

УДК 004.58

СРАВНЕНИЕ СИНХРОННОГО И АСИНХРОННОГО МЕТОДОВ ЮЗАБИЛИТИ-ТЕСТИРОВАНИЯ

Сорокин М.В. (Университет ИТМО)

Научный руководитель – доцент, кандидат технических наук Балканский А.А.
(Университет ИТМО)

Введение. В рамках работы над НИР была определена тема будущей магистерской работы. Были рассмотрены статьи, посвященные сравнению методов юзабилити-тестирования. В статье, посвященной сравнению синхронного и асинхронного дистанционного тестированию [1] исследователи выявили, что для респондентов из Саудовской Аравии асинхронное тестирование является более эффективным, чем использование синхронного метода. В качестве подтверждения они приводят различные статьи, в которых обнаружилось, что респонденты из таких стран как Саудовская Аравия предпочитают высказывать косвенную критику из-за вежливости [2]. Далее, они отмечают, что данные результаты являются противоположными тем, что были получены в схожем исследовании, проведенном на западной аудитории [3]. В данном исследовании были получены результаты, что синхронное тестирование является более эффективным чем асинхронное тестирование и позволяет обнаружить больше проблем с удобством использования. Кроме этого, в рамках данного исследования сравнивались лабораторное и синхронное дистанционное тестирование. В ходе исследования было получено, что два этих метода показывают схожие показатели в количестве обнаруженных ошибок и практически не имеют отличия. При этом в другой статье, посвященной сравнению дистанционного, лабораторного и очного тестирования без лабораторных условий были получены отличающиеся результаты [4]. Все это натолкнуло на вывод, что наиболее эффективный метод юзабилити-тестирования для респондентов из разных стран может отличаться в связи с культурными особенностями. В связи с этим было принято решение провести аналогичное исследование на русскоговорящих респондентах с целью выявления наиболее эффективного метода.

Основная часть. В эксперименте приняло участие 80 респондентов. Все респонденты являются русскоговорящими, так как цель исследования выявить наиболее эффективный метод для русскоговорящей аудитории. В качестве сравниваемых методов были выбраны синхронное и асинхронное дистанционное тестирование [5]. Исследование состояло из двух этапов. Первым этапом был пилотный эксперимент, в рамках которого были поставлены гипотезы, а также определен размер выборок для подтверждения каждой из них. Так же был определен план эксперимента. Каждый респондент проходит тестовое задание на двух сайтах и с использованием обоих методов. Порядок сайтов и методов у респондентов отличался. После проведения пилотного эксперимента начался основной. Всего в нем поучаствовало 80 респондентов. Собирались такие данные как время прохождения всех заданий, количество ошибок и уровень знакомства с сайтом. Каждый из респондентов проходит тестирование с использованием обоих методов на двух разных сайтах, аналогично пилотному эксперименту. После того, как все респонденты прошли тестовые задания был начат анализ результатов. Были определены статистические данные такие как мощность, вероятность групповой ошибки, p -value, а также необходимый размер выборки для проверки каждой из гипотез. Также был определен коэффициент корреляции для двух гипотез о корреляции. Затем для всех гипотез была проведена обработка методом Холма-Бонферрони для проверки статистической значимости поставленных гипотез.

Выводы. В ходе проведенного эксперимента были проверены 3 гипотезы. Одна гипотеза о влиянии и две о корреляции. Для расчетов был принят групповой уровень значимости 1.5%. Для всех гипотез получились статистически значимые результаты. Основная гипотеза о

влиянии наблюдателя в синхронном методе на количество обнаруженных ошибок получила подтверждение согласно конкурирующей гипотезе на уровне значимости $p\text{-value} = 0.0019\%$ при $\alpha = 0.75\%$. Данная гипотеза проверялась с использованием тестов знаков. Были получены 63 ненулевые разности, из которых 48 положительных и 15 отрицательных. Еще 17 оказались нулевыми. Для проверки использовалась левосторонняя проверка. Гипотезы о корреляции также получили подтверждение согласно конкурирующим гипотезам. У первой гипотезы о корреляции между временем прохождения тестового задания и количеством обнаруженных ошибок прослеживается небольшая корреляция в сторону увеличения количества обнаруженных ошибок при увеличении времени прохождения задания. Данная гипотеза проверялась с помощью рангового метода Спирмана и левосторонней проверки. Уровень значимости $p\text{-value} = 0.0175\%$ при $\alpha = 1.5\%$, коэффициент корреляции $r = 0.285$. Третьей проверяемой гипотезой была гипотеза о существовании корреляции между временем прохождения задания и уровнем знакомства с сайтом. Аналогично предыдущей гипотезе она проверялась с использованием рангового метода Спирмана и левосторонней проверки. Наблюдается умеренная корреляция между временем прохождением задания и уровнем знакомства с сайтом. Уровень значимости $p\text{-value}$ оказался меньше, чем $2 * 10^{-24}\%$ при $\alpha = 0.5\%$, коэффициент корреляции $r = 0.678$. Таким образом, по результатам эксперимента можно сделать вывод, что для русскоговорящей аудитории использование синхронного метода юзабилити тестирования позволит определить больше ошибок в интерфейсе, чем с использованием асинхронного метода. Следовательно, если при исследовании того или иного интерфейса на русскоговорящей аудитории необходимо определить как можно больше ошибок, стоит использовать синхронный метод.

Список использованных источников:

1. O. Alhadreti. A Comparison of Synchronous and Asynchronous Remote Usability Testing Methods //International Journal of Human-Computer Interaction, vol. 38, no. 3. – 2022. – С. 289–297. <https://doi.org/10.1080/10447318.2021.1938391>.
2. M. Hall, M. de Jong, and M. Steehouder. Cultural differences and usability evaluation: Individualistic and collectivistic participants compared //Technical Communication, vol. 51, no. 4. – 2004. – С. 489–503.
3. M. S. Andreasen, H. V. Nielsen, S. O. Schroder, and J. Stage. What Happened to Remote Usability Testing? An Empirical Study of Three Methods //Proceedings of the 2007 Conference on Human Factors in Computing Systems. – 2007. – С. 1624. <https://doi.org/10.1145/1240624.1240838>.
4. J. Sauer, A. Sonderegger, K. Heyden, J. Biller, J. Klotz, and A. Uebelbacher. Extralaboratorial usability tests: An empirical comparison of remote and classical field testing with lab testing //Applied Ergonomics, vol. 74. – 2019. – С. 85–96. <https://doi.org/10.1016/j.apergo.2018.08.011>.
5. J. M. C. Bastien. Usability testing: a review of some methodological and technical aspects of the method //International Journal of Medical Informatics, vol. 79, no. 4. – 2010. – С. 18–23. <https://doi.org/10.1016/j.ijmedinf.2008.12.004>.