

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»»
Факультет компьютерных наук
Департамент программной инженерии

Домашнее задание
по курсу «Эргономика человеко-машинного взаимодействия»
по направлению подготовки бакалавров
09.03.04 «Программная инженерия»

Выполнили:
студенты группы БПИ153
образовательной программы
09.03.04 «Программная инженерия»
Плесовских А.И.
Нагаев А.О.

Москва 2018

Оглавление

Приложение.....	3
Предполагаемые изменения.....	4
Тестируемые версии приложения	4
Гипотеза.....	5
Переменные	5
Планирование эксперимента	6
Обоснование выбора стиля тестирования	7
Статистические тесты.....	7
Исследование	8
Проведение исследования.....	8
Результаты исследования	9
Корреляционный анализ	10
Проверка гипотезы о времени	11
Проверка гипотезы о эффективности	12
Заключение.....	13
Ссылки.....	14

Приложение

Для исследования было выбрано приложение для Апрельской международной ежегодной конференции (AIAC). Данное приложение используется для просмотра расписания конференции, добавления отдельных сессий в избранное с целью составления своего собственного расписания посещения выбранных мероприятий.

Исследование призвано показать, какой интерфейс отдельных элементов наиболее удобен для пользователей, а также позволяет выполнять необходимые действия быстрее, эффективнее и без ошибок.

Предполагаемые изменения

В данном исследовании мы планируем изучить изменение эффективности и скорости выполнения операций поиска сессий. В приложении мы сделаем новый интерфейс экрана поиска – все сессии будут упорядочены по датам на одном экране, без необходимости переключаться между датами.

Тестируемые версии приложения

В качестве основной версии взяли старый экран поиска. Его отображение в приложении представлено на рис. 1.

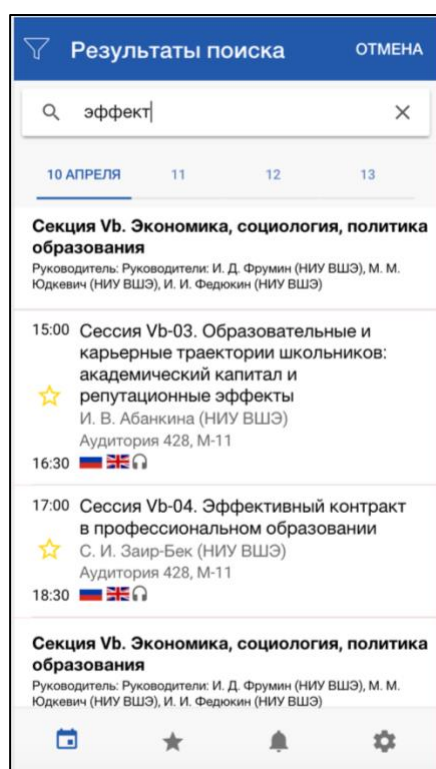


Рис.1. Старый интерфейс экрана поиска

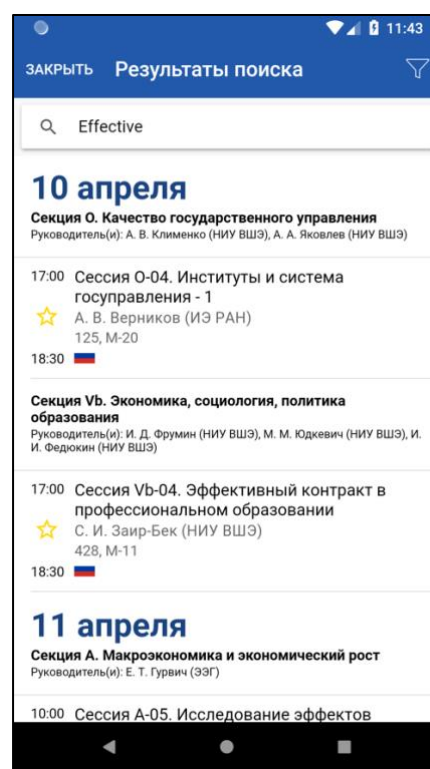


Рис.2.Новый интерфейс экрана поиска

В новой версии экрана поиска (рис.2) мы изменили отображение списка сессий. Теперь все сессии упорядочены на одном экране по датам. Данные изменения были сделаны путём изменения программного кода приложения.

Гипотеза

Нулевая гипотеза – новый интерфейс экрана поиска не позволяет пользователю быстрее и эффективнее искать сессии, то есть должно наблюдаться равенство средних значений времени поиска и количества правильно выполненных заданий на разных группах испытуемых.

Альтернативная гипотеза – новый интерфейс экрана поиска позволяет пользователю быстрее и эффективнее искать сессии.

Переменные

Независимые:

- Используемая испытуемым для тестирования версия экрана поиска, категориальный признак.

Зависимые:

- Время выполнения серии заданий испытуемым, номинальный признак, измеряется системой аналитики.
- Процент успешно выполненных заданий пользователем

Контрольные:

- Пол
- Возраст

Планирование эксперимента

Самый оптимальный вариант для сравнения двух версий экрана поиска - A/B тестирование. В данном случае мы каждому участнику тестирования предложим старый и новый дизайн экрана поиска в случайном порядке, и предложим им найти определенные сессии в приложении.

Для эксперимента выбираем 20 испытуемых, из числа студентов и преподавателей университетов, как наиболее близких к портрету пользователя нашего приложения. Испытуемых разбиваем на две равные группы, каждая будет использовать свою версию программы первой, то есть первой группе сначала предлагаем старый интерфейс, а затем новый, а второй группе наоборот.

Каждому испытуемому выдаем одинаковую для всех серию заданий по поиску отдельных сессий и добавлении их в избранное.

Для каждого человека замеряем время выполнения серии заданий, и процент выполненных заданий, затем считаем среднее по группе для дальнейшего сравнения. Также в набор заданий добавлены сессии, которые нельзя найти, то есть попытка поиска и ответ «нет такой сессии» допустим, тем более, что пользователь часто сталкивается с этим при использовании приложения. Дополнительно это позволит исследовать эффективность нового интерфейса на предмет успешности выполнения заданий.

(Время засекается с помощью программных средств, встроенных в приложение для этого эксперимента.)

Список заданий

Для всех сессий необходимо найти её через поиск и добавить в избранное.

1. Сессия «Общественные палаты», 11 апреля
2. Сессия «Бюджетная политика и управление», 10 апреля
3. Доклад «Инвестирование в цифровую экономику» *
4. Сессия «Российские торгово-инвестиционные связи», 11 апреля
5. Доклад «Устаревает ли военная сила?», 13 апреля
6. Сессия «Медиа и социология» *
7. Сессия «Право: новые риски и результаты», 13 апреля

**не существует*

Обоснование выбора стиля тестирования

Мы считаем, что дизайн within-subjects в нашем случае будет оптимален, так как у нас небольшая группа испытуемых и нам необходимо, ее увеличить и уменьшить дисперсию ошибки, так как тесты будут проходить при абсолютно одинаковых условиях для каждого человека. Стоит отметить, что при данном способе проведения теста большое влияние на результат оказывает усталость человека, но данную отрицательную черту мы попытаемся исправить, давая 50% испытуемых сначала новый дизайн, а потом старый дизайн сайта, а другой половине сначала старый, а потом новый.

Статистические тесты

В данной работе мы будем сравнивать среднее время (продолжительность) выполнения серии заданий на старой и новой версии приложения, а также успешность выполнения заданий (количество), поэтому для проверки нашей гипотезы мы будем использовать t-Критерий Стьюдента.

Исследование

Проведение исследования

После проведения эксперимента мы получили данные, представленные в файле dataset_all.csv, где:

- index: порядковый номер испытуемого (от 1 до 20)
- first_old: порядок тестирования версий (0 – сначала новая версия, 1 – сначала старая версия)
- time_old, time_new: время выполнения задания
- task_old, task_new: количество выполненных заданий (целое число, от 0 до 7, всего 7 заданий было предложено).
- age: возраст в годах
- gender: пол (1 - мужчина, 0 - женщина)

В данном исследовании выбросов не было обнаружено. На полученных данных были проведены следующие тесты:

- построены графики рассеивания данных по времени, приведены интерпретации результатов
- рассчитана корреляция между временем выполнения задания и возрастом испытуемых;
- проведён тест Стьюдента между данными времени с новой версией экрана поиска и со старой версией;
- проверена гипотеза о эффективности (количество успешно выполненных заданий) с помощью t-Статистики.

Результаты исследования

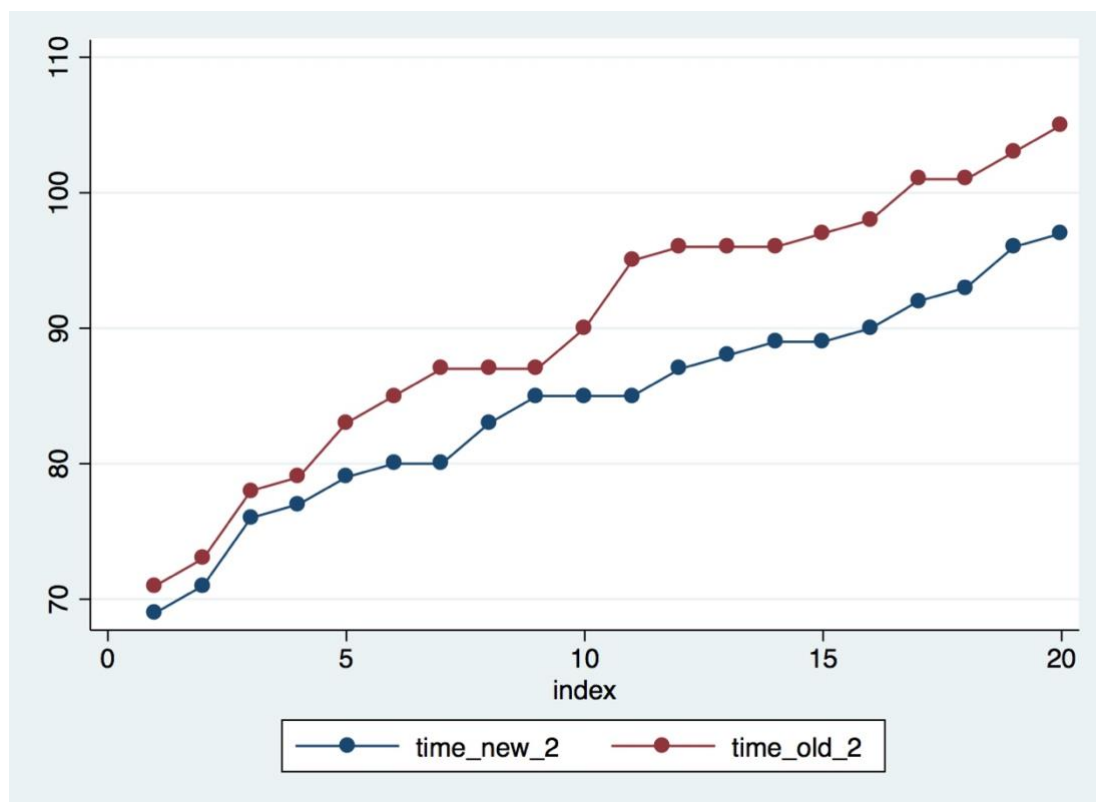


Рис. 3. Отсортированный график рассеивания времени

На рисунке 3 изображен график рассеивания данных, где по оси Y - это время, которое респонденты затратили на выполнение набора заданий, в секундах. Данные обозначенные красным были замерены на старой версии, синим - на новой. По данному графику видно, что в данных нет выбросов, а также, что в среднем значения измерений для новой версии ниже, чем для старой.

Корреляционный анализ

Кроме того, мы проверили взаимосвязь между временем выполнения заданий и возрастом испытуемых, а также между количеством успешно выполненных заданий и возрастом испытуемых.

```
. pwcorr age time_old time_new task_old task_new
```

	age	time_old	time_new	task_old	task_new
age	1.0000				
time_old	0.3303	1.0000			
time_new	0.3709	0.0680	1.0000		
task_old	-0.1642	0.2026	-0.2394	1.0000	
task_new	-0.1929	0.2464	-0.3061	0.8808	1.0000

Рисунок 4. Корреляция Пирсона

Нам не удалось выявить положительную корреляцию между этими переменными на 5% уровне значимости, таким образом возраст испытуемого не влияет на скорость или успешность выполнения заданий. Это можно объяснить высоким уровнем подготовки всех испытуемых, так как это студенты, аспиранты и преподаватели университетов.

Проверка гипотезы о времени

Теперь проверим различия в двух группах с помощью t-теста Стьюдента. Будем сравнивать время, затраченное на поиск сессий на новой и старой версии приложения. Сравнение показало, что время поиска на новой версии (mean = 84.55, n = 20) и на старой версии (mean = 90.4, n = 20) имеет значимые различия (t = 2.07, p-value = 0.045).

Следовательно, мы отклоняем нулевую гипотезу о равенстве среднего значения времени на новой и старой версии сайта на 5% уровне значимости. И принимаем альтернативную гипотезу о том, что в новом интерфейсе время уменьшилось (P-value – 0.02) на 5% уровне значимости.

```
. ttest time_old ==time_new, unpaired
```

Two-sample t test with equal variances

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
time_old	20	90.4	2.233006	9.986306	85.72626	95.07374
time_new	20	84.55	1.725162	7.715159	80.93919	88.16081
combined	40	87.475	1.469339	9.292918	84.50298	90.44702
diff		5.85	2.82179		.1375844	11.56242

```
diff = mean(time_old) - mean(time_new)          t = 2.0732
Ho: diff = 0                                   degrees of freedom = 38
```

```
Ha: diff < 0                                Ha: diff != 0                                Ha: diff > 0
Pr(T < t) = 0.9775                        Pr(|T| > |t|) = 0.0450                        Pr(T > t) = 0.0225
```

Рисунок 5. Проверка гипотезы о времени с помощью t-Статистики.

Проверка гипотезы о эффективности

А теперь сравним количество успешно выполненных заданий на новой и старой версии приложения. Сравнение показало, что количество выполненных заданий на новой версии (mean = 6.55, n = 20) и на старой версии (mean = 5.9, n = 20) имеет значимые различия (t = -2.035, p-value = 0.0488).

Следовательно, мы отклоняем нулевую гипотезу о равенстве среднего значения выполненных заданий на новой и старой версии сайта на 5% уровне значимости. И принимаем альтернативную гипотезу о том, что в новом интерфейсе количество ошибок уменьшилось (P-value – 0.024) на 5% уровне значимости.

```
. ttest task_old==task_new, unpaired
```

Two-sample t test with equal variances

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
task_old	20	5.9	.2605662	1.165287	5.354629	6.445371
task_new	20	6.55	.1846048	.8255779	6.163618	6.936382
combined	40	6.225	.1659761	1.049725	5.889282	6.560718
diff		-.65	.3193332		-1.296456	-.0035438

diff = mean(task_old) - mean(task_new)

t = -2.0355

Ho: diff = 0

degrees of freedom = 38

Ha: diff < 0

Ha: diff != 0

Ha: diff > 0

Pr(T < t) = 0.0244

Pr(|T| > |t|) = 0.0488

Pr(T > t) = 0.9756

Рисунок 6. Проверка гипотезы с помощью t-Статистики.

Заключение

В рамках полученных данных наши изменения (переработка отображения результатов на экране поиска) позволили сократить время поиска и уменьшить количество ошибок, совершаемых пользователями.

Нулевая гипотеза отвергнута в пользу альтернативной на уровне значимости в 5%.

Также в данном исследовании была исследована зависимость между возрастом респондентов и временем выполнения заданий. Корреляция между этими переменными оказалась незначимой на 5% уровне значимости. Что объясняется высоким уровнем образования и подготовки испытуемых.

Можно предположить, что приведенные нами изменения могут положительно сказаться на времени поиска и эффективности для всех пользователей приложения, так как выбор респондентов для исследования осуществлялся с учетом особенностей портрета целевого пользователя нашего приложения.

Ссылки

1. XIX April Conference. [Электронный ресурс]// URL: <https://conf.hse.ru/en/2018/>
2. Google Play. Андроид приложение «XIX April Conference» [Электронный ресурс]// URL: <https://play.google.com/store/apps/details?id=ru.hse.AIAC>
3. App Store. IOS приложение XIX April Conference» [Электронный ресурс]// URL: <https://itunes.apple.com/ru/app/april-conf/id1357616201>
4. Wikipedia. A/B-тестирование. [Электронный ресурс]// URL: <https://ru.wikipedia.org/wiki/A/B-тестирование>
5. Psychology World. Within-subject Design. [Электронный ресурс]// URL: https://web.mst.edu/~psyworld/within_subjects.htm
6. Wikipedia. Критерий Стьюдента [Электронный ресурс]// URL: https://ru.wikipedia.org/wiki/T-Критерий_Стьюдента