Combinação de técnicas de interação humano-computador para avaliação de sistemas de recuperação de imagens por conteúdo

Ana Lúcia Filardi

Embrapa Monitoramento por Satélite
Av. Soldado Passarinho, 303
13070-115 - Campinas, SP - Brazil
+55 19-3211-6200

ana@cnpm.embrapa.br

Agma Juci Machado Traina

Universidade de São Paulo
Av. do Trabalhador São-Carlense, 40
13560-970 - São Carlos, SP - Brazil
+55 16-3373-9693

agma@icmc.usp.br

ABSTRACT

The human interaction is a valuable tool to evaluate the performance of a Content-based Image Retrieval (CBIR) system regarding its efficiency and efficacy, as well as the user satisfaction. The fast growing and the strength achieved in the CBIR research field has brought relevant improvement to the pictorial information retrieval, but it is not supported by techniques or guidelines to help evaluating the usability of such systems. Thus, a challenge of CBIR systems is to provide the users intuitive ways to ask queries and to receive their answers increasing the usefulness of such systems. The goal of this paper is to present an approach to evaluate the human-computer interaction of CBIR systems. To do so, we employed a combination of evaluation techniques considering the final users and the human factors that influence their work, taking also into account the large amount of images and data stored in these systems. As a case study, we used a CBIR system for medical images, which was developed aimed at supporting the decision making when the physician is drawing the diagnosis. Our study aimed at improving the diagnoses, as well as the evolution of the patients' treatment, making both more precise and efficient.

RESUMO

A interação humana pode contribuir efetivamente para avaliar o desempenho de sistemas de recuperação de imagens baseada em conteúdo (*Content-based Image Retrieval* - CBIR), tanto em relação à eficiência e eficácia, quanto a satisfação do usuário. O crescimento, aprimoramento e a excelência da pesquisa em técnicas CBIR, avaliados pela sua performance na recuperação de informações, contradizem-se com a ausência de procedimentos específicos de avaliação por meio da percepção visual do usuário. Um dos desafios decorrentes desses sistemas consiste na necessidade em prover aos usuários uma interface de alta qualidade na formulação de perguntas e exibição dos resultados. O objetivo deste trabalho foi avaliar a interação

humano-computador para sistemas CBIR que atuam em modo compartilhado usando intranet, aplicando uma combinação de técnicas de avaliação, levando em consideração o usuário final e os fatores humanos que influenciam o trabalho com grandes volumes de dados e imagens. Tal avaliação foi baseada em um estudo de caso aplicado à medicina, visando tornar o diagnóstico médico e a evolução dos tratamentos cada vez mais precisos e eficientes.

Categories and Subject Descriptors

H.5.2 [Information Systems]: User Interfaces – evaluation/methodology, user-centered design.

General Terms

Design, Human Factors.

Keywords

Human-computer interaction, HCI, content-based image retrieval, CBIR, evaluation.

1. INTRODUÇÃO

A área de Interação Humano-Computador (*Human-Computer Interaction* – HCI) visa estudar como as pessoas interagem com os sistemas computacionais e explorar e desenvolver técnicas para projetar interfaces com o usuário de alta usabilidade. A medicina é uma das áreas beneficiadas com o estudo e interação da ciência da computação, visando tornar o diagnóstico médico e a evolução dos tratamentos cada vez mais precisos e eficientes.

Para dar suporte a diagnósticos, imagens de exames gerados por meio de equipamentos médicos (tomografia, mamografia, ultrassonografia, radiografia, etc.) têm sido incorporadas nos Sistemas de Informações Hospitalares (SIH), para consolidar em um só sistema as informações dos pacientes. O interesse em utilizar imagens nos diagnósticos vem aperfeiçoando o desenvolvimento de novas técnicas que se preocupam com a recuperação desses tipos de dados.

As consultas tradicionais sobre informações armazenadas em bases de dados são efetuadas através de uma chave de busca que permite recuperar o registro associado. As chaves de buscas de dados em bases convencionais são normalmente no formato numérico ou textual. Entretanto, nos sistemas de recuperação para

bases de imagens, a chave de busca utilizada é a própria imagem. A busca, utilizando imagens como chave, é um processo complexo, pois o propósito nem sempre é buscar a imagem exata, o que poderia ser realizado através da comparação *pixel a pixel*, e sim, buscar a imagem parecida ou similar. A consulta por similaridade recupera objetos dentro de um certo grau de similaridade e um número fixo de objetos mais similares. A comparação de imagens por similaridade é essencial para tratar dados complexos e capturar o que há de mais representativo, de modo a extrair informação que as represente de modo mais fiel.

Essas técnicas, denominadas CBIR, propiciam a recuperação de imagens baseada em seu conteúdo por meio de suas características visuais e são muito utilizadas nas áreas de processamento de imagens e base de dados.

Sistemas que utilizam as técnicas CBIR têm emergido nas últimas décadas devido à necessidade de manipular grandes volumes de dados e imagens armazenadas e ao interesse de recuperar informações visuais de maneira eficiente e precisa.

Um dos desafios decorrentes desses sistemas é o processo de validação das imagens em relação à sua semelhança. Contudo, as técnicas adotadas ainda permanecem sob o senso de análise automática de recuperação de informações. Uma das possíveis razões para isso pode ser que muitos estudos têm focado, principalmente, sobre os processos de extração de características e consultas por similaridade, sem levar em consideração o usuário final e os fatores humanos que influenciam a indexação, navegação e recuperação da imagem.

Dentro desse contexto, o estudo envolvendo o usuário caracteriza-se pela sua interação com a interface do sistema, permitindo que o usuário possa consultar imagens similares a uma dada imagem de referência e visualizar os resultados.

Para atender às necessidades particulares desse estudo, visando avaliar a funcionalidade e a usabilidade do sistema, identificar os problemas específicos e melhorar a interação humano-computador, foi adotada uma combinação de técnicas de avaliação que apresentam bons resultados em relação à performance de sistemas com grande complexidade, consistindo de uma técnica de inspeção de usabilidade e agregando testes de usabilidade.

2. MEDIDAS DE USABILIDADE

Um bom design de interface de usuário deve prover uma interação fácil, natural e atraente entre o usuário e o sistema e permitir que o usuário possa realizar suas tarefas com segurança, de maneira eficiente, eficaz e com satisfação [1, 2]. Esses aspectos são conhecidos como usabilidade, e são aplicados para medir o desempenho e a satisfação do usuário.

As medidas de desempenho e satisfação do usuário avaliam o sistema de trabalho como um todo, fornecendo informações sobre a usabilidade daquele produto dentro de um contexto específico de uso. O contexto de uso, constituído por usuários, tarefas, equipamentos (hardware, software e materiais) e o ambiente físico e social, influencia o nível de usabilidade desejado de um produto, dependendo das circunstâncias particulares nas quais o produto é usado.

Essas medidas de usabilidade são importantes para visualizar a complexidade das interações entre os usuários e às necessidades em alcançar seus objetivos. Desse modo, para determinar quanto

que um produto é usável dentro de um contexto particular é necessário medir o desempenho (eficácia e eficiência) e a satisfação dos usuários.

Para alcançar as necessidades e os objetivos dos usuários é preciso saber como fazer a transição do que pode ser feito (funcionalidade) para como deve ser feito (usabilidade). Conhecer as tecnologias também é essencial para compreender a psicologia humana e as características particulares dos usuários. Deve-se levar em consideração o aspecto físico, que relaciona o tipo de ambiente que será usado o sistema; o aspecto social, como as pessoas interagem ou compartilham as tarefas e o aspecto organizacional, que consiste na hierarquia e nas diferentes formas de trabalhar.

O modelo de design HCI deve realçar a qualidade da interação entre usuários e sistemas computacionais. Para compreender melhor a organização social e os aspectos físicos do ambiente de trabalho dos usuários é necessário extrair conhecimento sobre o objetivo e o desempenho humano em relação aos conhecimentos sobre a capacidade e a limitação dos recursos computacionais. Para isso, avaliações da interação humano-computador são extremamente necessárias para verificar se as idéias do projetista são realmente o que os usuários necessitam ou desejam.

Geralmente, a avaliação é realizada para conhecer o que os usuários desejam e minimizar os possíveis erros no projeto do sistema, identificando as dificuldades provenientes dos usuários que irão repercutir diretamente na acessibilidade ao sistema, sendo que, quanto mais o projetista estiver informado sobre seus usuários, melhor será o design de seus produtos. O design da interface deve visar à simplicidade acima de tudo, obtendo o menor número de distrações possíveis e possuir uma arquitetura de informação muito clara [3].

É essencial que os projetistas de interface saibam porque é importante avaliar, o que avaliar e quando avaliar [4]. Pode-se dizer que a avaliação possui três objetivos principais, que são [5]: avaliar a funcionalidade do sistema, avaliar o efeito da interface junto ao usuário e identificar problemas específicos do sistema.

Em geral, a avaliação tende a ocorrer durante todo o ciclo de vida do design, auxiliando na criação de um produto útil e usável. A avaliação pode ser realizada em laboratórios experimentais ou no próprio local de trabalho do usuário. Pode envolver a participação ativa do usuário ou também pode ser realizada através da análise de especialistas em avaliação de design [6].

Projetar sistemas usáveis requer o conhecimento sobre quem irá usar o sistema, o que será usado, o contexto do trabalho, o ambiente em que será usado e quais as técnicas e logísticas praticáveis. Para isso, o processo de avaliação constitui o envolvimento dos seguintes componentes:

- **Pessoas:** usuários, participantes, observadores, avaliadores e especialistas em HCI.
- **Trabalho:** tarefas ou atividades específicas ou gerais.
- **Ambiente:** consiste em uma série de circunstâncias ou condições em que o usuário trabalha, relacionado aos aspectos físicos, organizacional e social do ambiente.
- **Tecnologia:** qualquer artefato incluindo hardware, software e materiais associados com computadores e equipamentos eletrônicos auxiliares.

Qualquer tipo de avaliação possui métodos e técnicas a eles associados. Tanto um método como uma técnica de avaliação em HCI é um procedimento de coleta de dados relevante no que diz respeito ao design da interface e a postura do usuário voltado para ele [7].

3. TÉCNICAS DE INTERAÇÃO HUMANO-COMPUTADOR

Existem vários métodos e técnicas de interação humano-computador que podem ser aplicados para avaliar a funcionalidade e a usabilidade de sistemas. Alguns métodos destinados à avaliação da interface baseiam-se no julgamento de especialistas e são conhecidos como inspeção de usabilidade (*predictive evaluation*), enquanto que outros contam com a participação efetiva de usuários reais e são denominados testes de usabilidade.

3.1 Inspeção de usabilidade

Vários métodos e técnicas têm sido propostos para avaliar a interatividade de sistemas através da análise de especialistas em HCI. Muitas dessas técnicas de avaliação são analíticas e envolvem a participação de especialistas para estimar o impacto do design em relação a um usuário típico, baseando nos processos cognitivos e nos princípios de usabilidade. Por não requerer o envolvimento de usuários, esses métodos são considerados relativamente baratos. No entanto, não estimam o uso real do sistema, somente preservam os princípios de usabilidade e por isso são conhecidos como inspeção de usabilidade [6].

3.2 Testes de usabilidade

Os testes de usabilidade são métodos de avaliação centrados no usuário e tem como premissa básica que as necessidades dos usuários devem ser levadas em conta durante todo o ciclo de desenvolvimento do projeto [4]. No entanto, a participação do usuário na avaliação geralmente tende a ocorrer nos últimos estágios de desenvolvimento, onde há pelo menos um protótipo do sistema para simular a capacidade interativa do sistema. Essas abordagens de avaliação com a participação do usuário são conhecidas como testes de usabilidade [6].

4. AVALIAÇÃO DA INTERFACE

Os métodos de avaliação devem questionar como, quem e quando podem ser úteis e assim determinar quais técnicas são mais apropriadas [8]. Cerca de metade do esforço gasto no desenvolvimento é focado na interface do usuário. As avaliações são necessárias para ajudar a encontrar e eliminar problemas obtidos na criação de interfaces dos usuários, por mais simples que os sistemas possam ser. Sendo assim, técnicas de testes tradicionais muitas vezes são inadequadas. Uma solução possível para isto é a combinação de métodos de avaliação convencional e a criação de novos métodos para avaliações específicas [9].

Cada projeto requer uma forma diferente de avaliação para ser empregada dentro do seu próprio contexto, conseqüentemente a escolha do método mais apropriado implica em um estudo detalhado de cada design. Em geral, uma ou mais técnicas são adotadas para atender às necessidades específicas da interface a ser avaliada.

Para atender às necessidades particulares deste estudo, foi adotada uma combinação de técnicas de avaliação que apresentam bons resultados em relação à performance de sistemas com grande

complexidade, consistindo de uma técnica de inspeção de usabilidade, baseada no julgamento de especialistas e agregando testes de usabilidade, com a participação efetiva de usuários.

Dentre as técnicas de inspeção de usabilidade, a avaliação heurística foi escolhida por ser uma das técnicas mais populares e eficientes em detectar problemas potenciais de usabilidade, e também por ser uma técnica intuitiva, fácil e rápida de ser ensinada, além de possuir baixo custo.

No entanto, por não envolver usuários reais ou representativos, em geral, é fácil cometer erros em prever como os usuários reais lidam com a interface. Visando minimizar essa limitação, a avaliação foi complementada pela técnica de observação do usuário que apresenta bons resultados práticos de testes com o usuário real.

As técnicas de inspeção heurística e observação do usuário têm o mesmo objetivo de estabelecer se o sistema é usável, porém têm algumas diferenças na maneira de preparar e conduzir a avaliação. Frequentemente, a inspeção heurística pode ajudar a tomar cuidado com defeitos de usabilidade mais óbvios, enquanto que a observação do usuário empreendida após a inspeção pode ser efetivamente mais útil em focar problemas da perspectiva do usuário. Conseqüentemente, as duas técnicas devem ser vistas como um complemento para avaliação.

5. RESULTADOS EXPERIMENTAIS

A avaliação de interface proposta neste trabalho constituiu em aplicar técnicas de interação humano-computador no protótipo "MRI Browser" (*Magnetic Resonance Imaging Browser*), que utiliza uma coleção de imagens de modalidade médica, visando auxiliar o usuário médico no processo de diagnóstico.

O protótipo MRI Browser é um sistema de navegação que consiste em uma ferramenta útil e prática para recuperar imagens similares baseadas em seu conteúdo, dentro da rotina hospitalar. Incorpora todas as funcionalidades de um sistema CBIR e permite realizar tanto consultas textuais como por conteúdo.

Para efetuar a consulta por imagens similares, o usuário deve selecionar uma imagem de interesse como referência, especificar a quantidade de imagens similares desejadas e pressionar o botão *Query* para iniciar o processo de busca. O resultado é apresentado na mesma janela *Result* que é usada para exibir tanto os *thumbnails* obtidos pelo resultado da busca textual, como para retornar o resultado da consulta de imagens por similaridade.

Figura 1: Resultado da consulta por similaridade a partir de uma dada imagem de referência.

A Figura 1 ilustra, na janela à direita, uma imagem em seu tamanho original contendo suas informações textuais, que foi utilizada como referência para uma operação de busca das dez imagens mais semelhantes a ela. Como resultado, os *thumbnails* que mais satisfizeram o critério da consulta em relação a sua similaridade são apresentados na janela central, permitindo que o usuário efetue sua escolha, e a imagem será apresentada em tamanho normal.

O sistema também incorpora recursos avançados para avaliar a eficiência dos algoritmos de CBIR. Esses recursos mais avançados foram desenvolvidos para as pesquisas de âmbito mais abrangente. Contudo, esse estudo utilizou o protótipo MRI Browser para explorar a interface e avaliar a sua funcionalidade e usabilidade, a fim de resolver algumas questões específicas relacionadas à recuperação de imagens similares a uma dada imagem de referência, empreendidas apenas no processo que auxilia o médico em seus diagnósticos.

5.1 Material

Para organizar e conduzir uma avaliação foi necessário criar uma estratégia a fim de planejar o que, quem, quando e onde avaliar. Para isso, foram consideradas algumas questões importantes:

- Qual o domínio: representa a área de conhecimento dos especialistas em relação à aplicação desenvolvida.
- Qual o propósito da avaliação: refere-se ao objetivo para empreender uma avaliação.
- O que será testado: refere-se ao artefato físico, ou seja, o produto, sistema ou protótipo que será testado.
- Quais são os usuários: são as pessoas que usam a aplicação como base para seu trabalho.
- Quais tipos de dados coletar: são resultados coletados na avaliação que podem ser usados de maneiras diferentes para auxiliar a tomada de decisões sobre o re-design da interface do usuário. Os dados podem ser numéricos ou não-numéricos.
- Quais as tarefas que os participantes vão empreender: referem-se às atividades relacionadas à natureza do trabalho.
- Quais as considerações ambientais: são fatores relacionados ao ambiente em que a aplicação será empreendida. Devem ser considerados o aspecto físico, social, organizacional e o ambiente de apoio ao usuário.
- Quais são as restrições: consiste nas limitações em relação ao custo, tempo, tecnologia e recursos humanos que devem ser consideradas para realizar uma avaliação.

Todos esses requisitos a respeito da elaboração de uma estratégia de avaliação foram traduzidos para refletir o contexto de uso do estudo de caso, a fim de especificar o domínio da aplicação do estudo em questão, os usuários, suas tarefas e o ambiente dentro do contexto em que o sistema será usado.

5.2 Métodos

Existe uma variedade de técnicas e métodos disponíveis para avaliar sistemas interativos. Cada método possui suas potencialidades e limitações e cada qual tem sua utilidade quando aplicado adequadamente. Para selecionar o método mais

apropriado, que atendesse às necessidades particulares, foram considerados alguns fatores para categorizar os diferentes métodos [6]:

- A fase do ciclo de desenvolvimento em que a avaliação é realizada: fase de design ou implementação.
- O estilo da avaliação: estudo de laboratório ou estudo de campo.
- O nível de subjetividade ou objetividade da técnica: baseia-se fortemente na interpretação do avaliador.
- O tipo de medidas providas: medidas quantitativas ou qualitativas.
- A informação provida: informação de baixo ou alto nível.
- Os recursos requeridos: são equipamentos de hardware e software, equipamentos de gravação de vídeo e áudio, tempo, dinheiro, tecnologia, participantes, experiência do avaliador, o contexto, entre outros.

6. RESULTADOS

O processo de análise e interpretação concentrou-se em transformar os dados coletados da avaliação em informações para auxiliar a tomada de decisões sobre o re-design da interface do usuário. Primeiramente, foram analisados e interpretados os dados da observação do usuário e em seguida os dados da inspeção heurística. O resultado da avaliação será apresentado a seguir.

6.1 Observação dos usuários

A observação do usuário é um modo simples para obter informações sobre o uso real de um sistema através da interação dos usuários com ele. Os usuários foram solicitados a completar uma série de tarefas pré-determinadas, enquanto foram observados cumprindo por completo as suas obrigações rotineiras. A função do avaliador foi observar os usuários, prestando atenção em suas ações e registrando-as.

Primeiramente, foi aplicado um questionário pré-sessão para obter um *feedback* do perfil e formação dos usuários. Muitas vezes, o perfil do usuário é descrito como um conjunto geral da população do usuário em termos dos atributos ou características relevantes para o design. Na prática, os primeiros participantes já conseguem prover as informações requeridas pelos avaliadores e estima-se que cerca de cinco participantes são frequentemente suficientes [10]. Porém, vale a pena persistir com um pouco mais de participantes, pois sempre tem uma informação extra a adicionar.

Visto que a proposta de testar o sistema deveria representar o cenário que mais se aproximasse da vida real, a observação dos usuários contou com a colaboração de sete participantes da área médica, constituindo por pesquisadores do Centro de Ciências de Imagens e Física Médica (CCIFM) do Hospital das Clínicas da USP de Ribeirão Preto, que reuniu quatro médicos radiologistas, dois radiologistas residentes e um técnico em informática biomédica.

Foram selecionados participantes que possuíam faixas etárias diferenciadas, sendo que 43% estavam na faixa de 18 a 30 anos, 14% entre 31 a 44 anos e 43% entre 45 a 64 anos. A partir dessa divisão foi possível classificar os participantes nos seguintes grupos de usuários:

- Adultos jovens: de 18 a 30 anos.

- Adultos de meia idade: de 31 a 44 anos.
- Adultos mais velhos: de 45 a 64 anos.

Os participantes foram recrutados devendo também possuir alguma habilidade computacional, um pouco de conhecimento na língua inglesa e nenhuma experiência em sistemas CBIR.

A experiência dos participantes em informática foi estabelecida de acordo com a frequência de uso de sistemas computacionais. Assim, foi possível notar que todos os participantes possuíam alguma familiaridade com sistemas computacionais. O idioma inglês adotado como padrão da interface não influenciou no desempenho dos participantes. No geral, os participantes possuíam um bom conhecimento na língua inglesa.

Além desses dados, anotações adicionais também foram feitas durante as sessões. Visto que participantes mais jovens demonstraram uma maior habilidade e agilidade computacional, foi possível classificar os participantes em alguns níveis de experiência em informática: principiantes, médios e experientes. Os níveis de experiência em informática indicaram que 14% corresponderam participantes principiantes, 43% médios e 43% experientes.

O tempo de atuação profissional dos participantes corresponderam a 42% aos participantes que possuíam menos de 5 anos de atuação profissional e 58% acima de 11 anos. A partir desses dados, os participantes foram classificados de acordo com o conhecimento do domínio clínico em novatos e experientes.

Foram aplicadas 5 tarefas e cada tarefa foi formulada visando alcançar mais de um objetivo. As duas primeiras tarefas constituíram de 3 objetivos, a terceira tarefa reuniu 5 objetivos e as duas últimas incorporaram 6 objetivos. Considerando que cada tarefa possuiu vários objetivos a serem alcançados, foi possível calcular a média da completude de cada tarefa. A maioria das tarefas foi completada em 100%.

Tabela 1. Tempo de execução de tarefas por participante.

Participante	Experiência em informática	Faixa etária	Tempo de tarefa (hora : minuto : segundo)					Tempo Total
			Tarefa 1	Tarefa 2	Tarefa 3	Tarefa 4	Tarefa 5	
P1	Principiante	Adulto mais velho	0:12:08	0:11:44	0:12:03	0:04:49	0:02:57	0:43:41
P2	Médio	Adulto mais velho	0:08:15	0:07:16	0:09:23	0:08:05	0:05:10	0:38:09
P3	Experiente	Adulto jovem	0:00:51	0:01:00	0:03:03	0:02:05	0:02:12	0:09:11
P4	Médio	Adulto mais velho	0:14:43	0:05:12	0:05:55	0:05:04	0:05:12	0:36:06
P5	Médio	Adulto de meia idade	0:03:35	0:02:15	0:04:10	0:03:44	0:03:08	0:16:52
P6	Experiente	Adulto jovem	0:01:23	0:00:45	0:01:31	0:02:03	0:01:15	0:06:57
P7	Experiente	Adulto jovem	0:01:13	0:01:21	0:01:43	0:01:59	0:03:36	0:09:52
Tempo Médio			0:06:01	0:04:13	0:05:24	0:03:58	0:03:21	

Na Tabela 1 é apresentado o tempo que cada participante levou para completar as tarefas, bem como seu perfil em relação ao nível de experiência em informática e sua faixa etária. Pode-se notar que:

- Adultos mais jovens e com boa experiência em informática, atingiram um desempenho melhor no uso

do sistema, levando menos de 10 minutos para completar todas as tarefas.

- Adultos de meia idade e com experiência média em informática levaram quase que 17 minutos para realizar as tarefas.
- Adultos mais velhos e com experiência média em informática levaram aproximadamente 40 minutos para completar as tarefas.
- Adultos mais velhos e com pouca experiência em informática levaram mais de 40 minutos para realizar as tarefas.

Percebeu-se que quando o participante é mais jovem e com maior habilidade computacional ele utiliza o sistema mais fácil e rapidamente. Porém, alguns termos e funções avançadas utilizadas no sistema eram mais familiares para os participantes experientes. Assim, pôde-se verificar que o tempo de atuação profissional dos participantes refletiu diretamente nas necessidades de uso de funções relevantes disponíveis no sistema.

Tabela 2. Quantidade de erros cometidos e o tempo gasto para corrigi-los.

	Quantidade de erros cometidos e o tempo gasto para corrigi-los										Total
	Tarefa 1		Tarefa 2		Tarefa 3		Tarefa 4		Tarefa 5		
	Qtd.	Tempo	Qtd.	Tempo	Qtd.	Tempo	Qtd.	Tempo	Qtd.	Tempo	
P1	3	0:09:03	4	0:09:56	3	0:04:36	1	—	1	—	12
P2	1	—			1	0:02:30			1	0:00:53	3
P3					1	0:01:59			1	0:00:45	2
P4	7	0:10:52	3	0:02:20	3	0:02:24			4	0:03:53	17
P5	1	0:01:47	3	0:01:03	2	0:02:48			1	—	7
P6	2	0:01:15					1	0:00:44			3
P7							1	—	2	0:01:45	3
Total	14		10		10		3		10		47

A Tabela 2 apresenta a média da acurácia de tarefas completadas, ou seja, a quantidade de erros cometidos durante as tarefas e o tempo gasto para corrigi-los.

Observa-se que, na maioria, o tempo requerido para recuperar um erro é muito pequeno, alguns desses erros foram realizados por engano e percebidos rapidamente pelos participantes. No geral, os erros de uma determinada tarefa foram cometidos sucessivamente, assim foi possível medir o tempo gasto na correção de cada erro. A representação do sinal “—” significa que o participante não percebeu o erro e por isso não foi corrigido. A maioria desses erros não corrigidos decorreu da falta de visibilidade da mensagem exibida na tela informando a quantidade de imagens recuperadas.

Apesar da facilidade que os jovens e experientes em informática demonstraram no desempenho da tarefa, pôde-se observar que não possuiu nenhuma relação quanto aos erros cometidos pelos participantes. Independente da idade e habilidade computacional, percebeu-se que participantes mais cautelosos e que tiveram mais atenção na formulação das tarefas obtiveram menos erros.

As tarefas foram elaboradas para conceber os processos de aprendizagem de uso do sistema em três etapas:

- Processo de exploração do sistema por tentativa e erro;
- Processo de busca por imagens similares; e

- Processo de reconhecimento de objetos e ações ao invés de lembrança.

A primeira etapa teve o objetivo de conhecer e aprender a usar o sistema por meio da exploração da interface por tentativa e erro. Essa etapa correspondeu à execução das duas primeiras tarefas. O tempo gasto de cada participante para explorar o sistema variou, aproximadamente, cerca de 2 a 24 minutos, obtendo uma média aproximada de 10 minutos. Os participantes mais jovens obtiveram um desempenho melhor nessa etapa.

A segunda etapa correspondeu à introdução da aprendizagem da busca por imagens similares a uma dada imagem de referência. Essa etapa correspondeu à execução de uma tarefa e levou um tempo médio de aproximadamente 5 minutos e meio.

A terceira etapa focou o uso do sistema por meio do reconhecimento de objetos e ações ao invés de lembrança e durou um tempo médio de aproximadamente 7 minutos e meio. Essa fase correspondeu à execução das duas últimas tarefas referentes a todo o processo de busca textual e por similaridade, bem como a execução de algumas ações relevantes.

Observou-se, também, que o tempo médio para completar cada tarefa foi diminuindo ao decorrer da sessão, devido ao aprendizado introduzido.

A satisfação do usuário pôde ser medida por meio do questionário pós-sessão, adaptado do modelo QUIS¹ (*Questionnaire for User Interaction Satisfaction*), que foi organizado de maneira modular, focando seis fatores de interesse particular da interface: reação do sistema, tela, terminologia e informação do sistema, aprendizado, capacidade do sistema e imagens. As respostas obtidas nos questionários foram convertidas em valores numéricos e realizadas análises estatísticas simples.

Com relação à avaliação da satisfação da interface, os participantes se mostraram bastante familiarizados com a interface. A grande maioria das tarefas foram cumpridas tanto na acurácia quanto na completude. Os resultados obtidos no QUIS, apresentados na Figura 2, foram os seguintes:

- **Reação do sistema:** a impressão geral que os participantes apontaram foi um sistema muito satisfatório, bem estimulante, muito fácil de usar e adequado.
- **Tela:** os participantes consideraram a forma e tamanho da letra fácil de ler. Contudo, apontaram que nem sempre os itens da tela são fáceis de encontrar. Embora a sequência de telas esteja apresentada de maneira clara, a organização da informação não está disposta tão claramente e o sistema não dispõe de muitos realces na tela em relação a cores e negrito para ressaltar as informações relevantes.
- **Terminologia e informação do sistema:** o uso de termos utilizados no sistema não demonstrou ser tão claro. Percebeu-se que profissionais iniciantes não têm conhecimento mais aprofundado sobre o assunto e o sistema não possui instruções para auxiliá-los. No entanto, às vezes, o sistema mantém o usuário informado sobre o que está fazendo. Além disso, as

poucas mensagens que apareceram na tela não são claras e sua localização um pouco confusa e muitas vezes despercebida. Constatou-se que as mensagens de erro não se mostraram úteis.

- **Aprendizado:** o sistema foi considerado muito fácil de aprender a operar e explorar por tentativa e erro. As tarefas quase sempre puderam ser executadas de uma maneira rápida e/ou lógica e demonstrou facilidade ao lembrar de termos e uso de comandos. A conclusão das tarefas mostrou-se bem clara. O sistema não provê mensagens de ajuda, o que poderia ser muito útil para esclarecer dúvidas durante o aprendizado.
- **Capacidade do sistema:** os participantes apontaram que o sistema possui uma velocidade rápida e quase sempre apresentaram respostas confiáveis. Demonstrou facilidade na correção dos erros. O sistema foi projetado facilmente para atender todos os níveis de usuários, tanto para iniciantes como experientes.
- **Imagens:** embora a qualidade das imagens tenha sido considerada boa, o tamanho não agradou muitos participantes, sendo um pouco pequeno para requerer uma visualização mais minuciosa dos detalhes da imagem. O processo de busca de imagens similares foi considerado muito interessante e relevante, embora o resultado demonstrou nem sempre ser confiável.

			Grau de concordância					
			1	2	3	4	5	
REAÇÃO DO SISTEMA		Frustrante					4,14	Satisfatório
		Tedioso					4,43	Estimulante
		Difícil					4,29	Fácil
		Inadequado					4	Adequado
TELA	Forma e tamanho das letras	Difícil de ler					4	Fácil de ler
	Realces na tela (cores e negrito)	Pouco					3,71	Bastante
	Organização da informação	Confusa					3,71	Clara
	Sequência de telas	Confusa					4	Clara
	Itens da tela são fáceis para encontrar	Nunca					3,71	Sempre
TERMINOLOGIA E INFORMAÇÃO DO SISTEMA	Uso dos termos utilizados no sistema	Confuso					3,57	Claro
	Mensagens que aparecem na tela	Confuso					3,14	Claro
	Localização das mensagens na tela	Confusa					2,71	Clara
	Instruções para o usuário	Nunca					2,71	Sempre
	Sistema mantém você informado sobre o que está fazendo	Nunca					3,29	Sempre
APRENDIZADO	Mensagens de erro	Indefinido					2,71	Útil
	Aprender a operar o sistema	Difícil					4,57	Fácil
	Explorar por tentativa e erro	Difícil					4,57	Fácil
	Lembrar termos e uso de comandos	Difícil					4,14	Fácil
	Tarefas podem ser executadas de uma maneira rápida e/ou lógica	Nunca					4,14	Sempre
CAPACIDADE DO SISTEMA	Condição da tarefa	Confuso					4,43	Claro
	Mensagens de ajuda	Confuso					2,86	Claro
	Velocidade do sistema	Lento					4,29	Rápido
	O sistema é confiável	Nunca					3,86	Sempre
	Corrigir seus erros	Difícil					4,14	Fácil
IMAGENS	Projetado para todos os níveis de usuários (iniciantes e experientes)	Com dificuldades					3,86	Facilmente
	Qualidade das imagens	Ruim					3,57	Bom
	Tamanho das imagens	Pequeno					2,86	Grande
	Processo de busca de imagens similares	Desinteressante					4,14	Interessante
	Recuperação de imagens similares	Não relevante					4	Relevante
	Resultado da recuperação de imagens similares	Não confiável					3,71	Confiável

Figura 2. Resultado das medidas de satisfação do QUIS.

6.2 Inspeção heurística

A série escolhida para o estudo em questão foram as heurísticas de usabilidade para design de interface propostas por Nielsen que

¹ <http://lap.umd.edu/QUIS/>

apresentam bons resultados e permitem uma avaliação global do sistema. As “10 Heurísticas de Nielsen” focam problemas de usabilidade de alto nível e constituem dos seguintes itens:

- **H1:** Visibilidade do *status* do sistema.
- **H2:** Compatibilidade do sistema com o mundo real.
- **H3:** Controle e liberdade para o usuário.
- **H4:** Consistência e padrões.
- **H5:** Prevenção de erros.
- **H6:** Reconhecimento ao invés de lembrança.
- **H7:** Flexibilidade e eficiência de uso.
- **H8:** Estética e design minimalista.
- **H9:** Ajuda aos usuários a reconhecer, diagnosticar e recuperar os erros.
- **H10:** Ajuda e documentação.

Os defeitos de usabilidade foram avaliados em relação ao grau de severidade correspondentes à gravidade de cada problema no que diz respeito à frequência com que os erros ocorrem, ao impacto do problema ocorrido e a persistência do problema. Os graus de severidade foram representados por:

- **G0:** Não é considerado um problema de usabilidade.
- **G1:** Problema cosmético somente.
- **G2:** Problema de usabilidade de pequeno impacto.
- **G3:** Problema de usabilidade de grande impacto.
- **G4:** Catástrofe de usabilidade.

A avaliação heurística envolveu a participação de quatro inspetores, reunindo um especialista em usabilidade, um especialista do domínio e dois especialistas do domínio com algum conhecimento em usabilidade.

Tabela 3. Quantidade de defeitos encontrados por heurística e distribuídos pelos graus de severidade.

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	Total
G0		1		1	1			2			5
G1						1			1		2
G2	1	2		3	2	2	1				11
G3	3	7	1	3	5	4	3	1	5	1	33
G4	1		2		2	1	1		3	2	12
Total	5	10	3	7	10	8	5	3	9	3	63

Os dados foram tabulados e resumidos, conforme mostra a Tabela 3, onde apresenta a quantidade de defeitos de usabilidade encontrados em cada heurística e distribuídos pelos graus de severidade.

Figura 3. Defeitos encontrados por grau de severidade.

A Figura 3 ilustra os defeitos de usabilidade encontrados, em que cerca de 53% dos defeitos referem-se ao grau de severidade 3, que correspondem a problemas de usabilidade de grande impacto, sendo importante consertar, requerendo alta prioridade de correção. Pode-se observar, também, que 19% dos defeitos estão relacionados ao grau de severidade 4 e são considerados como catástrofe de usabilidade, por isso devem ser consertados antes que o produto seja lançado. Para os defeitos de usabilidade de pequeno impacto, de grau 2, foram registrados 17%, sendo que a prioridade deste conserto deve ser baixa. Apenas 3% dos defeitos classificados com grau 1 foram considerados como sendo somente problemas cosméticos, não tendo tanta urgência em serem consertados. Finalmente, 8% foram determinados com grau 0, isto é, não foram considerados como um problema de usabilidade. Percebe-se que 89% dos defeitos encontrados devem ser consertados independentemente da intensidade do impacto.

Figura 4. Defeitos de usabilidade encontrados.

A Figura 4 mostra que os maiores defeitos de usabilidade, correspondendo ao mesmo valor de 15,9%, são atribuídos a duas heurísticas:

- referente à compatibilidade do sistema com o mundo real, não expressando uma linguagem familiar para os usuários (H2); e
- atribuída à prevenção de erros, onde o sistema não previne que o erro ocorra antes de acontecer (H5).

A falta de ajuda aos usuários a reconhecer, diagnosticar e recuperar seus erros representou 14,3% dos defeitos de usabilidade (H9).

O reconhecimento ao invés de lembrança, não possuindo instruções de uso do sistema e sobrecarregando a memória do usuário indicou 12,7% dos problemas encontrados (H6).

A consistência e padrões, seguindo as convenções de plataforma computacional, representaram 11,1% (H4).

Para as duas heurísticas H1 e H7 obteve-se o mesmo valor de 7,9% de defeitos referentes:

- à visibilidade do *status* do sistema, mantendo o usuário informado sobre o que está acontecendo (H1); e

Os menores defeitos foram encontrados em três heurísticas com 4,8% cada, a saber:

- referente ao controle e liberdade para o usuário quando os usuários, freqüentemente, escolhem funções por engano e precisam sair do estado indesejado sem ter que percorrer um extenso diálogo (H3);
- correspondem a estética e design minimalista apresentando um design simples e prático, sem informações irrelevantes ou desnecessárias (H8); e
- prover ajuda ao usuário e documentação completa do sistema. (H10).

Constatou-se que cada avaliador identificou problemas de usabilidade diferentes. Neste estudo, foi possível perceber que avaliadores com conhecimento do domínio tiveram uma visão mais geral do sistema, comparando com outros sistemas semelhantes, enquanto que o especialista em usabilidade teve uma interpretação mais crítica e detalhada de cada ponto da interface. Desse modo, deve-se ressaltar que existe uma relação entre o perfil dos avaliadores e os problemas encontrados. Assim sendo, é importante recrutar inspetores com áreas de conhecimento diferentes, pois o perfil do avaliador tem uma influência direta na descoberta de problemas.

7. CONCLUSÕES

Este estudo introduziu um esforço para testar o desempenho de um sistema CBIR típico em relação à funcionalidade e usabilidade por meio da interação de usuários reais dentro do domínio da aplicação. Por esse motivo, o desempenho de tais sistemas pode ser avaliado por critérios reais, fundamentando uma base sólida para julgar a percepção humana e prover meios para melhorar o desenvolvimento de sistemas que apóiam usuários reais em suas tarefas.

Para alcançar esse desafio, foram adotadas uma combinação de técnicas de avaliação da interação humano-computador, a fim de identificar abordagens promissoras no desempenho e satisfação dos usuários. Primeiramente, foi empreendida a técnica de observação de usuário, com o envolvimento efetivo de usuários reais, completando a seguir com a inspeção heurística, conduzida por meio da participação de especialistas.

A observação de usuários atingiu o objetivo proposto, pois pôde prover uma grande percepção de como os usuários pensam sobre a interface. Independente do perfil do usuário, a interface demonstrou facilidade de uso, assegurando que, no geral, os participantes cumprissem as tarefas tanto na acurácia como na completude. Com relação à satisfação da interface, os usuários superaram suas expectativas que foram observadas antes e depois da interação.

Contudo, o sistema submetido à avaliação heurística sugeriu um escopo considerável de melhorias. Pôde-se perceber que é desejável que os avaliadores possuam uma certa experiência em usabilidade para abstrair todo o conhecimento requerido.

complemento visual para a análise e a análise de usabilidade as técnicas para avaliar a funcionalidade e a usabilidade de sistemas CBIR aplicados à medicina, são primordiais para torná-los mais usáveis, eficientes e com elevado grau de satisfação para o usuário médico. Assim, os sistemas CBIR poderão ser melhor empregados pela comunidade médica, contribuindo para aprimorar os diagnósticos e tratamentos médicos.

Ainda há muito trabalho a ser realizado nessa área. É preciso ter um bom conhecimento dos fatores humanos que afetam as estratégias de busca dentro do contexto de uso, levando em consideração as necessidades particulares dos usuários e enfatizando sempre os aspectos da interação humano-computador.

8. AGRADECIMENTOS

A pesquisa apresentada neste artigo foi apoiada pela EMBRAPA, FAPESP e CNPq.

9. REFERÊNCIAS

- [1] ISO-9241-11. 1998. Ergonomic requirements for office work with Visual Display Terminals (VDTs): guidance on usability. Geneva.
- [2] Carvalho, A. F. P., Anacleto, J. C. "Conceitos de Usabilidade e sua Abrangência no Desenvolvimento de Sistemas Interativos", Revista Eletrônica de Iniciação Científica, Brasil, v. IV, n. IV, p. 3, 2004.
- [3] Nielsen, J. 2000. *Projetando websites*. Rio de Janeiro: Campus.
- [4] Preece J., Rogers Y., and Sharp H. 2005. *Design de Interação: além da interação homem-computador*. Porto Alegre: Bookman.
- [5] Rocha H. V., Baranauskas M. C. C. 2003. *Design e Avaliação de Interfaces Humano-Computador*. Campinas: NIED/UNICAMP.
- [6] Dix A., Finlay J., Abowd G. D., and Beale R. 2004. *Human-Computer Interaction*, 3 ed. Essex: Prentice Hall.
- [7] Preece, J. 1995. *A Guide to Usability: human factors in computing*. Wokingham: Addison-Wesley.
- [8] Kosara, R., Healey, C. G., Interrante, V., Laidlaw, D. H., and Ware, C. 2003. User Studies: Why, How, and When?. *IEEE Comput. Graph. Appl.* 23, 4 (Jul. 2003), 20-25. DOI= <http://dx.doi.org/10.1109/MCG.2003.1210860>
- [9] Marsh, T. 1999. Evaluation of virtual reality systems for usability. In *CHI '99 Extended Abstracts on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, May 15 - 20, 1999). CHI '99. ACM, New York, NY, 61-62. DOI= <http://doi.acm.org/10.1145/632716.632756>
- [10] Stone, D., Jarret, C., Woodroffe, M., and Minocha, S. 2005. *User Interface Design and Evaluation*. San Francisco: Morgan Kaufmann.