICOS 00213 j

第3页

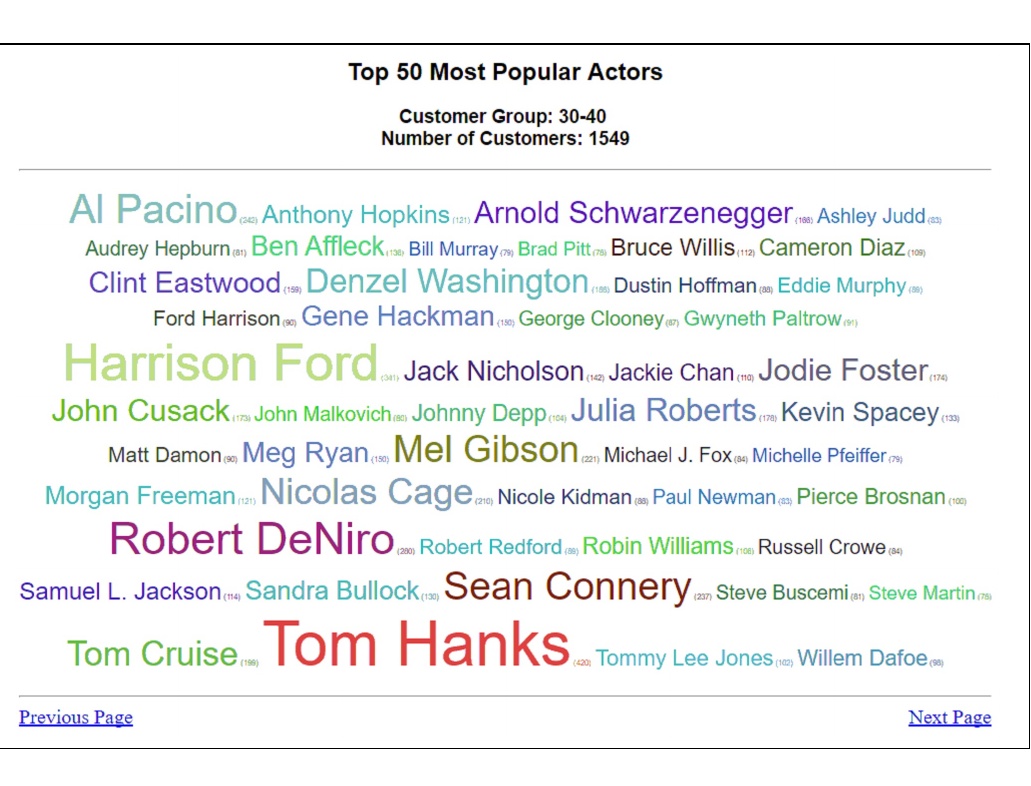
**构建IT系统**



**2019年秋天,**

*桑德拉·布洛克只有130。*在图1中，Tom Hanks使用的字体明显大于Sandra Bullock使用的字体。每个演员的名字后面都跟着喜欢这个演员的电影客户的数量，用一种较小的、不打扰的字体显示。例如，在图1中，Tom Hanks与420联系在一起，因为他是420个客户最喜欢的演员。

在这个任务中，我们也使用不同的颜色来展示不同的演员。同样受欢迎的演员以同样的字体大小和同样的颜色显示。例如，下面例子中的达斯汀·霍夫曼和妮可·基德曼字体大小和颜色都相同，因为他们都有88位粉丝。不同受欢迎度的演员用不同的字体和不同的颜色，例如下面例子中的汤姆·汉克斯和桑德拉·布洛克用不同的字体和不同的颜色。(一个可能



优化的任务是联系颜色，使更受欢迎的演员得到一个更强的颜色，但你不需要这样做。

**图1:演员人气(前50名)，粉丝年龄30 - 40岁(Test\_02, p2)**

您的工作是开发一个Python程序，该程序使用SQL数据库movie\_survey中的数据生成这样的页面。如这些说明所附的Python模板文件中所示，您必须定义一个名为show\_popularity的函数，该函数给出一个客户类别列表，生成可以在web浏览器中显示的HTML文件，每个客户类别一个。

CRICOS 00213 j

第4页

**构建IT系统**



**2019年秋天,**

图1中显示的文档的另一个特性是，在页面的顶部必须显示客户的类别，例如，在本例中为“30-40”岁的客户，以及该组中的客户数量，例如，上面的1549。

此外，您的程序必须将每个页面超链接在一起，以便用户能够方便地搜索它们。在上面的图1中，我们可以在页面底部看到到列表中的前一个和下一个类别的超链接。但是，不会为用户列表中的第一个类别生成“previous”链接，也不会为列表中的最后一个页面生成“next”链接。例如，如果给定的客户组列表是[' Female '， ' Male '， ' 30-40 ']，那么您的函数必须生成三个HTML页面，一个用于女性客户，一个用于男性客户，一个用于30-40岁之间的客户(包括30-40岁)。在男性电影客户页面，“上一个”链接和“下一个”链接将分别引导我们进入女性客户页面和30 - 40岁客户页面。然而，对于女性客户的页面，没有“先前”链接，对于30到40岁的客户的页面，没有“下一个”链接。

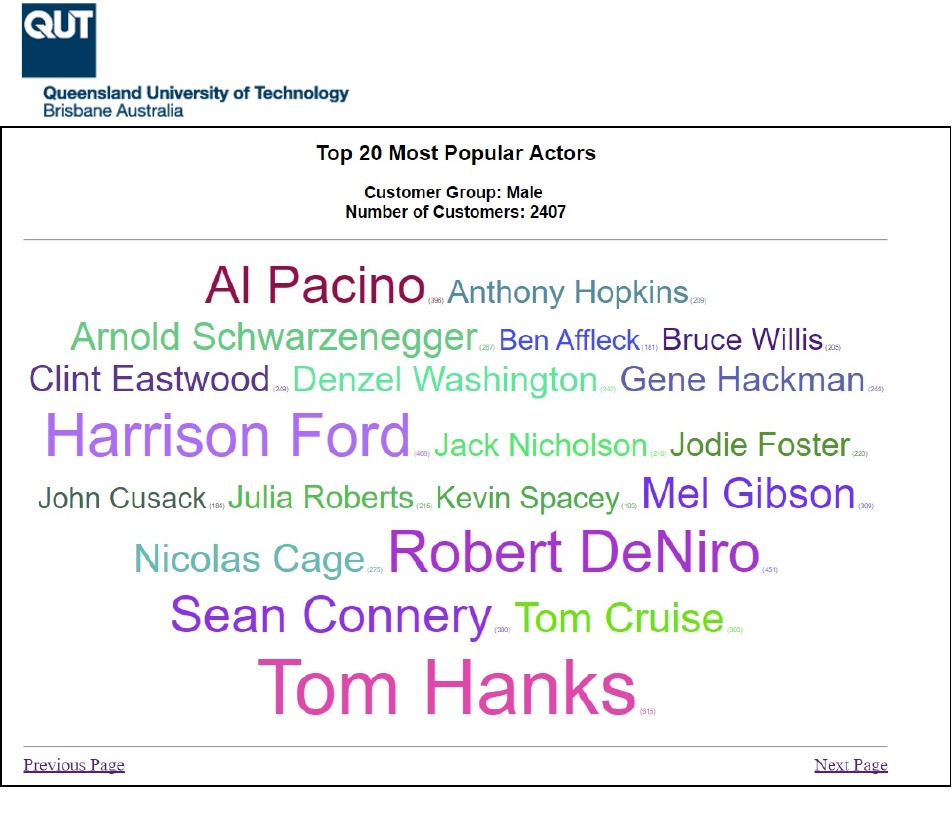
下面的图2到图4显示了我们在分类列表[' Female '， ' Male '， ' 30-40 ']中询问粉丝最喜欢的20位演员时的预期输出。(这三个页面是模板文件popularity.py中测试1的预期输出。)在每种情况下，每个特定客户类别中的参与者的相对受欢迎程度都被清楚地显示出来，并且每个页面都按照客户类别列表的顺序被超链接在一起。



**图2:演员人气(前20)**

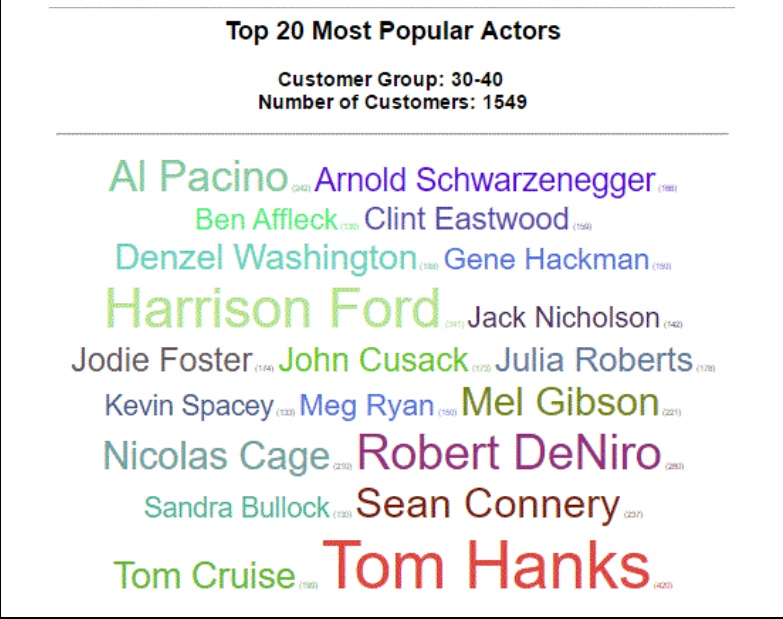
CRICOS 00213 j

第5页



**JIT104构建IT系统，2019年秋季**

**图3:演员人气(前20)(Test\_01, p2)**



**图4:演员人气(前20)**

CRICOS 00213 j

第6页

**构建IT系统**



**2019年秋天,**

***特定的任务***

完成这个项目组合任务必须做的具体事情如下。这是一项艰巨的任务。您应该将解决方案模块化为几个函数，以使程序代码易于管理和阅读。您必须使用提供的模板文件普及度。以py为起点。

1.恢复movie\_survey movie\_survey.sql

数据库

使用

的

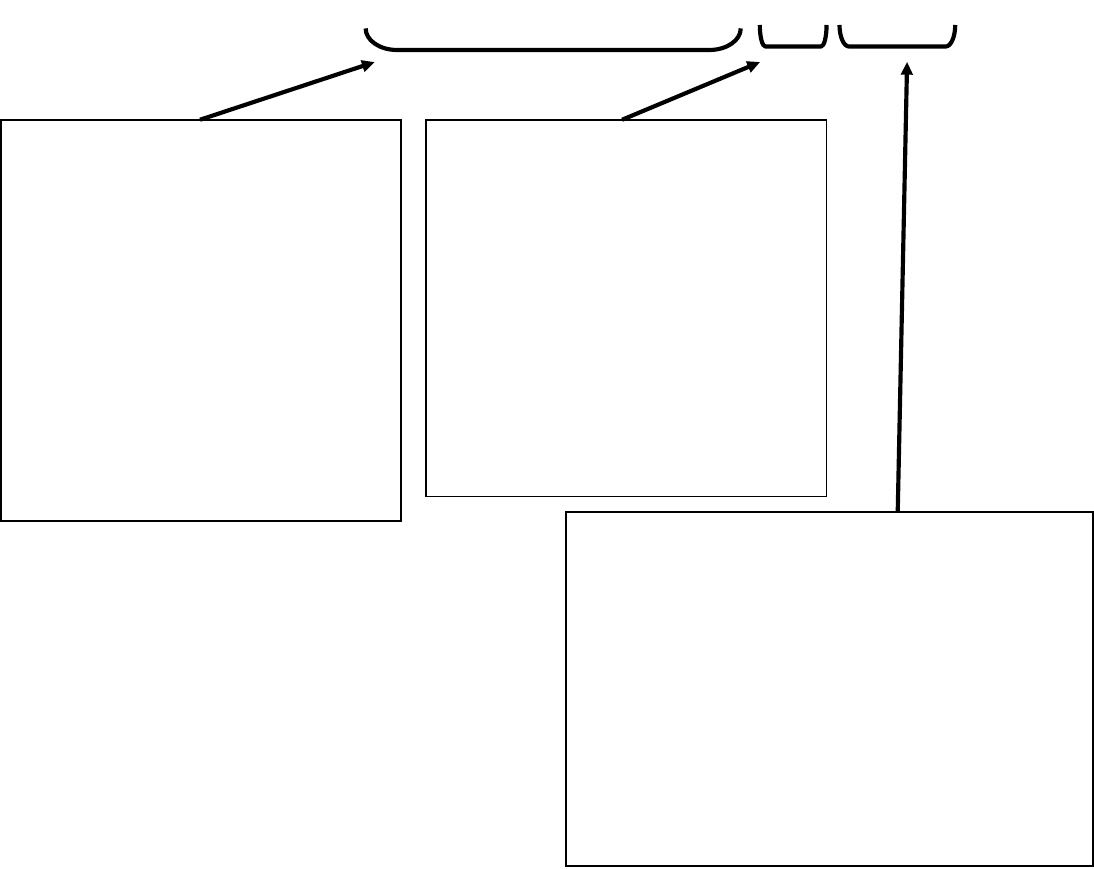
提供

转储

文件

2.编写一个名为show\_popularity的Python函数,有三个参数:(a)一个字符串列表,代表客户组的列表,(b)一个整数值代表有多少演员要显示在页面上,和一个字符串(c)给这个特殊的多页文档名称。

例如，考虑下面的函数调用:



show\_popularity([20 - 40、40 - 80, '所有'],30日Test03)

第一个参数(列表)表示所需的客户类别，此示例表明需要创建三个页面:一个用于20到40岁的客户(包括20岁到40岁的客户)，一个用于40到80岁的客户(包括40岁到80岁的客户)，一个用于所有客户。(其他有效的客户类别为“女性”和“男性”。)

第二个参数是一个整数值，它指定每个页面上要显示多少参与者的名字。对于特定的客户群体，那些显示的必须是最受欢迎的演员。(提示:您希望根据流行程度对数据库查询返回的结果集进行排序。)

第三个参数只是用来唯一标识这个HTML页面集合的名称。您应该将它与客户类别一起使用，为生成的HTML文件创建名称。在这个特殊的例子中，你的程序必须产生以下三个HTML文件:

Test03\_20-40。html test03\_40 - 80。html Test03\_All.html

CRICOS 00213 j

第7页

**构建IT系统**



**2019年秋天,**

无论演员的知名度如何，都必须按字母顺序排列。如前所述，数据库中每个参与者的名字由一个给定的名字和一个“given\_name姓氏”格式的姓氏组成，因此只需将它们作为单个字符串排序。例如，在图1中，阿尔·帕西诺、安东尼·霍普金斯和阿诺德·施瓦辛格按这个顺序显示，尽管阿诺德·施瓦辛格的受欢迎程度比安东尼·霍普金斯高。每个演员的名字后面必须跟着从数据库中获得的粉丝的实际数量。确保数据库查询生成的数字与上面的数字匹配。

在每个页面的底部，如果有客户类别列表，您必须有到上一页和下一页的超链接。

3.有些查询根本不会产生任何结果。在这种情况下，您仍然应该生成一个HTML页面，即使没有列出参与者。下面的图6显示了一个示例(这是测试10在提供的模板文件中生成的页面之一)。



**图5:没有返回结果的页面(Test\_10, p2)**

4.如前所述，有效的客户类别是“女性”、“男性”、“所有人”和年龄范围(如30 - 50。如果在作为show\_popularity的第一个参数提供的列表中发现任何其他类别，则应该打印一条消息，指出错误，然后忽略该项目。但是，仍然必须为列表中的所有有效情况生成页面。有关示例，请参阅提供的模板文件中的单元测试。

***发展线索***

•在开始这项任务之前，您必须确保能够访问SQLite DB浏览器，以便能够创建数据库。您还必须能够访问必要的SQLite-Python模块，以便能够从Python代码中调用SQLite函数。

•

给定一个电影客户类别，您需要对数据库执行两个查询，一个查询这个类别中有多少客户，另一个查询演员列表和他们的受欢迎程度。第一个查询相对简单，只涉及到customers表。第二个涉及到customer和favorite\_actors表的连接。建议您在尝试将数据库查询合并到Python程序之前，先在DB浏览器中开发和调试数据库查询。将查询保存到.sql文件中，以便以后需要时可以引用它们。

CRICOS 00213 j

8页



**构建IT系统**

**2019年秋天,**

•使用上述截图作为指导，但你不需要复制字体大小和颜色的精确选择。然而，重要的是，您的SQL查询必须返回与上面示例中相同数量的fan。

•这是一个很大的项目，所以建议您循序渐进地解决它。在开始编写代码之前，您应该使用DB浏览器来确定Python程序需要执行哪些查询。然后开发代码来生成一个HTML页面。最后，编写代码迭代客户组列表以生成所有文档。

•你需要解决如何定义字体大小和颜色的文本显示在网页上。这可以通过改变字体的“样式”来实现。关于如何做到这一点，请参考演讲演示和研讨会练习。

***可交付成果***

您应该通过完成并提交所提供的模板文件popularity来开发您的解决方案。py如下。

1.这个文件必须包含一个函数show\_popularity，它具有上面定义的所有特征。当在适当配置的SQLite数据库中执行时，该程序必须生成所需的HTML文件。

2.在Python文件的开头填写“作者声明”，通过在指定的地方插入您的姓名和学号来确认这是您自己的个人作品。**我将假定没有完成的陈述的提交不是你自己的工作，他们将不被标记。**

您不能提交任何SQL数据库脚本或转储。提供的SQL数据库，以及可能的数据库更新，将用于测试您的软件。因此，您的程序不能以任何方式硬编码，因为它将需要与任何类似结构的数据库一起工作。

***要求及评分指引***

您需要通过完成show\_popularity函数(并根据需要添加其他函数)来扩展所提供的模板，以便它能够从数据库中存储的数据生成所需的HTML页面。您的代码必须适用于模板文件中提供的所有数据集和相同格式的任何其他数据集。

您的解决方案必须通过模板文件中提供的所有测试用例，并使用相同格式的任何其他数据库。特别是，你的解决方案必须具备以下特点:

•SQL查询(4%)。SQL查询必须格式良好，对于任何给定的类别值，都必须从数据库中生成准确的结果。

CRICOS 00213 j

9页



**构建IT系统**

**2019年秋天,**

•HTML页面(2%)。根据提供的第一个参数(即类别列表)生成所需的HTML页面数量。

•顶级演员(2%)。根据第二个参数中提供的值，每个HTML页面只包含最受欢迎的参与者的数量。

•HTML内容(10%)每个HTML页面应包括:

oan适当的标题(在浏览器选项卡中)

oa标题包括演员名单的数量

oa子标题包括:

 客户 group; 的 类型和

 group. 客户 的 数量

超连结至上一页及下一页(如有)

•HTML主体(10%)。HTML页面的主体必须按字母顺序显示最受欢迎的演员，包括:

演员名字的大小反映了他的受欢迎程度(更受欢迎的演员用更大的字体显示);

具有相同知名度的oactors以相同大小的字体显示;oa不同的字体颜色为困难的人群;和

以相同字体颜色显示具有相同人气的oactors;每个演员的受欢迎程度根据他的名字而定。HTML中使用的字体和颜色必须是可读的。

Python代码质量和表示(2%)你的程序代码必须以专业的方式呈现。有关如何实现这一点的建议，请参阅JIT104代码标记指南中的编码指南。所有Python和HTML代码必须易于阅读和理解，感谢:

oclear整洁的布局;

简明的英文注释，解释每一个重要的代码段的目的;阐明其作用的变量名;

避免“魔力数字”;

避免不必要的代码重复的ocode;

适当地使用函数来模块化代码;

纠正语法和拼写错误。

如果你不能解决整个问题，提交任何你可以工作的部分。不完整的答案会给你打几分。

CRICOS 00213 j

第10页



**构建IT系统**

**2019年秋天,**

***学术诚信***

此作业仅供个人评估。这意味着你提交的作品必须是你自己的作品，而不是别人的作品。除了讨论高层次的策略外，你不能和你的同事合作完成这项任务。你不被允许要求或委托他人为你写作业，或帮助你写作业。

如果你对什么是允许的，什么是违反学术诚信的，有任何疑问，请尽快与教学人员交谈。

作者:修订:

CRICOS 00213 j

科林·菲奇(QUT)唐娜·蒂格(QUTIC 2019)

11页