Avaliação dos mecanismos de Privacidade e Personalização na Web

Luanna L. Lobato
Universidade Federal de São Carlos, Departamento de Computação
São Carlos-SP, Brazil, 13565-905
luanna_lobato@dc.ufscar.br

e

Sérgio D. Zorzo
Universidade Federal de São Carlos, Departamento de Computação
São Carlos-SP, Brazil, 13565-905
zorzo@dc.ufscar.br

Abstract

In the services of the Web should be considered two conflicting goals: the user's right to keep guaranteed their privacy and the benefits of providing personalized services. This paper proposes the construction of a system, capable to analyze privacy and personalization tools, sites and use sceneries, targeting to quantify the offered privacy and personalization. We also present a taxonomy for classifying privacy and personalization in layers. The analysis of the privacy and personalization, by the proposed system, aims offering the user quantitative values that helps in her context of use of services.

Keywords: Privacy, Personalization, Internet, Web, Framework, Pseudonym, Anonymity.

Resumo

Nos serviços da Web devem ser considerados dois pontos que acabam se conflitando: o direito do usuário em ter sua privacidade garantida e os benefícios em ter serviços personalizados. Este artigo propõe a construção de um sistema que seja capaz de analisar as ferramentas de privacidade e personalização existentes, sites e cenários de utilização, de forma a quantificar a privacidade e personalização oferecida. Apresenta-se uma taxonomia para classificação de privacidade e personalização em camadas. A análise de privacidade e personalização, pelo sistema proposto, objetiva fornecer ao usuário valores quantitativos que o auxiliem em seu contexto de utilização de serviços.

Palavras chaves: Privacidade, Personalização, Web, Framework, Pseudônimo, Anonimato.

1. INTRODUÇÃO

A Internet é um artefato mundial de comunicação e divulgação de informações. Em decorrência do avanço da Internet, muitos benefícios têm sido encontrados, no entanto apesar de benefícios serem gerados com seu crescimento, alguns problemas também são detectados decorrentes de sua utilização [25].

Ao navegar pela Internet, às vezes sem saber, os usuários deixam registros de sua navegação que podem ser utilizados para seu benefício ou não. Se tais informações são utilizadas sem o consentimento do usuário, isso poderá ser caracterizado como invasão de privacidade. No entanto, deve ser ressaltado, que desde que o usuário tenha conhecimento de que suas informações estão sendo coletadas, não é caracterizado como invasão de privacidade [29].

A privacidade pode ser caracterizada como o direito que o usuário tem de proteger sua intimidade. Dentre várias definições, pode-se ressaltar que de uma forma geral a privacidade é o direito da pessoa de não ser importunada se não o desejar [5].

No entanto, mesmo que haja o desejo em ter a privacidade garantida, se houver necessidade de personalização, é necessário a divulgação de algumas informações para que seja possível prover serviços personalizados [17].

Personalização é tornar algo adaptável a alguém, adequando-o a suas vontades, necessidades e preferências. É apresentar algo de forma diferente a cada pessoa, pois cada uma tem um gosto definido, um perfil formado [21].

Se nenhuma informação referente ao usuário puder ser disponibilizada para fins de personalização, torna-se inviável o uso desse serviço, pois poderão ser relacionados ao usuário informações que nada tem haver com seu perfil, ocasionando em aborrecimento ao invés de facilidades de serviços.

Pode ocorrer conflito, quando a privacidade e personalização necessitam ser oferecidas ao usuário, pois a privacidade requer sigilo das informações e a personalização necessita de disponibilização. É necessário que haja uma ponderação entre a importância da privacidade e a necessidade em ter serviços personalizados.

Para solucionar essa questão conflitante, diversas ferramentas podem ser empregadas de forma a tornar as informações dos usuários seguras e ainda assim prover serviços personalizados.

Neste artigo ressalta-se a importância de um sistema que seja capaz de prover a avaliação e quantificação, retornando valores, sobre o quanto de privacidade e personalização está sendo oferecido ao usuário quando em interação com a Web, É avaliado o nível de privacidade e personalização para: (i) as ferramentas que provêem privacidade e personalização; (ii) sites de interesse do usuário e (iii) para cenários de utilização, sendo esses definidos como necessidades que os usuários têm, por exemplo; cadastro em listas de discussão, compra online. Apresentando camadas referentes às necessidades do usuário e sugerindo possíveis ferramentas para utilização.

A seção 2 apresenta os mecanismos de personalização existentes e a seção 3 apresenta os mecanismos de privacidade disponíveis na literatura.

Para desenvolvimento do trabalho, utilizou-se uma classificação, já existente, para a taxonomia de privacidade [15], porém essa foi reorganizada para se adequar às necessidades do sistema e, foi definida uma taxonomia para a personalização, ainda não existente na literatura, de acordo com as necessidades que devem ser observadas para a existência da mesma.

A seção 4 do artigo detalha a taxonomia proposta para o tratamento de privacidade e personalização deste trabalho.

Através dessas taxonomias, é possível a criação de ferramentas que contenham as principais características que devem ser apresentadas, para que a privacidade e personalização possam estar sendo oferecidas aos usuários, podendo essas ferramentas serem criadas de forma a auxiliar o usuário leigo em sua navegação, apresentando ao mesmo, soluções mais claras quanto às desejadas por ele, apresentada, na seção 5 deste artigo, e finalmente, a seção 6 apresenta as conclusões da proposta apresentada.

2. MECANISMOS DE PERSONALIZAÇÃO

Os navegadores (*browsers*) são programas utilizados para exibir o conteúdo disponível na Internet sendo através deles que os usuários se comunicam remotamente com os servidores, um dos locais onde as informações requeridas pelos usuários podem estar armazenadas.

Esses enviam aos servidores informações necessárias para estabelecerem uma comunicação, como a data e a hora da requisição, o tipo de navegador, o sistema operacional utilizado e a *Uniform Resource Locator* (URL) onde estão as informações sobre o interesse do usuário [27].

Além dessas informações, alguns navegadores armazenam na máquina dos usuários, informações que servem como um identificador do usuário na rede, coletando informações pessoais com o intuito de prover serviços personalizados.

Existem diferentes mecanismos de personalização para tratar das necessidades impostas pelos usuários. A seguir são apresentados alguns desses mecanismos.

2.1 Cookies

Cookies são pequenas informações armazenadas em um arquivo texto, que são trocadas entre o navegador do usuário e o servidor de páginas quando é feita uma requisição ao servidor. Os *cookies* podem ser armazenados na maquina do usuário para que o mesmo seja reconhecido quando em contato com a Web [18].

A troca das informações entre as páginas do servidor e o computador originário da requisição, são realizadas através do protocolo HTTP (*Hypertext Transfer Protocol*), que não possui estados. Os *cookies* surgiram para solucionar a característica de ausência de estado do protocolo HTTP, possibilitando que sessões pudessem ser criadas para os usuário, com o objetivo de gravar dados, ações e preferências do usuário, podendo identificá-lo quando retornasse ao site.

Para que o navegador receba e armazene o nome e o valor do *cookie* na memória, geralmente em sua criação é inserido um cabeçalho de *cookie* nos pacotes de comunicação do protocolo HTTP. Depois de criados, os *cookies* devem ser transmitidos facilmente entre servidor/cliente, cliente/servidor através desses cabeçalhos.

Os *cookies* não são necessariamente ferramentas para invasão de privacidade, foram criados para melhorar a interação entre as aplicações Web e seus usuários, porém a forma como alguns sites os utilizam é que os tornam perigosos.

2.2 Web bugs

Web bugs são pequenas imagens inseridas em mensagens de correio eletrônico ou em páginas *Web* com o objetivo de monitorar os usuários, coletando seus dados. São imagens transparentes e para visualizá-los é necessário verificar *tags* de imagens no código HTML (*Hypertext Markup Language*) da mensagem ou da página que os contêm. [20].

Se uma mensagem que possui *web bugs* é acessada, esses se instalam na máquina do usuário com o objetivo de coletar informações pessoais como, o nome do usuário, IP (*Internet Protocol*) da máquina e outros dados de identificação e então as encaminham à máquina a qual esses *web bugs* estão relacionados.

É possível saber quais são as páginas visitadas pelo usuário, perfis dos usuários, relação entre propagandas e compras efetuadas, dentre outros dados que podem ser utilizados para tomadas de decisão sobre as preferências do usuário [24].

2.3 Clickstream

Outra maneira encontrada para analisar as ações feitas pelos usuários enquanto navegam na Internet é através do uso de *clickstream*. Esses são capazes de informar o caminho que o usuário percorreu durante a navegação através de um ou mais *sites*.

Os caminhos obtidos pela análise dessa seqüência, que nada mais são que os rastros dos cliques do *mouse*, ou *clickstream*, podem ser usados para proporcionar diversas informações à empresa [1].

Existem alguns termos utilizados para descrever a navegação na Web e que são essenciais para análise de *clickstream* e sua melhor compreensão: requisição de página, visão de página e seção [22].

Quando o usuário faz uma requisição a uma página, está sendo feita a requisição através da URL de seu programa de navegação, a requisição aparece como uma marcação no arquivo de *log* do servidor. Se o usuário desejar visualizar a mesma página depois, será gerada uma nova visão de página e não será feita uma nova requisição.

Para isso, o programa de navegação retornará uma cópia da página previamente armazenada, podendo essa ser apresentada um pouco diferente se alguma mudança tiver ocorrido, como por exemplo, quando se tratando de anúncios na página. Nesse caso uma seção é definida como um período de navegação *Web* assistida pelo usuário ou uma seqüência de visões de página [22].

Os dados de *clickstream* podem ser coletados através de várias formas, arquivos de *log* de servidores, dados de painel e de Provedor de Serviço de Internet (ISP). O mais utilizado é a coleta de dados através de arquivos de *log* de servidores dos *sites* que estão sendo visitados [22], identificando assim informações que são registradas nos *logs*, como, endereço IP, última URL e tipo do navegador.

Porém esses tipos de dados de clickstream são vastos em tamanho e potencialmente muito complexos, sendo necessários métodos estatísticos e de mineração de dados para tratá-los [22].

2.4 Data Mining

Diariamente, grandes volumes de informações são coletadas e armazenadas nas bases de dados de empresas. Essas são importantes fontes de informações, porém, muitas vezes, não são exploradas dadas às dificuldades inerentes a esse grande volume, ultrapassando assim a habilidade técnica e a capacidade humana em sua interpretação [2].

Esses dados tornam-se relevantes à empresa quando são tratados e explorados de forma adequada, podendo ser recuperadas as informações que são relevantes, podendo contar com a utilização de tecnologias para esse fim.

Um exemplo dessas tecnologias é o *data mining* (mineração de dados) que tem por objetivo analisar informações em um banco de dados à procura de padrões que tenham valor significativo para a empresa, usando ferramentas que procuram tendências ou anomalias sem o conhecimento do significado dos dados [13].

Fayyad aponta *data mining* como: “o processo não-trivial de identificar, em dados, padrões válidos, potencialmente úteis e ultimamente compreensíveis” [6].

São chamadas de *data mining* algumas das técnicas que permitem extrair conhecimento de uma massa de dados que, de outra maneira, permaneceria escondido nas grandes bases de dados. Esse processo permite que se investiguem esses dados à procura de padrões que tenham valor significativo para a empresa [23].

A primeira etapa dentro do processo é a compreensão e definição de um domínio. Logo após, é necessário selecionar, dentro desse domínio, os dados nos qual o descobrimento será realizado, esses dados devem ser limpos, removendo informações desnecessárias e transformadas. É preciso, ainda, definir a técnica e o algoritmo de *data mining* a ser utilizado.

Assim, os dados selecionados do domínio devem ser então transformados de acordo com as características da técnica e do algoritmo. Nesse ponto, os dados já podem ser submetidos ao processo de mineração propriamente dito.

A partir daí, com o resultado gerado, pode-se analisar o conhecimento descoberto. E caso os resultados não sejam satisfatórios, várias etapas do processo podem ser realizadas novamente.

3. MECANISMOS PARA PROTEÇÃO DE PRIVACIDADE

Os mecanismos para garantia de privacidade têm a finalidade de manter as informações pessoais dos usuários seguras não sendo possível identificar o usuário e nem relacionar informações pessoais a ele. Entretanto, o uso incorreto desses mecanismos podem não permitir que a coleta de informação do usuário seja realizada, o que é fundamental para a personalização.

3.1 Criptografia

A criptografia tem o objetivo de dificultar ou impedir o entendimento de informações não autorizadas. A idéia básica é aplicar algum método matemático sobre o que se deseja criptografar, cifrando ou transformando uma mensagem de tal forma que ela não possa ser entendida por todos, exceto por aqueles que possuam a chave de deciframento que permite decifrar a informação ilegível em seu estado inicial.

Com relação à proteção da privacidade de usuários na Internet, as ferramentas utilizadas para encriptar informações são as mais utilizadas e as que obtiveram mais sucesso.

Porém esse método não é totalmente eficiente quando se tratando de mineração de dados, pois mesmo não podendo decifrar o conteúdo de uma mensagem, ainda é possível saber o endereço IP do cliente e do servidor, saber o horário em que a comunicação foi realizada, a freqüência das transmissões e o comprimento dos dados permutados. Por isto, a criptografia deve ser utilizada em conjunto com outras tecnologias de proteção à privacidade [28].

Existem diversas maneiras de criptografar as informações, porém nosso interesse é apenas ressaltar sua existência e utilidade, já que o principal foco neste trabalho é o uso conjunto com outros mecanismos de segurança..

3.2 Anonimato

Através da utilização do anonimato, é possível que o usuário interaja com *sites* na *Web* sem que seja identificado, ocultando suas características e garantindo a proteção de sua identidade.

Quando em contato com a *Web,* a característica principal que se deve proteger é o endereço IP da máquina do usuário, através do qual pode-se encontrar diversas informações referentes ao usuário, e assim como feito pelos *cookies,* podem ser utilizadas para correlacionar atividades através de diferentes *sites* [19].

Alguns usuários com medo de sua privacidade ser invadida, optam por utilizarem serviços que possibilitam uma navegação e o envie de mensagens de forma anônima na Internet, como o *anonymizer.com*[1] ou *the-cloak.com*[2] .

Nesses serviços, quando o usuário faz uma requisição, as informações referentes a ele são apagadas das mensagens, sendo fornecido ao usuário um endereço anônimo, para onde as mensagens são encaminhadas de volta.

O maior problema em um sistema de anonimato é quando existem terceiros capazes de observar todos ou vários nós de uma comunicação, necessitando assim, de sistemas mais complexos para manter a segurança sobre a privacidade do usuário.

Existe o anonimato do transmissor, onde o emissor da mensagem não pode ser identificado; o anonimato do receptor que é a não identificação do recebimento da mensagem; e a não ligação entre o transmissor e o receptor, que é a não identificação da mensagem por um terceiro.

Para solucionar o problema, novas propostas foram criadas utilizando-se *proxy* de anonimato [26]. Existem dois tipos de abordagem para a utilização de um *proxy* enquanto o usuário navega anonimamente na *Web*, sendo eles: *proxies* de anonimato de único nó e *proxies* de anonimato de vários nós.

Nos *proxies* de único nó, o próprio *proxy* é o responsável pelo envio das mensagens do remetente ao destinatário. O *Anonymizer*, exemplo de *proxy Web* de único nó*,* filtra todas as identificações do navegador do usuário, permitindo que eles naveguem anonimamente pela *Web*, sem revelar suas identidades ao servidor final [11].

Já os *proxies* de vários nós utilizam uma rede chamada *mix net*, onde o objetivo é fazer com que a origem da requisição seja perdida, de forma a fornecer o anonimato. Dentre as várias ferramentas e tecnologias de anonimato, ressaltam-se as seguintes: Onion Routing[3] e Crowd [10].

3.3 Pseudônimos

Os pseudônimos criam uma camuflagem, um apelido para o usuário de sua verdadeira identidade. Sendo essa, um disfarce utilizado de forma a impedir que o usuário seja identificado, permitindo dessa forma o anonimato e a personalização.

[1] http://www.anonymizer.com
[2] http://www.the-cloak.com
[3] http://www.onion-router.net

Quando se trata de Internet, é necessário camuflar informações que poderiam identificar o usuário, como por exemplo, o IP da máquina que o usuário utiliza, e ainda assim permitir que sites possam disponibilizar um serviço personalizado.

Ao contrário do anonimato onde o objetivo é não se identificar, com o pseudônimo é feita uma identificação, porém de forma a não revelar a identidade real de quem o utiliza, sendo caracterizada por um anonimato de identidade real, mas com revelação do pseudônimo utilizado. Podendo ser utilizado mais de uma vez um mesmo pseudônimo pelo usuário, entretanto se a verdadeira identidade associada ao pseudônimo for descoberta, todas as ações realizadas sob esse poderão ser ligadas ao usuário que o utiliza.

Para o funcionamento do pseudônimo é necessária a utilização de um terceiro intermediando a troca de informações. Há dois problemas para essa abordagem: (i) decisão de quem será essa entidade intermediária, tendo essa que ser confiável e (ii) o centralizador, que é essencial ao funcionamento da rede, pode se tornar um ponto vulnerável para ataque [14].

A vantagem dessa tecnologia é que permite aos sites oferecerem serviços personalizados de acordo com perfis dos usuários, sem identificá-los e sem que suas informações sejam combinadas com outros sites.

Existem diversos mecanismos baseados no uso de pseudônimos, um exemplo é o JPWA, hoje chamado de LPWA (*Lucent Personalized Web Assistant*). O LPWA foi projetado pra utilizar um único pseudônimo toda vez que o usuário interagir com um mesmo *site*, porém usa pseudônimos diferentes quando o usuário interage com *sites* diferentes [9].

Os apelidos dos usuários são transferidos do proxy LPWA ao sites e se um terceiro interceptar essa comunicação e conhecer o apelido, esse poderá se passar por dono do apelido e acessar o site. Para isso, o uso de SSL[4] (*Security Sockets Layer*) pode ser necessário como mecanismo de proteção para as conexões, tornando o LPWA viável [4].

3.4 Máscaras

O propósito da máscara é produzir uma impressão definida nos outros e, muitas vezes, embora não obrigatoriamente, dissimula a natureza real do indivíduo. As máscaras utilizadas na Web tem o objetivo de camuflar os usuários de forma a esconder algumas de suas características, sendo essa máscaras identificações temporárias que podem ser assumidas pelos usuários em interações com *sites*.

Podem ser utilizados dois tipos de máscaras: as físicas e as psicológicas. As físicas têm por objetivo o anonimato, de forma a não permitir o reconhecimento da identificação das pessoas. As máscaras psicológicas ou *personae* buscam esconder a personalidade real dos usuários, em respostas às convenções da sociedade e às suas próprias necessidades [12].

Entretanto esse esquema desperta várias preocupações: (i) é preciso garantir que o usuário tenha controle sobre suas informações pessoais; (ii) a compatibilidade com protocolos padrões da Web, e (iii) como utilizar os serviços de máscaras sem causar um *delay* percebível ao usuário em sua navegação.

Devido a esses fatores, surgiu a arquitetura MASKS (*Managing Anonymity while Sharing Knowledge to Servers*) proposta por Ishitani, utiliza como base o conceito de máscara e atende às necessidades impostas na utilização de máscaras pelos usuários, oferecendo privacidade e personalização ao mesmo [14].

O MASKS é uma arquitetura onde a idéia básica é colocar uma barreira entre os dados privados e a *Web*, controlando as informações que podem atravessar essa barreira. Essa arquitetura minimiza a divulgação de dados pessoais sem impedir uma análise contínua desses dados [14].

A arquitetura do MASKS possui dois componentes principais: o agente de privacidade e segurança PSA (*Privacy and Security Agent*) e o servidor de máscaras MASKS *Server*.

O PSA é um programa intermediário entre os usuários e o MASKS *Server* que é executado junto ao navegador do usuário, oferecendo ao usuário as duas primeiras camadas de proteção à privacidade encontradas no MASKS, sendo responsável por: (i) cifrar as requisições dos usuários; (ii) mantê-los informados sobre os riscos de ter sua privacidade invadida e sobre as máscaras que lhe estão sendo atribuídas; (iii) permitir que os usuários desliguem o processo de atribuir máscaras a eles, se optarem pela interação direta com o *site* sem anonimato e (iv) bloquear e filtrar métodos conhecidos de invasão de privacidade, como os cookies de terceirose os *web bugs* dentre outros.

O MASKS *Server* trabalha como um *proxy*, sendo o ponto intermediário entre os usuários e os sites, responsável pela criação dos grupos e pelo gerenciamento e atribuição das máscaras aos usuários, garantindo o anonimato. A atribuição das máscaras é baseada no conceito de grupo, sendo que cada grupo representa um tópico de interesse.

[4] SSL é uma camada que tem o objetivo de promover um tráfego seguro na Internet. http://www.openssl.org/

3.5 Rede Mix

A rede *mix* é baseada no mecanismo de anonimato através de pseudônimos e tem a função de garantir a privacidade, utilizando pseudônimos, e oferecer serviços personalizados em sistemas adaptáveis ao usuário (*user-adaptative system*).

O sistema adaptável ao usuário é a responsável pela obtenção das informações do usuário e pelo gerenciamento, podendo essa coleta ser feita através dos sites, ou com a utilização de uma aplicação adaptável ao usuário (*user-adaptative system*).

Os sistemas adaptáveis ao usuário coletam mais informações pessoais sobre os usuários que os tradicionais sites e, com isso, o rastreamento do usuário pode ser feito de forma mais eficaz.

As aplicações adaptáveis ao usuário tratam das necessidades individuais de cada usuário, adaptando-se da melhor forma possível, sendo a associação com a rede *mix* importante para seu anonimato. Para isso, essas aplicações resgatam as informações sobre as características individuais do usuário e de sua interação, transmitindo-as ao servidor que modela o usuário (*user modeling server*) [7].

Nesse, as informações são armazenadas em modelos individuais de usuário, chamados de *User models*, que são ligados ao usuário, persistindo durante diferentes sessões [16]. Esse servidor cria um modelo para cada usuário que será o repositório de informações, armazenando as informações que são relevantes ao sistema de acordo com o perfil do usuário, podendo adaptá-lo a grupos que tratem de interesses similares.

Algumas exigências devem ser seguidas para tratar da segurança em sistemas adaptáveis ao usuário, como a linguagem e o protocolo de comunicação que permitem a troca de informações sobre o usuário com o servidor que modela usuário.

3.6 Políticas de Privacidade

Geralmente a privacidade é discutida como um problema social, devendo ser aplicados métodos que tratem desses problemas. As políticas de privacidade são utilizadas pare esse fim, devendo ser definidas de forma clara [Ishitani 2003].

Com base em estudos, desenvolveu-se o P3P (*Plataform for Privacy Preference Project)* da W3C *(World Wide Web Consortium*) [3].

O P3P é uma tentativa de fazer uma padronização da linguagem de especificação da política de privacidade, fácil de ser localizada e capaz de ser lida pelo computador, permitindo que sites negociem com o usuário quais informações serão coletadas, onde serão utilizadas e de que forma serão coletadas [3].

Isso é possível, pois o P3P possui um protocolo projetado em um formato XML (*Extensible Markup Language*), o qual permite que administradores de *sites* publiquem a política de privacidade, em um formato padrão que pode ser recuperado automaticamente.

Dessa forma, quando o usuário interage com o *site*, o navegador recupera a política de privacidade desse *site*, que está no formato do P3P e então avalia se essa política atende aos requisitos impostos pelo usuário, comparando com as definições de segurança configuradas por ele. Se as políticas atenderem a essas especificações, o navegador continua a requisição, senão, dúvidas podem ser resolvidas através de interação com o usuário [4].

Embora o P3P forneça um mecanismo técnico de avaliação das políticas de privacidade dos *sites*, ele não é capaz de garantir que esses estão agindo de acordo com o que é especificado em suas políticas.

Para isso são utilizadas as chamadas entidades certificadoras, que são responsáveis por analisar se as práticas de segurança conduzidas pelos *sites* estão de acordo com o que foi proposto, dando ao usuário uma maior garantia de que o mesmo cumpre com as políticas de privacidade definidas [8].

4. TAXONOMIA PARA O TRATAMENTO DA PRIVACIDADE E PERSONALIZAÇÃO

Devido ao uso indevido das informações pessoais dos usuários, aos problemas de privacidade decorrentes do uso de técnicas de personalização e devido à diversidade de ferramentas encontradas que tratam da privacidade do usuário e da personalização dos serviços oferecidos a ele, fica evidente a necessidade de um sistema que possa avaliar tais ferramentas, *sites* e cenários de utilização.

O artigo consiste na apresentação de um sistema que seja capaz de avaliar de forma quantitativa, retornando valores reais, o quanto de privacidade e personalização as ferramentas e os *sites* oferecem e apresentar de acordo com o que o usuário deseja possíveis ferramentas para sua utilização.

Para a criação do sistema utilizou-se uma taxonomia de privacidade e criou-se uma de personalização. Com entradas submetidas ao sistema será possível avaliar quais níveis, expressos em camadas, estarão presentes nas ferramentas e, a partir de valores determinados às camadas, definir o nível de privacidade e personalização que elas oferecem.

Essas camadas irão referenciar as características que devem ser apresentadas para um mecanismo prover privacidade sobre os dados do usuário e personalização nos serviços oferecidos, objetivando a construção de *sites* e

ferramentas que apresentem tais camadas, para que estejam cada vez mais próximos do que espera ser encontrado pelo usuário, quando se tratando de privacidade e personalização.

Como nos traz Ishitani, os usuários podem ter sua privacidade protegida através de diferentes camadas de proteção de privacidade [14].

Com a utilização do termo proteção de privacidade é possível fazer a divisão em diferentes camadas para cada abordagem, justificando-se a utilização dessa divisão, onde cada ferramenta pode ser classificada em uma ou em mais.

Para tanto, foi proposta por Ishitani, uma taxonomia de proteção à privacidade dividida em 5 camadas que possuem responsabilidade entre o usuário e a sociedade [14].

Apesar da taxonomia proposta por Ishitani estar próxima às necessidades desse estudo, decidiu-se apresentar uma nova organização, visualizada na
Figura 1, de forma que algumas camadas deveriam existir para fazer valer a existência de outras, que vão de funcionalidades próximas ao hardware até benefícios oferecidos ao usuário, atendendo as necessidades do usuário de forma a não comprometer o funcionamento do sistema.

Figura 1. Nova arquitetura para as camadas de privacidade

Por exemplo, não é julgado necessário a existência da camada, certificação de privacidade (camada 3), se não houver a camada política de privacidade (camada 2), a quem a certificação é aplicada.

Visa-se essa nova organização de camadas devido à necessidade de sua classificação, referente à sua importância dentro do escopo, pois a partir dessa classificação é possível definir quais pesos serão atribuídos a cada camada de acordo com sua relevância e em contrapartida, avaliar o nível de privacidade e personalização oferecida.

Com base no estudo feito sobre taxonomia de privacidade, foi definida uma taxonomia de personalização, mostrada na
Figura 2 que enfatiza as características que podem ser apresentadas para prover serviços personalizados.

Quanto maior quantidade de camadas o método abranger e quanto maior relevância a camada tiver, maior será a qualidade dos serviços personalizados disponibilizados por ele.

Figura 2. Taxonomia de personalização

A Figura 3 mostra o significado de cada camada dentro do escopo de utilização.

Assim como na taxonomia de privacidade, a arquitetura da taxonomia de personalização não foi criada devendo obrigatoriamente haver uma interligação entre as camadas para que o serviço de personalização possa ser oferecido, já que se uma ferramenta se enquadra a alguma dessas camadas, a personalização está sendo aplicada. Porém, se houver a necessidade da personalização ser feita de forma exclusiva e precisa a um usuário, a utilização de mais de uma camada poderá ser significante.

Objetiva-se através dessas taxonomias permitir a construção de ferramentas que possam quantificar o valor de privacidade e personalização oferecida.

Leis de proteção de privacidade: obrigam as empresas a protegerem a privacidade dos usuários.	**Filtro de códigos maliciosos:** possibilita ao usuário configurar o que deve ser permitido em sua máquina, filtrando códigos maliciosos, como *web bugs*
Políticas de privacidade: anunciada pelas empresas visando conquistar a confiança dos usuários.	**Formulários:** é submetido ao usuário formulários, que deve ser preenchido e retornado ao *site*, o qual, coleta as informações necessárias para personalização.
Certificação de privacidade: verificação da política de privacidade anunciada pela empresa.	**Armazenamento na máquina do usuário:** dados podem ser armazenados na máquina do usuário permitindo a identificação desse, com o uso de *cookies*
Notificação: informa os usuários sobre os riscos de ter sua privacidade invadida.	**Utilização de *proxy*:** as informações dos usuários se encontram em *proxies*, que intercedem a comunicação entre usuário e *site*.
Controle: aumenta o controle dos usuários sobre sua privacidade.	**Armazenamento em Servidor:** as informações estáticas dos usuários são armazenadas no servidor, em algum Banco de dados.
Ferramentas para proteção de privacidade: melhora o nível de privacidade dos usuários.	**Históricos:** é possível saber o caminho feito pelo usuário durante sua navegação na *Web*, a partir de técnicas como o *clickstream* .

Figura 3. Significado de cada uma das camadas da taxonomia de Privacidade e Personalização

5. FERRAMENTA DE AVALIAÇÃO

Para ser possível essa avaliação, deve-se utilizar um mecanismo que tenha sido construído para esse propósito, onde ele analise se tais camadas estão presentes nas ferramentas, *sites* ou cenários de utilização, e através de pesos definidos a cada camada, seja possível avaliar de forma quantitativa, o quanto de privacidade e personalização está sendo oferecido.

Pretende-se com a elaboração do sistema fazer a interação com o usuário sobre as características de privacidade e personalização desejadas, retornando a ele um valor quantificado, que denotará o grau de privacidade e personalização apresentado pela ferramenta ou *site*.

Uma representação do modo de funcionamento do sistema é apresentada na Figura 4, onde deve ser determinado pelo usuário qual entrada será submetida ao sistema, ou seja, o que deverá ser avaliado. A partir dessa entrada, o sistema será encarregado de fazer a avaliação e retornar a resposta ao usuário de forma clara e concisa.

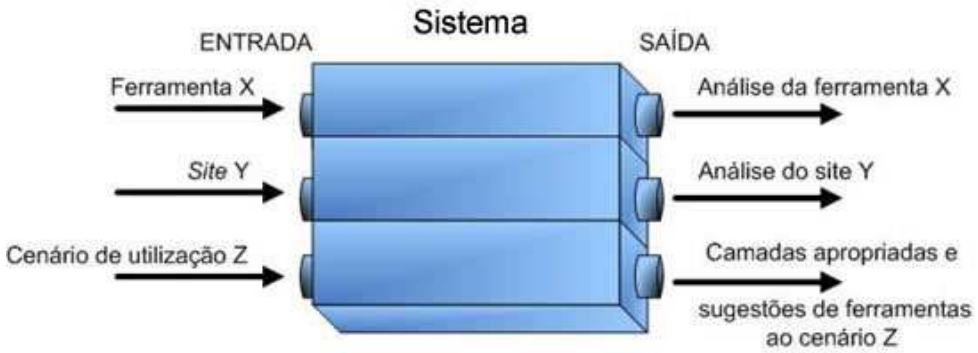

Figura 4. Exemplo do funcionamento do sistema

Por exemplo, no caso da submissão de um cenário de utilização Z, o sistema retornará quais as camadas e ferramentas são aplicáveis em seu uso.

Esse mecanismo deve ser capaz de identificar a relevância de cada camada para a avaliação e, não apenas a quantidade de camadas que o objeto avaliado abranger.

O sistema, inicialmente terá cadastrado algumas ferramentas e cenários de utilização, podendo ser adaptado a diversas situações e ferramentas, devido a sua característica inerente de ser extensível. Já a avaliação dos *sites*, será feita apenas informando a URL do *site* desejado.

Em relação ao processo de identificação de privacidade e personalização referente ao que as ferramentas apresentam, ao que os *sites* possibilitam e ao que o usuário deseja, são ilustradas as seguintes metodologias:

O sistema precisa inicialmente de uma entrada com a opção desejada para análise do tipo:

- Uma ferramenta que trata de privacidade e personalização, dada uma lista previamente cadastrada no sistema, tal lista poderá ser atualizada pelo usuário;
- Uma URL válida de uma página;
- Um cenário de utilização que será montado a partir de opções de interação do usuário, apresentadas pelo sistema.

Avaliação das Ferramentas

O sistema apresenta uma base de dados de ferramentas, que será apresentada ao usuário. De acordo com sua escolha, a ferramenta será analisada e será retornado o nível de privacidade e personalização provido por ela.

A Figura 5 mostra um esboço de uma aplicação implementada no sistema para análise de ferramentas que tratam do controle de privacidade e serviços de personalização.

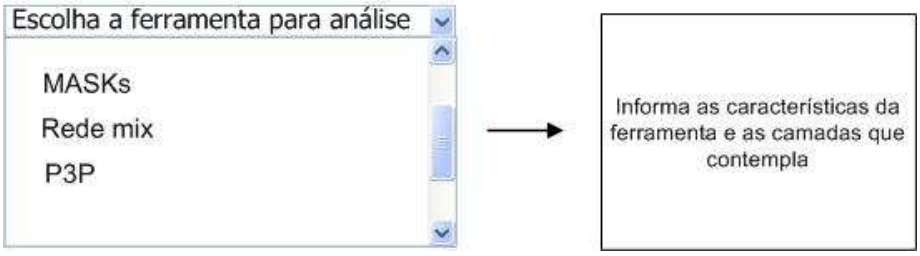

Figura 5. Escolha da ferramenta para análise

O usuário então pode escolher uma ferramenta e ver detalhes de sua utilização, bem como, o grau de privacidade e personalização de forma quantificada e, se desejado, quais camadas ela aborda. Também poderá consultar informações técnicas, como o fornecedor, tipo de distribuição (código livre, proprietária) e onde fazer o download.

Avaliação dos Sites

A partir de uma URL informada pelo usuário, o sistema acessa as páginas referentes ao *site*, de maneira automatizada, processando o código HTML de cada uma a procura de palavras ou *tags*[5] que apresentem alguma relevância sobre o que está sendo procurado, e que possam identificar a existência ou não de algumas das camadas proposta na taxonomia de privacidade e personalização.

Para cada camada existem palavras que são relevantes à sua existência no *site* e devem existir regras de análise gramatical (*parsing*) para a análise do código.

Utiliza-se expressões regulares próprias para o encontro (*matching*) de palavras e expressões como: "leis" (camada 1) que regem o país; "políticas de privacidade" (camada 2) que o *site* apresenta; "entidade certificadora" (camada 3) mostrando qual é a entidade que garante segurança no *site*; "<form>, CPF, cartão de crédito" (camada 4), notificando o usuário sobre possíveis coletas; dentre outros.

A Figura 6 mostra um exemplo de procura pela *tag form* (formulário) que pode estar sendo usada para coletar dados pessoais dos usuários.

Na figura, o resultado da análise é informado ao usuário, mostrando o perigo (quando existe a coleta de seus dados) ou segurança (no caso do uso de entidades certificadoras) apresentada pelo *site*.

[5] *Tags* são marcações de códigos que fazem a formatação e programação da página HTML

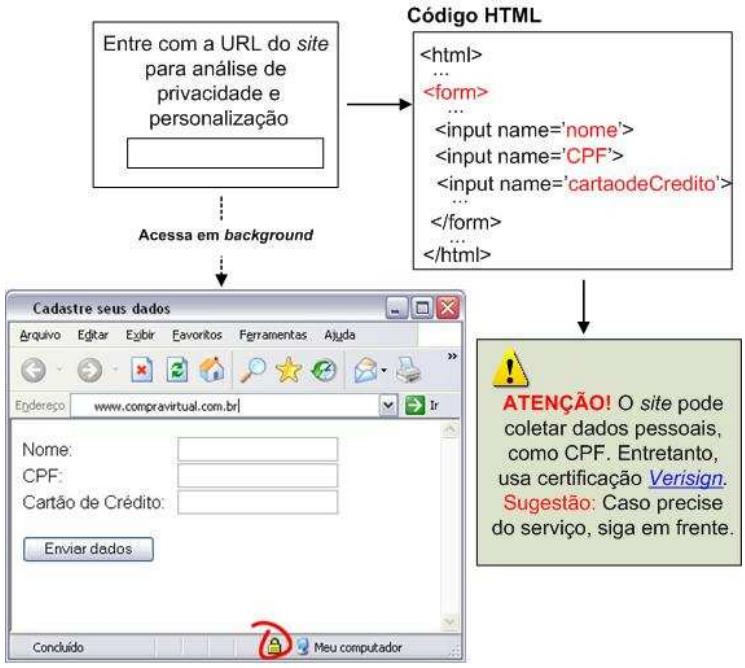

Figura 6. Análise de *site*

Avaliação dos Cenários de Uso

O sistema previamente terá cadastrado as opções mais comuns de interação do usuário na *Web*, como por exemplo: efetuação de compras, cadastro de e-mail em listas de discussão, cadastro de dados em *sites*, transações bancárias, login em serviços em geral, dentre outros, mostrando as camadas que os serviços atingem e as possíveis ferramentas a serem utilizadas, como mostrado na
Figura 7.

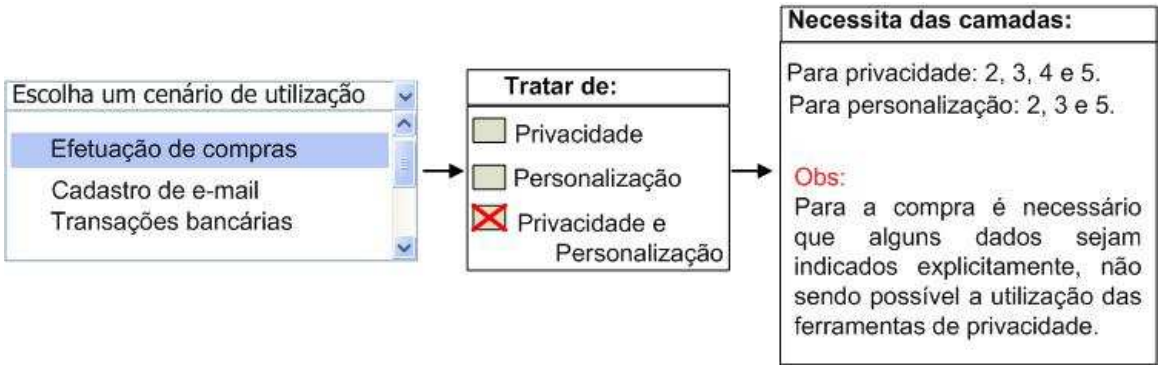

Figura 7.Análise de cenário de utilização

A partir da avaliação do que o usuário deseja, é retornado a ele, de forma quantificada, o nível de privacidade e personalização que poderá ser atingido.

6. CONCLUSÕES

Através da quantificação da privacidade e da personalização oferecida, o usuário terá mais segurança sobre o que é feito com seus dados, fazendo com que a interação do usuário com os *sites* aconteça de uma forma mais confiante e principalmente mais transparente.

Inicia-se um processo de formalização de camadas de personalização, o que ainda era vago na literatura, enfatizando as características que devem ser observadas para a análise da privacidade e personalização que esta sendo oferecida ao usuário.

Esse trabalho também serve como base, auxiliando no desenvolvimento de sistemas eficientes e usáveis nesse domínio.

Referências

[1] Bucklin, R. E. *et al.* (2002) *Choice and the Internet: from Clickstream to Research Stream*, U.C. Berkeley 5th Invitational Choice Symposium, Mareting Letters, 13(3), 245 -258, Last Revised February.

[2] Carvalho, D. R. (1999) *Data Mining através de indução de regras e algoritmos genéticos*. Dissertação (Mestrado em Informática Aplicada) - Pontifícia Universidade Católica do Paraná, PR

[3] Coyle, K. (1999) *A social analysis of the platform for privacy preferences (P3P*). Platform for Privacy Preferences (P3P) Project. Disponível em: <http://www.w3.org /P3P/>. Acesso em: 07 out. 2004.

[4] Cranor, L. F., Byers, S. e Kormann, D. (2003) *An Analysis of P3P Deployment on Commercial, Government, and Children's Web Sites*. AT&T Labs-Research and Florham Park, NJ, May 2003. Disponível em: <http://www.research.att.com/projects/p3p/p3p-census-may03.pdf>. Acesso em: 29 jan. 2006.

[5] Cretella Jr, J. (1997) *Comentários à Constituição Brasileira de 1998*. Rio de Janeiro/RJ: Forense Universitária.

[6] Fayyad, U., Piatetski-shapiro, G. e Smyth, P. (1996) *The KDD process for extracting useful knowledge from volumes of data*. Commun. ACM 39, 11 (Nov. 1996), 27-34. Disponível em: <http://doi.acm.org/10.1145/240455.240464>. Acesso em: 02 nov. 2005.

[7] Fink, J. e Kobsa, A. (2000) *A review and analysis of commercial using modeling servers for personalization on the world wide web*. User modeling and User-Adapted Interaction. Disponível em:<http://www.ics.uci.edu/~kobsa/papers/ 2000/UMUAI-kobsa.pdf>. Acesso em: 22 jan. 2006.

[8] Friedman, B., Khan Jr., P. H. e Howe, D. C. (2000) *Trust online*. Commun. ACM 43, 12 (Dec. 2000), 34-40. Disponível em:<http://doi.acm.org/10.1145/355112.355120> Acesso em: 08 out. 2005.

[9] Gabeer, E. *et al.* (1998) *LPWA Lucent personalized web assistant*. Bell Laboratories, Information Sciences Research Center, Lucent Technologies. Murray Hill, NJ. Disponível em: <http://www.bell-labs.com/projects/lpwa> Acesso: 25 jan. 2006.

[10] Garfinkel, S. (1978) *Web Security, Privacy & Commerce*. O'Reilly, 2nd edition, jan. 2002

[11] Goldberg, I., Wagner, D. e Brewer, E. (1997) *Privacy-enhancing technologies for the Internet*. Proc. of IEEE Spring CompCon. Disponível em: <http://citeseer.nj.nec.co m/54687.html> Acesso: 03 dez. 2005.

[12] Hall, C. S. e Lindzey, G. (1978) *Theories of Personality*. John Wiley & Sons, 3rd edition.

[13] Han, J. e Kamber, M. (2001) *Data Mining: concepts and techniques*.1.ed. New York: Morgan Kaufmann, 2001.

[14] Ishitani,L. *Uma Arquitetura para Controle de Privacidade na Web*. 92 f. Tese (Doutorado em Ciência da Computação) - Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, M.G, 2003.

[15] Ishitani, L., Almeida, V. e Meira Jr., W. *MASKS: Bringing Anonymity and Personalization Together*. IEEE Security & Privacy, vol. 1, no. 3, May/June 2003. Disponível em: <http://ieeexplore.ieee.org/xpl/abs_free.jsp?arNumber=1203218> Acesso em: 24 out. 2004.

[16] Kobsa, A (2001). *Tailoring privacy to user's needs*. Lecture Notes in Computer Science Volume 2109, Jan 2001. Page 303.

[17] Koch, M. (2003) *User Representation in eCommerce and Collaboration Applications*. Department of Informatics, Technische Universitaet Muenchen, Germany. Disponível em: <http://www11.informatik.tu-muenchen.de/publications/pdf/Koch20 03a.pdf> Acesso: 29 jan. 2006.

[18] Kristol, D. M. (2001) *HTTP cookies: Standards, privacy, and politics*. ACM Trans. Inter. Tech. 1, 2 (Nov. 2001), 151-198. Disponível em: <http://doi.acm.org/10.1145/ 502152.502153>. Acesso em: 11 agosto 2005.

[19] Lucena Neto, C. (2002) *Função social da privacidade*. Módulo Security. Disponível em: <http://www.modulo.com.br/pdf/funcaosocialpriv.pdf>. Acesso em: 19 out. 2005.

[20] Martin, D. (2003) *Detecting Web Bugs with Bugnosis: Privacy Advocacy through Education*. Boston University Computer Science Department. Disponível em: < http://www.bugnosis.org/faq.html> Acesso: 29 nov. 2005.

[21] Mayer, A. (1997) *How to Make Personalized Web Browsing Simple, Secure, and Anonymous*. Bell Laboratories, Information Sciences Research Center, Lucent Technologies, Murray Hill, NJ. Disponível em: <http://www.bell-labs.com/project/lp wa/papers.html>. Acesso em: 26 jan. 2006.

[22] Montgomery, A. L. (2003*) Using Clickstream Data to Predict WWW Usage*. WebShop, University of Maryland, June 2003.

[23] Quonian, L. *et al.* (2001) *Inteligência obtida pela aplicação de Data Mining em base de teses francesas sobre o Brasil*. Ciência da Informação, Brasília, Maio/Ago. 2001. Disponível em: <http://www.ibict.br/cionline/include/getdoc.php?id=505&article= 216&mode=pdf>.Acesso em: 16 jan. 2006.

[24] Rocha, B. G. *et al.* (2002) *Disclosing users' data in an environment that preserves privacy*. In Proceedings of the 2002 ACM Workshop on Privacy in the Electronic Society (Washington, DC, November 21 - 21, 2002). WPES '02. ACM Press, New York, NY, 71-80. Disponível em: <http://doi.acm.org/10.1145/644527.644535>. Acesso em: 12 out. 2005.

[25] Rosa, E. H. P. e Fanhani, E. E. (2004) *Algumas reflexões sobre o Comércio Eletrônico: vantagens da comercialização pela Internet*. Revista de Administração Nobel, Nº 03, Jan./jun.2004. p. 71-76. Disponível em: <http://www.nobel.br/template s/administracao/revista/2005-03-03/08_artigo_06.pdf>.Acesso em: 12 agosto 2005.

[26] Shubina, A. M. e Smith S. W. (2003) *Using caching for browsing anonymity*. Dartmouth Computer Science Technical Report TR2003-470, July 2003. Disponível em: <http://www.cs.dartmouth.edu/~sws/pubs/ss03.pdf>. Acesso em: 13 jan. 2006.

[27] Silva, J. A. *apud* Vianna, C. S. M. (2002) *Software e privacidade: uma defesa do código-fonte aberto na preservação do direito constitucional à vida privada*. Jus Navigandi, 2002.Disponível em:<http://jus2.uol.com.br/doutrina/texto.asp?id=2931> Acesso em: 19 set. 2005.

[28] Wang, H., Lee, M.K.O. e Wang, C. (1998) *Consumer privacy concerns about Internet marketing*. Commun. ACM 41, 3 (Mar. 1998), 63-70. Disponível em: <http://doi.ac m.org/10.1145/272287.272299>. Acesso em: 03 dez. 2005.

[29] Warren, S. D. e Brandeis L. D. (1980) *The right to privacy*. Harvard Law Review, 4(5), 1890. Disponível em: <http://chnm.gmu.edu/aq/photos/texts/4hlr193.htm>. Acesso em: 05 set. 2005