

# Montando questionários para medir a satisfação do usuário: Avaliação de interface de um sistema que utiliza técnicas de recuperação de imagens por conteúdo

Ana Lúcia Filardi

Embrapa Monitoramento por Satélite  
Av. Soldado Passarinho, 303  
13070-115 - Campinas, SP - Brasil  
ana@cnpm.embrapa.br  
+55 19-3211-6200

Agma Juci Machado Traina

Universidade de São Paulo  
Av. do Trabalhador São-Carlense, 40  
13560-970 - São Carlos, SP - Brasil  
agma@icmc.usp.br  
+55 16-3373-9693

## ABSTRACT

An attractive and intuitive design of the human-computer interface is essential to the success of any computational system. Content-based image retrieval (CBIR) systems are inherently interactive. However, little attention has been paid to their interfaces. A major difficulty for designing the interface of these systems is to ally high quality means to integrate the querying on the images and the presentation and handling of the results. The goal of this paper was to analyze the user interaction in CBIR systems regarding their functionality and usability. This was achieved by following the actual interaction of users while working with a CBIR system. This task allowed us to suggest a questionnaire that can properly reflect the evaluated aspects of the users interaction regarding their satisfaction degree. This result can help on improving the development of new CBIR systems, which are a valuable tool in medical systems.

## RESUMO

O design intuitivo e atraente da interface com o usuário é essencial para o sucesso de qualquer sistema computacional. Dentro do escopo de sistemas CBIR (*Content-Based Image Retrieval*), que são sistemas inerentemente interativos de recuperação de imagens baseada em conteúdo, a interface do usuário ainda é um elemento constituído de pouca pesquisa e desenvolvimento. Um dos maiores obstáculos para a eficácia de design desses sistemas consiste da necessidade em prover aos usuários uma interface de alta qualidade na formulação de perguntas e exibição dos resultados. O objetivo principal deste trabalho visou analisar a interação do usuário em sistemas

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IHC 2008 – VIII Simpósio Sobre Fatores Humanos em Sistemas Computacionais. October 21-24, 2008, Porto Alegre, RS, Brazil. Copyright 2008 SBC. ISBN 978-85-7669-203-4

CBIR em relação à funcionalidade e usabilidade por meio da interação de usuários reais dentro do domínio da aplicação e propor questionário que reflita fielmente os pontos avaliados. Para alcançar esse desafio, a interação humano-computador foi avaliada, adotando técnicas de questionamento a fim de mensurar a satisfação do usuário em relação ao uso do sistema.

## Author Keywords

Human-computer interaction, HCI, evaluation, QUIS, CBIR.

## ACM Classification Keywords

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

## INTRODUÇÃO

A interação da ciência da computação com diversas áreas da pesquisa científica vem crescendo muito nos últimos anos e contribuindo com novos desafios. A medicina é uma das áreas beneficiadas tornando o diagnóstico médico e a evolução dos tratamentos cada vez mais precisos e eficientes [1]. Para dar suporte a diagnósticos, imagens de exames gerados por meio de equipamentos médicos têm sido incorporadas nos Sistemas de Informações Hospitalares (SIH), visando consolidar em um só sistema as informações dos pacientes.

O interesse em utilizar imagens nos diagnósticos vem impulsionando o desenvolvimento de novas técnicas que se preocupam com a recuperação desses tipos de dados. Nesse sentido, as consultas por similaridade tornam-se muito interessantes, pois, ao analisar uma imagem, o usuário pode desejar recuperar as imagens mais semelhantes a ela com o objetivo de efetuar cruzamentos entre diagnósticos e tratamentos já realizados e auxiliar o profissional da área médica na elaboração de laudos e novos diagnósticos. Geralmente, para recuperar tais dados, as consultas são realizadas por meio de critérios de similaridades baseados nas características intrínsecas extraídas das imagens, ou seja, em seu conteúdo. Para isso, o sistema deve prover uma

interface interativa e intuitiva para facilitar a recuperação e exibição desses dados.

Este trabalho teve como objetivo analisar a interação do usuário em sistemas CBIR e avaliar sua funcionalidade e usabilidade, aplicando técnicas de questionamento a fim de identificar abordagens promissoras no desempenho e satisfação dos usuários.

#### TÉCNICAS DE INTERAÇÃO HUMANO-COMPUTADOR

Um bom design de interface de usuário deve prover uma interação fácil, natural e atraente entre o usuário e o sistema. A área da computação que estuda como as pessoas interagem com os sistemas computacionais denomina-se Interação Humano-Computador, do inglês *Human-Computer Interaction* (HCI), e baseia-se em estabelecer técnicas e guias mestras para projetar e desenvolver interfaces com o usuário que alcancem alta usabilidade, permitindo que o usuário possa realizar suas tarefas com segurança, de maneira eficiente, eficaz e com satisfação.

Esses aspectos são conhecidos como usabilidade e são aplicados para medir o desempenho (eficácia e eficiência) e a satisfação do usuário, a fim de determinar quanto um produto é usável dentro de um contexto particular [2].

As medidas de usabilidade refletem os resultados da interação dos usuários com o sistema proposto. Os dados relacionados a cada componente de usabilidade podem ser obtidos a partir de medidas objetivas ou subjetivas:

- **Medidas objetivas:** são medidas de observação direta, tipicamente em relação ao desempenho do usuário na execução de testes enquanto usa a interface. As medidas objetivas podem prover dados como medidas de tempo, velocidade ou ocorrência de eventos particulares [3].
- **Medidas subjetivas:** representam opiniões, usualmente dos usuários, no que diz respeito à usabilidade da interface. Respostas subjetivas provêm dados que expressam sentimentos, crenças, atitudes e preferências dos usuários [3].

Enquanto que as medidas objetivas fornecem indicações diretas de eficácia e eficiência do comportamento do usuário, as medidas subjetivas estão relacionadas diretamente com a satisfação do usuário, derivadas das opiniões que os usuários expressam sobre seus trabalhos e seus resultados. Entretanto, a satisfação também pode ser inferida por medidas objetivas do comportamento do usuário, assim como as estimativas de eficácia e eficiência podem ser derivadas de opiniões subjetivas expressadas pelos usuários.

A escolha das medidas depende dos objetivos e do contexto em que a avaliação está inserida, pois freqüentemente possuem diferenças significativas dos tipos de usuários, tarefas ou ambientes. Normalmente, é necessário fornecer pelo menos uma medida para estimar a eficácia, eficiência e satisfação.

#### Eficácia

As medidas de eficácia estão relacionadas a acurácia e completude com que os usuários atingem objetivos específicos [2]. A acurácia pode ser especificada ou medida pela quantidade de erros, enquanto que a completude pelo cumprimento da tarefa a ser realizada, ou seja, pela proporção que a tarefa foi alcançada.

#### Eficiência

As medidas de eficiência relacionam o nível de eficácia alcançada no dispêndio de recursos, como esforço mental ou físico, tempo, custos materiais ou financeiros [2]. Ou seja, a eficiência é medida atribuindo o nível de eficácia alcançada com os recursos usados. Por exemplo, a eficiência temporal pode ser definida como a proporção entre a medida de eficiência em alcançar um objetivo específico e o tempo para alcançar tal objetivo.

#### Satisfação

A satisfação é uma resposta do usuário na interação com o produto. Permite medir a extensão pela qual os usuários estão livres de desconforto e as atitudes em relação ao uso do produto [2]. A satisfação pode ser especificada e medida pela avaliação subjetiva em escalas de desconforto experimentado, gosto pelo produto, satisfação com o uso do produto, aceitação da carga de trabalho quando realizam diferentes tarefas ou a extensão com os quais objetivos particulares de usabilidade foram alcançados. Outras medidas de satisfação podem incluir o número de comentários positivos e negativos registrados durante o uso do produto.

A satisfação pode ser avaliada ou estimada através de medidas subjetivas ou objetivas. Medidas objetivas podem ser baseadas na observação do comportamento do usuário como, por exemplo, a postura corporal, movimento do corpo, freqüência de distração ou através do monitoramento de respostas psicológicas do usuário. As medidas subjetivas de satisfação são adquiridas quantificando subjetivamente a intensidade das reações, atitudes ou opiniões expressadas pelo usuário. Esse processo de quantificação pode ser correspondente à intensidade de seus sentimentos ou classificando produtos na ordem de preferência ou atribuindo uma escala de atitudes baseada em questionário.

O grande desafio em projetar sistemas para pessoas usarem é saber como fazer a transição do que pode ser feito (funcionalidade) para como deve ser feito (usabilidade), buscando alcançar as necessidades e os objetivos dos usuários. Desse modo, avaliações da interação humano-computador são necessárias para verificar se as idéias do projetista são realmente o que os usuários necessitam ou desejam. 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 [5]:

- avaliar a funcionalidade do sistema;
- avaliar o efeito da interface junto ao usuário; e
- identificar problemas específicos do sistema.

O primeiro objetivo visa avaliar a eficiência do sistema na execução das tarefas pelo usuário. O design da interface deve permitir que o usuário efetue a tarefa pretendida de modo fácil e eficiente. Esse tipo de avaliação envolve medir a performance do usuário em relação ao sistema. O segundo objetivo da avaliação é medir o impacto do design junto ao usuário, levando em consideração os aspectos da facilidade de aprendizado do sistema (*learnability*), sua flexibilidade (*flexibility*), bem como sua robustez (*robustness*). O último objetivo da avaliação requer identificar problemas específicos do design, verificando os aspectos usados no contexto do design que ocasionam resultados inesperados ou confusos para os usuários.

Vários métodos e técnicas têm sido propostos para avaliar a interatividade de sistemas. Muitas dessas técnicas de avaliação são analíticas e envolvem o julgamento de especialistas em HCI para estimar o impacto do design em relação a um usuário típico. Essas técnicas baseiam-se 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 aspectos relacionados à usabilidade da interface e, por isso, são conhecidos como inspeção de usabilidade [6].

Essas técnicas são úteis para filtrar e refinar o design, porém não substituem os testes com pessoas para as quais os sistemas foram projetados, que são os usuários. As avaliações de design centrado no usuário tem como premissa básica que as necessidades dos usuários devem ser levadas em conta durante todo o ciclo de desenvolvimento do projeto [4]. Essas abordagens de avaliação com a participação do usuário são conhecidas como testes de usabilidade.

Os testes de usabilidade são métodos de avaliação centrados no usuário e têm o objetivo de auxiliar como o design pode ser melhorado durante o seu desenvolvimento e avaliar a qualidade global da interface utilizando-se medidas de performance. Os testes de usabilidade determinam medidas quantitativas de desempenho que são importantes para avaliar se os objetivos de usabilidade foram atendidos, para comparar produtos competitivos e para realizar pesquisas em fatores humanos.

Em geral, os testes de usabilidade são considerados como os métodos mais eficazes em detectar problemas mais graves, porém são também os mais caros.

#### TÉCNICAS DE QUESTIONAMENTO

As técnicas de questionamento são aplicadas nos testes de usabilidade, que envolvem a participação efetiva do usuário e constituem em perguntar diretamente ao usuário sobre a interface com o objetivo de descobrir se o sistema está de

acordo com as suas necessidades. Essas técnicas podem ser úteis para extrair detalhes do ponto de vista do usuário em relação ao sistema. Normalmente, são utilizadas para coletar informações sobre as necessidades dos usuários e podem revelar resultados que não teriam sido considerados pelo designer. Além disso, é um método relativamente simples, barato de administrar e provê materiais suplementares úteis para serem usados por outros métodos.

As principais técnicas de questionamento são entrevistas e questionários [6]. Tais técnicas provêm dados sobre a preferência dos usuários, porém diferem no tipo de preparação requerida, no estilo de apresentação e na flexibilidade das respostas. Os dados coletados nas entrevistas tendem a ser qualitativos enquanto que os questionários são geralmente quantitativos. Os questionários também oferecem a vantagem de atingir um grande número de pessoas, podendo obter resultados estatísticos significantes [7].

#### QUESTIONÁRIOS

A aplicação de questionário é uma técnica bem estabelecida de coleta de dados demográficos e de opiniões de usuários. Os questionários podem ser utilizados isolados ou em conjunto com outros métodos de avaliação, a fim de esclarecer ou aprofundar algum assunto. Também podem ser aplicados em várias fases do processo de design, incluindo o levantamento de requisitos, análise e avaliação das tarefas.

Todavia, os questionários devem ser bem elaborados. Deve-se estabelecer qual o propósito do questionário, como as respostas serão analisadas e especificar quais as variáveis a serem mensuradas. Deve-se assegurar que as perguntas sejam claras e que os dados coletados possam ser analisados adequadamente. É prudente realizar um estudo piloto antes de aplicar efetivamente o questionário para checar se as questões são compreensíveis e os resultados são os esperados [6]. Isso permite corrigir qualquer problema com a estruturação do questionário antes de ser aplicado.

Geralmente, as respostas obtidas nos questionários são convertidas em valores numéricos e são realizadas análises estatísticas. Média e desvio padrão são as principais medidas estatísticas usadas nas análises da maioria dos dados da pesquisa. Se houver a necessidade de utilizar técnicas mais avançadas deve-se ter o auxílio de um estatístico durante a fase de planejamento e elaboração dos questionários.

Outro uso bem conhecido dos questionários é o estudo “antes e depois” do desempenho do usuário. Alguns temas são elaborados em forma de questionário para extrair as expectativas do usuário ou testar seu desempenho antes de ser realizada uma experiência em particular. Então, depois da experiência, o mesmo questionário é novamente aplicado. Esses tipos de questionários, conhecidos como pré e pós-questionários, são capazes de identificar como a

atitude e o desempenho dos usuários podem mudar de perspectiva [7].

**COLETANDO DADOS DA AVALIAÇÃO**

A utilização de questionários para coleta de dados durante as sessões de avaliação possui as seguintes vantagens:

- todas as questões que precisam ser perguntadas podem ser descritas no questionário, diminuindo a chance de esquecer de coletar algum dado importante;
- todos os participantes recebem as mesmas questões, possibilitando comparar as respostas entre os participantes diferentes; e
- pode coletar tanto dados quantitativos como qualitativos.

A avaliação subjetiva é um componente importante nas avaliações de usabilidade. Essa preocupação levou ao desenvolvimento e padronização de alguns instrumentos gerais para avaliação de usuários em sistemas de computadores interativos. Esses instrumentos são compostos de questionários que possuem uma abordagem hierárquica de fatores de usabilidade, onde os dados são armazenados e analisados. Métodos de testes psicológicos são aplicados para assegurar uma validação empírica dos itens e estimar sua confiabilidade. Tem como propósito:

- orientar o design ou re-design de sistemas;
- oferecer uma ferramenta para estimar áreas potenciais de melhorias de sistemas;
- prover um instrumento de validação para conduzir avaliações comparativas; e
- servir como um instrumento de teste em laboratórios de usabilidade.

Vários questionários estão disponíveis como parte da avaliação de usabilidade, tanto para uso acadêmico como pessoal. Os questionários de avaliação de usabilidade mais conhecidos e que mais se destacam como produtos comerciais são [4, 8]:

- **QUIS**<sup>1</sup> (*Questionnaire for User Interaction Satisfaction*);
- **SUMI**<sup>2</sup> (*Software Usability Measurement Inventory*);
- **WAMMI**<sup>3</sup> (*Website Analysis and Measurement Inventory*); e
- **SUS**<sup>4</sup> (*System Usability Scale*).

Todos esses tipos de questionários são úteis, porém possuem propósitos diferentes.

<sup>1</sup> <http://lap.umd.edu/QUIS/>

<sup>2</sup> <http://sumi.ucc.ie>

<sup>3</sup> <http://www.wammi.com>

<sup>4</sup> <http://www.mindd.com>

**QUIS**

O QUIS é uma ferramenta que foi desenvolvida por uma equipe multidisciplinar de pesquisadores do *Human-Computer Interaction Laboratory* (HCIL) da *University of Maryland*, com a finalidade de estimar a satisfação subjetiva dos usuários focando aspectos específicos da interface humano-computador.

Foi estruturado de maneira modular e organizado hierarquicamente, podendo ser configurado de acordo com as necessidades de análise de cada interface, podendo incluir somente as seções que são de interesse do usuário [9]. Cada sessão especifica alguns pontos de interesse da interface. Foi projetado para medir a satisfação global do sistema, abordando 11 fatores específicos de interface: fatores da tela, terminologia e *feedback* do sistema, fatores de aprendizagem, capacidade do sistema, manuais técnicos, tutoriais *on-line*, multimídia, reconhecimento de voz, ambiente virtual, acesso a internet e instalação do software.

As questões são apresentadas na forma de afirmações utilizando as escalas de diferencial semântico, que baseiam-se em explorar uma faixa de atitudes bipolares representada por um par de adjetivos. As questões são respondidas em uma escala que varia de 0 a 9, onde o zero representa um adjetivo negativo e os demais representam adjetivos positivos. Por ser um questionário geral utilizado para uma ampla variedade de produtos, também inclui a opção N/A (não-aplicável). A Figura 1 mostra um exemplo de uma questão com escalas de satisfação específica.

QUIS	1	2	3	4	5	6	7	8	9	N/A
Mensagens que aparecem na tela	confusa									clara

Figura 1. Exemplo de uma questão do QUIS que utiliza escala de diferencial semântico.

O QUIS tem demonstrado alta eficácia e confiabilidade, sendo um dos questionários de usabilidade mais amplamente empregados na avaliação de interfaces. Embora tenha sido desenvolvido para avaliar a satisfação do usuário, é freqüentemente aplicado a outros aspectos do design de interação. Tem como vantagem ter passado por vários ciclos de qualificação e ter sido utilizado por centenas de estudos de avaliação. Trata-se, portanto, de um instrumento bem experimentado e testado.

**SUMI**

O SUMI foi desenvolvido pelo *Human Factors Group* (HFC) da *University College*. É um método consistente para estimar a qualidade do software do ponto de vista do usuário final e pode ajudar na detecção de falhas de usabilidade antes de um produto ser lançado. É um método rigorosamente testado e comprovado. É apoiado por uma extensa base de dados embutida em uma ferramenta de análise efetiva e geração de relatórios.

Tem sido utilizado eficazmente para [10]:

- estimar novos produtos durante a avaliação do produto;

- comparar produtos ou versões de produtos;
- estabelecer metas para desenvolvimento de aplicações futuras;
- definir metas para a qualidade de uso; e
- destacar bons e maus aspectos da interface.

Consiste de 50 questões de afirmações escalares correspondentes a três níveis: "concordo", "não sei" e "não concordo", como mostra a Figura 2.

SUMI	Concordo	Não sei	Não concordo
As mensagens de prevenção de erro não são adequadas	[ ]	[ ]	[ ]

Figura 2. Exemplo de uma questão do SUMI.

O SUMI tem sido normalizado por uma base de dados substancial de resultados, permitindo que o software de análise possa comparar, dentro de uma escala global, a percepção da qualidade de uso e medir, dentro de uma sub-escala, aspectos da satisfação do usuário referentes aos fatores de eficiência, preferência, utilidade, controle e facilidade de aprendizagem.

**WAMMI**

O WAMMI é um serviço exclusivo para avaliação de *Websites on-line*, com o propósito de ajudar os proprietários do site a cumprir suas metas corporativas através da medição e monitoramento das reações do usuário sobre sua facilidade de uso. Através de um botão colocado no *site*, é disponibilizado um questionário com a estrutura de um formulário para ser preenchido. Os dados do questionário são armazenados e analisados a partir de uma base de dados padronizada com escores normalizados. São utilizados para avaliar os seguintes aspectos: atratividade, controle, eficiência, utilidade, aprendizagem e usabilidade global.

O WAMMI tem como objetivo [11]:

- medir a satisfação do usuário sobre o site baseado na reação do usuário;
- gerar dados objetivos de gestão em um formato fácil de entender;
- prover uma base para mudanças do *Website* e melhorias de design;
- comparar seu site em relação aos demais em termos de satisfação do usuário; e
- acompanhar o desempenho do *Website* para verificar se as metas estão sendo cumpridas.

O WAMMI iniciou em 1996 como um serviço comercial para análise de *Website*, visando apresentar um relatório com rapidez e objetivo sobre quem visitou o site, o que acharam e o que precisa melhorar. Atualmente, com foco nos mercados globais, as reações e opiniões dos usuários variam entre os diferentes países. WAMMI é

frequentemente utilizado internacionalmente e está disponível na maioria dos idiomas europeu.

O serviço de análise de *Website* WAMMI é constituído por um questionário contendo 20 afirmações e uma única base de dados internacional. Um relatório eletrônico é gerado automaticamente pelo servidor no final de um período de avaliação, que inclui:

- uma medida objetiva das reações do usuário para seu *Website*;
- perfil do usuário WAMMI;
- um mapeamento da análise do visitante incluindo as principais razões pelas quais os usuários visitam seu site;
- uma análise de melhoria do site;
- uma análise estatística e tabulações cruzadas; e
- dados numéricos para análises posterior.

**SUS**

O SUS foi desenvolvido como parte do programa de engenharia de usabilidade integrado a *Digital Equipment Co Ltd.*, localizado em Reading, no Reino Unido. É um questionário simples que aborda uma visão global de estimativas subjetivas de usabilidade. As questões consistem em 10 afirmações que utilizam o formato da escala *Likert*, onde é mensurada a intensidade de concordância dentro de uma escala de cinco pontos. Abaixo, a Figura 3 apresenta um exemplo de uma questão do SUS na escala *Likert*.

SUS	Discordo fortemente				Concordo fortemente
Achei o sistema fácil de usar					
	1	2	3	4	5

Figura 3. Exemplo de uma questão do SUS que utiliza escala *Likert*.

Esse questionário provê um alto nível de subjetividade e é frequentemente usado para comparação de usabilidade entre sistemas. Tem se mostrado como uma ferramenta valiosa de avaliação, por ser robusta, confiável e por correlacionar bem com outras medidas subjetivas de usabilidade [12].

O SUS tem sido disponível gratuitamente para uso em avaliações de usabilidade e tem sido utilizado em uma variedade de projetos de pesquisa e avaliações na área industrial.

**CONDUZINDO A SESSÃO DA AVALIAÇÃO**

Cada design requer um modo 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. Cada método possui suas potencialidades e limitações e cada qual tem sua utilidade quando aplicado adequadamente.

Parte da avaliação foi analisar os questionários de usabilidade disponíveis e escolher, compor e adaptar as técnicas a partir de um conjunto disponível para ser aplicado em sistemas CBIR. Tal avaliação foi baseada em um estudo de caso aplicado à medicina para auxiliar o médico no processo de diagnóstico, visando melhorar a interação humano-computador.

Os SIH's vem proporcionando a integração de todas as informações dos pacientes dentro do ambiente hospitalar com imagens médicas dos pacientes, provenientes de exames gerados por equipamentos médicos, tal como Ressonância Magnética (RM), Medicina Nuclear (MN), Tomografia Computadorizada (CT), etc. Desse modo, os diagnósticos podem ser efetuados utilizando as descrições textuais e tendo o apoio visual das imagens dos exames médicos.

Para análises médicas é importante cruzar diagnósticos previamente realizados e comparar imagens de exames de pacientes, a fim de comparar casos anteriores ou verificar o diagnóstico efetuado. Dificilmente imagens de dois exames médicos serão exatamente iguais, mesmo se as anomalias tiverem a mesma classificação. Assim, é muito mais interessante realizar consultas em que as imagens sejam similares à do paciente em estudo. Nesse caso, o critério mais adequado para ser usado é o da semelhança, que utiliza como chave de busca a própria imagem.

Para obter um auxílio efetivo à recuperação de imagens e informações de pacientes de maneira rápida e descentralizada, as técnicas CBIR têm sido incorporadas nos sistemas médicos [13, 14].

Sistemas que utilizam técnicas CBIR propiciam o armazenamento e manipulação de grandes volumes de dados e imagens e processam operações de consultas de imagens a partir de características visuais extraídas automaticamente por meio de métodos computacionais e ao interesse de recuperar informações visuais de maneira eficiente e precisa.

As principais ferramentas conceituais utilizadas nos sistemas CBIR são denominadas consultas por similaridade e são realizadas a partir do significado do conteúdo do objeto, ao invés das informações descritivas associadas às imagens. A comparação de imagens por similaridade é essencial para tratar dados complexos e capturar o que há de mais representativo a fim de extrair informação que as represente de modo mais fiel.

Um sistema CBIR consiste de três fases de procedimentos, que incluem: processamento de imagens, extração de vetores de características e indexação [15]. A Figura 4 apresenta uma visão geral de um sistema CBIR.

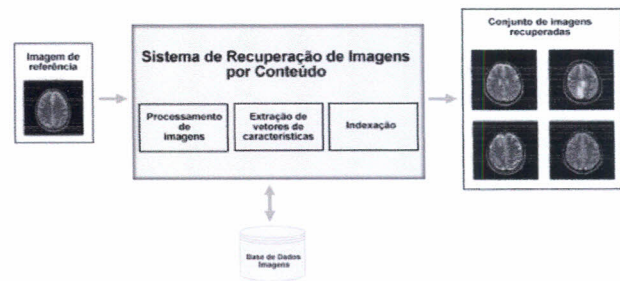


Figura 4. Visão geral de um sistema CBIR.

Este trabalho utilizou o protótipo “MRI Browser” (*Magnetic Resonance Imaging Browser*) que incorpora todas as funcionalidades de um sistema CBIR e permite realizar tanto consultas textuais como por conteúdo.

O MRI Browser é um sistema de navegação desenvolvido pelo Grupo de Base de Dados e Imagens (GbdI) do Instituto de Ciências Matemáticas e de Computação (ICMC) da Universidade de São Paulo (USP), que consiste em uma ferramenta útil e prática para recuperar imagens similares baseadas em seu conteúdo, utilizando-se de uma coleção de imagens de modalidade médica para auxiliar o usuário médico em seus diagnósticos.

A Figura 5 exibe o resultado de uma consulta por imagens similares efetuada no protótipo MRI Browser, que ilustra um conjunto de miniaturas (*thumbnails*) das imagens que mais satisfizeram o critério de busca tendo como referência uma imagem de interesse selecionada pelo usuário.

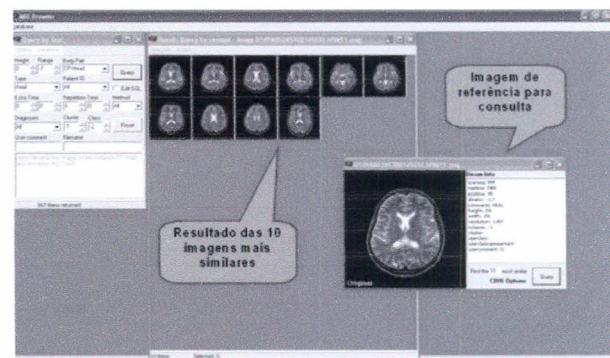


Figura 5. Resultado da consulta por similaridade a partir de uma dada imagem de referência.

O protótipo MRI Browser foi utilizado para explorar a interface e avaliar a sua funcionalidade e usabilidade, aplicando técnicas de HCI, 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. Desse modo, para atender às necessidades particulares desse trabalho, parte dessa avaliação compreendeu a utilização da técnica de observação do usuário que apresenta bons resultados práticos de testes com o usuário real.

A observação do usuário é um modo simples para obter informação sobre o uso real da interação do usuário com o sistema. Nessa avaliação, os usuários foram solicitados a completarem uma série de tarefas pré-determinadas, sendo observados por um avaliador no cumprimento por completo de suas obrigações rotineiras, prestando atenção em suas ações e registrando-as.

A avaliação constitui da aplicação de dois questionários, um questionário antes do teste, denominado pré-sessão, para obter um feedback do perfil e formação do usuário e um questionário pós-sessão para obter um background dos sentimentos do usuário em relação ao uso do sistema.

Participante nº: \_\_\_\_\_

**Questionário**

1. Sexo:             [ ] Feminino  
                      [ ] Masculino

2. Faixa de idade:   [ ] até 17 anos  
                      [ ] 18 a 24 anos  
                      [ ] 25 a 44 anos  
                      [ ] 45 a 64 anos  
                      [ ] acima de 65 anos

3. Grau de instrução:             Especialidade: \_\_\_\_\_  
      [ ] Superior completo.       Tempo de atuação: \_\_\_\_\_  
      [ ] Superior incompleto.   Especifique: \_\_\_\_\_

4. Experiência em informática:	Frequência de uso			
	Diária	Semanal	Eventual	Nunca
Uso de sistemas médicos informatizados	[ ]	[ ]	[ ]	[ ]
Uso de correio eletrônico (e-mails)	[ ]	[ ]	[ ]	[ ]
Uso de internet (por ex. pesquisa e diversão)	[ ]	[ ]	[ ]	[ ]
Uso de editores de texto (por ex. Word) e/ou planilhas eletrônicas (por ex. Excel)	[ ]	[ ]	[ ]	[ ]

5. Conhecimento da língua inglesa:	Pouco	Razoavelmente	Bem
Compreende	[ ]	[ ]	[ ]
Fala	[ ]	[ ]	[ ]
Lê	[ ]	[ ]	[ ]
Escreve	[ ]	[ ]	[ ]

6. O que você espera de um sistema para auxiliar o diagnóstico médico, que tem o objetivo de recuperar imagens de exames médicos similares?  
\_\_\_\_\_

Figura 6. Questionário pré-sessão.

O questionário pré-sessão, referente ao perfil do participante, conforme mostra a Figura 6, constituiu de dois tipos de questões:

- **Questões fechadas:** onde o participante pôde selecionar uma resposta dentre uma escolha de alternativas. Foram abordadas questões sobre o sexo, faixa de idade, grau de instrução, especialidade médica, tempo de atuação, experiência em informática e conhecimento da língua inglesa.
- **Questões abertas:** onde o participante ficou livre para prover sua própria resposta. Foi elaborada apenas uma questão complementar referente à expectativa do participante sobre o uso de um sistema para recuperar imagens de exames médicos similares. Essa questão teve

o propósito de obter uma perspectiva do usuário antes e depois do seu desempenho. Questões abertas são úteis para obter informação subjetiva geral, mas são difíceis de analisar ou comparar e podem ser consideradas apenas como um suplemento. No entanto, podem identificar erros ou fazer sugestões que não foram consideradas pelo designer.

**RESULTADOS**

Em geral, não são necessários muitos participantes para empreender uma avaliação. Na prática, os primeiros participantes já conseguem prover as informações requeridas pelos avaliadores e estima-se que cerca de cinco participantes são freqüentemente suficientes [8]. 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 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.

Os participantes foram divididos por faixa etária, sendo que 43% corresponderam os adultos jovens (faixa etária entre 18 a 30 anos), enquanto que os adultos de meia idade corresponderam a 14% (idade entre 31 a 44 anos), e os adultos mais velhos corresponderam a 43% (idade entre 45 a 64 anos).

A experiência dos participantes em informática foi estabelecida de acordo com a freqüência de uso de sistemas computacionais. Os níveis de experiência em informática indicaram que 14% foram representados por participantes principiantes, 43% médios e 43% experientes.

Os participantes foram também classificados em novatos e experientes de acordo com o conhecimento do domínio clínico em função do tempo de atuação profissional. Os novatos representaram 42% dos participantes que possuem menos de 5 anos de atuação profissional e os experientes consistiram em 58% acima de 11 anos.

O questionário pós-sessão, apresentado nas Figuras 7 e 8, objetivou mensurar a satisfação do usuário em relação ao uso do sistema, a fim de se obter uma avaliação subjetiva. Esse tipo de avaliação é considerado um componente importante nas avaliações de usabilidade.

Analisando todos os questionários de usabilidade disponíveis, o QUIS foi o que mais se aproximava do contexto em questão.

O questionário QUIS original foi aplicado no teste piloto e demonstrou ser muito longo, cansativo e tedioso. Os participantes tiveram dúvidas durante o preenchimento em relação a algumas questões que pareciam ser bastante similares ou por falta de compreensão das afirmações.

Visando reduzir a carga de esforço dos participantes e também a encorajá-los a responder um segundo questionário depois de empreender o teste de avaliação da interface e não deixar que a sessão ultrapassasse o tempo determinado, a melhor opção para registrar a satisfação da interação do usuário no estudo em questão foi elaborar um outro questionário baseado no QUIS.

de claro a confuso. Esse tipo de escala é bastante utilizada em pesquisas de HCI e é útil para indicar a preferência dos usuários, tendo a vantagem de ser fácil de analisar.

A fim de não confundir o respondente, o grau de concordância da escala foi reduzido de dez para cinco itens, variando de 1 (grau mais baixo) a 5 (grau mais alto) ou N/A (não se aplica).

Das 50 questões propostas no QUIS original foram consideradas as 25 questões que mais se aproximavam do estudo em questão. As 5 últimas questões foram adicionadas, baseadas no fator multimídia, para focar os aspectos de imagens, totalizando 30 questões.

A nova versão do QUIS foi projetada para medir a satisfação global do sistema em uma organização hierárquica, focando seis fatores específicos de interface: reação do sistema, tela, terminologia e informação do sistema, aprendizado, capacidade do sistema e imagens.

Na última parte do formulário foi incluída uma questão aberta para que os respondentes pudessem registrar seus comentários e sugestões.

Como resultado da avaliação obtido no QUIS, os participantes se mostraram bastante familiarizados com a interface. Os resultados resumidos e apresentados na Figura 9 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 seqüência das 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 prevê

Participante nº: \_\_\_\_\_

QUIS - Questionário da satisfação da interação do usuário						
Por favor, em cada questão abaixo, assinale com um "X" a alternativa que mais reflete a sua impressão sobre o uso deste sistema, indicando o grau de concordância dentro de uma escala de 1 (grau mais baixo) a 5 (grau mais alto) ou N/A (Não se aplica).						
PARTE A – REAÇÃO DO SISTEMA						
		1	2	3	4	5
1.	Frustrante					Satisfatório
2.	Tedioso					Estimulante
3.	Difícil					Fácil
4.	Inadequado					Adequado
PARTE B – TELA						
		1	2	3	4	5
5.	Forma e tamanho das letras					Fácil de ler
6.	Realces na tela (cores e negrito)					Bastante
7.	Organização da informação					Clara
8.	Seqüência de telas					Claro
9.	Itens da tela são fáceis para encontrar					Sempre
PARTE C – TERMINOLOGIA E INFORMAÇÃO DO SISTEMA						
		1	2	3	4	5
10.	Uso dos termos utilizados no sistema					Claro
11.	Mensagens que aparecem na tela					Claro
12.	Localização das mensagens na tela					Clara
13.	Instruções para o usuário					Sempre
14.	Sistema mantém você informado sobre o que está fazendo					Sempre
15.	Mensagens de erro					Útil
PARTE D – APRENDIZADO						
		1	2	3	4	5
16.	Aprender a operar o sistema					Fácil
17.	Explorar por tentativa e erro					Fácil
18.	Lembrar termos e uso de comandos					Fácil
19.	Tarefas podem ser executadas de uma maneira rápida e/ou lógica					Sempre
20.	Conclusão da tarefa					Claro
21.	Mensagens de ajuda					Claro
PARTE E – CAPACIDADE DO SISTEMA						
		1	2	3	4	5
22.	Velocidade do sistema					Rápido
23.	O sistema é confiável					Sempre
24.	Corrigir seus erros					Fácil
25.	Projetado para todos os níveis de usuários (iniciantes e experientes)					Facilmente
PARTE F – IMAGENS						
		1	2	3	4	5
26.	Qualidade das imagens					Boa
27.	Tamanho das imagens					Grande
28.	Processo de busca de imagens similares					Interessante
29.	Recuperação de imagens similares					Relevante
30.	Resultado da recuperação de imagens similares					Confiável

Figura 7. Questionário pós-sessão (frente).

Participante nº: \_\_\_\_\_

PARTE G – SUGESTÕES E COMENTÁRIOS	

Figura 8. Questionário pós-sessão (verso).

A proposta foi obter uma versão resumida do QUIS, constituindo de afirmações associadas a uma escala de classificação de diferencial semântico, que utiliza termos opostos colocados na mesma dimensão, onde os participantes selecionam o valor mais apropriado. Esses termos são adjetivos bi-polares colocados nos pontos extremos da escala como, por exemplo, de fácil a difícil ou



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

Figura 9. Resultado das medidas de satisfação do QUIS.

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. Percebeu-se que os usuários mais jovens demonstraram uma maior habilidade e agilidade computacional, conseguindo desempenhar as tarefas com mais eficiência. No entanto, os usuários que possuem mais conhecimento profissional desempenharam as tarefas com mais eficácia.

Com relação à satisfação da interface, os usuários superaram suas expectativas que foram observadas antes e depois da interação. Apesar do tempo de resposta do sistema ser um fator primordial que influencia a expectativa do usuário, a qualidade da resposta, referente à quão similar as imagens são perante a uma dada imagem de referência, é tão ou mais importante que o tempo. Desse modo, a avaliação atingiu o objetivo proposto, pois pôde prover uma grande percepção de como os usuários pensam sobre a interface.

### CONCLUSÕES

Este trabalho adotou técnicas de questionamento para analisar a interação do usuário em sistemas CBIR e avaliar sua funcionalidade e usabilidade. O objetivo principal deste trabalho foi compor um questionário de avaliação de usabilidade a partir de um conjunto de questionários disponíveis, a fim de mensurar a satisfação dos usuários em relação ao uso de sistemas que apresentam grande complexidade. O desempenho de tais sistemas pôde ser avaliado por critérios reais, fundamentando uma base sólida para julgar a percepção humana de como os usuários pensam sobre a interface e prover meios para melhorar o desenvolvimento de sistemas que apóiam usuários reais em suas tarefas.

A facilidade de utilização de sistemas CBIR por meio de interfaces intuitivas e também da confiabilidade e credibilidade da resposta provida pelo sistema determinou um fator importante para aceitação dos especialistas em utilizar sistemas de apoio a diagnósticos. De maneira geral, pode-se concluir que o uso de técnicas de interação humano-computador 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 tornar o diagnóstico médico e a evolução dos tratamentos cada vez mais precisos e eficientes.

Neste estudo pôde-se concluir que a experiência de trabalhar com usuários reais pode mudar totalmente a percepção de análise das tarefas dentro do domínio da aplicação. Não importa quão bom ou eficiente seja a aplicação, a aceitação ou não da aplicação depende, principalmente, das atitudes dos usuários em relação ao uso do computador no trabalho. Nesse contexto, como sugestão

para trabalhos futuros é necessário dar continuidade a pesquisas contando sempre com o envolvimento do usuário real.

### AGRADECIMENTOS

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

### REFERÊNCIAS

- Müller, H., Michoux, N., Bandon, D., and Geissbuhler, A. A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *International Journal of Medical Informatics* 73, 1 (2004), 1-23.
- ISO-9241-11. *Ergonomic requirements for office work with Visual Display Terminals (VDTs): guidance on usability*. Genebra, 1998.
- Hix, D., Hartson, H.R. *Developing User Interfaces: Ensuring Usability Through Product & Process*. John Wiley & Sons, New York, 1993.
- Preece, J., Rogers, Y., and Sharp, H. *Design de Interação: além da interação homem-computador*. Bookman, Porto Alegre, RS, Brasil, 2005.
- Rocha, H.V.M. and Baranauskas, C.C. *Design e Avaliação de Interfaces Humano-Computador*. NIED/UNICAMP, Campinas, SP, Brasil, 2003.
- Dix, A., Finlay, J., Abowd, G.D., and Beale, R. *Human-Computer Interaction*. 3 ed. Prentice Hall, Essex, 2004.
- Preece, J., Rogers, Y., and Sharp, H. *Human-Computer Interaction*. Addison-Wesley, Wokingham, 1994.
- Stone, D., Jarret, C., Woodroffe, M., and Minocha, S. *User Interface Design and Evaluation*. Morgan Kaufmann, San Francisco, 2005.
- Questionnaire for User Interaction Satisfaction. <http://lap.umd.edu/QUIS/>.
- Software Usability Measurement Inventory. <http://sumi.ucc.ie>.
- Website Analysis and MeasurMent Inventory. <http://www.wammi.com>.
- System Usability Scale. <http://www.mindd.com>.
- Bueno, J.M., Chino, F.J.T., Traina, A.J.M., Traina Jr., C., and Marques, P.M.A. 2002. How to Add Content-based Image Retrieval Capability in a PACS. In *Proc. CBMS'02* (2002).
- Traina, A.J.M., Castañón, C.A.B., and Traina Jr., C. 2003. MultiWaveMed: A System for Medical Image Retrieval through Wavelets Transformations. In *Proc. CBMS'03* (2002).
- Traina Jr., C., Traina, A.J.M., and Figueiredo, J.M., 2004. Including Conditional Operators in Content-Based Image Retrieval in Large Sets of Medical Exams. In *Proc. CBMS'04* (2004).