为什么我们要设计task：usability test。

文本

低可信度描述已自动生成

问卷中的启发式“task”（实际上我们并不能知道用户到底做了没有，也无从客观地评估effectiveness，efficiency这些measures），这些要依赖用户对问卷后面题目的打分。这些“task”的提出是为了引导他们在网页上停留更长时间，进行更多的操作，并且帮助他们思考，为后面的答题做准备。

Please explore it freely, learn about its features and consider the tasks below.

1. Think of a kind of vegetable, fruit or other food (for example, spinach, strawberry, bread or milk). Is this product available on the website? Try to find it.

目的：引导用户思考和评价网站是否提供了便利的浏览、筛选和搜索功能。

1. Try adding and removing items to and from the shopping basket.

目的：引导用户思考和评价添加和移除商品的操作是否便利。

1. Can you keep track of the items in your basket and the total price?

目的：引导用户思考和评价网站是否对操作提供了灵活的交互和响应。

1. Try proceeding to the checkout page.

目的：引导用户尝试在不同页面和操作阶段之间跳转。

1. Look at the navigation bar, floating windows and other components, what do you think of them?

目的：引导用户思考和评价网页的features、外观、设计和布局等。

与lab study中task的对比：问卷中的task比较简单，数量较少，耗时短，描述比较宽泛，自由度高，这是因为这些task是启发性的，目的是让用户多尝试一些操作，从而对effectiveness，efficiency和satisfaction进行主观的思考和评价。

Lab study中的task更复杂，耗时长，描述具体，自由度低，是对上面几种操作的综合（定位商品，加入/移出购物篮，实时检查操作是否成功、加入的商品和份数是否正确，跳转到checkout页面，等）。由于实验者可以现场观察被试的操作并进行录像、计时、检查结果等，必要时还可以与被试交谈，比起问卷，lab study可以对effectiveness，efficiency、satisfaction进行更客观和全面的评价。

Lab Study Task: Using the website provided to you, please add the items (with specified quantity) to the basket and proceed to the checkout page:

1. Bread \* 1
2. Eggs \* 4
3. Milk \* 8
4. Orange juice \* 3
5. Sausages \* 1
6. Strawberries \* 2
7. Blueberries \* 2
8. Cherries \* 2
9. Potatoes \* 1
10. Carrots \* 1
11. Mushrooms \* 2
12. Spinach \* 5

文本

中度可信度描述已自动生成

Hypothesis: our new website is a) more usable on the desktop than the prototype.

b) as usable on a mobile phone（同问卷想要用统计学方法分析打分证明的。）

* One-tailed: 我们假设新版总是比旧版更好。

Experimental design:

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

Choice of subjects：你已经写了

Determine independent and dependent variables: 独立变量和因变量

* 独立变量：用户使用的网页版本（新/旧）和操作设备（手机/电脑）（为了控制变量，我们让所有的被试使用同一部手机（iPhone 8 Plus，Safari浏览器）或同一部笔记本电脑（ThinkPad，Google Chrome浏览器）进行实验。

1. Time/effort to learn: 实验第一阶段，即给定上限为五分钟，请被试自由探索网站的功能，如果认为已经学会了，可以提前停止计时。在这个过程中，观察被试进行了哪些操作，是否对实验者提出问题，提出了什么问题。
2. Time to perform tasks：第二阶段，记录被试完成上面task（根据实验者提供的shopping list将指定商品加入购物篮并checkout）所花费的时间。
3. Errors committed：由实验者观察被试在操作过程中的失误，以及检查最后task的完成情况（加入购物篮的商品和份数是否正确）。一般情况下，要求实验者尽可能少地介入，如果被试在执行任务时遇到困难，让他们先自行探索可能的解决方法，只有在被试完全卡住，实验无法继续时才进行提示。
4. User satisfaction：两阶段实验结束后，让用户对网站进行满意度打分。

Organization and the problem of training: 由于场地和时间限制，我们携带设备对被试上门拜访，让他们在家中完成实验。为了评估网站自身的Learnability，我们并不对被试进行网站使用的教学，而是让他们自行探索用法（第一阶段）。

Between subjects allocates subjects to different conditions: 按照你说的，participants来自不同gender, age等群体，随机分配，以减小个体差异对实验结果的影响。

实验第一阶段补充：

图形用户界面

低可信度描述已自动生成

在被试“自学”的过程中，关注他们的操作。

Which buttons or links are they interested in? Which functions did they try out? (For example, did they try clicking on “Contact” or “About” for the old website? Did they try “Help” and “About & Contact” for the new one? If so, how long did they spend on the pop-up windows before closing them, did they actually read the content?) Did anyone actually read the privacy policy, or did they all just close the window directly?

文本

中度可信度描述已自动生成

基于通过cooperative evaluation来enrich lab study的目标，我们鼓励被试在实验过程中“think aloud”，说出自己实时的想法，比如抱怨找不到某件商品，关不掉弹窗，字太小看不清等。实验者必须忠实地记录这些context，因为他们是对网页操作体验最真实、具体的反馈。比起全部实验结束后再询问“您对网页有什么意见和建议？”，被试在执行任务过程中遇到挫折发出的抱怨更能诚实地反映网页最细节处的缺陷和漏洞。然而，为了尽可能少地affect normal behaviour，实验者在实验过程中一般不回答被试提出的问题（如“xx按钮在哪儿？”，“为什么我点击这个链接没有反应？”），除非问题已经严重阻碍实验的继续进行。

文本

描述已自动生成

尽量启发被试自己想办法解决，不到万不得已不直接回答或者上手帮忙。

文本

描述已自动生成

同样也在鼓励被试“think aloud”的过程中得到实现：观察，他们遇到了哪些困难，哪些是可以自己解决的，哪些需要提示，哪些需要实验者上手操作？

实验结果（定量分析）：

* 对于每个网页版本和每个设备，各有10名被试，共40人。
* Time to perform tasks: (Efficiency)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Condition | Prototype + Computer | Prototype + Mobile Phone | Our Version + Computer | Our Version + Mobile Phone |
| Average Time | 2 min 35 s | 2 min 59 s | 2 min 4 s | 2 min 21 s |

可以看到在不同的版本上，使用手机完成任务所需的时间略长于电脑，这可能是由于手机屏幕较小，一次性能显示的商品较少，上下划动操作频繁，手指点击容易误触等。而对于同一设备，改进后的版本所需时间明显短于prototype。

* Errors committed: (Effectiveness)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Condition | Prototype + Computer | Prototype + Mobile Phone | Our Version + Computer | Our Version + Mobile Phone |
| No Error | 6 | 4 | 8 | 7 |
| Minor Error | 3 | 4 | 2 | 3 |
| Major Error | 1 | 2 | 0 | 0 |

Definition

* No Error：被试顺利按照提供的shopping list加入了指定份数的商品。
* Minor Error：发生了以下的任意一种情况：
  + 被试漏掉了一件商品；
  + 被试加入了一件购物清单上不存在的商品；
  + 被试有一件的商品加入购物篮的份数不对；
* Major Error：两种及以上的Minor Error组合，或者以下的任何情况：
  + 被试漏掉了两件及以上的商品；
  + 被试加入了两件购物清单上不存在的商品；
  + 被试有两件及以上的商品加入购物篮的份数不对；

观察到常见的Minor Error，单件商品遗漏或加入份数不对，除了被试本人的粗心大意（没有一一核对shopping list和basket中的商品）之外通常是由于被试点击添加按钮未点中或重复点击（在手机上更容易误触）。Major Error除了多次此类小失误的叠加还有一种较为严重的情形——被试误以为在旧版网站中，点击加号添加商品至指定份数后，还要再点一次“Add Basket”才完成加购操作，导致该被试最后添加到basket中的商品每种都比shopping list上多了一件。

* User satisfaction: (Satisfaction)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Condition | Prototype + Computer | Prototype + Mobile Phone | Our Version + Computer | Our Version + Mobile Phone |
| Excellent | 6 | 4 | 9 | 10 |
| Acceptable | 3 | 6 | 1 | 0 |
| Unacceptable | 1 | 0 | 0 | 0 |

当被问及整体使用体验时，绝大多数被试，无论使用的是电脑还是手机，认为新版网站的使用体验excellent，而在旧版网站上，只有半数用户选excellent，其中电脑略多于手机，其他人大多数选择acceptable，还有一人认为unacceptable。

实验结果（定性分析/case study）：

从每组选出具有代表性（完成任务用时最接近平均值）的被试进行分析：

被试A：使用Prototype + Computer

实验阶段1观察：

1. 被试尝试按“+”、“-”添加和移除商品，抱怨按钮重叠；
2. 被试连续点击了三次“Contact”链接，询问为什么没有反应；
3. 被试再次连续点击三次“About”链接没有反应；
4. 被试继续试图点击“Welcome, Test User”和“Checkout”；
5. 被试进入Checkout页面，点击“Name”输入框，询问为什么不能输入；
6. 被试点击“Pay by Credit Card”进入付款界面后停滞了几秒，随后连续点了两次浏览器上方的“返回”键回到网站主页面；
7. 被试尝试搜索“apple”，商品显示，随后被试试图通过再点一次“search”来取消搜索，搜索框消失，被试说“嗯？回不去了！”

实验阶段2：

被试通过搜索框依次搜索（只输入单词的前几个字母）每一件商品，按“+”加入购物车（注意到被试不清楚“Add Basket”按钮的用途，总是在按“+”获得正确的份数后又点“Add Basket”，以至于每种商品都比指定数量多加了一件）。

被试在完成任务过程中不得不需要实验者的帮助：搜索商品名称后，整个网页此时只有这一件商品，此时商品的“Add Basket”、“+”和数字输入框都被Cookie Message遮挡，即使拖动到网页最下方也无法正常显示。该版本Cookie Message的“Accept”与背景同色，此时几乎没有一个被试能找到并关闭浮窗，导致实验无法继续进行，实验者必须介入，帮助被试关掉Cookie Message。

结论：

1. 网站看似提供但实际上并没有实现的功能会对用户造成困扰，如点击没有反应的“About”、“Contact”链接和设置为禁用的Customer Detail输入框；
2. 网站相邻商品的“+”和“-”重叠，对用户点击按钮制造了困难；
3. 网站在checkout和payment页面之间缺乏好的返回机制；
4. 网站的搜索框功能较令人费解，用户并不总是清楚如何使它出现、隐藏、在搜索完毕后如何取消搜索，重新显示全部商品等；
5. 网站的“Add Basket”按钮功能与“+”重复，容易使用户产生误会；
6. 网站的Cookie Message找不到“Accept”按钮无法关闭，遮挡屏幕，很有可能会妨碍用户进行基本操作，对网站的可用性有严重影响。

可见这些问题我们都已在优化过程中提出并一一解决过了。

被试B：使用Prototype + Mobile Phone

实验阶段1观察：

1. 被试同样反复试图点击“Contact”、“About”等；
2. 被试试图搜索“strawberry”，但由于网站上的商品名称是“strawberries”，搜索并没有成功，被试改搜“potato”，这一次成功地显示了商品；
3. 被试在点击按钮和打开输入框时屏幕经常自动放大，导致用户不得不又用手指拖动缩放回正常大小，从而继续浏览，对此被试抱怨“真麻烦；

实验阶段2观察：

被试上下划动页面并用肉眼寻找每一件商品，并抱怨字太小找起来眼睛很累。找到每件商品后，被试连续按“+”添加指定份数至购物篮，点击时网页仍然经常局部缩放。此外，当被试想要添加位于最后一行的商品cherries时，商品下方按钮再次被遮挡，无法点击，实验者不得不再次介入关闭Cookie Message。

结论：

1. 网站的搜索功能不支持对同一商品的单复数形式统一对待（类似地，也不支持模糊搜索或拼写错误），这在实际应用中很不利，因为用户并不总是能按网站设置的商品名称进行搜索，可能会误以为网站根本没有这件商品；
2. 网站在手机上字体过小，不利于阅读，而且交互时的缩放降低了使用体验。

其中第二点我们已经进行了一定的改进，第一点还可以继续优化。

被试C：使用Our Version + Computer

实验阶段1观察：

1. 被试第一步就点击“Accept & Close”关闭了Cookie Message；
2. 被试点进“Pumpkins”下方的输入框，试图输入“1000”，出现弹窗“Insufficient stock: only 90 left”，被试重新输入“90”；
3. 被试点进“Basket”，点击Pumpkins右侧的“X”将该商品删除。

实验阶段2观察：

被试通过搜索框（只输入单词的前几个字母）搜索每件商品，并通过商品下方的输入框输入件数，此时用户抱怨（当输入份数合法时）按下Enter没有提示说明加入购物篮成功。在进入Checkout页面前，用户试图检查Basket中的商品及份数是否和任务提供的shopping list匹配，此时用户抱怨Basket内商品的排列没有规律，即不是按拼写也不是按添加的顺序，以至于很难一一检查。

结论：

1. 像这位被试一样，大多数用户在使用新版网页时进行的第一项操作都是关闭Cookie Message，可见去除浮窗的遮挡确实是一项重要改进；
2. 当用户的商品份数输入不合法时，弹出提示信息是有必要的。然而，当输入合法，除了右上角Basket的数字变化，用户需要更多反馈确认加购成功；
3. 在进入Checkout页面之前，用户可以方便地在Basket弹窗中核对和修改加购的商品，这对任务的执行是有帮助的。然而，Basket内商品的排序缺乏规律，未来还应该提供按加入时间、商品名称、数量、价格等排序的支持，方便用户在checkout前更好地查看和核对商品。

被试D：Our Version + Mobile Phone

实验阶段1观察：

1. 被试点击help打开弹窗后，并未将内容全部读完，而是试图通过再点一次help来关闭弹窗（失败），最终只能拖到弹窗底部点close。

实验阶段2观察：

被试依次对三个分类进行筛选，划动肉眼寻找商品，连续按“+”添加商品到购物篮；对Basket进行核对时发现milk遗漏了，通过搜索找到该商品加入。注意到即使同样是肉眼搜索，提供Filter对搜索速度有所提高。此外，用户在使用新版网页时，手指点击按钮没有触发页面缩放，只在搜索框输入时仍会触发。

结论：

1. 新版网页实现了字体、按钮大小等的自适应，提供更流畅的使用体验；
2. 用户并不总是有耐心读完全部内容再点击最下方的close关闭弹窗，未来有必要在弹窗右上角添加“X”按钮，并支持用户再次点击链接关闭弹窗；
3. 注意到用户在使用电脑执行任务时更倾向于使用搜索框寻找商品，而使用手机的用户更倾向于上下拖动网页并肉眼查找，可能的原因有：
   1. 手机端输入框小，键盘小，输入麻烦；
   2. 手机端每页能容纳的商品数量少，用户在浏览过程中回到页面顶部使用搜索框需要大量翻页（观察到极少人会使用“Top”按钮返回顶部）。

在未来的开发中，或许应该将搜索框也置于顶部悬浮navigation bar中，使得用户无论向下翻页浏览至何处都能随时方便地使用。