测试报告

### 测试目的

在开发大型软件的漫长过程中，面对极其错综复杂的问题，人的主观认识不可能完全符合客观现实，与工程密切相关的各类人员之间的通信和配合也不可能完美无缺。因此，在软件生命周期的每个阶段都不可避免地会产生差错。尤其对于毕业生管理信息系统这类会影响学生工作的工程软件，必须尽量减少差错，以免造成严重的损失。测试是“为了发现程序中的错误而执行程序的过程”。测试的目的就是在软件投入生产性运行之前，尽可能多的发现软件中的错误。目前软件测试仍然是保证软件质量的关键步骤，它是对软件规格说明．设计和编码的最后复审，也是必不可少的关键步骤。

### 项目背景

测试项目：景点推荐系统

项目开发者：褚思源，林靖清，王炜康

项目形式：网页

### 测试概要

项目测试分为两方面：网站兼容性测试和推荐算法测试。

#### 网站兼容性测试

需要在主流桌面系统（Windows，macOS，Linux）和主流桌面浏览器（Chrome，FireFox，Edge，IE）测试网站的兼容性，包括网页响应速度，网站布局，网站交互是否正确等。

#### 推荐算法测试

###### 人工评测

顾名思义，邀请一帮人来对推荐系统的结果进行评测。这个方法非常直白的做到了人工评测，但是缺点也很明显。首先是人工评测的成本比较高；其次就是无法指定详细的规则，掺杂了较高用户主观性；另外就是推荐之前的用户行为很难模拟；最后是如何做到和基准的对比等等。所以人工评测应作为辅助。

###### 指标评估

指标化推荐结果，也就是将推荐的结果用不同的指标来进行说明，通过这些指标，我们可以更加全面了解推荐系统，部分指标不一定越高越好，需要让它保持在一定的范围内。下面我们看下这些指标：

**覆盖率**

定义：推荐系统能够推荐出来的 “商品 / 类目” 占“总商品 / 类目”集合的比例。假设系统的用户集合为 U，推荐系统给每个用户推荐一个长度为 N 的物品列表 R(u) ，总物品为 N。

意义：描述推荐系统对物品发掘的能力。

**基尼系数**

覆盖率反应出的分布情况是比较有限的，只能知道哪些类目覆盖，哪些没有覆盖，那类目之间究竟哪个类目占的多，哪个类目占的少？为了更细致地描述推荐系统发掘长尾的能力，我们需要统计推荐列表中不同类目出现次数的分布，引入基尼系数来评价。

定义：按照类目的流行度（曝光次数）从大到小排序后进行统计后进行洛伦茨曲线的绘制。

做法：以类目分布基尼系数为例，算出所有的类目被曝光的次数，需要以天周期为单位进行数据的统计。

**发现性**

定义：推荐系统对用户未产生过关系的商品的发现能力。

以类目为例，用户点击了一个感兴趣的景点，而这个景点的类似恰恰是用户前一周都没有点击过的内容，这就说明推荐系统的为我推荐了一个我之前都没有关注过并且我感兴趣的内容，也就是系统的发现性，在算出每个人的值之后，再进行求平均计算。

### 测试结果

在主流桌面系统（Windows，macOS，Linux）和主流桌面浏览器（Chrome，FireFox，Edge，IE）下，网站兼容性表现良好。由于网站未适配移动端，除 iPad 之外，小屏移动设备上的体验均不佳。推荐算法应用效果良好。