



Grupo 3

BarlSTa

Testes com Utilizadores

Guião de Testes

Objetivo: saber se atingimos os objetivos estabelecidos no início do desenvolvimento, isto é, os critérios de usabilidade para cada tarefa.

Onde: numa sala meio-cheia, tentando simular um ambiente de convívio de bar.

Quando: no dia 17/05/2016, entre as 14:00 e as 16:30.

Duração: Aproximadamente 10 minutos, não ultrapassando os 15 minutos.

Equipamento: portátil com ecrã tátil para melhor replicar situação real.

Software: apenas é necessário o nosso protótipo e um browser, para o correr.

Estado: os utilizadores começarão no menu principal, sem produtos no carrinho.

Tempo de resposta: carga normal, sem anomalias em tempo de resposta.

Coordenador e Observador: Nuno Anselmo e André Mendes, respetivamente.

Utilizadores: 15 alunos do IST, faixa etária entre os 18 e 23 anos, e preferencialmente com igual distribuição entre os sexos (8 homens e 7 mulheres ou vice-versa).

Tarefas: as tarefas específicas estabelecidas previamente, de “pedir”, “votar” e “medir”.

Fim correto: a tarefa está terminada quando o utilizador a completar.

Ajuda: o utilizador apenas pode utilizar a ajuda presente no protótipo.

Ajudar: não será dada qualquer ajuda durante os testes para além de relembrar as tarefas ainda por executar.

Dados: serão recolhidas todas as medidas de usabilidade referidas previamente, para analisar e realizar testes estatísticos.

Sucesso: consideraremos que tivemos sucesso caso todos os objetivos sejam cumpridos, e caso existam dados suficientes para suportar reprodutibilidade dessa conclusão

Análise da tarefa de “Pedir”

Tarefa: pedir 2 gins e 1 hamburger, somando um total de 54 euros.

Eficácia: a nossa medida de eficácia foi o sucesso na realização do pedido corretamente, tendo como objetivo que 80% dos utilizadores conseguissem efetuá-lo. Este objetivo foi cumprido, com 100% dos utilizadores a terem conseguido efetuar a tarefa com sucesso. Concluímos, portanto, que o nosso critério foi demasiado pouco restrito, eliminando apenas a possibilidade de a interface ser inutilizável, e não nos dando informação útil que permita a sua melhoria.

Eficiência: a nossa medida de eficiência foi o tempo necessário para realizar a tarefa, tendo como objetivo um tempo de 20 segundos. Os nossos testes tiveram como resultado um tempo médio de 16.5 segundos, e como desvio padrão 5.48 segundos. O intervalo de confiança (confiança de 95%) é de 16.5 ± 3.03 segundos, estando abaixo do nosso valor objetivo, podendo dizer que este foi cumprido com sucesso.

Satisfação: a nossa medida de satisfação foi a fluidez no processo de pedir, sendo avaliada numa escala de 1 a 5, e tendo como objetivo que a percentagem de classificações abaixo de 3 fosse inferior a 10%. Todas as classificações obtidas foram superiores a 3, tendo como moda e mediana o valor de 5, sendo que concluímos que o objetivo foi cumprido.

Análise da tarefa de “Entretenimento”

Tarefa: votar na música “Menina Surfista” dos “The Beach Boys”, para que esta seja adicionada à playlist e para que mais tarde seja escolhida ou rejeitada.

Eficácia: a nossa medida de eficácia foi o número de ecrãs percorridos, tendo como objetivo o número 2, e sendo utilizada para detetar erros. Tal como na medição de eficácia da tarefa de pedir, todos os utilizadores executaram a tarefa sem cometer erros, com 2 ecrãs percorridos, não conseguindo, portanto, extrair conclusões úteis para além de que os testes efetuados não parecerem ter colocado os utilizadores numa situação de limite propícia a erros.

Eficiência: a nossa medida de eficiência foi o tempo necessário para realizar a tarefa, tendo como objetivo um tempo de 15 segundos. Os nossos testes tiveram como resultado um tempo médio de 10.1 segundos, e como desvio padrão 4.02 segundos. O intervalo de confiança (confiança de 95%) é de 10.1 ± 2.23 segundos, encontrando-se sempre abaixo do tempo limite, tendo cumprido o objetivo.

Satisfação: a nossa medida de satisfação foi a facilidade com que os utilizadores encontraram a música que procuravam, sendo avaliada numa escala de 1 a 5, e tendo como objetivo que a percentagem de classificações abaixo de 3 fosse inferior a 10%. Todas as classificações obtidas foram superiores a 3, tendo como moda e mediana o valor de 4, sendo que concluímos que o objetivo foi cumprido.

É de notar que alguns dos dados recolhidos quanto à tarefa de entretenimento, e quanto à tarefa de pedir, se aplicam a ambos, pois a estrutura de menus é igual, bem como todo o design. Assim, podem-se desenhar paralelos entre os problemas descobertos numa funcionalidade e problemas existentes (mesmo que não descobertos durante os testes) na outra funcionalidade.

Análise da tarefa de “Medições”

Tarefa: medir o nível de álcool no sangue, e averiguar se está ou não apto para conduzir.

Eficácia: a nossa medida de eficácia foi a necessidade de pedir ajuda ou carregar no botão de ajuda para completar a tarefa. Tal como na medição de eficácia das tarefas anteriores, todos os utilizadores executaram a tarefa com sucesso, sem necessitar de ajudas. Este é o resultado esperado, pois a funcionalidade testada nesta tarefa é simples, sem margem para erros.

Eficiência: a nossa medida de eficiência foi o tempo necessário para realizar a tarefa, tendo como objetivo um tempo de 15 segundos. Os nossos testes tiveram como resultado um tempo médio de 8.1 segundos, e como desvio padrão 1.12 segundos. O intervalo de confiança (confiança de 95%) é de 8.1 ± 0.62 segundos, encontrando-se muito abaixo dos tempos esperados e cumprindo o objetivo com grande margem.

Satisfação: a nossa medida de satisfação foi quão direta e fácil de utilizar foi a funcionalidade, sendo avaliada numa escala de 1 a 5, e tendo como objetivo que a percentagem de classificações abaixo de 3 fosse inferior a 10%. Todas as classificações obtidas foram superiores a 3, tendo como moda e mediana o valor de 5, sendo que concluímos que o objetivo foi cumprido.

Conclusão

Concluímos, após os testes, que no geral estes foram feitos com sucesso, e que os critérios foram também atingidos com sucesso.

Os testes de eficácia para todas as tarefas efetuadas foram demasiado pouco específicos, fazendo com que as nossas conclusões fossem pouco informativas, limitando-se apenas a concluir que a interface apresenta ser funcional e utilizável, não tendo sido possível encontrar problemas com o auxílio desta medição.

Os testes de eficiência foram os que nos deram informação mais útil, não pelos resultados obtidos que cumpriram os nossos objetivos, e que não nos deram muita informação útil, mas sim ao observar os utilizadores durante a utilização do protótipo, descobrindo onde estes perderam mais tempo, e onde se poderia melhorar a interface para aumentar a fluidez com que esta pode ser utilizada.

Os testes de satisfação revelaram que os utilizadores ficaram satisfeitos com a utilização do nosso protótipo, não sendo possível concluir nada para além de que não existem problemas que tenham afetado os utilizadores de maneira notável.

Ao se observar os utilizadores, e também com as sugestões dos mesmos, ficámos com um conjunto de melhorias a realizar, sendo que destacamos que ao utilizarem o menu da playlist, em que tinham a tarefa de votar numa dada música, praticamente todos os utilizadores ficaram sem saber se a tarefa tinha sido concluída com sucesso, por não existir nenhum feedback. Parámos o cronómetro quando os utilizadores completaram a tarefa, não sendo algo que afetou os resultados, mas foi algo que notámos.

Para além disso, notámos também a pouca utilização da funcionalidade de pesquisa, bem como da funcionalidade de *sorting*, sendo que quase todos os utilizadores encontraram os produtos apenas percorrendo toda a lista de produtos (nomeadamente, a lista de bebidas e a lista de comidas).