信息管理与存储项目计划书

Steam数据查询平台

10164601140彭钧涛

10164800305郭源杰

10161900218李尚真

1. 背景与意义

随着经济的快速发展和物质生活水平的提高，人们的精神文化需求日益增长，互联网的高速进步使游戏逐渐成为了人们精神生活中的一部分。根据《2019中国游戏产业半年度报告》显示，2019年上半年，中国游戏市场实际销售收入1163.1亿元。游戏行业已经成为了中国经济中重要的一部分。

游戏是一种基于物质需求满足之上的，在一种特定时间、空间范围内遵循某种特定规则的，追求精神需求满足的社会行为方式。现在的游戏多指各种平台上的游戏。Steam平台作为现在市场上最大的综合性数字发行平台，玩家可以在该平台购买、下载、讨论、上传和分享游戏和软件。Steam平台上的数据能在一定程度上反应出玩家的特征、喜好以及游戏的发展趋势。

本数据管理项目研究Steam各款游戏的数据，由于各款游戏之间互相独立，不存在传统的关系型数据库中表与表之间的复杂查询逻辑，不会因为数据量庞大而产生重复的冗余数据，所以本项目通过非关系型数据库存储海量数据，可视化数据，通过研究价格走势来预测游戏的未来价格，根据用户的标签喜好推荐游戏，对研究游戏行业的发展有着重大意义。

1. 数据管理问题

下图是我们用爬虫从Steam公司网页爬取的《Half-Life》游戏的数据，可以看到数据由种类、名称、ID、年龄限制、是否免费、DLC情况、游戏介绍、支持语言、各种系统的配置要求、开发公司、促销包信息、平台信息、价格信息、分类标签、发售日期、历史价格记录。



1. 采用的数据库

由于我们准备存储的是是游戏的各项属性，而这些数据的关系基本上可以被解构为以游戏为单位的文档型数据。在考察现今流行的非关系数据库后我们发现MongoDB不以具体表进行存储，也没有传统关系型数据库中行和列的概念，同时以文档形式存储，每一个文档中用键值对进行存储。且文档利用JSON格式进行存储，这样的存储方式更有利于数据的传递和交互。MongoDB灵活性体现在其良好的可扩展性，我们可以向数据库中插入任何JSON形式的文档而不受模式的约束。因此我们决定使用MongoDB进行数据的存储。我们设计数据库存储的形式为以一个游戏为建立一个文档，文档中包含该游戏的各项信息。

1. 数据源

数据源来自steamdb.info，使用Selenium获取各个游戏的历史价格数据

1. 技术路线

由于使用MongoDB存储数据，我们将每个游戏以appid作为标识，存储为一个JSON文件，每个JSON文件存储游戏的名称、标签、随时间变化的价格等信息，查询时，直接根据要查询的appid找到对应的JSON文件，取出价格数据，绘制成图表进行显示。

同时我们提供根据游戏的类别标签进行游戏推荐，根据用户所选择的标签，筛选出用户可能感兴趣的游戏展示给用户。

由于Steam的各类促销周期比较稳定，因此每一类游戏的价格有相对固定的变化趋势，我们将为不同的打折原因（例如：万圣节、圣诞节、发行商周末、夏季和冬季促销）赋予不同的权重，并根据历史的价格变化数据，综合各方因素考虑，对下一次可能的打折日期进行预测。