**小学期CPP课报告之Model part**

课题：篮球数据可视化（俱乐部数据以及球员个人数据的收集与可视化分析）

本模块负责人：谭泽源 – 计科1601 – 3160103832

其他同组组员：梁家宁 - 计科1601 – 3160105677

林灿东 - 计科1601 – 3160103968

1. 文件处理
2. 文件格式

爬虫获取的文件为JSON格式的文件，而且大多数信息的格式如{“name”: “James”, “pts”:”20”}这样的成员结构。因此，这里的结构一一对应了我们所想要了解的每个球员或球队的各项数据。我们考察经典的C++解析JSON文件的库，最终选取了开源的RapidJSON代码库，即一个轻量级的JSON解析库，作为解析JSON文件的方式。

1. 文件读取

我们把球员和球队的信息分别记录在两个不同的文件夹中，两个文件夹里面分别用每个球员或球队的名字命名一个JSON文件。所以在读取文件的时候，我们采用了调用windows文件系统的函数。这个函数可以找到指定文件夹下所有文件的名字，对他们进行遍历。而我们分别对球员文件夹和球队文件夹分别遍历，就可以找到所有的文件信息。

1. 写入内存

我们在内存里不但定义了球队和球员这两个结构体，在内存中，我们采用哈希表的形式来储存球员和球队的结构体数据。所以，在写入内存的时候，首先要确定是哪个球员或球队，即选用哪个值作为哈希表的键值，最终我们简单选取了球员或球队的名字作为键值。另外一个重要的问题是如何将JSON文件变为结构体，这里我们用了RapidJSON的成员解析函数，即把每个球员或球队的各项数据通过JSON格式中的成员得到，例如”pts”:”20“就可以通过解析函数表示为有一个名为pts的成员，它的值是”20”，然后再通过字符串数字转化，最终确定该球员的得分成员的数据为20。这样，在球员和球队的哈希表中，以名字为索引，索引到对应的结构体向量。与此同时，我们也记录了结构体构成的向量，作为以后排序用。

1. 数据处理
2. 总数据

前端需要得到一段时间内的球员或球队的综合数据时，我们通过名字索引结构，先找到结构体向量中符合那一段时间限制的结构体，然后将他们的均值不断地用新读取的值更新，最终得到的就是总的均值。

1. 序列数据

前端需要得到一段时间内的球员或球队的一段时间内的序列数据时，我们还是类似于总数据的获得方式。只不过我们并没有对数据进行平均处理，而是按照时间逐一记录了数据返回给前端。

1. 排序数据

我们把所有的球员或结构体按照前端给定的指标（某项数据，或者数据的组合）进行排序，利用c++向量排序的方法，得到前15个结果返回。