# 6 数据模拟与测试

## 6.1 系统用户规模

在做数据模拟之前，首先必须要考虑到系统的用户规模。一般来说，为了保证模拟数据的真实性和可靠性，需要尽可能接近或超过实际系统的用户规模。具体而言，如果系统用户规模为百万级别，那么模拟数据的规模也应该达到百万级别或以上，以便充分测试系统的性能、稳定性和可扩展性。

系统的用户规模影响诸多方面的内容，比如用户关系。用户规模越大，用户关系的复杂程度就越高，数据处理的难度越大，输出的结果可能会变得不可控。

用户规模受到用户信息的限制。本测试内容主要针对用户关系的发现，模拟数据中，尽管用户信息的细致性和真实性并非是重点内容，但是好的用户案例可以更好的帮助测试人员评估预期结果的合理性，因此，在充分保证用户信息的真实意义前提下，模拟的数据不能过高，过高将会导致测试数据失真。

综合上述内容，考虑到模拟数据的生成和处理成本以及数据模拟的实际目的，针对用户发现的数据模拟数据量在十万级别。这个级别的数据量可以保证充分的测试接口，同时兼顾到数据的生成和处理成本，保证测试效果和数据质量。

## 6.2 用户关系的模拟

成功模拟一定规模的用户数据后，下一步内容就是用户关系数据的模拟。与用户数据的模拟不同，用户关系数据是驱动用户发现算法的关键。在进行用户关系数据模拟时，需要考虑到用户之间的互动方式、频率和强度等因素，并根据实际情况生成相应的数据。

有一个基本定律，对于内容的访问遵循80/20原则，也就是20%的内容，会占有80%的访问量。对于用户关系的一个描述是，20%的用户，会占据系统80%的粉丝数据。在数学中有一种分布规律可以近似描述这种规律：Zipf定律(Zipf's law)。以下模型能够很好的适合用户关系的模拟：

其中，u是一个[0,1)的随机数，α是幂律分布的指数。当α = 1时，它的分布情况比较理想，分布规律近似相近于函数：

图形工具可以直观的表现它的规律。

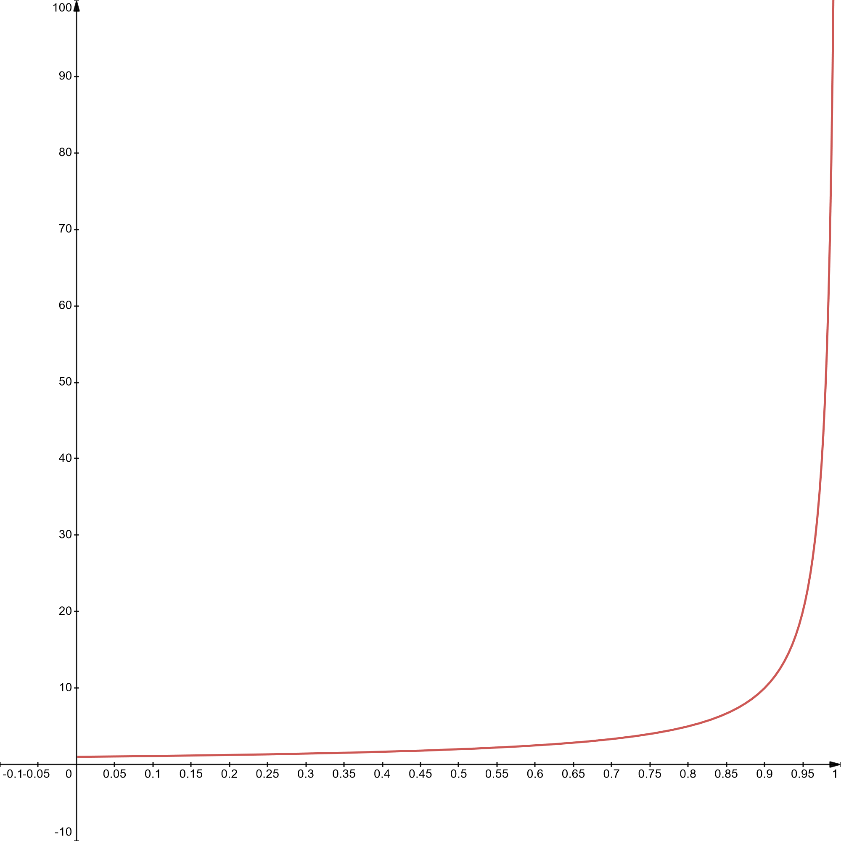


图6.1 幂函数的曲线规律

系统把上述公式作为用户关系的概率函数，产生了超10万用户关系记录。它们被存入数据库，并形成事务集，用来后续算法处理。



图6.2 事务集文件

算法的输出是频繁事务集。在最小支持度为1000的参数模型下，共生成了241条频繁项集。这是基于用户关系的用户推荐的核心数据。



图6.3 频繁项集文件

## 6.3 系统主要功能测试

功能测试是黑盒测试中的一种。本系统庞大，功能众多，本章节展示系统主要功能测试。

1、系统标识

系统标识是打造品牌的重要一步。一个好的标识能让用户留下深刻印象，增加系统的辨识度和品牌力。本系统设计了系统标识，以“PETEN 宠物园”以及形象小狗的标志作为系统标识，简洁且美观。



图6.4 系统标识图

2、系统概览

本系统以内容为主，在系统设计上大量使用卡片化设计，各模块之间功能明确，内容主体清晰。下图展示了系统首页测试图，内容为主的文章内容占据了页面是大部分，富有冲击力。左下角是用户推荐，推荐结果符合预期。

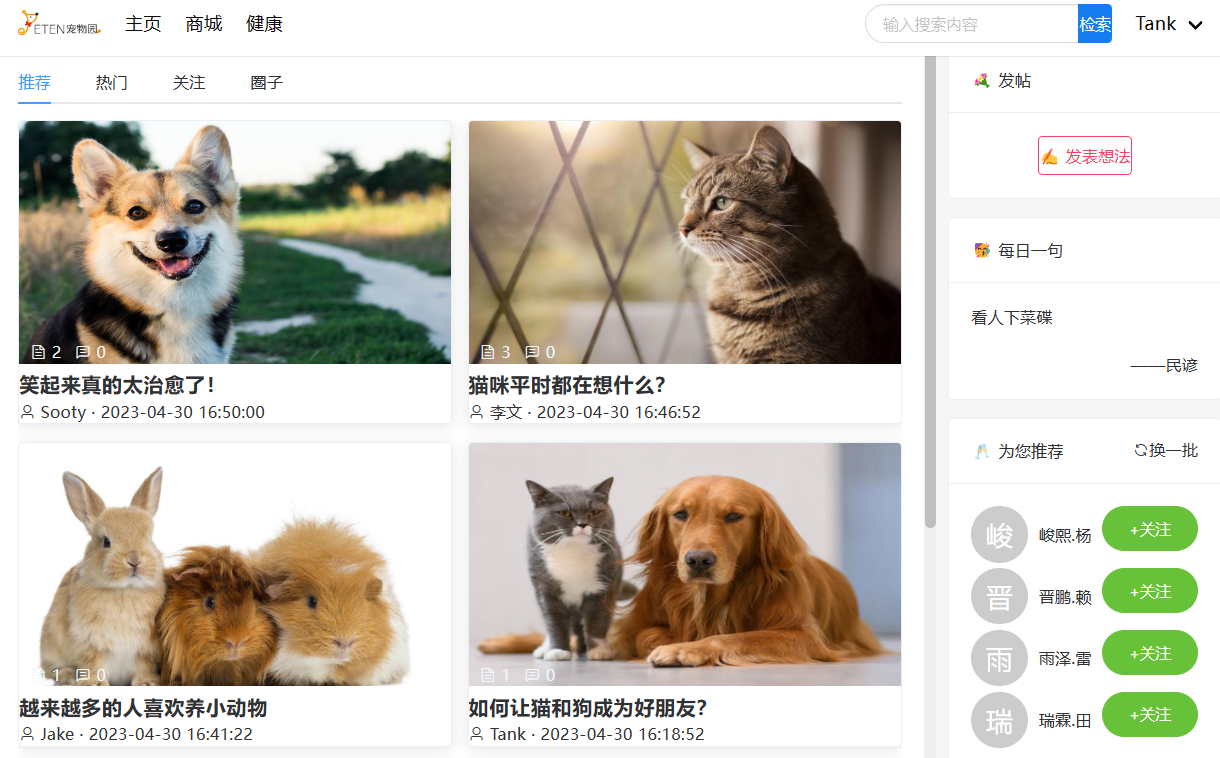


图6.5 系统首页概览图

3、文章编辑与发布

内容编辑和发布是系统的重要内容，系统采用了富文本编辑器，支持markdown的语法，用户的编辑体验将更加丰富。本测试案例采用markdown语法，进行文章编辑的测试，测试内容包含富文本文字、图片。编辑框中包含预览的内容，可以实时预览编辑结果。



图6.6 文章编辑测试图

内容发布后将被作为数据流转到多个地方，比如推荐、热门、动态等。文字内容的预览展示依然是卡片方式，卡片展示的好处是视觉冲击力强，内容清晰。



图6.7 文章发布测试图

4、文章详情与推荐

文章详情的展示关系到用户的预览体验。同时，文章内容与相关推荐的紧密结合能够很好的留住用户，吸引用户继续点击。相关推荐是基于文本特征的内容推荐算法的输出结果。



图6.8 文章详情与相关推荐测试图

5、用户评论

用户评论能够很好的促进用户之间的互动，本系统的评论模块是基于树结构的二级评价机制。测试案例使用了“评论-回复”的测试方法，结果符合预期。

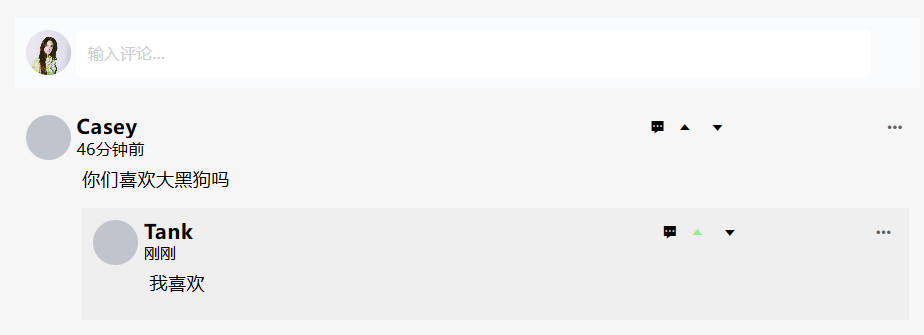


图6.9 用户评论测试图

6、热门文章

热门文章是根据文章热度增长与消退处理的结果，数据的展示方式与系统首页的内容展示相同，首页展示的推荐数据，本项目展示的热门文章数据信息。

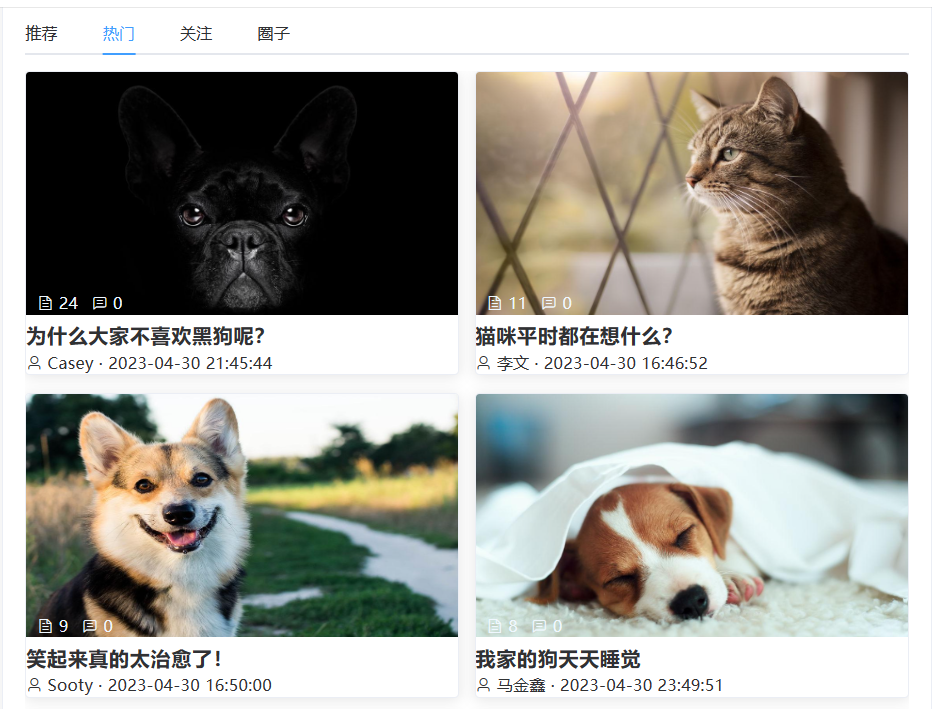


图6.10 热门文章测试图

7、用户关注与动态

用户关注动态是指被关注的用户的动态信息，一般来说，被关注的用户动态向关注用户公开。页面中展示5个被关注者，更多的被关注者可以点击“更多”展开查看。动态信息根据发布的时间信息倒序排序。



图6.11 用户关注动态测试图

8、文章搜索

文章搜索是系统体验和高效数据查找的保证。本系统的搜索功能基于Elasticsearch搭建，对于匹配的关键字进行了高亮处理。本搜索功能的基本逻辑是首先对关键字进行分词，然后进行高效查找。区别于传统数据的模糊查询，Elasticsearch的能力可以接受更加模糊的查询关键字。在本案例中，选用“养只猫养只狗”作为查询关键字。



图6.12 文章搜索测试图

9、圈子与帖子

圈子是吸引特定兴趣用户的保证。在设计中首先要保证圈子的主题要突出。内容的展示要重点突出客体。简单来说，圈子有基本的“装饰”；帖子的展示要突出内容主体；功能入口要清晰可见。本测试采用“柴犬”作为测试案例。

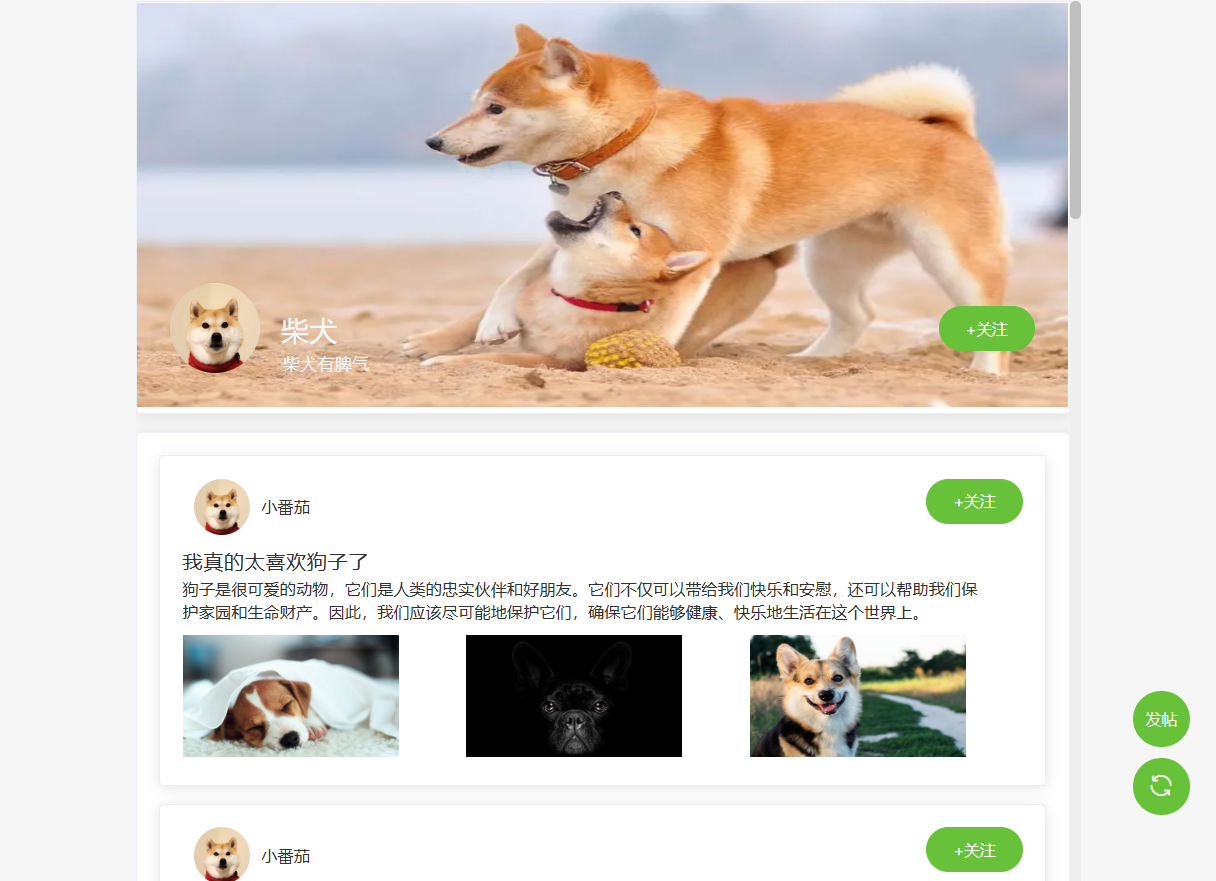


图6.13 宠物圈子测试图



图6.14用户帖子测试图

10、商城概览

商城包含了商品的展示，包括商品总览和商品分类；商城相关功能的入口标识，包括订单、购物车等入口信息展示。

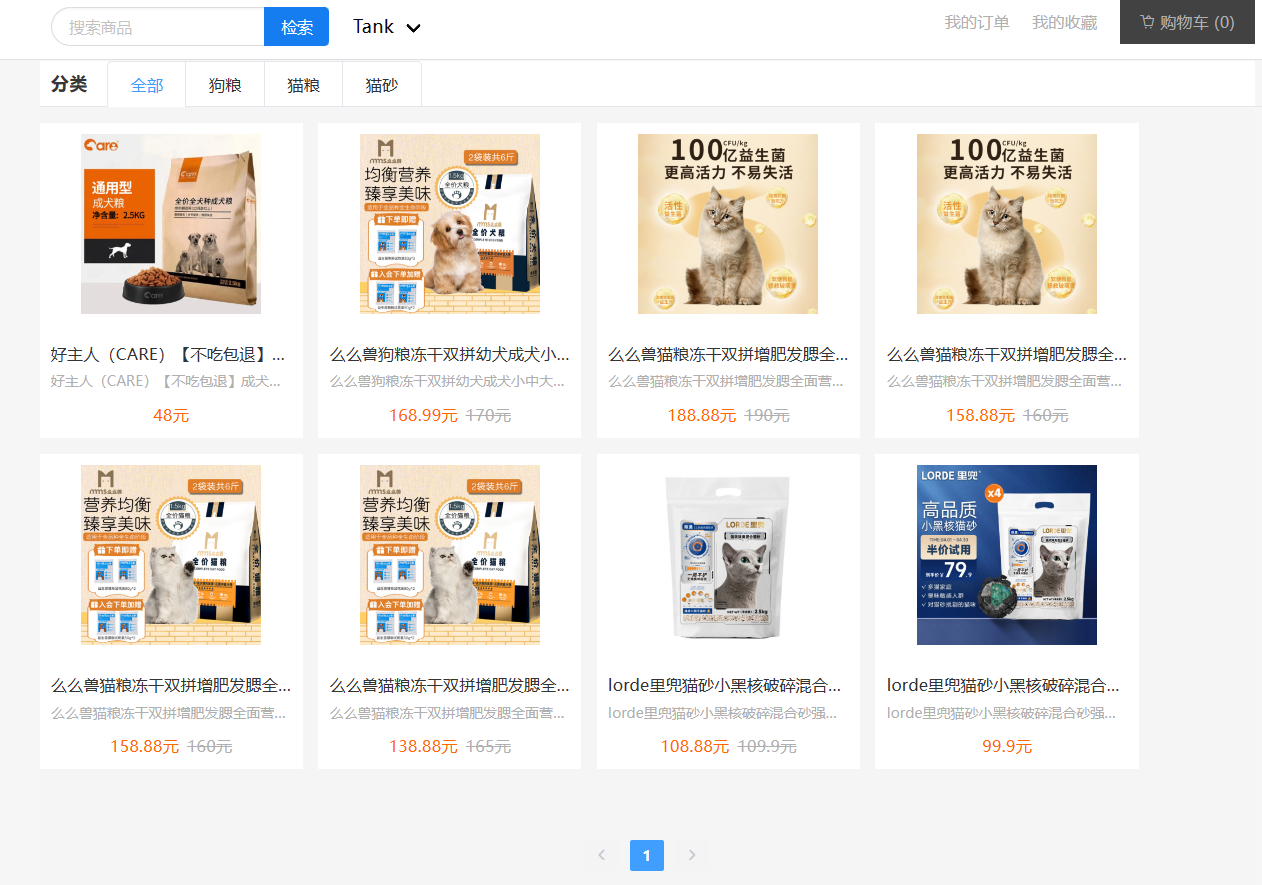


图6.15商城概览测试图

11、商品概览

商品概览详细展示了商品的信息，包括描述、广告图、价格说明等内容。本测试是“么么兽狗粮”，测试案例中展示了商品的全部信息，包含一些功能入口，测试结果符合预期。



图6.16 商品详情测试图

12、购物车概览

购物车是方便用户体验的优化设计，也是商品购买的入口。本测试案例是“么么兽狗粮”，测试案例中展示待购买信息，以及商品结算的入口。测试结果符合预期。



图6.17 用户购物车测试图

13、宠物服务发现

宠物服务发现是指用户在系统上查询和浏览宠物服务机构。本系统不提供第三方机构的宠物服务，但是提供搜索发现功能，为用户提供便捷入口。下图是宠物医院的测试案例。

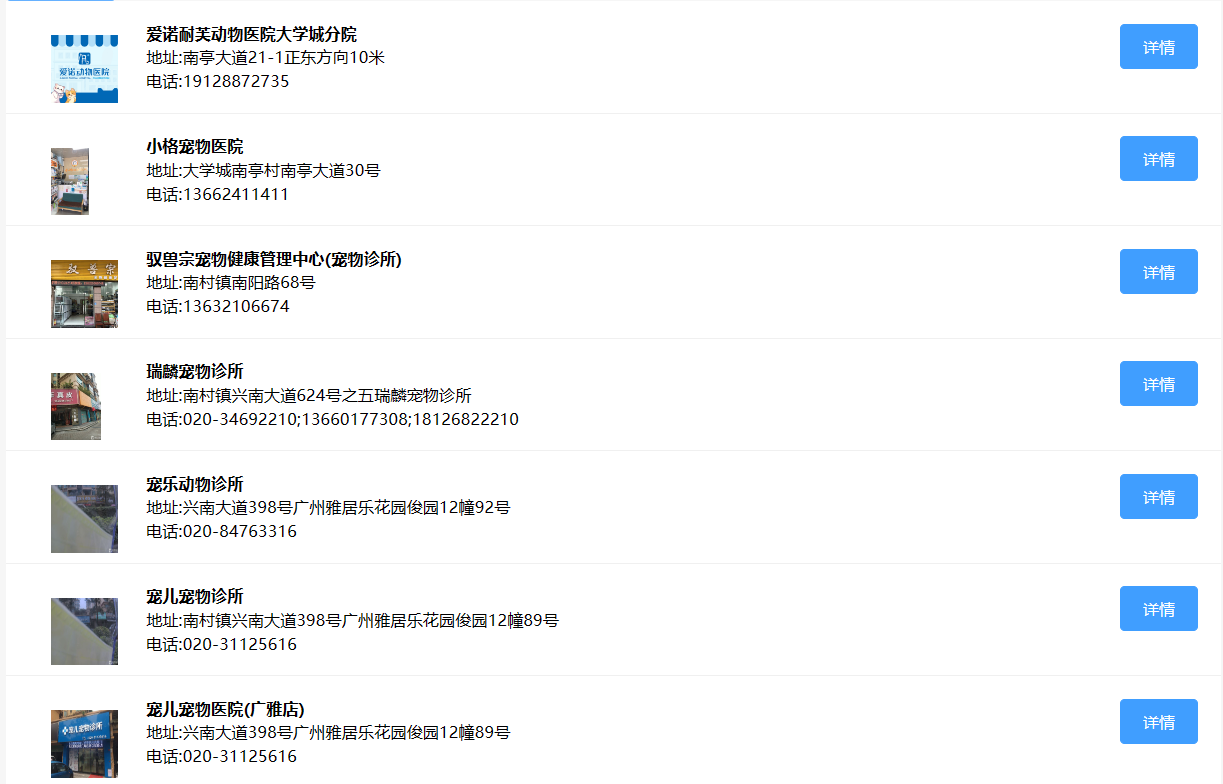


图6.18 宠物医院测试图

14、宠物服务机构详情

宠物服务机构详情是指宠物服务机构的详细信息。本系统使用高德地图服务，使系统具备定位与导航能力，给用户提供良好的使用体验。下图是“广州爱诺耐芙动物医院大学城分院”的测试案例。



图6.19 宠物医院详情测试图

## 6.4 系统测试小结

本测试包含和系统功能测试和接口性能测试。其中接口的性能测试包括了接口的单元测试、响应速度测试、边界测试等内容，属于白盒测试的一种。接口性能测试基于Spring Boot的能力，使用图形化测试工具实现。

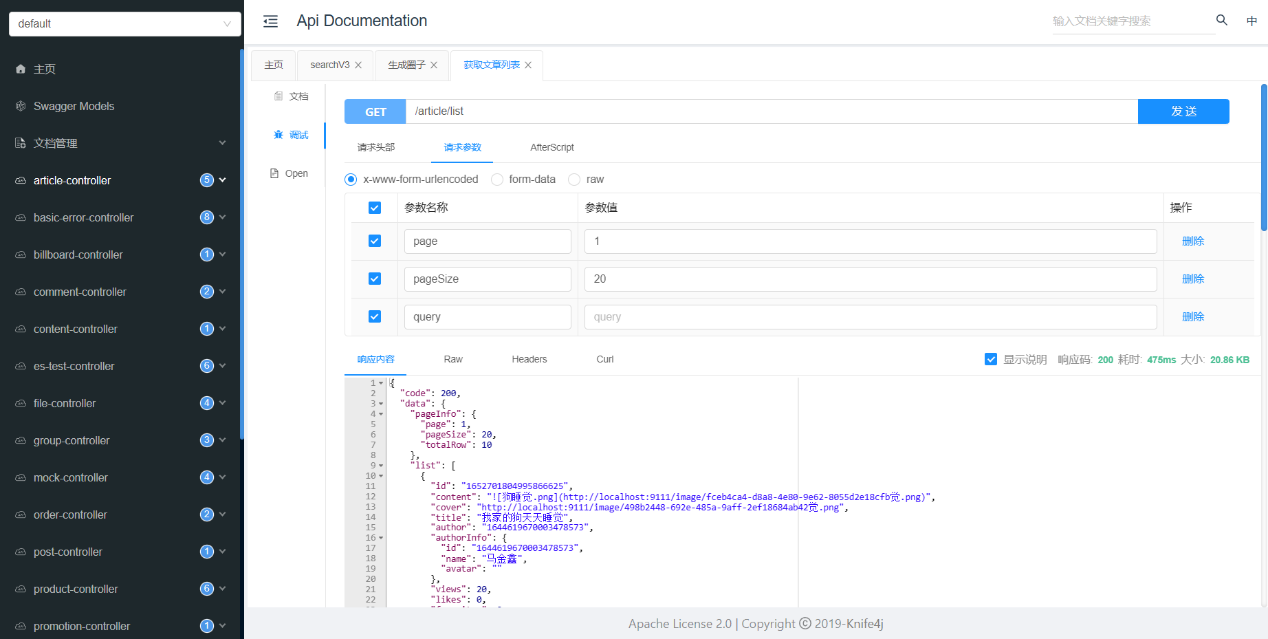


图6.20 性能测试图

系统中所只用的接口相应速度全都保持在500毫秒以内，系统的性能、数据处理的准确性、响应的速度得到保障。