FSE598 项目 2

（作业 3：50 分 和 作业 4：50 分）

企业数据处理

引言

此作业的目的是确保您理解并熟悉课程中涵盖的概念，包括数据结构、算法、XML、JSON 及他们的处理和实现。当您完成此作业后，您应该已将这些概念和技术应用于创建可运行的程序以处理这些数据结构。这是个人作业。每个学生都必须完成并提交独立作业。不允许合作完成。不能重复教材或课程幻灯片中给定的示例。可以遵循这些示例，但必须进行重大更改。此项目可以在 Windows, Mac, 或 Linux 机器上完成。

实践练习和准备（这部分无需提交）

这部分练习无需提交。但是，做这些练习可以帮助您更好地理解概念，从而帮助您完成测验或考试。

1. 在以下网址上查看并下载样本 XML 文件：

   http://venus.sod.asu.edu/WSRepository/xml/Courses.xml

2. 在本次作业中，我们将使用公共服务器来托管文件。有不同的服务器可用，例如阿里云、腾讯云、Azure 云、AWS 云和 Google 云，在本次作业中，我们建议您使用 GitHub 服务器：https://pages.github.com/ 来托管您的 .txt 文件。

3. 将您在作业 3 中生成的 Movies.txt 文件部署到您选择的服务器中。试图用游览器访问 Movies.txt 文件。

项目作业（这部分需要提交，100 分）

作业 3 （50 分）

本项目有两个作业。两个作业的截止日期相同，必须一起提交。 但是，两个作业的分数将分别输入。 去掉最低作业分数的规定只会去掉一个作业分数。

在此项目中，您将创建一个电影目录(Movies)，它将表示为对象树（键和值）。图 1 为此项目中创建的电影目录的所需结构。所有“Movie”对象都具有相同的结构。请注意，不同形状和颜色的方框具有不同的含义。蓝色圆角矩形表示键，白色矩形表示最终值，即键的文本内容。图 1 中给定的键结构必须按描述实现，而图中给定的文本内容（即白色方框）是示例值，在您的文件中可以不同。实线箭头表示父-子对象关系，虚线箭头表示最终键-值关系。对于每个 Movie 对象，必须允许多个 Director 对象。该项目包括以下问题（任务）。

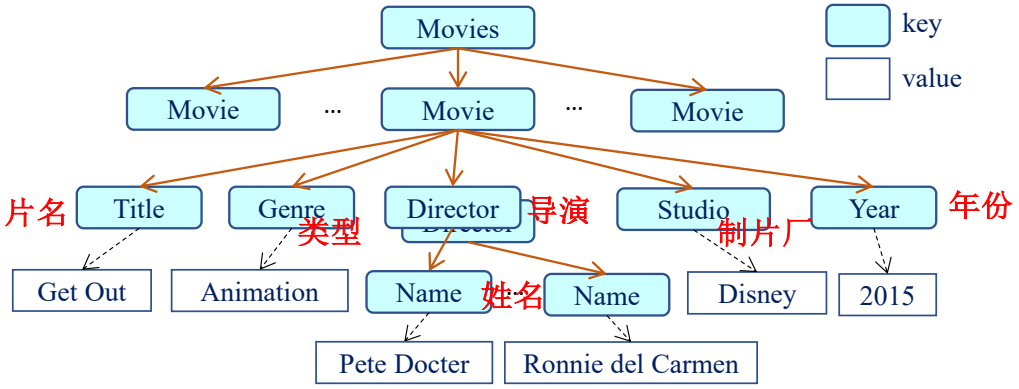

图 1.电影信息树

1. 使用无格式文本编辑工具（例如记事本）创建（手动键入）一个文本文件，并将文件命名为 Movies.txt。此文件必须包括至少四种不同电影类型（题材）的至少 10 部电影，例如动作(action)、动画(animation)、喜剧(comedy)、剧情(drama)、虚构(fiction)、恐怖(horror)等。该文档必须包含图 1 中包括的信息。所有电影片名都必须不同。输入电影时，不要按片名对电影进行排序。您可以按照您想要的方式组织文本文件。 但是，您必须以一种可以轻松读取文件的方式组织它，以构造要在以下问题中创建的数据结构。 但是，文件格式不得为 XML 或 JSON。 例如，您可以采用以下格式：

   Tile: Get Out
   Genre: Animation
   Director Name: Pete Doctor
   Director Name: Ronnie Carmen
   Studio: Disney
   Year: 2015

   将 Movies.txt 部署到服务器中（请参阅“实践练习和准备”部分中的练习）使您的程序可以在线读取。您也可以将文件放在程序的本地目录中。 [25 分]

2. 编写一个 Python 程序，其中包括一个函数，该函数可从服务器或本地目录中读取在问题 1 中部署的文本文件 Movies.txt。 然后，生成 JSON 数据结构的**列表**（JSON 对象和列表组合）。生产数据结构的代码可与读取该文件的代码写在同一函数中，也可写在另一个函数里。您可以将数据存储在内存中并输出(print)数据，或者将其保存在文件中。可使用 Python 字典结构存储 JSON 结构，该结构必须遵循图 1 中的树结构。JSON 文件应如下所示： [25 分]

```
Movies =
    [
        { “Movie”: [
                        {“Title”: “Get out”},
                        {“Genre”: “Animation”},
                        {“Director”: {“Name”: “Pete Doctor”}},
                        {“Director”: {“Name”: “Ronnie del Carmen”}},
                        {“Studio”: “Disney”},
                        {“Year”: 2015}
                    ]
        },
        { “Movie”: [
                                    …
                    ]
        },
        …
    ]
```

注意，第一行“Movies =”不是数据结构的一部分。可以使用单独的打印语句来打印。

作业 4 （50 分）

在此作业中，您将扩展前一个作业的工作。

3. 在 Python 程序中添加 Sort 函数，该函数应按照在上一个问题中创建的 JSON（字典）列表数据中数据中的给定键对电影列表进行排序。 键应该是 Sort 函数的参数。 键可以是以下值之一：Title, Genre, Director（将按第一位导演排序），Studio, or Year。 结果，Sort 函数生成一个新的 JSON（字典）数据列表，该列表按所选键值排序。 您可以将数据存储在内存或文件中。 [25 分]

4. 在 Python 程序中添加 XmlGeneration 函数，该函数应读取排序后的 JSON 数据并生成一个与图 1 中的树相同的 XML 树。XML 文件的根元素是”Movies”。您的程序应将 XML 树保存至名为“Movies.xml”的文件中。 [25 分]

作业 3 和作业 4 应一起提交。 提交清单应包括：

- 在问题 2、3 和 4 中创建的 Python 程序 Movies.py。
- 问题 1 中的 Movies.txt。即使文件在服务器中，您仍需要在此处提交文件。

- 问题 2 中的打印输出或文件中的 JSON 数据 Movies.json
- 问题 3 中的打印输出或文件中的按电影片名排序的 JSON 数据 MoviesSorted.json
- 问题 4 中的 Movies.xml

将在所有问题中创建的所有文件放入一个文件夹中，压缩该文件夹，然后提交 zip 文件。

提交要求和说明

请确保您整理后提交的文件夹中包含所有的文件，包括解决方案文件、项目文件、和代码。这样您提交的项目就可以直接被打开并测试，而不需要重新组织和整理您的文件。

然后，将您所有要提交的文件复制到一个文件夹中，并将此文件夹压缩，在Canvas界面上上传并提交您的作业。

提交作业注意事项：该作业由多个分布式任务和组件组成。当您创建它们时，它们可能存储在您电脑的不同位置。您必须在创建项目时选择一个新的位置来存储这些文件。然后，您可以将这些文件复制到一个文件夹中，以便在Canvas上提交。为了确保您在压缩文件中包含了所有的文件，并且它们能一起工作，您必须在提交前测试它们。您也必须从Canvas上下载您提交后的文件。然后在不同的机器上解压并测试您的作业，看看您是否能在不同的机器上运行解决方案，因为助教会在不同的机器上测试您的应用程序。

如果您提交了一个空的文件夹，或者一个不完整的文件夹，我们将不能对您在截止日期后重新提交的部分进行评分！我们只对您在截止日期前提交的作业进行评分。

分数和评判标准

每个小问题（编程任务）都有一定的分数。我们将按照以下步骤对您的程序进行评分：

(1) 编译代码。如果不能编译，将扣除50%的编译代码的分数。然后，我们将阅读代码，并在50%和0之间给分，如下面的评分表的右边部分所示。

(2) 如果代码通过了编译，我们将使用测试案例执行和测试代码。我们将根据评分表的左边部分来给分。

在这两种情况下（编译通过和编译失败），我们将阅读您的程序，并根据分配给每个子问题的分数、您的代码的可读性（代码的组织和注释）、逻辑、所需功能的包含以及每个功能的实现的正确性来给分。

请注意，我们不会对您的程序进行调试以弄清错误的大小。您可能会因为一个小错误而失去50%的分数，比如少了一个逗号或一个空格!

我们将对作业中列出的**每个子问题**采用以下评分标准。Pts是用来表示每个子问题的分数。

评分表

评分项	通过编译的代码	代码编译失败

	分数 * 100%	分数 * 90%	分数* 80%	分数* 70% -60%	分数 * 50% - 40%	分数* 30% -10%	0 or 1
针对于每一个子问题	满足所有要求，注释良好，在所有测试案例中都能正常工作。	在所有测试案例中工作正常。没有提供注释来解释代码的每一部分是做什么的。	可以通过测试但出现一些小问题，如不写注释，代码在某些不常见的边界条件下不工作。	在大多数测试案例中可以工作，但有重大问题，如代码不能通过一个常见的测试。	编译或不能正常工作，但显示出认真努力解决这个问题的态度。	编译失败，显示了一些努力，但代码没有实现所需的工作。	0 分，如果没有提交。 1分，如果提交的文件不能打开、没有内容或没有回答任何问题